Linear Simultaneous Equations Systems

Recall: Econometricians treat RHS regressors as stochastic and condition on them; statisticians often treat them as predetermined (or deterministic, exogenous).

This leaves possibility that, in econometric analyses LHS and RHS variables may be simultaneously determined, leading to endogenous regressors.

Implications for OLS: So far, $\mathbf{y} = \mathbf{X}\beta_0 + \epsilon$, and a.s.

$$\begin{aligned} \mathbb{E}[\mathbf{y}|\mathbf{X}] &= \mathbf{X}\beta_0 \\ \Leftrightarrow \mathbb{E}[\epsilon|\mathbf{X}] &= \mathbb{E}[\mathbf{y} - \mathbf{X}\beta_0|\mathbf{X}] = 0 \\ \Rightarrow \mathbb{E}[\mathbf{X}'\epsilon] &= \mathbb{E}[\mathbf{X}'\mathbb{E}[\epsilon|\mathbf{X}]] = 0 \end{aligned}$$

so residuals ϵ and regressors $\mathbf{X}$ are uncorrelated.

Now: RHS variables $\mathbf{X}$ and residuals ϵ may be correlated.

$\Rightarrow$ Violation of Gauss-Markov assumption $\mathbb{E}[\epsilon|\mathbf{X}] = \mathbf{0}$ a.s.

$\Rightarrow$ OLS estimator no longer BLUE.

Classical (Stylized) Example: Market Demand

market-level demand equation

$$q_t^d = \beta p_t + \epsilon_{1t},\ \beta < 0, t = 1, \cdots, T. \quad (1)$$

Want to estimate β.

Economic theory stipulates: Prices and quantities are simultaneously determined whenever price equilibrates demand and supply.

To model this, consider also supply equation:

$$q_t^s = \gamma p_t + \epsilon_{2t},\ \gamma > 0, t = 1, \cdots, T. \quad (2)$$

Equations (1) and (2) constitute the structural form of the system.

Assume $\mathbb{E}[\epsilon_{it}] = 0, \mathbb{E}[\epsilon_{it}^2] = \sigma_i^2$, $i = 1, 2$, $\mathbb{E}[\epsilon_{1t}\epsilon_{2t}] = 0$, and independence across t.

Market equilibrium requires $q_t^d = q_t^s =: q_t$ for all t. This yields reduced form system:

$$\begin{aligned} p_t &= \frac{1}{\gamma - \beta}(\epsilon_{1t} - \epsilon_{2t}), \\ q_t &= \frac{\gamma}{\gamma - \beta}\epsilon_{1t} - \frac{\beta}{\gamma - \beta}\epsilon_{2t}, t = 1, \cdots, T. \end{aligned}$$

Hence, $\mathbb{E}[p_t\epsilon_{1t}] = \frac{1}{\gamma-\beta}\mathbb{E}[\epsilon_{1t}^2 - \epsilon_{1t}\epsilon_{2t}] = \frac{\sigma_1^2}{\gamma-\beta} > 0$.

Implications

It cannot be that $\mathbb{E}[\epsilon_{1t}|p_t] = 0$, since this would imply $\mathbb{E}[p_t\epsilon_{1t}] = 0$, by iterated expectations.

What does this imply if q_t were regressed on p_t in order to estimate β in (1)?

To facilitate exposition, strengthen assumptions slightly: Suppose

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right), \ \forall t, \text{ indep. across } t.$$

This implies for the reduced from:

$$\begin{pmatrix} p_t \\ q_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\sigma_1^2+\sigma_2^2}{(\gamma-\beta)^2} & \frac{\gamma\sigma_1^2+\beta\sigma_2^2}{(\gamma-\beta)^2} \\ . & \frac{\gamma^2\sigma_1^2+\beta^2\sigma_2^2}{(\gamma-\beta)^2} \end{pmatrix}\right), \quad \forall t, \text{ indep. across } t. \tag{3}$$

This implies for the conditional distribution of q_t (see handout), given p_t:

$$q_t|p_t \overset{i.i.d.}{\sim} N\left(\frac{\gamma\sigma_1^2+\beta\sigma_2^2}{\sigma_1^2+\sigma_2^2}p_t, \mathrm{var}(q_t|p_t)\right).$$

Let $\mathbf{q}' = (q_1, \cdots, q_T)$, $\mathbf{p}' = (p_1, \cdots, p_T)$. Then, the OLS estimator for β in (1) is

$$\hat{\beta} = (\mathbf{p'p})^{-1}\mathbf{p'q},$$

Therefore,

$$\mathbb{E}\left[\hat{\beta}|\mathbf{p}\right] = (\mathbf{p'p})^{-1}\mathbf{p}'\mathbb{E}[\mathbf{q}|\mathbf{p}] = \frac{\gamma\sigma_1^2 + \beta\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \neq \beta.$$

This is referred to as *simultaneity bias*.

If $\sigma_1^2 \to 0$ or $\sigma_2^2 \to +\infty$, then $\hat{\beta}$ is unbiased for β.

If $\sigma_1^2 \to +\infty$ or $\sigma_2^2 \to 0$, then $\hat{\beta}$ is unbiased for γ.

In intermediate cases, $\hat{\beta}$ represents a mixture of β and γ.

How much can be uncovered about $(\beta, \gamma, \sigma_1^2, \sigma_2^2)$ from data on q and p?

This is a question about identifiability of $(\beta, \gamma, \sigma_1^2, \sigma_2^2)$.

