Lecture 16
SEM

Simultaneous Equations Models

• *Simultaneous equations models* (SEM) differ from those we have seen so far because in each equation there are *two* or more dependent variables:

$$y_j = Y_j\gamma_j + X_j\beta_j + \varepsilon_j \qquad j = 1,2,.., M$$

• In this model, we have *M* endogeneous variables –i.e., *M* equations-, and *K* exogeneous variables. We can write the model as:

$$y_j = Y_j\gamma_j + X_j\beta_j + \varepsilon_j = W_j\delta + \varepsilon_j \qquad j = 1,2,.., M$$

• A convenient way of writing the SEM for empirical work is stacking:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & W_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} = W\delta + \varepsilon$$

Simultaneous Equations Models

• The system looks like a system of regression equations:

$$\mathbf{y} = \mathbf{W}\,\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

But, there are endogenous variables => OLS will not be consistent. IV should be used.

• We can also write the SEM in matrix form as

$$\mathbf{Y}\,\Gamma + \mathbf{X}\,\mathbf{B} = \mathbf{E}$$

Dimensions:

$\mathbf{Y}$ and $\mathbf{E}$ are TxM matrices

$\mathbf{X}$ is a TxK matrix

Γ and $\mathbf{B}$ are MxM and KxM matrices, respectively.

Canonical example: Supply and Demand systems (P & Q endogenous).

Simultaneous Equations Models

• We write the SEM in matrix form as

$$\mathbf{Y}\,\boldsymbol{\Gamma} + \mathbf{X}\,\mathbf{B} = \mathbf{E} \qquad (*)$$

• The first equation, describing $\mathbf{y}_1$ –i.e., the first column of the TxM matrix $\mathbf{Y}$- is given by:

$$\mathbf{Y}\,\boldsymbol{\Gamma}_1 = \mathbf{X}\,\mathbf{B}_1 + \boldsymbol{\varepsilon}_1$$

Dimensions:

$\mathbf{Y}$ and $\boldsymbol{\varepsilon}_1$ (1st column of $\mathbf{E}$) are TxM and Tx1 matrices, respectively.

$\mathbf{X}$ is a TxK matrix

$\boldsymbol{\Gamma}_1$ (1st column of $\boldsymbol{\Gamma}$) is an Mx1 matrix.

$\mathbf{B}_1$ (first column of $\mathbf{B}$) is a Kx1 matrix.

• Economic theory will determine the structure of these models –i.e., the structure of $\boldsymbol{\Gamma}$ and $\mathbf{B}$.

Simultaneous Equations Models

• We can follow the CLM and write the SEM's assumptions:

• The DGP is

$$\mathbf{Y}\,\Gamma + \mathbf{X}\,\mathbf{B} = \mathbf{E} \qquad (*) \qquad (\mathbf{A1})$$

Model (*) is called the *structure or structural (behavioral) model*. It describes the structure or behavior of the economy. The parameters Γ and **B** are called *structural parameters*.

• Let **ε** be the MTx1 stack vector of **E**, or vec(**E**). Assume the $\boldsymbol{\varepsilon}_t$ are drawn from an M-variate distribution with

$$E[\boldsymbol{\varepsilon}_t|\mathbf{X}_t] = \mathbf{0} \qquad - E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0} \qquad (\mathbf{A2})$$

$$E[\boldsymbol{\varepsilon}_t\,\boldsymbol{\varepsilon}_t'|\mathbf{X}_t] = \boldsymbol{\Sigma} \qquad - E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}] = \boldsymbol{\Sigma}_{M\text{x}M} \otimes \mathbf{I}_T \qquad (\mathbf{A3})$$

• More complicated forms of heteroscedasticity for **ε** are possible.

Simultaneous Equations Models

• We assume that **X** is a well-behaved TxK matrix, with full rank:

$$\text{rank}(\mathbf{X}) = K \qquad (\mathbf{A4})$$

• But the traditional (**A2**) is incomplete, since **X** are not all the explanatory variables. They are only the exogenous variables. We also have endogenous variables. That is,

$$E[\boldsymbol{\varepsilon}_t|\mathbf{Y}_t] \neq \mathbf{0}$$

• We have endogenous explanatory variables, OLS *is not* appropriate (*biased* and *inconsistent*) in these equations.

• Even if we are only interested in one particular equation -say, equation 1–, we may have to consider it as part of a system of equations.

SEM: Reduced Form

• We want to solve the SEM for $\mathbf{Y}$:

$$\mathbf{Y}\,\boldsymbol{\Gamma} + \mathbf{X}\,\mathbf{B} = \mathbf{E} \qquad (*)$$

Assume Γ is nonsingular. Then, post multiply (*) by $\boldsymbol{\Gamma}^{-1}$:

$$\mathbf{Y}\,\boldsymbol{\Gamma}\,\boldsymbol{\Gamma}^{-1} + \mathbf{X}\,\mathbf{B}\,\boldsymbol{\Gamma}^{-1} = \mathbf{E}\,\boldsymbol{\Gamma}^{-1}$$

$$\Rightarrow \mathbf{Y} = \mathbf{X}\,\boldsymbol{\Pi} + \mathbf{V} \qquad (**) \qquad \text{–Note: } \boldsymbol{\Pi}\,\boldsymbol{\Gamma} = -\mathbf{B}$$

• The formulation (**) is called *reduced form*. The reduced form of a model expresses each $\mathbf{y}$ variable only in terms of the exogenous variables, $\mathbf{X}$. The $\boldsymbol{\Pi}$ matrix is called the *reduced form parameter matrix*.

• The reduced form model $\mathbf{Y} = \mathbf{X}\,\boldsymbol{\Pi} + \mathbf{V}$ can be estimated using OLS. The reduced form estimates KM parameters in $\boldsymbol{\Pi}$. But, there are more parameters in the reduced form model. The covariance parameters.

SEM: Reduced Form

• For the first equation, $\mathbf{y}_1$:

$$\mathbf{y}_1 = \mathbf{X}\,\boldsymbol{\Pi}_1 + \mathbf{v}_1$$

where $\boldsymbol{\Pi}_1$ is the first column of the KxM matrix $\boldsymbol{\Pi}$.

$\mathbf{v}_1$ is the first column of the TxM matrix $\mathbf{V} = \mathbf{E}\,\boldsymbol{\Gamma}^{-1}$

• The reduced form estimates KM parameters in $\boldsymbol{\Pi}$. But, there are more parameters in the reduced form model: The covariance parameters.

• Let $\boldsymbol{v}_t$ be a draw from the M-variate distribution of $\mathbf{E}\,\boldsymbol{\Gamma}^{-1}$, with

$$\mathrm{E}[\boldsymbol{v}_t\,\boldsymbol{v}_t'|\mathbf{X}_t] = \Gamma^{-1}\boldsymbol{\Sigma}\,\Gamma^{-1} = \boldsymbol{\Omega}.$$

Then, $\mathrm{E}[\mathbf{v}\,\mathbf{v}'|\mathbf{X}] = \mathrm{E}[\mathrm{vec}(\mathbf{E}\Gamma^{-1})\,\mathrm{vec}(\mathbf{E}\Gamma^{-1})'|\mathbf{X}] = \Gamma^{-1}\boldsymbol{\Sigma}\,\Gamma^{-1} \otimes \mathbf{I}_T$

• Here we have $1/2\ M(M + 1)$ covariance reduced form parameters.

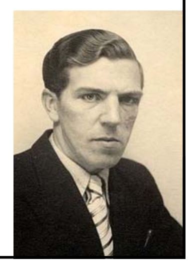
SEM: Reduced Form

• The reduced form estimates a $KM + 1/2\,M(M + 1)$ parameters. But, the structural model has $M^2 + KM + 1/2\,M(M + 1)$.

• We have a problem. This is the *identification* problem.

• Q: Can we go from the OLS estimates of $\{\Pi, \mathbf{\Omega}\}$ to $\{\mathbf{B}, \Gamma, \mathbf{\Sigma}\}$?

• Nobel Prize winner Haavelmo (1943) wrote the classic paper on SEM: "The Statistical Implications of a System of Simultaneous Equations," *Econometrica.*



Trygve Haavelmo (Norway, 1911– 1999)

Demand and Supply: Example

• Simple supply and demand model (M=2, K=1):

$$\begin{aligned} D &= \alpha_o + \alpha_1 P + \alpha_2 Y + \varepsilon_d \\ S &= \beta_o + \beta_1 P + \varepsilon_s \\ D &= S \end{aligned}$$

D and **S**: Quantity demanded and supplied in the market, respectively.
P: Market price of the product.
Y: Consumers' income
$\boldsymbol{\varepsilon}_d$ and $\boldsymbol{\varepsilon}_S$: unobservable ($T$x1) error terms. Let $\boldsymbol{\varepsilon}$ be the stack ($2T$x1) vector of errors ($\boldsymbol{\varepsilon}$=vec([$\boldsymbol{\varepsilon}_d$ $\boldsymbol{\varepsilon}_S$])).

