NBER WORKING PAPER SERIES

FIML ESTIMATION OF RATIONAL DISTRIBUTED LAG STRUCTURAL FORM MODELS

Kent D. Wall*

Working Paper No. 77

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

March 1975

Preliminary: not for quotation

*NBER Computer Research Center. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

Abstract

The Rational Distributed Lag Structural Form (RSF) representation of an econometric model is introduced and its relationship to several standard forms of representation discussed. The FIML estimation problem for the RSF is then considered and formulated as a nonlinear, unconstrained optimization problem. A solution to the relation optimization problem is then obtained by an application of the Davidon-Fletcher-Powell variable metric method using simple first difference approximations for the necessary gradients. This approach requires a minimum of effort on the part of the model builder since there is no longer any need to analytically determine, and then program, the gradient expressions. The feasibility of the method is demonstrated with several examples.

Acknowledgments

The author wishes to thank Ms. Sally Donovan and Mr. Berc Rustem, both of the Programme of Research into Econometric Methods at Imperial College (University of London), for their help in the initial coding and testing of the computer program used in the examples.

Contents

Figures

1. INTRODUCTION

The rapid advances made in recent years in digital computer technology have provided the econometrician with computational capabilities unheard of a scant ten years ago. The researcher in the field of economics is learning to take advantage of these developments as witnessed by the increasing influence of the computer in the development of economic and econometric methodology. Larger and more nonlinear econometric models are now treated as almost passé, especially with respect to forecasting and simulation. Similar developments are now taking place in the area of parameter estimation [12], [13], [17], and [23]. Systems of linear equations with linear parameterizations have been considered, together with a relaxation of the uncorrelatedness of the error process to allow special first or second order autoregressive structure. In each case the parameter estimation was formulated as a maximization problem and the computer used in conjunction with some sort of function minimization algorithm. However, more powerful and efficient function minimization algorithms are available; and nonlinear parameterizations yielding very general and more flexible model representations can be employed without exceeding the computational capabilities of the modern digital computer. It is these considerations which provide the motivation for this paper.

The purpose of this paper is twofold: First, the Rational Distributed Lag Structural Form (RSF) representation is to be introduced. This one representation, while containing a nonlinear parameterization of the model, is very general and encompasses all standard linear, stationary, constant coefficient econometric models as special cases. In addition, the RSF allows for a very flexible class of error processes which admit almost all types of

recast as a nonlinear unconstrained optimization problem and the variable metric method applied to obtain an iterative solution. The details of the implementation, especially with respect to the residual generation, are also discussed. Section 5 contains an exposition of several numerical examples which illustrate the method and illuminate its applicability to econometric problems. Finally, Section 6 contains a summary of the main points of the paper and some conclusions.

2. THE RATIONAL STRUCTURAL FORM (RSF) REPRESENTATION

The RSF representation can be characterized in two ways. First, the general form of each individual equation in the model may be presented. This description focuses attention on the precise way each of the variables (endogenous and exogenous) interact with one another, and is useful in discussions concerned with the computational problems involved in FIML estimation. It is also useful in discussing the relationship between the RSF and other single equation econometric models already in wide use. Second, the entire system of equations may be considered as a whole and presented in vector matrix notation. This description is analogous to the more standard forms of simultaneous econometric models and, although of less computational utility, is of value in examining the relationship between the RSF and the various standard forms of simultaneous model representations. A familiarity with both characterizations contributes to a fuller understanding of the implications of the RSF representation; therefore both are presented here.

The exact form of each rational operator is quite arbitrary, being solely determined by the degrees of each polynomial involved, i.e., by $\{\rho_{ij}, \sigma_{ij}, \delta_{ij}\}$, $\{R_{ij}, S_{ij}, D_{ij}\}$, $\{P_i, Q_i\}$. The leading coefficients of the $\alpha_{ij}(L)$, $a_{ij}(L)$, and $d_i(L)$ are fixed at unity in order to satisfy the conventional normalization rule.[3]

By "stacking" each equation of the model, as given in (2.1) - (2.3), on top of one another and resorting to vector-matrix notation, the RSF may alternately be written as[4]

$$(2.4) \qquad T(L)y_t = U(L)x_t + \underline{V}(L)\underline{e}_t,$$

$$\underline{V}(L) = \begin{bmatrix} I & 0 \\ 0 & V(L) \end{bmatrix},$$

$$\underline{e}_t' = [0 \mid e_t'],$$

$$I = G_{id} \times G_{id} \text{ identity matrix}$$

where y_t is now a G x 1 vector of endogenous variables observed at time t, x_t is a K x 1 vector of exogenous variables observed at time t, and e_t is a G_b x 1 vector of random distrubances at time t. The rational matrix operators T(L), U(L), and V(L) are dimensioned respectively as G x G, G x K, and G_b x G_b. In view of (2.1) and (2.2), it is clear that the ij^{th} elements of these matrices are given by

$$[T(L)]_{ij} = \begin{cases} 1 & ; i=j \\ \dfrac{-\beta_{ij}(L)}{\alpha_{ij}(L)} L^{\delta_{ij}} & ; i \neq j \end{cases}$$

$$[U(L)]_{ij} = \frac{b_{ij}(L)}{a_{ij}(L)} L^{D_{ij}} \quad ; \text{all } i, j$$

$$[V(L)]_{ij} = \begin{cases} \dfrac{c_i(L)}{d_i(L)} & ; i=j \\ 0 & ; i \neq j \end{cases}$$

2.1 Single Equation Models (G=1)

In the univariate model the first subscript, i, on the left-hand variable becomes immaterial so it is dispensed with. Moreover, all of the right-hand side variables becomes exogenous so that the first term on this side in (2.2) may be dropped. Thus (2.2) reduces to

$$(2.5) \qquad y_t = \sum_{j=1}^{K} \frac{b_j(L)}{a_j(L)} L^{D_j} x_{jt} + \frac{c(L)}{d(L)} e_t + k.$$

(i) *The Univariate Time Series (ARMA) Model.* By specifying K=0 all purely exogenous variables are excluded leaving

$$y_t = \frac{c(L)}{d(L)} e_t + k$$

or,

$$d(L)\, \tilde{y}_t = c(L)\, e_t$$

where $\tilde{y}_t \triangleq y_t - k$. This is just the autoregressive-moving average univariate time series model treated so extensively by Box & Jenkins [6], Phillips [29], and Aigner [1].

