Seemingly Unrelated Regression (SUR)

Suppose we have m equations:

$$\underset{T\times 1}{y^j} = \underset{T\times k_j}{X^j}\underset{k_j\times 1}{\beta^j} + \underset{T\times 1}{\varepsilon^j}, \quad j = 1, \dots, m$$

The superscript j denotes the equation.

We make the following assumptions: $\forall j,k \in (1,\dots,m)$

1) $E(\varepsilon^j \mid X^1, \dots, X^m) = 0$ (Orthogonality)
2) $V(\varepsilon^j) = \sigma_{jj} I_T$ (Spherical Disturbances within Equation)
3) $E(\underset{T\times 1}{\varepsilon^j}\underset{1\times T}{\varepsilon^{k\prime}}) = \sigma_{jk} I_T$ (Errors across equations Correlated in the same time periods)

<u>EX 1</u>: You have two firms, IBM and Microsoft. You have profits over time, $t = 1990, \dots 2000$

$$\underset{10\times 1}{y^I} = \underset{10\times k_I}{X^I}\underset{k_I\times 1}{\beta^I} + \underset{10\times 1}{\varepsilon^I} \quad ; \quad \underset{10\times 1}{y^M} = \underset{10\times k_m}{X^m}\underset{k_m\times 1}{\beta^m} + \underset{10\times 1}{\varepsilon^m}$$

X^I, X^m do not necessarily include the same variables or have the same dimension.

Our assumptions are 1) $E(\varepsilon^I \mid X^I, X^m) = 0$
$E(\varepsilon^m \mid X^I, X^m) = 0$

2) $V(\varepsilon^I) = \sigma_I^2 I_{10\times 10}$; $V(\varepsilon^m) = \sigma_m^2 I_{10\times 10}$
3) $E(\varepsilon^I \varepsilon^{m\prime}) = \sigma_{Im} I_{10\times 10}$

Lets stack the System.

$$\bar{y} = \begin{bmatrix} y^1 \\ \hline \vdots \\ \hline y^m \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} y^1_1 \\ \vdots \\ y^1_T \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y^m_1 \\ \vdots \\ y^m_T \end{pmatrix} \end{bmatrix} \; ; \quad \bar{X} = \begin{bmatrix} X^1 & 0 & \cdots & 0 \\ 0 & X^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & X^m \end{bmatrix}$$

$Tm \times 1$ $\qquad$ $Tm \times \sum_{j=1}^{m} k_j$

$$\bar{\beta} = \begin{bmatrix} \beta^1 \\ \vdots \\ \beta^m \end{bmatrix} \; ; \quad \bar{\varepsilon} = \begin{bmatrix} \varepsilon^1 \\ \vdots \\ \varepsilon^m \end{bmatrix}$$

$\sum_{j=1}^{m} k_j \times 1$ $\qquad$ $Tm \times 1$

This gives: $\quad \bar{y} = \bar{X}\bar{\beta} + \bar{\varepsilon}$

Kronecker Product: Recall that

$$\underset{(m \times n)}{A} \otimes \underset{(r \times s)}{B} = \underbrace{\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & \ddots & & \vdots \\ \vdots & & & \vdots \\ a_{m1}B & \cdots & \cdots & a_{mn}B \end{bmatrix}}_{(mr \times ns)}$$

1) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ for AC, BD well-defined (matrices need to be comformable)

2) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ for A, B invertible

3) $(A \otimes B)' = A' \otimes B'$

Let's examine the stacked equation

$$\bar{y} = \bar{X}\bar{\beta} + \bar{\varepsilon}$$

$E(\bar{\varepsilon} \mid \bar{X}) = 0$ follows from $E(\varepsilon^j \mid X^1, \dots, X^m) = 0$, $j = 1, \dots, m$.

$$V(\bar{\varepsilon}) = E(\bar{\varepsilon}\bar{\varepsilon}' \mid \bar{X}) = E\left[\begin{pmatrix} \varepsilon^1 \\ \vdots \\ \varepsilon^m \end{pmatrix} (\varepsilon^{1\prime} \cdots \varepsilon^{m\prime}) \,\Big|\, \bar{X}\right]$$

(drop $\bar{X}$)
$$= E\begin{bmatrix} \varepsilon^1\varepsilon^{1\prime} & \varepsilon^1\varepsilon^{2\prime} & \cdots & \varepsilon^1\varepsilon^{m\prime} \\ \varepsilon^2\varepsilon^{1\prime} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \varepsilon^m\varepsilon^{1\prime} & \cdots & \cdots & \varepsilon^m\varepsilon^{m\prime} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1m} I_T \\ \sigma_{21} I_T & \sigma_{22} I_T & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{m1} I_T & \cdots & \cdots & \sigma_{mm} I_T \end{bmatrix}$$

$$= \underbrace{\underset{m \times m}{\Sigma} \otimes \underset{T \times T}{I_T}}_{mT \times mT}$$

$\bar{\varepsilon}$ is not spherical $\Rightarrow$ GLS is BLUE.

$$\hat{\bar{\beta}}_{GLS} = (\bar{X}'(\Sigma \otimes I_T)^{-1}\bar{X})^{-1}(\bar{X}'(\Sigma \otimes I_T)^{-1}\bar{y})$$

Of course, Σ is unknown so we resort to FGLS - feasible generalized least squares. Fortunetly, the restrictions we're imposed on the errors give us only M^2 variables in Σ that need estimating. Recall that without any restrictions, $V(\bar{\varepsilon})$ is $TM \times TM$ and FGLS is not consistent - estimating $V(\bar{\varepsilon})$ in this case and plugging into FGLS would give inconsistent estimates.

