

IDENTIFICATION AND ESTIMATION
IN ECONOMETRIC SYSTEMS: A SURVEY*

Gregory C. Chow

Econometric Research Program
Research Memorandum No. 167
May 1974

The research described in this paper
was supported by NSF Grant GS 32003X.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

IDENTIFICATION AND ESTIMATION IN ECONOMETRIC SYSTEMS:
A SURVEY*

Gregory C. Chow

This is an introductory survey of some of the ideas and methods in the identification and estimation of simultaneous equation systems in econometrics. After pointing out the special features of econometric systems, it defines the problem of identification and presents several methods for estimating the parameters in such systems. Hopefully such a survey will be useful to research workers in related fields including control engineering and statistics who are interested in the estimation of dynamic systems and wish to find out whether the works of econometricians are relevant to their own research. This is not a substitute for a treatise in econometrics,¹ but it may help a researcher in a related field decide whether the techniques developed for dynamic econometric systems are useful for his purpose, whether he should study them in depth, and whether he can contribute to improving them. Some research topics will be suggested later in our discussion.

1. Nature of Simultaneous Econometric Systems

Consider a linear system of p equations determining a vector y_t of p dependent or endogenous variables at time t :

$$(1.1) \quad By_t = \Gamma z_t + \epsilon_t \quad (t=1, \dots, T)$$

where z_t is a vector of q predetermined variables defined to include lagged endogenous variables y_{t-1}, y_{t-2}, \dots (so that the system may be dynamic and of a high order) as well as current and lagged exogenous variables (some exogenous variables being control variables), and ϵ_t is a random vector with mean zero and covariance matrix $E\epsilon_t\epsilon_t' = \delta_{ts} \Sigma$. B and Γ are matrices of coefficients, B being nonsingular. The predetermined variables z_t are assumed to be distributed independently of ϵ_t , with a second-order moment matrix $\frac{1}{T} \sum_{t=1}^T z_t z_t'$ converging in probability to a constant matrix as T increases. Several features of the system (1.1) should be noted.

(1) Each equation, or row of (1.1), usually contains several endogenous variables which are correlated with the residual vector ϵ_t . The system (1.1) is called structural equations, to be distinguished from the reduced-form equations

$$(1.2) \quad y_t = B^{-1} \Gamma z_t + B^{-1} \epsilon_t = \Pi z_t + \eta_t$$

each of which contains only one endogenous variable as a function of the predetermined variables and ϵ_t , the coefficients Π being defined as $B^{-1} \Gamma$. The reduced form (1.2) shows that all elements of y_t are correlated with ϵ_t . As an example of a structural equation (a consumption function), we may have total consumption expenditures and national income as two of the endogenous variables and consumption at $t-1$ as one of the predetermined variables. For examples of econometric systems, the reader may refer to Nerlove (1966).

(2) Although the system (1.1) may be fairly large, sometimes containing over one hundred equations, the number of variables in each equation is usually small, most often below 10. The matrices B and Γ are thus sparse.

(3) The endogenous variables, which correspond to the state variables of the control engineers, are most frequently assumed to be directly observed without measurement errors.

(4) It is important to estimate the parameters B , Γ and Σ of the structural equations (1.1), and not only the parameters Π and $E\eta_t\eta_t'$ of the reduced-form equations (1.2). The main reason is that economic hypotheses such as the relation of consumption expenditures to after-tax income are formulated in the form of structural equations. If there is any change in economic institutions, technological relations, or behavioral patterns of the economic agents as described by the structural equations, one is able to assess its impact only by modifying the structural equation affected. For example, if consumption expenditures y_{1t} depends on after-tax income $(1-g)y_{2t}$, where y_{2t} is before-tax income and g is the tax rate, knowledge of this structural equation is required to assess the effects of a change in the tax rate from g_1 to g_2 . Knowledge of the reduced-form equations (1.2) is not sufficient for this purpose.

(5) Economic hypotheses are mainly qualitative in character. They help specify the important variables which should appear

in each structural equation and frequently also the signs of their coefficients but not the magnitudes. Limited historical data are employed for the estimation of all the unknown parameters in the system (1.1).

To summarize, by not dealing with possible observation errors in the variables, the econometrician has made his estimation problem easier than that of the control scientist. Note, however, that errors of measurement were considered an important problem in the early development of the simultaneous equations model in econometrics, as evidenced in Chapter I of Koopmans, ed. (1950).² On the other hand, features (1) and (2) mentioned above make the estimation of B and Γ more difficult than the parameters Π in the reduced-form equations (1.2). Even if only one endogenous variable appears in each equation of a system, different sets of predetermined variables appearing in different equations will make the method of least squares applied to each equation separately inefficient, as pointed out by Zellner (1962), unless the error terms in separate equations are themselves uncorrelated. In addition, with the presence of several endogenous variables in each equation which are correlated with the residual, the method of least squares yields inconsistent estimates.

2. The Identification Problem

Before studying methods for estimating the unknown parameters B , Γ and Σ in (1.1), one needs to impose restrictions on the parameters to insure their identifiability. A set of structural parameters is said to be identifiable iff there exists no other set

which will give rise to the same probability distribution of the endogenous variables. For the linear structure (1.1), the probability distribution of the endogenous variables is given by the reduced form (1.2), with mean vector Πz_t and covariance matrix $E\eta_t \eta_t'$. If there exist two sets of values for the structural parameters from which the same reduced-form is deduced, the structural parameters are unidentifiable. In this case, no consistent estimator for the set of parameters exists. One can estimate the reduced-form parameters Π and $E\eta_t \eta_t'$ consistently by the method of least squares, but from these parameters one cannot infer uniquely the parameters B , Γ and Σ of the structure.

