Multiple Equation Linear GMM

Eric Zivot

Winter 2013

Multiple Equation Linear GMM

Notation

$$y_{iM},\ i = \text{individual};\ M = \text{ equation}$$

There are M linear equations,

$$y_{im} = \underset{(1\times L_m)}{\mathbf{z}'_{im}}\underset{(L_m\times 1)}{\boldsymbol{\delta}_m} + \varepsilon_{im},\ i = 1,\ldots,n;\ m = 1,\ldots,M;$$

In matrix notation

$$\underset{n\times 1}{\mathbf{y}_m} = \underset{n\times L_m}{\mathbf{Z}_m}\underset{L_m\times 1}{\boldsymbol{\delta}_m} + \underset{n\times 1}{\boldsymbol{\varepsilon}_m}$$

Remarks:

1. Balanced (square) system. Same number of observations, n, in each equation

2. No a priori assumptions about cross equation error correlation or homoskedasticity

3. No cross equation parameter restrictions on $\boldsymbol{\delta}_m$ $(m = 1, \ldots, M)$

4. There could be variables in common across equations

Giant Regression Representation

$$\begin{bmatrix} \underset{n\times 1}{\mathbf{y}_1} \\ \vdots \\ \underset{n\times 1}{\mathbf{y}_M} \end{bmatrix} = \begin{bmatrix} \underset{n\times L_1}{\mathbf{Z}_1} & & \\ & \ddots & \\ & & \underset{n\times L_M}{\mathbf{Z}_M} \end{bmatrix} \begin{bmatrix} \underset{L_1\times 1}{\boldsymbol{\delta}_1} \\ \vdots \\ \underset{L_M\times 1}{\boldsymbol{\delta}_M} \end{bmatrix} + \begin{bmatrix} \underset{n\times 1}{\boldsymbol{\varepsilon}_1} \\ \vdots \\ \underset{n\times 1}{\boldsymbol{\varepsilon}_M} \end{bmatrix}$$

or

$$\begin{aligned} \underset{nM\times 1}{\underline{\mathsf{y}}} &= \underset{nM\times L}{\underline{\underline{\mathsf{Z}}}}\,\underset{L\times 1}{\boldsymbol{\delta}} + \underset{nM\times 1}{\underline{\mathsf{e}}} \\ L &= \sum_{m=1}^{M} L_m \end{aligned}$$

Main Issues and Questions:

1. Why not just estimate each equation separately?

   (a) Joint estimation may improve efficiency, but...

   (b) Joint estimation is sensitive to misspecification of individual equations

2. Theory may provide cross equation restrictions

   (a) Improve efficiency if restriction are imposed during estimation

   (b) test restrictions imposed by theory

Example (Hayashi): 2 equation wage equation

$$
\begin{aligned}
LW_i &= \phi_1 + \beta_1 S_i + \gamma_1 IQ_i + \pi_1 EXPR_i + \varepsilon_{i1},\ L_1 = 4 \\
&= \mathbf{z}_{i1}' \boldsymbol{\delta}_1 + \varepsilon_{i1} \\
KWW_i &= \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \varepsilon_{i2},\ L_2 = 3 \\
&= \mathbf{z}_{i2}' \boldsymbol{\delta}_2 + \varepsilon_{i2} \\
\mathbf{z}_{i1} &= (1, S_i, IQ_i, EXPR_i)',\ \boldsymbol{\delta}_1 = (\phi_1, \beta_1, \gamma_1, \pi_1)' \\
\mathbf{z}_{i2} &= (1, S_i, IQ_i)',\ \boldsymbol{\delta}_2 = (\phi_2, \beta_2, \gamma_2)'
\end{aligned}
$$

Note, ε_{i1} and ε_{i2} may be correlated (eg. due to common omitted variable ability); i.e.,

$$E[\varepsilon_{i1}\varepsilon_{i2}] \neq 0$$

Example (Hayashi): Panel data for wage equation

$$\begin{aligned} LW69_i &= \phi_1 + \beta_1 S_i + \gamma_1 IQ_i + \pi_1 EXPR_i + \varepsilon_{i1}, \\ LW80_i &= \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \pi_2 EXPR_i + \varepsilon_{i2}, \end{aligned}$$

If all coefficients do not change over time then

$$\begin{aligned} \phi_1 &= \phi_2, \beta_1 = \beta_2, \gamma_1 = \gamma_2, \pi_1 = \pi_2 \\ \varepsilon_{im} &= \alpha_i + \eta_{im} \\ \alpha_i &= \text{unobserved individual fixed effect} \end{aligned}$$

If α_i is uncorrelated with S_i, IQ_i and $EXPR_i$ then we have the *random effects* set-up

If α_i is correlated with S_i, IQ_i and $EXPR_i$ then we have the *fixed effects* set-up

Instruments

$$\begin{aligned}
\underset{(K_m \times 1)}{\mathbf{x}_{im}} &= \text{instruments for } m\text{th equation} \\
E[\mathbf{x}_{im}\varepsilon_{im}] &= \mathbf{0},\ m = 1, \ldots, M \\
&\Rightarrow K = \sum_{m=1}^{M} K_m \text{ orthogonality conditions}
\end{aligned}$$

Note: We are not assuming cross-equation orthogonality conditions. That is, we may have

$$E[\mathbf{x}_{im}\varepsilon_{ik}] \neq 0 \text{ for } m \neq k$$

unless $\mathbf{x}_{im}$ and $\mathbf{x}_{ik}$ have variables in common.

Example (Hayashi): 2 equation wage equation

Assume

1. IQ is endogenous in both equations

2. EXP, MED and S are exogenous in both equations (MED is the excluded exog variable)

Then both equations have the same set of instruments

$$\begin{aligned} E[MED_i \varepsilon_{i1}] &= 0, \; E[MED_i \varepsilon_{i2}] = 0 \\ \mathbf{x}_i &= \mathbf{x}_{i1} = \mathbf{x}_{i2} = (1, S_i, EXPR_i, MED_i)' \end{aligned}$$

GMM Moment Conditions and Identification

Define

$$\underset{(L\times 1)}{\boldsymbol{\delta}} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_M')', \ L = \sum_{m=1}^{M} L_m$$

$$\underset{K\times 1}{\mathbf{g}_i(\boldsymbol{\delta})} = \begin{bmatrix} \mathbf{g}_{i1}(\boldsymbol{\delta}_1) \\ \vdots \\ \mathbf{g}_{iM}(\boldsymbol{\delta}_M) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{i1}\varepsilon_{i1} \\ \vdots \\ \mathbf{x}_{iM}\varepsilon_{iM} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{i1}(y_{i1} - \mathbf{z}_{i1}'\boldsymbol{\delta}_1) \\ \vdots \\ \mathbf{x}_{iM}(y_{iM} - \mathbf{z}_{iM}'\boldsymbol{\delta}_M) \end{bmatrix}$$

