Chapter Three

Econometric Issues

3.1 INTRODUCTION

Most of the econometric issues that pertain to this study are discussed in this chapter. The three main issues that are discussed are the treatment of serial correlation problems, the computation of the two stage least squares (TSLS) estimates, and the computation of the full information maximum likelihood (FIML) estimates. The model is nonlinear in both variables and parameters, and so one cannot rely directly on the standard textbook procedures for estimating linear models in computing the TSLS and FIML estimates of the model.

The following notation will be used for the discussion in this chapter. Let G denote the total number of equations in the model, M the number of stochastic equations, N the total number of predetermined (exogenous and lagged endogenous) variables, and T the number of observations. Write the gth equation of the model as:

$$\phi_g(y_{1t}, \ldots, y_{Gt}, x_{1t}, \ldots, x_{Nt}, \beta_g) = u_{gt}, \, (g = 1, \ldots, G), \, (t = 1, \ldots, T), \qquad (3.1)$$

where the y_{it} are the endogenous variables, the x_{it} are the predetermined variables, β_g is the vector of unknown coefficients in equation g, and u_{gt} is the error term corresponding to equation g. For identities, u_{gt} is zero for all t. It will be assumed without loss of generality that the stochastic equations occur first in the model. The first M equations in the model are thus stochastic, with the remaining $G - M$ equations being identities. For the model as presented in the last chapter, M is 26 and G is 83. The basic period of estimation is 1954I–1974II, which gives a value of T of 82.

Counting the strike dummies, there are 78 exogenous variables in the model plus the constant term. There are also a number of lagged

endogenous and lagged exogenous variables that appear as explanatory variables in the stochastic equations and in the identities. The value of N for the model is thus some number greater than 78. The error terms in some of the equations show evidence of first order serial correlation, and, as mentioned in section 1.1, the serial correlation assumption was retained for 12 of the 26 stochastic equations. There are 166 unknown coefficients to estimate in the 26 stochastic equations, counting the serial correlation coefficients, but not counting the variances and covariances of the error terms.

It will be useful in the following discussion to consider a particular example of one of the equations in (3.1). Assume that the first equation is:

$$\log \frac{y_{1t}}{x_{1t}} = \beta_{11} + \beta_{12} \log \frac{y_{1t-1}}{x_{1t-1}} + \beta_{13} \log y_{2t} + \beta_{14} \log y_{3t} + \beta_{15} x_{2t} + u_{1t},$$

$$(3.2)$$

where

$$u_{1t} = \rho_{11} u_{1t-1} + \varepsilon_{1t}, \qquad (t = 1, \ldots, T). \tag{3.3}$$

The functional form of Equation (3.2) is common to a number of the stochastic equations in the model. Equation (3.2) is nonlinear in variables, but linear in the unknown coefficients. The first order serial correlation assumption in (3.3) is, as just mentioned, common to 12 of the stochastic equations. The error term ε_{1t} in Equation (3.3) is assumed not to be serially correlated.

3.2   THE TREATMENT OF SERIAL CORRELATION PROBLEMS

A convenient way of handling an equation with a first order serially correlated error term is to convert the equation into one that is nonlinear in coefficients, but that has a serially uncorrelated error term. Lagging Equation (3.2) once, multiplying through by ρ_{11}, and subtracting the resulting expression from Equation (3.2) yields, after some rearranging:

$$\log \frac{y_{1t}}{x_{1t}} = \rho_{11} \log \frac{y_{1t-1}}{x_{1t-1}} + \beta_{11} - \rho_{11}\beta_{11} + \beta_{12} \log \frac{y_{1t-1}}{x_{1t-1}} - \rho_{11}\beta_{12} \log \frac{y_{1t-2}}{x_{1t-2}}$$

$$+ \beta_{13} \log y_{2t} - \rho_{11}\beta_{13} \log y_{2t-1} + \beta_{14} \log y_{3t} - \rho_{11}\beta_{14} \log y_{3t-1}$$

$$+ \beta_{15} x_{2t} - \rho_{11}\beta_{15} x_{2t-1} + \varepsilon_{1t}. \tag{3.4}$$

Considering ρ_{11} to be just another coefficient to estimate, Equation (3.4) differs from Equation (3.2) by the inclusion of more explanatory variables

and by the inclusion of nonlinear restrictions on the coefficients of these variables. The error term in the equation is, however, not serially correlated. The nonlinear restrictions on the coefficients result from the treatment of ρ_{11} as an unknown coefficient.