(ii) *The Univariate Autoregressive-Moving Average Model with Exogenous Inputs (ARMAX).* By specifying K=1 and c(L) = d(L) = 1 there results

$$\tilde{y}_t = \frac{b(L)}{a(L)} x_{t-D} + e_t.$$

This is the ARMAX model treated by Jorgenson [24], Steiglitz & McBride [35], and Dhrymes, Klein, & Steiglitz [15] (where reference to many other works are cited).

(i) *The Polynomial Structural Form (PSF).* Perhaps the most widely investigated simultaneous model is that which may be classed as a polynomial structural form. This is obtained as a special case of (2.4) by constraining each equation of the model to have a polynomial form similar to that of (2.6). In this situation each equation in (2.2) is constrained such that $\alpha_{ij}(L) = a_{ij}(L) = d_i(L) \triangleq a_i(L)$, i.e., the denominator polynomials <u>across</u> any one equation are identical, but not <u>between</u> equations. Thus each equation of (2.4) may be written as

$$a_i(L)y_{it} = \sum_{\substack{j=1 \\ j\neq i}}^{G_i} \beta_{ij}(L)L^{\delta_{ij}} y_{jt} + \sum_{j=1}^{K_i} b_{ij}(L)L^{D_{ij}} x_{jt} + c_i(L)e_{ij} + k_i.$$

where $a_i(L) \triangleq 1 + a_i^1 L + \ldots + a_i^{R_i} L^{R_i}$. If all the coefficients of like powers of L are collected together in coefficient matrices then (2.4) takes on the special PSF form:

$$A(L)y_t = B(L)x_t + C(L)e_t + \kappa \tag{2.7}$$

where κ is G x 1 vector of constants and,

$$A(L) = \sum_{k=0}^{R} A_k L^k \quad , \quad \det A_0 \neq 0, \quad R \triangleq \max_{i,j} \{R_i, \delta_{ij}+\sigma_{ij}\}$$

$$B(L) = \sum_{k=0}^{S} B_k L^k \quad , \quad S \triangleq \max_{i,j} \{S_i, D_{ij}+S_{ij}\}$$

$$C(L) = \sum_{k=0}^{Q} C_k L^k \quad , \quad C_0 = I \quad , \quad Q \triangleq \max_{i} \{Q_i\}.$$

The interpretation of the $\{A_k, B_k$ and $C_k\}$ in terms of the coefficients of the polynomials $a_i(L)$, $\beta_{ij}(L)$, $b_{ij}(L)$ and $c_i(L)$ is somewhat complicated because of the pure delays, $L^{\delta_{ij}}$ and $L^{D_{ij}}$. However, in general, it may be

(iv) *The Final Form (FF).* Returning to (2.4) in its original rational description, it is seen that by specifying T(L)=I the traditional final form (FF) model is obtained,[5]

$$y_t = U(L)\, x_t + V(L)e_t$$

A model such as the above could be estimated directly, however the extremely complicated and dense nature of U(L) might make estimation impracticable. Thus both T(L) and U(L) are estimated (each usually rather sparse), and the equivalent FF obtained by inversion (when the inverse exists) of T(L) and premultiplication of (2.4) by this inverse:

$$y_t = T^{-1}(L)U(L)x_t + T^{-1}(L)V(L)e_t.$$

In terms of polynomial form models, the FF is computed by inversion of A(L) (when it exists) and premultiplication of (2.7) by $A^{-1}(L)$:

$$y_t = A^{-1}(L)B(L)x_t + A^{-1}(L)C(L)e_t.$$

Note that now a rational form has been obtained.

(v) *The Multivariate Time Series Model.* With the model (2.4) or (2.7) it becomes possible to talk of multivariate time series analysis - by excluding all exogenous explanators. Thus, with U(L)=0 and no identities there results

$$T(L)y_t = V(L)e_t.$$

Since V(L) has been constrained in (2.4) to be a diagonal operator, T(L) must be included if it is desired to remove any inter-equation correlation in the residuals. A similar model may be constructed using the polynomial forms above. Namely,

$$A(L)y_t = C(L)e_t$$

econometric model are the numerical values for the parameters. These are provided by the solution of an estimation problem, the formulation of which is the subject of the next section.

3. THE FIML ESTIMATION PROBLEM

The parameter estimation problem associated with the RSF representation (2.1)-(2.2) will be approached using the method of maximum likelihood. This requires reality to approximate an idealized environment characterised by several assumptions concerning the nature of the model as well as the dominating statistical properties of the process being modelled. Since the success of the method depends on the relative validity of each of these underlying assumptions, they will be briefly discussed in order that the limitations inherent in the proposed approach be more fully understood. The conditions guaranteeing the uniqueness and consistency of the estimates are then reviewed. Finally, the likelihood function is constructed, and the estimation problem stated in precise mathematical terms.

