Instumental Variables

Endogeneity: $Y = X^T\beta + e \quad E(Xe)\neq 0$

Structural parameter β: $Y=X^T\beta+e$, $E(Xe)\neq0$

Linear projection coefficient β^*: $Y = X^T\beta^* + e^*$, $E(Xe^*)=0$

LSE $\hat\beta \xrightarrow{p} E(XX^T)^{-1}E(XY) = \beta^* = \beta + E(XX^T)^{-1}E(Xe) \neq \beta$

Inconsistency of LSE: Endogeneity bias

Example: Measurement Error Data(Y, X)

$Y = Z^T\beta + e$, $E(Ze)=0$

$X = Z+u$ latent variable $\overset{e \perp\!\!\!\perp u}{Z \perp\!\!\!\perp u}$. $E(u)=0$ classical m.. error

$\Rightarrow Y = X^T\beta + (e - u^T\beta) \quad E(X(e-u^T\beta)) = -E(uu^T)\beta \neq 0$

$\beta^* = E(XX^T)^{-1}E(XY) = \beta[I_k - E(XX^T)^{-1}E(uu^T)]$

Measurement error bias / attenuation bias

Example: Supply & Demand Data(Q, P)

$Q = -\beta_1 P + e_1 \quad e=[e_1\ e_2]^T$, $E(e)=0$ $E(ee^T)=I_2$

$Q = \beta_2 P + e_2$

$$\Rightarrow \begin{bmatrix}1 & \beta_1\\ 1 & \beta_2\end{bmatrix}\begin{bmatrix}Q\\P\end{bmatrix} = \begin{bmatrix}e_1\\e_2\end{bmatrix} \Rightarrow \begin{bmatrix}Q\\P\end{bmatrix} = \frac{1}{\beta_1+\beta_2}\begin{bmatrix}\beta_2e_1+\beta_1e_2\\ e_1-e_2\end{bmatrix}$$

Reg: $Q=\beta^*P+e^*$. $E(Pe^*)=0$

$\beta^* = E(P^2)^{-1}E(PQ) = \frac{\beta_2-\beta_1}{2} \neq -\beta_1$ or β_2

Simultaneous equations bias

Example: Choice Variables as Regressors

$\ln(wage) = \beta\ edu. + e \quad Cov(edu., e) > 0$

$\overset{+}{\leftarrow} u \overset{+}{\rightarrow} \quad \beta^* > \beta$

$Y = X_1^T\beta_1 + X_2^T\beta_2 + e$. $E(X_1e)=0 \quad E(X_2e)\neq0 \quad X_1: k_1\times1 \quad X_2: k_2\times1$

Z is an instrumental variable if: Z $l\times1$

$E(Ze)=0$; $E(ZZ')>0$; $\operatorname{rank}(E(ZX'))=k \Rightarrow l\geq k$

Normalization Relevance condition

$Z=\begin{bmatrix}Z_1\\Z_2\end{bmatrix} = \begin{bmatrix}X_1\\Z_2\end{bmatrix}$ $k_1\times 1$ Inclued exogenous variables; $l_2\times 1$ Exclued exogenous variables

$Y_1 = Z_1^T\beta_1 + Y_2^T\beta_2 + e$ where Z exogenous, $\vec Y$ endogenous

Reduced Form:

$Y_2 = \Gamma^T Z + u_2 = [\Gamma_{12}^T\ \Gamma_{22}^T]\begin{bmatrix}Z_1\\Z_2\end{bmatrix} + u_2 = \Gamma_{12}^TZ_1 + \Gamma_{22}^TZ_2+u_2$

$E(Zu_2^T)=0 \Rightarrow \Gamma = E(ZZ^T)^{-1}E(ZY_2^T) \quad \Gamma = \begin{bmatrix}\Gamma_{12}\\\Gamma_{22}\end{bmatrix}$

$Y_1 = Z_1^T\beta + Y_2^T\beta_2 + e = Z_1^T(\beta_1+\Gamma_{12}\beta_2) + Z_2^T\Gamma_{22}\beta_2 + (u_2^T\beta_2+e)$

$= Z^T\lambda + u_1$

$\lambda = \begin{bmatrix}I_{k_1} & \Gamma_{12}\\ 0 & \Gamma_{22}\end{bmatrix}\begin{bmatrix}\beta_1\\\beta_2\end{bmatrix} = \begin{bmatrix}\begin{matrix}I_{k_1}\\0\end{matrix} & \Gamma\end{bmatrix}\begin{bmatrix}\beta_1\\\beta_2\end{bmatrix} = \bar\Gamma\beta$

$u_1 = u_2^T\beta_2 + e$

$\Rightarrow E(Zu_1)=0 \Rightarrow \lambda = E(ZZ^T)^{-1}E(ZY_1)$

$\bar\Gamma = E(ZZ^T)^{-1}E(ZX^T)$

$\begin{cases} Y_1 = \lambda^T Z + u_1 \\ Y_2 = \Gamma Z + u_2 \end{cases} \quad \vec Y = \begin{bmatrix}\lambda_1^T & \lambda_2^T\\ \Gamma_{12}^T & \Gamma_{22}^T\end{bmatrix} Z + u$. $E(Zu^T)=0$

LSE: $\hat\lambda = (\sum_{i=1}^n Z_iZ_i^T)^{-1}(\sum_{i=1}^n Z_iY_{1i}) \quad \hat\lambda = \hat{\bar\Gamma}\hat\beta$

$\hat{\bar\Gamma} = (\sum_{i=1}^n Z_iZ_i^T)^{-1}(\sum_{i=1}^n Z_iY_{2i}^T)$ Indirect Least Squares

$\lambda = \bar\Gamma\beta$, $\lambda = E(ZZ^T)^{-1}E(ZY_1) \Rightarrow \underset{l\times1}{E(ZY_1)} = \underset{l\times k}{E(ZX^T)}\underset{k\times1}{\beta}$

$\bar\Gamma = E(ZZ^T)^{-1}E(ZX^T)$

If $l=k$: $\beta = \bar\Gamma^{-1}\lambda = E(ZX^T)^{-1}E(ZY_1)$ just-identified

If $l>k$: $\beta = (\bar\Gamma^T\bar\Gamma)^{-1}\bar\Gamma^T\lambda$ over-identified

$\begin{cases} \lambda_1 = \beta_1 + \Gamma_{12}\beta_2 \\ \lambda_2=\Gamma_{22}\beta_2 \end{cases} \Rightarrow \begin{cases} \beta_1 = \lambda_1 - \Gamma_{12}(\Gamma_{22}^T\Gamma_{22})^{-1}\Gamma_{22}^T\lambda_2 \\ \beta_2 = (\Gamma_{22}^T\Gamma_{22})^{-1}\Gamma_{22}^T\lambda_2 \end{cases}$

$\hat\beta_{iv} = (\sum_{i=1}^n Z_iX_i^T)^{-1}(\sum_{i=1}^n Z_iY_{1i}) = \hat\beta_{ils} = \hat{\bar\Gamma}^{-1}\hat\lambda$

Orthogonality: $\sum_{i=1}^n Z_i\hat e_i = 0$, $\hat e_i = Y_{1i} - X_i^T\hat\beta_{iv}$

Wald Estimator: $Z\in\{0,1\}$.