Saw above:

$$\begin{pmatrix} p_t \\ q_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\sigma_1^2+\sigma_2^2}{(\gamma-\beta)^2} & \frac{\gamma\sigma_1^2+\beta\sigma_2^2}{(\gamma-\beta)^2} \\ . & \frac{\gamma^2\sigma_1^2+\beta^2\sigma_2^2}{(\gamma-\beta)^2} \end{pmatrix}\right), \quad \forall t, \text{ indep. across } t. \qquad (3)$$

The DGP is completely specified by (3), i.e. by var(q_t), var(p_t) and cov(q_t, p_t). Given a large sample, these can be inferred with high precision, but not more.

Since these 3 moments depend on 4 parameters, there is no way to uncover the parameters from estimates of the moments.

For example, (1,2,3,4) and(2,1,4,3) imply the same estimable moments for the DGP.

If different values for a parameter vector θ imply the same probability distribution for the DGP, then θ is unidentifiable.

Definition: Let $\mathcal{X}$ be the set of all conceivable data. Two structures θ_1 and θ_2, $\theta_1 \neq \theta_2$, of a model defined by the probability distribution $F_X(x;\theta)$, $x \in \mathcal{X}$, are said to be *observationally equivalent* if

$$F_X(x;\theta_1) = F_X(x;\theta_2) \forall x \in \mathcal{X}.$$

Definition: A structure θ_1 is *identifiable* if there is no observationally equivalent structure, i.e. if

$$\theta_1 \neq \theta_2 \Rightarrow \Pr\left(F_X(X;\theta_1) \neq F_X(X;\theta_2)\right) > 0.$$

General Representation

Structural Form of Linear Simultaneous Equation System:

$$\underset{n\times n}{\mathbf{B}}\ \underset{n\times 1}{\mathbf{y}_t} + \underset{n\times m}{\Gamma}\ \underset{m\times 1}{\mathbf{x}_t} = \underset{n\times 1}{\mathbf{u}_t}, \quad t = 1, \cdots, T.$$

n equations with n endogenous variables $\mathbf{y}_t$ and m exogenous variables $\mathbf{x}_t$, satisfying $\mathbb{E}[\mathbf{x}_t \mathbf{u}_t'] = \mathbf{0}$.

Assume also $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}, \mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \Sigma$, and $\mathbb{E}[\mathbf{u}_t \mathbf{u}_s'] = 0$, $\forall t \neq s$.

Reduced Form:

$$\begin{aligned} \mathbf{y}_t &= -\mathbf{B}^{-1}\Gamma \mathbf{x}_t + \mathbf{B}^{-1}\mathbf{u}_t \\ &= \Pi \mathbf{x}_t + \mathbf{v}_t, \text{ where } \underset{n\times m}{\Pi} = -\mathbf{B}^{-1}\Gamma, \mathbf{v}_t = \mathbf{B}^{-1}\mathbf{u}_t, \end{aligned}$$

so $\mathbb{E}[\mathbf{v}_t] = \mathbf{0}, \mathbb{E}[\mathbf{v}_t \mathbf{v}_t'] = \mathbf{B}^{-1}\Sigma \mathbf{B}^{-1'} =: \Omega$, and $\mathbb{E}[\mathbf{v}_t \mathbf{v}_s'] = \mathbf{0}$, $\forall t \neq s$.

Π is estimable consistently, using OLS equation by equation.

Ω can be estimated from regression residuals $\hat{\mathbf{v}}_t$.

Identification

Can the values of $\mathbf{B}, \Gamma$ and Σ be deduced from Π and Ω?

Structural Form (SF) parameters:

B: n^2

Γ: nm

Σ: $n(n+1)/2$

Total: $n^2 + nm + n(n+1)/2$.

Reduced Form (RF) parameters:

Π: nm

Ω: $n(n+1)/2$

Total: $nm + n(n+1)/2$.

$\Rightarrow$ Structural form involves n^2 parameters in excess, so without further restrictions, identification of SF from RF is impossible.

Possible restrictions:

- normalizations; e.g. coefficient=1;
- exclusions: coefficient=0;
- linear restrictions among coefficients;
- covariance restrictions; etc.

Notation

n= number of endogenous variables in all equations;

n_i= number of endogenous variables in equation i;

m= number of exogenous variables in all equations;

m_i= number of exogenous variables in equations i;

$\mathbf{B}_i$= columns of $\mathbf{B}$ corresponding to endogenous variables excluded from equation i;

Γ_i= columns of Γ corresponding to exogenous variables excluded from equation i; $i = 1, \cdots, n$.

Order Condition: A necessary condition for the ith equation to be identifiable is that

$$\begin{array}{rcl} g_i := n - n_i + m - m_i & \geq & n - 1 \\ (\text{no. of exl. variables} & \geq & n - 1) \\ \Leftrightarrow \qquad\qquad m - m_i & \geq & n_i - 1 \\ (\text{no. of instruments} & \geq & \text{no. of endog. RHS variables}) \end{array}$$

Rank Condition: A necessary and sufficient condition for the ith equation to be to be identifiable is that

$$R_i := \mathrm{rk}\left([B_i : \Gamma_i]\right) = n - 1.$$

NB: There exist also other ways to define the rank condition.

This yields the following possibilities:

$g_i < n-1$ or rank condition fails: ith equation is not identified;

$g_i = n-1$ and rank condition is met: ith equations is just identified.

