CHAPTER 8

Estimation of Structural Econometric Equations

The classical linear simultaneous-equation econometric model is a system which relates M output or endogenous variables to K input or exogenous variables. The special feature of the model is that each of the output variables in the vector $y_{t.} = [y_{t1}, y_{t2}, \ldots, y_{tM}]$ is a function not only of some of the K input variables of the vector $x_{t.} = [x_{t1}, x_{t2}, \ldots, x_{tK}]$ but also of some of the other variables in $y_{t.}$.

One way of describing this feature is to say that there is instantaneous feedback from the output side of the system to the input side. Thus, the jth structural equation, which describes the output y_{tj} in terms of the elements of $x_{t.}$ and the remaining elements of $y_{t.}$, can be written as

$$y_{tj} = y_{t.}c_{.j} + x_{t.}\beta_{.j} + \varepsilon_{tj}, \tag{1}$$

where $c_{.j}$ and $\beta_{.j}$ are the parameter vectors of this equation and where it is understood that $c_{jj} = 0$ to prevent y_{tj} from appearing on both the LHS and the RHS.

An alternative way of writing the structural equation, which places y_{tj} in the company of the other endogenous variables of the wider system, is to express it as

$$y_{t.}\gamma_{.j} + x_{t.}\beta_{.j} + \varepsilon_{tj} = 0. \tag{2}$$

Then $\gamma_{.j}$ and $c_{.j}$ differ only in respect of their jth elements which are $\gamma_{jj} = -1$ and $c_{jj} = 0$ respectively. The condition $\gamma_{jj} = -1$, which serves to identify the dependent variable of the structural equation, is called the normalisation rule.

The M structural equations may be compiled to give the following system:

$$\begin{aligned}[y_{t1}, y_{t2}, \ldots, y_{tM}] = y_{t.}[c_{.1}, c_{.2}, \ldots, c_{.M}] &+ x_{t.}[\beta_{.1}, \beta_{.2}, \ldots, \beta_{.M}] \\ &+ [\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tM}],\end{aligned} \tag{3}$$

which becomes

$$y_{t.} = y_{t.}C + x_{t.}B + \varepsilon_{t.} \tag{4}$$

in a summary notation. This can be written alternatively as

$$y_{t.}\Gamma + x_{t.}B + \varepsilon_{t.} = 0, \tag{5}$$

where $\Gamma = [\gamma_{.1}, \gamma_{.2}, \ldots, \gamma_{.j}]$.

The Reduced-Form Transformation

If we are prepared to overlook the structural details of the econometric model, then we might accept a description of each of the output variables in $y_{t.} = [y_{t1}, y_{t2}, \ldots, y_{tM}]$ which is in terms of the input variables alone. Such a description is called the reduced form of the model. The reduced form is obtained from equation (5) by postmultiplying it by the inverse of the matrix Γ which describes the instantaneous feedback connections. This gives

$$y_{t.} = x_{t.}\Pi + \eta_{t.} \quad \text{with} \quad \Pi = -B\Gamma^{-1} \quad \text{and} \quad \eta_{t.} = -\varepsilon_{t.}\Gamma^{-1}. \tag{6}$$

Assumptions must now be made regarding the stochastic elements of the model. We shall assume that the elements of the vector $\varepsilon_{t.} = [\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tM}]$, which are the M structural disturbances, are distributed independently of time such that, for every t, there are

$$E(\varepsilon_{t.}) = 0 \quad \text{and} \quad D(\varepsilon_{t.}) = E(\varepsilon'_{t.}\varepsilon_{t.}) = \Sigma_{\varepsilon\varepsilon}. \tag{7}$$

It is also assumed that the structural disturbances are distributed independently of the exogenous variables so that $C(\varepsilon_{t.}, x_{s.}) = 0$ for all t and s.

It follows that the vector $\eta_{t.} = -\varepsilon_{t.}\Gamma^{-1}$ of reduced-form disturbances has

$$E(\eta_{t.}) = 0 \quad \text{and} \quad D(\eta_{t.}) = \Gamma'^{-1}D(\varepsilon_{t.})\Gamma^{-1} = \Gamma'^{-1}\Sigma_{\varepsilon\varepsilon}\Gamma^{-1} = \Omega. \tag{8}$$

The transformed disturbances retain their independence of $x_{t.}$, which gives the condition that $C(\eta_{t.}, x_{s.}) = 0$ for all t and s.

