QuantFit Estimator SOP - Complete Set

QuantFit Estimator Standard Operating Procedure

SOP DOCUMENT: 01_OLS

SOP: Ordinary Least Squares (OLS)

The workhorse linear regression estimator

=> Use OLS as the baseline whenever residuals are well-behaved and X is exogenous.

1. Purpose

OLS estimates the linear conditional mean of Y given X by minimising the sum of squared residuals. It is BLUE (Best Linear Unbiased Estimator) under the Gauss-Markov assumptions: linearity in parameters, strict exogeneity, no perfect collinearity, homoskedasticity, and no autocorrelation of the error term.

2. When to use this estimator

Cross-sectional data with a single equation and exogenous regressors.

Time series where residuals are not autocorrelated and the data is stationary.

As the first benchmark before moving to richer estimators (FE, IV, ARDL).

When R2 and parsimony are the goal and residual diagnostics are clean.

3. Required data structure

Variables in numeric columns; categorical predictors must be dummy-encoded.

Single dependent variable Y; one or more X variables; no panel grouping required.

Sample size n >= k + 30 as a soft rule of thumb (k = number of regressors).

Missing data must be handled before estimation (listwise deletion or imputation).

4. Mathematical formulation

The model Y = Xbeta + epsilon is estimated by minimising epsilon'epsilon. The closed-form estimator and standard errors are:

beta? = (X'X)^-1 X'Y

Var(beta?) = sigma?2 (X'X)^-1

sigma?2 = epsilon?'epsilon? / (n - k)

t_j = beta?_j / sqrt(Var(beta?_j))

5. Pre-estimation diagnostics

Stationarity check (ADF/PP/KPSS) for time series - non-stationary regressors

invalidate inference.

VIF < 5 for all regressors to rule out multicollinearity.

Examine the dependent variable’s distribution; log-transform if heavy-tailed.

Identify and address obvious outliers via winsorising or robust regression where needed.

6. Estimation procedure

Build the design matrix X (with intercept) and response vector Y.

Solve beta? = (X’X)^-1 X’Y via QR or Cholesky decomposition.

Compute fitted values ? = Xbeta? and residuals epsilon? = Y - ?.

Estimate residual variance sigma?2 and the variance-covariance matrix.

Derive standard errors, t-statistics, and p-values for each coefficient.

Report R2, adjusted R2, F-statistic, AIC, BIC, and Durbin-Watson.

7. Output produced

8. Output interpretation

beta?_j is the expected change in Y per unit change in X_j, holding others constant.

p < 0.05 => reject H0: beta_j = 0 at 5% level. Stars: *** 1%, ** 5%, * 10%.

R2 near 1 means the model explains most variation; near 0 means little.

F-statistic with p < 0.05 indicates at least one regressor is jointly significant.

Durbin-Watson far from 2 (< 1.5 or > 2.5) signals serial correlation - switch to HAC SE.

9. Post-estimation diagnostics

Heteroskedasticity (Breusch-Pagan, White) - if rejected, use HC0-HC3 robust SE.

Autocorrelation (Durbin-Watson, Breusch-Godfrey) - if rejected, use HAC SE or ARDL.

Normality of residuals (Jarque-Bera, Shapiro-Wilk) - if rejected for small n, treat inference cautiously.

Specification (RESET) - if rejected, consider non-linear terms.

Influence (Cook’s D, leverage) - investigate top-3 influential points.

10. Common pitfalls

Spurious regression: applying OLS to non-stationary data produces inflated R2 and t-stats.

Endogeneity: if X is correlated with epsilon, beta? is biased and inconsistent - use 2SLS / GMM.

Omitted variable bias: confounders soak up explanatory power and bias coefficients.

Multicollinearity inflates SEs without biasing beta? - but makes individual t-tests unreliable.

Autocorrelation produces under-stated SEs; reported significance is misleading.

11. Reporting checklist

Sample period and N reported.

Coefficient table with beta?, SE, t, p, stars, and 95% CI.

R2, adjusted R2, F-statistic, AIC, BIC, Durbin-Watson.

All Card A diagnostics (BG, BP, JB, RESET) with verdict.

Robust SE method named (classical / HC0 / HC1 / HC3 / HAC) with reason.

Plain-language interpretation paragraph naming the dependent variable.

12. References

Greene, W. H. (2018). Econometric Analysis, 8th ed. Pearson.

Wooldridge, J. M. (2020). Introductory Econometrics, 7th ed. Cengage.

Davidson, R., MacKinnon, J. G. (2004). Econometric Theory and Methods. Oxford.

Field | Meaning

coefficients | beta?_j keyed by variable name

standardErrors | SE(beta?_j) - classical or robust depending on SE type

tStatistics | beta?_j / SE(beta?_j)

pValues | Two-sided p-value of the t-test

rSquared / adjustedRSquared | Goodness of fit

fStatistic | Joint significance of all regressors

residuals / fitted | epsilon? and ? at each observation

durbinWatson | Autocorrelation diagnostic; DW ~ 2 indicates no AR(1)

aic / bic | Information criteria for model comparison

SOP DOCUMENT: 02_FixedEffects

SOP: Fixed Effects (Within Estimator)

Panel data estimator that controls for time-invariant unobserved heterogeneity

=> Use FE when entity-specific time-invariant factors might be correlated with the regressors.

1. Purpose

FE eliminates time-invariant entity heterogeneity by within-transformation: subtracting the entity mean from each variable. Coefficients are identified from variation over time within each entity, not across entities.

2. When to use this estimator

Panel data with N entities x T periods, where T >= 3 ideally.

Suspected correlation between unobserved entity effects and regressors (Hausman test rejects RE).

Interest in within-entity dynamics rather than between-entity differences.

Country, firm, household, or region panels with sticky structural attributes.

3. Required data structure

A column identifying the entity (group ID) - this drives the within-transformation.

Optional time identifier for two-way FE.

Numeric Y and X with sufficient within-entity variation (low ICC => FE may be inefficient).