Then there are K linear moment equations such that

$$E[\mathbf{g}_i(\boldsymbol{\delta})] = \mathbf{0} \text{ (Here } \boldsymbol{\delta} \text{ is true value)}$$
$$E[\mathbf{g}_i(\tilde{\boldsymbol{\delta}})] \neq \mathbf{0} \text{ for } \tilde{\boldsymbol{\delta}} \neq \boldsymbol{\delta}$$

Now,

$$
\begin{aligned}
E[\mathbf{g}_i(\boldsymbol{\delta})] &= \begin{bmatrix} E[\mathbf{x}_{i1}y_{i1}] \\ \vdots \\ E[\mathbf{x}_{iM}y_{iM}] \end{bmatrix} - \begin{bmatrix} E[\mathbf{x}_{i1}\mathbf{z}'_{i1}]\boldsymbol{\delta}_1 \\ \vdots \\ E[\mathbf{x}_{iM}\mathbf{z}'_{iM}]\boldsymbol{\delta}_M \end{bmatrix} \\
&= \begin{bmatrix} E[\mathbf{x}_{i1}y_{i1}] \\ \vdots \\ E[\mathbf{x}_{iM}y_{iM}] \end{bmatrix} - \begin{bmatrix} E[\mathbf{x}_{i1}\mathbf{z}'_{i1}] & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & E[\mathbf{x}_{iM}\mathbf{z}'_{iM}] \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} \\
&= \underset{(K\times 1)}{\boldsymbol{\sigma}_{xy}} - \underset{(K\times L)}{\boldsymbol{\Sigma}_{xz}} \underset{(L\times 1)}{\boldsymbol{\delta}}
\end{aligned}
$$

Note: The above moment conditions for each equation are the same as the moment conditions we derived for the single equation linear GMM model.

If $K_m = L_m$ for $m = 1, \ldots, M$ then $\boldsymbol{\delta}$ is exactly identified and

$$\begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1 z_1}^{-1} \boldsymbol{\sigma}_{x_1 y_1} \\ \vdots \\ \boldsymbol{\Sigma}_{x_M z_M}^{-1} \boldsymbol{\sigma}_{x_M y_M} \end{bmatrix}$$

If $K_m > L_m$ for some m then to solve

$$E[\mathbf{g}_i(\boldsymbol{\delta})] = \underset{(K\times 1)}{\boldsymbol{\sigma}_{xy}} - \underset{(K\times L)}{\boldsymbol{\Sigma}_{xz}} \underset{(L\times 1)}{\boldsymbol{\delta}} = \mathbf{0}$$

we require the rank condition

$$\text{rank } \underset{(K_m\times L_m)}{\boldsymbol{\Sigma}_{x_m z_m}} = L_m \text{ for } m = 1, \ldots, M$$

That is, the single equation rank condition must hold for each equation.

Asymptotics

1. Let $\mathbf{w}_i$ denote the unique and nonconstant elements of $y_{i1}, \ldots, y_{iM}$, $z_{i1}, \ldots, z_{iM}$, $x_{i1}, \ldots, x_{iM}$.

2. Assume $\{\mathbf{w}_i\}$ is jointly ergodic-stationary such that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \xrightarrow{p} E[\mathbf{w}_i]$$

3. Assume $\{\mathbf{g}_i, I_i\}$ is an ergodic-stationary MDS satisfying

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{g}_i(\boldsymbol{\delta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$$

$$\underset{K\times K}{\mathbf{S}} = E[\mathbf{g}_i(\boldsymbol{\delta})\mathbf{g}_i(\boldsymbol{\delta})']$$

Note: $\mathbf{g}_i(\boldsymbol{\delta}) = (\mathbf{x}'_{i1}\varepsilon_{i1}, \ldots, \mathbf{x}'_{iM}\varepsilon_{iM})'$ and so

$$\begin{aligned}
\mathbf{S} &= E[\mathbf{g}_i(\boldsymbol{\delta})\mathbf{g}_i(\boldsymbol{\delta})'] \\
&= \begin{bmatrix} E[\mathbf{x}_{i1}\mathbf{x}'_{i1}\varepsilon^2_{i1}] & E[\mathbf{x}_{i1}\mathbf{x}'_{i2}\varepsilon_{i1}\varepsilon_{i2}] & \cdots & E[\mathbf{x}_{i1}\mathbf{x}'_{iM}\varepsilon_{i1}\varepsilon_{iM}] \\ E[\mathbf{x}_{i2}\mathbf{x}'_{i1}\varepsilon_{i2}\varepsilon_{i1}] & E[\mathbf{x}_{i2}\mathbf{x}'_{i2}\varepsilon^2_{i2}] & \cdots & E[\mathbf{x}_{i2}\mathbf{x}'_{iM}\varepsilon_{i2}\varepsilon_{iM}] \\ \vdots & \vdots & \ddots & \vdots \\ E[\mathbf{x}_{iM}\mathbf{x}'_{i1}\varepsilon_{iM}\varepsilon_{i1}] & E[\mathbf{x}_{iM}\mathbf{x}'_{i2}\varepsilon_{iM}\varepsilon_{i2}] & \cdots & E[\mathbf{x}_{iM}\mathbf{x}'_{iM}\varepsilon^2_{iM}] \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1M} \\ \mathbf{S}'_{12} & \mathbf{S}_{22} & \cdots & \mathbf{S}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}'_{1M} & \mathbf{S}'_{2M} & \cdots & \mathbf{S}_{MM} \end{bmatrix}
\end{aligned}$$

Note: The single equation $\mathbf{S}$ matrices are along the diagonal. Potential cross-equation correlations imply that the off diagonal $\mathbf{S}$ matrices may be non-zero.

