7

Estimating and Testing Complete Models

7.1 Notation

This chapter discusses the estimation and testing of complete models. Some additional notation is needed from that used in Chapter 4 to handle the complete model case. The model will continue to be written as in 4.1. The additional notation is as follows. J_t denotes the $n \times n$ Jacobian whose ij element is $\partial f_i / \partial y_{jt}$, $(i, j = 1, \ldots, n)$. u denotes the $m \cdot T$–dimensional vector $(u_{11}, \ldots, u_{1T}, \ldots, u_{m1}, \ldots, u_{mT})'$. G' denotes the $k \times m \cdot T$ matrix:

$$G' = \begin{bmatrix} G_1' & 0 & . & . & . & 0 \\ 0 & G_2' & & & & \\ . & & . & & & \\ . & & & . & & \\ . & & & & . & \\ 0 & & & & & G_m' \end{bmatrix}$$

where G_i' is defined in Section 4.1. Finally, u_t denotes the m–dimensional vector $(u_{1t}, \ldots, u_{mt})$, and Σ denotes the $m \times m$ covariance matrix of u_t.

7.2 3SLS and FIML[1]

Two full information estimation techniques are three stage least squares (3SLS) and full information maximum likelihood (FIML). 3SLS estimates of α are obtained by minimizing

$$S = u'[\hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z']u = u'Du \tag{7.1}$$

with respect to α, where $\hat{\Sigma}$ is a consistent estimate of Σ and Z is a $T \times K$ matrix of predetermined variables. An estimate of the covariance matrix of the 3SLS coefficient estimates (denoted $\hat{V}_3$) is

$$\hat{V}_3 = (\hat{G}'D\hat{G})^{-1} \tag{7.2}$$

where $\hat{G}$ is G evaluated at the 3SLS estimate of α. Σ is usually estimated from the 2SLS estimated residuals, which is done for the computational work in the next chapter.

Under the assumption that u_t is independently and identically distributed as multivariate normal $N(0, \Sigma)$, FIML estimates of α are obtained by maximizing

$$L = -\frac{T}{2}\log|\Sigma| + \sum_{t=1}^{T}\log|J_t| \tag{7.3}$$

with respect to α. An estimate of the covariance matrix of the FIML coefficient estimates (denoted $\hat{V}_4$) is

$$\hat{V}_4 = -\left(\frac{\partial^2 L}{\partial\alpha\,\partial\alpha'}\right)^{-1} \tag{7.4}$$

where the derivatives are evaluated at the optimum.

7.3 Stochastic Simulation[2]

Some of the methods in this chapter and in Chapter 10 require stochastic simulation, and so it will be useful to give a brief review of it. Stochastic simulation requires that an assumption be made about the distribution of u_t. It is usually assumed that u_t is independently and identically distributed as

[1] See Fair (1984), Sections 6.3.3, 6.3.4, 6.5.2, and 6.5.3, for a more detailed discussion of the 3SLS and FIML estimators.

[2] See Fair (1984), Section 7.3, for a more detailed discussion of stochastic simulation.

multivariate normal $N(0, \Sigma)$, although other assumptions can clearly be used. Alternative assumptions simply change the way the error terms are drawn. For the results in this book, the normality assumption has always been used. Stochastic simulation also requires that consistent estimates of α_i be available for all i. Given these estimates, denoted $\hat{\alpha}_i$, consistent estimates of u_{it}, denoted $\hat{u}_{it}$, can be computed as $f_i(y_t, x_t, \hat{\alpha}_i)$. The covariance matrix Σ can then be estimated as $(1/T)\hat{U}\hat{U}'$, where $\hat{U}$ is the $m \times T$ matrix of the values of $\hat{u}_{it}$.

Let u_t^* denote a particular draw of the m error terms for period t from the $N(0, \hat{\Sigma})$ distribution. Given u_t^* and given $\hat{\alpha}_i$ for all i, one can solve the model for period t. This is merely a deterministic simulation for the given values of the error terms and coefficients. Call this simulation a "repetition." Another repetition can be made by drawing a new set of values of u_t^* and solving again. This can be done as many times as desired. From each repetition one obtains a prediction of each endogenous variable. Let y_{it}^j denote the value on the jth repetition of variable i for period t. For J repetitions, the stochastic simulation estimate of the expected value of variable i for period t, denoted $\tilde{\mu}_{it}$, is

$$\tilde{\mu}_{it} = \frac{1}{J}\sum_{j=1}^{J} y_{it}^j \tag{7.5}$$

Let

$$\sigma_{it}^{2j} = (y_{it}^j - \tilde{\mu}_{it})^2 \tag{7.6}$$

The stochastic simulation estimate of the variance of variable i for period t, denoted $\tilde{\sigma}_{it}^2$, is then

$$\tilde{\sigma}_{it}^2 = \frac{1}{J}\sum_{j=1}^{J} \sigma_{it}^{2j} \tag{7.7}$$

Given the data from the repetitions, it is also possible to compute the variances of the stochastic simulation estimates and then to examine the precision of the estimates. The variance of $\tilde{\mu}_{it}$ is simply $\tilde{\sigma}_{it}^2/J$. The variance of $\tilde{\sigma}_{it}^2$, denoted $var(\tilde{\sigma}_{it}^2)$, is

$$var(\tilde{\sigma}_{it}^2) = \left(\frac{1}{J}\right)^2 \sum_{j=1}^{J} (\sigma_{it}^{2j} - \tilde{\sigma}_{it}^2)^2 \tag{7.8}$$

In many applications, one is interested in predicted values more than one period ahead, i.e., in predicted values from dynamic simulations. The above discussion can be easily modified to incorporate this case. One simply draws

values for u_t for each period of the simulation. Each repetition is one dynamic simulation over the period of interest. For, say, an eight quarter period, each repetition yields eight predicted values, one per quarter, for each endogenous variable.

It is also possible to draw coefficients for the repetitions. Let $\hat{\alpha}$ denote, say, the 2SLS estimate of all the coefficients in the model, and let $\hat{V}$ denote the estimate of the $k \times k$ covariance matrix of $\hat{\alpha}$. Given $\hat{V}$ and given the normality assumption, an estimate of the distribution of the coefficient estimates is $N(\hat{\alpha}, \hat{V})$. When coefficients are drawn, each repetition consists of a draw of the coefficient vector from $N(\hat{\alpha}, \hat{V})$ and draws of the error terms as above.

An important conclusion that can be drawn from stochastic simulation studies using macroeconometric models is that the values computed from deterministic simulations are quite close to the mean predicted values computed from stochastic simulations. In other words, the bias that results from using deterministic simulation to solve nonlinear models appears to be small.[3]

It may be the case that the forecast means and variances do not exist, although in practice the possible nonexistence of moments is generally ignored. Results in Fair (1984), Section 8.5.5, suggest that the possible nonexistence of moments is not an important problem for macroeconometric models. Alternative measures of dispersion that are robust to the nonexistence problem give very similar results to those obtained using variances.

7.4 Median Unbiased Estimates[4]

The estimator considered in this section will be called the "median unbiased" (MU) estimator. It has been known since the work of Orcutt (1948) and Hurwicz (1950) that least squares estimates of lagged dependent variable (LDV) coefficients are biased even when there are no right hand side endogenous variables. Macroeconometric model builders have generally ignored this problem, perhaps because they feel that the bias is likely to be small for the typical number of observations that are used. Hurwicz's estimates of the bias in an equation with only the LDV as an explanatory variable were small after about 100 observations. For example, for 100 observations the ratio of the expected value of the LDV coefficient estimate to the true value was .9804 (for small values of the coefficient). However, the results in Orcutt and Winokur (1969, Table IV) for 10, 20, and 40 observations show biases larger than those of Hurowitz

[3] See Fair (1984), Section 7.3.4, for references.

[4] The material in this section is taken from Fair (1994a).

for the case in which there is a constant term in the equation, suggesting that the bias in this case is also larger for, say, 100 observations. Furthermore, Andrews (1993) has recently shown that the bias is further increased when a time trend is added to the equation. For example, for 100 observations and a true coefficient of .8, the ratio of the median of the LDV coefficient estimate to the true value is .9388 in the equation with the constant term and time trend added.

Typical macroeconomic equations are, of course, more complicated than the equations just discussed. They have more explanatory variables; some of the explanatory variables are likely to be endogenous; the error terms are sometimes serially correlated; and the equations may be nonlinear in both variables and coefficients. It is important to know how the size of the biases for these types of equations compare to those estimated for simpler equations. The following stochastic simulation procedure provides a way of obtaining median unbiased estimates in macroeconometric models. From these estimates the bias for a coefficient, defined as the difference between the base estimates and the MU estimates, can be computed. For the work here the 2SLS estimates will be taken to be the base estimates. This procedure is an extension of Andrews' (1993) idea of computing exact median unbiased estimates in an equation with a constant term, time trend, and lagged dependent variable.

The procedure requires that one coefficient per stochastic equation be singled out for special treatment. The interest here is on the coefficient of the lagged dependent variable, but other coefficients could be considered. Let α_{1i} denote the coefficient of interest in equation i.

