14.383

Fall 2003

Handout #3

TA: Alexis León

aleon@mit.edu

E52-351

(Office hours: Tue., 4pm
in E52-398)

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)

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

$$\underset{10\times 1}{y^I} = \underset{10\times K_I}{X^I}\,\underset{K_I\times 1}{\beta^I} + \underset{10\times 1}{\varepsilon^I} \; ; \quad \underset{10\times 1}{y^M} = \underset{10\times K_M}{X^M}\,\underset{K_M\times 1}{\beta^M} + \underset{10\times 1}{\varepsilon^M}$$

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.

$$\bar{y} = \begin{bmatrix} y^1 \\ \hline \vdots \\ \hline y^m \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} y_1^1 \\ \vdots \\ y_T^1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y_1^m \\ \vdots \\ y_T^m \end{pmatrix} \end{bmatrix} ; \quad \bar{X} = \begin{bmatrix} X^1 & 0 & \cdots & 0 \\ 0 & X^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & X^m \end{bmatrix}$$

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

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

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

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

<u>Kronecker Products</u>: Recall that

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

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

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

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

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

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

$\bar{\varepsilon}$ 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've 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

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

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

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, ($\sigma_{ij} = 0 \;\; \forall\, i \neq j$), then GLS = OLS. However, FGLS will not be numerically equal to OLS.
2. If the equations have identical explanatory variables ($X^i = X^j \;\; \forall\, i,j$), 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 $V(\varepsilon^{j}) = \sigma_{jj} I_T$, $E(\varepsilon^{j}\varepsilon^{k\prime}) = \sigma_{jk} I_T$) 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):

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

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

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

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

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

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

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

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

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

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

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.

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

$$= \frac{1}{T}\underset{(\sum_{j=1}^{m} k_j \times Tm)}{\begin{bmatrix} S_{11}X^{1'} & S_{12}X^{1'} & \cdots & S_{1m}X^{1'} \\ S_{21}X^{2'} & S_{22}X^{2'} & \cdots & S_{2m}X^{2'} \\ \vdots & & \ddots & \vdots \\ S_{m1}X^{m'} & S_{m2}X^{m'} & \cdots & S_{mm}X^{m'} \end{bmatrix}}\underset{(Tm\times 1)}{\begin{bmatrix} \varepsilon^1 \\ \varepsilon^2 \\ \vdots \\ \varepsilon^m \end{bmatrix}}$$

$$= \underset{(\sum_{j=1}^{m} k_j \times 1)}{\begin{bmatrix} \frac{1}{T}(S_{11}X^{1'}\varepsilon^1 + S_{12}X^{1'}\varepsilon^2 + \cdots + S_{1m}X^{1'}\varepsilon^m) \\ \frac{1}{T}(S_{21}X^{2'}\varepsilon^1 + S_{22}X^{2'}\varepsilon^2 + \cdots + S_{2m}X^{2'}\varepsilon^m) \\ \vdots \\ \frac{1}{T}(S_{m1}X^{m'}\varepsilon^1 + S_{m2}X^{m'}\varepsilon^2 + \cdots + S_{mm}X^{m'}\varepsilon^m) \end{bmatrix}}\begin{matrix} \}k_1 \\ \}k_2 \\ \vdots \\ \}k_m \end{matrix}$$

Trade-off (Cont)

Consistency of $\hat{\hat{\beta}}_{FGLS}$ 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 $\frac{1}{T} X^{2\prime} \varepsilon^j \not\to 0$, 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: $\text{plim } \frac{1}{T} X^{j\prime} \varepsilon^k = 0 \quad \forall \ j, k$

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

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

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

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

$$\xrightarrow{d} \chi^2(\textstyle\sum k_i).$$

Three-Stage Least Squares (3SLS)

$$\underset{T\times 1}{y^j} = \underset{T\times r_j}{Y^j}\,\underset{r_j\times 1}{\delta^j} + \underset{T\times s_j}{Z^j}\,\underset{s_j\times 1}{\gamma^j} + \underset{T\times 1}{\varepsilon^j}, \quad j=1,\dots,M$$

$$= \underset{T\times k_j}{[Y^j \;\; Z^j]}\underset{k_j\times 1}{\begin{bmatrix}\delta^j\\ \gamma^j\end{bmatrix}} + \underset{T\times 1}{\varepsilon^j}, \quad k_j = r_j + s_j$$

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

This is the same setup as the Seemingly Unrelated Regressors model, but now X^j includes a set of variables Y^j that are endogenous.

This is our simultaneous equations system with normalization and exclusion restrictions.

