MAXIMUM LIKELIHOOD ESTIMATION OF LINEAR EQUATION SYSTEMS WITH AUTO-REGRESSIVE RESIDUALS[1]

BY Gregory C. Chow AND Ray C. Fair

This paper applies Newton's method to solve a set of normal equations when the residuals follow an autoregressive scheme. It is shown that this technique for computing maximum likelihood estimates can be applied to the "seemingly unrelated regression" model. An eight equation quarterly forecasting model of the U.S. economy is then used to illustrate the method described in the paper.

1. Introduction

The problem considered in this paper is the maximum likelihood estimation of a system of linear stochastic equations in which the residuals follow an autoregressive scheme. This problem has been studied previously by Sargan [10] and more recently by Hendry [5]. The former formulated the problem and provided numerical solution to a special case. The latter applied an algorithm of Powell [8] to this problem, an algorithm that does not require the use of first or second derivatives. The present paper provides an alternative method of computing the maximum likelihood estimates. It applies Newton's method to solve a set of normal equations and is a generalization of the well-known Cochran–Orcutt technique to deal with autoregressive residuals in a regression. Thus our method is traditional in conception. Our experience, which is partly reported below, is that this method works well. However, whether it is computationally better than Hendry's or other methods remains to be investigated.

In Section 2 a set of normal equations for the unknown coefficients in a linear econometric system is presented for the case in which the residuals are serially uncorrelated. The equations are first set forth without the imposition of linear restrictions, and then a method to deal with linear restrictions is discussed. A previous work, Chow [2], dealt only with linear restrictions on the coefficients within a single equation, and the method in Section 2 deals with linear restrictions on coefficients possibly belonging to different equations. The normal equations are nonlinear in the unknown coefficients, and both a direct iterative method and Newton's method have been tried for solving them. As discussed in Chow [2], Newton's method appears to converge more often and faster than the direct iterative method, and it is the method considered in this paper.

In Section 3 the analysis is expanded to the case in which the residuals follow an auto-regressive scheme. The main point of this section is that this more general statistical problem can be decomposed into two sub-problems, each of which can be solved by the method in Section 2. The decomposition is based on the observation that, given the coefficients of the auto-regressive scheme, the coefficients of the structural equations can be estimated by the method of Section 2, and, given

the latter coefficients, the former coefficients can be estimated likewise. The resulting solution of the more general problem is thus merely a two-step application of Newton's method and poses no additional computational difficulties.

A special case of the model considered in Section 3 is the case where the coefficient matrix of the endogenous variables is an identity matrix. The model then reduces to the "seemingly unrelated regression" model analyzed by Zellner [11], Parks [7], and others. Neither the two-step procedure suggested by Zellner for the serially uncorrelated case nor the three-step procedure suggested by Parks for the first order serially correlated case is a maximum likelihood procedure, but it can easily be shown, as is done in Section 4, that both of the procedures become maximum likelihood procedures if one does not stop after the second or third step but continues to iterate until convergence is reached. It is also shown in Section 4 that iterating with the Zellner or Parks procedure is equivalent to solving the set of normal equations of the system by the direct iterative method. Since Newton's method appears to be more useful in practice than the direct iterative method, the better way of obtaining the maximum likelihood estimates of the seemingly unrelated regression model appears to be to use the method discussed in Sections 2 and 3, which is based on Newton's method, rather than to iterate with the Zellner or Parks procedure.

The method described in this paper is quite general and can handle most of the problems associated with estimating linear equations systems. Linear restrictions on the coefficients can be handled, first and higher order auto-regressive properties of the residuals can be handled, and various special cases can be considered. Some of the more interesting special cases are the seemingly unrelated regression model, the case where the residuals obey a first-order auto-regressive scheme with a diagonal coefficient matrix, and the case where identities are present.

In order to illustrate the use of the method described in this paper, a numerical example is provided in Section 5. An eight equation model is estimated in which the residuals obey a first-order auto-regressive process with a diagonal coefficient matrix. There are also linear restrictions on the coefficients of one of the equations in the model, and one of the equations in the model is an identity. The model has 33 structural parameters and 7 auto-regressive parameters to be estimated.

