

## Section 11 Simultaneous Equations

The most crucial of our OLS assumptions (which carry over to most of the other estimators that we have studied) is that the regressors be *exogenous*—uncorrelated with the error term. This assumption is violated if we have “reverse causality” in which  $e \uparrow \rightarrow y \uparrow \rightarrow x \uparrow \downarrow$ .

### System estimation vs. single-equation

- The first essential question to ask in a situation where the regressor may be endogenous is “What is the model that determines the endogenous regressor?”
  - This question, which must be answered at least partially to use any of the techniques in this section, suggests that our single econometric equation should be thought of as part of a *system of simultaneous equations* that jointly determine both our  $y$  and our endogenous  $x$  variables.
  - For example, one of the most common applications in economics is attempting to estimate a demand curve: quantity is a function of price.
    - However, shocks to demand ( $e$ ) affect price, so price cannot generally be taken as exogenous.
    - The demand curve is part of a system of simultaneous equations along with the supply curve that jointly determine quantity and price.
  - Thinking of the joint determination of  $y$  and (at least some)  $x$  focuses our attention on a crucial set of variables: the exogenous variables that are in the “other” equation that determines  $x$  but that are not in the equation as separate determinants of  $y$ .
    - Whether we end up modeling the second equation explicitly or not, these variables are crucial to *identifying* the effects of  $x$  on  $y$ .
- The two main approaches to endogeneity revolve around our degree of interest in the determination of the endogenous regressors:
  - **System estimation** involves estimating a full set of equations with two or more dependent variables that are on the left-hand side of one equation and the right-hand side of others. (Example: both the supply and demand equations.)
  - **Single-equation estimation** involves estimating only the one equation of interest, but we still need to consider the variables that are in the other equation(s). (Example: estimate only the demand equation, but the exogenous variables in the supply equation are used as instruments.)

### Simultaneous equations and the identification problem

- In the simple case above, we had one endogenous variable on the right-hand side and one exogenous variable available to act as an instrument.

- In the more general case, there may be multiple endogenous variables and multiple instruments.
- This forces us to think about the problem of whether there is sufficient exogenous variation to *identify* the coefficients we want to estimate: the **identification problem**.
- We will examine an extended example of a set of supply and demand curves to explore the identification problem.

- **Model I:**

Demand curve:  $Q = \alpha_0 + \alpha_P P + u$

Supply curve:  $Q = \beta_0 + \beta_P P + v$

- Solving for the **reduced form**:

$$\beta_0 + \beta_P P + v = \alpha_0 + \alpha_P P + u$$

$$(\beta_P - \alpha_P)P = (\alpha_0 - \beta_0) + (u - v)$$

$$P = \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{u - v}{\beta_P - \alpha_P} \equiv \pi_{P0} + \varepsilon_P,$$

$$Q = \alpha_0 + \alpha_P \left[ \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{u - v}{\beta_P - \alpha_P} \right] + u$$

$$Q = \frac{\alpha_0(\beta_P - \alpha_P) + \alpha_P(\alpha_0 - \beta_0)}{\beta_P - \alpha_P} + \frac{u(\beta_P - \alpha_P) + \alpha_P(u - v)}{\beta_P - \alpha_P}$$

$$Q = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} + \frac{\beta_P u - \alpha_P v}{\beta_P - \alpha_P} \equiv \pi_{Q0} + \varepsilon_Q.$$

- The equations

$$P = \pi_{P,0} + \varepsilon_P$$

$$Q = \pi_{Q,0} + \varepsilon_Q$$

are called the reduced-form equations. We have solved the system of simultaneous linear equations for separate linear equations each of which has an endogenous variable on the left and none on the right.

