Causal Inference with Functional Data

Hyemin Yeon

Department of Statistics
Iowa State University

June 2, 2023

Causal Inference + Functional Data Analysis: example

- Electroencephalography (EEG) dataset (analyzed in [THZY22])
- Treatment $\{X(t) : t \in \mathcal{T}\}$: intensity of neuronal activities (represented by the frontal asymmetry) on the frequency domain $\mathcal{T}$
- Outcome Y: a measure of severity of major depressive disorder
- Covariate $\boldsymbol{W} = (W_1, W_2, W_3)^\top$
  - W_1: age
  - W_2: sex
  - W_3: Edinburgh Handedness Inventory score

Causal Inference + Functional Data Analysis: literature

- [MXZ20] Miao, Xue, and Zhang (2020) ATE Estimation in Observational Studies with Functional Covariates, ArXiv
- [ZXW21] Zhang, Xue, and Wang (2021) Covariate Balancing Functional Propensity Score for Functional Treatments, CSDA
- [THZY22] Tan, Huang, Zhang, and Yin (2022) Causal Effect of Functional Treatment, ArXiv
- [LKW22] Lin, Kong, and Wang (2022) Causal Inference on Distribution Functions
- [GWHS23] Gao, Wang, Hu, and Sun (2023) Functional Causal Inference with Time-to-Event Data

Causal Inference + Functional Data Analysis: literature

paper	Y	X	W	theory	
[MXZ20]	$\mathbb{R}$	$\{0,1\}$	$\mathbb{H}$	$\times$	relatively easy
[ZXW21]	$\mathbb{R}$	$\mathbb{H}$	$\mathbb{R}^p$	$\times$	following [FHI18]
[THZY22]	$\mathbb{R}$	$\mathbb{H}$	$\mathbb{R}^p$	$\checkmark$	following [HZ23]
[LKW22]	$\mathbb{S}$	$\{0,1\}$	$\mathbb{R}^p$	$\checkmark$	random object
[GWHS23]	$\mathbb{R}$	$\mathbb{H}$	$\mathbb{R}^p$	$\times$	survival, following [ZXW21]

- $\mathbb{H} = L^2([0,1])$
- $\mathbb{S} = \mathbb{W}^2([0,1])$

Causal Inference with Functional Treatment

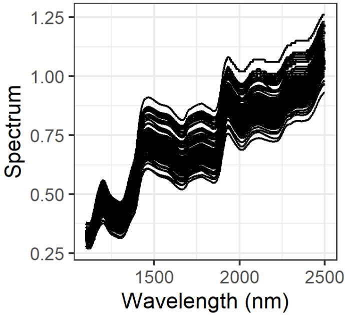
- Causal effect of (∞-dim'l) functional treatment on scalar outcome
- [ZXW21] applies
  - the methods by **[FHI18]** to the first J FPC scores to estimate the *stablized weights* $\pi_0^J(\boldsymbol{A}_i, \boldsymbol{W}_i)$ and
  - the **basis expansion** approach to estimate the causal effect $\beta \in \mathbb{H}$
- [THZY22] applies
  - the methods by **[HZ23]** to estimate the *stablized weights* $\pi_0(X_i, \boldsymbol{W}_i)$ and
  - the **FPCR** method to estimate the *causal effect* $\beta \in \mathbb{H}$.

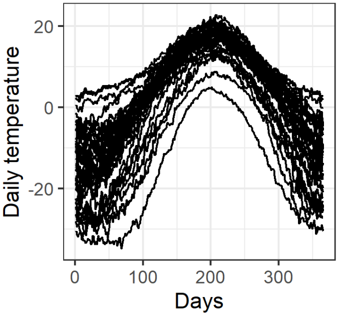
Contents

Functional data analysis (FDA)

- Function-valued random variables: curves, surfaces, ...
- Function space $\mathbb{H}$ with $\dim \mathbb{H} = \infty$.

