Doubly-Robust Functional Average Treatment Effect Estimation

Reading Group Presentation

Lorenzo Testa, Tobia Boschi, Francesca Chiaromonte, Edward H. Kennedy, Matthew Reimherr

March 28, 2025

Presented by Yuankang Zhao

Introduction

- **Motivation:** Understanding causal effects in observational studies with functional data.
- Many traditional methods target scalar outcomes; however, many applications (e.g., longitudinal medical data) require handling functional outcomes.
- The paper addresses the estimation of the Functional Average Treatment Effect (FATE) over a continuous domain.

Background

- **Challenges:**
  - Functional data is infinite-dimensional and exhibits complex dependence across the domain.
  - Standard causal inference methods (e.g., AIPW) are designed for scalar outcomes.
- **Existing Methods:**
  - Function-on-scalar regression models are often non-robust.
  - Some approaches (e.g., Liu et al. (2024)) require strong parametric assumptions.

Introduction: Need for a New Approach

- **DR-FoS:** A novel estimator that extends double robustness to the functional data setting.
- **Key Idea:** Combine outcome regression and propensity score weighting to achieve consistency if either model is correctly specified.
- **Additional Feature:** Uses cross-fitting and provides simultaneous confidence bands over the continuous domain.

Definition of FATE

- The target parameter is the **Functional Average Treatment Effect (FATE)**:

$$\beta(t) = E\left[Y^{(1)}(t) - Y^{(0)}(t)\right]$$

- Here, $Y^{(a)}(t)$ denotes the potential outcome function at time t under treatment $a \in \{0, 1\}$.
- The estimation target is a function over a continuous domain (e.g., time interval $T = [0, 1]$).
- **Identifiability Assumption**
- **Consistency** The potential outcome of a treatment is the same regardless of the mechanism by which the treatment is administered; that is, $Y^{(a)}(t)$ if $A = a$ for any $t \in T$.
- **No unmeasured confounding.** $Y^{(a)}(t) \perp\!\!\!\perp A \mid X$ for any $t \in T$.
- **Positivity.** $0 < \mathbb{P}[A = 1 \mid X] < 1$ for any $t \in T$.

Estimation: Model Setup

- Two main components are modeled:
  - **Propensity Score:**

$$\pi^{(a)}(X) = P(A = a \mid X)$$

  - **Outcome Regression:**

$$\mu^{(a)}(X, t) = E[Y^{(a)}(t) \mid X, A = a]$$

- Both models can be estimated using flexible methods (e.g., logistic regression, nonparametric regression).
- **Double Robustness:** The estimator remains consistent if either the propensity model or the outcome model is correctly specified.

Estimation: The DR-FoS Estimator

- **Corrected Regression Function:**

$$\gamma^{(a)}(D,t) = \begin{cases} \mu^{(a)}(X,t) + \dfrac{Y(t) - \mu^{(a)}(X,t)}{\pi^{(a)}(X)} & \text{if } A = a, \\ \mu^{(a)}(X,t) & \text{if } A \neq a. \end{cases}$$

  where $D = (A, X, Y)$

- Then FATE could be rewritten as:

$$\beta(t) = E\left[\gamma^{(1)}(D,t) - \gamma^{(0)}(D,t)\right]$$

Estimation: The DR-FoS Estimator

Proof:

$$E\left[\gamma^{(1)}(D,t)\right] = E\left[\mu^{(1)}(X,t) + \frac{A(Y(t) - \mu^{(1)}(X,t))}{\pi^{(1)}(X)}\right]$$

$$= E\left[\mu^{(1)}(X,t)\right] + E\left[\frac{A(Y(t) - \mu^{(1)}(X,t))}{\pi^{(1)}(X)}\right]$$

$$= E\left[E[Y(t)|X, A=1]\right] + E\left[E\left[\frac{A(Y(t) - \mu^{(1)}(X,t))}{\pi^{(1)}(X)}\middle| X\right]\right]$$