- The  $\pi$  coefficients are the **reduced-form coefficients**: they are nonlinear combinations of the **structural coefficients**  $\alpha$  and  $\beta$ .
- We can estimate the reduced-form coefficients by OLS because there are no endogenous variables on the right-hand side.
- In this case, there are no variables at all on the RHS! We can estimate  $\pi_{P,0}$  and  $\pi_{Q,0}$  as the means of  $P$  and  $Q$ .
  - Does this give us enough information to **identify** the  $\alpha$  and  $\beta$  parameters?
  - No. There are four structural coefficients (two  $\alpha$  and two  $\beta$ ) and only two reduced-form coefficients ( $\pi$ ). There is no way to

construct a unique estimator of and of the  $\alpha$  or  $\beta$  coefficients from the estimate of  $\pi$ .

- Thus, in Model I *neither of the equations is identified*.

- Show graph: all variation in  $P$  and  $Q$  are due to unobserved error terms.

○ **Model II:**

Demand curve:  $Q = \alpha_0 + \alpha_P P + \alpha_M M + u$ , where  $M$  is income and is exogenous

Supply curve:  $Q = \beta_0 + \beta_P P + v$

- Solving for the reduced form:

$$\beta_0 + \beta_P P + v = \alpha_0 + \alpha_P P + \alpha_M M + u$$

$$P = \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M}{\beta_P - \alpha_P} M + \frac{u - v}{\beta_P - \alpha_P} \equiv \pi_{P0} + \pi_{PM} M + \varepsilon_P,$$

$$Q = \beta_0 + \beta_P \left[ \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M}{\beta_P - \alpha_P} M + \frac{u - v}{\beta_P - \alpha_P} \right] + v$$

$$Q = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M \beta_P}{\beta_P - \alpha_P} M + \frac{\beta_P u - \alpha_P v}{\beta_P - \alpha_P} \equiv \pi_{Q0} + \pi_{QM} M + \varepsilon_Q.$$

- Suppose we estimate the four reduced-form coefficients  $\pi_{P0}$ ,  $\pi_{PM}$ ,  $\pi_{Q0}$ ,  $\pi_{QM}$  by OLS. Can we identify the five structural coefficients?

- Obviously not: can't identify five coefficients uniquely from four.
- However, we can identify *some of them*:

$$\frac{\pi_{QM}}{\pi_{PM}} = \frac{\frac{\alpha_M \beta_P}{\beta_P - \alpha_P}}{\frac{\alpha_M}{\beta_P - \alpha_P}} = \beta_P$$

$$\pi_{Q0} - \beta_P \pi_{P0} = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} - \beta_P \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} = \beta_0.$$

- This is called **indirect least squares** and is an antiquated method for estimating these models.
- The presence of the income term in the demand equation *identifies* the slope and intercept of the supply equation. Changes in income affect demand but not supply, so we can use these changes to trace out the slope of the supply curve. How much does an increase income affect  $P$  and how much does it affect  $Q$ ?
  - The supply equation is **just identified** because there is only one way of extracting the structural parameters from the reduced-form parameters.
  - 2SLS of the supply equation using income as an instrument gives us the same estimator as ILS in the just-identified case.

- The demand equation is not identified: the only variation in the supply curve is the unobserved random shock.
- What would happen if income also affected supply?

○ **Model III:**

Demand curve:  $Q = \alpha_0 + \alpha_P P + \alpha_M M + u$

Supply curve:  $Q = \beta_0 + \beta_P P + \beta_M M + v$

- Solving for the reduced form:

$$\beta_0 + \beta_P P + \beta_M M + v = \alpha_0 + \alpha_P P + \alpha_M M + u$$

$$P = \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M - \beta_M}{\beta_P - \alpha_P} M + \frac{u - v}{\beta_P - \alpha_P} \equiv \pi_{P0} + \pi_{PM} M + \varepsilon_P,$$

$$Q = \beta_0 + \beta_P \left[ \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M - \beta_M}{\beta_P - \alpha_P} M + \frac{u - v}{\beta_P - \alpha_P} \right] + v$$

$$Q = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M \beta_P - \beta_M \alpha_P}{\beta_P - \alpha_P} M + \frac{\beta_P u - \alpha_P v}{\beta_P - \alpha_P} \equiv \pi_{Q0} + \pi_{QM} M + \varepsilon_Q.$$

