# ECON 5360 Class Notes
# Systems of Equations

## 1 Introduction

Here, we consider two types of systems of equations. The first type is a system of Seemingly Unrelated Regressions (SUR), introduced by Arnold Zellner (1962). Sets of equations with distinct dependent and independent variables are often linked together by some common unmeasurable factor. Examples include systems of factor demands by a particular firm, agricultural supply-response equations, and capital-asset pricing models. The methods presented here can also be thought of as an alternative estimation framework for panel-data models.

The second type is a Simultaneous Equations System, which involve the interaction of multiple endogenous variables within a system of equations. Estimating the parameters of such as system is typically not as simple as doing OLS equation-by-equation. Issues such as identification (whether the parameters are even estimable) and endogeneity bias are the primary topics in this section.

## 2 Seemingly Unrelated Regressions (SUR) Model

Consider the following set of equations

$$y_i = X_i\beta_i + \epsilon_i \tag{1}$$

for $i = 1, ..., M$, where the matrices $y_i$, $X_i$ and $\beta_i$ are of dimension $(T\times 1)$, $(T\times K_i)$ and $(K_i\times 1)$, respectively[1]. The stacked system in matrix form is

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} = X\beta + \epsilon.$$

Although each of the $M$ equations may seem unrelated (i.e., each has potentially distinct coefficient vectors, dependent variables and explanatory variables), the equations in (1) are linked through their (mean-zero)

[1] Although each equation typically represents a separate time series, it is possible that $T$ instead denotes the number of cross sections within an equation.

error structure

$$E(\epsilon\epsilon') = \Omega = \Sigma \otimes I_T = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1M}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & & \sigma_{2M}I_T \\ \vdots & & \ddots & \vdots \\ \sigma_{M1}I_T & \sigma_{M2}I_T & \cdots & \sigma_{MM}I_T \end{bmatrix}_{MT\times MT}$$

where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2M} \\ \vdots & & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}_{M\times M}$$

is the variance-covariance matrix for each $t = 1, ..., T$ error vector.

## 2.1 Generalized Least Squares (GLS)

The system resembles the one we studied in chapter 10 on nonshperical disturbances. The efficient estimator in this context is the GLS estimator

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) = [X'(\Sigma^{-1} \otimes I)X]^{-1}[X'(\Sigma^{-1} \otimes I)Y].$$

Assuming all the classical assumptions hold (other than that of spherical disturbances), GLS is the best linear unbiased estimator. There are two important conditions under which GLS does not provide any efficiency gains over OLS:

- $\sigma_{ij} = 0$. When all the contemporaneous correlations across equations equal zero, the equations are not linked in any fashion and GLS does not provide any efficiency gains. In fact, one can show that $b = \hat{\beta}$.
- $X_1 = X_2 = \cdots = X_M$. When the explanatory variables are identical across equations, $b = \hat{\beta}$.

As a rule, the efficiency gains of GLS over OLS tend to be greater when

- the contemporaneous correlation in errors across equations ($\sigma_{ij}$) is greater and
- there is less correlation between $X$ across equations.

## 2.2 Feasible Generalized Least Squaures (FGLS)

Typically, $\Sigma$ is not known. Assuming that a consistent estimator of $\Sigma$ is available, the feasible GLS estimator

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y) = [X'(\hat{\Sigma}^{-1} \otimes I)X]^{-1}[X'(\hat{\Sigma}^{-1} \otimes I)Y] \quad (2)$$

will be a consistent estimator of $\beta$. The typical estimator of $\Sigma$ is

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1M} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & & \hat{\sigma}_{2M} \\ \vdots & & \ddots & \vdots \\ \hat{\sigma}_{M1} & \hat{\sigma}_{M2} & \cdots & \hat{\sigma}_{MM} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}e_1'e_1 & \frac{1}{T}e_1'e_2 & \cdots & \frac{1}{T}e_1'e_M \\ \frac{1}{T}e_2'e_1 & \frac{1}{T}e_2'e_2 & & \frac{1}{T}e_2'e_M \\ \vdots & & \ddots & \vdots \\ \frac{1}{T}e_M'e_1 & \frac{1}{T}e_M'e_2 & \cdots & \frac{1}{T}e_M'e_M \end{bmatrix}, \tag{3}$$