$\begin{cases} E(Y|Z=1) = E(X|Z=1)\beta+\alpha \\ E(Y|Z=0)=E(X|Z=0)\beta+\alpha \end{cases} \Rightarrow \beta = \frac{E(Y|Z=1)-E(Y|Z=0)}{E(X|Z=1)-E(X|Z=0)} = \beta_{iv}$

$\hat\beta = \frac{\bar Y_1 - \bar Y_0}{\bar X_1 - \bar X_0} \quad \bar Y_1 = \sum_{i=1}^n Z_iY_{1i} / \sum_{i=1}^n Z_i \quad \bar Y_0 = \frac{\sum_{i=1}^n(1-Z_i)Y_{1i}}{\sum_{i=1}^n(1-Z_i)}$

$\hat\beta_{iv} = \frac{\sum Z_i(Y_i-\bar Y)}{\sum X_i(X_i-\bar X)} = \frac{\bar Y_1 - \bar Y}{\bar X_1-\bar X} = \frac{(1-\bar Z)(\bar Y_1-\bar Y_0)}{(1-\bar Z)(\bar X_1-\bar X_0)} = \hat\beta$

Two stage Least Squares $(l\geq k)$

$Y_1 = Z^T\bar\Gamma\beta + u_1 \quad E(Zu_1)=0$

If $\bar\Gamma$ is known: $\hat\beta = (\bar\Gamma^TZ^TZ\bar\Gamma)^{-1}(\bar\Gamma^TZ^TY_1)$

Estimate $\bar\Gamma$: $\hat{\bar\Gamma} = (Z^TZ)^{-1}(Z^TX)$:

$\hat\beta_{2SLS} = (X^TZ(Z^TZ)^{-1}Z^TX)^{-1}(X^TZ(Z^TZ)^{-1}Z^TY_1)$

$= (Z^TX)^{-1}Z^TZ(X^TZ)^{-1}X^TZ(Z^TZ)^{-1}Z^TY_1$

$= (Z^TX)^{-1}Z^TY_1 = \hat\beta_{iv} \quad (l=k)$

① Reg $X\sim Z$: $\hat\Gamma = (Z^TZ)^{-1}(Z^TX)$, $\hat X = Z\hat\Gamma = P_ZX$

② Reg $Y_1\sim\hat X$: $\hat\beta_{2SLS} = (\hat X^T\hat X)^{-1}\hat X^TY_1$

$Y_1 = Z_1\beta_1 + \hat Y_2\beta_2 + e$.

FWL: β_2: Reg $M_1Y_1 \sim M_1\hat Y_2$. $M_1 = I_n - P_1 = I - Z_1(Z_1^TZ_1)^{-1}Z_1^T$

$\hat\beta_2 = (\hat Y_2^TM_1\hat Y_2)^{-1}\hat Y_2^TM_1Y_1 \quad \hat Y_2 = P_ZY_2 = Z(Z^TZ)^{-1}Z^T$

$= (Y_2^TP_Z(I-P_1)P_ZY_2)^{-1}Y_2^TP_Z(I-P_1)Y_1$

$= (Y_2^T(P_Z-P_1)Y_2)^{-1}Y_2^T(P_Z-P_1)Y_1$ since $P_ZP_1=P_1$

Overidentification: $Z^T\hat e\neq0$ but $\hat X^T\hat e=0$

$\hat\beta_{2SLS}\xrightarrow{p}\beta$, $\sqrt n(\hat\beta_{2SLS}-\beta)\xrightarrow{d} N(0,V_\beta)$

$\hat\beta_{2SLS} = (\hat X^T\hat X)^{-1}\hat X^T(X\beta+e) = \beta + (X^TP_ZX)^{-1}X^TP_Ze$

$=\beta + (X^TZ(Z^TZ)^{-1}Z^TX)^{-1}X^TZ(Z^TZ)^{-1}Z^Te$

$=\beta + [(\frac1n\sum X_iZ_i^T)(\frac1n\sum Z_iZ_i^T)^{-1}(\frac1n\sum Z_iX_i^T)]^{-1}(\frac1n\sum X_iZ_i^T)(\frac1n\sum Z_iZ_i^T)^{-1}\frac{\sum Z_i^Te_i}{n}$

$(EX_iZ_i^T\ (EZ_iZ_i^T)^{-1}\ EZ_iX_i^T)^{-1}\ EX_iZ_i^T\ (EZ_iZ_i^T)^{-1}\ EZ_i^Te_i$

$V_\beta = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}Q_{XZ}Q_{ZZ}^{-1}E(Z_i^TZ_ie_i^2)Q_{ZZ}^{-1}Q_{XZ}(\cdots)^{-1}$

Conditional homoskedastic: $V_\beta = V_\beta^0 = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}\sigma^2$

$E(e^2|Z)=\sigma^2$

$\theta = r(\beta)$: $\hat\theta_{2SLS} = r(\hat\beta_{2SLS})$

$\hat\theta_{2SLS}\xrightarrow{p}\theta$. $\sqrt n(\hat\theta_{2SLS}-\theta)\xrightarrow{d}N(0,V_\theta)$

$V_\theta = R^TV_\beta R \quad R = \frac{\partial}{\partial\beta}r(\beta)^T$

$H_0: \theta=\theta_0 \quad H_1:\theta\neq\theta_0$

Wald statistic $W = n(\hat\theta-\theta_0)^T\hat V_\theta^{-1}(\hat\theta-\theta_0) \quad W|H_0\xrightarrow{d}\chi^2_q$

Asymptotic critical value c: $\alpha = 1-G_q(c)$

Reject H_0 if $W>c$ has aysmptotic size α.