- It's no longer possible to identify either equation. None of the six structural coefficients can be identified from estimates of the four reduced-form coefficients.
- We can no longer use changes in  $M$  to trace out either curve because it affects both curves.
- Note that nothing in the data has changed: we have merely changed our *assumption* (lens analogy) about how the data were generated.
  - If the assumption in Model II that income does not affect supply is incorrect, our estimates of the supply curve would be nonsense.

○ **Model IV:**

Demand curve:  $Q = \alpha_0 + \alpha_P P + \alpha_M M + u$

Supply curve:  $Q = \beta_0 + \beta_P P + \beta_R R + v$ , where  $R$  is rainfall (exogenous)

- Solving for the reduced form:

$$\beta_0 + \beta_P P + \beta_R R + v = \alpha_0 + \alpha_P P + \alpha_M M + u$$

$$P = \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M}{\beta_P - \alpha_P} M - \frac{\beta_R}{\beta_P - \alpha_P} R + \frac{u - v}{\beta_P - \alpha_P} \equiv \pi_{P0} + \pi_{PM} M + \pi_{PR} R + \varepsilon_P,$$

$$Q = \beta_0 + \beta_P \left[ \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M}{\beta_P - \alpha_P} M - \frac{\beta_R}{\beta_P - \alpha_P} R + \frac{u - v}{\beta_P - \alpha_P} \right] + \beta_R R + v$$

$$Q = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M \beta_P}{\beta_P - \alpha_P} M - \frac{\beta_R \alpha_P}{\beta_P - \alpha_P} R + \frac{\beta_P u - \alpha_P v}{\beta_P - \alpha_P} \equiv \pi_{Q0} + \pi_{QM} M + \pi_{QR} R + \varepsilon_Q.$$

- There are now six estimable coefficients and six structural coefficients we would like to estimate. Just identification of all coefficients is possible based on the numbers.
- In fact, as before,

$$\frac{\pi_{QM}}{\pi_{PM}} = \frac{\frac{\alpha_M \beta_P}{\beta_P - \alpha_P}}{\frac{\alpha_M}{\beta_P - \alpha_P}} = \beta_P$$

$$\pi_{Q0} - \beta_P \pi_{P0} = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} - \beta_P \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} = \beta_0.$$

- Now, we can do the same thing with the rainfall coefficients:

$$\frac{\pi_{QR}}{\pi_{PR}} = \frac{\frac{\alpha_P \beta_R}{\beta_P - \alpha_P}}{\frac{\beta_R}{\beta_P - \alpha_P}} = \alpha_P$$

$$-\pi_{P0}(\beta_P - \alpha_P) + \beta_0 = \alpha_0.$$

$$\pi_{PM}(\beta_P - \alpha_P) = \alpha_M$$

$$-\pi_{RM}(\beta_P - \alpha_P) = \beta_R$$

- Both equations are just identified:
  - Rainfall identifies the demand equation because it is exogenous, affects the endogenous variable price, and is not in the demand equation on its own.
  - Income identifies the supply equation because it is exogenous, affects the endogenous variable price, and is not in the supply equation on its own.
- Again, 2SLS gives us the same estimators as ILS in the just-identified case:
  - ivregress 2sls q m (p = r) to estimate the demand equation
  - ivregress 2sls q r (p = m) to estimate the supply equation

○ **Model V:**

Demand curve:  $Q = \alpha_0 + \alpha_P P + \alpha_M M + u$

Supply curve:  $Q = \beta_0 + \beta_P P + \beta_R R + \beta_W W + v$ , where  $W$  is wages (exogenous)

- We now have *two* exogenous variables in the supply equation that are not in the demand equation. Two alternative ways of identifying the demand curve.