4. Mathematical formulation

The within-transformation removes entity means alpha_i, leaving the slope beta identified:

Y_it - ?_i = (X_it - X?_i)' beta + (epsilon_it - epsilon?_i)

beta?_FE = (SumSum (X_it - X?_i)(X_it - X?_i)')^-1 SumSum (X_it - X?_i)(Y_it - ?_i)

alpha?_i = ?_i - X?_i' beta?_FE

sigma?2 = epsilon?'epsilon? / (NT - N - k)

5. Pre-estimation diagnostics

Confirm the panel is balanced or near-balanced; flag unequal T per entity.

Stationarity per variable (CIPS / Fisher-ADF) since panel pooling assumes stationarity.

Cross-sectional dependence (Pesaran CD) - if rejected, use Driscoll-Kraay SE or CCE/CS-ARDL.

Inspect within R2 vs between R2 to gauge how much variation is within-entity.

6. Estimation procedure

Demean Y and each X by entity: Y_it - ?_i, X_it - X?_i.

Apply OLS to the demeaned data.

Recover entity intercepts alpha?_i = ?_i - X?_i' beta?.

Compute residuals epsilon?_it = Y_it - ?_it; standard errors clustered by entity by default.

Run pooled-vs-FE F-test (H0: all alpha_i equal) to confirm FE is needed.

7. Output produced

8. Output interpretation

beta?_j is the change in Y_it per unit change in X_jit holding entity-specific factors constant.

Significant pooled-vs-FE F-test => entity heterogeneity is real; pooled OLS would be biased.

Compare to RE via Hausman: H_p < 0.05 => FE preferred.

Entity effects alpha?_i can be plotted as a heat-map to spot structural outliers.

9. Post-estimation diagnostics

Cluster-robust SE by entity is the default; verify clustering choice matches data structure.

Heteroskedasticity (modified Wald for groupwise) - if rejected, retain clustered SE.

Serial correlation in residuals (Wooldridge AR(1) test) - if rejected, use HAC or AR(1) FE.

Cross-sectional dependence (Pesaran CD on residuals) - if rejected, prefer CCE / CS-ARDL.

10. Common pitfalls

FE wipes out time-invariant regressors - coefficients on gender, ethnicity, etc. are not identified.

If T is small relative to N, FE is consistent but inefficient; consider RE if Hausman accepts it.

Panel cointegration is not addressed by FE - non-stationarity can still produce spurious results.

FE on non-balanced panels with selection on time can introduce attrition bias.

11. Reporting checklist

Number of entities N, average T, total observations.

Within R2, between R2, overall R2.

Pooled-vs-FE F-test and Hausman test (FE vs RE).

Standard error type clearly named (clustered, Driscoll-Kraay, etc.).

Pesaran CD and slope-homogeneity diagnostics.

12. References

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data, 2nd ed. MIT.

Baltagi, B. H. (2021). Econometric Analysis of Panel Data, 6th ed. Springer.

Hsiao, C. (2014). Analysis of Panel Data, 3rd ed. Cambridge.

Field | Meaning

coefficients | Within-estimator slopes beta?_j

entityEffects | Per-entity intercept alpha?_i

rSquared | Within R2 (variation explained net of entity means)

metadata['betweenR2'] | Between-entity R2 for diagnostics

metadata['overallR2'] | Pooled R2 for completeness

metadata['pooledFTest'] | F-stat / p for pooled-vs-FE test

residuals / fitted | Within-transformed residuals

SOP DOCUMENT: 03_RandomEffects

SOP: Random Effects (RE / GLS)

Panel estimator that treats entity heterogeneity as a random component

=> Use RE when entity effects are uncorrelated with the regressors - confirmed by Hausman.

1. Purpose

RE assumes entity heterogeneity alpha_i is a random draw from a population, uncorrelated with X. The estimator is feasible GLS: a weighted combination of within and between variation, more efficient than FE when its assumptions hold.

2. When to use this estimator

Panel data where entity heterogeneity is plausibly random (sample drawn from a larger population).

Hausman test fails to reject H0: cov(alpha_i, X_it) = 0.

Need to estimate effects of time-invariant regressors (gender, sector, etc.).

Small T relative to N where FE is inefficient.

3. Required data structure

Panel data with entity ID column.

Numeric Y and X; time-invariant regressors are fine (RE retains them).

Adequate within and between variation in regressors.

4. Mathematical formulation

Quasi-demeaning with weight theta derived from variance components:

Y_it - theta ?_i = (X_it - theta X?_i)’ beta + nu_it

theta = 1 - sqrt(sigma2_epsilon / (T sigma2_alpha + sigma2_epsilon))

beta?_RE = OLS on the quasi-demeaned data

5. Pre-estimation diagnostics

Stationarity per variable (panel unit root tests).

Compute Hausman test (H_p < 0.05 => FE preferred over RE).

Breusch-Pagan LM test (H0: sigma2_alpha = 0; rejection => RE preferred over pooled OLS).

VIF < 5 across regressors.

6. Estimation procedure

Run pooled OLS to obtain sigma?2_epsilon.

Run between-entity regression on entity means to obtain sigma?2_b.

Solve for sigma?2_alpha = sigma?2_b - sigma?2_epsilon / T?.

Compute theta? and quasi-demean Y and X.

Run OLS on quasi-demeaned data; SE come from the GLS variance formula.

7. Output produced

8. Output interpretation

beta?_j is interpreted similarly to OLS: average effect on Y per unit X.

Hausman p >= 0.05: RE is consistent and more efficient than FE.

Hausman p < 0.05: RE is inconsistent; switch to FE.

BP LM rejects: heterogeneity exists; pooled OLS is inefficient.

9. Post-estimation diagnostics

Hausman test result must accompany the RE estimate.

Examine residual autocorrelation and clustering.

Robust SE: clustered by entity for valid inference.

10. Common pitfalls

RE is biased and inconsistent when alpha_i is correlated with X - always run Hausman.

Reporting RE without Hausman is a red flag for a serious reviewer.

