

# FULL INFORMATION ESTIMATION AND STOCHASTIC SIMULATION OF MODELS WITH RATIONAL EXPECTATIONS

RAY C. FAIR

*Cowles Foundation, Yale University, New Haven, CT 06520, USA*

AND

JOHN B. TAYLOR

*Council of Economic Advisors, Executive Office of the President, Washington, D.C. 20500, USA*

## SUMMARY

A computationally feasible method for the full information maximum-likelihood estimation of models with rational expectations is described in this paper. The stochastic simulation of such models is also described. The methods discussed in this paper should open the way for many more tests of the rational expectations hypothesis within macroeconomic models.

## 1. INTRODUCTION

In an earlier paper (Fair and Taylor, 1983) we presented methods for the solution and full information estimation of models with rational expectations. The basic solution method, called the 'extended path' method, has come to be widely used for deterministic simulations of rational expectations models<sup>1</sup> but, probably because of the expense, the full information estimation method has not to our knowledge been tried by others. We discussed in this earlier paper a 'less expensive' method for obtaining full information estimates, but our preliminary results using the method were mixed. We have since experimented more with the less expensive method, and it seems much more promising than we originally thought.

This paper has two objectives. First, we examine the results that we have obtained using the less expensive method, and argue that full information estimation now seems feasible. In the process of doing this we correct some errors in our earlier paper regarding the treatment of models with rational expectations and autoregressive errors. Second, we examine methods for *stochastic* simulation of rational expectations models, something we only briefly touched on in the earlier paper.

---

<sup>1</sup>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.

## 2. THE SOLUTION METHOD

The model we consider is

$$f_i(y_t, y_{t-1}, \dots, y_{t-p}, {}_{t-1}E y_t, {}_{t-1}E y_{t+1}, \dots, {}_{t-1}E y_{t+h}, x_t, \alpha_i) = u_{it}, \quad (1)$$

$$u_{it} = \rho_i u_{it-1} + \varepsilon_{it}, \quad (i = 1, \dots, n), \quad (2)$$

where  $y_t$  is an  $n$ -dimensional vector of endogenous variables,  $x_t$  is a vector of exogenous variables,  ${}_{t-1}E$  is the conditional expectations operator based on the model and on information through period  $t-1$ ,  $\alpha_i$  is a vector of parameters,  $\rho_i$  is the serial correlation coefficient for the error term  $u_{it}$ , and  $\varepsilon_{it}$  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). 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  $\alpha_i$  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 we will always take to be zero). If the expectations  ${}_{s-1}E y_s, {}_{s-1}E y_{s+1}, \dots, {}_{s-1}E y_{s+h}$  were known, (1) could be solved in the usual ways (usually by the Gauss-Seidel algorithm). The model would be simultaneous, but future predicted values would not affect current predicted values. The extended path (EP) method iterates over solution *paths*. Values of the expectations through period  $s+h+k$  are first guessed, where  $k$  is a fairly large number relative to  $h$ .<sup>2</sup> 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 on one iteration are within a prescribed tolerance criterion of the solution values on 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 non-trivial 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)

<sup>2</sup> Gussed 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.

can either be assumed to be the actual values (if available and known by agents) or be projected from an assumed stochastic process. It is also possible to assume that agents have incorrect expectations about the exogenous variables, in which case 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  ${}_{s-1}Ey_s, {}_{s-1}Ey_{s+1}, \dots, {}_{s-1}Ey_{s+h}$ . Given these values, (1) 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,  ${}_{s-1}Ey_s$  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 (1), 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  $t - 1$  ( $u_{it-1}, i = 1, \dots, n$ ) 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 our earlier paper.

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, (1) 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 the expectations are solved for, (1) 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 starting in period  $s - j - 1$  and assuming that the error terms for period  $s - j - 2$  are zero. 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  $\varepsilon_{is-1} = 0$  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, (1) and (2) can be used to solve for  $\varepsilon_{is-1}$ .<sup>3</sup> 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  $\varepsilon_{is-1}/\rho_i$ . 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  $\varepsilon_{is-1}$  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.

### 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 algorithm 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

<sup>3</sup> 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.

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.

