Causal Inference on Quantiles in High Dimensions: A Bayesian Approach

Duong Trinh

Research Summary

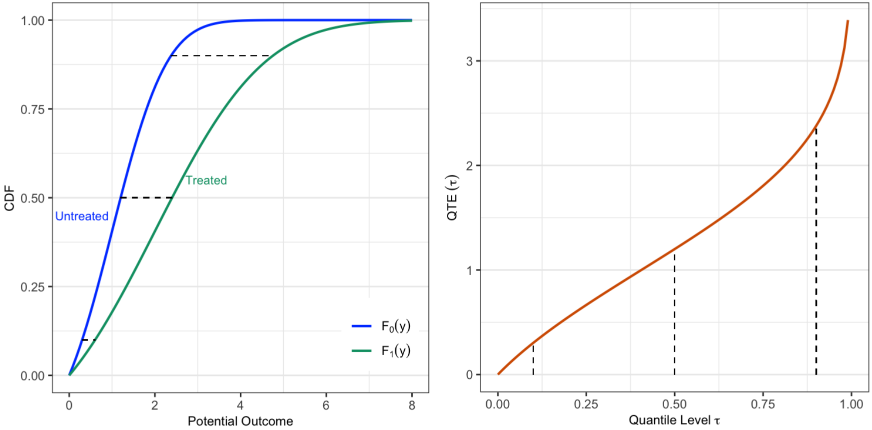
June 2024

University of Glasgow

Introduction

- **Q:** How can we investigate causal effects of an intervention on the entire *distribution* of outcomes when randomized experiments are unavailable?
- *Quantile treatment effects* matter: if impact is heterogeneous, average treatment effect may hide large positive and negative impacts.
- Relevant applications entail social welfare implications: financial interventions, education programs, public health policies, etc.
- I develop *Bayesian* tools to estimate quantile treatment effects in an *observational study* with potentially high-dimensional covariates.

Beyond Average: Quantile Treatment Effects



- For each quantile level $\tau \in [0, 1]$

$$QTE(\tau) := F_1^{-1}(\tau) - F_0^{-1}(\tau) = q_1(\tau) - q_0(\tau)$$

Identification: Conditional on Observables

- **Q:** Can't observe counterfactual outcomes, how do we even identify QTEs?

$$\underbrace{Y(1), Y(0)}_{\text{potential outcomes}} \perp \underbrace{T}_{\text{treatment}} \mid \underbrace{\mathbf{X}}_{\text{controls}}$$

$$\implies F_t(y) = \int_{\mathcal{X}} \underbrace{G(y \mid T = t, \mathbf{X} = \mathbf{x})}_{\text{conditional distribution}} dF_{\mathbf{X}}(\mathbf{x}), \quad \text{for } t \in \{0, 1\}.$$

- **Remained obstacles:**
  - The number of *possible* controls is large, but *specific* controls needed are unknown.
  - Conditional distribution is itself a *complex* function.

  $\implies$ this forces us to consider high dimensions.

This Paper

♣ **Main contribution**:
A Bayesian Analog of Doubly Robust (BADR) approach for estimation and inference on unconditional QTEs in presence of potentially high-dimensional covariates.

- *Double Robustness*
  utilize both treatment assignment model and outcome model to provide *double* protection against model misspecification.
- *Adaptability to High Dimensions*
  incorporate Bayesian flexible data-driven methods to accommodate high-dimensional covariates and non-linear relationships while achieving proper uncertainty quantification.

Related Literature and Contributions

Doubly Robust Estimators

Frequentist:
Scharfstein et al. (1999), Bang and Robins (2005), Farrell (2015), Van Der Laan and Rubin (2006), Chernozhukov et al. (2022), etc.

Bayesian:
Saarela et al. (2016), Stephens et al. (2023), Breunig et al. (2022), Luo et al. (2023), Antonelli et al. (2022), Shin and Antonelli (2023), etc.

ATE; CATE

Causal Machine Learning

Bayesian:
Confounding Adjustment (Hahn et al., 2018; Antonelli et al., 2019) with shrinkage priors, BART-based causal modeling (Hill, 2011; Spertus and Normand, 2018; Hahn et al., 2020), etc.

Frequentist:
Post-Double-Selection Lasso (Belloni et al., 2014), Double/De-biased Machine Learning (Chernozhukov et al., 2018), Approximate Residual Balancing (Athey et al., 2018), etc.

This paper: BADR-QTE

Bayesian:
Xu et al. (2018): Bayesian Non-parametric Estimation.

Frequentist:
Firpo (2007): Inverse Probability Weighted,
Zhang et al. (2012): Augmented Inverse Probability Weighted,
Díaz (2017): Targeted Maximum Likelihood Estimation,
Kallus et al. (2024): Localized Debiased Machine Learning.

Unconditional QTE *(selection-on-observables)*

Outline

Outline

Setup

- T and Y are a binary treatment and the outcome of interest;
  $\mathbf{X}$ is a p-dimensional vector of covariates.

- $Y(1)$ and $Y(0)$ are the *treated* and *untreated* potential outcomes;
  $F_1(y)$ and $F_0(y)$ are corresponding cumulative distribution functions.

- For each quantile level $\tau \in [0, 1]$

$$\boxed{QTE(\tau) := F_1^{-1}(\tau) - F_0^{-1}(\tau) = q_1(\tau) - q_0(\tau)}$$

which can be identified from observational data under *Conditional-on-Observables*, *Positivity*, and *SUTVA* assumptions.

Setup

- General QTE estimation problem from observational data involves *nuisance* functions:

  $G(y \mid 1, \mathbf{X}) = \mathbb{P}[Y \leq y \mid T = 1, \mathbf{X}]$ and $G(y \mid 0, \mathbf{X}) = \mathbb{P}[Y \leq y \mid T = 0, \mathbf{X}]$ are the *conditional distributions* of outcome Y given $(T, \mathbf{X})$.

  $\pi(\mathbf{X}) = \mathbb{P}(T = 1 \mid \mathbf{X})$ is the *propensity score* - the probability of receiving active treatment given covariates $\mathbf{X}$.