Time-invariant regressors are kept but may absorb part of alpha_i; interpret with care.

11. Reporting checklist

Hausman test statistic, df, p-value alongside coefficients.

Breusch-Pagan LM result.

Within / between / overall R2 for comparison with FE.

Cluster-robust SE.

12. References

Wooldridge (2010). Econometric Analysis of Cross Section and Panel Data, 2nd ed.

Hausman, J. A. (1978). Specification Tests in Econometrics. Econometrica.

Breusch, T. S., Pagan, A. R. (1980). The Lagrange Multiplier Test. Review of Economic Studies.

Field | Meaning

coefficients | GLS slopes beta?_j

metadata['hausmanH'] | Hausman ?2 statistic

metadata['hausmanP'] | Hausman p-value (small => prefer FE)

metadata['bpLM'] | Breusch-Pagan LM statistic

metadata['bpP'] | BP p-value (small => prefer RE over OLS)

metadata['theta'] | Quasi-demeaning weight

residuals / fitted | Standard residuals and fitted values

SOP DOCUMENT: 04_2SLS

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 >= 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? + u => 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)^-1Z'

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 X?1.

Stage 2: regress Y on (X?1, 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

SOP DOCUMENT: 05_GMM

SOP: Generalized Method of Moments (GMM)

Moment-based estimator including Arellano-Bond and Blundell-Bond panel GMM

=> Use GMM for dynamic panels where the lagged dependent variable is a regressor.

1. Purpose

GMM minimises a quadratic form in sample moment conditions E[Z'epsilon] = 0. Difference GMM (Arellano-Bond) first-differences the model and uses lagged levels as instruments. System GMM (Blundell-Bond) augments with the equation in levels, using lagged differences as additional instruments - more efficient when persistence is high.

2. When to use this estimator

Dynamic panels: Y_it = rho Y_{i,t-1} + X_it beta + alpha_i + epsilon_it.

Endogenous regressors with no exogenous instruments outside the panel.

Short T, large N panels (typical in microeconomic applications).

3. Required data structure

Panel with N entities x T periods (T >= 3 for AB, >= 4 ideal for BB).

At least one lagged dependent variable as regressor.

No requirement for external instruments - uses internal lags.

4. Mathematical formulation

Sample moment vector and GMM criterion:

g_N(theta) = (1/N) Sum Z_i' epsilon_i(theta)

theta?_GMM = argmin_theta g_N(theta)' W g_N(theta)

Two-step optimal W = (Avar(g_N))^-1

5. Pre-estimation diagnostics

Stationarity tests on the dependent variable.

Hausman or Sargan/Hansen for instrument validity.

Sargan J test of over-identifying restrictions.

Arellano-Bond AR(1) and AR(2) tests on differenced residuals - AR(2) should not reject.

6. Estimation procedure

First-difference the equation to remove alpha_i (AB) or stack difference + level (BB).

Build instrument matrix Z from lagged levels (AB) and lagged differences (BB).

Run one-step GMM with weighting matrix W1.

Use one-step residuals to update W to optimal W2; run two-step GMM.

Compute Sargan-Hansen J and AR(1)/AR(2) tests; report Windmeijer-corrected SE.

7. Output produced

8. Output interpretation

rho? on lagged Y measures persistence; |rho?| < 1 implies stationary autoregression.

AR(2) p >= 0.10 is required for instrument validity; rejection invalidates GMM.

Sargan-Hansen J p >= 0.05 supports the instrument set.

Number of instruments should not exceed the number of entities (avoid weak-instrument problem).

9. Post-estimation diagnostics

Always report AR(1) and AR(2) test results.

Sargan-Hansen J in over-identified setting.

Windmeijer correction in two-step SE.

Discuss instrument count vs N.

10. Common pitfalls

Too many instruments overfit and weaken Hansen J.

AR(2) rejection is fatal - return to OLS-based dynamic models or reduce the lag set.

BB (system GMM) requires stationarity of initial conditions.

Two-step SE without Windmeijer correction are severely downward biased.

11. Reporting checklist

Estimator: AB or BB; one-step or two-step.

Number of entities N, periods T, instruments.

Hansen J statistic and p-value.

AR(1) and AR(2) z-statistics and p-values.

Windmeijer-corrected SE if two-step.

12. References

Arellano, M., Bond, S. (1991). Some tests of specification for panel data. Review of Economic Studies.

Blundell, R., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM. Stata Journal.

Field | Meaning

coefficients | GMM slopes including lagged Y

metadata['jStat'] / jStatP / jStatDf | Sargan-Hansen over-id test

metadata['ar1Z'] / ar1P | AR(1) test (expected to reject - first differences)

metadata['ar2Z'] / ar2P | AR(2) test (must NOT reject)

metadata['nInstr'] / N | Instrument and entity counts

metadata['abStep'] | AB / BB and one- / two-step indicator

SOP DOCUMENT: 06_FMOLS

SOP: Fully Modified OLS (FMOLS)

Phillips-Hansen cointegration estimator with semi-parametric corrections

=> Use FMOLS for cointegrated I(1) variables when you want a single long-run estimate.

1. Purpose

FMOLS corrects OLS on cointegrated variables for endogeneity and serial correlation in the long-run residuals. The result is a consistent and asymptotically unbiased estimate of the cointegrating vector with mixed-normal asymptotic distribution suitable for inference.

2. When to use this estimator

All variables are I(1) and confirmed cointegrated (Engle-Granger or Johansen).

Single cointegrating equation with one normalising variable.

Both time-series and panel data variants exist.

3. Required data structure

All Y and X variables I(1) per Stage 3 unit-root tests.

At least one cointegrating relationship confirmed at Stage 6.

Long-enough sample for kernel-based long-run variance estimation (T >= 50).