$g_i > n-1$ and rank condition is met: ith equation is over-identified.

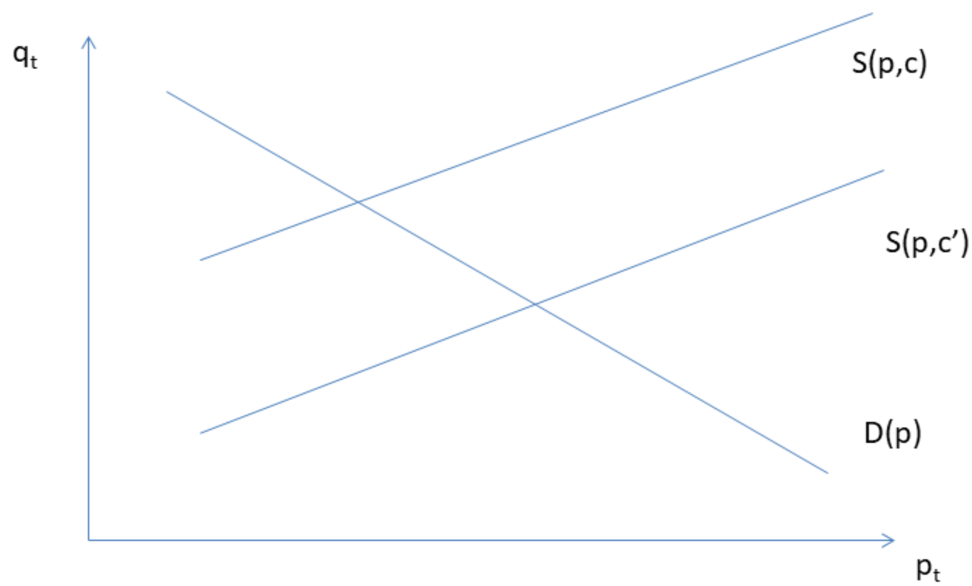
Interpretation of Order Condition: Variation due to variables that are extraneous to the relationship of interest (so-called "instruments", denoted by $\mathbf{Z}$) helps to trace out this relationship.

In previous example: Variation in costs, if part of (2), shifts supply curve while leaving demand unaffected. This traces out (or identifies) the demand curve.

Requirement: $\mathbb{E}[\mathbf{Z}'\mathbf{u}_i] = \mathbf{0}$.

Will see later how this orthogonality can be used to construct unbiased and consistent estimators in presence of endogenous regressors.

Role of Instrument in Identifying Demand Function:
instrument cost (c) excluded from demand D(p),
but cost shock c -> c' moved supply S(p,c) and traces out demand



Informal Interpretation of Rank Condition

ith equation is identifiable if there does not exist a linear combination of the other equations in the system that is indistinguishable from the ith equation.

B_i and Γ_i capture the impact of all other variables (that are excluded from the ith equation) on these other equations in the system.

$[B_i : \Gamma_i]$ has n rows, one of which is zero (the ith row). So its rank can be at most $n - 1$.

If, and only if, it has rank $n - 1$, there does not exist a linear combination of its rows that yields the zero vector.

This means that, under this condition, any linear combination of all the other equations is impacted by the variation in the variables excluded from the ith equation.

Hence, any linear combination of all other equations in the system is distinct from the ith equation.

Formal Interpretation of Rank Condition

W.l.o.g. consider 1st equation and arrange variables such that

$$
\begin{aligned}
\mathbf{B} &= \left[\underset{n\times n_1}{\bar{\mathbf{B}}} \;:\; \underset{n\times(n-n_1)}{\mathbf{B}_1} \right], \Gamma = \left[\underset{n\times m_1}{\bar{\Gamma}} \;:\; \underset{n\times(m-m_1)}{\Gamma_1} \right] \\
\mathbf{y}' &= \left(\underset{1\times n_1}{\mathbf{y}_1'} , \underset{1\times(n-n_1)}{\mathbf{y}_{-1}'} \right) \\
\mathbf{x}' &= \left(\underset{1\times m_1}{\mathbf{x}_1'} , \underset{1\times(m-m_1)}{\mathbf{x}_{-1}'} \right) \\
\mathbf{u}' &= (u_1, \mathbf{u}_{-1}').
\end{aligned}
$$

Note that

$$
\mathbf{B}_1 = \begin{bmatrix} \underset{1\times(n-n_1)}{\mathbf{0}'} \\ \underset{(n-1)\times(n-n_1)}{\tilde{\mathbf{B}}_1} \end{bmatrix}, \Gamma_1 = \begin{bmatrix} \underset{1\times(m-m_1)}{\mathbf{0}'} \\ \underset{(n-1)\times(m-m_1)}{\tilde{\Gamma}_1} \end{bmatrix}.
$$

So, system of equations can be written in partitioned form as

$$
\begin{aligned}
\begin{bmatrix} u_1 \\ \mathbf{u}_{-1} \end{bmatrix} &= \left[\bar{\mathbf{B}} : \mathbf{B}_1 \right] \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_{-1} \end{bmatrix} + \left[\bar{\Gamma} : \Gamma_1 \right] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_{-1} \end{bmatrix} \\
&= \begin{bmatrix} \bar{\mathbf{B}}_{1,\cdot} & \mathbf{0}' \\ \bar{\mathbf{B}}_{-1,\cdot} & \tilde{\mathbf{B}}_1 \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_{-1} \end{bmatrix} + \begin{bmatrix} \bar{\Gamma}_{1,\cdot} & \mathbf{0}' \\ \bar{\Gamma}_{-1,\cdot} & \tilde{\Gamma}_1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_{-1} \end{bmatrix}
\end{aligned}
$$

