Semiparametric Adaptive Estimation in Survey Sampling

BIRS Workshop @ UBC Okanagan

May 24th, 2022

Kosuke Morikawa

Graduate School of Engineering Science, Osaka University, Japan
Earthquake Research Institute, The University of Tokyo, Japan

This talk is joint work with

Jae Kwang Kim

Department of Statistics, Iowa State University, U.S.A.

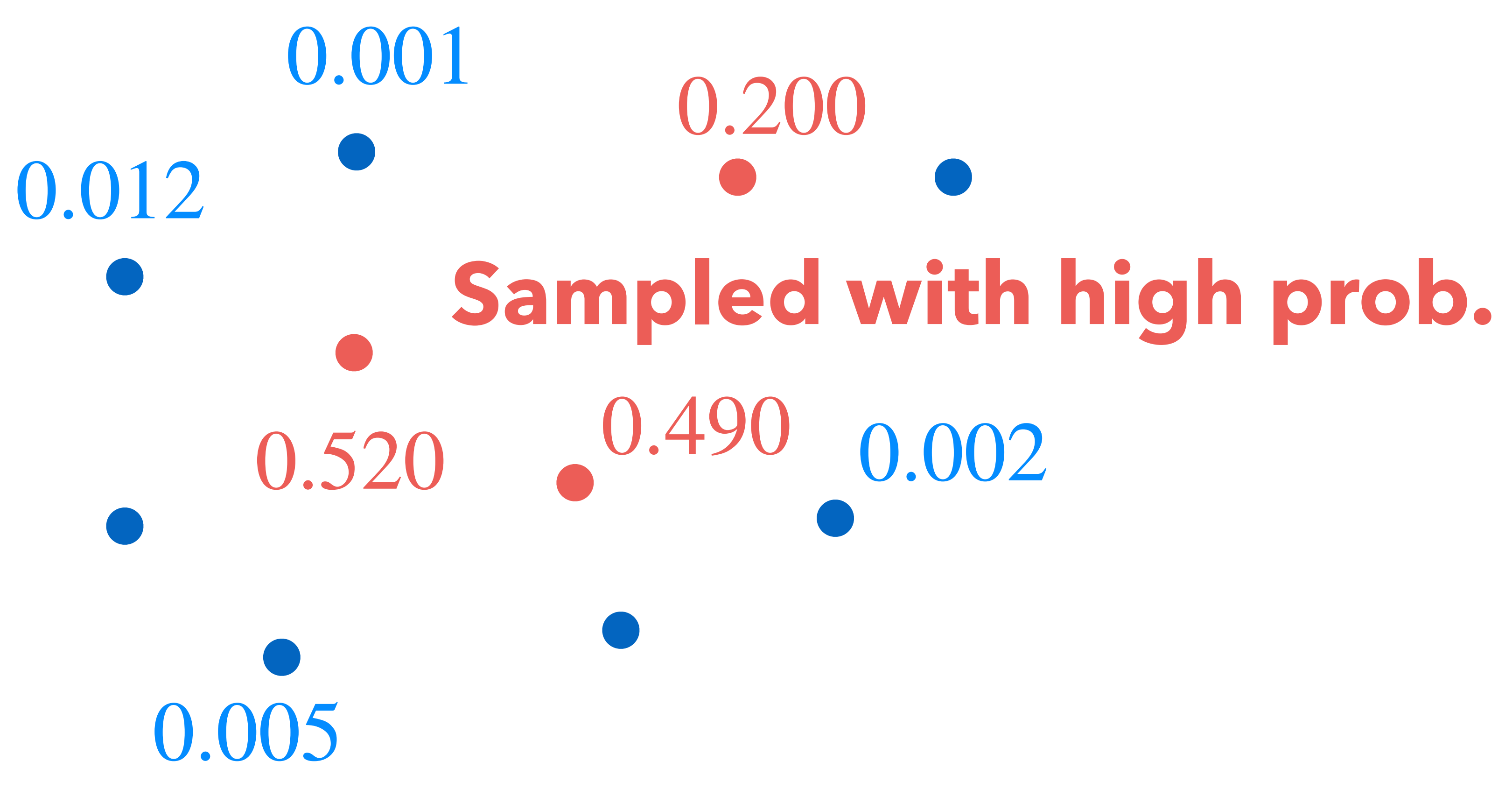
Brief Summary

- In survey sampling, some data are sampled according to inclusion probabilities instead of using all the data from the target population

- The inclusion probability (or weight) plays an important role to conduct valid statistical analysis

- However, classical weighting methods are unstable especially when the weights are extremely large

- We propose an estimator that attains the semiparametric efficiency bound by using a model on the weighting mechanism



illlustration of inclusion probability

Contents

- Introduction
- Proposed Estimator
- Simulation
- Real Data Analysis

Contents

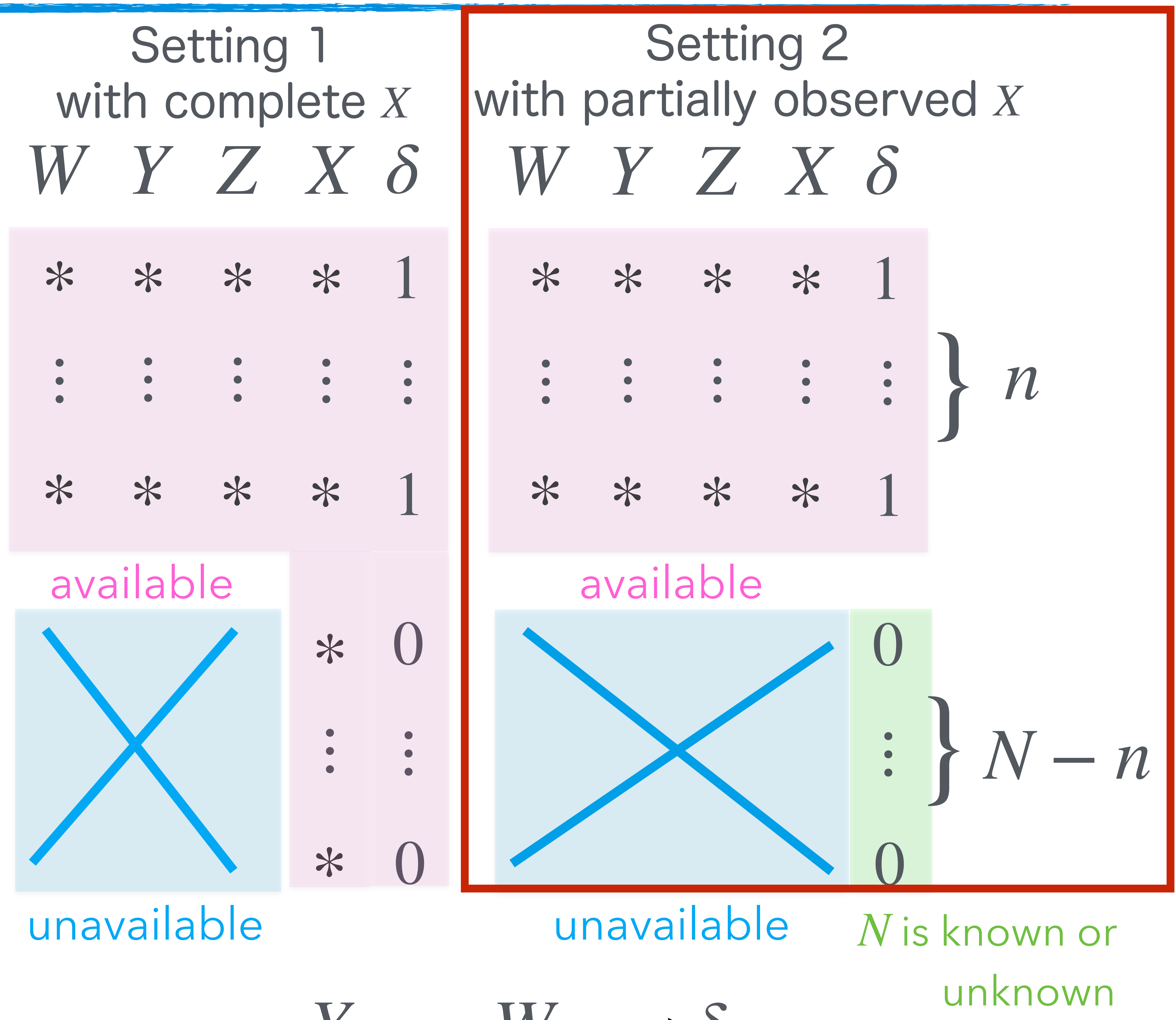
- Introduction
- Proposed Estimator
- Simulation
- Real Data Analysis

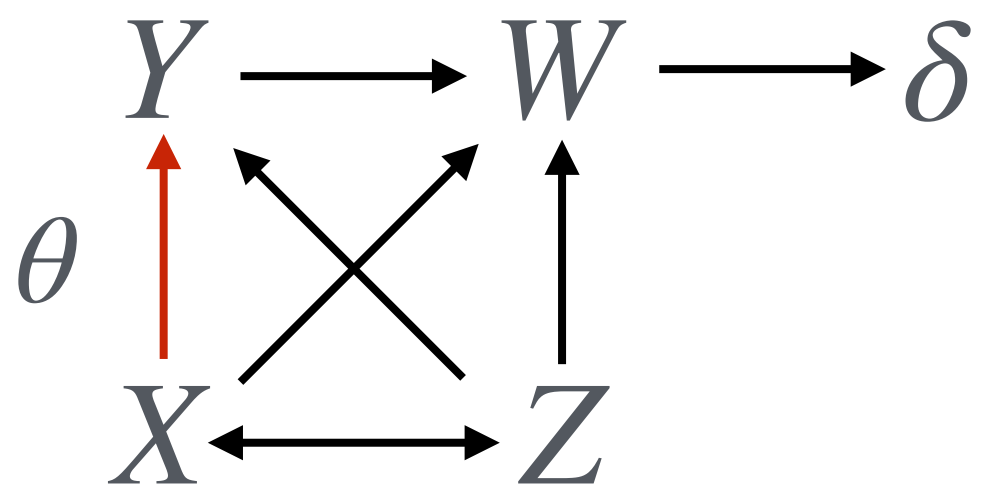
Setup

We consider this setting in this talk

- Variables: $(X_i, Y_i, Z_i, W_i, \delta_i)_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} F$
  - Y: response variable
  - X: (interesting) covariate
  - Z: other covariates
  - W: inverse of inclusion probability
  - δ: sampling indicator

    takes 1 if data are sampled
  - n: size of sampled dataset $\sum_{i=1}^{N} \delta_i = n$
- Target: $E(Y)$, $E(Y \mid x; \theta)$, $f(y \mid x; \theta)$

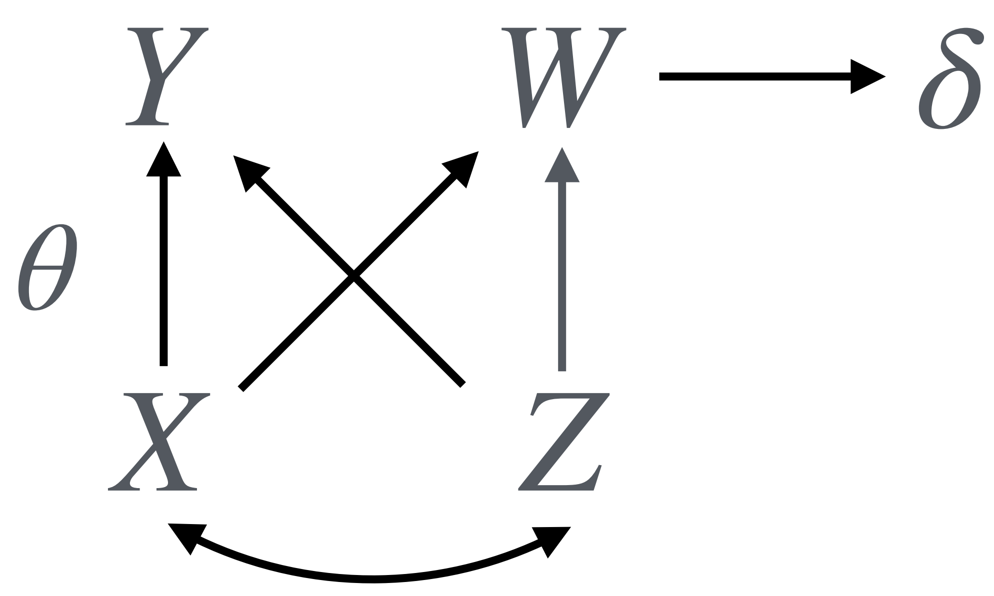




Sampling Mechanism

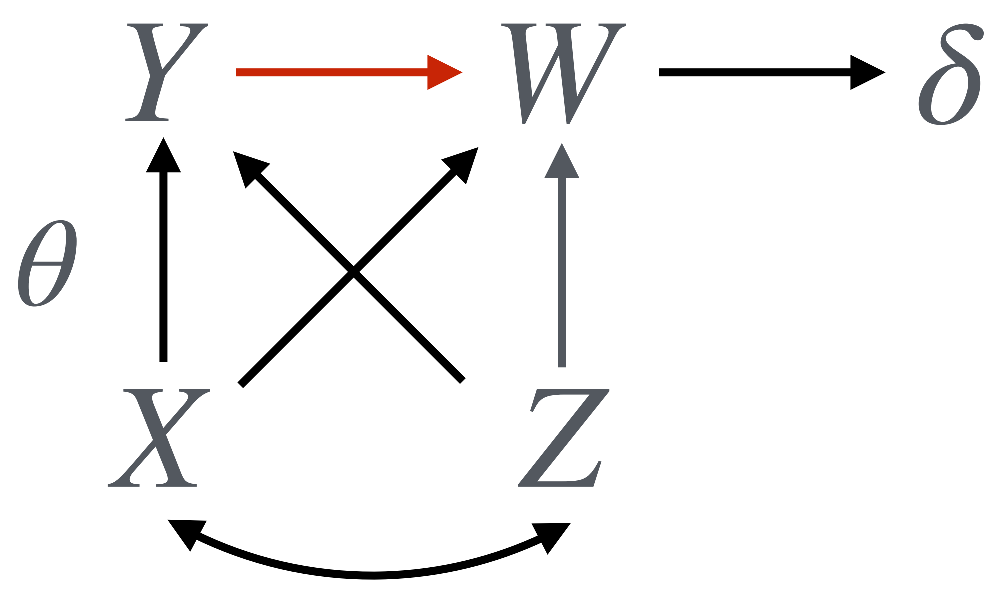
- Non-informative sampling (MAR)

$$W \perp Y \mid (X, Z)$$



- Informative sampling (NMAR)

$$W \not\perp Y \mid (X, Z)$$



We consider **informative sampling** in this talk

Example: The Canadian Workplace and Employee Survey (Fuller, 2009)

