Advanced Econometrics: Three-Stage Least Squares (3SLS)

Prepared by

Aditya Korekallu Srinivasa

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute

New Delhi

April 25, 2025

Contents

1 Introduction: Why 3SLS?

1.1 Motivation

In many economic applications, especially in agriculture and development, we encounter systems of equations with:

- Endogenous variables appearing as regressors in multiple equations,
- Correlation among error terms across equations (contemporaneous correlation),
- The need for efficient and consistent estimation of all parameters in the system.

Three-Stage Least Squares (3SLS) is a system estimation method that generalizes 2SLS by exploiting both the endogeneity structure and the correlation among error terms, leading to more efficient estimates.

1.2 When to Use 3SLS

- When you have a system of simultaneous equations (not just one equation).
- When errors across equations are correlated (e.g., due to unobserved shocks affecting multiple outcomes).
- When you want to estimate all equations jointly for efficiency.

2 The Structure of 3SLS Estimation

2.1 General System

Suppose we have a system of G linear equations, each with n observations:

$$y_{gi} = x'_{gi}\beta_g + u_{gi}, \quad g = 1, \ldots, G; \quad i = 1, \ldots, n$$

where:

- y_{gi}: Dependent variable in equation g,
- x_{gi}: Vector of regressors (including endogenous and exogenous variables),
- β_g: Parameter vector,
- u_{gi}: Error term.

2.2 Error Covariance Structure

The vector of errors across equations for observation i is:

$$u_i = (u_{1i}, u_{2i}, \ldots, u_{Gi})'$$

Assume $E[u_i] = 0$, $E[u_i u_i'] = \Omega$ (a $G \times G$ positive definite matrix, constant across i), and $E[u_i u_j'] = 0$ for $i \neq j$.

2.3 Stacked System Representation

Stack all equations for all observations:

$$Y = X\beta + u$$

where Y is $nG \times 1$, X is block-diagonal, β is stacked, and u is $nG \times 1$.

3 Three-Stage Least Squares (3SLS): Step-by-Step

3.1 Stage 1: Obtain Fitted Values for Endogenous Regressors (2SLS)

1. For each equation, regress each endogenous regressor on all exogenous variables (instruments) in the system.
2. Obtain fitted values for all endogenous regressors.
3. Replace endogenous regressors in each equation with their fitted values.

3.2 Stage 2: Estimate Each Equation by 2SLS

1. Using the fitted values, estimate each equation by 2SLS.
2. Obtain residuals for each equation.

3.3 Stage 3: Generalized Least Squares (GLS) on the System

1. Compute the sample covariance matrix of the residuals across equations ($\hat{\Omega}$).
2. Use $\hat{\Omega}$ to perform GLS on the system, exploiting the cross-equation error correlation.
3. The 3SLS estimator is:

$$\hat{\beta}_{3SLS} = (X'(\Omega^{-1} \otimes I_n)X)^{-1}X'(\Omega^{-1} \otimes I_n)Y$$

where $\otimes$ is the Kronecker product, and I_n is the $n \times n$ identity matrix.

3.4 Interpretation

- 3SLS combines the consistency of 2SLS with the efficiency of Seemingly Unrelated Regressions (SUR).
- If errors are uncorrelated across equations, 3SLS reduces to 2SLS.

4 Mathematical Derivation of 3SLS Estimator

4.1 Notation

Let $Y = (Y_1', Y_2', ..., Y_G')'$, $X = \text{blockdiag}(X_1, ..., X_G)$, $\beta = (\beta_1', ..., \beta_G')'$, $u = (u_1', ..., u_G')'$.

4.2 System Covariance Structure

Let Ω be the $G \times G$ covariance matrix of the errors for each observation. The full covariance of u is $\Sigma = \Omega \otimes I_n$.

4.3 3SLS Estimator

The Generalized Least Squares (GLS) estimator for the system is:

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

But since X contains fitted values from the first stage (instruments), this is 3SLS.

4.4 Properties

- 3SLS is consistent and asymptotically efficient among estimators using linear combinations of instruments.
- If the system is recursive (no endogenous regressors on the right), 3SLS reduces to SUR.
- If errors are uncorrelated, 3SLS reduces to 2SLS.

4.5 Assumptions

- Correct specification of the system.
- Valid instruments (relevance and exogeneity).
- Homoskedasticity and no serial correlation within equations (or use robust versions).

5 Identification in 3SLS Systems

5.1 Order and Rank Conditions

- Each equation must be identified (order and rank conditions) for 3SLS to be valid.
- Overidentification allows for testing instrument validity.

5.2 Testing Overidentifying Restrictions

- Sargan-Hansen tests can be applied equation-by-equation or system-wide.
- Failure to reject indicates instruments are valid.

6 Practical Implementation: Example from Agriculture

6.1 Example: Farm Production and Consumption System

Suppose we have a two-equation system for smallholder farms:

$$\text{Production}_i = \alpha_0 + \alpha_1 \text{Labor}_i + \alpha_2 \text{Fertilizer}_i + \alpha_3 \text{Consumption}_i + u_{1i}$$
$$\text{Consumption}_i = \beta_0 + \beta_1 \text{Production}_i + \beta_2 \text{Income}_i + \beta_3 \text{FamilySize}_i + u_{2i}$$

Here, Production and Consumption are endogenous in both equations.

