System of Regression and Seemingly Unrelated Regression (SUR)

Chayanee Chawanote

Semester 2/2014

Outline

- System of equations: introduction and examples
- Recall GLS
- System of equations: characteristics
- Kronecker products
- Seemingly Unrelated Regression (SUR)

Introduction

- We might want to estimate a similar specification for a number of different groups
  - a cost function for each industry
- Or, we want to estimate different dependent variables for the same cross-sectional units
  - Determinants of housing expenditure, food expenditure, and clothing expenditure
- We may estimate the equations separately if the equation for a given group/a given dependent variable meets the zero conditional mean assumption
- However, we may want to estimate the equations jointly
  - to allow cross-equation restrictions to be imposed or tested
  - to gain efficiency since we might expect the error terms across equations to be contemporaneously correlated

Systems of equations

- The population model is a set of G linear equations:

  $y_1 = \mathbf{x_1}\beta_\mathbf{1} + u_1$

  $y_2 = \mathbf{x_2}\beta_\mathbf{2} + u_2$

  ...

  $y_G = \mathbf{x_G}\beta_\mathbf{G} + u_G$

  where $\mathbf{x_g}$ is 1 $\times K_g$ and $\beta_\mathbf{g}$ is $K_g \times 1$, g $=$ 1, 2, ..., G

- $\mathbf{x_g}$ can be the same for all g, but the general model allows the elements and the dimension of $\mathbf{x_g}$ to vary across equations

- The system above represents a generic person, firm, city, etc. from the population, called "seemingly unrelated regressions (SUR)" model.

- In SUR model, the system has its own vector $\beta_\mathbf{g}$, so the equations are unrelated. However, correlation across the errors in difference equations can provide links that can be exploited in estimation.

- We have independent, identically distributed cross section observations $\{(\mathbf{X_i}, \mathbf{y_i}) : i = 1, 2, ..., N\}$ where $\mathbf{X_i}$ is a $G \times K$ matix and $\mathbf{y_i}$ is a $G \times 1$ vector.

- $\mathbf{y_i}$ contains the dependent variables for all G equations.

  - $\mathbf{X_i}$ contains the explanatory variables appearing anywhere in the system.

- The multivariate linear model: $\mathbf{y_i} = \mathbf{X_i}\beta + \mathbf{u_i}$

  - β is the $K \times 1$ parameter vector and $\mathbf{u_i}$ is a $G \times 1$ vector of unobservables (errors)

Systems of equations: Characteristics

- Case 1: if $E[\mathbf{u}_i\mathbf{u}_j'] = 0$, error of equation i and j are uncorrelated. We can estimate separately or by stacking all equations together.

  - If $E[\mathbf{u}_i\mathbf{u}_i'] = \sigma_i^2 I$, OLS is BLUE
  - If $E[\mathbf{u}_i\mathbf{u}_i'] = \sigma_i^2 \Omega_i$, GLS is BLUE

- Case 2: if $E[\mathbf{u}_i\mathbf{u}_j'] \neq 0$, there is contemporaneous correlation. If not stacking them together and estimating, β_i is still consistent but not efficient.

  - $E[\mathbf{uu}'] = \Sigma \otimes I_N$

Kronecker Products

- Let $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{p\times q}$
- $A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}_{mp\times nq}$
- $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$
- $(A \otimes B)(C \otimes D) = AC \otimes BD$ if AC and BD can be defined.
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if both A and B are invertible.
- $(A \otimes B)' = A' \otimes B'$

Seemingly Unrelated Regression (SUR)

- We see that variance-covariance matrix of the error terms is not homoskedasticity, and maybe also shows contemporaneous correlation.
- The efficient way to estimate the model is to use GLS estimator.
- SUR is GLS with specific variance-covariance matrix. Find H such that $H'H = E[uu']^{-1} = \Omega^{-1}$
  - $\hat{\beta}_{GLS} = (X*'X*)^{-1}X*'Y* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$
  - $\hat{\beta}_{SUR} = (X'(\Sigma \otimes I_N)^{-1}X)^{-1}X'(\Sigma \otimes I_N)^{-1}Y$
- Theorem: $\hat{\beta}_{SUR} = \begin{bmatrix} \hat{\beta}_{1,OLS} \\ \vdots \\ \hat{\beta}_{G,OLS} \end{bmatrix}$ if

  (1) $\sigma_{ij} = 0$, $\forall i \neq j$

  (2) $X_1 = X_2 = ... = X_G = Z$ (All k variables are the same)
- The gain in efficiency depends on the magnitude of the cross-equation contemporaneous correlations of the residuals. The higher are those correlations, the greather the gain. Furthermore if the X_i matrices' columns are highly correlated across equations, the gains will be smaller.
- If we do not know Ω, we need a feasible SUR estimator.

– Step 1: The representative element σ_{ij}, the contemporaneous correlation between u_i and u_j, may be estimated from equation-by-equation OLS residuals

– Step 2: Calculate $\hat{\sigma}_{ij} = \frac{\hat{u}_i'\hat{u}_j}{N}$ and put into $\hat{\Sigma}$

– Step 3: Estimate by $\hat{\beta}_{FSUR} = (X'(\hat{\Sigma}^{-1} \otimes I_N)X)^{-1}X'(\hat{\Sigma}^{-1} \otimes I_N)Y$

– Step 4: Calculate $\hat{u} = Y - X\hat{\beta}_{FSUR}$

– Step 5: Use $\hat{u}$ from step 4 back to start step 2 all over again and again (iteration) until $\hat{\beta}_{FSUR}$ converges to some value of $\hat{\beta}$.