1 ATE under exogeneity

1.1 Potential outcome framework

- In a structural model, we observe $y_i = g(x_i, u_i)$
- Potential outcome: $y_i(x^0) = g(x^0, u_i)$
- TE $= y_i(x^1) - y_i(x^0)$, ATE $= \mathbb{E}[y_i(x^1) - y_i(x^0)]$, ME $= \frac{\partial}{\partial x} g(x^0, u_i)$, AME $= \mathbb{E}[\frac{\partial}{\partial x} g(x^0, u_i)]$
- ASF$(x^0) = \mathbb{E}[y_i(x^0)] \neq \mathbb{E}[y_i \,|\, x_i = x^0]$, if x_i and u_i not independent

1.2 Causality in the linear model

- Linear model: $y_i = \beta_1 x_i + \beta_2 + u_i \implies$ TE $= \beta_1(x^1 - x^0)$, ME $= \beta_1$
- OLS: Slope coefficient is consistent for β_1 only if $\mathbb{E}[u_i \,|\, x_i] = 0$
- Random coefficient model: $y_i = \beta_{1,i} x_i + \beta_2 + u_i \implies$ ME $= \beta_{1,i}$ heterogeneous
- Assuming $\mathbb{E}[u_i \,|\, x_i] = 0$ and $\mathbb{E}[\beta_{1,i} \,|\, x_i] = \mathbb{E}[\beta_{1,i}]$, OLS consistently est. AME $= \mathbb{E}[\beta_{1,i}]$

1.3 ATE under random assignment

- Observed outcome: $y_i = y_i(1)x_i + y_i(0)(1 - x_i) = \Delta_i x_i + y_i(0)$
- ATE $= \mathbb{E}[\Delta_i] = \mathbb{E}[y_i(1) - y_i(0)]$
- Random assignment: $(y_i(0), y_i(1)) \perp\!\!\!\perp x_i$, e.g. when $x_i \perp\!\!\!\perp u_i$
- Identification: $\mathbb{E}[\Delta_i] = \mathbb{E}[y_i \,|\, x_i = 1] - \mathbb{E}[y_i \,|\, x_i = 0]$
- Rewrite $y_i = \Delta_i x_i + \mathbb{E}[y_i(0)] + e_i$ with $e_i = y_i(0) - \mathbb{E}[y_i(0)]$
- Checking assumptions $\implies$ OLS consistently estimates ATE

1.4 ATE under unconfoundedness

- $x_i = (d_i, w_i')'$ with binary treatment indicator d_i and control variables w_i
- Unconfoundedness: $(y_i(0), y_i(1)) \perp\!\!\!\perp d_i \,|\, w_i$, e.g. when $d_i \perp\!\!\!\perp u_i \,|\, w_i$
- Overlap: $\mathbb{P}(d_i = 1 \,|\, w_i) \in (0, 1)$
- Identification: CATE$(w_i) = \mathbb{E}[\Delta_i \,|\, w_i] = \mathbb{E}[y_i \,|\, d_i = 1, w_i] - \mathbb{E}[y_i \,|\, d_i = 0, w_i]$, ATE via LIE
- ATT $= \mathbb{E}[\Delta_i \,|\, d_i = 1]$ is identified if $y_i(0) \perp\!\!\!\perp d_i \,|\, w_i$ and $\mathbb{P}(d_i = 1 \,|\, w_i) < 1$
- Assume $y_i(0) = \alpha + \beta w_i + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i \,|\, w_i] = 0$. If $\Delta_i \equiv \Delta$, then $y_i = \alpha + \Delta d_i + \beta w_i + \varepsilon_i$ where unconfoundedness is equivalent to $\varepsilon_i \perp\!\!\!\perp d_i \,|\, w_i$. Hence, OLS estimates Δ. However, for heterogeneous TE, $\mathbb{E}[\Delta_i \,|\, d_i, w_i] = \mathbb{E}[\Delta_i \,|\, w_i] \neq \mathbb{E}[\Delta_i]$ s.t. OLS is inconsistent for $\mathbb{E}[\Delta_i]$.
- Under unconfoundedness, using OLS in the linear regression $y_i = \alpha_0 + \alpha_1 w_i + \beta d_i + e_i$ with $\mathbb{E}[e_i \,|\, w_i, d_i] = 0$ consistently estimates $\beta = \mathbb{E}[\omega(w_i)\text{CATE}(w_i)]$, $\omega(w_i) = \frac{\text{Var}(d_i \,|\, w_i)}{\mathbb{E}[\text{Var}(d_i \,|\, w_i)]} \geq 0$. If $\Delta_i \equiv \Delta$, we get $\beta = \Delta$. (Proof idea: $\widetilde{d}_i = d_i - \mathbb{E}[d_i \,|\, w_i]$, $\beta = \mathbb{E}[\widetilde{d}_i y_i]/\mathbb{E}[\widetilde{d}_i^{\,2}]$, LIE, U.A.)