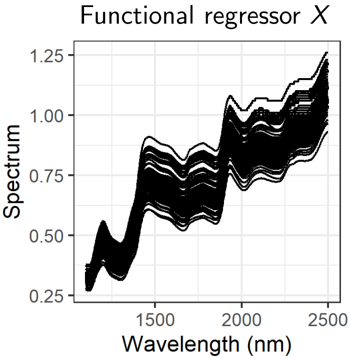
- e.g., $\mathbb{H} = L^2([0,1])$ with $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$.





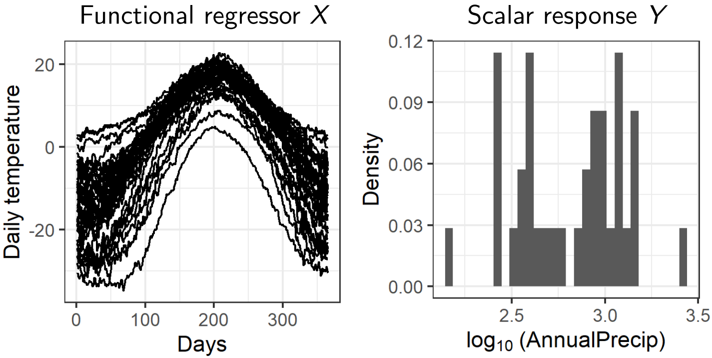
Functional linear regression models (FLRMs)

$$Y = \langle \beta, X \rangle + \varepsilon.$$



Functional linear regression models (FLRMs)

$$Y = \langle \beta, X \rangle + \varepsilon.$$



Estimation approaches in FLRMs

- Naive approach using basis expansion
- Functional Principal component regression (FPCR) estimation: an infinite-dimensional space version of PCR

Estimation: Basis Expansion

- $\{\psi_j\}_{j=1}^{\infty}$: a fixed known orthonormal basis of $\mathbb{H}$
  $\implies \beta = \sum_{j=1}^{\infty} \beta_j \psi_j \approx \sum_{j=1}^{J} \beta_j \psi_j$ where $\beta_j = \langle \beta, \psi_j \rangle$
  $\implies \langle \beta, X_i \rangle = \sum_{j=1}^{\infty} \beta_j \langle X_i, \psi_j \rangle \approx \sum_{j=1}^{J} \beta_j \langle X_i, \psi_j \rangle$

- $\hat{\beta}_J \equiv \sum_{j=1}^{J} \hat{\beta}_j \psi_j$, where

$$(\hat{\beta}_1, \ldots, \hat{\beta}_J)^\top \equiv \underset{(\beta_1,\ldots,\beta_J)^\top \in \mathbb{R}^J}{\operatorname{argmin}} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{J} \beta_j \langle X_i, \psi_j \rangle \right)^2$$

FPCR estimation: normal equations

- Model: $Y = \langle \beta, X \rangle + \varepsilon$
- Normal equation:

$$\underbrace{\mathrm{E}[XY]}_{\equiv \Delta} = \underbrace{\mathrm{E}[X \otimes X]}_{\equiv \Gamma} \beta$$

- Sample: $Y_i = \langle \beta, X_i \rangle + \varepsilon_i$, $i = 1, \ldots, n$
- Normal equation:

$$\underbrace{n^{-1} \sum_{i=1}^{n} X_i Y_i}_{\equiv \hat{\Delta}_n} = \underbrace{\left(n^{-1} \sum_{i=1}^{n} X_i \otimes X_i \right)}_{\equiv \hat{\Gamma}_n} \beta + \underbrace{n^{-1} \sum_{i=1}^{n} X_i \varepsilon_i}_{\equiv U_n}$$

FPCR estimation: ill-posed inverse problem

- The inversion of Γ is an ill-conditioned inverse problem.
- $\text{rank}(\hat{\Gamma}_n) < \infty \implies \hat{\Gamma}_n^{-1}$ does not exist.
- Regularization/Truncation!

