Simultaneous Equations

1 Motivation and Examples

Now we relax the assumption that $EX'u = 0$. This wil require new techniques:

1. Instrumental variables
2. 2- and 3-stage least squares
3. Limited (LIML) and full (FIML) information maximum likelihood

Also it is no longer clear if you can estimate parameters at all. This is the "identification problem."

1.1 Example 1: Consumption

$$\begin{aligned} c_t &= \alpha + \beta y_t + u_t; \\ y_t &= c_t + i_t. \end{aligned}$$

The second equation is called an identity because there is no error. The two equations together are the structure of the model or the structural equations. In this model, c_t and y_t are endogenous because they are determined inside the model, and i_t and u_t are exogenous because they are determined outside the model. In the first equation, 1 and y_t are explanatory variables. If c_{t-1} were in the model, it would be predetermined (and therefore maybe have many of the characteristics of exogenous variables), but it would not be exogenous.

In this model, there is correlation between y_t (one of the explanatory variables) and u_t (the error). Why? We can write our structure as

$$\begin{aligned} c_t &= \alpha + \beta y_t + u_t \\ &= \alpha + \beta (c_t + i_t) + u_t \\ &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} i_t + \frac{1}{1-\beta} u_t; \end{aligned} \tag{1}$$

$$\begin{aligned} y_t &= c_t + i_t \\ &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} i_t + i_t + \frac{1}{1-\beta} u_t \\ &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} i_t + \frac{1}{1-\beta} u_t. \end{aligned} \tag{2}$$

Equations (1) and (2) are called reduced form equations because they represent each endogenous variable as a function of only exogenous variables. We can write the reduced form equations as

$$\begin{aligned} c_t &= \pi_{01} + \pi_{11} i_t + v_{1t} \\ y_t &= \pi_{02} + \pi_{12} i_t + v_{2t} \end{aligned}$$

where

$$\begin{aligned}\pi_{01} &= \frac{\alpha}{1-\beta}, \\ \pi_{11} &= \frac{\beta}{1-\beta}, \\ \pi_{02} &= \frac{\alpha}{1-\beta}, \\ \pi_{12} &= \frac{1}{1-\beta}\end{aligned}$$

are reduced form parameters and

$$v_{1t} = v_{2t} = \frac{1}{1-\beta} u_t$$

are reduced form errors. Note the restrictions on the reduced form parameters and errors such as (but not limited to)

$$\pi_{01} = \pi_{02}; v_{1t} = v_{2t}.$$

Also note that we can solve for (possibly nonunique) values of the structural parameters in terms of the reduced form parameters. For example,

$$\begin{aligned}\beta &= \frac{\pi_{11}}{\pi_{12}}; \\ \alpha &= \pi_{01}\left(1 - \frac{\pi_{11}}{\pi_{12}}\right).\end{aligned} \tag{3}$$

So a feasible estimation approach is to estimate $\widehat{\pi}$ using OLS. OLS provides a consistent estimate of π. Why? Then solve for $\left(\widehat{\alpha}, \widehat{\beta}\right)$ using equation (3). This is called the indirect least squares estimate of (α, β), and it is consistent. Why? However, $\left(\widehat{\alpha}, \widehat{\beta}\right)$ is biased (why?) and nonunique (why?).

If, instead, we used OLS on the structural equation for c_t, we would get

$$\widehat{\beta} = \frac{\sum_t \widetilde{y}_t \widetilde{c}_t}{\sum_t \widetilde{y}_t^2} = \frac{\sum_t \widetilde{y}_t \left(\beta \widetilde{y}_t + \widetilde{u}_t\right)}{\sum_t \widetilde{y}_t^2} = \beta + \frac{\sum_t \widetilde{y}_t \widetilde{u}_t}{\sum_t \widetilde{y}_t^2}.$$

But

$$\widetilde{y}_t = \frac{1}{1-\beta}\widetilde{i}_t + \frac{1}{1-\beta}\widetilde{u}_t$$

(from the reduced form equation). So

$$E\widetilde{y}_t\widetilde{u}_t = E\left(\frac{1}{1-\beta}\widetilde{i}_t + \frac{1}{1-\beta}\widetilde{u}_t\right)\widetilde{u}_t = \frac{1}{1-\beta}E\widetilde{u}_t^2 = \frac{\sigma_u^2}{1-\beta}.$$

