L1 Neyman-Rubin Model

Unit: 人 or 一群物体的集合 (classroom, market)

action: 操作

Treatment (Active) ⟷ Control

causal effect: $Y_t - Y_c$ or Y_t / Y_c. Y_t, Y_c: potential outcome

Unit level: $\tau_i = Y_i(t) - Y_i(c) = Y_i(1) - Y_i(0)$

Average ce over finite sample $\tau_{fs} = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))$

SUTVA:

① No interference: $Y_i(\vec{t}) = Y_i(t_i)$ 潜在结果不受其它 unit 的 treatment 影响.

② No Hidden Variations of Treatment:

$Y_i(\vec{t}) = Y_i(t_i) \in \{Y_i(t), Y_i(c)\}$.

fs vs sp: finite units $\{1, 2, \cdots N\}$

$(Y_i(0), Y_i(1), W_i, X_i)$ follow (ϕ, θ, ϕ)

$f_{W|Y(0),Y(1),X}(W_i | Y_i(0), Y_i(1), X_i, \phi) \cdot f_{Y(0),Y(1)|X}(Y_i(0), Y_i(1) | X_i, \theta) \cdot f_X(X_i | \phi)$

$\tau_{sp} = E(Y_i(1) - Y_i(0))$

sp下 propensity score:

$e(x) = E_Y[E_W(W_i | X_i = x, Y(0), Y(1))] = E_{sp}(W_i | X_i = x)$

$= Pr(W_i = 1 | X_i = x, \phi, \theta)$.

两重随机性 ① 从人群中重复抽样 ② W_i 随机分配.

$CI^{1-\alpha}(\tau_{fs}) = \left(\hat{\tau}^{dif} - z_{\alpha/2}\sqrt{\hat{V}},\ \hat{\tau}^{dif} + z_{\alpha/2}\sqrt{\hat{V}}\right)$

区间估计.

$\hat{V}^{Neyman} = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$

$\hat{V}^{const} = s^2\left(\frac{1}{N_c} + \frac{1}{N_t}\right)$

$\hat{V}^{\rho_{tc}=1} = S_c^2 \cdot \frac{N_t}{N \cdot N_c} + S_t^2 \cdot \frac{N_c}{N \cdot N_t} + S_t S_c \cdot \frac{2}{N}$

Neyman方法.

t检验: $H_0: \frac{1}{N}\sum(Y_i(1) - Y_i(0)) = 0$ 检验量 $t = \frac{\bar{Y}_t^{obs} - \bar{Y}_c^{obs}}{\sqrt{\hat{V}^{Neyman}}}$ 平均作用.

L2 Assignment Mechanism.

Def: $W \in \{0,1\}^N$ 且满足 $\sum_{W \in \{0,1\}^N} Pr(W | X, Y(0), Y(1)) = 1$ for all $X, Y(0), Y(1)$.

Unit level: $E(W_i | X, Y(0), Y(1)) = p_i(X, Y(0), Y(1)) = \sum_{W: W_i = 1} Pr(W | X, Y(0), Y(1))$

Propensity Score: $e(x) = \frac{1}{N(x)}\sum_{i: X_i = x} p_i(X, Y(0), Y(1))$

$N(x) = \#\{i = 1, 2, \cdots N | X_i = x\}$ the number of units with $X_i = x$

$N(x) = 0$ 时 propensity score 也为 0. e.g.

$$\begin{array}{c|cc} & b & w \\ \hline 1 & a & b \\ 0 & c & d \end{array} \qquad b: \frac{a}{a+c} \quad w: \frac{b}{b+d}.$$

propensity score: 协变量均衡 → 避免 Simpson

Assumption:

① Individual assignment

$Pr(W | X, Y(0), Y(1)) = c \cdot \prod_{i=1}^{N} q(X_i, Y_i(0), Y_i(1))^{W_i}(1 - q(X_i, Y_i(0), Y_i(1)))^{1-W_i}$

$p_i(X, Y(0), Y(1)) = q(X_i, Y_i(0), Y_i(1))$ 类似于独立. $\underline{e(x) = q(x)}$

② Probabilistic assignment

$0 < p_i(X, Y(0), Y(1)) < 1$, 不能取等.