The procedure for obtaining median unbiased estimates of the α_{1i} coefficients ($i = 1, \ldots, m$) using the 2SLS estimator is as follows:

1. Estimate each equation i by 2SLS. Let $\hat{\alpha}_{1i}$ denote the 2SLS estimate of α_{1i}.

2. Guess the bias of $\hat{\alpha}_{1i}$, denoted b_{1i}. Add b_{1i} to $\hat{\alpha}_{1i}$ to obtain a first estimate of the true value of α_{1i}. Let α_{1i}^* denote this estimate: $\alpha_{1i}^* = \hat{\alpha}_{1i} + b_{1i}$. Constrain α_{1i} to be equal to α_{1i}^* and reestimate the other elements of α_i by 2SLS. Let α_i^* denote this estimate of α_i ($i = 1, \ldots, m$). Use the estimated residuals from these constrained regressions to estimate the covariance matrix Σ. Let Σ^* denote this estimate of Σ.

3. Draw T values of the vector u_t^*, $t = 1, \ldots, T$, from the distribution $N(0, \Sigma^*)$. Use these values and the values α_i^* ($i = 1, \ldots, m$) to solve

the model dynamically for $t = 1, \ldots, T$. This is a dynamic simulation of the model over the entire estimation period using the drawn values of the error terms and the coefficient values α_i^*. The lagged endogenous variable values in x_t in 4.1 are updated in the solution process. After this solution, update Z_{it} to incorporate the new lagged endogenous variable values (if lagged endogenous variable values are part of Z_{it}). Let Z_{it}^*, $t = 1, \ldots, T$, denote this update. Given the new data (i.e., the solution values of the endogenous and lagged endogenous variables), estimate each equation by 2SLS, and record the estimate of α_{1i} as $\alpha_{1i}^{(1)}$ $(i = 1, \ldots, m)$. This is one repetition. Do a second repetition by drawing another T values of u_t^*, using these values and the values α_i^* to solve the model, using the new data to estimate each equation by 2SLS, and recording the estimate of α_{1i} as $\alpha_{1i}^{(2)}$ $(i = 1, \ldots, m)$. Do this J times, and then find the median α_{1i}^m of the J values of $\alpha_{1i}^{(j)}$ $(j = 1, \ldots, J)$, $(i = 1, \ldots, m)$.

4. If for each i α_{1i}^m is within a prescribed tolerance level of $\hat{\alpha}_{1i}$, go to step 6. If this condition is met, it means that for the particular coefficient values used to generate the data (the α_i^*'s), the median 2SLS estimates are within a prescribed tolerance level of the original estimates based on the historical data. If this condition is not met, take the new value of α_{1i}^* to be the previous value plus $\hat{\alpha}_{1i} - \alpha_{1i}^m$ for each i. Then constrain α_{1i} to be equal to this new value of α_{1i}^* and reestimate the other elements of α_i by 2SLS using the historical data. Let α_i^* denote this estimate of α_i $(i = 1, \ldots, m)$. Again, use the estimated residuals from these constrained regressions to estimate the covariance matrix Σ. Let Σ^* denote this estimate of Σ. Now repeat step 3 for these new values.

5. Keep doing steps 3 and 4 until convergence is reached and one branches to step 6.

6. Take the median unbiased estimate of α_{1i} to be α_{1i}^*, and take the other coefficient estimates to be those in α_i^* $(i = 1, \ldots, m)$. α_{1i}^* is the median unbiased estimate in that it is the value of α_{1i} that generates data that lead to the median 2SLS estimate equaling (within a prescribed tolerance level) the 2SLS estimate based on the historical data. The estimated bias of $\hat{\alpha}_{1i}$ is $\hat{\alpha}_{1i} - \alpha_{1i}^*$.

Confidence intervals for α_{1i}^m can be computed from the final set of values of $\alpha_{1i}^{(j)}$ $(j = 1, \ldots, J)$. For a 90 percent confidence interval, for example,

5 percent of the smallest values and 5 percent of the largest values would be excluded.

As noted above, this procedure does not require the normality assumption. Other distributions could be used to draw the u_t^* values. Also, the basic estimator need not be the 2SLS estimator. Other estimators could be used. The model in 4.1 can also consist of just one equation. In this case Σ is a scalar and the "solution" of the model simply consists of solving the particular equation (dynamically) over the sample period.

The procedure does, however, have two limitations. First, as noted above, it focuses on just one coefficient per equation. No other coefficient estimate in an equation necessarily has the property that its median value in the final set of values is equal to the original estimate. The focus, of course, need not be on the coefficient of the LDV, but it must be on one particular coefficient per equation.

Second, there is no guarantee that the procedure will converge. Remember that overall convergence requires that convergence be reached for each equation, and achieving this much convergence could be a problem. For the results in the next chapter, however, as will be seen, convergence was never a problem.

7.5 Examining the Accuracy of Asymptotic Distributions[5]

It is possible using stochastic simulation and reestimation to examine whether the asymptotic approximations of the distributions of estimators that are used for hypothesis testing are accurate. If some variables are not stationary, the asymptotic approximations may not be very good. In fact, much of the recent literature in time series econometrics has been concerned with the consequences of nonstationary variables.

The procedure proposed here for examining asymptotic distribution accuracy is similar to the procedure of the previous section. Take an estimator, say 2SLS, 3SLS, or FIML, and estimate the model. Take these coefficient estimates, denoted $\hat{\alpha}$, as the base values, and compute $\hat{\Sigma}$ using these estimates. From the $N(0, \hat{\Sigma})$ distribution (assuming the normality assumption is used), draw a vector of the m error terms for each of the T observations. Given these error terms and $\hat{\alpha}$, solve the model for the entire period 1 through T. As in step 3 of the previous section, this is a dynamic simulation of the model over the entire estimation period. The lagged endogenous variable values in 4.1 are

[5] As in the previous section, the material in this section is taken from Fair (1994a).

updated in the solution process. Also, the matrices of first stage regressors, Z_{it}, are updated to incorporate the new lagged endogenous variable values if the matrices are used in the estimation, as for 2SLS. The predicted values from this solution form a new data set. Given this data set, estimate the model by the technique in question, and record the set of estimates. This is one repetition. Repeat the draws, solution, and estimation for many repetitions, and record each set of estimates. (Remember that the draws of the errors are always from the $N(0, \hat{\Sigma})$ distribution and that the coefficient vector used in the solution is always $\hat{\alpha}$.)

If J repetitions are done, one has J values of each coefficient estimate, which are likely to be a good approximation of the exact distribution. For ease of exposition, this distribution of the J values will be called the "exact distribution," although it is only an approximation because Σ is estimated rather than known. The asymptotic distribution can then be compared to this exact distribution to see how close the two distributions are.

There are a number of ways to examine the closeness of the asymptotic distribution to the exact distribution. For the empirical work in the next chapter, the median of the exact distribution for a coefficient was first compared to the coefficient estimate from the technique in question, which is 2SLS in this case. Remember that these coefficient estimates are the ones used to generate the data. One can then examine the bias of a coefficient estimate, defined as the difference between the median and the coefficient estimate. The coefficient estimates of the lagged dependent variables, for example, are likely to be biased downward, as discussed in the previous section.

Next, given the median from the exact distribution and given the estimated standard error of the coefficient estimate from the asymptotic distribution, one can compute the value above which, say, 20 percent of the coefficient estimates should lie if the asymptotic distribution is correct. For 20 percent, this value is the median plus 0.84 times the estimated asymptotic standard error. One can then compute the actual percent of the coefficient estimates from the exact distribution that lie above this value and compare this percent to 20 percent. For the work in the next chapter, this comparison was made for 20, 10, and 5 percent values and for both left and right tails. It will be seen that the exact and asymptotic distributions are generally quite similar regarding their tail properties.

7.6 VAR and AC Models for Comparison Purposes

When testing complete models, it is useful to have benchmark models to use for comparison purposes. Vector autoregressive (VAR) models provide useful benchmarks. As will be seen in the next chapter, however, if the interest is in GDP predictions, "autoregressive components" (AC) models appear to be better benchmarks than VAR models in the sense of being more accurate. An AC model is one in which each component of GDP is regressed on its own lagged values and lagged values of GDP. GDP is then determined from the GDP identity, as the sum of the components. AC models do not have the problem, as VAR models do, of adding large numbers of parameters as the number of variables (components in the AC case) is increased.

Two VAR Models

Two seven variable VAR models are used in the next chapter for comparison with the US model. The seven variables are (in the notation of the variables in the US model) 1) the log of real GDP, log $GDPR$, 2) the log of the GDP deflator, log $GDPD$, 3) the log of the wage rate, log WF, 4) the log of the import price deflator, log PIM, 5) the log of the money supply, log $M1$, 6) the unemployment rate, UR, and 7) the bill rate, RS. These are the same variables used by Sims (1980) with the exception of RS, which has been added here.

For the first VAR model, denoted VAR4, each of the seven variables is taken to be a function of the constant, a time trend, its first four lagged values, and the first four lagged values of each of the other variables. There are thus 30 coefficients to estimate per each of the seven equations. For the second VAR model, denoted VAR5/2, each of the seven variables is taken to be a function of the constant, a time trend, its first five lagged values, and the first two lagged values of each of the other variables, for a total of 19 coefficients per equation.

It is possible to decrease the number of unrestricted coefficients to estimate in VAR models by imposing various priors on the coefficients. For the work in Fair and Shiller (1990) three sets of Bayesian priors were imposed on VAR4. The results using these versions were similar to the results using VAR4, and very little gain seemed to result from the use of the priors in terms of making the VAR models more accurate. Therefore, although no priors were imposed for the work in this book, the results using VAR4 are likely to be close to the results that would be obtained using priors.[6]

[6]Sims (1993) considers a nine variable VAR model with five lags and imposes an elab-

The AC Model[7]

There are 19 components of $GDPR$ in the US model (counting the statistical discrepancy $STATP$), and the AC model used in the next chapter consists of estimated equations for each of these components.[8] Each component is taken to be a function of the constant, a time trend, its first five lagged values, and the first two lagged values of $GDPR$. The final equation of the AC model is the $GDPR$ identity.

