DEPARTMENT OF ECONOMICS

THE CLASSICAL ECONOMETRIC MODEL

D.S.G. Pollock, University of Leicester, UK

Working Paper No. 08/33
September 2008

THE CLASSICAL ECONOMETRIC MODEL

by D.S.G. Pollock (University of Leicester)
Email: stephen_pollock@sigmapi.u-net.com

A compendium is presented of the various approaches that may be taken in deriving the estimators of the simultaneous-equations econometric model according to the principle of maximum likelihood. The structural equations of the model have the character both of a regression equation and of an errors-in-variables equation. This partly accounts for way in which the various approaches that have been followed appear to differ widely. In the process of achieving a synthesis of the methods of estimation, some elements that have been missing from the theory are supplied.

1. Introduction

Recently, interest in the classical simultaneous-equations model of econometrics has been revived through a paper by Anderson (2005) concerning the origins of the LIML and 2SLS estimators. The paper tells how Anderson and Rubin (1950) derived the asymptotic distribution of LIML estimator by finding the asymptotic distribution of what is essentially the 2SLS estimator.

For many years, the classical model served as a paradigm for econometric theory and it continues to exercise a strong influence. However, the contributions to the theory are disparate and widely scattered, and it seems that it would be helpful if an account were given that summarises this theory and provides some of the missing elements. This is the purpose of the present paper.

The classical simultaneous-equations econometric model has had a long and complicated history. The statistical problems to which it is addressed began to interest economists in the 1920's, who were considering how to untangle estimates of equations of supply and demand from market data.

The fundamental difficulty of determining the two functions from observations on prices and quantities sold was highlighted by the brothers Holbrook Working, 1895–1985 and Elmer Joseph Working, 1900–1968. Whereas the elder brother was the first to grapple with the perplexities of the problem (in H. Woking, 1925), it was the younger brother who discovered the true nature of the so-called identification problem in econometrics and who suggested the solution to it (in E. Woking, 1927).

The full solution of the identification problem, and the provision of the appropriate statistical estimators, had to wait until the late 1940's and the early 1950's. These results were primarily the work of members of the *Cowles Commission for Research in Economics,* located at the University of Chicago. The statistical methods that were employed in solving the estimation problem of the classical model owed much to the pioneering work of Harold Hotelling (1933), (1936), in connection with principal components and canonical correlations.

Economists had difficulty in assimilating the work of the Cowles Commission. Nor did the theory result in any significant body of empirical work, given the inadequacy of the contemporary computing facilities. However, by the end

of the 1950's, there had emerged, from a Dutch–American school of econometrics, an alternative framework for understanding the problems of estimating the simultaneous-equations model.

Within this framework, were developed the two-stage least-squares estimator of Basmann (1957) and Theil (1958) and the three-stage least-squares estimator of Zellner and Theil (1962). These were to become the mainstay of most of the empirical studies that invoked the classical model. The estimators were depicted as elaborations of the familiar methods of least-squares regression; and there was virtually no rapprochement with the kindred estimators of the Cowles Commission.

Nevertheless, in an influential book, Malinvaud (1966), who was an erstwhile member of the Cowles Commission, did assert the commonality of the rival estimators, and he placed them firmly within the context of the linear errors-in-variables model. Indeed, the affinity of an isolated structural equation of the classical model to the equation of an error-in-variables model had long been apparent from the nature of the limited-information maximum-likelihood (LIML) estimator. Theil (1961) had already achieved a partial synthesis by defining the k-class of estimators that comprised both the two-stage least-squares estimator and the limited-information maximum-likelihood estimator.

Latterly, methods that are closely related to those of the Cowles Commission have been used by econometricians in estimating systems of co-integrated dynamic equations—see Johansen (1995). The maximum likelihood estimator of such systems owes much to Hotelling's method of canonical correlations. Moreover, in the form developed by Anderson and Rubin (1949), which is that of a reduced-rank regression (Anderson 1951), the LIML estimator is essentially the same as the one that is employed in estimating a cointegrating vector.

It is clear that the methods that have been engendered by the classical model have a protean complexion; and it is a challenging exercise to show how their various aspects are related to those of estimators which serve other purposes, which seem, at first sight, to be quite different. However, this will not be the task of the present paper. Instead, the paper aims at achieving a retrospective synthesis of the estimators of the classical model that has not hitherto been available in any accessible form.

In the process, an element that has been missing from the Cowles Commission's exposition will be supplied. This is a formulation of the estimator of multi-equation subsystems. Such an estimator stands midway between the full-information maximum-likelihood estimator (FIML) of a complete system and the limited-information maximum-likelihood estimator (LIML) of a single structural equation. Since it can accommodate either of the special cases, this estimator has the greatest generality.

Before this conclusion can be reached, it is necessary to expose the statistical structure of the simultaneous-equations model and to reveal the fundamental basis of the estimators. In doing so, we shall be able to join the Cowles Commission estimators seamlessly with the two-stage and three-stage least-squares estimators and to show how some strikingly different formulations of the LIML estimator are the same thing in various disguises.

2. The Classical Model

The classical model describes a stochastic relationship between a set of K exogenous variables contained in the vector $x' = [x_1, \ldots, x_K]$ and a set of G endogenous variables contained in the vector $y' = [y_1, \ldots, y_G]$. A single realisation of the model can be written as

$$y'\Gamma + x'B = \varepsilon', \tag{1}$$

where $\varepsilon' = [\varepsilon_1, \ldots, \varepsilon_G]$ is a normally distributed disturbance vector which has an expected value of $E(\varepsilon) = 0$ and a dispersion matrix of $D(\varepsilon) = \Psi$. The vectors x and ε are assumed to be uncorrelated, so that their covariance matrix is $C(x, \varepsilon) = 0$.

The diagonal elements of the nonsingular parameter matrix Γ are assumed to be all equal to minus one, and this indicates that y_j is to be regarded as the dependent variable of the jth equation. If e_j is the vector with a unit in the jth position and zeros elsewhere, then these restrictions, which are called the normalisation rules, can be written as $e_j'\Gamma e_j = -1; j = 1, \ldots, G$.

In addition, we assume that certain of the elements of Γ and B are zeros. Thus, if $\gamma_j = \Gamma e_j$ and $\beta_j = Be_j$ are the parameter vectors of the jth equation, we can write the full set of restrictions affecting them as

$$\begin{bmatrix} R_{\gamma j}' & 0 \\ 0 & R_{\beta j}' \end{bmatrix} \begin{bmatrix} \gamma_j \\ \beta_j \end{bmatrix} = \begin{bmatrix} r_j \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} R_{\gamma j}' & 0 \\ 0 & R_{\beta j}' \end{bmatrix} \begin{bmatrix} \gamma_j + e_j \\ \beta_j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{2}$$

where R_{β_j} comprises a selection of columns from the identity matrix I_K of order K, $R_{\gamma j} = [e_j, H_{\gamma j}]$ comprises, likewise, a set of columns from the identity matrix I_G of order G, and r_j is a vector containing zeros and a single element of minus one corresponding to the normalisation rule.

