

# Leave-cluster-out and variance estimation

## Abstract

We introduce the leave-cluster-out (LCO) machinery for clustered samples, a generalization of leave-one-out methods that prove useful for independent data. We use LCO to construct an estimator of the asymptotic variance of the OLS estimator in a linear regression characterized by possibly numerous regressors and arbitrary within-cluster heteroskedasticity. We show consistency of the LCO variance estimator when regressors may be many, regression errors may be heteroskedastic, clusters may be unbalanced and heterogeneous, and cluster sizes may be moderately large. Simulations reveal amazing robustness of the LCO estimator to regressor numerosity and heteroskedasticity.

## Keywords

Linear regression, ordinary least squares, heteroskedasticity, many regressors, leave-out estimation, variance estimation.