Thus,

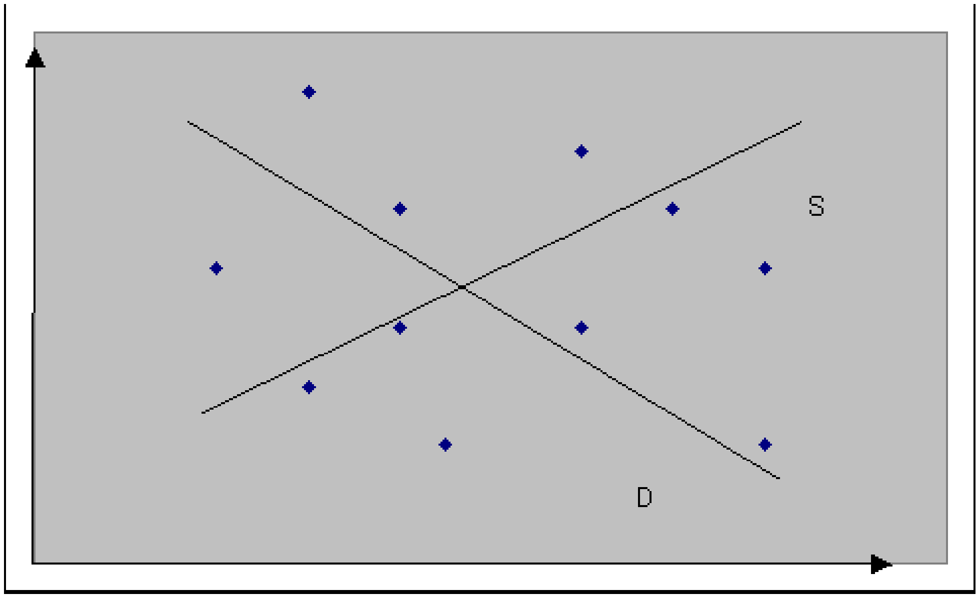
$$\begin{aligned}
plim\ \widehat{\beta} &= \beta + plim \frac{\sum_t \widetilde{y}_t \widetilde{u}_t}{\sum_t \widetilde{y}_t^2} \\
&= \beta + \frac{plim \frac{1}{T} \sum_t \widetilde{y}_t \widetilde{u}_t}{plim \frac{1}{T} \sum_t \widetilde{y}_t^2} \\
&= \beta + \frac{\frac{\sigma_u^2}{1-\beta}}{\left(\frac{1}{1-\beta}\right)^2 \sigma_i^2 + \left(\frac{1}{1-\beta}\right)^2 \sigma_u^2} \\
&= \beta + \frac{(1-\beta)\,\sigma_u^2}{\sigma_i^2 + \sigma_u^2} > \beta.
\end{aligned}$$

1.2 Example 2: Bananas

Consider the structural model below describing the supply and demand for bananas:

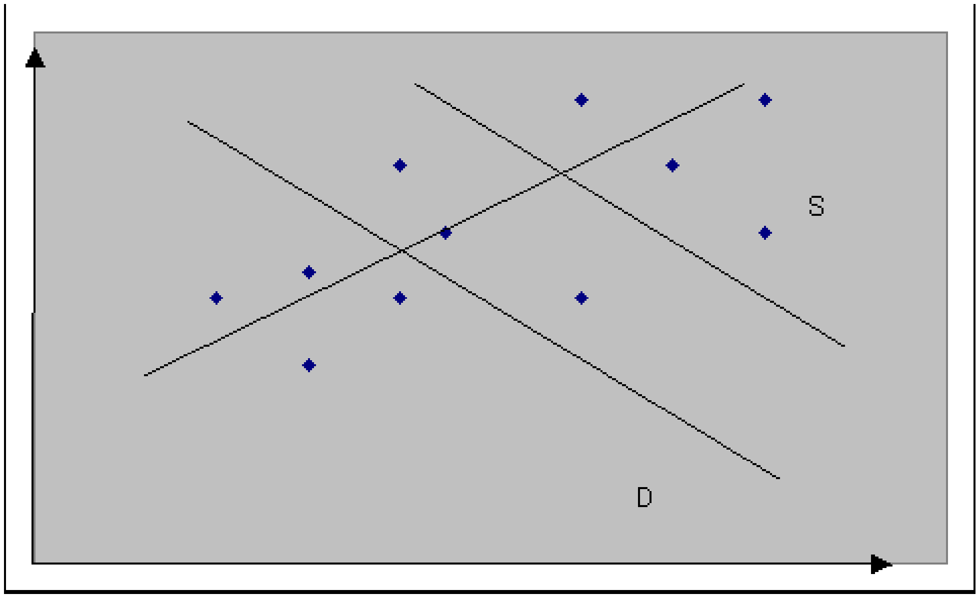
$$\begin{aligned}
q_t^d &= \alpha + \beta p_t + e_t \\
q_t^s &= \gamma + \delta p_t + u_t \\
q_t^d &= q_t^s.
\end{aligned}$$

We never observe supply or demand; we observe only the equilibrium price and quantity. Consider the scatter plot of combinations of equilibrium prices and



quantities. It is not at all obvious how to fit a supply and demand curve to them especially when both are moving around over time because of the errors (e_t and u_t). We try to analyze this problem more rigorously below so we don't slip on the banana problem.