### 3. FULL INFORMATION MAXIMUM-LIKELIHOOD (FIML) ESTIMATION

Assume that the estimation period is 1 through  $T$ . The objective function that FIML maximizes (assuming normality) is

$$L = -(T/2)\log |S| + \sum_{t=1}^T \log |J_t|, \quad (3)$$

where  $S$  is the covariance matrix of the error terms and  $J_t$  is the Jacobian matrix for period  $t$ .  $S$  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  $S$  is  $(1/T)\sum_{t=1}^T \varepsilon_{it}\varepsilon_{jt}$ . 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  $\alpha$  will be used to denote the vector of all the coefficients in the model. In the serial correlation case  $\alpha$  also includes the  $\rho_i$  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 Fair and Parke (1980), and this algorithm will not be discussed here. Suffice it to say that FIML estimation of large-scale models is computationally feasible. What any algorithm needs to do is to evaluate  $L$  many times for alternative values of  $\alpha$  in the search for the value that maximizes  $L$ .

In the standard case computing  $S$  for a given value of  $\alpha$  is fairly inexpensive. One simply solves (1) and (2) for the  $\varepsilon_{it}$  error terms given the data and the value of  $\alpha$ . 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  $\alpha$ . Therefore, to compute  $S$  for a given value of  $\alpha$ , 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  $S$  and thus one evaluation of  $L$ . In the 25-coefficient problem below, the Parke algorithm required 2817 evaluations of  $L$  to converge, which would be over 3 trillion passes if done this way.<sup>4</sup>

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  $S$  (and thus  $L$ ) each time the algorithm requires a value of  $L$  for a given value of  $\alpha$ .

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

<sup>4</sup> Note that these solutions of the error term  $\varepsilon_{it}$  are only approximations when  $f_i$  is nonlinear. Hence, the method gives an approximation of the likelihood function.

$\alpha$ . To compute all the derivatives requires  $K + 1$  solutions of the model over the  $T$  number of observations, where  $K$  is the dimension of  $\alpha$ . One solution is for the base values, and the  $K$  solutions are for the  $K$  changes in  $\alpha$ , one coefficient change per solution. From these  $K + 1$  solutions,  $K \cdot T \cdot (h + 1)$  derivatives are computed and stored for each expectations variable, one derivative for each length ahead for each period for each coefficient.<sup>5</sup> Once these derivatives are computed, they can be used in the computation of  $S$  for a given change in  $\alpha$ , and no further solutions of the model are needed. In other words, when the maximization algorithm changes  $\alpha$  and wants the corresponding value of  $L$ , the derivatives are first used to compute the expectations, which are then used in the computation of  $S$ . Since one has an estimate of how the expectations change when  $\alpha$  changes (from the derivatives), 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 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 I 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):

$$y_{1t} = \alpha_{11}y_{1t-1} + \alpha_{12}y_{1t-2} + \alpha_{13t-1}Ey_{1t+1} + \alpha_{14t-1}Ey_{1t+2} \\ + \alpha_{15t-1}Ey_{2t} + \alpha_{16t-1}Ey_{2t+1} + \alpha_{17t-1}Ey_{2t+2} + u_{1t}, \quad (4)$$

$$y_{2t} = \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{1t-2} + u_{2t}, \quad (5)$$

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,  $\alpha_{15}$  and  $\alpha_{21}$ . Data for this model were generated using normally distributed serially independent errors with zero correlation between equations. Values of  $\alpha_{15}$  and  $\alpha_{21}$  of 0.0333333 and -0.333333 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

<sup>5</sup> 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 in this paper two-sided derivatives seemed unnecessary. For the results below each coefficient was increased by 5 per cent from its base value when computing the derivatives. Five per cent seemed to give slightly better results than 1 per cent, although no systematic procedure of trying to find the optimal percentage size was undertaken.

$L$  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  $a_{15} = 0.0260125$  and  $a_{21} = -0.3916$ .

Table I 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

Table I. 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 function evaluations
Starting values	0.0333333	-0.333333		508.6022	
Iteration					
1	0.0252994	-0.391662	509.0470	509.0462	61
2	0.0260233	-0.391609	509.0467	509.0467	50
3	0.0260117	-0.391612	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 function evaluations
Starting value	0.0200000	0.600	-0.200000		501.8234	
Iteration						
1	0.0335672	0.635	-0.210860	505.5016	531.1740	77
2	0.0289718	0.673	-0.321878	532.0178	531.7876	166
3	0.0495646	0.745	-0.321324	532.1676	531.8590	103
4	0.0778620	0.837	-0.322183	532.3424	531.9918	103
5	0.0886905	0.878	-0.322699	532.1248	531.9346	96
6	0.0903430	0.889	-0.322646	531.9557	531.9032	90

## Model 3: Six-Equation Model, 25 Coefficients

	$\hat{L}$ using derivatives	$\hat{L}$ using full solution	No. of function evaluations
Starting values		170.3100	
Iterations			
1	189.1670	184.3381	2817
2	189.2047	189.0098	1103
3	189.0450	189.0297	538
4	189.0784	189.0784	258

Notes: DFP algorithm used for models 1 and 2. Parke algorithm used for model 3. Derivatives recomputed after each iteration for models 1 and 2, not for model 3.

