Kernel Regression Utilizing External Information as Constraints

Chi-Shian Dai

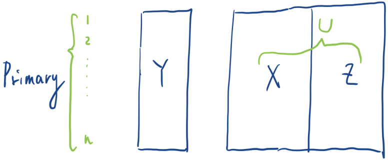
ENAR 2024

3/10/2024

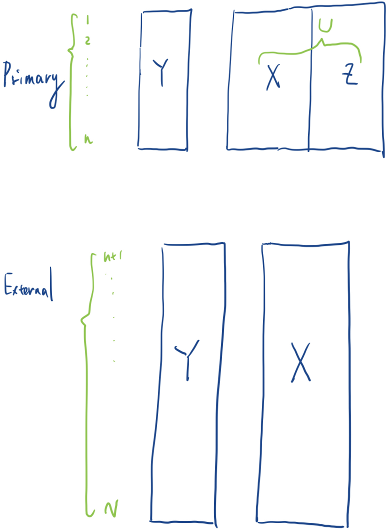
Agenda

1. Motivation
2. Summary-level External Data
3. Individual-level External Data
4. Real Data
5. Summary

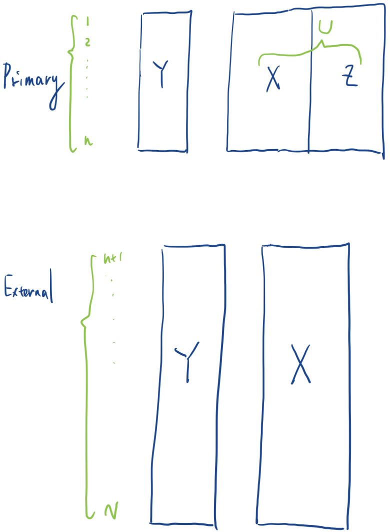
Motivation



Motivation

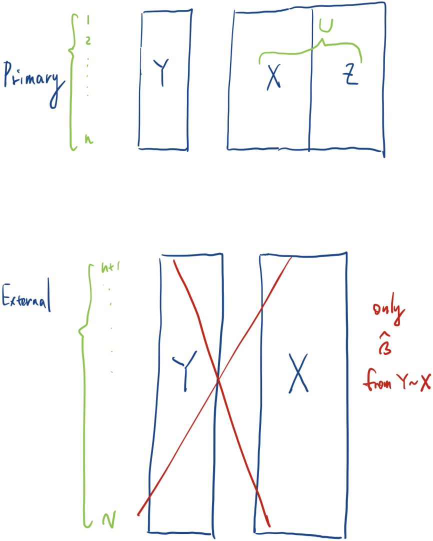


Motivation



- External data: historical data, populational census, preclinical trials.
- Primary data : Extra covariates $\boldsymbol{Z}$.
- Sample size of the external data $N - n$ is much larger than the primary data.
- Advantage: Information about Y and X.
- Challenge: Safety of the external data.

Motivation



- Summary-level information
- Only observe $\widehat{\beta}$, a coefficient from a linear partial model $Y \sim X$.
- Not a Meta-analysis.
- Not a traditional missing data problem.
- Goal: Use $\boldsymbol{X}$ and $\boldsymbol{Z}$ to predict Y. Find the predictor $\mu(\boldsymbol{X}, \boldsymbol{Z})$ for Y.

Idea for utilizing external information

- What does the summary-level external information $\widehat{\beta}$ tell us?

Idea for utilizing external information

- What does the summary-level external information $\widehat{\beta}$ tell us?
- It tells us how the outcome Y responds to the partial covariate $\boldsymbol{X}$.

Idea for utilizing external information

- What does the summary-level external information $\widehat{\beta}$ tell us?
- It tells us how the outcome Y responds to the partial covariate $\boldsymbol{X}$.
- Fit a regression of Y against $\boldsymbol{X}$, we should get $\widehat{\beta}$.

Idea for utilizing external information

- What does the summary-level external information $\widehat{\beta}$ tell us?
- It tells us how the outcome Y responds to the partial covariate $\boldsymbol{X}$.
- Fit a regression of Y against $\boldsymbol{X}$, we should get $\widehat{\beta}$.
- The way that predictor μ responds to the $\boldsymbol{X}$ should be the same.