The treatment of the serial correlation problem in this way means that the u_{gt} error terms in (3.1) can be considered to be serially uncorrelated, where any initial serial correlation of the error terms has been solved out in the manner just described. The interpretation of (3.1) in this way means that the β_g vector should be considered as including the serial correlation coefficient when serial correlation is present in the gth equation. When serial correlation is present in an equation, the number of predetermined variables in the equation should also be considered to be larger than it otherwise would be, and the equation should be considered to be nonlinear in coefficients as well as, possibly, in variables.

If observations on the endogenous and predetermined variables are available for $t = 0, 1, \ldots, T$, then Equation (3.4) must be estimated for $t = 1, \ldots, T$. There are ways of using information on the first observation more efficiently than the approach just described allows, but this added complication was not considered here. Ignoring the extra information on the first observation has no detrimental effect on the large sample properties of the estimators.

The present treatment has also not considered the case where an error term in one equation is directly correlated with the lagged value of an error term in some other equation. This complication would introduce nonlinear restrictions on the coefficients across, as well as within, equations. Since no experimentation with cross-serial correlation effects was carried out in this study, this added complication will not be considered in this chapter. For the linear model case, see Chow and Fair [9] and Fair [17] for a treatment of cross-serial correlation, as well as serial correlation of higher than first order.

3.3 THE COMPUTATION OF THE TWO STAGE LEAST SQUARES ESTIMATES

Since the model is nonlinear, explicit expressions for the reduced form equations cannot be derived. Consequently, consistent estimates of the reduced form coefficients cannot be obtained from any type of first stage regressions. Fortunately, the two stage least squares (TSLS) procedure does not require that consistent estimates of the reduced form coefficients be obtained in order to obtain consistent estimates of the structural coefficients in the second stage.

Consider, for example, the estimation of Equation (3.4) by TSLS. The endogenous terms on the right-hand side of the equation are

$\log y_{2t}$ and $\log y_{3t}$. If ε_{1t} is assumed not to be correlated with any variables on the right-hand side of the equation except $\log y_{2t}$ and $\log y_{3t}$, then consistent estimates of the coefficients of the equation can be obtained by the following two stage procedure. In the first stage, regress $\log y_{2t}$ and $\log y_{3t}$ on a common set of variables. The variables in this set should be variables that one feels, from knowledge of the overall model, have an effect, either directly or indirectly, on $\log y_{2t}$ and $\log y_{3t}$ and are not correlated with ε_{1t}. In other words, these variables should be correlated with $\log y_{2t}$ and $\log y_{3t}$, but not with ε_{1t}. The variables in this set must include the predetermined variables that appear on the right-hand side of Equation (3.4) in the form in which they appear in the equation: the constant, $\log \dfrac{y_{1t-1}}{x_{1t-1}}$, $\log \dfrac{y_{1t-2}}{x_{1t-2}}$, $\log y_{2t-1}$, x_{2t}, and x_{2t-1}. Let $\widehat{\log y_{2t}}$ and $\widehat{\log y_{3t}}$ denote the predicted values of $\log y_{2t}$ and $\log y_{3t}$ from the two regressions, and let $\hat{v}_{2t}$ and $\hat{v}_{3t}$ denote the estimated residuals from the two regressions. By definition, $\hat{v}_{2t} = \log y_{2t} - \widehat{\log y_{2t}}$ and $\hat{v}_{3t} = \log y_{3t} - \widehat{\log y_{3t}}$.

Now replace $\log y_{2t}$ and $\log y_{3t}$ in Equation (3.4) with their predicted values:

$$\log \frac{y_{1t}}{x_{1t}} = \rho_{11} \log \frac{y_{1t-1}}{x_{1t-1}} + \beta_{11} - \rho_{11}\beta_{11} + \beta_{12} \log \frac{y_{1t-1}}{x_{1t-1}} - \rho_{11}\beta_{12} \log \frac{y_{1t-2}}{x_{1t-2}}$$

$$+ \beta_{13} \widehat{\log y_{2t}} - \rho_{11}\beta_{13} \log y_{2t-1} + \beta_{14} \widehat{\log y_{3t}} - \rho_{11}\beta_{14} \log y_{3t-1}$$

$$+ \beta_{15} x_{2t} - \rho_{11}\beta_{15} x_{2t-1} + (\varepsilon_{1t} + \beta_{13} \hat{v}_{2t} + \beta_{14} \hat{v}_{3t}). \qquad (3.5)$$