3.1 Basic Assumptions

There are six underlying assumptions governing the success of the FIML estimation method described below:

(A.1) *Stationarity.* It is assumed that all the random distrubances acting on the model be representable as at least wide-sense (covariance) stationary stochastic processes. If nonstationary disturbances are found to occur, then it is assumed that stationarity can be induced by some simple data transformation such as differencing or exponential trend removal.

the economic specification together with the statistical properties of the data guarantee the existence of a solution. However, the uniqueness and consistency of the estimates cannot be established without the satisfaction of conditions concerned with the structure specification and admissible parameter values:

(C.1) *Identification.* The parameterization of the model, implicit in the structure specification, must be such that no other parameterization is observationally equivalent [32]. For models specified in the PSF, a complete set of identification conditions has been given by Hannan [21], [22]. Conditions for the RSF specification have not yet been established in the global sense of Hannan's results for the PSF, however, local results are available using the information matrix (see the work of Rothenberg [32]). For the present work, simple sufficient conditions, requiring the absence of polynomial factor cancelations across any equation or between equations in (2.4), will be imposed.

(C.2) *Stability.* The parameters of the model must be such that the solution of (2.4) for y_t remains finite for finite values of x_t and e_t as $t \to \infty$. This is equivalent to requiring that the poles of $\det(T^{-1}(L)U(L)) = 0$ and $\det(T^{-1}(L)\underline{V}(L)) = 0$ lie outside the unit.[7]

(C.3) *Inverse-Stability (Invertibility).* The parameters of the model must be such that the solution of (2.4) for e_t remains finite for finite values of y_t and x_t as $t \to \infty$. This is equivalent to requireing that the poles of $\det(\underline{V}^{-1}(L)T(L)) = 0$ and $\det(\underline{V}^{-1}(L)U(L)) = 0$ lie outside the unit circle.[7]

In effect, conditions (C.2) and (C.3) restrict the set of admissible parameters to some subset of the parameter space. Specifically, assume that there are a total of m unknown parameters for a particular RSF

The likelihood function for the estimation of the RSF model is obtained by constructing the joint probability density function of the observed output sequences, conditioned on the observed exogenous variables and the model parameters. As a matter of notation define

$$Y_t = [y_1, y_2, \cdots, y_t]$$

and

$$X_t = [x_1, x_2, \cdots, x_t].$$

Then the joint density function may be constructed recursively according to

$$\begin{aligned} p(Y_T|\theta, X_T) &= p(y_1, y_2, \cdots, y_T|\theta, X_T) \\ &= p(y_T|Y_{T-1}, \theta, X_T) \cdot p(Y_{T-1}|\theta, X_T) \\ &= \prod_{t=1}^{T} p(y_t|Y_{t-1}, \theta, X_T). \end{aligned}$$ [9]

Since e_t is assumed to be normally distributed, and since y_t may be expressed as a linear operation on e_t, y_t will also be normally distributed. Therefore, the required joint density function may be constructed once expressions can be determined for the conditional mean of the y_t process,

$$y_{t|t-1} = E\{y_t|Y_{t-1}, \theta, X_T\},$$

and the conditional variance-covariance matrix,

$$S_{t|t-1} = E\{[y_t - y_{t|t-1}][y_t - y_{t|t-1}]'\} = E\{\tilde{y}_{t|t-1}\tilde{y}'_{t|t-1}\}.$$

The derivation of the conditional mean, or equivalently, the one-period prediction of y_t, is extremely cumbersome using the representation of (2.4). A much more lucid development, equally valid, can be obtained using the equivalent PSF representation (2.7).[10] In terms of this

it is not difficult to conclude that,

$$(3.4) \qquad A_0 y_{t|t-1} = -\sum_{k=1}^{R} A_i\, y_{t-k} + \sum_{k=0}^{S} B_i\, x_{t-k} + \sum_{k=1}^{Q} C_i\, e_{t-k|t-1}.$$

The variance-covariance function of this estimate is computed using the expression for the one-period prediction error obtained by subtracting (3.4) from (3.2):

$$(3.5) \qquad S_{t|t-1} = A_0^{-1} R (A_0')^{-1} = S.$$

The likelihood function can now be given by combining the expression for $p(Y_T|\theta, X_T)$ with (3.4) and (3.5):

$$(3.6) \qquad L(\theta) = \prod_{t=1}^{T} \frac{|\det A_0|}{(2\pi)^{G/2}(\det R)^{1/2}} \exp\{-\frac{1}{2}\, \tilde{y}'_{t|t-1}\, S^{-1}\, \tilde{y}_{t|t-1}\}$$

where $L(\theta)$ has been used in place of $p(\cdot|\cdot)$ to emphasize the dependence on θ once Y_T and X_T have been observed.

3.4 The FIML Estimation Problem

In terms of the actual computation of estimates for θ, it is more convenient to work with the negative of the natrual logarithm of $L(\theta)$, denoted by $\ell(\theta)$:

$$\ell(\theta) = \frac{GT}{2} \ln(2\pi) + \frac{T}{2} \ln(\det R) + \frac{1}{2} \sum_{t=1}^{T} \tilde{y}'_{t|t-1}\, S^{-1} \tilde{y}_{t|t-1} - T \ln(|\det A_0|).$$

The maximum likelihood estimate for θ will be that value of θ which <u>minimizes</u> $\ell(\theta)$.

unknown parameters ϕ are then recovered with the aid of (3.7) using the estimated residual sequence $\{e_{t|t};\ 1 \leq t \leq T\}$ evaluated at $\theta = \hat{\theta}$.

Problem (P.1) constitutes a constrained parameter optimization problem which can be solved numerically using the theory of mathematical programming. Before discussing such a solution, the subject of identities is reconsidered.

3.5 The Inclusion of Identities

Suppose the first G_{id} equations of the model now represent identities (2.1). Then it would be possible to eliminate these exact equations, via substitution, in order to obtain a reduced system of G_b equations. The previous results could then be used exactly as they stand. However, this tedious substitution is unnecessary: The expression (3.9) can still be used if R is interpreted as the $G_b \times G_b$ variance-covariance matrix of the $G_b \times 1$ vector e_t entering into only the behavioral equations (2.2). Whereas before both A_o and R were $G_b \times G_b$, now A_0 is G x G and R is $G_b \times G_b$. These observations can be substantiated as follows.