We can represent the general solution to the combined restrictions by

$$\begin{bmatrix} \gamma_j \\ \beta_j \end{bmatrix} = \begin{bmatrix} P_{\gamma j} & 0 \\ 0 & P_{\beta j} \end{bmatrix} \begin{bmatrix} \gamma_{\diamond j} \\ \beta_{\diamond j} \end{bmatrix} - \begin{bmatrix} e_j \\ 0 \end{bmatrix}, \tag{3}$$

where $\gamma_{\diamond j}$ and $\beta_{\diamond j}$ are composed of the G_j and K_j unrestricted elements of γ_j and β_j respectively, and where $P_{\gamma j}$ and $P_{\beta j}$ are the complements of $R_{\gamma j}$ and $R_{\beta j}$ within I_G and I_K respectively.

On taking account of the so-called exclusion restrictions, which set some of the parameters to zeros, and by using the normalisation rule to isolate the dependent variable, we may write the jth structural equation of the system, which has hitherto been represented by $y'\gamma_j + x'\beta_j = \varepsilon_j$, as

$$y_j = y_{\diamond j}'\gamma_{\diamond j} + x_{\diamond j}'\beta_{\diamond j} - \varepsilon_j. \tag{4}$$

This is the regression form of the structural equation.

The essential departure of the structural equation from an ordinary regression equation lies in the fact that the so-called endogenous regressors within $y_{\diamond j}$ are statistically correlated with the structural disturbance term ε_j. This

relationship is manifest in equation (1), which depicts the system as a whole. The starting point of Basmann (1957) and Theil (1958) in their derivation of the two-stage least-squares estimator was to ask what needs to be done to purge the endogenous regressors of their dependence on the disturbance term.

The reduced form of equation (1) is written as

$$\begin{aligned} y' &= -x'B\Gamma^{-1} + \varepsilon'\Gamma^{-1} \\ &= x'\Pi + \eta', \end{aligned} \tag{5}$$

where $\Pi = -B\Gamma^{-1}$ and $\eta' = \varepsilon'\Gamma^{-1}$ with $D(\eta) = \Gamma'^{-1}\Psi\Gamma^{-1} = \Omega$. We shall denote the dispersion matrices of the data vectors by $D(x) = \Sigma_{xx}$, $D(y) = \Sigma_{yy}$ and their covariance matrix by $C(x, y) = \Sigma_{xy}$, and we shall assume that all of these attain the maximum rank.

Substituting the reduced form expression for y' into the structural equation $y'\gamma_j + x'\beta_j = \varepsilon_j$ gives $(x'\Pi + \eta')\gamma_j + x'\beta_j = \varepsilon_j$. But $\eta'\gamma_j = \varepsilon_j$. Therefore, on setting $x'\Pi = y' - \eta'$, the equation can be written as

$$(y - \eta)'\gamma_j + x'\beta_j = 0. \tag{6}$$

The equation may be reduced by elimination the parameters that are know to be zeros, but the normalisation rule may be held in abeyance.

It appears from (6) that the structural equation is a species of errors-in-variables equation in which the errors extend over the set of endogenous variables within y but fall short of the exogenous variables in x. The errors-in-variables formulation is the one that was adopted preponderantly by the Cowles Commission in their development of the limited-information maximum-likelihood estimator. This is apparent in the expositions of Anderson and Rubin (1949) and Koopmans and Hood (1953).

From the existing assumptions regarding the system as a whole, it follows that

$$\begin{bmatrix} \Gamma' & B' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} \Psi & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}, \tag{7}$$

and, from this, equivalent expressions may be obtained in the forms of

$$\begin{aligned} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma & 0 \\ B & I \end{bmatrix} &= \begin{bmatrix} \Gamma'^{-1} & \Pi' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Psi & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} \\ &= \begin{bmatrix} \Omega\Gamma & \Pi'\Sigma_{xx} \\ 0 & \Sigma_{xx} \end{bmatrix} \end{aligned} \tag{8}$$

and

$$\begin{aligned} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} &= \begin{bmatrix} \Gamma'^{-1} & \Pi' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Psi & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma^{-1} & 0 \\ \Pi & I \end{bmatrix} \\ &= \begin{bmatrix} \Pi'\Sigma_{xx}\Pi + \Omega & \Pi'\Sigma_{xx} \\ \Sigma_{xx}\Pi & \Sigma_{xx} \end{bmatrix}. \end{aligned} \tag{9}$$

The above identities provide us with the fundamental equations that relate the structural parameters γ_j, $\beta_j; j = 1, \ldots, G$ to the moment matrices of the data variables. We can write these equations in two alternative forms:

$$\begin{aligned}\begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \begin{bmatrix} \Sigma_{yy} - \Omega & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \gamma_j \\ \beta_j \end{bmatrix} \\ &= \begin{bmatrix} \Pi'\Sigma_{xy} & \Pi'\Sigma_{xx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \gamma_j \\ \beta_j \end{bmatrix}. \end{aligned} \tag{10}$$

The first of these follows directly from (8). The second is obtained via the identities $\Sigma_{yy} = \Pi'\Sigma_{xx}\Pi + \Omega$ and $\Sigma_{xy} = \Sigma_{xx}\Pi$ that are contained in (9).

It is clear, from the second expression under (10), that all the information relating to γ_j and β_j that is provided by that equation is contained in its constituent part $\Sigma_{xy}\gamma_j + \Sigma_{xx}\beta_j = 0$. On substituting the solutions $\gamma_j = P_{\gamma_j}\gamma_{\diamond j} - e_j$ and $\beta_j = P_{\beta_j}\beta_{\diamond j}$ from equation (3) into this expression, we get

$$\Sigma_{xy}P_{\gamma j}\gamma_{\diamond j} + \Sigma_{xx}P_{\beta j}\beta_{\diamond j} = \Sigma_{xy}e_j. \tag{11}$$

This is a set of K equations in $G_j + K_j$ unknowns; and, given that the matrix $[\Sigma_{xy}\ ,\ \Sigma_{xx}]$ is of full rank, it follows that the necessary and sufficient condition for the identifiability of the parameters of the jth equation is that $K \geq G_j + K_j$. If this condition is fulfilled, then any subset of $G_j + K_j$ of the equations of (11) will serve to determine $\gamma_{\diamond j}$ and $\beta_{\diamond j}$. However, we shall be particularly interested in a set of $G_j + K_j$ independent equations in the form of

$$\begin{bmatrix} P'_{\gamma j}\Pi'\Sigma_{xy}P_{\gamma j} & P'_{\gamma j}\Pi'\Sigma_{xx}P_{\beta j} \\ P'_{\beta j}\Sigma_{xy}P_{\gamma j} & P'_{\beta j}\Sigma_{xx}P_{\beta j} \end{bmatrix} \begin{bmatrix} \gamma_{\diamond j} \\ \beta_{\diamond j} \end{bmatrix} = \begin{bmatrix} P'_{\gamma j}\Pi'\Sigma_{xy}e_j \\ P'_{\beta j}\Sigma_{xy}e_j \end{bmatrix} \tag{12}$$

which are derived by premultiplying equation (11) by the matrix $[\Pi P_{\gamma_j}, P_{\beta_j}]'$. These equations, which we have derived solely by considering the relationship between the parameters of the model and the moments of the data vectors x and y, must be the basis of any reasonable estimator of the parameters of the individual structural equations, regardless of the principles from which it is derived.