Let's stack the equations, as before:

$$\underset{(TM\times \sum_{j=1}^{m} k_j)}{\bar{X} = \begin{bmatrix} X^1 & & 0 \\ & X^2 & \\ 0 & & \ddots \;\; X^m \end{bmatrix}}, \quad \underset{TM\times 1}{\bar{y} = \begin{bmatrix} y^1 \\ \vdots \\ y^m \end{bmatrix}}, \quad \underset{TM\times 1}{\bar{\varepsilon} = \begin{bmatrix} \varepsilon^1 \\ \vdots \\ \varepsilon^m \end{bmatrix}}, \quad \underset{\sum_{j=1}^{m} k_j \times 1}{\bar{\beta} = \begin{bmatrix} \beta^1 \\ \vdots \\ \beta^m \end{bmatrix}}$$

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

To do system IV, we need an instrumental variable $\underset{TM\times \sum_{j=1}^{m} k_j}{\bar{W}}$ so that $\hat{\beta}_{IV} = (\bar{W}'\bar{X})^{-1}(\bar{W}'\bar{y})$

Where does $\bar{W}$ come from? Note that

$$(I_m \otimes Z)'\bar{\varepsilon}/T = \frac{1}{T}\begin{bmatrix} Z' & & 0 \\ & Z' & \\ 0 & & \ddots & Z' \end{bmatrix}\begin{bmatrix} \varepsilon' \\ \vdots \\ \varepsilon^m \end{bmatrix} = \begin{bmatrix} \frac{1}{T}Z'\varepsilon' \\ \vdots \\ \frac{1}{T}Z'\varepsilon^m \end{bmatrix} \xrightarrow{p} 0$$

under the orthogonality assumption. So,

$\underset{\substack{m\times m \\ (Tm\times km)}}{I_m} \otimes \underset{T\times k}{Z}$ is the set of instruments to be used to form $\bar{W}$.

We need $mk \geq \sum_{i=1}^{m} k_j$. This holds if each equation is identified since the order condition for the i^{th} equation is $k \geq q_i$.

$$\underset{Tm\times\sum_{j=1}^m k_j}{\bar{W}} = \underset{Tm\times km}{(I \otimes Z)}\ \underset{km\times\sum_{j=1}^m k_j}{\bar{A}}.$$

What is the best $\bar{A}$?

$$\bar{A} = (\hat{\Sigma}^{-1} \otimes I)\cdot(I \otimes (Z'Z)^{-1}Z')\bar{X}$$

where

$Var(\bar{\varepsilon}) = (\Sigma \otimes I)$ and $(\hat{\Sigma}^{-1} \otimes I)$ is the inverse of the estimate of $Var(\bar{\varepsilon})$.

$$\bar{W} = (I \otimes Z)\bar{A} = (\hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z')\bar{X}$$
$$= (\hat{\Sigma}^{-1} \otimes P_Z)\bar{X}$$

(1) $$\hat{\beta}_{3SLS} = [\bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z)\bar{X}]^{-1}[\bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z)\bar{y}]$$

Let's step back a minute and find the 3SLS estimator in a different way, relating it more to SUR.

Consider the IV estimator formed from

$$\bar{W} = \hat{\bar{X}} = \begin{bmatrix} Z(Z'Z)^{-1}Z'X^1 & 0 & \cdots & 0 \\ 0 & Z(Z'Z)^{-1}Z'X^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & Z(Z'Z)^{-1}Z'X^m \end{bmatrix}$$

$$= \begin{bmatrix} \hat{X}^1 & & 0 \\ & \hat{X}^2 & \\ 0 & & \ddots \ \hat{X}^m \end{bmatrix}$$

The IV estimator $\hat{\bar{\beta}}_{IV} = (\hat{\bar{X}}'\bar{X})^{-1}(\hat{\bar{X}}'\bar{y})$ is nothing more than equation by equation 2SLS. By analogy with the SUR model, we would expect this estimator to be less efficient than a GLS estimator. A natural candidate is

$$\hat{\bar{\beta}}_{3SLS,I} = (\hat{\bar{X}}'(\Sigma^{-1}\otimes I)\bar{X})^{-1}(\hat{\bar{X}}'(\Sigma^{-1}\otimes I)\bar{y})$$

Infeasible since Σ not known.

Plugging in an estimate for Σ gives

$$(2) \quad \hat{\bar{\beta}}_{3SLS} = (\hat{\bar{X}}'(\hat{\Sigma}^{-1}\otimes I)\bar{X})^{-1}(\hat{\bar{X}}'(\hat{\Sigma}^{-1}\otimes I)\bar{y})$$

Noting that $\hat{\bar{X}}'(\hat{\Sigma}^{-1} \otimes I) = ((I \otimes P_Z)\bar{X})'(\hat{\Sigma}^{-1} \otimes I)$
$= \bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z') = \bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z)$, the 3SLS estimator in Eqns (1) and (2) are identical.