1.5 Propensity score weighting

- Propensity score: $p(w_i) = \mathbb{P}(d_i = 1 \,|\, w_i)$
- Under unconfoundedness and overlap: ATE $= \mathbb{E}[y_i(1)] - \mathbb{E}[y_i(0)] = \mathbb{E}[\frac{y_i d_i}{p(w_i)}] - \mathbb{E}[\frac{y_i(1-d_i)}{1-p(w_i)}] = \mathbb{E}[\omega_i y_i]$
- Missing data: $y_i = d_i y_i^*$ with binary response indicator d_i

- MCAR: $y_i^* \perp\!\!\!\perp d_i$, $\mathbb{E}[y_i^*] = \mathbb{E}[y_i/\mathbb{P}(d_i = 1)]$
- MAR: $y_i^* \perp\!\!\!\perp d_i \,|\, w_i$, $\mathbb{E}[y_i^*] = \mathbb{E}[y_i/p(w_i)]$

2 Instrumental variables

2.1 IV estimator in the univariate setup

- Structural model $y_i = \beta_1 x_i + \beta_2 + e_i$ with endogeneity $\mathrm{Cov}(x_i, e_i) \neq 0$
- $\hat{\beta}_{1,\mathrm{OLS}} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \xrightarrow{p} \frac{\mathrm{Cov}(x_i, y_i)}{\mathrm{Var}(x_i)} = \beta_1 + \frac{\mathrm{Cov}(x_i, e_i)}{\mathrm{Var}(x_i)}$ inconsistent
- IV z_i satisfies $\mathbb{E}[e_i \,|\, z_i] = 0$ or $\mathrm{Cov}(z_i, e_i) = 0$ (exogeneity) and $\mathrm{Cov}(z_i, x_i) \neq 0$ (relevance)
- Under IV assumptions: $\beta_1 = \frac{\mathrm{Cov}(z_i, y_i)}{\mathrm{Cov}(z_i, x_i)}$, $\hat{\beta}_{1,\mathrm{IV}} = \frac{\sum_{i=1}^n (z_i - \bar{z}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (z_i - \bar{z}_n)(x_i - \bar{x}_n)}$

2.2 Random coefficient model

- $y_i = \beta_{1,i} x_i + \beta_2 + e_i$, $x_i = \delta_{1,i} z_i + \delta_2 + u_i$ under IV assumptions and $z_i \perp\!\!\!\perp (\beta_{1,i}, \delta_{1,i}, u_i)$
- $\hat{\beta}_{1,\mathrm{IV}} \xrightarrow{p} \frac{\mathrm{Cov}(z_i, y_i)}{\mathrm{Cov}(z_i, x_i)} = \mathbb{E}[\omega_i \beta_{1,i}]$ with $\omega_i = \delta_{1,i}/\mathbb{E}[\delta_{1,i}]$ no a priori sign

2.3 Local average treatment effects

- Potential outcome framework: $y_i = \Delta_i x_i + \mathbb{E}[y_i(0)] + e_i$ with $x_i \not\perp\!\!\!\perp u_i$ and $z_i \perp\!\!\!\perp u_i$
- No global interpretation: $\hat{\beta}_{1,\mathrm{IV}} \xrightarrow{p} \frac{\mathrm{Cov}(z_i, y_i)}{\mathrm{Cov}(z_i, x_i)} = \mathbb{E}[\omega_i \Delta_i]$, $\omega_i = \frac{x_i(z_i - \mathbb{E}[z_i])}{\mathbb{E}[x_i(z_i - \mathbb{E}[z_i])]}$ no a priori sign
- If IV z_i is binary, we get Wold representation $\hat{\beta}_{1,\mathrm{IV}} \xrightarrow{p} \frac{\mathrm{Cov}(z_i, y_i)}{\mathrm{Cov}(z_i, x_i)} = \frac{\mathbb{E}[y_i \,|\, z_i = 1] - \mathbb{E}[y_i \,|\, z_i = 0]}{\mathbb{E}[x_i \,|\, z_i = 1] - \mathbb{E}[x_i \,|\, z_i = 0]}$
- Assume $x_i = h(z_i, v_i)$, $z_i \perp\!\!\!\perp (u_i, v_i)$ and define $x_i(1) = h(1, v_i)$, $x_i(0) = h(0, v_i)$
- If $x_i(1) \geq x_i(0)$, then $\hat{\beta}_{1,\mathrm{IV}} \xrightarrow{p} \mathbb{E}[\Delta_i \,|\, x_i(1) > x_i(0)] = \mathrm{LATE}$. If $\Delta_i \equiv \Delta$, $\mathrm{LATE} = \Delta$.