3. The Maximum-Likelihood Estimator of Single Equations

In describing the procedures by which inferences can be made about the parameters of the model, we shall assume that there is a set $T \geq G + K$ observations on the data vectors which may be denoted by $x_t, y_t; t = 1, \ldots, T$. From these observations, a set of moment matrices can be constructed in the forms of $S_{xx} = \sum_t x_t x_t'/T, S_{xy} = \sum_t x_t y_t'/T$ and $S_{yy} = \sum_t y_t y_t'/T$ which are assumed to have full rank. These are the empirical counterparts of the moment matrices Σ_{xx}, Σ_{xy} and Σ_{yy} which characterise the distribution of the population.

The principle of estimation known as the method of moments suggests that we should obtain the estimating equations by replacing the population moments within the equations (9) and (11) by the empirical moments. According to this principle, equation (9) yields an estimate $S_{xx}^{-1}S_{xy}$ of the reduced-form

parameter matrix Π which is simply the unrestricted least-squares estimator. Likewise, in the case where $G_j + K_j = K$, equation (11) yields the so-called indirect least-squares estimators of the structural parameters $\gamma_{\diamond j}$ and $\beta_{\diamond j}$.

However, the simple method of moments is not adequate in the case of the overidentified model where $G_j + K_j < K$. For then the restrictions on the vectors γ_j and β_j imply restrictions on Σ_{xy} and Σ_{yy}; and the effect is that, in the context of the set of all values that might be assumed by S_{xy} and S_{yy}, the subset consisting of the values that will render the equations in (11) algebraically consistent has a zero measure.

There are two alternative approaches that may be followed in the attempt to overcome the problem of overidentification. The first approach is to resolve the algebraic inconsistency of the equations

$$S_{xy}P_{\gamma j}\gamma_{\diamond j} + S_{xx}P_{\beta j}\beta_{\diamond j} = S_{xy}e_j. \tag{13}$$

via a method of least-squares regression. Thus, by using a regression metric defined in terms of the matrix S_{xx}^{-1}, we may derive the two-stage least-squares estimator of Theil (1958) and Basmann (1957). What is rather remarkable is the fact that this is precisely the estimator which is also derived from the equation (12) when Σ_{xy} and Σ_{yy} are replaced by S_{xy} and S_{yy} respectively and Π is represented by its unrestricted ordinary least-squares estimator.

Another way of overcoming the problem of overidentification is to find a set of admissible and mutually conformable estimates of Σ_{xy}, Σ_{yy}, γ_j and β_j which satisfy equation (11) exactly. This is what the method of maximum likelihood achieves.

3.1 The Π *Method*

There are a number of different yet equivalent criteria from which the maximum-likelihood estimators of the structural parameters can be derived. We shall begin, as the author has done in Pollock (1984), (1985), by using, as the criterion function, an expression of the likelihood function which is in terms of the reduced-form parameters:

$$\begin{aligned} L(\Pi, \Omega) =& (2\pi)^{-GT/2}|\Omega|^{-T/2}\times \\ & \exp\left\{-\frac{T}{2}\mathrm{Trace}\left(S_{yy} - S_{yx}\Pi - \Pi' S_{xy} + \Pi' S_{xx}\Pi\right)\Omega^{-1}\right\}. \end{aligned} \tag{14}$$

This is also the starting point for the classical derivation of Anderson and Rubin (1949).

In order to convey to the reduced-form parameters the effects of the structural-form restrictions under (2), we must use the condition $\Pi\gamma_j + \beta_j = 0$ which comes directly from the identity $\Pi = -B\Gamma^{-1}$ by which the reduced-form parameters have been defined. The appropriate criterion function is given by the expression

$$C(\Pi, \Omega, \gamma_j, \beta_j) = \log \mathrm{L}(\Pi, \Omega) - \kappa'(\Pi\gamma_j + \beta_j) - \mu'\big(Q'_{\gamma j}\gamma_j + Q'_{\beta j}\beta_j\big), \tag{15}$$

where $Q_{\gamma j} = [R_{\gamma j}, 0]$, $Q_{\beta j} = [0, R_{\beta j}]$, and where κ and λ are vectors of Lagrangean multipliers.

By differentiating the criterion function with respect to Π, Ω and κ and setting the results to zero for a maximum, the conditions are obtained from which we may derive the estimating equations for Π and Ω. The latter are respectively

$$\Pi = S_{xx}^{-1}S_{xy} - S_{xx}^{-1}(S_{xy}\gamma_j + S_{xx}\beta_j)\big(\gamma_j'W\gamma_j\big)^{-1}\gamma'_jW \tag{16}$$

and

$$\begin{aligned}\Omega = W + W\gamma_j\left(\gamma_j'W\gamma_j\right)^{-1}\left(\gamma_j'S_{yx} + \beta_j'S_{xx}\right)S_{xx}^{-1} \\ \times\left(S_{xy}\gamma_j + S_{xx}\beta_j\right)\left(\gamma_j'\,W\gamma_j\right)^{-1}\gamma_j'W,\end{aligned} \tag{17}$$

where $W = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$ is the ordinary unrestricted estimate of $\Omega = \Sigma_{yy} - \Pi'\Sigma_{xx}\Pi$. It is straightforward to confirm that the restricted estimator of Π does satisfy the condition $\Pi\gamma_j + \beta_j = 0$.

Given that Σ_{xx} is always estimated by S_{xx}, regardless of the restrictions on γ_j and β_j, it follows that the restricted estimator of $\Sigma_{xy} = \Sigma_{xx}\Pi$ is given by $S_{xx}\Pi$, where Π stands for the expression in (16). When these estimates are used to represent Σ_{xy} and Σ_{xx} within the expression under (11), we get

$$S_{xx}\Pi P_{\gamma j}\gamma_{\diamond j} + S_{xx}P_{\beta j}\beta_{\diamond j} = S_{xx}\Pi e_j. \tag{18}$$

This expression represents nothing more than an algebraic identity. Therefore, unlike the equation (13) upon which the two-stage least-squares estimator is based, there is no question of any algebraic inconsistency when $K > K_j + G_j$.