- Solving for the reduced form:

$$\beta_0 + \beta_P P + \beta_R R + \beta_W W + \nu = \alpha_0 + \alpha_P P + \alpha_M M + u$$

$$P = \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M}{\beta_P - \alpha_P} M - \frac{\beta_R}{\beta_P - \alpha_P} R - \frac{\beta_W}{\beta_P - \alpha_P} W + \frac{u - \nu}{\beta_P - \alpha_P}$$

$$P \equiv \pi_{P0} + \pi_{PM} M + \pi_{PR} R + \pi_{PW} W + \varepsilon_P,$$

$$Q = \beta_0 + \beta_P \left[ \frac{\alpha_0 - \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M}{\beta_P - \alpha_P} M - \frac{\beta_R}{\beta_P - \alpha_P} R - \frac{\beta_W}{\beta_P - \alpha_P} W + \frac{u - \nu}{\beta_P - \alpha_P} \right] + \beta_R R + \beta_W W + \nu$$

$$Q = \frac{\beta_P \alpha_0 - \alpha_P \beta_0}{\beta_P - \alpha_P} + \frac{\alpha_M \beta_P}{\beta_P - \alpha_P} M - \frac{\beta_R \alpha_P}{\beta_P - \alpha_P} R - \frac{\beta_W \alpha_P}{\beta_P - \alpha_P} W + \frac{\beta_P u - \alpha_P \nu}{\beta_P - \alpha_P}$$

$$Q \equiv \pi_{Q0} + \pi_{QM} M + \pi_{QR} R + \pi_{QW} W + \varepsilon_Q.$$

- There are now eight estimable reduced-form coefficients and seven structural coefficients.
- All six of the equations that we used in Model IV to get the those six coefficients still work.
- Now we can estimate  $\alpha_P$  *either* as  $\alpha_P = \frac{\pi_{QR}}{\pi_{PR}}$  or as  $\alpha_P = \frac{\pi_{QW}}{\pi_{PW}}$ .
  - The demand equation is now **overidentified** because there are two exogenous variables that affect the single endogenous variable  $P$  that are not separately in the demand equation.
  - Will they be the same? Will  $\frac{\pi_{QR}}{\pi_{PR}} = \frac{\pi_{QW}}{\pi_{PW}}$ ?
  - Generally they won't be identical even if the model is correct because of sampling error. Is there more inequality than would be expected randomly?
  - We can test this nonlinear null hypothesis.
    - If the model is valid, we should not be able to reject this null hypothesis.
    - Rejecting these **overidentifying restrictions** suggests that the model is not valid.
- There are two different ILS estimates for the coefficients of the demand equation. 2SLS will be a combination of them.
  - Estimate demand equation by ivregress 2sls q m (p = r w)
  - The instrument used is the prediction of  $Q$  based on  $R$  and  $W$ .
- Note several properties of identification
  - Identification is usually by equation/coefficient, not necessarily of the whole system.
  - It's possible to have one equation that is identified with others not.

- If there are multiple endogenous regressors it is possible to have one identified and others not.
- Identification depends crucially on three assumptions:
  - That the instrument is exogenous
  - That the instrument does not itself appear in the equation
  - That the instrument does appear in another equation that influences the endogenous regressor
  - If any of these assumptions is violated, then the 2SLS estimator is biased and inconsistent.
- In general, there needs to be one omitted exogenous variable for each included endogenous variable. (Order condition for identification)
  - However, if you have two instruments that are correlated with one endogenous variable but neither is correlated with the other, then identification of the second endogenous regressor fails.
  - Order is not enough; the rank condition applies as well.
- Matrix notation of the 2SLS estimator
  - Consider our 2-equation system of Model V.
    - Let  $y = [Q \ P]$  be an  $N \times 2$  matrix of the two endogenous variables.
    - Let  $Z = [1 \ M \ R \ W]$  be an  $N \times 4$  matrix of the four exogenous variables (which are instruments for one equation and included exogenous variables for the other).
    - Let  $e = [u \ v]$  be an  $N \times 2$  matrix of error terms, which are probably correlated within a single observation.
    - Let  $\Gamma$  be the  $2 \times 2$  matrix of coefficients applied to the endogenous variables (which will often be 1 or  $-1$  by normalization)
    - Let  $B$  be the  $4 \times 2$  matrix of coefficients applied to the exogenous variables, which must have some elements that are known to be zero in order for identification to be achieved.
    - The two equations of the model can be written as  $Y\Gamma + ZB = e$ , where