2.4 General IV setup

- $y_i = x_i'\beta + e_i$, $x_i \in \mathbb{R}^k$, $z_i \in \mathbb{R}^l$, $\mathbb{E}[e_i \,|\, z_i] = 0$ or $\mathbb{E}[z_i e_i] = 0$, $\mathbb{E}[z_i x_i']$ full rank k, $\mathbb{E}[z_i z_i']$ SPD
- Univariate setup as a special case: $\widetilde{x}_i = (1, x_i)'$, $\widetilde{z}_i = (1, z_i)'$
- Moment equations $\mathbb{E}[z_i y_i] = \mathbb{E}[z_i x_i']\beta$ can only identify β if $l \geq k$ and $\mathbb{E}[z_i x_i']$ full rank k
- Just-identified ($l = k$): $\beta = \mathbb{E}[z_i x_i']^{-1}\mathbb{E}[z_i y_i]$, $\hat{\beta}_{\mathrm{IV}} = \hat{\beta}_{\mathrm{MM}} = \left(\frac{1}{n}\sum_{i=1}^n z_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n z_i y_i$
- Consider the reduced form $x_i = \Gamma' z_i + u_i$ by defining $\Gamma = \mathbb{E}[z_i z_i']^{-1}\mathbb{E}[z_i x_i']$, i.e. $\mathbb{E}[z_i u_i'] = 0$. This yields $y_i = z_i'\Gamma\beta + v_i$ with $v_i = u_i'\beta + e_i$. We get $\mathbb{E}[\Gamma' z_i y_i] = \mathbb{E}[\Gamma' z_i z_i' \Gamma]\beta$.
- $\hat{\beta}_{\mathrm{2SLS}} = \left(\frac{1}{n}\sum_{i=1}^n \hat{\Gamma}' z_i z_i' \hat{\Gamma}\right)^{-1} \frac{1}{n}\sum_{i=1}^n \hat{\Gamma}' z_i y_i = (\hat{X}'\hat{X})^{-1}\hat{X}'Y = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y$ with $\hat{\Gamma} = (Z'Z)^{-1}Z'X$ and $\hat{X} = Z\hat{\Gamma} = P_Z X$. Coincides with $\hat{\beta}_{\mathrm{IV}}$ for $l = k$.
- Biased: $\mathbb{E}[\hat{\beta}_{\mathrm{2SLS}} \,|\, X, Z] = \beta + (X'P_Z X)^{-1} X' P_Z \mathbb{E}[e \,|\, X, Z]$ with $\mathbb{E}[e \,|\, X, Z] \neq 0$
- Consistency: $\hat{\beta}_{\mathrm{2SLS}} \xrightarrow{p} \beta$ by LLN and CMT
- Asymp. normality: $\sqrt{n}(\hat{\beta}_{\mathrm{2SLS}} - \beta) \xrightarrow{d} \mathcal{N}(0, V_\beta(\mathbb{E}[z_i z_i']^{-1}))$ by CLT, LLN, CMT and Slutsky
- Covariance matrix can be consistently est. using sample analogues and $\hat{e}_i = y_i - x_i'\hat{\beta}_{\mathrm{2SLS}}$; $V_\beta(\mathbb{E}[z_i z_i']^{-1}) = Q^{-1}\Omega(Q')^{-1}$ for $l = k$; if homosc. $\mathbb{E}[e_i^2 \,|\, z_i] = \sigma^2$: $\sigma^2\left(Q'\mathbb{E}[z_i z_i']^{-1}Q\right)^{-1}$

3 Panel data

- Structural model: $y_{it} = x_{it}'\beta + u_i + \varepsilon_{it}$, $X_i = (x_{i1}, \ldots, x_{iT})' \in \mathbb{R}^{T\times k}$, $\mathbb{E}[\varepsilon_{it} \,|\, X_i] = 0$

