A Semiparametric Bayesian Approach for Estimating Extreme Conditional Quantile Treatment Effect for Heavy-tailed Data

Arnab Aich[a], Indrabati Bhattacharya[a]

[a]*Department of Statistics, Florida State University, 117 N Woodward Ave, Tallahassee, FL, 32306, USA*

ARTICLE INFO

Keywords:
Bayesian inference
Dirichlet process mixture
Extreme value theory
Generalized Pareto distribution
Causal inference
Quantile treatment effect

ABSTRACT

Estimation of extreme quantile treatment effects has substantial applications in various domains. In this article, we propose a novel semi-parametric Bayesian approach for estimation and inference of extreme quantile treatment effects, which provides a direct mechanism for uncertainty quantification, unlike the existing frequentist methods. In addition, the proposed method leverages density estimation and does not suffer from the "quantile crossing" issue. The modeling framework combines a Dirichlet process mixture and a generalized Pareto distribution, which accommodates the scarcity of data points and can extrapolate beyond the data range to estimate the extreme quantile treatment effects. Simulation studies show that our method can effectively capture the causal effects corresponding to both extreme and regular (non-extreme) quantiles. We illustrate the performance of our model by estimating the quantile treatment effect of obesity on Type II Diabetes using a real dataset.

Highlights

- Semiparametric Bayesian model for extreme conditional quantile treatment effects with heavy-tailed outcomes.
- Dirichlet process mixture for the central and a Generalized Pareto Distribution for the tail; no quantile crossing.
- Confounding adjustment via propensity scores; diabetes application with public data.

Quick Links: Introduction | Identification | Methodology | Simulation | Real Data | Discussion

1. Introduction

Causal inference seeks to quantify cause-and-effect relationships beyond mere association. In many applications, including medicine, finance, and climate science, inference on tail outcomes is of primary interest. We propose a novel Bayesian framework for estimating CQTEs in observational studies, with particular emphasis on extreme quantile levels. Within the potential outcomes framework, a substantial literature, both frequentist (Hirano et al. (2003); Imbens (2004); Lopez and Gutman (2017)) and Bayesian (Ray and Szabó (2019); Li et al. (2022); Oganisian and Roy (2021); Blette et al. (2023)), focuses on estimating average treatment effects. However, mean-based summaries can obscure important distributional heterogeneity, particularly in the tails, and may be sensitive to heavy-tailed outcomes. Quantile-based measures, such as the Quantile Treatment Effect (QTE) and its conditional analogue, provide a more complete characterization of treatment effects across the outcome distribution (Chernozhukov and Hansen (2005); Firpo (2007); Xu (2018); Powell (2020)). The Conditional Quantile Treatment Effect (CQTE) provides more detailed insights than the Quantile Treatment Effect (QTE) by accounting for covariate-specific heterogeneity. While QTE reflects average distributional (quantile) changes across the population, CQTE demonstrates how treatment effects vary across subgroups, supporting targeted policies and personalized treatments.

E aaaich@fsu.edu (A. Aich); ib22g@fsu.edu (I. Bhattacharya)
ORCID(s):

In randomized experiments, QTE can be identified by directly comparing outcome quantiles across treatment groups. For example, Venturini et al. (2015) proposed quantile ratio smoothing to estimate distributional differences. In observational studies, however, treatment assignment depends on confounders, requiring additional identification assumptions (see Section 2). A common approach is to use propensity scores (Rosenbaum and Rubin (1983)) to adjust for confounding and recover causal effects.

In extreme value theory, quantiles are classified according to their asymptotic order as the sample size n increases. Specifically, a quantile index τ is termed extreme if $n\tau \to 0$, moderately extreme if $n\tau \to d > 0$, and intermediate if $n\tau \to \infty$, with $\tau \to 0$. While intermediate quantiles can be estimated using standard quantile regression methods, estimation of moderately extreme and extreme quantiles is challenging due to data sparsity in the tails. As noted by de Haan and Ferreira (2007), reliable inference in such settings typically requires extrapolation beyond the observed data. A large body of work addresses extreme quantile estimation using extreme value theory. Classical approaches include block maxima methods and peak-over-threshold (POT) techniques, which model tail behavior using limiting distributions. Bayesian extensions have incorporated flexible modeling strategies, including Dirichlet process mixtures (do Nascimento et al. (2011)) and generalized Pareto–based formulations (Cabras and Castellanos (2011)). However, these approaches primarily focus on marginal quantiles rather than treatment effects. In the context of quantile treatment effects, Zhang et al. (2018) propose a parametric approach based on extreme-value approximations for intermediate and moderately extreme quantiles, while Deuber et al. (2022) extend this framework to fully extreme regimes with asymptotic guarantees. Despite these advances, Bayesian methods for estimating extreme quantile treatment effects remain largely underdeveloped, leaving a significant gap in the literature.

We propose a semiparametric Bayesian framework, MixGPD, that combines a Dirichlet process mixture model with peak-over-threshold (POT) methods to estimate CQTEs across the entire outcome distribution, including both intermediate and extreme quantiles. To adjust for confounding, propensity scores are estimated first, and incorporated in the outcome model as additional covariates. The framework provides full uncertainty quantification through posterior credible intervals for CQTEs. Unlike existing approaches (Fúquene Patiño (2015); Zhang et al. (2018); Deuber et al. (2022)), our method does not rely on shifting quantile indices and provides a unified framework for modeling central and tail behavior. In particular, the bulk of the distribution is modeled using a Dirichlet process mixture, while tail behavior is captured via a generalized Pareto distribution (GPD). The use of the GPD is supported by extreme value theory, which shows that exceedances over a high threshold converge to a generalized Pareto distribution under broad regularity conditions. The proposed approach also mitigates the quantile crossing issue commonly encountered in standard quantile regression models.

