

SOLUTION AND MAXIMUM LIKELIHOOD ESTIMATION OF DYNAMIC NONLINEAR RATIONAL EXPECTATIONS MODELS

BY RAY C. FAIR AND JOHN B. TAYLOR¹

A solution method and an estimation method for nonlinear rational expectations models are presented in this paper. The solution method can be used in forecasting and policy applications and can handle models with serial correlation and multiple viewpoint dates. When applied to linear models, the solution method yields the same results as those obtained from currently available methods that are designed specifically for linear models. It is, however, more flexible and general than these methods. The estimation method is based on the maximum likelihood principle. It is, as far as we know, the only method available for obtaining maximum likelihood estimates for nonlinear rational expectations models. The method has the advantage of being applicable to a wide range of models, including, as a special case, linear models. The method can also handle different assumptions about the expectations of the exogenous variables, something which is not true of currently available approaches to linear models.

1. INTRODUCTION

CONSIDER THE DYNAMIC RATIONAL expectations model given by

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

where y_t is an n -dimensional vector of endogenous variables at time t , x_t is a vector of exogenous variables at time t , 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, and u_{it} is a stationary scalar random variable which has mean zero and which may be correlated across equations ($Eu_{it}u_{jt} \neq 0$ for $i \neq j$) and over time ($Eu_{it}u_{is} \neq 0$ for $t \neq s$). The model is *nonlinear* in that the function f_i may be nonlinear in the variables, parameters, and expectations, although we will require certain regularity conditions on these functions and their derivatives with respect to y_t and α_i . It is a *rational expectations* model in that expectations of future endogenous variables are conditional forecasts based on the model itself, and it is *dynamic* in that the lags and expected leads of the endogenous variables appear in the equations.² The main objectives of this paper are to describe and

¹The research described in this paper was financed by Grants SOC77-03274 and SES79-26724 from the National Science Foundation. The authors are indebted to Ates Dagli for computational assistance.

²Several properties of the general form represented in model (1) should be noted. By appropriate construction, the model can include *expectations* of nonlinear functions of the endogenous variables. For example, if $y_{2t} = y_{1t}^2$, then the appearance of $E_{t-1} y_{2t}$ in one of the equations indicates that the agents are concerned with the conditionally expected variance of y_{1t} . Also, the model permits nonlinear restrictions on the α_i parameters both within and across equations. However, the model does not explicitly include expectations based on *current* period (t) information. The incorporation of such variables does not cause difficulties for the solution of the model (as we describe below), but it does cause difficulties for estimation since the Jacobian of the transformation from the u_t to the y_t is altered.

investigate numerically (i) a method for solving the model for the vector y_t in terms of its past values and the values of the exogenous variables x_t and (ii) a method for obtaining the maximum likelihood estimates of the parameters α_t and the covariance structure of the u_{it} given a series of observations on y_t and x_t , $t = 1, \dots, T$.

The solution method is an extension of the iterative technique used in Fair [3]. In addition to dealing with serial correlation and multiple viewpoint dates, the extension involves an iterative procedure (called Type III in the following discussion) designed to insure numerical convergence to the rational expectations solution.

The estimation method is an extension to the nonlinear case of full information maximum likelihood techniques designed for linear rational expectations models, as described by Wallis [15] and Hansen and Sargent [7, 8].³ Applications to particular economic problems are found in Sargent [12] and Taylor [14]. Full information estimation techniques are particularly useful for rational expectations models because of the importance of cross equations restrictions, where most of the testable implications of the rational expectations hypothesis lie. For linear models one can explicitly calculate a reduced form of model (1), in which the expectations variables are eliminated and nonlinear restrictions are placed on the parameters. Under the assumption that the u_{it} are normally distributed this restricted reduced form can be used to evaluate the likelihood function in terms of the structural parameters. The maximum of the likelihood function with respect to the structural parameters is found using numerical nonlinear maximization routines.

For nonlinear models the reduced form cannot be calculated explicitly, but it can be calculated numerically. Our estimation strategy is to replace the calculation of the restricted reduced form in linear models with the numerical solution in nonlinear models. This permits one to evaluate the likelihood function in terms of the unknown structural parameters, much like in the linear case.

While we feel that the nonlinear methods described will expand the range of empirical problems that can be approached using rational expectations, there is a limitation that may affect their general applicability. Because of computational costs it is necessary in some applications to approximate the conditional expectations that appear in (1) by setting the future disturbances u_{it} equal to their conditional means in a deterministic simulation of the model. In nonlinear rational expectations models the conditional expectations will involve higher order moments of the u_{it} in addition to their means. (See Lucas and Prescott [10], for example.) As we describe in the paper, it is possible to use *stochastic*

³The connection between our problem and the one considered by Hansen and Sargent appears in the f_t functions, which would represent first-order conditions for the linear-quadratic optimization problem they consider. Chow [1] has proposed an alternative approach that leads to the same functional relationship between the structural parameters and the likelihood function as does the Hansen and Sargent approach.

simulation to obtain the actual conditional forecasts, but this is computationally expensive. For some macroeconomic applications the deterministic approximation may be fairly good. Some evidence that this is true is contained in Fair [4], where differences between forecasts using stochastic simulation and deterministic simulation are very small in a nonlinear model without rational expectations. For other applications, however, the approximation may be inaccurate, and the stochastic simulation procedure will need to be used, however expensive.

2. A NUMERICAL METHOD FOR SOLVING NONLINEAR RATIONAL EXPECTATIONS MODELS

In this section we consider the numerical solution of model (1) for a particular period s and for a given set of values of the α_i parameters. The model without serial correlation of the errors is considered first, and then the modifications needed for the serial correlation case are discussed.