To illustrate, consider the identifiability of the coefficients in the i^{th} structural equation through the absence of certain variables in that equation. Let the coefficients be elements of the vector

$$(2.1) \quad (1 \quad \beta_i \quad 0; \quad \gamma_i \quad 0)$$

where we have normalized the coefficient of y_{it} to be unity and rearranged the variables y_t and z_t in the system (1.1) so that y_{it} appears first and those absent from equation i appears last. After the rearrangement (2.1) becomes the first row of $(B \quad \Gamma)$. Let p_2 endogenous variables and q_2 predetermined variables be absent from this equation, with $p = p_1 + p_2$ and $q = q_1 + q_2$. Using the first row of the relation

$$(2.2) \quad B\Pi = \Gamma$$

between the structural and reduced-form parameters and partitioning Π accordingly, we have

$$(2.3a) \quad \begin{matrix} p_1 \times q_1 \\ (1 \ \beta_i) \end{matrix} \Pi_{11} = \begin{matrix} 1 \times q_1 \\ \gamma_i \end{matrix} ;$$

$$(2.3b) \quad \begin{matrix} p_1 \times q_2 \\ (1 \ \beta_i) \end{matrix} \Pi_{12} = \begin{matrix} 1 \times q_2 \\ 0 \end{matrix} .$$

Given β_i and Π_{11} , (2.3a) determines γ_i . To determine β_i uniquely from Π_{12} using (2.3b), there are (p_1-1) unknowns in q_2 linear equations. It is therefore necessary that $q_2 \geq p_1-1$, or $q_2 + p_2 \geq p_1 + p_2 - 1$. Thus the number $p_2 + q_2$ of variables excluded from the equation must equal or exceed the total number of endogenous variables (or simultaneous equations) minus one. If the number of excluded variables is smaller, the equation is unidentifiable. If it equals $p-1$, the equation is just-identified. If it exceeds $p-1$, the equation is overidentified. The identification problem has been studied in a more general setting covering non-linear econometric systems by Fisher (1966) Wegge (1965) and Rothenberg (1971).

3. Extensions of the Method of Least Squares

The method of least squares can be applied to the reduced form (1.2) to obtain a consistent estimate of Π . If the random residual ϵ_t , and thus $\eta_t = B^{-1}\epsilon_t$, is assumed to be normal, the least squares estimate is also the maximum likelihood estimate.

However, it has not incorporated the nonlinear restrictions on the elements of Π resulting from linear restrictions on B and Γ through the relation (2.2). For example, if equation i is overidentified with $q_2 > p_1 - 1$, the rank of Π_{12} in (2.3b) is restricted to be $p_1 - 1$.

To estimate the parameters of the i^{th} structural equation, one can first obtain the least-squares estimate $\hat{\Pi}$ of the reduced-form parameters Π and, using these, solve equation (2.3) for β_i and γ_i . If the equation is just-identified, (2.3b) contains exactly $p_1 - 1$ linear equations for the same number of unknowns in β_i . The method just described is known as indirect least squares. If the equation is overidentified, there are more equations (or columns of $\hat{\Pi}_{12}$) than unknowns. One can arbitrarily select any $p_1 - 1$ columns of $\hat{\Pi}_{12}$ to solve for β_i . This method is still consistent, but it does not utilize efficiently all the information in $\hat{\Pi}_{12}$.

An asymptotically more efficient method of estimating the parameters of one over-identified structural equation is that of two-stage least squares proposed by Theil (1953) and Basmann (1957). By this method one first obtains "estimated" values of the endogenous variables $\hat{y}_t = \hat{\Pi}z_t$ from the least squares estimates $\hat{\Pi}$ of Π . In the second stage, least squares is applied to the structural equation using the variables \hat{y}_{jt} and z_{jt} included in that equation. The motivation is that, replacing y_{jt} by \hat{y}_{jt} , one avoids the correlations between the explanatory variables and the random residual of that equation. Let the T observations

of the i^{th} structural equation be written as

$$(3.1) \quad Y_i = Y_i \beta_i + Z_i \gamma_i + \epsilon_i \equiv X_i \alpha_i + \epsilon_i .$$

Note the change of notations from here on to conform to the more standard notations in the estimation literature, with β_i and γ_i denoting column vectors of unknown coefficients in equation i , Y_i denoting a $T \times 1$ vector of observations on the first endogenous variable, Y_i and Z_i respectively denoting $T \times (p_i - 1)$ and $T \times q_i$ matrices of observations on the remaining endogenous variables and the predetermined variables included in equation i , and ϵ_i denoting a $T \times 1$ vector of observations on the residual. Letting Z be the $T \times q$ matrix of all predetermined variables, we have

$$(3.2) \quad \hat{Y}_i = Z(Z'Z)^{-1}Z'Y_i ;$$

$$\hat{X}_i = (\hat{Y}_i \quad Z_i) = Z(Z'Z)^{-1}Z'(Y_i \quad Z_i) .$$

The two-stage least squares (2SLS) estimate of α_i is

$$(3.3) \quad \alpha_i^{(2)} = (\hat{X}_i' \hat{X}_i)^{-1} \hat{X}_i' Y_i = (\hat{X}_i' X_i)^{-1} \hat{X}_i' Y_i .$$

Solving (3.3) for the two components of α_i , we have

$$(3.4) \quad \gamma_i^{(2)} = (Z_i' Z_i)^{-1} Z_i' (Y_i - Y_i \beta_i)$$

and, letting $M = I - Z(Z'Z)^{-1}Z'$ and $M_i = I - Z_i(Z_i'Z_i)^{-1}Z_i'$,

$$\begin{aligned}
 (3.5) \quad \beta_i^{(2)} &= [\hat{Y}_i' \hat{Y}_i - Y_i' Z_i (Z_i' Z_i)^{-1} Z_i' Y_i]^{-1} \hat{Y}_i' Y_i \\
 &= [Y_i' M_i Y_i - Y_i' M Y_i]^{-1} Y_i' [M_i - M] Y_i .
 \end{aligned}$$

(3.4) and (3.5) will be useful for comparison with other estimation methods to be presented later on.

