February 1995

The Bank of Canada's New Quarterly Projection Model

Part 2

A Robust Method For Simulating Forward-Looking Models

by John Armstrong, Richard Black,
Douglas Laxton and David Rose

ISSN 0713-7931

ISBN 0-662-22985-1

Printed in Canada on recycled paper

CONTENTS

ACKNOWLEDGMENTS

We thank Leo Butler, Peter Hollinger, Sean Holly, Tiff Macklem, Hope Pioro, Steven Symansky, Robert Tetlow and Ralph Tryon for helpful comments and discussion. Quynh-Loan Le and Hope Pioro have provided us with excellent programming assistance in implementing the methods described in this report. Helen Meubus's editorial suggestions helped improve the exposition of difficult material. Any remaining errors are our responsibility.

John Armstrong's contribution to this research was made while he was working for the Bank of Canada on part-time secondment from his regular position at Statistics Canada.

ABSTRACT

In this report, we describe methods for solving economic models when expectations are presumed to have at least some element of consistency with the predictions of the model itself. We present analytical results that establish the convergence properties of alternative solution procedures for linear models with unique solutions. Only one method is guaranteed to converge, whereas most widely used methods, including the popular Fair-Taylor approach, do not have this property. This method, which we have implemented for simulation of the Bank of Canada's models of the Canadian economy, involves solving simultaneously the full problem, "stacked" to represent each endogenous variable at each time point with a separate equation, using a Newton algorithm.

We discuss briefly the extension of our convergence results to applications with non-linear models, but the strong analytical conclusions for linear systems do not necessarily carry over to non-linear systems.

We illustrate the analytical discussion and provide some evidence on comparative solution times and on the robustness of the procedures, using simulations of a simple, linear model of a hypothetical economy and of two much larger, non-linear models of the Canadian economy developed at the Bank of Canada. The examples show that the robustness of our procedure does carry over to applications with working, non-linear economic models. They also suggest that the limitations of iterative methods are of practical importance to economic modellers.

RÉSUMÉ

Dans le présent rapport, les auteurs décrivent des méthodes de résolution des modèles économiques sous l'hypothèse que les anticipations ont un élément de cohérence avec les prédictions du modèle en cause lui-même. Ils présentent des résultats, obtenus par la voie analytique, qui établissent les propriétés de convergence que revêtent certaines méthodes appliquées aux modèles linéaires à solution unique. Toutefois, la propriété de convergence n'est garantie que dans le cas d'une seule méthode, tandis qu'elle ne l'est pas dans celui des méthodes les plus couramment utilisées, comme celle de Fair et Taylor. La méthode en question est celle qui a été mise à contribution dans la simulation des modèles de l'économie canadienne mis au point à la Banque du Canada; elle implique la résolution simultanée, à l'aide d'un algorithme de Newton, d'un système d'équations où chacune des variables endogènes est représentée par une équation distincte à chaque période de temps.

Les auteurs traitent brièvement de l'extension de l'analyse des propriétés de convergence à des modèles non linéaires, mais les résultats analytiques ne s'appliquent pas nécessairement à ce type de modèle.

Pour illustrer les résultats analytiques, fournir des mesures comparatives du temps nécessaire à la solution des modèles et des témoignages au sujet du degré de solidité des méthodes évaluées, les auteurs présentent des simulations d'un modèle linéaire simple d'une économie fictive et de deux modèles non linéaires beaucoup plus grands de l'économie canadienne mis au point à la Banque du Canada. Ces exemples montrent que leur méthode demeure robuste lorsqu'elle est appliquée à des modèles économiques non linéaires. Ces exemples laissent aussi supposer que les limites des méthodes itératives revêtent une importance pratique pour les constructeurs de modèles économiques.

1 INTRODUCTION

In this report, we document the properties of alternative numerical solution procedures for economic models with forward-looking behaviour and some degree of model consistency in expectations of future values of endogenous variables. In particular, we describe the advantages of a procedure, which we call the Integrated Dynamic Simulator (IDS), that we have implemented for solving the Bank of Canada's new Quarterly Projection Model (QPM).[1]

In IDS, time-dependent equations are "stacked" such that each endogenous variable at each time point is represented by an independent equation. Time is effectively removed from the problem by "integrating" it into the stacked structure. The entire resulting system is solved simultaneously using a Newton procedure.

More precisely, then, we label our procedure IDS-N, to distinguish it from other approaches to solving the stacked problem. The stacking incorporated in IDS-N eliminates what is known as type II iterations in the method of Fair and Taylor (1983) and substitutes a second-order method based on derivatives.

We show that IDS-N has a strong advantage over other suggested procedures. In the case of a linear system with a unique solution, convergence to that solution is theoretically guaranteed under IDS-N, where it is not with other methods commonly used. In the case of non-linear systems, we can offer no conclusions that hold universally, but the advantages of IDS-N do carry over to solutions started in the region of the true values. Moreover, we offer some evidence, based on our experience with using this procedure to solve QPM (which is highly non-linear), that IDS-N is quite robust in practice.

The idea of stacking equations to form a single system is by no means original. Indeed, many of the papers that consider alternative

1. See Poloz, Rose and Tetlow (1994) for a summary of the model and its use. Details on the underlying theory are provided in Black, Laxton, Rose and Tetlow (1994).

solution methods describe stacked systems and possible first-order methods for solving them.[2] However, most of the practical focus has been on finding ways to break the problem down into a number of smaller sub-problems that can be solved separately and then linked in some iterative scheme. There has been virtually no attention to the possibility of solving the entire system simultaneously using a second-order method.[3]

Because of the stacking to remove time from the problem, the IDS method generates matrices that expand rapidly with the length of time that must be simulated, as well as with the size of the model. This is one reason why second-order methods have not been considered practicable in the past.[4]

For the type of model used regularly in policy analysis at institutions like the Bank of Canada, this point is of great importance. To simulate QPM, for example, we regularly generate problems with more than 30 000 pseudo-equations. The IDS-N method gains the advantage of robustness at the cost of increasing substantially the scale of the core matrix inversion problem.