GMM Sample Moment

$$\mathbf{g}_n(\tilde{\boldsymbol{\delta}}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i(\tilde{\boldsymbol{\delta}})$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1}y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{iM}y_{iM} \end{bmatrix} - \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1}z'_{i1}\tilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{iM}z'_{iM}\tilde{\boldsymbol{\delta}}_M \end{bmatrix}$$

$$= \frac{1}{n}\begin{bmatrix} \mathbf{X}'_1 & & \\ & \ddots & \\ & & \mathbf{X}'_M \end{bmatrix}\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} - \frac{1}{n}\begin{bmatrix} \mathbf{X}'_1\mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{X}'_M\mathbf{Z}_M \end{bmatrix}\begin{bmatrix} \tilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \tilde{\boldsymbol{\delta}}_M \end{bmatrix}$$

$$= \frac{1}{n}\underline{\mathsf{X}}'\underline{\mathsf{y}} - \frac{1}{n}\underline{\mathsf{X}}'\underline{\mathsf{Z}}\tilde{\boldsymbol{\delta}}$$

$$= \underset{(K\times 1)}{\underline{\mathsf{S}}_{xy}} - \underset{(K\times L)}{\underline{\mathsf{S}}_{xz}}\underset{(L\times 1)}{\tilde{\boldsymbol{\delta}}}$$

GMM Estimator Defined

If $K_m = L_m$ for $m = 1, \ldots, M$ then $\mathbf{X}'_m \mathbf{Z}_m$ is a square matrix and so

$$\hat{\boldsymbol{\delta}}_m = \mathbf{S}^{-1}_{x_m z_m} \mathbf{S}_{x_m} m, \; m = 1, \ldots, M$$

Therefore,

$$\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\delta}}'_1, \ldots, \hat{\boldsymbol{\delta}}'_M)'$$

solves

$$\underset{(K\times 1)}{\underline{\mathsf{S}}_{xy}} - \underset{(K\times L)}{\underline{\mathsf{S}}_{xz}} \underset{(L\times 1)}{\tilde{\boldsymbol{\delta}}} = \mathbf{0}$$

If each equation is identified and $K_m > L_m$ for some m then we cannot find some $\hat{\boldsymbol{\delta}}$ that solves

$$\underset{(K\times 1)}{\underline{\mathsf{S}}_{xy}} - \underset{(K\times L)}{\underline{\mathsf{S}}_{xz}} \underset{(L\times 1)}{\tilde{\boldsymbol{\delta}}} = \mathbf{0}$$

For the overidentified model, let $\hat{\mathbf{W}}$ be a $K \times K$ pd symmetric matrix with elements

$$\hat{\mathbf{W}} = \begin{bmatrix} \hat{\mathbf{W}}_{11} & \hat{\mathbf{W}}_{12} & \cdots & \hat{\mathbf{W}}_{1M} \\ \hat{\mathbf{W}}'_{12} & \hat{\mathbf{W}}_{22} & \cdots & \hat{\mathbf{W}}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{W}}'_{1M} & \hat{\mathbf{W}}'_{2M} & \cdots & \hat{\mathbf{W}}_{MM} \end{bmatrix}$$

such that $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ pd.

Then, the GMM estimator solves

$$\begin{aligned} \hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) &= \arg\min_{\tilde{\boldsymbol{\delta}}} J(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}}) = n\mathbf{g}_n(\tilde{\boldsymbol{\delta}})'\hat{\mathbf{W}}\mathbf{g}_n(\tilde{\boldsymbol{\delta}}) \\ &= \arg\min_{\tilde{\boldsymbol{\delta}}} n \left(\underset{\sim}{\mathsf{S}}_{xy} - \underset{\sim}{\mathsf{S}}_{xz}\tilde{\boldsymbol{\delta}} \right)' \hat{\mathbf{W}} \left(\underset{\sim}{\mathsf{S}}_{xy} - \underset{\sim}{\mathsf{S}}_{xz}\tilde{\boldsymbol{\delta}} \right) \end{aligned}$$

Straightforward but tedious algebra gives

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = \left(\underset{\sim}{\mathsf{S}}'_{xz}\hat{\mathbf{W}}\underset{\sim}{\mathsf{S}}_{xz}\right)^{-1}\underset{\sim}{\mathsf{S}}'_{xz}\hat{\mathbf{W}}\underset{\sim}{\mathsf{S}}_{xy} =$$

$$\begin{pmatrix} \mathbf{S}'_{x_1z_1}\hat{\mathbf{W}}_{11}\mathbf{S}_{x_1z_1} & \mathbf{S}'_{x_1z_1}\hat{\mathbf{W}}_{12}\mathbf{S}_{x_2z_2} & \cdots & \mathbf{S}'_{x_1z_1}\hat{\mathbf{W}}_{1M}\mathbf{S}_{x_Mz_M} \\ \mathbf{S}'_{x_2z_2}\hat{\mathbf{W}}'_{12}\mathbf{S}_{x_1z_1} & \mathbf{S}'_{x_2z_2}\hat{\mathbf{W}}_{22}\mathbf{S}_{x_2z_2} & \cdots & \mathbf{S}'_{x_2z_2}\hat{\mathbf{W}}_{2M}\mathbf{S}_{x_Mz_M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}'_{x_Mz_M}\hat{\mathbf{W}}'_{1M}\mathbf{S}_{x_1z_1} & \mathbf{S}'_{x_Mz_M}\hat{\mathbf{W}}'_{2M}\mathbf{S}_{x_2z_2} & \cdots & \mathbf{S}'_{x_Mz_M}\hat{\mathbf{W}}_{MM}\mathbf{S}_{x_Mz_M} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \mathbf{S}'_{x_1z_1}\sum_{m=1}^{M}\hat{\mathbf{W}}_{1m}\mathbf{S}_{x_my_m} \\ \mathbf{S}'_{x_2z_2}\sum_{m=1}^{M}\hat{\mathbf{W}}_{2m}\mathbf{S}_{x_my_m} \\ \vdots \\ \mathbf{S}'_{x_Mz_M}\sum_{m=1}^{M}\hat{\mathbf{W}}_{Mm}\mathbf{S}_{x_my_m} \end{pmatrix}$$

Note: Eviews handles this type of model very easily.

Remark

If $\hat{\mathbf{W}}$ is block diagonal, $\hat{\mathbf{W}}$ =**diag**$(W_{11},\ldots,W_{MM})$, then system GMM reduces to single-equation GMM:

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = \left(\underset{\sim}{\mathsf{S}}'_{xz}\hat{\mathbf{W}}\underset{\sim}{\mathsf{S}}_{xz}\right)^{-1}\underset{\sim}{\mathsf{S}}'_{xz}\hat{\mathbf{W}}\underset{\sim}{\mathsf{S}}_{xy} =$$

$$\begin{pmatrix} \mathbf{S}'_{x_1z_1}\hat{\mathbf{W}}_{11}\mathbf{S}_{x_1z_1} & 0 & \cdots & 0 \\ 0 & \mathbf{S}'_{x_2z_2}\hat{\mathbf{W}}_{22}\mathbf{S}_{x_2z_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{S}'_{x_Mz_M}\hat{\mathbf{W}}_{MM}\mathbf{S}_{x_Mz_M} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \mathbf{S}'_{x_1z_1}\hat{\mathbf{W}}_{11}\mathbf{S}_{x_1y_1} \\ \mathbf{S}'_{x_2z_2}\hat{\mathbf{W}}_{22}\mathbf{S}_{x_1y_2} \\ \vdots \\ \mathbf{S}'_{x_Mz_M}\hat{\mathbf{W}}_{MM}\mathbf{S}_{x_My_M} \end{pmatrix}$$