We assume $\Sigma = E[\varepsilon_t \varepsilon_t'] = \begin{bmatrix} \sigma_{dd} & \sigma_{dS} \\ \sigma_{dS} & \sigma_{SS} \end{bmatrix} \Rightarrow Var[\varepsilon] = \Sigma \otimes I_T$

• The structural model is given by the supply and demand equations above. The parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are the structural parameters.

Demand and Supply: Example

• The model for the covariance matrix Var[**ε**|**X**] is very simple:

$$\Sigma = E[\varepsilon_t \varepsilon_t'] = \begin{bmatrix} \sigma_{dd} & \sigma_{dS} \\ \sigma_{dS} & \sigma_{SS} \end{bmatrix} \quad \Rightarrow Var[\varepsilon] = \Sigma \otimes I_T$$

• We assume no autocorrelation or time-varying cross-correlations across equations. We have a SUR-type structure for Var[**ε**|**X**].

Demand and Supply: Example

• We cannot separately observe **D** and **S**, we only observe equilibrium quantities. Let's call this **Q**. Replacing by **Q**:

$$Q = \alpha_1 P + \alpha_2 Y + \varepsilon_d$$
$$Q = \beta_1 P + \varepsilon_S$$

• We have a two equation model with two endogenous variables (**Q**, **P**) and one exogenous variable (**Y**).

• Since E[$\mathbf{P}'\boldsymbol{\varepsilon}_S$] ≠ **0** (& plim($\mathbf{P}'\boldsymbol{\varepsilon}_S/T$) ≠ 0), OLS in the supply equation would result in a *biased* (& *inconsistent*) estimator of β_1.

• We express the system in matrix form and solve for **Q** and **P**:

$$\begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\beta_1 \end{bmatrix} \begin{bmatrix} Q \\ P \end{bmatrix} = \begin{bmatrix} \alpha_2 Y + \varepsilon_d \\ \varepsilon_S \end{bmatrix} \quad => \begin{bmatrix} Q \\ P \end{bmatrix} = \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\beta_1 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_2 Y + \varepsilon_d \\ \varepsilon_S \end{bmatrix}$$

Demand and Supply: Example

• Solving for **Q** and P as a function of **Y**:

$$\begin{bmatrix} Q \\ P \end{bmatrix} = \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\beta_1 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_2 Y + \varepsilon_d \\ \varepsilon_S \end{bmatrix} = \begin{bmatrix} \pi_1 Y + w_1 \\ \pi_2 Y + w_2 \end{bmatrix} = \pi Y + w$$

where ***w*** is the new error term.

• This formulation is the *reduced form*. The **π** vector is the *reduced form parameter vector*.

• Since **Y** is uncorrelated with the error **w** –i.e., E[**Y'w**] = **0**–, we can use OLS to obtain unbiased (and consistent) estimates of **π**.

• Recall the identification problem: Can we go from the OLS estimates of **π** to estimates of **α** and **β**?

Demand and Supply: Example

• It is easy to derive the exact relation between **π** and the structural parameters **α** and **β**. We work out the solution to the system:

$$\begin{bmatrix} Q \\ P \end{bmatrix} = \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\beta_1 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_2 Y + \varepsilon_d \\ \varepsilon_S \end{bmatrix} = \frac{1}{\alpha_1 - \beta_1} \begin{bmatrix} -\beta_1 & \alpha_1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_2 Y + \varepsilon_d \\ \varepsilon_S \end{bmatrix}$$

• Then, equating terms with the reduced form, we get the following relationships between **π** and the structural parameters, **α** and **β**:

$$\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \frac{1}{(\beta_1 - \alpha_1)} \begin{bmatrix} \alpha_2 \beta_1 \\ \alpha_2 \end{bmatrix}$$

• As expected, we also get a relation between **w** and **ε**:

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{\alpha_1 - \beta_1} \begin{bmatrix} -\beta_1 \varepsilon_d + \alpha_1 \varepsilon_S \\ -\varepsilon_d + \varepsilon_S \end{bmatrix}$$

Demand and Supply: Example - Bias

• With this framework, we can calculate the bias in the OLS estimation of equation 2. Recall, the bias is given by E[**P'ε**$_S$].

Proof

$$E(P\varepsilon_S) = E[(\pi_2 Y + w_2)(\varepsilon_S)] = \pi_2 E[Y\varepsilon_S] + E[w_2\varepsilon_S] = E[w_2\varepsilon_S]$$

Now since $w_2 = \frac{-\varepsilon_d + \varepsilon_S}{\alpha_1 - \beta_1}$

$$E[w_2\varepsilon_S] = E\left[\left(\frac{-\varepsilon_d + \varepsilon_S}{\alpha_1 - \beta_1}\right)\varepsilon_S\right] = \frac{E[(\varepsilon_S)^2] - E[\varepsilon_d\varepsilon_S]}{\alpha_1 - \beta_1} = \frac{\sigma_S^2 - \sigma_{Sd}}{\alpha_1 - \beta_1} \neq 0$$

• Since $\beta_1>0$, $\alpha_1<0$, unless σ_{Sd} is high and positive, the bias is likely negative.

Demand and Supply: Example - Identification

• We can also calculate the relation between **Σ** and **Ω:**

$$E[w_t w_t'] = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} = \frac{1}{(\alpha_1 - \beta_1)^2}\begin{bmatrix} \beta_1^{\ 2}\sigma_{dd} + \alpha_1^{\ 2}\sigma_{SS} - 2\beta_1\alpha_1\sigma_{Sd} & \beta_1\sigma_{dd} + \alpha_1\sigma_{SS} - (\beta_1 + \alpha_1)\sigma_{Sd} \\ \beta_1\sigma_{dd} + \alpha_1\sigma_{SS} - (\beta_1 + \alpha_1)\sigma_{Sd} & \sigma_{dd} + \sigma_{SS} - 2\sigma_{Sd} \end{bmatrix}$$

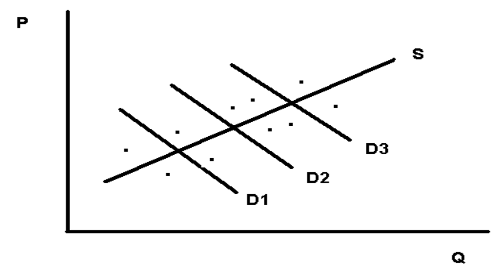
• Going back to the issue of identification. We want to use **π** to identify **α** and **β.** We have 2 equations and 3 unknowns (β_1, α_1 , α_2). Complicated. But, we can recover β_1 from π_1 and π_2:

$$\frac{\pi_1}{\pi_2} = \frac{\frac{\alpha_2\beta_1}{(\beta_1 - \alpha_1)}}{\frac{\alpha_2}{(\beta_1 - \alpha_1)}} = \beta_1$$

• In this case, we will say that the structural form parameter β_1 is *identified*. That is, it can be recovered through the vector **π**.

Demand and Supply: Example - Identification

• Intuition: We were able to identified –i.e., recover from the reduced form estimation- the slope of the supply equation, β_1. Our model has an implicit assumption: the exogenous variable, Y, only affects demand.



Note: The points cluster around S, due to shifts in D as Y varies.

Demand and Supply: Example - Identification

• Using the OLS estimates of $\boldsymbol{\pi}$, we estimate β_1 (*Indirect LS*). Steps:

(1) Estimate the reduced form

(2) Solve for the structural form parameters mathematically.

Note: This method does not provide standard errors for the structural form parameter estimates. The *delta method* should be used.

• Given that we have $E[\mathbf{P}'\boldsymbol{\varepsilon}_S] \neq \mathbf{0}$, why not use 2SLS to estimate β_1?

2SLS Steps:

(1) Regress the endogenous variable, **P**, on the exogenous variables, **Y**.

(2) Regress **Q** against $\hat{P}$.

Note: We cannot use 2SLS to estimate α_1. The 2[nd] step will not work.

Demand and Supply: Example - Identification

• Delta Method – Review

When using Indirect LS, to calculate SE we need to use the delta method. In this case, the multivariate version.