Control Function Regression

$Y = X_1^T\beta_1 + X_2^T\beta_2 + e \qquad E(Ze)=0 \qquad E(X_1e)=0$

$X_2 = \Gamma_{12}^T Z_1 + \Gamma_{22}^T Z_2 + u_2 \qquad E(Zu_2)=0 \qquad E(X_1u_2)=0 \qquad (Z_1 = X_1)$

$e = u_2^T\alpha + v : \quad \alpha = E(u_2u_2^T)^{-1}E(u_2e) \qquad Eu_2v=0$

$\Rightarrow Y = X_1^T\beta_1 + X_2^T\beta_2 + u_2^T\alpha + v \qquad E(X_1v)=0 \quad EX_2v=0 \quad Eu_2v=0$

① Reg $X_2 \sim Z$: $\hat{u}_2 = X_2 - \hat{\Gamma}_{12}^T Z_1 - \hat{\Gamma}_{22}^T Z_2 = (I_n - P_Z)X_2$

② Reg $Y \sim (X_1, X_2, \hat{u}_2)$: $Y = X\hat{\beta} + \hat{u}_2\hat{\alpha} + \hat{v}$

FWL: $\hat{\beta}$: Reg $M_uY \sim M_uX \qquad M_u = I_n - \hat{u}_2(\hat{u}_2^T\hat{u}_2)^{-1}\hat{u}_2^T$

$\hat{\beta}_{CF} = (X^TM_uX)^{-1}X^TM_uY . \qquad M_uX = \begin{bmatrix} X_1 \\ P_ZX_2 \end{bmatrix} = P_ZX$

$= (X^TP_ZX)^{-1}X^TP_ZY = \tilde{\beta}_{2SLS}$

$\hat{\alpha}$: Reg $M_XY \sim M_X\hat{u} \qquad M_X = I_n - X(X^TX)^{-1}X^T$

Endogeneity Tests

$H_0: E(X_2e)=0 \qquad H_1: E(X_2e)\neq 0$

$\Leftrightarrow H_0: \alpha = 0 \qquad H_1: \alpha \neq 0$

Local Average Treatment Effects (Both X, Z binary)

$Y = h(X,U) \quad X = g(Z,U) \quad \Rightarrow Y = Y(X).\ X = X(Z)$

ATE $= E(Y(1) - Y(0)) \qquad X(0)=0 \quad X(0)=1$

Monotonicity condition: $X(1) \geq X(0) \qquad X(1)=0$ Never Takers ~~Defiers~~

LATE $= E(Y(1)-Y(0) \mid X(1) > X(0)) \qquad X(1)=1$ Compliers Always Takers

With Assumptions ① $U \perp\!\!\!\perp Z$ ② monotone $P(X(1)<X(0))=0$:

LATE $= E(C \mid X(1)-X(0)=1) = \dfrac{E(Y|Z=1) - E(Y|Z=0)}{E(X|Z=1) - E(X|Z=0)} = \beta_{Wald} = \beta_{IV}$

$X = (1-Z)X(0) + ZX(1) = X(0) + Z[X(1)-X(0)]$

$Y = Y(0) + X[Y(1)-Y(0)] = Y(0) + XC \qquad C$: causal effect

$= Y(0) + X(0)C + Z[X(1)-X(0)]C$

$U \perp\!\!\!\perp Z \Rightarrow (Y(0), X(0), Y(1), X(1)) \perp\!\!\!\perp Z$

$E(Y|Z=1) = EY(0) + E(X(0)C) + E[(X(1)-X(0))C]$

$E(Y|Z=0) = EY(0) + E(X(0)C)$

$\Rightarrow E(Y|Z=1) - E(Y|Z=0) = E[(X(1)-X(0))C]$

$= P(X(1)-X(0)=1)\ E(C \mid X(1)-X(0)=1)$

$= E(X(1)-X(0))\ E(C \mid X(1)-X(0)=1)$

$= [E(X|Z=1) - E(X|Z=0)]\ E(C \mid X(1)-X(0)=1)$

Regression Discontinuity

Sharp regression discontinuity : treatment is discontinuous

Fuzzy regression discontinuity: P(treatment) is discontinuous

$y_i = T_iy_{1i} + (1-T_i)y_{0i} \qquad T_i \in \{0,1\}$

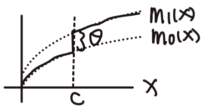
$TE = \theta_i = y_{1i} - y_{0i}. \qquad ATE = E(\theta_i) = E(y_{1i}) - E(y_{0i})$

Conditional ATE $= \theta(x) = E(\theta_i \mid X_i = x) = E(y_{1i} - y_{0i} \mid X_i = x)$

$T_i = 1\{X_i \geq c\} \qquad X_i$: running variable. c: cutoff

Core Identification Thm.

Suppose $m_0(x) = E(y_{0i} \mid X_i = x)$. $m_1(x) = E(y_{1i} \mid X_i = x)$ are continuous at $x=c$. Then $\theta = \theta(c) = m_1(c) - m_0(c) = m(c+) - m(c-)$



$\theta(c) = E(y_{1i} \mid X_i = c) - E(y_{0i} \mid X_i = c)$

$= m_1(c) - m_0(c)$

LL Estimation:

$Z_i(x) = \begin{bmatrix} 1 \\ X_i - x \end{bmatrix}$

$\hat{\beta}_0(x) = \left(\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Z_i(x) Z_i(x)^T 1\{X_i < c\}\right)^{-1} \left(\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Z_i(x) y_i 1\{X_i < c\}\right)$

$\hat{\beta}_1(x) = \qquad 1\{X_i \geq c\} \qquad 1\{X_i \geq c\}$

$\hat{m}(x) = [\hat{\beta}_0(x)]_1 1\{x<c\} + [\hat{\beta}_1(x)]_1 1\{x \geq c\}$

$\hat{\theta} = \hat{m}(c+) - \hat{m}(c-) = [\hat{\beta}_1(c)]_1 - [\hat{\beta}_0(c)]_1$

RDD with Covariates

$E(y_{0i} \mid X_i = x, Z_i = z) = m_0(x) + z^T\beta$

$E(y_{1i} \mid X_i = x, Z_i = z) = m_1(x) + z^T\beta$

$E(y_i \mid X_i = x, Z_i = z) = m(x,z) = m_0(x)1\{x<c\} + m_1(x)1\{x \geq c\} + z^T\beta$

Conditional ATE $\theta = \theta(c,z) = m(c+,z) - m(c-,z)$

① RDD LL reg $y_i \sim X_i \Rightarrow \hat{m}(X_i)$, $y_i - \hat{m}(X_i)$

② LL reg $Z_i \sim X_i \Rightarrow \hat{g}(X_i)$, $Z_i - \hat{g}(X_i)$

③ OLS reg $y_i - \hat{m}(X_i) \sim Z_i - \hat{g}(X_i) \Rightarrow \hat{\beta}$, $\hat{e}_i = y_i - Z_i^T\hat{\beta}$

