Volume Title: Annals of Economic and Social Measurement, Volume 5, number 1

Volume Author/Editor: NBER

Volume Publisher:

Volume URL: http://www.nber.org/books/aesm76-1

Publication Date: 1976

Chapter Title: FIML Estimation of Rational Distributed Lag Structural Form Models

Chapter Author: Kent D. Wall

Chapter URL: http://www.nber.org/chapters/c10427

Chapter pages in book: (p. 53 - 63)

Annals of Economic and Social Measurement, 5/1, 1976

FIML ESTIMATION OF RATIONAL DISTRIBUTED LAG STRUCTURAL FORM MODELS

BY KENT D. WALL*

The Rational Distributed Lag Structural Form (RSF) representation of an econometric model is introduced, and its associated FIML estimation problem formulated. When viewed as a nonlinear unconstrained optimization problem, FIML parameter estimates can be obtained via an application of the Davidon–Fletcher–Powell variable metric method. Simple first difference approximations are employed in place of the necessary gradients, thus requiring a minimum of effort on the part of the model builder by obviating the analytical derivation, and coding, of the gradient expressions. The feasibility of estimating such complicated model representations is demonstrated with a realistic example using a nine equation variant of the Fair short-term macroeconomic forecasting model.

1. INTRODUCTION

The rapid advances made in recent years in digital computer technology have provided the econometrician with computational capabilities unheard of scant ten years ago. 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 [3] and [10]. 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.

The primary purpose of this paper is to demonstrate the feasibility of estimating complicated (nonlinearly parameterized) econometric models when full advantage is taken of both the computational power of modern digital computers and the most efficient optimization algorithms. The expositional vehicle for this purpose is the Rational Distributed Lag Structural Form (RSF) representation; and its introduction and use may be viewed as a secondary purpose. This one representation (at the cost of a possible nonlinear parameterization) encompasses all standard linear, stationary, constant-coefficient econometric models. In addition the RSF representation admits autoregressive-moving average error processes of arbitrary degree.[1]

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

[1] A detailed description of the relationship between the RSF representation and the more traditional forms of econometric models is given by the author in [14].

There is nothing conceptually new in what is presented below. The contribution of the paper is mainly methodological and is written much in the spirit of [3]. The Davidon–Fletcher–Powell function minimization algorithm employed here is not new; it just does not appear to be in as wide a use as it should be in econometrics. Perhaps the RSF representation may be considered new, but then it is just a multivariate extension of the univariate models first treated by Astrom and Bohlin [1] and Box and Jenkins [2]. Indeed, multivariate versions of these earlier models have already been considered by Hannan [9] and most recently by Zellner and Palm [16]. The true value of the present work is the idea of combining a very advanced minimization technique with a very general model representation.

2. The Rational Structural From (RSF) Representation

Consider a general linear econometric model with G_{id} identities (definitional equations) and G_b behavioural (stochastic) equations. Let there be a total of K purely exogenous variables acting as inputs. Then each equation of the RSF representation takes the following form:[2]

$$y_{it} = \sum_{\substack{j=1 \\ j \neq i}}^{G_i} \frac{\beta_{ij}(L)}{\alpha_{ij}(L)} L^{\delta_{ij}} y_{jt} + \sum_{j=1}^{K_i} \frac{b_{ij}(L)}{a_{ij}(L)} L^{D_{ij}} x_{jt} + k_i \tag{2.1}$$

for $1 \leq i \leq G_{id}$, and

$$y_{it} = \sum_{\substack{j=1 \\ j \neq i}}^{G_i} \frac{\beta_{ij}(L)}{\alpha_{ij}(L)} L^{\delta_{ij}} y_{jt} + \sum_{j=1}^{K_i} \frac{b_{ij}(L)}{a_{ij}(L)} L^{D_{ij}} x_{jt} + \frac{c_i(L)}{d_i(L)} e_{it} + k_i \tag{2.2}$$

