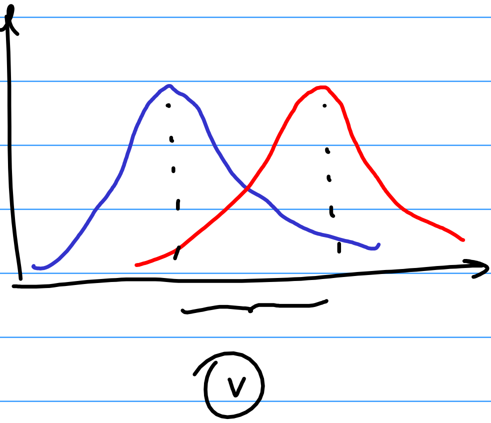
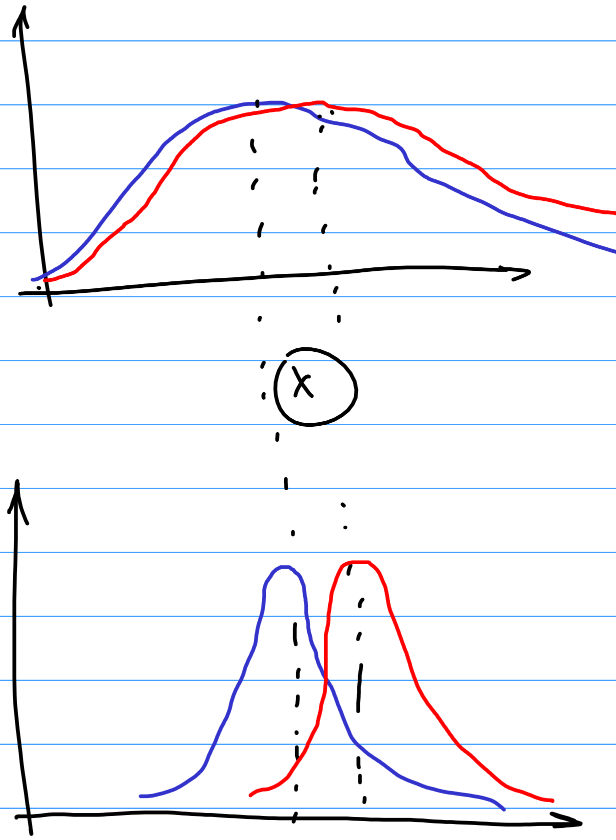
Variance: Stratification





1. CUPED (Controlled Experiment using Pre-Experimental Data)
2. Diff-in-diff

Naive Approach $\Delta = \bar{y}_1 - \bar{y}_0$

Example: y – employment level

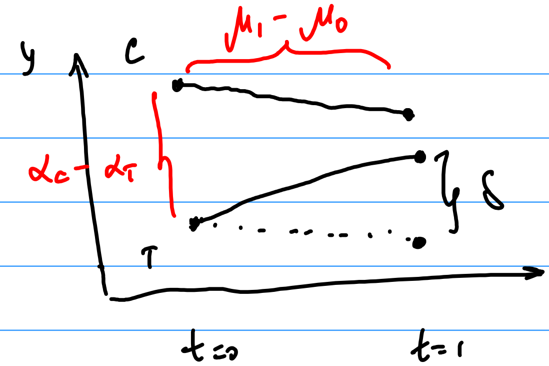
$t=1$ policy increasing min wage

(1) $\Delta = \bar{y}_1 - \bar{y}_0$ other factors changing over time aren't controlled for

t

② $\Delta = \bar{y}_T - \bar{y}_c$

selection Bias



$\shortparallel$

Diff. in. diff

$$E(y \mid T, t=0) = \mu_0 + \alpha_T$$

$$E(y \mid T, t=1) = \mu_1 + \alpha_T + \delta$$

$\uparrow$ unknown TE

$$\Delta_T = \mu_1 - \mu_0 + \delta$$

$$E(y \mid C, t=0) = \mu_0 + \alpha_c$$

$$E(y \mid C, t=1) = \mu_1 + \alpha_c$$

$$\Delta_c = \mu_1 - \mu_0$$

$$\underbrace{\Delta_T} - \underbrace{\Delta_c} = \delta$$

$$\Delta y = \beta + \delta \cdot D_T + \varepsilon$$

② CUPED $\quad y_1 \mid X$ – covariate : $y_{t=-2w}$

1) $y^{cuped} = y - \theta \cdot X$

$$\theta = \frac{cov(y, x)}{Var(x)}$$

Frisch-Waugh

$y \mid D_T, X \quad \hat{\theta}$

1) $y \mid X \Rightarrow \tilde{y}$

2) $\tilde{y} \mid D_T \Rightarrow \hat{\theta}$

2) t-test for y^{cuped};

$$E(y) = E(\bar{y}) =$$

$$= E(\bar{y} - \theta X) + \theta E(X) =$$

$$= E(\bar{y} - \theta \bar{X}) + \theta E(X)$$

$$\tilde{y}^{cup} = \bar{y} - \theta \bar{x} + \theta E(X) \to \min_{\theta}$$

$y|X$

$$\theta = \frac{cov(x,y)}{Var(x)} \qquad \hat{\theta} = \frac{\widehat{cov}(x,y)}{Var(x)}$$

$$Var(\bar{y}^{cup}) = (1-\rho^2) \cdot Var(\bar{y}) \leq Var(\bar{y})$$

unexplained var. ratio

Generalization: $y = \alpha + \delta_1 D_{T_1} + \dots + \delta_k D_{T_k} +$

$$+ \theta_1 X_1 + \dots + \theta_k X_k + \varepsilon$$

stratification = Cuped with X- categorical

$$Var(\bar{y}) = \underbrace{\sum \frac{w_k}{n} \sigma_k^2}_{\text{Within}} + \underbrace{\sum \frac{w_k}{n} \cdot (\mu_k - \mu)^2}_{\text{between}} \geq$$

$$\sum \frac{w_k}{n} \cdot \sigma_k^2 = Var(\hat{y}_{strat})$$

$E(X) = w$

$Var(X) = (1-w)w$

$$X = \begin{cases} 1, & w \\ 0, & 1-w \end{cases}$$

$$\hat{y}_{strat} = w \cdot \bar{y}_1 + (1-w) \cdot \bar{y}_0$$

$$\hat{y}_{cuped} = \bar{y} - \hat{\theta} \cdot \bar{X} + \hat{\theta} \cdot \underbrace{E(X)}_{=w}$$

$$y = \beta + \delta \cdot X + \varepsilon$$

$$\delta = \bar{y}_1 - \bar{y}_0$$

$$\hat{\theta} = \frac{\hat{Cov}(X,Y)}{\hat{Var}(X)} = \bar{y}_1 - \bar{y}_0$$

$$\hat{y}_{cv} = \bar{y} - (\bar{y}_1 - \bar{y}_0) \cdot \bar{X} + (\bar{y}_1 - \bar{y}_0) \cdot w =$$

$$\left\{ \bar{y} = \bar{X} \cdot \bar{y}_1 + (1-\bar{X}) \cdot \bar{y}_0 \right\} \quad = (1-\bar{X}) \cdot \bar{y}_0 + \bar{y}_0 \cdot \bar{X} + (\bar{y}_1 - \bar{y}_0) w =$$

$$= w \cdot \bar{y}_1 + (1-w) \cdot \bar{y}_0 = \hat{y}_{strat}$$