$$Y = \begin{pmatrix} Q_1 & P_1 \\ Q_2 & P_2 \\ \vdots & \vdots \\ Q_n & P_n \end{pmatrix}, \quad \Gamma = \begin{pmatrix} -1 & -1 \\ \alpha_P & \beta_P \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & M_1 & R_1 & W_1 \\ 1 & M_2 & R_2 & W_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & M_n & R_n & W_n \end{pmatrix},$$

$$B = \begin{pmatrix} \alpha_0 & \beta_0 \\ \alpha_M & 0 \\ 0 & \beta_R \\ 0 & \beta_W \end{pmatrix}, \quad e = \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \\ \vdots & \vdots \\ u_n & v_n \end{pmatrix}.$$

- The reduced-form equations are obtained by post-multiplying the equation by the inverse of  $\Gamma$  (which must exist for the model to be

solvable):  $Y\Gamma^{-1} + ZB\Gamma^{-1} = e\Gamma^{-1}$ , or  $Y = -ZB\Gamma^{-1} + e\Gamma^{-1} \equiv Z\Pi + \varepsilon$ , where

$$\Pi \equiv -B\Gamma^{-1} \text{ and } \varepsilon \equiv e\Gamma^{-1} = \begin{pmatrix} \varepsilon_{Q1} & \varepsilon_{P1} \\ \varepsilon_{Q2} & \varepsilon_{P2} \\ \vdots & \vdots \\ \varepsilon_{Qn} & \varepsilon_{Pn} \end{pmatrix}.$$

- If the system is identified, then there are enough restrictions on the  $\Gamma$  and  $B$  matrices (five in the model above—two  $-1$ s and three  $0$ s) to assure that the remaining elements can be obtained uniquely from the  $\Pi$  matrix.

## Estimation of systems of equations (not in text)

- The method of instrumental variables offers us a means to estimate a single equation from a larger system of simultaneous equations. Sometimes we want or need to estimate the entire system.
  - Estimates are generally more efficient if all equations are estimated together.
    - Taking account of the correlation between the error terms is beneficial
    - Suppose that we know that the error terms are positively correlated and that equation 2 seems to have a large positive error for observation  $i$
    - Joint estimation allows us to take account of the likely positive error in equation 1 and not attempt to fit the outlying observation too closely
    - Adds information and thus improves efficiency
  - We may want to impose and/or test coefficient restrictions across the equations of a system.
    - Demand equations derived from a common utility function (or factor demands from a common cost function) have cross-equation “symmetry” restrictions. (The Slutsky condition for demand says that the income-compensated cross-price elasticity of demand for  $x$  with respect to the price of  $y$  equals the elasticity of demand for  $y$  with respect to the price of  $x$ .)
    - Might want to test whether the income elasticity of demand for apples exceeds that of bananas.
    - Might want to test whether all of the coefficients of the demand for apples are the same as those of bananas so that we can aggregate them together
- Two kinds of joint-system estimation
  - **Seemingly unrelated regressions (SUR)** (also called Zellner-efficient regression)
    - System of equations with no endogenous variables on right-hand side
    - Efficiency gains from taking account of correlation of error
    - Possibility of testing/imposing cross-equation coefficient restrictions
  - **Three-state least squares (3SLS)**