The three-stages of computing the 3SLS estimator are:

1. Compute $\hat{\bar{X}}$ by regressing each component of $\bar{X}$ on Z:
$$\hat{\bar{X}} = \begin{bmatrix} Z(Z'Z)^{-1}Z'X^1 & & 0 \\ & \ddots & \\ 0 & & Z(Z'Z)^{-1}Z'X^m \end{bmatrix}$$

2. Compute $\hat{\beta}^j_{2SLS}$ for each equation j by doing 2SLS on each equation.
$$\begin{pmatrix} \hat{\beta}^1_{2SLS} \\ \vdots \\ \hat{\beta}^m_{2SLS} \end{pmatrix} = \hat{\bar{\beta}}_{2SLS} = (\hat{\bar{X}}'\bar{X})^{-1}(\hat{\bar{X}}'y)$$
Define $S_{ij} = \dfrac{\hat{\varepsilon}^{i\prime}\hat{\varepsilon}^j}{T}$ where $\hat{\varepsilon}^i$ are the residuals from equation i, $\hat{\varepsilon}^i = y^i - X^i\hat{\beta}^i_{2SLS}$

3. $$\hat{\Sigma} = S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & & \vdots \\ \vdots & & \ddots & \\ s_{m1} & & \cdots & s_{mm} \end{pmatrix}$$
Compute $\hat{\bar{\beta}}_{3SLS} = [\bar{X}'(S^{-1} \otimes P_Z)\bar{X}]^{-1}[\bar{X}'(S^{-1} \otimes P_Z)\bar{y}]$

Optimality of 3SLS

Let's find the variance of the general IV estimator using $\bar{W} = (I \otimes Z)\bar{A}$

$$\sqrt{T}(\hat{\bar{\beta}}_{IV} - \bar{\beta}) = \left[\frac{\bar{A}'(I \otimes Z)'\bar{X}}{T}\right]^{-1}\left[\frac{\bar{A}'(I \otimes Z)'\bar{\varepsilon}}{\sqrt{T}}\right]$$

$$(I \otimes Z)'\bar{\varepsilon}/\sqrt{T} \xrightarrow{d} N\left(0, \operatorname{plim} \frac{(I \otimes Z)'(\Sigma \otimes I)(I \otimes Z)}{T}\right)$$

$$\xrightarrow{d} N\left(0, \operatorname{plim} \Sigma \otimes \left(\frac{Z'Z}{T}\right)\right)$$

$$\xrightarrow{d} N(0, \Sigma \otimes M_Z), \quad M_Z = \operatorname{plim}\left(\frac{Z'Z}{T}\right)$$

$$\frac{\bar{A}'(I \otimes Z)'\bar{X}}{T} = \bar{A}' \begin{bmatrix} \frac{Z'X^1}{T} & & & 0 \\ & \frac{Z'X^2}{T} & & \\ & & \ddots & \\ 0 & & & \frac{Z'X^n}{T} \end{bmatrix}$$

$$\frac{Z'X^j}{T} = \frac{1}{T} Z'[Y^j \; Z^j] = \frac{1}{T} Z'[Z\Pi_j + V_j, \; Z^j]$$

$$\xrightarrow{p} M_Z [\Pi_j \; C_j]$$

where Π_j is the matrix of coefficients given from the reduced form for $Y^j = Z\Pi_j + V_j$ and where C_j is a selection matrix that selects from M_Z the part that represents the plim of $\frac{1}{T} Z'Z^j$.

Optimality of 3SLS (Cont)

Let $D_j = [\pi_j \quad C_j]$, $\quad A = \text{plim}\ \bar{A}$.

Then $\dfrac{\bar{A}'(I \otimes Z)'\bar{X}}{T} \xrightarrow{p} A' \begin{bmatrix} M_Z D_1 & & 0 \\ & M_Z D_2 & \\ 0 & & \ddots \ M_Z D_m \end{bmatrix}$

$$= A'(I \otimes M_Z) \begin{bmatrix} D_1 & & 0 \\ & D_2 & \\ 0 & & \ddots \ D_m \end{bmatrix}$$

$$= A'(I \otimes M_Z)\bar{D}$$

Putting this together, we have

$$\sqrt{T}(\hat{\bar{\beta}}_{IV} - \bar{\beta}) \xrightarrow{d} (A'(I \otimes M_Z)\bar{D})^{-1} A' N(0, \Sigma \otimes M_Z)$$

so that

$$AVar(\hat{\bar{\beta}}_{IV}) = (A'(I \otimes M_Z)\bar{D})^{-1}[A'(\Sigma \otimes M_Z)A](\bar{D}'(I \otimes M_Z)A)^{-1}$$