3.1 Random effects

- Key assumption: $\mathbb{E}[u_i \,|\, X_i] = 0$, i.e. $\mathbb{E}[e_{it} \,|\, X_i] = 0$ with $e_{it} = u_i + \varepsilon_{it}$
- Rank assumption: $\mathbb{E}[\sum_{t=1}^T x_{it}x_{it}']$ has full rank
- $\hat\beta_{\text{pool}} = (\sum_i \sum_t x_{it}x_{it}')^{-1} \sum_i \sum_t x_{it}y_{it} = \beta + (\sum_i X_i'X_i)^{-1} \sum_i X_i'e_i$
- Unbiased: $\mathbb{E}[\hat\beta_{\text{pool}} \,|\, X_1, \ldots, X_n] = \beta$, Consistency: $\hat\beta_{\text{pool}} \xrightarrow{p} \beta$
- Asymp. normality: $\sqrt{n}(\hat\beta_{\text{pool}} - \beta) \xrightarrow{d} \mathcal{N}(0, V_{\text{pool}} = \mathbb{E}[X_i'X_i]^{-1}\mathbb{E}[X_i'e_ie_i'X_i]\mathbb{E}[X_i'X_i]^{-1})$
- Further assumptions: $\mathbb{E}[\varepsilon_{is}\varepsilon_{it} \,|\, X_i] = \mathbb{1}\{s = t\}\sigma_\varepsilon^2$, $\mathbb{E}[u_i^2 \,|\, X_i] = \sigma_u^2$, $\mathbb{E}[u_i\varepsilon_{it} \,|\, X_i] = 0$
- Unfeasible GLS: $\hat\beta_{\text{gls}} = (\sum_i X_i'\Omega^{-1}X_i)^{-1} \sum_i X_i'\Omega^{-1}y_i$ with $\Omega(\sigma_u^2, \sigma_\varepsilon^2) \equiv \mathbb{E}[e_ie_i' \,|\, X_i]$
- Minimal variance: $\sqrt{n}(\hat\beta_{\text{gls}} - \beta) \xrightarrow{d} \mathcal{N}(0, V_{\text{gls}} = \mathbb{E}[X_i'\Omega^{-1}X_i]^{-1})$ with $V_{\text{gls}} \le V_{\text{pool}}$
- Feasible GLS $\hat\beta_{\text{re}}$: Plug in $\hat\Omega \xrightarrow{p} \Omega$, similar properties as $\hat\beta_{\text{gls}}$

3.2 Fixed effects

- Key difference: $\mathbb{E}[u_i \,|\, X_i] \neq 0$ or $\mathbb{E}[x_{it}u_i] \neq 0$
- $\hat\beta_{\text{pool}} \xrightarrow{p} \beta + \mathbb{E}[X_i'X_i]^{-1} \sum_t \mathbb{E}[x_{it}u_i]$ inconsistent
- Within trafo: $\bar x_i = \frac{1}{T}\sum_t x_{it}$, $\dot x_{it} = x_{it} - \bar x_i$, $\dot y_{it} = \dot x_{it}'\beta + \dot\varepsilon_{it}$, $\mathbb{E}[\dot\varepsilon_{it} \,|\, X_1, \ldots, X_n] = 0$
- Rank assumption: $\mathbb{E}[\sum_{t=1}^T \dot x_{it}\dot x_{it}']$ has full rank
- $\hat\beta_{\text{fe}} = (\sum_i \sum_t \dot x_{it}\dot x_{it}')^{-1} \sum_i \sum_t \dot x_{it}\dot y_{it} = \beta + (\sum_i \dot X_i'\dot X_i)^{-1} \sum_i \dot X_i'\dot\varepsilon_i$
- Unbiased: $\mathbb{E}[\hat\beta_{\text{fe}} \,|\, X_1, \ldots, X_n] = \beta$, Consistency: $\hat\beta_{\text{fe}} \xrightarrow{p} \beta$
- Asymp. normality: $\sqrt{n}(\hat\beta_{\text{fe}} - \beta) \xrightarrow{d} \mathcal{N}(0, V_{\text{fe}} = \mathbb{E}[\dot X_i'\dot X_i]^{-1}\Omega_T\mathbb{E}[\dot X_i'\dot X_i]^{-1})$
- V_{fe} and $\Omega_T = \mathbb{E}[\dot X_i'\dot\varepsilon_i\dot\varepsilon_i'\dot X_i] = \mathbb{E}[\dot X_i'\varepsilon_i\varepsilon_i'\dot X_i] = \mathbb{E}[\dot X_i'e_ie_i'\dot X_i]$ estimated via sample analogues as well as $\hat e_{it} = y_{it} - x_{it}'\hat\beta_{\text{fe}}$
- First difference: $x_{it}^\Delta = x_{i,t} - x_{i,t-1}$, $y_{it}^\Delta = x_{it}^{\Delta\prime}\beta + \varepsilon_{it}^\Delta$, $\hat\beta_\Delta = (\sum_i \sum_t x_{it}^\Delta x_{it}^{\Delta\prime})^{-1} \sum_i \sum_t x_{it}^\Delta y_{it}^\Delta$
- Unbiased, consistent, asymp. normal due to $\mathbb{E}[\varepsilon_{it}^\Delta \,|\, X_1, \ldots, X_n] = 0$; $\hat\beta_\Delta = \hat\beta_{\text{fe}}$ for $T = 2$