where $e_i$, $i = 1, ..., M$ represent the OLS residuals. Degrees of freedom corrections for the elements in $\hat{\Sigma}$ are possible, but will not generally produce unbiasedness. It is also possible to iterate on (2) and (3) until convergence, which will produce the maximum likelihood estimator under multivariate normal errors. In other words, $\hat{\beta}_{FGLS}$ and $\hat{\beta}_{ML}$ will have the same limiting distributions such that

$$\hat{\beta}_{ML,FGLS} \overset{asy}{\sim} N(\beta, \Psi)$$

where $\Psi$ is consistently estimated by

$$\hat{\Psi} = [X'(\hat{\Sigma}^{-1} \otimes I)X]^{-1}.$$

### 2.3 Maximum Likelihood

Although asymptotically equivalent, maximum likelihood is an alternative estimator to FGLS that will provide different answers in small samples. Begin by rewriting the model for the $t^{th}$ observation as

$$Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{Mt} \end{bmatrix}' = x_t^* \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_M \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{M,t} \end{bmatrix}' = x_t^* \Pi + \epsilon_t'$$

where $x_t^*$ is the row vector of all different explanatory variables in the system and each $\pi_i$ is the column vector of coefficients for the $i^{th}$ equation (unless each equation contains all explanatory variables, there will be zeros in $\pi_i$ to allow for exclusion restrictions). Assuming multivariate normally distributed errors, the log likelihood function is

$$\log L = \sum\nolimits_{t=1}^{T} \log L_t = -\frac{MT}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| - \frac{1}{2}\sum\nolimits_{t=1}^{T} \epsilon_t'\Sigma^{-1}\epsilon_t \tag{4}$$

where, as defined earlier, $\Sigma = E(\epsilon_t \epsilon_t')$. The maximum likelihood estimates are found by taking the derivatives of (4) with respect to $\Pi$ and $\Sigma$, setting them equal to zero and solving.

## 2.4 Hypothesis Testing

We consider two types of tests – tests for contemporaneous correlation between errors and test for linear restrictions on the coefficients.

### 2.4.1 Contemporaneous Correlation

If there is no contemporaneous correlation between errors in different equations (i.e., $\Sigma$ is diagonal), then OLS equation-by-equation is fully efficient. Therefore, it is useful to test the following restriction

$$H_0 \quad : \quad \sigma_{ij} = 0 \; \forall i \neq j$$
$$H_A \quad : \quad H_0 \text{ false.}$$

Breusch and Pagan suggest using the Lagrange multiplier test statistic

$$\lambda = T \sum_{i=2}^{M} \sum_{j=1}^{i-1} r_{ij}^2$$

where $r_{ij}$ is calculated using the OLS residuals as follows

$$r_{ij} = \frac{e_i' e_j}{\sqrt{(e_i' e_i)(e_j' e_j)}}.$$

Under the null hypothesis, $\lambda$ is asymptotically chi-squared with $M(M-1)/2$ degrees of freedom.

### 2.4.2 Restrictions on Coefficients

The general $F$ test presented in chapter 6 can be extended to the SUR system. However, since the statistic requires using $\hat{\Sigma}$, the test will only be valid asymptotically. Consider testing the following $J$ linear restrictions

$$H_0 \quad : \quad R\beta = q$$
$$H_A \quad : \quad H_0 \text{ false}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_M)'$. Within the SUR framework, it is possible to test coefficient restrictions across equations. One possible test statistic is

$$W = (R\hat{\beta}_{FGLS} - q)'[Rvar(\hat{\beta}_{FGLS})R']^{-1}(R\hat{\beta}_{FGLS} - q)$$

which has an asymptotic chi-square distribution with $J$ degrees of freedom.

### 2.5 Autocorrelation

Heteroscedasticity and autocorrelation are possibilities within the SUR framework. I will focus on autocorrelation because SUR systems are often comprised of time series observations for each equation. Assume the errors follow

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{it}$$

where $\nu_{it}$ is white noise. The overall error structure will now be

$$E(\epsilon\epsilon') = \Omega = \begin{bmatrix} \sigma_{11}\Omega_{11} & \sigma_{12}\Omega_{12} & \cdots & \sigma_{1M}\Omega_{1M} \\ \sigma_{21}\Omega_{21} & \sigma_{22}\Omega_{22} & & \sigma_{2M}\Omega_{2M} \\ \vdots & & \ddots & \vdots \\ \sigma_{M1}\Omega_{M1} & \sigma_{M2}\Omega_{M2} & \cdots & \sigma_{MM}\Omega_{MM} \end{bmatrix}_{MT\times MT}$$

where

$$\Omega_{ij} = \begin{bmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & & \rho_j^{T-2} \\ \vdots & & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \cdots & 1 \end{bmatrix}_{T\times T}.$$

