COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 564

SOLUTION AND MAXIMUM LIKELIHOOD ESTIMATION OF DYNAMIC NONLINEAR RATIONAL EXPECTATIONS MODELS

Ray C. Fair and John B. Taylor

October 1980

SOLUTION AND MAXIMUM LIKELIHOOD ESTIMATION OF DYNAMIC NONLINEAR RATIONAL EXPECTATIONS MODELS

by

Ray C. Fair and John B. Taylor*

1. Introduction

In this paper we consider the solution and maximum likelihood estimation of nonlinear rational expectations models. A general representation of the class of models we consider is given by

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

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, $\underset{t-1}{E}$ 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

---

*Yale University and Princeton University. The 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.

are conditional forecasts based on the model itself, and it is dynamic in that lags and expected leads of the endogenous variables appear in the equations.[1] The two main objectives of this paper are to describe and analyze: (1) 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 (2) a method for obtaining the maximum likelihood estimates of the parameters α_i and the covariance structure of the u_{it} given a series of observations on y_t and x_t, $t = 1, \ldots, T$.

The solution method is an extension of the iterative technique used in Fair (1979). The extension involves an additional iterative round to insure numerical convergence to the rational expectations solution without the need for off-line sensitivity checks. As will be seen, the solution method must be used to solve a given model a large numbers of times in any single estimation problem, and so it is necessary that convergence be as automatic as possible. In addition, the technique is extended to handle serial correlation of the errors and multiple viewpoint dates for the expectations variables.

---

[1] By appropriate construction of the functions f_i it is possible to make some elements of the vector y_t nonlinear functions of other elements of y_t. For example, if $y_{2t} = \alpha_{21}y_{1t} + \alpha_{22}y_{1t}^2$, then the appearance of $E_{t-1} y_{2t}$ in one of the equations indicates that agents are concerned not only with the conditionally expected mean but also with the conditionally expected variance of y_{1t}. This may be appropriate in an asset demand equation, for example, where y_{1t} represents a rate of return. Hence, although the representation of the model in (1) may appear to involve only expectations of linear functions of the endogenous variables, such transformations permit one to consider expectations of nonlinear functions. It should also be noted that the estimation technique we consider permits nonlinear restrictions on the α_i parameters both within and across equations.

The estimation method is an extension to the nonlinear case of full information maximum likelihood techniques previously designed for linear rational expectations models. Wallis (1980) and Hansen and Sargent (1980a, 1980b) have described methods for estimating linear models using full information maximum likelihood techniques,[2] and applications to particular economic problems are found in Sargent (1978) and Taylor (1980). Full information estimation techniques are particularly useful for rational expectations models because of the importance of cross equation 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. This restricted reduced form usually has a vector ARMA representation that can be used for forecasting, policy evaluation, and estimation. Under the assumption that the u_{it} are normally distributed, the likelihood function can be evaluated in terms of the structural parameters. Maximum likelihood estimation then entails finding the maximum of this function with respect to the structural parameters, using numerical nonlinear maximization routines. This, for example, is the approach used by Taylor (1980) for estimating small linear rational expectations models.

---

[2] In the applications considered by Hansen and Sargent the f_i represent first order conditions for a linear-quadratic optimization problem. Chow (1980) has proposed an alternative approach to estimating parameters in the type of linear-quadratic problem considered by Hansen and Sargent. His technique bypasses the step of obtaining first order conditions, as in Hansen and Sargent, and instead directly computes the likelihood function by matrix Riccati iteration. Because of this, the connection between Chow's problem and ours is not so transparent. Chow, has, however, demonstrated that his approach leads to the same functional relationship between the structural parameters and the likelihood function as does the Hansen and Sargent approach.

For nonlinear models the reduced form cannot be calculated explicitly. It is possible, however, using the solution method described in this paper, to solve a nonlinear model numerically taking into account the restrictions imposed by rational expectations. The estimation strategy proposed in this paper is to replace the calculation of the restricted reduced form in linear models with 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, even though the restricted reduced form is never explicitly calculated.

While the nonlinear solution and estimation methods described here should expand the range of economic problems that can be approached using rational expectations, there are two approximations that may affect their general applicability. First, we assume (as an approximation) that the economic agents being described by the model have knowledge of the parameters α_i and the functions f_i and use this information in computing forecasts. When the solution method is used for forecasting or policy simulations, this assumption pertains to the simulation period. When the method is used for estimation, the assumption pertains to the entire sample period. Taken literally, this is a strong assumption, for it ignores the learning of agents as observations accumulate. Some attempts have been made to avoid the approximation in empirical work by using rolling regressions or Bayes procedures to obtain expectations proxies in the first stage of limited information estimation methods. The appropriate treatment of learning in a full information setup, however, does not yet appear practical. Most previous research has relied on this approximation. Sargent (1979) outlines some of the complications that arise when the assumption is relaxed in a full information context.

Second, we 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 models the conditional expectations will in principle involve higher order moments of the u_{it} in addition to their means. A better approximation of the conditional forecasts would therefore require stochastic simulation, but this is computationally expensive. A strict interpretation of rational expectations clearly requires taking into account the higher order moments. This is the interpretation given to rational expectations in the theoretical nonlinear model of Lucas and Prescott (1970), for example. For many applications, however, the less expensive approximation may be fairly good. Some evidence that this is true is contained in Fair (1980a), where differences between forecasts using stochastic simulation and deterministic simulation are very small in a nonlinear model without rational expectations. We plan in future research to explore the effects of taking the higher order moments into account in rational expectations models, but for the solution and estimation methods described in this paper the less expensive approximation is used.[3]

The paper proceeds as follows. In Section 2 the solution method is described and the results of some experiments using the method on

---

[3] It should be also noted that model (1) is not general enough to include expectations based on current period (t) information. The incorporation of such variables does not cause difficulties for the solution of the model, but it does cause difficulties for estimation since the Jacobian of the transformation from the u_t to the y_t is altered. We are able to handle the case where expectations at different viewpoint dates (e.g., $E_{t-1} y_t$, $E_{t-2} y_t$) appear in the same model, as long as the viewpoint dates are not period t. Serial correlation of the u_{it} leads to models with different viewpoint dates.

a large nonlinear model are reported. In Section 3 the method when applied to linear models is shown to converge to the rational expectations solution from an arbitrary set of initial guesses under a fairly general set of conditions. In Section 4 the maximum likelihood estimation method is described and the results of some experiments using the method on a small linear model are reported.