for $G_{id} + 1 \leq i \leq G = G_{id} + G_b$. Thus the i-th endogenous variable at time t, y_{it}, is related to $G_i (\leq G)$ other endogenous variables, $y_{jt} (i \neq j)$; $K_i (\leq K)$ exogenous variables, x_{jt}; a constant bias or intercept term, k_i; and, in the case of the last G_b equations, a random disturbance or error, e_{it}. The relationships between all of these variables take the form of rational distributed lags in the "lag operator" L (i.e., $L^l z_t = z_{t-l}$). Each rational operator is characterized by the following polynomials in L:

$$\begin{aligned}
\alpha_{ij}(L) &= 1 + \alpha_{ij}^1 L + \ldots + \alpha_{ij}^{P_{ij}} L^{P_{ij}} \\
\beta_{ij}(L) &= \beta_{ij}^0 + \beta_{ij}^1 L + \ldots + \beta_{ij}^{\sigma_{ij}} L^{\sigma_{ij}} \\
a_{ij}(L) &= 1 + a_{ij}^1 L + \ldots + a_{ij}^{R_{ij}} L^{R_{ij}} \\
b_{ij}(L) &= b_{ij}^0 + b_{ij}^1 L + \ldots + b_{ij}^{S_{ij}} L^{S_{ij}} \\
c_i(L) &= 1 + c_i^1 L + \ldots + c_i^{Q_i} L^{Q_i} \\
d_i(L) &= 1 + d_i^1 L + \ldots + d_i^{P_i} L^{P_i}.
\end{aligned} \tag{2.3}$$

[2] As a matter of convention it is assumed that all the identities appear first in the system of G equations comprising the model. This ordering of the equation is also computationally convenient when solving the model for its residuals in the parameter estimation algorithm.

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]

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

$$\underline{V}(L) = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & V(L) \end{array}\right],$$

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

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

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

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

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

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

3. The FIML Estimation Problem

The estimation of the parameters associated with (2.1)–(2.3) can be approached using the method of maximum likelihood by assuming that the additive random disturbances represent zero mean "white noise" with a jointly normal distribution, i.e.,

$$e_t \sim N(0, R) \quad E\{e_k e_l'\} = R \cdot \Delta_{kl} \tag{3.1}$$

[3] The leading coefficient in the $c_i(L)$ polynomial is not fixed at unity to satisfy any normalization. The constraint is imposed to allow the unique estimation of the residual variance-convariance matrix (see Hannan [7] and [8]).

[4] In order to simplify the notation the constant terms, k_i. have been eliminated from explicity mention in the model. It is assumed that they have been absorbed into the $b_{ij}(L)$ coefficients with aid of dummy variables.

where Δ_{kl} is the Kronecker delta function and R is the symmetric positive definite variance-covariance matrix.[5] The likelihood function is then given by (see Wall [15] for details):

$$L(\theta) = \prod_{t=1}^{T} \frac{|\det A_0|}{(2\pi)^{G/2}(\det R)^{1/2}} \exp\{-\tfrac{1}{2} e_{t/t}' S^{-1} e_{t/t}\} \tag{3.2}$$

where the equivalent "structure matrix" A_0 is given by

$$A_0 = \begin{bmatrix} 1 & -\beta_{12}^0 & \cdots & -\beta_{1G}^0 \\ -\beta_{21}^0 & 1 & \cdots & -\beta_{2G}^0 \\ \vdots & \vdots & & \vdots \\ -\beta_{G1}^0 & -\beta_{G2}^0 & & 1 \end{bmatrix}$$

(If $\delta_{ij} \neq 0$ then the ij-th location in A_0 contains a zero since no contemporaneous endogenous interaction occurs between the i-th and j-th endogenous variables.)

and

$$S = A_0^{-1} R (A_0')^{-1}$$

The unknown parameters (the coefficients of the polynomials in (2.3), say m in number, together with the unknown elements of R, $G_b(G_b+1)/2$ in number) are collected into a vector Θ according to

$$\Theta' = [\theta' \mid \phi']$$

where,

θ' = contains all unknown polynomial coefficients,[6]

[5] It is also assumed that the parameterization implicit in the particular lag structure being estimated is identified. The identification problem for ARMA versions of the RSF have been solved by Hannan [8], but identification conditions for the general rational form are yet to be derived—a possible approach, however, may stem from the work of Rothenberg [12]. A more complete list and discussion of all the assumptions upon which the success of the estimation depends can be found in [15].

