CHAPTER 2
THE FRAMEWORK OF LINEAR ECONOMETRIC MODELS AND THEIR ESTIMATION

1. Introduction: Construction of an econometric model is a pre-requisite for any econometric research. A model abstractly represents reality by bringing out what is relevant to a particular question neglecting all other unnecessary details. These models may be of different sizes and complexities. The most simple of them may involve single relationship among economic variables and, therefore, can be summed up in a single equation such as:

\[ y = a_0 x_0 + a_1 x_1 + \ldots + a_k x_k + e \]

where \( x_0 \) is a unit (column) vector. In the matrix form this model is simply expressed as \[ y = Xa + e. \] Such a model is popularly known as the multi-variate regression model. However, this simple relationship is the most fundamental building block of more complex models known as the multi-equation models that may consist of an impressive array of relations.

Multi-equation models almost invariably exhibit simultaneity relations in different variables of the model. Simultaneity occurs in an econometric model within which more than one causal relationship among the variables is specified. When interrelations exist among the variables, the single equation specification (as illustrated above) with its one way implied causality from predetermined to one endogenous variable is neither an accurate nor sufficient representation. In such cases, the economic variables are determined by a complete system of equations. A complete system is the representation
(model) in which there are as many equations as endogenous variables (whose formation is to be 'explained' by the equations). The equations are usually of four types: equations of economic behaviour, institutional rules, technological laws of transformation and identities. The term 'structural equation' is used to comprise all four types of equations.

Systems of structural equation may be composed entirely on the basis of economic theory or on the dual basis of economic theory combined with systematically collected statistical data for the relevant variable for a given period. These models serve as an indispensable aid to forecasting and an invaluable guide to policy-making whether for a firm or a governmental agency.

Since a characteristic feature of most economic models is interdependence and joint determination of several variables, the use of simultaneous equation model is more meaningful and logical. The seminal papers from which the simultaneous equation model developed establish the importance of joint endogeneity for statistical analysis of economic relationships. Haavelmo (1943, 1944) realised that in the presence of jointly endogenous variables, a joint probability distribution is necessary to analyse the data.

II. The General Simultaneous Equation Model: Before proceeding further let us develop the general simultaneous equation model. For this let us take a linear simultaneous equation model containing m structural relations. The model may be best described in a matrix form as $YA + XB + E = 0$, where, Y represents m number of endogenous variables, the X represents k number of pre-determined variables, E symbolizes the stochastic disturbance terms m in number and A and B are the matrices of structural coefficients of compatible order. For the purpose of estimation, n is the sample size of data on Y and X. The underlying theory in general will satisfy that only a small
number of all the variables of the model will occur in any one equation, and the matrices
A and B will have a considerable number of zeros. Further in each structural equation,
one endogenous variable can by convention be regarded as the dependent variable, the
coefficient of which is equal to unity. This is known as the normalization rule. By a
suitable arrangement of the equations and the endogenous variables in the model, this
rule helps to make the main diagonal elements such that \( a_i = -1 \) for all \( i \) in the matrix A.
To take care of the constant terms in the equations, the first column of X is a unit vector
such that \( x_n = 1 \) for all \( i = 1, 2, \ldots, n \).

\[
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1m} \\
a_{21} & a_{22} & \cdots & a_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nm}
\end{bmatrix};
B = \begin{bmatrix}
b_{11} & b_{12} & \cdots & b_{1m} \\
b_{21} & b_{22} & \cdots & b_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
b_{k1} & b_{k2} & \cdots & b_{km}
\end{bmatrix};
0 = \begin{bmatrix}
0_{11} & 0_{12} & \cdots & 0_{1m} \\
0_{21} & 0_{22} & \cdots & 0_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
0_{n1} & 0_{n2} & \cdots & 0_{nm}
\end{bmatrix}
\]

\[
Y = \begin{bmatrix}
y_{11} & y_{12} & \cdots & y_{1m} \\
y_{21} & y_{22} & \cdots & y_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
y_{n1} & y_{n2} & \cdots & y_{nm}
\end{bmatrix};
X = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1k} \\
x_{21} & x_{22} & \cdots & x_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nk}
\end{bmatrix};
E = \begin{bmatrix}
e_{11} & e_{12} & \cdots & e_{1m} \\
e_{21} & e_{22} & \cdots & e_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
e_{n1} & e_{n2} & \cdots & e_{nm}
\end{bmatrix}
\]

In the model described above, A is a square matrix and is assumed to be non-
singular, since if it were not, one or more of the structural relations would merely be a
linear combination of other structural relations, thus being redundant, or if the rows of B
did not obey the same linear restrictions as the rows of A, the m structural equations
would be inconsistent. The B matrix is generally not a square matrix. The m equations in
the structural from jointly determine for each observation, the value of m endogenous
variables, given k pre-determined variables, m stochastic disturbance term and m(m+k-1)
coefficients of the system. The stochastic disturbance terms are assumed to be identically
and independently distributed over the samples with zero mean and constant covariance matrix.

