Causal Inference on Quantiles in High Dimensions: A Bayesian Approach

Duong Trinh*

March 31, 2025

Job Market Paper

Please click here for the latest version.

Abstract

This paper proposes a novel approach, Bayesian Analog of Doubly Robust (BADR) estimation, to estimate unconditional Quantile Treatment Effects (QTEs) in observational studies. By augmenting the proposed estimator with shrinkage priors, this framework can account for high-dimensional covariates and feature a flexible Bayesian modeling strategy with favorable frequentist properties in finite samples, even when either the treatment assignment or outcome models are misspecified. The proposed approach offers a straightforward and adaptable implementation for incorporating probabilistic machine learning techniques to fit the propensity score and conditional cumulative distribution function, followed by combining posterior draws. This enables the effective handling of high-dimensional covariate spaces or nonlinear relationships to achieve better accuracy and appropriate uncertainty quantification. The simulation results show that BADR estimators yield a substantial improvement in bias reduction for QTE estimates compared with popular alternative estimators found in the literature. We revisit the role of microcredit expansion and loan access on Moroccan household outcomes, demonstrating how the new method adds value in characterizing heterogeneous distributional impacts on outcomes and detecting changes in overall economic inequality, which is also appealing to other applied contexts.

Keywords: Quantile treatment effects; Bayesian modeling; Regularization methods; Doubly robust estimation.

JEL classifications: C11, C31, D63, G51.

*Adam Smith Business School, University of Glasgow, University Avenue, G12 8QQ, United Kingdom. E-mail address: Duong.Trinh@glasgow.ac.uk. I am very grateful to my advisors, Dimitris Korobilis, Kenichi Shimizu, and Santiago Montoya-Blandón for their guidance and encouragement throughout my work on this project. I would like to thank Alberto Ciancio and Cecilia Balocchi for their valuable comments and suggestions, as well as seminar participants at the University of Glasgow, the 5th Panmure House Annual PhD Workshop in Edinburgh and the 2024 IAAE Annual Conference in Thessaloniki for helpful discussions. All remaining errors are mine.

1 Introduction

When evaluating the causal effect of policy interventions, the distributional impact appeals to researchers and policymakers rather than the average impact alone. It helps to gain more comprehensive and nuanced understanding of the complex effects, ultimately leading to more effective decision-making. In many instances, uniform policies may benefit certain individuals while adversely affecting others. If the effects are considerably heterogeneous, the average treatment effect may not be a sufficient measure, as it likely masks substantial positive and negative effects. Consequently, it is crucial to determine whether certain individuals are worse off as a result of the policy. Even if multiple programs generate positive effects for all individuals, the one that offers the greatest benefits to those at the lower tail of the distribution of the outcome variable might be the most favorable. To illustrate, consider two job training programs with identical mean net impact that is positive. The first program, which increases wages at the bottom of the wage distribution, would be more appreciated than the second program, which only raises the top of the wage distribution. This necessitates the advancement of econometric techniques to enable studies on distributional treatment effects in the presence of heterogeneity. This goal has received special interest and has become increasingly relevant in economic applications[1].

Sets of quantile treatment effects (QTEs) can characterize the heterogeneous impacts of the treatment on different points of the outcome distribution. With a binary treatment, as originally defined by Doksum (1974) and Lehmann et al. (1974), QTEs measure the difference between the unconditional quantiles of the potential outcome distribution under treatment and the potential outcome distribution under nontreatment. Put differently, this captures any difference between the two cumulative distribution functions of treated and untreated potential outcomes. Moreover, the quantile method may be employed, even by those not primarily interested in distributional consequences, to enhance the robustness of their analysis. This is particularly relevant in light of the well-established fact that median regression is more resistant to outliers than mean regression, while many economic data sets involve heavy tails. One notable example in development economics is household welfare measures, including consumption and business outcomes.

The difference between these two unconditional distributions of the potential outcomes itself might be appealing to policymakers. This reflects the change in the distribution function as a whole when the treatment could be exogenously shifted between two distinct counterfactual scenarios: universal treatment and no treatment. As the entire distribution function often yields insights into inequality or social welfare analysis, computing QTEs serves as a convenient way to summarise noteworthy aspects. For instance, this enables the detection of changes in overall inequality in the distribution of outcomes, which is a critical concern given the potentially negative consequences of social and economic inequality in the contemporary world[2].

By definition, QTEs can reveal heterogeneity in the causal effects on different quantiles. However, individual results are only interpretable under a rank preservation assumption on the underlying

[1] A variety of relevant applications include, but are not limited to, financial interventions (Meager 2022; Callaway and Li 2019), educational subsidies (Duflo, Dupas, and Kremer 2021), public health policies (Schiele and Schmitz 2016), and local migration incentives (Chetty, Hendren, and Katz 2016; Bryan, Chowdhury, and Mobarak 2014); all entail social welfare implications and garner substantial attention in public discourse.

[2] This interest is at the core of the econometric literature strand on distributional counterfactual analysis (see, e.g., Firpo and Pinto 2016; Chernozhukov, Fernández-Val, and Melly 2013; Rothe 2010).

treatment effect distribution. This assumption asserts that an observed individual would maintain their position (rank) in the distribution regardless of their treatment status. As a result, the set of quantile treatment effects is equivalent to the quantiles of the distribution of individual treatment effects. Nonetheless, rank preservation is a strong assumption because it requires the relative value of the potential outcome for a given individual to be unchanged, whether that individual is treated or untreated. Even when rank preservation is violated, heterogeneity in the effects across various quantiles shows evidence of heterogeneity in these individual effects, making QTEs remain a meaningful parameter for policy purposes (see, e.g., Meager 2022; Angrist and Pischke 2009).

In this paper, the primary focus is on unconditional QTEs, which are separate from conditional QTEs. Although both are standard parameters of interest in the program evaluation literature, it is important to highlight the distinction between them. An unconditional (marginal) quantile function is a one-dimensional function of the quantile level τ only. Defined as the difference between the unconditional quantiles of the treated and untreated potential outcome distributions, unconditional QTEs describe the effects of treatment status on the overall outcome distribution without conditioning on the covariates. In contrast, conditional quantile functions are multi-dimensional, depending on not only a chosen quantile level but also values of the covariates. Conditional QTEs thereby express the effects on the outcome distribution within sub-populations characterized by covariate values. More specifically, an individual may rank high in the unconditional distribution of the outcome, meanwhile possessing a low rank in the conditional distribution of the outcome. This is possible if that person has values of observed characteristics that are associated with a large value of outcome overall, yet within the group of people sharing identical values of the observed characteristics, he or she has a comparatively low outcome[3]. Conditional QTEs enable examination of the heterogeneity of the effects with respect to the observables; however, they might be sensitive to the choice of covariates to be included. Unconditional QTEs, on the other hand, aggregate the conditional effects across the entire population, thereby being more easily conveyed to the policymakers and the public, at the cost of not providing any information about the relationship between the covariates and the outcome. Further discussion can be found in surveys by Glewwe and Todd (2022) and Frölich and Melly (2013). The unconditional quantile treatment effects are appropriate estimands to focus on when the ultimate objective is related to the marginal distribution, for example, the welfare of the (unconditionally) poor. This unconditional effect has been a central estimand of interest in the literature on micro-credit expansion and housing outcomes, reinforcing the need for robust methods suited to recovering and exploring heterogeneity in these effects.

When the target causal estimand is an unconditional quantile treatment effect (QTE), several identification strategies have been developed in the literature. One common approach relies on the assumption of exogenous treatment, typically in the form of a randomized controlled experiment where all participants comply with their treatment assignment. In this ideal scenario, implementing an unconditional QTE estimator is a straightforward process, similar to estimating the average treatment effect (ATE) directly from the realized outcomes of control and treatment groups. However, when such experimental data is unavailable, inferring causal relationships from observational data poses challenges because the observed treatment status is not assigned randomly. This gives rise to the

[3]Consider a simple example involving wages and years of education, the median income of all individuals with doctoral degrees may be greater than the top quantile for high school dropouts, presuming a strong positive association between education levels and earnings.

second approach based on the *selection-on-observables* assumption, which implies that the treatment is as good as randomly assigned once we condition on observables. It is worth noting that, although our ultimate goal is to obtain an unconditional QTE, covariate information serves to correctly identify the unconditional quantiles and remove selection bias.

In this paper, we maintain the identifying assumption of *selection-on-observables*, which is widely applicable to empirical studies in economics. This is due to the fact that randomized controlled trials (RCTs) are often intricate and expensive, rendering them infeasible in many cases. In addition, this assumption is justifiable in various contexts, such as when the treatment is randomly allocated within demographic groups. As elaborated in the subsequent sections, employing covariates for the sake of identification involves identifying the entire conditional cumulative distribution function (CDF) of each potential outcome conditional on potentially high-dimensional covariates. This CDF is then a nuisance function for the identification of QTEs. More often than not, applied researchers encounter a vast set of possible covariates, but they are uncertain about which specific ones are necessary to control for when recovering treatment effects. In addition, the conditional CDF can itself be a complex function. This necessitates the consideration of high-dimensional models to estimate quantile treatment effects.

This paper aims to circumvent such obstacles and contribute to the emerging econometric literature on identification and estimation of QTEs. We propose a novel framework, the Bayesian Analog of Doubly Robust (BADR) approach, for estimating QTEs in an observational study while accounting for the presence of potentially high-dimensional covariates. Briefly, we employ Bayesian techniques to specify and estimate both the treatment assignment and the outcome models, obtaining posterior draws that are then plugged into the doubly robust estimator for QTEs. This estimator is derived as the solution to efficient influence functions, leading to its double-robustness property. The resulting BADR framework comprises two ingredients. First, to effectively accommodate high-dimensional covariates and nonlinear relationships while achieving proper uncertainty quantification, we incorporate various Bayesian regularization methods including sparsity-inducing priors and Bayesian nonparametric methods to generate auxiliary estimators for both the propensity score and the conditional distribution of the outcome variable. We leverage multiple Bayesian quantile regressions to address the unique challenge in quantile estimation, which differs from previous studies on ATE. Second, our method provides double protection against model specification by employing posterior predictive distributions of parameters from both the treatment assignment and outcome models. In the absence of high-dimensional covariates, this approach collapses to a doubly-robust Bayesian estimator for the QTEs without shrinkage priors, which itself has not been explored in previous literature. Overall, the proposed strategy enables us to obtain QTE estimators which showcase highly flexible Bayesian modeling manner coupled with favorable frequentist properties in finite samples.

The advantages of the proposed estimators are demonstrated in Monte Carlo simulations, which consider difficult settings such as high-dimensional covariate spaces or complex nonlinear effects of covariates. Through numerical evidence, we observe substantial gains in bias reduction for QTE estimates across all scenarios, highlighting the strong estimation and inferential features of our methodology in comparison with the naive estimator and popular approaches considered in the literature. The proposed methodology introduces a fresh perspective to empirical research by offering a novel approach for estimating unconditional QTEs in microeconomic applications. The new estimator,

Bayesian Doubly Robust with shrinkage prior, allows us to revisit the microcredit experiment originating from the work of Crépon et al. (2015) and explore the impact of household financial access on household welfare. Unlike previous studies that strictly rely on the randomization of microcredit availability at the village level, we employ a new causal estimand and identification strategy that utilizes observed, non-random borrowing patterns at the household level as well as observable household characteristics. Our findings indicate an overall positive effect, with heterogeneous impacts across the different points of each outcome distribution of interest. It is anticipated that universal financial access will result in an ex-post increase in economic inequality among households, mostly attributed to the significant improvements in consumption and business outcomes at the upper quantiles. Notably, there is evidence of systematic harm in terms of total profit, as a segment of households may experience adverse effects that extend the left tail of the distribution to the left.

The remainder of this paper is organized as follows. Section 2 presents a brief review of existing studies relevant to our paper and situates the paper within existing literature. In Section 3, we formally define quantile treatment effects in a causal framework along with key identification assumptions. In Section 4, we present the proposed approach for estimating quantile treatment effects. Next, we evaluate the performance of our methods using simulations in Section 5 and use the proposed method to examine the causal impact of loan access on the distribution of household outcomes in Section 6. Finally, we conclude the paper in Section 7 with brief final remarks on the method and policy recommendations based on our results.

2 Related Literature

2.1 Causal Inference on Quantiles

Firpo (2007) first considered efficient estimation of unconditional quantile treatment effects (QTEs) and proposed an inverse propensity weighting (IPW) estimator based on propensity scores estimated using a sieve approach, specifically a logistic power series approximation. Under strong smoothness conditions, this IPW estimator is $\sqrt{N}$-consistent[4] and achieves the semiparametric efficiency bound, which is reminiscent of analogous results for the IPW estimator for the average treatment effect (ATE) with nonparametrically estimated propensities (Hirano, Imbens, and Ridder 2003). Although these purely weighted methods circumvent the estimation of nuisances that depend on the estimand, their desired behavior is restricted to certain nonparametric weight estimators and requires strong smoothness assumptions. Extending the IPW estimator to high-dimensional settings runs into issues due to the fact that its convergence rate can be slowed down by that of the propensity score and its error may depend heavily on the particular method used to learn the propensity score. The properties prohibit the use of general machine learning methods and potentially leading to unstable estimates. In this sense, Firpo (2007)'s IPW estimator lacks the double robustness and flexibility of our proposal.

Zhang et al. (2012) developed several nonparametric methods that resemble those used for ATE and proved that the augmented inverse probability weighted (AIPW) estimator, by augmenting a term that involves the residuals from the outcome regression model, enhances the efficiency of the IPW estimator. The AIPW estimator is expected to be locally efficient and doubly robust under regularity conditions. Díaz (2017) proposed a semiparametric approach using targeted maximum likelihood estimation (TMLE) for marginal quantiles. While sharing the same asymptotic properties as the standard AIPW, the TMLE estimator demonstrates better finite-sample performance when analyzing causal effects on the quantiles, similar to improvements in the mean effect (e.g., Van der Laan, Rose, et al. 2011). Our proposed approach relates to both AIPW and TMLE estimators when solving the estimating equation derived from the efficient influence function, a core concept for achieving double robustness. However, Zhang et al. (2012)'s AIPW method assumes strong distributional assumption (e.g., a normal linear model after a Box-Cox transformation of the outcome for each treatment), limiting its application to cases of positive outcomes. In contrast, our Bayesian Analog of Doubly Robust estimation framework employs Bayesian data-adaptive estimation to flexibly fit nuisance functions. Díaz (2017)'s TMLE approach can be considered quite general and closest to our approach among frequentist methods. While both opt for estimating the conditional distribution as an important middle step, the distinct feature of our modeling strategy lies in utilizing multiple Bayesian quantile regressions and thus can incorporate regularization seamlessly.

Unlike previous studies that did not explicitly consider the case of potentially high-dimensional covariates, Kallus, Mao, and Uehara (2024) proposed Localized Debiased Machine Learning (LDML) to enable efficient inference on QTEs in this scenario. For ATE estimation, nuisance functions do not depend on the estimand and can therefore be estimated independently using flexible, data-driven, machine learning methods and plugged into the estimating equation. This Debiased Machine Learning (DML) approach is, however, far more challenging for QTEs estimation, as the efficient influence function involves nuisances that depend on the estimand of interest. Specifically, DML requires we learn

[4] N is the sample size.

the whole conditional cumulative distribution function of a real-valued outcome, potentially conditioned on high-dimensional covariates, evaluated at the quantile of interest. To obviate this cumbersomeness, Kallus, Mao, and Uehara (2024) localize the nuisance estimation step to a single initial rough guess of the estimand, such as the IPW estimate, thereby enabling the standard use of machine learning methods in this DML-extended framework. Despite also aiming for a flexible modeling manner, our paper takes a different approach, estimating the whole continuum of the estimand-dependent nuisances by discretizing a hypothetical continuum of quantile regression estimators. The rationale for our choice is based on the advantages of Bayesian quantile regression over the frequentist alternative.

While most studies on unconditional QTEs are based on frequentist methods, Xu, Daniels, and Winterstein (2018) proposed a Bayesian nonparametric approach (BNP) that utilizes a Bayesian additive regression tree (BART) model to estimate the propensity score, followed by a Dirichlet process mixture (DPM) of normals model to construct the distribution of potential outcomes conditional on the estimated propensity score. A key advantage of this approach over frequentist methods is the simultaneous estimation of multiple quantiles of interest. However, it can be regarded as a propensity score analysis which avoids directly modeling the conditional distribution of potential outcomes given the covariates. In contrast, we propose Bayesian Analog of Doubly Robust estimators that can handle a large number of covariates and are more robust to misspecifications.

2.2 Causal Inference in High Dimensions

This paper fits into a broader literature on high-dimensional causal inference with observational data. High-dimensional settings are becoming increasingly prevalent, presenting challenges for causal inference. This problem involves either a large number of available covariates or an outcome model with an infinite or large number of parameters, such as nonparametric and semiparametric models. Regularization, a popular technique originally designed to perform prediction in high-dimensional data analysis, has garnered substantial attention in causal inference. It gives rise to numerous causal machine-learning techniques which provide high-quality inference on treatment effects (Belloni, Chernozhukov, and Hansen 2014; Chernozhukov et al. 2018; Farrell 2015; Athey, Imbens, and Wager 2018). While the majority of studies are frequentist regularization-based approaches, there has been growing interest in adopting Bayesian regularization-based techniques into causal inference, as Bayesian inference is a natural probabilistic framework for quantifying uncertainty and learning about model parameters. It is known that many frequentist penalized likelihood estimators can be considered equivalent to the posterior modes of Bayesian estimators under certain choices of shrinkage priors such as spike-and-slab prior (Mitchell and Beauchamp 1988; Ishwaran and Rao 2005), Bayesian Lasso prior (Park and Casella 2008), and Horseshoe prior (Carvalho, Polson, and Scott 2009, 2010).[5] Recent studies have successfully deployed these techniques for confounding adjustment to estimate average treatment effects in the presence of high-dimensional controls (Antonelli, Parmigiani, and Dominici 2019; Antonelli, Papadogeorgou, and Dominici 2022; Hahn et al. 2018). Bayesian nonparametric methods are also powerful tools utilized for regularization within the Bayesian paradigm. Among them, the Bayesian Additive Regression Tree (BART) has emerged as a workhorse widely used for

[5]Thorough reviews and well-designed simulations could be found in Korobilis and Shimizu (2022), Van Erp, Oberski, and Mulder (2019) and Polson and Sokolov (2019), who advocate the merits of Bayesian sparsity-inducing priors in comparison to frequentist counterparts.

causal inference. Introduced by Chipman, George, and McCulloch (2010; 2006), BART models offer several advantages over linear models, such as automatic adaptation to nonlinearity. Regarding implementation, BART is also preferred due to its fast computation, good performance of default choices of hyperparameters and available software (Linero and Antonelli 2023). When there is sufficient covariate overlap, BART has been shown to outperform numerous Frequentist machine learning methods in prediction problems, including random forests. Hill (2011) further proposed the use of BART in causal inference and demonstrated its efficacy in flexibly modeling the response surface. To mitigate the regularization-induced confounding issue (Hahn et al. 2018) when using a BART outcome model, Hahn, Murray, and Carvalho (2020) developed the Bayesian causal forest model, a BART-based approach that includes a fixed estimate of the propensity score for additional adjustment in the outcome model. This model yields excellent performance in estimating heterogeneous treatment effects, making BART a strong default choice for integrating Bayesian nonparametric methods into causal effect modeling. Subsequent studies by Spertus and Normand (2018) and Xu, Daniels, and Winterstein (2018) employed BART models to fit the propensity score in the first stage, enabling the use of Bayesian propensity score analysis to estimate ATE (with high-dimensional data) and QTEs, respectively.

2.3 Double Robustness

This paper also adds to the development of doubly robust estimators, which have gained extensive use in the causal inference literature (Scharfstein, Rotnitzky, and Robins 1999; Bang and Robins 2005) owing to their desirable property of providing consistent inference even under misspecification of either the treatment assignment or outcome regression models (but not both). For a comprehensive survey of doubly robust estimators and their properties, we refer interested readers to Daniel (2014).

Doubly robust estimators have been extended to accommodate nonparametric or high-dimensional settings by enabling data-adaptive estimation of treatment and outcome models. This includes doubly robust estimators with the group Lasso (Farrell 2015), double machine learning estimators (Chernozhukov et al. 2018), doubly robust matching estimators (Antonelli et al. 2018) and targeted maximum likelihood estimators (Van der Laan, Rose, et al. 2011), among others. In these complex settings, doubly robust estimators offer an extra benefit that parametric convergence rates ($\sqrt{N}$) can be achieved even when each of the propensity score or outcome regression models converges at slower rates ($\sqrt[4]{N}$ or faster). Roughly speaking, this echoes the insights from recent advances in the double machine learning literature (Chernozhukov, Newey, and Singh 2022). Specifically, penalizing either the propensity score model or the outcome model alone would be insufficient for valid causal inference, but combining the two as nuisance functions achieves a desirable convergence rate and finite-sample performance in high-dimensional causal analyses.