3.3 Two-way fixed effects

- $y_{it} = x_{it}'\beta + u_i + v_t + \varepsilon_{it}$, $\mathbb{E}[\varepsilon_{it} \,|\, X_i] = 0$
- Within trafo: $\tilde x_t = \frac{1}{n}\sum_i x_{it}$, $\bar x = \frac{1}{nT}\sum_i \sum_T x_{it}$, $\ddot x_{it} = x_{it} - \bar x_i - \tilde x_t + \bar x$, $\ddot y_{it} = \ddot x_{it}'\beta + \ddot\varepsilon_{it}$, $\mathbb{E}[\ddot\varepsilon_{it} \,|\, X_1, \ldots, X_n] = 0$
- $\hat\beta_{2\text{WFE}} = (\sum_i \sum_t \ddot x_{it}\ddot x_{it}')^{-1} \sum_i \sum_t \ddot x_{it}\ddot y_{it}$ unbiased, consistent, asymp. normal

3.4 Difference in difference

- $T = 2$, binary treatment d_i received between $t = 1$ and $t = 2$
- Potential outcome framework: $y_{it} = d_iy_{it}(1) + (1 - d_i)y_{it}(0)$
- Want ATT in period 2: $\Delta_2 = \mathbb{E}[y_{i2}(1) - y_{i2}(0) \,|\, d_i = 1]$

- Parallel trends: $\mathbb{E}[y_{i2}(0) - y_{i1}(0) \,|\, d_i = 1] = \mathbb{E}[y_{i2}(0) - y_{i1}(0) \,|\, d_i = 0]$
- No anticipation: $y_{i1}(0) = y_{i1}(1)$
- Identification: $\text{DiD} = \mathbb{E}[y_{i2} - y_{i1} \,|\, d_i = 1] - \mathbb{E}[y_{i2} - y_{i1} \,|\, d_i = 0] = \Delta_2$
- $\hat{\Delta}_2 = (\bar{y}_{t=2,d=1} - \bar{y}_{t=1,d=1}) - (\bar{y}_{t=2,d=0} - \bar{y}_{t=1,d=0}) = \frac{\sum_i d_i(y_{i2}-y_{i1})}{\sum_i d_i} - \frac{\sum_i (1-d_i)(y_{i2}-y_{i1})}{\sum_i 1-d_i}$
- Assume structural model $y_{it} = d_{it}\beta + u_i + v_t + \varepsilon_{it}$, $d_{i1} = 0$, $d_{i2} = d_i$. (Then $\Delta_2 = \beta$.)
- One can actually show that $\hat{\Delta}_2 = \hat{\beta}_{2\text{WFE}}$. But imagine we don't know this ...
- If $\mathbb{E}[\varepsilon_{it} \,|\, d_i] = 0$, then $\hat{\beta}_{2\text{WFE}} \xrightarrow{p} \beta$. If additionally $\mathbb{E}[v_2 - v_1 \,|\, d_i = 1] = \mathbb{E}[v_2 - v_1 \,|\, d_i = 0]$ (parallel trends), then $\Delta_2 = \text{DiD} = \beta$, i.e. coeff. on d_{it} in 2WFE id. ATT in period 2.

4 Limited dependent variables

4.1 Binary outcome

- Quantity of interest: $\mathbb{E}[y_i \,|\, x_i] = \mathbb{P}(y_i = 1 \,|\, x_i) = p(x_i)$
- Underlying model: $y_i = p(x_i) + e_i$, $\mathbb{E}[e_i \,|\, x_i] = 0$, $\text{Var}(e_i \,|\, x_i) = p(x_i)(1 - p(x_i))$ heterosc.
- LPM: $\mathbb{P}(y_i = 1 \,|\, x_i) = x_i'\beta$. $+$: estimation via OLS, β marginal effect; $-$: misspecified