So that

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) &= \begin{pmatrix} \left(\mathbf{S}'_{x_1 z_1}\hat{\mathbf{W}}_{11}\mathbf{S}_{x_1 z_1}\right)^{-1}\mathbf{S}'_{x_1 z_1}\hat{\mathbf{W}}_{11}\mathbf{S}_{x_1 y_1} \\ \left(\mathbf{S}'_{x_2 z_2}\hat{\mathbf{W}}_{22}\mathbf{S}_{x_2 z_2}\right)^{-1}\mathbf{S}'_{x_2 z_2}\hat{\mathbf{W}}_{22}\mathbf{S}_{x_1 y_2} \\ \vdots \\ \left(\mathbf{S}'_{x_M z_M}\hat{\mathbf{W}}_{MM}\mathbf{S}_{x_M z_M}\right)^{-1}\mathbf{S}'_{x_M z_M}\hat{\mathbf{W}}_{MM}\mathbf{S}_{x_M y_M} \end{pmatrix} \\
&= \begin{pmatrix} \hat{\boldsymbol{\delta}}_1(\hat{\mathbf{W}}_{11}) \\ \hat{\boldsymbol{\delta}}_2(\hat{\mathbf{W}}_{22}) \\ \vdots \\ \hat{\boldsymbol{\delta}}_M(\hat{\mathbf{W}}_{MM}) \end{pmatrix}
\end{aligned}
$$

Efficient GMM

The efficient GMM estimator uses

$$\begin{aligned}\hat{\mathbf{W}} &= \hat{\mathrm{S}}^{-1} \\ \hat{\mathrm{S}} \xrightarrow{p} \mathbf{S} &= E[\mathbf{g}_i(\boldsymbol{\delta})\mathbf{g}_i(\boldsymbol{\delta})']\end{aligned}$$

and solves

$$\begin{aligned}\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}^{-1}) &= \arg\min_{\tilde{\boldsymbol{\delta}}} J(\tilde{\boldsymbol{\delta}}, \hat{\mathrm{S}}^{-1}) = n\mathbf{g}_n(\tilde{\boldsymbol{\delta}})'\hat{\mathrm{S}}^{-1}\mathbf{g}_n(\tilde{\boldsymbol{\delta}}) \\ &= \arg\min_{\tilde{\boldsymbol{\delta}}} n\left(\underset{\sim}{\mathsf{S}}_{xy} - \underset{\sim}{\mathsf{S}}_{xz}\tilde{\boldsymbol{\delta}}\right)' \hat{\mathrm{S}}^{-1}\left(\underset{\sim}{\mathsf{S}}_{xy} - \underset{\sim}{\mathsf{S}}_{xz}\tilde{\boldsymbol{\delta}}\right)\end{aligned}$$

giving

$$\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}^{-1}) = \left(\underset{\sim}{\mathsf{S}}_{xz}'\hat{\mathbf{S}}^{-1}\underset{\sim}{\mathsf{S}}_{xz}\right)^{-1}\underset{\sim}{\mathsf{S}}_{xz}'\hat{\mathbf{S}}^{-1}\underset{\sim}{\mathsf{S}}_{xy}$$

As with single equation GMM, one can compute

1. 2-step efficient estimator

2. Iterated efficient estimator

3. Continuous updating estimator

Estimation of $\mathbf{S}$

$$\hat{\mathbf{S}} = \begin{bmatrix} \hat{\mathbf{S}}_{11} & \hat{\mathbf{S}}_{12} & \cdots & \hat{\mathbf{S}}_{1M} \\ \hat{\mathbf{S}}'_{12} & \hat{\mathbf{S}}_{22} & \cdots & \hat{\mathbf{S}}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{S}}'_{1M} & \hat{\mathbf{S}}'_{2M} & \cdots & \hat{\mathbf{S}}_{MM} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_1^n \mathbf{x}_{i1}\mathbf{x}'_{i1}\hat{\varepsilon}_{i1}^2 & \sum_1^n \mathbf{x}_{i1}\mathbf{x}'_{i2}\hat{\varepsilon}_{i1}\hat{\varepsilon}_{i2} & \cdots & \sum_1^n \mathbf{x}_{i1}\mathbf{x}'_{iM}\hat{\varepsilon}_{i1}\hat{\varepsilon}_{iM} \\ \sum_1^n \mathbf{x}_{i2}\mathbf{x}'_{i1}\hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1} & \sum_1^n \mathbf{x}_{i2}\mathbf{x}'_{i2}\hat{\varepsilon}_{i2}^2 & \cdots & \sum_1^n \mathbf{x}_{i2}\mathbf{x}'_{iM}\hat{\varepsilon}_{i2}\hat{\varepsilon}_{iM} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_1^n \mathbf{x}_{iM}\mathbf{x}'_{i1}\hat{\varepsilon}_{iM}\hat{\varepsilon}_{i1} & \sum_1^n \mathbf{x}_{iM}\mathbf{x}'_{i2}\hat{\varepsilon}_{iM}\hat{\varepsilon}_{i2} & \cdots & \sum_1^n \mathbf{x}_{iM}\mathbf{x}'_{iM}\hat{\varepsilon}_{iM}^2 \end{bmatrix} \times \frac{1}{n}$$

Here, $\hat{\mathbf{S}}_{hm} = n^{-1}\sum_1^n \mathbf{x}_{ih}\mathbf{x}'_{im}\hat{\varepsilon}_{ih}\hat{\varepsilon}_{im}$

Here,

$$\begin{aligned}\hat{\varepsilon}_{im} &= y_{im} - \mathbf{z}'_{im}\hat{\boldsymbol{\delta}}_m,\ m = 1, \ldots, M \\ &\hat{\boldsymbol{\delta}}_m \xrightarrow{p} \boldsymbol{\delta}_m\end{aligned}$$

Potential initial consistent estimators of $\boldsymbol{\delta}$:

1. $\hat{\boldsymbol{\delta}}(\mathbf{I}_K) = (\hat{\boldsymbol{\delta}}_1(\mathbf{I}_{K_1})', \ldots, \hat{\boldsymbol{\delta}}_M(\mathbf{I}_{K_m})')'$

2. Single equation efficient GMM estimators:

$$\hat{\boldsymbol{\delta}}_m(\hat{\mathbf{S}}_{mm}^{-1}),\ m = 1, \ldots, M$$

Single Equation vs. Multiple Equation GMM

Result: Single equation GMM is a special case of multiple equation GMM

Single equation GMM estimation: Do GMM on each equation individually with equation specific weight matrix $\hat{\mathbf{W}}_{mm}$:

$$\begin{aligned} \hat{\boldsymbol{\delta}}_m(\hat{\mathbf{W}}_{mm}) &= (\mathbf{S}'_{x_m z_m}\hat{\mathbf{W}}_{mm}\mathbf{S}_{x_m z_m})^{-1}\mathbf{S}'_{x_m z_m}\hat{\mathbf{W}}_{mm}\mathbf{S}_{x_m y_m} \\ m &= 1,\ldots,M \end{aligned}$$

This is multiple equation GMM with a block diagonal weight matrix

$$\hat{\mathbf{W}} = \text{diag}(\hat{\mathbf{W}}_{11},\ldots,\hat{\mathbf{W}}_{MM})$$

Q: When is multiple equation efficient GMM equivalent to single equation efficient GMM?