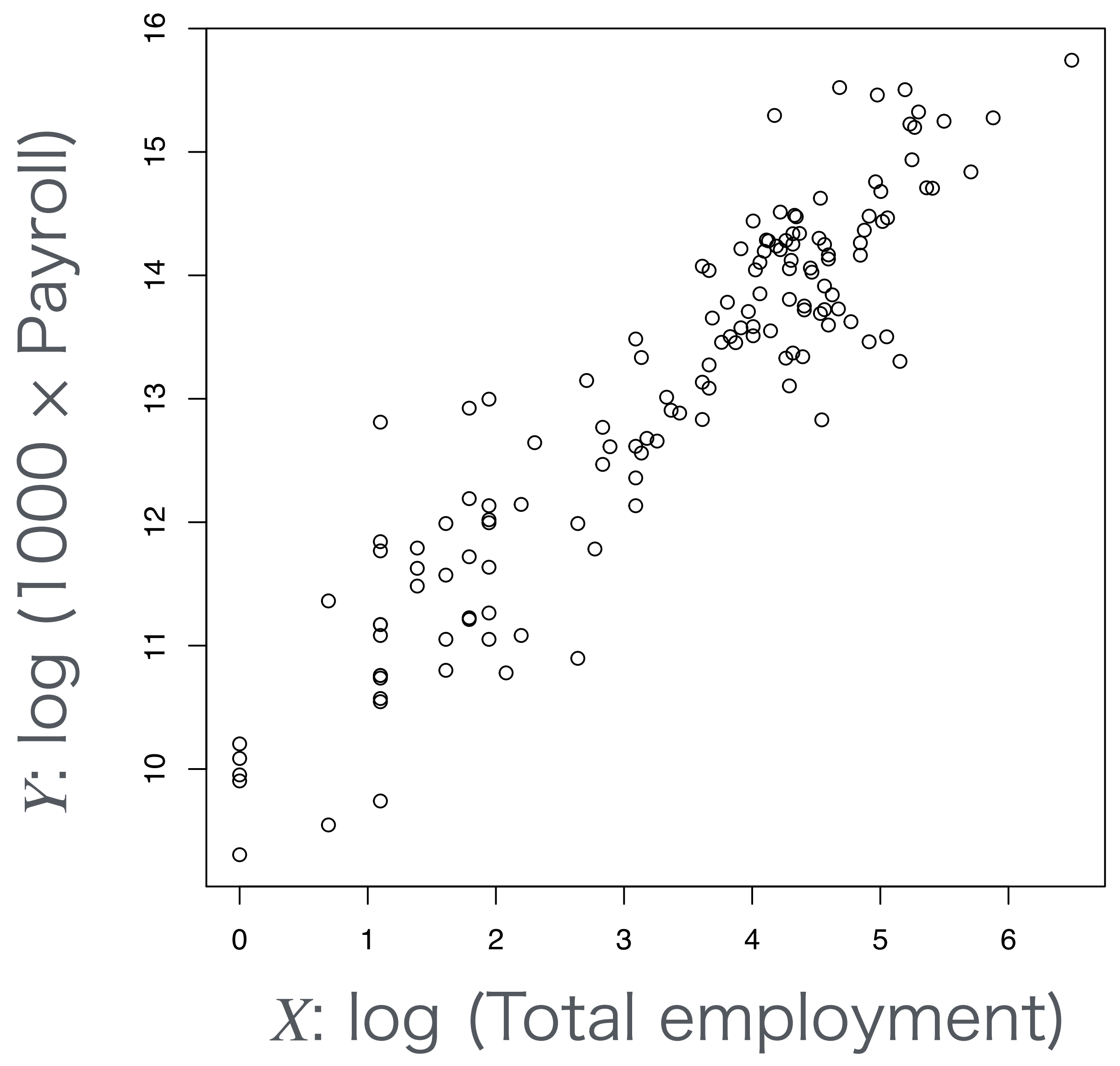
- We want to know the relationship between Payroll (Y) and total Employment (X)

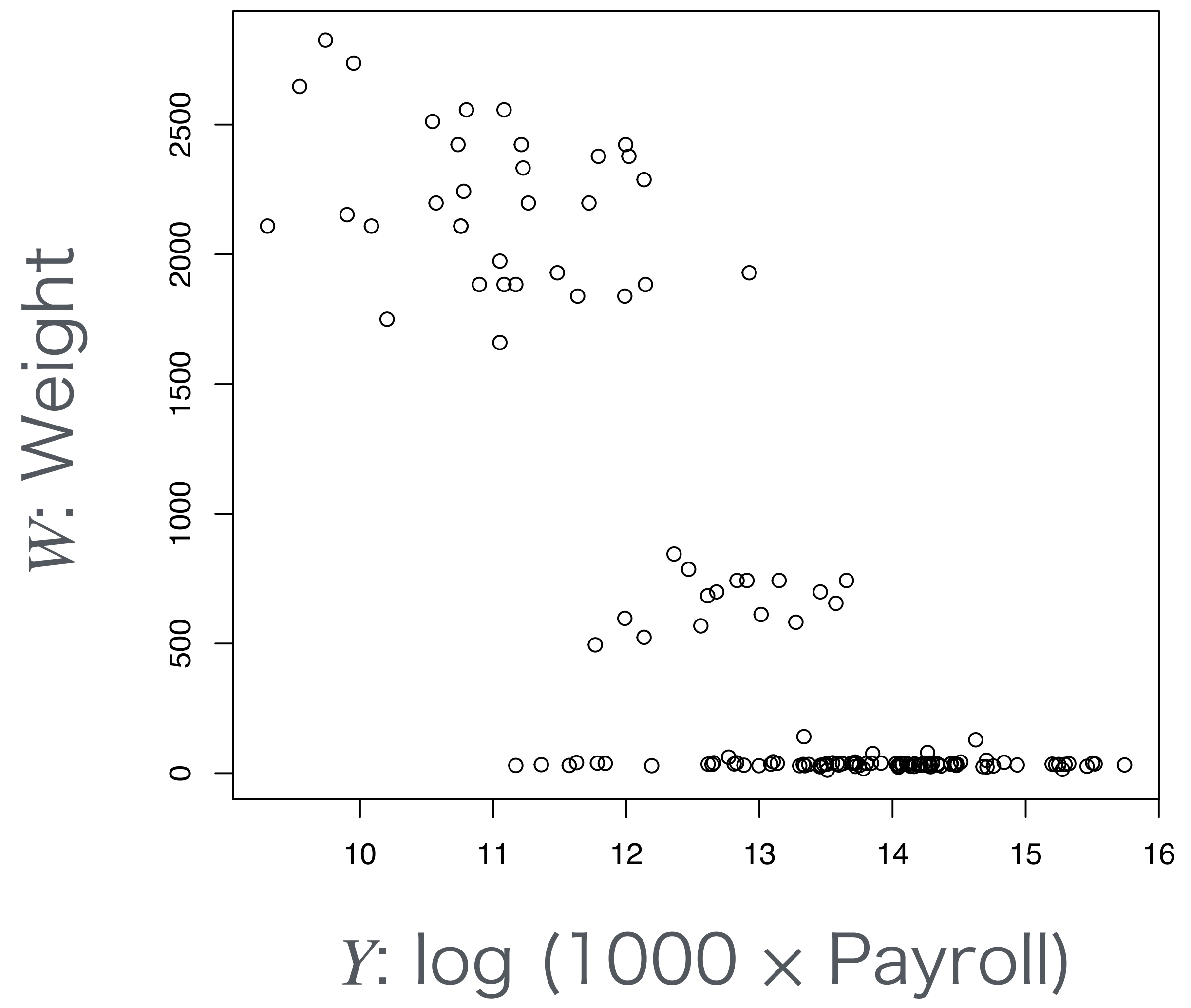
- Size of population (N): 2029 workplaces

- Sampled size (n): 142 workplaces
  - Stratified sampling (3 strata)
    + simple random sampling
    with nonresponse adjustment

- Model:

$$Y \mid X = x \sim N(a + bx,\ \sigma^2), \quad \theta = (a, b, \sigma^2)^\top$$





Z-estimator

(Semiparametric) Z-estimator θ: Unique solution to

$$E\{U(X, Y; \theta)\} = 0$$

$U(\cdot)$ depends on θ as follows..

Mean of response variable: $\theta = E(Y) \;\Rightarrow\; U(X, Y; \theta) = \theta - Y$

Regression parameter: $\mu(X; \theta) = E(Y \mid X) \;\Rightarrow\; U(X, Y; \theta) = A(X)\{Y - \mu(X; \theta)\}$

arbitrary function

Outcome model: $f(Y \mid X; \theta) \;\Rightarrow\; U(X, Y; \theta) = \frac{\partial}{\partial \theta} \log f(Y \mid X; \theta)$

$$\frac{\partial}{\partial \theta} \log f(Y \mid X; \theta) = S_\theta(X, Y)$$

Score function

Horvitz-Thompson Estimator

- Horvitz-Thompson (HT) estimator: the solution to

$$\sum_{i=1}^{n} W_i U(X_i, Y_i;\ \theta) = 0,$$

Available when N is unknown

where $E\{U(X, Y; \theta)\} = 0$

- The most well known method in survey sampling
- No additional assumptions are required
- Theoretical validity: Unbiased estimating equation $\Rightarrow$ moment method

Proof for Unbiasedness

$$E\left[\sum_{i=1}^{n} W_i U(X_i, Y_i; \theta)\right] = E\left[\sum_{i=1}^{N} \delta_i W_i U(X_i, Y_i; \theta)\right]$$

$$= E\left[\sum_{i=1}^{N} \underbrace{P(\delta_i = 1 \mid X_i, Y_i, W_i)}_{= \frac{1}{W_i}} W_i U(X_i, Y_i; \theta)\right]$$

$$= N \times E\left[U(X, Y; \theta)\right]$$

$$= 0$$

Smoothing Weight

- Smoothing weight: $\tilde{W} := E(W \mid x, y, \delta = 1)$

- Beaumont (2008, Biometrika) shows that using $\tilde{W}$ instead of W is more efficient in the context of regression analysis
  - $\tilde{W}(x, y)$ is to be estimated
  - Misspecification of the model causes bias

- Kim and Skinner (2013, Biometrika) proposed an optimal weight in the same setup.

There are possibilities that we can construct more efficient estimator than HT!!

Preparation: Bayes' Theorem

- Let $f_1(y \mid x) = f(y \mid x, \delta = 1)$ and $\pi(x, y) = P(\delta = 1 \mid x, y)$

- Transformation of $f_1 \to f$

$$f_1(y \mid x) = f(y \mid x, \delta = 1) = \frac{f(y, \delta = 1 \mid x)}{P(\delta = 1 \mid x)} = \frac{f(y \mid x)\pi(x, y)}{\int f(y \mid x)\pi(x, y)\mathrm{d}y}$$

- Transformation of $f \to f_1$

$$f(y \mid x) = \frac{f_1(y \mid x)\pi^{-1}(x, y)}{\int f_1(y \mid x)\pi^{-1}(x, y)\mathrm{d}y}$$

Conditional Maximum Likelihood (CML) for Outcome model

- Assume that
  - $f(y \mid x; \theta)$ is of our interest
  - response probability $\pi(x, y) = P(\delta = 1 \mid x, y)$ is known

$$f_1(y \mid x) = f(y \mid x, \delta = 1) = \frac{f(y \mid x)\pi(x, y)}{\int f(y \mid x)\pi(x, y)\mathrm{d}y}$$

- Then, the conditional maximum likelihood (CML) estimator is the efficient: the solution to

$$\sum_{i=1}^{n} S_{1,\theta}(X_i, Y_i) := \sum_{i=1}^{n} \frac{\partial \log f_1(Y_i \mid X_i)}{\partial \theta} = 0$$

$$= \sum_{i=1}^{n} \left[S_\theta(X_i, Y_i) - \frac{\int S_\theta(x, y)\pi(x, y) f(y \mid x; \theta)\mathrm{d}y}{\int \pi(x, y) f(y \mid x; \theta)\mathrm{d}y} \right]$$

$$= \sum_{i=1}^{n} \left[S_\theta(X_i, Y_i) - E_1\{S_\theta(x, Y) \mid x; \theta\} \right]$$

How to Handle When $\pi(x, y)$ is Unknown??