There have been attempts to propose doubly robust Bayesian recipes, however, this area is still less established than its Frequentist counterpart. This is mainly due to a lack of consensus on propensity score adjustment in the Bayesian causal modeling framework (Robins and Ritov 1997; Robins, Hernán, and Wasserman 2015; Li, Ding, and Mealli 2023), despite its central role being well recognized in the literature (Rosenbaum and Rubin 1983; Zigler 2016). Pioneering work on the Bayesian approach for doubly robust causal inference was done by Saarela, Belzile, and Stephens (2016), in which the authors formalized the problem and addressed it by combining the posterior predictive distribution of

parameters with the Bayesian bootstrap. The idea of utilizing posterior predictive distribution was later advanced in the line of work by Antonelli, Papadogeorgou, and Dominici (2022) and Shin and Antonelli (2023), who aimed to improve inference for doubly robust estimators for the average treatment effect (ATE) and the conditional average treatment effect (CATE), respectively. The general strategy involves estimating both the propensity score and the conditional outcome mean using Bayesian methods. Posterior draws from their respective posterior predictive distributions are then obtained and plugged into a doubly robust estimator. While this approach is not fully Bayesian because there is no joint likelihood for all parameters stated, it effectively integrates Bayesian modeling techniques and Frequentist inferential procedures for causal analysis with proper uncertainty quantification. These features are particularly important in high-dimensional scenarios, where handling large numbers of covariates and quantifying uncertainty can be challenging.

In spite of the progress made in Bayesian literature, most studies have focused on the doubly robust estimation of either unconditional or conditional average treatment effect. Our paper presents a distinctive contribution by concentrating on unconditional quantile treatment effects (QTE). This causal estimand is of independent interest as it offers a different and complementary approach to uncover treatment effect heterogeneity. Although the conditional average treatment effect (CATE) also characterizes treatment effect heterogeneity, the effects vary across individuals or subgroups defined by observed characteristics. In contrast, unconditional QTE focuses on the effect heterogeneity of the treatment across different outcome ranks without conditioning on individual characteristics or covariates. Evaluating the impact on the entire outcome distribution of interest makes this approach particularly relevant to distributional concerns and inequality, offering valuable insights when used alongside other estimands, such as CATE, to understand the potential consequences of treatments and policies.

With regard to methodology, we build on the work of Antonelli, Papadogeorgou, and Dominici (2022), who combined the posterior predictive distribution of nuisance parameters with the Frequentist doubly robust estimator initially proposed for the ATE. We develop a Bayesian Analog of Doubly Robust estimators for the QTE, tackling unique challenges that arise in the quantile setting. First, we address the need for different doubly robust estimators for the QTE compared to the ATE setting by solving an estimating equation built upon an efficient influence function specifically tailored to quantile functionals. Second, the QTE estimation problem involves new nuisance parameters, including the entire conditional cumulative distribution function (CDF) of each potential outcome conditional on potentially high-dimensional covariates, which increases the computational complexity. We overcome this hurdle by employing multiple Bayesian quantile regressions that incorporate shrinkage priors. This helps us explicitly estimate the conditional distribution while accounting for high dimensionality. This approach has not been pursued in previous studies, making our contribution unique in the literature.

3 Notation and Causal Estimand

3.1 Notation

Let T and Y be the treatment and outcome of interest, respectively, while $\mathbf{X}$ is a p-dimensional vector of potential controls. Denote $\mathcal{P}_o$ as the joint distribution of the observed data. Assume that we observe an independent and identically distributed (i.i.d.) sample $\mathbf{Z}_i = \{Y_i, T_i, \mathbf{X}_i\}$ for $i = 1, \ldots, N$ with empirical distribution $\mathcal{P}_N$, where we collect all observations into $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$. For $t \in \{0, 1\}$, let

- $Y^{(t)}$ be the potential outcome for a generic subject under treatment t.
- $F_t(y) \coloneqq \mathbb{P}[Y^{(t)} \leq y]$ be the cumulative distribution function (CDF) of $Y^{(t)}$, and $q_t(\tau) \coloneqq F_t^{-1}(\tau) = \inf\{y \mid F_t(y) \geq \tau\}$ be its τ^{th} quantile, where $\tau \in (0, 1)$.
- $F_{t|1}(y) \coloneqq \mathbb{P}[Y^{(t)} \leq y \mid T = 1]$ be the cumulative distribution function (CDF) of $Y^{(t)}$ given $T = 1$, and $q_{t|1}(\tau) \coloneqq F_{t|1}^{-1}(\tau) = \inf\{y \mid F_{t|1}(y) \geq \tau\}$ be its τ^{th} quantile, where $\tau \in (0, 1)$.

3.2 Causal Estimand

Quantile Treatment Effects (QTEs) are defined as the difference between the τ^{th} quantiles (for a particular value of τ) of the treated potential outcome distribution and the untreated potential outcome distribution. For $\tau \in (0, 1)$,

$$QTE(\tau) \coloneqq F_1^{-1}(\tau) - F_0^{-1}(\tau) = q_1(\tau) - q_0(\tau). \tag{3.1}$$

Likewise, Quantile Treatment Effects on the Treated (QTET) are defined as the difference between the quantiles of the distribution of treated potential outcomes for the treated group and the distribution of untreated potential outcomes for the treated group. For $\tau \in (0, 1)$,

$$QTET(\tau) \coloneqq F_{1|1}^{-1}(\tau) - F_{0|1}^{-1}(\tau) = q_{1|1}(\tau) - q_{0|1}(\tau). \tag{3.2}$$

For identification, a fundamental issue is whether each of the distributions is identified. For the QTE, some assumptions need to be invoked to identify both $F_1(y)$ and $F_0(y)$. For the QTET, $F_{1|1}(y)$ is identified directly from the data because we observe the distribution of treated outcomes for the treated group. However, identifying $F_{0|1}(y)$ requires some identifying assumptions because we do not observe untreated potential outcomes for the treated group. We make the following assumptions in our setup:

1. *Unconfoundedness (Selection-on-Observables)*

$$Y^{(1)}, Y^{(0)} \perp\!\!\!\perp T | \mathbf{X}, \tag{3.3}$$

where $\perp\!\!\!\perp$ denotes statistical independence.

2. *Strong Overlap*

$$\exists \delta \in \mathbb{R}: \quad 0 < \delta < \mathbb{P}[T = 1 \mid \mathbf{X}] < 1 - \delta < 1. \tag{3.4}$$

3. *Stable Unit Treatment Value Assumption (SUTVA)*

$$T_i = t \text{ implies } Y_i^{obs} = Y_i^{(t)}, \quad \text{for } t \in \{0, 1\}. \tag{3.5}$$

The conditional distributions of potential outcomes are therein identified by determining the conditional distribution of observed outcomes for individuals within each group, as expressed by: $\mathbb{P}[Y(t) \leq y \mid \mathbf{X}] = \mathbb{P}[Y \leq y \mid T = t, \mathbf{X}]$. Consequently, the marginal distribution of potential outcomes can be identified and calculated as

$$F_t(y) = \int \mathbb{P}[Y \leq y \mid T = t, \mathbf{X} = \mathbf{x}] dF_{\mathbf{X}}(\mathbf{x}), \quad \text{for } t \in \{0, 1\}, \tag{3.6}$$

where $F_{\mathbf{X}}(\mathbf{x})$ is the marginal distribution of covariates $\mathbf{X}$ in the population of interest.

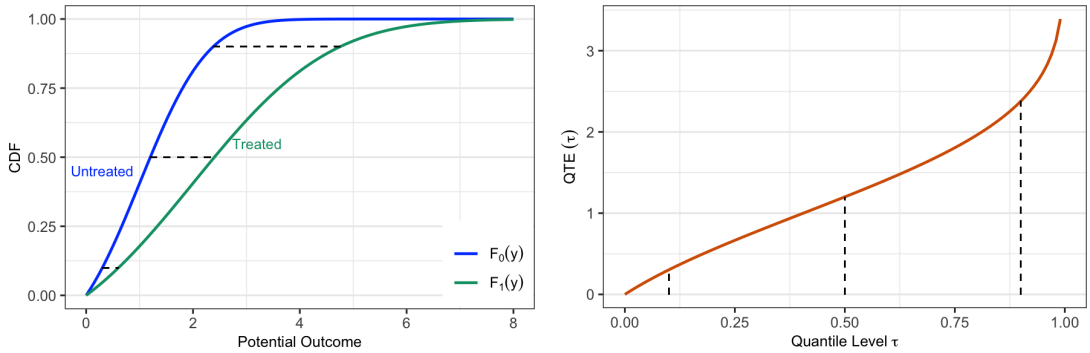

Figure 3.1: Illustration of Quantile Treatment Effects (QTEs). The left-hand figure demonstrates unconditional distributions of Treated and Untreated potential outcomes, which are colored in green and blue, respectively. The horizontal distance between these two distributions yields QTEs. For instance, QTE(0.95), QTE(0.5), and QTE(0.05) are represented by three dashed lines in the figure. QTEs across all values of quantile levels are plotted in the right-hand figure.

4 Proposed Estimation Approach

4.1 Justification

With the primary parameter of interest being Quantile Treatment Effects (QTE), we develop the Bayesian Analog of Doubly Robust (BADR) estimation framework for this target causal estimand. Our approach draws inspiration from the work by Antonelli, Papadogeorgou, and Dominici (2022), originally proposed for the Average Treatment Effect (ATE), to combine Bayesian modeling methods for treatment assignment and outcome models with Frequentist doubly robust estimators using posterior predictive distributions. By tackling extra unique challenges that arise in the quantile setting, our framework aims to offer enhanced finite-sample performance without strict reliance on correct model specifications.

The implementation procedure is straightforward as follows:

1. Specify separate Bayesian treatment assignment and outcome models;
2. Draw the propensity score and the conditional distribution of each potential outcome from their corresponding posterior predictive distributions; and
3. Plug these values into a doubly robust estimator associated with the parameter of interest.

While our estimation approach is applicable in general, it is particularly useful in handling high dimensionality, addressing challenges posed by potentially large numbers of controls and the involvement of the entire conditional potential outcome distribution as a nuisance parameter. In Section 4.2, we propose a modeling framework in high dimensions that flexibly incorporates Bayesian regularization techniques. Before doing so, we first establish a foundation by defining the doubly robust estimator of QTE. Then, we demonstrate a promising avenue for estimation and inference within a Bayesian framework.

4.1.1 The Doubly Robust Estimator for Quantile Treatment Effects

Let $\pi(\mathbf{X}) \coloneqq \mathbb{P}(T = 1 \mid \mathbf{X}; \Theta_\pi)$ be the propensity score (i.e., the probability of receiving active treatment given covariates $\mathbf{X}$), which is associated with the treatment assignment model; and $G(y \mid t, \mathbf{X}) \coloneqq \mathbb{P}[Y \leq y \mid T = t, \mathbf{X}; \Theta_G]$ (for $t \in \{0, 1\}$) be the conditional distributions of Y given the treatment status and covariates, which is associated with the outcome model. Let $\Theta = \Theta_\pi \cup \Theta_G$ represent the parameters of the treatment assignment and outcome models. A general QTE estimation problem involves $\pi(\mathbf{X})$ and $G(y \mid t, \mathbf{X})$ as nuisance functions. We denote $\hat{\pi}(\mathbf{X})$ and $\hat{G}(y \mid t, \mathbf{X})$ as estimators of $\pi(\mathbf{X})$ and $G(y \mid t, \mathbf{X})$ (for $t \in \{0, 1\}$), respectively.

For a chosen quantile level $\tau \in (0, 1)$, a doubly robust (DR) estimator of the QTE for binary treatments is given by

$$\widehat{QTE}^{dr}(\tau) = \hat{q}_1^{dr}(\tau) - \hat{q}_0^{dr}(\tau), \tag{4.1}$$

where for a sample of size N

- $\hat{q}_1^{dr}(\tau)$ is a DR estimator of the τ-quantile of *treated* potential outcome and can be derived as the solution to

$$N^{-1}\sum_i \left\{ \frac{T_i}{\hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}\{Y_i \leq q_1\} - \hat{G}(q_1 \mid 1, \mathbf{X}_i) \right] + \hat{G}(q_1 \mid 1, \mathbf{X}_i) \right\} = \tau. \tag{4.2}$$

- $\hat{q}_0^{dr}(\tau)$ is a DR estimator of the τ-quantile of *untreated* potential outcome and can be derived as the solution to

$$N^{-1}\sum_i \left\{ \frac{1-T_i}{1-\hat{\pi}(\mathbf{X}_i)} \left[\mathbb{1}\{Y_i \leq q_0\} - \hat{G}(q_0 \mid 0, \mathbf{X}_i) \right] + \hat{G}(q_0 \mid 0, \mathbf{X}_i) \right\} = \tau. \tag{4.3}$$

Formal derivation and discussion regarding $\hat{q}_1^{dr}(\tau)$ are presented in Appendix A, where (A.14) is equivalent to equation (4.2). Estimating equations (4.2) and (4.3) are built upon the efficient influence function for quantiles of each potential outcome distribution. The efficient influence function captures the first-order sensitivity of the target parameter to small perturbations in the underlying distributions. In our estimation problem, which involves two nuisance models – treatment assignment and outcome – the efficient influence function exhibits a double robustness property, resulting in a doubly robust estimator. This estimator is consistent provided that either the propensity score $\hat{\pi}(\mathbf{X})$ or the conditional outcome distribution $\hat{G}(y \mid t, \mathbf{X})$ is consistent, but not necessarily both.

Estimating equations (4.2) and (4.3) are also closely connected to the Neyman orthogonal moment conditions, which is extensively leveraged in the debiased machine learning literature (see, e.g., Belloni et al. 2017; Chernozhukov et al. 2018; Kallus, Mao, and Uehara 2024). Neyman orthogonality is a desirable property that ensures the final estimate of the target parameter remains robust even when there are small errors in the estimation of nuisance parameters. This property is particularly relevant when regularization methods are needed to handle high-dimensional covariates in estimating the nuisances. In such cases, employing Neyman orthogonal moment conditions helps correct for the first-order biases that may arise from plugging in estimates of the nuisance parameters.

4.1.2 The Bayesian Analog of Doubly Robust Estimator

The population parameters Θ are typically unknown and need to be estimated. In this setting, we consider a Bayesian framework to estimate the parameters associated to both the treatment assignment and outcome models. This enables uncertainty in parameter estimation to be directly captured from the posterior distribution.

Let $\mathbb{P}_{\Theta|\mathbf{Z}}$ denote the posterior distribution, and $\{\Theta^{(b)}\}_{b=1}^{B}$ be a sequence of B draws obtained from this posterior distribution. The point estimator $\hat{\Delta}$ for the estimand of interest Δ takes a form of the posterior mean

$$\hat{\Delta} := \mathbb{E}_{\Theta|\mathbf{Z}}[\Delta(\mathbf{Z}, \Theta)] \approx B^{-1}\sum_b \Delta(\mathbf{Z}, \Theta^{(b)}), \tag{4.4}$$

where $\Delta(\mathbf{Z}, \Theta^{(b)})$ is evaluated using the observed data $\mathbf{Z}$ and parameters $\Theta^{(b)}$. Therefore, our point estimator for the QTE at a chosen quantile level $\tau \in (0,1)$ is the average value of the quantity in (4.1) with respect to the posterior distribution of model parameters.

Regarding inference, our variance of interest corresponds to the variance of the estimator's sampling distribution and can be defined as follows

$$\mathbb{V}_{\mathbf{Z}}(\hat{\Delta}) \coloneqq \mathbb{V}_{\mathbf{Z}}(\mathbb{E}_{\Theta|\mathbf{Z}}[\Delta(\mathbf{Z}, \Theta)]). \tag{4.5}$$

Variance estimation can be implemented using the nonparametric bootstrap (Tibshirani and Efron 1993) to account for uncertainty in all stages of the estimator. Specifically, L multiple datasets $\{\mathbf{Z}^{(l)}\}_{l=1}^{L}$ are created by sampling with replacement from the empirical distribution of the data. For each resampled dataset, we re-estimate the posterior distribution of Θ and then recalculate $\mathbb{E}_{\Theta|\mathbf{Z}}[\Delta(\mathbf{Z}, \Theta)]$ accordingly. Finally, we compute variance of this quantity across all bootstrap samples. It is worth noting that there are two main sources of uncertainty arising throughout our QTE estimation procedure: first, the sampling variability stemming from the data even if we know the true outcome and treatment assignment models; second, the variability in parameter estimation for the propensity score and conditional outcome distributions. An alternative inference scheme from Antonelli, Papadogeorgou, and Dominici (2022) can be adopted by targeting these two parts separately

$$\mathbb{V}_{\mathbf{Z}}(\hat{\Delta}) = \underbrace{\mathbb{V}_{\mathbf{Z}^{(l)}}\{\mathbb{E}_{\Theta|\mathbf{Z}}[\Delta(\mathbf{Z}^{(l)}, \Theta)]\}}_{\substack{\text{uncertainty stemming} \\ \text{from the data}}} + \underbrace{\mathbb{V}_{\Theta|\mathbf{Z}}[\Delta(\mathbf{Z}, \Theta)]}_{\substack{\text{uncertainty in estimation} \\ \text{of nuisance parameters}}} . \tag{4.6}$$

The first term resembles the true variance, except for the outer moment, which is associated with a resampled version $\mathbf{Z}^{(l)}$ of the original data $\mathbf{Z}$. The posterior samples of parameters Θ estimated using the original data are maintained, but the point estimator $\mathbb{E}_{\Theta|\mathbf{Z}}[\Delta(\mathbf{Z}^{(l)}, \Theta)]$ is recalculated for each resampled dataset. In this way, it captures only the uncertainty of the data, not that resulting from parameter estimation. In contrast, the second term accounts for the latter type of uncertainty based on the variability of the full posterior samples of $\Delta(\mathbf{Z}, \Theta)$ given the observed data $\mathbf{Z}$.

4.2 Modeling Framework in High Dimensions

At this juncture, we present the high-dimensional modeling framework we use to provide scalable estimation algorithms. We employ Bayesian techniques to specify and estimate the treatment assignment and outcome models separately. To address the challenges posed by high-dimensional feature spaces, we integrate various Bayesian regularization methods into our proposed framework to yield estimators of the nuisance functions corresponding to the propensity score and the conditional distribution of the outcome, denoted as $\hat{\pi}(\mathbf{X})$ and $\hat{G}(y \mid t, \mathbf{X})$, respectively. For the treatment assignment model, we adopt Bayesian Additive Regression Trees (BART) priors, whose merits have been increasingly recognized (see, e.g., H. A. Chipman, George, and McCulloch 2010; Hill 2011; Hahn, Murray, and Carvalho 2020; Linero and Antonelli 2023). For the outcome model, we leverage multiple Bayesian quantile regressions combined with shrinkage priors to explicitly estimate the conditional distribution while accommodating potentially high-dimensional covariates. This combined strategy allows us to develop a flexible and doubly-robust Bayesian estimator with desirable finite-sample frequentist properties. To the best of our knowledge, this modeling framework has not been previously

pursued in the literature, resulting in a novel approach to estimation of heterogeneous treatment effects.

4.2.1 Treatment Assignment Model

We fit a binary Bayesian Additive Regression Trees (BART) model on the observations $\{T_i, \mathbf{X}_i\}_{i=1}^n$ to model the regression of the treatment assignment on control variables, that is,

$$\pi(\mathbf{X}_i) = \mathbb{P}\left(T_i = 1 \mid \mathbf{X}_i\right) = H\left[f_{\text{BART}}\left(\mathbf{X}_i\right)\right], \tag{4.7}$$

where the link function H is either the CDF of the standard normal distribution for probit BART or the CDF of the logistic distribution for the logit BART, and

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

For $m \in \{1, \ldots, M\}$, Γ_m is a tree structure that consists of a set of splitting rules and a set of terminal nodes; and $\mu = (\mu_{m,1}, \ldots, \mu_{m,b_m})$ is a vector of parameters associated with b_m terminal nodes of Γ_m, such that $f_{\text{tree}}(\mathbf{X}_i; \Gamma_m, \mu_m) = \mu_{m,l}$ if $\mathbf{X}_i$ is corresponding to the l^{th} terminal node of Γ_m.

The modeling choices used to implement the BART specification are presented in Appendix B. Once a sequence of B posterior draws for the underlying BART parameters has been obtained, B posterior samples of the fitted propensity score $\{\pi^{(b)}(\mathbf{X})\}_{b=1}^B$ can be calculated by

$$\pi^{(b)}(\mathbf{X}_i) = H\left[\sum_{m=1}^{M} f_{\text{tree}}\left(\mathbf{X}_i; \Gamma_m^{(b)}, \mu_m^{(b)}\right)\right] \quad \text{for } i = 1, \ldots, N \text{ and } b = 1, \ldots, B. \tag{4.8}$$

4.2.2 Outcome Model

In contrast to the literature's extensive coverage of conditional expectation estimation, data-adaptive estimation of conditional distributions has received considerably less attention. For each $t \in \{0, 1\}$, we estimate the conditional distribution based on fitted conditional quantiles, employing the sample analog of the following alternative representation of the conditional distribution

$$F_{Y|\mathbf{X}}(y) = \int_0^1 \mathbb{1}\left\{F_{Y|\mathbf{X}}^{-1}(\tau) \leq y\right\} d\tau = \int_0^1 \mathbb{1}\left\{\mathcal{Q}_{Y|\mathbf{X}}(\tau) \leq y\right\} d\tau, \tag{4.9}$$

where $F_{Y|\mathbf{X}}(\cdot)$ and $\mathcal{Q}_{Y|\mathbf{X}}(\cdot)$ are conditional distribution and conditional quantiles, respectively.