The results in Fair and Shiller (1990) show that going from a few components to 17 improves the accuracy of the AC model,[9] but that going beyond this does not. The results also show that adding lagged values of $GDPR$ (versus not having $GDPR$ in the equations at all) leads to a slight improvement in accuracy. As with different versions of the VAR model, however, the results are not highly sensitive to different versions of the AC model (i.e., alternative choices of number of components, the length of the lag, and whether or not lagged values of $GDPR$ are included).

7.7 Comparing Predictive Accuracy[10]

As discussed in Section 7.3, stochastic simulation allows one to compute forecast error variances. Let $\tilde{\sigma}^2_{itk}$ denote the stochastic simulation estimate of the variance of the forecast error for a k period ahead forecast of variable i from a simulation beginning in period t. This estimate is presented in equation 7.7 except that a k subscript has been added to denote the length ahead of the forecast. If the estimated variance is based on draws of both the error terms and coefficients, then the uncertainty from both of these sources has been accounted for.

One might think that forecast error variances computed in this way could

orate set of priors on the coefficients. In future work it would be interesting to see how well this model does compared to, say, VAR4, but at the present time it would be extremely difficult to try to duplicate Sims' procedures.

[7] AC models were first proposed in Fair and Shiller (1990).

[8] The 19 components in alphabetical order are CD, CN, COG, COS, CS, EX, IHB, IHF, IHH, IKB, IKF, IKG, IKH, IM, IVF, IVH, $PROG$, $PROS$, and $STATP$. $PROG$ and $PROS$ are combined in the US model in such a way that they do not appear as separate variables. They are, however, raw data variables and are defined in Table A.4.

[9] The number of components in the US model at the time of this work was 17, hence 17 instead of 19 components were used.

[10] The method discussed in this section was briefly outlined in Section 1.2, and it is discussed in more detail in Fair (1980a) and in Fair (1984), Chapter 8.

simply be compared across models to see which variances are smaller. There are, however, two additional problems. The first is controlling for different sets of exogenous variables across models (VAR and AC models, for example, have no exogenous variables, whereas a model like the US model has many). This can be done in a variety of ways. One is to estimate autoregressive equations for each exogenous variable and add these equations to the model. The expanded model can then be stochastically simulated to get the variances. The expanded model in effect has no exogenous variables. Another way is to estimate in some manner the forecast error variance for each exogenous variable (perhaps using past errors made by forecasting services in forecasting the variable) and then use these estimates and, say, the normality assumption to draw exogenous variable values for the stochastic simulation.

The second problem is the possibility of data mining. A model may have small estimated variances of the structural error terms and small estimated variances of the coefficient estimates (which leads to small forecast error variances from the stochastic simulation) because it has managed to spuriously fit the sample well. A further step is needed to handle this problem, which is to compare variances estimated from outside sample forecast errors with variances estimated from stochastic simulation. 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 adjust the forecast error variances for the effects of misspecification.

Let the prediction period begin one period after the end of the estimation period, and call this period s. Consider stochastic simulation with both error terms and coefficients drawn. From a stochastic simulation beginning in period s one obtains an estimate of the variance of the forecast error, $\tilde{\sigma}^2_{isk}$, in equation 7.7, where again k refers to the length ahead of the forecast. From this simulation one also obtains an estimate of the expected value of the k period ahead forecast of variable i, $\tilde{\mu}_{isk}$, in equation 7.5. The difference between this estimate and the actual value, y_{is+k-1}, is the mean forecast error, denoted $\hat{\epsilon}_{isk}$:

$$\hat{\epsilon}_{isk} = y_{is+k-1} - \tilde{\mu}_{isk} \tag{7.9}$$

If it is assumed that $\tilde{\mu}_{isk}$ exactly equals the true expected value, then $\hat{\epsilon}_{isk}$ in equation 7.9 is a sample draw from a distribution with a known mean of zero and variance σ^2_{isk}, where σ^2_{isk} is the true variance. The square of this error, $\hat{\epsilon}^2_{isk}$, is thus under this assumption an unbiased estimate of σ^2_{isk}. One therefore has two estimates of σ^2_{isk}, one computed from the mean forecast error and one

computed by stochastic simulation. Let d_{isk} denote the difference between these two estimates:

$$d_{isk} = \hat{\epsilon}_{isk}^2 - \tilde{\sigma}_{isk}^2 \tag{7.10}$$

If it is further assumed that $\tilde{\sigma}_{isk}^2$ exactly equals the true value (i.e., $\tilde{\sigma}_{isk}^2 = \sigma_{isk}^2$), then d_{isk} 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_{isk} is zero for a correctly specified model.

If a model is misspecified, it is not in general true that the expected value of d_{isk} is zero. Misspecification has two effects on d_{isk}. 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_{isk} 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_{isk} to be zero for a misspecified model.

Because of the common practice in macroeconometric work of searching for equations that fit the data well (data mining), it seems likely that the estimated means of d_{isk} will be positive in practice for a misspecified model. If the model fits the data well within sample, the stochastic simulation estimates of the variances will be small 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_{isk} is thus likely to be positive: $\tilde{\sigma}_{isk}^2$ is small because of small estimated covariance matrices, and $\hat{\epsilon}_{isk}^2$ is large because of large outside sample forecast errors.

The procedure described so far uses only one estimation period and one prediction period, where the estimation period ends in period $s-1$ and the prediction period begins in period s. It results in one value of d_{isk} for each variable i and each length ahead k. Since one observation is obviously not adequate for estimating the mean of d_{isk}, 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 150. The model can be estimated through, say, period 100, with the prediction beginning with period 101. Stochastic simulation for the prediction period will yield for each i and k a value of d_{i101k} in equation 7.10. The model can then be reestimated through period 101, with the prediction period now beginning with period 102. Stochastic simulation for this prediction period will yield for each i and k a value of d_{i102k}. This process can be repeated through the estimation period ending with period 149. For the one period ahead forecast ($k = 1$) the procedure will yield for each variable i 50 values of d_{is1} ($s = 101, \ldots, 150$); for the two period ahead forecast ($k = 2$) it will yield 49 values of d_{is2}, ($s = 101, \ldots, 149$); and so on.

The final step in the process is to make an assumption about the mean of d_{isk} that allows the computed values of d_{isk} to be used to estimate the mean. A variety of assumptions are possible, which are discussed in Fair (1984), Chapter 8. The assumption made for the empirical work in the next chapter is that the mean is constant across time. In other words, misspecification is assumed to affect the mean in the same way for all s. Given this assumption, the mean, denoted as $\bar{d}_{ik}$, can be estimated by merely averaging the computed values of d_{isk}. Note that calculating the individual d_{isk} values that are needed to calculate $\bar{d}_{ik}$ is computer intensive in that it requires estimating and stochastic simulating many times.

Given $\bar{d}_{ik}$, it is possible to estimate the total variance of the forecast error. Assume that the period of interest begins in period t, and let $\tilde{\tilde{\sigma}}^2_{itk}$ denote the stochastic simulation estimate of the variance based on draws of error terms, coefficients, and exogenous variables. The total variance, denoted $\hat{\sigma}^2_{itk}$, is the sum of the stochastic simulation estimate plus $\bar{d}_{ik}$:

$$\hat{\sigma}^2_{itk} = \tilde{\tilde{\sigma}}^2_{itk} + \bar{d}_{ik} \tag{7.11}$$

Since the procedure in arriving at $\hat{\sigma}^2_{itk}$ takes into account the four main sources of uncertainty of a forecast, it can be compared across models for a given i, t, and k.

7.8 Comparing Information in Forecasts[11]

Introduction

This section discusses an alternative way of comparing models from the method of comparing variances in the previous section. It focuses on the

[11]The material in this section is taken from Fair and Shiller (1990).

information contained in each model's forecast. Econometric models obviously differ in structure and in the data used, and so their forecasts are not perfectly correlated with each other. How should one interpret the differences in forecasts? Does each model have a strength of its own, so that each forecast represents useful information unique to it, or does one model dominate in the sense of incorporating all the information in the other models plus some?

Structural econometric models make use of large information sets in forecasting a given variable. The information set used in a large scale macroeconometric model is typically so large that the number of predetermined variables exceeds the number of observations available for estimating the model. Estimation can proceed effectively only because of the large number of *a priori* restrictions imposed on the model, restrictions that do not work out to be simple exclusion restrictions on the reduced form equation for the variable forecasted.

VAR models are typically much smaller than structural models and in this sense use less information. The above question with respect to VAR models versus structural models is thus whether the information not contained in VAR models (but contained in structural models) is useful for forecasting purposes. In other words, are the *a priori* restrictions of large scale models useful in producing derived reduced forms that depend on so much information, or is most of the information extraneous? The same question can be asked of AC models versus structural models.

One cannot answer this question by doing conventional tests of the restrictions in a structural model. These restrictions might be wrong in important ways and yet the model contain useful information. Even ignoring this point, however, one cannot perform such tests with most large scale models because, as noted above, there are not enough observations to estimate unrestricted reduced forms.