To obtain the maximum-likelihood estimating equations for the structural parameters, we differentiate the criterion function in respect of γ_j and β_j and set the results to zero. Then, by substituting for $\kappa = (S_{xy}\gamma_j + S_{xx}\beta_j)(\gamma_j'W\gamma_j)^{-1}$, we obtain the equation

$$\begin{bmatrix}\Pi'S_{xy} & \Pi'S_{xx}\\ S_{xy} & S_{xx}\end{bmatrix}\begin{bmatrix}\gamma_j\\ \beta_j\end{bmatrix} + \begin{bmatrix}R_{\gamma j} & 0\\ 0 & R_{\beta j}\end{bmatrix}\begin{bmatrix}\phi_\gamma\\ \phi_\beta\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}, \tag{19}$$

where $[\phi_\gamma', \phi_\beta']' = \phi = \mu(\gamma_j'W\gamma_j)$ is a vector related to the lagrangean multipliers. On substituting $\gamma_j = P_{\gamma j}\gamma_{\diamond j} - e_j$ and $\beta_j = P_{\beta_j}\beta_{\diamond j}$ from equation (3) into this expression, and by premultiplying it by the transpose of the matrix in (3), we get

$$\begin{bmatrix}P'_{\gamma j}\Pi'S_{xy}P_{\gamma j} & P'_{\gamma j}\Pi'S_{xx}P_{\beta j}\\ P'_{\beta j}S_{xy}P_{\gamma\, j} & P'_{\beta j}S_{xx}P_{\beta j}\end{bmatrix}\begin{bmatrix}\gamma_{\diamond j}\\ \beta_{\diamond j}\end{bmatrix} = \begin{bmatrix}P'_{\gamma j}\Pi'S_{xy}e_j\\ P'_{\beta j}S_{xy}e_j\end{bmatrix}. \tag{20}$$

This is precisely what one derives from the equation (12) when Σ_{xx} and Σ_{xy} are replaced by S_{xx} and S_{xy} respectively and Π is represented by its restricted estimator from (16).

If we take the above equation together with equation (16) for the restricted estimator of Π, then we have a complete estimating system for $\gamma_j = P_{\gamma_j}\gamma_{\diamond j} -$

e_j and $\beta_j = P_{\beta_j}\beta_{\diamond j}$ which can be solved iteratively. That is to say, given an initial value Π_0, we can obtain first-round estimates of $\gamma_{\diamond j}$ and $\beta_{\diamond j}$ from equation (20). On taking these values to equation (16), we can obtain a revised estimate for Π for use in the second round; and we can extend this procedure through an indefinite number of iterations in the expectation, that, eventually, the estimates from successive rounds will become virtually identical. We shall describe this procedure as the Π method of estimation.

3.2 The Ω Method

To obtain the second version of the estimating equations, we use the identity

$$(S_{yy} - \Omega)\gamma_j + S_{yx}\beta_j = \Pi' S_{xy}\gamma_j + \Pi' S_{xx}\beta_j \tag{21}$$

which holds when Ω and Π are represented by the expressions in (16) and (17). This identity enables us to rewrite the estimating equations for $\gamma_{\diamond j}$ and $\beta_{\diamond j}$ as

$$\begin{bmatrix} P'_{\gamma j}(S_{yy} - \Omega)P_{\gamma j} & P'_{\gamma j}S_{yx}P_{\beta j} \\ P'_{\beta j}S_{xy}P_{\gamma j} & P'_{\beta j}S_{xx}P_{\beta j} \end{bmatrix} \begin{bmatrix} \gamma_{\diamond j} \\ \beta_{\diamond j} \end{bmatrix} = \begin{bmatrix} P'_{\gamma j}(S_{yy} - \Omega)e_j \\ P'_{\beta j}S_{xy}e_j \end{bmatrix}. \tag{22}$$

This form corresponds to the first of the two equations under (10) which describe the relationship between the moments of the data variables and the parameters of the jth equation. If we take these equations (22) together with equation (17) for the restricted estimate of Ω, then we have an alternative estimating system that can be solved iteratively for Ω, γ_j and β_j. Thus we have what we shall call the Ω method of estimation.

3.3 The λ Method

To obtain the third estimating system, we use an identity which can be written as

$$\Omega\gamma_j = \lambda W\gamma_j, \tag{23}$$

where W is the unrestricted estimator of Ω and where

$$\lambda = \big(\gamma'_j S_{yy}\gamma_j + \gamma'_j S_{yx}\beta_j + \beta'_j S_{xy}\gamma_j + \beta'_j S_{xx}\beta_j\big)\big(\gamma'_j W\gamma_j\big)^{-1}. \tag{24}$$

On substituting the identity into (22), the estimating equations for $\gamma_{\diamond j}$ and $\beta_{\diamond j}$ become

$$\begin{bmatrix} P'_{\gamma j}(S_{yy} - \lambda W)P_{\gamma j} & P'_{\gamma j}S_{yx}P_{\beta j} \\ P'_{\beta j}S_{xy}P_{\gamma j} & P'_{\beta j}S_{xx}P_{\beta j} \end{bmatrix} \begin{bmatrix} \gamma_{\diamond j} \\ \beta_{\diamond j} \end{bmatrix} = \begin{bmatrix} P'_{\gamma j}(S_{yy} - \lambda W)e_j \\ P'_{\beta j}S_{xy}e_j \end{bmatrix}. \tag{25}$$

The sequence of estimates of γ_j and β_j is now obtained by solving the equations (24) and (25) in successive rounds. This constitutes the λ method of estimation.

The equations of the λ method come closest to the limited-information maximum-likelihood estimating equations that are common to the expositions of Anderson and Rubin (1949) and Koopmans and Hood (1953). The outstanding difference is that the latter equations are written in a homogeneous

form which results from the suppression of the normalisation rule $\gamma_j e_j = -1$. In consequence of this difference, the factor λ assumes the role of a latent root which can be extracted by the power method.

When λ assumes the role of a latent root, an alternative interpretation of the structural equation is engendered. For then the estimating equations come to resemble those of an errors-in-variables model. In this context, it is appropriate to seek to estimate γ_j and β_j by minimising the function λ subject to the restrictions in (3). This approach to the derivation of the estimating equations, which may, of course, be subsumed under the maximum-likelihood principle, is what Koopmans and Hood have described as the minimum variance-ratio method.

3.4 Computations

In order to specify completely any of the foregoing methods of estimation, we need to choose the initial conditions with which to begin the various iterative procedures. The natural choices for the initial values of Π and Ω are their unrestricted estimates which are given by $S_{xx}^{-1}S_{xy}$ and $S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$ respectively. The natural choice for a starting value for λ is unity, since this is the value to which it must tend asymptotically as the sample moments tend to the values of the population moments. It is remarkable that, with these choices, the first rounds of the Π, Ω and λ methods of estimation give rise to identical estimates of γ_j and β_j which are simply the two-stage least-squares estimates of Theil (1958) and Basmann (1957).