FPCR estimation: regularization

- Regularization of $\hat{\Gamma}_n^{-1}$:

$$\hat{\Gamma}_h^{-1} = \sum_{j=1}^{h} \hat{\gamma}_j^{-1} (\hat{\phi}_j \otimes \hat{\phi}_j),$$

where $(\hat{\gamma}_j, \hat{\phi}_j)$ is the j-th eigenpair of $\hat{\Gamma}_n$.

- The estimator $\hat{\beta}_{h_n}$ of β:

$$\hat{\beta}_{h_n} = \hat{\Gamma}_{h_n}^{-1} \hat{\Delta}_n = \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1} \left\langle n^{-1} \sum_{i=1}^{n} X_i Y_i, \hat{\phi}_j \right\rangle \hat{\phi}_j$$

Contents

Notation

- Y: outcome valued in $\mathbb{R}$
- X: treatment valued in $\mathbb{H} = L^2([0,1])$
- $Y(x)$: potential outcome for $x \in \mathbb{H}$
- $\boldsymbol{W}$: covariate valued in $\mathbb{R}^p$
- Causal estimand of interest: $\mathrm{E}[Y(x)]$

Contents

Identification assumptions

- (Consistency) $Y_i = Y_i(x)$ if $X = x$
- (Ignorability) $\forall x \in \mathbb{H},\ Y(x) \perp X | \boldsymbol{W}$
- (Positivity) $f_{X|\boldsymbol{W}}(x|\boldsymbol{w}) > 0$...?

Issues

- No density for functional data (Delaigle and Hall, 2010)
- Need a new approach of defining propensity score (or stablized weight)

- [ZXW21] truncate X and use density for truncated X (use the densities of FPC scores)

- [THZY22] use conditional density of $\boldsymbol{W}$ given X instead of the one of X given $\boldsymbol{W}$

Positivity condition and stablized weight in [ZXW21]

- Karhunen–Loéve expansion:

$$X \overset{d}{=} \sum_{j=1}^{\infty} \gamma_j^{1/2} A_j \phi_j,$$

where $A_j = \gamma_j^{-1/2} \langle X, \phi_j \rangle \sim (0, 1)$ are uncorrelated.
Set $\boldsymbol{A} = (A_1, \ldots, A_J)^\top$ for a **fixed** $J \in \mathbb{N}$.

- Standardized covariate: $\mathrm{E}[\boldsymbol{W}] = 0$, $\mathrm{var}[\boldsymbol{W}] = I_p$
$\implies W_k \sim (0, 1)$ are uncorrelated.
- (Positivity) $f_{\boldsymbol{A}|\boldsymbol{W}}(\boldsymbol{a}|\boldsymbol{w}) > 0$ for all $\boldsymbol{a} \in \mathbb{R}^J$ and $\boldsymbol{w} \in \mathbb{R}^p$
- Stablized weight:

$$\pi_0^J(\boldsymbol{a}, \boldsymbol{w}) = f_{\boldsymbol{A}}(\boldsymbol{a}) / f_{\boldsymbol{A}|\boldsymbol{W}}(\boldsymbol{a}|\boldsymbol{w})$$

Positivity condition in [ZXW21]

- (Positivity) $f_{\boldsymbol{A}|\boldsymbol{W}}(\boldsymbol{a}|\boldsymbol{w}) > 0$ for all $\boldsymbol{a} \in \mathbb{R}^J$ and $\boldsymbol{w} \in \mathbb{R}^p$
- Identification:

$$\mathrm{E}[Y(x)|\boldsymbol{W}] = \mathrm{E}[Y(x)|X = x, \boldsymbol{W}] = \mathrm{E}[Y|X = x, \boldsymbol{W}]?$$
$$\mathrm{E}[Y(x)|\boldsymbol{W}] = \mathrm{E}[Y(x)|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{W}] \neq \mathrm{E}[Y|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{W}]!$$