4.2 Probit and logit model

- $\mathbb{P}(y_i = 1 \,|\, x_i) = F(x_i'\beta)$ for $F\colon \mathbb{R} \to [0,1]$, strictly increasing, $F(u) = 1 - F(-u)$
- Probit: $F = \Phi$, Logit: $F(u) = \frac{1}{1+\exp(-u)}$
- Underlying model: $y_i = \mathbb{1}\{x_i'\beta + u_i \geq 0\}$ with $u_i \sim F$, $u_i \perp\!\!\!\perp x_i$
- MLE: $\mathcal{L}(\theta) = \prod_i f(x_i, \theta)$, $\hat{\theta}_{\text{ML}} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$; consistent, asymp. normal, efficient
- $f_{y|x}(y_i \,|\, x_i) = F(x_i'\beta)^{y_i}(1 - F(x_i'\beta))^{1-y_i}$, $l(\beta) = \sum_i y_i \log F(x_i'\beta) + (1 - y_i)\log(1 - F(x_i'\beta))$
- Of interest: $F(x_i'\hat{\beta}_{\text{ML}})$; $\text{ME} = \frac{\partial}{\partial x}\mathbb{P}(y_i = 1 \,|\, x_i = x) = f(x'\beta)\beta$, $\widehat{\text{AME}} = \frac{1}{n}\sum_i f(x_i'\hat{\beta}_{\text{ML}})\hat{\beta}_{\text{ML}}$
- Potential outcome framework (parametric): $y_i = g(d_i, w_i, u_i) = \mathbb{1}\{d_i\beta_1 + w_i'\beta_{-1} + u_i \geq 0\}$, $u_i \sim F$, $u_i \perp\!\!\!\perp (d_i, w_i')'$, $\mathbb{P}(y_i = 1 \,|\, d_i, w_i) = F((d_i, w_i')\beta)$. Under unconfoundedness, $\text{ATE} = \mathbb{E}[F(\beta_1 + w_i'\beta_{-1}) - F(w_i'\beta_{-1})]$, can be estimated using sample mean and $\hat{\beta}_{\text{ML}}$.
- Normalization: $y = \mathbb{1}\{\tilde{\beta}_1 + \tilde{\beta}_2 x + \tilde{u} \geq 0\} = \mathbb{1}\{\frac{\tilde{\beta}_1+\mu}{\sigma} + \frac{\tilde{\beta}_2}{\sigma}x + \frac{\tilde{u}-\mu}{\sigma} \geq 0\}$ with $\tilde{u} \sim \mathcal{N}(\mu, \sigma^2)$, $\frac{\tilde{u}-\mu}{\sigma} \sim \mathcal{N}(0,1)$ give the same observational data. Thus μ and σ^2 are non-identifiable. Normalizing them fixes the scale of β (no interpretive meaning). But (estimated) conditional choice probabilities and marginal effects remain invariant.
- Random utility model: $v_{ij} = x_i'\beta_j + u_{ij}$, $y_i = \mathbb{1}\{v_{i1} > v_{i0}\}$, $x_i \perp\!\!\!\perp (u_{i1}, u_{i0})$, $\mathbb{P}(y_i = 1 \,|\, x_i) = F_{u_{i0}-u_{i1}}(x_i'(\beta_1 - \beta_0))$
- If $(u_{i0}, u_{i1})' \sim \mathcal{N}(\mu, \Sigma)$ with $\mu_0 - \mu_1 = 0$, $\sigma_0^2 + \sigma_1^2 - 2\sigma_{01} = 1$, then $u_{i0} - u_{i1} \sim \mathcal{N}(0,1)$ and we get probit model for $\beta_1 - \beta_0$
- If u_{i0}, u_{i1} independent EVT1, i.e. $F(u) = \exp(-\exp(-u))$, then $u_{i0} - u_{i1}$ standard logistic

4.3 Multinomial logit

- Latent utility model: $v_{ij} = x_i'\beta_j + u_{ij}$, $y_i = \arg\max_j v_{ij}$, $x_i \perp\!\!\!\perp (u_{i1}, \ldots, u_{iJ})$
- Normalization: Set $\beta_J = 0$ and interpret other coefficients as differences. Fix scale of u_{ij}, but then scale of β_j has no interpretive meaning.

- If u_{ij} iid. EVT1, then $\mathbb{P}(y_i = j \mid x_i) = \frac{\exp(x_i'\beta_j)}{\sum_l \exp(x_i'\beta_l)}$. For $J = 2$, we get binary logit.
- $\text{ME}_j = \frac{\partial}{\partial x}\mathbb{P}(y_i = j \mid x_i = x) = \mathbb{P}(y_i = j \mid x_i = x)(\beta_j - \sum_l \beta_l \mathbb{P}(y_i = l \mid x_i = x))$ not same sign as β_j, $\widehat{\text{AME}}_j = \frac{1}{n}\sum_i \hat{P}_j(x_i)(\hat{\beta}_j - \sum_l \hat{\beta}_l \hat{P}_l(x_i))$
- Conditional logit: $v_{ij} = w_i'\beta_j + x_{ij}'\gamma + u_{ij}$, $y_i = \arg\max_j v_{ij}$, $(w_i, X_i) \perp\!\!\!\perp (u_{i1}, \ldots, u_{iJ})$; γ and $\beta_j - \beta_l$ identified up to scale; if u_{ij} iid. EVT1, then $\mathbb{P}(y_i = j \mid w_i, X_i) = \frac{\exp(w_i'\beta_j + x_{ij}'\gamma)}{\sum_l \exp(w_i'\beta_l + x_{il}'\gamma)}$
- MLE: $l(\beta_1, \ldots, \beta_{J-1}, \gamma) = \sum_i \sum_j \mathbb{1}\{y_i = j\} \log \mathbb{P}(y_i = j \mid w_i, X_i)$
- IIA: strange substitution patterns