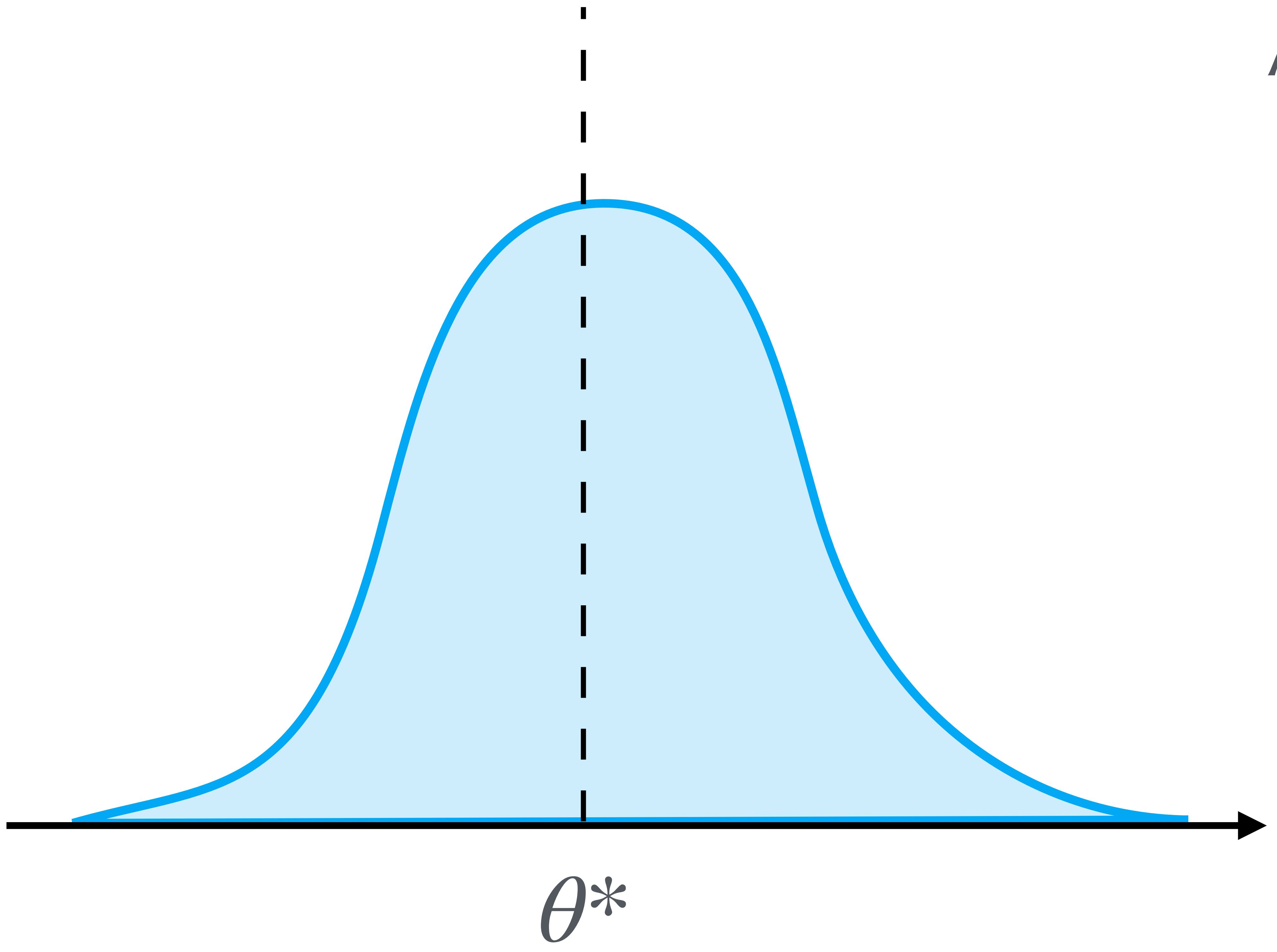
- Sverchkov and Pfeffermann (1999, Sankya B) shows that

$$
\begin{aligned}
E_1(W \mid x, y) &= \int w f_1(w \mid x, y) \mathrm{d}w \\
&= \frac{\int w \overbrace{P(\delta = 1 \mid w, x, y)}^{\frac{1}{w}\,=} f(w \mid x, y) \mathrm{d}w}{\int P(\delta = 1 \mid w, x, y) f(w \mid x, y) \mathrm{d}w} \\
&= \frac{1}{P(\delta = 1 \mid x, y)} =: \frac{1}{\pi(x, y)}
\end{aligned}
$$

- π can be estimated by the regression W on (X, Y) with sampled data
- If π is misspecified, the estimator causes bias

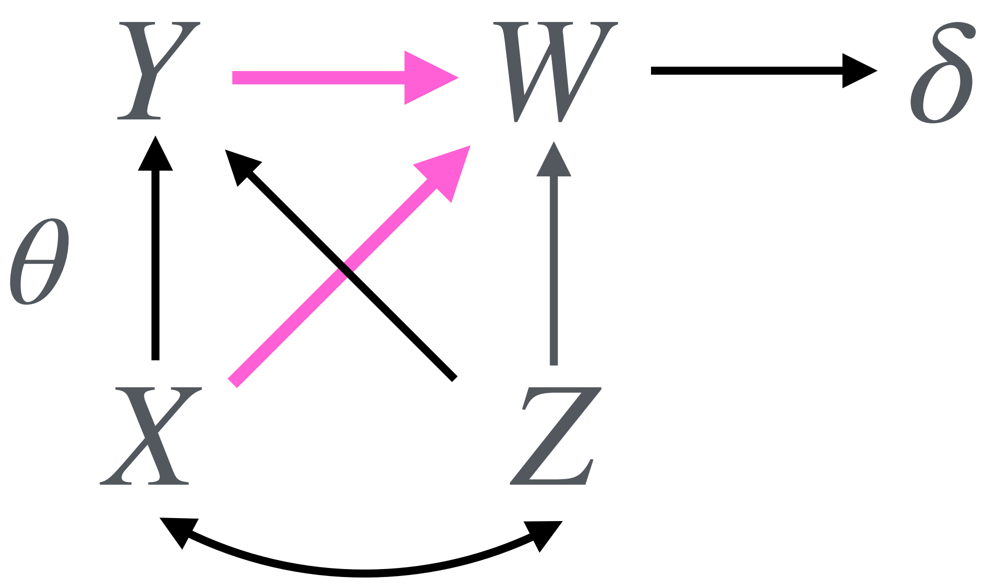
Conditional Maximum Likelihood (CML)

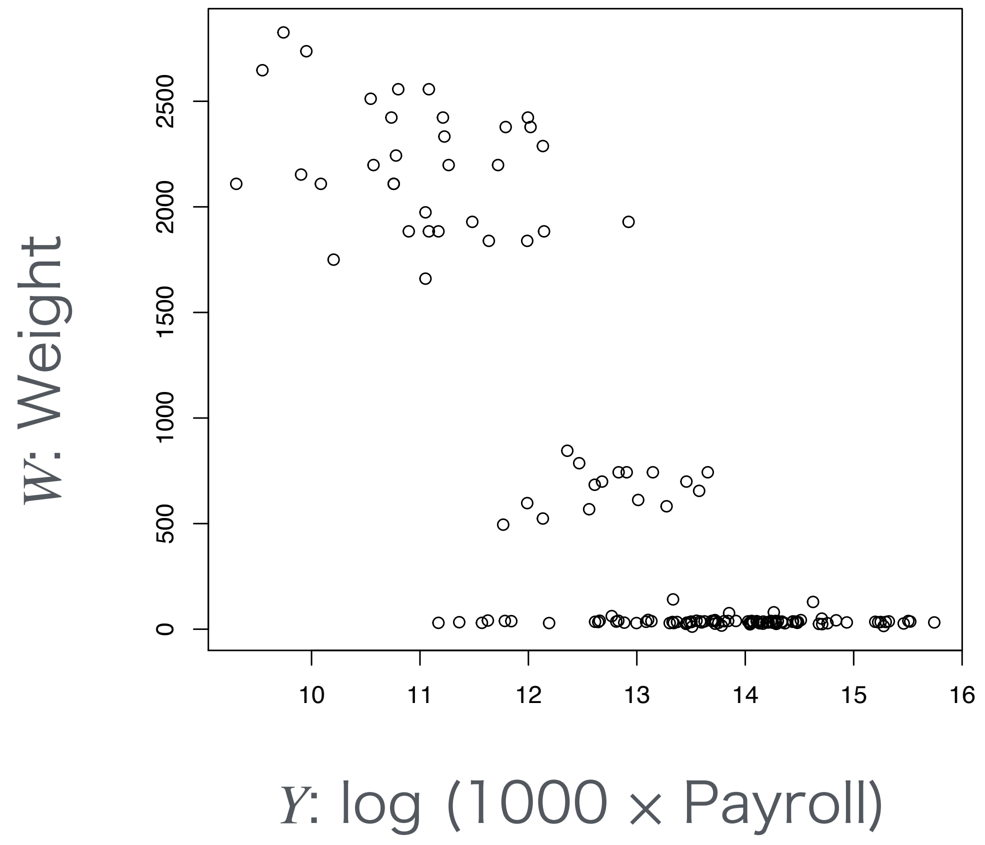
Dist. of HT estimator



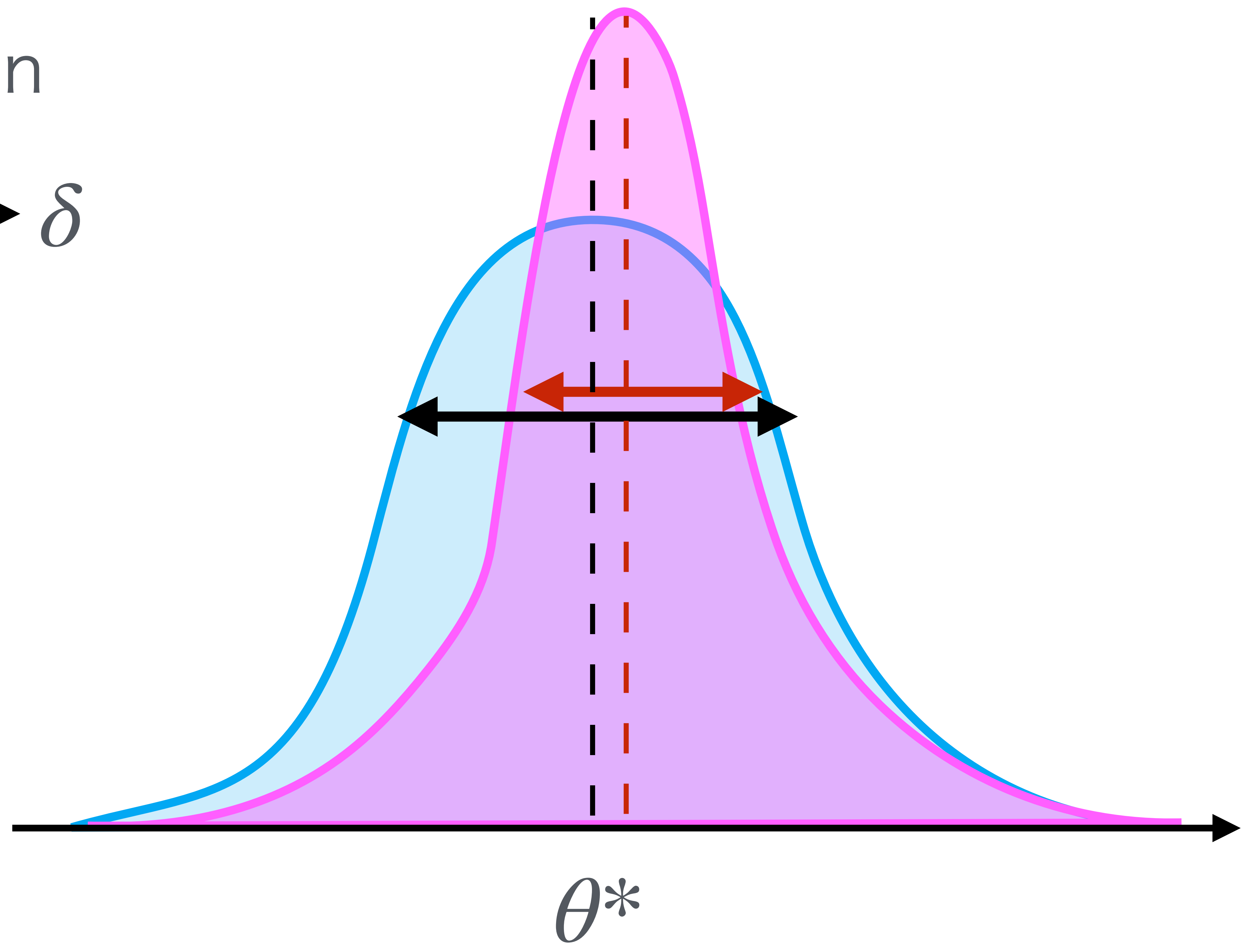
- ✔ Consistency
- ✔ Asymptotic normality