• We have a vector $x_n = [\pi_1\ \pi_2]'$, with

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} \xrightarrow{a} N\left(\begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix}, \begin{bmatrix} \omega_{xx} & \omega_{xy} \\ \omega_{yx} & \omega_{yy} \end{bmatrix} \right)$$

• The multivariate delta method formula:

$$g(x_n) \xrightarrow{a} N(g(\theta), [g'(\theta)]' \, \Sigma \, [g'(\theta)]).$$

• We have a ratio of random variables: $g(x_n) = \beta_1 = \dfrac{\pi_1}{\pi_2}$

Demand and Supply: Example - Identification

Let
$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} \xrightarrow{a} N\left(\begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix}, \begin{bmatrix} \omega_{xx} & \omega_{xy} \\ \omega_{yx} & \omega_{yy} \end{bmatrix} \right)$$

Define $R = x_n / y_n$.

Q: What is the Var(R)= ?

(1) Calculate the plims of $g(x_n)$ and $g'(x_n)$:

$g(R_n) = x_n / y_n$ => plim $g(R_n) = (\theta_x / \theta_y)$

$g'(R_n) = [(1/y_n)\ \ (-x_n / y_n^2)]$ => plim $g'(R_n) = [(1/\theta_y)\ \ (-\theta_x / \theta_y^2)]'$

(2) Multivariate delta method: $g(x_n) \xrightarrow{a} N(g(\theta), [g'(\theta)]' \, \Sigma \, [g'(\theta)])$.

$$Var(R_n) = \begin{bmatrix} \dfrac{1}{\theta_y} & -\dfrac{\theta_x}{\theta_y^2} \end{bmatrix} \begin{bmatrix} \omega_{xx} & \omega_{xy} \\ \omega_{yx} & \omega_{yy} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\theta_y} \\ -\dfrac{\theta_x}{\theta_y^2} \end{bmatrix} = \frac{\omega_{xx}}{\theta_y^2} - \frac{\theta_x \omega_{xy}}{\theta_y^3} - \frac{\theta_x \omega_{xy}}{\theta_y^3} + \frac{\theta_x^2 \omega_{yy}}{\theta_y^4}$$

Demand and Supply: Example - Identification

• We use the delta method for the ratio: $g(x_n) = \beta_1 = \frac{\pi_1}{\pi_2}$

$$g(x_n) \xrightarrow{a} N(g(\theta), [g'(\theta)]' \, \Sigma \, [g'(\theta)]).$$

Let $R_n = x_n / y_n$. Then, the variance of R_n is given by:

$$Var(R_n) = \begin{bmatrix} \frac{1}{\theta_y} & -\frac{\theta_x}{\theta_y^2} \end{bmatrix} \begin{bmatrix} \omega_{xx} & \omega_{xy} \\ \omega_{yx} & \omega_{yy} \end{bmatrix} \begin{bmatrix} \frac{1}{\theta_y} \\ -\frac{\theta_x}{\theta_y^2} \end{bmatrix} = \frac{\omega_{xx}}{\theta_y^2} - \frac{\theta_x \omega_{xy}}{\theta_y^3} - \frac{\theta_x \omega_{xy}}{\theta_y^3} + \frac{\theta_x^2 \omega_{yy}}{\theta_y^4}$$

To do inferences, we replace the θ's and ω's by the estimated values from the Indirect LS:

$$Var(\hat{R}_n) = \frac{s_{\pi_1}^2}{\hat{\pi}_2^2} - \frac{2\hat{\pi}_1 s_{\pi_1\pi_2}}{\hat{\pi}_2^3} + \frac{\hat{\pi}_1^2 s_{\pi_2}^2}{\hat{\pi}_2^4}$$

Demand and Supply: Example - Identification

• We have used two equations to identify the structural parameters:

$$\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \frac{1}{(\beta_1 - \alpha_1)} \begin{bmatrix} \alpha_2 \beta_1 \\ \alpha_2 \end{bmatrix}$$

• We could have also used the estimates of $\mathbf{\Omega}$ to help us in the identification process. The estimates of $\mathbf{\Omega}$ add 3 more equations:

$$\omega_{11} = (\beta_1^2 \sigma_{dd} + \alpha_1^2 \sigma_{SS} - 2\beta_1 \alpha_1 \sigma_{Sd}) / (\beta_1 - \alpha_1)^2$$
$$\omega_{12} = (\beta_1 \sigma_{dd} + \alpha_1 \sigma_{SS} - (\beta_1 + \alpha_1) \sigma_{Sd}) / (\beta_1 - \alpha_1)^2$$
$$\omega_{22} = (\sigma_{dd} + \sigma_{SS} - 2\sigma_{Sd}) / (\beta_1 - \alpha_1)^2$$

• Now, we have 5 equations, with 6 unknowns ($\beta_1, \alpha_1, \alpha_2, \sigma_{dd}, \sigma_{SS}, \sigma_{dS}$).

• We also know β_1 –from the ratio π_1/π_2. But, this does not help for $\boldsymbol{\alpha}$; given the structure of the equations, we cannot solve for α_1 and α_2.

Identification

• In general, we say that the structural form parameters are *identified* if we can solve for them algebraically using information about the reduced form parameters.

• We say that an equation is:
- under-identified (or not identified) if its structural parameters cannot be expressed in terms of the reduced form parameters.
- exactly identified if its structural parameters can be uniquely expressed in terms of the reduced form parameters.
- over-identified if there is more than one solution for expressing its structural parameters in terms of the reduced form parameters.

In the previous example, the supply equation is exactly identified but the demand function is not identified.

Identification - Restrictions

• Recall that we estimate $KM + 1/2\, M(M + 1)$ reduced form parameters. But, the structural model has $M^2 + KM + 1/2\, M(M + 1)$.

• We need to reduce the number of unknown structural parameters we want to estimate. That is, we need restrictions to identify the model.

• **First Order Restrictions**
1. Normalization: Reduces the number of unknown parameters in $\mathbf{\Gamma}$ to $M^2 - M$.
2. Identities: Reduces the number of known parameters by a structure specific amount. Example, D=S=Q.
3. Exclusion restrictions: Reduces the number of unknown parameters by a structure specific amount. Example, **Y** is not in the supply equation.

Identification - Restrictions

4. Linear restrictions –for example, coefficients add to a given value.

• **Second Order Restrictions**

5. Restrictions on $\boldsymbol{\Sigma}$. Apart from VAR systems, this restriction is not used very much.

• **Higher Order Moments Restrictions**

6. Non-linearities: Outside of this course. They have been used to deal with the errors-in-variables problem.

Identification – Single Equation

• WLOG, we study identification of a single equation, say equation 1:

$$\mathbf{Y}\,\Gamma_1 = \mathbf{X}\,\mathbf{B}_1 + \boldsymbol{\varepsilon}_1$$

$$\Rightarrow \qquad \mathbf{y}_1 = \mathbf{Y}_{(-1)}\,\Gamma_{(-1)} + \mathbf{X}\,\mathbf{B}_1 + \boldsymbol{\varepsilon}_1$$

where $\mathbf{Y}_{(-1)}$ is the Tx$(M-1)$ matrix that excludes the first column of $\mathbf{Y}$.

• The normalization condition will be imposed. We will add exclusion restrictions. Excluded variables will have an * added.

Then, we have the equivalences:

exogenous (predetermined) variables = $k_1 + k_1^* = K$

endogenous variables = $m_1 + m_1^* = M - 1 \quad -\gamma_{11}=1$.

• We partition the first equation accordingly

$$\mathbf{y}_1 = \mathbf{Y}_{(-1)}\,\Gamma_{(-1)} + \mathbf{Y}_{(-1)}^*\,\Gamma_{(-1)}^* + \mathbf{X}_1\,\mathcal{B}_1 + \mathbf{X}_1^*\,\mathcal{B}_1^* + \boldsymbol{\varepsilon}_1$$

Identification – Single Equation

$$\mathbf{y}_1 = \mathbf{Y}_{(-1)}\,\Gamma_{(-1)} + \mathbf{Y}_{(-1)}{}^*\,\Gamma_{(-1)}{}^* + \mathbf{X}_1\,\mathcal{B}_1 + \mathbf{X}_1{}^*\,\mathcal{B}_1{}^* + \boldsymbol{\varepsilon}_1$$

Be careful with the dimensions. For example, $\mathbf{Y}_{(-1)}{}^*$ is a Tx$m_1{}^*$ matrix, $\mathbf{X}_1{}^*$ is a Tx$k_1{}^*$ matrix, etc.