Dealing routinely with such large-scale problems has become feasible only recently with advances in computing technology and depends critically on the availability of efficient methods for inverting sparse matrices. Even with these methods, solving QPM for 100 quarters requires about 34 megabytes of memory, most of which is for inverting the core matrix.[5]

The report also contains some examples documenting the relative performance of the Fair-Taylor (FT) and IDS-N methods under different

2. See, for example, Hall (1985), Fisher and Hughes Hallett (1988), Holly and Hughes Hallett (1989), Fisher (1992) and Press, Teukolsky, Vetterling and Flannery (1992).

3. Hall (1985) illustrates an iterative, Gauss-Seidel procedure to solve a stacked system. However, Hall's application involves a model with just one forward-looking variable in two equations and a quite short simulation period (10 quarters).

4. This view is expressed in Fisher, Holly and Hughes Hallett (1986), for example.

5. We use the Harwell MA30 routines in our implementation. See J. K. Reid (1977) for further information.

conditions.[6] Where conditions are especially favourable, FT may produce answers in less time. However, our examples show that the speed of IDS-N is relatively independent of the conditions of the problem, whereas the speed of FT deteriorates sharply as conditions become less favourable. Moreover, we report several examples where FT cannot find the solution, but where IDS-N has no difficulty in doing so. It is in this sense that IDS-N is robust. We have found it to be very reliable in regular use under widely varying conditions.

The paper is organized as follows. In Section 2, we describe the problem we consider and the general nature of solution procedures that have been suggested. A number of solution procedures are described in Section 3 in greater detail, and some closed-form results concerning the convergence properties of these methods in the linear case are derived in Section 4. Section 5 contains some summary remarks about linear systems and an overview of extension to non-linear systems. In Section 6, we describe a simple linear model of an economy and use it to illustrate the main analytical points of Section 4. We also describe our experience in using alternative solution procedures for two much larger models of the Canadian economy. Section 7 concludes the report.

6. We apply the FT label to a variety of methods that use iterative methods which do not treat forward expectations as endogenous variables at each stage of the solution. The specific suggestion of Fair and Taylor (1983) is one example, perhaps the most widely known, of such a procedure. Our examples use a version of FT that applies a Newton procedure for "inner-loop" calculations. In the terminology we establish in Section 3, we use an FT-N algorithm in our comparisons.

2 THE RATIONAL EXPECTATIONS PROBLEM

The starting point for the analytical discussion is a model including forward expectations given by

$$f_i(y_t, y_{t-1}, E_{t-1}y_{t+1}, x_t) = u_{it}, \qquad i = 1, \ldots, n, \tag{1}$$

where y_t is an $n \times 1$ vector of endogenous variables at time t, x_t is a vector of exogenous variables at time t, E_{t-1} denotes an expectation conditional on the model and information available through time $t-1$, and the disturbance terms u_{it} are independent scalar random variables with zero means and finite variances. This model is a rational expectations model in the sense that $E_{t-1}y_{t+1}$ is the forecast of y_{t+1} obtained by solution of the model conditional on information available through time period $t-1$. The expectation is formed in period t, or the start of period t, and could, in principle, take account of the values of exogenous variables in period t. However, in our representation of the problem in discrete time, the random disturbances, and hence the solution of the endogenous variables for period t, are not known when the expectation is formed. Note that a model involving lags or leads of more than one period can be written in the form of equation (1) by introducing additional variables.

Without loss of generality, we can assume that the derivative of f_i with respect to endogenous variable i is non-zero. Using this assumption, the model can be rewritten as

$$y_{it} = g_i(y_t, y_{t-1}, E_{t-1}y_{t+1}, x_t, u_{it}), \qquad i = 1, \ldots, n. \tag{2}$$

In the linear case, model (2) can be written as

$$B_0 y_t = B_1 y_{t-1} + A_1 E_{t-1} y_{t+1} + C x_t + u_t, \tag{3}$$

where the $n \times n$ matrices B_0, B_1 and A_1 do not depend on y_t, y_{t-1} or $E_{t-1}y_{t+1}$, and B_0 has been normalized to have a unit main diagonal.

Blanchard and Kahn (1980) and Buiter (1984) provide closed-form solutions for some examples of model (3) and discuss conditions for the uniqueness and stability of the solutions.

A large number of methods have been used to solve model (2). They generally involve a reformulation of the problem as a two-point boundary-value problem, which necessitates the provision of a terminal condition. We describe a classification scheme for these methods in the next section.[7] We focus on what are called extended-path methods, to use the terminology of Fair and Taylor (1983), for imposing the terminal condition. Extended-path methods involve computing a solution for a finite number of time periods beyond the last point for which the solution is to be reported. The computation horizon is extended until the impact on the solution of further extensions is small relative to some convergence tolerance over the entire period of interest.

7. Our classification scheme uses many of the elements of the presentation in Fisher (1992).

3 SOLUTION METHODS

Eight methods for solution of model (2), as well as two variants on these methods, are described in this section. The convergence properties of these methods are considered in Section 4.

Each method fits into the same basic iterative framework, which can be described as follows. Suppose that it is necessary to solve a rational expectations model for T time periods. To begin, it is necessary to choose an initial value of $k \geq T + 1$, the number of time periods for which expectations must be computed to obtain a solution within a prescribed tolerance. There are then four steps involved in the iterative framework:

(i) Choose a value, h_{k+1}, for the terminal condition, $E_{k-1}y_{k+1} = h_{k+1}$.

(ii) Solve model (2) for time period r, $r = 1, 2, \ldots, k$, subject to the rational expectations constraints $E_{r-1}y_{r+1} = y_{r+1}$, $r = 1, 2, \ldots, k-1$, and the terminal condition from step (i). Let $e_r(k)$, $r = 1, 2, \ldots, k$, denote the $n \times 1$ vectors of endogenous variable values corresponding to the solution at time r.[8]

For the first pass, repeat steps (i) and (ii), replacing k by $k+1$ to obtain an initial comparison for step (iii).

(iii) Compare $e_r(k+1)$ and $e_r(k)$ for $r = 1, 2, \ldots, T+1$. If any of the differences between these vectors is greater than the prescribed tolerance, increase k by one and repeat steps (i), (ii) and (iii). Otherwise, let e_r, $r = 1, 2, \ldots, T+1$, denote the converged values and go to step (iv).