Steps:

1. Use exogenous variables (e.g., Labor, Fertilizer, Income, FamilySize) as instruments.
2. Stage 1: Predict endogenous variables using instruments.
3. Stage 2: Estimate each equation by 2SLS.
4. Stage 3: Estimate the system by GLS using the covariance of residuals.

Interpretation:

- 3SLS provides more efficient estimates of the effect of, say, fertilizer on production, accounting for the feedback between production and consumption and their error correlation.

6.2 Software Implementation

- **R:** Use the `systemfit` package.
- **Stata:** Use the `reg3` command.
- **Python:** Use `linearmodels.system` (limited support).

7 Comparison: 2SLS, SUR, 3SLS, FIML

Method	Endogeneity	Error Correlation	System Estimation
2SLS	Yes	No	No
SUR	No	Yes	Yes
3SLS	Yes	Yes	Yes
FIML	Yes	Yes	Yes (full info)

8 Advantages and Limitations of 3SLS

8.1 Advantages

- More efficient than 2SLS if errors are correlated across equations.
- Allows joint hypothesis testing across equations.
- Useful in policy analysis where multiple outcomes are jointly determined.

8.2 Limitations

- Sensitive to misspecification in any equation (system estimation).
- Requires correct identification of all equations.
- Computationally more intensive.
- If instruments are weak or invalid, 3SLS can be inconsistent.

9 Agricultural Economics Papers Using 3SLS: Practical Examples

9.1 Example 1: Technology Adoption and Productivity

Source: Matuschke, I., Qaim, M. (2009). "The impact of social networks on hybrid seed adoption in India." *Agricultural Economics*, 40(5), 493–505.

Endogeneity: Adoption of hybrid seed and farm productivity are jointly determined—more productive farmers may be more likely to adopt, and adoption may increase productivity.

How 3SLS Solved It: The authors specified a system where adoption and productivity equations are estimated jointly, using exogenous variables (e.g., access to extension, social network characteristics) as instruments. 3SLS captured the feedback and improved efficiency.

9.2 Example 2: Fertilizer Demand and Credit Access

Source: Guirkinger, C., Boucher, S. (2008). "Credit constraints and productivity in Peruvian agriculture." *Agricultural Economics*, 39(3), 295–308.

Endogeneity: Fertilizer demand and access to credit are endogenous—credit affects input use, and productive farms are more likely to get credit.

How 3SLS Solved It: The authors estimated a system with fertilizer demand and credit access equations, using exogenous variables (e.g., land size, household characteristics) as instruments, and 3SLS to account for error correlation.

9.3 Example 3: Market Participation and Farm Income

Source: Barrett, C. B. (2008). "Smallholder market participation: Concepts and evidence from eastern and southern Africa." *Food Policy*, 33(4), 299–317.

Endogeneity: Market participation and income are jointly determined—higher income enables participation, and participation increases income.

How 3SLS Solved It: System of equations for participation and income, using household characteristics and asset variables as instruments, estimated by 3SLS.

9.4 Example 4: Irrigation, Crop Choice, and Yield

Source: Dillon, B., Barrett, C. B. (2017). "Agricultural factor markets in Sub-Saharan Africa: An updated view with formal tests for market failure." *Food Policy*, 67, 64–77.

Endogeneity: Irrigation use, crop choice, and yield are jointly determined.

How 3SLS Solved It: Joint estimation of equations for irrigation, crop choice, and yield, using exogenous weather and policy variables as instruments.

10 Practice Questions

1. Explain in detail the three stages of 3SLS and why each is necessary.
2. Discuss the efficiency gains of 3SLS over 2SLS and SUR.
3. In the context of smallholder farm systems, design a 3SLS system for production, consumption, and credit access.
4. How do you test for the validity of instruments in a 3SLS system?
5. Using one of the agricultural economics examples above, explain the identification strategy and the role of 3SLS.
6. What are the limitations of 3SLS in empirical research?
7. Compare and contrast 3SLS and FIML.

Glossary

- **3SLS**: Three-stage least squares, a system estimation method for simultaneous equations.
- **SUR**: Seemingly unrelated regressions, a system method for equations with correlated errors but no endogeneity.
- **2SLS**: Two-stage least squares, single-equation IV method.
- **GLS**: Generalized least squares, efficient estimation in the presence of correlated errors.
- **Instrument**: Exogenous variable used to identify endogenous regressors.
- **Identification**: Ability to uniquely estimate model parameters.
- **Sargan-Hansen Test**: Test for overidentifying restrictions (instrument validity).

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.
- Matuschke, I., & Qaim, M. (2009). "The impact of social networks on hybrid seed adoption in India." *Agricultural Economics*, 40(5), 493–505.

- Guirkinger, C., & Boucher, S. (2008). "Credit constraints and productivity in Peruvian agriculture." *Agricultural Economics*, 39(3), 295–308.
- Barrett, C. B. (2008). "Smallholder market participation: Concepts and evidence from eastern and southern Africa." *Food Policy*, 33(4), 299–317.
- Dillon, B., & Barrett, C. B. (2017). "Agricultural factor markets in Sub-Saharan Africa: An updated view with formal tests for market failure." *Food Policy*, 67, 64–77.