- We need a stronger consistency assumption: $Y_i = Y_i(x)$ if $\boldsymbol{A} = \boldsymbol{a}$. This implies

$$\mathrm{E}[Y(x)] = \mathrm{E}[\mathrm{E}[Y|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{W}]]$$

Positivity condition and stablized weight in [THZY22]

- The conditional density $f_{\boldsymbol{W}|X}(\boldsymbol{w}|x)$ of covariate $\boldsymbol{W}$ given (functional) treatment X is well-defined.
- (Positivity) $f_{\boldsymbol{W}|X}(\boldsymbol{w}|x) > 0$ for all $\boldsymbol{w} \in \mathbb{R}^p$ and $x \in \mathbb{H}$
- Stablized weight:

$$\pi_0(x, \boldsymbol{w}) = f_{\boldsymbol{W}}(\boldsymbol{w})/f_{\boldsymbol{W}|X}(\boldsymbol{w}|x)$$

- If $\mathbb{H} = \mathbb{R}^q$, this will be the classical stablized weight:

$$\begin{aligned}\pi_0(x, \boldsymbol{w}) = f_{\boldsymbol{W}}(\boldsymbol{w})/f_{\boldsymbol{W}|X}(\boldsymbol{w}|x) &= f_{\boldsymbol{W}}(\boldsymbol{w})/\{f_{\boldsymbol{W},X}(\boldsymbol{w}, x)/f_X(x)\} \\ &= f_X(x)/f_{X|\boldsymbol{W}}(x|\boldsymbol{w}).\end{aligned}$$

- Identification:

$$\begin{aligned}&\mathrm{E}[Y(x)|\boldsymbol{W}] = \mathrm{E}[Y(x)|X = x, \boldsymbol{W}] = \mathrm{E}[Y|X = x, \boldsymbol{W}] \\ \Longrightarrow\ &\mathrm{E}[Y(x)] = \mathrm{E}[\mathrm{E}[Y|X = x, \boldsymbol{W}]]\end{aligned}$$

Contents

Estimate the stablized weights in [ZXW21]

- Following the methods by [FHI18] (Fong, Hazlett, and Imai, 2018)
- [FHI18] focuses on continuous treatment X (valued in $\mathbb{R}$)
  $\Longrightarrow$ [ZXW21] extends the methods by [FHI18]
  to the case when the treatment variable lies in $\mathbb{R}^J$
  - Parametric: $\boldsymbol{A}$ and $\boldsymbol{W}$ are jointly normal and use moment equations
  - Nonparametric: use an empirical likelihood approach based on moment equations
- Almost same as [FHI18]
- No theory

Estimate the stablized weights in [THZY22]

- For each $x \in \mathbb{H}$, $\pi(x, \boldsymbol{W}) = \pi_0(x, \boldsymbol{W})$ almost surely if and only if

$$\mathrm{E}[\pi(x, \boldsymbol{W}) u(\boldsymbol{W}) | X = x] = \mathrm{E}[u(\boldsymbol{W})], \quad \forall u \in L^1(\mathcal{W}), \qquad (1)$$

  where $\mathcal{W} \equiv \mathrm{supp}(\boldsymbol{W})$.

- Use the Nadaraya-Watson estimator of RHS in (1)
  to estimate $\{\pi_0(x, \boldsymbol{W}_{i'})\}_{i' \neq i}$ at each $x = X_i$:

$$\frac{\sum_{i' \neq i} \pi(x, \boldsymbol{W}_{i'}) u(\boldsymbol{W}_{i'}) K(d(X_{i'}, x)/h)}{\sum_{i' \neq i} K(d(X_{i'}, x)/h)} = (n-1)^{-1} \sum_{i' \neq i} u(\boldsymbol{W}_{i'}).$$