Doubly Robust Estimator

$$\widehat{QTE}^{dr}(\tau) = \hat{q}_1^{dr}(\tau) - \hat{q}_0^{dr}(\tau), \quad \text{for each } \tau \in [0,1]. \tag{*}$$

- $\hat{q}_1^{dr}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{T_i}{\hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}(Y_i \le q_1) - \hat{G}(q_1 \mid 1, \mathbf{X}_i)\right] + \hat{G}(q_1 \mid 1, \mathbf{X}_i) - \tau = 0. \tag{1}$$

- $\hat{q}_0^{dr}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1 - T_i}{1 - \hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}(Y_i \le q_0) - \hat{G}(q_0 \mid 0, \mathbf{X}_i)\right] + \hat{G}(q_0 \mid 0, \mathbf{X}_i) - \tau = 0. \tag{2}$$

Doubly Robust Estimator

$$\widehat{QTE}^{dr}(\tau) = \hat{q}_1^{dr}(\tau) - \hat{q}_0^{dr}(\tau), \quad \text{for each } \tau \in [0,1]. \tag{*}$$

- $\hat{q}_1^{dr}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{T_i}{\hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}(Y_i \leq q_1) - \hat{G}(q_1 \mid 1, \mathbf{X}_i)\right] + \hat{G}(q_1 \mid 1, \mathbf{X}_i) - \tau = 0. \tag{1}$$

- $\hat{q}_0^{dr}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1 - T_i}{1 - \hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}(Y_i \leq q_0) - \hat{G}(q_0 \mid 0, \mathbf{X}_i)\right] + \hat{G}(q_0 \mid 0, \mathbf{X}_i) - \tau = 0. \tag{2}$$

Doubly Robust Estimator

$$\widehat{QTE}^{dr}(\tau) = \hat{q}_1^{dr}(\tau) - \hat{q}_0^{dr}(\tau), \quad \text{for each } \tau \in [0,1]. \tag{*}$$

- $\hat{\theta}^{dr}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{T_i}{\hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}(Y_i \le \theta) - \hat{\eta}(\mathbf{X}_i, \theta)\right] + \hat{\eta}(\mathbf{X}_i, \theta) - \tau = 0. \tag{1}$$

- $\hat{q}_0^{dr}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1 - T_i}{1 - \hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}(Y_i \le q_0) - \hat{G}(q_0 \mid 0, \mathbf{X}_i)\right] + \hat{G}(q_0 \mid 0, \mathbf{X}_i) - \tau = 0. \tag{2}$$

Doubly Robust Estimator

- **Q:** Why Double Robustness property? What does $\hat{\theta}^{dr}$ estimate?
  - Estimating equation is built upon an efficient influence function (Robins and Rotnitzky, 1995; Tsiatis, 2006; Kennedy, 2022) tailored to quantiles $\rightarrow$ first-order insensitivity to perturbations in nuisances.
  - $\mathbf{1^{st}}$ view: Augmented OR (outcome regression) estimating equation

$$0 = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\eta}(\mathbf{X}_i,\theta) - \tau\right] + \frac{1}{n}\sum_{i=1}^{n}\frac{T_i\left[\mathbb{1}(Y_i \le \theta) - \hat{\eta}(\mathbf{X}_i,\theta)\right]}{\hat{\pi}(\mathbf{X}_i)}.$$

  - $\mathbf{2^{nd}}$ view: Augmented IPW (inverse-probability-weighting) estimating equation

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{T_i\left[\mathbb{1}(Y_i \le \theta) - \tau\right]}{\hat{\pi}(\mathbf{X_i})} + \frac{1}{n}\sum_{i=1}^{n}\frac{\left[T_i - \hat{\pi}(\mathbf{X_i})\right]\left[\hat{\eta}(\mathbf{X}_i,\theta) - \tau\right]}{\hat{\pi}(\mathbf{X_i})}.$$

Doubly Robust Estimator

- **Q:** Why Double Robustness property? What does $\hat{\theta}^{dr}$ estimate?

  - $\mathbf{1^{st}}$ view: Augmented OR estimating equation; *when* $n \to \infty$, if outcome regression model is corrected specified

$$0 = \mathbb{E}\left[\eta_{\text{true}}(\mathbf{X}, \theta) - \tau\right] + \overbrace{\mathbb{E}\left\{\frac{T\left[\mathbb{1}(Y \leq \theta) - \eta_{\text{true}}(\mathbf{X}, \theta)\right]}{\pi(\mathbf{X})}\right\}}^{\mathbf{0}}.$$

  - $\mathbf{2^{nd}}$ view: Augmented IPW estimating equation; *when* $n \to \infty$, if treatment assignment model is corrected specified

$$0 = \mathbb{E}\left\{\frac{T\left[\mathbb{1}(Y \leq \theta) - \tau\right]}{\pi_{\text{true}}(\mathbf{X})}\right\} + \overbrace{\mathbb{E}\left\{\frac{\left[T - \pi_{\text{true}}(\mathbf{X})\right]\left[\eta(\mathbf{X}, \theta) - \tau\right]}{\pi_{\text{true}}(\mathbf{X})}\right\}}^{\mathbf{0}}.$$

➡ $\hat{\theta}^{dr}$ remains consistent even if *one* of the treatment assignment model or the outcome regression model is misspecified.

High-dimensional Modeling

1. Propensity score: $\pi(\mathbf{X}) = \mathbb{P}(T = 1 \mid \mathbf{X})$

   Fit a binary **Bayesian Additive Regression Trees (BART)** model:

$$\mathbb{P}(T_i = 1 \mid \mathbf{X}_i) = H\left[f_{\text{BART}}(\mathbf{X}_i)\right], \tag{1}$$

   where H is either the probit link or logistic link function, and

$$f_{\text{BART}}(\mathbf{X}_i) = \sum_{m=1}^{M} f_{\text{tree}}\left(\mathbf{X}_i; \Gamma_m, \mu_m\right) \text{ are sum of } M \text{ Bayesian regression trees.}$$

Hierarchical priors:
1. independent priors over tree structures Γ_m
2. independent priors over leaf parameters μ_m, given the tree.

B *posterior* draws $b = 1, \ldots, B$:

$$\pi^{(b)}(\mathbf{X}_i) = H\left[\sum_{m=1}^{M} f_{\text{tree}}\left(\mathbf{X}_i; \Gamma_m^{(b)}, \mu_m^{(b)}\right)\right]$$

High-dimensional Modeling

2. Conditional distributions:
   $G(y \mid 1, \mathbf{X}) = \mathbb{P}[Y \leq y \mid T = 1, \mathbf{X}]$ and $G(y \mid 0, \mathbf{X}) = \mathbb{P}[Y \leq y \mid T = 0, \mathbf{X}]$.