The method of 2SLS is a "limited-information" method in the sense that information on the specification of all other structural equations is not utilized in the estimation of equation i , except for the list of all predetermined variables Z appearing in the system. A "full-information" method requires specifying all structural equations, and it is applied to the estimation of all parameters simultaneously. One such method is three-stage least squares of Zellner and Theil (1962). Let the p equations be set up by the notation of (3.1) as

$$(3.6) \quad \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ & & \ddots & \\ 0 & \dots & & X_p \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{bmatrix} .$$

By replacing each X_i in (3.4) by its least-squares estimated value $\hat{X}_i = Z(Z'Z)^{-1}Z'X_i$ as given by (3.2), and by applying Aitken's generalized least squares to the resulting system, using the Kronecker product $\Sigma \otimes I_T$ as the covariance matrix of the residuals, one obtains the estimating equations

$$(3.7) \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \stackrel{(3)}{=} \begin{bmatrix} \sigma^{11} \hat{X}'_1 \hat{X}_1 & \sigma^{12} \hat{X}'_1 \hat{X}_2 & \dots & \sigma^{1p} \hat{X}'_1 \hat{X}_p \\ \sigma^{21} \hat{X}'_2 \hat{X}_1 & \sigma^{22} \hat{X}'_2 \hat{X}_2 & \dots & \sigma^{2p} \hat{X}'_2 \hat{X}_p \\ \dots & \dots & \dots & \dots \\ \sigma^{p1} \hat{X}'_p \hat{X}_1 & \sigma^{p2} \hat{X}'_p \hat{X}_2 & \dots & \sigma^{pp} \hat{X}'_p \hat{X}_p \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11} \hat{X}'_1 & \sigma^{12} \hat{X}'_1 & \dots & \sigma^{1p} \hat{X}'_1 \\ \sigma^{21} \hat{X}'_2 & \sigma^{22} \hat{X}'_2 & \dots & \sigma^{2p} \hat{X}'_2 \\ \dots & \dots & \dots & \dots \\ \sigma^{p1} \hat{X}'_p & \sigma^{p2} \hat{X}'_p & \dots & \sigma^{pp} \hat{X}'_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

where σ^{ij} denotes the i - j element of Σ^{-1} . Since Σ is unknown, the sample covariance matrix S of the residuals obtained by applying 2SLS to the structural equations can be used to approximate Σ . If $s^{-1} = (s^{ij})$ replaces (σ^{ij}) in the estimating equations (3.7) the resulting estimates are known as three-stage least squares (3SLS) estimates.

4. Applications of the Method of Maximum Likelihood

By assuming the residual vector ϵ_t in (1.1) to be p -variate normal, one can apply the method of maximum likelihood to estimate the unknown parameters. The limited-information maximum likelihood (LIML) method of Anderson and Rubin (1949) estimates the parameters pertaining to one structural equation, subject to the identifiability restrictions. The likelihood function is that of a multivariate normal regression, given by the reduced-form, of the endogenous variables included in the i^{th} structural equation to be estimated on all predetermined variables, with Π_{11} and Π_{12} of equation (2.3) as coefficients. This likelihood function is combined with the restrictions given by (2.3b) to form a Lagrangian

expression which can be maximized with respect to Π_{11} , Π_{22} , and β_i ; γ_i can be determined by (2.3a). As pointed out by Hood and Koopmans (1953, p. 168), this constrained maximization problem is equivalent to that of minimizing the ratio

$$(4.1) \quad \lambda_i = \frac{\|y_i \beta_{ii} - Y_i \beta_i - z_i \gamma_i\|^2}{\|M(y_i \beta_{ii} - Y_i \beta_i)\|^2}.$$

The numerator of (4.1) is the variance of the residual of (3.1) prior to the normalization $\beta_{ii} = 1$. The denominator is the variance of the residuals of the linear combinations $y_i \beta_{ii} - Y_i \beta_i$ of the endogenous variables in equations i around their estimated values by the reduced-form equations. Setting to zero the derivatives of (4.1) with respect to γ_i , β_{ii} and β_i , we obtain

$$(4.2) \quad \gamma_i = (z_i' z_i)^{-1} z_i' (y_i \beta_{ii} - Y_i \beta_i),$$

which should be compared with (3.4), and

$$(4.3) \quad [(y_i \quad Y_i)' M_i (y_i \quad Y_i) - \lambda_i (y_i \quad Y_i)' M (y_i \quad Y_i)] \begin{pmatrix} \beta_{ii} \\ -\beta_i \end{pmatrix} = 0.$$

The minimization problem is solved by finding λ_i as the smallest root of the determinantal equation

$$(4.4) \quad |(y_i \quad Y_i)' M_i (y_i \quad Y_i) - \lambda_i (y_i \quad Y_i)' M (y_i \quad Y_i)| = 0$$

and obtaining β_i from equation (4.3) with β_{ii} set equal to 1,

$$(4.5) \quad \beta_i = [Y_i' M_i Y_i - \lambda_i Y_i' M Y_i]^{-1} Y_i' [M_i - \lambda_i M] y_i.$$

If λ_i equals 1, the LIML estimator (4.5) will be identical with the 2SLS estimator of (3.5).