III. Estimation of Single Equation Econometric Models: Single equation econometric models are often estimated by the method of Least Squares that minimizes the squared Euclidean norm of the error vector, \( S \). It is tantamount to minimization of the Euclidean norm of the error vector.

\[
S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_k x_{ik})\}^2.
\]

To minimize \( S \) we differentiate \( S \) partially with respect to \( a_i \) (\( i=0, 1, 2, \ldots, k \)) and set those partial derivatives equal to zero.

\[
\frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_k x_{ik})\}(1) = 0
\]

\[
\frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_k x_{ik})\}(x_{i1}) = 0
\]

\[
\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots
\]

\[
\frac{\partial S}{\partial a_k} = -2 \sum_{i=1}^{n} \{y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_k x_{ik})\}(x_{ik}) = 0
\]  \( \ldots \)  (2.1)

Simplifying the above equations we obtain the so-called normal equations as follows:

\[
\sum_{i=1}^{n} y_i = n a_0 + a_1 \sum_{i=1}^{n} x_{i1} + a_2 \sum_{i=1}^{n} x_{i2} + \ldots + a_k \sum_{i=1}^{n} x_{ik}
\]

\[
\sum_{i=1}^{n} y_i x_{i1} = a_0 \sum_{i=1}^{n} x_{i1} + a_1 \sum_{i=1}^{n} x_{i1}^2 + a_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \ldots + a_k \sum_{i=1}^{n} x_{i1} x_{ik}
\]

\[
\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots
\]

\[
\sum_{i=1}^{n} y_i x_{ik} = a_0 \sum_{i=1}^{n} x_{ik} + a_1 \sum_{i=1}^{n} x_{i1} x_{ik} + a_2 \sum_{i=1}^{n} x_{i1} x_{ik} + \ldots + a_k \sum_{i=1}^{n} x_{ik}^2
\]  \( \ldots \)  (2.2)
This system of normal equations has \( k+1 \) unknowns in \( k+1 \) equations. Assuming linear independence of these equations, the system may be solved for \( \mathbf{a} \). This method of obtaining 'best' \( \mathbf{a} \) \((= \hat{\mathbf{a}})\) is called the Principle of Least Squares.

A purely Algebraic approach to obtain \( \hat{\mathbf{a}} \) : The linear model described above may also be expressed in terms of (sample) observations. It is an inconsistent system of \( n \) linear equations in \( k+1 \) variables.

\[
\begin{align*}
y_1 &= a_0 x_{10} + a_1 x_{11} + a_2 x_{12} + \ldots + a_k x_{1k} \\
y_2 &= a_0 x_{20} + a_1 x_{21} + a_2 x_{22} + \ldots + a_k x_{2k} \\
y_3 &= a_0 x_{30} + a_1 x_{31} + a_2 x_{32} + \ldots + a_k x_{3k} \\
& \vdots \\
y_n &= a_0 x_{n0} + a_1 x_{n1} + a_2 x_{n2} + \ldots + a_k x_{nk}
\end{align*}
\]

\[\text{... (2.3)}\]

The new variable \( x_0 = [x_{10}, x_{20}, \ldots, x_{n0}]' \) appearing in the first term with the coefficient \( a_0 \) is a unitary (column) vector taking on the value of unity (that is, 1) in all equations. The system of linear equations in (2.3) above is inconsistent since no value of \( \mathbf{a} = [a_0, a_1, \ldots, a_k]' \) can satisfy the equality relations. Alternatively expressed in matrix notations, the system of equations \( \mathbf{y} = \mathbf{Xa} \) is an inconsistent system of (linear) equations. Pre-multiplying the equation \( \mathbf{y} = \mathbf{Xa} \) by \( G \), we have \( Gy = GXa \). Now suppose \( GX = I \), an identity matrix, and the solution vector \( \hat{\mathbf{a}} = Gy \). If \( G = (X'X)^{-1}X' \), then obviously \( GX = (X'X)^{-1}X'X = I \). Here \( G \) is the (generalized) inverse of \( X \) (Krishnamurthy & Sen, 1976, p. 155, pp. 196-198; Rao & Mitra, 1971). Therefore, \( \hat{\mathbf{a}} = Gy = (X'X)^{-1}X'y \). The property of \( \hat{\mathbf{a}} \) thus obtained is that it minimizes the squared Euclidean norm \( S = e'e \).
\((1 - X\hat{a})^T (1 - X\hat{a})\) of the disturbance vector \(e\). It also implies that the Euclidean norm of \(e\) is minimized by \(\hat{a}\). If we considered \(y^* = (y - e)\) instead of \(y\), the system \(y^* = Xa\) is over-determined, though consistent. The over-determined system of equations will admit a generalized solution

\[
\hat{a}^* = (X'X)^{-1} X'y = (X'X)^{-1} X'y - (X'X)^{-1} X'e.
\]

Now if \((X'X)^{-1} X'e = 0\), we have \(\hat{a} = \hat{a}^*\).