Since the first row of both $\mathbf{B}_1$ and Γ_1 is a zero vector,

$$\begin{aligned} \operatorname{rk}\left([\mathbf{B}_1 : \Gamma_1]\right) &= \operatorname{rk}\left([\tilde{\mathbf{B}}_1 : \tilde{\Gamma}_1]\right) \\ &\leq \min\{n-1, n-n_1+m-m_1\}. \end{aligned}$$

Since the order condition requires $m - m_1 \geq n_1 - 1 \Leftrightarrow n - n_1 + m - m_1 \geq n - 1$, it follows that

$$\operatorname{rk}\left([\tilde{\mathbf{B}}_1 : \tilde{\Gamma}_1]\right) \leq n - 1.$$

From the partitioned representation of the system,

(1) 1st eqn: $\underset{1\times n_1}{\bar{\mathbf{B}}_{1,\cdot}}\mathbf{y}_1 + \underset{1\times m_1}{\bar{\Gamma}_{1,\cdot}}\mathbf{x}_1 = u_1$.

(2) 2nd - nth eqn:

$$\begin{aligned}
\mathbf{u}_{-1} &= \underset{(n-1)\times n_1}{\bar{\mathbf{B}}_{-1,\cdot}}\mathbf{y}_1 + \underset{(n-1)\times(n-n_1)}{\tilde{\mathbf{B}}_1}\mathbf{y}_{-1} \\
&\quad + \underset{(n-1)\times m_1}{\bar{\Gamma}_{-1,\cdot}}\mathbf{x}_1 + \underset{(n-1)\times(m-m_1)}{\tilde{\Gamma}_1}\mathbf{x}_{-1} \\
\Leftrightarrow \bar{\mathbf{B}}_{-1,\cdot}\mathbf{y}_1 + \bar{\Gamma}_{-1,\cdot}\mathbf{x}_1 &= \mathbf{u}_{-1} - \tilde{\mathbf{B}}_1\mathbf{y}_{-1} - \tilde{\Gamma}_1\mathbf{x}_{-1} \\
&= \mathbf{u}_{-1} - \left[\tilde{\mathbf{B}}_1 : \tilde{\Gamma}_1\right]\begin{bmatrix}\mathbf{y}_{-1} \\ \mathbf{x}_{-1}\end{bmatrix}.
\end{aligned}$$

Suppose $\operatorname{rk}\left([\tilde{\mathbf{B}}_1 : \tilde{\Gamma}_1]\right) < n-1 \Rightarrow \exists\gamma \in \mathbb{R}^{n-1}, \gamma \neq \mathbf{0} : \gamma'[\tilde{\mathbf{B}}_1 : \tilde{\Gamma}_1] = \mathbf{0}$. This implies:

$$\begin{aligned}
\underbrace{\gamma'\bar{\mathbf{B}}_{-1,\cdot}}\mathbf{y}_1 + \underbrace{\gamma'\bar{\Gamma}_{-1,\cdot}}\mathbf{x}_1 &= \gamma'\mathbf{u}_{-1} \\
\underset{1\times n_1}{\mathbf{b}}{}'\mathbf{y}_1 + \underset{1\times m_1}{\mathbf{g}}{}'\mathbf{x}_1 &= v_1;
\end{aligned}$$

i.e. linear combination of system (2) is indistinguishable from (1), unless $\operatorname{rk}([\mathbf{B}_1 : \Gamma_1]) = n-1$.

Example: $n = 2 \Rightarrow n^2 = 4$ restrictions

$$
\begin{aligned}
(1)\ \ q_t &= a_1 + b_1 p_t + c_1 m_t + u_{1t} \text{ (demand)} \\
(2)\ \ q_t &= a_2 + b_2 p_t + c_2 r_t + u_{2t} \text{ (supply)} \\
\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} &= \underbrace{\begin{bmatrix} 1 & -b_1 \\ 1 & -b_2 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} q_t \\ p_t \end{bmatrix}}_{\mathbf{y}_t} + \underbrace{\begin{bmatrix} -a_1 & -c_1 & 0 \\ -a_2 & 0 & -c_2 \end{bmatrix}}_{\Gamma} \underbrace{\begin{bmatrix} 1 \\ m_t \\ r_t \end{bmatrix}}_{\mathbf{x}_t} \\
\mathbf{u}_t &= \mathbf{B} \quad \mathbf{y}_t \quad + \quad \Gamma \quad \mathbf{x}_t.
\end{aligned}
$$

g_1= no. of excl. variables in 1st eqn. $= 1 \geq n - 1$;

g_2= no. of excl. variables in 2nd eqn. $= 1 \geq n - 1$;

$\Rightarrow$ order condition satisfied for both equations.

rank condition:

$$
\begin{aligned}
R_1 &= \operatorname{rk}\left([B_1 : \Gamma_1]\right) = \operatorname{rk}\left(\begin{bmatrix} 0 \\ -c_2 \end{bmatrix}\right) = 1, \text{ iff } c_2 \neq 0; \\
R_2 &= \operatorname{rk}\left([B_2 : \Gamma_2]\right) = \operatorname{rk}\left(\begin{bmatrix} -c_1 \\ 0 \end{bmatrix}\right) = 1, \text{ iff } c_1 \neq 0.
\end{aligned}
$$

$\Rightarrow$ Both equations are just identified, provided there exist instruments (income m_t for supply equation, interest rate r_t for demand equation).