Add information on





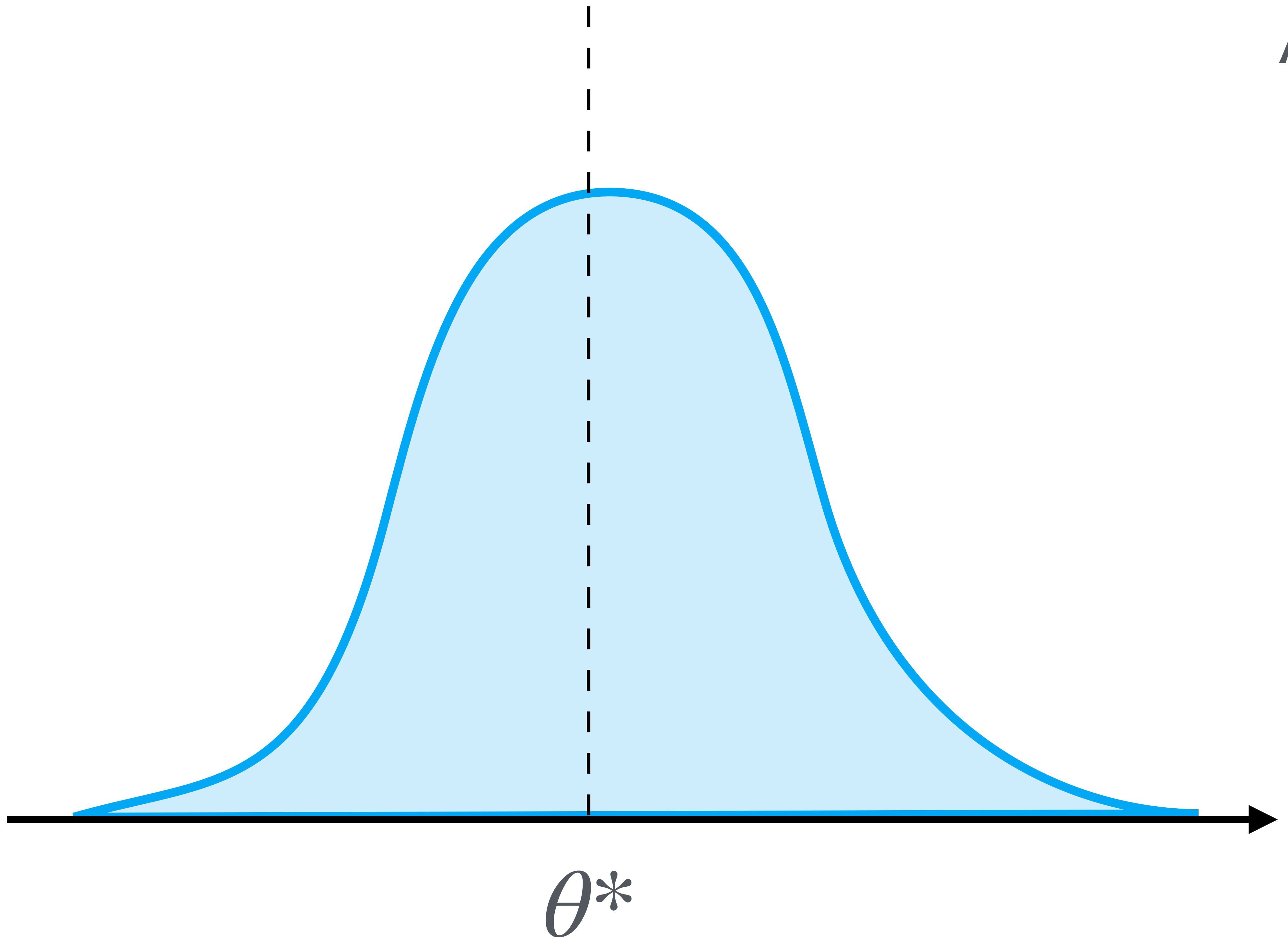
Dist. of CML estimator



- ✔ Consistency
- ✔ Asymptotic normality
- ✔ Efficiency

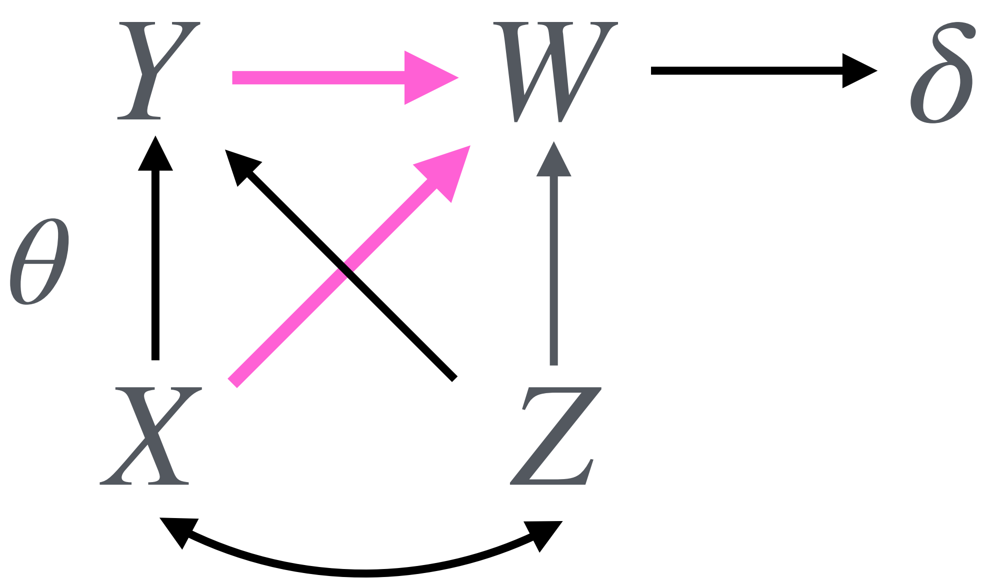
Our Goal

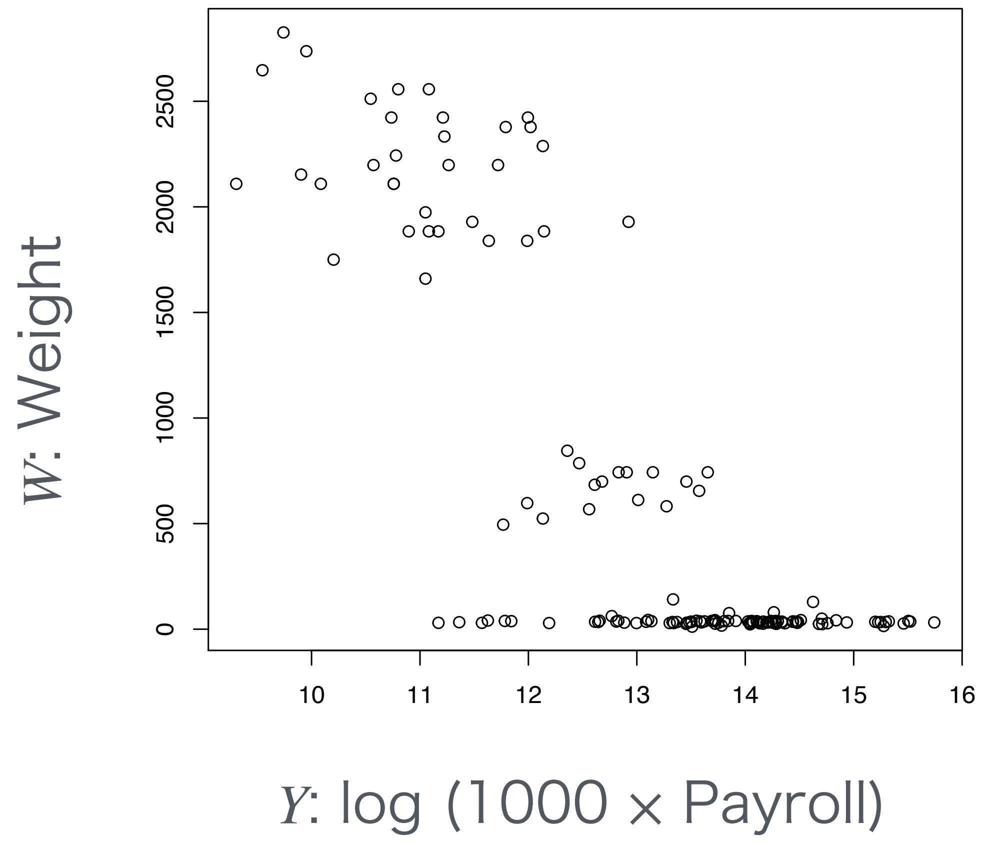
Dist. of HT estimator



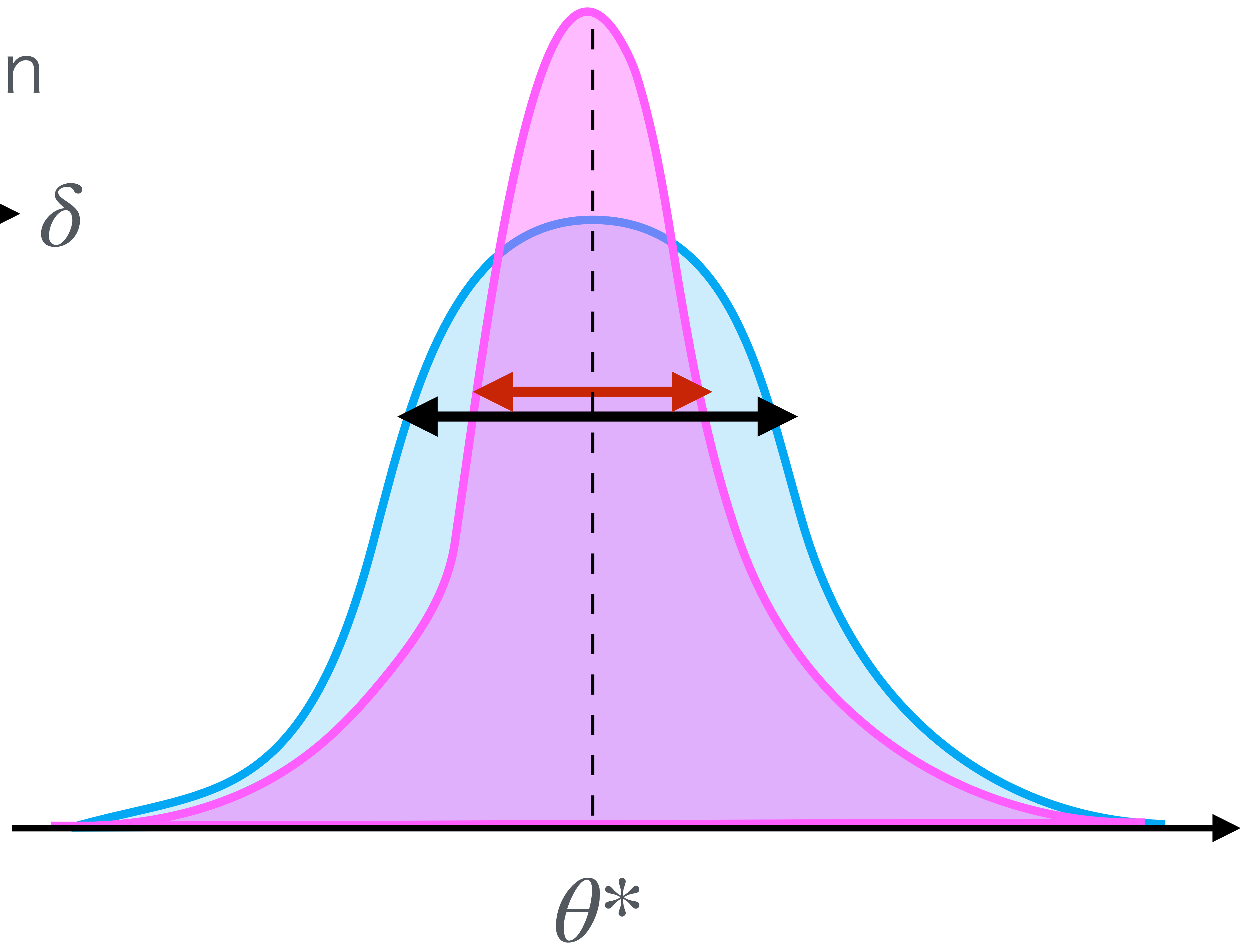
- ✔ Consistency
- ✔ Asymptotic normality