Strictly speaking, the method of Least Squares to obtain \(\hat{a}\) is a mathematical rather than statistical method because it does not incorporate randomness or probability considerations in obtaining \(\hat{a}\) nor does it give us any hint on the statistical properties of \(\hat{a}\), although \(\hat{a}\), the best estimator of \(a\), is obtained by minimizing the sample \(S = e'e\) which is stochastic in nature. Since \(\hat{a} = (X'X)^{-1} X'y\) where \(y\) is stochastic, \(\hat{a}\) assumes stochasticity because of the simple fact that the function of a random variable is in turn a random variable.

It may be shown (Johnston, pp. 171-174) that under certain assumptions \(\hat{a}\) \(= (X'X)^{-1} X'y\) is the best linear unbiased estimator (blue) of \(a\) in \(y = Xa + e\). These assumptions are known as the Gauss-Markov assumptions of (linear) regression model \(y = Xa + e\) (Intriligator, pp. 106-109). These assumptions are: (i) \(E(e_i) = 0 \Rightarrow E(y_i) = \sum_{j=0}^{n} a_j x_y; \forall i, i = 1, 2, \ldots, n\). In general, \(E(e) = 0 \Rightarrow E(y_i) = Xa\). Here \(E(.)\) means the expectation of the random variable (.). (ii) \(E(e'e) = \sigma^2 I\). This assumption asserts that each disturbance \(e_i\) is distributed with the same positive and finite variance \(\sigma^2\) and all
disturbances are pair-wise uncorrelated. This is also called the assumption of homoskedasticity and non-autocorrelation (or geometrically, the spherical disturbances).

(iii) The matrix of sample values of the explanatory variables, $X$, is a fixed non-stochastic matrix of full rank. Although the assumption of normality of $e$ is not required to prove blue properties of $\hat{a}$, such an assumption makes $\hat{a}$ the most efficient estimator of $a$ in $y = Xa + e$ since in that case $\hat{a}$ is also a maximum likelihood estimator of $a$, attaining the Cramer-Rao bound (Intriligator, p. 108).

However, reality often breaks away with the Gauss-Markov assumptions that generously bestow upon the Least Squares estimator, $\hat{a}$, the properties of an ideal estimator. Very often disturbances are non-spherical. It has been shown that if $e$ has a non-spherical distribution, the Aitken estimator (or Generalized Least Squares - GLS - estimator) $\tilde{a}$ of $a$ is blue. The GLS estimator $\tilde{a}$ is given by

$$\tilde{a} = (X^{\prime} \Omega^{-1} X)^{-1} X^{\prime} \Omega^{-1} y \quad \ldots \quad (2.4)$$

The estimator is 'general' in the sense that for spherical disturbances $\Omega = I$, $\tilde{a}$ reduces to the (ordinary) Least Squares estimator $\hat{a}$, that is, $\tilde{a} = \hat{a}$. In practice, however, $\Omega$ is not known and therefore it has to be estimated from the sample observations (Intriligator, pp. 168-173) and consequently the GLS estimator is only consistent (Theil, pp. 398-400).

Occasionally, in single equation models, the assumption regarding the non-stochasticity of the sample values of the explanatory variables may not be satisfied. That makes the (ordinary) Least Squares estimator, $\hat{a}$, an inconsistent estimator (Theil, p. 608). There are two suitable estimators that deserve a mention in this regard. First, the
Wald-Hooper-Theil estimator (Hooper, JW & H Theil (1958)) and the second, the Instrumental Variable method (Reiersol, O. (1950)). In multi-equation (or simultaneous equation) models stochasticity of explanatory variables is invariably met with. Therefore, (ordinary) Least Squares estimator is an inconsistent estimator of the structural parameters of multi-equation models. Other methods such as Two-Stage Least Squares, Limited Information Maximum Likelihood method, etc. are applied to estimation of parameters of these models.

The violation of the first Gauss-Markov assumption is no less serious. It introduces bias into the Least Squares estimator. If the estimator continues to be biased even if the sample size is increased indefinitely (although in practice it is often impossible to increase the sample size indefinitely), it is futile to investigate into other desirable properties of the estimator. The presence of outliers in the errors often leads to the violation of the first Gauss-Markov assumption and the method of Least Squares yields unreliable estimates of the parameters. This fact prompts us to investigate into the alternative methods of estimation such as the Least Absolute Deviation Estimation.