Restrictions: 2 normalizations and 2 exclusions.

Example (cont'd)

Reduced form:

$$
\begin{aligned}
q_t &= \pi_1 + \pi_2 m_t + \pi_3 r_t + v_{1t} \\
p_t &= \pi_4 + \pi_5 m_t + \pi_6 r_t + v_{2t},
\end{aligned}
$$

where

$$
\begin{aligned}
\pi_1 &= \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1}, \pi_2 = \frac{c_1 b_2}{b_2 - b_1} \\
\pi_3 &= -\frac{c_2 b_1}{b_2 - b_1}, \qquad \pi_4 = \frac{a_1 - a_2}{b_2 - b_1} \\
\pi_5 &= \frac{c_1}{b_2 - b_1}, \qquad \pi_6 = -\frac{c_2}{b_2 - b_1}.
\end{aligned}
$$

In the just identified case, one can recover SF parameters from RF parameters:

$$
\begin{aligned}
b_1 &= \frac{\pi_3}{\pi_6}, \qquad\qquad b_2 = \frac{\pi_2}{\pi_5} \\
c_1 &= \pi_5(b_2 - b_1), c_2 = -\pi_6(b_2 - b_1) \\
a_1 &= \pi_1 - b_1\pi_4, \quad a_2 = \pi_1 - b_2\pi_4.
\end{aligned}
$$

Recovering SF parameters from RF estimates in just identified case is called Indirect Least Squares.

Problem: Calculation of standard errors is cumbersome.

Two-Stage Least Squares (2SLS)

W.l.o.g. consider 1st equation of the system,

$$\begin{aligned}
y_{1t} &= \underset{1\times(n_1-1)}{\mathbf{Y}_{1t}}{}'\beta_1 + \underset{1\times m_1}{\mathbf{X}_{1t}}{}'\gamma_1 + u_{1t}, t = 1,\cdots,T; \\
\underset{T\times 1}{\mathbf{y}_1} &= \underset{T\times(n_1-1)}{\mathbf{Y}_1}\beta_1 + \underset{T\times m_1}{\mathbf{X}_1}\gamma_1 + \mathbf{u}_1 \\
&= \mathbf{Z}_1\delta_1 + \mathbf{u}_1, \quad \underset{T\times(n_1+m_1-1)}{\mathbf{Z}_1} = [\mathbf{Y}_1 : \mathbf{X}_1], \delta_1' = (\beta_1', \gamma_1').
\end{aligned}$$

Assume $m - m_1 \geq n_1 - 1$ (just or over-identification). Let $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$, for $\mathbf{X}_2$ an array of $m - m_1$ instruments for $\mathbf{Y}_1$.

Consider two-step procedure:

1. Regress columns of $\mathbf{Y}_1$ on $\mathbf{X}$, to get $\hat{\mathbf{Y}}_1 = \mathbf{P}_X\mathbf{Y}_1$, $\mathbf{P}_X = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}'$.

   Note that

$$\begin{aligned}
\mathbf{y}_1 &= \hat{\mathbf{Y}}_1\beta_1 + \mathbf{X}_1\gamma_1 + \mathbf{u}_1 + (\mathbf{Y}_1 - \hat{\mathbf{Y}}_1)\beta_1 \\
&= \hat{\mathbf{Z}}_1\delta_1 + \mathbf{w}_1, \text{ where} \\
\hat{\mathbf{Z}}_1 &= [\hat{\mathbf{Y}}_1 : \mathbf{X}_1] \in \mathrm{Col}(\mathbf{X}) \\
\mathbf{w}_1 &= \mathbf{u}_1 + (\mathbf{I} - \mathbf{P}_X)\mathbf{Y}_1\beta_1 \\
\text{where } \mathbf{X} &\perp\!\!\!\perp \mathbf{u}_1, (\mathbf{I} - \mathbf{P}_X)\mathbf{Y}_1\beta_1 \in \mathrm{Col}(\mathbf{X})^{\perp} \Rightarrow \mathbb{E}[\hat{\mathbf{Z}}_1'\mathbf{w}_1] = \mathbf{0}.
\end{aligned}$$

2. Regress $\mathbf{y}_1$ on $\hat{\mathbf{Z}}_1$, to get

$$\begin{aligned}
\hat{\delta}_1 &= \hat{\delta}_{1,2SLS} = (\hat{\mathbf{Z}}_1'\hat{\mathbf{Z}}_1)^{-1}\hat{\mathbf{Z}}_1'\mathbf{y}_1 \\
&= \left(\mathbf{Z}_1'\mathbf{P}_X\mathbf{Z}_1\right)^{-1}\mathbf{Z}_1'\mathbf{P}_X\mathbf{y}_1.
\end{aligned}$$

Comments:

- In just identified case:

$$\begin{aligned} m - m_1 &= n_1 - 1, \text{ so that} \\ m_1 + n_1 - 1 &= \text{no. of col.s of } \mathbf{Z}_1 = m \\ \Rightarrow \hat{\delta}_1 &= \Big(\underbrace{\mathbf{Z}_1'\mathbf{X}}_{m\times m} \underbrace{(\mathbf{X'X})^{-1}}_{m\times m} \mathbf{X'Z}_1 \Big)^{-1} \mathbf{Z}_1'\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y} \\ &= (\mathbf{X'Z}_1)^{-1}(\mathbf{X'X})(\mathbf{Z}_1'\mathbf{X})^{-1}\mathbf{Z}_1'\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y} \\ &= (\mathbf{X'Z}_1)^{-1}\mathbf{X'y} \\ &= \text{instrumental variables (IV) estimator.} \end{aligned}$$