In the following discussion $E_{t-1}x_{t+j}$ will be used to denote the expected value of x_{t+j} based on information through period $t-1$. Both the actual realizations of x_t and the expected values are assumed to be known. If there are any exogenous variables that are not known but can be described by a known stochastic process, then these are treated as endogenous and incorporated in the y_t vector. In this section, all simulations of the model are deterministic and are subject to the approximation mentioned in the Introduction.

2.1. Models without Serial Correlation: The Basic Method

If one were given numerical values for the expected endogenous variables in model (1) for all periods from s on, then it would be straightforward to solve the model for period s using the Gauss-Seidel iterative technique. The numerical method described here entails a series of iterations that converge from an arbitrary initial path of values for these expectations to a path of *rational* expectations, consistent with the forecasts of the model itself. Let the initial set of values for the expected endogenous variables, $E_{s-1}y_{s+r}$, be represented as g_r , $r = 0, 1, \dots$. Since in general the model will have no natural termination date, an infinite number of these values need to be specified in principle. In practice, however, only a finite number of these will be used in obtaining a solution with a given finite tolerance range. We require that the initial values be bounded: $|g_r| < M$ for every r , where M is not a function of r .

The solution method can be described in terms of 5 steps:

(i) Choose an integer k , which is an initial guess at the number of periods beyond the horizon h for which expectations need to be computed in order to obtain a solution within a prescribed tolerance level δ . Set $E_{s-1}y_{s+r}$ equal to g_r , $r = 0, 1, \dots, k + 2h$. For the purpose of describing the iterations, call these initial values $e_r(1, k)$, $r = 0, 1, \dots, k + 2h$; the values at later iterations will then be called $e_r(i, k)$, $i > 1$.

(ii) Obtain a new set of values for $E_{s-1}y_{s+r}$, $r = 0, 1, \dots, k + h$, by solving the model dynamically for y_{s+r} , $r = 0, 1, \dots, k + h$. This is done by setting the disturbances to their expected values (usually zero), using the values $E_{s-1}x_s, \dots, E_{s-1}x_{s+h+k}$ in place of the actual x 's, and using the values $e_r(i, k)$ in place of $E_{s-1}y_{s+r}$. Call these new guesses $e_r(i + 1, k)$, $r = 0, 1, \dots, k + h$. If the model is nonlinear, then the solution for each period requires a series of Gauss-Seidel iterations. Call each of these a *Type I* iteration.

(iii) Compute for each expectation variable and each period the absolute value of the difference between the new guess and the previous guess, i.e., compute the absolute value of the difference between each element of the $e_r(i + 1, k)$ vector and the corresponding element of the $e_r(i, k)$ vector for $r = 0, 1, \dots, h + k$. If any of these differences are not less than a prescribed tolerance level (i.e., if convergence has not been achieved), increase i by 1 and return to step (ii). If convergence has been achieved, go to step (iv). Call this iteration (performing steps (ii) and (iii)) a *Type II* iteration.⁴ Let $e_r(k)$ be the vector of the convergent values of a series of Type II iterations ($r = 0, 1, \dots, k + h$).⁵

(iv) Repeat steps (i) through (iii), replacing k by $k + 1$.⁶ Compute the absolute value of the difference between each element of the $e_r(k + 1)$ vector and the corresponding element of the $e_r(k)$ vector, $r = 0, 1, \dots, h$. If any of these differences are not less than δ , increase k by 1 and repeat steps (i) through (iv). If convergence has been achieved, go to step (v). Call this iteration (performing steps (i) through (iv)) a *Type III* iteration. Let e_r be the vector of the convergent values of a series of Type III iterations ($r = 0, 1, \dots, h$).

(v) Use e_r for $E_{s-1}y_{s+r}$, $r = 0, 1, \dots, h$, and the actual values for x_t to solve the model for period s . This gives the desired solution, say \hat{y}_s , and concludes the solution method.

To summarize, the above method iterates on future *paths* of the expected endogenous variables $E_{s-1}y_{r+s}$. Starting from an initial guess at the path g_r , $r = 0, 1, 2, \dots, k + 2h$, the path is *extended* beyond $k + 2h$ until further extensions do not affect the solution by more than δ and convergence is obtained. Since we will need to refer to the procedure outlined in Steps (i) through (v) several times in the remainder of the paper, it will be convenient to use the shorthand terminology *extended path* (EP) method for this series of computational steps.

Computation costs for the EP method are determined by the total number of "passes" through the model required for convergence. A "pass" is simply a single

⁴The Type II tolerance level should be smaller than δ , which is the overall tolerance level. Similarly, the Type I tolerance level should be smaller than the Type II tolerance level.

⁵Note that in the process of achieving Type II convergence the initial guesses $e_r(1, k)$, $r = k + h + 1, \dots, k + 2h$, never get changed. These guesses are needed to allow the model to be solved through period $s + h + k$.