- Sieve approximation to $L^1(\mathcal{W})$ by $\{u_k\}_{k=1}^{K}$
- An empirical likelihood method
- Almost same as [HZ23]

Proof of Proposition 3.1 in [THZY22] (Equation (1))

For each $x \in \mathbb{H}$, $\pi(x, \boldsymbol{W}) = \pi_0(x, \boldsymbol{W})$ almost surely if and only if

$$\mathrm{E}[\pi(x, \boldsymbol{W})u(\boldsymbol{W})|X = x] = \mathrm{E}[u(\boldsymbol{W})], \quad \forall u \in L^1(\mathcal{W}),$$

where $\mathcal{W} \equiv \mathrm{supp}(\boldsymbol{W})$.

Proof of $(\Rightarrow)$.

$$\begin{aligned} & \mathrm{E}[\pi_0(x, \boldsymbol{W})u(\boldsymbol{W})|X = x] \\ = & \int \pi_0(x, \boldsymbol{w})u(\boldsymbol{w})f_{\boldsymbol{W}|X}(\boldsymbol{w}|x)d\boldsymbol{w} \\ = & \int u(\boldsymbol{w})f_{\boldsymbol{W}}(\boldsymbol{w})d\boldsymbol{w} = \mathrm{E}[u(\boldsymbol{W})] \end{aligned}$$

■

Proof of Proposition 3.1 in [THZY22] (Equation (1))

For each $x \in \mathbb{H}$, $\pi(x, \boldsymbol{W}) = \pi_0(x, \boldsymbol{W})$ almost surely if and only if

$$\mathrm{E}[\pi(x, \boldsymbol{W})u(\boldsymbol{W})|X = x] = \mathrm{E}[u(\boldsymbol{W})], \quad \forall u \in L^1(\mathcal{W}),$$

where $\mathcal{W} \equiv \mathrm{supp}(\boldsymbol{W})$.

Proof of $(\Leftarrow)$.

For each $x \in \mathbb{H}$, by putting $u(\boldsymbol{W}) = \pi(x, \boldsymbol{W}) - \pi_0(x, \boldsymbol{W})$, we have

$$\begin{aligned} 0 =& \mathrm{E}[\{\pi(x, \boldsymbol{W}) - \pi_0(x, \boldsymbol{W})\}u(\boldsymbol{W})|X = x] \\ =& \mathrm{E}[\{\pi(x, \boldsymbol{W}) - \pi_0(x, \boldsymbol{W})\}^2|X = x], \end{aligned}$$

which implies that $\pi(x, \boldsymbol{W}) - \pi_0(x, \boldsymbol{W}) = 0$ almost surely.
For each $x \in \mathbb{H}$, is $\pi(x, \cdot) - \pi_0(x, \cdot)$ really contained in $L^1(\mathcal{W})$? ■

Contents

Estimate the causal effect β in [ZXW21]

- Linear model: $\mathrm{E}[Y|X] = \alpha + \langle \beta, X \rangle$
- β measures the causal effect of the functional treatment
- Use the basis expansion approach with weighted regression
- $\{\psi_j\}_{j=1}^{\infty}$: an orthonormal basis of $\mathbb{H}$
  $\implies \beta = \sum_{j=1}^{\infty} \beta_j \psi_j \approx \sum_{j=1}^{J} \beta_j \psi_j$ where $\beta_j = \langle \beta, \psi_j \rangle$
  $\implies \langle \beta, X_i \rangle = \sum_{j=1}^{\infty} \beta_j \langle X_i, \psi_j \rangle \approx \sum_{j=1}^{J} \beta_j \langle X_i, \psi_j \rangle$
- $\hat{\beta}_J \equiv \sum_{j=1}^{J} \hat{\beta}_j \psi_j$, where

$$(\hat{\beta}_1, \ldots, \hat{\beta}_J)^\top \equiv \underset{(\beta_1,\ldots,\beta_J)^\top \in \mathbb{R}^J}{\operatorname{argmin}} \sum_{i=1}^{n} \hat{\pi}^J(\boldsymbol{A}_i, \boldsymbol{W}_i) \left(Y_i - \sum_{j=1}^{J} \beta_j \langle X_i, \psi_j \rangle \right)^2$$