- Assuming normality of $\mathbf{y}_1$, conditional on $\mathbf{X}, \mathbf{Z}_1$,

$$\hat{\delta}_1 | \mathbf{X,Z}_1 \sim N\left(\delta_1, \sigma_1^2 \left(\mathbf{Z}_1'\mathbf{P}_X\mathbf{Z}_1\right)^{-1}\right),$$

  where $\sigma_1^2 = \text{var}(u_{1t}|\mathbf{X,Z}_1)$ for all t. This parameter can be estimated by

$$s_1^2 = \frac{\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1}{T - (m_1 + n_1 - 1)},$$

  where $\hat{\mathbf{u}}_1 = \mathbf{y}_1 - \mathbf{Z}_1\hat{\delta}_1$, i.e. residuals from *original* equation.

- Note: 2SLS uses non-orthogonal projector

$$\begin{aligned} \mathbf{P}_{Z_1 \perp X} &= \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{P}_X\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{P}_X \\ \hat{\mathbf{y}}_1 &= \mathbf{P}_{Z_1 \perp X}\mathbf{y}_1. \end{aligned}$$

  It projects onto $\text{Col}(\mathbf{Z}_1)$, preserving $\text{Col}(\mathbf{Z}_1)$ and annihilating $\text{Col}(\mathbf{X})^{\perp}$.

Comparison: OLS vs. IV/2SLS in the linear regression model $\mathbf{y} = \underset{(T\times K)}{\mathbf{X}}\beta + \mathbf{u}$.

- OLS:

  population: $\mathbb{E}[\mathbf{X'u}] = \mathbb{E}[\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)] = \mathbf{0}$.

  sample: $\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}_{OLS}) = 0$.

  If $\mathrm{rk}(\mathbf{X}) = K$, then $\hat{\beta}_{OLS} = (\mathbf{X'X})^{-1}\mathbf{X'y}$.

- IV:

  population: $\mathbb{E}[\mathbf{X'u}] \neq \mathbf{0}$, but there exists array of instruments $\underset{T\times K}{\mathbf{Z}}$ such that

  $\mathbb{E}[\mathbf{Z'u}] = \mathbb{E}[\mathbf{Z}'(\mathbf{y} - \mathbf{X}\beta)] = \mathbf{0}$.

  sample: $\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\beta}_{IV}) = 0$.

  If $\mathrm{rk}(\mathbf{Z'X}) = K$, then $\hat{\beta}_{IV} = (\mathbf{Z'X})^{-1}\mathbf{Z'y}$.

- 2SLS:

  population: $\mathbb{E}[\mathbf{X'u}] \neq \mathbf{0}$, but there exists array of instruments $\underset{T\times M}{\mathbf{Z}}$, $M > K$, such that $\mathbb{E}[\mathbf{Z'u}] = \mathbb{E}[\mathbf{Z}'(\mathbf{y} - \mathbf{X}\beta)] = \mathbf{0}$.

  sample: $\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS}) \approx 0$. But $\mathbf{X}'\mathbf{P}_Z(\mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS}) = 0$ (K equations in K unknowns).

  If $\mathrm{rk}(\mathbf{Z}) = M$, $\mathrm{rk}(\mathbf{X}) = \mathrm{rk}(\mathbf{Z'X}) = K$, then $\hat{\beta}_{2SLS} = (\mathbf{X'P}_Z\mathbf{X})^{-1}(\mathbf{X'P}_Z\mathbf{y})$.

Testing

1. Hausman Test

   Know: $\hat{\beta}_{OLS}$ is BLUE (unbiased and efficient) if $\mathbb{E}[\mathbf{X'u}] = \mathbf{0}$ and biased if $\mathbb{E}[\mathbf{X'u}] \neq \mathbf{0}$.

   If $\exists$ instruments $\mathbf{Z}$ s.t. $\mathbb{E}[\mathbf{Z'u}] = \mathbf{0}$, then $\hat{\beta}_{IV/2SLS}$ is unbiased, but inefficient if, in fact, $\mathbb{E}[\mathbf{X'u}] = \mathbf{0}$.

$\Rightarrow$ Both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV/2SLS}$ are unbiased if $\mathbb{E}[\mathbf{X'u}] = \mathbf{0}$, but $\hat{\beta}_{OLS}$ is relatively efficient.

   $\hat{\beta}_{IV/2SLS}$ is unbiased if $\mathbb{E}[\mathbf{X'u}] \neq \mathbf{0}$, but $\hat{\beta}_{OLS}$ is not.

   This is basis of a test of $H_0 : \mathbb{E}[\mathbf{X'u}] = \mathbf{0}$ (exogeneity of $\mathbf{X}$).

   Assuming conditional normality, under H_0,

$$\hat{\beta}_{OLS} - \hat{\beta}_{IV/2SLS} | \mathbf{X,Z} \sim N(\mathbf{0}, \mathbf{V}),$$

   where $\mathbf{V} = \text{var}\left(\hat{\beta}_{OLS} - \hat{\beta}_{IV/2SLS} | \mathbf{X,Z}\right)$, symmetric and positive semi-definite.