2. A Method of Maximum Likelihood Estimation of Linear Equation Systems with Linear Restrictions on the Coefficients

Let the linear system of structural equations be

$$(2.1) \qquad\qquad YB' = Z\Gamma' + U,$$

with Y and Z denoting $T \times G$ and $T \times K$ matrices of observations on the G dependent variables and the K predetermined variables, U denoting a $T \times G$ matrix of residuals, and B' and Γ' (prime for transpose) denoting $G \times G$ and $K \times G$ matrices of coefficients (the ith columns of B' and Γ' being the coefficients of the ith equation). Assume that the T rows of U are uncorrelated, and that the G elements of each row satisfy a multivariate normal distribution with mean 0 and covariance matrix Σ. Then the log concentrated likelihood function can be

18

written as [1, 2]

$$L = \text{const.} - \frac{T}{2} \log \left\{ \left| \frac{1}{T}(BY' - \Gamma Z')(YB' - Z\Gamma') \right| \Big/ \left| \frac{1}{T} BY'YB' \right| \right\}$$

(2.2)

$$= \text{const.} - \frac{T}{2} \log \{|S|/|W|\}.$$

If all the variables with zero coefficients in the ith equation are excluded and if β_{ii} is set equal to 1, then the ith equation of (2.1) can be written as

(2.3) $$y_i = Y_i\beta_i' + Z_i\gamma_i' + u_i, \qquad (i = 1, \ldots, G),$$

where β_i' and γ_i' are column vectors of the remaining unknown coefficients in the ith equation.

Setting the partial derivatives of (2.2) with respect to these unknown coefficients equal to zero yields the following system of normal equations [2, equation (2.8)],

(2.4)
$$\begin{bmatrix} q^{11}Y_1'Y_1 \ldots q^{G1}Y_1'Y_G & s^{11}Y_1'Z_1 \cdots s^{G1}Y_1'Z_G \\ \cdots & \cdots \\ q^{1G}Y_G'Y_1 \ldots q^{GG}Y_G'Y_G & s^{1G}Y_G'Z_1 \ldots s^{GG}Y_G'Z_G \\ s^{11}Z_1'Y_1 \ldots s^{G1}Z_1'Y_G & s^{11}Z_1'Z_1 \ldots s^{G1}Z_1'Z_G \\ \cdots & \cdots \\ s^{1G}Z_G'Y_1 \ldots s^{GG}Z_G'Y_G & s^{1G}Z_G'Z_1 \ldots s^{GG}Z_G'Z_G \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_G' \\ \gamma_1' \\ \vdots \\ \gamma_G' \end{bmatrix} = \begin{bmatrix} Y_1'\sum_h q^{h1}y_h \\ \vdots \\ Y_G'\sum_h q^{hG}y_h \\ Z_1'\sum_h s^{h1}y_h \\ \vdots \\ Z_G'\sum_h s^{hG}y_h \end{bmatrix}$$

or

$$f(\alpha) = 0,$$

where s^{ij} and w^{ij} are respectively the $i - j$ elements of the inverses of S and W as defined by (2.2), $q^{ij} = (s^{ij} - w^{ij})$, and α stands for the vector of all of the unknown coefficients in the system.

Newton's method can now be applied to solve the system of normal equations (2.4). Let F be the matrix of partial derivatives of the elements of f with respect to the elements of α, as given explicitly in Chow [2], equations (4.8)–(4.10), and let α^r be the value of α in the rth iteration. Newton's method iterates by the formula[2]

(2.5) $$\alpha^{r+1} = \alpha^r - [F(\alpha^r)]^{-1} f(\alpha^r).$$

If there are linear restrictions on the elements of α (these elements may be coefficients in different structural equations), one has to modify the vector $f(\alpha)$

---

[2] In the programming of Newton's method for the work in Chow [2] and for the work here, the actual value of α for the $r + 1$ iteration is taken to be $\alpha^r + h(\alpha^{r+1} - \alpha^r)$. If the likelihood is larger for $h = 1$, then $h = 1.25, (1.25)^2, \ldots$ is tried until the likelihood decreases. If the likelihood is not larger for $h = 1$, then $h = 0.8, -0.8, (0.8)^2, -(0.8)^2 \ldots$ is tried in an attempt to find a larger likelihood. If a larger likelihood is not found and if the difference between α^{r+1} (as computed in (2.5)) and α^r is still sizeable, then the program breaks down.