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. As mentioned earlier, all simulations of the model are deterministic.

2.1. Models without Serial Correlation

If one were given numerical values for the expected endogenous variables in model (1) in all periods, then it would be straightforward to solve the model using the Gauss-Seidel iterative technique, as is typically done in nonlinear econometric models. However, unless these numerical values were equal to the rational expectations values, the predictions from the model would be inconsistent with these values.

The numerical method described here entails a series of iterations that move away from an arbitrary and generally inconsistent set of expectations (the initial guesses) toward a consistent set of rational expectations. Throughout these iterations the model is solved repeatedly using the Gauss-Seidel technique, each time for a different set of expectation values.

The solution method applied to a given period s can be defined in terms of the following 5 computational steps:

[1] Choose an iteger k, which is the number of periods beyond the horizon h for which expectations are to be computed, and guess an initial set of values for $\underset{s-1}{E}\, y_{s+r}$, $r = 0, 1, \ldots, k+2h$. Call these initial values $e_r(i,k)$, $r = 0, 1, \ldots, k+2h$, where $i = 1$ represents the first guess.

[2] Obtain a new set of guesses for $\underset{s-1}{E}\, y_{s+r}$, $r = 0, 1, \ldots, k+h$ by solving the model dynamically for y_{s+r}, $r = 0, 1, \ldots, k+h$. This is done by setting the disturbances to zero, using the values $\underset{s-1}{E}\, x_s, \ldots, \underset{s-1}{E}\, x_{s+h+k}$ in place of the actual x's, and using the values $e_r(i,k)$ in place of $\underset{s-1}{E}\, y_{s+r}$. Call these new guesses $e_r(i+1, k)$, $r = 0, 1, \ldots, k+h$. 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.

[3] 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 $r = 0, 1, \ldots, h+k$. If any of these differences are not less than the prescribed tolerance level (i.e., if convergence has not been achieved), increase i by 1 and return to step [2]. If convergence has been achieved, go to step [4]. Call this iteration (performing steps [2] and [3]) a Type II iteration.[4] Let $e_r(k)$ be the vector of the convergence values of a series of Type II iterations $(r = 0, 1, \ldots, k+h)$.

[4] Repeat steps [1] through [4] replacing k by $k+1$, using as initial guesses $e_r(1, k+1) = e_r(k)$, $r = 0, 1, \ldots, k+2h$, and an initial guess for $\underset{s-1}{E}\, y_{s+2h+k+1}$. 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, $r = 0, 1, \ldots, h$. If any of these differences are not less than a prescribed tolerance level, increase k by 1 and repeat steps [1] through [4]. If convergence has been achieved, go to step [5]. Call this iteration (performaing steps [1] through [4]) a Type III iteration. Let e_r be the vector of the convergent values of a series of Type III iterations $(r = 0, 1, \ldots, h)$.

---

[4] Note that in this process the initial guesses $e_r(1,k)$, $r = k+h+1, \ldots, k+2h$ never get changed. They are merely used to allow the model to be solved through period $s+h+k$. There is, however, nothing wrong with changing these values after each Type II iteration (to, say, the last predicted values $\underset{s-1}{E}\, y_{s+h+k}$). As will be seen, the final solution values are not sensitive to end-of-horizon assumptions.

[5] Use e_r for $\underset{s-1}{E}\, y_{s+r}$, $r = 0, 1, \ldots, h$, 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 steps.

It is illustrative and useful for judging computation costs to examine the total number of "passes" through the model required to obtain a solution. A "pass" is simply a single evaluation of the "left hand side" endogenous variables in terms of the "right hand side" variables. The computation time for one pass is directly related to the number of basic arithmetic operations required for a single evaluation, which in turn depends on the size and complexity of the model. Note that one Type I (Gauss-Seidel) iteration requires one pass through the model.

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 $N_2 \times N_1 \times (h+k+1)$, where h is the expectations horizon and k is the number of additional periods for which the model is solved. This number is 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

$$\sum_{q=k}^{k+N_3-1} [N_2 \times N_1 \times (h+q+1)] = N_1 \times N_2 \times N_3 \times (h+k+1) + N_1 \times N_2 \times \frac{N_3(N_3-1)}{2}$$

passes through the model. For example, if $N_1 = 10$, $N_2 = 3$, and $N_3 = 2$ starting from $k = 10$, then a model with an expectations horizon $h = 5$ would require 990 passes. The final evaluation for the solution (step [5]), after Type III convergence is achieved, would require an additional $N_1 = 10$ passes.

Type III iterations are used to insure that the guesses about the distant future have a negligible effect on the predicted (i.e., expected) values through period $s+h$. This is shown for linear models in Section 3 below. Given Type III convergence, the solution for a linear model is exact, subject to the tolerance criteria. For nonlinear models the solution is only approximate because of the bias introduced by the deterministic simulation. As noted in the Introduction, this bias appears to be small for many econometric applications.

Two further 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.

The method just described is an extension of the technique used in Fair (1979), where only Type I and Type II iterations were performed. The technique was used to solve a model that had an infinite expectations horizon ($h = \infty$), and for solution purposes particular assumptions were

made about the expectation values beyond a certain future date. Although in this case, as reported in Fair (1979, p. 550), a sensitivity analysis showed that the solution values were not sensitive to the assumptions, in general there is no way of insuring this without achieving Type III convergence.

It is sometimes necessary when using the Gauss-Seidel technique to solve models to "damp" the successive solution values in order to achieve convergence. In other words, it is sometimes necessary 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 $n-1$. This type of damping can also be done for the various iterations involved in the present solution method, including the extensions of the method discussed in the next two sections.