⁶When repeating steps (i) through (iii) for $k + 1$, it may be possible to speed convergence by using some information from iteration k . The most obvious thing to do is to use as initial guesses $e_r(1, k + 1) = e_r(k)$, $r = 0, 1, \dots, k + h$. The values g_r would then be used for $e_r(1, k + 1)$, $r = k + h + 1, \dots, k + 2h + 1$.

evaluation of the “left hand side” endogenous variables in terms of the “right hand side” variables. Let N_1 be the number of Type I iterations required for convergence, and let N_2 be the number of Type II iterations required for convergence. Then the number of passes through the model required for one Type III iteration is given by the product of the number of passes for one Type II iteration ($N_1 \times (h + k + 1)$) and the number of Type II iterations required for convergence (N_2). The total number of passes through the model to obtain Type III convergence is given by the sum of this expression from k to $k + N_3 - 1$, where N_3 is the number of Type III iterations required for convergence. In other words, Type III convergence requires $\sum_{q=k}^{k+N_3-1} [N_2 \times N_1 \times (h + q + 1)]$ passes through the model.⁷

Two points about the solution method should be noted. First, it can be easily modified to handle the case in which the expectations are based on information through period s rather than through period $s - 1$. Just replace E_{s-1} by E_s everywhere. Second, if the expectations horizon is infinite ($h = \infty$), then it must be truncated first. For most models the error introduced by this truncation for reasonably large values of h is likely to be small. A large value of h means, of course, that a large number of calculations are required per Type II iteration, and so in practice there may be a tradeoff between truncation error and computational cost.

For a general nonlinear model there is no guarantee that any of the iterations will converge. If convergence is a problem, it is sometimes helpful to “damp” the successive solution values. This means to take the value of a variable at, say, the start of iteration n to be some fraction of the difference between the value actually computed on iteration $n - 1$ and the value used at the start of iteration $n - 1$.

In special cases a problem may have terminal conditions. If, say, the values $E_{s-1}y_{s+r}$, $r = k + h + 1, \dots, k + 2h$, are known, then the present method gives the correct answer after Type II convergence. No Type III convergence tests are needed because the values for periods $s + k + h + 1$ through $s + k + 2h$ are known. Cases with terminal conditions are referred to as two-point boundary problems. They have been used to study rational expectations models when one can approximate the terminal conditions with steady state values, which may be derived in certain situations.⁸ The approximation that comes from equating the terminal conditions with the steady state values does not arise with the method used in this paper. Moreover, our method does not require that one compute steady state values beforehand.

⁷This formula is only approximate because it is based on the assumption of the same number of Type I iterations for each period and the same number of Type II iterations for each Type III iteration. In practice this is usually not the case.

⁸See Lipton, Poterba, Sachs, and Summers [9]. These authors use a “multiple shooting” method to solve the two-point boundary value problem. The above remarks should make it clear that the method proposed in this paper can be used to solve two-point boundary value problems with an arbitrary number of terminal conditions.

2.2. Models with Serial Correlation: Forecasting and Policy Applications

We focus on the case where the error terms can be described by the first order process:

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

where the ρ_i are serial correlation coefficients. In this section the EP method is modified for applications where there are enough data prior to the solution period s to permit calculation of the solution values with only a negligible effect of the errors prior to period $s - 1$. This situation is likely to occur in forecasting or policy applications, where a large sample prior to the simulation period is usually available. In Section 2.3 the method is modified for estimation applications, where sufficient prior data are generally not available.

First note that (1) and (2) can be combined to yield:

$$(3) \quad f_i(y_t, y_{t-1}, \dots, y_{t-p}, y_{t-p-1}, {}_{t-1}E y_t, {}_{t-1}E y_{t+1}, \dots, {}_{t-1}E Y_{t+h}, \\ {}_{t-2}E y_{t-1}, {}_{t-2}E y_t, \dots, {}_{t-2}E y_{t+h-1}, x_t, x_{t-1}, \alpha_i, \rho_i) = \epsilon_{it}, \\ (i = 1, \dots, n),$$

where the ρ_i can be thought of as structural coefficients. For solution purposes the important difference between (1) and (3) is the addition in (3) of an extra viewpoint date ($t - 2$). This requires an additional type of iteration denoted Type IV.

If one were given values for the expectations with viewpoint date $s - 2$, then model (3) could be solved using the EP method in Section 2.1. These expectations could be obtained by solving the model one period earlier at time $s - 1$, but this in turn would require values for the expectations with viewpoint date $s - 3$, and so on. By working backwards in this way, however, it is possible to insure that these initial values have negligible influence on the current period s .

The procedure is as follows: (a) Choose an integer j , which is an initial guess at the number of periods before period s for which the model needs to be solved in order to achieve the prescribed tolerance level. Set $E_{s-j-2} y_{s-j-1+r}$, $r = 0, 1, \dots, h$, to an initial set of values. (As with the basic method, we require that the initial guesses be bounded.)

(b) Given the values from (a), solve the model for period $s - j$ using the EP method. For this solution the viewpoint date for the expectations for x_{s-j} and beyond is $s - j - 1$. Actual values are used for x_{s-j-2} . The solution yields values for $E_{s-j-1} y_{s-j+r}$, $r = 0, 1, \dots, h$.

(c) Given the expectations with viewpoint date $s - j - 1$ from (b), solve the model for period $s - j + 1$ using the EP method. For this solution the viewpoint date for the expectations for x_{s-j+1} and beyond is $s - j$. Actual values are used

for x_{s-j-1} . This solution yields values for $E_{s-j} y_{s-j+1+r}$, $r = 0, 1, \dots, h$. Continue this procedure (using the EP method to solve for the next period, given the solved-for expectations from the previous period) through period s . The solution for period s yields values for $E_{s-1} y_{s+r}$, $r = 0, \dots, h$.

(d) Increase j by 1 and repeat (a) through (c). This yields new values for $E_{s-1} y_{s+r}$, $r = 0, 1, \dots, h$. Compare these values to the values obtained by using the smaller j . If any new value is not within the prescribed tolerance level⁹ of the old value, increase j by 1 and repeat steps (a) through (c). Keep doing this until convergence is reached. Call this iteration (performing steps (a) through (c)) a *Type IV* iteration.