(iv) Solve the model equations using e_r in place of $E_{r-2}y_r$ for $r = 2, 3, \ldots, T+1$.

8. The set of equations in (ii) is what we refer to as the "stacked" model. It consists of (2) repeated for times $t = 1, 2, \ldots, k$, with $E_{t-1}y_{t+1}$ being replaced by y_{t+1} and $E_{k-1}y_{k+1}$ by h_{k+1}. Obviously, as (2) itself represents a system of n equations, the stacked model is a system of nk equations.

One repetition of step (ii) results in a type II solution. One repetition of steps (i), (ii) and (iii) is called a type III iteration by Fair and Taylor.

For our purposes, it is the different methods for obtaining type II solutions that distinguish alternative approaches to solving the rational expectations problem. Two general approaches to calculation of type II solutions are examined. They are (a) direct application of an algorithm to the complete set of simultaneous equations in (ii), and (b) iterative procedures involving calculation of a sequence of solutions for endogenous variables conditional on various values for forward expectations. In what follows, we refer to these as type (a) and type (b) methods.

Any algorithm that can be applied to a system of simultaneous equations can be applied to the stacked model. Such algorithms include the Jacobi (J), Gauss-Seidel (GS) and Successive Over-Relaxation (SOR) procedures, as well as the Newton algorithm (N). These algorithms provide four type (a) methods for calculation of type II solutions and hence four methods for solution of a rational expectations model. We refer to these as IDS-J, IDS-GS, IDS-SOR and IDS-N.

As far as we know, Hall (1985) is the only example of an attempt to treat the problem as we do – as a single, large matrix inversion problem. Citing, among other things, the lack of robustness of Fair-Taylor methods, Hall proposes that GS be applied to the complete system of equations – IDS-GS in our classification scheme.

Calculation of the solution of model (2) for k time periods using a type (b) method involves additional iterations that can be defined in two parts as follows.

(b1) To start iteration $i+1$, set the forward expectations $E_{r-1}y_{r+1}$ to $e_{r+1}(i, k)$ for $r = 1, 2, \ldots, k-1$. Then, compute a new solution for the endogenous variables, conditional on these expectations, using an algorithm for the solution of simultaneous equations. Denote the solution by $y_r(i+1, k)$, for $r = 1, 2, \ldots, k$.

(b2) If $e_{r+1}(i, k)$ and $y_{r+1}(i+1, k)$ do not differ by more than a prescribed tolerance for $r = 1, 2, \ldots, k-1$, declare convergence. Otherwise, set $e_{r+1}(i+1, k)$ to $y_{r+1}(i+1, k)$ for $r = 1, 2, \ldots, k-1$, increase i by one and repeat step (b1).

One repetition of (b1) is called an "inner-loop" iteration and one repetition of (b1) and (b2) is called an "outer-loop" iteration. The new solution for the endogenous variables calculated in step (b1) can be obtained in one of two ways. First, an algorithm for the solution of simultaneous equations (J, GS, SOR or N) can be applied to the equations for each time period separately. In this case, lagged endogenous variables are replaced by $y_{r-1}(i+1, k)$ during solution of the equations for time period r at iteration $i+1$. Alternatively, the algorithm can be applied to the stacked model, but with forward expectations held fixed at the previous outer-loop solution values.

In this report, only type (b) methods that involve a separate solution process for each time period at step (b2), using the same algorithm, will be considered. The first example of a type (b) method is the procedure proposed by Fair and Taylor (1983). The FT method involves separate inner-loop solution processes for each time period. They suggest use of the Gauss-Seidel algorithm. The implementation of the FT method in the TROLL software (Hollinger 1988), which we use for the computing in the examples reported in Section 6, employs the Newton algorithm to obtain inner-loop solutions. Four type (b) methods involving separate inner-loop solution processes using the same algorithm for each time period – FT-J, FT-GS, FT-SOR and FT-N – will be examined here.

Two variations on type (b) methods will also be discussed. First, since outer-loop damping is available in TROLL, it is a variation of considerable practical interest. At step (b2), $e_{r+1}(i+1, k)$ is replaced by the linear combination $\gamma y_{r+1}(i+1, k) + (1-\gamma)e_{r+1}(i, k)$ for $\gamma > 0$.

Our second variant is a method described by Fisher (1992, 54). At step (b2), the vectors $e_{r+1}(i+1, k)$, $r = 1, 2, \ldots, k-1$ are obtained by solving for expectations using the Newton algorithm. The solution is computed for all time periods simultaneously, conditional on the current best estimates of the endogenous variables. This procedure is a special case of the "penalty function" method for solving non-linear rational expectations models suggested by Holly and Zarrop (1983) in the context of optimal control. It is also related to a method proposed by Don and Gallo (1987) for modifying Newton iterations used for the solution of models that do not include forward expectations.

Type (b) methods that feature incomplete inner-loop iterations have received considerable attention in the literature. These methods, called type (c) methods by Fisher and Hughes Hallett (1988), include as special cases the type (b) methods we describe as well as first-order type (a) methods. However, IDS-N cannot be formulated as a type (c) method. We find it clearer, therefore, to think of these methods as variants of type (a) and type (b) procedures rather than as a distinct category.

There are a number of proposed solution procedures that may not appear to fit into our classification scheme. For example, Lipton, Poterba, Sachs and Summers (1982) suggest solving model (2) using a multiple-shooting method based on the methods available for non-linear difference equations (for example, Roberts and Shipman 1972). Given initial values for the endogenous variables, the state-space form of the model is used to integrate forward. The difference between the results of forward integration and a terminal condition is used to update the initial values. Fisher and Hughes Hallett (1988) indicate that this method can be written in a form such that it can be interpreted as a type (a) method.