Add information on





Dist. of Proposed estimator



- ✔ Consistency
- ✔ Asymptotic normality
- ✔ Efficiency

Contents

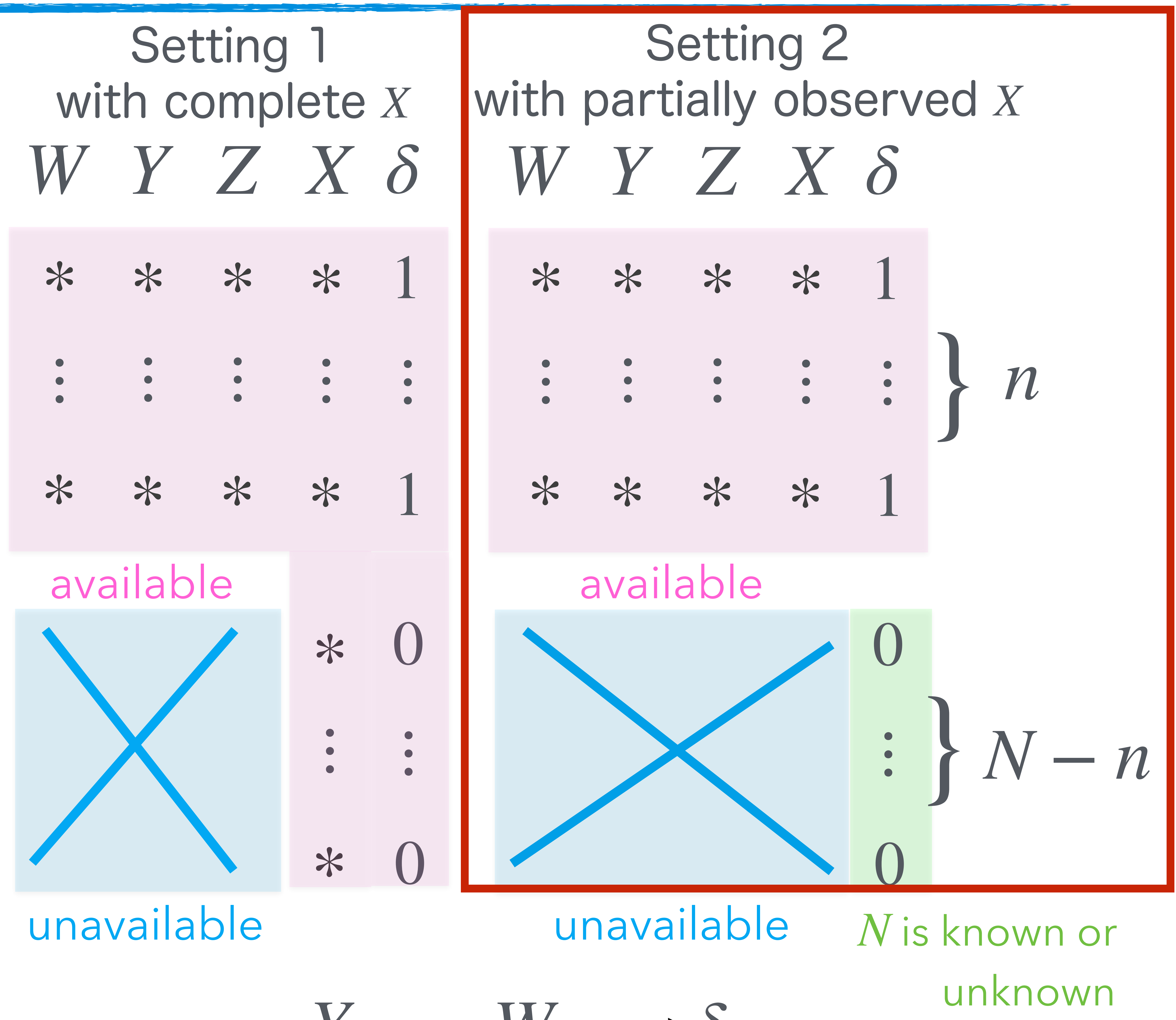
- Introduction
- Proposed Estimator
- Simulation
- Real Data Analysis

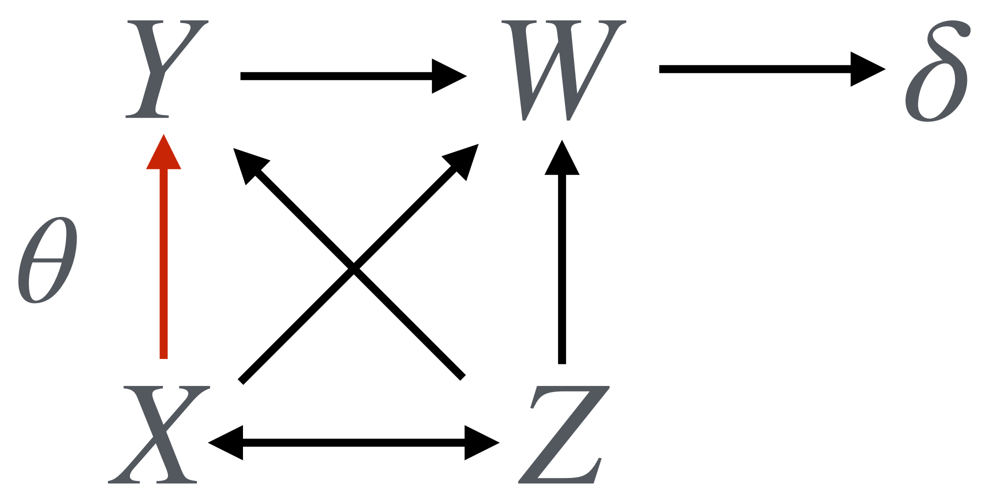
Setup

We consider this setting in this talk

- Variables: $(X_i, Y_i, Z_i, W_i, \delta_i)_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} F$
  - Y: response variable
  - X: (target) covariate
  - Z: other covariates
  - W: inverse of inclusion probability
  - δ: sampling indicator

    takes 1 if data are sampled
  - n: size of sampled dataset $\sum_{i=1}^{N} \delta_i = n$
- Target: $E(Y)$, $E(Y \mid x; \theta)$, $f(y \mid x; \theta)$



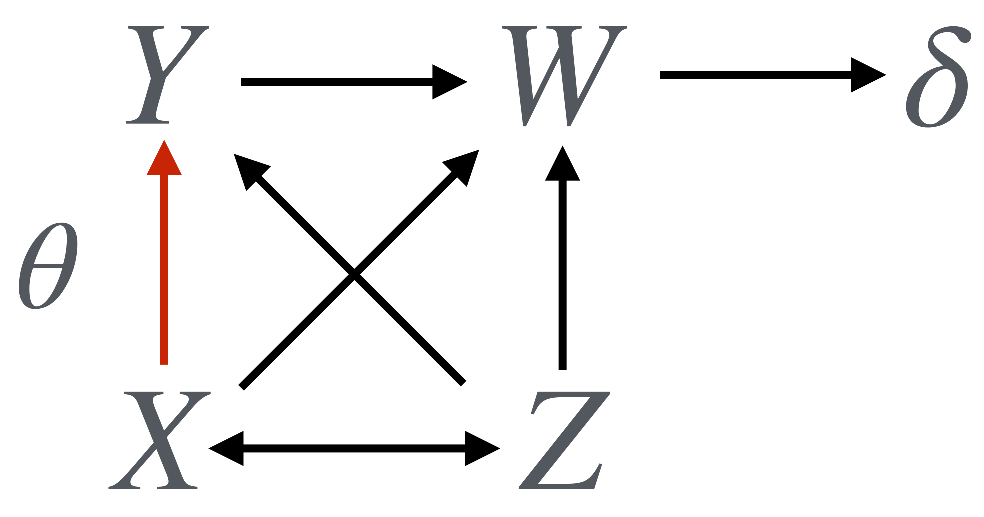


Key Idea: Regard W as a Random Variable

- $W^{-1} = P(\delta = 1 \mid X, Y, Z, W)$ is a probability (propensity score)

- However, **we treat W as a random variable** and construct a semiparametric model

$$
\begin{aligned}
&f(x, y, z, w \mid \delta = 1; \theta, \eta_1, \eta_2, \eta_3) \\
&= \frac{P(\delta = 1 \mid x, y, z, w) f(z, w \mid x, y; \eta_1) f(y \mid x; \theta, \eta_3) f(x; \eta_2)}{\int P(\delta = 1 \mid x, y, z, w) f(z, w \mid x, y; \eta_1) f(y \mid x; \theta, \eta_3) f(x; \eta_2) \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}w} \\
&= \frac{w^{-1} f(z, w \mid x, y; \eta_1) f(y \mid x; \theta, \eta_3) f(x; \eta_2)}{\int w^{-1} f(z, w \mid x, y; \eta_1) f(y \mid x; \theta, \eta_3) f(x; \eta_2) \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}w}
\end{aligned}
$$



  - η_1, η_2, η_3: infinite dimensional nuisance parameters
  - NOTE: If our interest is estimating outcome model $f(y \mid x; \theta)$, then $f(y \mid x; \theta) = f(y \mid x; \theta, \eta_3)$

- **Goal**: Estimate θ that is not affected by η_1, η_2, η_3

Lemma: Rotnitzky and Robins (1997, Stat. Med.)

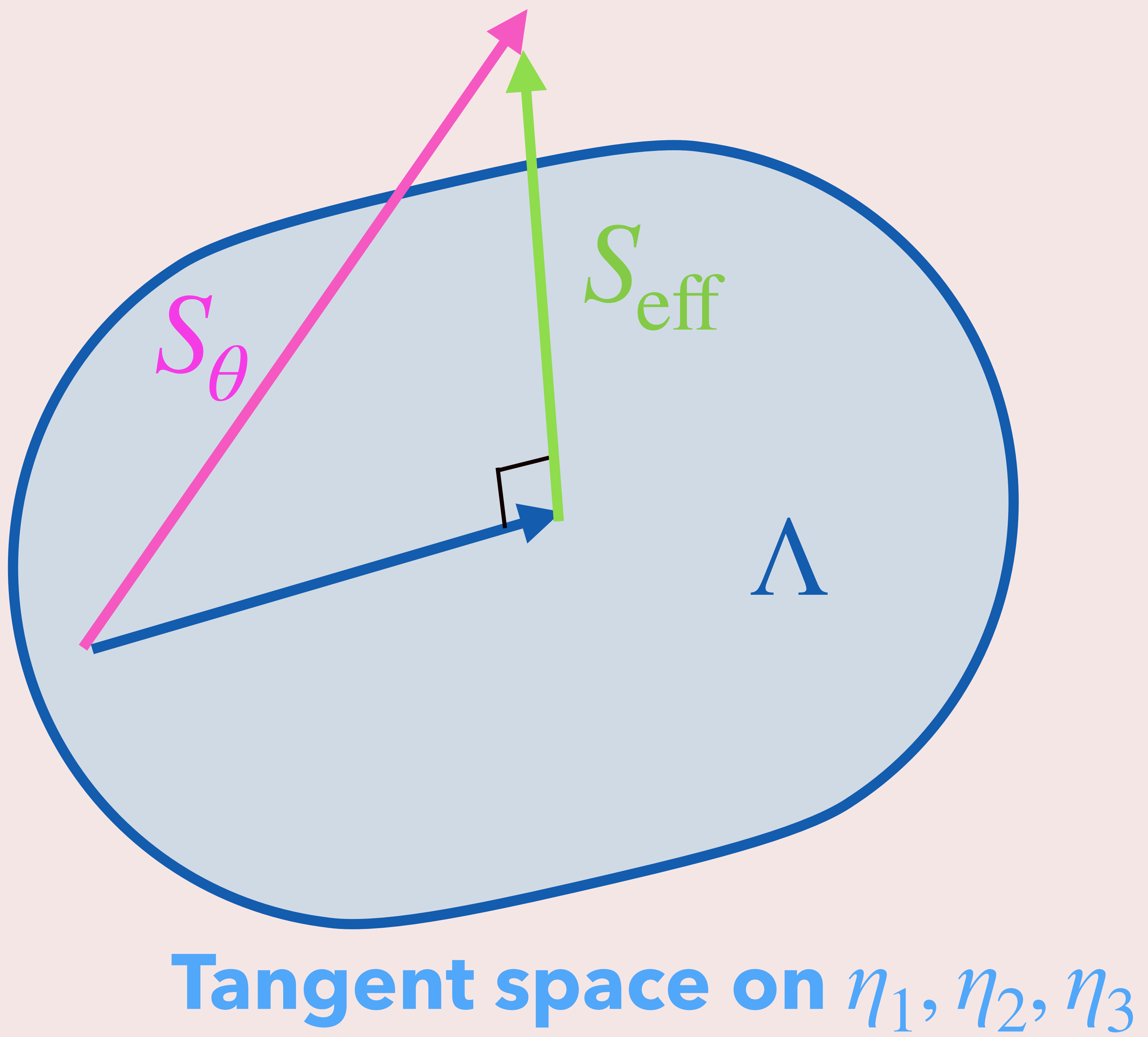
Lemma 1. When N is known

The efficient score S_{eff} is given by

$$S_{\text{eff}} = \underbrace{\delta W D^*_{\text{eff}}}_{\text{IPW}} + \underbrace{(1-\delta W)\frac{E\{(W-1)D^*_{\text{eff}}\}}{E(W-1)}}_{\text{Augmented term}},$$

where $D^*_{\text{eff}} \in \Lambda^{F,\perp}$ is the unique solution to

$$\Pi\left(W D^*_{\text{eff}} - (W-1)\frac{E\{(W-1)D^*_{\text{eff}}\}}{E(W-1)} \,\middle|\, \Lambda^{F,\perp} \right) = S^F_{\text{eff}}$$



Then, the semiparametric efficiency bound for θ is $\{E(S_{\text{eff}}^{\otimes 2})\}^{-1}$

Target Parameter

1. Z-estimator: Solution to $E\{U(X,Y;\theta)\} = 0$

$$\theta = E(Y) \;\; \Rightarrow \;\; U(X,Y;\theta) = \theta - Y$$

2. Regression parameter: $\mu(X;\theta) = E(Y \mid X)$

3. Outcome model: $f(Y \mid X;\theta)$

Semiparametric Efficiency Bound for θ with partially observed X

Theorem 1. When N is known

The efficient score for θ is

$$S_{\mathrm{eff}} = \underbrace{\delta W D^*_{\mathrm{eff}}}_{\text{IPW}} + \underbrace{(1-\delta W) c^*_{\mathrm{eff}}}_{\text{Augmented term}},$$

where D^*_{eff} and c^*_{eff} are different according to the target parameters.