③ Unconfounded assignment.

$Pr(W | X, Y(0), Y(1)) = Pr(W | X)$ 分配机制与潜在结果无关.

Classification:

Randomized Experiment: Bernoulli, CRE, SRE, PRE

Observation: RAM.

若每个个体因果作用一致 $V_W(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$ 放大了.

令 $S_{tc}^2 = S_c^2 + S_t^2 - 2\rho_{tc} S_t S_c$ 则 $V_W(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}) = \frac{N_t}{N \cdot N_c}S_c^2 + \frac{N_c}{N \cdot N_t}S_t^2 + \frac{2}{N}\rho_{tc} S_t S_c$

其中 $\rho_{tc} = \frac{1}{(N-1)S_t S_c}\sum_{i=1}^{N}(Y_i(1) - \bar{Y}(1))(Y_i(0) - \bar{Y}(0))$ 正相关反而方差大.

$\rho_{tc} = 1$ $V_W = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{(S_c - S_t)^2}{N}$ $\rho_{tc} = 0$: $V_W = \frac{N_t}{N \cdot N_c}S_c^2 + \frac{N_c}{N \cdot N_t}S_t^2 - \frac{2}{N}S_c S_t$

$\hat{V}^{Neyman} = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} \xRightarrow{S_c^2 = S_t^2} \hat{V} = S^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t}\right)$ 在CR中的方差 基于const假设

$\hat{V}^{\rho_{tc}} = S_c^2 \cdot \frac{N_t}{N \cdot N_c} + S_t^2 \frac{N_c}{N \cdot N_t} + \rho_{tc} S_c \cdot S_t \cdot \frac{2}{N}$

$\hat{V}^{\rho_{tc}=1} = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{(S_c - S_t)^2}{N}$ 保守 无假设时更友好.

L3 FEP & Neyman.

其它统计量 $T = |S_t^2 - S_c^2|$ e.g.

fisher exact p-value:

① null hyp:

$H_0: Y_i(1) = Y_i(0) + C \quad i = 1, 2, \cdots N$

$H_0: Y_i(1) / Y_i(0) = C \quad i = 1, \cdots N$

$H_0: Y_i(1) = Y_i(0) + C_i \quad i = 1, \cdots N$

② choose T

$T(W, Y^{obs}, X)$

$\underline{\underline{T^{dif}}} = |\bar{Y}_t^{obs} - \bar{Y}_c^{obs}|$

$= \left|\frac{\sum_{i: W_i = 1}^{N} Y_i^{obs}}{N_t} - \frac{\sum_{i: W_i = 0}^{N} Y_i^{obs}}{N_c}\right|$

$\underline{\underline{T^{rank}}} = |\bar{R}_t - \bar{R}_c| = \left|\frac{\sum_{i: W_i = 1} R_i}{N_t} - \frac{\sum_{i: W_i = 0} R_i}{N_c}\right|$

R从小到大排, 并列时取均值 (1, 2, 2.5)

Model方法: 适用范围窄但功效高.

Rank方法: 功效高, 鲁棒性好.

③ measure p-value

$\hat{p} = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}_{T^{dif,k} \geq T^{dif,obs}}$

K为 $\begin{cases} \text{模拟次数} - \text{simulation} \\ O\binom{N_t + N_c}{N_t} - \text{exact} \end{cases}$

$\underbrace{\sqrt{\hat{p}^*(1-\hat{p}^*)/K}}_{\hat{s.e.}} \leq \frac{1}{2\sqrt{K}}$ $\hat{s.e.}$ 用来衡量 $\hat{p}$ 估计的准确度

Gain score: $Y_i'(w) = Y_i(w) - X_i$

Neyman Approach.

Focus on Average, unbiased point & interval estimators.