Partition the orgiinal specification of G equations, as given in its PSF representation (3.4), according to whether an equation is behavioral or definitional. Then the structure matrix will take the following form

$$A_o = \left[\begin{array}{c|c} A_{011} & A_{012} \\ \hline A_{021} & A_{022} \end{array}\right]$$

where A_{011} is $G_{id} \times G_{id}$ and A_{022} is $G_b \times G_b$. The elimination of the identities via substitution leads to a reduced system of G_b equations similar to (3.4),

4. NUMERICAL SOLUTION OF THE ESTIMATION PROBLEM

The general constrained sequential minimization of $J(\theta)$ is almost impossible to solve because of the generally nonlinear form of the restrictions on θ which define A. If only linear restrictions existed (which would be the case if only first order lags appeared in the model), then a true nonlinear programming solution could be obtained by coupling an efficient optimization algorithm to some constraint incorporation technique such as Rosen's gradient projection method [19]. Because a solution to the general problem is sought, such specialized constrained methods can not be used. Instead, it is suggested that the restrictions on θ be incorporated only in an implicit fashion - that the actual hill-climb on $J(\theta)$ be executed in a completely unconstrained fashion, with the progress and final results monitored by the researcher to determine if any constraints have been violated.

There are several reasons justifying the relaxation of the constraints imposed by conditions (C.2)-(C.3). First, and most important, the stationarity assumption should guarantee the existence of stable and invertible parameter estimates if the model has been adequately specified. Second, if unstable and/or uninvertible estimates arise, the residual series generation will become unbounded and result in numerical overflows which automatically terminate the estimation. Thus, an indication that something is wrong with the model is automatically provided. In other words, the need to explicitly impose stability and invertibility constraints implies misspecification; i.e., violation of the underlying assumptions or hypotheses upon which the model has been based.

$$\theta^{k+1} = \theta^k + \delta\theta^k = \theta^k - \alpha^k H(\theta^k)\, J_\theta(\theta^k)$$

where α^k is a scalar "step size", $H(\theta^k)$ an approximation to the inverse Hessian, $J_{\theta\theta}$, evaluated at θ^k, and $J_\theta(\theta^k)$ is the gradient of J evaluated at θ^k. Each improvement of θ^k constitutes an iteration, and these iterations proceed until covergence criteria are met, i.e., until

$$||H(\theta^k)J_\theta(\theta^k)|| \leq \epsilon_1$$

or,

$$||\delta\theta^k|| \leq \epsilon_2 .$$

The description of the exact details is beyond the scope of this paper and the interested reader is directed to works dealing solely with this algorithm [14], [18], [19]. The algorithm automatically computes x^k and $H(\theta^k)$ - the user is only required to supply the convergence parameters ϵ_1, ϵ_2, and the expressions for $J(\theta)$ and $J_\theta(\theta)$.

The major demand placed upon the researcher by the algorithm is the need to evaluate $J_\theta(\theta)$. For simple models analytic expressions can be derived and programmed with ease, but the desire to estimate a general RSF makes such derivation and coding extremely involved. It is possible to generate the required gradient vector by numerical differencing, if care is taken in choosing the form of differencing and in the parameter perturbations so that numerical accuracy is preserved. A very rapid and trivially implemented gradient generation scheme can be obtained using a first difference approximation to the partial derivatives.[11] In particular, if $\partial J/\partial\theta_i$ denotes the i^{th} component of J_θ then

Initialization. This operation consists of the bookkeeping necessary to keep track of all the information provided as input to the program. This includes the economic specification, the structure as defined by the integer set Π, and the initial guess for $\hat{\theta}$. In practice the required accuracies of the final estimate, ε_1 and ε_2, are input at this time. Experience has shown that $\varepsilon_1 = \varepsilon_2 = 10^{-4}$ is sufficient for most problems.

DFP. This operation contains the actual Davidon-Fletcher-Powell algorithm. In the process of execution, repeated calls are made to the external function which define the values of $J(\theta)$ and $J_\theta(\theta)$ given θ.

Function and Gradient Evaluation. This operation is called upon by DFP and evaluates $J(\theta)$ according to (3.9) given θ and the observed data series $\{y_t;\ 1 \leq t \leq T\}$ and $\{x_t;\ 1 \leq t \leq T\}$. When a request is made for $J_\theta(\theta)$, this operation repeatedly computes $J(\theta)$ for each perturbation of θ and evaluates (4.1). In order to evaluate (3.9) the estimated residual series $\{e_{t|t};\ 1 \leq t \leq T\}$ must be provided. This is accomplished by calling upon the simulation operation.

Simulation. The function of this operation is to solve the system of equations representing the model for the estimated residual series. The implementation of this operation is crucial, and therefore is discussed separately in the following subsection.

Output. This operation is entered after exit from the DFP algorithm. The results of the iterative minimization are used to produce the final parameter estimates, their estimated standard errors, and their estimated variance-covariance matrix. In order to obtain these statistics, the

However, the PSF representation (3.2) was only employed to expedite the derivation of the likelihood function, and the construction of the structure matrix, A_0, in particular. In general, the use of (3.2) is to be avoided for two reasons. First, this approach requires additional work in converting the general RSF representation to its equivalent PSF representation before a solution for $e_{t|t}$ is embarked upon. Second, and more important, such a procedure applied to the estimation of RSF representation parameters leads to a highly complex set of constraints between the estimated parameters of the PSF.[12] The general coding to incorporate these constraints is prohibitive for all but the simplest models. It is far easier, and more efficient, to utilize the original RSF specification.[13]