The following three-step approach is recommended

1. Run OLS equation-by-equation. Compute consistent estimate of $\rho_i$ (e.g., $\hat{\rho}_i = (\sum_{t=2}^{T} e_{i,t}e_{i.t-1})/(\sum_{t=1}^{T} e_{i.t}^2)$). Transform the data, using either Prais-Winsten or Cochrane-Orcutt, to remove the autocorrelation.
2. Estimate $\Sigma$ using the transformed data as suggested in (3).
3. Use $\hat{\Sigma}$ and equation (2) to calculate the FGLS estimates.

# 3 SUR Gauss Application

Consider the data taken from Woolridge (2002). The model attempts to explain wages and fringe benefits for 616 workers:

$$\begin{aligned} Wages_{1t} &= X_{1t}\beta_1 + \epsilon_{1t} \\ Benefits_{2t} &= X_{2,t}\beta_2 + \epsilon_{2t} \end{aligned}$$

where $X_{1t} = X_{2t}$ so that OLS and FGLS will produce equivalent results. Although OLS and FGLS are equivalent, one advantage of FGLS within a SUR framework is that it allows you test coefficient restrictions across equations. Doing OLS equation-by-equation would not allow such tests. The variables are defined as follows:

Dependent Variables

- Wages. Hourly earnings in 1999 dollars per hour.
- Benefits. Hourly benefits (vacation, sick leave, insurance and pension) in 1999 dollars per hour.

Explanatory Variables

- Education. Years of schooling.
- Experience. Years of work experience.
- Tenure. Years with current employer.
- Union. One if union member, zero otherwise.
- South. One if live in south, zero otherwise.
- Northeast. One if live in northeast, zero otherwise.
- Northcentral. One if live in northcentral, zero otherwise.
- Married. One if married, zero otherwise.
- White. One if white, zero otherwise.
- Male. One if male, zero otherwise.

See Gauss example 9 for further details.

## 4 The Simultaneous Equation Model

The simultaneous system can be written as

$$Y\Gamma + XB = E \tag{5}$$

where the variable matrices are

$$Y_{T\times M} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1M} \\ Y_{21} & Y_{22} & & Y_{2M} \\ \vdots & & \ddots & \vdots \\ Y_{T1} & Y_{T2} & \cdots & Y_{TM} \end{bmatrix};\ X_{T\times K} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1K} \\ X_{21} & X_{22} & & X_{2K} \\ \vdots & & \ddots & \vdots \\ X_{T1} & X_{T2} & \cdots & X_{TK} \end{bmatrix};\ E_{T\times M} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1M} \\ \epsilon_{21} & \epsilon_{22} & & \epsilon_{2M} \\ \vdots & & \ddots & \vdots \\ \epsilon_{T1} & \epsilon_{T2} & \cdots & \epsilon_{TM} \end{bmatrix}$$

and the coefficient matrices are

$$\Gamma_{M \times M} = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{M1} \\ \gamma_{12} & \gamma_{22} & & \gamma_{M2} \\ \vdots & & \ddots & \vdots \\ \gamma_{1M} & \gamma_{2M} & \cdots & \gamma_{MM} \end{bmatrix}; \ B_{K \times M} = \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{M1} \\ \beta_{12} & \beta_{22} & & \beta_{M2} \\ \vdots & & \ddots & \vdots \\ \beta_{1K} & \beta_{2K} & \cdots & \beta_{MK} \end{bmatrix}.$$

Some definitions.

- $Y_{t,j}$ is the $jth$ endogenous variable.
- $X_{t,j}$ is the $jth$ exogenous or predetermined variable
- Equations (5) are referred to as structural equations. $\Gamma$ and $B$ are the structural parameters.