The question whether one model's forecast of a variable, for example, real GDP, carries different information from another's can be examined by regressing the actual change in the variable on the forecasted changes from the two models. This procedure, which is discussed below, is related to the literature on encompassing tests[12] and the literature on the optimal combination of forecasts.[13] This procedure has two advantages over the standard procedure of computing root mean squared errors (RMSEs) to compare alternative forecasts. First, if the RMSEs are close for two forecasts, little can be

[12] See, for example, Davidson and MacKinnon (1981), Hendry and Richard (1982), Chong and Hendry (1986), and Mizon and Richard (1986). See also Nelson (1972) and Cooper and Nelson (1975) for an early use of encompassing like tests.

[13] See, for example, Granger and Newbold (1986).

concluded about the relative merits of the two. With the current procedure one can sometimes discriminate more. Second, even if one RMSE is much smaller than the other, it may still be that the forecast with the higher RMSE contains information not in the other forecast. There is no way to test for this using the RMSE framework.

It should be stressed that the current procedure does not allow one to discover whether all the variables in a model contribute useful information for forecasting. If, say, the regression results reveal that a large model contains all the information in smaller models plus some, it may be that the good results for the large model are due to a small subset of it. It can only be said that the large model contains all the information in the smaller models that it has been tested against, not that it contains no extraneous variables.

The procedure requires that forecasts be based only on information available prior to the forecast period. Assume that the beginning of the forecast period is t, so that only information through period $t-1$ should be used for the forecasts. There are four ways in which future information can creep into a current forecast. The first is if actual values of the exogenous variables for periods after $t-1$ are used in the forecast. The second is if the coefficients of the model have been estimated over a sample period that includes observations beyond $t-1$. The third is if information beyond $t-1$ has been used in the specification of the model even though for purposes of the tests the model is only estimated through period $t-1$. The fourth is if information beyond period $t-1$ has been used in the revisions of the data for periods $t-1$ and back, such as revised seasonal factors and revised benchmark figures.

One way to handle the exogenous variable problem is to estimate, say, an autoregressive equation for each exogenous variable in the model and add these equations to the model. The expanded model effectively has no exogenous variables in it. This method of dealing with exogenous variables in structural models was advocated by Cooper and Nelson (1975) and McNees (1981). McNees, however, noted that the method handicaps the model: "It is easy to think of exogenous variables (policy variables) whose future values can be anticipated or controlled with complete certainty even if the historical values can be represented by covariance stationary processes; to do so introduces superfluous errors into the model solution." (McNees, 1981, p. 404). For the work in the next chapter autoregressive equations have been estimated for each exogenous variable in the US model, although, as McNees notes, this may bias the results against the US model.

The coefficient problem can be handled by doing rolling estimations for each model. For the forecast for period t, for example, the model can be

estimated through period $t-1$; for the forecast for period $t+1$, the model can be estimated through period t; and so on. By "model" in this case is meant the model inclusive of any exogenous variable equations. If the beginning observation is held fixed for all the regressions, the sample expands by one observation each time a time period elapses. This rolling estimation was followed for the work in the next chapter.

The third problem—the possibility of using information beyond period $t-1$ in the specification of the model—is more difficult to handle. Models are typically changed through time, and model builders seldom go back to or are interested in "old" versions. For the work in Fair and Shiller (1990), however, a version of the US model was used that existed as of the second quarter of 1976, and all the predictions were for the period after this. For the work in the next chapter the current version of the US model has been used, and so this potential problem has been ignored here. This may bias the results in favor of the US model, although the changes in the model that have been made since 1976 are fairly minor.

The data revision problem is very hard to handle, and almost no one tries. It is extremely difficult to try to purge the data of the possible use of future information. It is not enough simply to use data that existed at any point in time, say period $t-1$, because data for period t are needed to compare the predicted values to the actual values. To handle the data revision problem one would have to try to construct data for period t that are consistent with the old data for period $t-1$, and this is not straightforward. For the work in the next chapter nothing has been done about this problem either.

Forecasts that are based only on information prior to the forecast period will be called "quasi ex ante" forecasts. They are not true ex ante forecasts if they were not issued at the time, but they are forecasts that could in principle have been issued had one been making forecasts at the time.

Quasi ex ante forecasts may, of course, have different properties from forecasts made with a model estimated with future data. If the model is misspecified (e.g., parameters change through time), then the rolling estimation forecasts (where estimated parameters vary through time) may carry rather different information from forecasts estimated over the entire sample.[14] The

[14]Even if the model is not misspecified, estimated parameters will change through time due to sampling error. If the purpose were to evaluate the forecasting ability of the true model (i.e., the model with the true coefficients), there would be a generated regressor problem. However, the interest here is in the performance of the model *and* its associated estimation procedure. If one were interested in adjusting for generated regressors, the correction discussed in Murphy and Topel (1985) could not be directly applied here because

focus here is on quasi ex ante forecasts.

It should also be noted that some models may use up more degrees of freedom in estimation than others, and with varied estimation procedures it is often very difficult to take formal account of the number of degrees of freedom used up. In the extreme case where there were so many parameters in a model that the degrees of freedom were completely used up when it was estimated (an obviously over parameterized model), it would be the case that the forecast value equals the actual value and there would be a spurious perfect correspondence between the variable forecasted and the forecast. One can guard against this degrees of freedom problem by requiring that no forecasts be within sample forecasts, which is true of quasi ex ante forecasts proposed here.[15]

The Procedure

Let ${}_{t-s}\hat{Y}_{1t}$ denote a forecast of Y_t made from model 1 using information available at time $t-s$ and using the model's estimation procedure and forecasting method each period. Let ${}_{t-s}\hat{Y}_{2t}$ denote the same thing for model 2. (In the notation above, these two forecasts should be quasi ex ante forecasts.) The parameter s is the length ahead of the forecast, $s > 0$. Note that the estimation procedure used to estimate a model and the model's forecasting method are considered as part of the model; no account is taken of these procedures here.

The procedure is based on the following regression equation:

$$Y_t - Y_{t-s} = \alpha + \beta({}_{t-s}\hat{Y}_{1t} - Y_{t-s}) + \gamma({}_{t-s}\hat{Y}_{2t} - Y_{t-s}) + u_t \qquad (7.12)$$

If neither model 1 nor model 2 contains any information useful for s period ahead forecasting of Y_t, then the estimates of β and γ should both be zero. In this case the estimate of the constant term α would be the average s period change in Y. If both models contain independent information[16] for s period ahead forecasting, then β and γ should both be nonzero. If both models contain information, but the information in, say, model 2 is completely contained in

the covariance matrix of the coefficient estimates used to generate the forecasts changes through time because of the use of the rolling regressions. Murphy and Topel require a single covariance matrix.

[15]Nelson (1972) and Cooper and Nelson (1975) do not stipulate that the forecasts be based only on information through the previous period.

[16]If both models contain "independent information" in the present terminology, their forecasts will not be perfectly correlated. Lack of perfect correlation can arise either because the models use different data or because they use the same data but impose different restrictions on the reduced form.

model 1 and model 1 contains further relevant information as well, then β but not γ should be nonzero.[17]

The procedure is to estimate equation 7.12 for different models' forecasts and test the hypothesis H_1 that $\beta = 0$ and the hypothesis H_2 that $\gamma = 0$. H_1 is the hypothesis that model 1's forecasts contain no information relevant to forecasting s periods ahead not in the constant term and in model 2, and H_2 is the hypothesis that model 2's forecasts contain no information not in the constant term and in model 1.

As noted above, this procedure bears some relation to encompassing tests, but the setup and interests are somewhat different. For example, it does not make sense in the current setup to constrain β and γ to sum to one, as is usually the case for encompassing tests. If both models' forecasts are just noise, the estimates of both β and γ should be zero. Also, say that the true process generating Y_t is $Y_t = X_t + Z_t$, where X_t and Z_t are independently distributed. Say that model 1 specifies that Y_t is a function of X_t only and that model 2 specifies that Y_t is a function of Z_t only. Both forecasts should thus have coefficients of one in equation 7.12, and so in this case β and γ would sum to two. It also does not make sense in the current setup to constrain the constant term α to be zero. If, for example, both models' forecasts were noise and equation 7.12 were estimated without a constant term, then the estimates of β and γ would not generally be zero when the mean of the dependent variable is nonzero.

It is also not sensible in the current setup to assume that u_t is identically distributed. It is likely that u_t is heteroskedastic. If, for example, $\alpha = 0$, $\beta = 1$, and $\gamma = 0$, u_t is simply the forecast error from model 1, and in general forecast errors are heteroskedastic. Also, if k period ahead forecasts are considered, where $k > 1$, this introduces a $k - 1$ order moving average process to the error term in equation 7.12.[18] Both heteroskedasticity and the moving average process can be corrected for in the estimation of the standard errors of the coefficient estimates. This can be done using the procedure given by Hansen (1982), Cumby, Huizinga, and Obstfeld (1983), and White and Domowitz (1984) for the estimation of asymptotic covariance matrices. Let $\theta = (\alpha \;\; \beta \;\; \gamma)'$. Also, define X as the $T \times 3$ matrix of variables, whose row t is $X_t = (1 \;\; {}_{t-s}\hat{Y}_{1t} - Y_{t-s} \;\; {}_{t-s}\hat{Y}_{2t} - Y_{t-s})$, and let $\hat{u}_t = Y_t - Y_{t-s} - X_t\hat{\theta}$.

[17] If both models contain the same information, then the forecasts are perfectly correlated, and β and γ are not separately identified.

[18] The error term in equation 7.12 could, of course, be serially correlated even for the one period ahead forecasts. Such serial correlation, however, does not appear to be a problem for the work in the next chapter, and so it has been assumed to be zero here.