(e) After *Type IV* convergence, one has final values of $E_{s-1} y_{s+r}$ and $E_{s-2} y_{s-1+r}$, $r = 0, 1, \dots, h$. Use these values and the actual values of x_s and x_{s-1} to solve the model for period s .

Each *Type IV* iteration requires solving the model for $j + 1$ starting points (i.e., achieving *Type III* convergence $j + 1$ times). The serial correlation case is thus considerably more expensive than the nonserial correlation case when solving the model for one period. However, no additional *Type IV* iterations are required for solving the model for periods later than s , once the solution for period s has been obtained. The predictions with viewpoint date $s - 1$ are known after solving for period s , for example, and they can be used in solving for period $s + 1$.

It should be emphasized that *Type IV* iterations can handle problems more general than the case of first-order autoregressive errors. In particular, the expectations variables with viewpoint dates $t - 2$ need not arise solely from the presence of autoregressive errors, and there can be more than two viewpoint dates. If, say, viewpoint date $t - 3$ were also included in the model, the only change in the procedure would be the addition of initial guesses for E_{s-j-3} values in step (a). One would merely need to keep track of three sets of expectations instead of two as the solutions proceeded from period $s - j$ to period s .

2.3. Models with Serial Correlation: Estimation Applications

The *Type IV* iterations discussed in Section 2.2 require sufficient data prior to the solution period so that the initial guesses have a negligible effect on the solution. In most estimation problems one would not want to lose as many observations from the beginning of the sample as would be required for *Type IV* convergence. Fortunately, there is a way around this problem, which is based on an assumption that is usually made when estimating multiple equation models with moving average residuals. This assumption is that the last presample *uncorrelated* error is zero, in particular that $\epsilon_{is-1} = 0$ in equation (2) when solving for period s . As before, we focus on the case of first-order autoregressive errors;

⁹The tolerance level for the *Type IV* iterations should be greater than δ , the tolerance level for the *Type III* iterations.

generalization to higher orders is fairly straightforward. The method requires data¹⁰ for period $s - 1$. Rather than first transforming model (1) into the form of model (3), the method works directly with equation (1), treating the error process (2) as another equation.

If u_{is-2} were known, then model (1) could be solved for period $s - 1$ and all subsequent periods using the EP method and the fact that $E_{s-2}u_{is+r} = \rho_i^{(r+2)}u_{is-2}$. In other words, in the dynamic simulations that underly the EP method, one would use $\rho_i^{(r+2)}u_{is-2}$ on the right-hand side of (1). The problem then becomes one of choosing an appropriate value for u_{is-2} . This is where the assumption about ϵ_{is-1} comes in. The idea is to choose u_{is-2} so that when the model is solved for period $s - 1$, it generates a value of $\epsilon_{is-1} = 0$; that is, $u_{is-1} = \rho_i u_{is-2}$. The rationale for this choice is simply that 0 is the unconditional mean of ϵ_{is-1} and so the actual value is likely to be relatively close to this value.

An iterative procedure for choosing u_{is-2} so that $\epsilon_{is-1} = 0$ can be described as follows (note that each calculation is performed for each equation $i = 1, \dots, n$):

(i) Guess values for the error terms u_{is-2} .

(ii) Given the values from (i), solve the model for period $s - 1$ using the EP method. Note that $E_{s-2}u_{is+r}$ is set to $\rho_i^{(r+2)}u_{is-2}$ in calculating the predicted values.

(iii) Given the predicted value of $y_{is-1}(\hat{y}_{is-1})$ from step (ii), calculate $\hat{\epsilon}_{is-1} = y_{is-1} - \hat{y}_{is-1}$ and $\hat{u}_{is-1} = \rho_i u_{is-2} + \hat{\epsilon}_{is-1}$, where u_{is-2} is the initial guess. If $\hat{\epsilon}_{is-1}$ is not within a prescribed tolerance level of 0, then convergence has not been reached (i.e., the solution is not consistent with the assumption that $\epsilon_{is-1} = 0$).

(iv) If convergence is not reached in (iii), set the new value of u_{is-2} equal to \hat{u}_{is-1}/ρ_i and do (ii) and (iii) over for these new values. Repeat this until convergence is reached.

(v) Using the converged iterate u_{is-2} , compute $u_{is-1} = \rho_i u_{is-2}$. Given these values, solve for period s using the EP method, where in this case $E_{s-1}u_{s+r} = \rho_i^{(r+1)}u_{is-1}$ is used in calculating the predicted values. This completes the solution for period s .

As was the case for the iterative procedure in Section 2.2, once the solution for period s has been obtained, the solutions for periods $s + 1$ and beyond do not require further iterations from those used by the EP method. Again, this is because the forecasts with viewpoint date $s - 1$ are known after solving for period s .

2.4. An Example

In order to see how the method worked for a large-scale nonlinear model, it was applied to two modified versions of the model described in Fair [2, 5]. This model is nonlinear in variables and coefficients, has 97 equations, 29 of which

¹⁰Data before period $s - 1$ will be needed if there are lagged endogenous or exogenous variables in the model. It is implicitly assumed here that sufficient data for the lagged variables are available for the solution for period $s - 1$.

are stochastic, and has first-order serial correlation in 12 of the stochastic equations. The regular version of the model does not have any rational expectations variables in it. For the first modified version four equations were changed: one consumption equation and three labor supply equations. The explanatory price and wage variables in these equations, which enter with no lags, were replaced with the one-period-ahead expected values of the variables. The same coefficients were used for the expectations variables as were used for the non-expectations variables in the regular version. Serial correlation was not present in any of the four equations, so no Type IV iterations were needed.¹¹ The expected values of all the exogenous variables in the model were assumed to be the actual values. For the second modified version two additional equations were changed: another consumption equation and an output equation. Both of these equations have first-order serially correlated errors. Again, the explanatory price and wage variables in the consumption equation were replaced with one-period-ahead expected values. The current sales variable in the output equation was replaced with the one-period-ahead expected value of sales. In both equations the same coefficients were used for the expectations variables as were used for the variables in the regular version.