Theil (1958, Chapter 6) proposed a k-class estimator defined by

$$(4.6) \quad \begin{bmatrix} Y_i' Y_i - k Y_i' M Y_i & Y_i' Z_i \\ Z_i' Y_i & Z_i' Z_i \end{bmatrix} \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} = \begin{bmatrix} (Y_i' - k Y_i' M) Y_i \\ Z_i' Y_i \end{bmatrix}$$

which implies (3.4) and

$$(4.7) \quad \beta_i = [Y_i' M_i Y_i - k Y_i' M Y_i]^{-1} Y_i' [M_i - k M] Y_i$$

Thus both 2SLS and LIML are members of the k-class, the former with $k=1$ and the latter with k equal to the smallest root of (4.4).

The method of full-information maximum likelihood (FIML) maximizes the log-likelihood function for the system

$$(4.8) \quad L = \text{const} - \frac{T}{2} \log |\Sigma| + T \log |B| - \frac{T}{2} \text{tr} \left[\frac{1}{T} \Sigma^{-1} (YB + Z\Gamma)' (YB + Z\Gamma) \right].$$

On differentiation with respect to Σ^{-1} , B and Γ , one obtains respectively

$$(4.9) \quad 2 \cdot \frac{\partial L}{\partial \Sigma^{-1}} = T \Sigma - (YB + Z\Gamma)' (YB + Z\Gamma) = 0$$

$$(4.10) \quad \frac{\partial L}{\partial B} = - Y' (YB + Z\Gamma) \Sigma^{-1} + T (B')^{-1}$$

$$(4.11) \quad \frac{\partial L}{\partial \Gamma} = - Z' (YB + Z\Gamma) \Sigma^{-1}$$

Equation (4.9) can be used to solve for Σ in terms of B and Γ . Note that in (4.10) and (4.11), only those derivatives of L with respect to the unknown elements of B and Γ are set equal to zero. Thus the derivatives with respect to the unknown coefficients in the i^{th} equation $y_i = Y_i \beta_i + Z_i \gamma_i + \epsilon_i$ are

$$(4.12) \quad \frac{\partial L}{\partial \beta_i} = - Y_i' \sum_{h=1}^p \sigma^{hi} (y_h - Y_h \beta_h - Z_h \gamma_h) + T \beta^{i(i)} = 0$$

where σ^{hi} is the h - i element of Σ^{-1} and $\beta^{i(k)}$ denotes a column vector consisting of those elements of the i^{th} row of B^{-1} which correspond to the unknown elements of β_k , and

$$(4.13) \quad \frac{\partial L}{\partial \gamma_i} = - Z_i' \sum_{h=1}^p \sigma^{hi} (y_h - Y_h \beta_h - Z_h \gamma_h) = 0.$$

As reported in Chow (1968), one computationally efficient method to solve (4.12) and (4.13) for $\alpha' = (\alpha_1' \cdots \alpha_p')$ = $(\beta_1', \gamma_1', \dots, \beta_p', \gamma_p')$ is Newton's method

$$(4.14) \quad \alpha^{r+1} = \alpha^r + h_r H(\alpha^r)^{-1} g(\alpha^r)$$

where α^r denotes the value of the unknown vector at the r^{th} iteration, $g(\alpha)$ is the gradient of L as given by (4.12) and (4.13), $H(\alpha)$ is the Hessian matrix given by

$$(4.15) \quad \frac{\partial^2 L}{\partial \beta_i \partial \beta_j} = - \sigma^{ji} Y_i' Y_j + \frac{1}{T} Y_i' \sum_{h=1}^p \epsilon_h \sum_{n=1}^p (\sigma^{hj} \sigma^{ni} + \sigma^{hn} \sigma^{ji}) \epsilon_n' Y_j - T \beta^{j(i)} \beta^{i(j)'}.$$

$$(4.16) \quad \frac{\partial^2 L}{\partial \beta_i \partial \gamma_i} = - \sigma^{ji} Y_i' Z_j + \frac{1}{T} Y_i' \sum_{h=1}^p \epsilon_h \sum_{n=1}^p (\sigma^{hj} \sigma^{ni} + \sigma^{hn} \sigma^{ji}) \epsilon_n' Z_j,$$

with (σ^{ij}) denoting $[\frac{1}{T}(YB + Z\Gamma)'(YB + Z\Gamma)]^{-1}$, and h_r is a scalar chosen to promote convergence of the iterative process. Of course, (σ^{ij}) changes from iteration to iteration. If σ^{ij} were replaced by a consistent estimate, as in the method of 3SLS, and if B^{-1} were also replaced by a consistent estimate, equations (4.12) and (4.13) would be a system of linear equations in α_i similar to the system (3.7) for the method of 3SLS. Rothenberg and Leenders (1964) has suggested a linearized FIML estimator obtained by performing only one iteration of Newton's method and using a consistent estimate of α as initial value for the iteration.

In the special case of a structure with a triangular B and a diagonal Σ , known as a recursive or causal chain system, the estimating equations (4.12) and (4.13) will be reduced to those of the method of least squares applied to each equation individually, since $\beta^{(i)i} = 0$ and $\sigma^{hi} = 0$ for $i \neq h$. If B is triangular (with $\beta^{(i)i} = 0$) but Σ is not diagonal, (4.12) and (4.13) are equivalent to the estimating equations of Aitken's generalized least squares applied to the entire system (3.6). The method of ordinary least squares applied to each equation separately will still be consistent but no longer efficient. The recursive system has been studied extensively by Wold (1954, 1964).

5. Method of Instrumental Variables

The method of instrumental variables originally suggested by Reiersøl (1945) and Geary (1949) employs an estimating equation of the form

$$(5.1) \quad \hat{\alpha} = (W'X)^{-1}W'y$$

where W is a matrix of instruments assumed to be uncorrelated with the residuals of the stochastic equations. For the estimation of β_i in equation (3.1) by a method belonging to the k -class, we can interpret the matrix $W_i = (M_i - kM)Y_i$ in equation (4.7) as a set of instruments. Here the matrix M may be replaced by $M^* = I - Z^*(Z'^*Z^*)^{-1}Z'^*$ where Z^* is a matrix of selected major principal components of Z as suggested by Klock and Mennes (1960). This method also admits of an instrumental variables interpretation.