The remainder of the paper is organized as follows. Section 2 outlines the identification framework and required assumptions. Section 3 presents the proposed model and estimation procedure. Section 4 reports simulation results, and Section 5 presents an empirical application. Section 6 presents a discussion.

2. Notation and assumptions

Let $A \in \{0, 1\}$ denote the binary treatment of interest and $Y(0)$ and $Y(1)$ denote the corresponding potential outcomes. Let Y_i denote the observed outcome of interest for the ith individual, $A_i \in \{0, 1\}$ denote the observed treatment, and $\boldsymbol{X}_i \in \mathbb{R}^p$ denote the vector of baseline covariates. The outcome is assumed to be supported on the positive real line. Our causal estimand of interest is the conditional quantile treatment effect (CQTE), defined as

$$q_\tau(\boldsymbol{x}) = Q^{(1)}(\tau \mid \boldsymbol{X} = \boldsymbol{x}) - Q^{(0)}(\tau \mid \boldsymbol{X} = \boldsymbol{x}), \tag{1}$$

where $Q^{(a)}(\tau \mid \boldsymbol{X} = \boldsymbol{x})$ denotes the conditional τ-quantile of $Y(a)$. Our primary focus is on upper-tail quantile treatment effects corresponding to sequences $\tau_n \to 1$, with $n(1 - \tau_n) \to 0$, following extreme value theory. To identify the estimand in (1), we impose the following standard assumptions:

1. **Ignorability.** $\{Y_i(1), Y_i(0)\} \perp A_i \mid \boldsymbol{X}_i$, for all $i = 1, \ldots, n$.
2. **Overlap.** $0 < P(A_i = 1 \mid \boldsymbol{X}_i) < 1$, for all $i = 1, \ldots, n$.
3. **SUTVA.** $Y_i(A_1, \ldots, A_n) = Y_i(A_i)$, for all $i = 1, \ldots, n$.

4. **Consistency.** $Y_i = Y_i(A_i)$, for all $i = 1, \ldots, n$.

Assumptions (1)–(4) imply that the conditional distribution of the potential outcomes is identified from the observed data. Let

$$F^{(a)}(y \mid \boldsymbol{X} = \boldsymbol{x}) := \Pr(Y(a) \leq y \mid \boldsymbol{X} = \boldsymbol{x})$$

denote the conditional distribution function of the potential outcome under treatment level a, and let

$$F_Y(y \mid \boldsymbol{X} = \boldsymbol{x}, A = a) := \Pr(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}, A = a)$$

denote the corresponding conditional distribution of the observed outcome. Under ignorability and consistency,

$$F^{(a)}(y \mid \boldsymbol{X} = \boldsymbol{x}) = F_Y(y \mid \boldsymbol{X} = \boldsymbol{x}, A = a), \qquad a \in \{0, 1\}.$$

Accordingly, the CQTE in (1) can be obtained from the conditional quantile functions

$$Q^{(a)}(\tau \mid \boldsymbol{X} = \boldsymbol{x}) = \inf\{y : F_Y(y \mid \boldsymbol{X} = \boldsymbol{x}, A = a) \geq \tau\}, \quad a \in \{0, 1\}.$$

Thus, estimation of the CQTE reduces to modeling the conditional distribution of the observed outcome given covariates and treatment assignment (Firpo, 2007). The estimation methodology is presented in the following section.

3. Proposed Methodology

We propose a two-stage Bayesian framework for estimating CQTEs, accommodating both intermediate and extreme quantiles. The framework separates the modeling of the treatment assignment from the outcome, mitigating feedback issues, while allowing for robust estimation in the tails of the outcome distribution.

3.1. Stage 1: Treatment Assignment Model

Let $\boldsymbol{x} \in \mathbb{R}^p$ denote the covariate vector and $A \in \{0, 1\}$ the treatment indicator. The propensity score (PS) summarizes the treatment assignment mechanism:

$$\rho_{\boldsymbol{x}} = \Pr(A = 1 \mid \boldsymbol{X} = \boldsymbol{x}).$$

Under Assumptions (1) and (2), treatment assignment is strongly ignorable given $\rho_{\boldsymbol{x}}$, and the PS satisfies the balancing property (Rosenbaum and Rubin, 1983). In practice, the true PS is unknown and must be estimated. While the PS was originally proposed as a low-dimensional summary of covariates, it is generally insufficient to model the outcome distribution alone:

$$\Pr(Y(a) \mid \boldsymbol{X} = \boldsymbol{x}) \neq \Pr(Y(a) \mid \rho_{\boldsymbol{x}}),$$

as discussed in Zigler (2016). To address this, we adopt a two-stage procedure described as follows. At the first stage, we estimate the PS using a Bayesian logistic regression:

$$A \mid \boldsymbol{x}, \boldsymbol{\beta}_{ps} \sim \text{Bernoulli}(\rho_{\boldsymbol{x}}), \quad \rho_{\boldsymbol{x}} = \frac{1}{1 + \exp(-\boldsymbol{x}^\top \boldsymbol{\beta}_{ps})}, \quad \boldsymbol{\beta}_{ps} \sim p_0(\boldsymbol{\beta}_{ps}). \tag{2}$$

At the second stage, we use the posterior estimates $\widehat{\rho}_{\boldsymbol{x}}$ along with the original covariates $\boldsymbol{x}$ as predictors in the outcome model. This approach mitigates feedback and allows flexible modeling of the outcome conditional on covariates and treatment assignment.