Idea for utilizing external information

- What does the summary-level external information $\widehat{\beta}$ tell us?
- It tells us how the outcome Y responds to the partial covariate $\boldsymbol{X}$.
- Fit a regression of Y against $\boldsymbol{X}$, we should get $\widehat{\beta}$.
- The way that predictor μ responds to the $\boldsymbol{X}$ should be the same.
- Fit a regression of the predictor μ against $\boldsymbol{X}$, should get the same $\widehat{\beta}$.

Idea for utilizing external information

- What does the summary-level external information $\widehat{\boldsymbol{\beta}}$ tell us?
- It tells us how the outcome Y responds to the partial covariate $\boldsymbol{X}$.
- Fit a regression of Y against $\boldsymbol{X}$, we should get $\widehat{\boldsymbol{\beta}}$.
- The way that predictor μ responds to the $\boldsymbol{X}$ should be the same.
- Fit a regression of the predictor μ against $\boldsymbol{X}$, should get the same $\widehat{\boldsymbol{\beta}}$.
- Constrained optimization:

$$
\begin{aligned}
\widehat{\mu}_1, \ldots \widehat{\mu}_n = \quad & \arg\min_{\boldsymbol{\mu}} \quad \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\
& \text{subject to} \quad \text{the summary-level external information,} \\
& \qquad\qquad\quad \boldsymbol{\mu} \sim \boldsymbol{X} \text{ get } \widehat{\boldsymbol{\beta}}.
\end{aligned}
$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$.

Idea for utilizing external information

- What does the summary-level external information $\widehat{\boldsymbol{\beta}}$ tell us?
- It tells us how the outcome Y responds to the partial covariate $\boldsymbol{X}$.
- Fit a regression of Y against $\boldsymbol{X}$, we should get $\widehat{\boldsymbol{\beta}}$.
- The way that predictor μ responds to the $\boldsymbol{X}$ should be the same.
- Fit a regression of the predictor μ against $\boldsymbol{X}$, should get the same $\widehat{\boldsymbol{\beta}}$.
- Constrained optimization:

$$
\begin{aligned}
\widehat{\mu}_1, \ldots \widehat{\mu}_n = \quad & \arg\min_{\boldsymbol{\mu}} && \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\
& \text{subject to} && \text{the summary-level external information,} \\
& && \boldsymbol{\mu} \sim \boldsymbol{X} \text{ get } \widehat{\boldsymbol{\beta}}.
\end{aligned}
$$

  where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$.
- We implement this idea by kernel regression. We call it constrained kernel regression (CK).

Loss function based on Kernel Regression

- The loss function for a single point $\boldsymbol{u}$.

$$\widehat{\mu}(\boldsymbol{u}) = \arg\min_{\mu} \sum_{j=1}^{n} \kappa_l(\boldsymbol{u} - \boldsymbol{U}_j)(Y_j - \mu)^2,$$

- Kernel regression estimate for $\mu_i = \mu(\boldsymbol{U}_i)$:

$$\begin{aligned} \widehat{\mu}_1, \ldots, \widehat{\mu}_n &= \arg\min_{\mu_1,\ldots,\mu_n} \frac{1}{n} \sum_{i=1}^{n} \frac{\boxed{\sum_{j=1}^{n} \kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)(Y_j - \mu_i)^2}}{\sum_{k=1}^{n} \kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_k)} \\ &\approx \arg\min_{\mu} E[(Y - \mu(\boldsymbol{U}))^2] \end{aligned}$$

Summay-level external information as constraints

Hence, the constrained optimization can be

$$
\begin{aligned}
\widehat{\mu}_1, \ldots, \widehat{\mu}_n = \arg \min_{\mu_1, \ldots, \mu_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)(Y_j - \mu_i)^2}{\sum_{k=1}^{n} \kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_k)} \\
\text{subject to} \quad \sum_{i=1}^{n} (\mu_i - \boldsymbol{X}_i^\top \widehat{\beta}) \boldsymbol{X}_i = 0.
\end{aligned}
\tag{1}
$$

(1) is a quadratic programming. Hence, it can be solved by Lagrange multiplier.