First define
$$q_t = q_t^d = q_t^s,$$
and rewrite the structural equations as
$$\begin{aligned} q_t &= \alpha + \beta p_t + e_t \\ q_t &= \gamma + \delta p_t + u_t. \end{aligned}$$
If we regress q_t as a function of p_t, what are we estimating? Consider the reduced form of the model:
$$\alpha + \beta p_t + e_t = \gamma + \delta p_t + u_t$$
$$\Rightarrow \quad p_t = \frac{\gamma - \alpha}{\beta - \delta} + \frac{u_t - e_t}{\beta - \delta}$$
$$\Rightarrow \quad q_t = \alpha + \beta \frac{\gamma - \alpha}{\beta - \delta} + e_t + \beta \frac{u_t - e_t}{\beta - \delta}.$$
We can estimate
$$\pi_p = \frac{\gamma - \alpha}{\beta - \delta}$$
and
$$\pi_q = \alpha + \beta \frac{\gamma - \alpha}{\beta - \delta}.$$



But knowledge of π_p and π_q will not provide us with any information about the structural parameters. We say that all of the structural parameters are not identified.

Now consider changing the structural model to

$$\begin{aligned} q_t^d &= \alpha + \beta p_t + \theta y_t + e_t \\ q_t^s &= \gamma + \delta p_t + u_t \\ q_t^d &= q_t^s. \end{aligned}$$

In the new model, we know that the demand curve is shifting in a way predicted by the exogenous variable, y_t. This is going to allow us to plot a supply curve. However we still won't be able to say much about the demand curve. Consider the problem more rigorously. First, as before, write

$$\begin{aligned} q_t &= \alpha + \beta p_t + \theta y_t + e_t \\ q_t &= \gamma + \delta p_t + u_t. \end{aligned}$$

Now when we solve for the reduced form parameters, we get

$$\begin{aligned} p_t &= \tfrac{\alpha-\gamma}{\delta-\beta} + \tfrac{\theta}{\delta-\beta} y_t + \tfrac{e_t - u_t}{\delta-\beta} &&= \pi_{0p} + \pi_{1p} y_t + v_{pt} \\ q_t &= \tfrac{\delta\alpha-\gamma\beta}{\delta-\beta} + \tfrac{\delta\theta}{\delta-\beta} y_t + \tfrac{\delta e_t - \beta u_t}{\delta-\beta} &&= \pi_{0q} + \pi_{1q} y_t + v_{qt} \end{aligned}.$$

So, given consistent estimates of π_{1p} and π_{1q} (e.g., from OLS), we can estimate

$$\widehat{\delta} = \widehat{\pi}_{1q} / \widehat{\pi}_{1p};$$

thus δ is identified. Similarly, we can set

$$\widehat{\gamma} = \widehat{\pi}_{0q} - \widehat{\delta}\widehat{\pi}_{0p} + \widehat{\pi}_{1q}.$$

The students should try (at home) the structural model

$$\begin{aligned} q_t^d &= \alpha + \beta p_t + \theta y_t + e_t \\ q_t^s &= \gamma + \delta p_t + \phi z_t + u_t \\ q_t^d &= q_t^s. \end{aligned}$$

Note that, if the parameter is not identified, then no amount of data will allow us to estimate it. Thus, before we can even discuss estimation, we must first determine identification.

Over the rest of this topic, we will discuss:

1. Concepts and Terminology
2. Identification
3. Estimation
4. Properties of Estimators

2 Concepts and Terminology

The general structural specification of our model is

$$\underset{n\times n}{B}\underset{n\times 1}{y_t} = \underset{n\times m}{C}\underset{m\times 1}{x_t} + \underset{n\times 1}{u_t}.$$

By stacking equations (over time) next to each other, we get

$$\underset{n\times n}{B}\underset{n\times T}{Y'} = \underset{n\times m}{C}\underset{m\times T}{X'} + \underset{n\times T}{U'}.$$

We assume that:

1. B is nonsingular;
2. All diagonal elements of B are 1 (scaling convention);
3. $u_t \sim iid\,(0, \Omega)$ with Ω positive definite;
4. x_t are exogenous;
5. $\exists$ a priori (from theory) restrictions on B and C.

The reduced form of the model is

$$Y' = B^{-1}CX' + B^{-1}U' = AX' + V'.$$

3 Identification

We observe $\widehat{A}$, but we want to observe $\widehat{B}$ and $\widehat{C}$. $\exists\, n^2 - n$ unknown parameters in B, nm unknown parameters in C, and nm known parameters in $\widehat{A}$. We have nm equations in

$$A = B^{-1}C$$

to explain $n^2 - n + nm$ parameters. We can't do it; we need to place restrictions (from theory) on B and C to reduce the number of unknown parameters.

Possible restrictions:

1. $B_{ij} = 0$ or $C_{ij} = 0$ (exclusion or zero restrictions);
2. $$\sum_{i,j} \gamma_{ij} B_{ij} + \sum_{i,j} \delta_{ij} C_{ij} = \alpha;$$
3. Restrictions on Ω or nonlinear restrictions.

By far, the most popular restrictions are exclusion restrictions.