3.2. Stage 2: Outcome model

Let $\boldsymbol{r}(\boldsymbol{x}) = (\boldsymbol{x}^\top, \widehat{\rho}_{\boldsymbol{x}})^\top$ denote the augmented covariate vector, including the estimated propensity score. For each treatment arm $a \in \{0, 1\}$, we model the conditional outcome distribution using a two-part approach: a Dirichlet process mixture (DPM) for the bulk of the distribution and a generalized Pareto distribution (GPD) for the tail.

Bulk Model: Covariate-Dependent Dirichlet Process Mixture. Let $k(y \mid \mu(\boldsymbol{x}), \eta)$ denote a kernel density with covariate-dependent location parameter $\mu(\boldsymbol{x})$ and additional parameter η. Each component is chosen from a right-skewed, moderate-tailed, positive-support family (Gamma, Lognormal, or Inverse-Gaussian (Folks and Chhikara, 1978)) to ensure that the bulk model does not interfere with the tail behavior. Covariate dependence is introduced through

$$\mu^{(a)}(\boldsymbol{x}) = g\big(\boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{\beta}_\mu^{(a)}\big),$$

where $\boldsymbol{\beta}_\mu^{(a)}$ are component-specific regression coefficients and $g(\cdot)$ is a link function determined by the kernel. The hierarchical DPM for treatment arm a is then specified as

$$\begin{aligned} Y \mid A = a, \boldsymbol{x};\, \boldsymbol{\beta}_\mu^{(a)}, \eta^{(a)} &\sim k\big(y \mid \mu^{(a)}(\boldsymbol{x}), \eta^{(a)}\big), \\ (\boldsymbol{\beta}_\mu^{(a)}, \eta^{(a)}) \mid H^{(a)} &\sim H^{(a)}, \\ H^{(a)} \mid \kappa^{(a)}, H_0^{(a)} &\sim DP\big(\kappa^{(a)}, H_0^{(a)}\big), \quad H_0^{(a)} = H_\beta^{(a)} \times H_\eta^{(a)}, \end{aligned}$$

where $\kappa^{(a)}$ is the concentration parameter and $H_0^{(a)}$ is the base measure. The component parameters are assumed independent under $H_0^{(a)}$. Marginalizing over $H^{(a)}$ yields the mixture density

$$f_{\mathrm{DP}}^{(a)}(y \mid \boldsymbol{x}) = \sum_{j=1}^{\infty} w^{(a,j)}\, k\big(y \mid \mu^{(a,j)}(\boldsymbol{x}), \eta^{(a,j)}\big), \qquad \sum_{j=1}^{\infty} w^{(a,j)} = 1,$$

where $\{w^{(a,j)}\}$ are the Dirichlet process weights and j indexes mixture components.

Threshold and Tail Model. To separate the bulk and tail regions, we introduce a covariate-dependent threshold

$$u^{(a)}(\boldsymbol{x}) \sim \mathcal{LN}\big(\boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{\beta}_u^{(a)}, s_{u^{(a)}}\big),$$

ensuring positivity of the threshold. For exceedances $y > u^{(a)}(\boldsymbol{x})$, the tail is modeled using a generalized Pareto distribution:

$$f_{\mathrm{GPD}}^{(a)}(y \mid \boldsymbol{x}; \boldsymbol{\beta}_u^{(a)}, \sigma^{(a)}, \xi^{(a)}) = \begin{cases} \dfrac{1}{\sigma^{(a)}}\left(1 + \dfrac{\xi^{(a)}}{\sigma^{(a)}}\big(y - u^{(a)}(\boldsymbol{x})\big)\right)^{-(1+1/\xi^{(a)})}, & \xi^{(a)} \neq 0, \\ \dfrac{1}{\sigma^{(a)}} \exp\left(-\dfrac{y - u^{(a)}(\boldsymbol{x})}{\sigma^{(a)}}\right), & \xi^{(a)} = 0, \end{cases}$$

with $\sigma^{(a)} > 0$. We restrict $\xi^{(a)} \geq -0.5$ to satisfy regularity conditions (Pickands, 1975).

Combined Conditional Density and Distribution. The full conditional density for arm a is

$$f^{(a)}(y \mid \boldsymbol{x}) = \begin{cases} f_{\mathrm{DP}}^{(a)}(y \mid \boldsymbol{x}), & y \leq u^{(a)}(\boldsymbol{x}), \\ \big[1 - F_{\mathrm{DP}}^{(a)}(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x})\big]\, f_{\mathrm{GPD}}^{(a)}(y \mid \boldsymbol{x}), & y > u^{(a)}(\boldsymbol{x}), \end{cases}$$

where $F_{\mathrm{DP}}^{(a)}(\cdot)$ is the CDF of the bulk mixture. The corresponding CDF is

$$F^{(a)}(y \mid \boldsymbol{x}) = \begin{cases} F_{\mathrm{DP}}^{(a)}(y \mid \boldsymbol{x}), & y \leq u^{(a)}(\boldsymbol{x}), \\ F_{\mathrm{DP}}^{(a)}\big(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x}\big) + \big[1 - F_{\mathrm{DP}}^{(a)}\big(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x}\big)\big] & \\ \quad \times F_{\mathrm{GPD}}^{(a)}(y \mid \boldsymbol{x}), & y > u^{(a)}(\boldsymbol{x}). \end{cases}$$