[6] For convenience in computation, the ordering of the unknown coefficient parameters follows in an equation-by-equation scheme, i.e., if m_i denotes the number of unknown coefficients in the i-th equation and $\underline{\theta}_i$ is the $m_i \times 1$ vector of unknowns for this equation, then θ is constructed as follows:

$$\theta' = [\underline{\theta}_1' \underline{\theta}_2' \ldots \underline{\theta}_G'].$$

$$\underline{\theta}_i = [(\alpha_{ij}^k: k = 1, \rho_{ij}; \beta_{ij}^k: k = 1, \sigma_{ij}); j = 1, G_i;$$
$$(a_{ij}^k: k = 1, R_{ij}; b_{ij}^k: k = 1, S_{ij}); j = 1, K_i;$$
$$(c_i^k: k = 1, Q_i; d_i^k: k = 1, P_i); k_i].$$

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

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.

and,

$$\phi' = \text{contains the } G_b(G_b+1)/2 \text{ unknown elements of } R.$$

The estimated one-step forecast errors, or residuals, $e_{t/t}$ are defined by $e_{t/t} = y_t - y_{t/t-1}$ with each component of $y_{t/t-1}$ computed from the right-hand side of (2.2) when e_t is set to its expected value of zero.

In terms of the actual computation of estimates for θ, it is more convenient to work with the "scaled" version of the concentrated negative log likelihood function:

$$J(\theta) = \ln(\det \hat{R}) - 2\ln(|\det A_0|) \tag{3.4}$$

where

$$\hat{R} = \frac{1}{T}\sum_{t=1}^{T} e_{t/t}e'_{t/t} \tag{3.5}$$

Thus the independent parameters ϕ have been removed from the set of unknowns, all constant terms not dependent upon θ subtracted out, and the log-likelihood multiplied by $-2/T$.[7] The minimization of $J(\theta)$ with respect to θ will produce the ML estimates of the coefficients in (2.3). This function minimization constitutes the core of the FIML estimation for the RSF representation:

(P.1) *Basic FIML Estimation Problem.* Given a particular structure specification, determine the vector θ such that $J(\theta)$ is minimized over θ, i.e., find $\theta = \hat{\theta}$ such that $J(\hat{\theta}) \leq J(\theta)$. The unknown parameters ϕ are then recovered with the aid of (3.5) using the estimated residual sequence $\{e_{t/t};\ 1 \leq t \leq T\}$ evaluated at $\theta = \hat{\theta}$.

4. Numerical Solution of the Estimation

The numerical minimization of the scaled negative log likelihood function with respect to θ has been considered earlier in the literature for some more standard representations of econometric models. Astrom and Bohlin [1] used a modified Newton–Rhapson algorithm, Box and Jenkins [2] employed a repeated relinearization coupled with a variation of the steepest descent algorithm, and Hendry [10] implemented a direct (grid type) search algorithm due to Powell. All of these numerical procedures possess certain advantages as well as disadvantages. These stem from the relative value of the numerical algorithms with respect to the dimension of θ, the initial guess for $\hat{\theta}$, the rate of convergence, and whether analytic or numerical expressions for the first and second partial derivatives of $J(\theta)$ are available. Today there exists an algorithm, due to Davidon and improved upon by Fletcher and Powell, known in the literature as DFP which appears to capture the best characteristics of each of those mentioned above [4], [5]. The DFP algorithm represents a blend of the convergence rate of Newton–Rhapson techniques with the initial stability of descent methods. This algorithm also strikes a balance between the extensive computational requirements of Newton–Raphson and the very meager requirements of direct search techniques like that of Powell. For these reasons the DFP algorithm is employed here.

[7] As noted by a referee, the concentration of the likelihood function is not accomplished without some loss in generality since concentration rules out restrictions on the contemporaneous residual variance-covariance matrix.

