Moshe Buchinsky
Department of Economics
UCLA

Economics 203C
Spring 2003

Lecture Note 4

Multivariate Seemingly Unrelated Regression (SUR) Model

I. MODEL

System of equations:

$$y_{ij} = x'_{ij}\beta_j + \epsilon_{ij} \qquad (i = 1, \ldots, n,\ j = 1, \ldots, J).$$

That is, we have J equations, each of which has a set of explanatory variables x_{ij}, not necessarily the same, with x_{ij} a $k_j \times 1$ vector, and β_j is a $k_j \times 1$ vector of parameters.

Altogether we have a set of J linear models:

$$y_j = X_j\beta_j + \epsilon_j,$$

where

$$y_j = (y_{1j}, \ldots, y_{nj})', \quad X_j = (x_{1j}, \ldots, x_{nj})', \quad \epsilon_i = (\epsilon_{1j}, \ldots, \epsilon_{nj})'.$$

II. MODEL ASSUMPTIONS

1. $E[\epsilon_j \mid x_k] = 0,\ \forall j, k,$
2. $\mathrm{Var}(\epsilon_j \mid x_k) = E[\epsilon_j\epsilon'_j \mid x_k] = \sigma_{jj}I,\ \forall j, k,$
3. $\mathrm{Cov}(\epsilon_j, \epsilon_k \mid x_j, x_k) = \sigma_{jk}I, \quad \forall j, k.$

   Therefore,

$$E[\epsilon_i] = E\begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iJ} \end{bmatrix} = 0, \quad \text{and } \mathrm{Var}(\epsilon_i) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1J} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{J1} & \sigma_{J2} & \cdots & \sigma_{JJ} \end{pmatrix} = \Sigma.$$

4. X_j is random with $\mathrm{rank}(X_j) = k_j$ with probability one.

Note, the errors are not correlated across individuals $i = 1, \dots, n$, but they are correlated across equations for the same individual.

III. STACKED MODEL

Let

$$y = X\beta + \epsilon,$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix}, \quad X = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_J \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_J \end{pmatrix}.$$

That is,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_J \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_J \end{pmatrix}.$$

Note, y and ϵ are $nJ \times 1$ vectors, X is an $nJ \times (\sum_{j=1}^{J} k_j)$ matrix and β is a $(\sum_{j=1}^{J} k_j) \times 1$ vector of parameters. Let $k = \sum_{j=1}^{J} k_j$.

$$E[\epsilon \mid X] = 0, \quad \text{and } \mathrm{Var}(\epsilon \mid X) = \begin{pmatrix} \sigma_{11} I & \sigma_{12} I & \dots & \sigma_{1J} I \\ \sigma_{21} I & \sigma_{22} I & \dots & \sigma_{2J} I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{J1} I & \sigma_{J2} I & \dots & \sigma_{JJ} I \end{pmatrix} = \Omega = \Sigma \otimes I_J.$$

IV. KRONECKER PRODUCT—NOTATION AND RULES

If A is an $m \times n$ matrix with elements a_{ij} and B is an $k \times l$ matrix, then the **Kronecker product** of A and B, denoted by $A \otimes B$, is an $mk \times nl$ matrix given by

$$A \otimes B = \begin{pmatrix} a_{11} B & a_{12} B & \dots & a_{1n} B \\ a_{21} B & a_{22} B & \dots & a_{2n} B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} B & a_{m2} B & \dots & a_{mn} B \end{pmatrix}.$$

The following results hold (provided that all matrix are comparable and all products are well defined):

1. $(A \otimes B)(C \otimes D) = AC \otimes BD$.

2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, if $\det(A) \neq 0$, $\det(B) \neq 0$.

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

4. $A \otimes (B + C) = A \otimes B + A \otimes C$.

V. LEAST SQUARES ESTIMATOR

The least squares estimator is given by

$$\begin{aligned} b = \begin{pmatrix} b_1 \\ \vdots \\ b_J \end{pmatrix} &= (X'X)^{-1}X'y \\ &= \begin{pmatrix} X_1'X_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_J'X_J \end{pmatrix}^{-1} \begin{pmatrix} X_1'y_1 \\ \vdots \\ X_J'y_J \end{pmatrix} \\ &= \begin{pmatrix} (X_1'X_1)^{-1}X_1'y_1 \\ \vdots \\ (X_J'X_J)^{-1}X_J'y_J \end{pmatrix}. \end{aligned}$$

That is, $b_1, \dots, b_J$ are the same as the LS estimators when run equation-by-equation.