It should be stressed that the two modified versions are not meant to be sensible alterations of the basic model. They are merely used to test the method. The method worked well for both versions. For the first version k was taken to be 8. (h is 1 in both versions.) Three Type II iterations were needed for the first Type III iteration. This required about 150 passes through the model (10 periods \times 3 Type II iterations \times about 5 Type I iterations per period).¹² Three Type II iterations were also needed for convergence for the second Type III iteration, which required about 165 passes through the model (11 periods \times 3 Type II \times about 5 Type I per period). Type III convergence was reached after two iterations using a tolerance level of 0.003 per cent.

For the second version k was taken to be 8 and j was taken to be 6. This required solutions for seven different starting points for the first Type IV iteration and eight different starting points for the second Type IV iteration. Each solution required two Type III iterations and about five Type II iterations per Type III iteration. Type IV convergence was reached after two iterations using a tolerance level of 0.004 per cent. The total number of passes through the model for the overall solution was about 7875.

3. CONVERGENCE CONDITIONS IN A LINEAR EXAMPLE

In this section we examine the conditions under which the solution method will converge from an arbitrary set of initial guesses to the rational expectations

¹¹Note that Type IV iterations are needed only if there are expectations variables with different viewpoint dates in the model. Therefore, the existence of serial correlation in a model requires Type IV iterations only if the equations with serially correlated errors have expectations variables as explanatory variables.

¹²The approximate time for one pass through the model on the Yale IBM370-158 is 0.05 seconds.

solution in a linear scalar example. The aim is to motivate the numerical method and to relate it to analytic techniques that have been used in previous research for solving and estimating rational expectations models.

A scalar linear version of model (1) with serial correlation is given by

$$(4) \quad y_t = \alpha E_{t-1} y_{t+1} + \gamma E_{t-1} x_t + u_{1t},$$

$$(5) \quad x_t = \lambda x_{t-1} + \epsilon_{2t},$$

$$(6) \quad u_{1t} = \rho u_{1t-1} + \epsilon_{1t},$$

where $\alpha, \gamma, \lambda,$ and ρ are scalar parameters and $(\epsilon_{1t}, \epsilon_{2t})$ is a serially uncorrelated vector. We assume that $|\lambda| < 1$ and $|\rho| < 1$. Equations (4) and (5) correspond to (1) when the exogenous variable x_t is assumed to follow a known stochastic process, and equation (6) corresponds directly to the autoregressive error assumption made in equation (2).

The rational expectations solution of equations (4) through (6) in period s is given by¹³

$$(7) \quad \begin{aligned} E_{s-1} y_s &= \sum_{i=0}^{\infty} \alpha^i \lambda^{i+1} x_{s-1} + \sum_{i=0}^{\infty} \alpha^i \rho^{i+1} u_{1s-1} \\ &= \frac{\gamma \lambda}{1 - \alpha \lambda} x_{s-1} + \frac{\rho}{1 - \alpha \rho} u_{1s-1}. \end{aligned}$$

Note that the last equality in (7) requires that $|\alpha \lambda| < 1$ and $|\alpha \rho| < 1$, which will be satisfied if $|\alpha| < 1$. Our objective is to show that the numerical solution method generates the same solution value as that given in (7). For now take u_{1s-1} as given. A procedure for calculating u_{1s-1} is described subsequently. Recall that $e_r(i, k)$ is the guess of $E_{s-1} y_{s+r}$ on Type II iteration i and Type III iteration k . We start each Type III iteration with an initial set of guesses $e_r(1, k), r = 0, 1, \dots, k + 2$ ($h = 1$ in this example). We want to show that $\lim_{i, k \rightarrow \infty} e_0(i, k)$ equals the right hand side of (7).

For a fixed k , the Type II iterations can be described by the set of equations

$$(8) \quad e_r(i + 1, k) = \alpha e_{r+1}(i, k) + \gamma \lambda^r x_{s-1} + \rho^r u_{1s-1},$$

$r = 0, 1, \dots, k + 1$. By repeated substitution we obtain

$$(9) \quad e_0(k + 3, k) = (\alpha)^{k+2} e_{k+2}(1, k) + \gamma \lambda \sum_{h=1}^{k+1} (\alpha \lambda)^h x_{1s-1} + \rho \sum_{h=1}^{k+1} (\alpha \rho)^h u_{1s-1},$$

which is the converged iterate of the Type II iterations for a fixed k . Equation (9) is not equal to the right hand side of (7). However, if $|\alpha| < 1$, then the limit of $e_0(k + 3, k)$ as $k \rightarrow \infty$ is equal to the right hand side of (7). This motivates our

¹³ An analytic solution method is discussed in Hansen and Sargent [8] and Taylor [14].

requirement that the initial values $e_{k+2}(1, k) \equiv g_{k+2}$ are bounded, and it shows that Type III iterations converge to the rational expectations solution.¹⁴

Note that the condition for this convergence ($|\alpha| < 1$) is identical to the condition needed to obtain a unique solution in rational expectations models (see Taylor [13]). This suggests that the numerical method will converge in the class of rational expectations models for which the uniqueness conditions hold, although a general proof is still open.