FWL

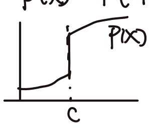
④ RDD LL reg $\hat{e}_i \sim X_i \Rightarrow \hat{m}(x,z)$

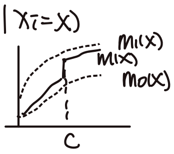
Fuzzy Regression Discontinuity FRD

$E(y_{1i} \mid X_i = x) = m_1(x) \qquad E(y_{0i} \mid X_i = x) = m_0(x)$

$p(x) = P(T_i = 1 \mid X_i = x) \qquad \theta = \theta(c) = E(y_{1i} - y_{0i} \mid X_i = c)$

$= m_1(c) - m_0(c)$





$m(x) = p(x)m_1(x) + (1-p(x))m_0(x)$

$= m_0(x) + p(x)[m_1(x) - m_0(x)]$

Thm: $m_0(x)$, $m_1(x)$ are continuous at $x=c$.

$p(x)$ is discontinuous at $x=c$.

$T_i \perp\!\!\!\perp \theta_i$ for X_i near c.

$\theta = \dfrac{m_1(c+) - m_0(c-)}{p(c+) - p(c-)}$

Estimation: LL reg : $y_i \sim X_i$: $\hat{m}_1(c+) - \hat{m}_0(c-) = [\hat{\beta}_1(c)]_1 - [\hat{\beta}_0(c)]_1$

LL reg: $T_i \sim X_i$: $\hat{p}_1(c+) - \hat{p}_0(c-)$

$\hat{\theta} = \dfrac{\hat{m}_1(c+) - \hat{m}_0(c-)}{\hat{p}(c+) - \hat{p}(c-)}$

Nonparametric Regression: $y_i = m(x_i)+e_i$

$E(e_i|x_i)=0 \quad E(e_i^2|x_i)=\sigma^2(x_i)$

Binned mean estimator $\hat{m}(x)=\dfrac{\sum_{i=1}^n 1(|x_i-x|\le h)y_i}{\sum_{i=1}^n 1(|x_i-x|\le h)}$

bandwidth h

Kernel regression

Kernel $K(u)$: $0\le k(u)\le \bar{k}<+\infty$, $k(u)=k(-u)$

$\int_{-\infty}^{+\infty} k(u)du=1 \qquad \int_{-\infty}^{+\infty}|u|^r k(u)du<+\infty \quad \forall r\in\mathbb{N}_+$

Normalized kernel: $\int u^2 k(u)du=1$

Nadaraya-Watson: $\hat{m}_{nw}(x)=\dfrac{\sum_{i=1}^n k(\frac{x_i-x}{h})y_i}{\sum_{i=1}^n k(\frac{x_i-x}{h})}$ (local constant)

$\lim_{h\to 0}\hat{m}_{nw}(x)=y|x \qquad \lim_{h\to+\infty}\hat{m}_{nw}(x)=\bar{y}$

NW: $y_i=m(x_i)+e_i\simeq m(x)+e_i$ for $x_i\simeq x$

$\hat{m}_{nw}(x)=\underset{m}{\arg\min}\sum_{i=1}^n k(\frac{x_i-x}{h})(y_i-m)^2$

LL (Local linear): $y_i=m(x_i)+e_i\simeq m(x)+m'(x)(x_i-x)+e_i$

$\{\hat{m}_{LL}(x),\hat{m}'_{LL}(x)\}=\underset{\alpha,\beta}{\arg\min}\sum_{i=1}^n k(\frac{x_i-x}{h})(y_i-\alpha-\beta(x_i-x))^2$

$\Leftrightarrow y_i=z_i(x)^T\beta(x)+e_i, \quad z_i(x)=\begin{bmatrix}1\\x_i-x\end{bmatrix} \quad \beta(x)=\begin{bmatrix}m(x)\\m'(x)\end{bmatrix}$

$\hat{\beta}_{LL}(x)=\left(\sum_{i=1}^n k(\frac{x_i-x}{h})z_i(x)z_i(x)^T\right)^{-1}\left(\sum_{i=1}^n k(\frac{x_i-x}{h})z_i(x)y_i\right)$

$=(Z^TKZ)^{-1}(Z^TKy)$

$K=\mathrm{diag}\{k(\frac{x_1-x}{h}),\dots,k(\frac{x_n-x}{h})\}$

$E(\hat{m}_{nw}(x)|x)=\dfrac{\sum k(\frac{x_i-x}{h})m(x)}{\sum k(\frac{x_i-x}{h})}$

$Var(\hat{m}_{nw}(x)|x)=\dfrac{\sum k(\frac{x_i-x}{h})^2\sigma^2(x)}{(\sum k(\frac{x_i-x}{h}))^2}$

Dynamic Panel AR(1) $Y_{it}=\alpha Y_{it-1}+u_i+\varepsilon_{it}$

Fixed effect: $\Delta Y_{it}=\alpha\Delta Y_{it-1}+\Delta\varepsilon_{it}$

$T=3, p=1$; $\hat{\alpha}_{FE}=(\sum_{i=1}^N \Delta Y_{i2}^2)^{-1}(\sum_{i=1}^N \Delta Y_{i2}\Delta Y_{i3})\to -\frac{1+\alpha}{2}$

Anderson-Hsiao Estimator

$\Delta Y_{it}=\alpha_1\Delta Y_{it-1}+\alpha_2\Delta Y_{it-2}+\dots+\alpha_p\Delta Y_{it-p}+\Delta X_{it}^T\beta+\Delta\varepsilon_{it}$

$E(\Delta Y_{it-1}\Delta\varepsilon_{it})=E[(Y_{it-1}-Y_{it-2})(\varepsilon_{it}-\varepsilon_{it-1})]=-\sigma_\varepsilon^2\neq 0$

$E(\Delta Y_{it-2}\Delta\varepsilon_{it})=E(Y_{it-2}\varepsilon_{it})-E(Y_{it-2}\varepsilon_{it-1})=0$

IV: $(\Delta Y_{it-2},\dots,\Delta Y_{it-p-1})$ for $(\Delta Y_{it-1},\dots,\Delta Y_{it-p})$

$T=3, p=1$ No X:

$\hat{\alpha}_{IV}=(\sum_{i=1}^N Y_{i1}\Delta Y_{i2})^{-1}(\sum_{i=1}^N Y_{i1}\Delta Y_{i3})\to\alpha$

Panel Data (y_{it}, x_{it}) $\quad y_i: T_i\times 1 \quad X_i: T_i\times K$

Pooled Regression : $y_{it}=x_{it}^T\beta+e_{it} \quad E(x_{it}e_i)=0$

$y_i=X_i\beta+e_i \quad E(X_i^Te_i)=0$

$Y=X\beta+e$

$\hat{\beta}_{pool}=(\sum_{i=1}^N\sum_{t\in S_i}x_{it}x_{it}^T)^{-1}(\sum_{i=1}^N\sum_{t\in S_i}x_{it}y_{it})$

$=(\sum_{i=1}^N X_i^TX_i)^{-1}(\sum_{i=1}^N X_i^Ty_i)=(X^TX)^{-1}X^TY$

$=\beta+(\sum_{i=1}^n X_i^TX_i)^{-1}(\sum_{i=1}^N X_i^Te_i)$

Strict mean independence: $E(e_{it}|X_i)=0 \Rightarrow \hat{\beta}_{pool}\xrightarrow{P}\beta$

$\sqrt{n}(\hat{\beta}_{pool}-\beta)\xrightarrow{d}N(0,V_{pool})$.

$V_{pool}=(EX_i^TX_i)^{-1}E(X_i^Te_ie_i^TX_i)(EX_i^TX_i)^{-1}$

$\hat{V}_{pool}=(\sum X_i^TX_i)^{-1}(\sum X_i^T\hat{e}_i\hat{e}_i^TX_i)(\sum X_i^TX_i)^{-1}\xrightarrow{P}\frac{1}{n}V_{pool}$

One-Way Error Component Model : $e_{it}=u_i+\varepsilon_{it}$

$y_{it}=x_{it}^T\beta+u_i+\varepsilon_{it}$

Random Effect: $E(u_i|X_i)=0 \quad E(u_i^2|X_i)=\sigma_u^2 \quad E(u_i\varepsilon_{it}|X_i)=0$

$E(\varepsilon_{it}|X_i)=0 \quad E(\varepsilon_{it}^2|X_i)=\sigma_\varepsilon^2 \quad E(\varepsilon_{it}\varepsilon_{js}|X_i)=0$

$E(e_i|X_i)=0 \quad E(e_ie_i^T|X_i)=\begin{bmatrix}\sigma_u^2+\sigma_\varepsilon^2 & \cdots & \sigma_u^2\\ \sigma_u^2 & \ddots & \vdots\\ \sigma_u^2 & \cdots\ \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2\end{bmatrix}=\sigma_\varepsilon^2\Omega_i$.

GLS $y_i=X_i\beta+e_i \quad E(e_i|X_i)=0 \quad E(e_ie_i^T|X_i)=\Omega_i\sigma_\varepsilon^2$.

$\Omega_i^{-\frac{1}{2}}y_i=\Omega_i^{-\frac{1}{2}}X_i\beta+\Omega_i^{-\frac{1}{2}}e_i$.

$\hat{\beta}_{gls}=(\sum_{i=1}^N X_i^T\Omega_i^{-1}X_i)^{-1}(\sum_{i=1}^N X_i^T\Omega^{-1}Y_i)$ if σ_ε^2, σ_u^2 are known

$\hat{\beta}_{gls}-\beta=(\sum_{i=1}^N X_i^T\Omega_i^{-1}X_i)^{-1}(\sum_{i=1}^N X_i^T\Omega_i^{-1}e_i)$

$\hat{\beta}_{gls}\xrightarrow{P}\beta$, $\sqrt{n}(\hat{\beta}_{gls}-\beta)\xrightarrow{d}N(0,V_{gls}) \quad V_{gls}=(EX_i^T\Omega_iX_i)^{-1}\sigma_\varepsilon^2$

$\hat{V}_{gls}=(\sum_{i=1}^N X_i^T\Omega_i^{-1}X_i)^{-1}\sigma_\varepsilon^2\xrightarrow{P}\frac{1}{n}V_{gls}$

$\hat{V}_{pool}\geq\hat{V}_{gls}$ $\quad\hat{\beta}_{gls}$ is more efficient.

"=" $\Leftrightarrow\sigma_u^2=0$ No individual effects

Cluster-robust $\hat{V}_{gls}=(\sum_{i=1}^N X_i^T\Omega_i^{-1}X_i)^{-1}(\sum_{i=1}^N X_i^T\Omega_i^{-1}\hat{e}_i\hat{e}_i^T\Omega_i^{-1}X_i)(\quad)^{-1}$

Fixed Effect: $Cov(u_i,X_{it})\neq 0 \quad Y_{it}=X_{it}^T\beta+u_i+\varepsilon_{it}$

$\hat{\beta}_{pool}$ & $\hat{\beta}_{gls}$ are biased due to endogeneity.

X_{it} is strictly exogenous : $E(X_{is}\varepsilon_{it})=0$.

$\Uparrow$

strict mean independent $E(\varepsilon_{it}|X_i)=0$

Within transformation: $\dot{Y}_{it}=Y_{it}-\bar{Y}_i$ i.e. $\dot{Y}_i=M_iY_i$

individual-specific mean $\bar{Y}_i=\frac{1}{T_i}\sum_{t=1}^{T_i}Y_{it}=(1_i^T1_i)^{-1}1_iY_{it}$

$Y_{it}=X_{it}^T\beta+u_i+\varepsilon_{it}\Rightarrow\bar{Y}_i=\bar{X}_i^T\beta+\bar{u}_i+\bar{\varepsilon}_i$

$\Rightarrow Y_{it}-\bar{Y}_i=(X_{it}-\bar{X}_i)^T\beta+(\varepsilon_{it}-\bar{\varepsilon}_i)$

$\dot{Y}_{it}=\dot{X}_{it}^T\beta+\dot{\varepsilon}_{it}$

$\hat{\beta}_{fe}=(\sum_{i=1}^N\sum_{t=1}^{T_i}\dot{X}_{it}\dot{X}_{it}^T)^{-1}(\sum_{i=1}^N\sum_{t=1}^{T_i}\dot{X}_{it}\dot{Y}_{it})$

$=(\sum_{i=1}^N\dot{X}_i^T\dot{X}_i)^{-1}(\sum_{i=1}^N\dot{X}_i^T\dot{Y}_i)=(\sum_{i=1}^N X_i^TM_iX_i)^{-1}(\sum_{i=1}^N X_i^TM_iY_i)$