IV. Estimation of Multi-equation or Simultaneous Equation Econometric Models: It is well-known now that in a system of simultaneous equations the method of (ordinary) least squares yields biased and inconsistent estimators (Intriligator, p. 376) of the structural coefficients. This was first established by Haavelmo (1943, 1947). He showed that simultaneity induces correlation between the regressors and the residuals giving inconsistent and biased estimates. In his path-breaking monograph, "The probability Approach in Econometrics" (1944) he developed a method for solving the problem of single equation bias.
Marschak (1953) has explained why we must try to establish the structural coefficients. The basic argument is that by its very definition each structural equation describes a specific, distinct and thereby autonomous link in the economic process. This autonomous character of structural relations also means that the structural coefficients are more stable, more like physical constants, than are the reduced form composites. Structural coefficients unlike the reduced form coefficients are more easily judged by intuition and their changes are better capable of reasonable discussion and interpretation. Structural coefficients are more easily comparable with the accumulated empirical evidence, which refers to the economic structure and not to the reduced form.

In the way to establishing the structural coefficients of a simultaneous equation model one has to estimate the reduced form equations. Haavelmo proposed to first estimate the reduced form of a system of simultaneous equations by ordinary least squares and then to derive estimates for the parameters of the structural form by indirect least squares. In the case of exactly identified systems, Haavelmo (1943, 1944, 1947) showed that this method is equivalent to the method of maximum likelihood. Considerable theoretical investigations were also carried out by the researchers in Cowles Foundation, e.g., Koopmans (1950) and Hood & Koopmans (1953), etc. on methods of estimating structural parameters of a system of simultaneous linear stochastic equations. The estimates have been derived from maximum likelihood considerations and proved to be asymptotically unbiased and efficient.

Haavelmo’s investigations made the ‘identification problem’ central to the estimation of simultaneous equation models. The identification problem may be cogently described as follows.
From the structural equations $YA + XB + E = 0$, we may obtain the reduced form equations through pre-multiplying the system of structural equations by $A^{-1}$. That is explicitly, $YAA^{-1} + XBA^{-1} + EA^{-1} = 0$ or $Y = XP + U$, where $\Pi = -BA^{-1}$ and $U = -EA^{-1}$.

Now since the reduced form equations are amenable to estimation by Ordinary Least Squares, we obtain the least squares estimates of $\Pi$ as $P = (X'X)^{-1}X'Y$. But $\Pi A = -B$ and hence $P \hat{A} = -\hat{B}$. This gives us a system of $k$ equations in $m(m+k-1)$ unknowns even after applying the normalization rule on $A$ such that $a_i = -1$ for all $i = 1,2,...,m$. Obviously, such a system of equations admits no determinate solution. For any particular structural equation, say $j^{th}$ equation, we have $P \hat{a}_j = -\hat{b}_j$, which is a system of $k$ equations in $m+k-1$ unknowns. Unless we assume some (at least as many as $m-1$) elements of $\hat{a}_j$ or $\hat{b}_j$ equal to zero, we cannot estimate other remaining elements of $\hat{a}_j$ and $\hat{b}_j$. Thus, the identifiability or the zero restriction on at least $m-1$ elements of $\hat{a}_j$ and $\hat{b}_j$ is the basic requirement for obtaining the estimates of the coefficients of the $j^{th}$ structural equation. Moreover, the system of equations $P \hat{a}_j = -\hat{b}_j$ must satisfy the rank condition which is the basic algebraic requirement for solving any system of linear equations.

In what follows we describe various methods to estimate the structural equations of a multi-equation (simultaneous equation) model. These methods minimize the Euclidean norm of errors in the dependent variable directly or indirectly. Identifiability of each equation in the model is presumed.

**Indirect Least Squares and Generalized Indirect Least Squares:** The Indirect Least Squares (ILS) method was first proposed by Girshick (ref. by Hood & Koopmans, 1953, p. 140). It was generalized by Khazzoom (1976), though he did not name it the
Generalized Indirect Least Squares (GILS). For a comprehensive treatment we will describe here the method of GILS only.

From the structural equations $YA + XB + E = 0$ we derive the reduced form equations $Y = X\Pi + U$ and estimate its coefficients by $P = (X'X)^{-1}X'y$ by OLS. Since $(X'X)^{-1}X' = X^{-8}$ (the generalized inverse of $X$), we may also write $P = X^{-8}Y$.