4. The Maximum-Likelihood Estimation of Subsystems

One of the more surprising aspects of the development of the theory of the classical model has been the lack of a complete treatment of the problem of estimating subsystems of the model which comprise more than one equation. This state of affairs prevailed in spite of the fact that, in their seminal account of the single-equation estimator, Koopmans and Hood (1953) had provided a criterion function from which the estimating equations of a multi-equation subsystem may be derived directly.

The advantage of the subsystem estimator is that it can be specialised easily in either direction to provide, on the one hand, the estimator of a single equation and, on the other hand, the estimator of the system as a whole. Thus, a full theory of estimation can be articulated by considering only the case of subsystems.

4.1 The Π Method for Subsystems

It transpires that the procedure which we have used in deriving the estimating equations for the Π method can be generalised very easily to cope with subsystems. The only complication is that, in order to derive the estimating equations of multi-equation subsystems, we have to deal in terms of the vectorisation operator and the algebra of Kronecker products. An account of this algebra was given in Pollock (1979).

Let us assume that the subsystem of interest has a total of $G_1 = G - G_2$ equations whose parameters are contained in the matrices $\Gamma_1 = \Gamma J_1$ and $B_1 = BJ_1$, where J_1 is a matrix consisting of G_1 columns selected from the identity matrix of order G. Then the restrictions on these parameters, including the normalisation rules, may be written as

$$R'\begin{bmatrix}\Gamma_1 + J_1\\ B_1\end{bmatrix}^c = 0 \qquad \text{and} \qquad (\Gamma_1 \otimes I)\Sigma_{xy}^c + (B_1' \otimes I)S_{xx}^c, \tag{26}$$

where the superscript c denotes the vectorisation operator. We can proceed to form a Lagrangean expression which is a straightforward generalisation of the expression under (15). Then, from the first-order conditions optimising the Lagrangean, we can obtain the following estimators of Π and Ω:

$$\Pi = S_{xx}^{-1}S_{xy} - S_{xx}^{-1}(S_{xy}\Gamma_1 + S_{xx}B_1)(\Gamma' W\Gamma_1)^{-1}\Gamma_1' W, \tag{27}$$

$$\begin{aligned}\Omega = {} & W + W\Gamma_1(\Gamma_1' W\Gamma_1)^{-1}\times \\ & (\Gamma_1' S_{yx} + B_1' S_{xx})S_{xx}^{-1}(S_{xy}\Gamma_1 + S_{xx}B_1)(\Gamma'_1 W\Gamma_1)^{-1}\Gamma_1' W.\end{aligned} \tag{28}$$

These are the natural generalisations of the expressions under (16) and (17). The estimating equations for Γ_1 and B_1, which are obtained in the same manner as in the single-equation case, are given by

$$\left\{\begin{bmatrix}\Pi' S_{xy} & \Pi' S_{xx}\\ S_{xy} & S_{xx}\end{bmatrix}\begin{bmatrix}\Gamma_1\\ B_1\end{bmatrix}(\Gamma'_1\Omega\Gamma_1)^{-1}\right\}^c + R\mu = 0, \tag{29}$$

where μ is a vector of Lagrangean multipliers associated with the restrictions on Γ_1 and B_1. This is just a generalisation of equation (19). Notice, however, that, whereas, in the single-equation case, we were able to absorb the scalar factor $\gamma_j'\Omega\gamma_j$ in the multiplier ϕ, in the multi-equation case, we have to represent the corresponding matrix factor $\Gamma_1'\Omega\Gamma_1$ explicitly. To find a compact form of the estimating equations to compare with those under (19), we solve the first set of restrictions under (26) to obtain

$$\begin{bmatrix}\Gamma_1\\ B_1\end{bmatrix}^c = PP'\begin{bmatrix}\Gamma_1\\ B_1\end{bmatrix}^c - \begin{bmatrix}J_1\\ 0\end{bmatrix}^c, \tag{30}$$

where P is the complement of R within the identity matrix of order $G_1(G+K)$ and $[\Gamma_1', B_1']'^c P$ is a vector of the unrestricted elements of Γ_1 and B_1. On Substituting this solution into (29) and on premultiplying that equation by P', we get

$$\begin{aligned}&\left(P'\left\{(\Gamma_1'\Omega\Gamma_1)^{-1} \otimes \begin{bmatrix}\Pi' S_{xy} & \Pi' S_{xx}\\ S_{xy} & S_{xx}\end{bmatrix}\right\}P\right)P'\begin{bmatrix}\Gamma_1\\ B_1\end{bmatrix}^c \\ &\qquad = \left(P'\left\{(\Gamma_1'\Omega\Gamma_1)^{-1} \otimes \begin{bmatrix}\Pi' S_{xy} & \Pi' S_{xx}\\ S_{xy} & S_{xx}\end{bmatrix}\right\}P\right)P'\begin{bmatrix}J_1\\ 0\end{bmatrix}^c.\end{aligned} \tag{31}$$

4.2 The Estimating Equations of the Ω Method

To derive the estimating equations of the Ω method directly, we may begin by reconsidering the likelihood function that is given under (14). The exponent in this expression can be written as

$$\begin{aligned}&(S_{yy} - \Pi' S_{xy} - S_{yx}\Pi - \Pi' S_{xx}\Pi)\,\Omega^{-1}\\ &\quad = \big(S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}\big)\,\Omega^{-1} + \big(S_{yx}S_{xx}^{-1}S_{xy} - \Pi' S_{xy} - S_{yx}\Pi + \Pi' S_{xx}\Pi\big)\Omega^{-1}.\end{aligned} \tag{32}$$

Using $\Sigma_{xy} = S_{xx}\Pi$ to denote the restricted estimator Σ_{xy}, we can write the second term on the RHS as

$$\begin{aligned}&\big(S_{yx}S_{xx}^{-1}S_{xy} - \Pi' S_{xy} - S_{yx}\Pi + \Pi' S_{xx}\Pi\big)\Omega^{-1}\\ &\quad = (S_{yx} - \Sigma_{yx})S_{xx}^{-1}(S_{xy} - \Sigma_{xy})\Omega^{-1}.\end{aligned} \tag{33}$$

If our object is to find an estimator of Σ_{xy} that is admissible with respect to the various restrictions on Γ_1 and B_1, then it is appropriate to find the value that minimises the following Lagrangean criterion function:

$$\begin{aligned}C(\Gamma_1, B_1, \Omega) &= \frac{1}{2}(S_{xy} - \Sigma_{xy})^{c\prime}(\Omega \otimes S_{xx})(S_{xy} - \Sigma_{xy})^{c}\\ &\quad + \Phi^{c\prime}\big\{\big(\Gamma_1{}' \otimes I\big)\Sigma_{xy}^{c} + (B_1' \otimes I)S_{xx}^{c}\big\}.\end{aligned} \tag{34}$$