Estimate the causal effect β in [ZXW21]

- Issue: the linear model $\mathrm{E}[Y|X] = \alpha + \langle \beta, X \rangle$ is reasonable?
- $\mathrm{E}[Y(x)]$ is identifiable only with $\boldsymbol{A}$ not X, so..
  (1) The model should be truncated at J: $\mathrm{E}[Y|X] = \alpha + \sum_{j=1}^{J} \beta_j \langle X, \psi_j \rangle$.
  (2) The estimation method should use $\psi_j = \phi_j$, which is unknown.

- simulation: comparison with the unweighted LSE
  (the method without considering the stablized weights)

Estimate the causal effect β in [THZY22]: FSW

$$\mathrm{E}[Y(x)] = \mathrm{E}[\pi_0(X, \boldsymbol{W})Y | X = x] \tag{2}$$

Proof.

$$\begin{aligned}
\mathrm{E}[Y(x)] &= \mathrm{E}[\mathrm{E}[Y|X = x, \boldsymbol{W}]] \\
&= \int \mathrm{E}[Y|X = x, \boldsymbol{W} = \boldsymbol{w}] f_{\boldsymbol{W}}(\boldsymbol{w}) d\boldsymbol{w} \\
&= \int \mathrm{E}[\pi_0(x, \boldsymbol{w})Y|X = x, \boldsymbol{W} = \boldsymbol{w}] f_{\boldsymbol{W}|X}(\boldsymbol{w}|x) d\boldsymbol{w} \\
&= \mathrm{E}[\mathrm{E}[\pi_0(X, \boldsymbol{W})Y|X = x, \boldsymbol{W}]|X = x] \\
&= \mathrm{E}[\pi_0(X, \boldsymbol{W})Y|X = x]
\end{aligned}$$

∎

Estimate the causal effect β in [THZY22]: FSW

- Linear model: $\mathrm{E}[Y(x)] = \alpha + \langle \beta, x \rangle$
- Based on Equation (2): $\mathrm{E}[Y(x)] = \mathrm{E}[\pi_0(X, \boldsymbol{W})Y | X = x]$
- This suggests estimating $\mathrm{E}[Y(x)]$ by using the FPCR estimator from the regression of $\{Z_i \equiv \hat{\pi}(X_i, \boldsymbol{W}_i)Y_i\}_{i=1}^n$ on $\{X_i\}_{i=1}^n$:

$$\hat{\beta}_{h_n}^{\mathrm{FSW}} = \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1} \left\langle n^{-1} \sum_{i=1}^{n} X_i Z_i, \hat{\phi}_j \right\rangle \hat{\phi}_j$$

$$\hat{\alpha}^{\mathrm{FSW}} = \bar{Z} - \langle \hat{\beta}_{h_n}, \bar{X} \rangle$$

Estimate the causal effect β in [THZY22]: OR

- Partially linear model:

$$Y = \alpha + \langle \beta, X \rangle + \boldsymbol{\theta}^\top \boldsymbol{W} + \varepsilon \tag{3}$$

- Based on $\mathrm{E}[Y(x)] = \mathrm{E}[\mathrm{E}[Y|X = x, \boldsymbol{W}]]$
- The backfitting algorithm: setting $\hat{\boldsymbol{\theta}}^{(0)} = 0$, do the following:
  (1) $(\hat{\alpha}^{(k)}, \hat{\beta}_{h_n}^{(k)})$: the FPCR estimators from the regression of $\{Y_i - (\hat{\boldsymbol{\theta}}^{(k-1)})^\top \boldsymbol{W}_i\}_{i=1}^n$ on $\{X_i\}_{i=1}^n$
  (2) $\hat{\boldsymbol{\theta}}^{(k)}$: the LS estimator from the regression of $\{Y_i - (\hat{\alpha}^{(k)} + \langle \hat{\beta}_{h_n}^{(k)}, X_i \rangle)\}_{i=1}^n$ on $\{\boldsymbol{W}_i\}_{i=1}^n$