4 CONVERGENCE ANALYSIS FOR LINEAR SYSTEMS

In this section, the convergence properties of the type II solution methods described above are analysed with reference to the linear model, model (3), under the assumption that a unique solution exists. We begin with the properties of type (a) methods and then turn to those of type (b) methods.[9]

Analysis of the convergence properties of type (a) methods is straightforward. The stacked system corresponding to k time periods can be written as

$$Bz = v, \tag{4}$$

where z is an $nk \times 1$ vector formed by stacking the vectors y_t, $t = 1, 2, \ldots, k$, where the $nk \times nk$ matrix B has a block tridiagonal structure reflecting the constraints $y_{t+1} = E_{t-1}y_{t+1}$, $t = 0, 1, \ldots, k-1$, and where the $nk \times 1$ vector v incorporates disturbance terms as well as terms involving exogenous variables and the terminal condition $E_{k-1}y_{k+1} = h_{k+1}$.

Under our assumption that there is a unique solution, the matrix B is non-singular and has a unique inverse, and the solution is simply $z = B^{-1}v$. The IDS-N algorithm is guaranteed to reach this solution in one iteration. The only problem is to compute the inverse matrix. While this may not be a trivial numerical problem, there is no logical difficulty.

To analyse the convergence properties of the first-order type (a) methods – IDS-J, IDS-GS and IDS-SOR – additional notation must be introduced. In particular, we need the decomposition

$$B = I_{nk} - L - U, \tag{5}$$

where I_{nk} is the identity matrix of dimension $nk \times nk$ and the matrices L and U are lower triangular and upper triangular, respectively.

9. See Fisher and Hughes Hallett (1988) for a similar analysis that focusses on type (b) methods.

Let $z(i, k)$ denote the best estimate of z after i iterations of a first-order type (a) method. The general form of a first-order type (a) iteration is given by

$$z(i+1, k) = P^{-1}Qz(i, k) + P^{-1}v \ , \tag{6}$$

where $B = P - Q$. Recursive application of equation (6) yields

$$z(i, k) = (P^{-1}Q)^{i}z(0, k) + \left(\sum_{m=0}^{i-1}(P^{-1}Q)^{m}\right)P^{-1}v \quad . \tag{7}$$

The spectral radius of a matrix is the modulus of its largest eigenvalue. As is evident from equation (7), the recursive solution of equation (6) will converge if and only if the spectral radius of $P^{-1}Q$ is less than unity.

For the algorithms IDS-J, IDS-GS and IDS-SOR, the matrix P is given by I_{nk}, $I_{nk} - L$ and $(I_{nk} - \alpha L)/\alpha$, respectively, where the relaxation parameter α can be chosen. Consequently, for example, a type (a) algorithm that uses IDS-J will converge if and only if the spectral radius of the matrix

$$G_{IDS-J} = I_{nk} - B \tag{8}$$

is less than one. Similarly, type (a) algorithms that use IDS-GS or IDS-SOR will converge if and only if the spectral radii of the matrices

$$G_{IDS-GS} = (I_{nk} - L)^{-1}U \tag{9}$$

and

$$G_{IDS-SOR} = (I_{nk} - \alpha L)^{-1}(\alpha U + (1-\alpha)I_{nk}) \tag{10}$$

are less than one, respectively.

It is important to note that the restrictions in equations (8), (9) and (10) are not trivial. There will be a class of models where a solution exists but where an iterative, first-order algorithm will fail to find it. Only IDS-N is guaranteed to find the solution.

Overall convergence of the FT-N, FT-J, FT-GS and FT-SOR solution methods depends on the convergence of both inner-loop and outer-loop iterations. We consider first the conditions required for inner-loop convergence and then the conditions required for outer-loop convergence.

In the case of FT-N, inner-loop convergence is guaranteed. To analyse inner-loop convergence for the first-order type (b) methods, the decomposition $B_0 = I_n - L_0 - U_0$ is required. The matrices L_0 and U_0 are lower- and upper-triangular, respectively. Let $y_r^{(s)}(i, k)$ denote the estimate of $y_r(i, k)$ obtained after s iterations of a first-order procedure. It follows that

$$\begin{aligned} y_r^{(s)}(i, k) = {} & P_0^{-1} Q_0 y_r^{(s-1)}(i, k) \\ & + P_0^{-1}(A_1 e_{r+1}(i, k) + B_1 y_{r-1}(i, k) + v_r), \end{aligned} \tag{11}$$

where v_r is the component of v corresponding to time period r, $B_0 = P_0 - Q_0$ and P_0 is given by I_n, $I_n - L_0$ and $(I_n - \alpha L_0)/\alpha$ for the J, GS and SOR algorithms, respectively. Convergence will be obtained for $y_r(i, k)$, $r = 1, 2, \ldots, k$, if and only if the spectral radius of the matrix $G_0 = P_0^{-1} Q_0$ is less than or equal to one.

Given inner-loop convergence, the conditions required for outer-loop convergence of type (b) methods do not depend on the algorithm used to compute inner-loop iterations. Converged inner-loop values must satisfy the equation

$$z(i, k) = (I_k \otimes B_0^{-1})(U^* f(i, k) + L^* z(i, k)) + \Delta \quad , \tag{12}$$

where the $nk \times 1$ vector $f(i, k)$ is formed by stacking the vectors $e_r(i, k)$, $r = 1, 2, \ldots, k$, Δ is an $nk \times 1$ vector that depends on exogenous variables, disturbance terms and the terminal condition, $\otimes$ is the Kronecker product operator, and the $nk \times nk$ matrices U^* and L^* are upper and lower block triangular, respectively:

$$J^* = \begin{bmatrix} 0 & A_1 & \dots & 0 \\ 0 & 0 & A_1 \dots & 0 \\ \dots & \dots & \dots & A_1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad L^* = \begin{bmatrix} 0 & 0 & 0 & 0 \\ B_1 & 0 & \dots & 0 \\ 0 & B_1 & \dots & 0 \\ 0 & 0 & \dots B_1 & 0 \end{bmatrix}. \tag{13}$$

Solving system (12) for $z(i, k)$ yields

$$(I_{nk} - (I_k \otimes B_0^{-1})L^*)z(i, k) = (I_k \otimes B_0^{-1})(U^* f((i, k) + \Delta)) \quad . \tag{14}$$

Recall that outer-loop iterations of a Fair-Taylor algorithm involve updating expectations using new best estimates of endogenous variables. That is, $f(i + 1, k)$ is set to $z(i, k)$, and the updating equation for the outer loop is then

$$f(i + 1, k) = (I_{nk} - (I_k \otimes B_0^{-1})L^*)^{-1}(I_k \otimes B_0^{-1})(U^* f(i, k) + \Delta) \quad . \tag{15}$$

Repeated substitution yields

$$f(i, k) = (SU^*)^i f(0, k) + S \sum_{m=0}^{i-1} (SU^*)^m \Delta \ , \tag{16}$$

where

$$S = (I_{nk} - (I_k \otimes B_0^{-1})L^*)^{-1}(I_k \otimes B_0^{-1}) \quad . \tag{17}$$

Thus, regardless of the algorithm used to solve the inner loop of a type (b) method, outer-loop convergence will be achieved if and only if the spectral radius of the matrix SU^* is less than one.