The question that arises now is whether we can obtain $A$ and $B$ from the knowledge of $P$. This is the problem of identification. If all the columns of $A$ and $B$ can be obtained then $A$ and $B$ can also be completely obtained. If a particular column of $A$ and $B$ (say, $a_j$ and $b_j$) can be obtained from the relation

$$Pa_j = -b_j \quad \ldots \quad (2.5)$$

then the $j^{th}$ equation is identifiable. Since this is true for any equation and (2.5) holds for any particular column of $A$ and $B$, dropping the subscript $j$ the equation (2.5) can be rewritten as

$$P_{k+m \times 1} a_{k+1} = -b_{k+1} \quad \ldots \quad (2.6)$$

The $j^{th}$ element of $a$ is $-1$ if (2.6) relates to the equation $j$. The system (2.6) is in $k$ equations and $k+m-1$ unknowns. So equation (2.6) cannot be solved for $a$ and $b$.

Now, suppose, from a-priori information the values of some (say $k_2$) elements of $b$ and some other elements of $a$ can be obtained such that $k_2$ elements of $b$ and $m_2$ elements of $a$ are known. In that case, out of $k$ elements of $b$ only $k-k_2 = k_1$ elements are unknown. Similarly out of $m$ elements of $a$, $m - m_2 = m_1$ elements are unknown. Thus, we proceed to identification by restriction on the structural coefficients matrix $a$ and $b$. 
Pre-multiplying (2.6) by a suitable permutation matrix R we may obtain

\[ RPa = -Rb \] ... (2.7)

such that \( b \) can be partitioned into two sub matrices, first of which, \( b_1 \) has \( k_1 \) unknown elements and the second, \( b_2 \) has \( k_2 \) known elements.

Correspondingly, \( R \) can also be partitioned into \( P_1 \) and \( P_2 \). It may noted that such a permutation effects only a re-shuffling of equations in (2.6) and does not have any bearing on its solution. Consequently, equation (2.7) can thus be re-written in a partitioned form as

\[
\begin{bmatrix}
P_1 \\
P_2
\end{bmatrix}
a = -
\begin{bmatrix}
b_1 \\
b_2
\end{bmatrix}; \quad R^{-1}\begin{bmatrix}
b_1 \\
b_2
\end{bmatrix} = b
\] ... (2.7a)

Here, \( P_1, P_2, b_1 \) and \( b_2 \) are \( k_1 \times m, k_2 \times m, k_1 \times 1 \) and \( k_2 \times 1 \) matrices respectively.

Again, pre-multiplying \( a \) by a suitable permutation matrix, S, and post-multiplying \( R \) by \( S^{-1} = S' \) in equation (2.7a) we obtain

\[
\begin{bmatrix}
P_1 \\
P_2
\end{bmatrix} S^{-1}\begin{bmatrix}a_1 \\
a_2
\end{bmatrix} = \begin{bmatrix}
P_1 \\
P_2
\end{bmatrix} S^{-1}\begin{bmatrix}a_1 \\
a_2
\end{bmatrix} ; \quad S^{-1}\begin{bmatrix}a_1 \\
a_2
\end{bmatrix} = a ; \quad S^{-1}S = I
\]

where \( a_1 \) contains \( m_1 \) unknown elements and \( a_2 \) contains \( m_2 \) known elements. Equation (2.7a) may now be rewritten as

\[ RPS^{-1}a = -Rb \] ... (2.7b)

This operation helps to shuffle and partition \( R \) into two column submatrices, without any effect to the solution of the system of equations in (2.6). In the partitioned form (2.7b) can be written as
\[
\begin{bmatrix}
    P_{11} & P_{12} \\
    P_{21} & P_{22}
\end{bmatrix}
\begin{bmatrix}
    a_1 \\
    a_2
\end{bmatrix} =
\begin{bmatrix}
    b_1 \\
    b_2
\end{bmatrix} 
\]  

... (2.8)

From (2.8) we obtain two equations,

\[ P_{11}a_1 + P_{12}a_2 = b_1 \]  

... (2.9)

\[ P_{21}a_1 + P_{22}a_2 = b_2 \]  

... (2.10)

Now, since \( a_2 \) and \( b_2 \) are known, we get

\[ P_{21}a_1 = -[b_2 + P_{22}a_2] \]  

... (2.11)

Here \( P_{21} \) is a \( k_2 \times m_1 \) matrix giving rise to \( k_2 \) equations in \( m_1 \) unknowns as in (2.10).

