Chapter 18
Seemingly Unrelated Regression Equations Models

A basic nature of the multiple regression model is that it describes the behaviour of a particular study variable based on a set of explanatory variables. When the objective is to explain the whole system, there may be more than one multiple regression equations. For example, in a set of individual linear multiple regression equations, each equation may explain some economic phenomenon. One approach to handle such a set of equations is to consider the set up of simultaneous equations model is which one or more of the explanatory variables in one or more equations are itself the dependent (endogenous) variable associated with another equation in the full system. On the other hand, suppose that none of the variables is the system are simultaneously both explanatory and dependent in nature. There may still be interactions between the individual equations if the random error components associated with at least some of the different equations are correlated with each other. This means that the equations may be linked statistically, even though not structurally – through the jointness of the distribution of the error terms and through the non-diagonal covariance matrix. Such behaviour is reflected in the seemingly unrelated regression equations (SURE) model in which the individual equations are in fact related to one another, even though superficially they may not seem to be.

The basic philosophy of the SURE model is as follows. The jointness of the equations is explained by the structure of the SURE model and the covariance matrix of the associated disturbances. Such jointness introduces additional information which is over and above the information available when the individual equations are considered separately. So it is desired to consider all the separate relationships collectively to draw the statistical inferences about the model parameters.

Example:

Suppose a country has 20 states and the objective is to study the consumption pattern of the country. There is one consumption equation for each state. So all together there are 20 equations which describe 20 consumption functions. It may also not necessary that the same variables are present in all the models. Different equations may contain different variables. It may be noted that the consumption pattern of the neighbouring states may have characteristics in common. Apparently, the equations may look distinct individually but there may be some kind of relationship that may be existing among the equations. Such equations can be used to examine the jointness of the distribution of disturbances. It seems reasonable to

assume that the error terms associated with the equations may be contemporaneously correlated. The equations are apparently or "seemingly" unrelated regressions rather than independent relationships.

Model:

We consider here a model comprising of M multiple regression equations of the form

$$y_{ti} = \sum_{j=1}^{k_i} x_{tij}\beta_{ij} + \varepsilon_{ti},\ t = 1,2,...,T;\ i = 1,2,...,M;\ j = 1,2,...,k_i$$

where y_{ti} is the t^{th} observation on the i^{th} dependent variable which is to be explained by the i^{th} regression equation, x_{tij} is the t^{th} observation on j^{th} explanatory variable appearing in the i^{th} equation, β_{ij} is the coefficient associated with x_{tij} at each observation and ε_{ti} is the t^{th} value of the random error component associated with i^{th} equation of the model.

These M equations can be compactly expressed as

$$y_i = X_i\beta_i + \varepsilon_i,\ i = 1,2,...,M$$

where y_i is $(T \times 1)$ vector with elements y_{ti}; X_i *is* $(T \times K_i)$ matrix whose columns represent the T observations on an explanatory variable in the i^{th} equation; β_i is a $(k_i \times 1)$ vector with elements β_{ij}; and ε_i is a $(T \times 1)$ vector of disturbances. These M equations can be further expressed as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & X_M \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{pmatrix}$$

or $y = X\beta + \varepsilon$

where the orders of y is $(TM \times 1)$, X is $(TM \times k^*)$, β is $(k^* \times 1)$, ε is $(TM \times 1)$ and $k^* = \sum_i k_i$.

Treat each of the M equations as the classical regression model and make conventional assumptions for $i = 1, 2, ..., M$ as

- X_i is fixed.
- $rank\left(X_i\right) = k_i$.
- $\lim_{T\to\infty}\left(\frac{1}{T} X_i' X_i\right) = Q_{ii}$ where Q_{ii} is nonsingular with fixed and finite elements.
- $E\left(u_i\right) = 0$.
- $E\left(u_i u_i'\right) = \sigma_{ii} I_T$ where σ_{ii} is the variance of disturbances in i^{th} equation for each observation in the sample.

Considering the interactions between the M equations of the model, we assume

- $\lim_{T\to\infty} \frac{1}{T} X_i' X_j = Q_{ij}$
- $E\left(u_i u_j'\right) = \sigma_{ij} I_T;\ i, j = 1, 2, ..., M$

where Q_{ij} is non-singular matrix with fixed and finite elements and σ_{ij} is the covariance between the disturbances of i^{th} and j^{th} equations for each observation in the sample.