1. Obvious case: Each equation is just identified (so weight matrix does not matter)

2. Not so obvious case: At least one equation is overidentified but $\mathbf{S} = E[\mathbf{g}_i(\boldsymbol{\delta})\mathbf{g}_i(\boldsymbol{\delta})']$ is block diagonal

$$\begin{aligned} \mathbf{S} &= \text{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{MM}) \\ &= \text{diag}(E[\mathbf{x}_{i1}\mathbf{x}_{i1}'\varepsilon_{i1}^2], \ldots, E[\mathbf{x}_{iM}\mathbf{x}_{iM}'\varepsilon_{iM}^2]) \end{aligned}$$

That is, if

$$E[\mathbf{x}_{im}\mathbf{x}_{ih}'\varepsilon_{im}\varepsilon_{ih}] = 0 \text{ for all } m \neq h$$

Multiple Equation GMM Can be Hazardous

1. Except for the two cases listed above, multiple equation GMM is asymptotically more efficient than single equation GMM

2. Finite sample properties of multiple equation GMM may be worse than single equation GMM

3. Multiple Equation GMM assumes that all equations are correctly specified. Misspecification of one equation can lead to rejection of jointly estimated equations (J-statistic is sensitive to any violation of orthogonality conditions)

Special Case of Multiple Equation GMM: 3SLS

Assumptions:

1. Conditional homoskedasticity

$$E[\varepsilon_{im}\varepsilon_{ih}|\mathbf{x}_{im},\mathbf{x}_{ih}] = \sigma_{mh}$$
$$\Rightarrow \mathbf{S}_{mh} = E[\mathbf{x}_{im}\mathbf{x}'_{ih}\varepsilon_{im}\varepsilon_{ih}] = \sigma_{mh}E[\mathbf{x}_{im}\mathbf{x}'_{ih}]$$

2. Use the same instruments across all equations

$$\mathbf{x}_{i1} = \mathbf{x}_{i2} = \cdots = \mathbf{x}_{iM} = \underset{k\times 1}{\mathbf{x}_i}$$

3SLS moment conditions

$$\underset{Mk\times 1}{\mathbf{g}_i(\boldsymbol{\delta})} = \begin{bmatrix} \mathbf{x}_i\varepsilon_{i1} \\ \vdots \\ \mathbf{x}_i\varepsilon_{iM} \end{bmatrix} = \boldsymbol{\varepsilon}_i \otimes \mathbf{x}_i, \quad \underset{M\times 1}{\boldsymbol{\varepsilon}_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix} = \begin{bmatrix} y_{i1} - \mathbf{z}'_{i1}\boldsymbol{\delta}_1 \\ \vdots \\ y_{iM} - \mathbf{z}'_{iM}\boldsymbol{\delta}_M \end{bmatrix}$$

3SLS Efficient Weight matrix

$$\underset{Mk\times Mk}{\mathbf{S}_{3SLS}} = \begin{bmatrix} \sigma_{11}E[\mathbf{x}_i\mathbf{x}_i'] & \sigma_{12}E[\mathbf{x}_i\mathbf{x}_i'] & \cdots & \sigma_{1M}E[\mathbf{x}_i\mathbf{x}_i'] \\ \sigma_{12}E[\mathbf{x}_i\mathbf{x}_i'] & \sigma_{22}E[\mathbf{x}_i\mathbf{x}_i'] & \cdots & \sigma_{2M}E[\mathbf{x}_i\mathbf{x}_i'] \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M}E[\mathbf{x}_i\mathbf{x}_i'] & \sigma_{2M}E[\mathbf{x}_i\mathbf{x}_i'] & \cdots & \sigma_{MM}E[\mathbf{x}_i\mathbf{x}_i'] \end{bmatrix}$$

$$= \underset{M\times M}{\mathbf{\Sigma}} \otimes \underset{k\times k}{E[\mathbf{x}_i\mathbf{x}_i']}$$

$$\mathbf{\Sigma} = E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{pmatrix}$$

Then

$$\mathbf{S}_{3SLS}^{-1} = \underset{M\times M}{\mathbf{\Sigma}^{-1}} \otimes \underset{k\times k}{E[\mathbf{x}_i\mathbf{x}_i']}^{-1}$$

Estimation of $\mathbf{S}_{3SLS}$

$$
\begin{aligned}
\hat{\mathbf{S}}_{3SLS} &= \hat{\boldsymbol{\Sigma}}_{2SLS} \otimes \mathbf{S}_{xx}, \\
\mathbf{S}_{xx} &= \frac{1}{n}\mathbf{X}'\mathbf{X} \\
\hat{\sigma}_{mh,2SLS} &= \frac{1}{n}(\mathbf{y}_m - \mathbf{Z}_m \hat{\boldsymbol{\delta}}_{m,2SLS})'(\mathbf{y}_h - \mathbf{Z}_h \hat{\boldsymbol{\delta}}_{h,2SLS}) \\
\hat{\boldsymbol{\delta}}_{m,2SLS} &= (\mathbf{Z}_m' \mathbf{P}_X \mathbf{Z}_m)^{-1} \mathbf{Z}_m' \mathbf{P}_X \mathbf{y}_m
\end{aligned}
$$

Analytic Formula for 3SLS Estimator

$$\hat{\boldsymbol{\delta}}(\hat{\mathsf{S}}_{3SLS}^{-1}) = \left(\underline{\mathsf{S}}_{xz}' \left(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{S}_{xx}^{-1}\right) \underline{\mathsf{S}}_{xz}\right)^{-1} \underline{\mathsf{S}}_{xz}' \left(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{S}_{xx}^{-1}\right) \underline{\mathsf{S}}_{xy}$$