The solution of the RSF representation defined by (2.1) - (2.2) for $e_{t|t}$ can be achieved simply and rapidly (on a digital computer) owing to the special structure inherent in the specification. In particular, the presence of one, and only one, random disturbance term in each behavioral equation permits the computation of the entire vector sequence for $e_{t|t}$ to be carried out one element at a time. Thus, for each i ($G_{id}+1 \leq i \leq G$) all that is required is the computation of $e_{i,t|t}$ for $1 \leq t \leq T$ where

$$(4.5) \qquad e_{i,t|t} = \frac{d_i(L)}{c_i(L)} \tilde{y}_{it}$$

$$(4.6) \qquad \tilde{y}_{it} = y_{it} - \sum_{j=1}^{G_i} \frac{\beta_{ij}(L)}{\alpha_{ij}(L)} L^{\delta_{ij}} y_{jt} - \sum_{j=1}^{K_i} \frac{b_{ij}(L)}{a_{ij}(L)} L^{D_{ij}} x_{jt}$$

$$= y_{it} - \sum_{j=1}^{G_i} \hat{y}^j_{it} - \sum_{j=1}^{K_i} \hat{y}^j_{it}$$

$$= y_{it} - \hat{y}_{it}.$$

This in turn requires the solution of a succession of K_i+G_i+1 simple single input-single output rational lag models: (i) G_i solutions for each endogenous

use of zero starting values introduces errors in the residual computation, however the invertibility condition together with sufficiently long data series insures that these errors remain negligible.[14]

Consider next the problem of solving for the forecasted output of the univariate rational transfer function model

$$\hat{y}^{j}_{it} = \frac{b_{ij}(L)}{a_{ij}(L)} L^{D_{ij}} x_{jt}.$$

This problem is equivalent to solving the difference equation:

$$\hat{y}^{j}_{it} = -\sum_{k=1}^{R_{ij}} a^{k}_{ij} \hat{y}^{j}_{i,t-k} + \sum_{k=0}^{S_{ij}} b^{k}_{ij} x_{j,t-k-D_{ij}} \tag{4.8}$$

for $1 \leq t \leq T$. As before, a starting value problem arises due to the lagged values of $\hat{y}^{j}_{i,t-k}$ and $x_{j,t-k-D_{ij}}$. Since x_{jt} is observed, its starting problem can be solved similar to that for $\tilde{y}_{it}$ above; the solution for $\hat{y}^{j}_{it}$ is started at $t = D_{ij}+S_{ij}$ instead of $t = 1$. The method of determining starting values for $\hat{y}^{j}_{it}$ is slightly more complex since it depends on whether there are more than one right-hand side input terms in (2.2). In the former case (one observed exogenous input) it is easy to see that if the solution is started at $t = R_{ij}$ then

$$\hat{y}^{j}_{i,t-k} = E\{y^{j}_{i,t-k} | Y_{R_{ij}}\} = y_{i,t-k}$$

for $1 \leq k \leq R_{ij}$. Thus starting the solution process at t_{oi} where

$$t_{oi} = \max\{R_{ij}, D_{ij}+S_{ij}\}$$

permits the use of observed quantities for the starting variables. In the latter case (more than one exogenous input) the forecasted output will be the sum of the forecasted outputs for each separate transfer function term,

of $e_{t/t}$ which begins at a different time period. The inconvenience of this "ragged" origin can be easily overcome by specifying a common origin over all behavorial equations according to

$$(4.10) \qquad t_o = \max_i \{t_{oi}\}$$

The methods described above can be immediately applied to the identities of the model, (2.1), in exactly the same manner. The only difference arises from the absence of the error term - the steps involved with solution of (4.5) or (4.7) are not required. The entire simulation operation described above may be summarized by the following outline.

Simulation Operation.

1. for $1 \leq i \leq G_{id}$ solve (2.1) for $y_{it} = \hat{y}_{it}$ over $t_o \leq t \leq T$:
   (a) for $1 \leq j \leq G_i$ solve (4.9) with zero starting values for $\hat{y}^j_{it}$ and observed values for y_{jt} $(j \neq i)$.
   (b) for $1 \leq j \leq K_i$ solve (4.8) with zero starting values for $\hat{y}^j_{it}$ and observed values for x_{jt}.
   (c) for $t_o \leq t \leq T$ compute $\hat{y}_{it}$ by summing the $\hat{y}^j_{it}$ and store results as y_{it} for use in the behavioral equations.
2. for $G_{id} + 1 \leq i \leq G$ solve (2.2) for $e_{i,t|t}$ over $t_o \leq t \leq T$:
   (a) for $1 \leq j \leq G_i$ solve (4.9) with starting values as above
   (b) for $1 \leq j \leq K_i$ solve (4.8) with starting values as above
   (c) for $t_o \leq t \leq T$ compute $\tilde{y}_{it}$ from (4.6)
   (d) for $t_o \leq t \leq T$ solve (4.7) with zero starting values for $e_{i,t|t}$ and store in $e_{t|t}$ for use in forming $J(\theta)$.

The model was first estimated over the entire set of observations $1 \leq t \leq 75$ with an initial parameter guess

$$\theta' = [-0.4,\ 0.1,\ -0.4,\ -0.1,\ 0.0,\ 0.0]$$

after 13 iterations the minimization of $J(\theta)$ converged to a stationary point with each element of the gradient vector on the order of 10^{-5}. The resulting model was:

$$(5.1) \qquad y_{1t} = \underset{(\pm.158)}{\frac{0.762}{1-0.618L}}\, x_{1t} - \underset{(\pm.043)}{\frac{1.076}{1-0.787L}}\, x_{2t} + \underset{(\pm.131)}{\frac{1+0.435L}{1-0.503}}\, e_{1t}$$

where the estimated standard errors of the parameters are given in parentheses. This result was obtained assuming zero starting values wherever required in the residual simulation. Since the entire observation interval was used, this gives an indication of the model which best fits the sample data and does not result from any errors due to unknown starting values.