Compactly, we can write

$$E\left(\varepsilon\right) = 0$$

$$E\left(\varepsilon\varepsilon'\right) = \begin{pmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1M} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T & \cdots & \sigma_{2M} I_T \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{M1} I_T & \sigma_{M2} I_T & \cdots & \sigma_{MM} I_T \end{pmatrix} = \Sigma \otimes I_T = \psi$$

where $\otimes$ denotes the Kronecker product operator, ψ is $\left(MT \times MT\right)$ matrix and $\Sigma = \left(\left(\sigma_{ij}\right)\right)$ *is* $\left(M \times M\right)$ positive definite symmetric matrix. The definiteness of Σ avoids the possibility of linear dependencies among the contemporaneous disturbances in the M equations of the model.

The structure $E\left(uu'\right) = \Sigma \otimes I_T$ implies that

- variance of ε_{ti} is constant for all t.
- contemporaneous covariance between ε_{ti} and ε_{tj} is constant for all t.
- intertemporal covariance between ε_{ti} and $\varepsilon_{t^* j}$ $\left(t \neq t^*\right)$ are zero for all i and j.

By using the terminologies “contemporaneous” and “intertemporal” covariance, we are implicitly assuming that the data are available in time series form but this is not restrictive. The results can be used for cross-section data also. The constancy of the contemporaneous covariances across sample points is a natural generalization of homoskedastic disturbances in a single equation model.

It is clear that the M equations may appear to be not related in the sense that there is no simultaneity between the variables in the system and each equation has its own explanatory variables to explain the study variable. The equations are related stochastically through the disturbances which are serially correlated across the equations of the model. That is why this system is referred to as SURE model.

The SURE model is a particular case of simultaneous equations model involving M structural equations with M jointly dependent variable and $k\left(\geq k_i \text{ for all } i\right)$ distinct exogenous variables and in which neither current nor logged endogenous variables appear as explanatory variables in any of the structural equations.

The SURE model differs from the multivariate regression model only in the sense that it takes account of prior information concerning the absence of certain explanatory variables from certain equations of the model. Such exclusions are highly realistic in many economic situations.

OLS and GLS estimation:

The SURE model is

$$y = X\beta + \varepsilon,\ E\left(\varepsilon\right) = 0,\ V\left(\varepsilon\right) = \Sigma \otimes I_T = \psi.$$

Assume that ψ is known.

The OLS estimator of β is

$$b_0 = \left(X'X\right)^{-1} X'y$$

Further

$$E\left(b_0\right) = \beta$$
$$V\left(b_0\right) = E\left(b_0 - \beta\right)\left(b_0 - \beta\right)'$$
$$= \left(X'X\right)^{-1} X'\psi X\left(X'X\right)^{-1}.$$

The generalized least squares (GLS) estimator of β

$$\hat{\beta} = \left(X'\psi^{-1}X\right)^{-1} X'\psi^{-1}y$$
$$= \left[X'\left(\Sigma^{-1} \otimes I_T\right)X\right]^{-1} X'\left(\Sigma^{-1} \otimes I_T\right)y$$
$$E\left(\hat{\beta}\right) = \beta$$
$$V\left(\hat{\beta}\right) = E\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'$$
$$= \left(X'\psi^{-1}X\right)^{-1}$$
$$= \left[X'\left(\Sigma^{-1} \otimes I_T\right)X\right]^{-1}.$$

Define

$$G = \left(X'X\right)^{-1} X' - \left(X'\psi^{-1}X\right)^{-1} X'\psi^{-1}$$

then $GX = 0$ and we find that

$$V\left(b_0\right) - V\left(\hat{\beta}\right) = G\psi G'.$$

Since ψ is positive definite, so $G\psi G'$ is atleast positive semidefinite and so GLSE is, in general, more efficient than OLSE for estimating β. In fact, using the result that GLSE best linear unbiased estimator of β, so we can conclude that $\hat{\beta}$ is the best linear unbiased estimator in this case also.