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 (4) is assumed to be serially correlated:

$$u_{1t} = \rho_1 u_{1t-1} + \varepsilon_{1t}, \quad (6)$$

where  $\rho_1$  was set equal to 0.7 to generate the data. The coefficient estimates using the factorization routine and the DFP algorithm are  $a_{15} = 0.0738367$ ,  $\rho_1 = 0.83545$ , and  $a_{21} = -0.32211$ . The results in Table I show that the derivative method of this paper got close, but not quite to, the 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 I. This basically seems to be a hard computational problem. The likelihood function is fairly flat near the top, especially with respect to  $a_{15}$  and  $\rho_1$ , and we have found at least one other local optimum.<sup>6</sup>

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 II ( $C$  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 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 (1982) method of moments estimator.<sup>7</sup> The estimation period was 1954 I–1984 IV, for a total of 124 observations. The Hansen estimates were then used as starting values for the FIML calculations.

The results in Table I 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

Table II. Six-equation model 3

1. $\log C_t$	$\text{cnst}, \log C_{t-1}, {}_{t-1}E \log Y_{t+2}, R_t$
2. $I_t - I_{t-1}$	$\text{cnst}, Y_t - Y_{t-1}, {}_{t-1}E(Y_{t+1} - Y_t), R_t, t, I_{t-1}$
3. $\log(M_t/P_t)$	$\text{cnst}, \log(M_{t-1}/P_{t-1}), \log Y_t, R_t$
4. $\log P_t$	$\text{cnst}, \log P_{t-1}, \log PM_t, (YS_t - Y_t)/YS_t,$ ${}_{t-1}E((YS_{t+1} - Y_{t+1})/YS_{t+1}), RHO$
5. $R_t$	$\text{cnst}, R_{t-1}, {}_{t-1}E100((P_{t+2}/P_{t+1})^4 - 1), 100((Y_t/Y_{t-1})^4 - 1),$ $100((M_{t-1}/M_{t-2})^4 - 1)$
6. $Y_t = C_t + I_t + Q_t$ .	

<sup>6</sup> Also, although not reported in Table I, 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.

<sup>7</sup> See Fair (1989) for a discussion of the use of Hansen's estimator in this context.



coefficients. The Parke algorithm was used for the maximization. It can be seen in Table I 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 estimate of the FIML covariance matrix of the coefficient estimates (say  $\hat{V}$ ) is

$$\hat{V} = -(\partial^2 L / \partial \alpha \partial \alpha')^{-1}, \quad (7)$$

where the derivatives are evaluated at the optimum.  $\hat{V}$  can be computed using numerical derivatives,<sup>8</sup> and this was done at the end of the estimation for each model. The covariance computations are feasible because the expectations derivatives can be used in calculating the derivatives in (7). In other words, no further solutions of the model are needed to compute  $\hat{V}$  in (7).  $\hat{V}$  for model 3 is used in the next section for the stochastic simulation results.

#### 4. STOCHASTIC SIMULATION

For models with rational expectations we must state very carefully what we mean by a stochastic simulation of the model and what we use stochastic simulation for. Here we do *not* use stochastic simulation to 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 for the model, and then (4) solving the model for periods  $s$  and beyond. These solution values are the agents' expectations.

What we mean by 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  $N(0, \hat{S})$ , where  $\hat{S}$  is the FIML estimates of  $S$  from the previous section. 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 'trial'. Another trial can be done using a new draw of the vector of error terms, and so on. Let  $y_{is}^j$  be the solution value of  $y_{is}$  on the  $j$ th trial. For  $J$  trials the estimate of the expected value of  $y_{is}$ , denoted  $\tilde{y}_{is}$ , is

$$\tilde{y}_{is} = (1/J) \sum_{j=1}^J y_{is}^j. \quad (8)$$

The estimated variance of the forecast error, denoted  $\hat{\sigma}_{is}^2$ , is

$$\hat{\sigma}_{is}^2 = (1/J) \sum_{j=1}^J (y_{is}^j - \tilde{y}_{is})^2. \quad (9)$$

<sup>8</sup> Again, there are some tricks needed to do this for large models; see Parke (1982) and Fair and Parke (1980).

