ECON 5350 Class Notes
Chapter 14. Systems of Regression Equations

1 Introduction

Arnold Zellner (1962) introduced the idea of Seemingly Unrelated Regressions (SUR). Often a set of equations with distinct dependent and independent variables, as well as different coefficients, are linked together by some common unmeasurable factor. Examples include systems of factor demands by a particular firm, agricultural supply-response equations, and capital-asset pricing models. The methods presented here can also be thought of as an alternative estimation framework for panel-data models.

2 Seemingly Unrelated Regressions (SUR) Model

Consider the following set of equations

$$y_i = X_i\beta_i + \epsilon_i \tag{1}$$

for $i = 1, ..., M$, where the matrices y_i, X_i and β_i are of dimension $(T \times 1)$, $(T \times K_i)$ and $(K_i \times 1)$, respectively[1]. The stacked system in matrix form is

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} = X\beta + \epsilon.$$

Although each of the M equations may seem unrelated (i.e., each has potentially distinct coefficient vectors, dependent variables and explanatory variables), the equations in (1) are linked through their (mean-zero) error structure

$$E(\epsilon\epsilon') = \Omega = \Sigma \otimes I_T = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1M}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & & \sigma_{2M}I_T \\ \vdots & & \ddots & \vdots \\ \sigma_{M1}I_T & \sigma_{M2}I_T & \cdots & \sigma_{MM}I_T \end{bmatrix}_{MT \times MT}$$

[1] Although each equation typically represents a separate time series, it is possible that T instead denotes the number of cross sections within an equation.

where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2M} \\ \vdots & & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}_{M\times M}$$

is the variance-covariance matrix for each $t = 1, ..., T$ error vector.

2.1 Generalized Least Squares (GLS)

The system resembles the one we studied in chapter 10 on nonshperical disturbances. The efficient estimator in this context is the GLS estimator

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) = [X'(\Sigma^{-1} \otimes I)X]^{-1}[X'(\Sigma^{-1} \otimes I)Y].$$

Assuming all the classical assumptions hold (other than that of spherical disturbances), GLS is the best linear unbiased estimator. There are two important conditions under which GLS does not provide any efficiency gains over OLS:

- $\sigma_{ij} = 0$. When all the contemporaneous correlations across equations equal zero, the equations are not linked in any fashion and GLS does not provide any efficiency gains. In fact, one can show that $b = \hat{\beta}$.
- $X_1 = X_2 = \cdots = X_M$. When the explanatory variables are identical across equations, $b = \hat{\beta}$.

As a rule, the efficiency gains of GLS over OLS tend to be greater when

- the contemporaneous correlation in errors across equations (σ_{ij}) is greater and
- there is less correlation between X across equations.

2.2 Feasible Generalized Least Squaures (FGLS)

Typically, Σ is not known. Assuming that a consistent estimator of Σ is available, the feasible GLS estimator

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y) = [X'(\hat{\Sigma}^{-1} \otimes I)X]^{-1}[X'(\hat{\Sigma}^{-1} \otimes I)Y] \quad (2)$$

will be a consistent estimator of β. The typical estimator of Σ is

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1M} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & & \hat{\sigma}_{2M} \\ \vdots & & \ddots & \vdots \\ \hat{\sigma}_{M1} & \hat{\sigma}_{M2} & \cdots & \hat{\sigma}_{MM} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}e_1'e_1 & \frac{1}{T}e_1'e_2 & \cdots & \frac{1}{T}e_1'e_M \\ \frac{1}{T}e_2'e_1 & \frac{1}{T}e_2'e_2 & & \frac{1}{T}e_2'e_M \\ \vdots & & \ddots & \vdots \\ \frac{1}{T}e_M'e_1 & \frac{1}{T}e_M'e_2 & \cdots & \frac{1}{T}e_M'e_M \end{bmatrix}, \tag{3}$$

where e_i, $i = 1, ..., M$ represent the OLS residuals. Degrees of freedom corrections for the elements in $\hat{\Sigma}$ are possible, but will not generally produce unbiasedness. It is also possible to iterate on (2) and (3) until convergence, which will produce the maximum likelihood estimator under multivariate normal errors. In other words, $\hat{\beta}_{FGLS}$ and $\hat{\beta}_{ML}$ will have the same limiting distributions such that

$$\hat{\beta}_{ML,FGLS} \overset{asy}{\sim} N(\beta, \Psi)$$

where Ψ is consistently estimated by

$$\hat{\Psi} = [X'(\hat{\Sigma}^{-1} \otimes I)X]^{-1}.$$

2.3 Maximum Likelihood

Although asymptotically equivalent, maximum likelihood is an alternative estimator to FGLS that will provide different answers in small samples. Begin by rewriting the model for the t^{th} observation as

$$Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{Mt} \end{bmatrix}' = x_t^* \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_M \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{M,t} \end{bmatrix}' = x_t^* \Pi + \epsilon_t'$$

where x_t^* is the row vector of all different explanatory variables in the system and each π_i is the column vector of coefficients for the i^{th} equation (unless each equation contains all explanatory variables, there will be zeros in π_i to allow for exclusion restrictions). Assuming multivariate normally distributed errors, the log likelihood function is

$$\log L = \sum\nolimits_{t=1}^{T} \log L_t = -\frac{MT}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| - \frac{1}{2}\sum\nolimits_{t=1}^{T} \epsilon_t'\Sigma^{-1}\epsilon_t \tag{4}$$

where, as defined earlier, $\Sigma = E(\epsilon_t \epsilon_t')$. The maximum likelihood estimates are found by taking the derivatives of (4) with respect to Π and Σ, setting them equal to zero and solving.