The covariance matrix of $\hat{\theta}$, $V(\hat{\theta})$, is

$$V(\hat{\theta}) = (X'X)^{-1}S(X'X)^{-1} \tag{7.13}$$

$$S = \Omega_0 + \sum_{j=1}^{s-1}(\Omega_j + \Omega_j') \tag{7.14}$$

$$\Omega_j = \sum_{t=j+1}^{T}(u_t u_{t-j})\hat{X}_t'\hat{X}_{t-j} \tag{7.15}$$

where $\hat{\theta}$ is the ordinary least squares estimate of θ and s is the forecast horizon. When s equals 1, the second term on the right hand side of 7.14 is zero, and the covariance matrix is simply White's (1980) correction for heteroskedasticity.

Note that as an alternative to equation 7.12 the *level* of Y could be regressed on the forecasted *levels* and a constant. If Y is an integrated process, then any sensible forecast of Y will be cointegrated with Y itself. In the level regression, the sum of β and γ will thus be constrained in effect to one, and one would in effect be estimating one less parameter. If Y is an integrated process, running the levels regression with an additional independent variable Y_{t-1} (thereby estimating β and γ without constraining their sum to one) is essentially equivalent to the differenced regression 7.12. For variables that are not integrated, the levels version of 7.12 can be used.

It should finally be noted that there are cases in which an optimal forecast does not tend to be singled out as best in regressions of the form 7.12, even with many observations. Say the truth is $Y_t - Y_{t-1} = aX_{t-1} + e_t$. Say that model 1 does rolling regressions of $Y_t - Y_{t-1}$ on X_{t-1} and uses these regressions to forecast. Say that model 2 always takes the forecast to be bX_{t-1} where b is some number other than a, so that model 2 remains forever an incorrect model. In equation 7.12 regressions the two forecasts tend to be increasingly collinear as time goes on; essentially they are collinear after the first part of the sample. Thus, the estimates of β and γ tend to be erratic. Adding a large number of observations does not cause the regressions to single out the first model; it only has the effect of enforcing that $\hat{\beta} + (\hat{\gamma}b)/a = 1$.

7.9 Estimating Event Probabilities[19]

Stochastic simulation can be used to calculate the probability of various events happening. This is straightforward once the stochastic simulation has been set

[19]The material in this section is taken from Fair (1993c)

up and the event defined. Consider a five quarter prediction period and the event that within this period there were two consecutive quarters of negative real GDP growth. Assume that 1000 repetitions are taken. For each repetition one can record whether or not this event occurred. If it occurred, say, 150 times out of the 1000 repetitions, its estimated probability would be 15 percent. Many events, of course, can be considered. The only extra work for each extra event is keeping track of how often each event occurs in the repetitions.

Government policy makers and business planners are obviously interested in knowing the probabilities of various economic events happening. Model builders who make forecasts typically do not directly answer probability questions. They typically present a "base" forecast and a few alternative "scenarios." If probabilities are assigned to the scenarios, they are subjective ones of the model builders.[20] An advantage of estimating probabilities from stochastic simulation is that they are objective in the sense that they are based on the use of estimated distributions. They are consistent with the probability structure of the model.

In estimating probabilities by stochastic simulation, it seems best to draw only error terms (not also coefficients). Although coefficient estimates are uncertain, the true coefficients are fixed. In the real world, the reason that economic events are stochastic is because of stochastic shocks (error terms), not because the coefficients are stochastic. (This is assuming, of course, that the true coefficients are fixed, which is the assumption upon which the estimation is based.)[21] For the estimation of probabilities in the next chapter, only error terms are drawn.

This procedure for estimating probabilities can also be used for testing purposes. It is possible for a given event to compute a *series* of probability estimates and compare these estimates to the actual outcomes. Consider an event A_t, such as two consecutive quarters of negative growth out of five for the period beginning in quarter t. Let P_t denote a model's estimate of the probability of A_t occurring, and let R_t denote the actual outcome of A_t, which is 1 if A_t occurred and 0 otherwise. If one computes these probabilities for $t = 1, \ldots, T$, there are T values of P_t and R_t available, where each value of P_t is derived from a separate stochastic simulation.

To see how good a model is at estimating probabilities, P_t can be compared

[20] Stock and Watson (1989) do present, however, within the context of their leading indicator approach, estimates of the probability that the economy will be in a recession six months hence.

[21] I am indebted to Gregory Chow for suggesting to me that one may not want to draw coefficients when estimating probabilities.

to R_t for $t = 1, \ldots, T$. Two common measures of the accuracy of probabilities are the quadratic probability score (QPS):

$$QPS = (1/T) \sum_{t=1}^{T} 2(P_t - R_t)^2 \tag{7.16}$$

and the log probability score (LPS):

$$LPS = -(1/T) \sum_{t=1}^{T} [(1 - R_t) \log(1 - P_t) + R_t \log P_t] \tag{7.17}$$

where T is the total number of observations.[22] It is also possible simply to compute the mean of P_t (say $\bar{P}$) and the mean of R_t (say $\bar{R}$) and compare the two means. QPS ranges from 0 to 2, with 0 being perfect accuracy, and LPS ranges from 0 to infinity, with 0 being perfect accuracy. Larger errors are penalized more under LPS than under QPS.

The testing procedure is thus simply to define various events and compute QPS and LPS for alternative models for each event. If model 1 has lower values than model 2, this is evidence in favor of model 1.

7.10 Full Information Estimation and Solution of Rational Expectations Models[23]

Introduction

The single equation estimation of equations with rational expectations was discussed in Section 4.3, where Hansen's method was described. It is also possible, however, to use FIML to estimate models with rational expectations. Methods for the solution and FIML estimation of these models were presented in Fair and Taylor (1983) and also discussed in Fair (1984), Chapter 11. The basic solution method, called the "extended path" (EP) method, has come to be widely used for deterministic simulations of rational expectations models,[24]

[22] See, for example, Diebold and Rudebusch (1989).

[23] The material in this section is taken from Fair and Taylor (1990).

[24] For example, the extended path method has been programmed as part of the TROLL computer package and is routinely used to solve large scale rational expectations models at the IMF, the Federal Reserve, the Canadian Financial Ministry, and other government agencies. It has also been used for simulation studies such as DeLong and Summers (1986) and King (1988). Other solution methods for rational expectations models are summarized in Taylor and Uhlig (1990). These other methods do not yet appear practical for medium size models and up.

but probably because of the expense, the full information estimation method has not been tried by others. This earlier work discussed a "less expensive" method for obtaining full information estimates, but the preliminary results using the method were mixed. Since this earlier work, however, more experimenting with the less expensive method has been done, and it seems much more promising than was originally thought.

This section has two objectives. The first is to discuss the new results using the less expensive method that have been obtained and to argue that full information estimation now seems feasible for rational expectations models. In the process of doing this some errors in the earlier work regarding the treatment of models with rational expectations and autoregressive errors are corrected. The second objective is to discuss methods for stochastic simulation of rational expectations models, something that was only briefly touched on in the earlier work.

The Solution Method

The notation for the model used here differs somewhat from the notation used in equation 4.1. The lagged values of the endogenous variables are written out explicitly, and x_t is now a vector of only exogenous variables. The model is written as

$$f_i(y_t, y_{t-1}, \ldots, y_{t-p}, E_{t-1}y_t, E_{t-1}y_{t+1}, \ldots, E_{t-1}y_{t+h}, x_t, \alpha_i) = u_{it} \tag{7.18}$$

$$u_{it} = \rho_i u_{it-1} + \epsilon_{it}, \quad (i = 1, \ldots, n) \tag{7.19}$$

where y_t is an n–dimensional vector of endogenous variables, x_t is a vector of exogenous variables, E_{t-1} is the conditional expectations operator based on the model and on information through period $t-1$, α_i is a vector of parameters, ρ_i is the serial correlation coefficient for the error term u_{it}, and ϵ_{it} is an error term that may be correlated across equations but not across time. The function f_i may be nonlinear in variables, parameters, and expectations. The following is a brief review of the solution method for this model. More details are presented in Fair and Taylor (1983) and in Fair (1984), Chapter 11. In what follows i is always meant to run from 1 through n.

Case 1: $\rho_i = 0$

Consider solving the model for period s. It is assumed that estimates of α_i are available, that current and expected future values of the exogenous variables

are available, and that the current and future values of the error terms have been set to their expected values (which will always be taken to be zero here). If the expectations $E_{s-1}y_s$, $E_{s-1}y_{s+1}$, ..., $E_{s-1}y_{s+h}$ were known, 7.18 could be solved in the usual ways (usually by the Gauss-Seidel technique). The model would be simultaneous, but future predicted values would not affect current predicted values. The EP method iterates over solution *paths*. Values of the expectations through period $s+h+k+h$ are first guessed, where k is a fairly large number relative to h.[25] Given these guesses, the model can be solved for periods s through $s+h+k$ in the usual ways. This solution provides new values for the expectations through period $s+h+k$—the new expectations values are the solution values. Given these new values, the model can be solved again for periods s through $s+h+k$, which provides new expectations values, and so on. This process stops (if it does) when the solution values for one iteration are within a prescribed tolerance criterion of the solution values for the previous iteration for all periods s through $s+h+k$.