FINITE SAMPLE PROPERTIES

By the same type of arguments we used in previous class note,

$$E[b \mid X] = \begin{pmatrix} E[b_1 \mid X] \\ \vdots \\ E[b_j \mid X] \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_J \end{pmatrix},$$

and

$$
\begin{aligned}
\operatorname{Var}(b \mid X) &= \operatorname{Var}\left(\begin{array}{c} A_1 y_1 \\ \vdots \\ A_J y_J \end{array} \middle| X \right) \\
&= \begin{pmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_J \end{pmatrix} \operatorname{Var}(y \mid X) \begin{pmatrix} A_1' & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_J' \end{pmatrix} \\
&= \begin{pmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_J \end{pmatrix} \begin{pmatrix} \sigma_{11} I & \dots & \sigma_{1J} I \\ \vdots & \ddots & \vdots \\ \sigma_{J1} I & \dots & \sigma_{JJ} I \end{pmatrix} \begin{pmatrix} A_1' & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_J' \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11} A_1 A_1' & \dots & \sigma_{1J} A_1 A_J' \\ \vdots & \ddots & \vdots \\ \sigma_{J1} A_J A_1' & \dots & \sigma_{JJ} A_J A_J' \end{pmatrix}.
\end{aligned}
$$

So, for each $b_j,\ j = 1, \dots, J$ we get

$$\operatorname{Var}(b_j \mid X) = \sigma_{jj} (X_j' X_j)^{-1}$$

and

$$\widehat{\operatorname{Var}}(b_j \mid X) = s_{jj} (X_j' X_j)^{-1},$$

where

$$s_{jj} = \frac{1}{n - k_j} e_j' e_j \xrightarrow{p} \sigma_{jj}.$$

VI. GENERALIZED LEAST SQUARES (Σ KNOWN)

The GLS estimation procedure exploits the fact that $\sigma_{jk} \neq 0$. The GLS estimator is given by

$$\begin{aligned}
b_{GLS} &= \left(X'(\Sigma \otimes I)^{-1}X\right)^{-1} X'(\Sigma \otimes I)^{-1} y \\
&= \left(X'(\Sigma^{-1} \otimes I)X\right)^{-1} X'(\Sigma^{-1} \otimes I) y \\
&= \left\{ \begin{pmatrix} X_1' & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_J' \end{pmatrix} \begin{pmatrix} \sigma^{11} I & \dots & \sigma^{1J} I \\ \vdots & \ddots & \vdots \\ \sigma^{J1} I & \dots & \sigma^{JJ} I \end{pmatrix} \begin{pmatrix} X_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_J \end{pmatrix} \right\}^{-1} \\
&\quad \times \begin{pmatrix} X_1' & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_J' \end{pmatrix} \begin{pmatrix} \sigma^{11} I & \dots & \sigma^{1J} I \\ \vdots & \ddots & \vdots \\ \sigma^{J1} I & \dots & \sigma^{JJ} I \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix} \\
&= \begin{pmatrix} \sigma^{11} X_1' X_1 & \dots & \sigma^{1J} X_1' X_J \\ \vdots & \ddots & \vdots \\ \sigma^{J1} X_J' X_1 & \dots & \sigma^{JJ} X_J' X_J \end{pmatrix}^{-1} \begin{pmatrix} X_1' \sum_{j=1}^{J} \sigma^{1j} y_j \\ \vdots \\ X_J' \sum_{j=1}^{J} \sigma^{Jj} y_j \end{pmatrix},
\end{aligned}$$

where σ^{ij} denote the ij elements of Σ^{-1} ($\Sigma^{-1} = [\sigma^{ij}]$).

What is the relationship between b and b_{GLS}?