$\hat{\beta}_{fe}-\beta=(\sum_{i=1}^N X_i^TM_iX_i)^{-1}(\sum_{i=1}^N X_i^TM_i\varepsilon_i)$

$\hat{\beta}_{fe}\xrightarrow{P}\beta$. $\sqrt{n}(\hat{\beta}_{fe}-\beta)\xrightarrow{d}N(0,V_{fe})$

$V_{fe}=(E\dot{X}_i^T\dot{X}_i)^{-1}E(\dot{X}_i^T\varepsilon_i\varepsilon_i^T\dot{X}_i)(E\dot{X}_i^T\dot{X}_i)^{-1}$

$\hat{V}_{fe}=Var(\hat{\beta}_{fe}|X)=(\sum_{i=1}^N\dot{X}_i^T\dot{X}_i)^{-1}(\sum_{i=1}^N\dot{X}_i^T\Sigma_i\dot{X}_i)(\sum_{i=1}^N\dot{X}_i^T\dot{X}_i)^{-1}$

$\Sigma_i=E(\varepsilon_i\varepsilon_i^T|X)$

$\hat{V}_{pool} \geq \hat{V}_{fe}$ since $\dot{x}_i$ has reduced variation relative to x_i

Differenced Estimator:

First-differencing transformation: $\Delta Y_{it} = Y_{it} - Y_{it-1}$

$\Delta Y_i = D_i Y_i \quad D_i : (T_i - 1) \times T_i \quad \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ & -1 & 1 & & \\ & & & \ddots & -1 \quad 1 \end{bmatrix}$

$\Rightarrow \Delta Y_i = \Delta X_i \beta + \Delta \varepsilon_i$

$\hat{\beta}_\Delta = \left(\sum_{i=1}^N \sum_{t \geq 2} \Delta X_{it} \Delta X_{it}^T\right)^{-1} \left(\sum_{i=1}^N \sum_{t \geq 2} \Delta X_{it} \Delta Y_{it}\right)$

$= \left(\sum_{i=1}^N \Delta X_i^T \Delta X_i\right)^{-1} \left(\sum_{i=1}^N \Delta X_i^T \Delta Y_i\right)$

$= \left(\sum_{i=1}^N X_i^T D^T D X_i\right)^{-1} \left(\sum_{i=1}^N X_i^T D^T D Y_i\right) \qquad T=2 \Leftrightarrow \hat{\beta}_\Delta = \hat{\beta}_{fe}$

If ε_{it} is iid: $\Delta \varepsilon_i = D_i \varepsilon_i$, $E(\Delta \varepsilon_i \Delta \varepsilon_i^T) = D_i D_i^T \sigma_\varepsilon^2 =: H \sigma_\varepsilon^2$.

GLS: $\tilde{\beta}_\Delta = \left(\sum_{i=1}^N \Delta X_i^T H^{-1} \Delta X_i\right)^{-1} \left(\sum_{i=1}^N \Delta X_i^T H^{-1} \Delta Y_i\right) = \hat{\beta}_{fe}$

since $M_i = D_i^T (D_i D_i^T)^{-1} D_i$

Dummy Variables Regression

$Y_{it} = X_{it}^T \beta + d_i^T u + \varepsilon_{it}$. $\quad u_i = d_i^T u$

$Y = X\beta + DU + \varepsilon$. $\quad D = \text{diag}\{I_{T_1}, I_{T_2}, \cdots, I_{T_N}\}$

$\hat{\beta}_D = (X^T M_D X)^{-1} X^T M_D Y = \hat{\beta}_{fe} = \beta + (X^T M_D X)^{-1} X^T M_D \varepsilon$

Between Estimator: $\bar{Y}_i = \bar{X}_i^T \beta + u_i + \bar{\varepsilon}_i$

At individual level: $\hat{\beta}_{be} = \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i^T\right)^{-1} \left(\sum_{i=1}^N \bar{X}_i \bar{Y}_i\right) = \beta + \left(\sum \bar{X}_i \bar{X}_i^T\right)^{-1} \left(\sum \bar{X}_i \bar{e}_i\right)$

At observation level: $\tilde{\beta}_{be} = \left(\sum_{i=1}^N T_i \bar{X}_i \bar{X}_i^T\right)^{-1} \left(\sum_{i=1}^N T_i \bar{X}_i \bar{Y}_i\right)$

$V_{be} = Var(\hat{\beta}_{be} | X) = \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i^T\right)^{-1} \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i^T \sigma_i^2\right) \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i^T\right)^{-1}$

$\sigma_i^2 = \sigma_u^2 + \frac{\sigma_\varepsilon^2}{T_i}$

Balanced panel: $\hat{\beta}_{be} = \tilde{\beta}_{be}$. $\quad V_{be} = \left(\sum_{i=1}^N \bar{X}_i \bar{X}_i^T\right)^{-1} \left(\sigma_u^2 + \frac{\sigma_\varepsilon^2}{T}\right)$

Hausman Test for Random vs Fixed Effects

H_0: Random effect $\qquad H_1$: Fixed effect.

Hausman statistic: $H = (\hat{\beta}_{fe} - \hat{\beta}_{re})^T \widehat{var}(\hat{\beta}_{fe} - \hat{\beta}_{re})^{-1} (\hat{\beta}_{fe} - \hat{\beta}_{re})$

$= (\hat{\beta}_{fe} - \hat{\beta}_{re})^T (\hat{V}_{fe} - \hat{V}_{re})^{-1} (\hat{\beta}_{fe} - \hat{\beta}_{re})$

Time Trends: $Y_{it} = X_{it}^T \beta + \gamma_i t + u_i + \varepsilon_{it}$

Two-way Error Component $\quad Y_{it} = X_{it}^T \beta + v_t + u_i + \varepsilon_{it}$

Two-way within transformation: $\ddot{Y}_{it} = Y_{it} - \bar{Y}_i - (\tilde{Y}_t - \bar{Y})$

time-specific mean $\tilde{Y}_t = \frac{1}{N_t} \sum_i Y_{it}$

$\ddot{Y}_{it} = \ddot{X}_{it}^T \beta + \ddot{\varepsilon}_{it}$

① $Y_{it} = X_{it}^T \beta + \tau_t^T v + u_i + \varepsilon_{it} \qquad \tau_t$: Dummy

② $\dot{Y}_{it} = \dot{X}_i^T \beta + \dot{\tau}_t^T v + \dot{\varepsilon}_{it} \qquad$ FWL: $\hat{\beta}$