As pointed out by Hausman (1974), the method of full-information maximum likelihood can also be interpreted as a method of instrumental variables. Substituting $(YB + Z\Gamma)'(YB + Z\Gamma)\Sigma^{-1}$ for $T \cdot I$ in (4.10) and simplifying, one obtains

$$(5.2) \quad \frac{\partial L}{\partial B} = (B')^{-1}\Gamma'Z'(YB + Z\Gamma)\Sigma^{-1} = (Z\Pi)'(YB + Z\Gamma)\Sigma^{-1}$$

where $\Pi = \Gamma B^{-1}$ is the matrix of reduced-form coefficients and $Z\Gamma B^{-1} = \tilde{Y}$ is a matrix of estimated values of the endogenous variables by the reduced-form. Selecting from (5.2) the derivatives with respect to the unknown vector β_i , we have

$$(5.3) \quad \frac{\partial L}{\partial \beta_i} = - \tilde{Y}'_i (\mathbf{YB} + \mathbf{Z}\Gamma) \Sigma^{-1} = - \tilde{Y}'_i \sum_{h=1}^p \sigma^{hi} (y_h - Y_h \beta_h - Z_h \gamma_h) = 0$$

where \tilde{Y}'_i is a $T \times (p_i - 1)$ matrix of estimated values of Y_i selected from the matrix $\tilde{Y} = \mathbf{Z}\Gamma\mathbf{B}^{-1}$. Denoting $(\tilde{Y}'_i \quad Z'_i)$ by \tilde{X}'_i , we can rewrite (5.3) and (4.13) as

$$(5.4) \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \sigma^{11} \tilde{X}'_1 X_1 & \sigma^{12} \tilde{X}'_1 X_2 & \dots & \sigma^{1p} \tilde{X}'_1 X_p \\ \sigma^{21} \tilde{X}'_2 X_1 & \sigma^{22} \tilde{X}'_2 X_2 & \dots & \sigma^{2p} \tilde{X}'_2 X_p \\ \dots & \dots & \dots & \dots \\ \sigma^{p1} \tilde{X}'_p X_1 & \sigma^{p2} \tilde{X}'_p X_2 & \dots & \sigma^{pp} \tilde{X}'_p X_p \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11} \tilde{X}'_1 & \sigma^{12} \tilde{X}'_1 & \dots & \sigma^{1p} \tilde{X}'_1 \\ \sigma^{21} \tilde{X}'_2 & \sigma^{22} \tilde{X}'_2 & \dots & \sigma^{2p} \tilde{X}'_2 \\ \dots & \dots & \dots & \dots \\ \sigma^{p1} \tilde{X}'_p & \sigma^{p2} \tilde{X}'_p & \dots & \sigma^{pp} \tilde{X}'_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

The estimating equation (5.4) for FIML is very similar to (3.7) for 3SLS. In (5.4), the reduced-form coefficients $\Pi = \Gamma\mathbf{B}^{-1}$ used to calculate \tilde{Y}'_i are obtained from the structural coefficients \mathbf{B} and Γ , whereas in (3.7) the reduced-form coefficients $\hat{\Pi} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}$ used to calculate \hat{Y}'_i are obtained by regression on the predetermined variables \mathbf{Z} . Note that in (3.7) $\hat{X}'_i \hat{X}'_j$ can be written as $\hat{X}'_i X_j$. Both (5.4) and (3.7) can be interpreted as methods of instrumental variables. The matrix \mathbf{W}' of instruments in (5.4) is the matrix in curly brackets. The matrix \mathbf{X} of explanatory variables for both is given in equation (3.6). The reduced-form coefficients $\Pi = \Gamma\mathbf{B}^{-1}$ employed in computing the instruments for FIML take fully into account the over-identifying restrictions imposed on the system whereas the coefficients $\hat{\Pi} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}$

employed in 3SLS do not. Brundy and Jorgenson (1971), Dhrymes (1971), and Lyttkens (1970) have suggested using some consistent estimates of B and Γ to compute $\Pi = \Gamma B^{-1}$ and Σ^{-1} (or simply the identity matrix as Σ in the case of Lyttkens) for use in the estimating equation (5.4). These are known as full-information instrumental variable estimators (FIVE). If one continues to iterate and the estimates converge, they become the FIML estimates. However, the convergence properties of using (5.4) for computing FIML, as compared with the Newton method described earlier, remain to be investigated.

6. Other Methods of Estimation

We have just surveyed three of the major approaches to the estimation of econometric systems within the framework of classical statistics. They are by no means exhaustive even within the classical framework. Other estimators can be found in the texts cited in footnote 1. Besides, we should at least mention two other approaches which have been considered by econometricians.

The first is the application of Bayesian statistical methods. Through the use of prior information in the form of a prior density function for the parameters in an econometric model, the Bayesian approach has provided a generalization of the concept of identifiability of the parameters and, when analytical or numerical integration can be performed, is capable of yielding exact, finite-sample results on the distribution of the unknown parameters as

expressed in their posterior density functions. However, for large-scale econometric models, the Bayesian approach has yet to produce computationally feasible method of estimation; its contribution so far has lied mainly in conceptual clarification and providing an alternative justification of some classical methods such as that of maximum likelihood when the prior density used is diffuse. The reader may consult Dreze and Morales (1970), Dreze (1971) and Zellner (1971) concerning the Bayesian approach.