To examine the assumptions about the error terms, rewrite the $E$ matrix as

$$\tilde{E} = vec(E) = (\epsilon_{11}, \epsilon_{21}, ..., \epsilon_{T1}, \epsilon_{12}, \epsilon_{22}, ..., \epsilon_{T2}, ..., \epsilon_{1M}, \epsilon_{2M}, ..., \epsilon_{TM})'.$$

We assume

$$\begin{aligned} E(\tilde{E}) &= 0 \\ E(\tilde{E}\tilde{E}') &= \Sigma \otimes I_T \end{aligned}$$

where the variance-covariance matrix for $\epsilon_t = (\epsilon_{t1}, \epsilon_{t2}, ..., \epsilon_{tM})'$ is

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21} & \cdots & \sigma_{M1} \\ \sigma_{12} & \sigma_{22} & & \sigma_{M2} \\ \vdots & & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{bmatrix}.$$

## 4.1 Reduced Form

The reduced-form solution to (5) is

$$Y = -XB\Gamma^{-1} + E\Gamma^{-1} = X\Pi + V$$

where $\Pi = -B\Gamma^{-1}$, $V = E\Gamma^{-1}$ and the error vector $\tilde{V} = vec(V)$ satisfies

$$\begin{aligned} E(\tilde{V}) &= 0 \\ E(\tilde{V}\tilde{V}') &= (\Gamma^{-1\prime}\Sigma\Gamma^{-1} \otimes I_T) = (\Omega \otimes I_T) \end{aligned}$$

where $\Sigma = \Gamma'\Omega\Gamma$.

### 4.2 Demand and Supply Example

Consider the following demand and supply equations

$$\begin{aligned} Q_t^s &= \alpha_0 + \alpha_1 P_t + \alpha_2 W_t + \alpha_3 Z_t + \epsilon_t^s \\ Q_t^d &= \beta_0 + \beta_1 P_t + \beta_3 Z_t + \epsilon_t^d \\ Q_t^s &= Q_t^d \end{aligned}$$

where $Q_t^s$, $Q_t^d$ and $P_t$ are endogenous variables and $W_t$ and $Z_t$ are exogenous variables. Let $Q = Q_t^s = Q_t^d$. In matrix form, the system can be written as

$$Y = \begin{bmatrix} Q_1 & P_1 \\ Q_2 & P_2 \\ \vdots & \vdots \\ Q_T & P_T \end{bmatrix}; \; X = \begin{bmatrix} 1 & W_1 & Z_1 \\ 1 & W_2 & Z_2 \\ \vdots & \vdots & \vdots \\ 1 & W_T & Z_T \end{bmatrix}; \; E = \begin{bmatrix} \epsilon_1^s & \epsilon_1^d \\ \epsilon_2^s & \epsilon_2^d \\ \vdots & \vdots \\ \epsilon_T^s & \epsilon_T^d \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} 1 & 1 \\ -\alpha_1 & -\beta_1 \end{bmatrix}; \; B = \begin{bmatrix} -\alpha_0 & -\beta_0 \\ -\alpha_2 & 0 \\ -\alpha_3 & -\beta_3 \end{bmatrix}$$

## 5 Identification

Identification Question. Given data on $X$ and $Y$, can we identify $\Gamma$, $B$ and $\Sigma$?

### 5.1 Estimation of $\Pi$ and $\Omega$

Begin by making the standard assumptions about the reduced form $Y = X\Pi + V$:

- $\text{plim}(\frac{1}{T}X'X) = Q$
- $\text{plim}(\frac{1}{T}X'V) = 0$

- $\text{plim}(\frac{1}{T}V'V) = \Omega$.

These assumptions imply that the equation-by-equation OLS estimates of $\Pi$ and $\Omega$ will be consistent.

## 5.2 Relationship Between $(\Pi, \Omega)$ and $(\Gamma, B, \Sigma)$

With these estimates ($\hat{\Pi}$ and $\hat{\Omega}$) in hand, the question is whether we can map back to $\Gamma$, $B$ and $\Sigma$? We know the following

1. $\Pi = -B\Gamma^{-1}$ and
2. $\Omega = \Gamma^{-1\prime}\Sigma\Gamma^{-1}$.

To see if identification is possible, we can count the number of known elements on the left-hand side and compare with the number of unknown elements on the right-hand side.

Number of Known Elements

- $KM$ elements in $\Pi$
- $\frac{1}{2}M(M+1)$ elements in $\Omega$

Total $= M(K + \frac{1}{2}(M+1))$.

Number of Unknown Elements

- $M^2$ elements in $\Gamma$
- $\frac{1}{2}M(M+1)$ elements in $\Sigma$
- $B = KM$

Total $= M(M + K + \frac{1}{2}(M+1))$.