- System of equations with endogenous regressors
- Example would be estimating both demand and supply equations together
- Adds efficiency gains (or cross-equation tests) to 2SLS/IV consistent estimator of equation(s) with endogenous regressors
- Estimation by seemingly unrelated regressions
  - Here we have a set of equations that have no endogenous regressors, but we want to estimate the equations jointly.
  - We can do this by “stacking” the regressions:
    - Suppose that there are 3 equations to be jointly estimated with dependent variables  $y_1$ ,  $y_2$ , and  $y_3$ , sets of regressors (which might overlap)  $X_1$ ,  $X_2$ , and  $X_3$ , and error terms  $e_1$ ,  $e_2$ , and  $e_3$ .
    - Separately, the equations can be written
 
$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{e}_1,$$

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}_2,$$

$$\mathbf{y}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{e}_3.$$
    - Let  $\mathbf{y}$  be the  $3N \times 1$  element column vector that stacks the 3  $\mathbf{y}$  vectors:
 
$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix}.$$
    - Let  $\mathbf{X}$  be the  $3N \times (K_1 + K_2 + K_3)$  matrix  $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 \end{pmatrix}.$
    - Let  $\mathbf{e}$  be the  $3N \times 1$  element column vector  $\mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix}.$
    - Let  $\boldsymbol{\beta}$  be the  $(K_1 + K_2 + K_3) \times 1$  element vector  $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{pmatrix}.$
    - We can then write the combined system of equations as  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$
  - Can we estimate this system by OLS?
    - Yes, except this not efficient because of probably correlation between the  $i$ th observation's error term across equations.
  - Specification of error term
    - If observations are IID, then,  $\text{cov}(e_{mi}, e_{ji}) = 0$  if  $i \neq j$ . (First subscript is equation; second is the observation.)
    - However, it is likely that within each observation,  $\text{cov}(e_{mi}, e_{ji}) = \sigma_{ml} \neq 0.$
    - Heteroskedasticity is also almost certain since we have different dependent variables for each third of the stacked regression.

- Let  $\Sigma_u \equiv \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$

- Assume that there is no correlation across observations either within any of the equations or between them.
- Then the covariance matrix of the stacked error term is the  $3n \times 3n$

$$\text{matrix } \Omega = \Sigma \otimes \mathbf{I}_N = \begin{pmatrix} \sigma_{11}\mathbf{I}_N & \sigma_{12}\mathbf{I}_N & \sigma_{13}\mathbf{I}_N \\ \sigma_{12}\mathbf{I}_N & \sigma_{22}\mathbf{I}_N & \sigma_{23}\mathbf{I}_N \\ \sigma_{13}\mathbf{I}_N & \sigma_{23}\mathbf{I}_N & \sigma_{33}\mathbf{I}_N \end{pmatrix}.$$

- Given the non-scalar covariance matrix, this is another potential application of generalized least square:  $\hat{\beta}_{GLS} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\mathbf{y}$ .
  - (This general formula specializes to weighted least-squares when there is heteroskedasticity but no autocorrelation. We will also see a GLS application for serial correlation of the error.)
- Of course, we don't know  $\sigma_{ml}$ , so we must estimate it.
  - We can do so based on OLS residuals because OLS is consistent (if not efficient).
- SUR is a two-step procedure
  - First estimate the three regressions by OLS and calculate the residual vectors  $\hat{e}_1, \hat{e}_2$ , and  $\hat{e}_3$ .
  - Next estimate  $\hat{\sigma}_{ml} = \frac{1}{n} \sum_{i=1}^n \hat{e}_{mi} \hat{e}_{li}$ ,  $m = 1, 2, 3$ ;  $l = 1, 2, 3$ , and assemble these estimators into  $\hat{\Sigma}_e \equiv \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} & \hat{\sigma}_{23} \\ \hat{\sigma}_{13} & \hat{\sigma}_{23} & \hat{\sigma}_{33} \end{pmatrix}$  and  $\hat{\Omega} \equiv \hat{\Sigma}_e \otimes \mathbf{I}_n$ .
    - Finally, use “feasible” GLS to estimate  $\beta$  as  $\hat{\beta}_{FGLS} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\Omega}^{-1}\mathbf{y}$ .
- This procedure can be iterated:
  - Because  $\hat{\beta}_{FGLS}$  is a more efficient estimator than the OLS estimator, we should get “better” residuals by calculating them based on  $\hat{\beta}_{FGLS}$  rather than on OLS.
  - Iterated seemingly unrelated regressions (ISUR) repeatedly re-estimates  $\sigma_{ml}$  based on the FGLS coefficient estimator, then recalculates  $\hat{\Omega}$  and re-estimates  $\beta$  by FGLS.
  - This can be repeated over and over until the elements of  $\hat{\Omega}$  do not change from iteration to iteration.
- SUR is more efficient than separate OLS except in two situations (in which they are identical):