$$= E\left[E[Y(t)|X, A=1]\right]$$

$$= E[Y^{(1)}(t)]$$

Estimation: The DR-FoS Estimator

The DR-FoS estimator is then given by:

$$\hat{\beta}_{\text{DR-FoS}}(t) = P_n\left[\hat{\gamma}^{(1)}(D,t) - \hat{\gamma}^{(0)}(D,t)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\gamma}^{(1)}(D_i,t) - \hat{\gamma}^{(0)}(D_i,t)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\left(\hat{\mu}^{(1)}(X_i,t) + \frac{A_i(Y_i(t) - \hat{\mu}^{(1)}(X_i,t))}{\hat{\pi}^{(1)}(X_i)}\right)\right.$$

$$\left. - \left(\hat{\mu}^{(0)}(X_i,t) + \frac{(1-A_i)(Y_i(t) - \hat{\mu}^{(0)}(X_i,t))}{1-\hat{\pi}^{(1)}(X_i)}\right)\right]$$

Cross-Fitting Procedure for DR-FoS Estimation

Step 1: Data Partitioning

Randomly partition dataset into J folds:

$$\{D_1, D_2, \ldots, D_n\} D_i = (A_i, X_i, Y_i(t))$$

Step 2: Training-Prediction Cycle (for each fold j)

For each fold $j = 1, \ldots, J$:

1. Train models using all data except fold j:
   - Estimate propensity score: $\hat{\pi}^{(a),[-j]}(X)$
   - Estimate outcome regression: $\hat{\mu}^{(a),[-j]}(X, t)$
2. Compute DR-FoS estimator on fold j:

$$\hat{\beta}^{[j]}_{\text{DR-FoS}}(t) = \frac{1}{n_j} \sum_{i \in \text{Fold } j} \left[\hat{\gamma}^{(1),[-j]}(D_i, t) - \hat{\gamma}^{(0),[-j]}(D_i, t) \right]$$

Step 3: Averaging Over Folds

$$\hat{\beta}_{\text{DR-FoS}}(t) = \frac{1}{J} \sum_{j=1}^{J} \hat{\beta}^{[j]}_{\text{DR-FoS}}(t)$$

Inference: Asymptotic Properties

- Under regularity conditions, the estimator converges in distribution:

$$\sqrt{n}\big(\hat{\beta}_{\text{DR-FoS}}(t) - \beta(t)\big) \Rightarrow \mathcal{GP}(0, \Sigma(s,t))$$

- $\mathcal{GP}$ denotes a Gaussian process, ensuring pointwise asymptotic normality.
- The covariance function is given by:

$$\Sigma(s,t) = E\left[\phi(D;s)\phi(D;t)\right],$$

where $\phi(D;t) = \gamma^{(1)}(D,t) - \gamma^{(0)}(D,t) - \beta(t)$.

Inference: Constructing Simultaneous Confidence Bands

- Two approaches are discussed:
  1. **Critical Value Function Approach:** Uses an estimated critical value $u^*_{\alpha/2}$ to form bands:

$$C_\alpha(t) = \hat{\beta}_{\text{DR-FoS}}(t) \pm \frac{u^*_{\alpha/2}\hat{\sigma}(t)}{\sqrt{n}}.$$

  2. **Parametric Bootstrap:** Repeatedly sample from the estimated Gaussian process to obtain empirical quantiles, forming the band:

$$C_\alpha(t) = [\hat{\beta}_{\alpha/2}(t), \hat{\beta}_{1-\alpha/2}(t)].$$

- These bands provide simultaneous coverage over the entire function domain.