Assumptions for the constraints

Assumption 1

- Let $\boldsymbol{\beta}_1$ be the asymptotic coefficient from $Y \sim \boldsymbol{X}$ in the primary data.
- Let $\boldsymbol{\beta}_0$ be the asymptotic coefficient from $Y \sim \boldsymbol{X}$ in the external data.

Assume that $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1$.

Test the assumption $\beta_0 = \beta_1$

- From the external summary information, get $\widehat{\beta}$ and $sd(\widehat{\beta})$ for β_0.
- Use the primary data to estimate β_1 and its standard deviation.
- Run a test for $\beta_0 = \beta_1$.

Simulation 1

Goal: Show that CK is better than standard kernel regression.

- X, Z are normal distribution with variance 1, covariance 0.5.
- $Y = \mu(X, Z) + \epsilon$; $\epsilon \sim N(0, \sigma^2)$
- $\mu = X - Z^2$.
- Goal: estimate μ.
- Primary sample size : 200 (Observed Y, X, Z)
- External sample size : 1000 (Observed Y, X)

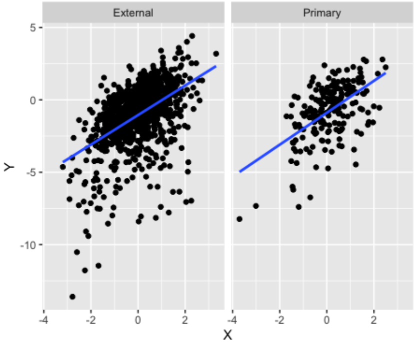
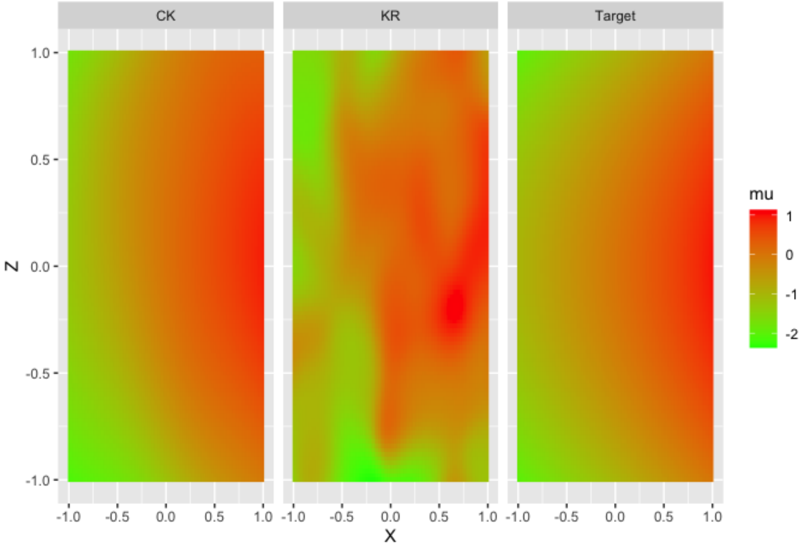


Figure: Scatter plot for Y, X.

Simulation 1:Prediction result



Simulation 2

What if the assumption is wrong? ($\beta_0 \neq \beta_1$)

- Generate normal distribution X, Z with variance 1, covariance 0.5, sample size 1200.
- Generate $Y = X - Z^3 + XZ + \epsilon$
- Generate $R = \{1, 0\}$ to indicate the data set.

