Lecture ~~#~~. Endogeneity.

Linear projection model. $Y_i = X_i \beta + \varepsilon_i$

by construction, $E(\varepsilon_i X_i) = 0$.

β is the projection coefficient.

it doesn't have a clear economic interpretation.

Sole a feature of the joint distribution of observable (X_i, Y_i).

eg. measurement error.

Suppose (Y_i, X_i^*)'s β coefficient is parameter of interest.

X^* unobservable. observe $X_i = X_i^* + u_i$.

Y_i: credit card consumption. X_i^*: expected income (unobs.)

X_i: consumption (obs.).

$$y_i = x_i^{*'} \beta + e_i$$

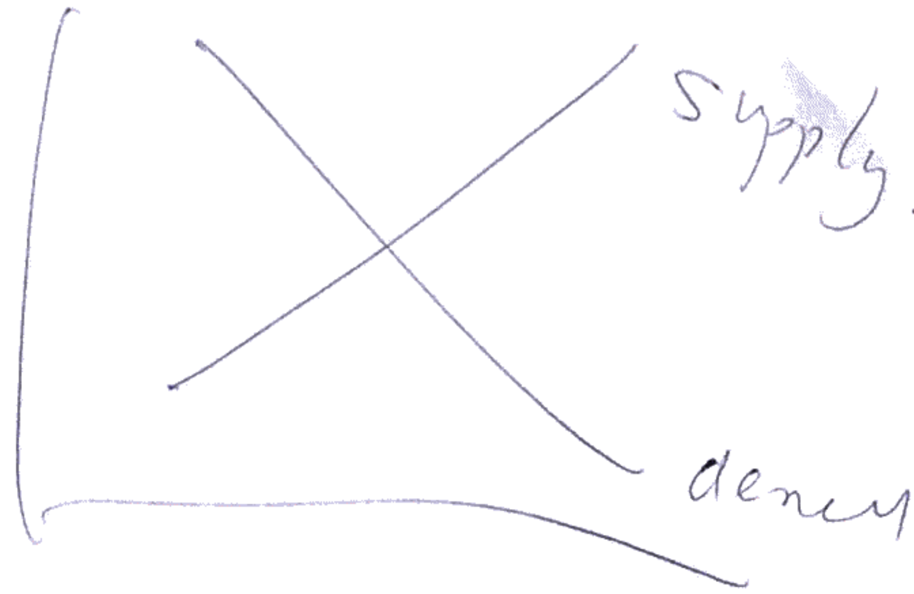
$$= (x_i - u_i)\beta + e_i$$

$$= x_i \beta + v_i, \text{ where } v_i = e_i - u_i \beta$$

$$E(X_i, v_i) = E[(X_i^* + u_i)(e_i - u_i\beta)] = -E(u_i^2)\beta.$$

Supply & demand.

$q_i \quad p_i$



$$\begin{cases} q_i = -\beta_1 p_i + e_{1i} \\ q_i = \beta_2 p_i + e_{2i} \end{cases}$$

To illustrate the point, impose some assumptions to facilitate the calculation. $\beta_1 + \beta_2 = 1$, $E e_i = 0$, $E e_i e_i' = I_2$.

$$\begin{pmatrix} 1 & \beta_1 \\ 1 & -\beta_2 \end{pmatrix} \begin{pmatrix} q_i \\ p_i \end{pmatrix} = \begin{pmatrix} e_{1i} \\ e_{2i} \end{pmatrix}$$

$$\begin{pmatrix} q_i \\ p_i \end{pmatrix} = \frac{1}{-\beta_1-\beta_2} \begin{pmatrix} -\beta_2 & -\beta_1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} e_{1i} \\ e_{2i} \end{pmatrix} = \begin{pmatrix} \beta_2 & \beta_1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} e_{1i} \\ e_{2i} \end{pmatrix}$$

Projection q_i on p_i. $q_i = \beta^* p_i + \varepsilon_i$. β^* is the projection coeff.