  until some convergence criteria hold.
- Outcome regression estimators: $\hat{\alpha}^{\mathrm{OR}}$, $\hat{\beta}_{h_n}^{\mathrm{OR}}$, $\hat{\boldsymbol{\theta}}^{\mathrm{OR}}$

Estimate the causal effect β in [THZY22]: DR

It holds that

$$\mathrm{E}[Y(x)] = \mathrm{E}\left[\pi(X, \boldsymbol{W})\{Y - m(X, \boldsymbol{W})\} + \int m(X, \boldsymbol{W}) dP_{\boldsymbol{W}} \middle| X = x\right]$$

if either $\pi = \pi_0$ or $m(x, \boldsymbol{W}) = \mathrm{E}[Y|X = x, \boldsymbol{W}]$ almost surely.

Proof.

If $m(x, \boldsymbol{W}) = \mathrm{E}[Y|X = x, \boldsymbol{W}]$ almost surely, then

$$\begin{aligned}
&\mathrm{E}[\pi(X, \boldsymbol{W})\{Y - m(X, \boldsymbol{W})\}|X = x] \\
=&\mathrm{E}[\mathrm{E}[\pi(X, \boldsymbol{W})\{Y - m(X, \boldsymbol{W})\}|X = x, \boldsymbol{W}]] \\
=&\mathrm{E}[\pi(x, \boldsymbol{W})\{\mathrm{E}[Y|X = x, \boldsymbol{W}] - m(x, \boldsymbol{W})\}] \\
=&0.
\end{aligned}$$

■

Estimate the causal effect β in [THZY22]: DR

It holds that

$$\mathrm{E}[Y(x)] = \mathrm{E}\left[\pi(X, \boldsymbol{W})\{Y - m(X, \boldsymbol{W})\} + \int m(X, \boldsymbol{W}) dP_{\boldsymbol{W}} \middle| X = x\right]$$

if either $\pi = \pi_0$ or $m(x, \boldsymbol{W}) = \mathrm{E}[Y|X = x, \boldsymbol{W}]$ almost surely.

Proof.

If $\pi = \pi_0$, then

$$\begin{aligned}
\mathrm{E}[\pi_0(X, \boldsymbol{W})m(X, \boldsymbol{W})|X = x] &= \int \pi_0(x, \boldsymbol{w})m(x, \boldsymbol{w})f_{\boldsymbol{W}|X}(\boldsymbol{w}|x)d\boldsymbol{w} \\
&= \int m(x, \boldsymbol{w})f_{\boldsymbol{W}}(\boldsymbol{w})d\boldsymbol{w} \\
&= \mathrm{E}\left[\int m(X, \boldsymbol{W})dP_{\boldsymbol{W}} \middle| X = x\right]
\end{aligned}$$

which implies that $\mathrm{E}[Y(x)] = \mathrm{E}[\pi_0(X, \boldsymbol{W})Y|X = x]$. ∎

Estimate the causal effect β in [THZY22]: DR

$$\mathrm{E}[Y(x)] = \mathrm{E}\left[\pi_0(X, \boldsymbol{W})\{Y - \mathrm{E}[Y|X = x, \boldsymbol{W}]\} + \int \mathrm{E}[Y|X, \boldsymbol{W}]dP_{\boldsymbol{W}} \middle| X = x\right] \tag{4}$$

- Based on Equation (4)
- Regress an estimator of V on X, where

$$V \equiv \pi_0(X, \boldsymbol{W})\{Y - \mathrm{E}[Y|X, \boldsymbol{W}]\} + \int \mathrm{E}[Y|X, \boldsymbol{W}]dP_{\boldsymbol{W}}$$