4. Mathematical formulation

Two semi-parametric corrections applied to the OLS regression in levels:

?_t* = Y_t - Omega?_yx Omega?_xx^-1 DeltaX_t (endogeneity correction)

beta?_FMOLS = (SumX_t X_t')^-1 (SumX_t Y_t* - T*gamma?) (autocorrelation correction)

Omega? = long-run variance from kernel (Bartlett / QS / Parzen)

5. Pre-estimation diagnostics

Confirm I(1) on every variable (Stage 3).

Confirm cointegration (Stage 6 - Engle-Granger / Johansen / Pedroni).

Choose kernel and bandwidth (Newey-West Bartlett or Andrews automatic).

6. Estimation procedure

Run OLS on the levels equation Y = Xbeta + u.

Estimate the long-run covariance matrix Omega? from residuals via kernel weighting.

Apply endogeneity correction: Y* = Y - Omega?_yx Omega?_xx^-1 DeltaX.

Apply autocorrelation correction term gamma?.

Compute beta?_FMOLS, mixed-normal SE, and t-statistics.

7. Output produced

8. Output interpretation

beta? is the long-run elasticity / multiplier between Y and X.

t-stats use mixed-normal critical values - same 1.96 cutoff applies.

Residuals must be I(0); rerun ADF on residuals for confirmation.

9. Post-estimation diagnostics

ADF/KPSS on residuals to confirm cointegration ex-post.

Recursive stability (CUSUM/CUSUMSQ) where T permits.

Compare to DOLS for robustness - coefficients should be similar.

10. Common pitfalls

Applying FMOLS to I(0) variables produces meaningless estimates - always check Stage 3.

Misspecified cointegration rank biases the long-run estimate.

Bandwidth too short under-estimates Omega?; too long inflates SE.

11. Reporting checklist

I(1) confirmation per variable.

Cointegration test result.

Kernel and bandwidth choice.

Coefficient table with mixed-normal SE.

Residual unit-root test on cointegrating residual.

12. References

Phillips, P. C. B., Hansen, B. E. (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. Review of Economic Studies.

Pedroni, P. (2001). Fully Modified OLS for heterogeneous cointegrated panels. Advances in Econometrics.

Field | Meaning

coefficients | Cointegrating-vector coefficients

standardErrors | Mixed-normal SE valid for inference

metadata['kernel'] / bandwidth | Long-run variance settings

residuals / fitted | Cointegrating residuals (should be I(0))

SOP DOCUMENT: 07_DOLS

SOP: Dynamic OLS (DOLS)

Stock-Watson cointegration estimator with leads and lags of DeltaX

=> DOLS is a parametric alternative to FMOLS - augment OLS with leads/lags of

differenced regressors.

1. Purpose

DOLS removes the long-run correlation between regressors and errors by augmenting the cointegrating regression with leads and lags of DeltaX. The OLS estimate on this augmented regression is asymptotically equivalent to FMOLS but parametrically simpler.

2. When to use this estimator

Cointegrated I(1) variables, single long-run relationship.

Smaller samples where FMOLS kernel choice is sensitive.

Robustness check against FMOLS.

3. Required data structure

All Y and X I(1); cointegration confirmed.

Sample large enough to spare 2(p_lead + p_lag) observations to the augmentation block.

4. Mathematical formulation

Augmented regression with q leads and p lags of DeltaX:

Y_t = alpha + beta'X_t + Sum_{j=-p}^{q} gamma_j' DeltaX_{t-j} + epsilon_t

beta?_DOLS = OLS slopes on X_t

Newey-West HAC SE for inference

5. Pre-estimation diagnostics

I(1) confirmation per variable.

Cointegration test.

Choose lead/lag order via AIC/BIC across a small grid.

6. Estimation procedure

Form DeltaX leads (j=1..q) and lags (j=1..p).

Augment the cointegrating regression with these leads/lags.

Run OLS and extract beta? on contemporaneous X.

Newey-West HAC SE with appropriate bandwidth.

7. Output produced

8. Output interpretation

beta? is the long-run elasticity (same interpretation as FMOLS).

Compare to FMOLS - large discrepancy suggests misspecification.

9. Post-estimation diagnostics

Residual unit-root test.

CUSUM / CUSUMSQ stability.

Compare to FMOLS estimates.

10. Common pitfalls

Too many leads/lags consume degrees of freedom; pick parsimoniously.

Sensitive to outliers in DeltaX leads/lags.

Same I(1)+cointegration prerequisites as FMOLS.

11. Reporting checklist

Lead and lag orders chosen and selection criterion.

HAC bandwidth.

FMOLS comparison column.

12. References

Stock, J. H., Watson, M. W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. Econometrica.

Field | Meaning

coefficients | Long-run coefficients

standardErrors | HAC SE

metadata['leads'] / lags | Augmentation orders chosen

residuals / fitted | Cointegrating residuals

SOP DOCUMENT: 08_ARDL

SOP: Autoregressive Distributed Lag (ARDL)

Pesaran-Shin-Smith bounds testing approach

=> ARDL handles mixed I(0)/I(1) regressors and yields long-run + short-run dynamics from a single equation.

1. Purpose

ARDL embeds the dependent variable's lags and the regressors' lags in a single equation. Re-parameterising into the unrestricted error correction model (UECM) gives long-run multipliers and the speed of adjustment to equilibrium. The bounds test (Pesaran, Shin, Smith 2001) tests for cointegration without pre-determining the integration order.

2. When to use this estimator

Time-series with mixed I(0)/I(1) regressors (none I(2)).

Need both long-run elasticities and short-run dynamics.

Smaller samples (T >= 30) where Johansen is unreliable.

Single dependent variable framework.

3. Required data structure

Time series with no I(2) variables (run unit-root tests pre-estimation).

T sufficient for AIC-selected lag orders (T >= 30 typical).

Optional dummies for structural breaks identified by Zivot-Andrews.