Specialize for $J = 2$. Then,

$$\begin{pmatrix} \sigma^{11}(X_1' X_1) & \sigma^{12}(X_1' X_2) \\ \sigma^{21}(X_2' X_1) & \sigma^{22}(X_2' X_2) \end{pmatrix} \begin{pmatrix} b_{1GLS} \\ b_{2GLS} \end{pmatrix} = \begin{pmatrix} \sigma^{11}(X_1' y_1) + \sigma^{12}(X_1' y_2) \\ \sigma^{21}(X_2' y_1) + \sigma^{22}(X_2' y_2) \end{pmatrix}.$$

Hence,

$$\sigma^{11}(X_1' X_1) b_{1GLS} + \sigma^{12}(X_1' X_2) b_{2GLS} = \sigma^{11}(X_1' y_1) + \sigma^{12}(X_1' y_2).$$

So,

$$\begin{aligned}
b_{1GLS} &= (X_1' X_1)^{-1} X_1' y_1 + \frac{\sigma^{12}}{\sigma^{11}} (X_1' X_1)^{-1} X_1' y_2 - \frac{\sigma^{12}}{\sigma^{11}} (X_1' X_1)^{-1} X_1' X_2 b_{2GLS} \\
&= b_1 + \frac{\sigma^{12}}{\sigma^{11}} A_1 (y_2 - X_2 b_{2GLS})
\end{aligned}$$

Similarly, we get that

$$b_{2GLS} = b_2 + \frac{\sigma^{21}}{\sigma^{22}} A_2 (y_1 - X_1 b_{1GLS}).$$

But,

$$\Sigma^{-1} = \frac{1}{\det(\Sigma)} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{pmatrix}.$$

$$\Longrightarrow \quad \frac{\sigma^{12}}{\sigma^{11}} = -\frac{\sigma_{12}}{\sigma_{22}} \quad \text{and} \quad \frac{\sigma^{21}}{\sigma^{22}} = -\frac{\sigma_{21}}{\sigma_{11}}.$$

So,

$$b_{1GLS} = b_1 - \alpha_1 A_1 (y_2 - X_2 b_{2GLS}),$$

$$b_{2GLS} = b_2 - \alpha_2 A_2 (y_1 - X_1 b_{1GLS}),$$

where $\alpha_1 = \sigma_{12}/\sigma_{22}$ and $\alpha_2 = \sigma_{21}/\sigma_{11}$.

INTERPRETATION: The GLS estimator is a combination of the LS estimator and a term which has zero expectation:

$$\begin{aligned} E[\alpha_1 A_1 (y_2 - X_2 b_{2GLS}) \mid X] &= E\left[\alpha_1 A_1 (X_2 b_{2GLS} + e_2 - X_2 b_{2GLS})\right] \\ &= E\left[\alpha_1 (X_1' X_1)^{-1} X_1' e_2\right] \\ &= E\left[\alpha_1 (X_1' X_1)^{-1} X_1' E\left[e_2 \mid X\right]\right] \\ &= 0. \end{aligned}$$

So, we get an estimator which is unbiased, since b_1 is unbiased, and it exploits the non-zero covariance between b_1 with an estimator of 0, to reduce its variance.

VII. SPECIAL CASES

1. ORTHOGONAL X_j's: $X_j' X_{j'} = 0$, if $j \neq j'$

$$\begin{aligned} b_{GLS} &= \begin{pmatrix} (\sigma^{11} X_1' X_1)^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (\sigma^{JJ} X_J' X_J)^{-1} \end{pmatrix} \begin{pmatrix} X_1' \sum_{j=1}^{J} \sigma^{1j} y_j \\ \vdots \\ X_J' \sum_{j=1}^{J} \sigma^{Jj} y_j \end{pmatrix}, \\ &= \begin{pmatrix} (X_1' X_1)^{-1} X_1' \sum_{j=1}^{J} \frac{\sigma^{1j}}{\sigma^{11}} y_j \\ \vdots \\ (X_J' X_J)^{-1} X_J' \sum_{j=1}^{J} \frac{\sigma^{Jj}}{\sigma^{JJ}} y_j \end{pmatrix} \\ &= \begin{pmatrix} b_1 + A_1 \sum_{j \neq 1} \frac{\sigma^{1j}}{\sigma^{11}} y_j \\ \vdots \\ b_J + A_J \sum_{j \neq J} \frac{\sigma^{Jj}}{\sigma^{JJ}} y_j \end{pmatrix}, \end{aligned}$$

where $A_j = (X_j' X_j)^{-1} X_j'$. That is, b_{jGLS} is equal to b_j plus the sum of LS estimators from the regressions of $y_1, \dots, y_J$ on X_j, weighted by σ^{jl}/σ^{jj}.