The Identification Problem and the Structural Model

The structural simultaneous-equation model is affected by the so-called identification problem which limits the possibilities of estimating the structural parameters. Given a sufficient set of observations, we shall always be able to estimate the parameters of the statistical relationship between the endogenous variables in $y_{t.}$ and the exogenous variables in $x_{t.}$, which is the reduced-form relationship. However, if we are to succeed in uncovering the parameters of the structural relationships, then we must have some prior information regarding the structure.

Let us assume that the statistical properties of the data can be described completely in terms of its first and second moments. We can denote the dispersion matrices of $x_{t.}$ and $y_{t.}$ by $D(x_{t.}) = \Sigma_{xx}$ and $D(y_{t.}) = \Sigma_{yy}$ and their covariance matrix by $C(x_{t.}, y_{t.}) = \Sigma_{xy}$. By combining the reduced-form regression relationship of (6) with a trivial identity in $x_{t.}$, we get the following system:

$$(9) \qquad [\, y_{t.} \quad x_{t.} \,] \begin{bmatrix} I & 0 \\ -\Pi & I \end{bmatrix} = [\, \eta_{t.} \quad x_{t.} \,].$$

Given the assumptions that $D(\eta_{t.}) = \Omega$ and that $C(\eta_{t.}, x_{t.}) = 0$, it follows that

$$(10) \qquad \begin{bmatrix} I & -\Pi' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Pi & I \end{bmatrix} = \begin{bmatrix} \Omega & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}.$$

Premultiplying this system by the inverse of the leading matrix gives an equivalent equation in the form of

$$(11) \qquad \begin{aligned} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Pi & I \end{bmatrix} &= \begin{bmatrix} I & \Pi' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Omega & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} \\ &= \begin{bmatrix} \Omega & \Pi'\Sigma_{xx} \\ 0 & \Sigma_{xx} \end{bmatrix}. \end{aligned}$$

From this system, the equations $\Sigma_{yy} - \Sigma_{yx}\Pi = \Omega$ and $\Sigma_{xy} - \Sigma_{xx}\Pi = 0$ may be extracted, from which are obtained the parameters that characterise the reduced-from relationship:

$$(12) \qquad \Pi = \Sigma_{xx}^{-1}\Sigma_{xy} \qquad \text{and} \qquad \Omega = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

These parameters can be estimated provided that the empirical counterparts of the moment matrices Σ_{xx}, Σ_{yy} and Σ_{xy} are available in the form of $M_{xx} = T^{-1}\sum_t x'_{t.}x_{t.}$, $M_{yy} = T^{-1}\sum_t y'_{t.}y_{t.}$ and $M_{xy} = T^{-1}\sum_t x'_{t.}y_{t.}$.

Now consider combining the structural equation of (5) with a trivial identity to form the counterpart of equation (9). This is the equation

$$(13) \qquad [\, y_{t.} \quad x_{t.} \,] \begin{bmatrix} \Gamma & 0 \\ B & I \end{bmatrix} = [\, \varepsilon_{t.} \quad x_{t.} \,].$$

Given that $D(\varepsilon) = \Sigma_{\varepsilon\varepsilon}$ and that $C(\varepsilon, x) = 0$, it follow that

$$(14) \qquad \begin{bmatrix} \Gamma' & B' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix},$$

and, from this, an equivalent expression can be obtained the form of

$$
\begin{aligned}
\begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma & 0 \\ B & I \end{bmatrix} &= \begin{bmatrix} \Gamma'^{-1} & \Pi' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} \\
&= \begin{bmatrix} \Omega\Gamma & \Pi'\Sigma_{xx} \\ 0 & \Sigma_{xx} \end{bmatrix}.
\end{aligned} \tag{15}
$$

This identity provides the fundamental equations that relate the structural parameters Γ, B to the moment matrices of the data variables. These equations can be written in two alternative forms:

$$
\begin{aligned}
\begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \begin{bmatrix} \Sigma_{yy} - \Omega & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma \\ B \end{bmatrix} \\
&= \begin{bmatrix} \Pi'\Sigma_{xy} & \Pi'\Sigma_{xx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} \Gamma \\ B \end{bmatrix}.
\end{aligned} \tag{16}
$$

The first equation follows directly from (15). The second is obtained via the identities $\Sigma_{yy} = \Pi'\Sigma_{xx}\Pi + \Omega$ and $\Sigma_{xy} = \Sigma_{xx}\Pi$, which are from (12). Indeed, by replacing Π' by $\Sigma_{yx}\Sigma_{xx}^{-1}$, we can express the matrix of the second equation in terms of the data moments alone.