By differentiating the function with respect to Σ_{xy}^{c} and setting the result to zero, we obtain the condition

$$(S_{xy} - \Sigma_{xy})^{c\prime}(\Omega \otimes S_{xx})^{-1} - \Phi^{c\prime}(\Gamma_1' \otimes \mathrm{I}) = 0. \tag{35}$$

This gives us

$$(S_{xy} - \Sigma_{xy})^{c\prime}(\Omega \otimes S_{xx})^{-1}(S_{xy} - \Sigma_{xy})^{c} = \Phi^{c\prime}(\Gamma_1'\Omega\Gamma_1 \otimes S_{xx})\Phi^{c}, \tag{36}$$

as well as

$$\Phi^{c\prime} = (S_{xy} - \Sigma_{xy})^{c\prime}(\Gamma_1 \otimes I)(\Gamma_1'\Omega\Gamma_1 \otimes S_{xx})^{-1}. \tag{37}$$

However, the condition, $\Sigma_{xy}\Gamma_1 + S_{xx}B_1 = 0$, can be used to rewrite the latter as

$$\Phi^{c\prime} = (S_{xy}\Gamma_1 + S_{xx}B_1)^{c\prime}\,(\Gamma_1'\Omega\Gamma_1 \otimes S_{xx})^{-1} \tag{38}$$

On substituting this into (35) and rearranging the result, we get,

$$\Sigma_{xy} = S_{xy} - (S_{xy}\Gamma_1 + S_{xx}B_1)\,(\Gamma_1'\Omega\Gamma_1)^{-1}\,\Gamma_1'\Omega. \tag{39}$$

This is the restricted estimator of Σ_{xy} and, of course, in the form of $\Pi = S_{xx}^{-1}\Sigma_{xy}$, it provides us again with the restricted estimator of Π.

On substituting the expression from (38) into (36), we get

$$\begin{aligned}&(S_{xy} - \Sigma_{xy})^{c\prime}(\Omega \otimes S_{xx})(S_{xy} - \Sigma_{xy})^{c}\\ &\quad = (S_{xy}\Gamma_1 + S_{xx}B_1)^{c\prime}(\Gamma_1'\Omega\Gamma_1 \otimes S_{xx})^{-1}(S_{xy}\Gamma_1 + S_{xx}B_1)^{c}.\end{aligned} \tag{40}$$

The function on the RHS is essentially a generalisation of the criterion function used by Keller (1975) in his derivation of the single-equation estimator. To derive the criterion for estimating Γ_1 and B_1, we subject the function to the restrictions in (26). By differentiating the resulting Lagrangean expression with respect to Γ_1 and B_1 and setting the results to zero, we obtain the condition

$$\left\{ \begin{bmatrix} S_{yx}S_{xx}^{-1}S_{xy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ B_1 \end{bmatrix} (\Gamma'_1 \Omega \Gamma_1)^{-1} + \begin{bmatrix} Q \\ 0 \end{bmatrix} \right\}^c + R\mu = 0, \tag{41}$$

wherein

$$\begin{aligned} Q &= \Omega\Gamma_1(\Gamma'\Omega\Gamma_1^{-1}(\Gamma'_1 S_{yx} + B'_1 S_{xx})S_{xx}^{-1}(S_{xy}\Gamma_1 + S_{xx}B_1)(\Gamma'\Omega\Gamma_1)^{-1} \\ &= (W\Gamma_1 - \Omega\Gamma_1)(\Gamma'_1\Omega\Gamma_1)^{-1}. \end{aligned} \tag{42}$$

Here the second equality follows from equation (28) via the identity

$$W\Gamma_1(\Gamma'_1 W\Gamma_1)^{-1} = \Omega\Gamma_1(\Gamma'_1\Omega\Gamma_1)^{-1}.$$

On substituting the expression for Q back into (41) and using $W = S_{yy} - S_{yx}S_{xx}^{-1}Sxy$, we get the estimating equation of the Ω method:

$$\left\{ \begin{bmatrix} S_{yy} - \Omega & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ B_1 \end{bmatrix} (\Gamma'_1 \Omega \Gamma_1)^{-1} \right\}^c + R\mu = 0. \tag{43}$$

4.3 The λ Method for Subsystems

As in the cases of the Π method and the Ω method, the estimating equations of the λ method can be derived directly from their own version of the maximum-likelihood criterion function. This is the concentrated likelihood function of Koopmans and Hood (1953). The process by which these authors derived their concentrated function is lengthy and difficult, and we shall give only a brief summary of it.

We may begin by considering the structural form of the likelihood function:

$$\begin{aligned} L(\Gamma, B, \Psi) = {} & (2\pi)^{-GT/2}|\Psi|^{-T/2}|\Gamma|^T \times \\ & \exp\left\{ -\frac{T}{2}\text{Trace}(\Gamma' S_{yy}\Gamma + \Gamma' S_{yx}B + B' S_{xy}\Gamma + B' S_{xx}B)\Psi^{-1} \right\}. \end{aligned} \tag{44}$$

This is obtained from the reduced form of the likelihood function under (14) via the identities $\Pi = -B\Gamma^{-1}$ and $\Omega^{-1} = \Gamma'\Psi^{-1}\Gamma$. The matrices of the structural parameters are partitioned as

$$\begin{bmatrix} \Gamma \\ B \end{bmatrix} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ B_1 & B_2 \end{bmatrix} \qquad \text{and} \qquad \Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}. \tag{45}$$

In addition to the parameters Γ_1, B_1 and Ψ_{11} of the subsystem in which we are interested, we are now encumbered with the parameters of a complementary

subsystem in which we have no direct interest. These additional parameters must either have acceptable arbitrary values assigned to them, or else they must be eliminated somehow from the likelihood function.