By one of the properties of least squares, all of the variables on the right-hand side of this equation are orthogonal within the sample period to $\hat{v}_{2t}$ and $\hat{v}_{3t}$. This is because a common set of regressors has been used for both first stage regressions and because this set includes all of the predetermined variables on the right-hand side of Equation (3.4) in the form in which they appear in the equation. ε_{1t} is uncorrelated with all of the right-hand side variables in Equation (3.5). It is uncorrelated with the two predicted value variables because these variables are merely linear combinations of variables that are uncorrelated with ε_{1t} by assumption. ε_{1t} is uncorrelated with all of the other variables in the equation by assumption. Consequently, the composite error term in parentheses in Equation (3.5) is uncorrelated with all of the right-hand side variables, and so consistent estimates of this equation can be obtained by minimizing the sum of squared residuals with respect to the six coefficients: β_{11}, β_{12}, β_{13}, β_{14}, β_{15}, and ρ_{11}.

Minimizing the sum of squared residuals in Equation (3.5) is a nonlinear minimization problem because of the presence of ρ_{11}. This problem

is not, however, very difficult to solve. One procedure that can be used is the iterative procedure outlined in Fair [22] (p. 509, fn. 3), which is merely the Cochrane-Orcutt [10] procedure adjusted to account for simultaneous equations bias. Since this minimization problem is not very difficult, other procedures could clearly be used. The question of which procedure one uses to minimize the sum of squared residuals in Equation (3.5) is a numerical question, not a statistical one.

The above analysis is also not limited to the particular kind of nonlinearity present in Equation (3.5). One could, for example, have a restriction that says that $\beta_{12} = \beta_{13}\beta_{14}$ and carry out the minimization incorporating this restriction as well. All this would do would be to change possibly the numerical procedure used to carry out the minimization. The Cochrane-Orcutt procedure and its various generalizations, for example, are more or less restricted to nonlinearities caused by the presence of serial correlation of the error terms.

In a very elegant paper, Amemiya [2] discusses the nonlinear two stage least squares estimator. He proves, for the case in which the equation being estimated is only nonlinear in coefficients, that the nonlinear two stage least squares estimator has the same asymptotic distribution as the limited information maximum likelihood estimator, providing that one uses all the predetermined variables in the model as regressors in the first stage regressions. (Amemiya considers only the case in which the predetermined variables are fixed.) For the nonlinear-in-variables case, no such theorem exists. The efficiency of the two stage least squares estimator in this case depends on how closely one has approximated the (unknown) reduced form equations in the first stage regressions.

The TSLS estimates of the model are presented in Table 2-3. The only nonlinearity in coefficients that existed in any of the equations was due to the presence of the serial correlation coefficient, and so the iterative procedure described in [22] was used to minimize the sum of squared errors when nonlinearity existed. A different set of first stage regressors was used for each equation estimated, depending on the predetermined variables and the right-hand side endogenous variables included in the equation. The regressors that were chosen for each equation were, in addition to the ones that were necessary to meet the orthogonality requirement discussed above, ones that seemed likely to have important effects on the included right-hand side endogenous variables.

The "t-statistics" that are presented in Table 2-3 are the absolute values of the ratios of the coefficient estimates to the estimates of their asymptotic standard errors. The estimates of the asymptotic standard errors for those equations that were linear in coefficients (no serial correlation) were computed in the usual way for the two stage least squares estimator. The estimates were computed as the square roots of the diagonal elements of

$\hat{\sigma}^2(\hat{Z}'\hat{Z})^{-1}$, where $\hat{\sigma}^2$ is the TSLS estimate of the variance of the error term in the equation being estimated and $\hat{Z}$ is the matrix of observations on the variables used in the second stage regression. A ^ is placed on Z to denote the fact that some of the variables in $\hat{Z}$ are variables of predicted values. $\hat{\sigma}^2$ is the estimate of the variance of the actual error term in the equation, not of the variance of the composite error term that is minimized in the second stage regression.