Things that can happen:

1. No structural parameters are identified;
2. Some structural parameters are identified;
3. Some structural parameters are identified, and there are restrictions on reduced form parameters (overidentified);
4. All structural parameters are identified, and there are restrictions on reduced form parameters (overidentified);
5. All structural parameters are identified (just identified).

We also can discuss identifying an equation. If all of the structural parameters in a structural equation are identified, then the equation is identified.

3.1 Necessary and Sufficient Conditions for Identification

Let

$$P = [B \mid C] = \begin{bmatrix} P_1' \\ P_2' \\ \vdots \\ P_n' \end{bmatrix}.$$

Write the jth restriction on the ith equation as

$$P_i' \phi_{ij} = 0.$$

Assume $\exists R_i$ restrictions on the ith equation, and let

$$\Phi_i = \left[\phi_{i1}, \phi_{i2}, .., \phi_{iR_i}\right].$$

Then we can write all of the restrictions on the ith equation as

$$P_i' \Phi_i = 0.$$

Example 1 *Example:*

$$\begin{aligned} y_{1t} &= \beta_{12} y_{2t} + u_{1t} \\ y_{2t} &= \gamma_{21} x_{1t} + \gamma_{22} x_{2t} + u_{2t}. \end{aligned}$$

Then

$$P = \begin{bmatrix} 1 & -\beta_{12} & \gamma_{11} & \gamma_{12} \\ -\beta_{22} & 1 & \gamma_{21} & \gamma_{22} \end{bmatrix}$$

with restrictions

$$\gamma_{11} = \gamma_{12} = \beta_{22} = 0.$$

$\Rightarrow$

$$\Phi_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \Phi_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Theorem 2 *The ith equation is identified iff Rank $(P\Phi_i) = n - 1$.*

An implication of this theorem is that a necessary condition for the ith equation to be identified is $R_i \geq n - 1$. This is called the order condition. While it is not sufficient for identification, it is necessary and very easy to verify.

Example 3 *(continued from Example 1)*

$$P\Phi_1 = \begin{bmatrix} 1 & -\beta_{12} & \gamma_{11} & \gamma_{12} \\ -\beta_{22} & 1 & \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}.$$

If the restrictions are true, then

$$\begin{aligned} P\Phi_1 &= \begin{bmatrix} 0 & 0 \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \\ &\Rightarrow \text{Rank}\,(P\Phi_1) = 1 \end{aligned}$$

unless $\gamma_{21} = \gamma_{22} = 0$. What would it mean for $\gamma_{21} = \gamma_{22} = 0$? Therefore the first equation is identified unless $\gamma_{21} = \gamma_{22} = 0$. Also

$$P\Phi_2 = \begin{bmatrix} 1 & -\beta_{12} & \gamma_{11} & \gamma_{12} \\ -\beta_{22} & 1 & \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -\beta_{22} \end{bmatrix}$$

whose rank is 1. $\Rightarrow$ the second equation is identified.

Example 4 *Consider a crime rate example developed by students in class.*

3.1.1 Some Formalism and Proof of the Identification Theorem

Definition 5 *A structure of a model is a set of parameters and functions which completely define the stochastic relationships between the endogenous variables and the exogenous variables. We can denote the true structure as* $\left(B^0, C^0, f^0\right)$ *where* f *is the joint density of the errors.*

Definition 6 *Two structures,* $\left(B^0, C^0, f^0\right)$ *and* $\left(B^1, C^1, f^1\right)$ *are equivalent if they imply the same joint distribution for the endogenous variables (iff they have the same reduced form).*

Definition 7 *Suppose* $\left(B^0, C^0, f^0\right)$ *is the true structure. Then if* α *is a parameter of the model and* $\alpha = \alpha^0$ *in all structures equivalent to* $\left(B^0, C^0, f^0\right)$*, then* α *is identified.*

Theorem 8 *The structure* $\left(B^1, C^1, f^1\right)$ *is equivalent to* $\left(B^0, C^0, f^0\right)$ *iff* $\exists$ *a nonsingular matrix* D :

$$\begin{aligned} B^1 &= DB^0, \\ C^1 &= DC^0, \end{aligned}$$

and u_t^1 *has the same distribution as* Du_t^0.

