# 2SLS

2SLS produces the two-stage least-squares estimation for a structure of simultaneous linear equations.

## Notation

The following notation is used throughout this chapter unless otherwise stated:

| | |
|---|---|
| $p$ | Number of predictors |
| $p_1$ | Number of endogenous variables among $p$ predictors |
| $p_2$ | Number of non-endogenous variables among $p$ predictors |
| $k$ | Number of instrument variables |
| $n$ | Number of cases |
| $\mathbf{y}$ | $n \times 1$ vector which consists of a sample of the dependent variable |
| $\mathbf{Z}$ | $n \times p$ matrix which represents observed predictors |
| $\beta$ | $p \times 1$ parameter vector |
| $\mathbf{X}$ | $n \times k$ matrix with element $x_{ij}$, which represents the observed value of the $j$th instrumental variable for case $i$ |
| $\mathbf{Z}_1$ | Submatrix of $\mathbf{Z}$ with dimension $n \times p_1$, which represents observed endogenous variables |
| $\mathbf{Z}_2$ | Submatrix of $\mathbf{Z}$ with dimension $n \times p_2$, which represents observed non-endogenous variables |
| $\beta_1$ | Subvector of $\beta$ with parameters associated with $\mathbf{Z}_1$ |
| $\beta_2$ | Subvector of $\beta$ with parameters associated with $\mathbf{Z}_2$ |

# Model

The structure equations of interest are written in the form

$$\mathbf{y} = \mathbf{Z}\beta = [\mathbf{Z}_1, \mathbf{Z}_2]\begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} + \varepsilon$$

$$\mathbf{Z}_1 = \mathbf{X}\gamma + \delta \qquad \textbf{(1)}$$

where

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2]\,,\ \beta = \begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix}$$

and $\varepsilon$ and $\delta$ are the disturbances with zero means and covariance matrices $\sigma^2\mathbf{I}_n$ and $\varsigma^2\mathbf{I}_n$, respectively.

# Estimation

The estimation technique used was developed by Theil (1953, a, b). Let us first premultiply both sides of equation **(1)** by $\mathbf{X}'$ to obtain

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{Z}\beta + \mathbf{X}'\varepsilon \qquad \textbf{(2)}$$

Since the disturbance vector has zero mean and covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})$, it is easy to see that $(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'\varepsilon$ would have a covariance matrix $\sigma^2\mathbf{I}_n$. Thus, multiplying $(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}$ to both sides of equation **(2)** results in a multiple linear regression model

$$(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'\mathbf{Z}\beta + (\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'\varepsilon \qquad \textbf{(3)}$$

The ordinary least-square estimator $\hat{\beta}$ for $\beta$ is

$$\hat{\beta} = \left(\mathbf{Z'X}(\mathbf{X'X})^{-1}\mathbf{X'Z}\right)^{-1}\mathbf{Z'X}(\mathbf{X'X})^{-1}\mathbf{X'y}$$

# Computation Details

- 2SLS constructs a matrix **R**,

  $$\mathbf{R} = \begin{bmatrix} \mathbf{1} & \mathbf{V'} \\ \mathbf{V} & \mathbf{M} \end{bmatrix}$$

  where

  $$\mathbf{M} = \mathbf{C}_{zx}(\mathbf{C}_{xx})^{-1}\mathbf{C}'_{zx}$$
  $$\mathbf{V} = \mathbf{C}_{zx}(\mathbf{C}_{xx})^{-1}\mathbf{C}'_{xy}$$

  and $\mathbf{C}_{zx}$ is the correlation matrix between **Z** and **X**, and $\mathbf{C}_{xx}$ is the correlation matrix among instrumental variables.

- Sweep the matrix **R** to obtain regression coefficient estimate for $\beta$.

- Compute sum of the squares of residuals (SSE) by

  $$\mathbf{y'y} - \mathbf{uZ'y} - \mathbf{y'Zu'} + \mathbf{uZ'Zu'}$$

  where

  $$\mathbf{u} = \mathbf{y'X}(\mathbf{X'X})^{-1}\mathbf{X'z}\left[\mathbf{z'X}(\mathbf{X'X})^{-1}\mathbf{X'z}\right]^{-1}$$

- Compute the statistics for the ANOVA table and for variables in the equation. Details can be found in REGRESSION.

# References

Theil, H. 1953a. *Repeated least square applied to complete equation systems*. The Hague: Central Planning Bureau.

Theil, H. 1953b. *Estimation and simultaneous correlation in complete equation systems*. The Hague: Central Planning Bureau.