Now

$$\begin{aligned}
\underline{\mathsf{S}}_{xz} &= \frac{1}{n}\underline{\mathsf{X}}'\underline{\mathsf{Z}}, \;\; \underline{\mathsf{S}}_{xy} = \frac{1}{n}\underline{\mathsf{X}}'\underline{\mathsf{y}} \\
\underset{nM\times Mk}{\underline{\mathsf{X}}} &= \text{diag}(\mathbf{X}, \ldots, \mathbf{X}) = \underset{M\times M}{\mathbf{I}_M} \otimes \underset{n\times k}{\mathbf{X}} \\
\underset{nM\times L}{\underline{\mathsf{Z}}} &= \text{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_M) \\
\underset{nM\times 1}{\underline{\mathsf{y}}} &= (\mathbf{y}_1', \ldots, \mathbf{y}_M')'
\end{aligned}$$

Then

$$\underline{\mathsf{S}}_{xz} = \frac{1}{n}(\mathbf{I}_M \otimes \mathbf{X})'\underline{\mathsf{Z}}, \;\; \underline{\mathsf{S}}_{xy} = \frac{1}{n}(\mathbf{I}_M \otimes \mathbf{X})'\underline{\mathsf{y}}$$

Rewriting the 3SLS estimator

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{3SLS}^{-1}) &= \left[\underline{\underline{Z}}'(\mathbf{I}_M \otimes \mathbf{X})(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes (\mathbf{X}'\mathbf{X})^{-1})(\mathbf{I}_M \otimes \mathbf{X})'\underline{\underline{Z}}\right]^{-1} \\
&\quad \times \underline{\underline{Z}}'(\mathbf{I}_M \otimes \mathbf{X})(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes (\mathbf{X}'\mathbf{X})^{-1})(\mathbf{I}_M \otimes \mathbf{X})'\underline{y} \\
&= \left[\underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\underline{\underline{Z}}\right]^{-1} \\
&\quad \times \underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\underline{y} \\
&= \left[\underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{P}_X)\underline{\underline{Z}}\right]^{-1} \underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{P}_X)\underline{y}
\end{aligned}
$$

where

$$
\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'
$$

3SLS Overidentification

Parameters to be estimated

$$L = \sum_{m=1}^{M} L_m$$

Total Moment conditions

$$MK, \ K = \# \text{ of common instruments per equation}$$

Number of overidentifying restrictions

$$MK - L$$

Distribution of 3SLS J-statistic

$$\begin{aligned} J(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{3SLS}^{-1}), \hat{\mathbf{S}}_{3SLS}^{-1}) &= \\ n\left(\underline{\mathsf{S}}_{xy} - \underline{\mathsf{S}}_{xz}\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{3SLS}^{-1})\right)' \hat{\mathbf{S}}_{3SLS}^{-1} \left(\underline{\mathsf{S}}_{xy} - \underline{\mathsf{S}}_{xz}\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{3SLS}^{-1})\right) &\sim \chi^2(MK - L) \end{aligned}$$

Special Case of Multiple Equation GMM: Seemingly Unrelated Regressions (SUR)

Assumptions:

1. Conditional homoskedasticity

$$E[\varepsilon_{im}\varepsilon_{ih}|\mathbf{x}_{im}, \mathbf{x}_{ih}] = \sigma_{mh}$$
$$\Rightarrow \mathbf{S}_{mh} = E[\mathbf{x}_{im}\mathbf{x}_{ih}'\varepsilon_{im}\varepsilon_{ih}] = \sigma_{mh}E[\mathbf{x}_{im}\mathbf{x}_{ih}']$$

2. Use the same instruments across all equations

$$\mathbf{x}_{i1} = \mathbf{x}_{i2} = \cdots = \mathbf{x}_{iM} = \underset{k\times 1}{\mathbf{x}_i}$$

3. $\mathbf{x}_i =$ union of $(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iM}) = \mathbf{z}_i \Rightarrow \mathbf{z}_i$ is not endogenous

$$E[\mathbf{z}_{im}\varepsilon_{ih}] = \mathbf{0},\ m, h = 1, \ldots, M$$

SUR Efficient Weight Matrix

$$
\begin{aligned}
\underset{Mk\times Mk}{\mathbf{S}_{SUR}} &= \begin{bmatrix} \sigma_{11}E[\mathbf{z}_i\mathbf{z}_i'] & \sigma_{12}E[\mathbf{z}_i\mathbf{z}_i'] & \cdots & \sigma_{1M}E[\mathbf{z}_i\mathbf{z}_i'] \\ \sigma_{12}E[\mathbf{z}_i\mathbf{z}_i'] & \sigma_{22}E[\mathbf{z}_i\mathbf{z}_i'] & \cdots & \sigma_{2M}E[\mathbf{z}_i\mathbf{z}_i'] \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M}E[\mathbf{z}_i\mathbf{z}_i'] & \sigma_{2M}E[\mathbf{z}_i\mathbf{z}_i'] & \cdots & \sigma_{MM}E[\mathbf{z}_i\mathbf{z}_i'] \end{bmatrix} \\
&= \mathbf{\Sigma} \otimes \underset{k\times k}{E[\mathbf{z}_i\mathbf{z}_i']}
\end{aligned}
$$

Estimating $\mathbf{S}_{SUR}$

$$
\begin{aligned}
\hat{\mathbf{S}}_{SUR} &= \hat{\mathbf{\Sigma}}_{OLS} \otimes \mathbf{S}_{zz},\ \mathbf{S}_{zz} = \frac{1}{n}\mathbf{Z}'\mathbf{Z} \\
\hat{\sigma}_{mh,OLS} &= \frac{1}{n}(\mathbf{y}_m - \mathbf{Z}_m\hat{\boldsymbol{\delta}}_{m,OLS})'(\mathbf{y}_h - \mathbf{Z}_h\hat{\boldsymbol{\delta}}_{h,OLS}) \\
\hat{\boldsymbol{\delta}}_{m,OLS} &= (\mathbf{Z}_m'\mathbf{Z}_m)^{-1}\mathbf{Z}_m\mathbf{y}_m
\end{aligned}
$$

Analytic Formula for SUR Estimator

$$\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}_{SUR}^{-1}) = \left(\underline{\mathsf{S}}_{xz}'\left(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1}\otimes\mathbf{S}_{zz}^{-1}\right)\underline{\mathsf{S}}_{xz}\right)^{-1}\underline{\mathsf{S}}_{xz}'\left(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1}\otimes\mathbf{S}_{zz}^{-1}\right)\underline{\mathsf{S}}_{xy}$$