If outer-loop damping is used, outer-loop convergence of a type (b) method will depend on the spectral radius of the matrix

$$S_\gamma = (1 - \gamma)I_{nk} + \gamma SU^* . \tag{18}$$

Hughes-Hallett (1981) provides a proof that some $\gamma > 0$ exists such that the matrix S_γ has spectral radius less than one if and only if all the eigenvalues of the matrix SU^* have real parts less than one.

Consider now the variant where updated estimates of expectation terms are calculated at outer-loop iterations through application of Newton's algorithm to all equations simultaneously, holding the current best estimates of the endogenous variables fixed.[10]

At step (b2), $f(i+1, k)$ is calculated, for the linear case, as the solution to

$$(I_k \otimes B_0 - (I_k \otimes B_1)L^*)z(i, k) = (I_k \otimes A_1)U^* f(i+1, k) + \Delta \quad . \tag{19}$$

Algebraic manipulation of (14) and (19) leads to

$$U^* f(i+1, k) + \Delta = ZU^* f(i, k) + Z\Delta, \tag{20}$$

where

$$Z = (I_k \otimes A_1)^{-1}(I_k \otimes B_0 - I_k \otimes B_1 L^*) \times (I_{nk} - (I_k \otimes B_0^{-1})L^*)^{-1}(I_k \otimes B_0^{-1}) \tag{21}$$

and where matrix pseudo-inverses are calculated when matrix inverses cannot be obtained owing to singularity. In this case, outer-loop iterations will converge if and only if the spectral radius of Z is less than one.

10. See Fisher (1992, 54). This is a special case of the penalty function method of Holly and Zarrop (1983).

5 THEORETICAL CONCLUSIONS AND EXTENSIONS TO NON-LINEAR SYSTEMS

The convergence results presented here can be summarized as follows. Type II convergence of IDS-N, a type (a) method in our classification scheme, is guaranteed if a unique solution exists. Convergence is not guaranteed for the other seven type II solution methods considered – IDS-J, IDS-GS, IDS-SOR, FT-N, FT-J, FT-GS and FT-SOR. We have provided necessary and sufficient conditions for each of these methods to converge on the solution. These conditions can always be written as a requirement that the spectral radius of one or more matrices be less than one.

Convergence of the three type (a) methods that do not use the Newton algorithm – IDS-J, IDS-GS and IDS-SOR – is determined by the spectral radius of a single matrix, which differs from one method to the next.

Convergence of FT-N also depends on the spectral radius of a single matrix, which differs from the matrices involved in convergence results for the IDS methods. Convergence of the other type (b) methods considered – FT-J, FT-GS and FT-SOR – will not be achieved unless the matrix involved in FT-N convergence has spectral radius less than one. In addition, convergence of FT-J, FT-GS and FT-SOR depends on the spectral radius of a second matrix, which differs from one method to the next.

Variations on these methods involving more complicated schemes for updating estimates of expectations terms have qualitatively similar convergence properties.

The importance of the type II convergence results described here for any particular application involving a linear rational expectations model will obviously depend on model characteristics. There is little we can say, in general, as to how important the limitations of the iterative methods are. However, in Section 6 we offer some examples that demonstrate that these limitations cannot be dismissed as curiosities.

In the non-linear case, both local and global convergence results that can be applied to type (a) methods are provided by Rheinboldt (1974). However, these results are not very powerful. The global results apply only to special cases, and application of the local results is limited by a requirement that starting values be “close enough” to the solution. The importance of starting values for convergence of iterative solution techniques for non-linear systems is well known. See, for example, Ralston (1965) or Dodes (1978).

We know of no general method that can be used to determine whether an application of IDS-N to a particular model using a given starting point will lead to a region of solution space in which the local convergence conditions apply. Similarly, the eigenvalues required to analyse the convergence properties of the other type II solution methods must be computed using a linearized version of the model and consequently depend on the particular iterative solution path as well as on the properties of the linear approximation of the model at the true solution values.

It is important to stress that the conditions for convergence of iterative type II methods established for linear systems apply in the non-linear case in the neighbourhood of the true solution values. There can be no guarantee that an iterative method will converge, even from arbitrarily good starting values. In addition, there is a path-dependency problem – the convergence properties will depend on the nature of the linear approximation to the non-linear model. Obviously, if the convergence conditions for a linear system are satisfied at all points along an iterative solution path, convergence will be achieved.[11] Unfortunately, no results are available, as far as we know, to indicate whether a particular combination of model, method and starting point will generate such a path.

Although the superior convergence properties of IDS-N for linear models suggest that this method should be preferred in the non-linear case

11. This is not a necessary condition – not meeting the convergence requirements in the linear approximation at a particular point in the sequence does not necessarily imply that those requirements will not be satisfied at subsequent points. Presumably, however, there must be some point in a sequence past which these sufficient conditions must hold.

as well, empirical evaluation is necessary. We now turn to some examples of the properties of alternative solution procedures applied to some models of economic systems.

6 EXAMPLES

In this section, we compare results from IDS-N with those from FT-N applied to three economic models. We have established (in Section 4) that the properties of outer-loop convergence are independent of the method used to obtain inner-loop convergence. For linear models, therefore, if FT-N fails to converge, then so will any other variant of the FT class. We return to this issue in subsection 6.4.