   Hausman test statistic

$$\mathcal{H} = \left(\hat{\beta}_{OLS} - \hat{\beta}_{IV/2SLS}\right)' \mathbf{V}^{-1} \left(\hat{\beta}_{OLS} - \hat{\beta}_{IV/2SLS}\right) \sim \chi^2_K.$$

   Reject H_0 (exogeneity) if $\mathcal{H} > \chi^2_{K,1-\alpha}$.

Recall Result: Orthogonality of Relatively Efficient Estimators

Under H_0, $\hat{\beta}_{OLS}$ is efficient, so

$$
\begin{aligned}
0 &= \text{cov}(\hat{\beta}_{OLS}, \hat{\beta}_{OLS} - \hat{\beta}_{IV/2SLS}) \\
\Leftrightarrow &= \text{var}(\hat{\beta}_{OLS}) = \text{cov}(\hat{\beta}_{OLS}, \hat{\beta}_{IV/2SLS}) \\
\Rightarrow \mathbf{V} &= \text{var}\left(\hat{\beta}_{OLS} - \hat{\beta}_{IV/2SLS}\right) \\
&= \text{var}(\hat{\beta}_{OLS}) + \text{var}(\hat{\beta}_{IV/2SLS}) \\
&\quad - \text{cov}(\hat{\beta}_{OLS}, \hat{\beta}_{IV/2SLS}) - \text{cov}(\hat{\beta}_{IV/2SLS}, \hat{\beta}_{OLS}) \\
&= \text{var}(\hat{\beta}_{IV/2SLS}) - \text{var}(\hat{\beta}_{OLS}) \\
&= \sigma_0^2 \left[\left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1} - (\mathbf{X}'\mathbf{X})^{-1} \right].
\end{aligned}
$$

Problem: Often, $\left[\left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1} - (\mathbf{X}'\mathbf{X})^{-1} \right]$ is singular, so $\mathbf{V}^{-1}$ does not exist.

2. Alternative Implementation: Durbin-Wu-Hausman Test

Include first-stage residuals as added regressors:

$$\begin{aligned} \hat{\mathbf{U}} &= \mathbf{X} - \mathbf{P}_Z\mathbf{X} \\ \mathbf{y} &= \mathbf{X}\beta + \hat{\mathbf{U}}\alpha + \epsilon, \end{aligned}$$

and carry out F-test of $H_0 : \alpha = \mathbf{0}$.

- Intuition:
  - (i) If $\mathbf{X}$ exogenous: residuals $\hat{\mathbf{U}}$ are mere data noise and uncorrelated with $\mathbf{y}$, so expect $\alpha =$ (and $\hat{\alpha} \approx$) 0; $\hat{\beta}$ is then OLS estimator, rel. inefficient due to redundant regressors $\hat{\mathbf{U}}$.
  - (ii) If $\mathbf{X}$ contains endogenous regressors: $\hat{\mathbf{U}}$ captures endogenous part of $\mathbf{X}$, expected to be correlated with $\mathbf{y}$, so expect α (and $\hat{\alpha}$) $\neq 0$.

    Also: by partitioned regression, $\hat{\beta}$ captures corr. of $\mathbf{y}$ with exogenous part of $\mathbf{X}$, so $\hat{\beta} = \hat{\beta}_{IV/2SLS}$.

By similar logic, can also carry out F-test in the regression

$$\begin{aligned} \hat{\mathbf{X}} &= \mathbf{P}_Z\mathbf{X} \\ \mathbf{y} &= \mathbf{X}\beta + \hat{\mathbf{X}}\alpha + \epsilon. \end{aligned}$$

Optional: Derivation of Durbin-Wu-Hausman Regression

Video for slides 23 - 24 (click here)

Consider the linear simultaneous equations system

$$\underset{N\times 1}{\mathbf{y}_1} = \underset{N\times 1}{\mathbf{y}_2}\underset{1\times 1}{\alpha} + \underset{N\times k_1}{\mathbf{Z}_1}\underset{k_1\times 1}{\delta} + \underset{N\times 1}{\mathbf{u}_1} \quad (1)$$

$$\underset{N\times 1}{\mathbf{y}_2} = \underset{N\times k}{\mathbf{Z}}\underset{k\times 1}{\pi} + \underset{N\times 1}{\mathbf{u}_2} \quad (2),$$

where $\underset{N\times k}{\mathbf{Z}} = \left[\underset{N\times k_1}{\mathbf{Z}_1} : \underset{N\times k_2}{\mathbf{Z}_2}\right]$, $k = k_1 + k_2$, and $\mathbb{E}[\mathbf{Z}'\mathbf{u}_i] = \mathbf{0}$, $i = 1, 2$.

Note that $\mathbf{y}_2$ is correlated with $\mathbf{u}_1$ in (1) above if, and only if, $\mathbf{u}_1$ is correlated with $\mathbf{u}_2$:

$$\mathbb{E}[\mathbf{y}_2'\mathbf{u}_1] = \mathbb{E}[\pi'\mathbf{Z}'\mathbf{u}_1 + \mathbf{u}_2'\mathbf{u}_1] = \mathbb{E}[\mathbf{u}_2'\mathbf{u}_1],$$

because $\mathbf{Z}$ is uncorrelated with $\mathbf{u}_1$.

If so, estimate α and δ by IV/2SLS.

Suppose $\mathbf{u}_1 = \mathbf{u}_2\rho + \epsilon$, where $\mathbb{E}[\mathbf{u}_2'\epsilon] = 0$ and ρ is a scalar parameter.