• By definition $\Gamma_{(-1)}{}^* = \mathbf{0}$ and $\mathcal{B}_1{}^* = \mathbf{0}$. Now, the number of parameters in the equation is $k_1 + m_1 - 1$.

• We write the reduced form of the structural equation.

$$[\mathbf{y}_1\ \ \mathbf{Y}_1\ \ \mathbf{Y}_1{}^*] = [\mathbf{X}_1\ \mathbf{X}_1{}^*\ ]\ \Pi_1 + \mathbf{V}_1$$

Using $\Pi\,\Gamma = -\mathbf{B}$, we have:

$$\begin{bmatrix} \pi_1 & \ddot{\Pi}_1 & \tilde{\Pi}_1 \\ \pi_1^* & \Pi_1^* & \tilde{\Pi}_1^* \end{bmatrix} \begin{bmatrix} 1 \\ -\Gamma_{(-1)} \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ 0 \end{bmatrix}$$

Identification – Single Equation

• Thus, we have two equations relating the parameters:

$$\pi_1 - \ddot{\Pi}_1\Gamma_{(-1)} = \boldsymbol{\beta}_1$$

$$\pi_1^* - \Pi_1^*\Gamma_{(-1)} = 0$$

• First equation: $k_1 + m_1 - 1$ unknowns with k_1 equations.

- If $\Gamma_{(-1)}$ is known, we have a k_1xk_1 system $\Rightarrow \mathcal{B}_1$ can be found.
- Otherwise, $\mathcal{B}_1$ and $\Gamma_{(-1)}$ can take on arbitrary values and still satisfy it.

• Second equation: k_1* equations with $m_1 - 1$ unknowns ($\Gamma_{(-1)}$).

- If we know π_1* and Π_1*, we can potentially recover $\Gamma_{(-1)}$. Once $\Gamma_{(-1)}$ is found, it can be used to get $\mathcal{B}_1$.

• This result is known as the *order condition* (a *necessary* condition).

We can recover $\Gamma_{(-1)}$ (and, then, $\mathcal{B}_1$) if $k_1* \geq m_1 - 1$.

"*The number of excluded exogenous variables must be at least as great as the number of RHS included endogenous variables.*"

Identification – Single Equation

• The order condition ($k_1* \geq m_1 - 1$) is only a necessary condition.

• A necessary and sufficient condition known as the *rank condition:*

$$\text{rank}(\pi_1* \;\; \Pi_1^*) = \text{rank}(\Pi_1^*) = m_1 - 1$$

• Using matrix notation, we can generalize the restrictions

$$\begin{bmatrix} \Pi & I \\ & \Phi \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} 0 \\ \Theta \end{bmatrix}$$

where Φ is a gx(m_1+k_1) matrix of linear restrictions, and Θ is a gx1 non-zero vector. Then, the rank condition can be expressed as:

$$rank \begin{bmatrix} \Pi & I \\ & \Phi \end{bmatrix} \geq m_1 + k_1 - 1$$

Identification - Conditions

• Order condition for identification of equation *j*:

$$k_j* \geq m_j - 1$$

The number of exogenous variables excluded from equation *j* must be at least as large as the number of endogenous variables included in equation *j*.

It ensures that the second equation has at least one solution. It does not ensure that it has only one solution. It is a necessary condition.

• Rank condition:

$$\text{rank}(\pi_j* \;\; \Pi_j*) = \text{rank}(\Pi_j*) = m_j - 1$$

The rank condition ensures that there is exactly one solution for the structural parameters given the reduced-form parameters.

Identification - Conditions

• Rank condition:

$$\text{rank}(\pi_j{*}\ \ \Pi_j{*}) = \text{rank}(\Pi_j{*}) = m_j - 1$$

• Consider the set of variables excluded from the equation *j*. The matrix of coefficients for these variables in the other equations must have full row rank.

"For each equation: Each of the variables excluded from the equation must appear in at least one of the other equations (no zero columns). Also, at least one of the variables excluded from the equation must appear in each of the other equations (no zero rows)."

The rank condition ensures that there is exactly one solution for the structural parameters given the reduced-form parameters.

Identification - Conditions

• Rank condition is complicated to calculate.

In practice, we follow these simple steps to calculate it:
1) Write down the system in tabular form (*Rank Condition Table*).
2) Strike out the coefficients corresponding to the equation to be identified.
3) Strike out the columns corresponding to those coefficients in Step 2 which are non-zero.
4) The entries left in the table will give only the coefficients of the variables included in the system, but not in the equation under consideration. From these coefficients form all possible A matrices of order M-1 and obtain a corresponding determinant. If at least one of these determinants is non-zero then that equation is identified.

Identification – Conditions – Example

• Suppose we have the following system, with M=3 and K=3:

$$Y_1 = \alpha_1 + \alpha_2 X_1 + \alpha_3 X_3 + \varepsilon_1$$
$$Y_2 = \beta_1 + \beta_2 Y_3 + \beta_3 X_1 + \beta_4 X_2 + \varepsilon_2$$
$$Y_3 = \gamma_1 + \gamma_2 Y_1 + \gamma_3 X_1 + \gamma_4 X_3 + \varepsilon_3$$

- Order condition -- There are 3 exogenous variables in the system (X_1;X_2;X_3) and no more than 3 slope coefficients in any one equation.

- Rank condition – Steps to get the rank of submatrices.
(1) We use a table (Rank Condition Table) in which an X indicates a variable appears in the given equation and a 0 indicates a variable does not appear in the given equation.
(2) Strike from the matrix the column of equation of interest.
(3) Keep only the rows with a 0.

Identification – Conditions – Example

• Rank Condition Table
X: a variable appears in the given equation
0: a variable does not appear in the given equation:

Equation	Y_{1t}	Y_{2t}	Y_{3t}	X_{1t}	X_{2t}	X_{3t}
1	×	0	0	×	0	×
2	0	×	×	×	×	0
3	×	0	×	×	0	×

• Then, for equation *j*:
1) Select the columns corresponding to the variables that do not appear in equation *j*. From this submatrix, delete row *j*.
2) If the rank of remaining submatrix is ≥ $m - 1$ => rank condition is satisfied for equation *j* --and parameters of equation *j* are identified.

Identification – Conditions – Example

• Rank Condition Table

Equation	Y_{1t}	Y_{2t}	Y_{3t}	X_{1t}	X_{2t}	X_{3t}
1	×	0	0	×	0	×
2	0	×	×	×	×	0
3	×	0	×	×	0	×

• Equation *1:* (Y_2;Y_3;X_2) are excluded, so the relevant submatrix is

$$\begin{matrix} \times & \times & \times \\ 0 & \times & 0 \end{matrix}$$

Rank of the submatrix is 2. Equation 1 is identified.

• Equation *3:* (Y_2;X_2) are excluded, so the relevant submatrix is

$$\begin{matrix} 0 & 0 \\ \times & \times \end{matrix}$$

Rank of the submatrix is 1. Equation 3 is not identified.

Identification – Conditions – Example

• What would restore identification? We need rank 2 for the submatrix. One way to do this is to replace X_1 with X_2.

The solution seems natural as now each of the 3 equations has different exogenous shifters.

• The other way is to add Y_2 to the equation *1*. (By adding Y_2 we have made Y_1 a function of X_2.)

Note: It is clear that the distribution of the exogenous variables depends on the distribution of the endogenous variables.

Identification - Terminology

• The usual terminology regarding identification in the SEM context is that equations or systems are *over identified*, *underidentified* - or *exactly identified*.

These are formally defined as:

- Underidentified Structure: $k_j* < m_j - 1$
- Exactly identified Structure: $k_j* = m_j - 1$
- Overidentified Structure: $k_j* > m_j - 1$

Identification – Covariance Restrictions

• So far, all the information provided by $\mathbf{\Omega}$ is used in the estimation of $\mathbf{\Sigma}$. But, for given Γ, the relationship between $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ is one-to-one. Recall that $\Gamma' \mathbf{\Omega}\, \Gamma = \mathbf{\Sigma}$.

• If restrictions are placed on $\mathbf{\Sigma}$, we have more information in the unrestricted $\mathbf{\Omega}$ than is needed for estimation of $\mathbf{\Sigma}$. In many cases, the excess information can be used to identify $\mathbf{B}$, Γ.

• Recall that the OLS bias arises because $E[\mathbf{y}_i\,'\boldsymbol{\varepsilon}_j\,] \neq \mathbf{0}$. This can happen in 2 ways:
(1) Direct relation: $\Gamma_{ij} \neq 0$.
(2) Indirect relation: $\Gamma_{ij} = 0$, but $E[\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_k] \neq \mathbf{0}$ and $\Gamma_{kj} \neq 0$.