Simulation Study: Design and Data Generation

- **Data Generating Process:**
  - True outcome model:

$$\mu^{(a)}(t) = a\beta(t) + \rho(t),$$

    where $\beta(t)$ and $\rho(t)$ are drawn from a Gaussian process with Matern covariance.
  - Observations:

$$Y_i^{(a)}(t) = \mu^{(a)}(t) + \epsilon_i(t),$$

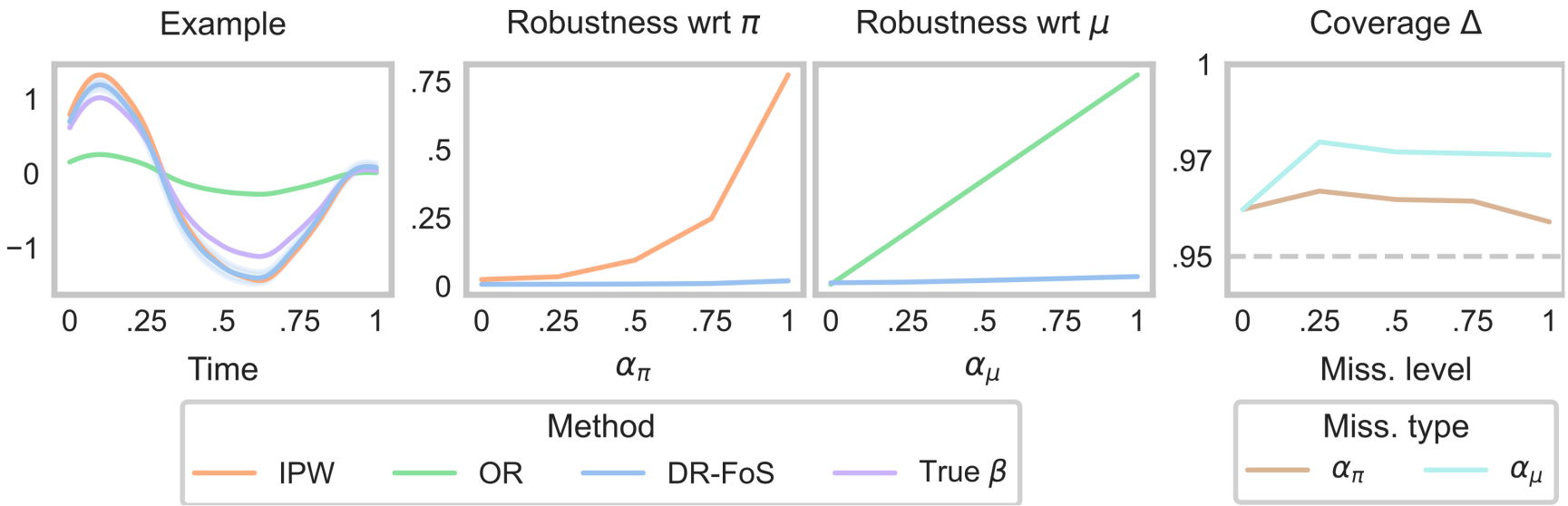
    and

$$Y_i(t) = A_i Y_i^{(1)}(t) + (1 - A_i) Y_i^{(0)}(t).$$

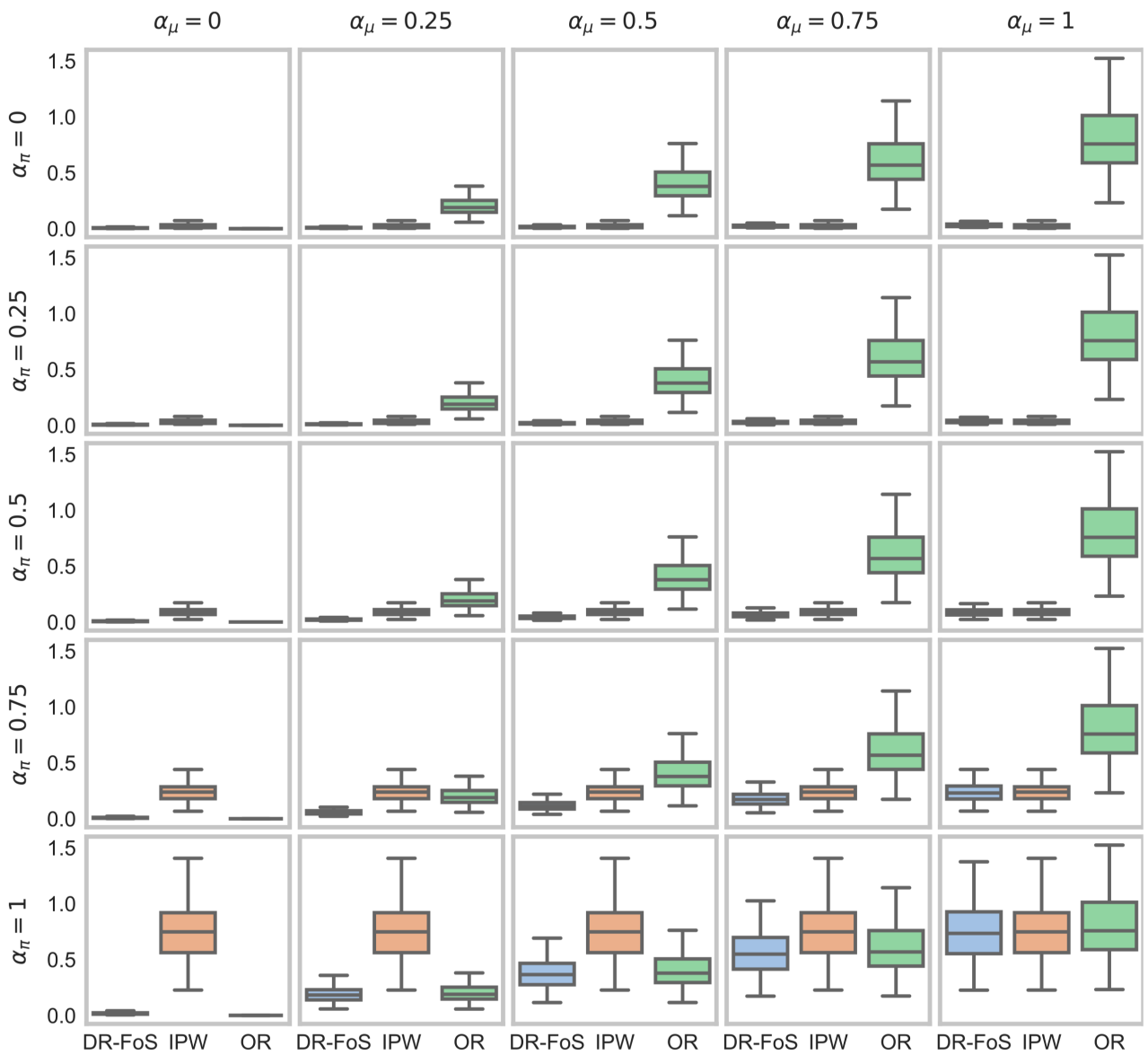
- **Controlling Model Misspecification:**
  - α_π controls the misspecification of the propensity score model.
  - α_μ controls the misspecification of the outcome regression model.

Simulation Study: Results and Implications

- Compared methods: DR-FoS, Outcome Regression (OR), and Inverse Probability Weighting (IPW) via MSE and Coverage probability of the simultaneous confidence bands.
- Key findings:
  - When one model is misspecified (either $\alpha_\pi > 0$ or $\alpha_\mu > 0$), DR-FoS remains robust.
  - When both models are misspecified, DR-FoS outperforms the other methods.