$$P(R = 1|X) = \{1+\exp(1.5+X^2)\}^{-1}.$$

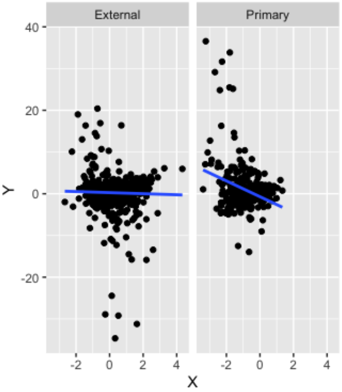
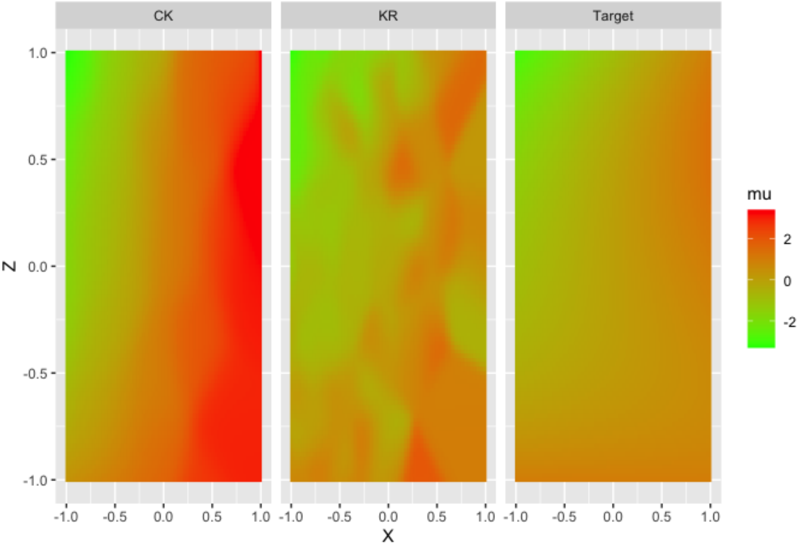


Figure: Scatter plot for Y, X.

Simulation 2: Prediction result



Individual-level external information

- The constraints for summary-level information:

$$0 = \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \boxed{\boldsymbol{X}_i^\top \widehat{\beta}})\boldsymbol{X}_i.$$

- $\boxed{\boldsymbol{X}^\top \widehat{\beta}}$ is a predictor function for Y given $\boldsymbol{X}$.
- Consider the robust constraints:

$$0 = \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \boxed{\widehat{h}(\boldsymbol{X}_i)})\boldsymbol{X}_i.$$

- $\boxed{\widehat{h}(\boldsymbol{X})}$ is a predictor function for $E(Y|\boldsymbol{X})$ based on a non-parametric method.
- We need individual-level external data to construct $\boxed{\widehat{h}(\boldsymbol{X})}$.

Assumptions for robust constraints

Assumption 2

$E(Y|\boldsymbol{X}, \text{Primary Data}) = E(Y|\boldsymbol{X}, \text{External Data}).$

CK for individual-level external information

- Solve the contrained optimization problem.

$$\widehat{\mu}_1 \ldots \widehat{\mu}_n = \arg\min_{\mu_1,\ldots,\mu_n} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)(Y_j - \mu_i)^2}{\sum_{k=1}^{n} \kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_k)}$$

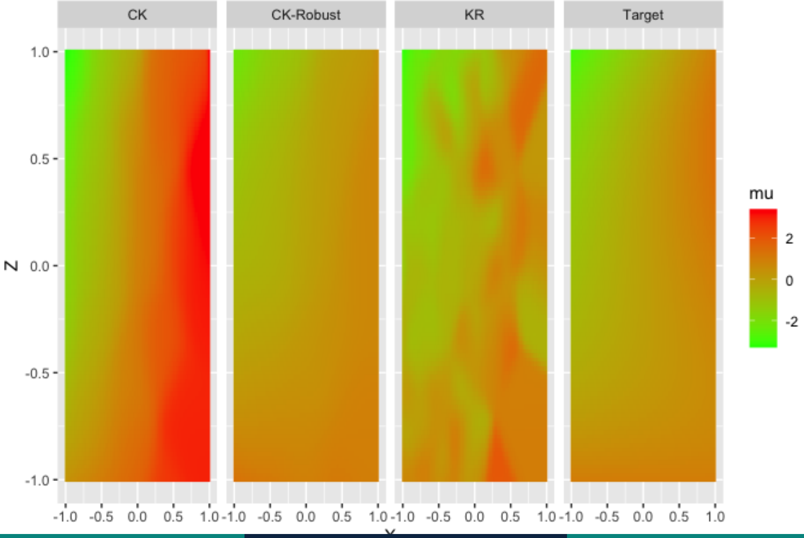
$$\text{subject to} \quad \sum_{i=1}^{n}(\mu_i - \widehat{h}(\boldsymbol{X}_i))\boldsymbol{X}_i = 0.$$

- For arbitrary $\boldsymbol{u} \in \mathcal{U}$, we apply additional kernel regression by replacing Y_i with $\widehat{\mu}_i$.

$$\widehat{\mu}_{CK}(\boldsymbol{u}) = \sum_{i=1}^{n} \widehat{\mu}_i \kappa_b(\boldsymbol{u} - \boldsymbol{U}_i) \Big/ \sum_{i=1}^{n} \kappa_b(\boldsymbol{u} - \boldsymbol{U}_i). \tag{2}$$

Simulation 2: Prediction result

Goal: Show that CK-Robust is robust.