The second approach is the use of robust estimation techniques. Rather than minimizing the sum of squares of the residuals or, in the case of a multivariate system, perhaps the determinant of the matrix consisting of sums of squares and cross-products of the residuals in different equations, a robust method replaces the square or cross-product terms by some other function of the residuals which are less sensitive to extreme values. The objective is to reduce the influence of extreme values on the estimates and to prevent a few outliers, or possibly erroneous observations, from seriously affecting the estimates. One approach to robust estimation is to replace the residuals in the various generalizations of least-squares or maximum likelihood for simultaneous econometric equations by a set of weighted residuals, with the weight depending on the size of the residual. Modifications to the estimating equations based on the method of maximum likelihood have been presented in Chow (1973) and some results on comparing several classical and robust estimators have been presented by Fair (1974).

7. Properties of the Estimators

Concerning large-sample properties of the estimators, the methods of full-information maximum likelihood, linearized FIML, three-stage least squares, and full-information instrumental variables are all asymptotically efficient, the method of 3SLS being less so if there are a priori restrictions on the covariance matrix Σ of the residuals. The asymptotic covariance matrix is given by the inverse of the matrix in the estimation equations for 3SLS or the inverse of the Hessian for FIML. The limited-information estimators including LIML, 2SLS and the k-class (with $\sqrt{n}(k-1)$ converges to zero) have the same asymptotic covariance matrix.

Results on small-sample properties have been obtained by Monte Carlo studies, several of which are summarized in Johnston (1972, pp. 408-420). The studies cited, together with Mikhail (1973), though somewhat inconclusive and subject to the special characteristics of the experiments, confirm the asymptotic result that the full-information methods are more efficient if the model is correctly specified and if the covariance matrix Σ is not very close to being diagonal. If Σ is diagonal, 3SLS reduces to 2SLS and the latter will be more efficient because it is equivalent to the former method combined with correct specification of the off-diagonal elements of Σ . Numerous analytical studies have also been made concerning finite-sample distributions of the limited-information estimators in the form of approximate or exact

distribution functions. The latter are mostly confined to special cases of two to three equations in the system, or only two endogenous variables in the equation to be estimated, as exemplified by the study of Anderson and Sawa (1973).

One interesting Monte Carlo study is that of Zellner (1971, pp. 276-287) in which classical estimators are compared with a Bayesian estimator using a very simple 2-equation model

$$(7.1) \quad \begin{aligned} Y_{1t} &= \gamma Y_{2t} + \epsilon_{1t} \\ Y_{2t} &= \beta x_t + \epsilon_{2t} \end{aligned} \quad (t=1, \dots, T)$$

with the reduced-form

$$(7.2) \quad \begin{aligned} Y_{1t} &= \pi_1 x_t + \eta_{1t} \\ Y_{2t} &= \pi_2 x_t + \eta_{2t} \end{aligned} \quad (t=1, \dots, T)$$

where $\pi_1 = \beta \gamma$, $\pi_2 = \beta$. Since this model is just-identified, the classical estimators for γ including indirect least squares, 2SLS, 3SLS, LIML and FIML are all identical; it is $\hat{\gamma} = \hat{\pi}_1 / \hat{\pi}_2$, $\hat{\pi}_1$ and $\hat{\pi}_2$ being ordinary least squares estimates of π_1 and π_2 in (7.2). To obtain a Bayesian estimate of γ , Zellner first finds the joint posterior density of π_1 and π_2 , transforms it to a joint posterior density of γ and β , and uses as a point estimate the modal value of the marginal posterior density of γ . For samples of size 20, the Bayesian estimator turns out to be distinctly superior to the classical estimator (in the sense of having a sampling distribution more highly concentrated about the true

parameter value). The superiority gradually disappears as the sample size increases from 40, 60, to 100.

Zellner (p. 286) concludes that "Bayesian procedures produced better results than sampling-theory estimation procedures." The main lesson from his experiments, however, appears to be concerned with alternative methods of forming sampling-theory estimates rather than the distinction between sampling-theory and Bayesian estimates. The reason why the classical or sampling-theory estimator behaves poorly in small samples is that it is a nonlinear function (in fact a ratio) of the least-squares estimates $\hat{\pi}_1$ and $\hat{\pi}_2$ of the reduced-form parameters. If $\hat{\pi}_1$ and $\hat{\pi}_2$ are efficient estimators of π_1 and π_2 respectively, their ratio $\hat{\pi}_1/\hat{\pi}_2$ is not an efficient estimator for $\gamma = \pi_1/\pi_2$ in small samples. The ratio is biased and does not have the smallest variance around the true parameter value as Zellner's experiment demonstrates. Therefore, in so far as the classical estimators for the structural parameters can be transformed through the relation $\Pi = \Gamma B^{-1}$ to produce nearly efficient estimators of Π , they themselves are not nearly efficient estimators for B and Γ . The development of estimators for structural parameters with optimal small-sample properties requires further research using both the Bayesian and the classical approaches. Note, however, that if the ultimate use of econometric model is to predict the behavior of the endogenous variables given the predetermined variables, with or without changes in structure, the main concern

is with the properties of the reduced-form parameters as they are obtained by solving the structural equations, rather than the properties of the structural parameters themselves.

The robust estimators mentioned in section 6 are only proposed procedures of estimation. Their sampling properties also remain to be fully investigated by analytical or Monte Carlo methods.

8. Nonlinear Econometric Systems

For a nonlinear system with additive random residuals, we write the t^{th} observation on the i^{th} structural equation as

$$(8.1) \quad f_i(y_{1t}, \dots, y_{pt}, z_{1t}, \dots, z_{qt}; \beta_i) \quad (i=1, \dots, p) \\ = -y_{it} + \phi_i(y_{1t}, \dots, y_{i-1,t}, y_{i+1,t}, \dots; \beta_i) = \epsilon_{it} \quad (t=1, \dots, T)$$

β_i being a vector of unknown parameters. Various methods for estimating a nonlinear system have been proposed. For a general treatment of nonlinear estimation in econometrics, the reader may refer to Goldfeld and Quandt (1973).