One can also use this approach to analyse the effects of uncertainty in the coefficients  $\alpha$  by assuming that the coefficients are distributed  $N(\hat{\alpha}, \hat{V})$ , where  $\hat{\alpha}$  is the FIML estimate of  $\alpha$  and  $\hat{V}$  is the estimated covariance matrix as computed in (7). In this case each draw also involves the vector of coefficients.

If  $u_{it}$  is serially correlated as in (2), 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 section), and so no further work is needed. The estimate of  $u_{is-1}$  is simply taken as predetermined for all the trials, and  $u_{is}$  is computed as  $\rho_i u_{is-1}$  plus the draw for  $\varepsilon_{is}$ . (Note that the  $\varepsilon$  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 trial 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 trial, where  $p$  is the length of the period.

Consider the multi-period 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  $s$ ).<sup>9</sup> 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  $\varepsilon_{is+1}$ . 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 trial. The overall process is then repeated for the second trial, and so on. Note that only one coefficient draw is used per trial, i.e. per dynamic simulation. After  $J$  trials 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 III. The FIML estimates of  $S$ ,  $\alpha$ , and  $V$  from the previous section were used for the draws. The length of the prediction was taken to be four, and 100 trials 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 macroeconomic 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

<sup>9</sup> 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 III. Stochastic simulation results for model 3

		1983			
		I	II	III	IV
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)	0.0046	0.0083	0.0110	0.0125
Interest Rate	(a)	8.39	8.57	8.75	8.94
	(b)	8.28	8.40	8.74	9.01
	(c)	0.79	0.96	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

Notes: (a) predicted values from deterministic simulation.  
 (b) estimated mean values from stochastic simulation.  
 (c) estimated standard deviations from stochastic simulation.

The results are based on 100 trials.

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 1 per cent 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  $\varepsilon_{it}$  were drawn, but not values of the coefficients. The vector of coefficients  $\alpha$  was taken to be fixed.

It should be clear that stochastic simulation as defined above is computationally feasible for models with rational expectations. It 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 eight periods and 100 trials, on the other hand, the model has to be solved using the EP method only 800 times.

## 5. CONCLUSION

The results in this paper 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 in section 4. 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 paper will open the way for many tests of models with rational expectations.<sup>10</sup>

#### REFERENCES

- Carlozzi, Nicholas, and John B. Taylor (1985), 'International capital mobility in the coordination of monetary policy rules', in Jagdeep Bandhori (ed.), *Exchange Rate Policy under Uncertainty*, MIT Press, Cambridge, MA.
- DeLong, J. Bradford and Lawrence H. Summers (1986), 'Is increased price flexibility stabilizing?', *American Economic Review*, **78**, 1031–1043.
- Fair, Ray C. (1989), 'Testing the rational expectations hypothesis in macroeconomic models', Mimeo.
- Fair, Ray C., and William R. Parke (1980), 'Full information estimates of a nonlinear macroeconomic model', *Journal of Econometrics*, **13**, 269–291.
- Fair, Ray, C., and John B. Taylor (1983). 'Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models', *Econometrica*, **50**, 1169–1185.
- Hansen, Lars P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica*, **50**, 1029–1054.
- King, Stephen R. (1988), 'Is increased price flexibility stabilizing?: comment', *American Economic Review*, **78**, 267–272.
- Kydland, Finn E., and Edward C. Prescott (1982), 'Time to build and aggregate fluctuations', *Econometrica*, **50**, 1345–1370.
- Parke, William R. (1982), 'An algorithm for FIML and 3SLS estimation of large nonlinear models', *Econometrica*, **50**, 81–96.
- Taylor, John B. (1980), 'Output and price stability: an international comparison', *Journal of Economic Dynamics and Control*, **2**, 109–132.
- Taylor, John B. (1988), 'Policy analysis with a multicountry model', NBER Working Paper.
- Taylor, John B., and Harald Uhlig (1990), 'Solving nonlinear stochastic growth models: a comparison of alternative solution methods', *Journal of Business and Economic Statistics*, **8**, 1–21.

---

<sup>10</sup> All the methods discussed in this paper are part of the Fair–Parke program. It is now quite easy to use these methods within the program, including the use of the derivatives for FIML estimation.