

04_2SLS

QuantFit Estimator Standard Operating Procedure

SOP: Two-Stage Least Squares (2SLS / IV)

Instrumental variables estimator for endogenous regressors

=> Use 2SLS when at least one regressor is correlated with the error and you have valid instruments.

1. Purpose

2SLS replaces an endogenous regressor with its projection onto a set of instruments Z, breaking the correlation between the regressor and the error. Identification requires Z to be relevant (correlated with X) and exogenous (uncorrelated with epsilon).

2. When to use this estimator

Endogeneity from omitted variables, simultaneity, or measurement error.

Reverse causality concerns (e.g. wages and education).

Available exclusion restrictions: instruments that affect Y only through X.

3. Required data structure

At least one column flagged as endogenous regressor.

Instrument set Z with cardinality \geq number of endogenous regressors (just- or over-identified).

Sample size large enough for first-stage F-statistic > 10 per Stock-Yogo rule.

4. Mathematical formulation

First stage projects each endogenous X onto instruments; second stage regresses Y on the projections plus exogenous regressors.

Stage 1: $X = Z\pi + u \Rightarrow X^* = Z(Z'Z)^{-1} Z'X$

Stage 2: $Y = X^*\beta + \epsilon$

$\beta_{2SLS} = (X'^*P_Z X)^{-1} X'^*P_Z Y$, $P_Z = Z(Z'Z)^{-1}Z'$

5. Pre-estimation diagnostics

Test instrument relevance: first-stage F per endogenous regressor (≥ 10).

Test instrument exogeneity: Sargan / Hansen J test in over-identified models.

Wu-Hausman / Durbin-Wu-Hausman for endogeneity (compare OLS vs 2SLS).

6. Estimation procedure

Partition X into endogenous X1 and exogenous X2; assemble instrument matrix Z.

Stage 1: regress each X1 on (X2, Z) and obtain X1*.

Stage 2: regress Y on (X₁, X₂) by OLS.

Compute heteroskedasticity-robust SE; correct for two-stage uncertainty.

Report first-stage F, Sargan J, and Wu-Hausman.

7. Output produced

8. Output interpretation

First-stage F < 10 => weak instruments; 2SLS estimates are unreliable.

Sargan J p < 0.05 => at least one instrument is invalid.

Wu-Hausman p < 0.05 => regressor is endogenous; 2SLS preferred over OLS.

9. Post-estimation diagnostics

Always report first-stage F by endogenous variable.

Heteroskedasticity-robust or cluster-robust SE.

If over-identified, Sargan / Hansen J statistic must be reported.

10. Common pitfalls

Weak instruments produce wide CIs and severe finite-sample bias toward OLS.

Many instruments can paradoxically increase finite-sample bias - use parsimoniously.

Exclusion restrictions are untestable in just-identified models; defend them theoretically.

11. Reporting checklist

First-stage F-statistic per endogenous regressor.

Sargan / Hansen J in over-identified specifications.

Wu-Hausman test of endogeneity.

List of instruments and the exclusion restriction rationale.

12. References

Stock, J. H., Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression.

Hayashi, F. (2000). Econometrics. Princeton.

Angrist, J. D., Pischke, J.-S. (2009). Mostly Harmless Econometrics. Princeton.

Field | Meaning

coefficients | Second-stage slopes beta?

metadata['firstStageF'] | First-stage F per endogenous regressor

metadata['sarganJ'] / sarganP | Over-identification test

metadata['wuHausman'] / whP | Endogeneity test