For 3SLS,

$$\bar{A} = (\hat{\Sigma}^{-1} \otimes I) \begin{bmatrix} (Z'Z)^{-1}Z'X^1 & & 0 \\ 0 & \ddots & (Z'Z)^{-1}Z'X^m \end{bmatrix}$$

$$= (\hat{\Sigma}^{-1} \otimes I) \begin{bmatrix} \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'X^1}{T}\right) & & 0 \\ & \ddots & \\ 0 & & \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'X^m}{T}\right) \end{bmatrix}$$

$$\xrightarrow{p} (\Sigma^{-1} \otimes I) \begin{bmatrix} M_Z^{-1}M_Z D_1 & & 0 \\ & \ddots & \\ 0 & & M_Z^{-1}M_Z D_m \end{bmatrix} = (\Sigma^{-1} \otimes I) \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_m \end{pmatrix}$$

Optimality of 3SLS (Cont)

For 3SLS,

$$\bar{A} \xrightarrow{p} (\Sigma^{-1} \otimes I)\bar{D}$$

Plugging in this plim for $AVar(\hat{\bar{\beta}}_{IV})$ gives

$$AVar(\hat{\bar{\beta}}_{3SLS}) = (\bar{D}'(\Sigma^{-1}\otimes Q)\bar{D})^{-1}(\bar{D}'(\Sigma^{-1}\otimes Q)\bar{D})(\bar{D}'(\Sigma^{-1}\otimes Q)\bar{D})^{-1}$$
$$= [\bar{D}'(\Sigma^{-1}\otimes Q)\bar{D}]^{-1}$$

The final step to showing optimality is to show that

$$AVar(\hat{\bar{\beta}}_{IV}) - AVar(\hat{\bar{\beta}}_{3SLS}) \text{ is p.s.d.}$$

This is left as an (unnecessary) excercise.

Comparing 2SLS, 3SLS

- When all equations are exactly identified, 2SLS $\equiv$ 3SLS. To see this, note:

$$\hat{\bar{\beta}}_{3SLS} = [\underbrace{\bar{X}'(\hat{\Sigma}^{-1}\otimes P_Z)\bar{X}}]^{-1}[\bar{X}'(\hat{\Sigma}^{-1}\otimes P_Z)\bar{y}]$$

$$\downarrow$$

$$\bar{X}'(I\otimes Z)'(\hat{\Sigma}^{-1}\otimes I)(I\otimes(Z'Z)^{-1})(I\otimes Z')\bar{X}$$

$$= \underbrace{\begin{bmatrix} x^{1\prime}z & & 0 \\ & x^{2\prime}z & \\ 0 & & \ddots \; x^{n\prime}z \end{bmatrix}}(\hat{\Sigma}^{-1}\otimes I)(I\otimes(Z'Z)^{-1})\underbrace{\begin{bmatrix} z'x^1 & & 0 \\ & z'x^2 & \\ 0 & & \ddots \; z'x^n \end{bmatrix}}$$

Each of these blocks is square and invertible.

Comparing 2SLS, 3SLS (Cont)

So, when all equations are exactly identified,

$$\hat{\beta}_{3SLS} = \begin{bmatrix} Z'X^1 & & 0 \\ & \ddots & \\ 0 & & Z'X^m \end{bmatrix}^{-1} (I \otimes (Z'Z)) (\hat{\Sigma} \otimes I) \begin{bmatrix} X^{1\prime}Z & & 0 \\ & \ddots & \\ 0 & & X^{m\prime}Z \end{bmatrix}^{-1}$$

$$\cdot \begin{bmatrix} X^{1\prime}Z & & 0 \\ & \ddots & \\ 0 & & X^{m\prime}Z \end{bmatrix} (\hat{\Sigma}^{-1} \otimes I)(I \otimes (Z'Z)^{-1}) \begin{bmatrix} Z'y^1 \\ \vdots \\ Z'y^m \end{bmatrix}$$

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

$$= \begin{bmatrix} (Z'X^1)^{-1} Z'y^1 \\ \vdots \\ (Z'X^m)^{-1} Z'y^m \end{bmatrix} = \text{2SLS eqn. by eqn.}$$

Also, as in the SUR model, you can see that misspecification in some equation effects in general the consistency in all equations. To see this, expand the term

$$\bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z)\,\bar{\varepsilon}$$

and argue as we did for the SUR estimator.