Lecture 7 – Three-Stage Least Squares

7.1 Introduction to Simultaneous Equation Techniques

OLS (Ordinary Least Squares), 2SLS (Two-Stage Least Squares) and LIML (Limited Information Maximum Likelihood) are all essentially limited-information estimators in that in the estimation of any structural equation complete information on all the other structural equations is not taken into account. In principle, information on the complete structure will yield estimators with greater asymptotic efficiency.

7.2 Three-Stage Least Squares

Three-Stage Least Squares (3SLS) was developed by Zellner & Theil in their 1962 paper (see ‘useful references’). It combines 2SLS + SURE.

The first step is to get an equation which is equivalent to the 2SLS estimator. Consider a general linear model containing $G$ jointly-dependent endogenous variables with $K$ predetermined variables where the $i^{th}$ equation is:

$$y_i = Y_i \beta_i + X_i \gamma_i + u_i$$

*Figure 2.1*

where $y_i$ is an $n \times 1$ vector of sample observations on the dependent variable in the $i^{th}$ equation, $Y_i$ is an $n \times g_i$ matrix of observations on the other endogenous variables in the equation, and $X_i$ is an $n \times k_i$ matrix. This can then be rewritten as:

$$y_i = Z_i \delta_i + u_i$$

where $Z_i = [Y_i \quad X_i]$ and $\delta_i = [\beta_i \quad \gamma_i]^T$  

*Figure 2.2*

We then premultiply the equation in Figure 2.2 by $X'$, the $n \times k$ matrix of all the predetermined variables in the model:

$$X' y_i = X' Z_i \delta_i + X' u_i, \quad i = 1, ..., G$$

*Figure 2.3*
The variance covariance matrix of this disturbance term is:

\[ E(X' u, u' X) = \sigma^2_u X X \]

**Figure 2.4**

on the standard assumption that \( E(u, u') = \sigma_u I \). The obvious way to estimate Figure 2.3 is by GLS (Generalised Least Squares)\(^1\). Hence the GLS estimator of \( \delta_i \) is:

\[ \delta_i = \left[ Z' X (X' X)^{-1} X' Z_i \right]^{-1} Z_i' X (X' X)^{-1} X' y_i \]

**Figure 2.5**

This is in fact just another way of writing the 2SLS of Figure 2.2. We also know from our previous work on the GLS estimators that it is possible to interpret it as equivalent to the estimator given by the application of OLS to suitably transformed data. This approach leads to a considerable simplification in the presentation of the 3SLS estimator. It can be shown (and has been used before in the notes on serial correlation – Lecture 4 of ES30027) that a non-singular matrix \( P \) exists such that:

\[ (X' X)^{-1} = PP' \]
\[ P' X X P = I \]

**Figure 2.6**

Intuitively I call \( P \) the square root matrix of \( (X' X)^{-1} \) Premultiplying the equation in Figure 2.3 by \( P' \):

\[ P' X' y_i = P' X Z_i \delta_i + P' X' u_i \]
\[ \text{or } w_i = W_i \delta_i + v_i \]

**Figure 2.7**

The variance matrix for the disturbance term is:

\[ E(v, v') = E(P' X' u, u' X P) \]
\[ = \sigma^2_u P' X P \]
\[ = \sigma^2_u I \]

---

\(^1\) In general the GLS estimator is, in the model \( y = X \beta + u \), \( \beta_* = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \), where \( E(u' u) = \sigma^2 \Omega \).
Figure 2.8

The application of OLS to the second equation in Figure 2.7 then gives:

$$\delta_i = (W_i'W_i)^{-1}W_i'w_i$$

Figure 2.9

This can be seen as the 2SLS estimator in Figure 2.5. This marks the end of step 1. We now apply SURE to these transformed equations.

Collecting all $G$ structural equations together gives:

$$
\begin{bmatrix}
w_1 \\
w_2 \\
\vdots \\
w_G 
\end{bmatrix}
=
\begin{bmatrix}
W_1 & 0 & \cdots & 0 \\
0 & W_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & W_G 
\end{bmatrix}
\begin{bmatrix}
\delta_1 \\
\delta_2 \\
\vdots \\
\delta_G 
\end{bmatrix}
+
\begin{bmatrix}
v_1 \\
v_2 \\
\vdots \\
v_G 
\end{bmatrix}
$$

Figure 2.10

Or more compactly:

$$w = W\delta + v$$

Figure 2.11

3SLS ‘simply’ involves estimating Figure 2.10 using SURE (see Lecture 7 of ES30027).

The variance matrix for the $V$ vector is:

$$
V = E(vv') = 
\begin{bmatrix}
\sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1G}I \\
\sigma_{21}I & \sigma_{22}I & \cdots & \sigma_{2G}I \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{GI}I & \sigma_{G2}I & \cdots & \sigma_{GG}I 
\end{bmatrix}
= \Sigma \otimes I
$$

Figure 2.12

Thus the basic assumptions are that (i) each structural equation has a homoskedastic non-autocorrelated error term and (ii) that the disturbances in different structural equations may be contemporaneously correlated. That is the error terms from the $i$th and $j$th equations, for example, are correlated. Provided that at least some of the $\sigma_{ij}$ are nonzero we should use SURE as there is then an efficiency gain. The only difficulty is that the $\Sigma$ matrix is unknown. Zellner & Theil suggest that we should first estimate each structural equation by 2SLS.
obtaining the residual vectors \( \hat{u}_i = y_i - Z_i \delta_i \) \((i = 1, \ldots, G)\), where \( \delta_i \) is the 2SLS estimator of \( \delta_i \).

The elements of \( \Sigma \) are then estimated by:

\[
S_{ij} = \frac{\hat{u}_i' \hat{u}_j}{n} \quad \forall i, j
\]

**Figure 2.13**

Note that although the maximum likelihood estimators of \( s_{ij} \) this will give biased estimators and arguably we should correct for that with not \( n \) in the denominator but \( [(n-k_i)(n-k_j)]^{0.5} \) where \( k_i \) is the number of parameters estimated in the \( i \)th equation and the superscript 0.5 denotes the square root:

In any case, this gives:

\[
\hat{V} = \hat{\Sigma} \otimes I
\]

**Figure 2.14**

The 3SLS estimator of \( \delta \) is then:

\[
\delta_{3SLS} = \left(W \hat{V}^{-1} W\right)^{-1} W \hat{V}^{-1} W
\]

**Figure 2.15**

The asymptotic variance matrix is estimated by:

\[
\text{asy var}(\delta_{3SLS}) = \left(W \hat{V}^{-1} W\right)^{-1}
\]

**Figure 2.16**

A necessary condition for the superior efficiency of 3SLS over 2SLS is that the specification of the model should be correct. Granted this, 3SLS will be equivalent to 2SLS if:

(a) \( \sigma_{ij} = 0 \) \( \forall i \neq j \)

or

(b) If all equations in the system are exactly identified.

Before attempting to apply 3SLS one must omit all underidentified equations and all identities, since the latter have zero disturbances, which would render the \( \Sigma \) (and hence \( V \)) matrix singular and non-invertible. Suppose that there remain \( G \)
identified equations, of which $G^*$ are exactly identified and $G'''$ overidentified. Zellner & Theil have shown that the 3SLS estimators of the $G''$ equations, treated as a group, are the same as that obtained from the application of 3SLS to the complete system of equations. Thus it is computationally efficient to obtain the 3SLS estimates in two steps. First compute the 3SLS estimates of the overidentified equations, and then compute using a formula involving the 2SLS estimates for the other equations (see Zellner & Theil (1962), p.67).

**Useful references:**