Feasible generalized least squares estimation:

When Σ is unknown, then GLSE of β cannot be used. Then Σ can be estimated and replaced by $\left(M \times M\right)$ matrix S. With such replacement, we obtain a feasible generalized least squares (FGLS) estimator of β as

$$\hat{\beta}_F = \left[X'\left(S^{-1} \otimes I_T\right)X\right]^{-1} X'\left(S^{-1} \otimes I_T\right)y.$$

Assume that $S = \left(\left(s_{ij}\right)\right)$ is a nonsingular matrix and s_{ij} is some estimator of σ_{ij}.

Estimation of Σ

There are two possible ways to estimate σ_{ij}'s.

1. Use of unrestricted residuals

Let K be the total number of distinct explanatory variables out of $k_1, k_2, ..., k_m$ variables in the full model

$$y = X\beta + \varepsilon,\ E(\varepsilon) = 0,\ V(\varepsilon) = \Sigma \otimes I_T$$

and let Z be a $T \times K$ observation matrix of these variables.

Regress each of the M study variables on the column of Z and obtain $(T \times 1)$ residual vectors

$$\begin{aligned}\hat{\varepsilon}_i &= y_i - Z(Z'Z)^{-1}Z'y_i \quad i = 1,2,...,M \\ &= \bar{H}_Z y_i\end{aligned}$$

where $\bar{H}_Z = I_T - Z(Z'Z)^{-1}Z'$.

Then obtain

$$\begin{aligned}s_{ij} &= \frac{1}{T}\hat{\varepsilon}_i'\hat{\varepsilon}_j \\ &= \frac{1}{T}y_i'\bar{H}_Z y_j\end{aligned}$$

and construct the matrix $S = \left(\left(s_{ij}\right)\right)$ accordingly.

Since X_i is a submatrix of Z, so we can write

$$X_i = ZJ_i$$

where J_i is a $(K \times k_i)$ selection matrix. Then

$$\begin{aligned}\bar{H}_Z X_i &= X_i - Z(Z'Z)^{-1}Z'X_i \\ &= X_i - ZJ_i \\ &= 0\end{aligned}$$

and thus

$$\begin{aligned}y_i'\bar{H}_Z y_j &= \left(\beta_i'X_i' + \varepsilon_i'\right)\bar{H}_Z\left(X_j\beta_j + \varepsilon_j\right) \\ &= \varepsilon_i'\bar{H}_Z\varepsilon_j.\end{aligned}$$

Hence

$$E\left(s_{ij}\right)=\frac{1}{T}E\left(\varepsilon_i'\bar{H}_Z\varepsilon_j\right)$$
$$=\frac{1}{T}\sigma_{ij}tr\left(\bar{H}_Z\right)$$
$$=\left(1-\frac{K}{T}\right)\sigma_{ij}$$
$$E\left(\frac{T}{T-K}s_{ij}\right)=\sigma_{ij}.$$

Thus an unbiased estimator of σ_{ij} is given by $\frac{T}{T-K}s_{ij}$.

2. Use of restricted residuals

In this approach to find an estimator of σ_{ij}, the residuals obtained by taking into account the restrictions on the coefficients which distinguish the SURE model from the multivariate regression model are used as follows.

Regress y_i on X_i, i.e., regress each equation, $i=1,2,...,M$ by OLS and obtain the residual vector

$$\tilde{u}_i=\left[I-X_i\left(X_i'X_i\right)^{-1}X_i'\right]y_i$$
$$=\bar{H}_{X_i}y_i.$$

A consistent estimator of σ_{ij} is obtained as

$$s_{ij}^*=\frac{1}{T}\tilde{u}_i'\tilde{u}_j$$
$$=\frac{1}{T}y_i'\bar{H}_{X_i}\bar{H}_{X_j}\tilde{y}_j$$

where

$$\bar{H}_{X_i}=I-X_i\left(X_i'X_i\right)^{-1}X_i$$
$$\bar{H}_{X_j}=I-X_j\left(X_j'X_j\right)^{-1}X_j.$$

Using s_{ij}^*, a consistent estimator of S can be constructed.

If T in s_{ij}^* is replaced by

$$tr\left(\bar{H}_{X_i}\bar{H}_{X_j}\right)=T-k_i-k_j+tr\left(X_i'X_i\right)^{-1}X_i'X_j\left(X_j'X_j\right)^{-1}X_j'X_i$$

then s_{ij}^* is an unbiased estimator of σ_{ij}.