4.4 Censored data

- $y_i^* = x_i'\beta + u_i$, $u_i \perp\!\!\!\perp x_i$, $u_i \sim \mathcal{N}(0, \sigma^2)$. Observe $y_i = y_i^*$ if $y_i^* > 0$.
- Two-step procedure: 1. Use MLE to estimate $\frac{\beta}{\sigma}$ in probit model $\mathbb{P}(y_i > 0 \mid x_i) = \Phi(\frac{x_i'\beta}{\sigma})$; 2. $\mathbb{E}[y_i \mid y_i > 0, x_i] = x_i'\beta + \sigma\lambda(\frac{x_i'\beta}{\sigma})$ since $\mathbb{E}[z \mid z > a] = \lambda(-a)$ for $z \sim \mathcal{N}(0,1)$, i.e. regress uncensored y_i on x_i and $\lambda(x_i'(\widehat{\frac{\beta}{\sigma}}))$ to get $\hat{\beta}$ and $\hat{\sigma}$
- $\mathcal{L}(\beta, \sigma) = \prod_i \mathbb{P}(y_i = 0 \mid x_i)^{\mathbb{1}\{y_i=0\}} (\frac{\partial}{\partial y}\mathbb{P}(y_i \leq y \mid x_i)|_{y=y_i})^{\mathbb{1}\{y_i>0\}} = \prod_i \Phi(-\frac{x_i'\beta}{\sigma})^{\mathbb{1}\{y_i=0\}} (\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma}))^{\mathbb{1}\{y_i>0\}}$

4.5 Sample selection

- Potential wage: $y_i = x_i'\beta + u_i$, reservation wage: $r_i^* = x_i'\delta + v_i$. Observe y_i if $y_i > r_i^*$.
- Outcome eq: $y_i = x_i'\beta + u_i$, selection eq: $T_i = \mathbb{1}\{z_i'\gamma \geq w_i\}$. Observe y_i if $T_i = 1$.
- $(u_i, w_i) \perp\!\!\!\perp (x_i, z_i)$, $\begin{pmatrix} u_i \\ w_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{uw} & 1 \end{pmatrix}\right)$, scale of γ no interpretation
- Heckman two-stage: 1. Use MLE to estimate γ in probit model $\mathbb{P}(T_i = 1 \mid z_i) = \Phi(z_i'\gamma)$; 2. $\mathbb{E}[y_i \mid T_i = 1, x_i = x, z_i = z] = x'\beta + \mathbb{E}[u_i \mid z'\gamma \geq w_i] = x'\beta - \sigma_{uw}\lambda(z'\gamma)$, i.e. regress obs. y_i on x_i and $\lambda(z_i'\hat{\gamma})$ to get $\hat{\beta}$ and $-\hat{\sigma}_{uw}$

5 Generalized method of moments

5.1 General GMM setup

- Moment conditions $\mathbb{E}[g_i(\theta_0)] = 0$ for a function $g_i \colon \mathbb{R}^k \to \mathbb{R}^l$ of observed data
- $G \colon \mathbb{R}^k \to \mathbb{R}^l$, $G(\theta) = \mathbb{E}[g_i(\theta)]$, $W \in \mathbb{R}^{l \times l}$ SPD, $\theta_0 = \arg\min_\theta G(\theta)'WG(\theta)$
- $\hat{\theta}_{\text{GMM}} = \arg\min_\theta \bar{g}_n(\theta)'\hat{W}\bar{g}_n(\theta)$ with $\bar{g}_n(\theta) = \frac{1}{n}\sum_i g_i(\theta)$ and $\hat{W} \xrightarrow{p} W$
- Method of moments: $\bar{g}_n(\hat{\theta}_{\text{MM}}) = 0$; if $\hat{\theta}_{\text{MM}}$ exists (uniquely), then $\hat{\theta}_{\text{GMM}} = \hat{\theta}_{\text{MM}}$