4. Mathematical formulation

ARDL(p, q1, ?, q_k) and its UECM re-parameterisation:

Y_t = alpha + Sum_{i=1}^{p} ?_i Y_{t-i} + Sum_{j=1}^{k} Sum_{l=0}^{q_j} beta_{jl} X_{j,t-l} + epsilon_t

DeltaY_t = alpha + phi Y_{t-1} + Sum theta_j X_{j,t-1} + Sum ?_i* DeltaY_{t-i} + Sum ?_{jl} DeltaX_{j,t-l} + epsilon_t

Long-run multiplier: lambda_j = - theta_j / phi

ECT_{t-1} = Y_{t-1} - Sum lambda_j X_{j,t-1}

5. Pre-estimation diagnostics

Confirm no variable is I(2) - bounds test invalid otherwise.

Stationarity tests (ADF, PP, KPSS, DF-GLS, Ng-Perron).

Zivot-Andrews to detect structural breaks.

VIF for multicollinearity diagnosis.

6. Estimation procedure

Search ARDL(p, q1, ?, q_k) lag combinations up to maxLag.

Select the AIC-minimising combination.

Estimate the UECM by OLS and extract phi? and theta?.

Compute long-run multipliers via the delta method for SE.

Run the bounds F-test (k+1 restrictions on UECM).

Optional CUSUM / CUSUMSQ for parameter stability.

7. Output produced

8. Output interpretation

phi? < 0 and significant => stable error correction; |phi?| is the fraction of disequilibrium corrected per period.

Half-life of disequilibrium: ln(0.5) / ln(1 + phi).

lambda_j is the long-run elasticity of Y w.r.t. X_j.

Bounds F above the upper bound => cointegration; below the lower bound => no cointegration.

All Card A diagnostics should pass for inference to be valid.

9. Post-estimation diagnostics

Bounds test F and t (Pesaran-Shin-Smith critical values).

Breusch-Godfrey for residual autocorrelation.

Breusch-Pagan / White for heteroskedasticity.

Jarque-Bera for residual normality.

RESET for functional form.

CUSUM / CUSUMSQ for parameter stability.

10. Common pitfalls

If any variable is I(2), the bounds test is invalid.

Positive phi? indicates explosive dynamics - model is misspecified.

AIC vs BIC can pick different lags; document the choice.

Do not interpret long-run multipliers without confirming bounds-test cointegration.

11. Reporting checklist

Selected ARDL order [p, q1, ?, q_k] and selection criterion.

ECT coefficient with stars and half-life.

Long-run coefficient table with delta-method SE.

Short-run RECM table.

Bounds test F and t with PSS critical values.

Card A diagnostics with verdict.

CUSUM / CUSUMSQ stability comments.

12. References

Pesaran, M. H., Shin, Y., Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics.

Pesaran, M. H., Shin, Y. (1999). An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis.

Field | Meaning

ardlOrder | [p, q1, ?, q_k] selected by AIC

ect | phi? - speed of adjustment (must be negative)

ectSE / ectTStat / ectPValue | ECT inference

longRunCoefficients / longRunSE | lambda_j and SE via delta method

shortRunCoefficients | Delta-form coefficients keyed Delta(X), Delta(L(X,1)), ?

boundsTest | F-stat, t-stat, conclusion vs PSS critical values

rSquared / aic / bic | Goodness of fit

diagnostics | Card A: BG, BP, JB, RESET, CUSUM, CUSUMSQ

SOP DOCUMENT: 09_NARDL

SOP: Nonlinear ARDL (NARDL)

Shin-Yu-Greenwood-Nimmo asymmetric ARDL

=> NARDL splits a regressor into positive and negative cumulative shocks to test asymmetric effects.

1. Purpose

NARDL replaces symmetric X with X? (sum of positive changes) and X^- (sum of negative changes) to test whether positive and negative shocks have different long-run and short-run effects. Wald tests for asymmetry give a formal statistical decision.

2. When to use this estimator

Theory predicts asymmetric responses (sticky prices, downside risk aversion, etc.).

Visual evidence of asymmetric reaction in residuals.

Same I(0)/I(1) data conditions as ARDL.

3. Required data structure

Same prerequisites as ARDL (no I(2)).

Identify which regressors to decompose into ?/^- asymmetries.

4. Mathematical formulation

Partial-sum decomposition of X plus standard ARDL UECM:

X_t? = Sum_{s<=t} max(DeltaX_s, 0)

X_t^- = Sum_{s<=t} min(DeltaX_s, 0)

DeltaY_t = alpha + phi Y_{t-1} + theta? X_{t-1}? + theta^- X_{t-1}^- + Sum ? DeltaY_{t-i} + Sum (?? DeltaX_{t-l}? + ?^- DeltaX_{t-l}^-) + epsilon_t

Long-run: L? = - theta? / phi; L^- = - theta^- / phi

Wald LR: H0: L? = L^-; Wald SR: H0: Sum?? = Sum?^-

5. Pre-estimation diagnostics

All ARDL pre-tests apply.

Decide which X to decompose; usually theoretical priors guide this.

6. Estimation procedure

Construct X? and X^- partial-sum series for each asymmetric regressor.

Estimate UECM with X? and X^- in place of X.

Compute long-run L? and L^- with delta-method SE.

Compute dynamic multipliers per horizon for both regimes.

Run Wald asymmetry tests (long-run, short-run).

Run bounds test on the extended set (k_extended = k + #asymmetric).

7. Output produced

8. Output interpretation

Wald LR p < 0.05 => asymmetric long-run response confirmed.

Wald SR p < 0.05 => asymmetric short-run response confirmed.

Dynamic multiplier chart shows the time profile of L? and L^-.

If both Wald tests fail to reject, ARDL (symmetric) is sufficient.

9. Post-estimation diagnostics

Bounds test on extended k.

Card A diagnostics.

CUSUM / CUSUMSQ stability.

Plot dynamic multipliers with CI bands.

10. Common pitfalls

Decomposing too many regressors thins degrees of freedom.

Asymmetry conclusions depend on bounds-test cointegration - confirm first.

Interpret L? / L^- only when phi? is significantly negative.

11. Reporting checklist

Which regressors are decomposed and theoretical motivation.

L? and L^- with stars.

Wald LR / SR asymmetry tests.