Proof. ($\Leftarrow$) Assume

$$\begin{aligned} B^1 &= DB^0, \\ C^1 &= DC^0, \end{aligned}$$

and u_t^1 has the same distribution as Du_t^0. Then

$$\left(B^1\right)^{-1} C^1 = \left(DB^0\right)^{-1} DC^0 = \left(B^0\right)^{-1} C^0,$$

and

$$\left(B^1\right)^{-1} u_t^1 \sim \left(B^0\right)^{-1} u_t^0.$$

($\Rightarrow$) Assume $\left(B^0, C^0, f^0\right)$ and $\left(B^1, C^1, f^1\right)$ are equivalent structures. Then they have the same reduced form

$$\begin{aligned} &\Rightarrow \left(B^1\right)^{-1} C^1 = \left(B^0\right)^{-1} C^0 \\ &\Rightarrow C^1 = DC^0 \end{aligned}$$

where

$$D = B^1 \left(B^0\right)^{-1}$$

$$\begin{aligned} &\Rightarrow B^1 = DB^0, \\ u_t^1 &\sim Du_t^0. \end{aligned}$$

■

Theorem 9 *Suppose the true structure is*

$$P^0 = \left[B^0 \mid C^0\right]$$

and the restriction matrix in the ith equation is Φ_i. *The the ith equation is identified iff*

$$Rank\left(P^0\Phi_i\right) = n - 1.$$

Proof. ($\Leftarrow$) Suppose $P^1 = DP^0$ is an equivalent structure satisfying

$$P_i^1\Phi_i = 0.$$

Then

$$D_i' P^0\Phi_i = P^1\Phi_i = 0 \Rightarrow D_i' P^1\Phi_i = 0$$

(where D_i is the ith row of D) and

$$e_i' P^0\Phi_i = e_i' P^1\Phi_i = 0$$

(where e_i is a vector with 1 in the ith element and 0 everywhere else). $\Rightarrow e_i'$ and D_i' are both elements of $Null\left[P^0\Phi_i\right]$. But $Null\left[P^0\Phi_i\right]$ has dimension 1 because $Rank\left(P^0\Phi_i\right) = n - 1$. $\Rightarrow D_i' = \lambda e_i'$ for some scalar constant λ. Normalization of B ensures that

$$\lambda = 1 \Rightarrow P_i^1 = P_i^0.$$

($\Rightarrow$) Let P^0 have the ith row identified, and let $Rank\left(P^0\Phi_i\right) < n-1$. Then $\exists D_i' \neq \lambda e_i'$:

$$D_i \in Null\left[P^0\Phi_i\right].$$

Define

$$P_i^1 = D_i P^0,$$

and note that $P^1 \neq P^0$. Note that

$$\left(D_i - e_i\right)' P^0 = 0$$

$\Rightarrow P^0$ has less than full rank. This is not possible because B has full rank $\Rightarrow P$ has full rank. ■

4 Instrumental Variables

Consider the structural equation

$$y = \underset{T\times k}{X}\beta + Q\gamma + u$$

where $EX'u \neq 0$ (i.e., X are endogenous explanatory variables). We want to find k regressors, Z, with the properties:

1. $plim\left(\frac{Z'X}{T}\right)$ is invertible;

2. $plim\left(\frac{Z'u}{T}\right) = 0$.

Note that an implication of $plim\left(\frac{Z'X}{T}\right)$ invertible is that $plim\left(\frac{Z'Z}{T}\right)$ is invertible. Z satisfying these two conditions are called instruments. Let

$$Z^* = (Z \mid Q)$$

be the set of exogenous variables. Let

$$X^* = (X \mid Q)$$

be the set of explanatory variables. Let

$$\beta^* = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$$

be the set of structural parameters (ignoring covariance parameters). Then our model can be written as

$$y = X^*\beta^* + u.$$

If we premultiply our equation by Z^*, we get

$$Z^{*\prime}y = Z^{*\prime}X^*\beta^* + Z^{*\prime}u.$$

Consider the estimator that solves the orthogonality condition

$$E\left[Z^{*\prime}y - Z^{*\prime}X^*\beta^*\right] = 0:$$

$$\widehat{\beta}^* = \left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}y.$$

This is the instrumental variables estimator of β^*.

It is too difficult to derive the moments of $\widehat{\beta}^*$, and in many cases they do not even exist. The trouble occurs because X^* (in the denominator) is endogenous and, therefore, random. But we can derive the asymptotic properties of $\widehat{\beta}^*$.

$$\begin{aligned}
\widehat{\beta}^* &= \left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}y \\
&= \left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}\left(X^*\beta^* + u\right) \\
&= \left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}X^*\beta^* + \left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}u \\
&= \beta^* + \left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}u.
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
plim\widehat{\beta}^* &= \beta^* + plim\left(Z^{*\prime}X^*\right)^{-1} Z^{*\prime}u \\
&= \beta^* + plim\left(\frac{Z^{*\prime}X^*}{T}\right)^{-1}\frac{Z^{*\prime}u}{T} \\
&= \beta^* + \left(plim\frac{Z^{*\prime}X^*}{T}\right)^{-1} plim\frac{Z^{*\prime}u}{T} \\
&= \beta^* + \left(plim\frac{Z^{*\prime}X^*}{T}\right)^{-1} 0 = \beta^*.
\end{aligned}$$