Equation (16) represents the basis from which we must infer the values the structural parameters Γ and B. It is clear that, as it stands, the system contains insufficient information for the purpose. In particular, the constituent equation $\Pi'\Sigma_{xy}\Gamma + \Pi'\Sigma_{xx}B = 0$ is a transformation of its companion $\Sigma_{xy}\Gamma + \Sigma_{xx}B = 0$; and, therefore, it contains no additional information. In fact, if the moment matrices are themselves unrestricted, apart from the inevitable conditions of symmetry and positive definiteness, then the number of unknown parameters that can be inferred from equation (16) cannot exceed MK, which is the number of parameters in the reduced form regression matrix Π.

In theory, the prior information affecting Γ and B can take many forms. In practice, we are liable to consider only linear parametric restrictions, which are usually the normalisation rules that set the diagonal elements of Γ to -1 and the exclusion restrictions that set certain of the elements of Γ and B to zeros. If none of restrictions affect more than one equation, then it is possible to treat each equation in isolation.

If the restrictions on the parameters of the jth equation are in the form of exclusion restrictions and a normalisation rule, then they can be represented by the equation

$$
\begin{bmatrix} R'_\diamond & 0 \\ 0 & R'_* \end{bmatrix} \begin{bmatrix} \gamma_{.j} \\ \beta_{.j} \end{bmatrix} = \begin{bmatrix} r_j \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} R'_\diamond & 0 \\ 0 & R'_* \end{bmatrix} \begin{bmatrix} \gamma_{.j} + e_j \\ \beta_{.j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{17}
$$

where R_* comprises a selection of columns from the identity matrix I_K of order K, $R_\diamond$ comprises, likewise, a set of columns from the identity matrix I_M

of order M, and r_j is a vector containing zeros and an element of minus one corresponding to the normalisation rule. The vector e_j is the jth column of I_M whose unit cancels with the normalised element of $\gamma_{.j}$.

We can represent the general solution to these restrictions by

$$\begin{bmatrix}\gamma_{.j}\\ \beta_{.j}\end{bmatrix} = \begin{bmatrix}S_\diamond & 0\\ 0 & S_*\end{bmatrix}\begin{bmatrix}\gamma_{\diamond j}\\ \beta_{*j}\end{bmatrix} - \begin{bmatrix}e_j\\ 0\end{bmatrix}, \tag{18}$$

where $\gamma_{\diamond j}$ and β_{*j} are composed of the M_j and K_j unrestricted elements of $\gamma_{.j}$ and $\beta_{.j}$ respectively, and where $S_\diamond$ and S_* are the complements of $R_\diamond$ and R_* within I_M and I_K respectively.

On substituting the solution of (18) into the equation $\Sigma_{xy}\gamma_{.j} + \Sigma_{xx}\beta_{.j} = 0$, which is from the jth equation of (16), we get

$$\Sigma_{xy}S_\diamond\gamma_{\diamond j} + \Sigma_{xx}S_*\beta_{*j} = \Sigma_{xy}e_j. \tag{19}$$

This is a set of K equations in $M_j + K_j$ 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 $K \geq M_j + K_j$.

If this condition is fulfilled, then any subset of $M_j + K_j$ of the equations of (19) will serve to determine $\gamma_{\diamond j}$ and β_{*j}. However, we shall be particularly interested in a set of $M_j + K_j$ independent equations in the form of

$$\begin{bmatrix}S'_\diamond\Pi'\Sigma_{xy}S_\diamond & S'_\diamond\Pi'\Sigma_{xx}S_*\\ S'_*\Sigma_{xy}S_\diamond & S'_*\Sigma_{xx}S_*\end{bmatrix}\begin{bmatrix}\gamma_{\diamond j}\\ \beta_{*j}\end{bmatrix} = \begin{bmatrix}S'_\diamond\Pi'\Sigma_{xy}e_j\\ S'_*\Sigma_{xy}e_j\end{bmatrix}, \tag{20}$$

which are derived by premultiplying equation (19) by the matrix $[\Pi S_\diamond, S_*]'$. These equations, which we have derived solely by considering the relationship between the parameters of our 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.