Quantile Estimation. Let $\tau \in (0,1)$ denote the target quantile level. When

$$\tau \leq F_{\mathrm{DP}}^{(a)}\big(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x}\big),$$

the quantile $Q^{(a)}(\tau \mid \boldsymbol{x})$ is obtained by numerically inverting $F_{\mathrm{DP}}^{(a)}$. When

$$\tau > F_{\mathrm{DP}}^{(a)}\big(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x}\big),$$

we define the covariate-dependent exceedance probability

$$\tilde{\tau}^{(a)}(\boldsymbol{x}) = \frac{\tau - F_{\mathrm{DP}}^{(a)}\big(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x}\big)}{1 - F_{\mathrm{DP}}^{(a)}\big(u^{(a)}(\boldsymbol{x}) \mid \boldsymbol{x}\big)},$$

and obtain the tail quantile by inverting the GPD CDF.

Although the proposed framework accommodates general kernels, we focus on Gamma, Lognormal, and Inverse-Gaussian distributions. In the next section, we discuss the prior distributions related to our model specifications.

3.3. Prior Selection

In this section, we specify prior distributions for all model parameters. Priors are assumed independent across stages and, within each stage, conditionally independent across treatment arms $a \in \{0,1\}$ and mixture components j, unless otherwise stated.

Propensity Score Model. For the propensity score model in (2), we assign a multivariate Gaussian prior

$$\boldsymbol{\beta}_{ps} \sim \mathcal{N}_p\big(\boldsymbol{\mu}_{\beta_{ps}}, \boldsymbol{\Sigma}_{\beta_{ps}}\big),$$

where $\boldsymbol{\mu}_{\beta_{ps}}$ and $\boldsymbol{\Sigma}_{\beta_{ps}}$ encode prior beliefs on the log-odds of treatment assignment.

Dirichlet Process Mixture (Bulk Model). For each arm a, the bulk distribution is modeled via a Dirichlet process prior with concentration parameter $\alpha^{(a)}$. Using the stick-breaking construction,

$$\alpha^{(a)} \sim \mathcal{G}(a_\alpha, b_\alpha), \qquad v_j^{(a)} \mid \alpha^{(a)} \sim \mathcal{B}(1, \alpha^{(a)}),$$

$$w_1^{(a)} = v_1^{(a)}, \quad w_j^{(a)} = v_j^{(a)} \prod_{k=1}^{j-1}(1 - v_k^{(a)}), \quad j \geq 2.$$

The regression coefficients for mixture components are assigned

$$\boldsymbol{\beta}_\mu^{(a,j)} \sim \mathcal{N}_{p+1}\big(\boldsymbol{\mu}_{\beta_\mu}, \boldsymbol{\Sigma}_{\beta_\mu}\big).$$

The specification of the location parameter depends on the kernel choice:

$$\mu^{(a,j)}(\boldsymbol{x}) = \begin{cases} \exp\big(\boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{\beta}_\mu^{(a,j)}\big), & \text{Gamma / Inverse-Gaussian,} \\ \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{\beta}_\mu^{(a,j)}, & \text{Lognormal.} \end{cases}$$

For all kernels, the additional parameter $\eta^{(a,j)}$ controls the shape and is assigned a Gamma prior:

$$\eta^{(a,j)} \sim \mathcal{G}(\alpha_\eta, \beta_\eta).$$

Under the independence assumption, the base measure $H_0^{(a)}$ factorizes as the product of priors for $\boldsymbol{\beta}_\mu^{(a,j)}$ and $\eta^{(a,j)}$.

Threshold Model. For each arm a, the covariate-dependent threshold is modeled as

$$u^{(a)}(\boldsymbol{x}) \sim \mathcal{LN}\Big(\boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{\beta}_u^{(a)},\, s_u^{(a)}\Big)\,,$$

ensuring positivity. To control the number of exceedances, we impose a lower truncation:

$$u^{(a)}(\boldsymbol{x}) \geq l_{u^{(a)}}(\boldsymbol{x}),$$

where $l_{u^{(a)}}(\boldsymbol{x})$ is typically set above the 60th percentile of observed outcomes in arm a.

The regression coefficients are assigned

$$\boldsymbol{\beta}_u^{(a)} \sim \mathcal{N}_{p+1}(\boldsymbol{\mu}_{\beta_u},\, \boldsymbol{\Sigma}_{\beta_u})\,.$$

GPD Tail Parameters. For the tail parameters $(\sigma^{(a)}, \xi^{(a)})$, we adopt a Jeffreys-type prior:

$$p_0(\sigma^{(a)}, \xi^{(a)}) \propto \frac{1}{\sigma^{(a)}(1+\xi^{(a)})\sqrt{1+2\xi^{(a)}}}\, \mathbb{I}\{\sigma^{(a)} > 0\}\, \mathbb{I}\{\xi^{(a)} > -0.5\}.$$

The constraint $\xi^{(a)} > -0.5$ ensures regularity (Pickands, 1975).