The DFP procedure requires an initial guess for θ, denoted θ^0, and an initial guess for the positive definite matrix $H(\theta^0)$. Then for $k = 1, 2, \ldots$ successively improved estimates are obtained for θ via

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

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 convergence criteria are met, i.e., until

$$\|H(\theta^k) J_\theta(\theta^k)\| \leq \varepsilon_1$$

or,

$$\|\delta\theta^k\| \leq \varepsilon_2.$$

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 [4], [5], [13]. The algorithm automatically computes α^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.[8] In particular, if $\partial J/\partial\theta_i$ denotes the i-th component of J_θ then

$$\partial J/\partial\theta_i = [J(\theta^k + \Delta\theta^{ki}) - J(\theta^k)]/\lambda_i, \tag{4.1}$$

where $\lambda_i = 10^{-4}|\theta_i^k| + 10^{-6}$ and $\Delta\theta^{ki}$ is defined as the $m \times 1$ null vector with the i-th element replaced by λ_i. The nonlinearities of any parameterization of the general RSF representation now become immaterial to the generation of J_θ; thus this scheme can be used for all structure specifications. The implementation of the DFP algorithm, within the context of the estimation problem (P.1), now requires only a means of computing $J(\theta^k)$.[9]

[8] 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.

[9] This operation is straightforward although somewhat messy. Essentially each of (2.1) are solved for the identity outputs, and then these are used in solving each of (2.2) for the residual $e_{i/i}$. The generation of these series involves some awareness of the lag structure of each equation so that starting values can be obtained directly from the data. Where no starting values are given (as with the residuals) these are set to zero in confidence that asymptotic behaviour will obtain, i.e. that the data sequences are long enough. The interested reader is directed to Wall [15] for a detailed exposition.

5. An Example

A reasonably realistic test problem is provided by the example considered in Chow and Fair [3] which represents a simultaneous block of seven stochastic equations and two identities.[10] Correlated errors are present requiring the estimation of additional parameters for the rational lags of the random disturbance terms. Thus many of the complications found in applied econometric modelling are encountered in this example. The economic and structure specifications were taken directly from Chow and Fair [3] with one exception: all the variables of the model were transformed to mean diviates of their first differences.[11] This insured stationarity and avoided the estimation of constant (bias) terms in many of the equations.

The simultaneous estimation problem was initialized with the results obtained from estimating the RSF representation of each equation in isolation.[12] The correlation structure for the errors of each equation was determined by examination of the residual autocorrelation functions produced 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):

$$(5.1)\quad \Delta \mathrm{GNP}_t = \Delta CD_t + \Delta CN_t + \Delta CS_t + \Delta IP_t + \Delta IH_t + \Delta^2 V_t - \Delta IMP_t + \Delta G_t$$

$$(5.2)\quad \Delta Z_t = \Delta CD_t + \Delta Cn_t$$

$$(5.3)\quad \underset{(\pm 0.016)}{\Delta CD_t = 0.136\,\Delta \mathrm{GNP}_t} + \underset{(\pm 0.049)\ (\pm 0.048)}{[0.137 + 0.058L]\,\Delta MOOD_{t-1}} \underset{(\pm 1.323)}{-2.593\,D_{644t}} + \underset{(\pm 1.103)}{3.611\,D_{651t}} + \underset{(\pm 0.139)}{[1 - 0.398L]\,e_{1t}}$$

$$(5.4)\quad \Delta CN_t = \underset{(\pm 0.118)}{\frac{0.073}{1 - 0.67L}}\,\Delta \mathrm{GNP}_t \underset{(\pm 0.0557)}{-0.1047}\ \Delta MOOD_{t-2} + \frac{1}{1 - 0.677L}\,e_{3t}$$

$$(5.5)\quad \Delta CS_t = \underset{(\pm 0.067)}{\frac{0.073}{1 - 0.672L}}\,\Delta \mathrm{GNP}_t \underset{(\pm 0.029)}{\frac{-0.029}{1 - 0.672L}}\,\Delta MOOD_{t-2} + \frac{1}{1 - 0.672L}\,e_{3t}$$

[10] The specification given in Chow and Fair [3] 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 β_{69} as unknown parameters.