If \( P_{21} \) is a square (rectangular) matrix in \( k_2 \) rows and \( m_1 \) column and rank \( (P_{21}) = r \leq m_1 \), then the Moore-Penrose inverse of \( P_{21} \) that is \( P_{21}^+ \) exists and is unique. Using this generalized inverse of \( P_{21} \) we obtain from (2.9) and (2.11)

\[ a_1 = -P_{21}^+[b_2 + P_{22}a_2] \]  

... (2.12)

\[ b_1 = -[P_{21}a_1 + P_{12}a_2] \]  

... (2.13)

This formulation and consequent technique in (2.12) and (2.13) is GILS. The ILS is only a special case of GILS when \( k_2 = m_1 = r \) and therefore, \( P_{21}^+ = P_{21}^{-1} \). However, in case \( k_2 \geq m_1 \) and \( r = m_1 \), then \( P_{21}^+ \) yields least squares g-inverse, \( P_{21}^{-1} \) or \( P_{21}^{g} \).

**GILS in Another Form:** Rewriting equation (2.8) as

\[
\begin{bmatrix}
    P_{11} & I \\
    P_{21} & 0
\end{bmatrix}
\begin{bmatrix}
    a_1 \\
    b_1
\end{bmatrix} =
\begin{bmatrix}
    P_{12}a_2 \\
    b_2 + P_{22}a_2
\end{bmatrix} 
\]  

... (2.14)

which may be again rewritten as
\[
\begin{bmatrix}
1 & P_{11} \\
0 & P_{21}
\end{bmatrix}
\begin{bmatrix}
b_1 \\
a_1
\end{bmatrix}
= -\begin{bmatrix}
P_{12}a_2 \\
b_2 + P_{22}a_2
\end{bmatrix}
\tag{2.15}
\]

From (2.15) we obtain
\[
\begin{bmatrix}
1 & P_{11} \\
0 & P_{21}
\end{bmatrix}^\top
\begin{bmatrix}
P_{12}a_2 \\
b_2 + P_{22}a_2
\end{bmatrix}
= -\begin{bmatrix}
b_1 \\
a_1
\end{bmatrix}
\tag{2.16}
\]

Now,
\[
\begin{bmatrix}
1 & P_{11} \\
0 & P_{21}
\end{bmatrix}
\begin{bmatrix}
I & -P_{11}P_{21}^* \\
0 & P_{21}
\end{bmatrix}
\begin{bmatrix}
I & P_{11} \\
0 & P_{21}
\end{bmatrix}
= \begin{bmatrix}
I & P_{11} \\
0 & P_{21}
\end{bmatrix}
\tag{2.17}
\]

or,
\[
\begin{bmatrix}
1 & 0 \\
0 & P_{21}P_{21}^*
\end{bmatrix}
\begin{bmatrix}
I & P_{11} \\
0 & P_{21}
\end{bmatrix}
= \begin{bmatrix}
I & P_{11} \\
0 & P_{21}
\end{bmatrix}
\tag{2.18}
\]

or,
\[
\begin{bmatrix}
1 & P_{11} \\
0 & P_{21}P_{21}^*
\end{bmatrix}
= \begin{bmatrix}
I & P_{11} \\
0 & P_{21}
\end{bmatrix}
\tag{2.19}
\]

Since \( P_{21}P_{21}^*P_{21} = P_{21} \) (Theil, pp. 269-70), (2.19) is equal to (2.17).

Thus,
\[
\begin{bmatrix}
I & P_{11} \\
0 & P_{21}
\end{bmatrix}^* = \begin{bmatrix}
I & -P_{11}P_{21}^* \\
0 & P_{21}^*
\end{bmatrix}
\tag{2.20}
\]

It goes without saying that since the least squares g-inverse is only a special case of Moore-Penrose inverse, the above proof is true for g-inverse \([\cdot]^g\) as well.

Equation (2.16) may now be written as
Two Stage Least Squares: The most widely used method for estimating the parameters of an equation of a simultaneous equation model is the 2-SLS developed by H. Theil and independently by Basmann who called it Generalized Classical Linear Estimation method. It is applicable to over-identified/exactly identified equations. In the case of exactly identified equation, the 2-SLS estimates are identical with the ILS estimates given by equations (2.12) and (2.13). The basic idea behind the 2-SLS is to substitute each endogenous explanatory variable (that is correlated with the residuals) by a corresponding composite variable (uncorrelated with the residuals) which is a linear function of all exogenous variables in the model. Since the surrogate (instrumental) variables are uncorrelated with the residuals, the estimates of the parameters will be consistent.