If the residuals ϵ_{it} ($i=1, \dots, p$) are multivariate normal with mean zero and covariance matrix Σ , and are serially uncorrelated, the method of full-information maximum likelihood can be applied. Let ϵ_{it} denote also the function f_i of β_i given by (8.1) evaluated at the t^{th} observation, and ϵ_i denote a $T \times 1$ vector consisting of $\epsilon_{i1}, \dots, \epsilon_{it}$. The maximum likelihood estimate

of Σ is obtained by $\Sigma = (\sigma_{ij}) = (\frac{1}{T} u_i' u_j)$, interpreted as a matrix function of the T observations and the parameter vectors β_1, \dots, β_p . The log-likelihood function is

$$(8.2) \quad L = \text{const} - \frac{T}{2} \log|\Sigma| + \sum_{t=1}^T \log|J_t|$$

where J is the matrix of the Jacobian with $\frac{\partial f_i}{\partial y_j}$ as its i - j element and J_t denotes this matrix evaluated at the t^{th} observation. One can find the gradient vector $\frac{\partial L}{\partial \beta_i}$ and the Hessian matrix $\frac{\partial^2 L}{\partial \beta_i \partial \beta_j}$ in terms of the various first and second derivatives of f_i and $\frac{\partial f_i}{\partial y_j}$, as given in Chow (1973). Newton's method can be applied for solving the equations as in the linear case.

Besides the method of maximum likelihood, Amemiya (1973) has recently proposed a generalization of the method of two-stage least squares for non-linear structural equations, and Hausman (1974) has suggested an application of the method of instrumental variables to certain type of nonlinear equations. Furthermore, attempts have been made to approximate each nonlinear function f_i in (8.1) by a linear function of β_i using a first-order Taylor expansion around a tentative value β_i^0 and to estimate the resulting linear system by the methods of 2SLS or 3SLS. One can iterate by revising the tentative value β_i^0 in each iteration. This approach has been programmed in the Troll System of the National Bureau of Economic Research (1974).

The methods just mentioned for nonlinear systems have not been tried extensively, and much more experimentation is required

to study their computational problems and their sampling properties.

An important special case of nonlinear systems is derived from a linear system with residuals ϵ_t which are serially correlated. For example, if ϵ_t in equation (1.1) follows a second-order autoregressive scheme

$$(8.3) \quad \epsilon_t = R_1 \epsilon_{t-1} + R_2 \epsilon_{t-2} + v_t ,$$

where v_t is serially uncorrelated, one can combine (8.3) and (1.1) to form a system

$$(8.4) \quad BY_t = \Gamma z_t + R_1 BY_{t-1} - R_1 \Gamma z_{t-1} + R_2 BY_{t-2} + R_2 \Gamma z_{t-2} + v_t$$

which is nonlinear in the parameters B , Γ , R_1 and R_2 . In addition to the methods mentioned above, Sargan (1961), Hendry (1971), Chow and Fair (1973), Fair (1972), Hannan and Terrell (1973) have studied the estimation of linear system with serially correlated residuals.

9. Concluding Remarks

Having reviewed some methods for identification and estimation of econometric systems and suggested some areas of further research, the author hopes that scholars in other disciplines especially the control scientists will find some of the econometric problems interesting and appealing. Economists are probably dealing with larger systems than the control engineers. The number of unknown parameters to be estimated is larger, and

the data on which statistical estimates can be derived are limited. Although the simultaneous-equations model is somewhat different from the models of the control engineer, the concept of identifiability of a set of parameters is applicable to both and problems of identification can provide an area of common interest. Furthermore, estimation methods and computational techniques in the two disciplines are based on similar basic ideas. Therefore exchanging experience in both topics will be fruitful. In this connection, the setting up of a control model and the proposal to apply certain econometric techniques for its estimation by Mehra (1974) should be mentioned as an attempt in this direction. Mention should also be made of the recent interest among econometricians in the estimation of time-varying coefficients in regression models, a topic sometimes treated by Kalman filtering. This topic has been surveyed in the October, 1973, issue of the Annals of Economic and Social Measurement.

Besides the uses of econometric models for testing economic hypothesis and for forecasting, an important application is in the setting of economic policies. The subject of estimation and control of econometric systems has received much interest among economists and control scientists, as evidenced by the papers from three joint conferences sponsored by the National Bureau of Economic Research in 1972, 1973, and 1974. Some of these papers have appeared in the October 1972 and January 1974 issues of the Annals of Economic and Social Measurement and some will appear in the April, 1975, issue of that journal. This certainly is a subject of mutual interest.

Hopefully, the present survey, by bringing out certain salient features of the identification and estimation of econometric systems, will help promote further cooperation between the two disciplines.

FOOTNOTES

* Research supported by NSF Grant GS 32003X and benefited by the work of the author as consultant to the NBER Computer Research Center, and by helpful comments from Ray C. Fair and Richard E. Quandt.

1. Text books in econometrics include Christ (1966), Dhrymes (1970), Goldberger (1964), Johnston (1972), Malinvaud (1966), Theil (1971), among others.
2. For surveys of recent works on errors in variables in econometrics, the reader may refer to Goldberger (1974) and Griliches (1973).