Dynamic multiplier chart.

Bounds test F and t.

Card A diagnostics.

12. References

Shin, Y., Yu, B., Greenwood-Nimmo, M. (2014). Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework.

Field | Meaning

ardlOrder | Selected lag structure

ect | phi?

lrPos / lrNeg / lrPosSE / lrNegSE | Asymmetric long-run multipliers

waldLRStat / waldLRPValue | Long-run asymmetry test

waldSRStat / waldSRPValue | Short-run asymmetry test

isAsymmetric | True when waldLRPValue < 0.05

dynamicMultipliersPos / dynamicMultipliersNeg | Per-horizon multipliers

dynamicMultipliersPosCI / dynamicMultipliersNegCI | Bootstrap CIs

uecmCoefficients / SE / TStats / PValues | Full UECM table

SOP DOCUMENT: 10_PMG

SOP: Pooled Mean Group (PMG)

Pesaran-Shin-Smith heterogeneous panel ARDL with common long-run

=> PMG pools the long-run coefficients across countries while letting short-run dynamics differ.

1. Purpose

PMG estimates a panel ARDL where the long-run coefficients are constrained to be equal across countries (consistent with theory) while the speed of adjustment, intercepts, and short-run dynamics are allowed to differ. Maximum likelihood is used because the long-run constraint is non-linear.

2. When to use this estimator

Panel ARDL where theory says LR effects are common but SR adjustment differs.

Hausman test fails to reject pooling of LR coefficients (PMG vs MG).

T sufficient per country (typically T >= 20).

3. Required data structure

Balanced or near-balanced panel.

Per-country lag orders selectable; pooled lag default supported.

Same I(0)/I(1) conditions as ARDL.

4. Mathematical formulation

Per-country UECM with pooled long-run vector:

DeltaY_{i,t} = alpha_i + phi_i (Y_{i,t-1} - lambda' X_{i,t-1}) + Sum ?_i DeltaY_{i,t-j} + Sum ?_{i,jl} DeltaX_{i,j,t-l} + epsilon_{i,t}

lambda same across countries; phi_i, ?_i, ?_{i,jl} country-specific

Estimated by Newton-Raphson MLE; convergence tolerance 1e-5.

5. Pre-estimation diagnostics

Stationarity, slope homogeneity (Pesaran-Yamagata), CSD diagnostics.

Determine maxP and maxQ for the panel lag search (typical maxP=1, maxQ=3).

Decide deterministic case (Case 2 / 3 / 4 per PSS).

6. Estimation procedure

Per-country: select ARDL order by AIC up to maxLag.

Stack per-country UECMs imposing common lambda.

Newton-Raphson MLE on the concentrated likelihood.

Iterate until |Deltalog L| < 1e-5 or 3000 iterations.

Compute long-run SE via the information matrix.

Hausman MG-vs-PMG to validate the pooling restriction.

7. Output produced

8. Output interpretation

lambda? is the long-run elasticity common to all countries.

Per-country phi?_i allows heterogeneous speeds of return to equilibrium.

Hausman p >= 0.05 => PMG pooling is consistent and more efficient than MG.

Hausman p < 0.05 => heterogeneity in lambda; switch to MG.

Variability across phi?_i highlights structural differences in adjustment.

9. Post-estimation diagnostics

Hausman MG-vs-PMG.

Per-country diagnostics where T permits.

Cross-section dependence check on residuals.

10. Common pitfalls

Imposing pooled lambda when heterogeneity exists biases the estimate severely.

Newton-Raphson can fail to converge with poor starting values.

Per-country lag selection: ensure each country has T > 2(p+q) observations.

11. Reporting checklist

Per-country selected ARDL orders.

Pooled long-run table with stars.

Mean ECT and per-country ECT distribution.

Hausman MG-vs-PMG.

Convergence and iteration count.

12. References

Pesaran, M. H., Shin, Y., Smith, R. P. (1999). Pooled Mean Group estimation of dynamic heterogeneous panels.

Pesaran, M. H., Smith, R. P. (1995). Estimating long-run relationships from dynamic heterogeneous panels.

Field | Meaning

longRunCoefficients | Pooled lambda? across countries

longRunSE / longRunTStats / longRunPValues | LR inference

ect | Mean of phi?_i across countries

shortRunCoefficients | Average per-country ?? at lag 0

unitResults | Per-country ARDL fit (lag order, ECT, RECM)

speedOfAdjustmentTable | Per-country half-life and adjustment %

converged | Convergence flag

SOP DOCUMENT: 11_MG

SOP: Mean Group (MG)

Pesaran-Smith heterogeneous panel ARDL with no pooling

=> MG runs ARDL country-by-country, then averages the long-run coefficients.

1. Purpose

MG estimator imposes no homogeneity. Each country's ARDL is estimated independently, then the long-run coefficient vector is the simple cross-sectional average. Standard error uses the variance of the country-specific estimates.

2. When to use this estimator

Slope heterogeneity confirmed (Pesaran-Yamagata p < 0.05).

Hausman MG-vs-PMG rejects PMG pooling.

T per country adequate (T >= 25 ideally).

3. Required data structure

Same as PMG; needs adequate per-country T.

4. Mathematical formulation

Per-country ARDL average:

lambda?_i estimated per country by ARDL.

lambda?_MG = (1/N) Sum lambda?_i

SE(lambda?_MG) = sqrt((1/(N(N-1))) Sum (lambda?_i - lambda?_MG)2)

5. Pre-estimation diagnostics

Slope homogeneity test (rejection motivates MG).

Per-country ARDL pre-checks.

6. Estimation procedure

Per-country ARDL with AIC lag selection.

Average the long-run coefficient vectors across countries.

Cross-sectional SE.

7. Output produced

8. Output interpretation

lambda?_MG is robust to slope heterogeneity but has wider SE than PMG.

Compare with PMG via Hausman; if pooling is rejected, MG is the consistent choice.

9. Post-estimation diagnostics

Per-country diagnostics where possible.