However, with our restrictions on ε^i, $i = 1, \dots, m$, we have $Var(\bar{\varepsilon}) = \underset{M \times M}{\Sigma} \otimes I_T$.

The least squares residuals may be used to estimate consistently the elements of Σ

$$\hat{\sigma}_{ij} = s_{ij} = \frac{\hat{e}^{i\prime}\hat{e}^{j}}{T}$$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & \ddots & & \vdots \\ \vdots & & & \vdots \\ s_{m1} & \cdots & \cdots & s_{mm} \end{bmatrix} \qquad S \text{ estimates } \Sigma.$$

The FGLS estimator is then:

$$\hat{\hat{\beta}}_{FGLS} = \left(\bar{X}'(S \otimes I_T)^{-1}\bar{X}\right)^{-1}\left(\bar{X}'(S \otimes I_T)^{-1}\bar{Y}\right).$$

We have some interesting results for the SUR model. The equations in the model are linked only by their disturbances, so it is interesting to ask how much efficiency is gained by using GLS instead of OLS.

1. If the equations are actually unrelated, ($\sigma_{ij} = 0 \quad \forall\, i \neq j$), then GLS = OLS. However, FGLS will not be numerically equal to OLS.
2. If the equations have identical explanatory variables ($X^i = X^j \quad \forall\, i,j$), OLS = GLS. This will be shown.
3. If the regressors in one block of equations are a subset of those in another, GLS brings no efficiency gain

3. (Cont) in estimation of the smaller equations.

We have two more results which apply to both the SUR model we have been looking at (with $V(\varepsilon^{j}) = \sigma_{jj} I_T$, $E(\varepsilon^{j}\varepsilon^{k'}) = \sigma_{jk} I_T$) and more general SUR models that allow some correlation in the disturbances.

4. The greater the correlation of the disturbances, the greater the efficiency gain accruing to GLS.
5. The less correlation there is between the X matrices, the greater the gain in GLS.

Property 2 applies to the reduced form of our structural equations model.

Let's prove it.

Proof of Prop (2):

$X^j = X^k \quad \forall \; k, j$ so that

$$\bar{X} = \underset{TM \times \sum_{j=1}^{m} k_j}{\begin{bmatrix} X^1 & 0 & \cdots & 0 \\ 0 & X^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & X^m \end{bmatrix}} = \underset{Tm \times km}{\begin{bmatrix} \underset{(T\times k)}{X} & 0 & \cdots & 0 \\ 0 & \underset{(T\times k)}{X} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \underset{(T\times k)}{X} \end{bmatrix}} = I_m \otimes X$$

$$\hat{\beta}_{GLS} = \left(\bar{X}'(\Sigma^{-1} \otimes I)\bar{X}\right)^{-1}\left(\bar{X}'(\Sigma^{-1} \otimes I)\bar{y}\right)$$

$$= \left[(I \otimes X)'(\Sigma^{-1} \otimes I)(I \otimes X)\right]^{-1}\left[(I \otimes X)'(\Sigma^{-1} \otimes I)\bar{y}\right]$$

$\{(A\otimes B)' = A' \otimes B'\}$ $= \left[(I' \otimes X')(\Sigma^{-1} \otimes I)(I \otimes X)\right]^{-1}\left[(I' \otimes X')(\Sigma^{-1} \otimes I)\bar{y}\right]$

$\{(A\otimes B)(C\otimes D) = AC \otimes BD\}$ $= \left[\Sigma^{-1} \otimes (X'X)\right]^{-1}\left[(\Sigma^{-1} \otimes X')\bar{y}\right]$

$\{(A\otimes B)^{-1} = A^{-1} \otimes B^{-1}\}$ $= \left[\Sigma \otimes (X'X)^{-1}\right]\left[(\Sigma^{-1} \otimes X')\bar{y}\right]$

$\{(A\otimes B)(C\otimes D) = AC \otimes BD\}$ $= \left[I_m \otimes (X'X)^{-1}X'\right]\bar{y}$

$$= \begin{bmatrix} (X'X)^{-1}X' & 0 & \cdots & 0 \\ 0 & (X'X)^{-1}X' & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & (X'X)^{-1}X' \end{bmatrix} \begin{bmatrix} y^1 \\ \vdots \\ \vdots \\ y^m \end{bmatrix}$$

$$= \begin{bmatrix} (X'X)^{-1}X'y^1 \\ (X'X)^{-1}X'y^2 \\ \vdots \\ (X'X)^{-1}X'y^m \end{bmatrix}$$

This is least squares equation by equation. ∎

Tradeoff between Consistency & Efficiency:

GLS is generally more efficient than equation-by-equation LS, but misspecification in one equation affects the consistency of the estimates of all equations.

$$\hat{\bar{\beta}}_{FGLS} - \bar{\beta} = \left(\frac{1}{T}\bar{X}'(S^{-1}\otimes I)\bar{X}\right)^{-1}\underbrace{\left(\frac{1}{T}\bar{X}'(S^{-1}\otimes I)\bar{\varepsilon}\right)}$$

Consistency of $\hat{\bar{\beta}}_{FGLS}$ depends on this term.

$$\underset{(\sum_{j=1}^{m} k_j \times Tm)}{\frac{1}{T}\bar{X}'}\underset{(Tm\times Tm)}{(S^{-1}\otimes I)}\underset{(Tm\times 1)}{\bar{\varepsilon}} = \frac{1}{T}\begin{bmatrix} X^{1\prime} & & 0 \\ & X^{2\prime} & \\ 0 & & \ddots \\ & & & X^{m\prime}\end{bmatrix}\begin{bmatrix} s_{11}I_T & s_{12}I_T & \cdots & s_{1m}I_T \\ s_{21}I_T & s_{22}I_T & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{m1}I_T & \cdots & & s_{mm}I_T\end{bmatrix}\begin{bmatrix}\varepsilon^1 \\ \varepsilon^2 \\ \vdots \\ \varepsilon^m\end{bmatrix}$$

$$= \frac{1}{T}\underset{(\sum_{j=1}^{m} k_j \times Tm)}{\begin{bmatrix} s_{11}X^{1\prime} & s_{12}X^{1\prime} & \cdots & s_{1m}X^{1\prime} \\ s_{21}X^{2\prime} & s_{22}X^{2\prime} & \cdots & s_{2m}X^{2\prime} \\ \vdots & & \ddots & \vdots \\ s_{m1}X^{m\prime} & s_{m2}X^{m\prime} & \cdots & s_{mm}X^{m\prime}\end{bmatrix}}\underset{(Tm\times 1)}{\begin{bmatrix}\varepsilon^1 \\ \varepsilon^2 \\ \vdots \\ \varepsilon^m\end{bmatrix}}$$

$$= \underset{(\sum_{j=1}^{m} k_j \times 1)}{\begin{bmatrix} \frac{1}{T}(s_{11}X^{1\prime}\varepsilon^1 + s_{12}X^{1\prime}\varepsilon^2 + \cdots + s_{1m}X^{1\prime}\varepsilon^m) \\ \frac{1}{T}(s_{21}X^{2\prime}\varepsilon^1 + s_{22}X^{2\prime}\varepsilon^2 + \cdots + s_{2m}X^{2\prime}\varepsilon^m) \\ \vdots \\ \frac{1}{T}(s_{m1}X^{m\prime}\varepsilon^1 + s_{m2}X^{m\prime}\varepsilon^2 + \cdots + s_{mm}X^{m\prime}\varepsilon^m)\end{bmatrix}}\begin{matrix}\}k_1 \\ \}k_2 \\ \vdots \\ \}k_m\end{matrix}$$

<u>Trade-off (Cont)</u>

Consistency of $\hat{\bar{\beta}}_{FGLS}$ requires (in general)

$$\text{plim } \frac{1}{T} X^{j\prime} \varepsilon^{k} = 0 \quad \forall\ j, k.$$

Note that if X^2 (X matrix for Eqn. 2) is correlated with ε^{j}, for any j, so that $\frac{1}{T} X^{2\prime} \varepsilon^{j} \not\rightarrow 0$, it is not only the β^2 part of $\bar{\beta} = \begin{pmatrix} \beta^1 \\ \beta^2 \\ \vdots \\ \beta^n \end{pmatrix}$ that is estimated inconsistently.

$\left(\frac{1}{T} \bar{X}'(S^{-1} \otimes I)\bar{\varepsilon}\right)$ is pre-multiplied by $\left(\frac{1}{T} \bar{X}'(S^{-1} \otimes I)\bar{X}\right)^{-1}$, thereby feeding the inconsistency into all of $\bar{\beta}$.

We can do a Hausman Specification Test to test H_0: $\text{plim } \frac{1}{T} X^{j\prime} \varepsilon^{k} = 0 \quad \forall\ j, k$.

$\hat{\bar{\beta}}_{FGLS}$ is consistent and efficient under H_0

$\hat{\bar{\beta}}_{OLS}$ (eqn by eqn) is consistent under H_0

$\hat{\bar{\beta}}_{FGLS}$, $\hat{\bar{\beta}}_{OLS}$ have different plims under H_1.

$$H = \left(\hat{\bar{\beta}}_{FGLS} - \hat{\bar{\beta}}_{OLS}\right)' \left(Var(\hat{\bar{\beta}}_{OLS}) - Var(\hat{\bar{\beta}}_{FGLS})\right)^{-1} \left(\hat{\bar{\beta}}_{FGLS} - \hat{\bar{\beta}}_{OLS}\right)$$

$$\xrightarrow{d} \chi^2\left(\sum^{m} k_i\right).$$