[11] Since all 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).

[12] Like all nonlinear iterative optimization algorithms, convergence is significantly influenced by the initial choice for θ. If easily computed, an initial consistent estimate of $\hat{\theta}$ should be used. The author has had great success using as initial estimates those obtained by considering each equation separately. This corresponds to a LIML estimation procedure serving as a starting place for the complete (simultaneous) FIML estimation.

$$(5.6)\quad \Delta IP_t = \underset{(\pm 0.019)}{0.124}\,\Delta \mathrm{GNP}_t + \underset{(\pm 0.088)}{0.416}\,\Delta PE2_t + \underset{(\pm 0.0132)}{\frac{1}{1+0.034L}}\,e_{1t}$$

$$(5.7)\quad \Delta IH_t = \underset{(\pm 0.0151)}{0.0246}\,\Delta \mathrm{GNP}_t + \underset{(\pm 0.011)\,(\pm 0.013)\quad (\pm 0.013)}{[0.061+0.088L+0.036L^2]}\,\Delta HSQ_t + \underset{(\pm 0.117)}{\frac{1}{1-0.290L}}\,e_{5t}$$

$$(5.8)\quad \Delta V_t = \underset{(\pm 0.130)(\pm 0.121)}{\frac{-0.235+0.547L}{1-0.612L}}\Delta Z_t + \underset{(\pm 2.53)}{\frac{+0.178}{1-0.612L}}D_{644t} + \underset{(\pm 0.689)}{\frac{+7.404}{1-0.612L}}D_{654t} + \underset{(\pm 0.080)}{\frac{1}{1-0.612L}}e_{6t}$$

$$(5.9)\quad \Delta IMP_t = \underset{(\pm 0.008)}{0.115}\,\Delta \mathrm{GNP}_t + \underset{(\pm 1.087)}{0.085}\,D_{644t} - \underset{(\pm 1.13)}{1.251}\,D_{651t} - \underset{(\pm 1.09)}{1.003}\,D_{684t} - \underset{(\pm 1.29)}{4.428}\,D_{691t} + \underset{(\pm 1.085)}{6.822}\,D_{692t} + \underset{(\pm 0.065)}{[1-0.684L]}\,e_{7t}.$$

The estimation of the entire 9 equation system, starting from the above estimates, took 56 iterations. Convergence of the DFP algorithm was considered achieved once the correction to each unknown parameter was computed to be less than 10^{-4}, which in this instance corresponded to each element of the gradient vector being of the order 10^{-5}. The method averaged approximately two function and gradient evaluations per iteration, or 4,250 simulations of the system of equations defining e_{it} as a function of y_t and x_t. Total central processor time amounted to less than 3 minutes on an IBM 370/168 computer.

The iteration history of $J(\theta)$ is depicted in Figure 1. The peculiar behavior of the minimization algorithm displayed during iterations 4, 5, and 6 is a consequence of numerical errors in updating the approximate inverse Hessian, H. In certain instances, implementation of the standard updating formulas can lead to indefinite Hessians, and uphill searches. This problem has been recognized and overcome by altering the updating formula for H [13], but the particular DFP version employed in this work was derived from a "straight" coding of the original formulas and therefore did not exclude this difficulty. It is interesting to note that once the algorithm recovers from these initial problems it proceeds to reduce $J(\theta)$ quite rapidly, achieving near-covergence by the 35th iteration.