Instrumental Variables

Logistic Regression

$y_{it} = 1\{X_{it}^T \beta + \alpha_i + \varepsilon_{it} \geq 0\}$

$\varepsilon_i | \alpha_i, X_i \sim$ Logistic $\qquad P(y_{it} = 1 | X_{i1}, X_{i2}, \alpha_i) = \frac{\exp\{X_{it}^T \beta + \alpha_i\}}{1 + \exp\{X_{it}^T \beta + \alpha_i\}}$

$A = \{y_{i1} = 0, y_{i2} = 1\}$

$P(A | X_{i1}, X_{i2}, \alpha_i) = \frac{1}{1 + \exp\{X_{i1}^T \beta + \alpha_i\}} \cdot \frac{\exp\{X_{i2}^T \beta + \alpha_i\}}{1 + \exp\{X_{i2}^T \beta + \alpha_i\}}$

$P(B | X_{i1}, X_{i2}, \alpha_i) = \frac{\exp\{X_{i1}^T \beta + \alpha_i\}}{1 + \exp\{X_{i1}^T \beta + \alpha_i\}} \cdot \frac{1}{1 + \exp\{X_{i2}^T \beta + \alpha_i\}}$

$P(A | A \cup B, X_{i1}, X_{i2}, \alpha_i) = \frac{\exp\{X_{i2}^T \beta + \alpha_i\}}{\exp\{X_{i2}^T \beta + \alpha_i\} + \exp\{X_{i1}^T \beta + \alpha_i\}}$

$= \left[1 + \exp\{(X_{i1} - X_{i2})^T \beta + \alpha_i\}\right]^{-1}$

$P(B | A \cup B, X_{i1}, X_{i2}, \alpha_i) = 1 - P(A | A \cup B, X_{i1}, X_{i2}, \alpha_i)$

Conditional likelihood function:

$L_n = \prod_{i=1}^n P(A | A \cup B, X_i, \alpha_i)^{(1-y_{i1}) y_{i2}} P(B | A \cup B, X_i, \alpha_i)^{y_{i1}(1-y_{i2})}$

Binary Choice Model: $\quad y_i = 1\{X_i^T \beta \geq u_i\} \quad u_i \perp\!\!\!\perp X_i$

$E(y_i | X_i) = P(y_i = 1 | X_i) = F_u(X_i^T \beta) \quad$ Index model

$\frac{\partial E(y|x)}{\partial x} = f_u(x^T \beta) \beta \Rightarrow \frac{\partial E(y|x) / \partial x_i}{\partial E(y|x) / \partial x_j} = \frac{\beta_i}{\beta_j}$

X_J is binary: $E(y|x) = F_u(X_c^T \beta_c + X_J \beta_J)$

$E(y | X_c^T \beta_c = a, X_J = 1) = F_u(a + \beta_J)$

$E(y | X_c^T \beta_c = b, X_J = 0) = F_u(b)$

If $F_u(a + \beta_J) = F_u(b)$ then $\beta_J = b - a$

$E(y|x) = \int y f(y|x) dy$, $\hat{E}(y|x) = \frac{\frac{1}{n} \sum_{i=1}^n y_i \frac{1}{h} k\left(\frac{X_i - X}{h}\right)}{\frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - X}{h}\right)}$

$\lim_{h \to 0} \hat{E}(y|x) = m(x) = E(y|x)$

12.10 $Y = X^T\beta + e$ $E(Ze)=0$

$X = \Gamma^T Z + u$ $E(Zu^T)=0$

Control function $e = u^T\gamma + v$ $E(uv)=0$ observed u

$Y = X^T\beta + u^T\gamma + v$ OLS $(\hat\beta, \hat\gamma)$

a. WTS: $E(Xv)=0$

$E(Xv) = E((\Gamma^T Z + u)v) = \Gamma^T E(Zv) + Euv$

$= \Gamma^T E(Z(e - u^T\gamma)) + Euv$

$= \Gamma^T E(Ze) - \Gamma^T E(Zu^T)\gamma + Euv = 0$

b. Asymptotic distribution $(\hat\beta, \hat\gamma)$

In matrix form: $Y = [X\ \ u]\begin{bmatrix}\beta\\ \gamma\end{bmatrix} + v$.

$$\begin{bmatrix}\hat\beta\\ \hat\gamma\end{bmatrix} = \begin{bmatrix}\beta\\ \gamma\end{bmatrix} + \begin{bmatrix}X^TX & X^Tu\\ u^TX & u^Tu\end{bmatrix}^{-1}\begin{bmatrix}X^Tv\\ u^Tv\end{bmatrix}$$

$$\sqrt{n}\left(\begin{bmatrix}\hat\beta\\ \hat\gamma\end{bmatrix} - \begin{bmatrix}\beta\\ \gamma\end{bmatrix}\right) = \begin{bmatrix}\frac1n\sum_{i=1}^n x_ix_i^T & \frac1n\sum_{i=1}^n x_iu_i^T\\ \frac1n\sum_{i=1}^n u_ix_i^T & \frac1n\sum_{i=1}^n u_iu_i^T\end{bmatrix}^{-1}\sqrt{n}\begin{bmatrix}\frac1n\sum_{i=1}^n x_iv_i\\ \frac1n\sum_{i=1}^n u_iv_i\end{bmatrix}$$

$$\xrightarrow{d} \begin{bmatrix}E(XX^T) & E(Xu^T)\\ E(uX^T) & E(uu^T)\end{bmatrix}^{-1} N\left(0, \begin{bmatrix}E(XX^Tv^2) & E(Xu^Tv^2)\\ E(uX^Tv^2) & E(uu^Tv^2)\end{bmatrix}\right)$$

12.11 $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$ $E(Xe)\neq 0$

Instrument Z: $E(e|Z)=0$, $\Rightarrow E(e)=E(Ze)=E(Z^2e)=0$

(a) X^2 is endogenous since X is endogenous.

(b) $X = \gamma_0 + \gamma_1 Z + u$ $u \perp\!\!\!\perp Z$. $E(u)=0$

$(1, Z, Z^2)$ is a sufficient number of instruments.