$\hat{\tau}^{dif} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$. 无偏性: $E_W[\hat{\tau}^{dif} | Y(0), Y(1)] = \frac{1}{N}\sum\left[\frac{E(W_i) Y_i(1)}{N_t/N} - \frac{(1 - E_W(W_i)) Y_i(0)}{N_c/N}\right]$

$= \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) = \tau_{fs}$.

sp下unbiased: $E(\hat{\tau}^{dif}) = E_Y(E_W[\hat{\tau}^{dif} | \vec{Y}]) = E_Y(\tau_{fs}) = \tau_{sp}$.

$E_{sp}(\tau_{fs}) = E_{sp}[\bar{Y}(1) - \bar{Y}(0)] = \frac{1}{N}\sum_{i=1}^{N}[Y_i(1) - Y_i(0)] = \tau_{sp}$ τ_{fs} 是 τ_{sp} 的无偏!

方差:

$V_W(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}$ $S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i(0) - \bar{Y}(0))^2$ $S_t^2 = \frac{1}{N-1}\sum(Y_i(1) - \bar{Y}(1))^2$

$S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i(1) - Y_i(0) - \tau_{fs})^2$

pf: $\hat{\tau}^{dif} = \frac{1}{N}\sum\left[\frac{N}{N_t}\left(D_i + \frac{N_t}{N}\right)Y_i(1) - \frac{N}{N_c}\left(\frac{N_c}{N} - D_i\right) \cdot Y_i(0)\right]$

$= \tau_{fs} + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{N}{N_t} \cdot Y_i(1) + \frac{N}{N_c} \cdot Y_i(0)\right) \cdots \to Y_i^+$ $D_i = W_i - \frac{N_t}{N} = \begin{cases} \frac{N_c}{N} & W_i = 1 \\ -\frac{N_t}{N} & W_i = 0 \end{cases}$

$E_W(D) = 0$

$V(D_i) = E(D_i^2) = \frac{N_c N_t}{N^2}$

$V_W(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}) = \frac{1}{N^2}E_W\left[\left(\sum_{i=1}^{N}(D_i \cdot Y_i^+)\right)^2\right]$ $E_W(D_i D_j) = \begin{cases} \frac{N_t N_c}{N^2} & i = j \\ -\frac{N_t \cdot N_c}{N^2(N-1)} & i \neq j \end{cases}$

$= \frac{1}{N^2}\left(\frac{N_t N_c}{N^2}\underbrace{\sum_{i=1}^{N} Y_i^{+2}}_{S_t^2 + S_c^2} - \frac{N_t N_c}{N^2(N-1)}\underbrace{\sum_{i=1}^{N}\sum_{j=1, i \neq j}^{N}(Y_i^+ Y_j^+)}_{S_{tc}^2}\right)$

sp下 Neyman 方差.

$$V(\hat\tau^{dif}) = E[(\bar Y_t^{obs} - \bar Y_c^{obs} - E(\bar Y_t^{obs} - \bar Y_c^{obs}))^2]$$

$$= E[(\bar Y_t^{obs} - \bar Y_c^{obs} - E_{sp}[\bar Y(1) - \bar Y(0)])^2]$$

$$= E[((\bar Y_t^{obs} - \bar Y_c^{obs}) - (\bar Y(1) - \bar Y(0)))^2] \rightarrow \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t} - \frac{\sigma_{tc}^2}{N}$$

$$+ E_{sp}[(\bar Y(1) - \bar Y(0)) - E_{sp}[\bar Y(1) - \bar Y(0)])^2] \rightarrow V_{sp}(\tau_{fs}) = \sigma_{tc}^2/N$$

$$+ 2 \cdot E[(\bar Y_t^{obs} - \bar Y_c^{obs} - (\bar Y(1) - \bar Y(0)))((\bar Y(1) - \bar Y(0)) - E_{sp}(\bar Y(1) - \bar Y(0)))] \Rightarrow E[(\hat\tau^{dif} - \tau_{fs})(\tau_{fs} - \tau_{sp})] = E_Y[E_W((\hat\tau^{dif} - \tau_{fs}) \mid Y)(\tau_{fs} - \tau_{sp})] = 0$$

$$\hat V_{sp} = V(\hat\tau^{dif}) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t}$$ 在 sp 下 Neyman 方法无偏但 $\rho_{tc}=1$ 偏紧.