and the matrix $F(\alpha)$ in equations (2.4) and (2.5). The modifications required can be seen by considering the restriction

$$(2.6) \qquad \alpha_i = c\alpha_j + d\alpha_k.$$

The unknown α_i will be eliminated, since it is a known linear function of two of the remaining unknowns α_j and α_k. The likelihood function L will be replaced by a new function L^* of a new set of variables α^* (having one fewer element than α), by substituting the right-side of (2.6) for α_i in L. By (2.6) and the chain rule of differentiation, the new $f^*(\alpha^*) = 0$ will contain the following equations

$$\frac{\delta L^*}{\delta \alpha_j} = \frac{\delta L}{\delta \alpha_j} + \frac{\delta L}{\delta \alpha_i} \cdot c = 0,$$

$$(2.7)$$

$$\frac{\delta L^*}{\delta \alpha_k} = \frac{\delta L}{\delta \alpha_k} + \frac{\delta L}{\delta \alpha_i} \cdot d = 0,$$

where it is understood that the argument α_i of any derivative of L is replaced by the right side of (2.6)—likewise for equations (2.8) and (2.9) below. If $f(\alpha)$ has n elements, say, then $f^*(\alpha^*)$ and $f(\alpha)$ are related by the equation

$$(2.8) \qquad f^*(\alpha^*) = Mf(\alpha),$$

where M is an $(n-1) \times n$ matrix which is constructed from the $n \times n$ identity matrix by (1) eliminating its ith row, (2) replacing the zero in the ith position of the jth row by c, and (3) replacing the zero in the ith position of the kth row by d.

By differentiating the elements of $f^*(\alpha^*)$ with respect to the remaining $n-1$ variables, one can obtain the new matrix $F^*(\alpha^*)$ of second partial derivatives:

$$(2.9) \qquad F^*(\alpha^*) = MF(\alpha)M'.$$

Equations (2.8) and (2.9) can then be used to modify equation (2.5) in order to perform iterations by Newton's method. If there is a second linear restriction, then another matrix, say M^*, can be used to multiply f^* and F^* in the same way as M was used in equations (2.8) and (2.9) to multiply f and F. This process can be repeated for any number of linear restrictions. Setting a coefficient equal to a constant c amounts to setting it equal to c times the dummy variable 1 in the list predetermined variables; similarly, non-homogeneous linear restrictions can be treated by using this dummy variable.

Two other points about the above method should be noted. First, as discussed in Chow [2, p. 107], identities can be quite easily handled by the above method. Secondly, the covariance matrix of the estimates of α can be consistently estimated by the inverse of $-F$ evaluated at the maximizing value of α.

3. Maximum Likelihood Estimation of Linear Equation Systems with Auto-Regressive Residuals

Now let the model (2.1) be modified by assuming that its residuals U obey an auto-regressive scheme such as

$$(3.1) \qquad U = U_{-1}R_1' + U_{-2}R_2' + E,$$

where the G columns of U_{-1} and U_{-2} are the residuals of the G structural equations lagged one period and two periods respectively,[3] R_1 and R_2 are matrices of co-efficients of the auto-regressive scheme, and the residuals E satisfy the same assumptions originally made for U in the model (2.1). It will be shown in this section that the method of Section 2 can be applied to obtain maximum likelihood estimates of the matrices B, Γ, R_1 and R_2 in this model. To simplify matters of exposition without loss of generality, R_2 will be assumed to be zero.

Since the model lagged one period satisfies

$$(3.2) \qquad Y_{-1}B' = Z_{-1}\Gamma' + U_{-1},$$

the equation system (2.1) and (3.1) can be written as (with $R_2 = 0$)

$$(3.3) \qquad YB' = Y_{-1}B'R_1' + Z\Gamma' - Z_{-1}\Gamma'R_1' + E$$

$$= Y_{-1}B_1' + Z\Gamma' - Z_{-1}\Gamma_1' + E.$$

The log concentrated likelihood function for this model, by (2.2), is simply

$$(3.4) \qquad L = \text{const.} - \frac{T}{2}\log\left\{\left|\frac{1}{T}E'E\right| \Big/ \left|\frac{1}{T}BY'YB'\right|\right\},$$

where E denotes

$$(3.5) \qquad YB' - Y_{-1}B'R_1' - Z\Gamma' + Z_{-1}\Gamma'R_1',$$

with B, Γ, and R_1 treated as unknowns and Y, Y_{-1}, Z, and Z_{-1} treated as given data.

To maximize (3.4) with respect to these unknowns, consider first the partial maximization with respect to B and Γ, given R_1. From the second line of (3.3), this amounts to maximization with respect to B, Γ, B_1, and Γ_1 subject to the linear restrictions

$$(3.6) \qquad B_1 = R_1 B; \qquad \Gamma_1 = R_1 \Gamma.$$

This problem can be solved by the method of Section 2.