We now use the example to illustrate the relationship between the procedure described in Section 2.3 (designed to choose initial conditions for estimation applications) and the conditional maximum likelihood estimates of linear ARMA models.

Substituting (7) into (4) results in

$$(10) \quad y_t = \frac{\gamma\lambda}{1 - \alpha\lambda} x_{t-1} + \frac{\rho}{1 - \alpha\rho} u_{1t-1} + \epsilon_{1t}.$$

Subtracting the lagged value of (10) multiplied by ρ from (10) results in the “quasi-differenced” expression

$$(11) \quad y_t = \rho y_{t-1} + \frac{\gamma\lambda}{1 - \alpha\lambda} (x_{t-1} - \rho x_{t-2}) + \frac{\alpha\rho^2}{1 - \alpha\rho} \epsilon_{1t-1} + \epsilon_{1t},$$

which when combined with (5) gives a two-dimensional vector ARMA (2, 1) model with nonlinear constraints on the parameters. For estimation of the parameters of this ARMA model it is necessary to calculate the residuals ($\epsilon_{1t}, \epsilon_{2t}$) in terms of the data and the parameters. For “conditional” maximum likelihood estimates, this calculation is started by setting $\epsilon_{1s-1} = 0$ and taking y_{s-1}, x_{s-1} , and x_{s-2} as given, where s is the beginning of the estimation period. The residual ϵ_{1s} is then computed by subtracting (11) with these values from the actual observation y_s . The residuals for later periods are calculated recursively using this computed residual ϵ_{1s} .

The procedure described in Section 2.3 is designed to calculate these “conditional” residuals numerically for linear as well as nonlinear models. This can be illustrated by showing that

$$(12) \quad \hat{y}_s = \rho y_{s-1} + \frac{\gamma\lambda}{1 - \alpha\lambda} (x_{s-1} - \rho x_{s-2})$$

when the value u_{1s-1} in (7) is chosen according to the procedure outlined in steps (i) through (v) in Section 2.3. We know from (7) that the basic numerical solution

¹⁴Note that in this model the solution is independent of all g_t values. Given that the g_t values are bounded, Type III iterations insure convergence to the correct answer. This may indicate that the EP method will work well for more complicated models, although, as noted next, this is still an open question.

method will generate

$$(13) \quad \hat{y}_{s-1} = \frac{\alpha\lambda}{1-\alpha\lambda} x_{s-2} + \frac{\rho}{1-\alpha\rho} u_{1s-2}$$

when applied in period $s-1$, as indicated in step (ii). Iterating steps (iii) and (iv) will yield a converged iterate of u_{1s-2} that has the property that $y_{s-1} - \hat{y}_{s-1} \equiv \epsilon_{1s-1} = 0$ to within the tolerance level. From (13) this value of u_{1s-2} is given by

$$(14) \quad u_{1s-2} = \frac{1-\alpha\rho}{\rho} \left(y_{s-1} - \frac{\gamma\lambda}{1-\alpha\lambda} x_{s-2} \right)$$

and therefore

$$(15) \quad u_{1s-1} = \rho u_{1s-2} = (1-\alpha\rho) \left(y_{s-1} - \frac{\gamma\lambda}{1-\alpha\lambda} x_{s-2} \right).$$

Substituting (15) into (7) yields (12), which is what is to be shown. Note that when analytic techniques can be used, it is trivial to choose u_{1s-2} according to (14), but when the solutions are calculated numerically, it is necessary to search for the value u_{1s-2} that gives $\epsilon_{1s-1} = 0$.

4. MAXIMUM LIKELIHOOD ESTIMATION

Assume that the first m equations of the model (1) are stochastic, with the remaining u_{it} ($i = m+1, \dots, n$) identically zero for all t . Given the model (1), let J_t be the $n \times n$ Jacobian matrix whose ij element is $\partial f_i / \partial y_{jt}$ ($i, j = 1, \dots, n$), and let S be the $m \times m$ matrix whose ij element is $(1/T) \sum_{t=1}^T u_{it} u_{jt}$ ($i, j = 1, \dots, m$). Also, let α denote the vector of all the unknown coefficients in the model. If the u_{it} are normally and independently distributed, then the full information maximum likelihood (FIML) estimates of α are obtained by maximizing

$$(16) \quad L = -\frac{T}{2} \log |S| + \sum_{t=1}^T \log |J_t|$$

with respect to α . An estimate of the covariance matrix of these estimates (say \hat{V}) is

$$(17) \quad \hat{V} = - \left(\frac{\partial^2 L}{\partial \alpha \partial \alpha'} \right)^{-1},$$

where the derivatives are evaluated at the optimum. If the u_{it} are correlated according to the relation $u_{it} = \rho_i u_{it-1} + \epsilon_{it}$, where the ϵ_{it} are normally and independently distributed, then the FIML estimates are obtained by maximizing (16) with S replaced by the matrix whose ij element is $(1/T) \sum_{t=1}^T \epsilon_{it} \epsilon_{jt}$. The maximization is then with respect to α and $\rho \equiv (\rho_1, \dots, \rho_m)$, and the derivatives are taken with respect to α and ρ in estimating \hat{V} .