   N1. General connection of the conditional distribution with the conditional quantiles

$$F_{Y|\mathbf{x}}(y) = \int_0^1 \mathbb{1}\{\mathcal{Q}_{Y|\mathbf{x}}(\tau) \leq y\} d\tau \overset{\text{discretization}}{\Longrightarrow} \hat{F}_{Y|\mathbf{x}}(y) = \epsilon + \sum_{s=1}^{S} \delta_s \mathbb{1}\{\hat{\mathcal{Q}}_{Y|\mathbf{x}}(\tau_s) \leq y\},$$

   N2. Conditional quantiles are predictable by fitting **Bayesian Quantile Regression** for S quantile levels τ_s

$$Y_i = \mathbf{X}_i \beta(\tau_s) + \epsilon_i(\tau_s) \text{ where } \epsilon_i(\tau_s) \sim \mathcal{A}\text{ssymetric } \mathcal{L}\text{aplace}\left(\tau, 0, \sigma(\tau_s)\right).$$

Hierarchical priors:
1. priors for error distribution: location $\&$ scale.
2. priors for regression coefficients:
shrinkage priors can be incorporated.

B *posterior* draws $b = 1, \ldots, B$:

$$\mathcal{Q}_i^{(b)}(\tau_s) \; = \mathbf{X}_i \beta^{(b)}(\tau_s)$$
$$G^{(b)}(y \mid 1, \mathbf{X}_i); G^{(b)}(y \mid 0, \mathbf{X}_i)$$

BADR-QTE Algorithm

Algorithm 1: Bayesian Analog Doubly Robust (BADR) estimation for QTEs

Data: $\{Y_i, T_i, \mathbf{X}_i\}_{i=1}^n, \tau \in (0,1)$

Result: $\widehat{QTE}^{dr}(\tau)$

Fit *treatment assignment model* on $\{T_i, \mathbf{X}_i\}_{i=1}^n$ and obtain B posterior samples $\{\pi^{(b)}(\mathbf{X})\}_{b=1}^B$

for $t = 0, 1$ **do**

  Fit *outcome model* on $\{Y_i, \mathbf{X}_i\}_{i:T_i=t}$ and obtain B posterior samples $\{G^{(b)}(y \mid t, \mathbf{X})\}_{b=1}^B$

end

for $b = 1, \ldots, B$ **do**

  Solve $q_1^{(b)}(\tau), q_0^{(b)}(\tau)$ using $\pi^{(b)}(\mathbf{X})$ and $G^{(b)}(y \mid t, \mathbf{X}_i)$, according to equations (1) and (2).

  Calculate $QTE^{(b)}(\tau) = q_1^{(b)}(\tau) - q_0^{(b)}(\tau)$.

end

Calculate $\widehat{QTE}^{dr}(\tau) = \frac{1}{B}\sum_{b=1}^B QTE^{(b)}(\tau)$

Algorithm 2

Outline

Monte Carlo Study

♣ Benchmark:
- Naive estimator (`Naive`)

♣ Proposed estimators:
- Bayesian Doubly Robust estimator (`BDR`)
- Bayesian Doubly Robust estimator with a shrinkage prior (`BDRS`)

♣ Popular alternative estimators:
- Firpo's (2007) Inverse Probability Weighted (`FIPW`)
- Díaz's (2017) Targeted Maximum Likelihood Estimation (`TMLE`)
- Xu et al.'s (2018) Bayesian Non-parametric (`BNP`)
- Kallus et al.'s (2024) Localized Debiased Machine Learning (`LDML`)

Data Generating Process

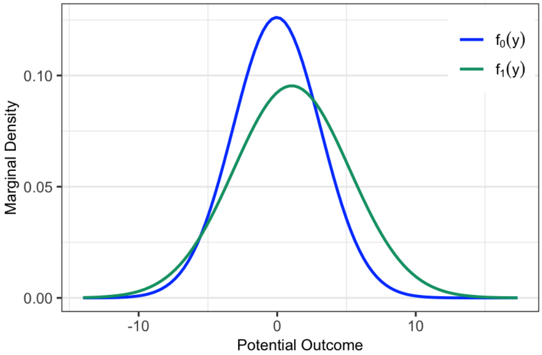
- Linear setting with varying p/N:

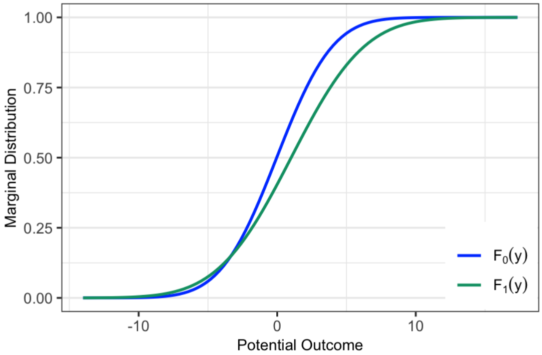
$$p = 40; N \in \{100, 500, 1000\}$$

$$\begin{aligned} X_1, X_2, \ldots, X_{40} &\overset{iid}{\sim} \text{Normal}\,(0, 1) \\ T \mid \mathbf{X} &\sim \text{Bernoulli}\,(\pi(\mathbf{X})) \\ Y(0) \mid \mathbf{X} &\sim \text{Normal}\left(\mu(\mathbf{X}), 2.5^2\right) \\ Y(1) \mid \mathbf{X} &\sim \text{Normal}\left(1 + \mu(\mathbf{X}), 3.75^2\right) \\ Y &= T \times Y(1) + (1 - T) \times Y(0) \end{aligned}$$

where
$$\begin{aligned} \pi(\mathbf{X}) &= \{1 + \exp\left[-(X_1 + X_2 + X_3)\right]\}^{-1}, \\ \mu(\mathbf{X}) &= X_1 + X_2 + X_4 + X_5 \end{aligned}$$

Data Generating Process





- True QTEs (population parameters of interest)

$$\Delta_{0.10} = -0.34, \Delta_{0.25} = 0.29, \Delta_{0.5} = 1, \Delta_{0.75} = 1.70, \Delta_{0.90} = 2.34$$

Table 1. Comparison of point estimates for QTEs and 95% CI across 100 replicates. $(N = 1000, p = 40)$