<u>Example</u>: Zero covariance restrictions σ_{ij}=0. In the Demand and Supply example this information can be used to identify the system.

Identification – Covariance Restrictions

• We incorporate the covariance restrictions to the system. That is,

$$\begin{bmatrix} \Pi_1 & & I \\ & \Phi & \\ \Omega & & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \\ B_1 \end{bmatrix} = \begin{bmatrix} 0 \\ \Theta \\ \Gamma_1^{-1}{}'\Sigma \end{bmatrix}$$

where Φ is a $(k_1{}^* \text{ x } m_1{+}k_1)$ matrix of exclusion restrictions.

• Note that the term $\Gamma^{-1}\mathbf{\Sigma}$ is related to the OLS bias:

$$\text{plim}(\mathbf{Y'E}/T) = \text{plim}((\mathbf{X\Pi+V})'\mathbf{E}\ /T) = \text{plim}((\mathbf{\Pi'X'E}/T) + \mathbf{V'E}/T)$$
$$= \mathbf{0} + \text{plim}(\Gamma^{-1}{}'\mathbf{E'E}/T) = \Gamma^{-1}\mathbf{\Sigma}$$

Identification – Covariance Restrictions

• The matrix $\Gamma^{-1}\mathbf{\Sigma}$ can identify if variable j can be an instrument in equation i.

Let's define $[\Gamma^{-1}{}'\ \mathbf{\Sigma}_i]_j = \Sigma_k\ (\Gamma^{-1})_{jk}{}'\ \sigma_{ki} = (\Gamma^{-1})_{j1}{}'\ \sigma_{1i} + (\Gamma^{-1})_{j2}{}'\ \sigma_{2i} + ...$

Then, if $[\Gamma^{-1}{}'\,\mathbf{\Sigma}_i]_j=0$, then $\mathbf{y}_j$ can be used as an instrument in equation i.

Identification – Remarks

• Showing that an equation in an SEM with more than two equations is identified is generally difficult. It is easy to see when certain equations are *not* identified.

• An equation in any SEM that does not satisfies the order condition, it is not identified. This is an easy condition to check.

• For identification, the order condition is only necessary, not sufficient, for identification.

• To obtain sufficient conditions, we need to extend the rank condition. For big systems, it can be complicated.

• In practice, it is usual to assume that an equation that satisfies the order condition is identified.

Single Equation Estimation – 2SLS

• WLOG, we assume we are interested in estimating equation 1.

$$\mathbf{y}_1 = \mathbf{Y}_1\,\Gamma_1 + \mathbf{X}_1\,\mathbf{B}_1 + \boldsymbol{\varepsilon}_1 = \mathbf{W}_1\,\boldsymbol{\delta}_1 + \boldsymbol{\varepsilon}_1 \quad (*)$$

• Given that $E[\mathbf{Y}_1'\boldsymbol{\varepsilon}_1] \neq \mathbf{0}$, OLS is inconsistent for Γ_1 and $\mathbf{B}_1$. We need IV –we need an instrument for $\mathbf{Y}_1$– to estimate Γ_1 and $\mathbf{B}_1$ consistently.

The reduced form $\mathbf{Y}_1 = \mathbf{X}\,\boldsymbol{\Pi}_1 + \mathbf{V}_1$ gives us a clue how to proceed: $\hat{Y}_1$.

$$\hat{Y}_1 = \mathbf{X}\,(\mathbf{X'X})^{-1}\mathbf{X'Y}_1 \qquad \Rightarrow E[\hat{Y}_1'\boldsymbol{\varepsilon}_1] = \mathbf{0}.$$

Then, $\mathbf{y}_1 = \hat{\mathbf{W}}_1\,\boldsymbol{\delta}_1 + \boldsymbol{\xi}_1$ with $\hat{\mathbf{W}}_1 = [\hat{Y}_1\ \mathbf{X}_1]$, $\boldsymbol{\delta}_1 = [\Gamma_1\ \mathbf{B}_1]'$

Note that $\boldsymbol{\delta}$ can be consistently estimated by OLS:

$$\mathbf{d}_{1,2SLS} = (\hat{\mathbf{W}}_1'\,\hat{\mathbf{W}}_1)^{-1}\,\hat{\mathbf{W}}_1'\,\mathbf{y}_1 \qquad \text{–this is the 2SLS estimator.}$$

To estimate the variance, we need $\sigma_{\varepsilon 1}^2$. It should be estimated using (*).

Single Equation Estimation – 2SLS

• To estimate the variance, we need $\sigma_{\varepsilon 1}^2$. It should be estimated using equation (*). That is, using $\mathbf{Y}_1$ not $\hat{\mathbf{Y}}_1$:

$$e_{1,2SLS} = y_1 - Y_1 \hat{\Gamma}_{1,2SLS} - X_1 \hat{B}_{1,2SLS}$$

• Note: The 2SLS is a limited information (LI) method, since it looks at information in one equation one at a time. Full information (FI) methods look at all equations jointly and simultaneously together.

• For example, suppose there is a correlation across equations. That is, $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j' | \mathbf{X}_i] \neq 0$, for $i \neq j$.

Then, from SUR theory, LI methods like 2SLS will not be efficient, in general. To gain efficiency, we need a FI method.

2SLS – Identification (again)

• Let's go back to the identification issue. Suppose that $\mathbf{X} = \mathbf{X}_1$ –i.e., there are no excluded exogenous variables; they are all in equation 1. Then,

$$\mathbf{y}_1 = \hat{\mathbf{Y}}_1 \Gamma_1 + \mathbf{X}_1 \mathbf{B}_1 + \boldsymbol{\xi}_1 = (\mathbf{X}\ \hat{\Pi}_1)\ \Gamma_1 + \mathbf{X}_1 \mathbf{B}_1 + \boldsymbol{\xi}_1$$

• We cannot estimate both Γ_1 and $\mathbf{B}_1$ since the same regressors are attached to each parameter (albeit recombined in the case of Γ_1).

• We must have more regressors in $\mathbf{X}$ than appear in $\mathbf{X}_1$. In particular, to estimate Γ_1 we must have at least $m_1 - 1$ more regressors, leading to the rule $K - k_1 \geq m_1 - 1$ (the order condition for identification).

• The rank condition is needed since having $m_1 - 1$ "extra regressors" is not enough, as $\hat{\Pi}_1$ may assign them zero weight, or induce dependencies.

2SLS – Identification (again)

• Identification as lack of IV

- To do 2SLS, we must have (at least) as many instruments as we have variables we need instruments for -i.e., dim(Z) = dim(X). Otherwise, in the identified case (Z'X) would not be square and, thus, invertible.

- In SEM, when wouldn't we have enough IVs? In **X** we have $\mathbf{X}_1$ (exogenous variables included in equation 1) and $\mathbf{X}_1$* (those excluded).

- If none were excluded, then, $\mathbf{X}=\mathbf{X}_1 \Rightarrow \hat{\Pi}_1\mathbf{X}_1=\hat{\Pi}_1\mathbf{X}$,

which are just linear combinations of the $\mathbf{X}_1$ –i.e., there are no extra IVs above $\mathbf{X}_1$ and (**Z'X**) would be singular.

- We need other instruments besides $\mathbf{X}_1$. We need them for $\mathbf{Y}_1$, thus, we need to have at least as many IV as dim(Γ_1).

2SLS – Identification (again)

• We need other instruments besides $\mathbf{X}_1$. We need them for Y_1, thus, we need dim(Γ_1). Then, the number of extra instruments needed dim($\mathbf{X}_1$*)= (K − k_1) must be at least as big as dim(Γ_1)= $m_1 - 1$.

This is the order condition.

• Also note that (**Z'X**) could be singular if the instruments were uncorrelated with the $\hat{\mathbf{W}} = [\hat{\mathbf{Y}}_1 \;\; \mathbf{X}_1]$.

This would be a failure of the rank condition.

"An equation (with associated restrictions) is identified if and only if there exists a consistent IV estimator for the parameters in the equation -i.e., if there are sufficient instruments for the RHS endogenous variables that are fully correlated with these variables."

2SLS – Application

• Epple and McCallum (2005) estimate the canonical example for the market for chickens. They add a price of a producing factor (W) in the supply. They use 1950-2001 USDA annual data to estimate the system:

$$Q_d = \alpha_o + \alpha_1 P + \alpha_2 Y + \varepsilon_d$$
$$Q_S = \beta_o + \beta_1 P + \beta_1 W + \varepsilon_s$$

• Two endogenous variables (Q and P), two exogenous variables (Y and W): M=2 and K=2.