Therefore, we are $M^2$ pieces of information shy of identifying the structural parameters. In other words, there is more than one set of structural parameters that are consistent with the reduced form. We say the model is underidentified.

## 5.3 Identification Conditions

There are several possibilities for obtaining identification:

1. Normalization (i.e., set $\gamma_{ii} = -1$ in $\Gamma$ for $i = 1, ..., m$).
2. Identities (e.g., national income accounting identity).
3. Exclusion restrictions (e.g., demand and supply shift factors).
4. Other linear (and nonlinear) restrictions (e.g., Blanchard-Quah long-run restriction).

### 5.3.1 Rank and Order Conditions

Begin by rewriting the $i^{th}$ equation from $\Pi = -B\Gamma^{-1}$ in matrix form as

$$\begin{bmatrix} \Pi & I_K \end{bmatrix} \begin{bmatrix} \Gamma_i \\ B_i \end{bmatrix} = 0 \tag{6}$$

where $\Gamma_i$ and $B_i$ represent the $i^{th}$ columns of $\Gamma$ and $B$, respectively. Since the rank of $[\Pi \; I_K]$ equals $K$, (6) represents a system of $K$ equations in $M+K-1$ unknowns (after normalization). In achieving identification, we will introduce linear restrictions as follows

$$R_i \begin{bmatrix} \Gamma_i \\ B_i \end{bmatrix} = 0 \tag{7}$$

where $Rank(R_i) = J$. Putting equations (6) and (7) together and redefining $\Delta_i = (\Gamma_i, B_i)'$ gives

$$\begin{bmatrix} (\Pi \vdots I_K) \\ R_i \end{bmatrix} \Delta_i = 0.$$

From this discussion, it is clear that $R_i$ must provide at least $M-1$ new pieces of information. Here are the formal rank and order conditions.

1. Order Condition. The order condition states that $Rank(R_i) = J \geq M-1$ is a necessary but not sufficient condition for identification. A situation where the order condition is not sufficient is when $R_i\Delta_j = 0$. More details on the order condition below.

2. Rank Condition. The rank condition states that $Rank(R_i\Delta) = M-1$ is a necessary and sufficient condition for identification.

We can now summarize all possible identification outcomes.

- Under Identification. If either $Rank(R_i) < M-1$ or $Rank(R_i\Delta) < M-1$, the $i^{th}$ equation is underidentified.

- Exact Identification. If $Rank(R_i) = M-1$ and $Rank(R_i\Delta) = M-1$, the $i^{th}$ equation is exactly identified.

- Over Identification. If $Rank(R_i) > M-1$ and $Rank(R_i\Delta) = M-1$, the $i^{th}$ equation is overidentified.

### 5.3.2 Identification Conditions in the Demand and Supply Example

Begin with supply and note that $M = 2$. The order condition is simple. Since all the variables are in the supply equation, there is no restriction matrix $R_s$ so that $Rank(R_s) = 0 < 1$. The supply equation is underidentified. There is no need to look at the rank condition.

Next, consider demand. The relevant matrix equations are

$$\begin{bmatrix} (\Pi \vdots I_K) \\ R_d \end{bmatrix} \Delta_d = \begin{bmatrix} \pi_{11} & \pi_{12} & 1 & 0 & 0 \\ \pi_{21} & \pi_{22} & 0 & 1 & 0 \\ \pi_{31} & \pi_{32} & 0 & 0 & 1 \\ -- & -- & -- & -- & -- \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -\beta_1 \\ -\beta_0 \\ -\beta_2 \\ -\beta_3 \end{bmatrix} = 0,$$

for which the order condition is clearly satisfied (i.e., $Rank(R_d) = 1 = M - 1$). For the rank condition, we need to find the rank of

$$R_d \Delta = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -\alpha_1 & -\beta_1 \\ -\alpha_0 & -\beta_0 \\ -\alpha_2 & 0 \\ -\alpha_3 & -\beta_3 \end{bmatrix} = \begin{bmatrix} -\alpha_2 & 0 \end{bmatrix}.$$

Clearly, $Rank(R_d\Delta) = 1 = M - 1$, so that the demand equation is exactly identified.

# 6 Limited-Information Estimation

We will consider five different limited-information estimation techniques – OLS, indirect least squares (ILS), instrumental variable (IV) estimation, two-stage least squares (2SLS) and limited-information maximum likelihood (LIML). The term limited information refers to equation-by-equation estimation, as opposed to full-information estimation which uses the linkages among the different equations.