Cross-section dependence (Pesaran CD).

10. Common pitfalls

Sensitive to outlier countries - inspect per-country lambda?_i distribution.

T must be large enough per country; otherwise individual ARDLs are unreliable.

11. Reporting checklist

Per-country lambda?_i histogram or table.

MG average with cross-sectional SE.

Hausman vs PMG.

12. References

Pesaran, M. H., Smith, R. P. (1995). Estimating long-run relationships from dynamic heterogeneous panels.

Field | Meaning

longRunCoefficients | Cross-country average of lambda?_i

longRunSE | Cross-sectional SE

unitResults | Per-country ARDL output

ect | Average of country-specific phi?

SOP DOCUMENT: 12_DFE

SOP: Dynamic Fixed Effects (DFE)

Pooled panel ARDL with country-specific intercepts

=> DFE pools every coefficient (LR and SR) but allows country-specific intercepts.

1. Purpose

DFE imposes full homogeneity of slope and adjustment coefficients across countries while allowing fixed country intercepts. It is the most restrictive of the panel ARDL family.

2. When to use this estimator

Strong theoretical reason to pool every coefficient.

Slope homogeneity not rejected.

Small N (where MG / PMG cross-sectional averages are noisy).

3. Required data structure

Balanced panel.

Same I(0)/I(1) conditions as ARDL.

4. Mathematical formulation

Pooled UECM with within-transformation:

DeltaY_{i,t} = alpha_i + phi Y_{i,t-1} + theta' X_{i,t-1} + Sum ? DeltaY_{i,t-j} + Sum ?' DeltaX_{i,j,t-l} + epsilon_{i,t}

Within-demean to remove alpha_i, then pooled OLS.

Driscoll-Kraay SE for cross-sectional dependence.

5. Pre-estimation diagnostics

Slope homogeneity test.

Pesaran CD on the pooled residuals.

6. Estimation procedure

Within-demean each variable per country.

Pooled OLS on demeaned data.

Driscoll-Kraay SE with bandwidth = floor(T^{1/3}).

7. Output produced

8. Output interpretation

Same single-equation interpretation as ARDL but applied to a panel mean.

Heterogeneity rejection => DFE biases all coefficients; switch to MG / PMG.

9. Post-estimation diagnostics

CSD on residuals.

Slope homogeneity ex-post check.

10. Common pitfalls

DFE is strongly biased when slopes are heterogeneous (Pesaran-Smith critique).

Driscoll-Kraay SE require T moderately large.

11. Reporting checklist

Slope homogeneity test verdict.

Pooled long-run table with Driscoll-Kraay SE.

Pooled ECT.

Comparison to MG / PMG.

12. References

Pesaran, M. H., Smith, R. P. (1995). Estimating long-run relationships from dynamic heterogeneous panels.

Driscoll, J. C., Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data.

Field | Meaning

longRunCoefficients | Pooled lambda? from -theta?/phi?

ect | Pooled phi?

shortRunCoefficients | Pooled DeltaX coefficients

standardErrors | Driscoll-Kraay

SOP DOCUMENT: 13_CSARDL

SOP: Cross-Sectionally Augmented ARDL (CS-ARDL)

Chudik-Pesaran heterogeneous panel ARDL with CSD correction

=> CS-ARDL augments each country's ARDL with cross-sectional means to absorb common factors.

1. Purpose

CS-ARDL adds the cross-sectional means of Y and X (and their lags) to each country's ARDL specification. This proxies for unobserved common factors driving cross-sectional dependence, restoring consistent estimation of the long-run coefficients.

2. When to use this estimator

Panel ARDL with confirmed CSD (Pesaran CD $p < 0.05$).

Slope heterogeneity present.

Large N and moderate T.

3. Required data structure

Balanced panel preferred.

Same I(0)/I(1) conditions as ARDL.

4. Mathematical formulation

Per-country UECM augmented with cross-sectional means:

DeltaY_{i,t} = alpha_i + phi_i (Y_{i,t-1} - lambda_i' X_{i,t-1}) + Sum ? DeltaX + Sum mu Z?_{t-l} + epsilon_{i,t}

Z?_{t-l} = cross-sectional means of (Y, X) at lag l

Per-country estimation; long-run = cross-country average.

5. Pre-estimation diagnostics

Pesaran CD => rejection motivates CS-ARDL.

Slope homogeneity test.

Per-country ARDL pre-checks.

6. Estimation procedure

Compute cross-sectional means of Y and X at each t.

Per-country ARDL with cross-sectional means and lags appended.

Cross-country average of lambda?_i.

Floor pT = floor(T^{1/3}) lags of cross-section means.

7. Output produced

8. Output interpretation

Long-run estimate consistent under CSD plus heterogeneity.

CS-mean coefficients absorb the influence of unobserved common factors.

9. Post-estimation diagnostics

Pesaran CD on residuals (should be eliminated).

Slope homogeneity ex-post.

10. Common pitfalls

Adding too many lags of cross-section means depletes degrees of freedom.

Strong CSD with small N can leave residual CSD.

11. Reporting checklist

Pesaran CD before and after CS augmentation.

Per-country selected lag orders.

Long-run table with cross-sectional SE.

Cross-section mean coefficient table.

12. References

Chudik, A., Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors.

Field | Meaning

longRunCoefficients | Average per-country lambda?

longRunSE | Cross-sectional SE

csMeanCoefficients | Cross-sectional mean coefficients (mu?)

unitResults | Per-country fit including CS columns

ect | Average per-country phi?

SOP DOCUMENT: 14_VAR

SOP: Vector Autoregression (VAR)

Multivariate time-series model with all variables endogenous

=> VAR treats every variable as endogenous; impulse responses and FEVD reveal system dynamics.

1. Purpose

VAR(p) regresses each of k variables on p lags of itself and all other variables. The reduced-form residuals are orthogonalised via Cholesky decomposition for structural interpretation. Impulse responses (IRF) trace shock propagation; forecast error variance decomposition (FEVD) attributes the share of forecast variance to each shock.