• Both equations are identified: There is a missing exogenous variables in each equation: $k_1* = k_2* = 1$.

2SLS – Application

• Epple and McCallum (2005) use the USDA per capita consumption of chicken as Q for demand. EM first report the OLS estimates (OLS estimates are biased and inconsistent).

OLS	Q_d = per capita consumption of chicken			ΔQ_d
Constant	-4.860 (0.67)	-4.679 (0.68)	5.939 (0.19)	-
Y (ΔY)	0.871 (.07)	0.852 (.07)	0.272 (0.27)	0.771 (0.15)
Price (ΔP)	-0.277 (0.07)	-0.264 (0.07)	-.307 (0.07)	-0.374 (0.06)
PBeef (ΔPb)	-	**-0.118 (0.08)**	0.247 (0.08)	0.251 (0.07)
ε(-1)	-	-	**0.997 (0.02)**	-
R^2	0.980	0.981	.995	.331
DW	0.343	0.443	2.396	2.380

2SLS – Application

• EM for supply they want to use a production aggregate. EM use $Q^A = Q + \log(\text{Population})$. The OLS estimates:

OLS	$Q^A = Q + \log(\text{Population})$		
Constant	9.185 (0.03)	2.652 (0.61)	2.478 (0.70)
W=Price Corn	-0.338 (.08)	-0029 (.02)	-
Price	-1.203 (0.11)	-0.143 (0.05)	**-.041 (0.05)**
Time (technology)	-	0.010 (0.003)	0.010 (0.004)
$Q^A(-1)$	-	0.629 (0.09)	0.647 (0.11)
W=Price Young	-	-	-0.083 (0.03)
R^2	0.942	0.997	.997
DW	0.591	2.054	1.883

2SLS – Application

• EM report the 2SLS estimates. In the last two columns, Q also is adjusted to account for exports (Exports = Qprod-Q^A).

2SLS	ΔQ_d	Q^A	$\boldsymbol{\Delta Q_d}$	**Qprod^A**
Constant	-	2.371 (0.77)	-	2.030 (0.70)
(ΔY)/W(PYoung)	0.843 (.14)	-0.113 (0.04)	0.841 (.14)	-0.146 (.05)
ΔP/Price	-0.404 (.09)	**0.105 (0.08)**	-0.397 (.09)	**0.221 (0.11)**
ΔPb	0.279 (0.09)	-	0.274 (0.09)	-
Time	-	0.012 (0.004)	-	0.018 (0.006)
$Q^A(-1)/Qpr^A(-1)$	-	0.640 (.12)	-	0.631 (.13)
R^2	0.291	0.996	0.299	0.996
DW	1.929	1.869	2.011	2.011

System Estimation – 3SLS

• In some cases, we want to estimate the whole system. We have M equations:

$$y_j = Y_j\gamma_j + X_j\beta_j + \varepsilon_j = W_j\delta + \varepsilon_j \qquad j = 1,2,..,M$$

• Writing the system is stack form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & W_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} = W\delta + \varepsilon$$

with $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$ and $E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}] = \mathbf{V} = \mathbf{\Sigma}_{MxM} \otimes \mathbf{I}_T$

• The OLS estimator (equation-by-equation) is inconsistent. We can do 2SLS. But, 2SLS ignores $\mathbf{V} \Rightarrow$ 2SLS will be inefficient.

System Estimation – 3SLS

• Similar to SUR, 3SLS allows us to estimate the error covariance matrix of dimension MxM, which will be used to do FGLS.

• Steps:

(1) 2SLS in each equation.

(1.a) Regress each endogenous variable (column of $\mathbf{Y}$) on *all* exogenous variables $\mathbf{X}$ to get $\hat{Y}_j = \mathbf{X}\,(\mathbf{X'X})^{-1}\mathbf{X'Y}_j = P_X\,\mathbf{Y}_j$

(1.b) Using $\hat{\mathbf{W}}_j = [\hat{Y}_j\ \mathbf{X}_j]$, estimate $\mathbf{d}_{j,2SLS}$. Keep residuals $\mathbf{e}_{j,2SLS}$

(2) Using $\mathbf{e}_{j,2SLS}$, calculate $\hat{\Sigma}$ (the MxM covariance matrix)

(3) Stack the data. Do FGLS as usual for the whole system:

$$\hat{\delta}_{3SLS} = (\hat{W}'V^{-1}\hat{W})^{-1}\hat{W}'V^{-1}y = (\hat{W}'[\hat{\Sigma}\otimes I]^{-1}\hat{W})^{-1}\hat{W}'[\hat{\Sigma}\otimes I]^{-1}y$$

System Estimation – 3SLS

• The FGLS third stage for the 3SLS estimator can be done by OLS on transformed data.

Steps:
(1) Cholesky decomposition of $\mathbf{\Sigma}^{-1}$ (from step 2) and P_X
Let **L** be a lower triangular Cholesky factor of $\mathbf{\Sigma}^{-1}$ (from step 2) and **Q** be a lower triangular Cholesky factor $\mathbf{X(X'X)^{-1}X'}$. That is,

$$(\mathbf{L}\otimes\mathbf{Q})(\mathbf{L}\otimes\mathbf{Q})' = \mathbf{\Sigma}^{-1}\otimes(\mathbf{X(X'X)^{-1}X'}).$$

(2) Transform the stacked system $(\mathbf{L}\otimes\mathbf{Q})\,\mathbf{y} = (\mathbf{L}\otimes\mathbf{Q})\,\mathbf{W}\,\boldsymbol{\delta} + \boldsymbol{\varepsilon}$
(3) Apply OLS to this system to get the 3SLS estimators.

• Main advantage of 3SLS over 2SLS: A gain in asymptotic efficiency.
• Main disadvantage: The estimators for a single equation are potentially less robust. They will be inconsistent if the IV assumptions that X is predetermined fail in any equation, not just a particular one.

System Estimation – 3SLS

• The 3SLS is given by:

$$\hat{\delta}_{3SLS} = [\hat{W}'(\hat{\Sigma}^{-1}\otimes I)\hat{W}]^{-1}\hat{W}'[\hat{\Sigma}^{-1}\otimes I]y$$

Notation:

$$\hat{W}=(I\otimes P_X)W \qquad \hat{W}=\begin{bmatrix}\hat{W}_1 & 0 & \ldots & 0\\ 0 & \hat{W}_2 & \ldots & 0\\ \ldots & \ldots & \ldots & \\ 0 & 0 & \ldots & \hat{W}_M\end{bmatrix}=\begin{bmatrix}P_X & 0 & \ldots & 0\\ 0 & P_X & \ldots & 0\\ \ldots & \ldots & \ldots & \\ 0 & 0 & \ldots & P_X\end{bmatrix}\begin{bmatrix}W_1 & 0 & \ldots & 0\\ 0 & W_2 & \ldots & 0\\ \ldots & \ldots & \ldots & \\ 0 & 0 & \ldots & W_M\end{bmatrix}$$

• Then substituting above

$$\begin{aligned}\hat{\delta}_{3SLS} &= [W'(I\otimes P_X)'(\hat{\Sigma}^{-1}\otimes I)(I\otimes P_X)W]^{-1}W'(I\otimes P_X)'(\hat{\Sigma}^{-1}\otimes I)y\\ &= [W'(\hat{\Sigma}^{-1}\otimes P_X)W]^{-1}W'(\hat{\Sigma}^{-1}\otimes P_X)y\end{aligned}$$

System Estimation – 3SLS

• Let $\mathbf{Z} = (\mathbf{\Sigma}^{-1} \otimes P_X)W$. Then,

$$\hat{\delta}_{3SLS} = [W'(\hat{\Sigma}^{-1} \otimes P_X)W]^{-1} W'(\hat{\Sigma}^{-1} \otimes P_X)y$$
$$= (Z'W)^{-1} Z'y$$

• That is, 3SLS is an IV estimator.

• Like all IV estimators, it is a consistent estimator. It is also efficient, relative to other IV estimators that use only sample information. The estimated asymptotic variance is:

$$Est.\ Asy.Var[\hat{\delta}_{3SLS} \mid X] = (\hat{W}'[\hat{\Sigma} \otimes I]^{-1} \hat{W})^{-1}$$

Note: It is also possible to iterate the 3SLS computation. But, unlike SUR, however, this method does not provide the MLE, nor does it improve the asymptotic efficiency.