For those equations that were nonlinear in coefficients because of the serial correlation assumption, the estimates of the asymptotic standard errors (including the estimates of the asymptotic standard errors of the estimates of the serial correlation coefficients) were computed in a manner analogous to that described in [22], p. 514, for the linear model case. Consider, for example, Equation (3.2). Let Z denote the matrix of observations on the right-hand side variables in this equation. Let $\hat{Z}$ denote the matrix that is obtained from Z by replacing $\log y_{2t}$ and $\log y_{3t}$ in Z with $\widehat{\log y_{2t}}$ and $\widehat{\log y_{3t}}$ $(t = 1, \ldots, T)$, the latter two series being obtained in the manner described above.

Define $\hat{Q}$ to be equal to $\hat{Z} - \hat{\rho}_{11}Z_{-1}$, where $\hat{\rho}_{11}$ is the TSLS estimate of ρ_{11} and Z_{-1} is the matrix Z lagged one period. (It is assumed that observations for $t = 0$ are available.) Then the estimates of the asymptotic standard errors of the coefficient estimates other than $\hat{\rho}_{11}$ were computed as the square roots of the diagonal elements of $\hat{\sigma}^2(\hat{Q}'\hat{Q})^{-1}$, where $\hat{\sigma}^2$ is the estimate of the variance of ε_{1t}, the nonserially correlated error term. The estimate of the asymptotic standard error of $\hat{\rho}_{11}$ was computed, as described in [22], as the square root of $(1 - \hat{\rho}_{11}^2)/T$.

The t-statistics and Durbin-Watson statistics presented in Table 2–3 are meant to be interpreted more as just summary measures of the regressions than as precise statistical tests of some hypothesis. Too many assumptions of classical statistical hypothesis testing have been violated in the process of arriving at the estimates in Table 2–3 for any rigorous interpretation of the statistics as test statistics to be warranted. The primary way that the model has been tested in this study is to compare, in the manner described in Chapter Eight, its prediction accuracy with the prediction accuracy of other models.

3.4   THE COMPUTATION OF THE FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATES

In by now a classic paper, Chow [7] provides an interpretation of the full information maximum likelihood (FIML) estimator of a linear simultaneous equations model as a natural generalization of least squares. The FIML

estimates are ones that minimize the generalized variance of the error terms in a model, subject to the restriction that the generalized variance of certain linear combinations of the endogenous variables be equal to a constant. The linear combination aspect of this procedure is the reason why the FIML estimator does not require, as do two and three stage least squares, that there be one natural left-hand side variable per equation.

In the present model there is a natural left-hand side variable for every equation except one, Equation 70 in Table 2–2. Equation 70 is, however, one of the key equations in the model, it being the equation that allows the bill rate to be implicitly determined. Therefore, because of Equation 70 and the implicit determination of the bill rate, the FIML estimator appears to be the natural one to use to estimate the model.

Under the assumption that the error terms for the stochastic equations in (3.1) are jointly normally distributed, the FIML estimates of the unknown coefficients in the model are obtained by maximizing:

$$L = -\tfrac{1}{2} T \log |S| + \sum_{t=1}^{T} \log |J_t| \qquad\qquad (3.6)$$

with respect to the unknown coefficients,[a] where

$$S = (s_{gh}),\ s_{gh} = \frac{1}{T} \sum_{t=1}^{T} u_{gt} u_{ht}, \qquad (g, h = 1, \ldots, M), \qquad (3.7)$$

$$J_t = \left(\frac{\partial \phi_g}{\partial y_{ht}} \right), \qquad (g, h = 1, \ldots, G). \qquad (3.8)$$

The matrix S is $M \times M$, and the Jacobian matrix J_t is $G \times G$.

The maximization of L in (3.6) is a computationally difficult problem for a model of even moderate size because of the presence of the Jacobian terms. For every evaluation of L, $T + 1$ determinants have to be computed. T of these determinants are for the J_t matrices, which are generally of much higher dimension than the dimension of S. Since, as just discussed, it seems important to obtain FIML estimates of the model, a considerable effort was put into this study in trying to do so.

It did turn out to be feasible to obtain a set of estimates of the model that may be close to the true set of FIML estimates. This set was obtained as follows. First, 78 of the 166 unknown coefficients were fixed at their TSLS estimates, leaving 88 coefficients to estimate. An attempt was made to choose for the coefficients to estimate by FIML those that appeared to be most important in the model. The coefficients of the strike dummy variables, for example, were never chosen to be estimated. Second, some of

the identities in the model were substituted out, decreasing the dimension of J_t in (3.8) to 48 × 48.