Let us suppose that we are interested in estimating the first equation of the general interdependent system of equations.

\[ y_1 = Y_1 a_1 + X_1 b_1 + e_1 \]  \hspace{1cm} \ldots (2.22)

In the notation above, \( y_1 \) is the dependent endogenous variable in the first structural equation and on account of normalization rule the coefficient associated with it is unity. \( Y_1 \) and \( X_1 \) are explanatory endogenous and exogenous variable (respectively) that appear in the first structural equation while \( a_1 \) and \( b_1 \) are the coefficients associated with them. In the first stage, reduced form regression of \( Y_1 \) on all the pre-determined variables in the
system \((X)\) is performed. The estimates \(\hat{Y}_i\) thus obtained are then used to replace the actual observations on the \(Y_1\) variables. Hence,

\[
\hat{Y}_i = X (XX)^{-1} XY_1
\]  

... (2.23)

Application of OLS in the second stage to the equation thus transformed yields the estimating equations.

\[
\begin{pmatrix}
\hat{Y}_i' \\
\hat{Y}_i \\
X_i' \\
X_i
\end{pmatrix}
\begin{pmatrix}
c \\
d
\end{pmatrix} =
\begin{pmatrix}
\hat{Y}_i' \ y_1 \\
X_i' \ y_1
\end{pmatrix}
\]

... (2.24)

The vector \(\begin{pmatrix} c \\ d \end{pmatrix}\) now denotes the 2SLS estimator of \(\begin{pmatrix} a_i \\ b_i \end{pmatrix}\). Since the above system contains the same number of equations and unknowns, i.e. \(m+k-1\) equations in \(m+k-1\) unknowns, the system in general will have a unique solution.

The form in which the 2SLS equations are usually presented can be derived in the following way:

The matrix \(Y_1\) can be written as

\[
Y_i = \hat{Y}_i + E_i
\]  

... (2.25)

Now, since the OLS residuals are orthogonal to the estimated value of the dependent variable and to each of the explanatory variables, therefore,

\[
\hat{Y}_i' E_i = 0 \quad \text{and} \quad X_i' E_i = 0
\]

Using equation (2.25) we get,
\[
\hat{Y}_i'X_i = (Y_i - E_i)'X_i = Y_i'X_i
\]

Thus for $\kappa = 1$ the 2SLS estimator can now be written as

\[
\begin{pmatrix}
Y_i' Y_i - \kappa (E_i' E_i) & Y_i' X_i \\
X_i' Y_i & X_i' X_i
\end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} (Y_i - \kappa E_i)' y_i \\ X_i'y_i \end{pmatrix}
\]  \qquad \text{(2.26)}

The form shows clearly how the 2SLS estimator differs from the inconsistent OLS estimator, which is given by

\[
\begin{pmatrix}
Y_i' Y_i & Y_i' X_i \\
X_i' Y_i & X_i' X_i
\end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} y_i' y_i \\ X_i'y_i \end{pmatrix}
\]  \qquad \text{(2.27)}

or

\[
\begin{pmatrix}
Y_i' Y_i - \kappa (E_i' E_i) & Y_i' X_i \\
X_i' Y_i & X_i' X_i
\end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} (Y_i - \kappa E_i)' y_i \\ X_i'y_i \end{pmatrix} \text{ for } \kappa = 0 \quad \text{(2.28)}
\]

**Limited-Information Maximum Likelihood Estimator:** The limited information maximum likelihood (LIML) method is another alternative approach that gives consistent
estimates and was developed by Anderson & Rubin. They estimated the parameters of an equation by maximizing the likelihood function subject to appropriate constraints using Lagrange multipliers. Hood & Koopmans have used a different method to develop the LIML estimator. They arrived at the estimates of the parameters of an equation by maximizing the likelihood function for the observations on the endogenous variables included in that equation disregarding the over identifying restrictions on other structural equations. This method assumes that the structural disturbances are normally distributed.

The structural equation to be estimated is again

\[
y_1 = Y_i q_i + X_i b_i + e_i
\]

... (2.29)

Equation (2.29) can be written as

\[
y_1^* = X_i b_i + e_i
\]

... (2.30)

where, \( y_1^* = y_1 - Y_i q_i \). Now \( y_1^* \) is a linear combination of the endogenous variables included in the first structural equation. The coefficient \( b_1 \) in equation (2.30) can be estimated by ordinary least squares method. Thus, the estimator is

\[
\hat{b}_1 = (X_1' X_1)^{-1} X_1' y_1^*
\]

... (2.31)

and the sum of squared errors is

\[
S_1 = (y_1^* - X_1 \hat{b}_1)' (y_1^* - X_1 \hat{b}_1)
\]

\[= y_1^* y_1^* - y_1^* X_1' (X_1' X_1)^{-1} X_1 y_1^*\]

... (2.32)

Now, let the relation which includes all the predetermined variables be given by

\[
y_1^* = X b + \varepsilon_1
\]

... (2.33)
where, $X$ is an nxk matrix of all pre-determined variables and

$$b = \begin{bmatrix} b_1 \\ O^* \end{bmatrix}$$

Application of OLS to equation (2.33) gives

$$\hat{b} = (X'X)^{-1}X'Y_i^*$$

... (2.34)

and the sum of squared errors is

$$S_2 = (y_i^* - Xb)(y_i^* - Xb)$$

$$= y_i^{**}y_i^* - y_i^{**}X(X'X)^{-1}X'y_i^*$$

... (2.35)