$$
\begin{aligned}
\sqrt{T}\left(\widehat{\beta}^{*}-\beta^{*}\right) &= \sqrt{T}\left(Z^{*\prime}X^{*}\right)^{-1}Z^{*\prime}u \\
& \left(\frac{Z^{*\prime}X^{*}}{T}\right)^{-1}\frac{Z^{*\prime}u}{\sqrt{T}}
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
D\left[\sqrt{T}\left(\widehat{\beta}^{*}-\beta^{*}\right)\right] &= plim\left(\frac{Z^{*\prime}X^{*}}{T}\right)^{-1}\frac{Z^{*\prime}u}{\sqrt{T}}\frac{u'Z^{*}}{\sqrt{T}}\left(\frac{X^{*\prime}Z^{*}}{T}\right)^{-1} \\
&= plim\left(\frac{Z^{*\prime}X^{*}}{T}\right)^{-1}\frac{Z^{*\prime}uu'Z^{*}}{T}\left(\frac{X^{*\prime}Z^{*}}{T}\right)^{-1} \\
&= \left(plim\frac{Z^{*\prime}X^{*}}{T}\right)^{-1}plim\frac{Z^{*\prime}uu'Z^{*}}{T}\left(plim\frac{X^{*\prime}Z^{*}}{T}\right)^{-1}.
\end{aligned}
$$

Note that

$$
EZ^{*\prime}uu'Z^{*}=Z^{*\prime}\sigma^{2}IZ^{*}
$$

$$
\Rightarrow plim\frac{Z^{*\prime}uu'Z^{*}}{T}=plim\frac{Z^{*\prime}\sigma^{2}IZ^{*}}{T}
$$

$$
\Rightarrow D\left[\sqrt{T}\left(\widehat{\beta}^{*}-\beta^{*}\right)\right]==\sigma^{2}\left(plim\frac{Z^{*\prime}X^{*}}{T}\right)^{-1}plim\frac{Z^{*\prime}Z^{*}}{T}\left(plim\frac{X^{*\prime}Z^{*}}{T}\right)^{-1}
$$

$$
\Rightarrow \sqrt{T}\left(\widehat{\beta}^{*}-\beta^{*}\right)\sim N\left[0,V\right]
$$

with

$$
V=\sigma^{2}\left(plim\frac{Z^{*\prime}X^{*}}{T}\right)^{-1}plim\frac{Z^{*\prime}Z^{*}}{T}\left(plim\frac{X^{*\prime}Z^{*}}{T}\right)^{-1}.
$$

Choice of instruments:

1. Excluded exogenous variables;
2. Lagged or leading regressors (maybe);
3. Excluded exogenous variables from implicit equations;
4. Out of thin air regressors;
5. 2SLS 1st stage predictors

5 Two Stage Least Squares

Two Stage Least Squares (2SLS) is a special case of instrumental variables with some nice optimality properties. Consider the model

$$
y=\underset{T\times k}{X}\beta+Q\gamma+u
$$

where X are endogenous explanatory variables. Let Z be a set of instruments with $Rank\,(Z) \geq k$. Let

$$Z^* = (Z \mid Q)\,,$$
$$X^* = (X \mid Q)\,,$$
$$\beta^* = \begin{pmatrix} \beta \\ \gamma \end{pmatrix},$$

and define

$$\widehat{X}_i = Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} X_i$$

$$\begin{aligned} \Rightarrow \quad & \widehat{X} = Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} X \\ \Rightarrow \quad & \widehat{X}^* = Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} X^*. \end{aligned}$$

$\widehat{X}$ satisfies the conditions for instruments for X (and $\widehat{X}^*$ for X^*) as long as Z^* satisfies them.

Now consider the 2SLS estimator

$$\widehat{\beta}^* = \left(\widehat{X}^{*\prime} X^*\right)^{-1} \widehat{X}^{*\prime} y.$$

First, this is equivalent to running a regression of

$$y = \widehat{X}\beta + Q\gamma + u. \tag{4}$$

To see this, write equation (4) as

$$y = \widehat{X}^* \beta^* + u$$

and

$$\begin{aligned} \widehat{\beta}^* &= \left(\widehat{X}^{*\prime} \widehat{X}^*\right)^{-1} \widehat{X}^{*\prime} y \\ &= \left[X^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} X^*\right]^{-1} \cdot \\ & \quad X^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} y \\ & \quad \left[X^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} X^*\right]^{-1} X^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} y \\ &= \left(\widehat{X}^{*\prime} X^*\right)^{-1} \widehat{X}^{*\prime} y \end{aligned}$$

which is the instrumental variables estimator.