	Percentiles				
	10th	25th	50th	75th	90th
True QTEs	**-0.34**	**0.29**	**1.00**	**1.71**	**2.34**
Methods					
BDR	-0.37 (-1.20, 0.46)	0.30 (-0.38, 0.97)	0.95 (0.37, 1.53)	1.64 (1.00, 2.27)	2.30 (1.50, 3.11)
BDRS	-0.34 (-1.16, 0.48)	0.31 (-0.34, 0.96)	0.95 (0.38, 1.53)	1.65 (1.03, 2.27)	2.34 (1.56, 3.12)
BNP	0.92 (0.24, 1.58)	1.55 (1.01, 2.09)	2.24 (1.74, 2.74)	2.93 (2.39, 3.48)	3.59 (2.91, 4.25)
LDML	0.29 (-0.74, 1.31)	0.96 (-0.08, 2.01)	1.63 (0.25, 3.01)	2.35 (-0.00, 4.70)	3.04 (-1.74, 7.82)
TMLE	-0.38 (-1.63, 0.86)	0.39 (-0.44, 1.22)	1.07 (0.36, 1.78)	1.75 (0.94, 2.56)	2.32 (1.15, 3.49)
FIPW	-0.38 (-1.71, 0.96)	0.27 (-0.93, 1.47)	0.92 (-0.18, 2.02)	1.64 (0.43, 2.85)	2.25 (0.88, 3.63)
Naive	0.94 (0.14, 1.74)	1.58 (0.97, 2.19)	2.25 (1.67, 2.84)	2.97 (2.35, 3.59)	3.65 (2.86, 4.43)

Figure 1.
Sampling distributions of bias for QTEs across 100 replicates.

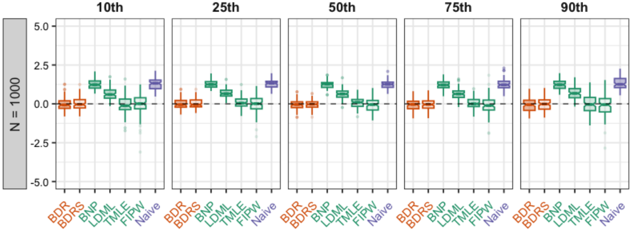
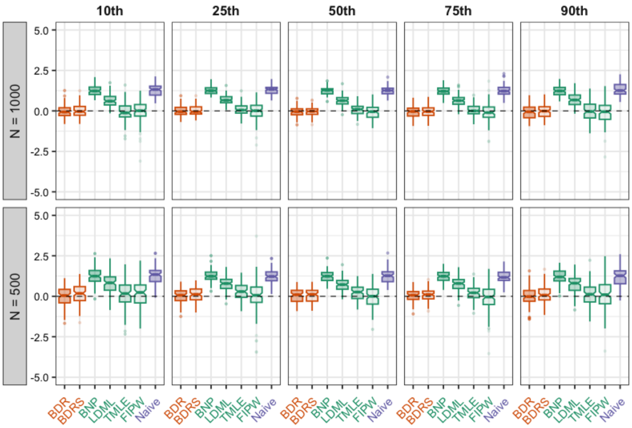


Figure 1.
Sampling distributions of bias for QTEs across 100 replicates.



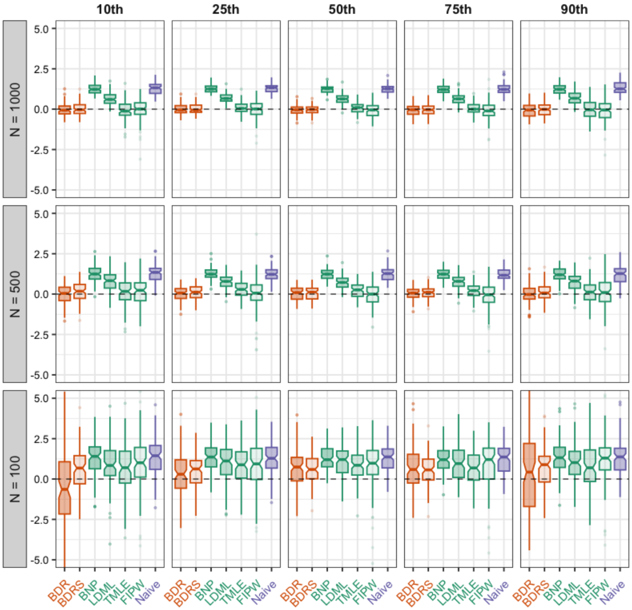


Figure 1.
Sampling distributions of bias for QTEs across 100 replicates.

Table 2. Average Bias.

Percentiles	N	Estimation Methods						
		BDR	**BDRS**	**BNP**	**LDML**	**TMLE**	**FIPW**	**Naive**
10th	1000	-0.022	**0.001**	1.261	0.63	-0.041	-0.034	1.282
	500	**0.008**	0.153	1.261	0.794	0.121	0.102	1.269
	100	-0.659	**0.56**	1.267	0.928	0.724	0.901	1.398
25th	1000	**0.003**	0.017	1.261	0.669	0.1	-0.025	1.288
	500	0.047	0.103	1.245	0.764	0.266	**-0.035**	1.237
	100	**0.265**	0.49	1.32	1.03	0.793	0.794	1.296
50th	1000	-0.049	**-0.045**	1.24	0.632	0.071	-0.08	1.25
	500	**0.035**	0.053	1.209	0.718	0.215	-0.054	1.241
	100	0.696	**0.595**	1.284	1.073	0.85	0.852	1.295
75th	1000	-0.071	-0.057	1.228	0.641	**0.045**	-0.068	1.266
	500	**0.022**	0.071	1.192	0.743	0.172	-0.111	1.226
	100	0.809	**0.574**	1.24	0.941	0.673	0.822	1.179
90th	1000	-0.039	**-0.006**	1.247	0.696	-0.024	-0.089	1.302
	500	**-0.008**	0.096	1.21	0.736	0.094	0.07	1.206
	100	**0.564**	0.712	1.293	1.036	0.825	1.093	1.314

Table 3. Relative Mean Absolute Error.

		Estimation Methods					
Percentiles	N	BDR	BDRS	BNP	LDML	TMLE	FIPW
10th	1000	1.67	1.645	**0.99**	1.169	1.741	1.741
	500	1.571	1.462	**0.964**	1.078	1.547	1.634
	100	2.449	1.194	**0.945**	1.123	1.272	1.346
25th	1000	0.999	0.996	0.969	**0.862**	0.978	1.077
	500	1.135	1.117	1.017	**0.947**	1.07	1.391
	100	1.182	0.986	**0.93**	1	0.992	1.107
50th	1000	**0.628**	**0.628**	0.994	0.739	0.64	0.688
	500	0.673	**0.666**	0.976	0.778	0.676	0.694
	100	0.876	**0.81**	0.964	0.943	0.876	1.012
75th	1000	**0.477**	0.479	0.981	0.704	0.513	0.523
	500	**0.547**	0.56	0.981	0.776	0.592	0.604
	100	0.927	**0.771**	1.004	0.912	0.81	0.992
90th	1000	**0.519**	0.529	0.979	0.771	0.539	0.534
	500	**0.572**	0.596	1.002	0.819	0.623	0.669
	100	1.109	**0.822**	0.988	0.919	0.851	1.06

Remarks

- `BDR` and `BDRS` showcase a substantial improvement in bias reduction for QTE estimates, proving beneficial in modeling the conditional distribution of potential outcomes given confounders.
- `BDRS` provides extra merit thanks to its adaptation to high-dimensional covariates by augmenting with shrinkage priors.