L_4 Regression

$\tau_{sp} = u_t - u_c = E_{sp}[Y_i(1) - Y_i(0)]$

model: $Y_i^{obs} = \alpha + \tau \cdot W_i + \varepsilon_i \rightarrow \varepsilon_i = \begin{cases} Y_i(0) - u_c & \text{if } W_i = 0 \\ Y_i(1) - u_t & \text{if } W_i = 1 \end{cases}$

$\alpha \downarrow u_c = E_{sp}[Y_i(0)]$

$V_{sp}[\varepsilon_i \mid W_i] = \sigma_t^2 W_i + \sigma_c^2(1 - W_i)$ if $\sigma_t^2 = \sigma_c^2 = \sigma^2$ $V_{sp}[\varepsilon_i \mid W_i] = \sigma^2$

$$\begin{cases} V_{sp}(\varepsilon_i) = E(V(\varepsilon_i \mid W)) + V(\underbrace{E(\varepsilon_i \mid W)}_{0}) \\ \operatorname{cov}(\varepsilon_i, X_i) = E(\varepsilon_i W_i) - E\varepsilon_i \, E W_i = 0 \end{cases} \Rightarrow \text{standard linear model.}$$

$E(\varepsilon_i W_i) = E(E(\varepsilon_i \mid W_i) W_i) = 0$, $E\varepsilon_i = 0$

$\sigma_t^2 = \sigma_c^2 = \sigma^2$ 时 $Y^{obs} = X\beta + \varepsilon$ $\beta = (\alpha, \tau)^T$ $X = \begin{pmatrix} 1_{N_c} & 0_{N_c} \\ 1_{N_t} & 1_{N_t} \end{pmatrix}$

$E(\varepsilon \mid X) = E(\varepsilon) = 0$ 且 $V_{sp}(\varepsilon \mid X) = \Sigma = \operatorname{diag}(\sigma^2 \cdots \sigma^2)$

$\hat\beta = (X^T X)^{-1} X^T Y^{obs} = \beta + H\varepsilon$ $\rightarrow$ CLT 正态 $H = (X^T X)^{-1} X^T = \begin{pmatrix} \frac{1_{N_c}}{N_c} & \frac{-1_{N_c}}{N_c} \\ \frac{0_{N_t}}{N_t} & \frac{1_{N_t}}{N_t} \end{pmatrix}^T$

$\operatorname{Var}(\hat\beta) = H \Sigma H^T \underset{(2,2)元}{\Rightarrow} \hat{V}ar(\hat\tau^{ols}) = \sigma^2\left(\frac{1}{N_c} + \frac{1}{N_t}\right)$ & $\hat\sigma^2 = \frac{1}{N-2}\sum \hat\varepsilon_i^2$

$\therefore \hat{V}ar(\hat\tau^{ols}) = \hat\sigma^2\left(\frac{1}{N_c} + \frac{1}{N_t}\right) \Longleftrightarrow \hat V^{const} = s^2\left(\frac{1}{N_c} + \frac{1}{N_t}\right)$

$\sigma_t^2 \neq \sigma_c^2$ 时 $E(\varepsilon \mid X) = 0$ $V_{sp}(\varepsilon \mid X) = \Sigma = \operatorname{diag}(\sigma_c^2 \cdot 1_{N_c}, \sigma_t^2 \cdot 1_{N_t})$

$\operatorname{Var}(\hat\beta^{ols}) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t}$ & $\hat\sigma_t^2 = \frac{1}{N_t - 1}\sum \hat\varepsilon_i^2 W_i$, $\hat\sigma_c^2 = \frac{1}{N_c - 1}\sum \hat\varepsilon_i^2 (1 - W_i)$

$\hat{V}ar(\hat\tau^{ols}) = \frac{\hat\sigma_c^2}{N_c} + \frac{\hat\sigma_t^2}{N_t} \Longleftrightarrow \hat V^{Neyman} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$

