03_RandomEffects

QuantFit Estimator Standard Operating Procedure

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