6.1 A simple linear model

This example takes the form of a simple linear model, which has a unique solution and is stable according to the Blanchard-Kahn (1980) conditions. We show that FT-N may not converge at all or may converge very slowly, depending on the value of a particular parameter. IDS-N will converge in one iteration at a speed that is not a function of the parameter value.

In addition to the issue of whether convergence is obtained, there is a question concerning the accuracy of the solution. A typical scheme is to compare the last two iterations and, if they are close under some metric, to declare convergence. We find that with FT methods this can be misleading, since convergence is often declared when the latest iteration is still some distance from the true solution, considerably more than the supposed tolerance.[12] For this linear example, we report the number of iterations required for FT to converge (within the stated tolerance) to the true solution. We discuss this issue in greater detail in subsection 6.3.

The example economy is represented in a four-equation model composed of an output equation, an inflation equation (Phillips curve), a monetary reaction function (designed to assure control of inflation around a fixed target rate) and a Fisher equation linking nominal interest rates to real interest rates and inflation. The equations are

12. Of course, the same problem can arise, at least in principle, for solutions of non-linear models with IDS methods, especially iterative IDS methods. Our experience suggests that this is not a practical problem for IDS-N, an observation consistent with the super-linear convergence property of Newton methods.

(output) $$y = 0.75y_{t-1} - \alpha r + \varepsilon, \quad (22)$$

(Phillips curve) $$\pi = 0.2y, \quad (23)$$

(monetary reaction function) $$i = 2\pi^e_{t+1}, \quad (24)$$

(Fisher equation) $$i = r + \pi, \quad (25)$$

where y is the deviation of actual output from potential output, r is the real interest rate, ε is a disturbance term, π is the inflation rate, i is the nominal interest rate and α is some positive parameter. This is a standard model augmented with a monetary authority whose current interest rate settings are based on the expected value of inflation next period. We have simplified the Fisher equation slightly, using the actual rather than the expected rate of inflation, to limit the problem to its simplest form, where the model has but one forward-looking variable appearing in only one equation.

Although this model is intended to provide a numerical example, not a realistic model of an economy, the structure and parameters do reflect some key properties of realistically calibrated models. It is similar, for example, to the model used in Laxton, Rose and Tetlow (1993). One simplification here is that the effect of the policy variable is assumed to be contemporaneous, whereas Laxton, Rose and Tetlow emphasize the importance of lags in the control process.

Table 1 reports the convergence properties of the outer-loop iterations of FT-N applied to this model for a simulation range of 50 periods and different values of α. In all cases described in the table, the Blanchard-Kahn conditions are satisfied and a unique solution exists. As is apparent, the greater is the effect of interest rates on aggregate demand, the slower is FT-N to converge; and for a range of values for which a solution does in fact exist, FT-N will fail to converge to that solution. It is worth noting, moreover, that the value of α used in the realistically calibrated model in Laxton, Rose and Tetlow (1993) falls in the range where FT-N fails to converge in these experiments.

The IDS-N procedure solves this problem in one iteration, regardless of the value of α.

Table 1
Convergence properties of FT-N for the simple linear model

Value of α	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70
Largest eigenvalue[a]	0.5	0.55	0.65	0.75	0.8	0.9	>1.0	>1.0
Number of iterations[b]	2	4	25	49	96	257	∞	∞

a. This is the modulus of the largest eigenvalue of the matrix *SU**, as defined in the discussion of equations (12)–(17) in Section 4. FT-N will converge if and only if this value is less than one.

b. We report the number of outer-loop iterations required before the absolute difference between the computed solution and the true solution is less than 0.001 for any variable at any time.

6.2 A dynamic, multisector model

This example is one involving a multisector, overlapping generations model of a small, open economy, calibrated to reflect the Canadian data, as described in Macklem (1993). The model has 109 equations. Although the model is non-linear, a linear approximation around its steady state satisfies the Blanchard-Kahn conditions. The model can be simulated without difficulty using IDS-N. However, it cannot be simulated reliably using FT-N. To obtain his reported results using FT-N, Macklem relaxed the outer-loop convergence criterion considerably relative to the usual setting. Although his reported solutions are close enough to the correct values over the horizon reported that the research conclusions are not seriously affected, our reworking of the problem suggests that Macklem's dynamic simulations generally diverge and are, hence, formally inconsistent with the reported long-run results from the steady-state version of the model. This finding appears to be robust to damping.[13]

13. In discussions with the authors, Macklem also confirmed that his FT-N solutions had long-lasting "nuisance" cycles that disappeared when the model was solved using IDS-N. This led to the discovery of a small coding error that had been obscured by the "noise" in the FT-N solution but which emerged clearly in the IDS-N results.

6.3 A relatively large model – QPM

The staff at the Bank of Canada have recently begun using a new computer simulation model of the Canadian economy for projections and policy analysis. That model is called QPM from Quarterly Projection Model (see Poloz, Rose and Tetlow 1994, for an overview of the model and its properties; see Black, Laxton, Rose and Tetlow 1994, for a discussion of the steady state of QPM; and see Coletti, Hunt, Rose and Tetlow 1994, for a more detailed discussion of some of the model's dynamic properties).

Two important features of this model are forward-looking behaviour by agents and dynamic stability around an explicit steady state with fully consistent stocks and flows. The model is also highly non-linear. It was predicted, during the model's developmental phase, that FT-N was likely to be inadequate in a production environment. Simulation times were long and convergence was often not obtained when no good (economic) reason could be found. The prediction turned out to be correct – the current, production version of QPM cannot be simulated reliably using the FT algorithm.

With IDS-N, QPM is routinely solved in a production environment. While QPM is of moderate size by the standards of applied macro models – it has about 320 equations – the stacked IDS version of the model contains over 32 000 equations.[14] Nevertheless, the simulation times are usually under half an hour on a Sun SPARC 10, model 52, computer, at least for solutions of shocks around an established control solution.