Now consider the maximization of (3.4) with respect to R_1, given B and Γ. With B and Γ treated as given, the model can be written as, by rearrangement of (3.3),

$$(3.7) \qquad (YB' - Z\Gamma') = (Y_{-1}B' - Z_{-1}\Gamma')R_1' + E,$$

with the terms in parentheses being treated as matrices of observed variables and R_1' being treated as a matrix of coefficients. Maximizing (3.4) partially with respect to R_1 amounts to maximizing

$$(3.8) \qquad L_1 = \text{const.} - \frac{T}{2}\log\left\{\left|\frac{1}{T}E'E\right|\right\}$$

since $|(1/T)BY'YB|$ is a constant. But (3.8) is precisely the log concentrated like-lihood function for the model (3.7), and the method of Section 2 can be applied to maximize this likelihood function with respect to the coefficient matrix R_1'. Of

---

[3] Because of (3.1), one observation is of course lost for each order of the auto-regressive scheme.

21

course, if there are no restrictions on R_1, the estimates are simply least squares estimates. In this case $(Y_{-1}B' - Z_{-1}\Gamma')$ is the matrix of the predetermined variables, and $(YB' - Z\Gamma')$ is the matrix of the dependent variables whose coefficient matrix is restricted to be the identity matrix.

The maximum likelihood estimates of B, Γ, and R_1 in the model (3.3) can be obtained as follows. Start with an initial value for R_1, possibly 0, and maximize the likelihood function with respect to B and Γ by the method of Section 2; take these values of B and Γ as given and maximize the likelihood function with respect to R_1, again by the method of Section 2; repeat this two-step process until convergence is reached. Convergence will be reached using this process if the method of Section 2 converges for the problem that it is supposed to solve, since the method of this section amounts simply to repeated applications of the method of Section 2.

Let δ denote the vector of all of the unknown coefficients in the system including the coefficients in R_1, let $h(\delta) = 0$ stand for the system of normal equations derived from differentiating the likelihood function (3.4) (as $f(\alpha) = 0$ stood for the system of normal equations derived from differentiating the likelihood function (2.2)), and let H be the matrix of partial derivatives of the elements of h with respect to the elements of δ. Then the covariance matrix of the estimator of δ can be consistently estimated by the inverse of $-H$ evaluated at the maximizing value of δ. The derivatives involved in such a procedure are quite complicated, however, and so an alternative procedure is recommended. This procedure is to compute the covariance matrix of the estimates of B and Γ under the assumption that R_1 is known (and equal to its estimate) and to compute the covariance matrix of the estimator of R_1 under the assumption that B and Γ are known (and equal to their estimates). These two estimates of the covariance matrices fall out of the two-step process above (since the matrix F^{-1} in (2.5) is computed in both steps) and so pose no further computational burden. These estimates will, of course, be an underestimate of the actual covariance matrix, since the stochastic nature of the estimator of B and Γ and the stochastic nature of the estimator of R_1 are not considered together.

The comments made in Section 2 about the ability of the method to handle various problems generally pertain to the two-step process in this section as well. In particular, linear restrictions on the coefficients B and Γ can be handled (including, of course, the restrictions in (3.6)), and linear restrictions on the coefficients of R_1 can be handled. The one type of restriction that cannot be handled by the two-step process is a restriction between the coefficients of B or Γ and the coefficients of R_1. The process cannot handle, in other words, a restriction that says that a given element of R_1 is a linear combination of given elements of B or Γ. In practice, however, this is not likely to be a serious limitation of the method. The matrix R_1 can, of course, be restricted to be diagonal, and for many problems it may be advisable to do this. Otherwise, with R_1 unrestricted, a large number of coefficients will have to be estimated for even moderately sized models, and it may be difficult to obtain estimates of this many coefficients.

4. A Special Case: Seemingly Unrelated Regressions

If B' is an identity matrix, then (2.1) reduces to the "seemingly unrelated regression" model analyzed by Zellner [11], Parks [7], and others. The basic

method proposed by Zellner for the case in which the residuals are serially uncorrelated consists in obtaining a consistent estimate of the variance-covariance matrix, S, of the residuals U and then using this estimate to compute the generalized least squares estimate of Γ. A consistent estimate of S can be obtained by estimating each equation of the model by ordinary least squares and using the estimated residuals from these equations to estimate S.