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:

$$u_{it} = \rho_i u_{it-1} + \epsilon_{it} , \quad i = 1, \ldots, n , \tag{2}$$

where the ρ_i are serial correlation coefficients. In this section the solution 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:

$$(3) \qquad f_i(y_t, y_{t-1}, \ldots, y_{t-p}, y_{t-p-1}, {}_{t-1}E\, y_t, {}_{t-1}E\, y_{t+1}, \ldots, {}_{t-1}E\, y_{t+h},$$

$${}_{t-2}E\, y_{t-1}, {}_{t-2}E\, y_t, \ldots, {}_{t-2}E\, y_{t+h-1}, x_t, x_{t-1}, \alpha_i, \rho_i) = \varepsilon_{it} ,$$

$$i = 1, \ldots, n ,$$

where the ρ_i can be thought of as structural coefficients. The model (3) differs from (1) by having more variables (the extra lagged values) and more nonlinear restrictions on the coefficients (because of the treatment of the ρ_i as structural coefficients). These differences are not fundamental, however, in the sense that model (1) already includes an arbitrary number of lags and nonlinear constraints on the 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 iterative procedure.

Some intuition behind this additional iterative procedure can be gained by noting that if one were given the expectations with viewpoint date $s-2$, then the model could be solved using steps [1]-[5] in Section 2.1. Model (3) would be in precisely the same form as model (1). The difficulty is that expectations with viewpoint date $s-2$ are not known. These expectations could be obtained by solving the model one period earlier at time $s-1$, but this would require expectations with viewpoint date $s-3$. Obtaining these expectations would require solving the model in period $s-2$, which in turn would require expectations at $s-4$, and so on. By working backwards in this way, however, it is possible to push the unknown information back to a time where

(assuming the model is stable and back data are available) it would have negligible influence on the current period s. The iterative procedure that will now be described is designed to insure that enough back periods have been used to satisfy this negligibility criterion.

The procedure is as follows:

[a] Choose an integer j, which is the number of periods before period s for which the model is to be solved, and guess an initial set of values for $\underset{s-j-2}{E}\, y_{s-j-1+r}$, $r = 0, 1, \ldots, h$.

[b] Given the guesses from [a], solve the model for period $s-j$ using steps [1]-[4] in Section 2.1. 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 $\underset{s-j-1}{E}\, y_{s-j+r}$, $r = 0, 1, \ldots, h$.

[c] Given the expectations with viewpoint date $s-j-1$ from [b], solve the model for period $s-j+1$ using steps [1]-[4]. 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 $\underset{s-j}{E}\, y_{s-j+1+r}$, $r = 0, 1, \ldots, h$. Continue this procedure (using steps [1]-[4] 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 $\underset{s-1}{E}\, y_{s+r}$, $r = 0, 1, \ldots, h$

[d] Increase j by 1 and repeat [a]-[c]. This yields new values for $\underset{s-1}{E}\, y_{s+r}$, $r = 0, 1, \ldots, h$. Compare these values

to the values obtained by using the smaller j. If any new value is not within a prescribed tolerance level of the old value, increase j by 1 and repeat steps [a]-[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 $\underset{s-1}{E}\, y_{s+r}$ and $\underset{s-2}{E}\, y_{s-1+r}$, $r = 0, 1, \ldots, h$. Use these values and the actual values of x_s and x_{s-1} to solve the model for period s.

The Type IV iterations are needed to insure that the guesses about the distant past have a negligible effect on the solution for period s. If convergence were reached after N_4 Type IV iterations, then the model would have been solved for

$$\sum_{q=j}^{j+N_4-1} (q+1) = jN_4 + N_4(N_4+1)/2$$

periods (where j denotes the first j chosen). The total number of passes would be this number times the number of passes required for Type III convergence. If $j = 6$ and $N_4 = 2$, this would be $15 \times 990 = 14850$ passes for the example considered above. This compares to only 990 passes for the case of no serial correlation. The serial correlation case is thus considerably more expensive than the non-serial correlation case when solving 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 forecasts with viewpoint date $s-1$ are known after solving for period

s (up to the approximation defined by the tolerance criteria) and can be used in solving for period s+1 ; and similarly for later periods.

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 .

Models with multiple viewpoint dates can arise from aggregation of economic agents who make their decisions at different points in time based on different information sets. Moreover, certain types of investment decisions need to be made with more lead time than others because of variations in gestation period, and this generates models with multiple viewpoint dates. Finally, higher order autoregressive errors gives rise to multiple viewpoint dates through transformations similar to that performed for model (3). Second order autoregressive errors, for example, would introduce viewpoint date t-3 into model (3).

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 conditions do not affect 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 $\varepsilon_{is-1} = 0$ in equation (2) when solving for period s. As before, we focus on the case of first-order autoregressive errors; generalization to higher orders is fairly straightforward. The method requires data[5] 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 steps [1]-[5] 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 steps [1]-[5], 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 $\varepsilon_{is-1} = 0$; 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 $\varepsilon_{is-1} = 0$ can be described as follows (note that each calculation is performed for each equation $i = 1, \ldots, n$) :

[i] Guess values for the error terms u_{is-2}.

[ii] Given the values from [i], solve the model for period $s-1$

---

[5] Data 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 $s-1$.

using steps [1]-[5]. Note that $E_{s-2} u_{s+r}$ 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 $\hat{\varepsilon}_{is-1} \equiv y_{is-1} - \hat{y}_{is-1}$ and $\hat{u}_{is-1} = \rho_i u_{is-2} + \hat{\varepsilon}_{is-1}$, where u_{is-2} is the initial guess. If $\hat{\varepsilon}_{is-1}$ 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 $\varepsilon_{is-1} = 0$).

[iv] If convergence is not reached in [iii], set the new value of u_{is-2} equal to $\hat{u}_{is-1}/\rho_i$ 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 $u_{is-1} = \rho u_{is-2}$. Given these values, solve for period s using steps [1]-[5], 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 merely require steps [1]-[5]. Again, this is because the forecasts with viewpoint date $s-1$ are known after solving for period s.

2.4. Experiments

Some results of solving a large-scale nonlinear model by the solution method are presented in Table 1. The model is described in Fair (1976, 1980b). It is nonlinear in variables and coefficients, has 97 equations, 29 of which 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. The results of solving this version for one quarter (1970I) are presented first in Table 1. Ten passes through the model were required for Type I convergence.

For the second version of the model four equations were modified: 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.[6] The expected values of all the exogenous variables in the model were assumed to be the actual values. The results of solving this version are presented next in Table 1. As noted in the table, the solution required about 315 passes through the model.