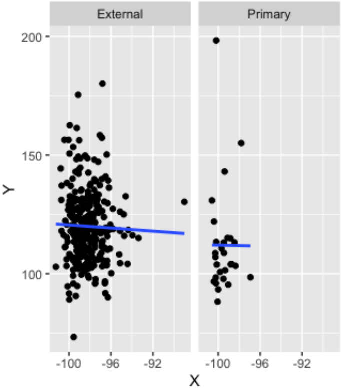


Real Data Application

- Goal: Predict the blood pressure(systolic).
- External Data:
  - Sample size is 54,060.
  - Covariates: heart rate, respiratory rate, blood oxygen saturation.
- Primary Data:
  - Sample size is 32.
  - Extra Covariates: oxygen concentration from the patient's mouth.

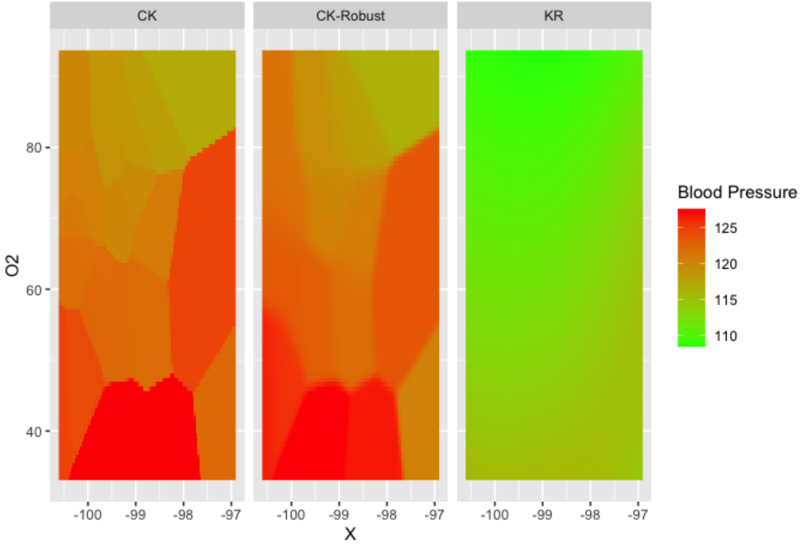
Real Data Application

- Apply sufficient dimention reduction algorithm (SAVE) to find a linear combination of common covariates as a one-dimensional covariate X.
- External Data: Blood Pressure, X.
- Primary Data: Blood Pressure, X, Oxygen Concentration.
- Use X and Oxygen Concentration to predict Blood Pressure.
- Test for $\beta_0 = \beta_1$



Coefficient	p-value
intercept	0.93
slope	0.94

Real Data Application



Summary

- Primary data: $Y, \boldsymbol{X}, \boldsymbol{Z}$.
- External information: how Y react to $\boldsymbol{X}$.
- The way μ reacts to $\boldsymbol{X}$ should be the same as how Y react to $\boldsymbol{X}$.
- Constrained optimization problem:

$$\begin{aligned} \arg\min_{\boldsymbol{\mu}} \quad & \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\ \text{subject to} \quad & \text{the external information,} \end{aligned}$$

- Use kernel regression to define the Loss.
- Summary-level external information: check assumption $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_1$.
- Individual-level external information: robust constraints.

Thank you

Thank you

References

[1] Nilanjan Chatterjee, Yi Hau Chen, Paige Maas, and Raymond J Carroll. Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, 111(513):107–117, 2016.

