Economics 312 — Spring 2020
Daily Problem #36 — April 20

Suppose that $y_{t,R}$ is the "yield rate" at Reed College in year t, the share of admitted students who choose to attend, and that $y_{t,LC}$ is the yield rate at Lewis & Clark College. We hypothesize the following model for yield at each school:

$$y_{t,i} = \beta_{0,i} + \beta_{1,i} p_{t,i} + \beta_{2,i} x_{t,i} + u_{t,i},\ i = R, LC,$$

where $p_{t,i}$ is inflation-adjusted tuition at college i and $x_{t,i}$ is a measure of perceived college quality (perhaps its *US News & World Report* rating from the previous year).

We assume that p and x are exogenous to the yield rate and ignore any possible correlation over time in $u_{t,i}$. We are interested in two hypotheses:

i. $H_0 : \beta_{1,R} = \beta_{1,LC}$ because we want to know if the tuition sensitivity of enrollment varies between the colleges, and
ii. $H_0 : \beta_{2,R} = \beta_{2,LC}$ because we think that Reed applicants might be less sensitive to *USNWR* rankings than those applying to Lewis & Clark due to Reed's famous non-participation in the *USNWR* survey.

In order to test these hypotheses, we must estimate the Reed and Lewis & Clark equations jointly using the method of "seemingly unrelated regressions" or SUR.

1. Let $\mathbf{y}_R$ be the $T \times 1$ vector of observations $y_{t,R}$ for $t = 1, 2, \ldots, T$ and $\mathbf{y}_{LC}$ be the corresponding vector for Lewis & Clark. Let $\mathbf{u}_i$ be the $T \times 1$ vector of the error term for school i and $\mathbf{X}_i$ be the $T \times 3$ matrix with the first column being ones, the second column being the vector of $p_{t,i}$ values, and the third column being the vector of $x_{t,i}$ values for the school. Show that we can write a "stacked" regression for both schools as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_{LC} \end{bmatrix},\ \mathbf{X} = \begin{bmatrix} \mathbf{X}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{LC} \end{bmatrix},\ \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_R \\ \boldsymbol{\beta}_{LC} \end{bmatrix},\ \mathbf{u} = \begin{bmatrix} \mathbf{u}_R \\ \mathbf{u}_{LC} \end{bmatrix}.$$

2. Why might we worry, even if $\sigma_{R,R} = \operatorname{var}(u_{t,R})$ and $\sigma_{LC,LC} = \operatorname{var}(u_{t,LC})$ are constant over time, that $\sigma_{R,R} \neq \sigma_{LC,LC}$? How would that affect the OLS estimator of the stacked regression?

3. Why might we worry that $\operatorname{cov}(u_{t,R}, u_{t,LC}) \equiv \sigma_{R,LC} \neq 0$? How would that affect the OLS estimator of the stacked regression?