The corresponding estimator is

$$\begin{aligned} \hat{F}_{Y|\mathbf{X}}(y) &= \int_0^1 \mathbb{1}\left\{\hat{\mathcal{Q}}_{Y|\mathbf{X}}(\tau) \leq y\right\} d\tau \\ &\approx \epsilon + \int_{\epsilon}^{1-\epsilon} \mathbb{1}\left\{\hat{\mathcal{Q}}_{Y|\mathbf{X}}(\tau) \leq y\right\} d\tau \\ &\approx \epsilon + \sum_{s=1}^{S} \delta_s \mathbb{1}\left\{\hat{\mathcal{Q}}_{Y|\mathbf{X}}(\tau_s) \leq y\right\}, \end{aligned} \tag{4.10}$$

where $\hat{\mathcal{Q}}_{Y|\mathbf{X}}(\tau_s) = \mathbf{X}^\top \hat{\beta}_{(\tau_s)}$ can be obtained by estimating S Bayesian quantile regression model for each $\{\tau_s\}_{s=1}^S$, where $\epsilon \leq \tau_0 < \ldots < \tau_s \leq 1-\epsilon$ and the width $\delta_s = \tau_s - \tau_{s-1} \to 0$ as $S \to \infty$. The second equation is adapted for tail trimming. The third equation aims to avoid estimating the whole quantile regression process. Our discretization technique is similar to some previous studies (Belloni et al. 2017; Chernozhukov, Fernández-Val, and Melly 2013 ; Frölich and Melly 2013), however, we use the Bayesian quantile regression model rather than Koenker and Bassett (1978)'s quantile regression from a frequentist viewpoint.

There are several advantages of this computational approach to the problem of estimating conditional distributions. First, it enables us to leverage the Bayesian quantile regression model, which not only suits our overall framework but also offers more flexibility than its frequentist counterpart. Especially, Bayesian shrinkage priors can be readily applied to this parametric quantile model with minor modifications, thereby handling better high-dimensional covariates. This feature is thoroughly reviewed by Korobilis and Shimizu (2022). In addition, while a very fine grid for values of τ_s (i.e., large S) is often required to gain accuracy, we can make use of parallel computation because the conditional posteriors are applied in each quantile level independently. While crossing or non-monotonic estimated quantiles are a valid concern when the regression for each quantile is estimated separately[6], the algorithm presented above is originally designed for the rearrangement of crossing quantiles. This ensures that our primary objective of interest, the conditional distribution, remains unaffected by these potential estimation issues.

Further discussion on the Bayesian shrinkage priors and Bayesian quantile regression can be found in Appendices C and D, respectively. By drawing a sequence of B posterior draws for the quantile regression parameters, we can obtain B posterior samples of the fitted conditional outcome distributions $\{G^{(b)}(y \mid 0, \mathbf{X})\}_{b=1}^B$ and $\{G^{(b)}(y \mid 1, \mathbf{X})\}_{b=1}^B$.

4.2.3 Algorithms for the Bayesian Analog of Doubly Robust (BADR) Estimation

Upon acquiring sequences of B posterior draws of the fitted propensity score $\{\pi^{(b)}(\mathbf{X})\}_{b=1}^B$ and the fitted conditional outcome distributions $\{G^{(b)}(y \mid t, \mathbf{X})\}_{b=1}^B$ for $t \in \{0, 1\}$, we can compute B values of the corresponding Quantile Treatment Effect (QTE) based on the full posterior distribution of these nuisance parameters. The BADR point estimate of the QTE used in this paper is derived as the average of these B values. The details of the implementation are presented in Algorithm 1. Utilizing B posterior samples of QTE, variance estimation can be proceeded according to (4.6).

It is noteworthy that following Algorithm 1 requires solving two estimation equations, (4.2) and (4.3), B times. This step may lead to intensive computations, particularly when bootstrapping is involved. An alternative approach to combining the posterior distribution of model parameters and the doubly robust estimator in Algorithm 1 would replace the nuisance parameters $\pi(\mathbf{X})$ and $G(y \mid t, \mathbf{X})$ with plug-in estimates, such as their posterior means, as outlined in Algorithm 2. This aligns with a frequentist modeling approach where doubly robust estimators are evaluated using plug-in estimates of the parameters Θ. While Algorithm 2 uses more compact information, there is a clear computational gain due to the fact that estimation equations only need to be solved once. The inference procedure

[6]The estimated conditional quantile functions may be non-monotonic in the sense that $\bar{\tau} > \tilde{\tau}$ does not necessarily imply $\hat{\mathcal{Q}}_{Y|\mathbf{X}}(\bar{\tau}) > \hat{\mathcal{Q}}_{Y|\mathbf{X}}(\tilde{\tau})$.

is conducted using the original bootstrap variance estimation in (4.5) for ease of implementation. Furthermore, our pilot Monte Carlo findings suggest that the alternative estimator yields similar results.

Algorithm 1:

Bayesian Analog of Doubly Robust (BADR) estimation for QTEs (*Full posterior samples*)

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 $q_1^{(b)}(\tau), q_0^{(b)}(\tau)$ based on $\pi^{(b)}(\mathbf{X})$ and $G^{(b)}(y \mid t, \mathbf{X})$, according to (4.2) and (4.3).

  Calculate $QTE^{(b)}(\tau) = q_1^{(b)}(\tau) - q_0^{(b)}(\tau)$, according to (4.1).

end

Calculate $\hat{\Delta}_\tau \equiv \widehat{QTE}^{dr}(\tau) = \frac{1}{B}\sum_{b=1}^B QTE^{(b)}(\tau)$, according to (4.4).

Algorithm 2:

Bayesian Analog of Doubly Robust (BADR) estimation for QTEs (*Posterior means*)

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

Derive posterior mean from B posterior samples

  $\hat{\pi}(\mathbf{X}) = \frac{1}{B}\sum_{b=1}^B \pi^{(b)}(\mathbf{X})$ and $\hat{G}(y \mid t, \mathbf{X}) = \frac{1}{B}\sum_{b=1}^B G^{(b)}(y \mid t, \mathbf{X})$, according to (4.4).

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 (4.2) and (4.3).

Calculate $\hat{\Delta}_\tau \equiv \widehat{QTE}^{dr}(\tau) = \hat{q}_1^{dr}(\tau) - \hat{q}_0^{dr}(\tau)$, according to (4.1).

5 Simulation Study

We assess the finite-sample performance of our proposed approach, Bayesian Analog of Doubly Robust (BADR) estimation, in two simulations with details described below. For each simulation, we specify the distribution of covariates, the treatment assignment mechanism and the distribution of potential outcomes. The first simulation focuses on a linear setting with varying covariate dimensionality to sample size ratio (p/N). In the second simulation, we consider a nonlinear setting and further examine the double robustness of our proposed estimators. Both of these data designs imply that assignment to the treatment is not completely random, but satisfies the *selection-on-observables* assumption. From a theoretical perspective, estimation of treatment effects that fails to account for the selection problem will inevitably produce inconsistent estimates. We regard this approach as a benchmark and consider the `Naive` estimator, which is an estimator of simple differences between empirical quantiles of treated and control groups, without any correction for selection bias.

We develop two versions of the estimators which represent our proposed methodology – Bayesian Doubly Robust estimator (`BDR`) and an extension that adds shrinkage priors (`BDRS`). Specifically, the former employs the original Bayesian Quantile Regression while the latter incorporates the Adaptive Lasso in order to account for sparsity and uncertainty in the outcome model. Both estimators fit the propensity score using a logit BART model in the first step. Furthermore, we also compare our proposed method with existing estimators. The Bayesian nonparametric counterpart (`BNP`) is a fully Bayesian approach developed in Xu, Daniels, and Winterstein (2018), where the propensity score is estimated using a logit BART, then the conditional distribution of the potential outcome given a BART posterior sample of the propensity score in each treatment group is estimated separately using a Dirichlet process mixture of multivariate normals. We additionally compare three frequentist methods – the Localized Debiased Machine Learning (`LDML`) method introduced in Kallus, Mao, and Uehara (2024), the Targeted Maximum Likelihood Estimation (`TMLE`) method proposed in Díaz (2017), and Firpo's Inverse Probability Weighted (`FIPW`) method developed in Firpo (2007). Among them, `LDML` and `TMLE` are two estimators that can leverage a variety of machine learning methods. Particularly, in our simulation exercise, Random Forest is incorporated into `LDML` and Lasso is integrated into `TMLE`. Implementation details of these methods can be found in Appendix E.1.

In each simulation design, we generate 100 synthetic datasets. For each simulated dataset, we calculate quantile causal effects for 5 quantile levels, $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$, and their 95% credible (or confidence) intervals (CIs). We compare all the different approaches in terms of average bias, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

5.1 Simulation Design 1 (SD1)

We first consider a linear setting in which the mean of potential outcomes is a linear combination of covariates. We draw 40-dimensional covariates $\mathbf{X}$ ($p = 40$) from the independent standard normal distributions and allow different sample sizes $N \in \{100, 500, 1000\}$ of the dataset. Accordingly, we could evaluate the estimation procedure across varying feature dimensionality (i.e., p/N ratio). The

exact form of the true model used to generate synthetic data is as follows:

$$
\begin{aligned}
T \mid \mathbf{X} &\sim \mathcal{B}ern\left(\pi(\mathbf{X})\right), \\
Y^{(0)} \mid \mathbf{X} &\sim \mathcal{N}\left(\mu(\mathbf{X}), 2.5^2\right), \\
Y^{(1)} \mid \mathbf{X} &\sim \mathcal{N}\left(1+\mu(\mathbf{X}), 3.75^2\right), \\
Y &= T \times Y^{(1)} + (1-T) \times Y^{(0)}; \\
\text{where} \qquad \pi(\mathbf{X}) &= \left\{1+\exp\left[-(X_1+X_2+X_3)\right]\right\}^{-1}, \\
\mu(\mathbf{X}) &= X_1 + X_2 + X_4 + X_5.
\end{aligned}
\tag{5.1}
$$

Under this specification, the unconditional distribution of potential outcomes are $Y^{(0)} \sim \mathcal{N}(0, 10.25)$ and $Y^{(1)} \sim \mathcal{N}(1, 18.0625)$. Figure 5.1 provides a visual illustration of the corresponding marginal densities and marginal distributions. As a result, the population quantile treatment effects can be computed analytically. In particular, the true 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} QTEs are $\Delta_{0.10} = (-4.447) - (-4.103) = -0.344$, $\Delta_{0.25} = (-1.866) - (-2.159) = 0.293$, $\Delta_{0.5} = 1 - 0 = 1$, $\Delta_{0.75} = 3.866 - 2.159 = 1.707$, and $\Delta_{0.90} = 6.447 - 4.103 = 2.344$, respectively.

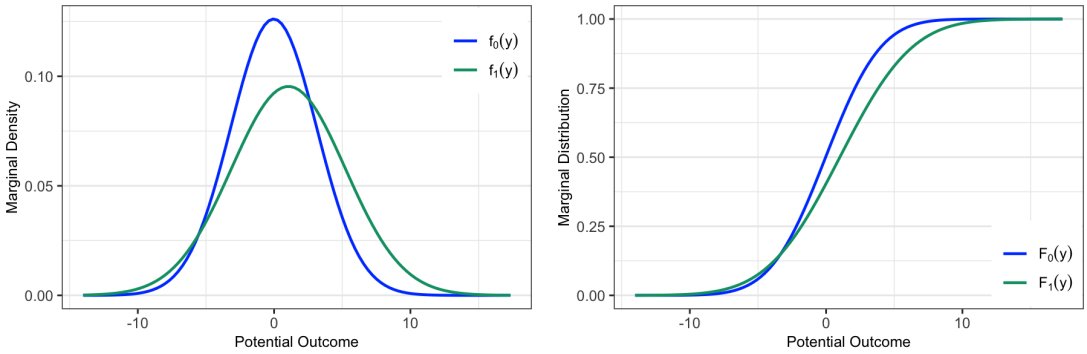

Figure 5.1: True marginal densities and marginal distributions of the treated and untreated potential outcomes in SD1. This design emulates a thought experiment relevant to policy evaluation literature. Hypothetically assigning the entire population either to treatment or to control induces a change in both location and shape of the outcome distribution.

Table 5.1 presents point estimates for the quantile treatment effects, along with the average lower and upper bounds of the corresponding 95% CIs across 100 simulated datasets. These computations are based on a sample size of $N = 1000$ and $p = 40$. It is clear that the `Naive` method exhibits substantial bias in its point estimates. This can be attributed to the absence of adjustment for confounders in $\mathbf{X}$, resulting in poor performance as expected. In comparison, all other methods considered in our current setting outperform the `Naive` method in terms of both bias and coverage, proving their effectiveness in correcting selection bias to some extent.

Our proposed estimators, `BDR` and `BDRS`, yield point estimates closest to the true values of QTEs. Notably, incorporating a shrinkage prior, as in `BDRS`, further enhances the performance of `BDR`, particularly when the object of interest is extreme tails (i.e. 10^{th} and 90^{th} percentiles).

Table 5.1: 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)

Notes: 95% CIs in parentheses correspond to 95% confidence intervals in Frequentist approach or 95% posterior credible intervals in Bayesian approach. To estimate these 95% CIs, LDML and FIPW use analytical standard errors, whereas others rely on the bootstrap method.

Despite sharing a probabilistic approach, the Bayesian nonparametric estimator, `BNP`, demonstrates differences from our proposed estimators. While the nonparametric method produces point estimates slightly better than `Naive` method, they are still far from the truth. Moreover, the 95% credible intervals associated with `BNP` fail to cover the true values of QTEs at any percentile. It aligns with the observation that `BNP` exhibits the smallest CI widths among all surveyed methods, posing challenges in achieving satisfactory coverage rates. It is worth noting that because both `BDR` and `BNP` use BART-logit to model the treatment assignment in the first stage, their distinct performance illustrates the role of modeling the conditional distribution of potential outcomes given confounders. Intuitively, `BNP` avoids directly modeling the conditional distribution of potential outcomes given confounders. Instead, it is grounded in the balancing property of the propensity score (Rosenbaum and Rubin 1983) to model the conditional distribution of the outcome given the propensity score alone. While this approach involves estimating a less complex distribution due to having only one binary regressor (i.e. the estimated propensity score) in the second stage, it becomes skeptical in the case of misspecified treatment assignment. According to Monte Carlo results in the original paper by Xu, Daniels, and Winterstein (2018), the inclusion of non-confounders in the treatment assignment equation entails less precise estimations of QTEs, thereby compromising the performance of the `BNP` method. Aside from efficiency loss, finite-sample bias is also a notable drawback of methods targeting a set of variables that best predict treatment assignment without accounting for how these variables are related to the outcome, as widely discussed in the context of average treatment effect (Zigler and Dominici 2014; Belloni, Chernozhukov, and Hansen 2014, etc.).

Among frequentist approaches, `TMLE` and `FIPW` perform reasonably well in terms of bias, although they do not surpass our proposed estimators. Whilst the bootstrapped standard errors of `TMLE` are smaller than the estimated asymptotic standard errors of `FIPW`[7], both methods provide corresponding 95% confidence intervals that contain the truth at any percentile. In contrast, `LDML` yields point estimates that are less favorable compared to `TMLE` and `FIPW`. However, its asymptotically calibrated confidence intervals still effectively capture the true QTEs, despite having the widest spans across all quantile levels.

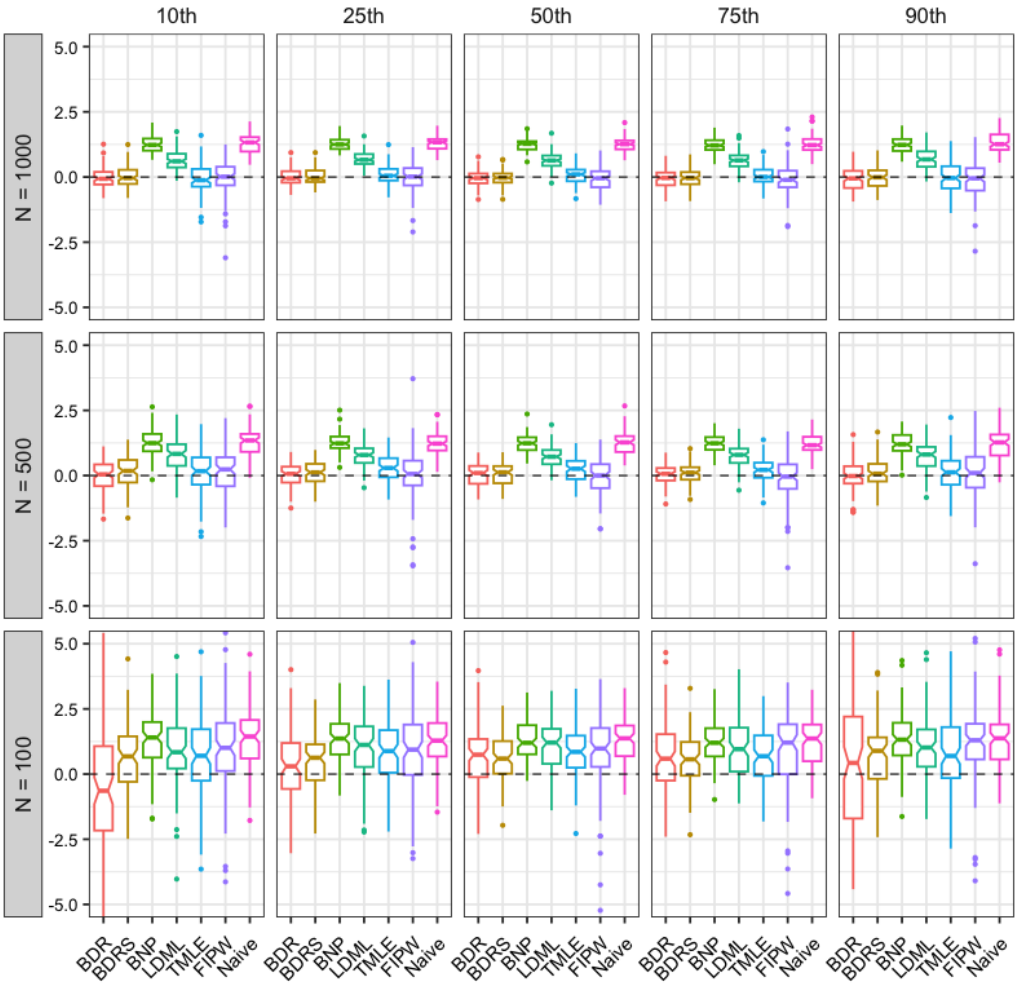

Figure 5.2: Sampling distributions of of the difference between the true and estimated quantities for 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} QTEs across 100 replicates. The dashed line indicates zero difference.

Boxplots in Figure 5.2 offer more insight on the sampling distributions of the difference between the true and estimated quantities which produced by all estimators across 100 simulated datasets. When $N = 1000$, `BDR` and `BDRS` showcase nearly zero median (or mean) bias as well as small variation, outperforming other methods. Their strong performance persists even in smaller sample size of $N = 500$. Interestingly, the advantage of `BDRS`, which is developed by adopting a hierarchical shrinkage prior, becomes prominent when $N = 100$. While the performance of `BDR` exhibits instability in the

[7] Firpo (2007) also recommends bootstrapping as possibly a good alternative to analytical standard errors estimation in `FIPW`.

presence of high-dimensional covariates, BDRS handles such settings more effectively, as evidenced by the remarkably reduced box widths observed for all QTEs of interest.

Table 5.2: Simulation Results for SD1, Average Bias

		Estimation Methods						
Percentiles	N	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

Notes: This table displays the average bias across 100 replicates of different estimation methods. The rows contain results for various percentile levels and for various sample size N.

Table 5.2 numerically validates our above findings on the pattern of mean bias. Both BDR and BDRS alternately secure the top rank, exhibiting the smallest average bias across all computed percentiles. While the challenge of pronounced average bias is inherent in the high-dimensional setting ($N = 100$ and $p = 40$), as the p/N ratio decreases, the average bias diminishes relatively fast for QTEs estimated by methods BDR, BDRS, TMLE, and FIPW. Additionally, LDML also exhibits a declining trend in average bias, albeit at a slower rate. This phenomenon is not observed with the Naive and BNP estimators.

With respect to the relative Mean Absolute Error (MAE), as presented in Table 5.3, our proposed approach outperforms all competitors at high percentiles 50^{th}, 75^{th}, and 90^{th}. Meanwhile, BDRS performs better than BDR in the majority of cases, especially in high-dimensional scenarios. In addition, line plots of raw MAE in Figure 5.3 illustrate a downward trend for both BDR and BDRS across all quantile levels as the p/N ratio decreases.

In conclusion, BDR and BDRS demonstrate similarly excellent performance in moderate dimensionality, thereby facilitating robustness checks in practical use. BDRS even provides extra merit thanks to its adaptation to high-dimensional settings. It is trivial that the Bayesian Adaptive Lasso in BDRS is only one option among a wide range of shrinkage priors which can be incorporated into our proposed framework. Thus, the results from this simulation exercise imply the great potential of our methodology in flexibly handling high dimensions when estimating quantile treatment effects.

Table 5.3: Simulation Results for SD1, relative MAE

		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

Notes: This table displays the relative Mean Absolute Error (MAE) of different estimation methods across 100 replicates. The rows contain results for various percentile levels and for various sample size N. The relative MAE is the MAE in comparison with the Naive method as the benchmark, where $MAE = R^{-1} \sum_{r=1}^{R} |\hat{\alpha}_r - \alpha|$ and $R = 100$.