Begin by writing the $i^{th}$ equation as

$$\begin{aligned} Y\Gamma_i + XB_i &= \epsilon_i \\ y_i &= Y_i\gamma_i + Y_i^*\gamma_i^* + X_i\beta_i + X_i^*\beta_i^* + \epsilon_i \end{aligned}$$

where $Y_i$ represents the vector of endogenous variables (other than $y_i$) in the $i^{th}$ equation, $Y_i^*$ represents the vector of endogenous variables excluded from the $i^{th}$ equation, and similarly for $X$. Therefore, $\gamma_i^* = 0$ and

$\beta_i^* = 0$ so that

$$\begin{aligned} y_i &= Y_i\gamma_i + X_i\beta_i + \epsilon_i \\ &= \begin{bmatrix} Y_i & X_i \end{bmatrix} \begin{bmatrix} \gamma_i \\ \beta_i \end{bmatrix} + \epsilon_i \\ &= Z_i\delta_i + \epsilon_i. \end{aligned}$$

### 6.1 Ordinary Least Squares (OLS)

The OLS estimator of $\delta_i$ is

$$\hat{\delta}_i^{OLS} = (Z_i'Z_i)^{-1}(Z_i'y_i).$$

The expected value of $\hat{\delta}_i$ is

$$E(\hat{\delta}_i^{OLS}) = \delta_i + E[(Z_i'Z_i)^{-1}Z_i'\epsilon_i].$$

However, since $y_i$ and $Y_i$ are jointly determined (recall $Z_i$ contains $Y_i$), we cannot expect that $E(Z_i'\epsilon_i) = 0$ or $\text{plim}(Z_i'\epsilon_i) = 0$. Therefore, OLS estimates will be biased and inconsistent. This is commonly known as simultaneity or endogeneity bias.

### 6.2 Indirect Least Squares (ILS)

The indirect least squares estimator simply uses the consistent reduced-form estimates ($\hat{\Pi}$ and $\hat{\Omega}$) and the relations $\Pi = -B\Gamma^{-1}$ and $\Omega = \Gamma^{-1\prime}\Sigma\Gamma^{-1}$ to solve for $\Gamma$, $B$ and $\Sigma$. The ILS estimator is only feasible if the system is exactly identified. To see this, consider the $i^{th}$ equation as given in (6)

$$\Pi\Gamma_i = -B_i$$

where $\hat{\Pi} = (X'X)^{-1}X'Y$. Substitution gives

$$(X'X)^{-1}X' \begin{bmatrix} y_i & Y_i \end{bmatrix} \begin{bmatrix} -1 \\ \hat{\gamma}_i \end{bmatrix} = \begin{bmatrix} -\hat{\beta}_i \\ 0 \end{bmatrix}.$$

Multiplying through by $(X'X)$ gives

$$\begin{aligned} -X'y_i + X'Y_i\hat{\gamma}_i &= -X'X \begin{bmatrix} -\hat{\beta}_i \\ 0 \end{bmatrix} \Rightarrow \\ X'y_i &= X'Y_i\hat{\gamma}_i + X'X_i\hat{\beta}_i \\ &= X'Z_i\hat{\delta}_i. \end{aligned}$$

Therefore, the ILS estimator for the $i^{th}$ equation can be written as

$$\hat{\delta}_i^{ILS} = (X'Z_i)^{-1}(X'y_i).$$

There are three cases:

1. If the $i^{th}$ equation is exactly identified, then $X'Z_i$ is square and invertible.
2. If the $i^{th}$ equation is underidentified, then $X'Z_i$ is not square.
3. If the $i^{th}$ equation is overidentified, then $X'Z_i$ is not square although a subset could be used to obtain consistent albeit inefficient estimates of $\delta_i$.

## 6.3 Instrumental Variable (IV) Estimation

Let $W_i$ be an instrument matrix (dimension $T \times (K_i + M_i)$) satisfying

- $\text{plim}(\frac{1}{T}W_i'Z_i) = \Sigma_{wz}$, a finite invertible matrix
- $\text{plim}(\frac{1}{T}W_i'W_i) = \Sigma_{ww}$, a finite positive-definite matrix
- $\text{plim}(\frac{1}{T}W_i'\epsilon_i) = 0$.