2. When to use this estimator

Multiple stationary variables with bidirectional causality.

Forecasting and structural analysis without imposing strict theoretical priors.

Identifying lead-lag relationships via Granger causality.

3. Required data structure

k stationary time series (apply unit-root tests first).

T sufficient for k2 x p parameter estimation (T ? k2 x p).

If variables are I(1) and cointegrated, switch to VECM.

4. Mathematical formulation

Reduced-form VAR(p):

Y_t = c + Sum_{i=1}^{p} A_i Y_{t-i} + u_t

Var(u_t) = Sum_u

Cholesky: Sum_u = L L' => structural shocks epsilon_t = L^-1 u_t

IRF_h = J A^h J' L, J = [I_k 0 ? 0] (companion-matrix form)

FEVD_h: share of forecast variance of variable i from shock j at horizon h.

5. Pre-estimation diagnostics

Stationarity per series.

Lag selection: minimise AIC / BIC / HQ over candidates.

Granger causality tests pre-IRF.

6. Estimation procedure

Build the design with p lags of each Y series.

Estimate each equation by OLS (efficient by GLS-equivalence).

Build companion matrix; verify spectral radius < 1 (stability).

Cholesky factorise Sum_u; compute IRF and FEVD over horizon H.

Bootstrap CIs for IRF (residual resampling).

Run Granger causality tests by F-statistic.

7. Output produced

8. Output interpretation

IRF: response of variable i to a one-standard-deviation shock in equation j over horizons.

FEVD at horizon h: percentage of variable i's forecast error variance attributable to shock j.

Granger causality p < 0.05 => X helps predict Y beyond Y's own past.

Spectral radius >= 1 => VAR is unstable; revisit lag order or use VECM.

9. Post-estimation diagnostics

Stability: spectral radius of companion matrix.

Residual autocorrelation: portmanteau (Ljung-Box) on residuals.

Normality: multivariate Jarque-Bera on residuals.

ARCH-LM for residual heteroskedasticity.

10. Common pitfalls

Cholesky ordering matters - reorder via theory; check robustness to alternative orderings.

Unit roots in any variable invalidate the reduced-form VAR - use VECM instead.

Too many lags overfit; AIC tends to over-pick relative to BIC.

11. Reporting checklist

Lag order with selection criterion.

Cholesky ordering and theoretical justification.

IRF panels with bootstrap CI.

FEVD panels with horizon table.

Granger causality matrix.

Stability diagnostic.

12. References

Sims, C. A. (1980). Macroeconomics and Reality. Econometrica.

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

Field | Meaning

varIRF | [H+1][shock][response] orthogonalised IRFs

varIRFLower / varIRFUpper | Bootstrap CIs

varFEVD | [H+1][variable][shock] variance decomposition

companionSpectralRadius | Stability check (< 1)

grangerTests | Per-pair F-statistic p-values

variableOrder | Cholesky ordering used

SOP DOCUMENT: 15_VECM

SOP: Vector Error Correction Model (VECM)

VAR for cointegrated I(1) systems

=> Use VECM when k>=2 I(1) variables are cointegrated - the levels VAR is non-stationary, the VECM is.

1. Purpose

VECM expresses a cointegrated VAR in error-correction form. The Johansen procedure identifies the cointegrating rank r ? [1, k-1] and estimates the cointegrating vectors beta and adjustment matrix alpha. Short-run dynamics are captured by lagged Delta-terms.

2. When to use this estimator

All variables I(1) with at least one cointegrating relationship (Johansen rank >= 1).

Multivariate framework where VAR in levels would be misspecified.

3. Required data structure

All k series I(1) per Stage 3.

Johansen test confirms r >= 1 cointegrating vectors.

T ? 2 x k x p typical.

4. Mathematical formulation

DeltaY_t = alphabeta' Y_{t-1} + Sum Gamma_i DeltaY_{t-i} + epsilon_t

beta: k x r matrix of cointegrating vectors (long-run equilibria).

alpha: k x r adjustment matrix; (alphabeta') is rank-r.

Gamma_i: short-run k x k coefficient matrices on lagged differences.

Reduced-rank regression solves alpha and beta simultaneously (Johansen).

5. Pre-estimation diagnostics

All variables I(1).

Johansen trace and max-eigenvalue tests for rank r.

Lag selection in the underlying VAR (AIC / BIC / HQ).

6. Estimation procedure

Determine VAR lag order p in levels.

Run Johansen reduced-rank regression to get r.

Estimate DeltaY_t = alphabeta' Y_{t-1} + Sum Gamma DeltaY_{t-i} + epsilon_t with rank r.

Reconstruct levels VAR for Cholesky IRF / FEVD.

Run residual diagnostics (autocorrelation, normality, stability).

7. Output produced

8. Output interpretation

beta'Y_{t-1} are the long-run equilibrium errors; alpha is the adjustment

toward them.

Sign and magnitude of alpha reveals which variables bear the burden of correction.

IRF / FEVD interpreted same as VAR after the levels reconstruction.

9. Post-estimation diagnostics

Stability (spectral radius < 1 after rank reconstruction).

Residual autocorrelation, normality, ARCH-LM.

Robustness across rank choice.

10. Common pitfalls

Mis-specifying the rank produces inconsistent alpha and beta.

Mixing I(0) and I(1) variables - drop I(0) ones or transform first.

Cholesky ordering still matters for IRF interpretation post-reconstruction.

11. Reporting checklist

Johansen trace / max-eig test results and chosen rank.

Cointegrating vectors beta (normalised).

Adjustment matrix alpha with significance.

IRF / FEVD panels.

Stability and residual diagnostics.

12. References

Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford.

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

Field | Meaning

metadata['cointegratingRank'] | Johansen-selected r

metadata['cointegratingVectors'] | beta as flattened text block

metadata['adjustmentMatrix'] | alpha

varIRF / varFEVD | Computed from the implied levels VAR

companionSpectralRadius | Stability

residuals | VECM residuals