$$\beta^* = \frac{\mathrm{Cov}(q_i, p_i)}{\mathrm{var}(p_i)} \overset{\text{zero mean}}{=\!=\!=} \frac{E(p_i q_i)}{E(q_i^2)} = \frac{E[(\beta_2 e_{1i} + \beta_1 e_{2i})(e_{1i} - e_{2i})]}{E[(e_{1i} - e_{2i})^2]}$$

$$= \frac{\beta_2 - \beta_1}{2}.$$

Structural equation. $y_i = x_i'\beta + e_i$

e_i structural error. if $E(x_i e_i) \neq 0$ endogenous.

if $Z: \ell \times 1$ $E(e_i z_i) = 0$. IV.

$$\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} \begin{matrix} k_1 \\ k_2 \end{matrix} \qquad z_i = \begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix} \begin{matrix} k_1 \\ \ell_2 \end{matrix} \qquad k_1 + \ell_2 = \ell.$$

included exog. excluded exog.

Reduced form: $y = Z\lambda + v$

$X = Z\Gamma + u$.

end. exogenous

Estimation method.

$$E(z_i e_i) = E[z(y_i - x\beta)] = 0.$$

$E(z x_i)\beta = E(z_i y_i)$. Sample version $\frac{1}{n}(\sum z x_i)\beta = \frac{1}{n}\sum z_i y_i$

~~if dim $\neq$ dim~~ $\dim(X) = k = \dim(Z) = \ell$.

then $\beta = \left(\frac{1}{n}\sum z x_i\right)^{-1} \frac{1}{n}\sum(z y_i)$. Just identified.

if $\ell > k$, overidentified.

No β makes the ℓ equations exactly hold.

$$\min \left(\frac{1}{n}\sum z(y - x\beta)\right) W \left(\frac{1}{n}\sum (y - x\beta) z\right).$$

W arbitrary p.d. weighting matrix.

$$\text{FOC: } \cancel{\left(\frac{1}{n}\sum (y-x\beta) z\right) W \left(\frac{1}{n}\sum xz\right) = 0}.$$

$$\left(\frac{1}{n}\sum xz\right) W \left(\frac{1}{n}\sum z(y - x\beta)\right) = 0$$

$$\hat{\beta} = \left[\left(\frac{1}{n}\sum xz\right) W \left(\frac{1}{n}\sum xz\right)\right]^{-1} \left(\frac{1}{n}\sum xz\right) W \left(\frac{1}{n}\sum zy\right)$$

$$= \left((X'Z) W Z'X\right)^{-1} \left(X'Z W Z'y\right).$$

Special case $W_n = (Z'Z)^{-1}$.

$$\hat{\beta}_{2SLS} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'y$$

regress X on Z. $\hat{X} = X(X'X)^{-1}X'Z$

$= P_X Z$.

then Y on $\hat{X}$.

$$\hat{\beta} = \cancel{X'Z(X'X)^{-1}}$$

$$(X'P_Z X)^{-1}(X'P_Z y).$$

(5-)

$$\cancel{\ } \hat{\beta} - \beta_0 = (X P_Z X)^{-1} (X \cancel{P_Z} Z)^{-1} (Z Z)^{-1} Z \varepsilon.$$

$$\sqrt{n}(\hat{\beta} - \beta_0) = N\Big(0, (Q_{ZX} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$
$$(Q_{ZX} Q_{ZZ}^{-1} \cancel{Q}\, \Omega\, Q_{ZZ}^{-1} Q_{ZX})$$
$$(Q_{ZX} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$

where $\Omega = E(z_i z_i' \varepsilon_i^2)$.

if $E(\varepsilon_i^2 | z_i) = \sigma^2$; then

$$\Omega = \cancel{\ } Q_{ZZ}\, \sigma^2.$$

$$\sqrt{n}(\hat{\beta} - \beta_0) = N\big(0, (Q_{Z\cancel{X}} Q_{ZZ}^{-1} Q_{Z\cancel{X}})^{-1} \sigma^2\big).$$