The estimates obtained from Zellner's procedure are not maximum likelihood estimates, but it can be shown that if one continued to iterate on S and achieved convergence, then the resulting estimates would be maximum likelihood estimates. When B' is an identity matrix, then the system of normal equations (2.4) reduces to the block of equations in the lower right-hand corner of (2.4). Solving this block for $\gamma'_1, \ldots, \gamma'_G$ yields:

$$
(4.1) \qquad
\begin{bmatrix} \gamma'_1 \\ \vdots \\ \gamma'_G \end{bmatrix}
=
\begin{bmatrix} s^{11} Z'_1 Z_1 \ldots s^{G1} Z'_1 Z_G \\ \vdots \\ s^{1G} Z'_G Z_1 \ldots s^{GG} Z'_G Z_G \end{bmatrix}^{-1}
\begin{bmatrix} Z'_1 \sum_h s^{h1} y_h \\ \vdots \\ Z'_G \sum_h s^{hG} y_h \end{bmatrix} .
$$

Equation (4.1) is the same as the equation for the generalized least squares estimator presented in Zellner [11], p. 351, equation (2.7). One possible way to try to solve this system of equations is to iterate on the elements of S.[4] This iterative method is equivalent to the direct iterative method discussed in Chow [2], and to the extent that the method converges, iterating in this manner produces maximum likelihood estimates.

As discussed in Chow [2], Newton's method appears to work better than the direct iterative method, and thus the better way of obtaining the maximum likelihood estimates of the seemingly unrelated regression model would appear to be to use Newton's method rather than the direct iterative method. The computational burden involved in computing the maximum likelihood estimates by Newton's method does not appear so great that one has to rely on Zellner's simpler two-step procedure to estimate the seemingly unrelated regression model.

Parks expanded the analysis of the seemingly unrelated regression model to include the case in which the residuals are first order serially correlated, that is, to include the case in which R_1 is diagonal and R_2 is zero in (3.1). His method consists in obtaining consistent estimates of the serial correlation coefficients, using these estimates to obtain a consistent estimate of the variance-covariance matrix, and then using both of these sets of estimates to compute the generalized least squares estimate of Γ. Estimates of the serial correlation coefficients can be obtained from the ordinary least squares residuals of each equation.

The estimates obtained from Parks' procedure are not maximum likelihood estimates, but again it can be shown, in a manner similar to that done above for Zellner's procedure, that iterating on the serial correlation coefficients and the elements of the variance-covariance matrix leads to maximum likelihood

---

[4] Iterating in this manner was suggested by Zellner and Theil [12], p. 78, within the context of the three-stage least squares technique.

estimates.[5] Again, the computational burden involved in computing the maximum likelihood estimates does not appear so great that one has to rely on Parks' three-step procedure, or some quasi-iterative version of it, to estimate the seemingly unrelated regression model with serially correlated residuals. The method proposed in Section 3 of this paper also has the advantage that linear restrictions on the coefficients can be easily handled and that more general auto-regressive properties of the residuals can be considered.

5. AN EXAMPLE

The model estimated in this section is the simultaneous part of the forecasting model developed in Fair [3]. The model is quarterly and consists of eight equations —seven equations explaining seven components of current dollar GNP and a GNP identity. The seven components are durable consumption, non-durable consumption, service consumption, plant and equipment investment, nonfarm housing investment, inventory investment, and imports. Government spending, exports, and farm housing investment are taken to be exogenous. The model is presented in Table I. A detailed description of the eight-equation model is presented in [3], along with a description of the overall forecasting model, and this description will not be repeated here.

The model was estimated for the 1960 I–1970 III period,[6] and the results are presented in Table II. The model was estimated both by the full information maximum likelihood technique described in Sections 2 and 3 of this paper and by the two-stage least squares technique adjusted to account for first order serial correlation of the residuals. A description of this latter technique can be found in Fair [4]. The two-stage least squares estimates were used as initial values for the maximum likelihood technique.

Given the initial two-stage least squares values, it took three iterations for the estimates of the β's and γ's to converge within a tolerance level of 0.1 percent (i.e., 0.001 percentage points). The values of h^7 for these three iterations were 0.191, 0.919, and 1.003. Given these new values of the β's and γ's, it then took two iterations for the estimates of the r's to converge within the same tolerance level. The values of h for these two iterations were 0.902 and 0.997. The resulting estimates from this first application of the two-step process are presented in Table II

[5] Parks did not propose any iterative procedure, but one of the methods considered by Kmenta and Gilbert [6] in this context is equivalent to iterating on the serial correlation coefficients. Kmenta and Gilbert did not propose iterating on the variance–covariance matrix, however. It should also be noted that Parks' procedure does not yield consistent estimates if there are lagged dependent variables among the predetermined variables in the system, since in this case not all of the estimates of the serial correlation coefficients are consistent. In this case one must resort to an iterative procedure on the serial correlation coefficients in order to achieve consistent estimates.