Outline

Welfare Impact of Household Financial Access

♣ **Q:** Unconditional quantile treatment effects of *loan access* on household *consumption* and *business outcome*?

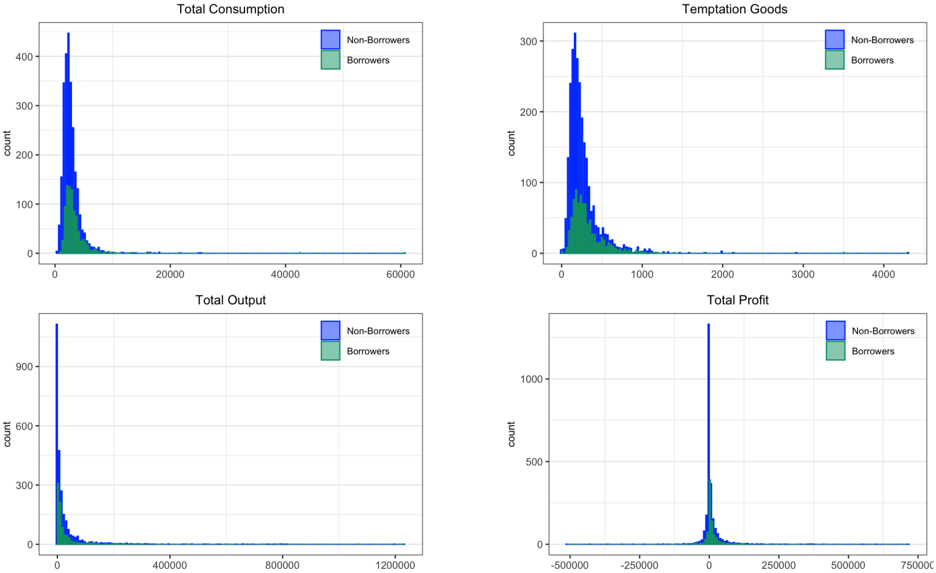
- Dynamic general equilibrium: Borrowers differ in their investment opportunities and productivity. $\rightarrow$ Potential winners and losers to financial market expansion (Kaboski and Townsend, 2011; Banerjee, 2013).
- Evaluation is often limited in *average* treatment effect or *randomization*.
- Revisit Crépon et al.'s (2015) microcredit study across 162 Moroccan villages.

Household Consumption and Business Outcomes

Table 4. Summary Statistics of Household Outcomes.

	Borrowers		Non-borrowers		Borrowers – Non-borrowers		
Outcome variables	Mean	St.Dev.	Mean	St.Dev.	Diff.Mean		t-statistic
(in MAD)							
Total Consumption	3268.62	(2956.01)	2863.49	(1792.97)	405.13	***	3.82
Temptation Goods	312.33	(229.91)	270.31	(219.33)	42.01	***	4.73
Total Output	32672.06	(85071.58)	30885.38	(85939.63)	1786.68		0.54
Total Profit	10081.86	(37986.07)	8409.95	(45277.88)	1671.91		1.07

Household Consumption and Business Outcomes



Control Variables

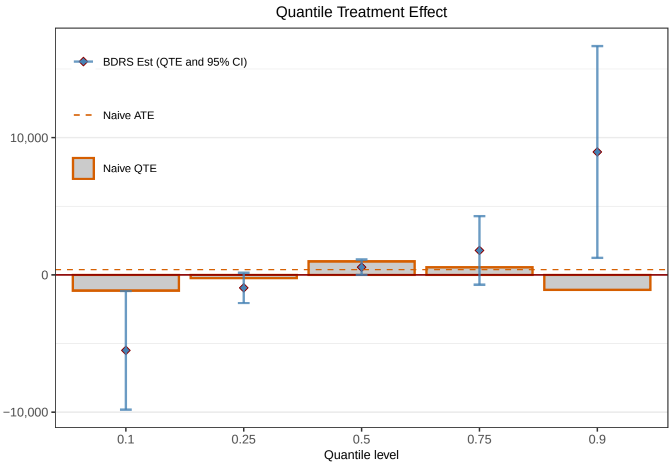
Table 5. Covariate Balance between Borrowers and Non-borrowers.

	Borrowers	Non-borrowers	Borrowers – Non-borrowers		
Control variables	Mean (sd)	Mean (sd)	Diff.Mean		t-statistic
Head age	49.01 (15.62)	48.53 (15.93)	0.49		0.79
Head with no education	0.68 (0.47)	0.68 (0.47)	0.00		0.05
Number of members	6.06 (2.46)	5.54 (2.48)	0.52	***	5.36
Number of adults	4.02 (2.01)	3.71 (1.92)	0.31	***	3.99
Number of members aged 6-16	1.36 (1.30)	1.19 (1.27)	0.16	**	3.23
Declared animal husbandry activities	0.59 (0.49)	0.57 (0.50)	0.02		1.23
Declared non-agricultural activities	0.23 (0.42)	0.18 (0.38)	0.05	**	3.17
Spouse of head responded	0.05 (0.23)	0.09 (0.29)	-0.04	***	-3.96
Member responded	0.05 (0.21)	0.05 (0.22)	0.00		-0.36
Microcredit availability	0.55 (0.50)	0.47 (0.50)	0.07	***	3.73

Table 6. Quantile Treatment Effects of Loan Access on Household Outcomes. Graphs

		BDRS			Naive
Outcomes	Percentiles	QTEs	Upper bound	Lower bound	QTEs
Total Consumption	10th	20.09	1469.21	-1429.02	232.80
	25th	9.24	173.51	-155.03	173.46
	50th	79.95	229.59	-69.69	229.75
	75th	132.22	273.97	-9.53	286.68
	90th	237.7	543.35	-67.96	685.44
Temptation Goods	10th	-8.69	65.66	-83.04	17.38
	25th	13.04	29.87	-3.80	21.73
	50th	**30.41**	45.96	14.87	43.45
	75th	**47.79**	79.22	16.37	60.83
	90th	**78.21**	129.14	27.28	78.21
Total Output	10th	0	13146.95	-13146.95	0.00
	25th	-330	19.77	-679.77	1093.45
	50th	50	1385.99	-1285.99	1787.50
	75th	1666	6933.21	-3601.21	2771.62
	90th	**27360**	52964.20	1755.80	2744.04
Total Profit	10th	**-5500**	-1183.54	-9816.46	-1142.70
	25th	-945	158.83	-2048.83	-241.88
	50th	**561**	1117.73	4.27	979.12
	75th	1780.77	4273.91	-712.38	549.37
	90th	**8954.38**	16664.23	1244.52	-1086.35

QTEs on Total Profit



- Evidence of systematic harm in terms of *total profit*: a segment of households may experience adverse effects that extend the lower tail of the distribution leftward.

