Bent E. Sørensen

Introduction to Clustered Standard Errors

Assume that you want to estimate the model

$$y_i = X_i\beta + u_i \,,$$

where $\text{var}(u_i) = \sigma_i^2$, y_i is a scalar, and X_i is $1 \times k$ vector containing k regressors. In vector form, the model is

$$y = X\beta + u \,,$$

where X is an $n \times k$ matrix and y and u are $n-$vectors.

The i index can stand for, say, individuals, time, or individuals and time (i.e. panel) and the problem is one of heteroskedasticity. If you have a model for the heteroskedasticity, you can deal with as taught in elementary econometrics and do GLS, or you can can control for it when calculating standard errors. In the clustering literature, the latter is always done. Recall that if the variance of the error vectore is Ω the variance of the estimator $\hat{\beta}$ is

$$(X'X)^{-1}X'\Omega X(X'X)^{-1}\,.$$

where Ω is $Var(u) = Euu'$. Clustering is the case where Ω is not diagonal, but observations in certain subgroups are correlated. That can be captured using a random effect (or fixed effect) model, but in the modern cluster literature, it is typically assumed that the error terms correlated with the regressors so the challenge is to estimate

$$E\{X'uu'X\}\,.$$

We need to rely on LLNs, so notice that we can write

$$X = \begin{pmatrix} x_1^1 & \dots & x_1^k \\ \vdots & \vdots & \vdots \\ x_N^1 & \dots & x_N^k \end{pmatrix},$$

or in partitioned form

$$X = \begin{pmatrix} X_1' \\ \vdots \\ X_i' \\ \vdots \\ X_N' \end{pmatrix},$$

where X_i is the vector of the k regressors for observation i.

Now. we have

$$X'uu'X = (X_1, ..., X_N) \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} (u_1, ..., u_N) \begin{pmatrix} X_1' \\ \vdots \\ X_i' \\ \vdots \\ X_N' \end{pmatrix}$$

or

$$X'uu'X = (X_1, ..., X_N) \begin{pmatrix} u_1^2 & \dots & u_1 u_N \\ \vdots & \dots & \vdots \\ u_N u_1 & \dots & u_N^2 \end{pmatrix} \begin{pmatrix} X_1' \\ \vdots \\ X_i' \\ \vdots \\ X_N' \end{pmatrix}$$

In the case of standard heteroskedasticity, the off-diagonal elements of the uu' matrix has mean zero and if we set those to zero and multiply, we have

$$X'uu'X = \Sigma_i X_i X_i' u_i^2 \,,$$

and White (and some before him) realized that (unless the Xs go crazy, which we assume they do not) this term satisfies a LLN and converges to, say, Σ, after dividing by N. So, using regression residuals for the true errors, we can estimate the variance of the estimator $\hat{\beta}$ is

$$(X'X)^{-1} \hat{\Sigma} \, (X'X)^{-1} \,,$$

where $\hat{\Sigma}$ is the finite sample estimate. (Note that the asymptotic variance of $\sqrt{N}\hat{\beta}$ is the limit of $(\frac{1}{N}X'X)^{-1}\frac{1}{N}X'uu'X(\frac{1}{N}X'X)^{-1}$, where each term now is normalized to satisfy a LLN). The beauty of this is that we can have heteroskedasticity of unknown form and variances and covariance can be correlated with the Xs as long as everything stays bounded enough that the LLN applies.

The same math, almost, can be done if there is correlation between the residuals within groups $g = 1..., G$ of size m, say. Assume with not loss of generality that the data are ordered so we first have group one, then group two etc. And collect the error terms in vectors $U_1, ..., U_G$ where

$$U_1 = (u_1, ...u_m)' .$$

we can then use partitioned matrix algebra and write

$$uu' = \begin{pmatrix} U_1 U_1' & \dots & U_1 U_G' \\ \vdots & \dots & \vdots \\ U_G U_1' & \dots & U_N U_N' \end{pmatrix} .$$

and

$$Euu' = \begin{pmatrix} U_1 U_1' & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & U_N U_N' \end{pmatrix} .$$

Partion X the same way, so that

$$X = \begin{pmatrix} X^{1'} \\ \vdots \\ X^{G'} \end{pmatrix} ,$$

where $X^{1'}$ now is the k regressors for observations $1, ..., G$ in a $k \times G$ matrix. Then, disregarding the mean zero terms, we have

$$X'uu'X = \Sigma_g \, X^g \, U_g U_g' \, X^{g'} \, ,$$

which for the number of clusters, G, satisfies an LLN when divided by G. The cluster-robust standard error is then

$$(X'X)^{-1} (\Sigma_g \, X^g \, U_g U_g' \, X^{g'}) \, (X'X)^{-1} \, .$$

In practise, you would of course use residuals for the error terms.