Now

$$\begin{aligned}
\underline{\mathsf{S}}_{xz} &= \frac{1}{n}\underline{\mathsf{X}}'\underline{\mathsf{Z}},\ \ \underline{\mathsf{S}}_{xy} = \frac{1}{n}\underline{\mathsf{X}}'\underline{\mathsf{y}} \\
\underset{nM\times Mk}{\underline{\mathsf{X}}} &= \mathsf{diag}(\mathbf{Z},\ldots,\mathbf{Z}) = \mathbf{I}_M\otimes\mathbf{Z} \\
\underset{nM\times L}{\underline{\mathsf{Z}}} &= \mathsf{diag}(\mathbf{Z}_1,\ldots,\mathbf{Z}_M) \\
\underset{nM\times 1}{\underline{\mathsf{y}}} &= (\mathbf{y}_1',\ldots,\mathbf{y}_M')'
\end{aligned}$$

Using the same algebraic tricks to derive the 3SLS estimator gives

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{SUR}^{-1}) = \left[\underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{P}_Z)\underline{\underline{Z}}\right]^{-1} \underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{P}_Z)\underline{y}$$

SUR Overidentification

Parameters to be estimated

$$L = \sum_{m=1}^{M} L_m$$

Total Moment conditions

$$MK, \ K = \# \text{ of common instruments per equation}$$

Number of overidentifying restrictions

$$MK - L$$

Distribution of SUR J-statistic

$$J(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{SUR}^{-1}), \hat{\mathbf{S}}_{SUR}^{-1})$$
$$= n\left(\underline{\mathsf{S}}_{xy} - \underline{\mathsf{S}}_{xz}\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{SUR}^{-1})\right)' \hat{\mathbf{S}}_{SUR}^{-1}\left(\underline{\mathsf{S}}_{xy} - \underline{\mathsf{S}}_{xz}\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{SUR}^{-1})\right) \sim \chi^2(MK - L)$$

Simplifying the 3SLS Estimator

$$
\begin{aligned}
& \underset{\sim}{Z}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{P}_X)\underset{\sim}{Z} \\
= & \begin{bmatrix} \mathbf{Z}_1' & & \\ & \ddots & \\ & & \mathbf{Z}_M' \end{bmatrix} \begin{bmatrix} \hat{\sigma}^{11}\mathbf{P}_X & \cdots & \hat{\sigma}^{1M}\mathbf{P}_X \\ \vdots & \ddots & \vdots \\ \hat{\sigma}^{1M}\mathbf{P}_X & \cdots & \hat{\sigma}^{MM}\mathbf{P}_X \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}_M \end{bmatrix} \\
= & \begin{bmatrix} (\mathbf{P}_X\mathbf{Z}_1)' & & \\ & \ddots & \\ & & (\mathbf{P}_X\mathbf{Z}_M)' \end{bmatrix} \begin{bmatrix} \hat{\sigma}^{11}\mathbf{I}_n & \cdots & \hat{\sigma}^{1M}\mathbf{I}_n \\ \vdots & \ddots & \vdots \\ \hat{\sigma}^{1M}\mathbf{I}_n & \cdots & \hat{\sigma}^{MM}\mathbf{I}_n \end{bmatrix} \\
& \times \begin{bmatrix} \mathbf{P}_X\mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{P}_X\mathbf{Z}_M \end{bmatrix} \quad (\text{use } \mathbf{P}_X \cdot \mathbf{P}_X = \mathbf{P}_X,\ \mathbf{P}_X = \mathbf{P}_X') \\
= & \hat{Z}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{I}_n)\hat{Z}
\end{aligned}
$$

Here

$$\hat{\mathsf{Z}} = \begin{bmatrix} \mathbf{P}_X\mathbf{Z}_1 & \\ & \mathbf{P}_X\mathbf{Z}_M \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{Z}}_1 & \\ & \hat{\mathbf{Z}}_M \end{bmatrix}$$

Therefore, the 3SLS estimator may be rewritten as

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{3SLS}^{-1}) = \left[\hat{\mathsf{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{I}_n)\hat{\mathsf{Z}}\right]^{-1} \hat{\mathsf{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{I}_n)\underline{\mathsf{y}}$$

Simplifying the SUR Estimator

$$
\begin{aligned}
& \underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{P}_Z)\underline{\underline{Z}} \\
= & \begin{bmatrix} (\mathbf{P}_Z\mathbf{Z}_1)' & & \\ & \ddots & \\ & & (\mathbf{P}_Z\mathbf{Z}_M)' \end{bmatrix} \begin{bmatrix} \hat{\sigma}^{11}\mathbf{I}_n & \cdots & \hat{\sigma}^{1M}\mathbf{I}_n \\ \vdots & \ddots & \vdots \\ \hat{\sigma}^{1M}\mathbf{I}_n & \cdots & \hat{\sigma}^{MM}\mathbf{I}_n \end{bmatrix} \\
& \times \begin{bmatrix} \mathbf{P}_Z\mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{P}_Z\mathbf{Z}_M \end{bmatrix} \\
= & \underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)\underline{\underline{Z}}
\end{aligned}
$$

Since

$$\mathbf{P}_Z\mathbf{Z}_m = \mathbf{Z}_m, \; m = 1, \ldots, M$$

Therefore, the SUR estimator can be rewritten as

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{SUR}^{-1}) = \left[\underline{\mathsf{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)\underline{\mathsf{Z}}\right]^{-1} \underline{\mathsf{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)\underline{\mathsf{y}}$$

Remarks

1. Compare SUR with 3SLS

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{3SLS}^{-1}) = \left[\hat{\mathsf{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{I}_n)\hat{\mathsf{Z}}\right]^{-1} \hat{\mathsf{Z}}'(\hat{\boldsymbol{\Sigma}}_{2SLS}^{-1} \otimes \mathbf{I}_n)\underline{\mathsf{y}}$$

3 steps of 3SLS

1) regress $\mathbf{Z}_m$ on $\mathbf{X}$ to get $\hat{\mathbf{Z}}_m$

2-3) compute 2-step SUR estimator of transformed system where $\hat{\mathbf{Z}}$ is data matrix.