2. IDENTICAL EXPLANATORY VARIABLES: $X_j = X_0,\ \forall j = 1, \dots, J.$

$$X = \begin{pmatrix} X_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_0 \end{pmatrix}.$$

$$\begin{aligned}
\Longrightarrow \quad b_{GLS} &= \Big(X'(\Sigma \otimes I)^{-1} X\Big)^{-1} X'(\Sigma \otimes I)^{-1} y \\
&= \Big((I \otimes X_0')(\Sigma^{-1} \otimes I)(I \otimes X_0)\Big)^{-1} (I \otimes X_0')(\Sigma^{-1} \otimes I) y \\
&= (\Sigma^{-1} \otimes X_0' X_0)^{-1} (\Sigma^{-1} \otimes X_0') y \\
&= (\Sigma \otimes (X_0' X_0)^{-1})(\Sigma^{-1} \otimes X_0') y \\
&= (I \otimes (X_0' X_0)^{-1} X_0') y \\
&= \begin{pmatrix} (X_0' X_0)^{-1} X_0' y_1 \\ \vdots \\ (X_0' X_0)^{-1} X_0' y_J \end{pmatrix} \\
&= \begin{pmatrix} b_1 \\ \vdots \\ b_J \end{pmatrix} \\
&= b.
\end{aligned}$$

That is, the GLS estimator is identical to the LS estimator, regardless of the structure of Σ!

3. UNCORRELATED DISTURBANCE:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{JJ} \end{pmatrix}.$$

$$\begin{aligned}
b_{GLS} &= \begin{pmatrix} (\sigma^{11} X_1' X_1)^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (\sigma^{JJ} X_J' X_J)^{-1} \end{pmatrix} \begin{pmatrix} \sigma^{11} X_1' y_1 \\ \vdots \\ \sigma^{JJ} X_1' y_J \end{pmatrix} \\
&= \begin{pmatrix} (X_1' X_1)^{-1} X_1' y_1 \\ \vdots \\ (X_J' X_J)^{-1} X_J' y_J \end{pmatrix} \\
&= b.
\end{aligned}$$

That is, the GLS and LS estimators are identical as in special case 2, although for a different reason.

VIII. ASYMPTOTIC DISTRIBUTION OF LS AND GLS ESTIMATORS

1. LS ESTIMATOR:

$$b = (X'X)^{-1}X'y.$$

$$\Longrightarrow \quad \sqrt{n}(b-\beta) \xrightarrow{D} N(0, \Delta^{-1}V\Delta^{-1}),$$

where

$$\Delta = \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}X'X\right) \quad \text{and } V = \operatorname*{plim}_{n\to\infty}{}_{n\to\infty}\left(\frac{1}{n}X'(\Sigma\otimes I)X\right).$$

2. GLS ESTIMATOR:

$$b_{GLS} = \left(X'(\Sigma\otimes I)^{-1}X\right)^{-1}X'(\Sigma\otimes I)^{-1}y.$$

$$\Longrightarrow \quad \sqrt{n}(b_{GLS}-\beta) \xrightarrow{D} N(0, \Lambda^{-1}),$$

where

$$\begin{aligned}\Lambda &= \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}X'(\Sigma\otimes I)^{-1}X\right)\\ &= \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}X'(\Sigma^{-1}\otimes I)X\right).\end{aligned}$$