Full Hierarchical Model. The complete model is

$$\begin{aligned}
&A_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}_{ps} \sim \mathrm{Ber}(\rho_{x_i}), \quad \mathrm{logit}(\rho_{x_i}) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_{ps},\\
&\boldsymbol{\beta}_{ps} \sim \mathcal{N}_p(\boldsymbol{\mu}_{\beta_{ps}}, \boldsymbol{\Sigma}_{\beta_{ps}}),\\
&\widehat{\rho}_{x_i} = \{1+\exp(-\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_{ps})\}^{-1}, \quad \boldsymbol{r}(\boldsymbol{x}_i) = (\boldsymbol{x}_i^\top, \widehat{\rho}_{x_i})^\top,\\
&y_i \mid A_i = a, \boldsymbol{x}_i, \Theta^{(a)} \sim \begin{cases} f_{\mathrm{DP}}^{(a)}(y_i \mid \boldsymbol{x}_i), & y_i \leq u^{(a)}(\boldsymbol{x}_i),\\ \left[1 - F_{\mathrm{DP}}^{(a)}(u^{(a)}(\boldsymbol{x}_i) \mid \boldsymbol{x}_i)\right] f_{\mathrm{GPD}}^{(a)}(y_i \mid \boldsymbol{x}_i), & y_i > u^{(a)}(\boldsymbol{x}_i), \end{cases}\\
&\Theta^{(a)} = \{\boldsymbol{\beta}_\mu^{(a)}, \boldsymbol{\eta}^{(a)}, \boldsymbol{\beta}_u^{(a)}, \sigma^{(a)}, \xi^{(a)}\}.
\end{aligned} \tag{3}$$

In the next section, we describe the likelihood for the proposed model and the corresponding posterior distribution,

3.4. Likelihood and Posterior Distributions

Stage 1. The log-likelihood for the propensity score model is

$$\log L(\boldsymbol{\beta}_{ps}) = \sum_{i=1}^{n} \Big[A_i \log \rho_{x_i} + (1-A_i)\log(1-\rho_{x_i})\Big], \tag{4}$$

with posterior

$$\log p(\boldsymbol{\beta}_{ps} \mid \boldsymbol{X}, \boldsymbol{A}) \propto \log L(\boldsymbol{\beta}_{ps}) + \log p_0(\boldsymbol{\beta}_{ps}). \tag{5}$$

Stage 2. For each arm a, define

$$\Theta^{(a)} = \Big\{\boldsymbol{\beta}_\mu^{(a)}, \boldsymbol{\eta}^{(a)}, \boldsymbol{\beta}_u^{(a)}, \sigma^{(a)}, \xi^{(a)}\Big\}\,.$$

The log-likelihood is

$$\ell(\Theta^{(a)}) = \sum_{i:\, A_i = a} \log f^{(a)}(y_i \mid \boldsymbol{x}_i)$$

$$= \sum_{i:\, A_i=a} \Big[\mathbb{I}(y_i \le u^{(a)}(\boldsymbol{x}_i)) \log f_{\mathrm{DP}}^{(a)}(y_i \mid \boldsymbol{x}_i) + \mathbb{I}(y_i > u^{(a)}(\boldsymbol{x}_i)) \Big\{ \log\big(1 - F_{\mathrm{DP}}^{(a)}(u^{(a)}(\boldsymbol{x}_i) \mid \boldsymbol{x}_i)\big) + \log f_{\mathrm{GPD}}^{(a)}(y_i \mid \boldsymbol{x}_i) \Big\} \Big]. \tag{6}$$

The full log-likelihood is

$$\ell(\Theta^{(1)}, \Theta^{(0)}) = \ell(\Theta^{(1)}) + \ell(\Theta^{(0)}). \tag{7}$$

Posterior Distribution. Combining likelihood and priors, the joint posterior is

$$\log p(\Theta^{(1)}, \Theta^{(0)} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{A}) \propto \ell(\Theta^{(1)}, \Theta^{(0)}) + \log p_0(\Theta^{(1)}, \Theta^{(0)}). \tag{8}$$

Posterior samples of $\Theta^{(0)}$ and $\Theta^{(1)}$ are obtained via MCMC (the steps are described in the supplement), and are subsequently used to estimate arm-specific conditional quantiles and CQTE.

3.5. Computation of conditional quantiles and the treatment effects

Here we describe posterior computation of conditional quantiles and Conditional Quantile Treatment Effects (CQTE). Let $\{\Theta^{(a,s)}\}_{s=1}^{S}$ denote posterior draws of the arm-specific parameter vector

$$\Theta^{(a,s)} = \Big\{ \boldsymbol{\beta}_{\mu}^{(a,s)},\, \boldsymbol{\eta}^{(a,s)},\, \boldsymbol{\beta}_{u}^{(a,s)},\, \sigma^{(a,s)},\, \xi^{(a,s)} \Big\}, \quad a \in \{0,1\}.$$

For each subject i with covariates $\boldsymbol{x}_i$, we proceed as follows.