For the third version of the model two additional equations were modified: 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

---

[6]Note 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.

TABLE 1

Computational Summary of Experiments with the Solution Method on a Large Scale Nonlinear Model*
(solution period is s = 11 (1970I))

Regular Version: no future expectations variables

10 Type I iterations (passes through the model) for solution.

Second Version: one-period-ahead expectations variables (h = 1) in four equations, no serial correlation

1. The initial value of k was taken to be 8, so the first simulation period was 11-20 (s through s+h+k). 3 Type II iterations were needed for convergence for this value of k . (Type I convergence was needed a total of 10 x 3 = 30 times.)
2. k was increased to 9, and the period 11-21 was solved. 3 Type II iterations were also needed for convergence for this value of k . (Type I convergence was needed a total of 11 x 3 = 33 times.) The expected values for period 12 from this simulation were within a prescribed tolerance level of the values from the first simulation, so Type III convergence was achieved after 2 iterations.
3. The average number of Type I iterations required for Type I convergence was about 5, so the total number of passes through the model for this problem was 5 x (30+33) = 315.

Third Version: one-period-ahead expectations variables (h = 1) in six equations, two with first order autoregressive errors

1. This problem was solved using the Type IV iterative procedure described in Section 2.2. The initial value of j was taken to be 6, and the initial value of k was taken to be 8. The first simulation period was 5-14 (s-j through s-j+h+k). 5 Type II iterations were needed for convergence. k was increased to 9, and the period 5-15 was solved. 6 Type II iterations were needed for convergence. Type III convergence was achieved.
2. The second simulation period was 6-15. Type III convergence was again achieved after 2 iterations. This process was repeated period by period through period 11-20. Each time, Type III convergence was achieved after 2 iterations. The average number of Type II iterations needed for Type II convergence each time was about 5.

*The model is described in Fair (1976, 1980b) and is summarized in the text. The regular version of the model has 97 equations, 29 of which are stochastic, and has first order serial correlation in 12 equations. The approximate time for one pass through the model on the Yale IBM370-158 is 0.05 seconds.

TABLE 1 (continued)

3. j was increased to 7, and the entire process was repeated. The first simulation period in this case was 4-13. The expected values for period 12 from the simulation that began with period 11 (the process starting with $j = 7$) were within a prescribed tolerance level of the previous values (the process starting with $j = 6$), so Type IV convergence was achieved after 2 iterations.

4. The first Type IV iteration required Type III convergence 7 times, and the second Type IV iteration required Type III convergence 8 times. Each time Type III convergence was achieved after 2 iterations. Each Type III iteration required about 5 Type II iterations. The total number of Type II iterations was thus about $15 \times 2 \times 5 = 150$. The average length of the simulation period for a Type II iteration was 10.5 periods, and the average number of Type I iterations needed for each period was about 5. The total number of passes through the model for this problem was thus about $150 \times 10.5 \times 5 = 7875$.

the expectations variables as were used for the variables in the regular version. The results of solving this version are presented last in Table 1. In this case the solution required about 7875 passes through the model.

Note from Table 1 that Type III and Type IV convergence were always reached in two iterations (the minimum possible). This means that cost savings may have resulted by using smaller starting values for k and j. In practice one should try to choose values of k and j no larger than the smallest values needed to allow Type III and Type IV convergence to be attained in 2 iterations. No experimentation with alternative values of k and j was done for the present example. Note also that the average number of Type I iterations (passes through the model) needed for convergence for the second and third versions of the model was smaller than the 10 iterations needed for the first version. This is because better starting values are generally available for the first Type I iteration for each period when the model is solved many times than when it is solved only once.

3. Convergence Conditions

In this section we establish conditions under which the solution method will converge from an arbitrary set of initial guesses to the rational expectations solution for a linear model. Because the results are confined to the linear case, they can only be applied locally to general nonlinear models--within a neighborhood of the solution for which a first order approximation is sufficiently accurate. It would be useful to establish global convergence conditions for nonlinear models, but this would likely entail restrictions on the f_i that are not satisfied in all economic applications. As a practical matter, there is

likely to be no guarantee that the solution method converges for general nonlinear models. A minimum requirement, however, would seem to be that convergence is guaranteed in the linear case. In addition, by examining the linear case it is possible to relate the numerical method suggested here to certain analytic techniques that have been used in previous research for solving and estimating rational expectations models.

It will be convenient to work through a simple scalar example. Multivariate generalizations are fairly straightforward. A linear version of model (1) with serial correlation is given by

$$y_t = \alpha \underset{t-1}{E} y_{t+1} + \gamma \underset{t-1}{E} x_t + u_{1t} , \tag{4}$$

$$x_t = \lambda x_{t-1} + \varepsilon_{2t} , \tag{5}$$

$$u_{1t} = \rho u_{1t-1} + \varepsilon_{1t} , \tag{6}$$

where α , γ , λ , and ρ are scalar parameters and $(\varepsilon_{1t}, \varepsilon_{2t})$ is a serially uncorrelated vector. Equations (5) and (6) are assumed to be stable: $|\lambda| < 1$ 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).

A rational expectations solution of equations (4) through (6) can be derived analytically as follows.[7] Take expectations on both sides of (4) and use the lag operator to obtain:

---

[7]The analytic techniques used here are discussed in Hansen and Sargent (1980b) and Taylor (1980).

$$(7) \qquad (1 - \alpha L^{-1}) \underset{t-1}{E} y_t = \gamma \underset{t-1}{E} x_t + \underset{t-1}{E} u_{1t} ,$$

where $L^{-i} \underset{t-1}{E} y_t \equiv \underset{t-1}{E} y_{t+i}$. Applying the operator $(1 - \alpha L^{-1})^{-1}$ to both sides of (7) yields:

$$(8) \qquad \underset{t-1}{E} y_t = (1 - \alpha L^{-1})^{-1} (\gamma \underset{t-1}{E} x_t + \underset{t-1}{E} u_{1t})$$

$$= \sum_{i=0}^{\infty} \alpha^i (\gamma \underset{t-1}{E} x_{t+i} + \underset{t-1}{E} u_{1t+i})$$

$$= \gamma \sum_{i=0}^{\infty} \alpha^i \lambda^{i+1} x_{t-1} + \sum_{i=0}^{\infty} \alpha^i \rho^{i+1} u_{1t-1}$$

$$= \frac{\gamma\lambda}{1-\alpha\lambda} x_{t-1} + \frac{\rho}{1-\alpha\rho} u_{1t-1} .$$

Note that the last equality in (8) requires that $|\alpha\lambda| < 1$ and $|\alpha\rho| < 1$, which under the stability assumptions made for (5) and (6) will be satisfied if $|\alpha| < 1$. In other words, the condition on the model necessary for this analytic derivation is that $|\alpha| < 1$.