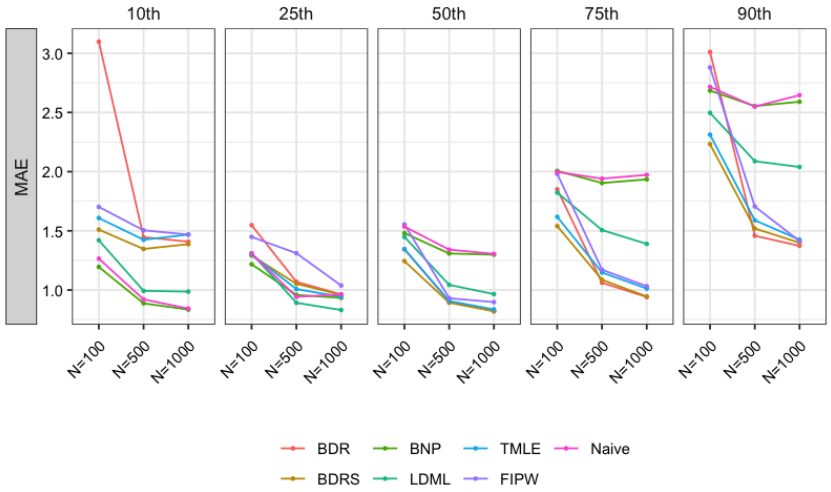

Figure 5.3: Line plots of raw MAE for 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} QTEs estimated based on 100 replicates.

5.2 Simulation Design 2 (SD2)

In this simulation design, we explore a setting with nonlinearities, where the mean of potential outcomes involves polynomial functions of covariates. We draw covariates $\mathbf{X}$ with $p = 5$ from the independent standard normal distributions. The sample size is fixed at $N = 1000$ for the remainder of the exercise

given small sample sizes are likely to be inadequate to explore nonlinearities. The true model for data generation takes the following form:

$$
\begin{aligned}
T \mid \mathbf{X} &\sim \mathcal{B}ern\left(\pi(\mathbf{X})\right), \\
Y^{(0)} \mid \mathbf{X} &\sim \mathcal{N}\left(\mu(\mathbf{X}), 1^2\right), \\
Y^{(1)} \mid \mathbf{X} &\sim \mathcal{N}\left(1+\mu(\mathbf{X}), 1.5^2\right), \\
Y &= T \times Y^{(1)} + (1-T) \times Y^{(0)}; \\
\text{where} \qquad \pi(\mathbf{X}) &= \left\{1+\exp\left[-(-0.6X_1 + 0.8X_2 + 1.2X_3)\right]\right\}^{-1}, \\
\mu(\mathbf{X}) &= -X_1 + X_2^2 + 1.5X_3X_4 + 1.5X_5^3.
\end{aligned}
\tag{5.2}
$$

Unlike the first simulation study, the true unconditional density and the true quantiles of the potential outcomes for this simulation are not analytically achievable. However, the true unconditional quantiles can be derived approximately from a large sample. At sample size of 10^7, the approximate values for true 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} QTEs are $\Delta_{0.10} = 0.705$, $\Delta_{0.25} = 0.794$, $\Delta_{0.5} = 1.022$, $\Delta_{0.75} = 1.205$, and $\Delta_{0.90} = 1.205$, respectively.

We introduce two simpler variants of our proposed framework in this simulation exercise. The first variant consists of Bayesian Outcome Modeling without and with shrinkage priors, represented by `BOM` and `BOMS` estimators, respectively. It could be regarded as an outcome-regression-based approach that omits the treatment assignment model fitted in the initial step of the BADR framework. Instead, it focuses solely on estimating the conditional distribution by using multiple Bayesian quantile regressions in the outcome model of each treatment group. Shrinkage priors, akin to the doubly robust approach, can be readily incorporated. In particular, the `BOMS` estimator considers the Adaptive Lasso prior. The second variant is Bayesian Propensity Score Analysis (`BPSA`), a treatment-assignment-based approach. Specifically, it involves fitting the treatment assignment using a logit BART model. Subsequently, it employs multiple Bayesian quantile regressions to model the conditional distribution of the outcome given the posterior mean of the propensity score in each treatment group. Further details on the implementation can be found in Appendix E.2.

We evaluate the performance of the methods with respect to two distinct modeling strategies: linear and nonlinear specification. For the linear specification, we use 5 raw covariates $X_1, \ldots, X_5$. For the nonlinear specification, we expand the covariates $\mathbf{X}$ to a 55-dimensional space by incorporating full cubic polynomials along with interaction terms. `BDR`, `BOM`, and `BNP` are excluded as competitors in the second specification since they are less suitable for high-dimensional contexts.

SD2a. Linear Specification

Table 5.4 illustrates simulation results when 5-dimensional covariates are employed as control variables. Overall, `BDR` and `BDRS` outperform all competing methods, achieving the smallest average bias at 25^{th} and 75^{th} percentiles. Frequentist methods including `TMLE`, `LDML`, and `FIPW`, individually rank first once at 10^{th}, 50^{th}, and 90^{th}, respectively. However, each of them is less superior to our proposed estimators in at least three of the five quantile levels of interest. The Bayesian nonparametric method, `BNP`, continues to register the lowest rank, offering only a marginal reduction in bias compared to the benchmark.

Table 5.4: Simulation Results for SD2a, Average Bias and relative MAE

	Bias					MAE				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Linear specification										
BDR	0.055	0.002	-0.032	-0.008	0.016	1.349	1.049	0.621	0.597	0.781
BDRS	0.061	0.002	-0.030	-0.006	0.024	1.331	1.056	0.624	0.595	0.790
BOM	0.040	0.092	-0.013	-0.094	-0.053	0.926	0.808	0.505	0.559	0.828
BOMS	0.059	0.102	-0.002	-0.076	-0.025	0.924	0.801	0.507	0.558	0.823
BPSA	0.078	-0.042	-0.044	0.061	0.112	0.752	0.945	0.493	0.555	0.725
BNP	0.335	0.368	0.343	0.328	0.284	1.114	1.005	0.861	0.939	1.095
LDML	0.009	0.010	-0.009	0.019	0.163	0.997	0.963	0.584	0.620	0.832
TMLE	0.000	-0.005	-0.022	-0.041	0.025	1.284	1.181	0.657	0.683	1.156
FIPW	0.019	-0.039	-0.034	-0.010	0.014	1.576	1.218	0.671	0.717	1.240
No covariates										
Naive	0.468	0.434	0.430	0.377	0.312	1.000	1.000	1.000	1.000	1.000

Notes: This table displays the average bias and the relative Mean Absolute Error (MAE) of different estimation methods across 100 replicates. The relative MAE is the MAE in comparison with the Naive method as the benchmark, where $MAE = R^{-1} \sum_{r=1}^{R} |\hat{\alpha}_r - \alpha|$ and $R = 100$.

Further investigation can unveil the mechanics of our analog doubly robust estimators. It is essential to note that, by considering only 5 raw covariates $\mathbf{X}$, there is a misspecification in the functional form of covariates in the outcome equation.

Bayesian Outcome Modeling estimators, `BOM` and `BOMS`, inherit the advantages of Bayesian Quantile Regression and shrinkage priors, as same as our primary approach. Nonetheless, since these estimators ignore the treatment assignment equation, they exhibit significantly higher average bias than doubly robust estimators, irrespective of whether penalization in covariate space is introduced or not. In contrast, by fitting the propensity score, doubly robust estimators gain another protective layer against misspecification of the outcome. A similar rationale applies to the favorable performance of `TMLE`, which is originally a doubly robust estimator from frequentist viewpoints. Other methods, `LDML` and `FIPW`, do not utilize both the whole conditional cumulative distribution function and the propensity score function as inputs in the doubly robust estimation procedure. However, their reliance on the treatment assignment equation from the outset makes them less affected by the misspecification of the outcome, resulting in reasonably good performance.

Bayesian Propensity Score Analysis estimator, `BPSA`, exhibits lower average bias than both `BOM` and `BOMS` when estimating 25^{th} and 75^{th} QTEs. Nevertheless, its performance is dominated by both `BDR` and `BDRS` in terms of average bias across all evaluated quantile levels. Despite sharing the first stage with doubly robust estimators when fitting the propensity score by a logit BART model, `BPSA` then uses posterior samples of propensity score rather than 5-dimensional covariates as control variables to estimate the conditional distribution of potential outcomes. Its inferiority compared to `BDR` underscores the doubly robust approach, suggesting that using the estimated propensity score alone is less favorable, especially when the treatment assignment and potential outcome equations contain different sets of control variables. This observation also aligns with the poor performance of `BNP` and reinforces our conclusion from the first simulation exercise.

SD2b. Nonlinear Specification

Table 5.5 presents simulation results when the 55-dimensional expansion of covariates is utilised as control variables. It can be seen that the performance of `BDRS` is noticeably improved, particularly in the extreme tails. `BDRS` outperforms all frequentist methods in terms of both average bias and MAE, across most quantile levels except for two instances when it ranks second after `TMLE`. This finding again highlights the superiority of `BDRS` in high dimensions. Continuing our previous discussion on the double robustness of `BDRS`, when considering this basis expansion of covariates, the treatment assignment equation is misspecified to some extent. Because the logit link is maintained across Bayesian methods, the use of high-order polynomials induces a nonlinear functional form of $\mathbf{X}$, whereas the true model involves only a linear combination of X_1, X_2, and X_3. `BPSA` produces larger average bias than `BOMS` across almost all quantile levels, other than 25^{th} QTEs, and remains persistently dominated by `BDRS`.

Table 5.5: Simulation Results for SD2b, Average Bias and relative MAE

	Bias					MAE				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Nonlinear specification										
BDRS	-0.014	0.015	0.019	0.027	0.011	0.724	0.859	0.503	0.527	0.550
BOMS	0.041	0.027	0.020	0.015	0.001	0.547	0.709	0.463	0.455	0.441
BPSA	0.140	-0.019	-0.029	0.077	0.099	0.744	0.913	0.494	0.562	0.724
LDML	0.111	0.052	0.047	0.062	0.182	0.880	0.922	0.663	0.683	0.820
TMLE	-0.020	0.009	0.023	0.023	0.064	0.728	0.863	0.500	0.533	0.667
FIPW	0.034	0.120	0.090	-0.050	-0.051	1.577	1.290	0.873	0.926	1.298
No covariates										
Naive	0.468	0.434	0.430	0.377	0.312	1.000	1.000	1.000	1.000	1.000

Notes: This table displays the average bias and the relative Mean Absolute Error (MAE) of different estimation methods across 100 replicates. The relative MAE is the MAE in comparison with the Naive method as the benchmark, where $MAE = R^{-1}\sum_{r=1}^{R}|\hat{\alpha}_r - \alpha|$ and $R = 100$.

In summary, our proposed doubly robust estimators (`BDR` and `BDRS`) consistently surpass at least one among outcome-regression-based estimators (`BOM` and `BOMS`) or treatment-assignment-based estimator (`BPSA`) regarding the average bias, when either the outcome equation or treatment equation is misspecified. By flexibly incorporating shrinkage priors, `BDRS` outperforms its Bayesian nonparametric counterpart and all frequentist competitors in high-dimensional settings. This result demonstrates that our proposed framework features not only adaptability to complexity but also robustness to misspecification.

6 Empirical Illustration

6.1 Overview

To demonstrate the applicability and usefulness of our proposed method, we revisit the microcredit study by Crépon et al. (2015), which was derived from a randomized experiment conducted in Morocco. The dataset enables us to examine the potential of our approach in two distinct contexts. In the first setting, we employ the *random treatment assignment* available in the original research to investigate the effect of microcredit availability on household borrowing activities, such as the total amount of loans. Our second setting deviates from randomization – we instead use *observational data* while assuming *selection-on-observables* to evaluate the welfare impact of household loans.

The evaluation was conducted across 162 Moroccan villages that were paired based on their observable similarities. The intervention was microcredit availability, which was randomly assigned to one village within each pair. These designated villages constituted the treated group, whereas the remaining villages formed the control group. In particular, a microfinance institution was established in the treated villages between 2006 and 2007. In 2009, a follow-up study surveyed 5551 households in both treated and control villages.

The expansion of microcredit, or access to loans in general, can have potentially heterogeneous effects on household welfare for several reasons. First, households have diverse loan take-up behaviors. They may differentially select into borrowing activities based on their characteristics, leading to varying outcomes. Those who do not take up loans may end up worse off due to effects on wages or the displacement of informal lending in a dynamic general equilibrium (Kaboski and Townsend 2011; Morduch 1999). Second, among borrowers, the effects may vary due to differences in the efficiency of loan use and uneven investment opportunities. Indeed, certain households may not benefit from loans if the requirements for investment purposes are restrictive or the term to maturity is too short (Banerjee 2013). Additionally, multiple microlenders in a community can engage in exploitative lending practices and "overlending" to households who cannot feasibly repay the loan (Schicks 2013; Ahmad 2003). This can result in high-productivity borrowers benefiting from the positive impact, whereas the most vulnerable borrowers are systematically harmed by the saturation of credit markets. In summary, there are potential winners and losers to financial market expansion. Even if disadvantaged groups are small, social welfare consequences could be substantial, particularly if economic inequality across households is exacerbated (Meager 2022).

The average treatment effect (ATE), which is most commonly utilized in empirical research, cannot reveal this heterogeneity. Even though loan access might have no impact on average, it could still have significant positive or negative effects on different types of households. This policy implication is particularly critical for developing countries. To gain a more comprehensive understanding of causal effects, it is worthwhile to estimate unconditional quantile treatment effects (QTEs), which offer a valid measurement that goes beyond the ATE for the entire population. Therefore, our proposed framework is well suited for this empirical context.

In contrast to the original paper and previous studies that typically rely on randomized controlled trial (RCT) design and ad hoc selection of baseline covariates, our approach offers more flexible specifications and data-driven estimation. This enables us to conduct new analyses using either

data from randomized experiments or observational data, as demonstrated in sections 6.2 and 6.3, respectively.

Specifically, our general strategy is to initially create a large set of covariates by combining village pair dummies and full cubic polynomials along with interaction terms of household observed characteristics. Once collinear columns are removed, this set serves as the baseline specification of $\mathbf{X}$ and can be readily integrated into our Bayesian Analogue of Doubly Robust (BADR) estimation framework. Given the high dimensionality of this empirical issue, we opt for the Bayesian Doubly Robust estimator with Adaptive Lasso (`BDRS`) due to its proven merits in our prior simulation study. To compare our results with the benchmark, we also include the Naive estimator (`Naive`) in our analysis.

6.2 Impact of Microcredit Availability on Loan Amount

We begin with the context of random treatment assignment, where our objective of interest is the effect of microcredit availability on the total amount of loans at the household level. To examine the balance between the treated and control groups, we select pre-treatment covariates which are observed characteristics for each household, including head age, education of the head, number of adults, total number of members in a household, indicators for households doing animal husbandry, doing other non-agricultural activities, and whether household spouse responded to the survey. Table 6.1 reports the mean values of these covariates in addition to the outcome and treatment variables, both for the whole sample and for each of the treated and control groups.

Table 6.1: Summary Statistics of Households

	Treated	Control	Treated – Control		
Treatment: Microcredit Availability (RCT)	Mean (sd)	Mean (sd)	Diff.Mean		t-statistic
Outcome variable					
Total amount of loans (in MAD)	2350.44 (10865.84)	1547.75 (7381.73)	802.69	*	2.54
Baseline covariates					
Head age	49.18 (15.83)	48.14 (15.85)	1.05	.	1.95
Head with no education	0.67 (0.47)	0.68 (0.47)	-0.01		-0.89
Number of members	5.70 (2.54)	5.64 (2.44)	0.06		0.71
Number of adults	3.81 (1.99)	3.76 (1.91)	0.05		0.83
Number of members aged 6-16	1.22 (1.29)	1.25 (1.26)	-0.03		-0.72
Declared animal husbandry activities	0.60 (0.49)	0.55 (0.50)	0.05	**	2.73
Declared non-agricultural activities	0.17 (0.37)	0.21 (0.41)	-0.04	**	-3.15
Spouse of head responded	0.09 (0.29)	0.07 (0.26)	0.02	*	2.27
Member responded	0.05 (0.22)	0.05 (0.21)	0.00		0.62

Data sources: Moroccan household survey (Crépon et al., 2015).

Although the randomization of microcredit availability and the absence of confounding factors leading to self-selection into treatment is plausible, there are slightly imbalances in covariates across the two groups. Regarding the unconditional means, the households' total loan amount for the treated group (2350.44) significantly exceeds that of the control group (1547.75). The potential heterogeneity of microcredit motivates us to further investigate this positive average treatment effect using quantile analysis.

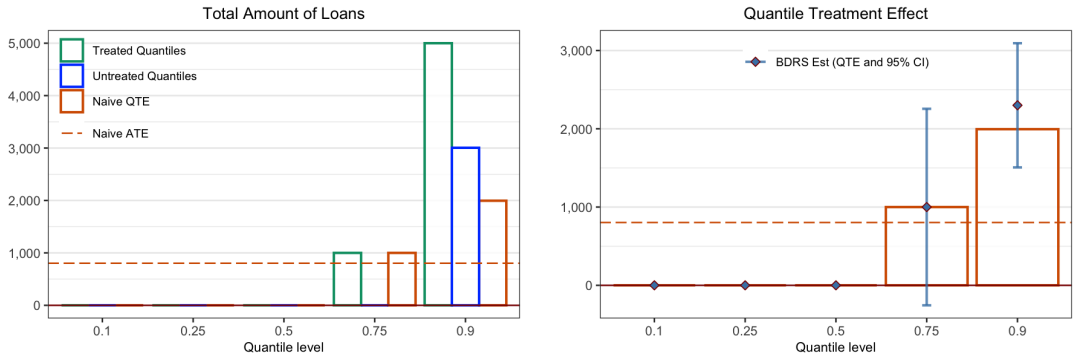

Figure 6.1: Quantile Treatment Effects (QTEs) of microcredit availability on households' total loan amount. The graph on the left demonstrates Naive estimation results. Red bar plots represent naive QTEs, which are differentials between empirical quantiles of treated group (in green) and control group (in blue). Red dashed line indicates naive Average Treatment Effects (ATE), which is simple mean difference between these two groups. Results obtained using BDRS method, QTE point estimates and corresponding 95% CI at five quantile levels based on 100 bootstrap replications, are plotted as error bars in the right-hand graph.

The results of Quantile Treatment Effects (QTEs), as estimated by the `Naive` and `BDRS` methods, are depicted in Figure 6.1. According to the findings, microcredit expansion has a precise zero effect below the 75$^{\text{th}}$ percentile of the distribution of total loan amounts, but exhibits positive effects above this threshold. In particular, at the 90$^{\text{th}}$ percentile, the positive effect is statistically significant (2300), contributing to the decomposition of the average treatment effect (802.69). Compared with naive estimates, `BDRS` produces similar results, only higher at the 90$^{\text{th}}$ percentile; however, the difference is insignificant. The result is robust after adjusting for the influence of covariate imbalance on the outcome.

Using the same dataset, findings in Chernozhukov et al. (2017) and Jacob (2021) also document the heterogeneity of the microcredit availability on total loan amount; however, their estimand is conditional ATE, different from this paper (QTEs).

6.3 Impact of Loan Access on Household Outcomes

Our second objective is to explore the causal impact of access to loans on household welfare, with a focus on the distribution of consumption and business outcomes, including total consumption, consumption of temptation goods, total output, and total profit. Unconditional QTEs provide deeper insights into the potential heterogeneity of causal effects across the distribution of each outcome interest, as well as the resulting change in household inequality. The binary treatment we consider is the actual borrowing status recorded at the household level. Table 6.2 indicates that the difference in mean between two groups of households (borrowers and non-borrowers) is highly statistically significant regarding consumption, but not for business outcomes. However, there are two caveats to these naive ATE estimates.

Table 6.2: 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

Data sources: Moroccan household survey (Crépon et al., 2015). *Definition:* Total Consumption is monthly total consumption (in MAD); Temptation Goods is monthly expenditure on temptation and entertainement(in MAD); Total Output is sum of agricultural, livestock, and non-agricultural business production over the 12 months prior to the survey (in MAD); Total Profit is total profit of self-employment activities over the 12 months prior to the survey (in MAD).

Firstly, all outcome variables in this empirical setting exhibit heavy tails and large variability, as illustrated in histograms in Figure 6.2. This is another motivation for quantile analysis since estimation results for a set of quantiles would be less susceptible to the influence of outliers than results for the mean.