We can write

$$\begin{aligned} \widehat{\beta}^* &= \left(\widehat{X}^{*\prime} X^*\right)^{-1} \widehat{X}^{*\prime} y \\ &= \left(\widehat{X}^{*\prime} X^*\right)^{-1} \widehat{X}^{*\prime} \left(X^* \beta + u\right) \\ &= \beta + \left(\widehat{X}^{*\prime} X^*\right)^{-1} \widehat{X}^{*\prime} u. \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} plim\ \widehat{\beta}^{*} &= \beta + plim\left(\widehat{X}^{*\prime}X^{*}\right)^{-1}\widehat{X}^{*\prime}u \\ &= \beta + plim\left(\frac{\widehat{X}^{*\prime}X^{*}}{T}\right)^{-1}\frac{\widehat{X}^{*\prime}u}{T} \\ &= \beta + \left(plim\frac{\widehat{X}^{*\prime}X^{*}}{T}\right)^{-1} plim\frac{\widehat{X}^{*\prime}u}{T}. \end{aligned}$$

Note that

$$\begin{aligned} & plim\frac{\widehat{X}^{*\prime}X^{*}}{T} \\ = \ & plim\begin{pmatrix} \frac{\widehat{X}'X}{T} & \frac{\widehat{X}'Q}{T} \\ \frac{Q'X}{T} & \frac{Q'Q}{T} \end{pmatrix} \\ = \ & plim\begin{pmatrix} \frac{X'Z^{*}\left(Z^{*\prime}Z^{*}\right)^{-1}Z^{*\prime}X}{T} & \frac{X'Z^{*}\left(Z^{*\prime}Z^{*}\right)^{-1}Z^{*\prime}Q}{T} \\ \frac{Q'X}{T} & \frac{Q'Q}{T} \end{pmatrix} \\ & \begin{pmatrix} plim\frac{X'Z^{*}\left(Z^{*\prime}Z^{*}\right)^{-1}Z^{*\prime}X}{T} & plim\frac{X'Z^{*}\left(Z^{*\prime}Z^{*}\right)^{-1}Z^{*\prime}Q}{T} \\ plim\frac{Q'X}{T} & plim\frac{Q'Q}{T} \end{pmatrix} \\ = \ & \begin{pmatrix} plim\frac{X'Z^{*}}{T}\left(\frac{Z^{*\prime}Z^{*}}{t}\right)^{-1}\frac{Z^{*\prime}X}{t} & plim\frac{X'Z^{*}}{T}\left(\frac{Z^{*\prime}Z^{*}}{T}\right)^{-1}\frac{Z^{*\prime}Q}{T} \\ plim\frac{Q'X}{T} & plim\frac{Q'Q}{T} \end{pmatrix} \\ = \ & \begin{pmatrix} plim\frac{X'Z^{*}}{T}\left(plim\frac{Z^{*\prime}Z^{*}}{t}\right)^{-1}plim\frac{Z^{*\prime}X}{t} & plim\frac{X'Z^{*}}{T}\left(plim\frac{Z^{*\prime}Z^{*}}{T}\right)^{-1}plim\frac{Z^{*\prime}Q}{T} \\ plim\frac{Q'X}{T} & plim\frac{Q'Q}{T} \end{pmatrix} \end{aligned}$$

where all *plim* terms exist. Also

$$\begin{aligned} plim\frac{\widehat{X}^{*\prime}u}{T} &= \begin{pmatrix} plim\frac{\widehat{X}'u}{T} \\ plim\frac{Q'u}{T} \end{pmatrix} \\ &= \begin{pmatrix} plim\frac{X'Z^{*}\left(Z^{*\prime}Z^{*}\right)^{-1}Z^{*\prime}u}{T} \\ plim\frac{Q'u}{T} \end{pmatrix} \\ &= \begin{pmatrix} plim\frac{X'Z^{*}}{T}\left(\frac{Z^{*\prime}Z^{*}}{T}\right)^{-1}\frac{Z^{*\prime}u}{T} \\ plim\frac{Q'u}{T} \end{pmatrix} \\ &= \begin{pmatrix} plim\frac{X'Z^{*}}{T}\left(plim\frac{Z^{*\prime}Z^{*}}{T}\right)^{-1}plim\frac{Z^{*\prime}u}{T} \\ plim\frac{Q'u}{T} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

because

$$plim\frac{Z^{*\prime}u}{T} = 0,$$
$$plim\frac{Q'u}{T} = 0.$$

Next,

$$\sqrt{T}\left(\widehat{\beta}^* - \beta^*\right) = \left(\frac{\widehat{X}^{*\prime}X^*}{T}\right)^{-1}\frac{\widehat{X}^{*\prime}u}{\sqrt{T}}$$

$\Rightarrow$

$$\begin{aligned}
& plimT\left(\widehat{\beta}^* - \beta^*\right)\left(\widehat{\beta}^* - \beta^*\right)' \\
= \;& plim\left(\frac{\widehat{X}^{*\prime}X^*}{T}\right)^{-1}\frac{\widehat{X}^{*\prime}u}{\sqrt{T}}\frac{u'\widehat{X}^*}{\sqrt{T}}\left(\frac{X^{*\prime}\widehat{X}^*}{T}\right)^{-1} \\
= \;& plim\left(\frac{\widehat{X}^{*\prime}X^*}{T}\right)^{-1}\frac{\widehat{X}^{*\prime}uu'\widehat{X}^*}{T}\left(\frac{X^{*\prime}\widehat{X}^*}{T}\right)^{-1} \\
= \;& \left(plim\frac{\widehat{X}^{*\prime}X^*}{T}\right)^{-1} plim\frac{\widehat{X}^{*\prime}uu'\widehat{X}^*}{T}\left(plim\frac{X^{*\prime}\widehat{X}^*}{T}\right)^{-1} \\
= \;& \sigma^2\left(plim\frac{\widehat{X}^{*\prime}X^*}{T}\right)^{-1} plim\frac{\widehat{X}^{*\prime}\widehat{X}^*}{T}\left(plim\frac{X^{*\prime}\widehat{X}^*}{T}\right)^{-1}
\end{aligned}$$