4.1. *Evaluating and Maximizing the Likelihood Function*

Given the solution method in Section 2, it is straightforward to compute L for a given value of α for rational expectations models. If there is no serial correlation, then for a given value of α one can solve for $E_{s-1}y_s, E_{s-1}y_{s+1}, \dots, E_{s-1}y_{s+h}$ for $s = 1, 2, \dots, T$ using the EP method. These values can then be used in conjunction with the y and x data to compute values of u_{it} ($s = 1, 2, \dots, T$) and thus the matrix S . The Jacobian determinants can also be computed, thereby completing the determination of L in (16). The extra work involved in the calculation of L for rational expectations models thus consists of using the solution model to compute the expected values for each of the T viewpoint dates. For models without rational expectations none of these calculations are needed. Given this extra work, however, FIML estimates can be obtained in the usual way by maximizing L numerically¹⁵ with respect to α .

When the u_{it} follow a first-order autoregression process, only one main change to the above procedure is necessary. In this case steps (i) through (iv) in Section 2.3 are needed to calculate the expected values for the first sample point (say, period 2). Given these expected values, which have a viewpoint date 1, the expected values for period 3 can then be obtained using the EP method. These expected values can then be used in the calculation of the expected values for period 4, and so on through the end of the sample point. The only extra work in the serial correlation case pertains to the first sample period. As noted above, numerical maximization in this case is with respect to α and ρ .

4.2. *An Example*

We have experimented with the method using a small linear model. This model, which can also be estimated using existing linear techniques, is a version of a wage contracting model estimated in Taylor [14]. It can be represented as

$$(18) \quad y_{1t} = \alpha_{11}y_{1t-1} + \alpha_{12}y_{1t-2} + \alpha_{13}E_{t-1}y_{1t+1} + \alpha_{14}E_{t-1}y_{1t+2} \\ + \alpha_{15}E_{t-1}y_{2t} + \alpha_{16}E_{t-1}y_{2t+1} + \alpha_{17}E_{t-1}y_{2t+2} + u_{1t},$$

$$(19) \quad y_{2t} = \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{1t-2} + u_{2t},$$

with the restrictions $\alpha_{11} = \alpha_{13} = 1/3$, $\alpha_{12} = \alpha_{14} = 1/6$, $\alpha_{15} = \alpha_{16} = \alpha_{17}$, $\alpha_{21} = \alpha_{22} = \alpha_{23}$. There are thus two free parameters to estimate, α_{15} and α_{21} . The data for this model were generated by simulating the model 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. 50 observations were generated.

¹⁵See Parke [11] and Fair and Parke [6] for a discussion of an algorithm that can be used in large-scale econometric applications and for an application of the algorithm to the model in Fair [2].

The likelihood function was evaluated in two ways. The first used the technique described in Taylor [14], which is based on a factorization procedure that calculates a restricted ARMA version of the model, and the second used the method of this paper. When evaluated at the same parameter values, both methods gave the same value of the likelihood function. A total of about 27,750 passes through the model was required for one evaluation of the likelihood function, which is estimated¹⁶ to take about 1 second on an IBM360/91. This compares to .004 seconds for the linear technique, and so the present method is about 250 times slower than the linear technique. The DFP algorithm was used to maximize the likelihood function, which required 90 function evaluations.

We also evaluated the likelihood function for the case where u_{1t} in (19) follows a first-order autoregressive process, with $\rho_1 = 0.7$. Steps (i) through (v) in Section 2.3 were used with a damping factor of 0.25 to solve for the first observation, with the EP method used thereafter. Again, for the same set of parameter values the same likelihood value was obtained using both the factorization technique and the method proposed in this paper. The total number of passes in this case was about 37,563, which is about 35 per cent greater than the number for the model without serial correlation.

4.3. *A Less Expensive Method for Maximizing the Likelihood Function*

We also experimented with a less expensive method, which is of interest to mention briefly. The basic method is expensive because many evaluations of L are needed for each estimation problem and the model must be solved T times for each evaluation of L . The less expensive method avoids most of the solutions of the model by obtaining a linear approximation to the relationship between the expectations variables and the coefficients. This is done by numerically evaluating the derivatives of the expectations variables with respect to the coefficients at the starting values and then maximizing the likelihood function as if this were the true relationship. This worked very well for the first example. Three solutions of the model were needed to get the derivatives (a base solution and one solution for each of the two coefficient perturbations). The DFP algorithm was then used to maximize L . This required 84 function evaluations,¹⁷ but no extra solutions of the model. The maximum was obtained, and further tests showed that the use of the derivatives in the calculation of L provided a very close approximation to the actual value (i.e. the value obtained by solving the model). For the serially correlated example, on the other hand, the approximation did not work.¹⁸ The

¹⁶These computation times were estimated in order to compare the two approaches. The actual iterations were computed on an IBM370-158. Because of the usual difficulties with such comparisons, these estimates should be viewed as approximate.

¹⁷A different DFP program was used for this problem from the one used in conjunction with the factorization procedure, and so the two uses of the DFP algorithm are not strictly comparable. Both uses led to the same answer (within the prescribed tolerance level), which is a useful check on the methods.

¹⁸For sake of completeness it should be noted that in the serial correlation case one must also calculate numerically the derivative of \hat{u}_{1s-1} with respect to each coefficient, where s is the first sample point.

expectations did not appear to be smooth functions of the coefficients. The results are thus mixed, and more experimentation with alternative models is needed before determining the likely usefulness of the less expensive method. Note that the less expensive method can always be checked by switching to the basic method after convergence has been achieved using the less expensive method.

5. SOLUTION AND ESTIMATION WITH STOCHASTIC SIMULATIONS

In this section we briefly describe how the EP method can be used to obtain more accurate solutions and estimates of rational expectations models by stochastic, rather than deterministic, simulations. Through stochastic simulations it is possible to avoid the approximation of taking expectations through nonlinear functions, though with a significant increase in computation costs. For the purposes of this presentation we focus entirely on the model presented in equation (1) without serial correlation.

