Lecture 5: Systems of Equations

1. (**SUR**) So, you've got a model with more than one equation: eg, consumer demand is about modelling how *all* expenditure shares vary with prices, expenditure and demographic characteristics.

   a. $Y_i^j = X_i^j b^j + e_i^j,\ j = 1,\ldots,M,\ i = 1,\ldots,N$.

      i. *i=1,..,N* indexes people, and *j=1,..,M* indexes equations. Note that the *X*'s can differ across equations, as well as, of course, the parameters b^j .

      ii. We will assume the normal things about the error terms: mean zero for each equation, homoskedastic for each equation.

      iii. We will allow for cross-equation correlations in the error terms, but no correlations across people *i*.

$$\begin{aligned} &e_i = [e_i^1,\ldots,e_i^M]' \\ &E[e_i] = 0_M\ \forall i, \\ &E[e_i'e_i] = \Sigma\ \forall i,\ E[e_i'e_j] = 0\ \forall i \neq j \end{aligned} \tag{1}$$

   b. Write out the regression equations as a stack of observations for each equations

   c.

$$Y = \begin{bmatrix} Y^1 \\ \vdots \\ Y^M \end{bmatrix} = \begin{bmatrix} Y_1^1 \\ \vdots \\ Y_N^1 \\ \vdots \\ Y_1^M \\ \vdots \\ Y_N^M \end{bmatrix},\ X = \begin{bmatrix} X^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & X^M \end{bmatrix}, b = \begin{bmatrix} b^1 \\ \ldots \\ b^M \end{bmatrix} e = \begin{bmatrix} e^1 \\ \vdots \\ e^M \end{bmatrix} = \begin{bmatrix} e_1^1 \\ \vdots \\ e_N^1 \\ \vdots \\ e_1^M \\ \vdots \\ e_N^M \end{bmatrix}$$

$$E[e] = 0_{MN},\ E[ee'] = \Omega$$

$$\Omega = \Sigma \otimes I_N = \begin{pmatrix} s_{11} I_N & \cdots & s_{M1} I_N \\ \vdots & \ddots & \vdots \\ s_{M1} I_N & \cdots & s_{MM} I_N \end{pmatrix}, \Sigma = \begin{pmatrix} s_{11} & \cdots & s_{M1} \\ \vdots & \ddots & \vdots \\ s_{M1} & \cdots & s_{MM} \end{pmatrix}$$

      i. If we knew Σ then, this could be estimated by GLS: premultiply everything by $\Omega^{-1/2}$ and run OLS on the stack. We would have all the small sample properties of OLS if we assumed normality.

d. In contrast, if we didn't know Σ, then we could find a consistent estimate of it, construct $\widehat{\Omega}$, and then premultiply everything by $\widehat{\Omega}^{-1/2}$, and run OLS on the stack. This is called feasible GLS (FGLS), and we can use only asymptotics (though we don't need to assume normality).

e. What is a consistent estimator for Ω? How about do OLS on the stack, collect the big vector of sample errors $e = \left[e^1 ' \ldots e^M ' \right]'$ (which is a vector of length *MN*—*N* errors in each of *M* equations), and then calculate

$$\Omega = \left\{ w^{jk} \right\}, \; w^{jk} = E[e^j e^k],$$

$$\widehat{\Omega} = \left\{ \widehat{w}^{jk} \right\}, \; \widehat{w}^{jk} = \frac{1}{N} \sum_{i=1}^{N} e_i^j e_i^k$$

f. This FGLS strategy is called "seemingly unrelated regression" (SUR) because the only connection across equations is in the disturbance terms—they are otherwise seemingly unrelated.

g. **Cool thing: If $X^j = X^K \; \forall j, k$, then FGLS is identical to eq-by-eq OLS.**