The structural equation is just-identified

(c) $X^2 = (\gamma_0+u)^2 + 2(\gamma_0+u)\gamma_1 Z + \gamma_1^2 Z^2$

$= \gamma_0^2 + Eu^2 + 2\gamma_0\gamma_1 Z + \gamma_1^2 Z^2 + \underbrace{2\gamma_1 Z u + 2\gamma_0 u + u^2 - Eu^2}_{v}$

$\Rightarrow Y = \beta_0 + \beta_1\gamma_0 + \beta_1\gamma_1 Z + \beta_1 u$

$+ \beta_2(\gamma_0^2 + Eu^2) + 2\beta_2\gamma_0\gamma_1 Z + \beta_2\gamma_1^2 Z^2 + \beta_2 v + e$

$= [\beta_0 + \beta_1\gamma_0 + \beta_2(\gamma_0^2 + Eu^2)] + [\beta_1\gamma_1 + 2\beta_2\gamma_0\gamma_1] Z$

$+ \beta_2\gamma_1^2 Z^2 + \underbrace{\beta_1 u + \beta_2 v + e}_{\varepsilon}$. $E(\varepsilon) = E(Z\varepsilon) = E(Z^2\varepsilon) = 0$

To ensure $(\beta_0, \beta_1, \beta_2)$ is identified:

$$\begin{bmatrix}1 & \gamma_0 & \gamma_0^2 + Eu^2\\ 0 & \gamma_1 & 2\gamma_0\gamma_1\\ 0 & 0 & \gamma_1^2\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} = \hat\alpha_{OLS} \text{ from Reg } (Y \sim (1, Z, Z^2))$$

Matrix has full rank. $\Leftrightarrow$ det $\neq 0$ $\Leftrightarrow$ $\gamma_1^3 \neq 0$ i.e. $\gamma_1 \neq 0$

12.13 $Y_1 = Z_1^T\beta_1 + Y_2^T\beta_2 + e$. $E(Ze)=0$

$Z = (Z_1, Z_2)$ $Y_1 = Z_1^T\pi_1 + Z_2^T\pi_2 + u_1$

$H_0: \beta_2 = 0$

Suppose Y_2 has reduced form equation:

$Y_2 = \Gamma^T Z + u_2 = \Gamma_{12}^T Z_1 + \Gamma_{22}^T Z_2 + u_2$. $\Gamma = \begin{bmatrix}\Gamma_{12}\\ \Gamma_{22}\end{bmatrix}$

$\Rightarrow Y_1 = Z_1^T\beta_1 + Z_1^T\Gamma_{12}\beta_2 + Z_2^T\Gamma_{22}^T\beta_2 + u_2^T\beta_2 + e$

$= Z_1^T\pi_1 + Z_2^T\pi_2 + u_1$

where $\pi_1 = \beta_1 + \Gamma_{12}\beta_2$; $\pi_2 = \Gamma_{22}^T\beta_2$; $u_1 = u_2^T\beta_2 + e$

Since rank$(ZX) = k \Leftrightarrow$ rank$(\Gamma_{22}) = k_2$, $\pi_2 = 0 \Leftrightarrow \beta_2 = 0$

$H_0: \beta_2 = 0 \Leftrightarrow H_0: \pi_2 = 0$ v.s. $H_1: \pi_2 \neq 0$

OLS: $Y_1 = Z_1^T\hat\pi_1 + Z_2^T\hat\pi_2 + \hat u_1$

FWL: $\hat\pi_2 = (Z_2^TM_1Z_2)^{-1}Z_2M_1Y_1$ $M_1 = I_n - Z_1(Z_1^TZ_1)^{-1}Z_1^T$

$= (Z_2^TM_1Z_2)^{-1}Z_2M_1(Z_1\pi_1 + Z_2\pi_2 + u_1)$

$= \pi_2 + (Z_2^TM_1Z_2)^{-1}Z_2M_1u_1$

$\sqrt{n}(\hat\pi_2 - \pi_2) = (Z_2^TM_1Z_2)^{-1}Z_2^TM_1\sqrt{n}u_1 \xrightarrow{d} N(0, V_{\pi_2})$

Wald statistic: $\hat W(\pi_2) = n(\hat\pi_2 - \pi_2)^T\hat V_{\pi_2}^{-1}(\hat\pi_2 - \pi_2)$

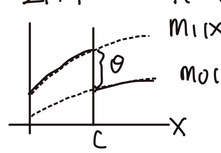
$\hat V_{\pi_2} = (Z_2^TM_1Z_2)^{-1}(\sum_{i=1}^n Z_{2i}^TM_1Z_{2i}\hat u_i^2)(Z_2^TM_1Z_2)^{-1}$

$\hat W(\pi_2 = 0)|H_0 \sim \chi^2_{k_2}$ with CDF $G_{k_2}(\cdot)$

Find asymptotic critical value c: $\alpha = 1 - G_{k_2}(c)$

Decision rule: "Reject H_0 iff $\hat W > c$" has asymptotic size α

21.1 $T_i = 1\{x_i \le c\}$



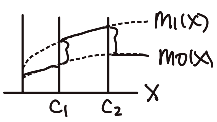
$E(y_{1i}|x_i = x) = m_1(x)$

$E(y_{0i}|x_i = x) = m_0(x)$

CATE $= \theta = \theta(c) = m_1(c) - m_0(c)$

$\theta = m_1(c-) - m_0(c+)$ if m_1, m_0 are continuous at $x = c$.

21.2 $T_i = 1\{c_1 \le x_i \le c_2\}$



CATE$_1 = \theta_1 = \theta(c_1) = m_1(c_1+) - m_0(c_1-)$

CATE$_2 = \theta_2 = \theta(c_2) = m_1(c_2-) - m_0(c_2+)$

21.3 WTS $m(x) = m_0(x)1\{x<c\} + m_1(x)1\{x\ge c\}$

Proof. $y_i = T_iy_{1i} + (1-T_i)y_{0i}$

$E(y_i|x_i=x) = E(T_iy_{1i}|x_i=x) + E((1-T_i)y_{0i}|x_i=x)$

$\Rightarrow m(x) = P(T_i=1|x_i=x)E(y_{1i}|x_i=x)$

$+ P(T_i=0|x_i=x)E(y_{0i}|x_i=x)$

$= 1\{x\ge c\}m_1(x) + 1\{x<c\}m_0(x)$ ∎

21.4 $y_i = \beta_0 + \beta_1x_i + \beta_3(x_i - c)T_i + \theta T_i + e_i$ on $|x_i - c| \le h$

$\Leftrightarrow$ LL regression with a Rectangular bandwidth

Proof.

$E(y_{1i}|x_i=x) = m_1(x) = \beta_0 + \beta_1x + \beta_3(x-c) + \theta$

$E(y_{0i}|x_i=x) = m_0(x) = \beta_0 + \beta_1x$

LL: $\theta = m_1(c+) - m_0(c-) = \beta_1c + \theta - \beta_1c = \theta$ ∎