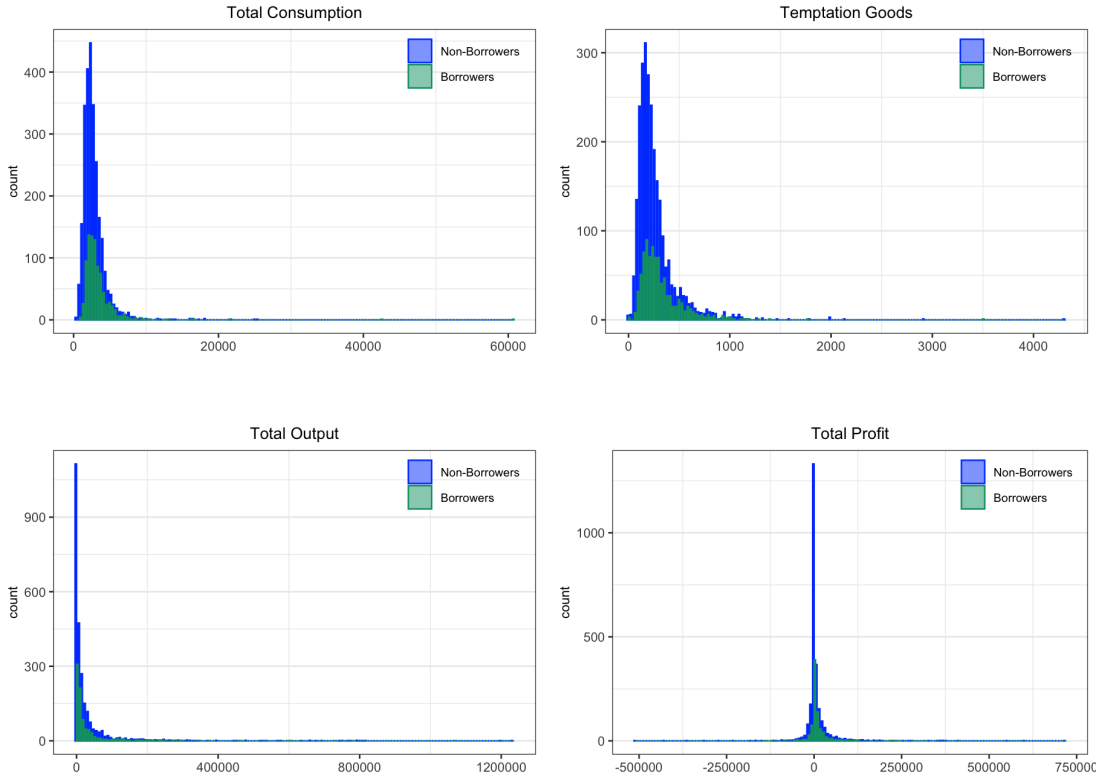

Figure 6.2: Histograms of various consumption and business outcomes of borrowing households (in green) and nonborrowing households (in blue). These graphs display raw data without any truncation applied.

Secondly, the treatment variable – borrowing pattern observed in the dataset – is no longer randomly assigned among households in the present context. This is confirmed by the imbalances between these two groups regarding the mean values of the observed characteristics, as shown in

Table 6.3. Specifically, borrowing households tend to have larger average household sizes. They are also more inclined to engage in non-agricultural self-employment activities and reside in villages where microcredit is available. The discrepancy observed is more than what would be expected by pure chance. Therefore, to identify causal effects using non-experimental data, we pursue the *selection-on-observables* assumption. That means, conditional on observed covariates, unmeasured factors that influence household loan access are independent of household outcomes.

Table 6.3: 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

Data sources: Moroccan household survey (Crépon et al., 2015).

Whilst a violation of randomization may threaten the performance of the naive estimator, the `BDRS` estimator serves as a debiasing device, as illustrated in our simulation using synthetic data. Table 6.4 presents the results for key outcome variables related to household consumption and business. Unlike the first setting, the `BDRS` estimates differ considerably from the naive estimates of QTEs because selection bias is accounted for in our proposed approach. Overall, the point estimates at extreme tails (10^{th} and 90^{th} percentiles) are fairly imprecise, as indicated by large credible intervals compared to the other quantile levels. Interestingly, the causal effect in the upper tail remains significantly positive in most cases.

Regarding *total consumption*, although all naive estimates of ATE and QTE are positive, estimation results obtained using the `BDRS` method reveal notably lower effects across all quantile levels. The effects of loan access are most pronounced at the 75^{th} and 90^{th} percentiles of the consumption distribution; however, they are insignificantly positive. While the naive method overestimates the effect of loan access compared to the `BDRS` estimator, the upward bias suggests a possible selection-on-gain pattern. Households inclined to borrow to support their consumption are more likely to gain higher total consumption when they have financial access.

Further examination of the impact of borrowing on *temptation consumption* shows a similar upward bias in the `Naive` method relative to the `BDRS` method. The effect is slightly negative at the lowest percentile (10^{th}) yet clearly insignificant. By contrast, significant positive effects are observed at the median and higher percentiles. This seems inconsistent with other works that have found a statistically significant reduction in nonessential expenditures. However, these studies used different treatment variables and designs compared to this paper.

Table 6.4: Quantile Treatment Effects of Loan Access on Household Outcomes.

		BDRS			**Naive**
Outcomes	Percentiles	QTEs	Upper bound	Lower bound	QTEs
Total Consumption	10th	20.093	1469.207	-1429.020	232.795
	25th	9.242	173.511	-155.027	173.456
	50th	79.949	229.587	-69.690	229.753
	75th	132.22	273.974	-9.534	286.680
	90th	237.699	543.355	-67.956	685.442
Temptation Goods	10th	-8.69	65.660	-83.040	17.380
	25th	13.035	29.871	-3.801	21.725
	50th	**30.415**	45.962	14.868	43.450
	75th	**47.795**	79.225	16.365	60.830
	90th	**78.21**	129.145	27.275	78.210
Total Output	10th	0	13146.948	-13146.948	0.000
	25th	-330	19.769	-679.769	1093.446
	50th	50	1385.992	-1285.992	1787.500
	75th	1666	6933.205	-3601.205	2771.616
	90th	**27360**	52964.198	1755.802	2744.044
Total Profit	10th	**-5500**	-1183.536	-9816.464	-1142.697
	25th	-945	158.825	-2048.825	-241.876
	50th	**561**	1117.727	4.273	979.125
	75th	1780.769	4273.915	-712.377	549.373
	90th	**8954.377**	16664.233	1244.520	-1086.350

Notes: Upper bound and Lower bound for BDRS method are associated with the estimates of 95% CI based on 100 bootstrap replications.

With respect to household business outcomes, there is a statistically significant increase in *total output*, concentrated only at the highest quantile level. For the rest of the community, specifically those below the 90th percentile, no systematic change appears to be taking place. Consequently, the majority of the total output distribution remains unchanged with or without universal access to loans.

Interestingly, there is notable evidence of heterogeneous effects on *total profit*. The effect on the median household estimated by the `BDRS` method is quite close to the naive ATE estimate, which is moderately positive. In general, access to loans has a favorable impact on households' profit by shifting the center of distribution towards the right. However, the impacts exhibited at extreme tails are more dramatic, with a negative effect at the lowest percentile (10^{th}) and a positive effect at the highest percentile (90^{th}), and both are statistically significant. If the rank invariance assumption is invoked, the rightward expansion of the upper tail means that high-profit households gain benefits, while the leftward expansion of the lower tail means that low-profit households experience loss when loans are accessible to everyone, compared to the opposite counterfactual scenario. While this assumption might be difficult to defend given the complexity and nonlinearities inherent to the financial market, interpretations about the shape change of distribution of household total profit remain valid. There do exist both winners and losers, even when we cannot identify the specific households that belong to each group. The outcome distribution disperses wider leading to the exacerbation of inequality across households.

Taken together, the estimated QTE patterns collectively indicate that broadening financial access is likely to result in an ex-post rise in economic inequality across households. Specifically, the increase in total output and consumption at the household level is solely attributable to the right tail of distributions expanding rightward, suggesting that certain households are likely to experience an improvement in their economic circumstances without incurring any systematic losses from others. Further investigation of total profit, however, reveals a more nuanced picture. While the overall impact of access to loans on total profit is positive, indicating a shift towards higher profits for many households, there is evidence of extreme heterogeneity. The effect is asymmetric as certain households may experience negative effects on their profits.

Although the treatment variable and identification strategies employed in this setting differ from those used in Crépon et al. (2015), the findings converge in several respects. In the original paper, both the reduced-form quantile regressions and instrumental variable (IV) estimates suggest substantial heterogeneity in the profitability of microfinance investments and emphasize the detrimental effects on certain households. Specifically, their reduced-form quantile analysis measures Intention-to-Treat (ITT) effects because the treatment variable is microcredit availability at the village level rather than actual borrowing at the household level. Additionally, the IV estimates in this study reveal changes in the unconditional distribution of total profit for those who take up microcredit (i.e., compliers only). These results are only valid when randomization holds. In contrast, the findings of this paper have broader implications for the understanding of economic inequality, as we focus on the entire population of households utilizing non-experimental data and a *selection-on-observables* assumption.

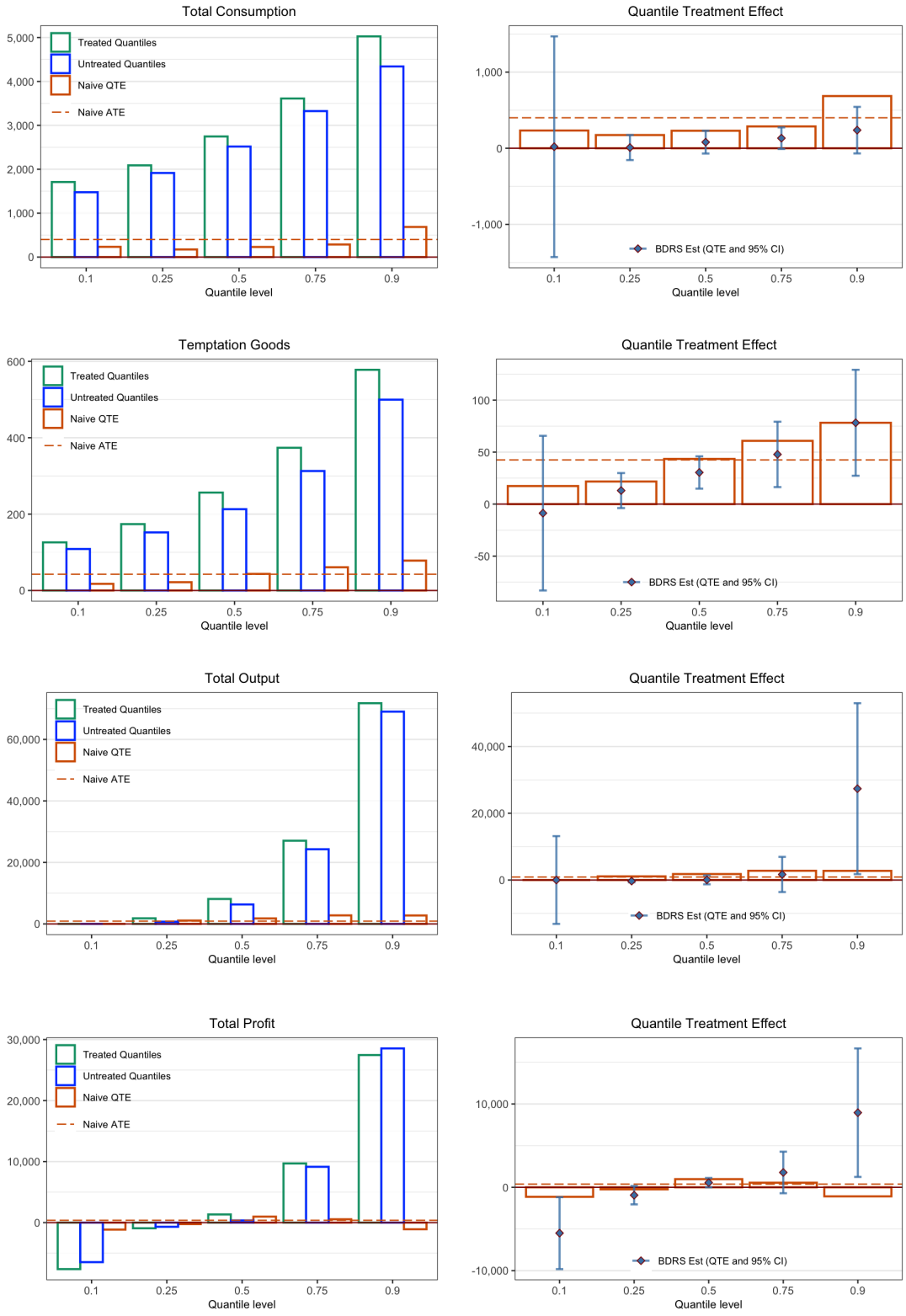

Figure 6.3: Quantile Treatment Effects (QTEs) of loan access on various household outcomes. Graph on the left demonstrates Naive estimation results. Red bar plots represent naive QTEs, which are differentials between empirical quantiles among borrowing households (in green) and nonborrowing households (in blue). Red dashed line indicates naive Average Treatment Effect (ATE), which is simple mean difference between these two groups. Results obtained using BDRS method, QTE point estimates and corresponding 95% CI at five quantile levels based on 100 bootstrap replications, are plotted as error bars in the right-hand graphs.

7 Conclusion

The goal of this paper was to address the challenges associated with estimating unconditional Quantile Treatment Effects (QTEs) in observational studies and to make a contribution to the burgeoning econometric literature on QTEs as well as causal machine learning. We introduced a novel approach, Bayesian Analog of Doubly Robust (BADR) estimation, which accounts for potentially high-dimensional covariates. The framework features a highly flexible Bayesian modeling scheme that showcases favorable frequentist properties in finite samples, even in the presence of high dimensions or model misspecifications, which has not been explored in previous literature. This approach, while not fully Bayesian in nature, offers a straightforward and versatile implementation for integrating probabilistic machine learning techniques into causal analysis on quantiles, with precise estimation and reliable uncertainty quantification. These attributes are particularly advantageous in complex, high-dimensional settings.

The performance of the proposed method was assessed through a simulation study in two different settings. The first simulation focused on a linear setting with varying feature dimensionality, whereas the second simulation considered a nonlinear setting and examined the double robustness of the proposed estimators. Through a comparison with both the naive approach and existing popular estimators, the simulation results consistently indicated a substantial improvement in bias reduction for QTE estimates when using the new method. This finding demonstrates that our proposed framework features not only the ability to adapt to high dimensions and complexity, but also robustness to misspecification.

The empirical illustration of estimating QTEs of financial access on household outcomes showed the potential benefits of using causal inference on quantiles to help characterize the heterogeneity or distributional impact of interventions, which is often appealing to researchers and easily conveyed to policymakers and stakeholders. Our proposed approach makes this possible even in the absence of experimental data. We found strong evidence for an overall positive effect yet heterogeneous across different points of outcome distributions. An ex-post rise in economic inequality among households is likely to occur, primarily driven by significant improvements in consumption and business outcomes at the top quantiles. However, certain households may experience adverse effects on their total profit.

An interesting extension of this framework to be explored in the future would be estimating QTEs when the *selection-on-observable* assumption is violated, that is, when there exists unmeasured confounding that drives endogenous selection into treatment. Another aspect is improving bootstrap inference scheme for doubly robust estimators. Although this objective could be achieved effectively in the BADR framework for average treatment effect, it is computationally demanding when applied to quantiles.

References

Ahmad, Mokbul Morshed. 2003. "Distant Voices: The Views of the Field Workers of NGOs in Bangladesh on Microcredit." *Geographical Journal* 169 (1): 65–74.

Albert, James H, and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 669–79.

Alhamzawi, Rahim, and Haithem Taha Mohammad Ali. 2020. "A New Gibbs Sampler for Bayesian Lasso." *Communications in Statistics-Simulation and Computation* 49 (7): 1855–71.

Alhamzawi, Rahim, Keming Yu, and Dries F Benoit. 2012. "Bayesian Adaptive Lasso Quantile Regression." *Statistical Modelling* 12 (3): 279–97.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Antonelli, Joseph, Matthew Cefalu, Nathan Palmer, and Denis Agniel. 2018. "Doubly Robust Matching Estimators for High Dimensional Confounding Adjustment." *Biometrics* 74 (4): 1171–79.

Antonelli, Joseph, Georgia Papadogeorgou, and Francesca Dominici. 2022. "Causal Inference in High Dimensions: A Marriage Between Bayesian Modeling and Good Frequentist Properties." *Biometrics* 78 (1): 100–114.

Antonelli, Joseph, Giovanni Parmigiani, and Francesca Dominici. 2019. "High-Dimensional Confounding Adjustment Using Continuous Spike and Slab Priors." *Bayesian Analysis* 14 (3): 805.

Athey, Susan, Guido W Imbens, and Stefan Wager. 2018. "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.

Banerjee, Abhijit Vinayak. 2013. "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.

Bang, Heejung, and James M Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61 (4): 962–73.

Belloni, Alexandre, Victor Chernozhukov, Iván Fernández-Val, and Christian Hansen. 2017. "Program Evaluation and Causal Inference with High-Dimensional Data." *Econometrica* 85 (1): 233–98.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects After Selection Among High-Dimensional Controls." *The Review of Economic Studies* 81 (2): 608–50.

Benoit, Dries F., and Dirk Van den Poel. 2017. "bayesQR: A Bayesian Approach to Quantile Regression." *Journal of Statistical Software* 76 (7): 1–32. https://doi.org/10.18637/jss.v076.i07.

Bhadra, Anindya, Jyotishka Datta, Nicholas G Polson, and Brandon Willard. 2019. "Lasso Meets Horseshoe: A Survey." *Statistical Science* 34 (3): 405–27.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica* 82 (5): 1671–1748.

Callaway, Brantly, and Tong Li. 2019. "Quantile Treatment Effects in Difference in Differences Models with Panel Data." *Quantitative Economics* 10 (4): 1579–1618.

Carvalho, Carlos M, Nicholas G Polson, and James G Scott. 2009. "Handling Sparsity via the Horseshoe." In *Artificial Intelligence and Statistics*, 73–80. PMLR.

———. 2010. "The Horseshoe Estimator for Sparse Signals." *Biometrika* 97 (2): 465–80.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney

Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." Oxford University Press Oxford, UK.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val. 2017. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." *arXiv e-Prints*, arXiv–1712.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81 (6): 2205–68.

Chernozhukov, Victor, Whitney K Newey, and Rahul Singh. 2022. "Automatic Debiased Machine Learning of Causal and Structural Effects." *Econometrica* 90 (3): 967–1027.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Chipman, Hugh A, Edward I George, and Robert E McCulloch. 2010. "BART: Bayesian Additive Regression Trees."

Chipman, Hugh, Edward George, and Robert McCulloch. 2006. "Bayesian Ensemble Learning." *Advances in Neural Information Processing Systems* 19.

Crépon, Bruno, Florencia Devoto, Esther Duflo, and William Parienté. 2015. "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): 123–50.

Daniel, Rhian M. 2014. "Double Robustness." *Wiley StatsRef: Statistics Reference Online*, 1–14.

Díaz, Iván. 2017. "Efficient Estimation of Quantiles in Missing Data Models." *Journal of Statistical Planning and Inference* 190: 39–51.

Doksum, Kjell. 1974. "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case." *The Annals of Statistics*, 267–77.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2021. "The Impact of Free Secondary Education: Experimental Evidence from Ghana." National Bureau of Economic Research.

Farrell, Max H. 2015. "Robust Inference on Average Treatment Effects with Possibly More Covariates Than Observations." *Journal of Econometrics* 189 (1): 1–23.

Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica* 75 (1): 259–76.

Firpo, Sergio, and Cristine Pinto. 2016. "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures." *Journal of Applied Econometrics* 31 (3): 457–86.

Frölich, Markus, and Blaise Melly. 2013. "Unconditional Quantile Treatment Effects Under Endogeneity." *Journal of Business & Economic Statistics* 31 (3): 346–57.

Glewwe, Paul, and Petra Todd. 2022. *Impact Evaluation in International Development: Theory, Methods, and Practice*. World Bank Publications.

Gramacy, Robert B, and Nicholas G Polson. 2012. "Simulation-Based Regularized Logistic Regression."

Guo, Bufei. 2019. "Variational Inference for Quantile Regression." PhD thesis, Washington University in St. Louis.

Hahn, P Richard, Carlos M Carvalho, David Puelz, and Jingyu He. 2018. "Regularization and Confounding in Linear Regression for Treatment Effect Estimation." *Bayesian Analysis* 13 (1):

163–82.

Hahn, P Richard, Jared S Murray, and Carlos M Carvalho. 2020. "Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion)." *Bayesian Analysis* 15 (3): 965–1056.

Hans, Chris. 2009. "Bayesian Lasso Regression." *Biometrika* 96 (4): 835–45.

Hastie, Trevor, and Robert Tibshirani. 2000. "Bayesian Backfitting (with Comments and a Rejoinder by the Authors." *Statistical Science* 15 (3): 196–223.

Hill, Jennifer L. 2011. "Bayesian Nonparametric Modeling for Causal Inference." *Journal of Computational and Graphical Statistics* 20 (1): 217–40.

Hines, Oliver, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. 2022. "Demystifying Statistical Learning Based on Efficient Influence Functions." *The American Statistician* 76 (3): 292–304.

Hirano, Keisuke, Guido W Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.

Ishwaran, Hemant, and J Sunil Rao. 2005. "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies." *The Annals of Statistics* 33 (2): 730–73.

Jacob, Daniel. 2021. "CATE Meets ML: Conditional Average Treatment Effect and Machine Learning." *Digital Finance* 3 (2): 99–148.

Kaboski, Joseph P, and Robert M Townsend. 2011. "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative." *Econometrica* 79 (5): 1357–1406.

Kallus, Nathan, Xiaojie Mao, and Masatoshi Uehara. 2024. "Localized Debiased Machine Learning: Efficient Inference on Quantile Treatment Effects and Beyond." *Journal of Machine Learning Research* 25 (16): 1–59.

Kennedy, Edward H. 2022. "Semiparametric Doubly Robust Targeted Double Machine Learning: A Review." *arXiv Preprint arXiv:2203.06469*.