Traditional Derivation of SUR Estimator

$$
\begin{aligned}
y_{im} &= \underset{(1\times L_m)}{\mathbf{z}'_{im}} \underset{(L_m\times 1)}{\boldsymbol{\delta}_m} + \varepsilon_{im},\ m = 1,\ldots,M; i = 1,\ldots,n \\
E[\mathbf{z}_{im}\varepsilon_{ih}] &= 0 \text{ (no endogeneity)} \\
E[\varepsilon_{im}\varepsilon_{ih}|\mathbf{z}_{im},\mathbf{z}_{ih}] &= \sigma_{mh} \text{ (homoskedasticity)}
\end{aligned}
$$

Giant Regression Representation

$$
\begin{bmatrix} \underset{n\times 1}{\mathbf{y}_1} \\ \vdots \\ \underset{n\times 1}{\mathbf{y}_M} \end{bmatrix} = \begin{bmatrix} \underset{n\times L_1}{\mathbf{Z}_1} & & \\ & \ddots & \\ & & \underset{n\times L_M}{\mathbf{Z}_M} \end{bmatrix} \begin{bmatrix} \underset{L_1\times 1}{\boldsymbol{\delta}_1} \\ \vdots \\ \underset{L_M\times 1}{\boldsymbol{\delta}_M} \end{bmatrix} + \begin{bmatrix} \underset{n\times 1}{\boldsymbol{\varepsilon}_1} \\ \vdots \\ \underset{n\times 1}{\boldsymbol{\varepsilon}_M} \end{bmatrix}
$$

$$
\underset{nM\times 1}{\underline{\mathsf{y}}} = \underset{nM\times L}{\underline{\mathsf{Z}}} \underset{L\times 1}{\boldsymbol{\delta}} + \underset{nM\times 1}{\underline{\mathsf{e}}},\ L = \sum_{m=1}^{M} L_m
$$

Error Covariance in Giant Regression

$$
\begin{aligned}
E[\underset{\sim}{e}\underset{\sim}{e}'] &= E\left[\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{pmatrix} \left(\varepsilon_1', \ldots, \varepsilon_M'\right)\right] \\
&= \begin{bmatrix} \sigma_{11}\mathbf{I}_n & \cdots & \sigma_{1M}\mathbf{I}_n \\ \vdots & \ddots & \vdots \\ \sigma_{1M}\mathbf{I}_n & \cdots & \sigma_{MM}\mathbf{I}_n \end{bmatrix} \\
&= \mathbf{\Sigma} \otimes \mathbf{I}_n \\
\mathbf{\Sigma} &= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{pmatrix}
\end{aligned}
$$

GLS and FGLS Estimation

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{GLS} &= \left[\underline{\underline{Z}}'(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\underline{\underline{Z}}\right]^{-1}\underline{\underline{Z}}'(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\underline{y} \\ &= \left[\underline{\underline{Z}}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n)\underline{\underline{Z}}\right]^{-1}\underline{\underline{Z}}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n)\underline{y}\end{aligned}$$

The feasible GLS (FGLS) estimator is

$$\hat{\boldsymbol{\delta}}_{FGLS} = \left[\underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1}\otimes\mathbf{I}_n)\underline{\underline{Z}}\right]^{-1}\underline{\underline{Z}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1}\otimes\mathbf{I}_n)\underline{y}$$

where

$$\hat{\sigma}_{mh,OLS} = (\mathbf{y}_m - \mathbf{Z}_m\hat{\boldsymbol{\delta}}_{m,OLS})'(\mathbf{y}_h - \mathbf{Z}_h\hat{\boldsymbol{\delta}}_{h,OLS})/n$$

SUR Model with Common Regressors

$$
\begin{aligned}
y_{im} &= \underset{(1\times L_m)}{\mathbf{z}_i'}\underset{(L_m\times 1)}{\boldsymbol{\delta}_m} + \varepsilon_{im},\ m = 1,\ldots,M; i = 1,\ldots,n \\
\mathbf{z}_{i1} &= \mathbf{z}_{i2} = \cdots = \mathbf{z}_{iM} = \mathbf{z}_i \text{ (common regressors)} \\
E[\mathbf{z}_i \varepsilon_{ih}] &= \mathbf{0} \text{ (no endogeneity)} \\
E[\varepsilon_{im}\varepsilon_{ih}|\mathbf{z}_i] &= \sigma_{mh} \text{ (homoskedasticity)}
\end{aligned}
$$

Giant Regression Representation

$$
\begin{aligned}
\begin{bmatrix} \underset{n\times 1}{\mathbf{y}_1} \\ \vdots \\ \underset{n\times 1}{\mathbf{y}_M} \end{bmatrix} &= \begin{bmatrix} \underset{n\times l}{\mathbf{Z}} & & \\ & \ddots & \\ & & \underset{n\times l}{\mathbf{Z}} \end{bmatrix} \begin{bmatrix} \underset{l\times 1}{\boldsymbol{\delta}_1} \\ \vdots \\ \underset{l\times 1}{\boldsymbol{\delta}_M} \end{bmatrix} + \begin{bmatrix} \underset{n\times 1}{\boldsymbol{\varepsilon}_1} \\ \vdots \\ \underset{n\times 1}{\boldsymbol{\varepsilon}_M} \end{bmatrix} \\
\underset{nM\times 1}{\underline{\mathsf{y}}} &= \underset{nM\times L}{\underline{\mathsf{Z}}}\ \underset{L\times 1}{\boldsymbol{\delta}} + \underset{nM\times 1}{\underline{\mathsf{e}}},\ L = Ml \\
\underline{\mathsf{Z}} &= \mathbf{I}_M \otimes \mathbf{Z}
\end{aligned}
$$

The efficient GMM estimator is the feasible GLS estimator

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{FGLS} &= \left[\underline{\underline{\mathsf{Z}}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)^{-1}\underline{\underline{\mathsf{Z}}}\right]^{-1} \underline{\underline{\mathsf{Z}}}'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)^{-1}\underline{\mathsf{y}} \\ &= \left[(\mathbf{I}_M \otimes \mathbf{Z})'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)(\mathbf{I}_M \otimes \mathbf{Z})\right]^{-1} (\mathbf{I}_M \otimes \mathbf{Z})'(\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{I}_n)\underline{\mathsf{y}} \\ &= \left[\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{Z}'\mathbf{Z}\right]^{-1} (\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{Z}')\underline{\mathsf{y}} \\ &= \left[\hat{\boldsymbol{\Sigma}}_{OLS} \otimes \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right] (\hat{\boldsymbol{\Sigma}}_{OLS}^{-1} \otimes \mathbf{Z}')\underline{\mathsf{y}} \\ &= (\mathbf{I}_M \otimes \left(\mathbf{Z}'\mathbf{Z}\right)^{-1} \mathbf{Z}')\underline{\mathsf{y}}\end{aligned}$$

Now

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}_{FGLS} &= \left(\mathbf{I}_M \otimes \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right)\underline{\mathrm{y}} \\
&= \begin{pmatrix} (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' & & \\ & \ddots & \\ & & (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_1 \\ \vdots \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_M \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\delta}}_{1,OLS} \\ \vdots \\ \hat{\boldsymbol{\delta}}_{M,OLS} \end{pmatrix}
\end{aligned}
$$

Result: When there are common regressors across equations, efficient GMM is numerically equivalent to OLS equation by equation!