Least-Squares Estimation of a Single Structural Equation

Now consider the identity $\eta_{t.}\Gamma = -\varepsilon_{t.}$ which expresses the relationship between the structural-form and reduced-form disturbances. This contains the equation $\eta_{t.}\gamma_{.j} = -\varepsilon_{tj}$ which can be used to rewrite the jth structural equation as

$$(y_{t.} - \eta_{t.})\gamma_{.j} + x_{t.}\beta_{.j} = 0. \tag{21}$$

The latter is the equation of an errors-in-variables model in which the errors extend only over a subset of the variables.

Let us ignore the subscript which indicates the location of the jth structural equation within the system of M equations. If the dispersion matrix $D(\eta_{t.}) = \Omega$ were known, then we should seek to estimate the parameters γ and β by finding admissible values which minimise the function

$$\sum_{t=1}^{T} \eta_{t.}\Omega^{-1}\eta'_{t.} = \sum_{t=1}^{T}(y_{t.} - \mu_{t.})\Omega^{-1}(y_{t.} - \mu_{t.})', \tag{22}$$

$$\text{where} \qquad \mu_{t.} = y_{t.} - \eta_{t.} = x_{t.}\Pi,$$

subject to the condition that

$$\mu_{t.}\gamma + x_{t.}\beta = 0. \tag{23}$$

The latter condition is implied by the relationship $\Pi\Gamma + B = 0$ which connects the reduced-form parameters to the structural parameters of the system as a whole.

The minimisation of (22) is accomplished in two stages. First, it may be noted that, for given values of γ and β, equation (23) defines a hyperplane within the space of dimension $K + M$ which contains the vectors $[y_{t.}, x_{t.}]'$ comprising the observations on the system's variables. The point $[\mu_{t.}, x_{t.}]'$ is contained within this hyperplane; and its distance from the corresponding vector of observations is given—in terms of the metric defined by Ω^{-1}—by

$$\|y_{t.} - \mu_{t.}\|_{\Omega^{-1}} = \sqrt{(y_{t.} - \mu_{t.})\Omega^{-1}(y_{t.} - \mu_{t.})'}. \tag{24}$$

First, we require to minimise this distance for given values of γ and β. Thereafter, we may find the values γ and β which minimise the sum of squares of the T distances which is expressed under (22). Let us therefore consider the following Lagrangean criterion function:

$$L = (y_{t.} - \mu_{t.})\Omega^{-1}(y_{t.} - \mu_{t.})' + 2\lambda(\mu_{t.}\gamma + x_{t.}\beta). \tag{25}$$

Differentiating this function with respect to $\mu'_{t.}$ and setting the result to zero for a minimum gives the condition

$$-(y_{t.} - \mu_{t.})\Omega^{-1} + \lambda\gamma' = 0, \tag{26}$$

from which

$$(y_{t.} - \mu_{t.})\Omega^{-1} = \lambda\gamma', \tag{27}$$

$$(y_{t.} - \mu_{t.}) = \lambda\gamma'\Omega. \tag{28}$$

Putting these two equations together gives

$$(y_{t.} - \mu_{t.})\Omega^{-1}(y_{t.} - \mu_{t.})' = \lambda^2\gamma'\Omega\gamma. \tag{29}$$

But, on postmultiplying (28) by γ and using the condition $-\mu_{t.}\gamma = x_{t.}\beta$, it is found that

$$y_{t.}\gamma - \mu_{t.}\gamma = y_{t.}\gamma + x_{t.}\beta = \lambda\gamma'\Omega\gamma, \tag{30}$$

which implies that

$$\lambda = \frac{y_{t.}\gamma + x_{t.}\beta}{\gamma'\Omega\gamma}.$$

Thus, (29) can be written as

$$(y_{t.} - \mu_{t.})\Omega^{-1}(y_{t.} - \mu_{t.})' = \frac{(y_{t.}\gamma + x_{t.}\beta)^2}{\gamma'\Omega\gamma}. \tag{31}$$