Loss functions

$$
\begin{aligned}
\operatorname*{arg\,min}_{\boldsymbol{\mu}} \quad & \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\
\text{subject to} \quad & \text{the external information,}
\end{aligned}
$$

Loss functions

$$\begin{aligned} \arg\min_{\boldsymbol{\mu}} \quad & \boxed{\mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu})} \\ \text{subject to} \quad & \text{the external information,} \end{aligned}$$

Loss functions

$$\begin{aligned} \arg\min_{\boldsymbol{\mu}} \quad & \boxed{\mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu})} \\ \text{subject to} \quad & \text{the external information,} \end{aligned}$$

- Paramatric models: Likelihood function.
- Semi-paramatric models: Psudo Likelihood function. Chatterjee et al. [1]
- Non-paramatric models:
  - Sequential basis: Fouriour Series
  - Neural network

External information

$$
\begin{aligned}
\arg\min_{\boldsymbol{\mu}} \quad & \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\
\text{subject to} \quad & \boxed{\text{the external information}},
\end{aligned}
$$

- Linear models.
- Likelihood models.
- Generalized estimating equation.

Assumptions for robust constraints

Assumption 2

$E(Y|\boldsymbol{X}, \text{Primary Data}) = E(Y|\boldsymbol{X}, \text{External Data})$

Assumption 3

$Y \mid X, \text{Primary Data} \overset{d}{=} Y \mid X, \text{External Data}.$

Assumption 3'

$Y \perp\!\!\!\perp R \mid \boldsymbol{X}$, where R indicates the data set.

Back

Loss functions

$$\begin{aligned} \operatorname*{arg\,min}_{\boldsymbol{\mu}} \quad & \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\ \text{subject to} \quad & \text{the external information,} \end{aligned}$$

Loss functions

$$\begin{aligned} \arg\min_{\boldsymbol{\mu}} \quad & \boxed{\mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu})} \\ \text{subject to} \quad & \text{the external information,} \end{aligned}$$

Loss functions

$$\begin{aligned} \arg\min_{\boldsymbol{\mu}} \quad & \boxed{\mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu})} \\ \text{subject to} \quad & \text{the external information,} \end{aligned}$$

- Paramatric models: Likelihood function.
- Semi-paramatric models: Psudo Likelihood function. Chatterjee et al. [1]
- Non-paramatric models:
  - Kernel regression-likelihood function.
  - Sequential basis: Fouriour Series
  - Neural network

External information

$$\begin{aligned} \arg\min_{\boldsymbol{\mu}} \quad & \mathrm{Loss}(\boldsymbol{Y}, \boldsymbol{\mu}) \\ \text{subject to} \quad & \boxed{\text{the external information}}, \end{aligned}$$

- Linear models.
- Likelihood models.
- Generalized estimating equation.
- Multiple external information: Multiple constraints.
- Chanllenge: Multiple external information from the same model.

Theorem

Theorem 1

Assume regularity conditions. As $n \to \infty$, we have the following asymptotic theorem.

$$\sqrt{nb^p}\{\widehat{\boldsymbol{\mu}}_{CK}(\boldsymbol{u}) - \mu(\boldsymbol{u})\} \to N\big(B_{CK}(\boldsymbol{u}), V_{CK}(\boldsymbol{u})\big) \text{ in distribution,} \tag{3}$$

$$B_{CK}(\boldsymbol{u}) = c^{1/2}[(1+\gamma^2)A(\boldsymbol{u}) - \gamma^2 \boldsymbol{x}^\top \boldsymbol{\Sigma}_X^{-1} E\{\boldsymbol{X} A(\boldsymbol{U})\}],$$

$$V_{CK}(\boldsymbol{u}) = V_{DK}(\boldsymbol{u}),$$

$$A(\boldsymbol{u}) = \int \kappa(\boldsymbol{v}) \left\{ \frac{1}{2} \boldsymbol{v}^\top \nabla^2 \mu(\boldsymbol{u}) \boldsymbol{v} + \nabla \mu(\boldsymbol{u})^T \boldsymbol{v} \boldsymbol{v}^T \nabla \log f_U(\boldsymbol{u}) \right\} d\boldsymbol{v}, \tag{4}$$

and f_U is the density of $\boldsymbol{U}$.