Outline

Conclusion

1. Quantile Treatment Effect, while posing technical challenges, is a worthwhile causal estimand to uncover treatment effect heterogeneity and distributional impacts.

2. This paper proposes flexible BADR-QTE estimation framework for observational studies:
   - *double robustness* & *adaptability* to high-dimensional covariates.
   - *substantial bias reduction* compared to popular alternative estimators.
   - *microcredit application*: value added in characterizing heterogeneous distributional impacts on outcomes and detecting changes in household inequality.

3. Future extensions: improve computation for inference; embrace unmeasured confounding.

Thank you!

https://duongtrinhss.github.io/

Bayesian Additive Regression Tree (BART)

- The prior of BART is specified for three components:
  - The ensemble structure $\{\Gamma_m\}_{m=1}^M$

$$\Pr(\text{split} \mid \overbrace{d}^{\text{tree depth}}) = \alpha(1+d)^{-\beta} \qquad \Rightarrow \qquad \Gamma_m \sim P_{\alpha,\beta}$$

where $\alpha \in (0,1), \beta \in (0,\infty)$.

  - The parameters $\{\mu_m\}_{m=1}^M$ associated with the terminal nodes given $\{\Gamma_m\}_{m=1}^M$

$$\mu_{m,l} \overset{iid}{\sim} N(0,v) \tag{2}$$

  - The error variance σ^2 that is independent with the former two

$$\sigma^2 \sim \text{Inv-Gamma}(r,s) \tag{3}$$

Bayesian Quantile Regression (BQR)

For i = 1, . . . , n

$$Y_i = \mathbf{X}_i \beta_{(\tau)} + \epsilon_{i,(\tau)}$$
$$Y_i = \mathbf{X}_i \beta_{(\tau)} + \theta_{(\tau)} z_{i,(\tau)} + \kappa_{(\tau)} \sqrt{\sigma_{(\tau)} z_{i,(\tau)}} u_i$$

- Prior specification

$$\beta_{(\tau)} \sim N_p \left(0, \lambda \times \mathbf{I}_p\right), \tag{4}$$
$$z_{i,(\tau)} \sim \text{Exp}\left(\sigma_{(\tau)}\right) \qquad \forall i = 1, \ldots, n, \tag{5}$$
$$\sigma_{(\tau)} \sim \text{Inv-Gamma}\left(r_{0,(\tau)}, s_{0,(\tau)}\right), \tag{6}$$

where λ is fixed and known for all τ.

- The conditional posteriors are of the form

$$\beta_{(\tau)} \mid \bullet \sim N_p\left(\mu_{\beta,(\tau)}, \mathbf{\Sigma}_{\beta,(\tau)}\right), \tag{7}$$

$$z_{i,(\tau)} \mid \bullet \sim \text{GIG}\left(\frac{1}{2}, a_{i,(\tau)}, b_{i,(\tau)}\right), \quad \forall i = 1, \dots, n, \tag{8}$$

$$\sigma_{(\tau)} \mid \bullet \sim \text{Inv-Gamma}\left(r_{\sigma,(\tau)}, s_{\sigma,(\tau)}\right), \tag{9}$$

where

$$\mathbf{\Sigma}_{\beta,(\tau)} = \left(\mathbf{X}^{\top}\mathbf{U}^{-1}\mathbf{X} + \mathbf{\Sigma}_{0,(\tau)}^{-1}\right)^{-1} \text{ and } \mu_{\beta,(\tau)} = \mathbf{\Sigma}_{\beta,(\tau)} \times \mathbf{X}^{\top}\mathbf{U}^{-1}\left(\mathbf{Y} - \theta_{(\tau)}\mathbf{z}_{(\tau)}\right),$$

$$\mathbf{U} = \left(\sigma_{(\tau)}\kappa_{(\tau)}^2\right) \times \text{diag}\left(\mathbf{z}_{(\tau)}\right), \quad \mathbf{z}_{(\tau)} = \left(z_{1,(\tau)}, \dots, z_{n,(\tau)}\right)'$$

$$a_{i,(\tau)} = \frac{1}{\sigma(\tau)}\left(2 + \frac{\theta_{(\tau)}^2}{\kappa_{(\tau)}^2}\right) \text{ and } b_{i,(\tau)} = \frac{\left(Y_i - \mathbf{X}_i\beta_{(\tau)}\right)^2}{\sigma_{(\tau)}\kappa_{(\tau)}^2},$$

$$r_{\sigma,(\tau)} = r_{0,(\tau)} + \frac{3n}{2} \text{ and } s_{\sigma,(\tau)} = s_{0,(\tau)} + \sum_{i=1}^{n} \frac{\left(Y_i - \mathbf{X}_i\beta_{(\tau)} - \theta_{(\tau)}z_{i,(\tau)}\right)^2}{2\kappa_{(\tau)}^2 z_{i,(\tau)}} + \sum_{i=1}^{n} z_{i,(\tau)}.$$

Bayesian Adaptive Lasso Quantile Regression

For i = 1, . . . , n

$$Y_i = \mathbf{X}_i\beta_{(\tau)} + \theta_{(\tau)} z_{i,(\tau)} + \kappa_{(\tau)}\sqrt{\sigma_{(\tau)} z_{i,(\tau)}}\, u_i$$

- Hierarchical Priors

$$\beta_{j,(\tau)}, v_{j,(\tau)} \mid \sigma_{(\tau)}, \lambda^2_{j,(\tau)} \sim \frac{1}{\sqrt{2\pi v_{j,(\tau)}}} \text{exp}\left\{-\frac{\beta^2_{j,(\tau)}}{2v_{j,(\tau)}}\right\} \frac{\sigma^{-1}_{(\tau)}}{2\lambda^2_{j,(\tau)}} \text{exp}\left\{\frac{-\sigma^{-1}_{(\tau)}}{2\lambda^2_{j,(\tau)}} v_{j,(\tau)}\right\}, \tag{10}$$

$$\lambda^2_{j,(\tau)} \sim \text{Inv-Gamma}\ \left(c_{0,(\tau)}, d_{0,(\tau)}\right), \tag{11}$$

$$\sigma_{(\tau)} \sim \text{Inv-Gamma}\ \left(r_{0,(\tau)}, s_{0,(\tau)}\right) \tag{12}$$