[6] The model could not be estimated before 1960 because of lack of good data on housing starts. Most of the equations in [3] were estimated beginning in 1956 I, but for the work here it was not possible to do this, since it was not possible to estimate some equations over different sample periods than others. For the work in [3], observations were omitted from all of the equations for the automobile strike in 1964 and observations were omitted from the import equation for the dock strike in 1968–1969. For the work here, no observations were omitted because of strikes, but rather dummy variables were used in those equations most affected by the strikes. The dummy variables used are listed in Table I. For the work in [3] the sample period ended in 1969 IV, but for the work here the sample period ended in 1970 III.

[7] See Footnote 2.

24

TABLE I

THE EIGHT EQUATION MODEL

Endogenous Variables								
	CD	CN	CS	IP	IH	$V - V_{-1}$	IMP	GNP
CD	1	0	0	0	0	0	0	β_{18}
CN	0	1	0	0	0	0	0	β_{28}
CS	0	0	1	0	0	0	0	β_{38}
IP	0	0	0	1	0	0	0	β_{48}
IH	0	0	0	0	1	0	0	β_{58}
$V - V_{-1}$	β_{61}	β_{62}	0	0	0	1	0	0
IMP	0	0	0	0	0	0	1	β_{78}
GNP	1	1	1	1	1	1	-1	1

Predetermined Variables								
	Cnst.	G	$MOOD_{-1}$	$MOOD_{-2}$	PE2	HSQ	HSQ_{-1}	HSQ_{-2}
CD	γ_{11}	0	γ_{13}	γ_{14}	0	0	0	0
CN	0	0	0	γ_{24}	0	0	0	0
CS	0	0	0	γ_{34}	0	0	0	0
IP	γ_{41}	0	0	0	γ_{45}	0	0	0
IH	γ_{51}	0	0	0	0	γ_{56}	γ_{57}	γ_{58}
$V - V_{-1}$	γ_{61}	0	0	0	0	0	0	0
IMP	γ_{71}	0	0	0	0	0	0	0
GNP	0	1	0	0	0	0	0	0

	CD_{-1}	CN_{-1}	CS_{-1}	V_{-1}	D644	D651	D684	D691	D692
CD	0	0	0	0	$\gamma_{1,13}$	$\gamma_{1,14}$	0	0	0
CN	0	$\gamma_{2,10}$	0	0	0	0	0	0	0
CS	0	0	$\gamma_{3,11}$	0	0	0	0	0	0
IP	0	0	0	0	0	0	0	0	0
IH	0	0	0	0	0	0	0	0	0
$V - V_{-1}$	γ_{69}	$\gamma_{6,10}$	0	$\gamma_{6,12}$	$\gamma_{6,13}$	$\gamma_{6,14}$	0	0	0
IMP	0	0	0	0	$\gamma_{7,13}$	$\gamma_{7,14}$	$\gamma_{7,15}$	$\gamma_{7,16}$	$\gamma_{7,17}$
GNP	0	0	0	0	0	0	0	0	0

Restrictions: $\beta_{62} = \beta_{61}$; $\quad \gamma_{6,10} = \gamma_{69}$.
R_1 in (3.1) is assumed to be a diagonal matrix with diagonal elements r_{11}, r_{22}, r_{33}, r_{44}, r_{55}, r_{66}, and r_{77}.
R_2 in (3.1) is assumed to be zero.
The eighth equation is an identity and has no error term associated with it.

Notation:
CD = Durable Consumption Expenditures
CN = Non-Durable Consumption Expenditures
CS = Service Consumption Expenditures
IP = Plant and Equipment Investment
IH = Nonfarm Housing Investment
$V - V_{-1}$ = Change in Total Business Inventories
IMP = Imports
GNP = Gross National Product
G = Government Expenditures plus Farm Housing Investment plus Exports
MOOD = Michigan Survey Research Center Index of Consumer Sentiment
PE2 = Two-quarter-ahead Expectation of Plant and Equipment Investment
HSQ = Quarterly Nonfarm Housing Starts
V = Stock of Total Business Inventories (arbitrary base period value of zero in 1953 IV)
D644 = Dummy variable that takes on a value of one in the fourth quarter of 1964 and zero otherwise. Similarly for dummy variables D651, D684, D691, and D692.

Note: The subscript -1 or -2 after a variable denotes the one-quarter or two-quarter lagged value of the variable.