The semiparametric efficiency bound for θ is $\{E(S_{\mathrm{eff}}^{\otimes 2})\}^{-1}$

$S_{\text{eff}} = \delta W D^*_{\text{eff}} + (1 - \delta W) c^*_{\text{eff}}$

(i) $E\{U(X, Y; \theta)\} = 0$:

$$\bar{\pi} = \bar{\pi}(x, y) = \frac{1}{E(W \mid x, y)}$$

$$D^*_{\text{eff}} = U(\theta), \qquad c^*_{\text{eff}} = \frac{E\{(W-1)U(\theta)\}}{E(W-1)}.$$

(ii) $\mu(x; \theta) = E(Y \mid x)$

$$D^*_{\text{eff}} = A^*_{\text{eff}}(X)\{\underbrace{Y - \mu(X; \theta)}_{\varepsilon}\}, \qquad c^*_{\text{eff}} = \frac{E\left[\frac{E(W\varepsilon \mid X)}{E(W\varepsilon^2 \mid X)} \frac{\partial}{\partial \theta} \mu(X; \theta)\right]}{E\left[E(W-1) - \frac{\{E(W\varepsilon \mid X)\}^2}{E(W\varepsilon^2 \mid X)}\right]},$$

where

$$A^*_{\text{eff}}(x) = \frac{1}{E(W\varepsilon^2 \mid x)} \left[E(W\varepsilon \mid x) c^*_{\text{eff}} + \frac{\partial}{\partial \theta} \mu(x; \theta)\right]$$

$S_{\text{eff}} = \delta W D^*_{\text{eff}} + (1 - \delta W) c^*_{\text{eff}}$

(iii) Outcome model $f(y \mid x; \theta)$:

$$\bar{\pi} = \bar{\pi}(x, y) = \frac{1}{E(W \mid x, y)}$$

$$D^*_{\text{eff}} = \bar{\pi} \left\{ S_\theta - \frac{E(\bar{\pi} S_\theta \mid x)}{E(\bar{\pi} \mid x)} \right\} + \left(1 - \frac{\bar{\pi}}{E(\bar{\pi} \mid x)} \right) c^*_{\text{eff}},$$

$$S_\theta = S_\theta(x, y) = \frac{\log f(y \mid x; \theta)}{\partial \theta}$$

$$c^*_{\text{eff}} = \frac{E\left\{ \frac{E(\bar{\pi} S_\theta \mid X)}{E(\bar{\pi} \mid X)} \right\}}{1 - E\left[\frac{1}{E(\bar{\pi} \mid X)} \right]}$$

$\bar{\pi}(x, y)$

and its conditional expectation

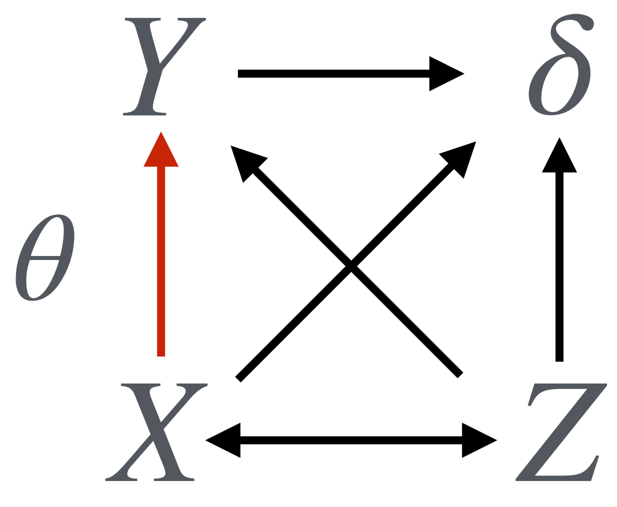
$E(\bar{\pi} \mid x)$ and $E(\bar{\pi} S_\theta \mid x)$

are unknown functions

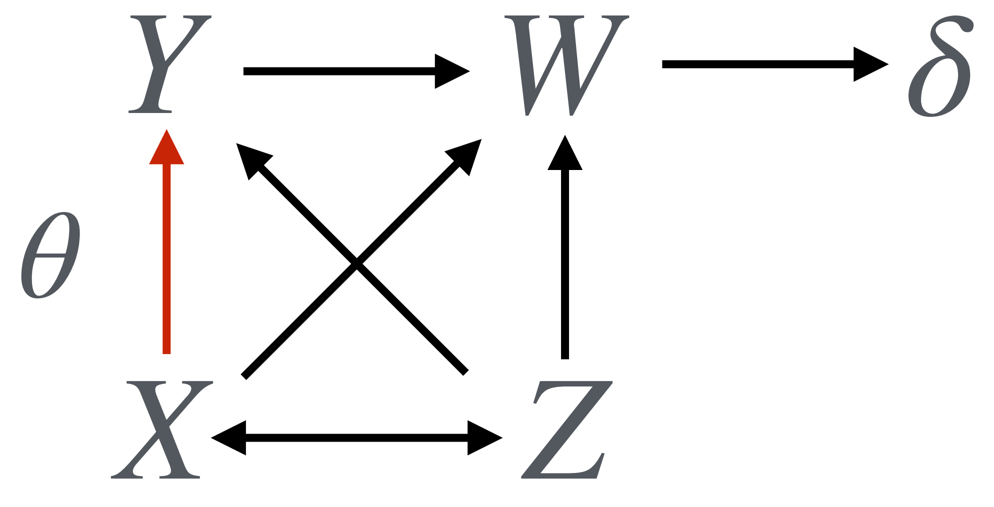
Remark. Z is Unnecessary

- Information of Z does NOT affect efficiency of θ at all
  - In missing data analysis, all the covariates that affect δ are required to be observed
  - However, in this case, observing W is enough to explain δ

- We do NOT need to sample Z even if it has an effect on W



Usual NMAR



Informative sampling

Example. Adaptive Estimator for $E(Y)$

- Estimating Equation: $S_{\text{eff}} = \delta W D^*_{\text{eff}} + (1-\delta W)c^*_{\text{eff}}, \quad U(\theta) = \theta - Y$

$$S_{\text{eff}}(\theta) = \sum_{i=1}^{N} \left\{ \delta_i W_i(\theta - Y_i) + (1-\delta_i W_i)\frac{E\{(W-1)(\theta - Y)\}}{E(W-1)} \right\} = 0$$

$$\Rightarrow \quad \hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} \left\{ \delta_i W_i Y_i + \left(1 - \delta_i W_i\right) \frac{E\{(W-1)Y\}}{E(W-1)} \right\}$$

Unknown value

$$\frac{E\{(W-1)Y\}}{E(W-1)} = \frac{E_1\{W(W-1)Y\}}{E_1(W(W-1))} \approx \frac{\sum_{\delta_j=1} W_j(W_j-1)Y_j}{\sum_{\delta_j=1} W_j(W_j-1)}$$

Estimator

Working Models

- Consider an adaptive estimator for (c) $f(y \mid x; \theta)$

- The optimal estimating equation involves estimation of unknown functions:

  1. $\bar{\pi}(x, y) = \{E(W \mid x, y)\}^{-1}$

     We give a reasonable model later.

  2. $E(\bar{\pi} \mid x) = \int \bar{\pi}(x, y) f(y \mid x; \theta) \mathrm{d}y$ and $E(\bar{\pi} S_\theta \mid x)$

     Because θ is estimable with the Horvitz-Thompson estimator (say, $\hat{\theta}_{\mathrm{HT}}$), this function can be computed by

$$\hat{E}_{\mathrm{HT}}(\bar{\pi} \mid x) = \int \bar{\pi}(x, y) f(y \mid x; \hat{\theta}_{\mathrm{HT}}) \mathrm{d}y$$

Parametric Model on W –1/2–

- $X \sim \mathrm{Beta}(\alpha, \beta) \;\Leftrightarrow\; 1 - X \sim \mathrm{Beta}(\beta, \alpha) \Leftrightarrow \dfrac{1-X}{X} \sim \mathrm{Beta}'(\beta, \alpha)$

- Assume that $W^{-1} \mid (x, y) \sim \mathrm{Beta}(m(x, y)\phi, \{1 - m(x, y)\}\phi)$
  - W^{-1} take values on $(0,\ 1)$
  - $E(W^{-1} \mid x, y) = m(x, y),\ V(W^{-1} \mid x, y) = \dfrac{m(x, y)\{1 + m(x, y)\}}{1 + \phi}$ (ϕ: precision parameter)
  - This is essentially same as the beta regression model (Ferrari and Chibari-Neto, 2004, J. Appl. Stat.)

- Thus, $O := W - 1 = \dfrac{1 - W^{-1}}{W^{-1}} \sim \mathrm{Beta}'(\{1 - m(x, y)\}\phi,\ m(x, y)\phi)$

Parametric Model on W –2/2–

$(W = O + 1)$

- Distribution on $O \mid (x, y, \delta = 1)$

$$f_1(o \mid x, y) \propto f(o \mid x, y)P(\delta = 1 \mid x, y, o) = f(o \mid x, y)\frac{1}{1+o}$$

$$= o^{\{1-m(x,y)\}\phi-1}(1+o)^{-\phi} \cdot \frac{1}{1+o}$$

$$\Rightarrow \quad O \mid (x, y, \delta = 1) \sim \mathrm{Beta}'(\{1 - m(x, y)\}\phi, m(x, y)\phi + 1)$$

- By using a property of the beta prime distribution,

$$E_1(W \mid x, y) = 1 + E_1(O \mid x, y) = \frac{1}{m(x, y)};$$

$$E(W \mid x, y) = 1 + E(O \mid x, y) = \frac{\phi - 1}{m(x, y)\phi - 1}$$

Parametric Model on W

$$E(W^{-1} \mid x, y) = m(x, y), \; V(W^{-1} \mid x, y) = \frac{m(x,y)\{1+m(x,y)\}}{1+\phi}$$

$\Uparrow$

Proposition 1.

Assume that $W^{-1} \mid (x, y) \sim \mathrm{Beta}(m(x, y)\phi, \{1 - m(x, y)\}\phi)$.

Then, $W - 1 =: O \mid (x, y) \sim \mathrm{Beta}'(\{1 - m(x, y)\}\phi, m(x, y)\phi)$ and

$$O \mid (x, y, \delta = 1) \sim \mathrm{Beta}'(\{1 - m(x, y)\}\phi, m(x, y)\phi + 1)$$

- The assumption is essentially same as the beta regression model (Ferrari and Chibari-Neto, 2004, J. Appl. Stat.)

- By using the properties of beta prime distribution, we have

$$E_1(W \mid x, y) = 1 + E_1(O \mid x, y) = \frac{1}{m(x, y)};$$

$$E(W \mid x, y) = 1 + E(O \mid x, y) = \frac{\phi - 1}{m(x, y)\phi - 1}$$

Proposed Adaptive Estimator for (c) $f(y \mid x; \theta)$

1. Assume a parametric model on $m(x, y)$, e.g.

$$W^{-1} \mid (x, y) \sim \mathrm{Beta}(m\phi, (1-m)\phi)$$

$$m(x, y; \beta) = \frac{\exp(\beta_0 + \beta_1 x + \beta_2 y)}{1 + \exp(\beta_0 + \beta_1 x + \beta_2 y)}$$

2. Estimate (ϕ, β) by ML based on the likelihood on $f_1(o \mid x, y)$ (beta prime distribution)

3. Let $\bar{\pi}(x, y; \hat{\beta}, \hat{\phi}) = \dfrac{m(x, y; \hat{\beta})\hat{\phi} - 1}{\hat{\phi} - 1}$

4. Solve the following estimating equation w.r.t. θ (say, $\hat{\theta}_{\mathrm{eff}}$):

$$S_{\mathrm{eff}}(\theta, \hat{\alpha}) := \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i W_i \hat{D}^*_{\mathrm{eff}}(X_i, Y_i; \theta, \hat{\alpha}) + (1 - \delta_i W_i)\hat{c}^*_{\mathrm{eff}}(\hat{\alpha})\right\},$$

where $\hat{\alpha} = (\hat{\beta}^\top, \hat{\phi}, \hat{\theta}_{\mathrm{HT}}^\top)^\top$ and $\hat{D}^*_{\mathrm{eff}}(\theta, \hat{\alpha})$ and $\hat{c}^*_{\mathrm{eff}}(\hat{\alpha})$ are obtained by replacing the unknown functions with the estimated ones.

Efficient Score When N is Unknown

- The efficient score when N is unknown is obtained by letting c^*_{eff} be 0

- For example, if the regression model is of our interest,

$$S_{\text{eff}} = \delta W D^*_{\text{eff}} + (1 - \delta W) \times 0,$$

where $D^*_{\text{eff}} = A^*_{\text{eff}}(X)\{Y - \mu(X;\theta)\}$ and $A^*_{\text{eff}}(x) = \dfrac{1}{E(W\varepsilon^2 \mid x)} \dfrac{\partial}{\partial \theta} \mu(x;\theta)$

This is exactly same as the result of Kim and Skinner (2013, Biometrika)

Summary of Efficient Score

$$S_{\text{eff}} = \delta W D^*_{\text{eff}} + (1 - \delta W) c^*_{\text{eff}}$$

Information		Target parameter θ			
N	X	Z-estimator	Regression	Outcome	
Known	Partial	✔	✔	✔	→ c^*_{eff}: constant
Unknown	Partial	✔	Kim and Skinner (2013, Biometrika)	✔	→ $c^*_{\text{eff}} \equiv 0$
Known	Complete	✔	✔	✔	→ c^*_{eff}: function of x

I focused on this part in this talk

Extension to Strata Mixed Model

- If the sampling mechanism is stratified sampling, it would be reasonable to assume that W^{-1} follows a beta distribution in each stratum h, e.g.

$$W^{-1} \mid (x, y, H = h) \sim \mathrm{Beta}(m_h(x, y)\phi_h, \{1 - m_h(x, y)\}\phi_h)$$

- However, we need an additional model on $P(H = h \mid x, y)$ such as the multinomial logit model
    - The parameters are computable by the EM algorithm

- We can compute $E(W \mid x, y)$ and $E_1(W \mid x, y)$ analogously

Large Sample Property of Proposed Estimator

Theorem 2.