- The conditional posteriors (Alhamzawi et al., 2012) are of the form

$$z_{i,(\tau)} \mid \bullet \sim \text{GIG}\left(\frac{1}{2}, a_{i,(\tau)}, b_{i,(\tau)}\right), \quad \forall i = 1, \ldots, n, \tag{13}$$

$$\beta_{j,(\tau)} \mid \bullet \sim N\left(\mu_{\beta_j,(\tau)}, \Sigma_{\beta_j,(\tau)}\right), \quad \forall j = 1, \ldots, p, \tag{14}$$

$$v_{j,(\tau)} \mid \bullet \sim \text{GIG}\left(\frac{1}{2}, \frac{\sigma_{(\tau)}^{-1}}{\lambda_{j,(\tau)}^2}, \beta_{j,(\tau)}^2\right), \tag{15}$$

$$\sigma_{(\tau)} \mid \bullet \sim \text{Inv-Gamma}\left(r_{\sigma,(\tau)}, s_{\sigma,(\tau)}\right), \tag{16}$$

$$\lambda_{j,(\tau)}^2 \mid \bullet \sim \text{Inv-Gamma}\left(c_{0,(\tau)} + 1, d_{0,(\tau)} + \sigma_{(\tau)}^{-1} v_{j,(\tau)}/2\right), \tag{17}$$

where

$$a_{i,(\tau)} = \frac{1}{\sigma(\tau)}\left(2 + \frac{\theta_{(\tau)}^2}{\kappa_{(\tau)}^2}\right) \text{ and } b_{i,(\tau)} = \frac{\left(Y_i - \mathbf{X}_i\beta_{(\tau)}\right)^2}{\sigma_{(\tau)}\kappa_{(\tau)}^2},$$

$$\Sigma_{\beta_j,(\tau)} = \left[\left(\sigma\kappa_{(\tau)}^2\right)^{-1}\sum_{i=1}^{n} x_{ij}^2 z_{i,(\tau)}^{-1} + v_{j,(\tau)}^{-1}\right]^{-1},$$

$$\mu_{\beta_j,(\tau)} = \Sigma_{\beta_j,(\tau)}\left(\sigma\kappa_{(\tau)}^2\right)^{-1}\sum_{i=1}^{n}\left(Y_i - \theta_{(\tau)} z_{i,(\tau)} - \sum_{k=1,k\neq j}^{p} x_{ij}\beta_{j,(\tau)}\right) x_{ij}^2 z_{i,(\tau)}^{-1},$$

$$r_{\sigma,(\tau)} = r_{0,(\tau)} + \frac{3n}{2} + p \text{ and } s_{\sigma,(\tau)} = s_{0,(\tau)} + \sum_{i=1}^{n}\frac{\left(Y_i - \mathbf{X}_i\beta_{(\tau)} - \theta_{(\tau)} z_{i,(\tau)}\right)^2}{2\kappa_{(\tau)}^2 z_{i,(\tau)}} + \sum_{i=1}^{n} z_{i,(\tau)} + \sum_{j=1}^{p}\frac{v_{j,(\tau)}}{2\lambda_j^2}.$$

BADR-QTE Algorithm

Algorithm 2: Bayesian Analog Doubly Robust (BADR) estimation for QTEs

Data: $\{Y_i, T_i, \mathbf{X}_i\}_{i=1}^n, \tau \in (0,1)$

Result: $\widehat{QTE}^{dr}(\tau)$

Fit *treatment assignment model* on $\{T_i, \mathbf{X}_i\}_{i=1}^n$ and obtain B posterior samples $\{\pi^{(b)}(\mathbf{X})\}_{b=1}^B$

for $\underline{t = 0, 1}$ **do**

  Fit *outcome model* on $\{Y_i, \mathbf{X}_i\}_{i:T_i=t}$ and obtain B posterior samples $\{G^{(b)}(y \mid t, \mathbf{X})\}_{b=1}^B$

end

Derive posterior mean from B posterior samples

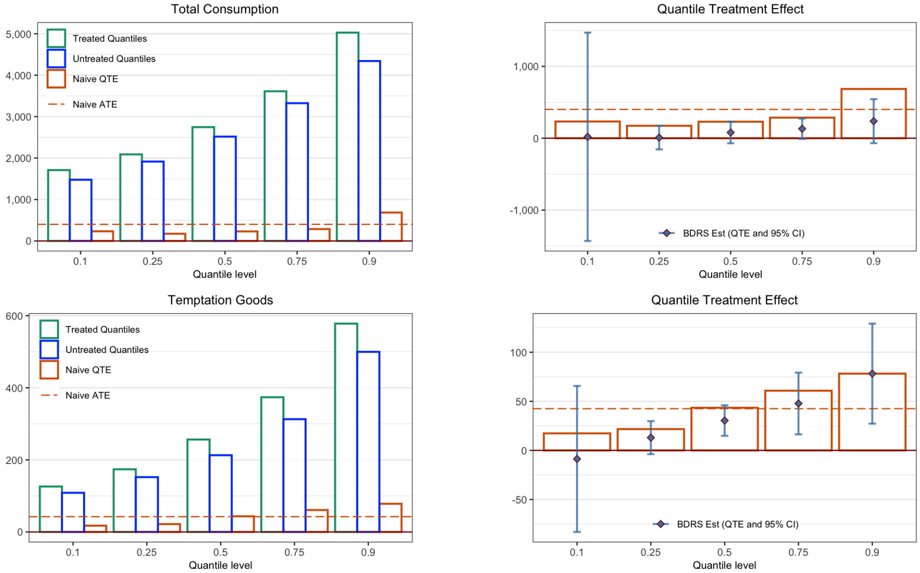
  $\hat{\pi}(\mathbf{X}) = \frac{1}{B}\sum_{b=1}^B \pi^{(b)}(\mathbf{X})$ and $\hat{G}(y \mid t, \mathbf{X}) = \frac{1}{B}\sum_{b=1}^B G^{(b)}(y \mid t, \mathbf{X})$

Solve $\hat{q}_1^{dr}(\tau), \hat{q}_0^{dr}(\tau)$ based on $\hat{\pi}(\mathbf{X})$ and $\hat{G}(y \mid t, \mathbf{X})$, according to equations (1) and (2).