We begin by attributing the following form to the structural dispersion matrix:

$$\Psi = \begin{bmatrix} \Psi_{11} & 0 \\ 0 & I \end{bmatrix}; \tag{46}$$

for, as Koopmans and Hood show, this choice will not affect the estimates of the sought-after parameters provided that the complementary subsystem is not rendered overidentified. With this specification, the likelihood function becomes

$$\begin{aligned} L(\Gamma, B, \Psi) &= (2\pi)^{-GT/2}|\Psi_{11}|^{-T/2}|\Gamma|^{T}\times \\ &\exp\Big\{-\frac{T}{2}\mathrm{Trace}\big(\Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 + B_1{}'S_{xy}\Gamma_1 + B_1'S_{xx}B_1\big)\Psi_{11}^{-1} \\ &\quad -\frac{T}{2}\mathrm{Trace}\big(\Gamma_2{}'S_{yy}\Gamma_2 + \Gamma_2{}'S_{yx}B_2 + B_2{}'S_{xy}\Gamma_2 + B_2'S_{xx}B_2\big)\Big\}. \end{aligned} \tag{47}$$

This device does not wholly succeed in separating the parameters of the two subsystems as we might desire, since we cannot resolve the determinant $|\Gamma| = |\Gamma_1, \Gamma_2|$ into the product of separate functions of Γ_1 and Γ_2 . Therefore we have to eliminate the remaining parameters of the second subsystem by partially maximising L, or its log, with respect to Γ_2 and B_2. It can be shown that, at the maximum, there are

$$\begin{gathered} \Gamma_2'S_{yy}\Gamma_2 + \Gamma_2'S_{yx}B_2 + B_2'S_{xy}\Gamma_2 + B_2'S_{xx}B_2 = I \qquad \text{and} \\ |\Gamma_1,\ \Gamma_2|^{T} = |\Gamma_1'W\Gamma_1|^{T/2}\,|W|^{-T/2}. \end{gathered} \tag{48}$$

On substituting these values into (47), we obtain a concentrated likelihood function in the form of

$$\begin{aligned} L(\Gamma_1, B_1, \Sigma_1) &= (2\pi)^{-GT/2}|\Psi_{11}|^{-T/2}|\Gamma_1'W\Gamma_1|^{T/2}|W|^{-T/2} \\ &\quad\times \exp\Big\{-\frac{\mathrm{T}}{2}\mathrm{trace}\big(\Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 \\ &\quad + B_1'S_{xy}\Gamma_1 + \mathrm{B}_1{}'S_{xx}B_1\big)\Psi_{11}^{-1} - \frac{G_2T}{2}\Big\}. \end{aligned} \tag{49}$$

On differentiating the latter function with respect to Ψ_{11} and setting the result to zero, a condition is obtained that provides the estimating equation

$$\Psi_{11} = \Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 + B_1{}'S_{xy}\Gamma_1 + B_1'S_{xx}B_1. \tag{50}$$

A further concentration of the likelihood function is available when (50) is

substituted back into (49) to give

$$L(\Gamma_1, B_1) = (2\pi)^{-GT/2}\left\{\frac{\left|\Gamma_1{}'W\Gamma_1\right|}{\left|\Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 + B_1{}'S_{xy}\Gamma_1 + B_1'S_{xx}B_1\right|}\right\}^{T/2} \times |W|^{-T/2}\exp\left(-\frac{GT}{2}\right). \tag{51}$$

Embedded in this concentrated criterion function is a ratio of determinants which is a generalisation of the ratio λ of (24) whose minimisation leads to a single-equation version of the estimating equations.

To derive the estimating equations of a multi-equation subsystem, we may subject the log of the concentrated likelihood function under (49) to the restrictions on Γ_1 and B_1. Then, by evaluating the first-order conditions for the optimisation of the resulting Lagrangean function, we obtain the following estimating equations:

$$\left\{\begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx}\end{bmatrix}\begin{bmatrix}\Gamma_1 \\ B_1\end{bmatrix}\Psi_{11}^{-1} - \begin{bmatrix} W & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}\Gamma_1 \\ B_1\end{bmatrix}\left(\Gamma_1{}'W\Gamma_1\right)^{-1}\right\}^c + R\mu = 0. \tag{52}$$

At first sight, these do not appear to resemble very closely the equations of the Ω method given under (43). However, the equivalence of the two sets of equations is readily established via the identities $W\Gamma_1(\Gamma_1'W\Gamma_1)^{-1} = \Omega\Gamma_1(\Gamma_1'\Omega\Gamma_1)^{-1}$ and $\Psi_{11} = \Gamma_1'\Omega\Gamma_1$.

The derivation of Koopmans and Hood owes its difficulty largely to their use of the structural form of the likelihood function in place of the reduced form. Let us take the latter function instead, with a given value of Ω, and let us also use the identity $\Psi = \Gamma'\Omega\Gamma$ to write the structural dispersion matrix in the form of

$$\Gamma'\Omega\Gamma = \begin{bmatrix}\Gamma_1'\Omega\Gamma_1 & 0 \\ 0 & \mathrm{I}\end{bmatrix}. \tag{53}$$

Then the likelihood function, which is conditional on Ω, can be written as

$$L(\Gamma, B|\Omega) = (2\pi)^{-GT/2}|\Omega|^{-T/2}\times \exp\left\{-\frac{T}{2}\mathrm{Trace}\left(\Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 + B_1{}'S_{xy}\Gamma_1 + B_1'S_{xx}B_1\right)\left(\Gamma_1'\Omega\Gamma_1\right)^{-1} - \frac{T}{2}\mathrm{Trace}\left(\Gamma_2{}'S_{yy}\Gamma_2 + \Gamma_2{}'S_{yx}B_2 + B_2{}'S_{xy}\Gamma_2 + B_2'S_{xx}B_2\right)\right\}. \tag{54}$$

It is clear that the estimates of Γ_1 and B_1 are to be found by minimising the the first of the trace expressions without reference to other parameters. In this way, we derive a set of estimating equations in the form of

$$\left\{\begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx}\end{bmatrix}\begin{bmatrix}\Gamma_1 \\ B_1\end{bmatrix}\left(\Gamma_1'\Omega\Gamma_1\right)^{-1} - \begin{bmatrix} Q \\ 0\end{bmatrix}\right\}^c + R\mu = 0, \tag{55}$$

where

$$Q = \Omega\Gamma_1(\Gamma_1'\Omega\Gamma_1)^{-1}\big(\Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 + B_1{}'S_{xy}\Gamma_1 + B_1'S_{xx}B_1\big)(\Gamma_1'\Omega\Gamma_1)^{-1} \tag{56}$$

If Ω is given by the expression in (28), then the identity

$$\Gamma_1'\Omega\Gamma_1 = \Gamma_1{}'S_{yy}\Gamma_1 + \Gamma_1{}'S_{yx}B_1 + B_1{}'S_{xy}\Gamma_1 + B_1'S_{xx}B_1 \tag{57}$$

prevails, and we obtain exactly the equations of the Ω method given under (43).

5. The Estimation of the System as a Whole

The virtue of the three derivations that we have provided in the previous section is that the resulting estimating procedures are applicable to subsystems of all sizes ranging from a single structural equation to the system as a whole. Therefore, our account of the useful methods could well be concluded at this point. Nevertheless, it may be interesting to consider the forms of some other criterion functions that are applicable only to the case of the system as a whole.