The model was then re-estimated using the same initial θ only over a shorter observation interval $5 \leq t \leq 75$. This situation introduces starting value errors since now any unknown variables required as lagged inputs from $t = 1,2,3,4$ will be arbitrarily set to zero. After 12 iterations the algorithm converged to a stationary point of $J(\theta)$ with a gradient whose elements were each on the order of 10^{-5}. The resulting model was:

$$(5.2) \qquad y_{1t} = \underset{(\pm.161)}{\frac{0.775}{1-0.571L}}\, x_{1t} - \underset{(\pm.044)}{\frac{1.083}{1-0.790L}}\, x_{2t} + \underset{(\pm.119)}{\frac{1+0.408L}{1-0.514L}}\, e_{1t}.$$

These estimates are easily within one standard error of the estimates of (5.1), and both are at least within two standard errors of the true model. The stationary point represented by (5.2) was found to be unique in the

when an equation was estimated assuming uncorrelated errors. In this respect, two of the equations were found to have moving average errors, perhaps as a result of the use of first differences, whereas Chow and Fair assume only autoregressive errors. The single equation estimates were as follows (with the estimated standard errors of each parameter appearing in parentheses):

$$(5.3)\qquad \Delta GNP_t = \Delta CD_t + \Delta CN_t + \Delta CS_t + \Delta IP_t + \Delta IH_t + \Delta^2 V_t - \Delta IMP_t + \Delta G_t$$

$$(5.4)\qquad \Delta Z_t = \Delta CD_t + \Delta CN_t$$

$$(5.5)\qquad \Delta CD_t = \underset{(\pm.016)}{0.136}\ \Delta GNP_t + \underset{(\pm.049)}{[0.137} + \underset{(\pm.048)}{0.058L}]\ \Delta MOOD_{t-1} \underset{(\pm1.323)}{-2.593}\ D_{644t} + \underset{(\pm1.103)}{3.611}\ D_{651t} + \underset{(\pm.139)}{[1-0.398L]}\ e_{1t}$$

$$(5.6)\qquad \Delta CN_t = \frac{\underset{(\pm.118)}{0.073}}{1-0.677L}\ \Delta GNP_t \underset{(\pm.022)}{-\ 0.104}\ \Delta MOOD_{t-2} + \frac{\underset{(\pm.0557)}{1}}{1-0.672L}\ e_{3t}$$

$$(5.7)\qquad \Delta CS_t = \frac{\underset{(\pm.067)}{0.073}}{1-0.672L}\ \Delta GNP_t - \frac{\underset{(\pm.029)}{0.029}}{1-0.672L}\ \Delta MOOD_{t-2} + \frac{1}{1-0.672L}\ e_{3t}$$

$$(5.8)\qquad \Delta IP_t = \underset{(\pm.019)}{0.124}\ \Delta GNP_t + \underset{(\pm.088)}{0.416}\ \Delta PE2_t + \frac{\underset{(\pm132)}{1}}{1+0.034L}\ e_{1t}$$

$$(5.9)\qquad \Delta IH_t = \underset{(\pm.0151)}{0.0246}\ \Delta GNP_t + \underset{(\pm.011)}{[0.061} + \underset{(\pm.013)}{0.088L} + \underset{(\pm.013)}{0.036L^2}]\ \Delta HSQ_t + \frac{\underset{(\pm.117)}{1}}{1-0.290L}\ e_{5t}\ ,$$

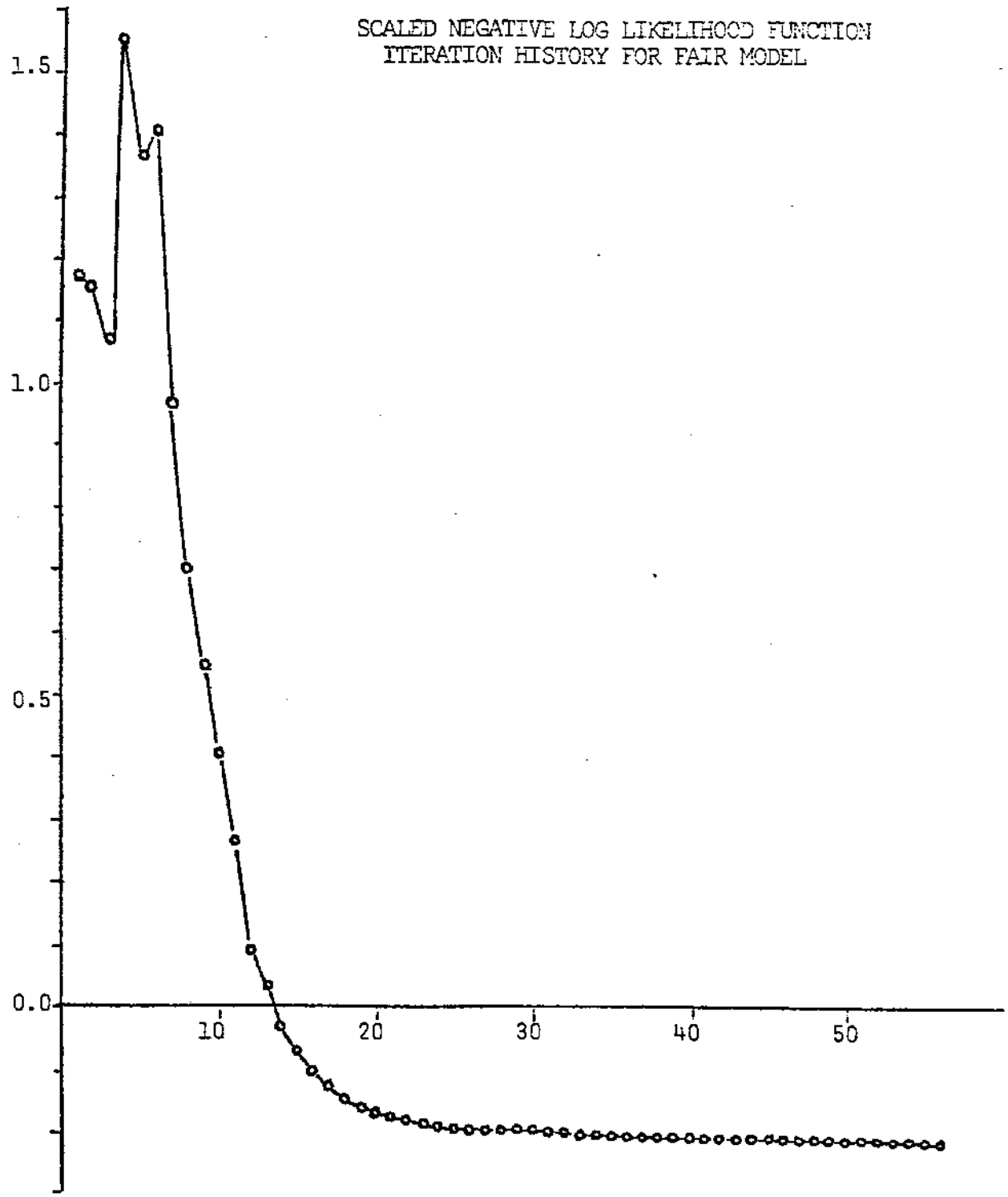
SCALED NEGATIVE LOG LIKELIHOOD FUNCTION
ITERATION HISTORY FOR FAIR MODEL
1.5
1.0
0.5
0.0
10
20
30
40
50