Now define the matrices $Y' = [y'_{1.}, y'_{2.}, \ldots, y'_{T.}]$ and $X' = [x'_{1.}, x'_{2.}, \ldots, x'_{T.}]$ which together comprise the complete set of observations on the system taken over the T periods. Then the expression for the sum of squares of the deviations of the observations from the regression hyperplane becomes

$$\sum_{t=1}^{T}(y_{t.} - \mu_{t.})\Omega^{-1}(y_{t.} - \mu_{t.})' = \frac{(Y\gamma + X\beta)'(Y\gamma + X\beta)}{\gamma'\Omega\gamma}. \tag{32}$$

The criterion function of (32) must be minimised subject to whatever prior information is available regarding the elements of γ and β. When it is combined with the sample information of Y and X, this prior information must be sufficient to render these parameters identifiable. The prior information regarding γ and β usually takes the form of *exclusion restrictions* which indicate that certain variables which are present in the wider system are, in fact, absent from the structural equation in question. There is also the *normalisation rule* to be taken into account which indicates that one of the elements of $y_{t.}$ is the dependent variable of the equation in question.

A general way or expressing the prior information regarding the parameters is to write an equation in the form of

$$R'_1\gamma + R'_2\beta = r. \tag{33}$$

In the case where the only restrictions are exclusion rules, the matrices R'_1 and R'_2 will contain elements which are either zeros or units, whilst the elements of the vector r will be zeros apart from a single element which corresponds to the normalisation rule and which takes a value of -1. Reference to equation (17) shows that $[R_\diamond, 0] = R_1$ and $[0, R_*] = R_2$.

The criterion function for the restricted estimation is given by the following Lagrangean expression:

$$L = \frac{(Y\gamma + X\beta)'(Y\gamma + X\beta)}{\gamma'\Omega\gamma} + 2\kappa'(R_1'\gamma + R_2'\beta - r). \tag{34}$$

Differentiating this with respect to γ via a product rule and setting the result to zero gives

$$\frac{(Y\gamma + X\beta)'Y}{\gamma'\Omega\gamma} - \frac{(Y\gamma + X\beta)'(Y\gamma + X\beta)\gamma'\Omega}{(\gamma'\Omega\gamma)^2} + \kappa' R_1' = 0. \tag{35}$$

Multiplying throughout by $\gamma'\Omega\gamma$ and defining a new multiplier $\mu = \kappa\gamma'\Omega\gamma$ gives an equation of which the transpose is

$$Y'Y\gamma + Y'X\beta - \left\{\frac{(Y\gamma + X\beta)'(Y\gamma + X\beta)}{\gamma'\Omega\gamma}\right\}\Omega\gamma + R_1\mu = 0. \tag{36}$$

Next, differentiating the Lagrangean function of (34) with respect to β and setting the result to zero gives

$$\frac{(Y\gamma + X\beta)'X}{\gamma'\Omega\gamma} + \kappa' R_2' = 0 \tag{37}$$

which, on multiplying by $\gamma'\Omega\gamma$ and transposing gives us

$$X'Y\gamma + X'Y\beta + R_2\mu = 0. \tag{38}$$

On combining equations (36) and (38) together with the equation (33) of the restrictions, we get an equation in the form of

$$\begin{bmatrix} Y'Y - \lambda\Omega & Y'X & R_1 \\ X'Y & X'X & R_2 \\ R_1' & R_2' & 0 \end{bmatrix} \begin{bmatrix} \gamma \\ \beta \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ r \end{bmatrix}, \tag{39}$$

wherein

$$\lambda = \left\{\frac{(Y\gamma + X\beta)'(Y\gamma + X\beta)}{\gamma'\Omega\gamma}\right\}. \tag{40}$$

In order to use these equations in estimating the parameters of the structural equation, a value must be given to Ω. This is the dispersion matrix of the reduced-form disturbances:

$$\Omega = D(\eta_{t.}) = E\big\{(y_{t.} - x_{t.}\Pi)'(y_{t.} - x_{t.}\Pi)\big\}. \tag{41}$$

A direct application of the method of moments suggests that an appropriate estimate would be given by

$$\begin{aligned}\hat{\Omega} &= \frac{1}{T}(Y - X\hat{\Pi})'(Y - X\hat{\Pi}) \\ &= \frac{1}{T}Y'\{I - X(X'X)^{-1}X'\}Y,\end{aligned} \tag{42}$$

where $\hat{\Pi} = (X'X)^{-1}X'Y$ is the estimate of the reduced-from coefficients from equation (6).