TABLE II
COEFFICIENT ESTIMATES OF THE MODEL

	Two-Stage Least Squares Estimates	Full Information Maximum Likelihood Estimates			Estimated Standard Errors on Eleventh Pass
		First Pass	Fourth Pass	Eleventh Pass	
β_{18}	0.1085	0.10869	0.10897	0.10902	0.00172
β_{28}	0.0446	0.045252	0.051593	0.054768	0.007782
β_{38}	0.0212	0.027339	0.028243	0.029302	0.005707
β_{48}	0.0801	0.073693	0.072075	0.073092	0.014802
β_{58}	0.0141	0.014285	0.014340	0.014357	0.000941
β_{61}	-0.2373	0.31191	0.49909	0.55695	0.15600
β_{78}	0.0859	0.084856	0.084032	0.083556	0.003980
γ_{11}	-34.62	-38.409	-41.204	-41.272	4.936
γ_{13}	0.1830	0.19866	0.20761	0.21723	0.05162
γ_{14}	0.0695	0.09280	0.11157	0.10237	0.05769
$\gamma_{1,13}$	-2.32	-2.0664	-1.9956	-2.0604	1.1298
$\gamma_{1,14}$	2.66	3.2085	3.3190	3.2581	1.1142
γ_{24}	0.0443	0.052629	0.064539	0.070305	0.012678
$\gamma_{2,10}$	0.8297	0.82354	0.79461	0.78022	0.03289
γ_{34}	-0.0233	-0.023319	-0.023373	-0.023511	0.002783
$\gamma_{3,11}$	0.9471	0.92279	0.91924	0.91513	0.02208
γ_{41}	-9.74	-7.8993	-7.7771	-7.7009	6.9687
γ_{45}	0.4625	0.51473	0.53308	0.51534	0.13076
γ_{51}	-2.92	-3.2580	-2.7355	-2.4767	1.4295
γ_{56}	0.0660	0.056493	0.055539	0.055719	0.009949
γ_{57}	0.0869	0.094790	0.088803	0.087797	0.011565
γ_{58}	0.0146	0.017996	0.019883	0.018326	0.010238
γ_{61}	-87.69	-156.78	-170.72	-166.57	21.753
γ_{69}	0.8296	0.67225	0.55839	0.47342	0.08858
$\gamma_{6,12}$	-0.3198	-0.51523	-0.53796	-0.52106	0.05949
$\gamma_{6,13}$	-1.27	-0.077041	0.43839	0.58868	1.70955
$\gamma_{6,14}$	6.47	0.61318	-0.68113	-1.26762	1.76805
γ_{71}	-25.96	-25.136	-24.336	-23.893	3.208
$\gamma_{7,13}$	0.55	0.23623	0.22515	0.22055	0.47947
$\gamma_{7,14}$	-1.68	-2.2084	-2.2363	-2.2541	0.4948
$\gamma_{7,15}$	-1.74	-1.7990	-1.7376	-1.7466	0.5411
$\gamma_{7,16}$	-6.24	-6.3386	-6.2616	-6.3141	0.6921
$\gamma_{7,17}$	1.66	1.4813	1.6556	1.5839	0.6908
r_{11}	0.3862	0.26808	0.22098	0.23671	0.10597
r_{22}	-0.2896	-0.23956	-0.28516	-0.27019	0.08099
r_{33}	0.0139	-0.09524	-0.12780	-0.12589	0.14360
r_{44}	0.8135	0.79745	0.85375	0.88649	0.10362
r_{55}	0.3829	0.43100	0.45379	0.44281	0.13285
r_{66}	0.9101	0.86736	0.77271	0.73130	0.04200
r_{77}	0.8931	0.88680	0.87756	0.87135	0.03993
10^{-12} (Likelihood Ratio)	1.283124	1.885751	2.124221	2.170796	—

under the heading "First Pass". This first pass increased the likelihood ratio by about 47 percent from what it was for the two-stage least squares estimates. By the fourth application of the two-step process, the successive estimates of the r's were within a tolerance level of 0.025 (i.e., the difference between the estimate of r_{ii} on the third pass and the estimate of r_{ii} on the fourth pass was less than 0.025 for each i). The estimates on the fourth pass are presented in Table II. The likelihood

ratio increased by a little over one percent between the first and fourth pass. By the eleventh application of the two-step process, the successive estimates of the r's were within a tolerance level of 0.002. The estimates on the eleventh pass are also presented in Table II. Within any one application of the two-step process, it never took more than three iterations for the estimates of the β's and γ's to converge within a tolerance level of 0.1 percent, and it never took more than two iterations for the estimates of the r's to converge. Near the end, the estimates were converging in one iteration. All of the values of h were very close to one after the first application of the two-step process.