So far the guessed values of the expectations for periods $s+h+k+1$ through $s+h+k+h$ (the h periods beyond the last period solved) have not been changed. If the solution values for periods s through $s+h$ depend in a nontrivial way on these guesses, then overall convergence has not been achieved. To check for this, the entire process above is repeated for k one larger. If increasing k by one has a trivial effect (based on a tolerance criterion) on the solution values for s through $s+h$, then overall convergence has been achieved; otherwise k must continue to be increased until the criterion is met. In practice what is usually done is to experiment to find the value of k that is large enough to make it likely that further increases are unnecessary for any experiment that might be run and then do no further checking using larger values of k.

The expected future values of the exogenous variables (which are needed for the solution) can either be assumed to be the actual values (if available and known by agents) or be projected from an assumed stochastic process. If the expected future values of the exogenous variables are not the actual values, one extra step is needed at the end of the overall solution. In the above process the expected values of the exogenous variables would be used for all the solutions, the expected values of the exogenous variables being chosen ahead of time. This yields values for $E_{s-1}y_s$, $E_{s-1}y_{s+1}$, ..., $E_{s-1}y_{s+h}$. Given these values,

[25] Guessed values are usually taken to be the actual values if the solution is within the period for which data exist. Otherwise, the last observed value of a variable can be used for the future values or the variable can be extrapolated in some simple way. Sometimes information on the steady state solution (if there is one) can be used to help form the guesses.

7.18 is then solved for period s using the *actual* value of x_s, which yields the final solution value $\hat{y}_s$. To the extent that the expected value of x_s differs from the actual value, $E_{s-1}y_s$ will differ from $\hat{y}_s$.

Two points about this method should be mentioned. First, no general convergence proofs are available. If convergence is a problem, one can sometimes "damp" the solution values to obtain convergence. In practice convergence is usually not a problem. There may, of course, be more than one set of solution values, and so there is no guarantee that the particular set found is unique. If there is more than one set, the set that the method finds may depend on the guesses used for the expectations for the h periods beyond $s + h + k$.

Second, the method relies on the certainty equivalence assumption even though the model is nonlinear. Since expectations of functions are treated as functions of the expectations in future periods in equation 7.18, the solution is only approximate unless f_i is linear. This assumption is like the linear quadratic approximation to rational expectations models that has been proposed, for example, by Kydland and Prescott (1982). Although the certainty equivalence assumption is widely used, including in the engineering literature, it is, of course, not always a good approximation.

Case 2: $\rho_i \neq 0$ and Data Before $s - 1$ Available

The existence of serial correlation complicates the problem considerably. The error terms for period $t - 1$ ($u_{it-1}, i = 1, \ldots, n$) depend on expectations that were formed at the end of period $t - 2$, and so a new viewpoint date is introduced. This case is discussed in Section 2.2 in Fair and Taylor (1983), but an error was made in the treatment of the second viewpoint date. The following method replaces the method in Section 2.2 of this paper.[26]

Consider again solving for period s. If the values of u_{is-1} were known, one could solve the model as above. The only difference is that the value of an error term like u_{is+r-1}would be $\rho_i^r u_{is-1}$ instead of zero. The overall solution method first uses the EP method to solve for period $s - j$, where $j > 0$, based on the assumption that $u_{is-j-1} = 0$. Once the expectations are solved for, 7.18 is used to solve for u_{is-j}. The actual values of y_{s-j} and x_{s-j} are used for this purpose (although the solution values are used for the expectations) because these are structural errors being estimated, not reduced form errors. Given the values for u_{is-j}, the model is solved for period $s - j + 1$ using the EP method, where an error term like u_{is-j+r} is computed as $\rho_i^r u_{is-j}$. Once

[26]The material in Fair and Taylor (1983) is also presented in Fair (1984), Chapter 11, and so the corrections discussed in this section pertain to both sources.

the expectations are solved for, 7.18 is used to solve for u_{is-j+1}, which can be used in the solution for period $s-j+2$, and so on through the solution for period s.

The solution for period s is based on the assumption that the error terms for period $s-j-1$ are zero. To see if the solution values for period s are sensitive to this assumption, the entire process is repeated with j increased by 1. If going back one more period has effects on the solution values for period s that are within a prescribed tolerance criterion, then overall convergence has been achieved; otherwise j must continue to be increased. Again, in practice one usually finds a value of j that is large enough to make it likely that further increases are unnecessary for any experiment that might be run and then do no further checking using larger values of j.

It should be noted that once period s is solved for, period $s+1$ can be solved for without going back again. From the solution for period s, the values of u_{is} can be computed, which can then be used in the solution for period $s+1$ using the EP method.

Case 3: $\rho_i \neq 0$ and Data Before Period $s-1$ not Available

This case is based on the assumption that $\epsilon_{is-1} = 0$ when solving for period s. This type of an assumption is usually made when estimating multiple equation models with moving average residuals. The solution problem is to find the values of u_{is-1} that are consistent with this assumption. The overall method begins by guessing values for u_{is-2}. Given these values, the model can be solved for period $s-1$ using the EP method and the fact that $u_{is+r-2} = \rho_i^r u_{is-2}$. From the solution values for the expectations, 7.18 and 7.19 can be used to solve for ϵ_{is-1}.[27] If the absolute values of these errors are within a prescribed tolerance criterion, convergence has been achieved. Otherwise, the new guess for u_{is-2} is computed as the old guess plus ϵ_{is-1}/ρ_i. The model is solved again for period $s-1$ using the new guess and the EP method, and so on until convergence is reached.

At the point of convergence u_{is-1} can be computed as $\rho_i u_{is-2}$, where u_{is-2} is the estimated value on the last iteration (the value consistent with ϵ_{is-1} being within a prescribed tolerance criterion of zero). Given the values of u_{is-1}, one can solve for period s using the EP method, and the solution is finished.

[27] These are again estimates of the structural error terms, not the reduced form error terms. Step (iii) on page 1176 in Fair and Taylor (1983) is in error in this respect. The errors computed in step (iii) should be the structural error terms.

Computational Costs

The easiest way to think about the computational costs of the solution method is to consider how many times the equations of a model must be "passed" through. Let N_1 be the number of passes through the model that it takes to solve the model for one period, given the expectations. N_1 is usually some number less than 10 when the Gauss-Seidel technique is used. The EP method requires solving the model for $h + k + 1$ periods. Let N_2 be the number of iterations it takes to achieve convergence over these periods. Then the total number of passes for convergence is $N_2 N_1 (h + k + 1)$. If, say, h is 5, k is 30, N_2 is 15, and N_1 is 5, then the total number of passes needed to solve the model for one period is 11,250, which compares to only 5 when there are no expectations. If k is increased by one to check for overall convergence, the total number of passes is slightly more than doubled, although, as noted above, this check is not always done.

For Case 2 above the number of passes is increased by roughly a factor of j if overall convergence is not checked. Checking for overall convergence slightly more than doubles the number of passes. j is usually a number between 5 and 10. If q is the number of iterations it takes to achieve convergence for Case 3 above, the number of passes is increased by a factor of $q+1$. In practice q seems to be between about 5 and 10. Note for both Cases 2 and 3 that the number of passes is increased relative to the non serial correlation case only for the solution for the first period (period s). If period $s + 1$ is to be solved for, no additional passes are needed over those for the regular case.

FIML Estimation

Assume that the estimation period is 1 through T. The objective function that FIML maximizes (assuming normality) is presented in equation 7.3 above and is repeated here for convenience

$$L = -\frac{T}{2} \log |\Sigma| + \sum_{t=1}^{T} \log |J_t| \tag{7.20}$$

Σ is the covariance matrix of the error terms and J_t is the Jacobian matrix for period t. Σ is of the dimension of the number of stochastic equations in the model, and J_t is of the dimension of the total number of equations in the model. The ij element of Σ is $(1/T)\Sigma_{t=1}^{T} \epsilon_{it}\epsilon_{jt}$. Since the expectations have viewpoint date $t-1$, they are predetermined from the point of view of taking derivatives for the Jacobian, and so no additional problems are involved for the

Jacobian in the rational expectations case. In what follows α will be used to denote the vector of all the coefficients in the model. In the serial correlation case α also includes the ρ_i coefficients.

FIML estimation of moderate to large models is expensive even in the standard case, and some tricks are needed to make the problem computationally feasible. An algorithm that can be used for large scale applications is discussed in Parke (1982), and this algorithm will not be discussed here. Suffice it to say that FIML estimation of large scale models is computationally feasible, and in fact FIML estimates of the US model are presented in the next chapter. What any algorithm needs to do is to evaluate L many times for alternative values of α in the search for the value that maximizes L.

In the standard case computing Σ for a given value of α is fairly inexpensive. One simply solves 7.18 and 7.19 for the ϵ_{it} error terms given the data and the value of α. This is only one pass through the model since it is the structural error terms that are being computed. In the rational expectations case, however, computing the error terms requires knowing the values of the expectations, which themselves depend on α. Therefore, to compute Σ for a given value of α one has to solve for the expectations for each of the T periods. If, say, 11,250 passes through the model are needed to solve the model for one period and if T is 100, then 1,125,000 passes are needed for one evaluation of Σ and thus one evaluation of L. In the 25 coefficient problem below, the Parke algorithm required 2,817 evaluations of L to converge, which would be over 3 trillion passes if done this way.[28]

It should be clear that the straightforward combination of the EP solution method and FIML estimation procedures is not likely to be computationally feasible for most applications. There is, however, a way of cutting the number of times the model has to be solved over the estimation period to roughly the number of estimated coefficients. The trick is to compute numerical derivatives of the expectations with respect to the parameters and use these derivatives to compute Σ (and thus L) each time the algorithm requires a value of L for a given value of α.