Equations (39) and (40) together represent a nonlinear system which must be solved by an iterative method in order to generate the estimates of the structural parameters. A straightforward iterative procedure can be defined which begins by setting $\lambda = \lambda_{(0)} = T$ in equation (39). The equation is then solved for the first-round estimates $\gamma_{(1)}$ and $\beta_{(1)}$. These estimates can be drafted into equation (40) to provide a revised coefficient $\lambda_{(1)}$ to replace $\lambda_{(0)} = T$. When equation (39) is solved a second time with $\lambda = \lambda_{(1)}$, the second-round estimates $\gamma_{(2)}$ and $\beta_{(2)}$ are generated.

It should be easy to see how this procedure can be taken through any number of iterations. The procedure can be halted when the values of γ and β emerging from successive cycles are virtually identical. In fact, four or five cycles should be a sufficient number. The values of $\gamma_{(1)}$ and $\beta_{(1)}$ which emerge from the first stage of the procedure are, in fact, the so-called two-stage least-squares (2SLS) estimates. The values upon which the procedure converges are the so-called limited-information maximum-likelihood (LIML) estimates.

The Conventional Forms of the Estimators

It may be useful to extract the conventional forms of the estimating equations for 2SLS and LIML from the system which comprises equations (39) and (40). For a start, we invoke the usual assumption, which is that, apart from the normalisation rule, which we shall monetarily ignore, the *a priori* restrictions on γ and β take the form of exclusion restrictions specifying that some of the variables which are present in the wider system are not present in the structural equation in question.

Let us adopt a notation which sets $Y = [Y_*, Y_{**}]$ and $X = [X_*, X_{**}]$, where Y_{**} and X_{**} are the matrices of the excluded variables, and let the leading column of Y_* be the vector of observations on the dependent variable of the structural econometric equation. The estimating equations in this case become

$$\begin{bmatrix} Y_*'Y_* - \lambda\hat{\Omega}_* & Y_*'X_* \\ X_*'Y_* & X_*'X_* \end{bmatrix}\begin{bmatrix}\gamma_* \\ \beta_*\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix}, \tag{43}$$

$$\lambda = \left\{ \frac{(Y_*\gamma_* + X_*\beta_*)'(Y_*\gamma_* + X_*\beta_*)}{\gamma_*'\Omega_*\gamma_*} \right\}, \tag{44}$$

$$\hat{\Omega}_* = \frac{1}{T} Y_*' \{ I - X(X'X)^{-1}X' \} Y_*. \tag{45}$$

Solving the second line of equation (43), which is the equation $X_*'Y_*\gamma_* + X_*'X_*\beta_* = 0$, gives

$$\beta_* = -(X_*'X_*)^{-1}X_*'Y_*\gamma_*. \tag{46}$$

Substituting this back into the top line of (43) gives

$$(Y_*'Y_* - \lambda\hat{\Omega}_*)\gamma_* + Y_*'X_*(X_*'X_*)^{-1}X_*'Y_*\gamma_* = 0. \tag{47}$$

This can be written as

$$\{Y_*'(I - P_*)Y_* - \lambda\hat{\Omega}_*\}\gamma_* = 0 \qquad \text{with} \quad \hat{\Omega}_* = \frac{1}{T}Y_*'(I - P)Y_*, \tag{48}$$

where $P = X(X'X)^{-1}X'$ and $P_* = X_*(X_*'X_*)^{-1}X_*'$.

Notice that the first equation under (48) is in the form of the estimating equation of an errors-in-variables model. Once this is recognised, alternative ways of finding the estimates of γ_* and of β_* suggest themselves. First the equation of (48) is solved by finding the characteristic root λ and the corresponding characteristic vector γ_* using one of the standard techniques such as the power method. Such methods find the value of γ_* subject to some arbitrary normalisation; and the normalisation rule which is appropriate to the present application is the one which gives the leading element of the vector γ_* the value of -1. By inserting γ_*, suitably normalised, back into equation (46), the estimate of β_* can be found.