As an example of the type of problem that we faced using the FT-N algorithm, consider a shock to aggregate demand in QPM (similar to a perturbation to ε in equation 22 of the simple linear model). It turns out that for very small shocks, of order 0.1 per cent, say, both FT-N and IDS-N converge. With convergence criteria, for each variable at each time point, of the form

14. This is the size of the model when QPM is simulated over 25 years (100 quarters). This is the length of an average simulation – it gives the stocks time to approach their steady-state levels – but frequently even longer simulations are required, which generates much larger systems. In test runs, we have solved problems twice this size.

$$\frac{\left|x^{(k+1)} - x^{(k)}\right|}{max(\left|x^{(k+1)}\right|, \left|x^{(k)}\right|) + 5} < 0.001 \tag{26}$$

for FT-N, where $x^{(k)}$ is the kth iteration for variable x at some time point, 268 type II iterations are required to obtain convergence.[15] However, this does not produce a solution that is accurate within 0.1 per cent. To obtain that degree of accuracy, 306 iterations are required. After 268 iterations, where convergence is declared under criterion (26), the maximum error is twice as large as the supposed convergence tolerance. The solution is not changing by enough to fail the test, but it has not reached the desired region of the true solution.

IDS-N is also much faster than FT-N in this application.[16] The total simulation time is about one hour for FT-N, for declared convergence under criterion (26), compared with about 35 minutes with IDS-N.[17] Furthermore, to obtain even this standard of convergence, it is necessary to use an advantageous damping factor of 0.7 with FT-N; if the damping factor is greater than 0.8 or less than 0.4, FT-N diverges, even for this very small shock. With other values for the damping factor, FT-N converges for this shock, but takes longer than reported above.

FT-N does not converge for larger shocks. For example, when the above experiment was repeated using a 1 per cent shock, which is still not

15. This is a standard convergence criterion. It is a mix of relative and absolute criteria. The 5 in the denominator is used to prevent division-by-zero problems – it makes the convergence criterion approximate an absolute criterion for small values of x and a relative criterion for large values. This is the form used in the TROLL software (Hollinger 1988).

16. The times reported here were generated using an implementation of IDS-N developed at the Bank of Canada as a C program that sets up the stacked problem and calls the Newton simulator of Portable TROLL to provide the solution. Recently, the IDS-N algorithm has been integrated into TROLL itself, in what is called a “stacked time” procedure, which has greatly reduced simulation times. For example, a 100-period solution of QPM under our previous system solved in roughly 90 minutes. This has been cut to 30 minutes with the new TROLL implementation.

17. Recall, moreover, that for equivalent accuracy in the solution, further iterations with FT-N would be required.

large by any means, FT-N did not converge under any damping setting. IDS-N converged in the same amount of time as for the small shock.[18]

6.4 What do the examples show?

We have shown that FT-N fails to converge in some cases that are of practical interest. In Section 4, we established that, for linear systems, FT-N inner-loop convergence is guaranteed and that outer-loop convergence properties are independent of the method used to obtain inner-loop convergence. For linear systems, it therefore follows that if FT-N fails to converge, then so will all other variants of the FT method. It is clearly the iterative procedure applied at the outer loop that causes the problem in such cases. This argument applies without qualification to our linear example. Although we do not have parallel analytical results for non-linear systems, we consider it most unlikely that other variants of the FT procedure would converge, where FT-N does not, in our non-linear examples.

We cannot claim that this establishes that IDS-N dominates all other methods. Indeed, there is an ongoing debate on whether Newton-based methods are superior for solving economic models to Gauss-Seidel methods, for example. See Don and Gallo (1987) and Hughes Hallett and Fisher (1990) for two different views on this subject. An IDS-GS application could, in principle, converge faster than IDS-N, perhaps even in cases where FT-N fails. Although there is no guarantee that IDS-GS will converge on the solution, it could be faster if it does.

While we cannot rule out the possibility that special characteristics of particular economic models could make the superior robustness

18. It is worth noting that, for IDS-N, the choice of a convergence tolerance and the size of a shock are relatively unimportant because of the super-linear nature of Newton methods. When a simulation of QPM converges under IDS-N, the solution is typically not affected by tightening the convergence tolerance. Moreover, the time required is typically unaffected by either the tolerance or the size of the shock. This should not be taken to apply absolutely. The model is non-linear and there could be failure if the shock was large enough. In practice, we have not found this to be a problem. For example, for testing purposes we occasionally simulate, without difficulty, shocks amounting to 10 per cent of aggregate demand.

property of the Newton method redundant, our examples illustrate some of the risks of alternative iterative methods and the problems that can occur. Moreover, we would speculate that the high intertemporal simultaneity in economic models with forward-looking behaviour and expectations makes the robustness of the Newton method more than simply a nice feature to have. It is our view that this robustness will prove important in facilitating progress in research and development of policy simulation models.

7 CONCLUDING REMARKS

We have documented the convergence properties of a number of methods for solving linear models with rational expectations. In particular, we have shown that none of these methods, including the popular Fair-Taylor approach, is guaranteed to converge on the true solution, with the exception of what we have called the IDS-N procedure.

In IDS, each variable at each time point is treated as a different variable. In effect, time is removed from the problem by stacking the equations for each variable at each time point into what is potentially a very large matrix. In IDS-N, the complete stacked system is solved with a Newton method that exploits analytical derivatives. We have shown that this procedure is very robust; however, this advantage comes at the cost of increasing the scale of the matrix that must be inverted in simulating a model. With small models, this need not be of any special practical concern. However, the size of the matrix expands very rapidly with the size of the model and the number of time periods that must be simulated.

For work with the QPM model, we deal regularly with simulations that require over 30 megabytes of memory, primarily for the matrix inversion. For this procedure to be practicable, efficient sparse-matrix inversion procedures are essential. Moreover, the method would not be as attractive if insufficient core memory was available to hold the full problem. If the problem must be solved with swapping to disk during the inversion, performance deteriorates substantially. On the other hand, IDS can be used on pieces of the problem, with an iterative procedure providing the overlap.[19]

---

19. In work with the Federal Reserve Board's multicountry model, Faust and Tryon (1994) solve individual country blocks and then iterate across blocks to obtain full convergence. This work uses the FT algorithm. Staff at the International Monetary Fund have informed us that IDS-N solves individual country blocks of their multicountry model, Multimod, about 24 times faster than FT-N, but that FT-N appears faster on the complete model. They have therefore implemented a procedure where IDS-N is used to solve the individual country blocks, and an iterative process is applied across the blocks. We have not been able to find any parallel results for solving QPM. Application of IDS-N to the complete problem seems to be the fastest method. This is understandable, since there is a high level of simultaneity within a country model, whereas the limited links between countries in a multicountry application may define an efficient partition of the full problem.