The estimating approach of Koopmans, Rubin and Leipnik (1950), which, for a considerable time, was regarded as the definitive approach to system-wide estimation, is based directly on the structural form of the likelihood function given under (44). The derivation of the estimating equations can be simplified by concentrating this function by substituting for Ψ its maximum-likelihood estimating equation which is

$$\Psi = \Gamma' S_{yy}\Gamma + \Gamma' S_{yx}B + B' S_{xy}\Gamma + B' S_{xx}B. \tag{58}$$

Then the concentrated likelihood function takes the form of

$$L(\Gamma, B) = (2\pi)^{-GT/2}\frac{|\Gamma|^{\mathrm{T}}}{\big|\Gamma' S_{yy}\Gamma + \Gamma' S_{yx}B + B' S_{xy}\Gamma + B' S_{xx}B\big|^{\mathrm{T}/2}}\exp\left(-\frac{GT}{2}\right). \tag{59}$$

The basic criterion is therefore to minimise the function

$$\frac{\big|\Gamma' S_{yy}\Gamma + \Gamma' S_{yx}B + B' S_{xy}\Gamma + B' S_{xx}B\big|}{|\Gamma'\Gamma|} \tag{60}$$

subject to the restrictions $R'[\Gamma' - I, B']'^c = 0$ affecting Γ and B. This gives rise to a set of estimating equations in the form of

$$\left\{\begin{bmatrix}\Gamma^{-1\prime}\\ 0\end{bmatrix} - \begin{bmatrix}S_{yy} & S_{yx}\\ S_{xy} & S_{xx}\end{bmatrix}\begin{bmatrix}\Gamma_1\\ B_1\end{bmatrix}\Psi^{-1}\right\}^c - R\mu = 0. \tag{61}$$

The explicit use of the inverse of Γ in these equations means that they cannot be used as a model for the estimating equations of a subsystem.

A variant of the criterion in (60) is to minimise the function

$$\frac{\left|\Gamma' S_{yy}\Gamma + \Gamma' S_{yx}B + B' S_{xy}\Gamma + B' S_{xx}B\right|}{|\Gamma' S_{yy}\Gamma|}. \tag{62}$$

The inclusion of the factor S_{yy} in the denominator has no effect upon the minimising values of Γ and B. Nevertheless, the resulting estimating equations take the altered form of

$$\left\{\begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}\begin{bmatrix} \Gamma \\ B \end{bmatrix}(\Gamma'\Omega\Gamma)^{-1} - \begin{bmatrix} S_{yy} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \Gamma \\ B \end{bmatrix}(\Gamma' S_{yy}\Gamma)^{-1}\right\}^{c} + R\mu = 0. \tag{63}$$

These equations, which were originally derived by Chow (1964), are the basis of a common textbook presentation (See Dhrymes 1970 and Klein 1974, *inter alia*). It is easy to see that they are equivalent to the previous equations under (61).

The final variant of the criterion function that should be considered is one that has been adopted by Scharf (1976) in deriving what he has described as the full-information K-matrix class estimator. This function is given by

$$\frac{\left|\Gamma' S_{yy}\Gamma + \Gamma' S_{yx}B + B' S_{xy}\Gamma + B' S_{xx}B\right|}{|\Gamma' W\Gamma|}. \tag{64}$$

wherein $W = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$ is the unrestricted estimate of Ω. The estimating equations that are derivable from this criterion function are nothing but the full-information version of the estimating equations of the λ method. It helps in understanding this result to recognise that the function in (64) is generalisation of the function in (24) from which, we have asserted, one can derive the single-equation version of the estimating equations.

References

Anderson, T.W., (2005), Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators, *Journal of Econometrics,* 127, 1–16.

Anderson, T.W., (1951), Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions, *The Annals of Mathematical Statistics,* 22, 327–351.

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.

Basmann, R.L., (1957), A Generalised Classical Method of Linear Estimation of Coefficients in a Structural Equation, *Econometrica,* 24, 77–83.

Chow, G.C., (1964), A Comparison of Alternative Estimators for Simultaneous Equations, *Econometrica,* 32, 532–553.

Chow, G.C., (1968), Two Methods of Computing Full-Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations, *International Economic Review,* 9, 100–112.

Dhrymes, P.J., (1970), *Econometrics: Statistical Foundations and Applications,* New York: Harper and Row.

Hotelling, H., (1933), Analysis of a Complex of Statistical Variables into Principal Components, *Journal of Educational Psychology,* 24, 417–441 and 498–520.

Hotelling, H., (1936), Relations Between Two Sets of Variables, *Biometrika,* 28, 321—377.

Johansen, S., (1995),*Likelihood-based Inference in Cointegrated Vector Autoregressive Models,* Oxford University Press, Oxford.

Keller, W.J., (1975), A New Class of Limited-Information Estimators for Simultaneous Equation Systems, *Journal of Econometrics,* 3, 71–92.

Klein, L.J., (1974), *A Textbook of Econometrics (Second Edition),* Englewood Cliffs: Prentice-Hall.

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950), Measuring the Equation Systems of Dynamic Economics', Chapter 2 in *Statistical Inference in Dynamic Economic Models,* T.C. Koopmans (editor), Cowles Foundation for Research in Economics, Monograph no. 10, New York: John Wiley and Sons.

Koopmans, T.C., and W.M. Hood (1953), The Estimation of Simultaneous Linear Economic Relationships, Chapter 6 in *Studies in Econometric Method,* W.C. Hood and T.C. Koopmans (editors), Cowles Foundation for Research in Economics, Monograph no. 14, New York: John Wiley and Sons.

Malinvaud, E., (1966), *Statistical Methods of Econometrics,* North-Holland Publishing Co. Amsterdam, originally published by Dunod, Paris under the title *Méthodes Statistiques de l' Économetrie.*

Neudecker, H., (1969), Some Theorems on Matrix Differentiation with Special Reference to Kronecker Products, *Journal of the American Statistical Association,* 64, 953–963.

Pollock, D.S.G., (1979), *The Algebra of Econometrics,* Chichester: John Wiley and Sons.

Pollock, D.S.G., (1983), Varieties of the LIML Estimator, *Australian Economic Papers,* 499–506.

Pollock, D.S.G., (1984), Two Reduced-Form Approaches to the Derivation of the Maximum-Likelihood Estimators for Simultaneous-Equation Systems, *Journal of Econometrics,* 24, 331–347.

Pollock, D.S.G., (1985), Tensor Products and Matrix Differential Calculus, *Linear Algebra and its Applications,* 67, 169–193.

Scharf, W., (1976), K-Matrix-Class Estimators and the Full-Information Maximum Likelihood Estimator as a Special Case, *Journal of Econometrics,* 4, 41–45.

Theil, H., (1958, 1961), *Economic Forecasts and Economic Policy,* (second edition 1961), Amsterdam: North-Holland Publishing Co.

Working, H., (1925), The Statistical Determination of Demand Curves *The Quarterly Journal of Economics,* 39, 503–543.

Working, E.J., (1927), What Do Statistical "Demand Curves" Show? *The Quarterly Journal of Economics,* 41, 212–235.

Zellner, A., and H. Theil (1962), Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations, *Econometrica,* 30, 54–78.