$\Rightarrow$

$$\Rightarrow \sqrt{T}\left(\widehat{\beta}^* - \beta^*\right) \sim N(0, V)$$

with

$$V = \sigma^2\left(plim\frac{\widehat{X}^{*\prime}X^*}{T}\right)^{-1} plim\frac{\widehat{X}^{*\prime}\widehat{X}^*}{T}\left(plim\frac{X^{*\prime}\widehat{X}^*}{T}\right)^{-1}.$$

Discuss testing.

Relationship Between 2SLS and IV

1. 2SLS is a way to use $k^* > k$ instruments optimally.
2. If $k^* = k$, then 2SLS=IV=ILS.

6 3SLS

Basic idea: Do 2SLS and use residuals to estimate Ω consistently. Then use $\widehat{\Omega}$ in GLS generalization of 2SLS.

Details: Consider

$$y_i = X_i^*\beta_i^* + u_i, \quad i = 1, 2, .., n$$

with common instruments Z^* for each equation. $\Rightarrow$

$$\begin{aligned}
\widehat{\beta}_i^* &= \left[X_i^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} X_i^*\right]^{-1} \left[X_i^{*\prime} Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime} y_i\right] \\
&= \left(X_i^{*\prime} P_{Z^*} X_i^*\right)^{-1} \left(X_i^{*\prime} P_{Z^*} y_i\right),
\end{aligned}$$

and residuals are

$$\begin{aligned}
\widehat{u}_i &= y_i - X_i^* \widehat{\beta}_i^* \\
&= \left[I - \left(X_i^{*\prime} P_{Z^*} X_i^*\right)^{-1} \left(X_i^{*\prime} P_{Z^*}\right)\right] y_i.
\end{aligned}$$

Now stack equations:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} X_1^* & 0 & \cdots & 0 \\ 0 & X_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^* \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_n^* \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

or

$$y = X^* \beta^* + u.$$

Let

$$I_n \otimes Z^*$$

be the instruments for X^*. Then

$$\left(I_n \otimes Z^*\right)' y = \left(I_n \otimes Z^*\right)' X^* \beta^* + \left(I_n \otimes Z^*\right)' u. \tag{5}$$

Note that

$$\begin{aligned}
D\left[\left(I_n \otimes Z^*\right)' u\right] &= \left(I_n \otimes Z^*\right)' D(u) \left(I_n \otimes Z^*\right) \\
&= \left(I_n \otimes Z^*\right)' \left(\Omega \otimes I_T\right) \left(I_n \otimes Z^*\right) \\
&= \Omega \otimes Z^{*\prime} Z^*.
\end{aligned}$$

Therefore, if we apply GLS to equation (5), we get

$$\begin{aligned}
\widehat{\beta}^* &= \left[X^{*\prime} \left(I_n \otimes Z^*\right) \left(\Omega \otimes Z^{*\prime} Z^*\right)^{-1} \left(I_n \otimes Z^{*\prime}\right) X^*\right]^{-1} \cdot \\
&\quad \left[X^{*\prime} \left(I_n \otimes Z^*\right) \left(\Omega \otimes Z^{*\prime} Z^*\right)^{-1} \left(I_n \otimes Z^{*\prime}\right) y\right] \\
&= \left[X^{*\prime} \left(\Omega^{-1} \otimes Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime}\right) X^*\right]^{-1} \cdot \\
&\quad \left[X^{*\prime} \left(\Omega^{-1} \otimes Z^* \left(Z^{*\prime} Z^*\right)^{-1} Z^{*\prime}\right) y\right] \\
&= \left[X^{*\prime} \left(\Omega^{-1} \otimes P_{Z^*}\right) X^*\right]^{-1} \left[X^{*\prime} \left(\Omega^{-1} \otimes P_{Z^*}\right) y\right].
\end{aligned}$$

Students should derive the asymptotic properties of 3SLS.

7 FIML

The basic idea in FIML is to write down the log likelihood function and maximize it over the parameters of interest. We describe the details later.

Advantages and Disadvantages of FIML:

1. FIML allows for cross equation restrictions.
2. FIML allows for nonlinearity more naturally.
3. FIML is more sensitive to misspecification of the model.
4. Small sample bias is larger for FIML than for IV than for OLS
5. All methods trade off efficiency against robustness.