Note that $\mathbf{y}_2$ is uncorrelated with ϵ:

$$\mathbb{E}[\mathbf{y}_2'\epsilon] = \mathbb{E}[\pi'\mathbf{Z}'\epsilon + \mathbf{u}_2'\epsilon] = \mathbb{E}[\pi'\mathbf{Z}'(\mathbf{u}_1 - \mathbf{u}_2\rho) + \mathbf{u}_2'\epsilon] = 0$$

because $\mathbb{E}[\mathbf{Z}'\mathbf{u}_i] = \mathbf{0}$, $i = 1, 2$, as well as $\mathbb{E}[\mathbf{u}_2'\epsilon] = 0$.

Inserting into (1): $\mathbf{y}_1 = \mathbf{y}_2\alpha + \mathbf{Z}_1\delta + \mathbf{u}_2\rho + \epsilon$. $\quad(\star)$

So (α, δ', ρ) could be estimated by OLS if $\mathbf{u}_2$ were observed.

$\mathbf{u}_2$ can be estimated by $\hat{\mathbf{u}}_2 = \mathbf{y}_2 - \mathbf{Z}\hat{\pi}$, where $\hat{\pi}$ is the OLS estimator of π in (2) (1st stage regression).

Use $\hat{\mathbf{u}}_2$ in lieu of $\mathbf{u}_2$ in the hypothetical OLS procedure contemplated above.

By partitioned regression,

$$\begin{aligned} \widehat{\begin{bmatrix} \alpha \\ \delta \end{bmatrix}} &= \left([\mathbf{y}_2 : \mathbf{Z}_1]'(I - P_{\hat{\mathbf{u}}_2})[\mathbf{y}_2 : \mathbf{Z}_1]\right)^{-1} \times \cdots \\ &\quad \cdots \times [\mathbf{y}_2 : \mathbf{Z}_1]'(I - P_{\hat{\mathbf{u}}_2})\mathbf{y}_1, \end{aligned}$$

where

$$\begin{aligned} P_{\hat{\mathbf{u}}_2} &= \hat{\mathbf{u}}_2 \left(\hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2\right)^{-1} \hat{\mathbf{u}}_2' \\ &= (\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2 \left(\mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2\right)^{-1} \mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}}) \\ &= (\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2\mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}})/\mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbf{y}_2'(\mathbf{I} - P_{\hat{\mathbf{u}}_2})\mathbf{y}_2 &= \mathbf{y}_2'\mathbf{y}_2 - \left(\mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2\right)^2 / \mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2 \\ &= \mathbf{y}_2'\mathbf{y}_2 - \mathbf{y}_2'(\mathbf{I} - P_{\mathbf{Z}})\mathbf{y}_2 \\ &= \mathbf{y}_2' P_{\mathbf{Z}} \mathbf{y}_2 \\ \mathbf{Z}_1'(\mathbf{I} - P_{\hat{\mathbf{u}}_2})\mathbf{Z}_1 &= \mathbf{Z}_1'\mathbf{Z}_1 = \mathbf{Z}_1' P_{\mathbf{Z}} \mathbf{Z}_1 \\ \mathbf{Z}_1'(\mathbf{I} - P_{\hat{\mathbf{u}}_2})\mathbf{y}_2 &= \mathbf{Z}_1'\mathbf{y}_2 = \mathbf{Z}_1' P_{\mathbf{Z}} \mathbf{y}_2, \end{aligned}$$

i.e. the OLS estimator of (α, δ') in $(\star)$ with $\hat{\mathbf{u}}_2$ in lieu of $\mathbf{u}_2$ is equal to the 2SLS estimator.

3. Test of validity of Over-Identifying Restrictions: Hansen-Sargan J-Test

Idea: If all $M-K$ over-identifying restrictions hold, then $\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS})$ should be small (close to zero vector, in the appropriate metric).

Under H_0 : "$M-K$ over-identifying restrictions are valid",

$$\begin{aligned}\mathbb{E}[\hat{\beta}_{2SLS}|\mathbf{X},\mathbf{Z}] &= \beta\\ \mathbb{E}[\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS}|\mathbf{X},\mathbf{Z}] &= \mathbf{0}\\ \mathbb{E}[\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS})] &\approx \mathbf{0}\end{aligned}$$

and

$$\begin{aligned}&\operatorname{var}\left(\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS})|\mathbf{X},\mathbf{Z}\right)\\ =\;& \sigma_0^2\left(\mathbf{Z}'(\mathbf{I}-\mathbf{X}(\mathbf{X'P}_Z\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}\right) =: \sigma_0^2\Sigma.\end{aligned}$$

Can show: $\operatorname{rk}\left(\mathbf{Z}'(\mathbf{I}-\mathbf{X}(\mathbf{X'P}_Z\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}\right) = \operatorname{rk}(\Sigma) = M-K$.

Construct quadratic form from moment functions $\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS})$, re-weighted inversely proportional to their variance,

$$\begin{aligned}\mathcal{J} &= \frac{(\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS})'\mathbf{Z}\Sigma^{-1}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\beta}_{2SLS})}{\sigma_0^2}\\ &\sim \chi^2_{M-K},\end{aligned}$$

and reject H_0 when $\mathcal{J} > \chi^2_{M-K,1-\alpha}$.

If test rejects, it does not permit any conclusion which of the moment condition fail(s).