Consider first the problem of solving a rational expectations model. Suppose that in addition to the α parameters in model (1) we also know Σ , the covariance matrix of the disturbances u_{it} . Assume that the u_{it} are normally distributed. The solutions procedure is modified as follows. First, the expected values computed in step (ii)— $E_{s-1} y_{s+r}$, $r = 0, 1, \dots, k + h$ —are computed by stochastic rather than deterministic simulations. Instead of setting the disturbances to their expected values and solving once, the model is solved for many different trials. Each trial consists of a set of draws of the disturbances u_{is+r} , $r = 0, 1, \dots, k + h$, from the $N(0, \Sigma)$ distribution (assuming the expected values of all the disturbances are zero). Each expected value is computed as the average across all the trials. Second, the final solution value \hat{y}_s computed in step (v) is also computed by a stochastic rather than a deterministic simulation. In this case only draws of the disturbances for period s are needed.

This method can be used to obtain maximum likelihood estimates of the parameters of model (1). Unlike in the deterministic case, however, the likelihood function cannot be “concentrated” as it is in equation (19). In the fully stochastic case, changes in Σ affect the solution of the model and thereby the computed residuals. Instead, we work directly with the “unconcentrated” (log) likelihood function, which except for a constant can be written

$$(20) \quad L^* = \sum_{t=1}^T \log |J_t| - \frac{T}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^T u_t' \Sigma^{-1} u_t,$$

where $u_t = (u_{1t}, \dots, u_{mt})'$. The maximum likelihood estimates can be obtained by maximizing the function L^* with respect to the parameters (α, Σ) . Each evaluation of L^* requires computing the expected values, $E_{t-1} y_{t+r}$, $r = 0, 1, \dots, k + h$, by stochastic simulation for $t = 1, \dots, T$ using draws from the $N(0, \Sigma)$ distribution. This allows u_t to be computed for each t . The determinants of the J_t can be obtained, and so the function L^* can be evaluated in terms of the parameters (α, Σ) . Nonlinear maximization routines can then be used to maximize L^* .

Because this estimation procedure requires maximization over the $(n + 1)n/2$ independent elements of Σ in addition to the elements of α and because of the stochastic simulation costs, the method is likely to be extremely expensive in practice. Given this, experiments with the method on small representative nonlinear models would be useful to try to gauge how much accuracy is likely to be gained by using stochastic simulation.

6. CONCLUSION

A numerical solution method and an estimation method for nonlinear rational expectations models have been presented in this paper. Our calculations indicate that, when applied to linear models, the solution method yields the same results as those obtained from currently available methods that are designed specifically for linear models. It is, however, more flexible and general than these methods. For large nonlinear models the experimental results in this paper indicate that the solution method works quite well.

The estimation method is based on the maximum likelihood technique. Although the method can be expensive, it is, as far as we know, the only method available for obtaining maximum likelihood estimates for nonlinear rational expectations models. The main advantage of the method is that it is applicable to a wide range of models. The method can also handle a wider range of assumptions about the expectations of the exogenous variables than is possible with currently available approaches to linear models.

Yale University
and
Princeton University

Manuscript received October, 1980; final revision received June, 1982.

REFERENCES

- [1] CHOW, G. C.: "Estimation of Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2(1980), 241–255.
- [2] FAIR, R. C.: *A Model of Macroeconomic Activity. Volume II: The Empirical Model*. Cambridge: Ballinger Publishing Co., 1976.
- [3] ———: "An Analysis of a Macro-Econometric Model with Rational Expectations in the Bond and Stock Markets," *American Economic Review*, 69(1979), 539–552.
- [4] ———: "Estimating the Expected Predictive Accuracy of Econometric Models," *International Economic Review*, 21(1980), 355–378.
- [5] ———: "The Fair Model as of May 1, 1980," mimeo, Yale University, 1980.
- [6] FAIR, R. C., AND W. R. PARKE: "Full Information Estimates of a Nonlinear Macroeconometric Model," *Journal of Econometrics*, 13 (1980), 269–291.
- [7] HANSEN, L. P., AND T. J. SARGENT: "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2(1980), 7–46.
- [8] ———: "Linear Rational Expectations Models for Dynamically Interrelated Variables," in *Rational Expectations and Econometric Practice*, ed. by R. E. Lucas, Jr. and T. J. Sargent. Minneapolis, Minnesota: University of Minnesota Press, 1981.
- [9] LIPTON, DAVID, JAMES POTERBA, JEFFREY SACHS, AND LAWRENCE SUMMERS: "Multiple Shooting in Rational Expectations Models," *Econometrica*, 50(1982), 1329–1333.

- [10] LUCAS, R. E., JR., AND E. C. PRESCOTT: "Investment Under Uncertainty," *Econometrica*, 39(1970), 659–681.
- [11] PARKE, W. R.: "An Algorithm for FIML and 3SLS Estimation of Large Nonlinear Models," *Econometrica*, 50(1982), 81–96.
- [12] SARGENT, T. J.: "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," *Journal of Political Economy*, 86(1978), 1009–1044.
- [13] TAYLOR, J. B.: "Conditions for Unique Solutions in Stochastic Macroeconomic Models with Rational Expectations," *Econometrica*, 45(1977), 1377–1386.
- [14] ———: "Output and Price Stability: An International Comparison," *Journal of Economic Dynamics and Control*, 2(1980), 109–132.
- [15] WALLIS, K. F.: "Econometric Implications of the Rational Expectations Hypothesis," *Econometrica*, 48(1980), 49–74.