Under some regularity conditions, $\hat{\theta}_{\text{eff}}$ has the following two properties:

(i) if all the working models are correct, $\hat{\theta}_{\text{eff}}$ attains the semiparametric efficiency bound;

(ii) even if all the working models are misspecified, $\hat{\theta}_{\text{eff}}$ has consistency and asymptotic normality. Let α be the parameter of the working models and $\tilde{\alpha}$ be the probability limit of α. Then, the asymptotic variance of $\hat{\theta}_{\text{eff}}$ is given by

$$V(\hat{\theta}_{\text{eff}}) = E\left\{\frac{\partial S_{\text{eff}}(\tilde{\alpha}, \theta^*)}{\partial \theta^\top}\right\}^{-1} E(S_{\text{eff}}^{\otimes 2}(\tilde{\alpha}, \theta^*)) E\left\{\frac{\partial S_{\text{eff}}(\tilde{\alpha}, \theta^*)}{\partial \theta^\top}\right\}^{-1}$$

- Property (ii) insists robustness of $\hat{\theta}_{\text{eff}}$ for model misspecification
- The asymptotic variance is independent of that of $\tilde{\alpha}$
- Model on $m(x, y)$ can be nonparametric

Semi- and Non-parametric Working Model

- Semparametric working model
  - We may keep assuming a beta regression, but with a nonparametric model on $m(x, y)$
- Nonparametric working model
  - By nonparametrically estimating $E_1(W \mid x, y)$ and $E_1(W^2 \mid x, y)$, we can estimate

$$\bar{\pi}(x, y) = \frac{1}{E(W \mid x, y)} = \frac{E_1(W \mid x, y)}{E_1(W^2 \mid x, y)}.$$

- We believe that we can show that estimators with above working models are also valid, but we have not finished to prove yet.

Contents

Numerical Study　Setup –1/2–

- Setup:
  - $X \sim N\left(0, \frac{1}{\sqrt{2}^2}\right)$, $Z \sim N\left(0, \frac{1}{\sqrt{2}^2}\right)$, $Y \mid (x, z) \sim N\left(x - z, \frac{1}{\sqrt{2}^2}\right)$
  - $W^{-1} \sim \mathrm{Beta}(m(x, y)\phi, \{1 - m(x, y)\}\phi)$ and $\phi = 2{,}500$
  - $\delta \mid w \sim \mathrm{Binom}(w^{-1})$
  - $N = 5{,}000$: size of a population
  - $B = 1{,}000$: number of iteration

- Model: $Y \mid x \sim N(a + bx, \sigma^2)$
  Target parameter $\theta = (a, b, \sigma^2)^\top$; True value $\theta^* = (0, 1, 1)^\top$

Numerical Study　Setup –2/2–

- Scenarios for $\mu(x, y)$: $n \approx 200$ in all cases

  S1. (No dependency) $\mathrm{logit}\{m(x,y)\} = -3.2$

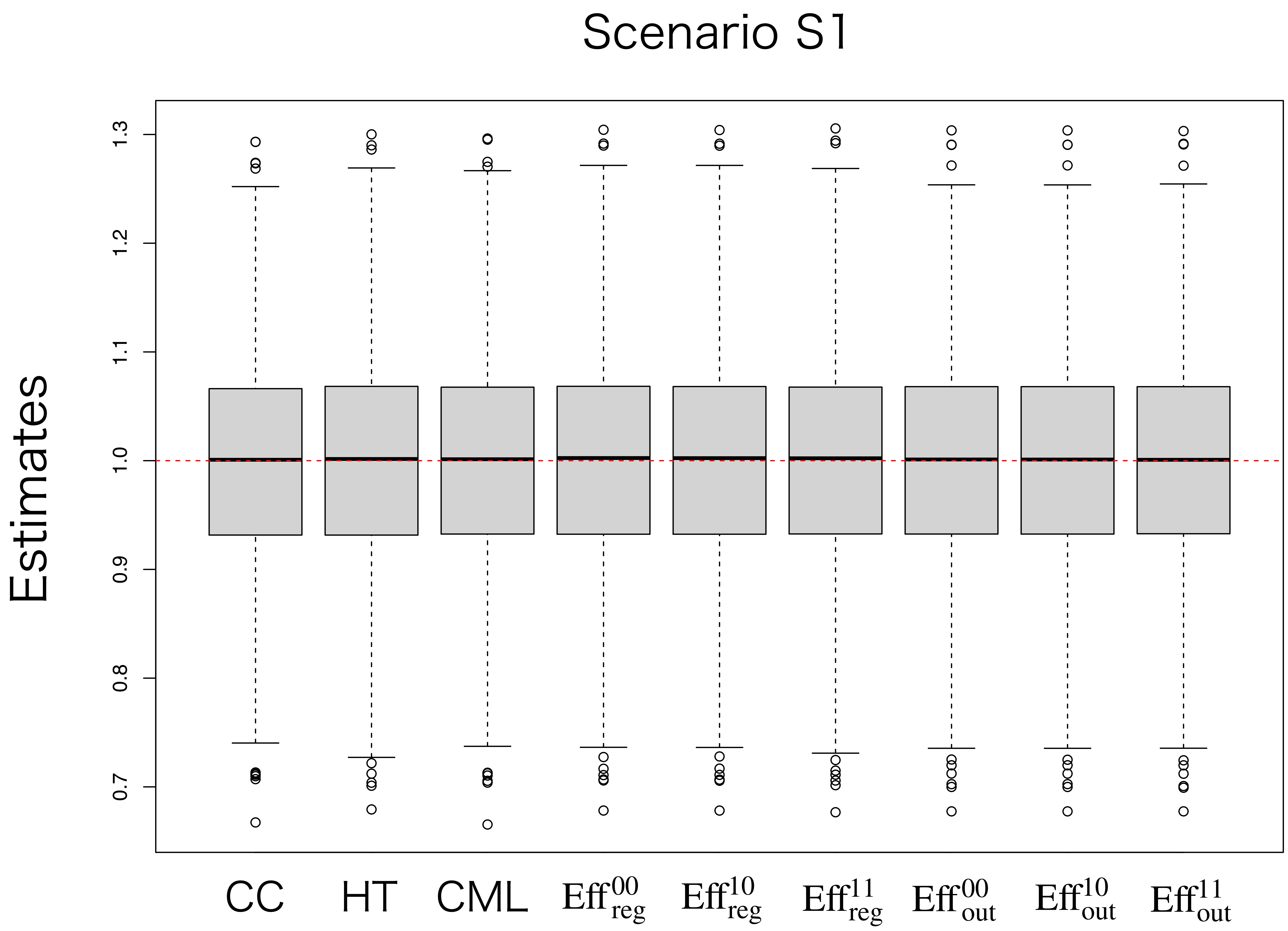
  S2. (Dependency) $\mathrm{logit}\{m(x,y)\} = -3.4 + 0.3x + 0.5y$

  S3. (Misspecified) $\mathrm{logit}\{m(x,y)\} = -3.4 + 0.25x + 0.25z + 0.1y^2$

- Parametric model on $m(x, y)$: $\mathrm{logit}\{m(x,y)\} = \alpha_0 + \alpha_1 x + \alpha_2 y$

- Methods:
  - CC: complete case analysis ($w_i \equiv 1$)
  - HT: Horvitz-Thompson type estimator
  - CML: Conditional Maximum Likelihood
  - $\mathrm{Eff}_{\mathrm{reg}}$, $\mathrm{Eff}_{\mathrm{out}}$: Proposed estimator
    - reg: adaptive estimator for **reg**ression model
    - out: adaptive estimator for **out**come model

Boxplot for $\hat{b}$ in Scenario S1

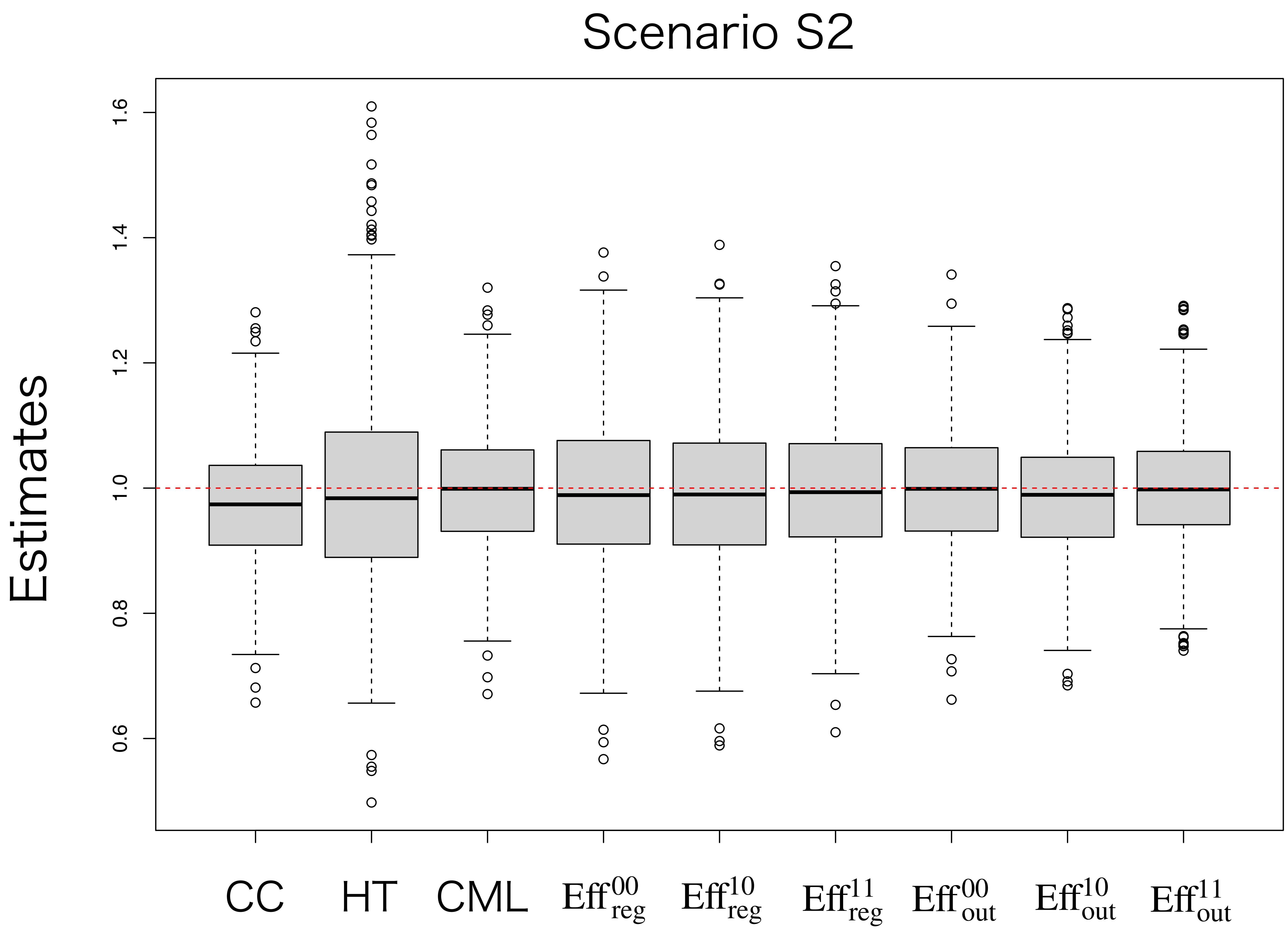


Eff^{ij}

i: N is known? (1/0)

j: X is completely observed?(1/0)

Boxplot for $\hat{b}$ in Scenario S2

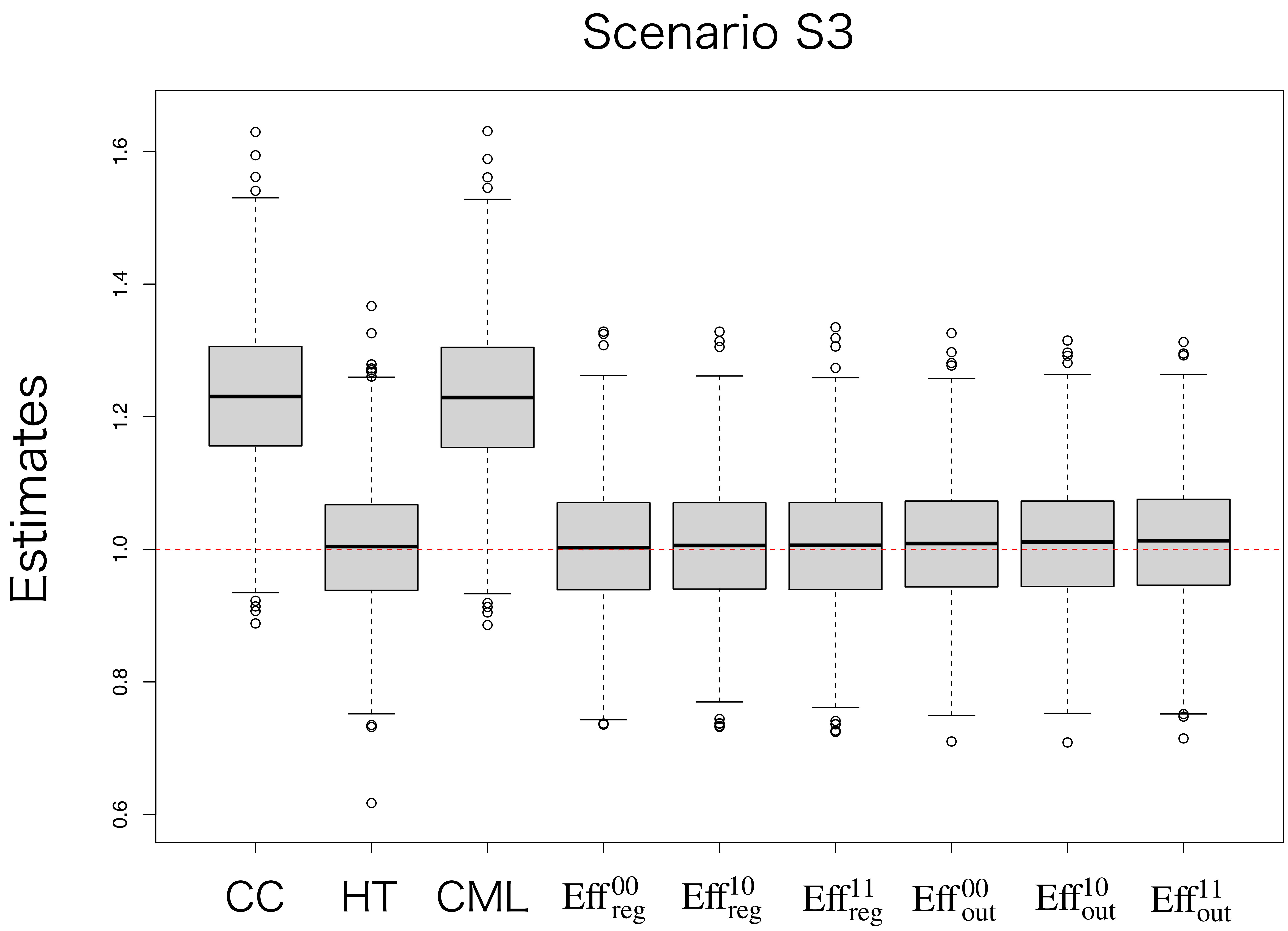


Eff^{ij}

i: N is known? (1/0)

j: X is completely observed?(1/0)