Third, J_t is a very sparse matrix, and advantage was taken of this fact in computing its determinant. Although J_t was 48 × 48, there were only 200 nonzero elements in it. There is a considerable literature, apparently largely unknown to economists, on dealing with sparse matrices,[b] and it turned out in the present case that considerable computational time could be saved by taking advantage of the fact that J_t is sparse. A good set of routines for dealing with sparse matrices is available from IBM [29], and when these routines were combined in the appropriate way to take the determinant of J_t, the computational time needed to take the determinant was decreased by a factor of 28 over the time that would otherwise be required. This is an enormous saving, and were it not for this saving, it would clearly not have been feasible to obtain the set of estimates that was in fact obtained.

Fourth, it turned out that a fairly good approximation to $\sum_{t=1}^{T} \log |J_t|$ is $\frac{T}{2}(\log |J_1| + \log |J_T|)$. When $\log |J_t|$ is plotted against t ($t = 1, \ldots, T$), the points come fairly close to lying on a straight line, so that the average of the first and last points multiplied by T/2 is a fairly close approximation to the sum of the T points. This approximation was used for the work here, which meant that the determinant of J_t only had to be computed twice per evaluation of L rather than 82 times. To give an example of the error introduced by the approximation, the sum of all 82 points using the TSLS estimates was −8056.3, whereas the average of the first and last points multiplied by 41 was −8105.5. This is an error of about 0.6 percent.

The above procedures decreased the computer time needed for one evaluation of L to about 0.4 of a second on the IBM 370-158 computer at Yale. (The 370-158 is not a particularly fast computer for this purpose relative to a number of other computers in existence.) The fifth and final step in the calculation of the estimates was to maximize L using algorithms for maximizing nonlinear functions of coefficients that do not require analytic derivatives. The two algorithms that were considered are the no-derivative algorithm of Powell [37], and a member of the class of gradient algorithms considered by Huang [28]. The gradient algorithm requires first derivatives, and for present purposes the derivatives were obtained numerically. The gradient algorithm that was used is the one that updates the approximation to the inverse of the matrix of second partial derivatives by means of the "rank one correction formula." These two algorithms were used successfully by the author in two other studies, one concerned with solving optimal control problems for econometric models [20] and one concerned with obtaining FIML and robust estimates of econometric models [19]. For the

optimal control work, a maximization problem in which there were 239 unknown coefficients to determine was solved using the gradient algorithm.

The TSLS estimates were used as starting points for both algorithms. From the results of some early experimentation, the no-derivative algorithm appeared to be more adept at increasing the value of the likelihood function, and so it was the one used in the final stages of the work. Even the use of the no-derivative algorithm did not, however, result in much of an increase in the value of the likelihood function from the value corresponding to the TSLS estimates. The value of the likelihood function for the TSLS estimates is 907.2. The value of the likelihood function for the "FIML" estimates presented in Table 2–3 is 924.6, which is a gain of only 1.9 percent.

It took the algorithm 24 iterations to achieve this value. The 24 iterations corresponded to 24,449 function evaluations (about 1,000 function evaluations per iteration), which at 0.4 seconds per evaluation took about two hours and 43 minutes of computer time. The value of the likelihood function was only changing in the fourth digit (the first decimal point) at the point that the algorithm was stopped (from having exhausted the computer budget for this project). The coefficient estimates were also changing by only small amounts.

It can be seen in Table 2–3 that the FIML estimates are in most cases quite close to the TSLS estimates. (Generally, only three significant digits are presented in Table 2–3, and in a number of cases the FIML and TSLS estimates are the same to three digits. Almost all the estimates, however, differed in at least the fifth digit.) This can mean either that the TSLS estimates are in fact quite close to the true FLML estimates, or that the algorithm did a poor job in maximizing the likelihood function. Cost considerations prevented any further experimentation to see if the true optimum had in fact been reached. Given the small increase in the value of the likelihood function that occurred, it is clear that more work needs to be done before one can have much confidence that the "FIML" estimates that have been obtained in this study are close to the true FIML estimates.

One final point about the computation of the FIML estimates should be noted. Constraining 78 of the coefficients to be equal to their TSLS estimates resulted in 13 of the 26 stochastic equations not having any coefficients left to be estimated by FIML. These 13 equations were *not*, however, dropped from the model when computing the FIML estimates. The predicted error terms for these equations (based on the TSLS estimates) were used, for example, in the computation of $|S|$ in (3.6). The Jacobian J_t was also not changed. This procedure allows the correlation between the error terms in the 13 unestimated equations and the error terms in the 13 estimated equations to have an effect on the coefficient estimates of the 13 estimated equations.