Although efficient sparse-matrix inversion procedures have been available for some time, the IDS-N idea has become feasible only recently, with advances in computer technology and the decline in the price of large-scale memory. Iterative procedures, like FT, require much less memory to run. However, it is important to understand that they are not guaranteed to solve even linear models for which a solution does exist. We think that this may have inhibited progress in developing policy simulation models. The QPM model of the Canadian economy, a highly non-linear model with forward-looking behaviour, is solved routinely with IDS-N. Based on our experience, the model simply could not be used if we were restricted to Fair-Taylor procedures.

Of course, this conclusion for QPM may not carry over to other models. It is certainly possible that in particular cases an iterative procedure would converge and that, owing to the large overhead of a Newton procedure, would run faster. However, we would argue that there is great danger in a methodology that can appear to indicate a modelling problem when there is none, or worse, that may encourage researchers to discard, as uninteresting, parameterizations of their models that result in slow solutions or systematic simulation failures, when such parameterizations may, in fact, be suggested by the data.

REFERENCES

Black, R., D. Laxton, D. Rose and R. Tetlow. 1994. *The Steady-State Model: SSQPM*. The Bank of Canada's New Quarterly Projection Model, Part 1. Technical Report No. 72. Ottawa: Bank of Canada.

Blanchard, O. J. and C. M. Kahn. 1980. "The Solution of Linear Difference Models Under Rational Expectations." *Econometrica* 48 (July): 1305–11.

Buiter, W. H. 1984. "Saddlepoint Problems in Continuous Time Rational Expectations Models: A General Method and Some Macroeconomic Examples." *Econometrica* 52 (May): 665–80.

Coletti, D., B. Hunt, D. Rose and R. Tetlow. 1994. "Some Dynamic Properties of QPM." Paper prepared for a presentation on the Bank of Canada's new Quarterly Projection Model at the meetings of the Canadian Economics Association, University of Calgary, Calgary, Alberta.

Dodes, I. A. 1978. *Numerical Analysis For Computer Science*. New York: North Holland.

Don, F. J. H. and G. M. Gallo. 1987. "Solving Large Sparse Systems of Equations in Econometric Models." *Journal of Forecasting* 6: 167–80.

Fair, R. C. and J. B. Taylor. 1983. "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models." *Econometrica* 51 (July): 1169–85.

Faust, J. and R. Tryon. 1994. "A Distributed Block Approach to Solving Near-Block-Diagonal Systems with an Application to a Large Macroeconometric Model." Board of Governors of the Federal Reserve System, Washington. Unpublished manuscript.

Fisher, P. 1992. *Rational Expectations in Macroeconomic Models*. Dordrecht (The Netherlands): Kluwer Academic Publishers.

Fisher, P. G. and A. J. Hughes Hallett. 1988. “Efficient Solution Techniques for Linear and Non-linear Rational Expectations Models.” *Journal of Economic Dynamics and Control* 12: 635–57.

Fisher, P. G., S. Holly and A. J. Hughes Hallett. 1986. “Efficient Solution Techniques for Dynamic Non-Linear Rational Expectations Models.” *Journal of Economic Dynamics and Control* 10: 139–45.

Hall, S. G. 1985. “On the Solution of Large Economic Models with Consistent Expectations.” *Bulletin of Economic Research* 37 (May): 157–61.

Hollinger, P. 1988. *TROLL Program: Forward-Looking Simulator.* Needham (MA): Intex Solutions.

Holly, S. and A. Hughes Hallett. 1989. *Optimal Control, Expectations and Uncertainty.* Cambridge (Eng.): Cambridge University Press.

Holly, S. and M. B. Zarrop. 1983. “On Optimality and Time-Consistency When Expectations Are Rational.” *European Economic Review* 20 (January): 23–40.

Hughes-Hallett, A. J. 1981. “Some Extensions and Comparisons in the Theory of Gauss-Seidel Iterative Techniques for Solving Large Equation Systems.” In *Proceedings of the 1979 Econometric Society Meeting: Essays in Honour of Stefan Valvanis*, edited by E. G. Charatsis. Amsterdam: North-Holland.

Hughes Hallett, A. J. and P. G. Fisher. 1990. “On Economic Structures and Model Solution Methods: Or Should Econometricians Use Newton Methods for Model Solution?” *Oxford Bulletin of Economics and Statistics* 52 (August): 317–30.

Laxton, D., D. Rose and R. Tetlow. 1993. *Monetary Policy, Uncertainty and the Presumption of Linearity.* Technical Report No. 63. Ottawa: Bank of Canada.

Lipton, D., J. Poterba, J. Sachs and L. Summers. 1982. "Multiple Shooting in Rational Expectations Models." *Econometrica* 50 (September): 1329–33.

Macklem, R. T. 1993. "Terms-of-Trade Shocks, Real Exchange Rate Adjustment, and Sectoral and Aggregate Dynamics." In *The Exchange Rate and the Economy: Proceedings of a Conference Held at the Bank of Canada, 22–23 June 1992.* Ottawa: Bank of Canada.

Poloz, S., D. Rose and R. Tetlow. 1994. "The Bank of Canada's New Quarterly Projection Model (QPM): An Introduction." *Bank of Canada Review* (Autumn): 23–38.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge (MA): Cambridge University Press.

Ralston, A. 1965. *A First Course In Numerical Analysis*. New York: McGraw-Hill.

Reid, J. K. 1977. "Sparse Matrices." In *The State of the Art in Numerical Analysis*, edited by D. Jacobs, 85–146. London: Academic Press.

Rheinboldt, W. C. 1974. *Methods for Solving Systems of Nonlinear Equations*. Philadelphia: Society for Industrial and Applied Mathematics.

Roberts, S. M. and J. S. Shipman. 1972. *Two-Point Boundary Value Problems: Shooting Methods.* New York: American Elsevier Publishing Co.