Leave-Out Clustered Standard Errors

Raffaele Saggio

October 9, 2023

This document explains the MATLAB function KSS_SE, which computes the leave-out (or "cross-fit") clustered SEs proposed by Kline et al. (2020)—KSS henceforth—for linear regression models.

1 Introduction

Consider a linear model of the form

$$y_{ij} = x_{ij}'\beta + \varepsilon_{ij} \qquad j = 1, ..., J; \qquad i = 1, ...n_j;$$

where i indexes a particular observation belonging to a cluster j and we have $N = \sum_j M_j$ total observations; x_{ij} is a vector of regressors of dimension $K \times 1$ and y_{ij} is the outcome of interest. The error terms, ε_{ij}, are potentially heteroskedastic and correlated across observations within the same cluster j, with a block-diagonal variance-covariance matrix given by

$$\Omega = \begin{bmatrix} \mathbf{\Omega}_1 & 0 & 0 & 0 \\ 0 & \mathbf{\Omega}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{\Omega}_J \end{bmatrix}$$

The variance of the OLS estimator of β, $\hat{\beta}$, is given by

$$\mathbb{V}[\hat{\beta}] = \left(\sum_{i,j}^{N} x_{ij}x_{ij}'\right)^{-1} \left[\sum_{j=1}^{J} \mathbf{x}_j'\mathbf{\Omega}_j\mathbf{x}_j\right] \left(\sum_{i,j}^{N} x_{ij}x_{ij}'\right)^{-1},$$

where $\mathbf{x}_j$ is a $n_j \times K$ matrix that stacks the regressors x_{ij} for the observations belonging to cluster j.

Estimates of $\mathbb{V}[\hat{\beta}]$ in most software packages (e.g., reghdfe) are given by

$$\tilde{\mathbb{V}}[\hat{\beta}] = d\left(\sum_{i,j}^{N} x_{ij}x_{ij}'\right)^{-1} \left[\sum_{j=1}^{J} \left(\sum_{i=1}^{n_j} x_{ij}\hat{e}_{ij}\right)\left(\sum_{i=1}^{n_j} x_{ij}\hat{e}_{ij}\right)'\right] \left(\sum_{i,j}^{N} x_{ij}x_{ij}'\right)^{-1},$$

where d is some degrees of freedom adjustment and $\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta}$ is the OLS residual.[1]

[1] For instance, reghdfe sets $d = \frac{J}{J-1}\frac{N}{N-K}$.

KSS introduced an unbiased variance estimator of the variance of $\mathbf{\Omega}_j$ given by

$$\hat{\mathbf{\Omega}}_j = \mathbf{y}_j(\mathbf{y}_j - \mathbf{x}_j\hat{\beta}_{-j})',$$

where $\hat{\beta}_{-j}$ is the OLS estimate of β obtained after fitting (1) leaving cluster j out; $\mathbf{y}_j$ is a $n_j \times 1$ vector that stacks the outcome variable y_{ij} for observations belonging in cluster j.[2] Note that this residual involves a cross-product rather than a square of the sort that arises in the HC2 or HC3 variance estimators studied by MacKinnon and White (1985).

Let $\hat{\eta}_{ij}$, represent the leave-cluster out residual, i.e. $\hat{\eta}_{ij} \equiv y_{ij} - x_{ij}\hat{\beta}_{-j}$. KSS showed that the following estimator of $\mathbb{V}[\hat{\beta}]$ is unbiased:

$$\hat{\mathbb{V}}[\hat{\beta}] = \left(\sum_{i,j}^{N} x_{ij}x_{ij}'\right)^{-1} \left[\sum_{j=1}^{J}\left(\sum_{i=1}^{n_j} x_{ij}y_{ij}\right)\left(\sum_{i=1}^{n_j} x_{ij}\hat{\eta}_{ij}\right)'\right]\left(\sum_{i,j}^{N} x_{ij}x_{ij}'\right)^{-1}.$$

An important consequence of unbiasedness is that some of the diagonal entries of $\hat{\mathbb{V}}[\hat{\beta}]$ may turn out to be negative. Negative variances are unlikely to arise in low dimensional models where interest centers on only a few parameters (e.g., event-study coefficients). In models with thousands or hundreds of thousands of coefficients, it becomes likely that at least one of the estimated variances will be negative. However, in settings where thousands of coefficients are being estimated, interest rarely centers on any particular coefficient, but rather linear combinations or quadratic functions of them (e.g. the variance of firm effects in a two-way worker-firm fixed effects model a la Abowd et al. (1999)). Estimation of linear and quadratic functions of this nature can be handled by the lincom and LeaveOutTwoWay routines respectively.

The software described in this vignette, `KSS_SE`, computes $\hat{\mathbb{V}}[\hat{\beta}]$. The readme in this repository illustrates how this function works in an example where one is interested in estimating an event-study design using a two-way fixed effects regression.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Kline, P., R. Saggio, and M. Sølvsten (2020). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

MacKinnon, J. G. and H. White (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics 29*(3), 305–325.

[2] To improve the finite sample performance of the estimator, we demean $\mathbf{y}_j$ before applying this formula. That is, we compute $\hat{\mathbf{\Omega}}_j = (\mathbf{y}_j - \bar{y})(\mathbf{y}_j - \mathbf{x}_j\hat{\beta}_{-j})'$, where $\bar{y} = \frac{1}{N}\sum_{i,j} y_{ij}$. This modification substantially reduces the variability of each $\hat{\mathbf{\Omega}}_j$ at the cost of introducing a negligible amount of bias.