2.4 Hypothesis Testing

We consider two types of tests – tests for contemporaneous correlation between errors and test for linear restrictions on the coefficients.

2.4.1 Contemporaneous Correlation

If there is no contemporaneous correlation between errors in different equations (i.e., Σ is diagonal), then OLS equation-by-equation is fully efficient. Therefore, it is useful to test the following restriction

$$\begin{aligned} H_0 &: \sigma_{ij} = 0 \; \forall i \neq j \\ H_A &: H_0 \text{ false.} \end{aligned}$$

Breusch and Pagan suggest using the Lagrange multiplier test statistic

$$\lambda = T \sum_{i=2}^{M} \sum_{j=1}^{i-1} r_{ij}^2$$

where r_{ij} is calculated using the OLS residuals as follows

$$r_{ij} = \frac{e_i' e_j}{\sqrt{(e_i' e_i)(e_j' e_j)}}.$$

Under the null hypothesis, λ is asymptotically chi-squared with $M(M-1)/2$ degrees of freedom.

2.4.2 Restrictions on Coefficients

The general F test presented in chapter 6 can be extended to the SUR system. However, since the statistic requires using $\hat{\Sigma}$, the test will only be valid asymptotically. Consider testing the following J linear restrictions

$$\begin{aligned} H_0 &: R\beta = q \\ H_A &: H_0 \text{ false} \end{aligned}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_M)'$. Within the SUR framework, it is possible to test coefficient restrictions across equations. One possible test statistic is

$$W = (R\hat{\beta}_{FGLS} - q)'[Rvar(\hat{\beta}_{FGLS})R']^{-1}(R\hat{\beta}_{FGLS} - q)$$

which has an asymptotic chi-square distribution with J degrees of freedom.

2.5 Autocorrelation

Heteroscedasticity and autocorrelation are possibilities within the SUR framework. I will focus on autocorrelation because SUR systems are often comprised of time series observations for each equation. Assume the errors follow

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{it}$$

where ν_{it} is white noise. The overall error structure will now be

$$E(\epsilon\epsilon') = \Omega = \begin{bmatrix} \sigma_{11}\Omega_{11} & \sigma_{12}\Omega_{12} & \cdots & \sigma_{1M}\Omega_{1M} \\ \sigma_{21}\Omega_{21} & \sigma_{22}\Omega_{22} & & \sigma_{2M}\Omega_{2M} \\ \vdots & & \ddots & \vdots \\ \sigma_{M1}\Omega_{M1} & \sigma_{M2}\Omega_{M2} & \cdots & \sigma_{MM}\Omega_{MM} \end{bmatrix}_{MT \times MT}$$

where

$$\Omega_{ij} = \begin{bmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & & \rho_j^{T-2} \\ \vdots & & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \cdots & 1 \end{bmatrix}_{T \times T} .$$

The following three-step approach is recommended

1. Run OLS equation-by-equation. Compute consistent estimate of ρ_i (e.g., $\hat{\rho}_i = (\sum_{t=2}^{T} e_{i,t}e_{i.t-1})/(\sum_{t=1}^{T} e_{i.t}^2)$). Transform the data, using either Prais-Winsten or Cochrane-Orcutt, to remove the autocorrelation.

2. Estimate Σ using the transformed data as suggested in (3).

3. Use $\hat{\Sigma}$ and equation (2) to calculate the FGLS estimates.

3 Gauss Application

Consider the data taken from Woolridge (2002). The model attempts to explain wages and fringe benefits for 616 workers:

$$
\begin{aligned}
Wages_{1t} &= X_{1t}\beta_1 + \epsilon_{1t} \\
Benefits_{2t} &= X_{2,t}\beta_2 + \epsilon_{2t}
\end{aligned}
$$

where $X_{1t} = X_{2t}$ so that OLS and FGLS will produce equivalent results. Although OLS and FGLS are equivalent, one advantage of FGLS within a SUR framework is that it allows you test coefficient restrictions across equations. Doing OLS equation-by-equation would not allow such tests. The variables are defined as follows:

Dependent Variables

- Wages. Hourly earnings in 1999 dollars per hour.
- Benefits. Hourly benefits (vacation, sick leave, insurance and pension) in 1999 dollars per hour.

Explanatory Variables

- Education. Years of schooling.
- Experience. Years of work experience.
- Tenure. Years with current employer.
- Union. One if union member, zero otherwise.
- South. One if live in south, zero otherwise.
- Northeast. One if live in northeast, zero otherwise.
- Northcentral. One if live in northcentral, zero otherwise.
- Married. One if married, zero otherwise.
- White. One if white, zero otherwise.
- Male. One if male, zero otherwise.

See Gauss example 14.1 for further details.