5.2 IV example

- $g_i(\theta) = z_i(y_i - x_i'\theta)$
- FOC: $\frac{\partial}{\partial\theta}(\bar{g}_n(\theta)'\hat{W}\bar{g}_n(\theta))|_{\theta=\hat{\theta}} = 2(\frac{\partial}{\partial\theta}\bar{g}_n(\theta))'\hat{W}\bar{g}_n(\theta)|_{\theta=\hat{\theta}} = -2(\frac{1}{n}\sum_i x_i z_i')\hat{W}\frac{1}{n}\sum_i z_i(y_i - x_i'\hat{\theta}) = 0$
- $\hat{\theta} = \left((\frac{1}{n}\sum_i x_i z_i')\hat{W}(\frac{1}{n}\sum_i z_i x_i')\right)^{-1}(\frac{1}{n}\sum_i x_i z_i')\hat{W}(\frac{1}{n}\sum_i z_i y_i)$
- Consistency: $\hat{\theta} \xrightarrow{p} \theta_0$, Asymptotic normality: $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_{\theta_0}(W))$
- $Q = \mathbb{E}[-\frac{\partial}{\partial\theta}g_i(\theta_0)] = \mathbb{E}[z_i x_i']$, $\Omega = \mathbb{E}[g_i(\theta_0)g_i(\theta_0)'] = \mathbb{E}[e_i^2 z_i z_i']$, $V_{\theta_0}(W) = (Q'WQ)^{-1}Q'W\Omega WQ(Q'WQ)^{-1}$

- For $l = k$: $\hat{\theta}_{\text{GMM}} = \hat{\theta}_{\text{MM}} = \hat{\theta}_{\text{IV}}$ independent of weight matrix
- Popular for $l > k$: $\hat{W} = \left(\frac{1}{n}\sum_i z_i z_i'\right)^{-1}$, giving $\hat{\theta}_{\text{2SLS}}$ (1 step)
- Optimal/efficient: $W_{\text{opt}} = \Omega^{-1}$, $V_{\theta_0}(\Omega^{-1}) = \left(Q'\Omega^{-1}Q\right)^{-1}$
- If $\mathbb{E}[e_i^2 \mid z_i] = \sigma^2$, then $W_{\text{opt}} = \sigma^{-2}\mathbb{E}[z_i z_i']^{-1}$. Thus, $\hat{\theta}_{\text{2SLS}}$ is only efficient in homosc. case.
- Two-step feasible efficient estimation: 1. Compute $\hat{\theta}(\hat{W}_1)$ for some feasible $\hat{W}_1$; 2. Utilize $\hat{\Omega} = \frac{1}{n}\sum_i g_i(\hat{\theta}(\hat{W}_1))g_i(\hat{\theta}(\hat{W}_1))' = \frac{1}{n}\sum_i \hat{e}_i^2 z_i z_i' \xrightarrow{p} \Omega$, with $\hat{e}_i = y_i - x_i'\hat{\theta}(\hat{W}_1)$, to get $\hat{\theta}(\hat{\Omega}^{-1})$.
- Consistent variance estimation: Use $\hat{W}$ and sample analogues $\hat{Q} = -\frac{\partial}{\partial\theta}\bar{g}_n(\hat{\theta}) = \frac{1}{n}\sum_i z_i x_i'$ and $\hat{\Omega} = \frac{1}{n}\sum_i g_i(\hat{\theta})g_i(\hat{\theta})' = \frac{1}{n}\sum_i \hat{e}_i^2 z_i z_i'$

5.3 Testing

- Wald test: If $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0, V_{\theta_0})$ with $\hat{V}_{\theta_0} \xrightarrow{p} V_{\theta_0}$, then $\hat{V}_{\theta_0}^{-1/2}\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0, I_k)$ and $n(\hat{\theta}-\theta_0)'\hat{V}_{\theta_0}^{-1}(\hat{\theta}-\theta_0) \xrightarrow{d} \chi_k^2$. We can test hypotheses and construct CI. To test $H_0\colon \theta_0 = \tilde{\theta}$ against $H_1\colon \theta_0 \neq \tilde{\theta}$, compute $W = n(\hat{\theta}-\tilde{\theta})'\hat{V}_{\theta_0}^{-1}(\hat{\theta}-\tilde{\theta}) \xrightarrow{H_0} \chi_k^2$ and reject H_0 if $W > c_{1-\alpha}$.
- Specification test: Test if model is correctly specified (IVs are valid). $H_0\colon \mathbb{E}[g_i(\theta_0)] = 0$ for some θ_0. $J = \min_\theta n\bar{g}_n(\theta)'\hat{\Omega}^{-1}\bar{g}_n(\theta) \xrightarrow{H_0} \chi_{l-k}^2$, reject H_0 if $J > c_{1-\alpha}$. Just-identified IV: H_0 not testable since $\bar{\theta} = \mathbb{E}[z_i x_i']^{-1}\mathbb{E}[z_i y_i]$ always satisfies moment equations, even if IVs invalid for $\theta_0 \neq \bar{\theta}$. Over-identification test possible for $l > k$.