Equation (8) is the rational expectation of y_t given x_{t-1} and u_{1t-1} . Because it is rational, it is also the current period forecast of y_t based on the model. Our objective is to show that the numerical solution method generates the same solution value as that given in (8). To correspond with the notation in Section 2, let period s be the solution period. We need to show that $\hat{y}_s$ computed from steps [1]-[5] is:

$$(9) \qquad \hat{y}_s = \frac{\alpha\lambda}{1-\alpha\lambda} x_{s-1} + \frac{\rho}{1-\alpha\rho} u_{1s-1} .$$

(Note that when u_{1s-1} is known, the procedure described in steps [1]-[5] can be used for serial correlation. In the following proof we 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 $e_r(1,k)$, $r = 0, 1, \ldots, k+2$ ($h = 1$ in this example). We need to show that $\lim_{i,k\to\infty} e_0(i,k)$ equals the right hand side of (9).

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

$$e_r(i+1, k) = \alpha e_{r+1}(i,k) + \gamma\lambda^r x_{s-1} + \rho^r u_{1s-1}, \tag{10}$$

$r = 0, 1, \ldots, k+1$. That is, given an initial set of guesses $e_r(1,k)$ we obtain a new set $e_r(2,k)$ by solving (10). We then replace the old set $e_r(1,k)$ with $e_r(2,k)$ and obtain $e_r(3,k)$, and so on. Note that $e_{k+2}(i,k) = e_{k+2}(1,k)$ for all $i \geq 1$. This implies that $e_{k+1}(i,k) = e_{k+1}(2,k)$ for all $i \geq 2$, which in turn implies that $e_k(i,k) = e_k(3,k)$ for $i \geq 3$. More generally we have that

$$e_r(i,k) = e_r(k+3-r, k) \quad \text{for } i \geq k+3-r, \quad r = 0, 1, \ldots, k+2, \tag{11}$$

and in particular that $e_0(i,k) = e_0(k+3, k)$ for $i \geq k+3$. Hence, the Type II iterations converge after $k+3$ iterations in this example. This fact can be used to determine the limit of these iterations by picking out the convergent equations from (10); that is

$$
\begin{aligned}
e_0(k+3,\ k) &= \alpha e_1(k+2,\ k) + \gamma\lambda x_{s-1} + \rho u_{1s-1} \\
e_1(k+2,\ k) &= \alpha e_2(k+1,\ k) + \gamma\lambda^2 x_{s-1} + \rho^2 u_{1s-1} \\
e_2(k+1,\ k) &= \alpha e_3(k,h) + \gamma\lambda^3 x_{s-1} + \rho^3 u_{1s-1} \qquad (12) \\
&\vdots \\
e_{k+1}(2,k) &= \alpha e_{k+2}(1,k) + \gamma\lambda^{k+2} x_{s-1} + \rho^{k+2} u_{1s-1} .
\end{aligned}
$$

By repeated substitution we obtain

$$
e_0(k+3,\ k) = (\alpha)^{k+2} e_{k+2}(1,k) + \gamma\lambda \sum_{h=1}^{k+1} (\alpha\lambda)^h x_{1s-1} + \rho \sum_{h=1}^{k+1} (\alpha\rho)^h u_{1s-1} , \qquad (13)
$$

which is the converged iterate of the Type II iterations for a fixed k. Note that (13) is not equal to the right hand side of (9). Hence, we could not expect the Type II iterations to lead to convergence to the rational expectations solution, even if these iterations converged. In this case there is convergence, but to the wrong value. Note also, however, that if $|\alpha| < 1$, then the limit of $e_0(k+3,\ k)$ as $k \to \infty$ is equal to the right hand side of (9). Hence, Type III iterations do converge to the rational expectations solution.

Note that the condition for this convergence ($|\alpha| < 1$) is identical to the condition needed for the analytic derivation of the solution. It is also useful to note that this condition is identical to the one needed to obtain unique saddle path solutions in rational expectations models (see Taylor (1977), for example). This indicates that the numerical method will work effectively in the wide class of rational expectations models for which unique saddle path conditions hold. This result is qualified by the fact that our argument is essentially a local one, since it is restricted to the linear case.

Since the model analyzed in this section has a serially correlated

error, it can be used to derive the relationship between the procedure described in Section 2.3 (designed to choose initial conditions for estimation applications) and the usual approach to choosing initial conditions in estimation based on linear ARMA models. This relationship will be useful in the discussion of maximum likelihood estimation in the next section.

Substituting (8) into (4) results in

$$(14) \qquad y_t = \frac{\gamma\lambda}{1-\alpha\lambda}x_{t-1} + \frac{\rho}{1-\alpha\rho}u_{1t-1} + \varepsilon_{1t} \ .$$

Subtracting the lagged value of (14) multiplied by ρ from (14) results in the "quasi-differenced" expression

$$(15) \qquad y_t = \rho y_{t-1} + \frac{\gamma\lambda}{1-\alpha\lambda}(x_{t-1} - \rho x_{t-2}) + \frac{\alpha\rho^2}{1-\alpha\rho}\varepsilon_{1t-1} + \varepsilon_{1t} \ ,$$

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 $(\varepsilon_{1t}, \varepsilon_{2t})$ in terms of the data and the parameters. In practice, this calculation is usually started by setting $\varepsilon_{1s-1} = 0$ 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 (15) with these values from the actual observation y_s. The residuals for later periods are calculated recursively using this computed residual ε_{1s}. In general an ARMA(p,q) model requires p initial endogenous variable values and q initial residuals set to zero. Because of these initial conditions, the calculated residuals are referred to as "conditional," and the likelihood function based on these residuals is called the "conditional" likelihood function.