The final estimates for the seven stochastic equations corresponding to (5.3)–(5.9) were as follows:

$$(5.10)\quad \Delta CD_t = \underset{(\pm 0.042)}{0.098}\,\Delta \mathrm{GNP}_t + \underset{(\pm 0.073)\,(\pm 0.066)}{[0.057+0.065L]}\,\Delta MOOD_t - \underset{(\pm 1.234)}{2.002}\,D_{644t} + \underset{(\pm 1.356)}{2.977}\,D_{651t} + \underset{(\pm 0.157)}{[1-0.340L]}\,e_{1t}$$

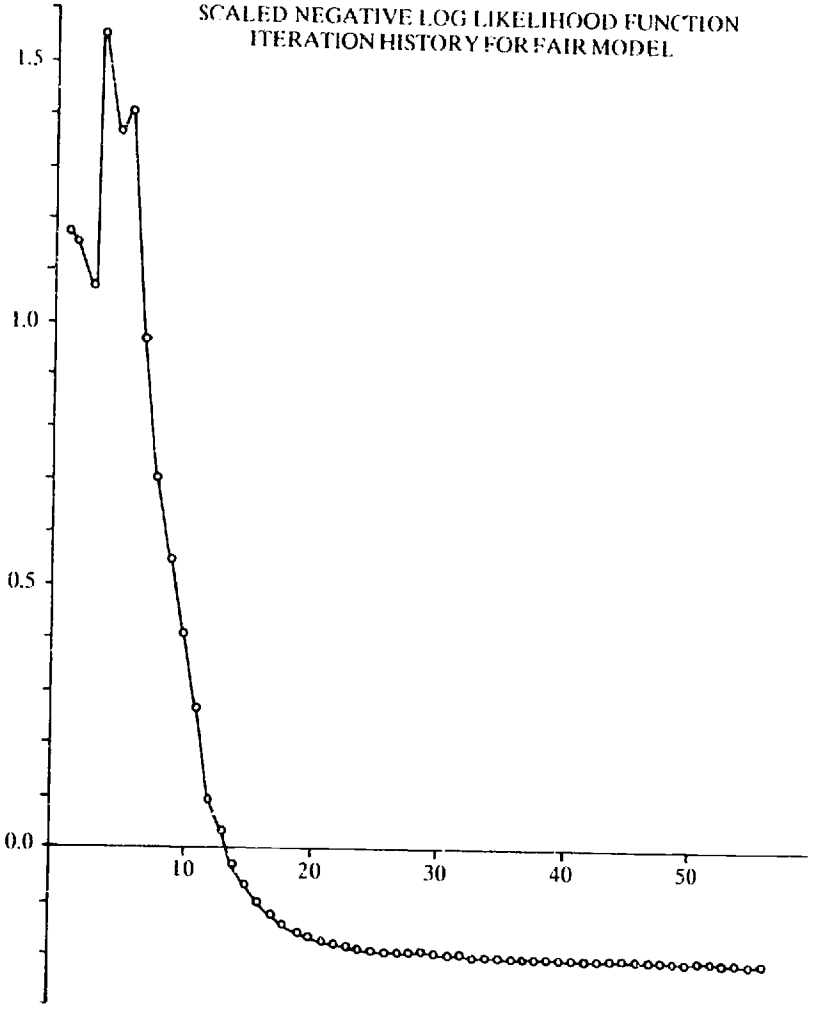

Figure 1

$$\Delta CN_t = \underset{(\pm 0.038)}{\frac{0.134}{1-0.542L}}\,\Delta \mathrm{GNP}_t - \underset{(\pm 0.071)}{0.066}\,\Delta MOOD_{t-2} + \underset{(\pm 0.126)}{\frac{1}{1+0.322L}}\,e_{2t} \tag{5.11}$$

$$\Delta CS_t = \underset{(\pm 0.034)}{\frac{0.071}{1-0.685L}}\,\Delta \mathrm{GNP}_t - \underset{(\pm 0.035)}{\frac{0.036}{1-0.685L}}\,\Delta MOOD_{t-2} + \underset{(\pm 0.103)}{\frac{1}{1-0.685L}}\,e_{3t} \tag{5.12}$$

$$\Delta IP_t = \underset{(\pm 0.135)}{-0.055}\,\Delta \mathrm{GNP}_t + \underset{(\pm 0.146)}{0.367}\,\Delta PE2_t + \underset{(\pm 0.286)}{\frac{1}{1-0.236L}}\,e_{4t} \tag{5.13}$$