Koenker, Roger, and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica: Journal of the Econometric Society*, 33–50.

Korobilis, Dimitris, and Kenichi Shimizu. 2022. "Bayesian Approaches to Shrinkage and Sparse Estimation." *Foundations and Trends® in Econometrics* 11 (4): 230–354. https://doi.org/10.1561/0800000041.

Kotz, Samuel, Tomasz Kozubowski, and Krzystof Podgorski. 2012. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer Science & Business Media.

Kozumi, Hideo, and Genya Kobayashi. 2011. "Gibbs Sampling Methods for Bayesian Quantile Regression." *Journal of Statistical Computation and Simulation* 81 (11): 1565–78.

Lehmann, Erich Leo et al. 1974. "Statistical Methods Based on Ranks." *Nonparametrics. San Francisco, CA, Holden-Day*.

Li, Fan, Peng Ding, and Fabrizia Mealli. 2023. "Bayesian Causal Inference: A Critical Review." *Philosophical Transactions of the Royal Society A* 381 (2247): 20220153.

Lim, Daeyoung, Beomjo Park, David Nott, Xueou Wang, and Taeryon Choi. 2020. "Sparse Signal Shrinkage and Outlier Detection in High-Dimensional Quantile Regression with Variational Bayes." *Statistics and Its Interface* 13 (2): 237–49.

Linero, Antonio R, and Joseph L Antonelli. 2023. "The How and Why of Bayesian Nonparametric

Causal Inference." *Wiley Interdisciplinary Reviews: Computational Statistics* 15 (1): e1583.

Luo, Chuji, and Michael J Daniels. 2021. "BNPqte: A Bayesian Nonparametric Approach to Causal Inference on Quantiles in r." *arXiv Preprint arXiv:2106.14599*.

Makalic, Enes, and Daniel F Schmidt. 2015. "A Simple Sampler for the Horseshoe Estimator." *IEEE Signal Processing Letters* 23 (1): 179–82.

Mallick, Himel, and Nengjun Yi. 2014. "A New Bayesian Lasso." *Statistics and Its Interface* 7 (4): 571–82.

Meager, Rachael. 2022. "Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature." *American Economic Review* 112 (6): 1818–47.

Mitchell, Toby J, and John J Beauchamp. 1988. "Bayesian Variable Selection in Linear Regression." *Journal of the American Statistical Association* 83 (404): 1023–32.

Morduch, Jonathan. 1999. "The Microfinance Promise." *Journal of Economic Literature* 37 (4): 1569–1614.

Park, Trevor, and George Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103 (482): 681–86.

Polson, Nicholas G, and James G Scott. 2010. "Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction." *Bayesian Statistics* 9 (501-538): 105.

Polson, Nicholas G, and Vadim Sokolov. 2019. "Bayesian Regularization: From Tikhonov to Horseshoe." *Wiley Interdisciplinary Reviews: Computational Statistics* 11 (4): e1463.

Robert, Christian P. 1995. "Simulation of Truncated Normal Variables." *Statistics and Computing* 5: 121–25.

Robins, James M, Miguel A Hernán, and Larry Wasserman. 2015. "On Bayesian Estimation of Marginal Structural Models." *Biometrics* 71 (2): 296.

Robins, James M, and Ya'acov Ritov. 1997. "Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models." *Statistics in Medicine* 16 (3): 285–319.

Rosenbaum, Paul R, and Donald B Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Rothe, Christoph. 2010. "Nonparametric Estimation of Distributional Policy Effects." *Journal of Econometrics* 155 (1): 56–70.

Rubin, Donald B. 1981. "The Bayesian Bootstrap." *The Annals of Statistics*, 130–34.

Saarela, Olli, Léo R Belzile, and David A Stephens. 2016. "A Bayesian View of Doubly Robust Causal Inference." *Biometrika* 103 (3): 667–81.

Scharfstein, Daniel O, Andrea Rotnitzky, and James M Robins. 1999. "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models." *Journal of the American Statistical Association*, 1096–1120.

Schicks, Jessica. 2013. "From a Supply Gap to a Demand Gap? The Risk and Consequences of over-Indebting the Underbanked." In *Microfinance in Developing Countries: Issues, Policies and Performance Evaluation*, 152–77. Springer.

Schiele, Valentin, and Hendrik Schmitz. 2016. "Quantile Treatment Effects of Job Loss on Health." *Journal of Health Economics* 49: 59–69.

Serfling, Robert J. 2009. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Shin, Heejun, and Joseph Antonelli. 2023. "Improved Inference for Doubly Robust Estimators of Heterogeneous Treatment Effects." *Biometrics*.

Sparapani, Rodney, Charles Spanbauer, and Robert McCulloch. 2021. "Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART r Package." *Journal of Statistical Software* 97: 1–66.

Spertus, Jacob V, and Sharon-Lise T Normand. 2018. "Bayesian Propensity Scores for High-Dimensional Causal Inference: A Comparison of Drug-Eluting to Bare-Metal Coronary Stents." *Biometrical Journal* 60 (4): 721–33.

Tibshirani, Robert J, and Bradley Efron. 1993. "An Introduction to the Bootstrap." *Monographs on Statistics and Applied Probability* 57 (1): 1–436.

Van der Laan, Mark J, Sherri Rose, et al. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Vol. 4. Springer.

Van der Vaart, Aad W. 2000. *Asymptotic Statistics*. Vol. 3. Cambridge university press.

Van Erp, Sara, Daniel L Oberski, and Joris Mulder. 2019. "Shrinkage Priors for Bayesian Penalized Regression." *Journal of Mathematical Psychology* 89: 31–50.

Xu, Dandan, Michael J Daniels, and Almut G Winterstein. 2018. "A Bayesian Nonparametric Approach to Causal Inference on Quantiles." *Biometrics* 74 (3): 986–96.

Zhang, Zhiwei, Zhen Chen, James F Troendle, and Jun Zhang. 2012. "Causal Inference on Quantiles with an Obstetric Application." *Biometrics* 68 (3): 697–706.

Zigler, Corwin Matthew. 2016. "The Central Role of Bayes' Theorem for Joint Estimation of Causal Effects and Propensity Scores." *The American Statistician* 70 (1): 47–54.

Zigler, Corwin Matthew, and Francesca Dominici. 2014. "Uncertainty in Propensity Score Estimation: Bayesian Methods for Variable Selection and Model-Averaged Causal Effects." *Journal of the American Statistical Association* 109 (505): 95–107.

APPENDIX

A Doubly Robust Estimator

To derive the doubly robust estimator for potential quantiles q_t $(t = 0, 1)$ as proposed in section 4 in the paper, we adopt the general strategy outlined by Kennedy (2022) and Hines et al. (2022). Without loss of generality, the following discussion focuses on the *τ-quantile of treated potential outcome*, denoted by $q_1(\tau)$. Let $\psi(\mathcal{P}_o)$ represent this estimand of interest, where $\mathcal{P}_o$ is the true join distribution of observed data $Z_i = \{Y_i, T_i, X_i\}$. The procedure first requires calculation of the estimand's efficient influence function[8]. Next, an estimator based on the efficient influence function is constructed. Finally, the asymptotic properties of the doubly robust estimator are briefly verified.

A.1 Deriving Influence Functions

Definition 1. For a given functional $\psi(.)$, the influence function for ψ is the function φ satisfying

$$\frac{\partial \psi(\mathcal{P} + \epsilon(\tilde{\mathcal{P}} - \mathcal{P}))}{\partial \epsilon}\bigg|_{\epsilon=0} = \int \varphi(z; \mathcal{P})\{\tilde{p}(z) - p(z)\}dz, \tag{A.1}$$

and $\int \varphi(z; \mathcal{P})p(z)dz = 0$ for any distribution $\mathcal{P}$ and $\tilde{\mathcal{P}}$ with densities p and $\tilde{p}$. The left-hand side measures the sensitivity of $\psi(\mathcal{P})$ to small changes (slight perturbations) in the underlying distribution $\mathcal{P}$, in the direction of a fixed, deterministic distribution $\tilde{\mathcal{P}}$. This quantity is known as the *Gateaux derivative* (Serfling 2009).

To simplify the calculation of the efficient influence function, we follow the "point mass contamination" strategy. In particular, we can isolate $\varphi(z; \mathcal{P})$ by setting $\tilde{\mathcal{P}}$ equal to a point mass at single observation $\tilde{z}$, denoted by $\mathbb{1}_{\tilde{z}}(z)$. Equation (A.1) reduces to

$$\frac{\partial \psi(\mathcal{P} + \epsilon(\mathbb{1}_{\tilde{z}} - \mathcal{P}))}{\partial \epsilon}\bigg|_{\epsilon=0} = \varphi(\tilde{z}; \mathcal{P}). \tag{A.2}$$

It should be noted that we focus here on perturbations in the direction parameterised via the one-dimensional mixture model

$$\mathcal{P}_\epsilon = \epsilon \mathbb{1}_{\tilde{z}} + (1 - \epsilon)\mathcal{P}, \quad \epsilon \in [0, 1], \tag{A.3}$$

which is called a parametric submodel. Hence, the efficient influence function at observation $\tilde{z}$ is

$$\varphi(\tilde{z}; \mathcal{P}) = \frac{d\psi(\mathcal{P}_\epsilon)}{d\epsilon}\bigg|_{\epsilon=0}. \tag{A.4}$$

Building on this general definition, we can calculate the efficient influence function in the context of estimating quantiles of treated potential outcome.

[8] Efficiency refers to locally minimax semiparametric efficiency

Theorem 1. *(Efficient Influence Function)*
Denote by $\psi_o \coloneqq \psi(\mathcal{P}_o)$ the τ-quantile of treated potential outcome under the true join distribution of observed data. The efficient influence function of ψ_o is equal to

$$\varphi(Z;\mathcal{P}_o) = -\frac{1}{f(\psi_o)}\left\{\frac{\mathbb{1}\{T=1\}}{\pi(X)}\left[\mathbb{1}\{Y\leq\psi_o\} - G(\psi_o \mid 1, X;\mathcal{P}_o)\right] + G(\psi_o \mid 1, X;\mathcal{P}_o) - \tau\right\}, \tag{A.5}$$

where $\pi(X;\mathcal{P}_o) = \mathbb{P}\left(T=1 \mid X;\mathcal{P}_o\right)$ and $G(\psi \mid 1, X;\mathcal{P}_o) = \mathbb{P}\left(Y\leq\psi \mid T=1, X;\mathcal{P}_o\right)$ are the propensity score and the conditional distribution of treated potential outcome, respectively.

Proof of Theorem 1.

- By definition, $\psi_\epsilon = \psi\left(\mathcal{P}_\epsilon\right)$ satisfies

$$\iint \mathbb{1}\left\{y\leq\psi_\epsilon\right\} f_\epsilon\left(y \mid 1, x\right) f_\epsilon\left(x\right) dydx = \tau. \tag{A.6}$$

- Denote

$$Q\left(\psi,\epsilon\right) = \iint \mathbb{1}\left\{y\leq\psi_\epsilon\right\} f_\epsilon\left(y \mid 1, x\right) f_\epsilon\left(x\right) dydx - \tau, \tag{A.7}$$

  which results in $Q\left(\psi_\epsilon,\epsilon\right) = 0$ and $Q\left(\psi_o, 0\right) = 0$.

- Also

$$Q\left(\psi,\epsilon\right) = \int_{-\infty}^{\psi} f_\epsilon\left(y\right) dy - \tau, \tag{A.8}$$

  hence

$$\left[\frac{\partial Q}{\partial\psi}\right]_{(\psi_o,0)} = f\left(\psi_o\right). \tag{A.9}$$

- By the Implicit Function Theorem

$$\frac{d\psi_\epsilon}{d\epsilon}\bigg|_{\epsilon=0} = -\left[\frac{\partial Q}{\partial\psi}\right]^{-1}_{(\psi_0,0)} \times \left[\frac{\partial Q}{\partial\epsilon}\right]_{(\psi_o,0)} = -\frac{1}{f\left(\psi_o\right)} \times \frac{dQ\left(\psi_o,\epsilon\right)}{d\epsilon}\bigg|_{\epsilon=0} \tag{A.10}$$

- By the Chain Rule

$$\begin{aligned}
\frac{dQ\left(\psi_o,\epsilon\right)}{d\epsilon}\bigg|_{\epsilon=0} &= \frac{d}{d\epsilon}\left\{\iint \mathbb{1}\left(y\leq\psi_o\right)\frac{f_\epsilon\left(y,1,x\right)f_\epsilon\left(x\right)}{f_\epsilon\left(1,x\right)}dydx\right\}\bigg|_{\epsilon=0} \\
&= \iint \mathbb{1}\left(y\leq\psi_o\right)\left\{\frac{f_\epsilon\left(x\right)}{f_\epsilon\left(1,x\right)}\frac{d}{d\epsilon}f_\epsilon\left(y,1,x\right)\bigg|_{\epsilon=0} - \frac{f_\epsilon\left(y,1,x\right)f_\epsilon\left(x\right)}{f_\epsilon\left(1,x\right)^2}\frac{d}{d\epsilon}f_\epsilon\left(1,x\right)\bigg|_{\epsilon=0}\right. \\
&\quad \left. + \frac{f_\epsilon\left(y,1,x\right)}{f_\epsilon\left(1,x\right)}\frac{d}{d\epsilon}f_\epsilon\left(x\right)\bigg|_{\epsilon=0}\right\} dydx \\
&= \iint \mathbb{1}\left(y\leq\psi_o\right)\frac{f(y,1,x)f(x)}{f(1,x)}\left(\frac{\mathbb{1}_{\tilde{y},\tilde{t},\tilde{x}}(y,1,x)}{f(y,1,x)} - \frac{\mathbb{1}_{\tilde{t},\tilde{x}}(1,x)}{f(1,x)} + \frac{\mathbb{1}_{\tilde{x}}(x)}{f(x)} - 1\right) dydx \\
&= \frac{\mathbb{1}_{\tilde{t}}(1)}{\pi(\tilde{x};\mathcal{P}_o)}\left[\mathbb{1}\left(\tilde{y}\leq\psi_o\right) - G\left(\psi_o \mid 1,\tilde{x};\mathcal{P}_o\right)\right] + G\left(\psi_o \mid 1,\tilde{x};\mathcal{P}_o\right) - \tau
\end{aligned} \tag{A.11}$$

- Hence

$$\varphi\left(\tilde{z};\mathcal{P}_o\right)=\frac{d\psi\left(\mathcal{P}_\epsilon\right)}{d\epsilon}\bigg|_{\epsilon=0}=-\frac{1}{f\left(\psi_o\right)}\left\{\frac{\mathbb{1}_{\tilde{t}}(1)}{\pi(\tilde{x};\mathcal{P}_o)}\left[\mathbb{1}\left(\tilde{y}\leq\psi_o\right)-G\left(\psi_o\mid 1,\tilde{x};\mathcal{P}_o\right)\right]+G\left(\psi_o\mid 1,\tilde{x};\mathcal{P}_o\right)-\tau\right\}. \tag{A.12}$$

A.2 Efficient Influence Function-Based Estimator

Let $\mathcal{P}_n$ denote the empirical distribution from a sample of size n. Denote $h_o(X,\psi)=G(\psi\mid 1,X;\mathcal{P}_o)$; and $\pi_o(X)=\pi(X;\mathcal{P}_o)$. Then, $h_o(X,\psi)$ and $\pi_o(X)$ are function-valued nuisance parameters of the estimation problem for the τ-quantile of treated potential outcome, which is our target parameter denoted by $\psi_o\coloneqq\psi(\mathcal{P}_o)$.

The moment condition associated with the efficient influence function in (A.5) to identify the target parameter value ψ_o is

$$\mathbb{E}\left[\varphi(Z;h_o(X,\psi_o),\pi_o(X)),\psi_o\right]=0, \tag{A.13}$$

in which the moment has zero derivative with respect to nuisances at ψ_o, h_o and π_o. Intuitively, this moment condition satisfies Neyman orthogonality, that is the first-order insensitivity of target parameter value to local perturbations of the values of nuisance parameters. This property is desirable because it helps ensure that the estimation of the parameter of interest remains robust even when there are small errors or uncertainties in the estimation of nuisance parameters. When regularization methods are needed to handle high-dimensional covariates or nonlinearities when estimating nuisance parameters, the use of Neyman orthogonal moment conditions help eliminate the first-order biases stemming from these plugging-in estimators (see e.g., Belloni et al. 2017; Chernozhukov et al. 2018; Kallus, Mao, and Uehara 2024).

Therefore, the efficient influence function-based estimator for $\psi(\mathcal{P}_o)$ is defined as a solution to the estimating equation

$$\begin{aligned}\mathbb{E}_{\mathcal{P}_n}\left[\varphi(Z;\hat{h}(X,\psi),\hat{\pi}(X)),\psi\right]&=0\\ \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}\{T_i=1\}}{\hat{\pi}(X_i)}\left[\mathbb{1}\{Y_i\leq\psi\}-\hat{h}(X_i,\psi)\right]+\hat{h}(X_i,\psi)-\tau&=0.\end{aligned} \tag{A.14}$$

Denote by $\widehat{\psi}^{dr}$ the resulting estimator from (A.14). As shown in the following section, $\widehat{\psi}^{dr}$ is a doubly robust estimator, which is consistent provided that either one of nuisance estimators – $\hat{h}$ or $\hat{\pi}$ – is consistent, but not necessarily both.

A.3 (Frequentist) Asymptotic Properties

Lemma 1. (*Double Robustness of Efficient Influence Function*)
Let $\eta^*=(h^*,\pi^*)$ with either $h^*=h_o$ or $\pi^*=\pi_o$. Then $\mathbb{E}_{\mathcal{P}_o}\left[\varphi(\eta^*,\psi_o)=0\right]$.

Sketch of Proof for Lemma 1.

- By the law of iterated expectation

$$\begin{aligned}\mathbb{E}_{\mathcal{P}_o}\left[\varphi(\eta^*,\psi)\right] &= -\frac{1}{f(\psi)}\mathbb{E}_{\mathcal{P}_o}\left[\frac{\pi_o}{\pi^*}\left(h_{o,\psi}-h^*_\psi\right)+h^*_\psi-\tau\right] \\ &= -\frac{1}{f(\psi)}\mathbb{E}_{\mathcal{P}_o}\left[\left(\frac{\pi_o}{\pi^*}-1\right)\left(h_{o,\psi}-h^*_\psi\right)+h_{o,\psi}-\tau\right].\end{aligned}$$

- When either $\pi^* = \pi_o$ or $h^* = h_o$, substituting ψ with ψ_o leads to $\mathbb{E}_{\mathcal{P}_o}\left[\varphi(\eta^*,\psi)\right] = 0$, and the lemma follows.

Theorem 2. (*Consistency of the Point Estimator*)
Under Identifying Assumptions 3.3-3.5 and additional Regularity Conditions, $\widehat{\psi}^{dr}$ is consistent if either nuisance estimator $\hat{h}$ or $\hat{\pi}$ is consistent.

Sketch of Proof for Theorem 2.

- By construction of $\widehat{\psi}^{dr}$ in (A.14) we have $\mathbb{E}_{\mathcal{P}_n}\varphi(\hat{\eta},\widehat{\psi}^{dr}) = 0$, where $\hat{\eta} = (\hat{h}(X,\widehat{\psi}^{dr}),\hat{\pi})$.
- By Lemma 1 we have $\mathbb{E}_{\mathcal{P}_o}\left[\varphi(\eta^*,\psi_o)\right] = 0$.
- An application of Theorem 5.9 of Van der Vaart (2000) yields $\widehat{\psi}^{dr} = \psi_o + o_{\mathcal{P}}(1)$, thereby $\widehat{\psi}^{dr}$ is consistent. This completes the proof.

A.4 (Bayesian) Posterior Contraction Rates

Let $\mathcal{P}_n$ denote the posterior distribution from a sample of size n. Let $\pi = (\pi_1,\ldots,\pi_n)$, $h = (h_1,\ldots,h_n)$ and let π_o and h_o denote their unknown, true values. Assume that the posterior distribution of the propensity score and the conditional distribution contract at rates ϵ_{nt} and ϵ_{ny}, respectively. There exist two sequences of numbers $\epsilon_{nt} \to 0$ and $\epsilon_{ny} \to 0$, and constants $M_t > 0$ and $M_y > 0$ such that

$$\sup_{\mathcal{P}_o} E_{\mathcal{P}_o}\mathcal{P}_n\left(\frac{1}{\sqrt{n}}\|\pi-\pi_o\| > M_t\epsilon_{nt} \mid Z\right) \to 0,$$

$$\sup_{\mathcal{P}_o} E_{\mathcal{P}_o}\mathcal{P}_n\left(\frac{1}{\sqrt{n}}\|h-h_o\| > M_y\epsilon_{ny} \mid Z\right) \to 0.$$

Theorem 3. (*Consistency of the Point Estimator*)
Under Identifying Assumptions 3.3-3.5 and additional Regularity Conditions, if the contraction assumptions hold, then

$$\sup_{\mathcal{P}_o} E_{\mathcal{P}_o}\mathcal{P}_n\left(\frac{1}{\sqrt{n}}\|\psi-\psi_o\| > M\epsilon_n \mid Z\right) \to 0,$$

with $\epsilon_n = \max(n^{-1/2}, \epsilon_{nt}\epsilon_{ny})$.

A.5 Regularity Conditions

1. The cumulative distribution function F has compact support $[a,b] \subset \mathbb{R}$ and is continuously differentiable on its support with strictly positive derivative f.