Consider the derivative of $E_{t-1}y_{t+r}$ with respect to the first element of α. One can first solve the model for a given value of α and then solve it again for the first element of α changed by a certain percent, both solutions using the EP method. The computed derivative is then the difference in the two solution values of $E_{t-1}y_{t+r}$ divided by the change in the first element of α. To compute

[28]Note that these solutions of the error term ϵ_{it} are only approximations when f_i is nonlinear. Hence, the method gives an approximation of the likelihood function.

all the derivatives requires $K+1$ solutions of the model over the T number of observations, where K is the dimension of α.[29] One solution is for the base values, and the K solutions are for the K changes in α, one coefficient change per solution. From these $K+1$ solutions, $K \cdot T(h+1)$ derivatives are computed and stored for each expectations variable, one derivative for each length ahead for each period for each coefficient.[30] Once these derivatives are computed, they can be used in the computation of Σ for a given change in α, and no further solutions of the model are needed. In other words, when the maximization algorithm changes α and wants the corresponding value of L, the derivatives are first used to compute the expectations, which are then used in the computation of Σ. Since one has (from the derivatives) an estimate of how the expectations change when α changes, one does not have to solve the model any more to get the expectations.

Assuming that the solution method in Case 3 above is used for the FIML estimates, derivatives of u_{it-1} with respect to the coefficients are also needed when the errors are serially correlated. These derivatives can also be computed from the $K+1$ solutions, and so no extra solutions are needed in the serial correlation case.

Once the $K+1$ solutions of the model have been done and the maximization algorithm has found what it considers to be the optimum, the model can be solved again for the T periods using the optimal coefficient values and then L computed. This value of L will in general differ from the value of L computed using the derivatives for the same coefficient values, since the derivatives are only approximations. At this point the new solution values (not computed using the derivatives) can be used as new base values and the problem turned over to the maximization algorithm again. This is the second "iteration" of the overall process. Once the maximization algorithm has found the new optimum, new base values can be computed, a new iteration performed, and so on. Convergence is achieved when the coefficient estimates from one iteration to the next are within a prescribed tolerance criterion of each other. This

[29] In the notation presented in Section 7.1 k rather than K is used to denote the dimension of α. K is used in this section since k has already been used in the description of the EP method.

[30] Derivatives computed this way are "one sided." "Two sided" derivatives would require an extra K solutions, where each coefficient would be both increased and decreased by the given percentage. For the work here two sided derivatives seemed unnecessary. For the results below each coefficient was increased by five percent from its base value when computing the derivatives. Five percent seemed to give slightly better results than one percent, although no systematic procedure of trying to find the optimal percentage size was undertaken.

procedure can be modified by recomputing the derivatives at the end of each iteration. This may improve convergence, but it obviously adds considerably to the expense. At a minimum, one might want to recompute the derivatives at the end of overall convergence and then do one more iteration. If the coefficients change substantially on this iteration, then overall convergence has not in fact been achieved.

Table 7.1 reports the results of estimating three models by FIML using the derivatives. The first model, Model 1, is a version of the wage contracting model in Taylor (1980):

$$y_{1t} = \alpha_{11} y_{1t-1} + \alpha_{12} y_{1t-2} + \alpha_{13} E_{t-1} y_{1t+1} + \alpha_{14} E_{t-1} y_{1t+2}$$

$$+ \alpha_{15} E_{t-1} y_{2t} + \alpha_{16} E_{t-1} y_{2t+1} + \alpha_{17} E_{t-1} y_{2t+2} + u_{1t} \tag{7.21}$$

$$y_{2t} = \alpha_{21} y_{1t} + \alpha_{22} y_{1t-1} + \alpha_{23} y_{1t-2} + u_{2t} \tag{7.22}$$

with the restrictions that $\alpha_{11} = \alpha_{13} = 1/3$, $\alpha_{12} = \alpha_{14} = 1/6$, $\alpha_{15} = \alpha_{16} = \alpha_{17}$, and $\alpha_{21} = \alpha_{22} = \alpha_{23}$. There are two free parameters to estimate, α_{15} and α_{21}. Data for this model were generated using normally distributed serially independent errors with zero correlation between equations. Values of α_{15} and α_{21} of .0333333 and $-.333333$ were used for this purpose. Fifty observations were generated.

Because this model is very small and linear, a factorization procedure can be used to evaluate L exactly. This procedure can in turn be used in the maximization of L using an algorithm like DFP. The coefficient estimates computed this way are $\hat{\alpha}_{15} = .0260125$ and $\hat{\alpha}_{21} = -.3916$.

Table 7.1 shows the results using the "derivative" method discussed above. The results for Model 1 show that convergence was essentially achieved after one iteration. Three solutions of the model over the 50 periods were needed for the derivatives for the first iteration, which compares to 61 that would have been needed had the derivatives not been used. The difference between L computed using the derivatives and L computed from the full solution after the first iteration is very small, and so the method worked quite well. The DFP algorithm was used for this problem since the model was not large enough to require the Parke algorithm. The two further iterations for Model 1, which were based on recomputing the derivatives, led to very small changes. The third iteration in particular was unnecessary.

For Model 2 the error term in equation 7.21 is assumed to be serially correlated:

$$u_{1t} = \rho_1 u_{1t-1} + \epsilon_{1t} \tag{7.23}$$

Table 7.1
FIML Results for Three Models

Model 1: Taylor Model, No Serial Correlation

	$\hat{\alpha}_{15}$	$\hat{\alpha}_{21}$	$\hat{L}$ using derivatives	$\hat{L}$ using full solution	No. of func. evals.
Start	.0333333	-.333333		508.6022	
Iteration:					
1	.0252994	-.391662	509.0470	509.0462	61
2	.0260233	-.391609	509.0467	509.0467	50
3	.0260117	-.391612	509.0467	509.0466	37

Model 2: Taylor Model, Serial Correlation

	$\hat{\alpha}_{15}$	$\hat{\rho}_1$	$\hat{\alpha}_{21}$	$\hat{L}$ using derivatives	$\hat{L}$ using full solution	No. of func. evals.
Start	.0200000	.600	-.200000		501.8234	
Iteration:						
1	.0335672	.635	-.210860	505.5016	531.1740	77
2	.0289718	.673	-.321878	532.0178	531.7876	166
3	.0495646	.745	-.321324	532.1676	531.8590	103
4	.0778620	.837	-.322183	532.3424	531.9918	103
5	.0886905	.878	-.322699	532.1248	531.9346	96
6	.0903430	.889	-.322646	531.9557	531.9032	90

Model 3: Six Equation Model, 25 Coefficients

	$\hat{L}$ using derivatives	$\hat{L}$ using full solution	No. of func. evals.
Start		170.3100	
Iteration:			
1	189.1670	184.3381	2817
2	189.2047	189.0098	1103
3	189.0450	189.0297	538
4	189.0784	189.0784	258

The DFP algorithm was used for Models 1 and 2.
The Parke algorithm was used for Model 3.
Derivatives were recomputed after each iteration for Models 1 and 2, but not for Model 3.

where ρ_1 was set equal to .7 to generate the data. The coefficient estimates using the factorization routine and the DFP algorithm are $\hat{\alpha}_{15}$ = .0738367, $\hat{\rho}_1$= .83545, and $\hat{\alpha}_{21} = -.32211$. This set of values will be called the “exact”

Table 7.2
Model 3: Six Equations

1. $\log C_t$	cnst, $\log C_{t-1}$, $E_{t-1} \log Y_{t+2}$, R_t
2. $I_t - I_{t-1}$	cnst, $Y_t - Y_{t-1}$, $E_{t-1}(Y_{t+1} - Y_t)$, R_t, t, I_{t-1}
3. $\log(M_t/P_t)$	cnst, $\log(M/P)_{t-1}$, $\log Y_t$, R_t
4. $\log P_t$	cnst, $\log P_{t-1}$, $\log PM_t$, $(YS_t - Y_t)/YS_t$, $E_{t-1}[(YS_{t+1} - Y_{t+1})/YS_{t+1}]$, RHO
5. R_t	cnst, R_{t-1}, $E_{t-1}100[(P_{t+2}/P_{t+1})^4 - 1]$, $100[(Y_t/Y_{t-1})^4 - 1]$, $100[(M_{t-1}/M_{t-2})^4 - 1]$
6. $Y_t =$	$C_t + I_t + Q_t$

answer. The results in Table 7.1 show that the derivative method got close, but not quite, to the exact answer. The largest value of L occurred after the fourth iteration, 531.9918, with coefficient estimates fairly close to the exact answer. On iterations 5 and 6, however, the method moved slightly further away from the answer. The derivatives were computed after each iteration for this problem. The value of L using the exact coefficient estimates (not reported in the table) was 532.0333. The method thus moved from L equal to 501.8234 to L equal to 531.9918, but it could not go the rest of the way to 532.0333. When the method was started off from the exact answer, it moved away from it slightly, like the case for iterations 5 and 6 in Table 1. This basically seems to be a hard computational problem. The likelihood function is fairly flat near the top, especially with respect to α_{15} and ρ_1, and one other local optimum was found in the course of this work.[31]

Model 3 is a simple six equation macroeconomic model with 25 coefficients, one of which is a serial correlation coefficient. The model is meant for computational exercises only; it is not meant to be a good approximation of the economy. The equations are shown in Table 7.2 (C is consumption, I is investment, M is the nominal money supply, P is the GNP deflator, R is the interest rate, Y is GNP, YS is an estimate of potential GNP, PM is the import price deflator, Q is government spending plus net exports, t is the time

[31] Also, although not reported in Table 7.1, Model 2 is much harder to solve than Model 1 in requiring a much larger value of k and many more iterations of the solution paths to converge.

trend, RHO means that the error term in the equation is first order serially correlated, and C, I, Y, YS, and Q are in real terms): The exogenous variables in the model are PM_t, YS_t, Q_t, and t. Future expected values are in equations 1, 2, 4, and 5, and the longest lead length is 2.