The numerical results of this procedure are precisely the same as those which would result from the previous iterative method when it is pursued to full convergence.

Now let us consider imposing the normalisation rule upon the estimating equations from the start. Let us define $[-1, \gamma_\diamond'] = \gamma_*'$ and $[y_0, Y_\diamond] = Y_*$ to conform in multiplication. Then the structural equation, which was formerly written as $Y_*\gamma_* + X_*\beta_* + \varepsilon = 0$, becomes $y_0 = Y_\diamond\gamma_\diamond + X_*\beta_* + \varepsilon$, and the corresponding system of estimating equations becomes

$$\begin{bmatrix} y_0'y_0 - \lambda\hat{\omega}_{00} & y_0'Y_\diamond - \lambda\hat{\omega}_{0\diamond} & y_0'X_* \\ Y_\diamond'y_0 - \lambda\hat{\omega}_{\diamond 0} & Y_\diamond'Y_\diamond - \lambda\hat{\Omega}_{\diamond\diamond} & Y_\diamond'X_* \\ X_*'y_0 & X_*'Y_\diamond & X_*'X_* \end{bmatrix} \begin{bmatrix} -1 \\ \gamma_\diamond \\ \beta_* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{49}$$

where

$$\hat{\Omega}_* = \begin{bmatrix} \hat{\omega}_{00} & \hat{\omega}_{0\diamond} \\ \hat{\omega}_{\diamond 0} & \hat{\Omega}_{\diamond\diamond} \end{bmatrix}. \tag{50}$$

Ignoring the first line of the system, which can serve to determine λ in terms of $\gamma_\diamond$ and β_*, and rearranging the remaining lines gives the system

$$\begin{bmatrix} Y_\diamond' Y_\diamond - \lambda\hat{\Omega}_{\diamond\diamond} & Y_\diamond' X_* \\ X_*' Y_\diamond & X_*' X_* \end{bmatrix} \begin{bmatrix} \gamma_\diamond \\ \beta_* \end{bmatrix} = \begin{bmatrix} Y_\diamond' y_0 - \lambda\hat{\omega}_{\diamond 0} \\ X_*' y_0 \end{bmatrix}. \tag{51}$$

Now let us set $\lambda = T$, which is the value which it assumes in the first round of the iterative procedure which has been described above. The solution of the resulting system was described as the 2SLS estimate. To derive a more common expression for the 2SLS estimating equations, we may use the definition under (42) to help in rewriting the equation. The definition indicates that

$$\begin{aligned} Y'Y - T\hat{\Omega} &= Y'Y - (Y - X\hat{\Pi})'(Y - X\hat{\Pi}) \\ &= Y'Y - Y'(I - P)Y = Y'PY \\ &= \hat{\Pi}'X'X\hat{\Pi}, \end{aligned} \tag{52}$$

where $P = X(X'X)^{-1}X$. Let us also note that $X'Y = X'\big\{X(X'X)^{-1}X\big\}Y = X'X\hat{\Pi}$. By using these results, we can rewrite the estimating equations of (51) as

$$\begin{bmatrix} \hat{\Pi}_{X\diamond}' X'X\hat{\Pi}_{X\diamond} & \hat{\Pi}_{X\diamond}' X'X_* \\ X_*'X\hat{\Pi}_{X\diamond} & X_*'X_* \end{bmatrix} \begin{bmatrix} \gamma_\diamond \\ \beta_* \end{bmatrix} = \begin{bmatrix} \Pi_{X\diamond}' X'X\hat{\Pi}_{X0} \\ X_*'X\hat{\Pi}_{X0} \end{bmatrix}, \tag{53}$$

where $X\hat{\Pi}_{X\diamond} = X(X'X)^{-1}X'Y_\diamond$ and $X\hat{\Pi}_{X0} = X(X'X)^{-1}X'y_0$. The latter indicate that the equation can also be written as

$$\begin{bmatrix} \hat{\Pi}_{X\diamond}' X'Y_\diamond & \hat{\Pi}_{X\diamond}' X'X_* \\ X_*'XY_\diamond & X_*'X_* \end{bmatrix} \begin{bmatrix} \gamma_\diamond \\ \beta_* \end{bmatrix} = \begin{bmatrix} \Pi_{X\diamond}' X'y_0 \\ X_*'Xy_0 \end{bmatrix}, \tag{54}$$

which corresponds directly to the equations of (20) which describe the relationship between the population moments of the data and the parameters of the structural equation.

