Seemingly Unrelated Regression (SUR)

Zellner (1962) introduced the concept of Seemingly Unrelated Regression (SUR).

When asked "How did you get the idea for SUR ? " Zellner responded:

"On a rainy night in Seattle in about 1956 or 1957, I somehow got the idea of algebraically writing a multivariate regression model in single equation form. When I figured out how to do that, everything fell into place because then many univariate results could be carried over to apply to the multivariate system and the analysis of the multivariate system is much simplified notationally, algebraically and, conceptually. Please do read the interview of Professor Arnold Zellner by Rossi (1989, p. 292).

This methodology is based on notion that the regression equations which seem apparently independent but in fact these regression equations are related because the error terms of these equations are correlated. So the equations are called are seemingly unrelated.

In SUR the more efficient estimates of regression coefficients and predictions are obtained from the joint analysis of these regression equations.

When the two or more unrelated dependent variables are regressed on a set of independent variables (which are not highly correlated and only error terms highly correlated) , than the separate equation by equation estimation through Ordinary Least Square (OLS).

The basic version of SUR model is presented as:

$$Y = X\beta + \mu$$

It also allows imposing the test restrictions on the parameters involved in equations (Moon and Peron, 2006).

Seemingly Unrelated Regressions (SUR)

Systems of equations examples

- Demand system for food items for individuals
- Expenditure system for several different types of expenditures
- Demand and supply models

Linear systems of equations

- Systems of equations include multiple equations instead of one equation.
    - Simultaneous Equation Models: Contain both endogenous and exogenous regressors.
    - Seemingly Unrelated Regression (SUR) Models: Contain only exogenous regressors.

SUR model

- SUR model is a system of linear equations with errors that are correlated across equations for a given individual but are uncorrelated across individuals.
- The model consists of j=1…m linear regression equations for i=1…N individuals. The jth equation for individual i is

$$y_{ij} = x'_{ij}\beta_j + u_{ij}$$

- With all observations stacked, the model for the jth equation can be written as

$$y_j = x'_j\beta_j + u_j$$

- We can stack the m equations into an SUR model:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ . & & \ddots & \vdots \\ 0 & & \cdots & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_m \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_m \end{bmatrix}$$

- The error terms are assumed to have zero mean and to be independent across individuals and homoskedastic.
- For a given individual, the errors are correlated across equations,

$$E\left(u_{ij}u_{ij'}\middle|X\right) = \sigma_{jj'}, and\ \sigma_{jj'} \neq 0\ where\ j \neq\ j'$$

- The error term u_j satisfies the following assumptions:
  - Mean of error term: $E\left(u_j \middle| X\right) = 0$
  - Variance of error term in equation j: $E\left(u_j u'_j \middle| X\right) = \sigma_{jj} I_N$
  - Covariance of error terms across equations j and j': $E\left(u_j u'_{j'} \middle| X\right) = \sigma_{jj'} I_N$ where $j \neq j'$
  - Overall variance-covariance matrix: $\Omega = E\left(uu'\right) = \Sigma \otimes I_N$

- Testing cross-equation restrictions and imposing constraints:
  - We can test whether the coefficients are jointly significantly different from zero $\beta_j = \beta_{j'} = 0$ or whether the coefficients are significantly different from each other $\beta_j = \beta_{j'}$.
  - We can also impose a cross-equation restriction $\beta_j = \beta_{j'}$ and then estimate the SUR model.

Properties of the SUR model

- The SUR model is used to gain efficiency when the equations are only related through the error term.
- The parameters in the SUR model generally vary from equation to equation.
- Regressors may or may not vary from equation to equation depending on the model.
- The SUR estimates result in equation-by-equation OLS estimates when:
  - The errors are uncorrelated across equations, so Σ is diagonal.
  - Each of the equations contains exactly the same set of regressors, so $X_j = X_{j'}$.

Seemingly Unrelated Regressions Example

- We want to study how math and reading scores are influenced by several factors.
- Data are from the High School and Beyond Study.
- Dependent variable for equation 1: math score.
- Dependent variable for equation 2: reading score.
- Independent variables for equation 1: female, program, and science score.
- Independent variables for equation 2: female and social sciences score.

- Because SUR reduces to OLS if the same regressors are used in both equations, we omit one regressor from each equation.

SUR results

	OLS math score	OLS reading score	SUR math score	SUR reading score
Female	0.97	-1.76	0.80	-1.71
Program	-0.41	-	-0.53	-
Science score	0.59*	-	0.53*	-
Social science score	-	0.60*	-	0.55*
Constant	21.97*	21.86*	25.60*	24.37*

- Coefficient interpretation: When a person has 1 unit of a higher science score, they have 0.59 or 0.53 units of higher math score using the OLS or SUR model.
- For this particular example, the OLS results are similar to the SUR results.
- The correlation between the errors in the two equations is positive (0.18).
- The Breusch-Pagan test for error independence shows a chi-square test statistic of 6.8 and a p-value of less than 0.05, which indicates that there is statistically significant correlation between the errors in the two equations (but correlation of 0.18 is not particularly strong).
- A test for the equality of the female coefficients in both equations cannot be rejected, so we can impose a cross-equation restriction of equal coefficients and estimate an SUR model.