Since $\text{plim}(Z_i'\epsilon_i) \neq 0$, we can instead examine

$$\begin{aligned}
\text{plim}(\frac{1}{T}W_i'y_i) &= \text{plim}(\frac{1}{T}W_i'Z_i)\delta + \text{plim}(\frac{1}{T}W_i'\epsilon_i) \\
&= \text{plim}(\frac{1}{T}W_i'Z_i)\delta.
\end{aligned}$$

Naturally, the <u>instrumental variable</u> estimator is

$$\hat{\delta}_i^{IV} = (W_i'Z_i)^{-1}(W_i'y_i)$$

which is consistent and has asymptotic variance-covariance matrix

$$asy.var.(\hat{\delta}_i^{IV}) = \frac{\sigma_{ii}}{T}[\Sigma_{wz}^{-1}\Sigma_{ww}\Sigma_{zw}^{-1}].$$

This can be estimated by

$$est.asy.var.(\hat{\delta}_i^{IV}) = \hat{\sigma}_{ii}(W_i'Z_i)^{-1}W_i'W_i(Z_i'W_i)^{-1}$$

and

$$\hat{\sigma}_{ii} = \frac{1}{T}(y_i - Z_i\hat{\delta}_i^{IV})'(y_i - Z_i\hat{\delta}_i^{IV}).$$

A degrees of freedom correction is optional. Notice that ILS is a special case of IV estimation for an exactly identified equation where $W_i = X$.

### 6.4 Two-Stage Least Squares (2SLS)

When an equation in the system is overidentified (i.e., $rows(X'Z_i) > cols(X'Z_i)$), a convenient and intuitive IV estimator is the two-stage least squares estimator. The 2SLS estimator works as follows:

- Stage #1. Regress $Y_i$ on $X$ and form $\hat{Y}_i = X\hat{\Pi}^{OLS}$.
- Stage #2. Estimate $\delta_i$ by an OLS regression of $y_i$ on $\hat{Y}_i$ and $X_i$.

More formally, let $\hat{Z}_i = (\hat{Y}_i, X_i)$. The 2SLS estimator is given by

$$\hat{\delta}_i^{2SLS} = (\hat{Z}_i'\hat{Z}_i)^{-1}(\hat{Z}_i' y_i)$$

where the asymptotic variance-covariance matrix for $\hat{\delta}_i^{2SLS}$ can be estimated consistently by

$$est.asy.var(\hat{\delta}_i^{2SLS}) = \hat{\sigma}_{ii}(\hat{Z}_i'\hat{Z}_i)^{-1}$$

and

$$\hat{\sigma}_{ii} = \frac{1}{T}(y_i - Z_i\hat{\delta}_i^{2SLS})'(y_i - Z_i\hat{\delta}_i^{2SLS}).$$

### 6.5 Limited-Information Maximum Likelihood (LIML)

Limited-information maximum likelihood estimation refers to ML estimation of a single equation in the system. For example, if we assume normally distributed errors, then we can form the joint probability distribution function of $(y_i, Y_i)$ and maximize it by choosing $\delta_i$ and the appropriate elements of $\Sigma$. Since the LIML estimator is more complex but asymptotically equivalent to the 2SLS estimator, it is not widely used.

Note. When the $i^{th}$ equation is exactly identified, $\hat{\delta}_i^{ILS} = \hat{\delta}_i^{IV} = \hat{\delta}_i^{2SLS} = \hat{\delta}_i^{LIML}$.

## 7 Full-Information Estimation

The five estimators mentioned above are not fully efficient because they ignore cross-equation relationships between error terms and any omitted endogenous variables. We consider two fully efficient estimators below.

## 7.1 Three-Stage Least Squares (3SLS)

Begin by writing the system (5) as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & Z_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} \Rightarrow y = \bar{Z}\delta + \epsilon$$

where $\epsilon = \tilde{E}$ and

$$E(\epsilon\epsilon') = \Sigma \otimes I_T.$$

Then, applying the principle from SUR estimation, the fully efficient estimator is

$$\hat{\delta} = (\bar{W}'(\Sigma^{-1} \otimes I)\bar{Z})^{-1}(\bar{W}'(\Sigma^{-1} \otimes I)y)$$

where $\bar{W}$ indicates an instrumental variable matrix in the form of $\bar{Z}$. Zellner and Theil (1962) suggest the following three-stage procedure for estimating $\delta$.