FIGURE 2

$$(5.18)\quad \Delta IMP_t = \underset{(\pm.010)}{0.147}\ \Delta GNP_t + \underset{(\pm.616)}{0.508}\ D_{644t} - \underset{(\pm.672)}{0.975}\ D_{651t}$$

$$- \underset{(\pm.758)}{0.822}\ D_{684t} - \underset{(\pm.770)}{4.407}\ D_{691t} + \underset{(\pm.656)}{6.647}\ D_{692t}$$

$$+ [1 - \underset{(\pm.094)}{0.713}L]\ e_{7t},$$

6. SUMMARY AND CONCLUSIONS

A unified approach to the representation and estimation of linear, discrete-time econometric models has been presented which takes full advantage of the computational capabilities provided by the combination of modern digital computers and the latest nonlinear minimization algorithms. The RSF representation was first introduced to provide a general framework within which the estimation of almost all other standard linear econometric models can be achieved through the use of just one estimation method. In particular, the RSF model can be used to represent five different single equation models:

(i) The univariate time series ARMA model,

(ii) The univariate ARMAX model,

(iii) The univariate distributed lag model,

(iv) The autoregressive multiple input model,

(v) The rational multiple input model,

and five different multiple equation models:

(vi) The polynomial structural form,

(vii) The polynomial reduced form model,

(viii) The simultaneous distributed lag model,

(ix) The final form model,

(x) The multivariate (simultaneous) time series model.

and therefore introduce error transients which do not decay to insignificance - even for long data series. In either case an alternative solution of the starting value problem is required. The concept of nuisance parameters seems the easiest to implement. Such an approach entails the estimation of the unknown starting values along with the other parameters of the model, demanding only minor alterations to any computer code originally designed for the zero value approximation.

Numerical examples have been presented demonstrating: (i) the validity of certain simplifying assumptions; and (ii) the practicality of the method when confronted with a realistic problem. The first example of Section 5 demonstrated the acceptibility of the zero starting value approximation on a model estimated over only 70 observations. The parameter estimates based on the sample were not noticeably different (considering both standard errors and the point estimates themselves) when the model was estimated with exact zero starting values (sample periods 1 through 75), or estimated with true starting values different from zero (sample periods 5 through 75). Experience with other models suggests that these observations hold for samples as short as 50 observations. The second example of Section 5 demonstrated the practicality in the light of a realistic simultaneous system. A total of 33 parameters were estimated without experiencing any difficulties. The DFP algorithm achieved convergence well within the number of iterations and computation time to be expected in a problem of such size and complexity.

The researcher in econometric modelling now has at his disposal advanced computational facilities and sophisticated optimization algorithms. By appropriate combination of these assets a very general class of estimation

REFERENCES

[1] Aigner, D. J., "A Compendium on Estimation of the Autoregressive-Moving Average Model from Time Series Data", Int'l Economic Rev, Vol. 12, No. 3, October 1971, pp. 348-371.

[2] Astrom, K. J., and Bohlin, T., "Numerical Identification of Linear Dynamical Systems from Operating Records", Proc. IFAC Conference on Self-Adaptive Control Systems, Teddington, U.K., 1965.

[3] Bohlin, T., "On the Maximum Likelihood Method of Identification", IBM J. Res. Develop., Vol. 14, January 1970. pp.

[4] Bohlin, T. and Wensmark, "Process Modelling and Control-Design Package", IBM Scientific Library Documentation 360D 13.4.004, IBM Nordic Laboratory, Sweden, August 1967.

[5] Bohlin, T., "On the Problem of Ambiguities in Maximum Likelihood Identification", Automatica, Vol. 7, 1971, pp. 199-210.

[6] Box, G.E.P. and Jenkins, G. M., Time Series Analysis: Forecasting and Control, San Fransisco: Holden-Day, 1970.

[7] Caines, P. E., and Rissanen, J., "Maximum Likelihood Estimation of Parameters in Multivariate Gaussian Stochastic Processes", IEEE Trans. Information Theory, Vol. IT-20, no. 1, January 1974.