The equations were first estimated using Hansen's method discussed in Section 4.3. The estimation period was 1954:1–1984:4, for a total of 124 observations. The Hansen estimates were then used as starting values for the FIML calculations.[32]

The results in Table 7.1 for Model 3 are based on only one set of calculations of the derivatives. The model was solved 26 times for the 124 observations to get the derivatives for the 25 coefficients. The Parke algorithm was used for the maximization. It can be seen in Table 7.1 that the use of the derivatives worked quite well. After the first iteration the difference between L computed using the derivatives and L computed from the full model solution is fairly large ($189.1670 - 184.3381$), but the differences are quite small for iterations 2, 3, and 4. Convergence had been achieved after iteration 4.

The good results for Model 3 are encouraging. Model 3 is probably more representative of models likely to be used in practice than is Model 2. Model 2 is probably extreme in the degree to which future predicted values affect current predicted values, and this may be one of the reasons results are not as good for it.

The FIML covariance matrix of the coefficient estimates ($\hat{V}_4$) was estimated for each model using the formula 7.4, where the derivatives are evaluated (numerically) at the optimum. These covariance computations are feasible because the expectations derivatives can be used in calculating the derivatives in 7.4. In other words, no further solutions of the model are needed to compute $\hat{V}_4$ in 7.4. $\hat{V}_4$ for Model 3 is used for the stochastic simulation results discussed next.

Stochastic Simulation

For models with rational expectations one must state very carefully what is meant by a stochastic simulation of the model and what stochastic simulation is to be used for. In the present case stochastic simulation is *not* used to

[32] The results for Model 3 in Tables 7.1 and 7.3 are the same as those in Fair and Taylor (1990). They have not been updated for present purposes. Since Model 3 is not part of the US model and is not used for any of the work in the following chapters, there was no need to update. Also, the results for Models 1 and 2 in Table 7.1 are the same as those in Fair and Taylor (1990).

improve on the accuracy of the solutions of the expected values. The expected values are computed exactly as described above—using the EP method. This way of solving for the expected values can be interpreted as assuming that agents at the beginning of period s form their expectations of the endogenous variables for periods s and beyond by 1) forming expectations of the exogenous variables for periods s and beyond, 2) setting the error terms equal to their expected values (say zero) for periods s and beyond, 3) using the existing set of coefficient estimates of the model, and then 4) solving the model for periods s and beyond. These solution values are the agents' expectations.

For present purposes stochastic simulation begins once the expected values have been solved for. Given the expected values for periods s through $s + h$, stochastic simulation is performed for period s. The problem is now no different from the problem for a standard model because the expectations are predetermined. Assume that the errors are distributed $N(0, \hat{\Sigma})$, where $\hat{\Sigma}$ is the FIML estimate of Σ from the last subsection. From this distribution one can draw a vector of error terms for period s. Given these draws (and the expectations), the model can be solved for period s in the usual ways. This is one repetition. Another repetition can be done using a new draw of the vector of error terms, and so on. The means and variances of the forecast values can be computed using equations 7.5 and 7.7 in Section 7.3.

One can also use this approach to analyze the effects of uncertainty in the coefficients by assuming that the coefficients are distributed $N(\hat{\alpha}, \hat{V}_4)$, where $\hat{\alpha}$ is the FIML estimate of α and $\hat{V}_4$ is the estimated covariance matrix of $\hat{\alpha}$. In this case each draw also involves the vector of coefficients.

If u_{it} is serially correlated as in 7.19, then an estimate of u_{is-1} is needed for the solution for period s. This estimate is, however, available from the solution of the model to get the expectations (see Case 2 in the previous subsection), and so no further work is needed. The estimate of u_{is-1} is simply taken as predetermined for all the repetitions, and u_{is} is computed as $\rho_i u_{is-1}$ plus the draw for ϵ_{is}. (Note that the ϵ errors are drawn, not the u errors.)

Stochastic simulation is quite inexpensive if only results for period s are needed because the model only needs to be solved once using the EP method. Once the expectations are obtained, each repetition merely requires solving the model for period s. If, on the other hand, results for more than one period are needed and the simulation is dynamic, the EP method must be used p times for each repetition, where p is the length of the period.

Consider the multiperiod problem. As above, the expectations with viewpoint date $s - 1$ can be solved for and then a vector of error terms and a vector of coefficients drawn to compute the predicted value of y_{is}. This is the first

step.

Now go to period $s+1$. An agent's expectation of, say, y_{is+2} is different with viewpoint date s than with viewpoint date $s-1$. In particular, the value of y_{is} is in general different from what the agent at the end of period $s-1$ expected it to be (because of the error terms that were drawn for period s).[33] A new set of expectations must thus be computed with viewpoint date s. Agents are assumed to use the original set of coefficients (not the set that was drawn) and to set the values of the error terms for periods $s+1$ and beyond equal to zero. Then given the solution value of y_{is} and the actual value of x_s, agents are assumed to solve the model for their expectations for periods $s+1$ and beyond. This requires a second use of the EP method. Given these expectations, a vector of error terms for period $s+1$ is drawn and the model is solved for period $s+1$. If equation i has a serially correlated error, then u_{is+1} is equal to $\rho_i^2 u_{is-1}$ plus the draw for ϵ_{is+1}. Now go to period $s+2$ and repeat the process, where another use of the EP method is needed to compute the new expectations. The process is repeated through the end of the period of interest. At the end, this is one repetition. The overall process is then repeated for the second repetition, and so on. Note that only one coefficient draw is used per repetition, i.e., per dynamic simulation. After J repetitions one can compute means and variances just as above, where there are now means and variances for each period ahead of the prediction. Also note that agents are always assumed to use the original set of coefficients and to set the current and future error terms to zero. They do not perform stochastic simulation themselves.

Stochastic simulation results for Model 3 are presented in Table 7.3. The FIML estimates of Σ, α, and V_4 from the previous subsection were used for the draws. The length of the prediction was taken to be four, and 100 repetitions were performed. This meant that the number of times the model had to be solved for the expectations was 400. Again, had the length been taken to be one, the number of solutions for the expectations would have been one. The results show, as is common with most macroeconometric models, that the stochastic simulation estimates of the means are quite close to the deterministic simulation estimates. The deterministic simulation estimates are simply based on setting the error terms to zero and solving once for each period (as the agents are assumed to do). The real use of stochastic simulation is to compute standard deviations or variances. The estimated standard deviations are presented in

[33] It may also be that the actual value of x_s differs from what the agent expected it to be at the end of $s-1$.

Table 7.3
Stochastic Simulation Results for Model 3

		1983			
		1	2	3	4
Consumption	a	2095.4	2111.9	2129.6	2146.5
	b	2094.0	2113.0	2130.8	2149.0
	c	13.1	17.9	23.1	29.4
Investment	a	259.3	264.2	268.2	272.5
	b	259.1	264.1	269.1	274.4
	c	6.7	8.7	9.8	12.3
Money Supply	a	521.5	532.2	543.2	554.5
	b	521.1	533.1	543.8	556.0
	c	5.5	8.4	10.9	11.7
Price Level	a	1.0293	1.0435	1.0587	1.0751
	b	1.0293	1.0437	1.0595	1.0762
	c	.0046	.0083	.0110	.0125
Interest Rate	a	8.39	8.57	8.75	8.94
	b	8.28	8.40	8.74	9.01
	c	.79	.96	1.09	1.21
Real GNP	a	3201.2	3243.9	3273.1	3305.4
	b	3199.6	3244.8	3275.2	3309.8
	c	17.6	23.3	28.1	35.7

a = predicted value from deterministic simulation
b = mean value from stochastic simulation
c = standard deviation from stochastic simulation

The results are based on 100 trials.
Units are billions of 1982 dollars for consumption, investment, and real GNP; billions of dollars for the money supply; 1982=1.0 for the price level; and percentage points for the interest rate.

row c in the table. For real GNP, for example, the estimated standard deviation of the four quarter ahead forecast error is $35.7 billion, which is about one percent of the mean value of $3309.8 billion.

Stochastic simulation has also been used to evaluate alternative international monetary systems using the multicountry models in Carloyzi and Taylor (1985) and Taylor (1988). For this work values of ϵ_{it} were drawn, but not values of the coefficients. The vector of coefficients α was taken to be fixed.

It seems that stochastic simulation as defined above is computationally feasible for models with rational expectations. Stochastic simulation is in fact likely to be cheaper than even FIML estimation using the derivatives.

If, for example, the FIML estimation period is 100 observations and there are 25 coefficients to estimate, FIML estimation requires that the model be solved 2600 times using the EP method to get the derivatives. For a stochastic simulation of 8 periods and 100 repetitions, on the other hand, the model has to be solved using the EP method only 800 times.

Conclusion

To conclude, the results in this section are encouraging regarding the use of models with rational expectations. FIML estimation is computationally feasible using the procedure of computing derivatives for the expectations, and stochastic simulation is feasible when done in the manner described above. FIML estimation is particularly important because it takes into account all the nonlinear restrictions implied by the rational expectations hypothesis. It is hoped that the methods discussed in this section will open the way for many more tests of models with rational expectations.