- Stage #1. Calculate $\hat{Y}_i$ for each equation $(i = 1, ..., M)$ using OLS and the reduced form.
- Stage #2. Use $\hat{Y}_i$ to calculate $\hat{\delta}_i^{2SLS}$ and $\hat{\sigma}_{ij} = \frac{1}{T}(y_i - Z_i\hat{\delta}_i^{2SLS})'(y_j - Z_j\hat{\delta}_j^{2SLS})$.
- Stage #3. Calculate the IV-GLS estimator

$$\begin{aligned} \hat{\delta}^{3SLS} &= [Z'(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1\prime})Z]^{-1}[Z'(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1\prime})y] \\ &= [\hat{Z}'(\hat{\Sigma}^{-1} \otimes I)\hat{Z}]^{-1}[\hat{Z}'(\hat{\Sigma}^{-1} \otimes I)y]. \end{aligned}$$

The asymptotic variance-covariance matrix can be estimated by

$$\begin{aligned} est.asy.var(\hat{\delta}^{3SLS}) &= [Z'(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1\prime})Z]^{-1} \\ &= [\hat{Z}'(\hat{\Sigma}^{-1} \otimes I)\hat{Z}]^{-1}. \end{aligned}$$

## 7.2 Full-Information Maximum Likelihood (FIML)

The full-information maximum likelihood estimator is asymptotically efficient. Assuming multivariate normally distributed errors, we maximize

$$\ln L(\delta, \Sigma | y, Z) = -\frac{MT}{2}\ln(2\pi) + \frac{T}{2}\ln|\Sigma|^{-1} + T\ln|\Gamma| - \frac{1}{2}(y - Z\delta)'(\Sigma^{-1} \otimes I_T)(y - Z\delta)$$

by choosing $\Gamma$, $B$ and $\Sigma$. The FIML estimator can be computationally burdensome and has the same asymptotic distribution as the 3SLS estimator. As a result, most researchers use 3SLS.

# 8 Simultaneous Equations Gauss Application

Consider estimating a traditional Keynesian consumption function using quarterly data between 1947 and 2003. The simultaneous system is

$$\begin{aligned} C_t &= \beta_0 + \beta_1 DI_t + \epsilon_t^c & (8) \\ DI_t &= \beta_2 + C_t + I_t + G_t + NX_t + \epsilon_t^y & (9) \end{aligned}$$

where the variables are defined as follows:

Endogenous Variables

- $C_t$ – Consumption.
- $DI_t$ – Disposable Income.

Exogenous Variables

- $I_t$ – Investment.
- $G_t$ – Government Spending.
- $NX_t$ – Net Exports.

The conditions for identification of (8), the more interesting equation to be estimated, are shown below. Begin by writing the system as

$$Y\Gamma + XB = Y \begin{bmatrix} 1 & -1 \\ -\beta_1 & 1 \end{bmatrix} + X \begin{bmatrix} -\beta_0 & -\beta_2 \\ 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} = E$$

where $Y_t = (C_t, DI_t)$ and $X_t = (1, I_t, G_t, NX_t)$. The order condition depends on the rank of the restriction matrix for (8),

$$R_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

which is obviously $Rank(R_1) = 3 > M - 1 = 1$. Therefore, the model is overidentified if the rank condition is satisfied (i.e., $Rank(R_1 \Delta) = M - 1$). The relevant matrix for the rank condition is

$$R_1 \Delta = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -\beta_1 & 1 \\ -\beta_0 & -\beta_2 \\ 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix},$$

which has $Rank(R_1 \Delta) = 1$. Therefore, equation (8) is overidentified. Another way to see that (8) is overidentified is to solve for the reduced-form representation of $C_t$

$$\begin{aligned} C_t &= \beta_0 + \beta_1[\beta_2 + C_t + I_t + G_t + NX_t + \epsilon_t^y] + \epsilon_t^c \\ &= \frac{1}{1-\beta_1}[(\beta_0 + \beta_1\beta_2) + \beta_1 I_t + \beta_1 G_t + \beta_1 NX_t + (\epsilon_t^c + \beta_1 \epsilon_t^y)] \\ C_t &= \pi_0 + \pi_1 I_t + \pi_1 G_t + \pi_1 NX_t + v_t. \end{aligned} \tag{10}$$

Clearly, estimation of (10) is likely to produce three distinct estimates of $\pi_1$, which reflects the overidentification problem.

See Gauss examples 10 and 11 for OLS and 2SLS estimation of (8).