System Estimation – FIVE

• Full Information Instrumental Variable (FIVE)
Simple idea:
(1) Do IV estimation in each equation, say j, using k_j instruments $\mathbf{Z}_j$:
$\mathbf{d}_{j,IV} = (\mathbf{Z}_j'\mathbf{W}_j)^{-1} \mathbf{Z}_j' \mathbf{y}_j$ $\Rightarrow$ get $\mathbf{e}_{j,IV}$

(2) Form $\hat{\Sigma}$

(3) Do FGLS

$$\hat{\delta}_{FIVE} = (\hat{W}'[\hat{\Sigma} \otimes I]^{-1} W)^{-1} \hat{W}'[\hat{\Sigma} \otimes I]^{-1} y \qquad \hat{W}_j = [X_j \hat{B}\hat{\Gamma}^{-1} \quad X_j]$$

• FIVE works in terms of the estimates of $\mathbf{B}$ and Γ^{-1}, while 3SLS works in terms of $\mathbf{\Pi}$.

System Estimation – FIML

• Full Information Maximum Likelihood (FIML)

- Write the full system

$$\mathbf{Y}\,\Gamma + \mathbf{X}\,\mathbf{B} = \mathbf{E}$$

- Assume $\mathbf{E}|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_T)$ –an MxM matrix.

- Write the likelihood function:

$$L(\mathbf{B}, \Gamma, \mathbf{\Sigma}|\mathbf{X}) = (2\pi)^{-T/2}\,|\mathbf{\Sigma}|^{-T/2} \exp[\text{tr}\{-\tfrac{1}{2}\,\mathbf{E}'\,\mathbf{\Sigma}^{-1}\,\mathbf{E}\}$$

- Or, in stacked form

$$L = (2\pi)^{-MT/2}\,|\mathbf{\Sigma}\otimes\mathbf{I}_T|^{-MT/2} \exp\{-\tfrac{1}{2}\,\mathbf{E}'\,(\mathbf{\Sigma}\otimes\mathbf{I}_T)^{-1}\,\mathbf{E}\}$$

- Write the log likelihood function:

$\mathcal{L} = -T/2 \ln(2\pi) - T/2 \ln|\mathbf{\Sigma}| - \tfrac{1}{2}\text{tr}\{(\mathbf{Y}\,\Gamma+\mathbf{X}\,\mathbf{B})'(\mathbf{Y}\,\Gamma+\mathbf{X}\,\mathbf{B})\,\mathbf{\Sigma}^{-1} + \ln|J|\}$

where $J=\mathbf{\Gamma}$ is the Jacobian from the change in variable (from $\mathbf{E}$ to $\mathbf{Y}$).

System Estimation – FIML

$\mathcal{L} = -T/2 \ln(2\pi) - T/2 \ln|\mathbf{\Sigma}| - \tfrac{1}{2}\text{tr}\{(\mathbf{Y}\,\mathbf{\Gamma}+\mathbf{X}\,\mathbf{B})'(\mathbf{Y}\,\mathbf{\Gamma}+\mathbf{X}\,\mathbf{B})\,\mathbf{\Sigma}^{-1} + \ln|\mathbf{\Gamma}|\}$

- Take derivatives of $\mathcal{L}$ w.r.t. Γ, $\mathbf{B}$ and $\mathbf{\Sigma}$ and set them equal to 0 (foc):

(1) $\delta\mathcal{L}/\delta\Gamma' = -\mathbf{Y}'\,(\mathbf{Y}\,\mathbf{\Gamma}+\mathbf{X}\,\mathbf{B})\,\mathbf{\Sigma}^{-1} + T\,\mathbf{\Gamma}^{-1}$

$\Rightarrow \mathbf{Y}'\mathbf{E}_{ML}/T = \mathbf{\Gamma}_{ML}^{-1}\,\mathbf{\Sigma}_{ML}$ – Bias in OLS due to SEM.

(2) $\delta\mathcal{L}/\delta\mathbf{B}' = -\mathbf{X}'\,(\mathbf{Y}\,\mathbf{\Gamma}+\mathbf{X}\,\mathbf{B})\,\mathbf{\Sigma}^{-1}$

$\Rightarrow \mathbf{Z}'\mathbf{E}_{ML}\,\mathbf{\Sigma}_{ML}^{-1} = \mathbf{0}$ – IV condition.

(3) $\delta\mathcal{L}/\delta\mathbf{\Sigma}^{-1\prime} = T/2\,(\mathbf{\Sigma}^{-1})^{-1} - \tfrac{1}{2}\,(\mathbf{Y}\,\mathbf{\Gamma}+\mathbf{X}\,\mathbf{B})'(\mathbf{Y}\,\mathbf{\Gamma}+\mathbf{X}\,\mathbf{B})$

$\Rightarrow \mathbf{\Sigma}_{ML} = \mathbf{E}_{ML}'\mathbf{E}_{ML}/T$ – Standard formula for $\mathbf{\Sigma}$.

Under FIML, we solve simulateneously (1)-(3) for $\mathbf{\Gamma}$, $\mathbf{B}$ and $\mathbf{\Sigma}$.

System Estimation – FIML

• Alternative FIML derivation – Greene's textbook

- Write reduced form system

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}$$

- Assume $\mathbf{V}|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Omega})$ –an MTxMT matrix.
- Write the log likelihood function as:

$\mathcal{L} = \ln L(\mathbf{\Pi}, \mathbf{\Omega}|\mathbf{X}) = -MT/2 \ln(2\pi) - T/2 \ln|\mathbf{\Omega}| - ½ \,\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{W})$

where $\mathbf{W}_{ij} = (\mathbf{y} - \mathbf{X}\boldsymbol{\pi}_j)'(\mathbf{y} - \mathbf{X}\boldsymbol{\pi}_i)$

- Make the substitutions: $\mathbf{\Pi} = -\mathbf{B}\,\Gamma^{-1}$ and $\mathbf{\Omega} = \Gamma^{-1\prime}\mathbf{\Sigma}\,\Gamma^{-1}$
- Then,

$\mathcal{L} = -MT/2 \ln(2\pi) - T/2 \ln|\Gamma^{-1\prime}\,\mathbf{\Sigma}\,\Gamma^{-1}| -$

$\quad - ½ \,\mathrm{tr}\{(\Gamma\,\mathbf{\Sigma}^{-1}\,\Gamma')(\mathbf{Y} + \mathbf{X}\,\mathbf{B}\Gamma^{-1})'(\mathbf{Y} + \mathbf{X}\,\mathbf{B}\Gamma^{-1})\}$

$= -MT/2 \ln(2\pi) - T/2 \ln|\mathbf{\Sigma}| - T/2 \times 2 \ln|\Gamma| -$

$\quad - ½ \,\mathrm{tr}\{\mathbf{\Sigma}^{-1}(\mathbf{Y}\Gamma + \mathbf{XB})'(\mathbf{Y}\Gamma + \mathbf{XB})\}$

System Estimation – FIML

$\mathcal{L} = -MT/2 \ln(2\pi) - T/2 \ln|\mathbf{\Sigma}| - T \ln|\Gamma| -$

$\quad - ½ \,\mathrm{tr}\{\mathbf{\Sigma}^{-1}(\mathbf{Y}\,\mathbf{\Gamma} + \mathbf{X}\,\mathbf{B})'(\mathbf{Y}\,\Gamma + \mathbf{X}\,\mathbf{B})\}$

$= -T/2\,\{M \ln(2\pi) + \ln|\mathbf{\Sigma}| - 2 \ln|\mathbf{\Gamma}| + \mathrm{tr}(\mathbf{\Sigma}^{-1}\,\mathbf{S})\}$

$\mathbf{S} = 1/T\,(\mathbf{Y}\,\mathbf{\Gamma} + \mathbf{X}\,\mathbf{B})'(\mathbf{Y}\,\mathbf{\Gamma} + \mathbf{X}\,\mathbf{B})$ --$s_{ij} = 1/T(\mathbf{Y}\,\Gamma_i + \mathbf{X}\,\mathbf{B}_i)'(\mathbf{Y}\,\Gamma_j + \mathbf{X}\,\mathbf{B}_j)$

- Take derivatives of $\mathcal{L}$ w.r.t. $\mathbf{\Gamma}$, $\mathbf{B}$ and $\mathbf{\Sigma}$, then get f.o.c.'s and solve for the MLE of Γ, $\mathbf{B}$ and $\mathbf{\Sigma}$.