均值无偏性证明: $\hat\tau^{ols} = \frac{\sum_{i=1}^N (W_i - \bar W) \cdot (Y_i^{obs} - \bar Y^{obs})}{\sum_{i=1}^N (W_i - \bar W)^2}$ $\rightarrow \sum_{i=1}^N W_i Y_i^{obs} - N\bar W \bar Y$; $\rightarrow \sum_{i=1}^N W_i - N\bar W = \frac{N_t N_c}{N}$

$$= \frac{N_t(\bar Y_t^{obs} - \bar Y)}{\frac{N_t N_c}{N}} = \bar Y_t^{obs} - \bar Y_c^{obs}$$

$\hat\alpha^{ols} = \bar Y^{obs} - \hat\tau^{ols} \cdot \bar W$

$\bar Y^{obs} = \frac{1}{N}\sum Y_i^{obs}$ $\bar W = \frac{1}{N}\sum_{i=1}^N W_i = \frac{N_t}{N}$.

有协变量: $Y_i^{obs} = \alpha + \tau \cdot W_i + X_i\beta + \varepsilon_i$ $(\hat\tau^{ols}, \hat\alpha^{ols}, \hat\beta^{ols}) = \underset{\tau, \alpha, \beta}{\arg\min} (Y_i^{obs} - \alpha - \tau W_i - X_i\beta)^2$

更精确 $\operatorname{Var}(Y \mid W) = \operatorname{Var}(Y \mid W, X) + \operatorname{Var}(E(Y \mid W)) \geq \operatorname{Var}(Y \mid W, X)$

L_5 Model Based

3 inputs $P(W \mid Y(0), Y(1))$ $f(Y(0), Y(1) \mid \theta)$ $p(\theta)$

step 0 $f(Y(0), Y(1), W \mid \theta) = P_r(W \mid Y(0), Y(1), \theta) \cdot f(Y(0), Y(1) \mid \theta)$

$$f(Y(0), Y(1) \mid W, \theta) = \frac{f(Y(0), Y(1), W \mid \theta)}{\int f(Y(0), Y(1), W \mid \theta)\, dY(0)\, dY(1)}$$

$\downarrow (Y^{mis}, Y^{obs}) = g(Y(0), Y(1), W)$

$f(Y^{mis}, Y^{obs} \mid W, \theta)$

step 1: $$f(Y^{mis} \mid Y^{obs}, W, \theta) = \frac{f(Y^{mis}, Y^{obs} \mid W, \theta)}{f(Y^{obs} \mid W, \theta)} = \frac{f(Y^{mis}, Y^{obs} \mid W, \theta)}{\int_{Y^{mis}} f(Y^{mis}, Y^{obs} \mid W, \theta)\, dY^{mis}}$$

step 2: $$P(\theta \mid Y^{obs}, W) = \frac{P(\theta) f(Y^{obs}, W \mid \theta)}{f(Y^{obs}, W)} = \frac{P(\theta) \int_{Y^{mis}} f(Y^{obs}, Y^{mis}, W \mid \theta)\, dY^{mis}}{\int_\theta P(\theta) \cdot f(Y^{obs}, W \mid \theta)\, d\theta}$$

step 3: $$\underbrace{f(Y^{mis} \mid Y^{obs}, W)}_{\text{target!}} = \int_\theta f(Y^{mis}, \theta \mid Y^{obs}, W)\, d\theta = \int_\theta f(Y^{mis} \mid Y^{obs}, W, \theta) f(\theta \mid Y^{obs}, W)\, d\theta$$

一些 trick: step②中 Normalize 项未必一定要算出来, step③积分也是, 看关键项求分布即可. 考试大概率正态.

L_6 SRE & PRE. SRE:

适用场景: CRE not feasible. more precise.

问题: sample size 过小, 不关心组内效果.