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

$$\hat{y}_s = \rho y_{s-1} + \frac{\gamma\lambda}{1-\alpha\lambda}(x_{s-1} - \rho x_{s-2}) \tag{16}$$

when the value u_{1s-1} in (9) is chosen according to the procedure outlined in steps [i] through [v] in Section 2.3. We know from (9) that the basic numerical solution method will generate

$$\hat{y}_{s-1} = \frac{\alpha\lambda}{1-\alpha\lambda}x_{s-2} + \frac{\rho}{1-\alpha\rho}u_{1s-2} \tag{17}$$

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 $y_{s-1} - \hat{y}_{s-1} \equiv \varepsilon_{1s-1} = 0$ to within the tolerance level. From (17) this value of u_{1s-2} is given by

$$u_{1s-2} = \frac{1-\alpha\rho}{\rho}\left(y_{s-1} - \frac{\gamma\lambda}{1-\alpha\lambda}x_{s-2}\right) \tag{18}$$

and therefore

$$u_{1s-1} = \rho u_{1s-2} = (1-\alpha\rho)\left(y_{s-1} - \frac{\gamma\lambda}{1-\alpha\lambda}x_{s-2}\right) . \tag{19}$$

Substituting (19) into (9) yields (16), 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 (18), but when the solutions are calculated numerically, it is necessary to search for the value u_{1s-2} that gives $\varepsilon_{1s-1} = 0$. In any case, this analysis has shown that when the procedure described in Section 2.3 is used to calculate residuals, the results will be equivalent to the "conditional" residuals used for "conditional" maximum likelihood estimation of multivariate ARMA models. This in fact is our main rationalefor the procedure.

4. Maximum Likelihood Estimation

Assume that the first m equations of the model (1) are stochastic, with the remaining u_{it} $(i = m+1, \ldots, n)$ identically zero for all t. Given the model (1), let J_t be the $n \times n$ Jacobian matrix whose ij element is $\partial f_i / \partial y_{jt}$ $(i, j = 1, \ldots, n)$, and let S be the $m \times m$ matrix whose ij element is $\frac{1}{T} \sum_{t=1}^{T} u_{it} u_{jt}$ $(i, j = 1, \ldots, m)$. 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

$$L = -\frac{T}{2} \log |S| + \sum_{t=1}^{T} \log |J_t| \tag{20}$$

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

$$\hat{V} = -\left(\frac{\partial^2 L}{\partial \alpha \partial \alpha'} \right)^{-1} , \tag{21}$$

where the derivatives are evaluated at the optimum. If the u_{it} are correlated according to the relation $u_{it} = \rho_i u_{it-1} + \varepsilon_{it}$, where the ε_{it} are normally and independently distributed, then the FIML estimates are obtained by maximizing (20) with S replaced by the matrix whose ij element is $\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \varepsilon_{jt}$. The maximization is then with respect to α and $\rho \equiv (\rho_1, \ldots, \rho_m)$, 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 $\underset{s-1}{E}\, y_s$, $\underset{s-1}{E}\, y_{s+1}$, ..., $\underset{s-1}{E}\, y_{s+h}$ for $s = 1, 2, \ldots, T$ using steps [1]-[4] in Section 2.1. This requires that steps [1]-[4] be done T times, once for each period of the sample. These values can then be used in conjunction with the y and x data to compute values of u_{is} $(s = 1, 2, \ldots, T)$ and thus the matrix S. The Jacobian determinants can also be computed, thereby completing the determination of L in (20). The extra work involved in the calculation of L for rational expectations models thus consists of using the solution method 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 numerically 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]-[iv] are needed to calculate the expected values for the first sample point (say, period 2). Given these expected values, which have viewpoint date 1, the expected values for period 3 can then be obtained from steps [1]-[4]. 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 period. Since steps [1]-[4] can be used for all sample points except the first, the only extra work in the serial correlation case pertains to the first sample point. In other words, steps [i]-[iv] are used to calculate initial conditions only once per solution over the entire sample period. Numerical

maximization in this case is with respect to α and ρ. The analysis of Section 3 indicates that in the linear case the numerical evaluation of L for a given α and ρ is equivalent (subject to the tolerance criteria) to the evaluation of the conditional likelihood function. For large samples, the estimates obtained from maximizing this conditional likelihood function will be close to the unconditional FIML estimates.

There are a number of algorithms available for numerically maximizing a nonlinear function of parameters given a procedure to evaluate the function for a given set of parameter values. For small models many of these algorithms can be used to maximize L, but for large models the only algorithm that appears capable of this is the Parke (1980) algorithm.[8] Once the estimates have been obtained, the covariance matrix (21) of the estimated parameters can be calculated by taking numerical derivatives with respect to α and ρ at the optimum.

Computation time is an important consideration in evaluating the proposed estimation method. We have performed limited experiments with the method on a small linear model, which can also be estimated using existing linear techniques. As would be expected, the computation time for the general nonlinear method described here is significantly greater than it is for the estimation techniques designed explicitly for this linear system. The computation time is not prohibative, however, and the results indicate some potential for practical applicability of the method on small systems. Even for linear systems the flexibility of the nonlinear method could make it attractive relative to linear techniques, despite the greater computation costs.

---

[8] See Parke (1980) and Fair and Parke (1980) for a discussion of this and for an application of the algorithm to the model in Fair (1976).

The model we experimented with is a version of a wage contracting model estimated in Taylor (1980). It can be represented as

$$(22) \qquad y_{1t} = \alpha_{11}y_{1t-1} + \alpha_{12}y_{1t-2} + \alpha_{13}{}_{t-1}E\, y_{1t+1} + \alpha_{14}{}_{t-1}E\, y_{1t+2} + \alpha_{15}{}_{t-1}E\, y_{2t} + \alpha_{16}{}_{t-1}E\, y_{2t+1} + \alpha_{17}{}_{t-1}E\, y_{2t+2} + u_{1t} ,$$

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

with the restrictions $\alpha_{11} = \alpha_{13} = 1/3$, $\alpha_{12} = \alpha_{14} = 1/6$, $\alpha_{15} = \alpha_{16} = \alpha_{13}$, $\alpha_{21} = \alpha_{22} = \alpha_{23}$. 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 randomly generated errors. Normally distributed serially independent errors with zero correlation between equations were used to generate the data. The true values of α_{15} and α_{21} (.0333333 and -.333333) were used for this purpose.