The above results thus indicate that the two-step process works quite well.[8] It is also encouraging to report that the process converged even when the initial values of the β's, γ's, and r's were all taken to be zero. In this case, it took 38 iterations for the estimates of the β's and γ's to converge the first time, with small values of h generally used for the first 34 iterations. Given these estimates of the β's and γ's, it then took eight iterations for the estimates of the r's to converge for the first time, with small values of h used for the first four iterations. The value of the likelihood ratio after this first pass was $0.750279(10^{12})$.

The above model was also estimated under the assumption that R_2 in (3.1) is a diagonal matrix. This meant that there were 47 coefficients to be estimated— 33 coefficients in B and Γ, 7 coefficients in R_1, and 7 coefficients in R_2. The full information maximum likelihood estimates in Table II were used as initial values for B, Γ, and R_1, and zeros were used as initial values for R_2. Given the initial values, it took three iterations for the r's to converge within a tolerance level of 0.1 percent. The values of h for these three iterations were 0.716, 1.000, and 1.000. After this first pass, it never took more than two iterations for the estimates of the β's and γ's or of the r's to converge. All of the values of h were very close to one. By the eighth pass the successive estimates of the r's were within a tolerance level of 0.008. The likelihood ratio after the eighth pass was $3.488035(10^{12})$.

The technique described in this paper thus appears capable of handling fairly large problems with no difficulty. No problems of convergence were encountered with any of the runs using the above model. There is no indication from the above results that the technique cannot handle problems even double or triple the size of the current problems. With respect to Hendry's use of Powell's algorithm [5], it should perhaps be mentioned that Powell [9, p. 34] has reported that the algorithm tends to be inefficient for more than about ten parameters. However, the computational efficiencies of the algorithm here proposed, as compared with possible alternatives, remain to be further investigated.[9]

6. Conclusion

Most of the problems involved in estimating linear econometric systems can be handled by the method described in Sections 2 and 3 of this paper. A user-oriented computer program has been written to implement the method and

[8] It is conceivable, of course, that one could achieve even faster overall convergence by using larger tolerance levels for the first few passes and then smaller levels after the estimates are close to converging.

[9] It should be noted with respect to Hendry's method that Hendry considered only the case of completely unrestricted autoregressive coefficient matrices (i.e., no zero elements).

is available from the authors on request.[10] To the extent that such a program is available, one should not have to rely on less satisfactory, but computationally easier, procedures to estimate linear econometric systems.

Princeton University

REFERENCES

[1] Chow, Gregory C. "A Comparison of Alternative Estimators for Simultaneous Equations," *Econometrica* XXXII (October, 1964), 532–553.
[2] ———. "Two Methods of Computing Full-Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations," *International Economic Review* IX (February, 1968), 100–112.
[3] Fair, Ray C. *A Short-Run Forecasting Model of the United States Economy*, Lexington: D. C. Heath and Company, 1971.
[4] ———. "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," *Econometrica* XXXVI (May, 1970), 507–516.
[5] Hendry, D. F. "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process," *International Economic Review* XII (June, 1971).
[6] Kmenta, Jan and Roy F. Gilbert. "Estimation of Seemingly Unrelated Regressions with Autoregressive Disturbances," *Journal of the American Statistical Association* LXV (March, 1970), 186–197.
[7] Parks, R. W. "Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated," *Journal of the American Statistical Association* LXII (June, 1967), 500–509.
[8] Powell, M. J. D. "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives," *Computer Journal* VII (1964), 155–162.
[9] Powell, M. J. D. "Recent Advances in Unconstrained Optimization," paper presented at the 7th Mathematical Programming Symposium 1970, The Hague, The Netherlands.
[10] Sargan, J. D. "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals," *Econometrica* XXXIX (July, 1961), 414–426.
[11] Zellner, A. "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* LVII (June, 1962), 348–368.
[12] Zellner, A. and H. Theil. "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica* XXX (January, 1962), 54–78.

[10] The current program (in FORTRAN IV) was written by Douglas R. Chapman. An earlier program, not incorporating restrictions across equations nor autoregressive properties of the error terms, was written by Richard Levitan. The current program is described in Douglas R. Chapman and Ray C. Fair, "Full-Information Maximum Likelihood Program: User's Guide," Research Memorandum No. 137, Econometric Research Program, Princeton University, April 1972.