Boxplot for $\hat{b}$ in Scenario S3



Eff^{ij}

i: N is known? (1/0)

j: X is completely observed?(1/0)

Contents

- Introduction
- Proposed Estimator
- Simulation
- Real Data Analysis

Example: The Canadian Workplace and Employee Survey

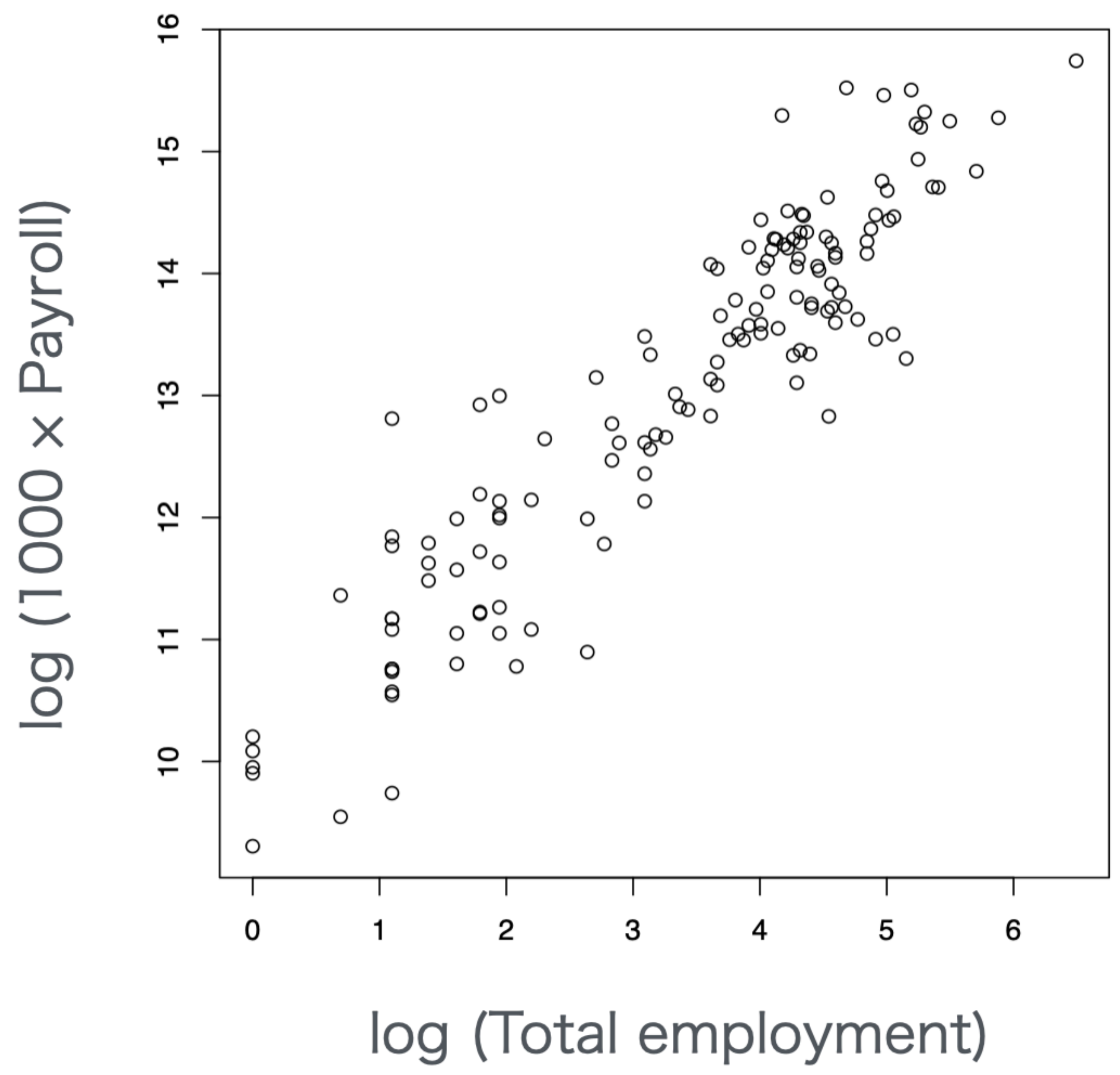
- We want to know the relationship between Payroll (Y) and total Employment (X)

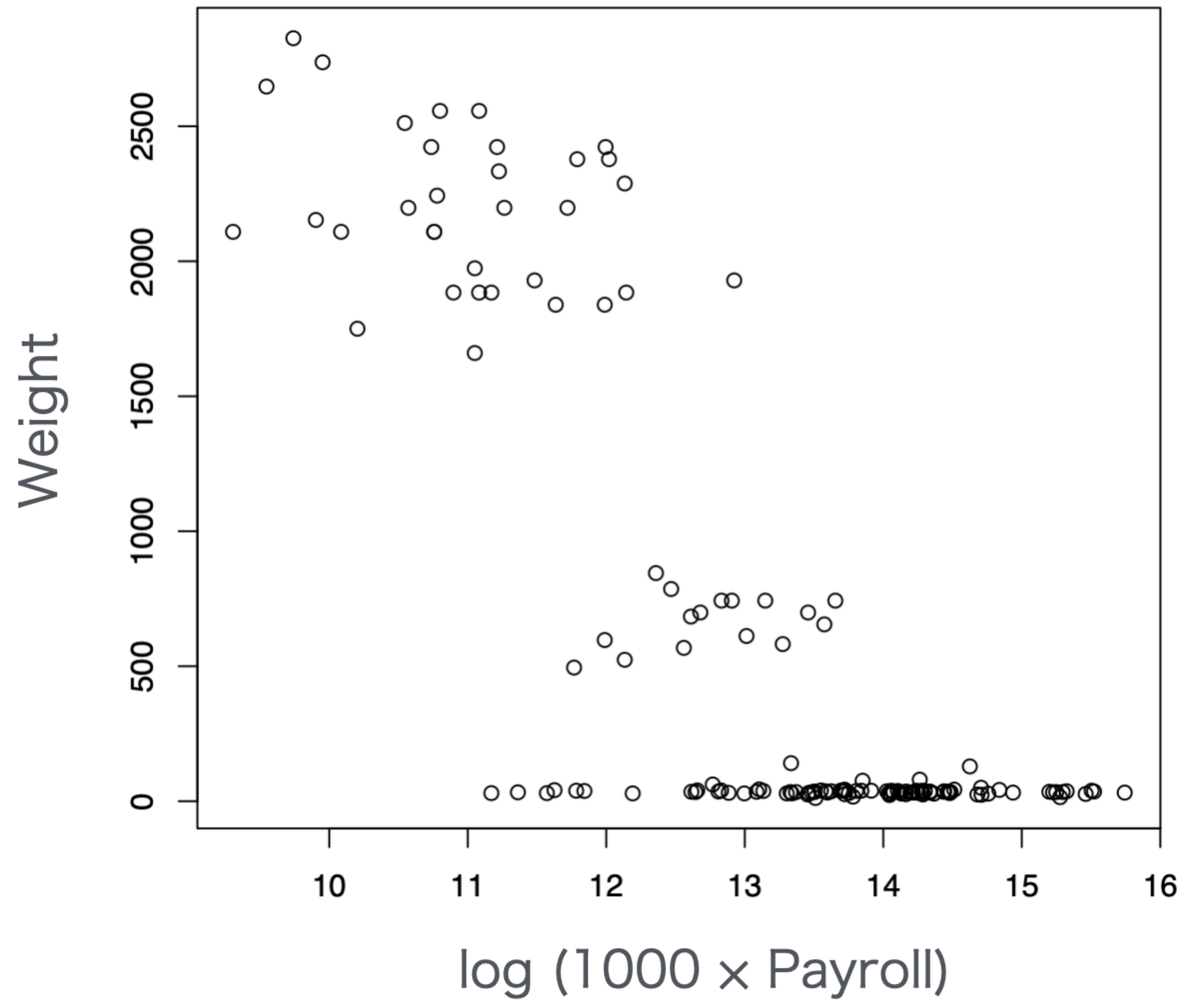
- Size of population (N): 2029 workplaces

- Sampled size (n): 142 workplaces
  - Stratified sampling (3 strata)
    + simple random sampling
      with nonresponse adjustment

- Model:

$$Y \mid X = x \sim N(a + bx, \sigma^2), \quad \theta = (a, b, \sigma^2)$$





Working model

- Mean function of $W^{-1} \mid (x, y, H = h)$:

$$m_h(x, y) = \beta_h \ (h = 1,2,3), \text{ where } 0 < \beta_h < 1$$

- Mixture probability of strata:

$$P(H = h \mid x, y; \gamma)$$
$$= \frac{I(h=1) + I(h=2)\exp(\gamma_0^{(1)} + \gamma_1^{(1)} y) + I(h=3)\exp(\gamma_0^{(2)} + \gamma_1^{(2)} y)}{1 + \exp(\gamma_0^{(1)} + \gamma_1^{(1)} y) + \exp(\gamma_0^{(2)} + \gamma_1^{(2)} y)}$$

Estimates for The Canadian Workplace and Employee Survey

Parameter	Methods		
	CC	HT	$\text{Eff}^{11}_{\text{out}}$
$\hat{a}$	13.082	12.889	12.898
	(0.0477)	(0.1140)	(0.0671)
$\hat{b}$	0.907	0.931	0.931
	(0.0327)	(0.0532)	(0.0370)
$\hat{\sigma}^2$	0.316	0.299	0.295
	(0.0428)	(0.2030)	(0.0666)

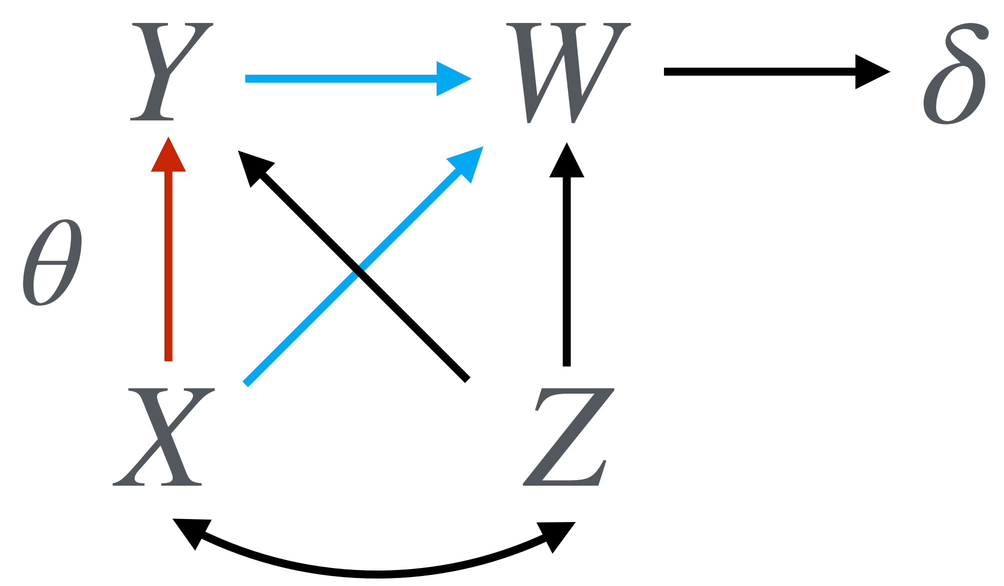
← **estimate**

← **estimated SE**

- Estimates of HT and Eff are very similar
- However, the standard error of Eff is much smaller than HT

Conclusion and Future Works

- In survey sampling, weights are known,
  but the information had NOT been fully utilized



- Our proposed estimator…
  - attains the semiparametric efficiency bound if the working models are correctly specified
  - is robust for misspecification of working models.

- Extension to nonparametric models of the working model

Thank you