The model was first estimated using the technique described in Taylor (1980), which is based on a factorization procedure that calculates a restricted ARMA version of the model. This ARMA model is used to calculate the likelihood function. Using a sample of 50 observations and the Davidon, Fletcher, Powell (DFP) algorithm,[9] the estimated parameters were $\hat{\alpha}_{15} = .02601$ and $\hat{\alpha}_{21} = -.39160$, with estimated asymptotic t-values of 1.18 and 6.33, respectively. Each evaluation of the likelihood function took .004 seconds of CPU time on an IBM360/91 using this factorization technique.

The likelihood function was also evaluated using the general nonlinear method discussed above. When evaluated at the same parameter values, the method gave the same value of the likelihood function as did the

[9]Because of the small size of this problem, the Parke algorithm was not used.

factorization technique, which serves as a useful check on both procedures. The details of the iterations of the method when evaluating this likelihood are summarized in the upper panel of Table 2. A total of about 27750 passes through the model were required for one function evaluation, which is estimated[10] to take about 1 second on an IBM360/91--about 250 times slower than the linear technique.

We also evaluated the likelihood function for the case where u_{1t} in (22) follows a first order autoregressive process, with $\rho_1 = 0.7$. Steps [i]-[v] in Section 2.3 were used with a damping factor of 0.25 to solve for the first observation, with steps [1]-[4] in Section 2.1 used thereafter. Some initial experimentation with no damping factor for calculating the initial condition indicated that convergence would either not be achieved or would be very slow. This is the only case where a damping factor was used for the empirical results in this paper. 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. A summary of the calculations for the method is presented in the lower panel of Table 2. The required number of passes in this case (37563) is about 35 percent greater than the number required for the model without serial correlation.

4.2. A Less Expensive Method for Maximizing the Likelihood Function

For other than small models the estimation method proposed in Section 4.1 may be prohibitive, depending on the speed and cost of one's computer. The problem is that the solution method of Section 2 must

---

[10] These 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.

TABLE 2

Computational Summary of Likelihood Function Evaluation for a Small Linear Model*

Model with no serial correlation (h = 2)

1. The initial value of k was taken to be 15. Type III convergence was almost always achieved after 2 iterations (i.e., for k = 16).
2. The average number of Type II iterations per Type III iteration was about 15.
3. Given the expectations, the model is recursive, so only one Type I iteration was needed for convergence each period. The average length of the simulation period for a Type II iteration was 18.5 periods, so the total number of Type I iterations for the solution for the 50 observations was about:

   (50 obs.) × (1 Type I iteration) × (15 Type II iterations)
   × (18.5 periods per Type II iteration) × (2 Type III iterations)
   = 27750.

4. One Type I iteration requires about 10 multiplications and 7 additions.

Model with serial correlation (h = 2)

1. Call the first period of the sample period, period s . Steps [i]-[iv] were first used to calculate u_{ls-1} . The initial guess for u_{ls-2} was zero. A damping factor of 0.25 was used. Convergence took 18 iterations to obtain u_{ls-1} . For these calculations the initial value of k was taken to be 15, and Type III convergence was always achieved after 2 iterations. The average number of Type II iterations per Type III iteration was about 15. Given the expectations, the model is recursive, and so only one Type I iteration was needed for convergence each period. The number of passes for these calculations was thus about

   (18 iterations) × (1 Type I) × (15 Type II)
   × (18.5 periods per Type II) × (2 Type III) = 9990.

2. Given u_{ls-1} , step [v] was used to solve for period s . This required 21 Type II iterations. The starting values for this step were the values computed in 1, and convergence was achieved after 1 Type III iteration. The number of passes for this step was thus (21 Type II) × (18 periods per Type II) = 378.

*The model is described in equations (22) and (23) in the text.

TABLE 2 (continued)

3. Given the solution for period s, the calculations for the remaining 49 periods are essentially the same as those above for the model with no serial correlation. These calculations thus required about $49 \times 1 \times 15 \times 18.5 \times 2 = 27195$ passes.

4. The total number of passes through the model was thus about $9990 + 378 + 27195 = 37563$.

be used T times for each evaluation of L, which requires many passes through the model; and many evaluations of L are required for any given maximization problem. In this section we consider a way of modifying the estimation method that requires fewer calls to the solution method. This modification is as follows:

[A] Given the initial value of α, solve for

$$E_{s-1} y_s,\ E_{s-1} y_{s+1},\ \ldots,\ E_{s-1} y_{s+h} \quad \text{for } s = 1, 2, \ldots, T \text{ by}$$

doing steps [1]-[4] T times. Call the solution values from this step the "base" values.

[B] Perturb each coefficient (one at a time) from its initial value and do steps [1]-[4] T times to get a new set of solution values. From these values and the base values, calculate numerically the derivatives of the expectations with respect to the coefficients. This step requires doing steps [1]-[4] T times for each coefficient.

[C] In the procedure that calculates L for a given value of α, use the base values and the derivatives to calculate new expected values for each new value of α. This allows the elimination of all step [1]-[4] calculations in the computation of L for a given value of α.

[D] Once the maximization algorithm has found the value of α that maximizes L, repeat steps [A]-[C] for the new value. Keep repeating these steps until each element of α on the last iteration is within a prescribed tolerance level of the same element on the previous iteration.

The advantage of this modification is that once the problem is turned over to the maximization algorithm, no further step [1]-[4] cal-

culations are needed. The use of the base values and derivatives in the calculation of L is very inexpensive relative to the use of steps [1]-[4], and given that the algorithms require many calculations of L, this modification is likely to result in considerable savings of time. There is, of course, no guarantee that the procedure will converge. If the expectations are not a well behaved function of α, then computing the derivatives at a given point may not be very helpful. It may be, in other words, that using the base values and derivatives to calculate new expected values yields values that are far away from the (correct) values that would be computed by doing steps [1]-[4] T times. For one of the two examples discussed below the use of the derivatives worked very well, but for the other example it did not. More experimentation is thus needed before the usefulness of the modification can be determined.