System Estimation – FIML

• Take derivatives of $\mathcal{L}$ w.r.t. Γ, **B** and **Σ**, then get f.o.c.'s and solve for the MLE of Γ, **B** and **Σ**. It turns out the FIML estimator is given by:

$$\hat{\delta}_{\text{FIML}} = [\hat{\mathbf{Z}}(\hat{\delta})'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}[\hat{\mathbf{Z}}(\hat{\delta})'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{y}] = [\hat{\hat{\mathbf{Z}}}'\mathbf{Z}]^{-1}\hat{\hat{\mathbf{Z}}}'\mathbf{y},$$

where

$$\hat{\mathbf{Z}}(\hat{\delta})'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}) = \begin{bmatrix} \hat{\sigma}^{11}\hat{\mathbf{Z}}_1' & \hat{\sigma}^{12}\hat{\mathbf{Z}}_1' & \cdots & \hat{\sigma}^{1M}\hat{\mathbf{Z}}_1' \\ \hat{\sigma}^{12}\hat{\mathbf{Z}}_2' & \hat{\sigma}^{22}\hat{\mathbf{Z}}_2' & \cdots & \hat{\sigma}^{2M}\hat{\mathbf{Z}}_2' \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\sigma}^{1M}\hat{\mathbf{Z}}_M' & \hat{\sigma}^{2M}\hat{\mathbf{Z}}_M' & \cdots & \hat{\sigma}^{MM}\hat{\mathbf{Z}}_M' \end{bmatrix} = \hat{\hat{\mathbf{Z}}}'$$

and

$$\hat{\mathbf{Z}}_j = [\mathbf{X}\hat{\boldsymbol{\Pi}}_j, \mathbf{X}_j].$$

$\hat{\boldsymbol{\Pi}}$ is computed from the structural estimates:

$$\hat{\boldsymbol{\Pi}}_j = M_j \text{ columns of } -\hat{\mathbf{B}}\hat{\boldsymbol{\Gamma}}^{-1}$$

and

$$\hat{\sigma}_{ij} = \frac{1}{T}(\mathbf{y}_i - \mathbf{Z}_i\hat{\delta}_i)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\delta}_j) \quad \text{and} \quad \hat{\sigma}^{ij} = (\hat{\boldsymbol{\Sigma}}^{-1})_{ij}.$$

System Estimation – FIML & 3SLS

• This result implies that the FIML estimator is an IV estimator. Not a surprise, it is one of the implications from the f.o.c.'s.

• The asymptotic covariance matrix for the FIML estimator follows directly from its form as an IV estimator (3SLS). If normality for errors is assumed, 3SLS has the same asymptotic distribution as ML. This result is due to Sargan (Econometrica, 1964).

• Implication: The 3SLS estimator is easier to compute. The easier computations comes at no cost in asymptotic efficiency.

• Small-sample properties remain ambiguous. We tend to find that 3SLS dominates FIML. The 3SLS estimator is robust to non-normality. The FIML estimator is not -- because of the term ln|J| in the log L.

System Estimation – FIML & 3SLS

• The 3SLS and FIML estimators are usually quite different numerically.

• Interesting result: If the system is just identified,

FIVE = 3SLS = FIML = 2SLS

System Estimation – GMM

• GMM is also possible both for single equation or for system estimation.

• As usual we need moments:

- For a single equation, assume : $y_{jt} = \mathbf{z'_{jt}}\, \boldsymbol{\delta_j} + \boldsymbol{\varepsilon}_{jt}$ $\quad z = [Y, x]$

$$g_T(\theta^*) = E[\mathbf{x}_t \boldsymbol{\varepsilon}_{jt}] = E[\mathbf{x}_t (y_{jt} - \mathbf{z'_{jt}}\, \boldsymbol{\delta})] = 0$$

- For a system:

$$q = \Sigma_j\, \Sigma_l\, [\mathbf{e}_t(\mathbf{w_t}, \boldsymbol{\delta_j})'\mathbf{X}/T]\; [\mathbf{W}]^{jl}\; [\mathbf{X'e}_t(\mathbf{w_t}, \boldsymbol{\delta_j})/T]$$
$$= \Sigma_j\, \Sigma_l\, \mathbf{m}(\boldsymbol{\delta_j})'\; [\mathbf{W}]^{jl}\; \mathbf{m}(\boldsymbol{\delta_j})$$

• **W** can incorporate different types of heteroscedasticity, White or NW are OK.

System Estimation – Application (Greene)

TABLE 15.3 Estimates of Klein's Model I (Estimated Asymptotic Standard Errors in Parentheses)

	Limited-Information Estimates				*Full-Information Estimates*			
	2SLS				**3SLS**			
C	16.6	0.017	0.216	0.810	16.4	0.125	0.163	0.790
	(1.32)	(0.118)	(0.107)	(0.040)	(1.30)	(0.108)	(0.100)	(0.033)
I	20.3	0.150	0.616	−0.158	28.2	−0.013	0.756	−0.195
	(7.54)	(0.173)	(0.162)	(0.036)	(6.79)	(0.162)	(0.153)	(0.038)
W^p	1.50	0.439	0.147	0.130	1.80	0.400	0.181	0.150
	(1.15)	(0.036)	(0.039)	(0.029)	(1.12)	(0.032)	(0.034)	(0.028)
	LIML				**FIML**			
C	17.1	−0.222	0.396	0.823	18.3	−0.232	0.388	0.802
	(1.84)	(0.202)	(0.174)	(0.055)	(2.49)	(0.312)	(0.217)	(0.036)
I	22.6	0.075	0.680	−0.168	27.3	−0.801	1.052	−0.146
	(9.24)	(0.219)	(0.203)	(0.044)	(7.94)	(0.491)	(0.353)	(0.30)
W^p	1.53	0.434	0.151	0.132	5.79	0.234	0.285	0.235
	(2.40)	(0.137)	(0.135)	(0.065)	(1.80)	(0.049)	(0.045)	(0.035)
	GMM (H2SLS)				**GMM (H3SLS)**			
C	14.3	0.090	0.143	0.864	15.7	0.068	0.167	0.829
	(0.897)	(0.062)	(0.065)	(0.029)	(0.951)	(0.091)	(0.080)	(0.033)
I	23.5	0.146	0.591	−0.171	20.6	0.213	−0.520	−0.157
	(6.40)	(0.120)	(0.129)	(0.031)	(4.89)	(0.087)	(0.099)	(0.025)
W^p	3.06	0.455	0.106	0.130	2.09	0.446	0.131	0.112
	(0.64)	(0.028)	(0.030)	(0.022)	(0.510)	(0.019)	(0.021)	(0.021)
	OLS				**I3SLS**			
C	16.2	0.193	0.090	0.796	16.6	0.165	0.177	0.766
	(1.30)	(0.091)	(0.091)	(0.040)	(1.22)	(0.096)	(0.090)	(0.035)
I	10.1	0.480	0.333	−0.112	42.9	−0.356	1.01	−0.260
	(5.47)	(0.097)	(0.101)	(0.027)	(10.6)	(0.260)	(0.249)	(0.051)
W^p	1.50	0.439	0.146	0.130	2.62	0.375	0.194	0.168
	(1.27)	(0.032)	(0.037)	(0.032)	(1.20)	(0.031)	(0.032)	(0.029)

System Estimation – Application (Greene)

• Practical Remarks
- It is often found that OLS estimates are very close to the structural estimates.
- OLS estimates can show smaller variances than 2SLS. (MSE issue?)
- Big numerical differences for all the methods. Even signs can be different.
- LI estimators can have smaller variances than the FI estimators. (Due to propagation of specification errors?)
- The calculation of the variance/weighting matrix matters.
- The gains from system estimation in finite samples may be modest.

• <u>Note</u>: All the remarks about OLS and LI estimators are done regarding the finite-sample properties; asymptotically IV and any FI estimator dominate.

DHW Specification Test

Estimator	$\mathbf{X}_j$ is exogenous	$\mathbf{X}_j$ is endogenous
$\boldsymbol{\delta}$	Consistent and Efficient	Inconsistent
$\boldsymbol{\delta}^*$	Consistent Inefficient	Consistent Possibly Efficient

• In Lecture 8 (IV estimation), under an H_0:plim($\mathbf{X'\varepsilon}/T$)=0, we have one estimator that is efficient (OLS) and one inefficient (IV). We can use a Durbin-Hausman-Wu test.

• In SEM, Hausman bases his version of the test on $\boldsymbol{\delta}$ being the 2SLS estimator and $\boldsymbol{\delta}^*$ being the 3SLS estimator. Shortcoming: we need to choose arbitrarily one equation where $\mathbf{X}_j$ is not present for the test.