3.5   THE SOLUTION OF THE MODEL

The model is solved by the use of the Gauss-Seidel technique. For the work in Chapter Eight and for most of the work in Chapter Nine, Equation 8 in Table 2-2, the equation explaining the value of demand deposits and currency of the household sector (DDH_t), was used to solve for the bill rate. Given values of the predetermined variables and given values of $SAVH_t$, CG_t, DDF_t, $BORR_t$, $SAVB_t$, LF_t, and $SECR_t$, Equations 45, 61, 62, 64, and 70 in Table 2-2 form a set of five equations in five unknowns that can be solved analytically. The five unknowns are: BR_t, A_t, DDB_t, $LBVBB_t$, and DDH_t.

These analytic solutions were obtained, and the five equations that resulted from these solutions were used as the equations explaining the five variables. This procedure means that Equation 70 is used in the solution of DDH_t, the DDH_t equation having been "used up" in determining the bill rate. Each of the remaining 76 equations in Table 2-2 was used to solve the variable that appears naturally on the left-hand side of the equation. Equations 77 and 80, which explain the hours and loan constraint variables, were modified slightly in the process of solving the model. These modifications are discussed in Chapters Four and Five.

There are other ways that the model could be solved, but this way was one of the most natural and proved to be quite satisfactory. The number of iterations needed to solve the model each quarter was generally between about 5 and 20, depending on the starting values used. The speed of convergence seemed to be maximized by damping the solution value of the bill rate by about 90 per cent on each iteration. In other words, if $\widehat{RBILL}_t^{(i)}$ denotes the solution value of $RBILL_t$ on the ith iteration and $\widetilde{RBILL}_t^{(i+1)}$ denotes the solution value of $RBILL_t$ that results from solving Equation 8 for $RBILL_t$ on the $(i+1)^{\text{st}}$ iteration, then the value of $\widehat{RBILL}_t^{(i+1)}$ was taken to be $\widehat{RBILL}_t^{(i)} + 0.1(\widetilde{RBILL}_t^{(i+1)} - \widehat{RBILL}_t^{(i)})$. Otherwise, no other damping was used in the solution of the model. One solution of the model for 82 quarters took about ten seconds of computer time on the IBM 370-158 computer at Yale.

The model was solved by setting all the structural error terms equal to zero (their expected value). It is well known that this procedure is incorrect for nonlinear models in the sense that it is not equivalent to setting the reduced form error terms equal to their expected values and then solving the reduced form equations. (See, for example, Howrey and Kelejian [27].) The proper way to solve the model would be by means of stochastic simulation, but this procedure is too costly to use in this study. Consequently, the

usual procedure for solving nonlinear models was followed, even though it is not quite right.

For some of the solutions in Chapter Nine, and for all the solutions in Chapter Ten, the bill rate was taken to be exogenous and VBG_t was added as the extra endogenous variable. In this case Equation 70 could be used to solve for VBG_t directly, and each of the other equations could be used to solve for the variable that appears naturally on the left-hand side. Convergence turned out to be somewhat faster in this case than in the endogenous bill rate case.

3.6 A POSSIBLE ESTIMATOR FOR FUTURE USE

The purpose of this section is to describe an estimator that may be of interest to consider in future work. The estimator is not computationally feasible on the IBM 370-158 computer, but it should be feasible on computers about ten to twenty times faster than the 370-158.

To motivate this estimator, consider first the estimation of a linear simultaneous equations model by FIML. Let V denote a $T \times M$ matrix of reduced form error terms, where M is the number of stochastic equations, and let $\hat{V}$ denote a $T \times M$ matrix of predicted reduced form error terms. Given values of the structural coefficients, one can obtain predictions of the reduced form error terms by "simulating" the model over the sample period. This simulation should be thought of for now as being a static simulation. In the linear model case, simulation does not require the use of any iterative procedure to solve the model each period because the reduced form coefficient matrix can be obtained directly from the structural coefficient matrices.

Consider now minimizing $|V'V|$ with respect to the structural coefficients. Since $|\hat{V}'\hat{V}|$ can be computed given a set of values of the structural coefficients, one of the algorithms discussed in section 3.4 could be used to carry out this minimization. If one were successful in this task, the values of the structural coefficients that minimized $|V'V|$ would be the FIML estimates. (See, for example, Malinvaud [33], Ch. 19, p. 677.) The FIML estimates are thus estimates that minimize the generalized variance of the reduced form error terms with respect to the structural coefficients.