Once the estimates have been obtained, the covariance matrix (21) can be calculated by taking numerical derivatives of L with respect to α (at the optimum). It may be possible to use the derivatives of the expectations with respect to α in the calculation of the values of L. This would allow the covariance matrix to be computed without having to do any step [1]-[4] calculations.

For the serial correlation case one must also calculate in step [B] the derivative of $\hat{u}_{is-1}$ with respect to α (for each i), where s is the first sample point. $\hat{u}_{is-1}$ is a function of α, and so if steps [i]-[iv] are to be bypassed in the calculation of L, the derivative of $\hat{u}_{is-1}$ with respect to α must also be calculated and used.

Using the model in (22) and (23), some experiments with the less expensive method were made. The true values of the parameters were used as starting values. Using these values, the model was first solved for

each of the 50 observations. As noted earlier, this solution requires about 27750 passes through the model. The model was then solved two more times to calculate the derivatives of the expectations with respect to the two coefficients. The DFP algorithm was then used to maximize L. This required 45 calls to the subroutine that calculates L for a given value of the coefficient vector. As can be seen in Table 3, convergence was essentially achieved after the first iteration. The program was allowed to run for three more iterations, where for each iteration the model was solved three times: once to get the base values and twice more to get the derivatives. The results in Table 3 show that the use of the derivatives provides a close approximation to the "true" value of L obtained by solving the entire model. Given that the DFP algorithm required 45 calls to the subroutine (for the first iteration), the use of the derivatives saved a considerable amount of time. The derivatives were also used in the calculation of the covariance matrix after the optimum was reached.

While the results in Table 3 are encouraging regarding the usefulness of the derivatives, another set of results that we have obtained is not. The procedure was also used to try to estimate the version of the model in (22) and (23) in which u_{1t} follows a first order autoregressive process. The use of the derivatives in this case did not work. The expectations did not appear to be well behaved functions of the coefficients, and quite different derivatives were obtained for different step sizes. The values of L computed using the derivatives were generally not very close to the values of L computed by solving the entire model. It thus appeared for this version that in order to estimate the model one would have to solve it each time that a new value

TABLE 3

Results of Estimating a Small Linear Model
Using the Less Expensive Method*

	$\hat{\alpha}_{15}$	$\hat{\alpha}_{21}$	Value of L Using Expected Values Computed from Derivatives	Value of L Using Expected Values Computed from Steps [1]-[4]	Number of Times L Was Computed by the DFP Algorithm
Initial Values	.0333333	-.333333		508.6022686	
Iteration 1	.0252715	-.391654	509.0471277	509.0460742	45
Iteration 2	.0260208	-.391609	509.0466651	509.0466725	39
Iteration 3	.0260044	-.391616	509.0466727	509.0466724	20
Iteration 4	.0260076	-.391612	509.0466725		71

Estimated standard error of $\hat{\alpha}_{15}$ = 0.0221141 .
Estimated standard error of $\hat{\alpha}_{21}$ = 0.0618044 .

*The model is described in equations (22) and (23) in the text. The method is described in Section 4.2 in the text.

of L was needed by the algorithm. More estimation of alternative models is needed before determining whether these difficulties with the less expensive method are specific to this example and whether the example is representative of the type of model that is likely to be estimated in practice.

5. Conclusion

A numerical solution method and an estimation method for nonlinear rational expectations models have been 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. For large nonlinear models the experimental results in this paper indicate that the method works quite well.

The estimation method is based on the maximum likelihood principal. 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.

The main disadvantage of the estimation method is that it requires many passes through the model to obtain the estimates. We had limited success with a modification of the method to cut down on the number of passes. The experimental results showed, however, that the basic method

is not prohibitive for small models. We plan in future research to estimate more models by the method to try to determine its practical limits. Two other limitations of the estimation method--that it is based on deterministic rather than stochastic simulation and that it cannot handle expectations with a current-period viewpoint date--are also items on our research agenda.

REFERENCES

Chow, G. C. (1980). "Estimation of Rational Expectations Models," Journal of Economic Dynamics and Control, 2, 3, 241-255.

Fair, R. C. (1976). A Model of Macroeconomic Activity. Volume II: The Empirical Model. Cambridge: Ballinger Publishing Co.

_____ (1979). "An Analysis of a Macro-Econometric Model with Rational Expectations in the Bond and Stock Markets," American Economic Review, 69, 4, 539-552.

_____ (1980a). "Estimating the Expected Predictive Accuracy of Econometric Models," International Economic Review, 21, 2, 355-378.

_____ (1980b). "The Fair Model as of May 1, 1980," mimeo.

_____ and W. R. Parke (1980). "Full Information Estimates of a Non-linear Macroeconometric Model," Journal of Econometrics, 13, 269-291.

Hansen, L. P. and T. J. Sargent (1980a). "Formulating and Estimating Dynamic Linear Rational Expectations Models," Journal of Economic Dynamics and Control, 2, 1, 7-46.

_____ (1980b). "Linear Rational Expectations Models for Dynamically Interrelated Variables," in R. E. Lucas, Jr. and T. J. Sargent (eds.), Rational Expectations and Econometric Practice. Minneapolis, MN: University of Minnesota Press.

Lucas, R. E., Jr. and E. C. Prescott (1970). "Investment under Uncertainty," Econometrica, 39, 659-681.

Parke, W. R. (1980). "An Algorithm for FIML and 3SLS Estimation of Large Nonlinear Models," mimeo.

Sargent, T. J. (1979). "Interpreting Economic Time Series," unpublished manuscript, University of Minnesota.

_____ (1978). "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," Journal of Political Economy, 86, 1009-1044.

Taylor, J. B. (1977). "Conditions for Unique Solutions in Stochastic Macroeconomic Models with Rational Expectations," Econometrica, 45, 6, 1377-1386.

_____ (1980). "Output and Price Stability: An International Comparison," Journal of Economic Dynamics and Control, 2, 1, 109-132.

Wallis, K. F. (1980). "Econometric Implications of the Rational Expectations Hypothesis," Econometrica, 48, 1, 49-74.