2. The class function $\{\varphi(\eta, \psi) : |\psi - \psi_o| < \delta, \|h_\psi - h^*_\psi\| < \delta, \|\pi_\psi - \pi^*_\psi\| < \delta\}$ is Donsker for some $\delta > 0$ and such that $\mathcal{P}_o\{\varphi(\eta, \psi) - \varphi(\eta^*, \psi_o)\}^2 \to 0$ as $(\eta, \psi) \to (\eta^*, \psi_o)$.

B Bayesian Additive Regression Tree (BART)

B.1 BART Model Specifications

BART is a nonparametric modeling technique that translates decision tree-based ensemble methods to a Bayesian framework. H. A. Chipman, George, and McCulloch (2010) present a comprehensive overview of the method. In essence, BART is a *sum-of-trees* model with prior distributions are placed over the parameters including *tree depth*, *splitting variables*, *splitting values*, and *terminal node* estimates.

Consider the regression problem that predicts a continuous Y_i using a p-dimensional vector of predictors $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^\top (i = 1, \ldots, N)$, BART model can be expressed as

$$Y_i = f_{\text{BART}}(\mathbf{X}_i) + \epsilon_i, \quad \epsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad f_{\text{BART}}(\mathbf{X}_i) = \sum_{m=1}^{M} f_{\text{tree}}(\mathbf{X}_i; \Gamma_m, \mu_m), \tag{B.1}$$

where $f_{\text{tree}}(\mathbf{X}_i; \Gamma_m, \mu_m)$ is a Bayesian single regression tree; Γ_m is a tree structure that consists of a set of splitting rules and a set of terminal nodes; and $\mu = (\mu_{m,1}, \ldots, \mu_{m,b_m})$ is a vector of parameters associated with b_m terminal nodes of Γ_m, such that $f_{\text{tree}}(\mathbf{X}_i; \Gamma_m, \mu_m) = \mu_{m,l}$ if $\mathbf{X}_i$ is corresponding to the l^{th} terminal node of Γ_m.

The prior of BART is specified for three components:

1. The ensemble structure $\{\Gamma_m\}_{m=1}^M$
   Independent regularization prior is placed on Γ_m. It consists of a Bernoulli distribution with probability
   $$\Pr(\text{split} \mid d) = \alpha(1 + d)^{-\beta}, \quad \alpha \in (0, 1), \quad \beta \in (0, \infty), \tag{B.2}$$
   for splitting a node at tree depth d ($d = \{0, 1, \ldots\}$ into two child nodes and two discrete uniform distributions for selecting a split variable and a split value given the selected split variable. This regularization prior helps prevent individuals from becoming too influential, thereby enhancing the overall fit and mitigating the risk of overfitting.

2. The parameters $\{\mu_m\}_{m=1}^M$ associated with the terminal nodes given $\{\Gamma_m\}_{m=1}^M$
   $$\mu_{m,l} \overset{iid}{\sim} \mathcal{N}(0, v) \tag{B.3}$$

3. The error variance σ^2 that is independent with the former two
   $$\sigma^2 \sim \text{Inv-Gamma}(r, s) \tag{B.4}$$

The process of sampling the posterior distribution is carried out using a Metropolis-within-Gibbs MCMC sampler, which can also be regarded as a special case of (generalized) Bayesian backfitting algorithm (Hastie and Tibshirani 2000), to update each tree iteratively. Estimated outcome is achieved by averaging the posterior samples of $f_{\text{BART}}(\mathbf{X}_i)$ after a burn-in period.

For binary outcome, the continuous BART model above has been extended to probit BART and

logit BART, which are specified as follows

$$\mathbb{P}(Y_i = 1 \mid \mathbf{X}_i) = H[f_{\text{BART}}(\mathbf{X}_i)], \tag{B.5}$$

where $f_{\text{BART}}(\mathbf{X}_i)$ is the sum-of-trees function in (B.1) and H is the link function with the probit link for probit BART and the logit link for logit BART. Both of the models maintain the same prior assigned to the ensemble structure and the parameters of the terminal nodes, i.e. $\{\Gamma_m, \mu_m\}_{m=1}^{M}$, but σ^2 is fixed for for the sake of identifiability.

Probit BART employs data augmentation of Albert and Chib (1993) to adapt the Bayesian backfitting sampler used in continuous BART. This involves introducing a latent variable Y_i^* such that $Y_i = \mathbb{1}\{Y_i^* > 0\}$ for each response variable Y_i. At each iteration of the MCMC algorithm, Y_i^* is imputed by sampling from the full conditional distribution of Y_i^* given Y_i and other parameters, which is essentially a truncated normal distribution. The imputed Y_i^*'s are then modelled using the continuous BART model with σ^2 set to 1, enabling the completion of the MCMC algorithm by performing the Bayesian backfitting algorithm of the continuous BART model on the imputed Y_i^*'s.

Logit BART also introduces latent variable Y_i^* that is instead assumed to follow a logistic distribution, which has a heavier tail than a normal distribution, thus improving estimation for extreme instances of $\mathbb{P}(Y_i = 1 \mid \mathbf{X_i})$. These latent Y_i^*'s are sampled using the method described by Gramacy and Polson (2012). Conditional on the imputed Y_i^*'s, the continuous BART model with given heteroskedastic variance σ_i^2's are fitted on Y_i^*'s, where σ_i^2's are obtained through the technique outlined by Robert (1995).

B.2 On Implementation

Computational details and implementation using **BART** R package can be found in Sparapani, Spanbauer, and McCulloch (2021). The actual sampling and computation are carried out in C++ code to maximize computational efficiency. Both options for Probit BART and Logit BART are available in this package, which can be utilised directly to fit the treatment assignment model in our proposed `BADR` framework.

C Bayesian Shrinkage Priors

C.1 Hierarchical Bayes for Linear Regression

Consider the linear regression model

$$Y_i = \mathbf{X}_i\beta + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \tag{C.1}$$

Assuming that interest lies in learning about the regression coefficients β, then a simple hierarchical specification takes the form

$$\begin{aligned} Y_i \mid \beta, \sigma^2 &\sim N(\mathbf{X}_i\beta, \sigma^2), i = 1 \ldots, n \\ \beta_j \mid \tau^2, \sigma^2 &\sim N(0, \sigma^2\tau^2), i = 1 \ldots, p \\ \tau^2 &\sim F(a, b) \\ \sigma^2 &\sim \frac{1}{\sigma^2} \end{aligned} \tag{C.2}$$

where $F(a, b)$ denotes some distribution function with hyper-parameters a, b. Due to the fact that choice of τ^2 is so crucial for the posterior outcome of β_j, the idea behind this hierarchical specification is to treat the hyper-parameter τ^2 as a random variable and learn about it via Bayes Theorem.

C.2 Bayesian Shrinkage Priors

The major goal of shrinkage priors is to shrink small coefficients to zero while maintaining true large coefficients, especially in high-dimensional settings. The possible variation in shrinkage amounts among those priors depends on their specific designs. In particular, the sharper the peak is around zero, the stronger shrinkage for small coefficients. Also, the heavier the tail, the lighter the shrinkage for large coefficients.

Bayesian Lasso

The Bayesian counterpart of the Lasso penalty is Laplace prior, which was first proposed by Park and Casella (2008). The Bayesian Lasso can be obtained as a scale mixture of a Normal density with an Exponential density as below:

$$\begin{aligned} \beta_j \mid \tau_j^2, \sigma^2 &\sim N\left(0, \sigma^2\tau_j^2\right) \\ \tau_j^2 \mid \lambda^2 &\sim \text{Exp}\left(\frac{\lambda^2}{2}\right), \text{ for } j = 1, \ldots, p \\ \lambda &\sim \text{half-Cauchy}(0, 1) \\ \sigma^2 &\sim \frac{1}{\sigma^2} \end{aligned} \tag{C.3}$$

Integrating τ_j^2 out leads to Double-exponential[9] or Laplace priors on the regression coefficients, i.e.,

$$\beta_j \mid \lambda, \sigma \sim \text{Double-exponential}\left(0, \frac{\sigma}{\lambda}\right), \text{ for } j = 1, \ldots, p \tag{C.4}$$

Although this version of Bayesian Lasso is the most popular form in literature so far; there are also some alternative formulations suggested by Hans (2009), Mallick and Yi (2014) and Alhamzawi and Taha Mohammad Ali (2020).

In addition to the overall shrinkage parameter λ, the Lasso prior has an additional predictor-specific shrinkage parameter τ_j. Therefore, the Lasso prior is more flexible than the Ridge prior, which only relies on the overall shrinkage parameter.

Horseshoe prior

A novel shrinkage prior in the Bayesian literature is the horseshoe prior Carvalho, Polson, and Scott (2010)[10]. This prior is particularly attractive for sparse signal recovery.

$$\begin{aligned} \beta_j \mid \tau_j^2 &\sim N\left(0, \tau_j^2\right) \\ \tau_j \mid \lambda &\sim \text{half-Cauchy}(0, \lambda), \text{ for } j = 1, \ldots, p \\ \lambda \mid \sigma &\sim \text{half-Cauchy}(0, \sigma) \end{aligned} \tag{C.5}$$

The half-Cauchy prior can be written as a mixture of inverse Gamma densities[11] (Makalic and Schmidt 2015), so that the horseshoe prior in Equation (C.5) can be equivalently specified as:

$$\begin{aligned} \beta_j \mid \tau_j^2 &\sim N\left(0, \tau_j^2\right) \\ \tau_j^2 \mid \omega &\sim \text{Inv-Gamma}\left(\frac{1}{2}, \frac{1}{\omega}\right) \\ \omega \mid \lambda^2 &\sim \text{Inv-Gamma}\left(\frac{1}{2}, \frac{1}{\lambda^2}\right) \\ \lambda^2 \mid \gamma &\sim \text{Inv-Gamma}\left(\frac{1}{2}, \frac{1}{\gamma}\right) \\ \gamma \mid \sigma^2 &\sim \text{Inv-Gamma}\left(\frac{1}{2}, \frac{1}{\sigma^2}\right) \end{aligned} \tag{C.6}$$

An expression for the marginal prior of the regression coefficients β_j is not analytically tractable, but a tight lower bound Carvalho, Polson, and Scott (2010) can be used instead.

[9] Mathematical representation:

$$\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 s_j}} e^{\left(-\frac{\beta_j^2}{2\sigma^2 s_j}\right)} \frac{\lambda^2}{2} e^{-\frac{\lambda}{2s_j}} ds_j = \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}}$$

[10] Note that Carvalho, Polson, and Scott (2010) explicitly include the half-Cauchy prior for λ in their specification, thereby implying a full Bayes approach. This formulation results in a horseshoe prior that is automatically scaled by the error standard deviation σ.

[11] If $x^2 \mid z \sim$ inv-Gamma $(1/2, 1/z)$ and $z \sim$ inv-Gamma $\left(1/2, 1/\alpha^2\right)$ then $x \sim C^+(0, \alpha)$

$$-\log p\left(\beta_i \mid \lambda\right) \geq -\log\log\left(1+\frac{2\lambda^2}{\beta_j^2}\right) \tag{C.7}$$

The key features for the appealing performance of horseshoe prior are its Cauchy-like tails and an asymptote at origin (unique advantage), which make horseshoe adaptive to sparsity and robust to large signals so outperform other shrinkage priors we have discussed.

In the search for intuitive reasons, we consider a common framework of shrinkage rules. Define $\kappa_j = 1/(1+\tau_j^2)$, then κ_j is a random shrinkage coefficient in $[0,1]$. Under a multivariate normal scale mixture prior (i.e. the general form of all shrinkage priors we are discussing), the posterior mean can be written as a linear function of the observation:

$$\mathbb{E}\left[\beta_j \mid Y_j\right] = \{1-\mathbb{E}\left[\kappa_i \mid Y_j\right]\}Y_j \tag{C.8}$$

Hence, $\mathbb{E}\left[\kappa_i \mid Y_j\right]$ implies the amount of weight that the posterior mean for β_j places on 0 once the data have been observed. A shrinkage coefficient κ_j that is close to zero leads to virtually no shrinkage, thus describes signals. A shrinkage coefficient κ_j that is close to one leads to nearly-total shrinkage, thus describes noises. Intuitively speaking, the behavior of a priori $p(\kappa_j)$ near $\kappa_j = 1$ will control the robustness of signal at tail, while near $\kappa_j = 0$ will control the shrinkage of noise toward 0. Because of difference choice of $p(\tau_j)$, each type of shrinkage prior has distinct $p(\kappa_j)$ reflecting its attempt to separate signal and noise. For horseshoe prior, the attempt is even implied in its name, which arises from the fact that for fixed values $\lambda = \sigma = 1$, $p(\kappa_j)$ is similar to a horseshoe-shaped Beta $(1/2, 1/2)$. This prior is symmetric and unbounded at both 0 and 1; thereby, small coefficients (noises) are heavily shrunken towards zero while substantial coefficients (signals) remain large. None of these common shrinkage priors above shares this characteristic. For instance, the Laplace prior, where $p(\kappa_j)$ is bounded at both 0 and 1, tends to over-shrink strong signals yet under-shrink noises. Carvalho, Polson, and Scott (2009), Carvalho, Polson, and Scott (2010) provide more explanation for other priors.

In fact, unlike local shrinkage priors above, the horseshoe prior is a member of a wider class of *global-local shrinkage priors* (Polson and Scott 2010; Bhadra et al. 2019) because it enables a clear separation between global and local shrinkage effects. Put another way, this class of priors adapt to sparsity by a global shrinkage parameter and recover signals by a local shrinkage parameter.

D Bayesian Quantile Regression (BQR)

D.1 Bayesian Quantile Regression

Consider the linear quantile regression model (Koenker and Bassett 1978) at a given quantile level $\tau \in (0,1)$

$$\mathcal{Q}_\tau\left(Y \mid \mathbf{X}\right) = \mathbf{X}\beta_{(\tau)}, \tag{D.1}$$

The quantile specific coefficient $\beta_{(\tau)}$ can be consistently estimated by

$$\hat{\beta}_{(\tau)} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau\left(Y_i - \mathbf{X}_i\beta\right), \tag{D.2}$$

where $\rho_\tau(u) = u\left(\tau - \mathbb{1}\{u<0\}\right)$ is the quantile loss function. As this functional form is an asymmetric L_1 loss function proportional to the negative log density of the asymmetric Laplace distribution ($\mathcal{ALD}$), the connection allows researcher to recast the quantile regression as a maximum likelihood problem of the linear model $Y_i = \mathbf{X}_i\beta_{(\tau)} + \epsilon_{i,(\tau)}$ where $\epsilon_{i,(\tau)} \sim \mathcal{ALD}\left(\tau, 0, \sigma_{(\tau)}\right)$[12]. The working likelihood is of the form

$$f\left(Y \mid \mathbf{X}, \beta_{(\tau)}, \sigma_{(\tau)}, \tau\right) = \frac{\tau^n(1-\tau)^n}{\sigma^n_{(\tau)}} \exp\left\{-\sum_{i=1}^{n} \frac{\rho_\tau\left(Y_i - \mathbf{X}_i\beta\right)}{\sigma_{(\tau)}}\right\}. \tag{D.3}$$

The asymmetric Laplace distribution is known to be expressible as a scale mixture of normals (Kotz, Kozubowski, and Podgorski 2012), we thus can rewrite $\epsilon_{i,(\tau)}$ as follows

$$\epsilon_{i,(\tau)} = \theta_{(\tau)} z_{i,(\tau)} + \kappa_{(\tau)}\sqrt{\sigma_{(\tau)} z_{i,(\tau)}}\, u_i, \quad \text{with } \theta_{(\tau)} = \frac{1-2\tau}{\tau(1-\tau)} \text{ and } \kappa^2_{(\tau)} = \frac{2}{\tau(1-\tau)},$$

where $z_{i,(\tau)} = \sigma_{(\tau)}\zeta_{i,(\tau)}$ with $\zeta_{i,(\tau)} \sim \text{Exp}(1)$, and $u_i \sim N(0,1)$.

As a result, the Bayesian quantile regression model has the following representation

$$Y_i = \mathbf{X}_i\beta_{(\tau)} + \theta_{(\tau)} z_{i,(\tau)} + \kappa_{(\tau)}\sqrt{\sigma_{(\tau)} z_{i,(\tau)}}\, u_i, \quad \text{for } i = 1, \ldots, n. \tag{D.4}$$

This leads to the following likelihood function:

$$f\left(Y \mid \mathbf{X}, \beta_{(\tau)}, \sigma_{(\tau)}, \mathbf{z}_{(\tau)}, \tau\right) \propto \exp\left\{-\sum_{i=1}^{n} \frac{\left(Y_i - \mathbf{X}_i\beta_{(\tau)} - \theta_{(\tau)} z_{i,(\tau)}\right)^2}{2\kappa^2_{(\tau)}\sigma_{(\tau)} z_{i,(\tau)}}\right\} \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma_{(\tau)} z_{i,(\tau)}}}. \tag{D.5}$$

[12] $\epsilon_{i,(\tau)}$ follows asymmetric Laplace distribution with density

$$f_{\mathcal{ALD}}\left(\epsilon_{(\tau)}\right) = \frac{\tau(1-\tau)}{\sigma_{(\tau)}} \exp\left\{-\rho_\tau\left(\epsilon_{(\tau)}\right)/\sigma_{(\tau)}\right\}.$$

We assume the priors as below (Kozumi and Kobayashi 2011)

$$\beta_{(\tau)} \sim N\left(0, \mathbf{\Sigma}_{0,(\tau)}\right), \tag{D.6}$$

$$z_{i,(\tau)} \sim \text{Exp}\left(\sigma_{(\tau)}\right) \propto \sigma_{(\tau)}^{-1} \exp\left\{-\sigma_{(\tau)}^{-1} z_{i,(\tau)}\right\} \forall i = 1, \ldots, n, \tag{D.7}$$

$$\sigma_{(\tau)} \sim \text{Inv-Gamma}\left(r_{0,(\tau)}, s_{0,(\tau)}\right) \propto \left(\sigma_{(\tau)}^{-1}\right)^{r_{0,(\tau)}+1} \exp\left\{-s_{0,(\tau)}\sigma_{(\tau)}^{-1}\right\}, \tag{D.8}$$

where for simplicity, $\mathbf{\Sigma}_{0,(\tau)} = \mathbf{D}_{\lambda,(\tau)} = \lambda \times \mathbf{I}_p$ where λ is fixed and known for all τ.

The conditional posteriors are of the form

$$\beta_{(\tau)} \mid \bullet \sim N_p\left(\mu_{\beta,(\tau)}, \mathbf{\Sigma}_{\beta,(\tau)}\right), \tag{D.9}$$

$$z_{i,(\tau)} \mid \bullet \sim \text{GIG}\left(\frac{1}{2}, a_{i,(\tau)}, b_{i,(\tau)}\right) \propto z_{i,(\tau)}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(a_{i,(\tau)} z_{i,(\tau)} + b_{i,(\tau)} z_{i,(\tau)}^{-1}\right)\right\}, \quad \forall i = 1, \ldots, n, \tag{D.10}$$

$$\sigma_{(\tau)} \mid \bullet \sim \text{Inv-Gamma}\left(r_{\sigma,(\tau)}, s_{\sigma,(\tau)}\right) \propto \left(\sigma_{(\tau)}^{-1}\right)^{r_{\sigma,(\tau)}+1} \exp\left\{-s_{\sigma,(\tau)}\sigma_{(\tau)}^{-1}\right\}, \tag{D.11}$$

where

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

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

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

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

D.2 Bayesian Quantile Regression with the Adaptive Lasso

Bayesian Quantile Regression with the Adaptive Lasso is a Bayesian hierarchical model given by

$$Y_i = \mathbf{X}_i\beta_{(\tau)} + \theta_{(\tau)} z_{i,(\tau)} + \kappa_{(\tau)}\sqrt{\sigma_{(\tau)} z_{i,(\tau)}}\, u_i, \quad \text{for } i = 1, \ldots, n, \tag{D.12}$$

$$u_i \sim N(0,1), \tag{D.13}$$

$$z_{i,(\tau)} \sim \text{Exp}\left(\sigma_{(\tau)}\right) \propto \sigma_{(\tau)}^{-1} \exp\left\{-\sigma_{(\tau)}^{-1} z_{i,(\tau)}\right\} \forall i = 1, \ldots, n \tag{D.14}$$

$$\beta_{j,(\tau)}, v_{j,(\tau)} \mid \sigma_{(\tau)}, \lambda_{j,(\tau)}^2 \sim \frac{1}{\sqrt{2\pi v_{j,(\tau)}}} \exp\left\{-\frac{\beta_{j,(\tau)}^2}{2v_{j,(\tau)}}\right\} \frac{\sigma_{(\tau)}^{-1}}{2\lambda_{j,(\tau)}^2} \exp\left\{\frac{-\sigma_{(\tau)}^{-1}}{2\lambda_{j,(\tau)}^2} v_{j,(\tau)}\right\}, \tag{D.15}$$

$$\lambda_{j,(\tau)}^2 \sim \text{Inv-Gamma}\ \left(c_{0,(\tau)}, d_{0,(\tau)}\right) \propto \left(\frac{1}{\lambda_{j,(\tau)}^2}\right)^{c_{0,(\tau)}+1} \exp\left\{-\frac{d_{0,(\tau)}}{\lambda_{j,(\tau)}^2}\right\}, \tag{D.16}$$

$$\sigma_{(\tau)} \sim \text{Inv-Gamma}\ \left(r_{0,(\tau)}, s_{0,(\tau)}\right) \propto \left(\sigma_{(\tau)}^{-1}\right)^{r_{0,(\tau)}+1} \exp\left\{-s_{0,(\tau)}\sigma_{(\tau)}^{-1}\right\} \tag{D.17}$$

The conditional posteriors (Alhamzawi, Yu, and Benoit 2012) are of the form

$$z_{i,(\tau)} \mid \bullet \sim \text{GIG}\left(\frac{1}{2}, a_{i,(\tau)}, b_{i,(\tau)}\right) \propto z_{i,(\tau)}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(a_{i,(\tau)} z_{i,(\tau)} + b_{i,(\tau)} z_{i,(\tau)}^{-1}\right)\right\}, \quad \forall i = 1, \ldots, n, \tag{D.18}$$

$$\beta_{j,(\tau)} \mid \bullet \sim N\left(\mu_{\beta_j,(\tau)}, \Sigma_{\beta_j,(\tau)}\right), \quad \forall j = 1, \ldots, p, \tag{D.19}$$

$$v_{j,(\tau)} \mid \bullet \sim \text{GIG}\left(\frac{1}{2}, \frac{\sigma_{(\tau)}^{-1}}{\lambda_{j,(\tau)}^2}, \beta_{j,(\tau)}^2\right) \propto v_{j,(\tau)}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{\sigma_{(\tau)}^{-1}}{\lambda_{j,(\tau)}^2} v_{j,(\tau)} + \beta_{j,(\tau)}^2 v_{j,(\tau)}^{-1}\right)\right\}, \tag{D.20}$$

$$\sigma_{(\tau)} \mid \bullet \sim \text{Inv-Gamma}\left(r_{\sigma,(\tau)}, s_{\sigma,(\tau)}\right) \propto \left(\sigma_{(\tau)}^{-1}\right)^{r_{\sigma,(\tau)}+1} \exp\left\{-s_{\sigma,(\tau)}\sigma_{(\tau)}^{-1}\right\}, \tag{D.21}$$

$$\lambda_{j,(\tau)}^2 \mid \bullet \sim \text{Inv-Gamma}\left(c_{0,(\tau)} + 1, d_{0,(\tau)} + \sigma_{(\tau)}^{-1} v_{j,(\tau)}/2\right), \tag{D.22}$$

where

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

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

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

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

D.3 On Implementation

The **bayesQR** R package (Benoit and Van den Poel 2017) provides the implementation of efficient Gibbs sampler for both Bayesian Quantile Regression and Bayesian Quantile Regression with the Adaptive Lasso outlined above. In addition, the core procedure is programmed in Fortran to speed up the computational time. Therefore, this package can be utilised straightforward to estimate multiple conditional quantiles, which then be used to approximate the condistional distributions of potential

outcomes in our proposed `BADR` framework.