Asymptotic Mean Integrated Square Error

- $\mathrm{AMISE}(\widehat{\mu}_{CK}) = E[\{B_{CK}(\boldsymbol{U})\}^2 + V_{CK}(\boldsymbol{U})]$

Theorem 2

Assume regularity conditions. $\mathrm{AMISE}(\widehat{\mu}_{CK}) < \mathrm{AMISE}(\widehat{\mu}_{K})$

Estimate $h_1 \neq h_0$

- If $h_1 \neq h_0$, find a weighting function $w(\boldsymbol{x}, y)$ on the external data such that

$$E\{Yw(\boldsymbol{X}, Y) \mid \boldsymbol{X}, \text{External Data}\} = E\{Y \mid \boldsymbol{X}, \text{Primary Data}\},$$

$$w(\boldsymbol{X}, Y) = f(Y \mid \boldsymbol{X}, \text{Primary Data}) / f(Y \mid \boldsymbol{X}, \text{External Data}),$$

$$\widehat{h}_1(\boldsymbol{x}) = \frac{\sum\limits_{i \in \text{Primary}} Y_i \kappa_b(\boldsymbol{x} - \boldsymbol{X}_i) + \sum\limits_{i \in \text{External}} w(\boldsymbol{X}_i, Y_i) Y_i \kappa_b(\boldsymbol{x} - \boldsymbol{X}_i)}{\sum_i \kappa_b(\boldsymbol{x} - \boldsymbol{X}_i)}$$

Estimate $w(\boldsymbol{X}, Y)$: Non-parametric method

- Primary Data: $R = 1$; External Data: $R = 0$;
- $$\widehat{f}(y|\boldsymbol{X} = \boldsymbol{x}, R = 1) = \sum_{i=1}^{n} \widetilde{\kappa}_{\widetilde{b}}(y - Y_i, \boldsymbol{x} - \boldsymbol{X}_i) \Big/ \sum_{i=1}^{n} \bar{\kappa}_{\bar{b}}(\boldsymbol{x} - \boldsymbol{X}_i),$$
  $$\widehat{f}(y|\boldsymbol{X} = \boldsymbol{x}, R = 0) = \sum_{i=n+1}^{N} \widetilde{\kappa}_{\widetilde{b}}(y - Y_i, \boldsymbol{x} - \boldsymbol{X}_i) \Big/ \sum_{i=n+1}^{N} \bar{\kappa}_{\bar{b}}(\boldsymbol{u} - \boldsymbol{U}_i),$$
- $$\widehat{w}(\boldsymbol{x}, Y) = \widehat{f}(y|\boldsymbol{X} = \boldsymbol{x}, R = 1) / \widehat{f}(y|\boldsymbol{X} = \boldsymbol{x}, R = 0).$$

Estimate $w(\boldsymbol{X}, Y)$: Semi-parametric method

- Assume a logistic model with unknown $\alpha(\boldsymbol{X})$ and γ.

$$\frac{P(R=0 \mid \boldsymbol{X}, Y)}{P(R=1 \mid \boldsymbol{X}, Y)} = \exp\{\alpha(\boldsymbol{X}) + \gamma Y\}$$

- From Bayes' rule:

$$w_\gamma(\boldsymbol{x}, Y) = \frac{f(Y|\boldsymbol{x}, R=1)}{f(Y|\boldsymbol{x}, R=0)} = e^{-\gamma Y} E(e^{\gamma Y} \mid \boldsymbol{X} = \boldsymbol{x}, R=1).$$

- Construct the predictor for $E(Y \mid \boldsymbol{X}, R=1)$ via $w_\gamma(\boldsymbol{x}, Y)$:

$$\begin{aligned} h_1(\boldsymbol{x}, \gamma) &= E\{Y w_\gamma(\boldsymbol{x}, Y) \mid \boldsymbol{X} = \boldsymbol{x}, R=0\} \\ &= E(Y e^{-\gamma Y} \mid \boldsymbol{X} = \boldsymbol{x}, R=0) E(e^{\gamma Y} \mid \boldsymbol{X} = \boldsymbol{x}, R=1) \end{aligned}$$

- Find the best $h_1(\boldsymbol{x}, \gamma)$ via MSE:

- $$\widehat{\gamma} = \arg\min_{\gamma} \frac{1}{N} \sum_{i=1}^{N} R_i \{Y_i - \widehat{h}_1(\boldsymbol{X}_i, \gamma)\}^2,$$
$$\widehat{w}_{\widehat{\gamma}}(\boldsymbol{x}, Y) = e^{-\widehat{\gamma} Y} \widehat{E}(e^{\widehat{\gamma} Y} \mid \boldsymbol{X} = \boldsymbol{x}, R=1).$$