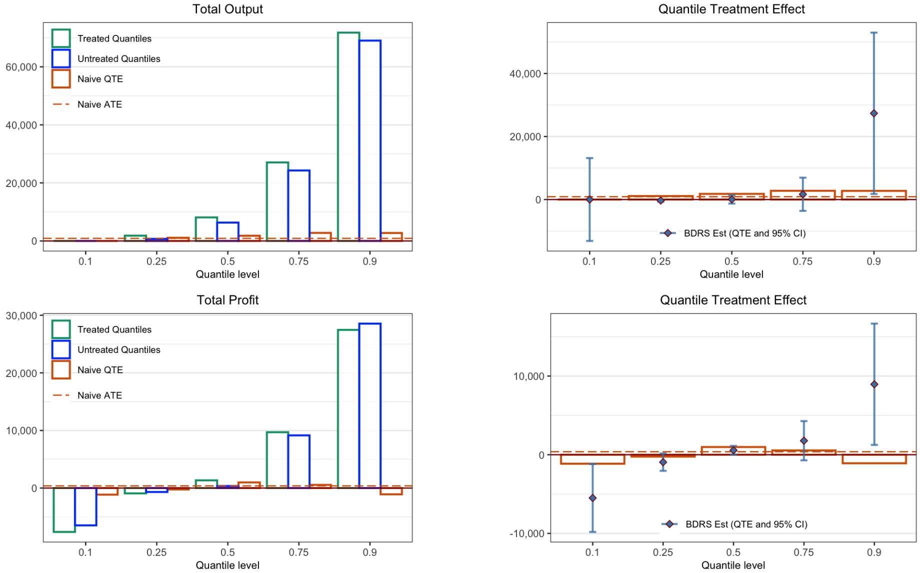
Calculate $\widehat{QTE}^{dr}(\tau) = \hat{q}_1^{dr}(\tau) - \hat{q}_0^{dr}(\tau)$.

Algorithm 1

Empirical Results: QTEs on Consumption back



Empirical Results: QTEs on Business Outcomes back



Reference I

R. Alhamzawi, K. Yu, and D. F. Benoit. Bayesian adaptive lasso quantile regression. Statistical Modelling, 12(3): 279–297, 2012.

J. Antonelli, G. Parmigiani, and F. Dominici. High-dimensional confounding adjustment using continuous spike and slab priors. Bayesian analysis, 14(3):805, 2019.

J. Antonelli, G. Papadogeorgou, and F. Dominici. Causal inference in high dimensions: A marriage between bayesian modeling and good frequentist properties. Biometrics, 78(1):100–114, 2022.

S. Athey, G. W. Imbens, and S. Wager. Approximate residual balancing: debiased inference of average treatment effects in high dimensions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(4):597–623, 2018.

A. V. Banerjee. Microcredit under the microscope: What have we learned in the past two decades, and what do we need to know? Annu. Rev. Econ., 5(1):487–519, 2013.

H. Bang and J. M. Robins. Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4):962–973, 2005.

Reference II

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies, 81(2):608–650, 2014.

C. Breunig, R. Liu, and Z. Yu. Double robust bayesian inference on average treatment effects. arXiv preprint arXiv:2211.16298, 2022.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

V. Chernozhukov, W. K. Newey, and R. Singh. Automatic debiased machine learning of causal and structural effects. Econometrica, 90(3):967–1027, 2022.

B. Crépon, F. Devoto, E. Duflo, and W. Parienté. Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in morocco. American Economic Journal: Applied Economics, 7(1): 123–150, 2015.

I. Díaz. Efficient estimation of quantiles in missing data models. Journal of Statistical Planning and Inference, 190:39–51, 2017.

Reference III

M. H. Farrell. Robust inference on average treatment effects with possibly more covariates than observations. Journal of Econometrics, 189(1):1–23, 2015.

S. Firpo. Efficient semiparametric estimation of quantile treatment effects. Econometrica, 75(1):259–276, 2007.

P. R. Hahn, C. M. Carvalho, D. Puelz, and J. He. Regularization and confounding in linear regression for treatment effect estimation. Bayesian Analysis, 13(1):163–182, 2018.

P. R. Hahn, J. S. Murray, and C. M. Carvalho. Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). Bayesian Analysis, 15(3): 965–1056, 2020.

J. L. Hill. Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics, 20(1):217–240, 2011.

J. P. Kaboski and R. M. Townsend. A structural evaluation of a large-scale quasi-experimental microfinance initiative. Econometrica, 79(5):1357–1406, 2011.

Reference IV

N. Kallus, X. Mao, and M. Uehara. Localized debiased machine learning: Efficient inference on quantile treatment effects and beyond. Journal of Machine Learning Research, 25(16):1–59, 2024.

E. H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint arXiv:2203.06469, 2022.

Y. Luo, D. J. Graham, and E. J. McCoy. Semiparametric bayesian doubly robust causal estimation. Journal of Statistical Planning and Inference, 225:171–187, 2023.

J. M. Robins and A. Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. Journal of the American Statistical Association, 90(429):122–129, 1995.

O. Saarela, L. R. Belzile, and D. A. Stephens. A bayesian view of doubly robust causal inference. Biometrika, 103(3):667–681, 2016.

D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. Journal of the american statistical association, pages 1096–1120, 1999.

H. Shin and J. Antonelli. Improved inference for doubly robust estimators of heterogeneous treatment effects. Biometrics, 2023.

Reference V

J. V. Spertus and S.-L. T. Normand. Bayesian propensity scores for high-dimensional causal inference: A comparison of drug-eluting to bare-metal coronary stents. Biometrical Journal, 60(4):721–733, 2018.

D. A. Stephens, W. S. Nobre, E. E. Moodie, and A. M. Schmidt. Causal inference under mis-specification: Adjustment based on the propensity score (with discussion). Bayesian Analysis, 18(2):639–694, 2023.

A. A. Tsiatis. Semiparametric theory and missing data, volume 4. Springer, 2006.

M. J. Van Der Laan and D. Rubin. Targeted maximum likelihood learning. The international journal of biostatistics, 2(1), 2006.

D. Xu, M. J. Daniels, and A. G. Winterstein. A bayesian nonparametric approach to causal inference on quantiles. Biometrics, 74(3):986–996, 2018.

Z. Zhang, Z. Chen, J. F. Troendle, and J. Zhang. Causal inference on quantiles with an obstetric application. Biometrics, 68(3):697–706, 2012.