Neyman 方法:

$$E_W(\hat\tau^{strat}) = \frac{N(f)}{N(f) + N(m)} \tau_{fs}(f) + \frac{N(m)}{N(f) + N(m)} \tau_{fs}(m)$$

$$V_W(\hat\tau^{strat}) = \left(\frac{N(f)}{N(f) + N(m)}\right)^2 \left(\frac{S_c^2(f)}{N_c(f)} + \frac{S_t^2(f)}{N_t(f)} - \frac{S_{ct}^2(f)}{N(f)}\right) + \left(\frac{N(m)}{N(f) + N(m)}\right)^2 \left(\frac{S_c^2(m)}{N_c(m)} + \frac{S_t^2(m)}{N_t(m)} - \frac{S_{ct}^2(m)}{N(m)}\right)$$

估计有问题.

令 $q = \frac{N_f}{N}$ $\hat\tau^{strat} = q\hat\tau^{dif}(f) + (1 - q)\hat\tau^{dif}(m)$

$p(c) = \frac{N_t(c)}{N(c)}$ $V_{sp}(\hat\tau^{strat}) = \frac{q}{N}\left(\frac{\sigma_t^2(f)}{p(f)} + \frac{\sigma_c^2(f)}{1 - p(f)}\right) + \frac{1 - q}{N} \cdot \left(\frac{\sigma_t^2(m)}{p(m)} + \frac{\sigma_c^2(m)}{1 - p(m)}\right)$

当 $p(m) = p(f) = p = \frac{N_t}{N}$ 时

$$N \cdot (V_{sp}(\hat\tau^{dif}) - V_{sp}(\hat\tau^{strat})) = q(1 - q) \cdot \left(\frac{(u_c(f) - u_c(m))^2}{1 - p} + \frac{(u_t(f) - u_t(m))^2}{p}\right) \geq 0$$

说明 SRE 更有效. pf: $\sigma_c^2 = \operatorname{Var}(Y_i(0))$

$= E(\operatorname{Var}(Y_i(0) \mid X)) + \operatorname{Var}(E(Y_i(0) \mid X))$

$= q\sigma_c^2(f) + (1 - q)\sigma_c^2(m) + q(1 - q)(u_c(m) - u_c(f))^2$

PRE:

$\tau_{fs} = \frac{2}{N}\sum_{j=1}^{N/2} \tau_{pair}(j)$ 本质上就是 2 个一组.

当需要将样本两两对比时使用, 问题: sample size 过小. S_c^2 S_t^2 不可现

$E_W(\hat\tau^{dif}) = \tau_{fs}$ 但 $V_W(\hat\tau^{dif}) = \frac{1}{N^2}\sum_{j=1}^{N/2}(Y_{j,A}(0) + Y_{j,A}(1) - (Y_{j,B}(0) + Y_{j,B}(1)))^2$ 认假设: 组内方差相同

$$\hat V^{pair}(\hat\tau^{dif}) = \frac{4}{N \cdot (N - 2)}\sum_{j=1}^{N/2}(\hat\tau^{pair}(j) - \hat\tau^{dif})^2$$

$$E(\hat V^{pair}(\hat\tau^{dif})) = V_W(\hat\tau^{dif}) + \frac{4}{N(N - 2)}\sum_{j=1}^{N/2}(\tau_{pair}(j) - \tau)^2$$ 更保守.

双样本均值 t-test

$T = \frac{(\bar X - \bar Y) - (u_1 - u_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$ 方差已知

$T = \frac{(\bar X - \bar Y) - (u_1 - u_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{\sum(X_i - \bar X)^2 + \sum(Y_i - \bar Y)^2}{n_1 + n_2 - 2}}} \sim t(n_1 + n_2 - 2)$ 方差未知但相等.

sp 推理 HW

$$E_{sp}[\tau W_i (X_i - u_x)\beta] = \tau\beta E_{sp}(W_i(X_i - u_x))$$

$$= \tau\beta E_{sp}((X_i - u_x) E_W(W_i \mid X, Y))$$

$$= \tau\beta E_{sp}((X_i - u_x) E_W(W_i \mid X))$$ Unconfounded 假设.

$$= \tau\beta E_{sp}((X_i - u_x) n_t / n) = 0$$