Two-Stage Least Squares and Instrumental Variables Estimation

The 2SLS estimating equations were derived independently by Theil and by Basmann, who followed a different line of reasoning from the one which we have pursued above. Their approach was to highlight the reason for the failure of ordinary least-squares regression to deliver consistent estimates of the parameters of a structural equation.

The failure is due to the violation of an essential condition of regression analysis which is that the disturbances must be uncorrelated with the explanatory variables on the RHS of the equation. Within the equation $y_o = Y_\diamond\gamma_\diamond + X_*\beta_* + \varepsilon$, there is a direct dependence of $Y_\diamond$ on the structural disturbances of ε. However, the disturbances are independent of the exogenous variables in X_*.

The original derivations of the 2SLS estimator were inspired by the idea that, if it were possible to purge the variables of $Y_\diamond$ of their dependence on ε, then ordinary least-squares regression would become the appropriate method of estimation. Thus, if $X\Pi_{X\diamond}$ were available, then this could be put in place of $Y_\diamond$; and the problem of dependence would be overcome.

Although $X\Pi_{X\diamond}$ is an unknown quantity, a consistent estimate of it is available in the form of $\hat{Y}_\diamond = X\hat{\Pi}_{X\diamond}$. Finding the estimate $\hat{\Pi}_{X\diamond}$ represents the first stage of the 2SLS procedure. Applying ordinary least-squares regression to the equation $y_o = \hat{Y}_\diamond\gamma_\diamond + X_*\beta_* + e$ is the second stage.

An alternative approach which leads to the same 2SLS estimator is via the method of instrumental-variables estimation. The method depends upon finding a set of variables which are correlated with the regressors yet uncorrelated with the disturbances.

In the case of the structural equation, the appropriate instrumental variables are the exogenous variables of the system as a whole which are contained in the matrix X. Premultiplying the structural equation by X' gives

$$X'y_0 = X'Y_\diamond\gamma_\diamond + X'X_*\beta_* + X'\varepsilon. \tag{55}$$

Within this system, the cross products correspond to a set of moment matrices which have the following limiting values:

$$\begin{aligned} \text{plim}(T^{-1}X'y_0) &= \Sigma_{xy}e_0, \\ \text{plim}(T^{-1}X'Y_\diamond) &= \Sigma_{xy}S_\diamond, \\ \text{plim}(T^{-1}X'X_*) &= \Sigma_{xx}S_*, \\ \text{plim}(T^{-1}X'\varepsilon) &= 0. \end{aligned} \tag{56}$$

When the moment matrices are replaced by their limiting values, we obtain the equation

$$\Sigma_{xy}e_0 = \Sigma_{xy}S_\diamond\gamma_\diamond + \Sigma_{xx}S_*\beta_*, \tag{57}$$

which has been presented already as equation (19). In this system, there are K equations in $M_\diamond + K_*$ parameters. We may assume that $[\Sigma_{xy}, \Sigma_{xy}]$ is of full rank. In that case, the necessary condition for the indentifiablity of the parameters $\gamma_\diamond$ and β_* is that $K \geq M_\diamond + K_*$, which is to say that the number of exogenous variables in the system as a whole must be no less that the number of structural parameters that need to be estimated.

The empircal counterpart of (57) is the equation

$$X'y_0 = X'Y_\diamond\gamma_\diamond + X'X_*\beta_*. \tag{58}$$

If $K = M_\diamond + K_*$, then this equation can be solved directly to provide the estimates. However, if $K > M_\diamond + K_*$, then the equation is bound to be algebraically inconsistent and the parameters are said to be overidentified. To resolve the inconsistency, we may apply to (55) the method of generalised least-squares regression. The disturbance term in (55), which is $X'\varepsilon$, had a dispersion matrix $D(X'\varepsilon) = \sigma^2 X'X$. When this is used in the context of the generalised least-squares estimator, we obtain, once again, the 2SLS estimates.