Using equations (2.32) and (2.35) we can obtain the ratio between the two sums as

$$l = \frac{S_1}{S_2} = \frac{y_i^{**}y_i^* - y_i^{**}X(X'X)^{-1}X'y_i^*}{y_i^{**}y_i^* - y_i^{**}X(X'X)^{-1}X'y_i^*}$$

... (2.36)

The ratio $l$ cannot be less than one simply because that the explanatory power of $X$ cannot be less than $X_1$ and therefore, the variance $S_1$ cannot be less than the variance $S_2$.

To estimate the elements of $a_i$, let us suppose that

$$y_i^* = Y_{i\alpha}a_{i\alpha}$$

... (2.37)

where $Y_{i\alpha} = [Y_1 \ Y_2 \ \cdots \ Y_m]$ and $a_{i\alpha} = \begin{bmatrix} 1 \\ a_{i12} \\ \vdots \\ a_{im} \end{bmatrix}$

The ratio $l$ can be written as
\[ l = \frac{a_{i\lambda}' W_i a_{i\lambda}}{a_{i\lambda}' W_i a_{i\lambda}} \quad \ldots (2.38) \]

where,

\[ W_i^* = Y_{i\lambda}'Y_{i\lambda} - (Y_{i\lambda}'X_\Lambda)(X_\Lambda'X_\Lambda)^{-1}X_\Lambda'Y_{i\lambda} \]

\[ W_1 = Y_{i\lambda}'Y_{i\lambda} - (Y_{i\lambda}'X)(X'X)^{-1}XY_{i\lambda} \]

The limited information maximum likelihood method is to choose such values of the coefficient that would minimize \( l \). Thus, on differentiating \( l \) with respect to \( a_{i\lambda} \) we have

\[ \frac{\partial l}{\partial a_{i\lambda}} = \frac{2(W_i a_{i\lambda})(a_{i\lambda}' W_i a_{i\lambda}) - 2(W_i a_{i\lambda})(a_{i\lambda}' W_i a_{i\lambda})}{(a_{i\lambda}' W_i a_{i\lambda})^2} \quad \ldots (2.39) \]

The necessary condition for minimization requires that the first order derivative is equated to zero. Therefore,

\[ \frac{2}{(a_{i\lambda}' W_i a_{i\lambda})} \left[ W_i a_{i\lambda} - \left( \frac{a_{i\lambda}' W_i a_{i\lambda}}{a_{i\lambda}' W_i a_{i\lambda}} \right) W_i a_{i\lambda} \right] = 0 \]

Using equation (2.38) we get,

\[ W_i a_{i\lambda} - lw_i a_{i\lambda} = 0 \quad \ldots (2.40) \]

Now, for \( a_{i\lambda} \neq 0 \), the determinant of the matrix formed must be equal to zero, i.e.,

\[ \left| W_i^* - lw_i \right| = 0 \quad \ldots (2.41) \]

The elements of \( W_i^* \) and \( W_i \) can be obtained from the sample observations (as defined for (2.38). The equation (2.41), therefore, becomes a polynomial of degree in \( l \)
which must be solved for the smallest root \( \hat{I} \). This may also be considered as an eigen value problem (Krishnamurthy & Sen, pp. 248-249). Substituting the value of \( \hat{I} \) thus obtained in equation (2.40) and solving the simplified equation

\[
\left(W_1^* - IW_1\right)\hat{a}_{1\alpha} = 0
\] ...

(2.42)

the estimator \( \hat{a}_{1\alpha} \) can be obtained. Next, the elements of \( b_1 \) can be obtained by

\[
\hat{b}_1 = (X_1'X_1)^{-1}(X_1'y_1) = (X_1'X_1)^{-1}X_1'y_{1\alpha}\hat{a}_{1\alpha}
\]

(2.43)

Equations (2.42) and (2.43) define the LIML estimates of the structural equation.

The equation-by-equation estimation methods, namely GILS, ILS, 2SLS and LIML discussed so far, are all essentially limited information estimators. In the estimation of any structural equation, these estimators do not take into account complete information on all structural equations in the model. Although the estimators give consistent estimates, they are not asymptotically efficient, as they do not consider the correlation of disturbances across the equation. This deficiency can be overcome by estimating all the equations of the system simultaneously. For this purpose the full information methods such as 3SLS or FIML can be used (Intriligator, pp. 402-416; Johnston, pp. 486-492; Theil, pp. 508-528). Since we have not used these estimators in our investigation, we do not consider it necessary to describe them here.