The minimization procedure just described could be carried out for a nonlinear model as well, where "simulation" would now require the use of something like the Gauss-Seidel procedure to solve the model each period. The predicted error terms that make up $\hat{V}$ would be the differences between the simulated and actual values of the endogenous variables. The values of the structural coefficients that corresponded to the minimum of $|\hat{V}'\hat{V}|$ would not be FIML estimates in this case because the true reduced form error terms are not additive in nonlinear models. There is, however,

at least some analogy between these estimates and the true FIML estimates.

The above minimization procedure can be carried out, in either the linear or nonlinear case, using dynamic simulation rather than static simulation, a dynamic simulation being defined as a simulation that uses generated values of the lagged endogenous variables rather than actual values. The dynamic simulation can either be over the entire sample period or just for a few periods ahead at a time. Let $\hat{V}$ denote the $T \times M$ matrix of predicted error terms obtained from dynamic simulation of the model. The values of the structural coefficients that correspond to the minimum of $|\hat{V}'\hat{V}|$ will be called full information dynamic (FDYN) estimates.

The suggestion here is that it may be of interest in future work to obtain FDYN estimates of the model. It is true, of course, that for a properly specified model the FIML estimates are asymptotically efficient, so that if one knew that the model was properly specified, there would be no reason to be concerned with obtaining FDYN estimates. It is almost never the case, however, that one has complete confidence in the specification of a model, especially regarding the specification of the lag structures. The reason for proposing the FDYN estimator here is the feeling that the estimator may—by taking into account in a somewhat more explicit way than does the FIML estimator the dynamic properties of a model—lessen the effects of misspecification. Whether this is true or not is, of course, unclear, but at least it does seem worthy of some experimentation.

As mentioned in section 3.5, the time taken to solve the model once for 82 quarters is about ten seconds on the IBM 370-158. The time taken does not vary much depending on whether the simulation is static or dynamic. The time that would be required to compute $|\hat{V}'\hat{V}|$ once the model is solved for the 82 quarters is less than one second. Consequently, if the algorithms discussed in section 3.4 were used to minimize $|\hat{V}'\hat{V}|$, the time taken per function evaluation would be about ten seconds. This compares to the time of 0.4 seconds for the evaluation of the likelihood function in (3.6) in the computation of the FIML estimates. The FDYN estimates are thus about 25 times more expensive to compute than the FIML estimates, which means that the problem is really not feasible on the IBM 370-158. It should, however, be feasible to compute the estimates on a computer about ten to twenty times faster.

Klein [30] has suggested that it might be useful to estimate dynamic models by minimizing some function of multiperiod prediction errors. He is not very explicit on what function should be used, although for the linear model case he does suggest in one place (p. 64) that one might use the sum of the variances of the predicted error terms, each variance being normalized by the variance of the endogenous variable to which the error term corresponds. A more natural function to use, however, for both linear and nonlinear models, would appear to be the function $|\hat{V}'\hat{V}|$ suggested

above. This function can be interpreted as a generalized variance of the predicted error terms, and it corresponds most naturally to the function that FIML minimizes in the static case.

Another reason for suggesting that some experimentation with the FDYN estimator be done is that fairly good results were obtained in Fair [12] using a single equation DYN estimator. The results in [12], while clearly tentative, do indicate that some gain in prediction accuracy may be attained by the use of DYN estimators. The results in [12] are all within-sample results. If in the future FDYN estimates are obtained, they will clearly have to be judged on grounds of outside-sample prediction accuracy, or at least on some criteria other than within-sample prediction accuracy, since the FDYN estimates are explicitly designed to minimize a generalized variance of within-sample prediction errors.

Another class of estimators that may be worth considering in future work is the class of robust estimators. As discussed in Fair [19], almost any estimator that is based on minimizing some function of the error terms in an equation or a model can be modified to be a robust estimator. Again, some encouraging results were obtained in [19] about the possibility of being able to increase prediction accuracy by the use of robust estimators. These results are also very tentative, but they do at least indicate that further experimentation with robust estimators of econometric models should be undertaken. Primarily because of cost considerations, robust estimators were not considered in this study.

NOTES

[a]See, for example, Chow [8].
[b]See Brayton, Gustavson, and Willoughby [5] for a fairly extensive bibliography on sparse matrices.