- First, there is no correlation between error terms across equations. In other words,  $\sigma_{ml} = 0$  for all  $m \neq l$ .
    - Second, the same regressors appear in all equations:  $\mathbf{X}_m = \mathbf{X}_l$  for all  $m, l$ .
  - In Stata, we use sureg (dvar1 indvars1) (dvar2 indvars2) (dvar3 indvars3)
    - The option isure iterates to convergence.
    - The option constraints ([dvar1]indvar1j = [dvar2]indvar2j) imposes the constraint that the indvar1j coefficient in the equation for dvar1 equals the indvar2j coefficient in the equation for dvar2.
    - If constraints are complex, can also use
      - constraint 1 [dvar1]indvar1j = [dvar2]indvar2j
      - constraint 2 [dvar2]indvar2j = [dvar3]indvar3j
- sureg (dvar1 indvars1) (dvar2 indvars2) (dvar3 indvars3), constraint(1 2)
- Estimation by three-stage least squares
  - If endogenous variables appear on the RHS of equations, then we must combine the system estimation of SUR with the instrumental variables method of 2SLS.
  - The resulting estimator is 3SLS:
    - Estimate the reduced-form equations by OLS.
      - Don't need SUR because all exogenous variables in the system appear in each equation, so the  $\mathbf{X}_m$  matrices are identical and OLS is equivalent to SUR.
      - Calculate fitted values of the endogenous variables based in the reduced-form regressions on the exogenous variables as in 2SLS.
    - Estimate the individual equations by 2SLS, using their fitted values in place of the endogenous regressors.
      - Calculate the residuals of each equation from the 2SLS regressions.
      - Calculate estimates of  $\sigma_{ml}$  and assemble them into  $\hat{\Omega}$ .
    - Estimate the system of equations jointly by FGLS using the estimated  $\hat{\Omega}$ .
      - As with SUR, this can be iterated.
  - 3SLS has the same advantages relative to 2SLS that SUR has relative to OLS:
    - Efficiency gain by taking account of cross-equation correlation of error (if it exists)
    - Possibility of imposing or testing cross-equation coefficient restrictions
  - Stata will do 3SLS using the reg3 command, which combines the forms of the ivregress (with endregs = instvars in the variable list of each regression) and multiple equations enclosed in parentheses.
- **Maximum-likelihood estimators for simultaneous equations**
  - Unlike OLS, 2SLS is *not* an MLE, nor is 3SLS.
  - There are MLEs that apply to these models under (usually) the normal distribution.

- **Limited-information maximum likelihood** is a single-equation MLE that is analogous to 2SLS.
- **Full-information maximum likelihood** is the multiple-equation MLE analog of 3SLS.
- Stata will do LIML (and GMM) by changing the 2SLS option in the ivregress command to liml or gmm.
  - Each of these has different options that will need to be set.
- I can't find any FIML procedure in Stata, but it may be possible to program it in the general-purpose ml command.