$$(5.14)\quad \Delta IH_t = \underset{(\pm 0.017)}{0.022}\,\Delta \mathrm{GNP}_t + \underset{(\pm 0.010)}{[0.055} + \underset{(\pm 0.012)}{0.087L.} + \underset{(\pm 0.011)}{0.024 1L.^2}]\Delta HSQ_t$$

$$+\underset{(\pm 0.137)}{\frac{1}{1+0.255L}}\,e_{5t}$$

$$(5.15)\quad \Delta V_t = \underset{(\pm 0.422)}{\frac{0.121+0.475L}{1-0.484L}}\,\Delta Z_t + \underset{(\pm 0.140)}{\frac{2.901}{1-0.484L}}\,D_{644t} + \underset{(\pm 2.443)}{\frac{2.324}{1-0.484L}}\,D_{651t}$$

$$+\underset{(\pm 2.538)}{\frac{1}{1-0.484L}}\,e_{6t}$$

$$(5.16)\quad \Delta IMP_t = \underset{(\pm 0.010)}{0.147}\,\Delta \mathrm{GNP}_t + \underset{(\pm 0.616)}{0.508}\,D_{644t} - \underset{(\pm 0.672)}{0.975}\,D_{651t}$$

$$-\underset{(\pm 0.758)}{0.822}\,D_{684t} - \underset{(\pm 0.770)}{4.407}\,D_{691t} + \underset{(\pm 0.656)}{6.647}\,D_{692t}$$

$$+\underset{(\pm 0.094)}{[1-0.713L]}\,e_{7t}$$

6. Conclusions

An 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 difficulties attendant to the estimation of the generally nonlinear parameterization inherent in the RSF representation have been overcome by the use of a very efficient minimization algorithm which, while attaining quadratic convergence rates, still only requires gradient information. Practical experience has shown that even the derivation of analytical expressions for the first partial derivatives (the gradient vector) may also be dispensed with, and replaced by a simple first difference approximation.

A numerical example has been presented demonstrating the practicality of the method when confronted with a realistic problem. A total of 33 parameters was 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 problems can be considered. One such combination has been presented in this paper; but other, perhaps more advantageous, combinations are certainly possible. It is hoped that the exposition here will stimulate others in this direction leading to unified estimation methods for even wider classes of econometric models.

NBER Computer Research Center

Bibliography

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

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

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

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

[5] Fletcher, R. and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization". *Computer J.*, Vol. 6, 1963, pp. 163–168.

[6] Goldfarb, E., 'Extension of Davidon's Variable Metric Method to Maximization under Linear Inequality and Equality Constraints", *SIAM Jour. Appl. Math.*, Vol. 17, No. 4, July 1969, pp. 739–764.

[7] Hannan, E. J., "Identification of Vector Mixed Autoregressive-Moving Average Systems". *Biometrika*, Vol. 57, No. 1, 1969, pp. 223–225.

[8] Hannan, E. J., "The Identification Problem for Multiple Equation Systems with Moving Average Errors", *Econometrica*, Vol. 39, No. 5, September 1971, pp. 751–765.

[9] Hannan, E. J., "The Estimation of Mixed Moving Average-Autoregressive Systems". Biometrika, Vol. 56, (1969), pp. 206–224.

[10] Hendry, D. F., "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process". *Int'l Economic Rev.*, Vol. 12, No. 2, June 1971.

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

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

[13] Shanno, D. F., "Conditioning of Quasi-Newton Methods for Function Minimization". *Math. Comp.*, Vol. 24, (1970), pp. 647–656.

[14] Wall, K. D., "The Rational Distributed Lag Structural Form—A Very General Econometric Model", NBER Working Paper Series No. 65, (October 1974).

[15] Wall, K. D., "FIML Estimation of Rational Distributed Lag Structural Form Models". NBER Working Paper Series No. 77, (March 1975).

[16] Zellner, A. and F. Palm, "Time Series Analysis and Simultaneous Econometric Models". *Journals of Econometrics*, Vol. 2, No. 1, (May 1974), pp. 17–54.

received May 1975