Step 1: Compute the fitted propensity score using posterior summaries of $\boldsymbol{\beta}_{ps}$:

$$\widehat{\rho}_{x_i} = \frac{1}{1 + \exp(-\boldsymbol{x}_i^{\top} \widehat{\boldsymbol{\beta}}_{ps})}, \quad \boldsymbol{r}(\boldsymbol{x}_i) = (\boldsymbol{x}_i^{\top}, \widehat{\rho}_{x_i})^{\top}.$$

Step 2: For each arm a and posterior draw s, compute:

- Threshold:
$$\widehat{u}^{(a,s)}(\boldsymbol{x}_i) = \exp\Big(\boldsymbol{r}(\boldsymbol{x}_i)^{\top} \boldsymbol{\beta}_{u}^{(a,s)} \Big).$$
- Mixture component means (for the DP bulk model):
$$\widehat{\mu}_{j}^{(a,s)}(\boldsymbol{x}_i) = \exp\Big(\boldsymbol{r}(\boldsymbol{x}_i)^{\top} \boldsymbol{\beta}_{\mu}^{(a,j,s)} \Big).$$
- Bulk CDF evaluated at the threshold:
$$F_{\mathrm{DP}}^{(a,s)}\Big(\widehat{u}^{(a,s)}(\boldsymbol{x}_i) \mid \boldsymbol{x}_i \Big).$$

Step 3: For a target quantile level $\tau \in (0,1)$, define the draw-wise conditional quantile $Q^{(a,s)}(\tau \mid \boldsymbol{x}_i)$ as follows.

(i) Bulk regime: If

$$\tau \le F_{\mathrm{DP}}^{(a,s)}\Big(\widehat{u}^{(a,s)}(\boldsymbol{x}_i) \mid \boldsymbol{x}_i \Big),$$

then

$$Q^{(a,s)}(\tau \mid \boldsymbol{x}_i) = \inf\Big\{ y \le \widehat{u}^{(a,s)}(\boldsymbol{x}_i) : F_{\mathrm{DP}}^{(a,s)}(y \mid \boldsymbol{x}_i) \ge \tau \Big\}.$$

(ii) Tail regime: Otherwise, define the exceedance probability

$$\tilde{\tau}^{(a,s)}(\boldsymbol{x}_i) = \frac{\tau - F_{\mathrm{DP}}^{(a,s)}\big(\widehat{u}^{(a,s)}(\boldsymbol{x}_i) \mid \boldsymbol{x}_i \big)}{1 - F_{\mathrm{DP}}^{(a,s)}\big(\widehat{u}^{(a,s)}(\boldsymbol{x}_i) \mid \boldsymbol{x}_i \big)}.$$

The tail quantile is obtained via inversion of the GPD:

$$Q^{(a,s)}(\tau \mid \boldsymbol{x}_i) = \begin{cases} \widehat{u}^{(a,s)}(\boldsymbol{x}_i) - \sigma^{(a,s)} \log\big(1 - \tilde{\tau}^{(a,s)}(\boldsymbol{x}_i)\big), & \xi^{(a,s)} = 0, \\ \widehat{u}^{(a,s)}(\boldsymbol{x}_i) + \dfrac{\sigma^{(a,s)}}{\xi^{(a,s)}} \left[\big(1 - \tilde{\tau}^{(a,s)}(\boldsymbol{x}_i)\big)^{-\xi^{(a,s)}} - 1\right], & \xi^{(a,s)} \neq 0. \end{cases}$$

Step 4: For each posterior draw, compute the CQTE as

$$q^{(s)}(\tau \mid \boldsymbol{x}_i) = Q^{(1,s)}(\tau \mid \boldsymbol{x}_i) - Q^{(0,s)}(\tau \mid \boldsymbol{x}_i).$$

Step 5: Calculate the posterior mean CQTE and $(1-\alpha)$ credible interval as

$$\widehat{q}(\tau \mid \boldsymbol{x}_i) = \frac{1}{S} \sum_{s=1}^{S} q^{(s)}(\tau \mid \boldsymbol{x}_i),$$
$$\widehat{CI}_{1-\alpha}(\tau \mid \boldsymbol{x}_i) = \big[\widehat{q}_r^{(\alpha/2)}(\tau \mid \boldsymbol{x}_i),\ \widehat{q}_r^{(1-\alpha/2)}(\tau \mid \boldsymbol{x}_i)\big].$$

For a new covariate vector $\tilde{\boldsymbol{x}}$, we compute $\widehat{\rho}_{\tilde{x}}$, form $\boldsymbol{r}(\tilde{\boldsymbol{x}})$, and repeat Steps 2–5 using posterior samples $\{\Theta^{(a,s)}\}_{s=1}^{S}$. The predicted CQTE and corresponding credible interval are obtained analogously:

$$\widehat{q}(\tau \mid \tilde{\boldsymbol{x}}) = \frac{1}{S} \sum_{s=1}^{S} q^{(s)}(\tau \mid \tilde{\boldsymbol{x}}),$$
$$\widehat{CI}_{1-\alpha}(\tau \mid \tilde{\boldsymbol{x}}) = \big[\widehat{q}_r^{(\alpha/2)}(\tau \mid \tilde{\boldsymbol{x}}),\ \widehat{q}_r^{(1-\alpha/2)}(\tau \mid \tilde{\boldsymbol{x}})\big].$$

These posterior summaries directly quantify uncertainty in CQTE through the empirical distribution of $\{q^{(s)}(\tau \mid \tilde{\boldsymbol{x}})\}_{s=1}^{S}$. For outcomes not naturally supported on $(0,\infty)$, a strictly monotone transformation can be applied prior to modeling. Conditional quantiles on the original scale are then obtained by back-transformation, and CQTE is computed as the difference between arm-specific quantiles after transformation.