Simulation Study: Results and Implications

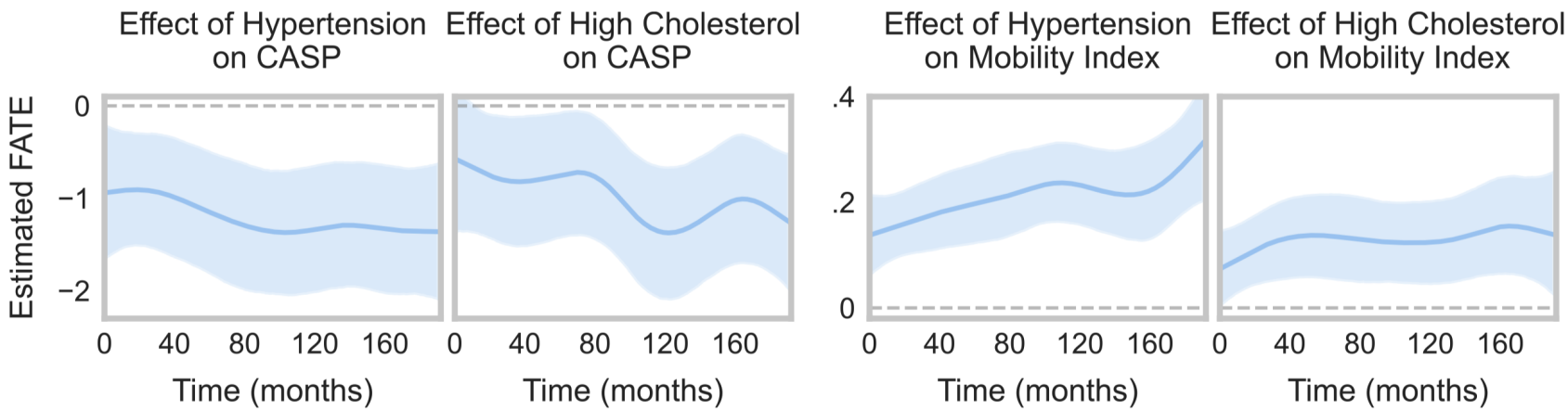


SHARE Application: Dataset and Variables

- **Dataset:** Survey of Health, Ageing and Retirement in Europe (SHARE).
- **Participants:** 1518 subjects with at least 7 out of 8 survey waves.
- **Key Variables:**
  - **Treatments:** High blood pressure and high cholesterol.
  - **Outcomes:** Functional variables such as the Mobility Index and CASP (quality of life measure).
  - **Covariates:** Age, education, numeracy score, gender, smoking status, childhood vaccinations.
- **Preprocessing:**
  - Time-varying outcomes are smoothed using cubic B-splines.
  - Functional representations are constructed to align data across survey waves.
- **Estimation Setup:**
  - Logistic regression for the propensity score.
  - Function-on-scalar least squares for the outcome regression.
  - 5-fold cross-fitting is employed.
  - Parametric bootstrap is used to build simultaneous confidence bands.

SHARE Application: Main Findings and Implications

- **Findings:**
  - Both high blood pressure and high cholesterol have a significant negative impact on quality of life (CASP) over time.
  - Effects on the Mobility Index indicate a deterioration in mobility (a higher index implies worse mobility).
  - The adverse effects become more pronounced with increasing age.



- **Implications:**
  - The DR-FoS method demonstrates practical utility in real-world, complex longitudinal data.
  - These insights can inform healthcare policy and personalized treatment planning.

Summary

- Proposed a **Doubly-Robust Functional Average Treatment Effect** (DR-FoS) estimator, and established rigorous asymptotic theory to make inference.
- Demonstrated excellent performance in simulation studies, and applied DR-FoS to real longitudinal data (SHARE study).

Concerns and Future Directions:

- **Parametric Assumptions and Robustness:**
  - Current simulations assume Gaussian processes. How robust is the DR-FoS estimator to deviations from such assumptions (e.g., non-Gaussian errors)?
  - Need to further assess robustness in diverse practical data-generating scenarios.
- **Limited Exploration of High-dimensional Covariates:**
  - Analyses involved relatively low-dimensional covariates. High-dimensional settings could challenge estimation accuracy.
  - Future work should explore regularization or dimensionality reduction approaches suitable for high-dimensional covariates.