Alternatively, Variational Inference algorithm could be used for Bayesian quantile regression with/without the regularisation (Lim et al. 2020; Guo 2019), which helps improving the speed of Gibbs sampling while maintaining a comparable accuracy in terms of MSE.

E Implementation of Other Estimators in Simulation Study

E.1 Existing Approaches

Firpo's Inverse Probability Weighted method (`FIPW`)

Firpo's Inverse Probability Weighted (`FIPW`) method (Firpo 2007) involves a two-step estimator. First, the propensity score is estimated nonparametrically as a logistic power series whose degree increases with sample size. In the second step, the quantiles are estimated by minimising an inverse probability weighted check loss function. These weights reflect the fact that the distribution of the covariates differs in the two groups.

Algorithm 3: FIPW Approach to Estimate QTEs

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

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

1 *Step(1).* Estimate propensity score $\hat{\pi}(x) = \text{expit}\left(H_K(x)'\hat{p}_K\right)$ where

$$\hat{p}_K \coloneqq \underset{p \in \mathbb{R}^K}{\text{argmax}} \frac{1}{N} \sum_{i=1}^{N} \{T_i \cdot \log\left(\text{expit}\left(H_K(X_i)'p\right)\right) + (1 - T_i) \cdot \log\left(1 - \text{expit}\left(H_K(X_i)'p\right)\right)\}$$

2 *Step(2).* Derive $\hat{q}_1(\tau)$ and $\hat{q}_0(\tau)$ as the solution to

$$\hat{q}_1(\tau) \coloneqq \underset{q}{\text{argmin}} \sum_{i=1}^{N} \frac{T_i}{N \cdot \hat{\pi}(\mathbf{X}_i)} \rho_\tau(Y_i - q) \text{ and } \hat{q}_0(\tau) \coloneqq \underset{q}{\text{argmin}} \sum_{i=1}^{N} \frac{1 - T_i}{N \cdot (1 - \hat{\pi}(\mathbf{X}_i))} \rho_\tau(Y_i - q)$$

where $\rho_\tau(a) = a \cdot (\tau - \mathbb{1}\{a \leq 0\})$ is the check function.

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

Targeted Maximum Likelihood Estimation method (`TMLE`)

The estimation procedure of Targeted Maximum Likelihood Estimation (`TMLE`) method (Díaz 2017) includes three steps. First, the propensity score and the conditional distribution of the outcome are estimated; second, the quantiles are estimated based on the current cdf of the outcome; and third, the conditional distribution of the outcome is updated based on an exponential submodel for the density of the outcome. The last two steps are iterated until convergence.

Algorithm 4: TMLE Approach to Estimate QTEs

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

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

Step (1). Initialize: Obtain initial estimates $\hat{\pi}$ and $\hat{G}$ of π_0 and G_0.

Step (2). Compute $\hat{q}_1(\tau)$: For the current estimate $\hat{G}$, compute

$$\hat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} \hat{G}(y \mid 1, X_i) \text{ and } \hat{q}_1(\tau) = \inf\{y : \hat{F}(y) \geq \tau\}$$

Step (3). Update $\hat{G}$: Let $\hat{g}$ denote the density associated to $\hat{G}$.

(a) Consider the exponential submodel:

$$\hat{g}_\epsilon(y \mid 0, x) = c(\epsilon, \hat{g}) \exp\{\epsilon \hat{H}_{\hat{\eta},\hat{\theta}(z)}\} \hat{g}(y \mid 0, x)$$

where $c(\epsilon, \hat{g})$ is a normalizing constant and

$$\hat{H}_{\hat{\eta},\hat{\theta}(z)} \coloneqq \frac{1}{\hat{\pi}(X)}\{\mathbb{1}_{(-\infty,\hat{\theta}]}(y) - \hat{G}(\hat{\theta} \mid 0, x)\}$$

is the score of the model.

(b) Estimate ϵ

$$\hat{\epsilon} = \text{argmax} \sum_{i=1}^{n} (1 - T_i) \log \hat{g}_\epsilon(Y_i \mid 0, X_i)$$

(c) Calculate $\hat{g}_\epsilon(Y_i \mid 0, X_i)$ as the updated estimator of g

Step (4). Iterate: Let $\hat{g} = \hat{g}_\epsilon$ and iterate steps 2-3 until convergence.

Derive $\hat{q}_0(\tau)$ similarly.

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

Localized Debiased Machine Learning method (`LDML`)

The Localized Debiased Machine Learning (`LDML`) method (Kallus, Mao, and Uehara 2024) is also motivated by the efficient estimation equation, but Inverse Probability Weighted (IPW) estimates are used as rough initial guessed values for $\hat{q}_1$ and $\hat{q}_0$. Then, these values are used to localize the estimation of conditional distributions $\hat{G}(y \mid 0, \mathbf{X})$ and $\hat{G}(y \mid 1, \mathbf{X})$, respectively. This approach aims to refine the IPW estimate while obviating the need to estimate a continuum of continuum nuisances. The main algorithm includes two parts: *three-way-cross-fold nuisance estimation* and *solving the estimating equation*.

Algorithm 5: LDML Approach to Estimate QTEs

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

Part(1). Three-way-cross-fold nuisance estimation

Fix integers $K \geq 3$ and $1 \leq K' \leq K-2$.

Randomly permute the data indices and let $\mathcal{D}_k$ be a random even K-fold split of the data.

for $k = 1, \ldots, K$ **do**

(a) Set $\mathcal{H}_{k,1} = \{1, \ldots, K' + \mathbb{1}[k \leq K']\} \setminus \{k\}$ and $\mathcal{H}_{k,2} = \{K' + \mathbb{1}[k \leq K'] + 1, \ldots, K\} \setminus \{k\}$.

(b) Use only the data in $\mathcal{D}_k^{C,1} = \{\mathbf{X}_i : i \in \bigcup_{k' \in \mathcal{H}_{k,1}}\}$ to construct $\hat{q}_{1,init}^{(k)}$.

(c) Use only the data in $\mathcal{D}_k^{C,2} = \{\mathbf{X}_i : i \in \bigcup_{k' \in \mathcal{H}_{k,2}}\}$ to construct $\hat{G}_1^{(k)}(\cdot, \hat{q}_{1,init}^{(k)})$.

(d) Use $\mathcal{D}_k^{C,1} \cap \mathcal{D}_k^{C,2}$ to construct estimator $\hat{\pi}^{(k)}$.

end

Part(2). Solving the average of the estimate in each fold to obtain $\hat{q}_1(\tau)$

$$\frac{1}{N} \sum_{k=1}^{K} \sum_{i \in \mathcal{D}_k} \psi\left(\mathbf{X}_i; q, \hat{G}_1^{(k)}(\mathbf{X}_i, \hat{q}_{1,init}^{(k)}), \hat{\pi}^{(k)}(\mathbf{X}_i)\right) = 0$$

Derive $\hat{q}_0(\tau)$ using two-part procedure similarly.

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

Bayesian nonparametric method (`BNP`)

Bayesian nonparametric (`BNP`) method (Xu, Daniels, and Winterstein 2018) is a fully Bayesian nonparametric (BNP) approach to estimate QTEs. The estimation procedure includes three steps. First, the propensity score is estimated using a logit BART model. Then, the conditional distribution of the potential outcome given a BART posterior sample of the PS in each treatment group is modelled separately using a Dirichlet process mixture (DPM) of multivariate normals model. Finally, marginalizing the estimated conditional distribution over the population distribution of the confounders using Bayesian bootstrap (Rubin 1981). Details of implementation using **BNPqte** `R` package can be found in Luo and Daniels (2021).

Algorithm 6: BNP Approach to Estimate QTEs

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

Fit a binary BART model on $\{T_i, X_i\}_{i=1}^n$ and obtain B posterior samples $\{H^{-1}(\pi_i^{\{b\}})\}_{i,b=1}^{n,B}$.

Create a set of grid points of Y values: $(q_1, \ldots, q_S)$.

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

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

    Fit a DPM of bivariate normals on $\{Y_i, H^{-1}(\pi_i^{\{b\}})\}_{i:T_i=t}$.

    Use Blocked Gibbs sampler to obtain L posterior samples.

    Calculate $\{F^{\{bl\}}(q_s \mid H^{-1}(\pi_i^b), T = t)\}_{i,s,l=1}^{n,S,L}$.

  end

  Sample $\left(u_1^b, \ldots, u_n^b\right)$ from $\mathcal{D}ir(1, \ldots, 1)$

  for $l = 1, \ldots, L$ **do**

    Calculate the CDF of $Y^{(t)}$ as follows:

$$SF_t^{bl}(q_s) = \sum_{i=1}^{n} u_i^b F^{kl}(q_s \mid H^{-1}(\pi_i^b), T = t), \text{ where } 1 \leq s \leq S \text{ and } t \in \{0,1\}$$

    Find a grid point $q_t^{bl}(\tau)$ such that $F_t^{kl}(q_t^{bl}(\tau)) = \tau$ for $t \in \{0,1\}$.

    The τ^{th} quantile from the CDF $F_t^{bl}(.)$ is $q_{t,\tau}^{bl}$ for the group $T = t$.

  end

end

Derive $\hat{q}_1(\tau)$ and $\hat{q}_0(\tau)$ as follows

$$\hat{q}_1(\tau) = \frac{1}{BL} \sum_{b}^{B} \sum_{l=1}^{L} q_1^{bl}(\tau) \quad \text{and} \quad \hat{q}_0(\tau) = \frac{1}{BL} \sum_{b}^{B} \sum_{l=1}^{L} q_0^{bl}(\tau)$$

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

E.2 Variants of the Proposed Approach

Bayesian Outcome Modelling (`BOM`)

Bayesian Outcome Modelling is an outcome-regression-based approach that omits the treatment assignment model. Instead, it solely focuses on estimating the conditional distribution through multiple Bayesian quantile regressions in the outcome model of each treatment group. Shrinkage priors, akin to the doubly-robust approach, can be readily incorporated.

Algorithm 7: Bayesian Outcome Modelling to estimate QTE

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

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

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

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

  Derive posterior mean $\hat{G}(y \mid t, \mathbf{X}) = \frac{1}{B}\sum_{b=1}^B G^{(b)}(y \mid t, \mathbf{X})$.

end

Derive $\hat{q}_1^{om}(\tau)$ and $\hat{q}_0^{om}(\tau)$ as the solutions to

$$\frac{1}{n}\sum_{i=1}^n \hat{G}(q_1 \mid 1, \mathbf{X}) = \tau \text{ and } \frac{1}{n}\sum_{i=1}^n \hat{G}(q_0 \mid 0, \mathbf{X}) = \tau$$

Calculate $\hat{\Delta}_\tau \equiv \widehat{QTE}^{om}(\tau) = \hat{q}_1^{om}(\tau) - \hat{q}_0^{om}(\tau)$.

Bayesian Propensity Score Analysis (`BPSA`)

Bayesian Propensity Score Analysis is a treatment-assignment-based approach, which involves fitting the treatment assignment and then employs multiple Bayesian quantile regressions to model the conditional distribution of the outcome given the posterior mean of the propensity score in each treatment group.

Algorithm 8: Bayesian Outcome Modelling to estimate QTE

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

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

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

Derive posterior mean from B posterior samples $\hat{\pi}(\mathbf{X}) = \frac{1}{B}\sum_{b=1}^B \pi^{(b)}(\mathbf{X})$.

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

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

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

end

Derive $\hat{q}_1^{ps}(\tau)$ and $\hat{q}_0^{ps}(\tau)$ as the solutions to

$$\frac{1}{n}\sum_{i=1}^n \hat{G}(q_1 \mid 1, \mathbf{X}) = \tau \text{ and } \frac{1}{n}\sum_{i=1}^n \hat{G}(q_0 \mid 0, \mathbf{X}) = \tau$$

Calculate $\hat{\Delta}_\tau \equiv \widehat{QTE}^{ps}(\tau) = \hat{q}_1^{ps}(\tau) - \hat{q}_0^{ps}(\tau)$.

F Additional Simulation Results

SD1

Table F.1: Simulation Results for SD1, relative RMSE

		Estimation Methods					
Percentiles	N	BDR	BDRS	BNP	LDML	TMLE	FIPW
10th	1000	1.71	1.689	**0.991**	1.234	1.771	1.789
	500	1.623	1.54	**0.946**	1.124	1.658	1.774
	100	2.459	1.226	**0.963**	1.193	1.332	1.529
25th	1000	1.052	1.044	0.978	**0.852**	1.022	1.136
	500	1.139	1.117	1.007	**0.92**	1.052	1.468
	100	1.188	0.977	**0.916**	0.995	1.001	1.169
50th	1000	**0.598**	0.599	0.99	0.744	0.623	0.659
	500	0.663	**0.657**	0.978	0.776	0.669	0.712
	100	0.916	**0.819**	0.971	0.944	0.886	1.114
75th	1000	**0.521**	0.525	0.979	0.758	0.568	0.577
	500	**0.586**	0.598	0.983	0.821	0.634	0.65
	100	0.99	**0.799**	0.986	0.92	0.835	1.032
90th	1000	**0.583**	0.593	0.98	0.803	0.605	0.606
	500	**0.635**	0.663	0.991	0.86	0.691	0.728
	100	1.233	**0.853**	0.974	0.928	0.897	1.055

Notes: This table displays the relative Root Mean Squared Error (RMSE) of different estimation methods across 100 replicates. The rows contain results for various percentile levels and for various sample size N. The relative RMSE is the RMSE in comparison with the Naive method as the benchmark, where $RMSE = \sqrt{R^{-1}\sum_{r=1}^{R}(\hat{\alpha}_r - \alpha)^2}$ and $R = 100$.

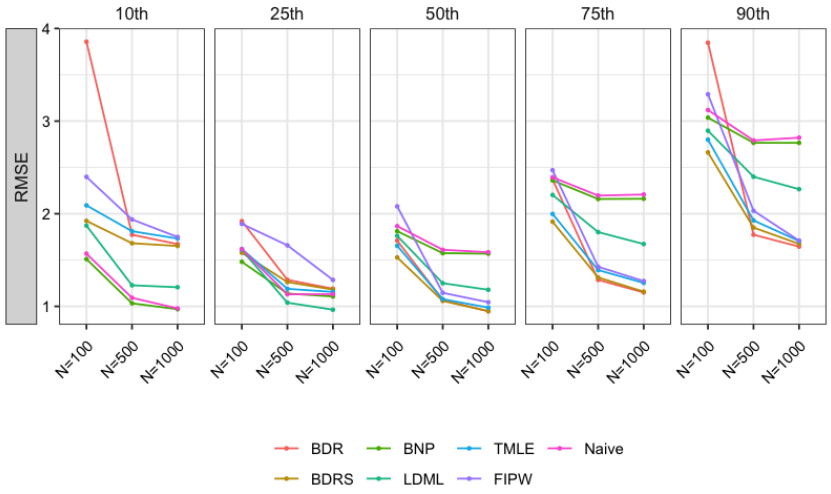

Figure F.1: Line plots of raw RMSE for 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} QTEs based on 100 replications.

SD2

Table F.2: Simulation Results for SD2a, Average Bias and relative RMSE

	Bias					RMSE				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Linear										
BDR	0.055	0.002	-0.032	-0.008	0.016	2.036	1.098	0.673	0.637	0.839
BDRS	0.061	0.002	-0.030	-0.006	0.024	2.028	1.098	0.678	0.638	0.846
BOM	0.040	0.092	-0.013	-0.094	-0.053	0.928	0.921	0.562	0.620	0.814
BOMS	0.059	0.102	-0.002	-0.076	-0.025	0.913	0.904	0.554	0.619	0.812
BPSA	0.078	-0.042	-0.044	0.061	0.112	0.735	0.980	0.523	0.595	0.719
BNP	0.335	0.368	0.343	0.328	0.284	1.025	1.042	0.917	0.966	1.038
LDML	0.009	0.010	-0.009	0.019	0.163	0.985	1.029	0.616	0.696	0.819
TMLE	0.000	-0.005	-0.022	-0.041	0.025	1.224	1.229	0.694	0.752	1.180
FIPW	0.019	-0.039	-0.034	-0.010	0.014	1.845	1.286	0.710	0.761	1.285
No covariates										
Naive	0.468	0.434	0.430	0.377	0.312	1.000	1.000	1.000	1.000	1.000

Notes: This table displays the average bias and the relative Root Mean Squared Error (RMSE) of different estimation methods across 100 replicates. The relative RMSE is the RMSE in comparison with the Naive method as the benchmark, where $RMSE = \sqrt{R^{-1}\sum_{r=1}^{R}(\hat{\alpha}_r - \alpha)^2}$ and $R = 100$.

Table F.3: Simulation Results for SD2b, Average Bias and relative RMSE

	Bias					RMSE				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Nonlinear										
BDRS	-0.014	0.015	0.019	0.027	0.011	0.680	0.878	0.520	0.570	0.543
BOMS	0.041	0.027	0.020	0.015	0.001	0.519	0.727	0.470	0.490	0.425
BPSA	0.140	-0.019	-0.029	0.077	0.099	0.715	0.955	0.538	0.604	0.705
LDML	0.111	0.052	0.047	0.062	0.182	0.852	1.006	0.722	0.730	0.831
TMLE	-0.020	0.009	0.023	0.023	0.064	0.694	0.898	0.522	0.578	0.669
FIPW	0.034	0.120	0.090	-0.050	-0.051	2.180	2.066	1.981	2.056	1.678
No covariates										
Naive	0.468	0.434	0.430	0.377	0.312	1.000	1.000	1.000	1.000	1.000

Notes: This table displays the average bias and the relative Root Mean Squared Error (RMSE) of different estimation methods across 100 replicates. The relative RMSE is the RMSE in comparison with the Naive method as the benchmark, where $RMSE = \sqrt{R^{-1}\sum_{r=1}^{R}(\hat{\alpha}_r - \alpha)^2}$ and $R = 100$.

G Additional Graphs in Empirical Illustration

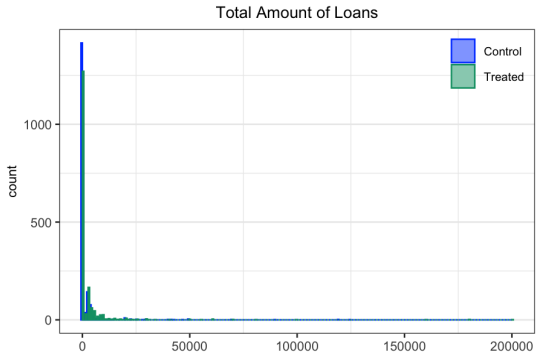

Figure G.1: Histogram of total amount of loans at household level in treated villages and control villages.