4. Simulation Study

In this section, we evaluate the finite-sample performance of the proposed approach (MixGPD) for conditional quantile treatment effects, with particular focus on extreme quantiles. We compare MixGPD against three benchmark approaches under varying sample sizes and model specifications, with the goal of assessing robustness to bulk misspecification and performance in the extreme tails.

4.1. Data Generating Process

We generate data from a mechanism that mirrors the proposed bulk–tail structure. Each observation consists of covariates $\boldsymbol{X}_i$, treatment A_i, and outcome Y_i.

Covariates. Each subject has a four-dimensional covariate vector

$$\boldsymbol{X} = (X_1, X_2, X_3, X_4)^\top,$$

where X_1 is discrete with $\Pr(X_1 = k) = 0.25$ for $k \in \{1,2,3,4\}$, and $X_2, X_3, X_4 \overset{iid}{\sim} \text{Uniform}(1,5)$.

Treatment assignment. Treatment is assigned via a logistic model,

$$\rho(\boldsymbol{x}) = \frac{1}{1 + \exp(-\boldsymbol{x}^\top \boldsymbol{\beta}_\rho)}, \quad \boldsymbol{\beta}_\rho = \begin{bmatrix} -0.03 \\ -0.40 \\ 0.12 \\ 0.55 \end{bmatrix}, \quad A \sim \text{Bernoulli}(\rho(\boldsymbol{x})).$$

Outcome model. Potential outcomes $Y(a)$ are generated separately for $a \in \{0,1\}$ using a two-part mixture: a covariate-dependent finite mixture (bulk), and a GPD tail above a covariate-dependent threshold. Let $u^{(a)}(\boldsymbol{x})$ denote the threshold and $p_u^{(a)}$ the exceedance probability. For each subject, draw $r \sim \mathcal{U}(0,1)$. If $r < p_u^{(a)}$, we generate from the bulk distribution; otherwise, we generate from the GPD tail. This construction ensures that the bulk component does not generate values exceeding the threshold.

Treatment arm ($a = 1$). The bulk distribution is a three-component mixture of reparameterized Gamma densities:

$$f_{\text{bulk}}^{(1)}(y \mid \boldsymbol{x}) = \sum_{k=1}^{3} \pi_k^{(1)}\, \mathcal{G}(y \mid \mu_k^{(1)}(\boldsymbol{x}), \eta_k),$$

with weights $(0.25, 0.35, 0.40)$ and shapes $(3, 2, 3)$. The Gamma density is

$$\mathcal{G}(y \mid \mu, \eta) = \frac{(\eta/\mu)^{\eta}}{\Gamma(\eta)} y^{\eta-1} \exp\!\left(-\frac{\eta}{\mu} y\right). \tag{9}$$

The component means depend on $(\boldsymbol{x}, \rho(\boldsymbol{x}))$ through a log-linear model:

$$\log \boldsymbol{\mu}^{(1)}(\boldsymbol{x}) = \begin{bmatrix} 0.625 & 0.4 & 0.55 \\ -0.125 & -0.1 & 0.15 \\ 0.2 & 0.1 & -0.075 \\ 0.15 & 0.05 & -0.1 \\ 0.4 & 0.25 & 0.2 \end{bmatrix}^{\top} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \rho(\boldsymbol{x}) \end{bmatrix}.$$

The threshold satisfies

$$\log u^{(1)}(\boldsymbol{x}) = \begin{bmatrix} 0.30 \\ -0.13 \\ 0.15 \\ 0.07 \\ 2 \end{bmatrix}^{\top} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \rho(\boldsymbol{x}) \end{bmatrix},$$

and exceedances follow $\text{GPD}(3.5, 0.3)$.

Control arm ($a = 0$). The bulk is a three-component mixture of lognormal distributions:

$$f_{\text{bulk}}^{(0)}(y \mid \boldsymbol{x}) = \sum_{k=1}^{3} \pi_k^{(0)}\, \mathcal{LN}(y \mid \mu_k^{(0)}(\boldsymbol{x}), \sigma_k),$$

with weights $(0.55, 0.20, 0.25)$ and log-scale standard deviations $(\sigma_1, \sigma_2, \sigma_3) = (0.40, 0.50, 0.35)$. Means and thresholds follow analogous log-linear structures as follows,

$$\log \boldsymbol{\mu}^{(0)}(\boldsymbol{x}) = \begin{bmatrix} 0.3 & 0.4 & 0.2 \\ -0.4 & -0.2 & 0.2 \\ 0.3 & 0.15 & -0.1 \\ -0.1 & 0.2 & 0.05 \\ 0.2 & 0.15 & 0.25 \end{bmatrix}^{\top} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \rho_x \end{bmatrix},$$

$$\log u^{(0)}(\boldsymbol{x}) = \begin{bmatrix} 0.075 \\ -0.05 \\ 0.125 \\ 0.05 \\ 1.5 \end{bmatrix}^{\top} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \rho_x \end{bmatrix}.$$

Finally, the tail follows GPD$(2, 0.25)$. For the treatment arm, the average exceedance probability is approximately $p_u^{(1)} \approx 0.85$, while for the control arm it is $p_u^{(0)} \approx 0.80$. For any $\boldsymbol{x}$ and quantile level τ, the true CQTE is

$$q(\tau \mid \boldsymbol{x}) = Q^{(1)}(\tau \mid \boldsymbol{x}) - Q^{(0)}(\tau \mid \boldsymbol{x}),$$

where $Q^{(a)}$ is obtained by numerically inverting the true mixture–GPD distribution. Figure (1) displays conditional quantile surfaces for the treatment and control groups, along with the corresponding CQTE, over the covariate pair (x_1, x_2), with (x_3, x_4) fixed. Results are shown for $\tau = 0.5$ and $\tau = 0.95$.