3. Asymptotic Covariance Matrices: Note, in special cases 2 and 3, the asymptotic covariance matrix simplifies to:

a. $X_j = X_0, \forall j = 1, \ldots, J$:

$$\text{Var}(b) = \Sigma\otimes\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}X_0'X_0\right)^{-1}.$$

b. $\Sigma = \text{diag}(\sigma_{11}, \ldots, \sigma_{JJ})$:

$$\text{Var}(b) = \begin{pmatrix}\sigma_{11}V_1^{-1} & \ldots & 0\\ \vdots & \ddots & \vdots\\ 0 & \ldots & \sigma_{JJ}V_J^{-1}\end{pmatrix},$$

where

$$V_j = \operatorname*{plim}_{n\to\infty}\frac{1}{n}(X_j'X_j),\ j = 1, \ldots, J.$$

IX. FEASIBLE GLS (FGLS)

Let $e_i = (e_{i1}, \ldots, e_{iJ}),\ i = 1, \ldots, n$ be the residuals $J \times 1$ vectors of the n observations across the J equations.

Define

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} e_i e_i'.$$

That is,

$$\hat{\sigma}_{jk} = \frac{1}{n} e_j' e_k, \qquad e_j \equiv y_j - X_j b_j.$$

It is easy to show that

$$\hat{\Sigma} \xrightarrow{p} \Sigma \quad \text{as } n \to \infty$$

by the usual arguments. Then,

$$b_{FGLS} = \left(X'(\hat{\Sigma}^{-1} \otimes I)X \right)^{-1} X'(\hat{\Sigma}^{-1} \otimes I)y.$$

It is easy to show that

$$\sqrt{n}(b_{FGLS} - b_{GLS}) \xrightarrow{p} 0,$$

under usual conditions. Hence,

$$\sqrt{n}(b_{FGLS} - \beta) \xrightarrow{D} N(0, \Lambda^{-1}),$$

where Λ is as defined before.

X. EXTENSION TO SUR MODEL (e.g., Judge et. al—Chapter 6)

1. UNBALANCED SAMPLE:

   $(y_{ij}, x_{ij}'),\ i = 1, \ldots, n_j,\ j = 1, \ldots, J.$ Note that n_j is different for different equations.

   The GLS estimator is the same if Σ is known. If Σ is unknown, there are several versions for FGLS (all asymptotically equivalent), based on alternative estimators $\hat{\Sigma}$.

2. COMMON REGRESSORS:

   $x_{ij}' = (x_{i1}', z_j'),\ j = 2, \ldots, J.$ That is x_{i1} is a subset of each $x_{ij},\ j = 2, \ldots, J.$

   The result is that

   $$b_{1GLS} = b_1 \quad \text{but } b_{jGLS} \neq b_j.$$

   That is, for the first equation it does not matter if we use the LS or the GLS, the resulting estimator is the same. For all other coefficients it may matter, if $\mathrm{Cov}(x_{i1}, z_j) \neq 0$.

3. SINGULAR COVARIANCE MATRIX:

   Say $\text{rank}(\Sigma) = J - 1$. This problem arises a lot in economic modeling. For example, y_{ij} may be budget shares, so $\sum_{j=1}^{J} y_{ij} = 1, \forall i$. Hence, $\sum_{j=1}^{J} e_{ij} = 0$.

   To correct, need to drop one equation and proceed with the remaining $J - 1$ equations as before. This would be the same as using a generalized inverse matrix Σ^- of Σ.

   (A^- is a unique generalized inverse of A if: (i) $AA^-A = A$; (ii) $A^-AA^- = A^-$; (iii) $(AA^-)' = AA^-$; and (iv) $(A^-A)' = A^-A$.)

4. SERIALLY CORRELATED ERRORS:

$$\epsilon_{tj} = \rho_j \epsilon_{t-1,j} + u_{tj}, \quad u \sim (0, \Sigma \otimes I).$$

$$\Longrightarrow \quad \epsilon \sim (0, \Omega),$$

   where

$$\Omega = [\Omega_{jk}], \qquad \Omega_{jk} = \frac{\sigma_{jk}}{1 - \rho_j \rho_k} \begin{pmatrix} 1 & \rho_k & \rho_k^2 & \dots & \rho_k^{T-1} \\ \rho_j & 1 & \rho_k & \dots & \rho_k^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T-2} & \dots & \rho_j & 1 & \rho_k \\ \rho_j^{T-1} & \dots & \rho_j^2 & \rho_j & 1 \end{pmatrix}.$$

   One can do generalized differencing and then SUR on the transformed data.