Summarize all the methods

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\mu_1,\ldots,\mu_n} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)(Y_j - \mu_i)^2}{\sum_{k=1}^{n} \kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_k)}$$

subject to $$\sum_{i=1}^{n}(\mu_i - \boldsymbol{X}_i^\top \widehat{\beta})\boldsymbol{X}_i = 0.$$

Summarize all the methods

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\mu_1,\dots,\mu_n} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)(Y_j - \mu_i)^2}{\sum_{k=1}^{n}\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_k)}$$

$$\boxed{\text{subject to} \quad \sum_{i=1}^{n}(\mu_i - \widehat{h}_1(\boldsymbol{X}_i))\boldsymbol{X}_i = 0.}$$

Summarize all the methods

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\mu_1,\ldots,\mu_n} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)(Y_j - \mu_i)^2}{\sum_{k=1}^{n} \kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_k)}$$

subject to $\sum_{i=1}^{n}(\mu_i - \widehat{h}_1(\boldsymbol{X}_i))\boldsymbol{X}_i = 0.$

Notation	Assumption
$\widehat{\mu}^{C1}$	$h_1 = h_0$
$\widehat{\mu}^{C2}$	No assumption
$\widehat{\mu}^{C3}$	Put a model on $w(\boldsymbol{X}, Y)$
$\widehat{\mu}_K$	No assumption, no constraint.

Bayes' Rule

$$\frac{f(Y|\boldsymbol{x}, R=1)}{f(Y|\boldsymbol{x}, R=0)} = \frac{P(R=1|\boldsymbol{X}=\boldsymbol{x}, Y)}{P(R=0|\boldsymbol{X}=\boldsymbol{x}, Y)} \frac{P(R=0|\boldsymbol{X}=\boldsymbol{x})}{P(R=1|\boldsymbol{X}=\boldsymbol{x})}, \tag{5}$$

Back

Conditions for Theorem 1

(A1) The response Y has a finite $\mathrm{E}|Y|^s$ with $s > 2 + p/2$ and $\boldsymbol{\Sigma}_g = \mathrm{E}\{\boldsymbol{g}(\boldsymbol{X})\boldsymbol{g}(\boldsymbol{X})^\top\}$ is positive definite. The covariate vector $\boldsymbol{U}$ has a compact support $\mathbb{U} \subset \mathbb{R}^p$. The density of $\boldsymbol{U}$ is bounded away from infinity and zero on $\mathbb{U}$, and has bounded second-order derivatives.

(A2) Functions $\mu(\boldsymbol{u}) = \mathrm{E}(Y|\boldsymbol{U} = \boldsymbol{u})$, $\sigma^2(\boldsymbol{u}) = \mathrm{E}[\{Y - \mu(\boldsymbol{U})\}^2|\boldsymbol{U} = \boldsymbol{u}]$, and $\boldsymbol{g}(\boldsymbol{x})$ are Lipschitz continuous; $\mu(\boldsymbol{u})$ has bounded third-order derivatives; and $\mathrm{E}(|Y|^s|\boldsymbol{U} = \boldsymbol{u})$ is bounded.

(A3) The kernel κ is a positive, bounded, and Lipschitz continuous density with mean zero and finite sixth moments.

(A4) The bandwidths b in (**??**) and l in (1) satisfy $b \to 0$, $l \to 0$, $l/b \to r \in (0, \infty)$, $nb^p \to \infty$, and $nb^{4+p} \to c \in [0, \infty)$, as the internal sample size $n \to \infty$.

(A5) The external sample size m satisfies $n = O(m)$, i.e., n/m is bounded by a fixed constant.

- Back

Conditions for Theorem 1

Theorem 3

*Assume the conditions in Theorem 1 with $r \leq 1$. Assume further that the function in (**??**) has continuous second-order derivative $\rho''(s) < 0$ for $0 < s < 1$. Then $\mathrm{AMISE}(\widehat{\mu}_{CK}) < \mathrm{AMISE}(\widehat{\mu}_K)$ if and only if*

$$c < \tau \frac{-\int_0^1 (1-t)^2 \rho''(rt) dt}{2(2+r^2)}.$$

◂ Back