- Regress $\{V_i\}_{i=1}^n$ on $\{X_i\}_{i=1}^n$, where

$$V_i \equiv \hat{\pi}(X_i, \boldsymbol{W}_i)\{Y_i - \hat{m}(X_i, \boldsymbol{W}_i)\} + n^{-1} \sum_{i'=1}^{n} \hat{m}(X_i, \boldsymbol{W}_{i'})$$

Estimate the causal effect β in [THZY22]: DR

- $\hat{\pi}$ is from FSW part
- $\hat{m}$ is from OR part:

$$\hat{m}(X, \boldsymbol{W}) = \hat{\alpha}^{\mathrm{OR}} + \langle \hat{\beta}^{\mathrm{OR}}, X \rangle + (\hat{\boldsymbol{\theta}}^{\mathrm{OR}})^\top \boldsymbol{W}$$

- Doubly robust estimators:

$$\hat{\beta}_{h_n}^{DR} = \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1} \left\langle n^{-1} \sum_{i=1}^{n} X_i V_i, \hat{\phi}_j \right\rangle \hat{\phi}_j$$

$$\hat{\alpha}^{\mathrm{DR}} = \bar{V} - \langle \hat{\beta}_{h_n}^{DR}, \bar{X} \rangle$$

Estimate the causal effect β in [THZY22]: theory

(4.1) $\sup_{x,\boldsymbol{w}} |\hat{\pi}(x, \boldsymbol{w}) - \pi_0(x, \boldsymbol{w})| = O(a_n + b_n)$ almost surely

(4.2) $\|\hat{\beta}^{\mathrm{FSW}}_{h_n} - \beta\|^2 = O_{\mathrm{P}}\left((a_n^2 + b_n^2) n^{\frac{\gamma+1}{\gamma+2\beta}}\right)$, when $h_n \asymp n^{\frac{1}{\gamma+2\beta}}$

(4.3) $\|\hat{\beta}^{\mathrm{OR}}_{h_n} - \beta\|^2 = O_{\mathrm{P}}\left(n^{\frac{-2\beta+1}{\gamma+2\beta}}\right)$, when $h_n \asymp n^{\frac{1}{\gamma+2\beta}}$

- $n^{\frac{-2\beta+1}{\gamma+2\beta}} = n^{-1} n^{\frac{\gamma+1}{\gamma+2\beta}}$ and $n^{-1} < a_n^2 + b_n^2$
  $\Longrightarrow$ $\hat{\beta}^{\mathrm{OR}}_{h_n}$ is faster than $\hat{\beta}^{\mathrm{FSW}}_{h_n}$
- Theorem 4.3 is the same result as the classical ones in FLRMs such as Hall and Horowitz (2007) and Shin (2009).

Contents

Selection of tuning parameters

- [ZXW21] Use the fraction of variance explained (FVE):

$$FVE(J) \equiv \frac{\sum_{j=1}^{J} \hat{\gamma}_j}{\sum_{j\geq 1} \hat{\gamma}_j}.$$

For example, use J such that $FVE(J) \approx 0.95$.

- [THZY22] Use the cross-validation based on the prediction error

Simulation

- Criterion: $MSE(\beta) = \mathrm{E}[\|\hat{\beta} - \beta\|^2]$

- [ZXW21] Compare the PCW with the unweighted LSE based on the basis expansion
  - PCW is better

- [THZY22] Compare the FSW, OR, DR with PCW and (naive) FPCR
  - All of FSW, OR, DR better than both PCW and FPCR.
  - Mostly, FSW is the best or competitive compared to OR and DR.

Future directions

- Statistical inference (e.g., confidence region, hypothesis testing) for β in the framework of causal inference with functional treatment

- Causal inference when the treatments are sparse functional data

The End

THANK YOU