REFERENCES

- Amemiya, T. (1973), "The Non-linear Two Stage Least Squares Estimator," Institute for Mathematical Studies in the Social Sciences, Working Paper No. 25, Stanford University.
- Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, 20, 46-63.
- Anderson, T.W. and T. Sawa (1973), "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," Econometrica, 41, 683-714.
- Basman, R.L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, 25, 77-83.
- Brundy, J. and D. Jorgenson (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables," Review of Economics and Statistics, 53, 207-224.
- Chow, G.C. (1968), "Two Methods of Computing Full-Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations," International Economic Review, 9, 100-112.
- _____ (1973), "A Family of Estimators for Simultaneous Equation Systems," Econometric Research Program Research Memorandum No. 155, Princeton, Princeton University, to appear in International Economic Review, October, 1974.
- Chow, G.C. and R.C. Fair (1973), "Maximum Likelihood Estimation of Linear Equation Systems with Auto-Regressive Residuals," Annals of Economic and Social Measurement, 2, 17-28.
- Christ, C.F. (1966), Econometric Models and Methods, New York, John Wiley and Sons, Inc.
- Dhrymes, P. (1970), Econometrics: Statistical Foundations and Applications. New York, Harper and Row, Publishers.
- _____ (1971), "A Simplified Structural Estimator for Large Scale Econometric Models," Austrian Journal of Statistics, 13 168-75.
- Dreze, J.H. (1971), "Limited Information Estimation from a Bayesian Viewpoint," Center for Operations Research and Econometrics Discussion Paper No. 7111, Heverlee, Catholic University of Louvain.

- Dreze, J.H. and J.A. Morales (1970), "Bayesian Full Information Analysis of the Simultaneous Equations Model," Center for Operations Research and Econometrics, Discussion Paper No. 7031, Heverlee, Catholic University of Louvain.
- Fair, R.C. (1972), "Efficient Estimation of Simultaneous Equations with Auto-Regressive Errors by Instrumental Variables," The Review of Economics and Statistics, 54, 444-449.
- _____ (1974), "On the Robust Estimation of Econometric Models," to appear in Annals of Economic and Social Measurement, October, 1974.
- Fisher, F.M. (1966), The Identification Problem in Economics, New York, McGraw Hill Publishing Company.
- Geary, R. (1949), "Determination of Linear Relations Between Systematic Parts of Variables with Errors of Observation the Variances of Which are Uninown," Econometrica, 17, 30-58.
- Goldberger, A.S. (1964), Econometric Theory, New York, John Wiley and Sons, Inc.
- _____ (1974), "Unobservable Variables in Econometrics," in Paul Zarembka, ed., Frontiers in Econometrics, 193-213.
- Goldfeld, S.M. and R.E. Quandt (1973), Nonlinear Methods in Econometrics, Amsterdam, North-Holland Publishing Company.
- Griliches, Z. (1973), "Errors in Variables and Other Unobservables," Harvard Institute of Economic Research Discussion Paper No. 333, to appear in Econometrica.
- Hannan, E.J. and D.R. Terrell (1973), "Multiple Equation Systems with Stationary Errors," Econometrica, 41, 299-320.
- Hausman, J. (1974), "An Instrumental Variable Approach to Full-Information Estimators for Linear and Non-linear Econometric Models," Department of Economics Working Paper No. 127, Cambridge, Massachusetts Institute of Technology.
- Hendry, D.F. (1971), "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Auto-Regressive Process," International Economic Review 12, 257-272.
- Johnston, J. (1972), Econometric Methods, 2nd Edition, New York, McGraw-Hill Book Company, Inc.

- Kloek, T. and L.B.M. Mennes (1960), "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables," Econometrica, 28, 45-61.
- Koopmans, T.C. ed. (1950), Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph No. 10, New York, John Wiley and Sons, Inc.
- Lyttkens, E. (1970), "Symmetric and Asymmetric Estimation Methods," in E. Mosback and H. Wold, Interdependent Systems, Amsterdam, North-Holland Publishing Company, 434-459.
- Malinvaud, E. (1966), Statistical Methods of Econometrics, Chicago, Rand McNally and Company.
- Mehra, R.K. (1974), "Identification in Control and Econometrics: Similarities and Differences," Annals of Economic and Social Measurement, 3, 21-48.
- Mikhail, W.M. (1973), "A Comparative Monte Carlo Study of the Properties of Econometric Estimators," mimeo.
- National Bureau of Economic Research, Inc. (1974), Troll Experimental Program, Cambridge, Mass.
- Nerlove, M. (1966), "A Tabular Survey of Macro-econometric Models," International Economic Review, 7, 127-175.
- Reiersøl, O. (1945), "Confluence Analysis by Means of Instrumental Sets of Variables," Arkiv for Matematik, Astronomi och Fysik, 32.
- Rothenberg, T.J. (1971), "Identification in Parametric Models," Econometrica, 39, 577-591.
- Rothenberg, T.J. and C. Leenders (1964), "Efficient Estimation of Simultaneous Equation Systems," Econometrica, 32, 57-76.
- Sargan, J.D. (1961), "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals," Econometrica, 39, 414-426.
- Theil, H. (1953), Estimation and Simultaneous Correlation in Complete Equation Systems. The Hague, Centraal Planbureau.
- _____ (1958), Economic Forecasts and Policy, Amsterdam, North-Holland Publishing Company.
- _____ (1971), Principles of Econometrics, New York, John Wiley and Sons, Inc.

- Wegge, L. (1965), "Identifiability Criteria for a System of Equations as a Whole," Australian Journal of Statistics, 7, 67-77.
- Wold, H. (1954), "Causality and Econometrics," Econometrica, 22, 162-177.
- _____ (1964), ed., Econometric Model Building: Essays in the Causal Chain Approach, Amsterdam, North-Holland Publishing Company.
- Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 57, 348-368.
- _____, (1971), An Introduction to Bayesian Inference in Econometrics, New York, John Wiley and Sons, Inc.
- Zellner, A. and H. Theil (1962), "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," Econometrica, 30, 54-78.