[8] Caines, P. E., and Rissanen, J., "Consistency of Maximum Likelihood Estimators for Multivariate Gaussian Processes with Rational Spectrum", Electronics Research Laboratory Memorandum No. ERL-M423, University of California, Berkeley, March 1974.

[9] Chow, G. C., "Optimal Stochastic Control of Linear Economic Systems", J. Money, Credit, and Banking, Vol. 2, August 1970, pp. 291-302.

[10] Chow, G. C., "Optimal Control of Linear Economic Systems with Finite Horizon Time", Int'l Economic Rev., Vol. 13, No. 1, February 1972, pp. 16-25.

[11] Chow, G. C., "How Much Could be Gained by Optimal Stochastic Control Policies", Ann. Econ. Soc. Meas., Vol. 1, No. 4, October 1972, pp. 391-406.

[12] Chow, G. C., "On the Computation of Full-Information Maximum Likelihood Estimates for Non-linear Equation Systems", Econometric Research Memorandum No. 142, Princeton University, 1972.

[13] Chow, G. C. and Fair, R. C., "Maximum Likelihood Estimation of Linear Equation Systems with Auto-Regressive Residuals", Ann. Econ. Soc. Meas., Vol. 2, No. 1, January 1973, pp. 17-28.

[14] Davidon, W. C., "Variable Metric Method for Minimization", A.E.C. Res. Develop. Report ANL-5990 (Rev.), 1959.

[30] Preston, A. J. and Wall, K. D., "An Extended Identification Problem for State Space Representations of Econometric Models", PREM Discussion Paper No. 6, Queen Mary College, London, June 1973.

[31] Rosenbrock, H. H., State-space and Multivariable Theory, New York: John Wiley & Sons, Inc., 1970.

[32] Rothenberg, T. J., "Identification in Parametric Models", Econometrica, Vol. 39, No. 3, May 1971, pp. 577-591.

[33] Rowe, I. H., "Bootstrap Method for the Statistical Estimation of Model Parameters", Int'l J. Control, Vol. 12.

[34] Solow, R. M., "On a Family of Lag Distributions", Econometrica. Vol. 28, No. 2, April 1960, pp. 393-406.

[35] Steiglitz, K. and MacBride, L.E., "A Technique for the Identification of Linear Systems", IEEE Transactions on Automatic Control, Vol. AC-10, No. 4, October 1965, pp. 461-464.

[36] Wall, K. D., "An Application of Simultaneous Estimation to the Determination of Causality Between Money and Income", PREM Discussion Paper No. 8, Queen Mary College, London, April 1974.

[37] Wall, K. D., Preston, A. J., Bray, J., and Preston, M. H., "Estimates for a Simple Control Model of the U.K. Economy", Chapter 13 in Modelling of the U.K.Economy, London: Heinemans Press, 1973.

[38] Zellner, A. "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Jour. American Stat. Assoc., Vol. LVII, June 1962, pp. 348-368.

$$\theta' = [\theta_1' \quad \theta_2' \quad \cdots\cdots \quad \theta_G'] ,$$

$$\theta_i^T = [(\alpha_{ij}^k : k=1, \rho_{ij} \; ; \beta_{ij}^k : k=1, \sigma_{ij}) \; ; j=1, G_i \; ;$$

$$(a_{ij}^k : k=1, R_{ij} \; ; b_{ij}^k : k=1, S_{ij}) \; ; j=1, K_i \; ;$$

$$c_i^k : k=1, Q_i \; ; d_i^k : k=1, P_i) \; ; \kappa_i] ,$$

$$m_i = \sum_j R_{ij} + \sum_j (1 + S_{ij}) + P_i + Q_i + 1(k_i),$$

where $1(k_i) = 1$ if k_i is estimated and zero otherwise.

For each equation, the ordering proceeds from one rational lag to another with the denominator coefficients appearing before the numerator coefficients. The rational lag parameters for the endogenous and exogenous variables are then followed by the rational lag parameters of the random error term, with the constant for the equation being positioned last in the subvector. This ordering is exactly that which is obtained by reading (2.1) - (2.2) from left to right, and then repeating the process for each i.

9. The notation $\prod_{k=1}^{N} z_k$ is introduced to denote the product operation, i.e., $z_1 \cdot z_2 \cdot z_3 \cdots \cdot z_N$.

10. The associated PSF can be obtained from (2.2) by multiplying each equation by its least common denominator. Assuming each denominator of (2.2) has roots distinct from all the others (a situtation almost always encountered in practice) the least common denominator becomes

$$a_i(L) = [\prod_{j=1}^{G_i} \alpha_{ij}(L)] \cdot [\prod_{j=1}^{K_i} a_{ij}(L)] \cdot d_i(L),$$

where $a_i(L)$ is as defined previous to (2.7). If any one equation is already constrained to have a polynomial form then $a_i(L)$ is given explicitly and does not have to be computed.

11. A variety of practical applications of the estimation method using a central differencing scheme has indicated that the gain in accuracy is not worth the additional computational burden. The simple first differencing approximation to the gradient using the parameter perturbation suggested in the text has proven more than satisfactory. The first difference scheme has always given the same rate of convergence for the DFP algorithm as the central difference scheme.

15. The specification given in Chow & Fair [13] contains only the identity for GNP. However, their parameter constraints $\beta_{62} = \beta_{61}$ and $\gamma_{6,10} = \gamma_{69}$ are equivalent to the addition of an identity aggregating durable and nondurable consumption while retaining only β_{61} and γ_{69} as unknown parameters.

16. Since all of the equations are linear, and assumed to adequately describe the phenomena in a linear fashion, this alteration of the data should not affect the final results. The only possible changes that could occur would be in the correlation structure of the error terms (which Chow and Fair assume to be first order autoregressive).