The treatment surfaces exhibit a stronger upward gradient with respect to both covariates compared to the control surfaces, indicating a more prominent covariate effect. This difference is reflected in the CQTE surfaces, where larger treatment effects appear in regions with steeper gradients. The contrast becomes more pronounced at higher quantiles, consistent with the heavier tail behavior induced by the GPD component. Additional results over (x_3, x_4) are provided in Appendix Figure (7), where increased curvature reflects nonlinear covariate effects built into the data-generating mechanism.

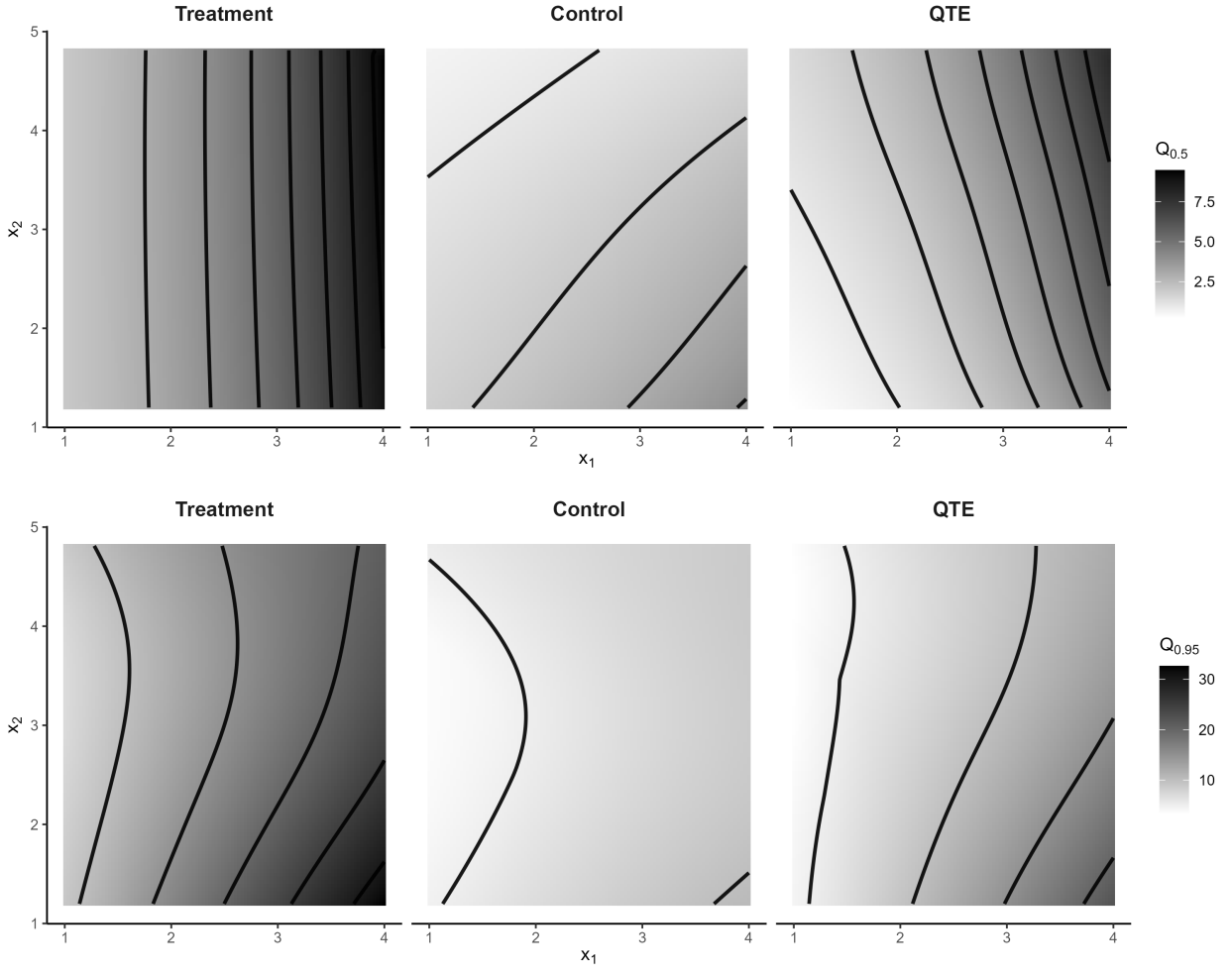

Figure 1: Conditional quantile surfaces over the covariate pair combination of x_1 and x_2 while other two covariates fixed at $x_3 = 3, x_4 = 3$. Top: Quantile index $(\tau) = 0.5$, Bottom: Quantile index $(\tau) = 0.95$. Each panel displays the treatment surface, the control surface, and the QTE surface. Color shading indicates the magnitude of the conditional quantile, with darker shades reflecting lower values and lighter shades representing higher values. Smooth white contour lines trace level curves of equal quantile/QTE height, revealing how quickly the surfaces change across the covariate space.

4.2. Simulation Design

We consider a sample size of $n = 500$ and generate $R = 100$ datasets per scenario. We evaluate the proposed bulk–tail mixture estimator and several competing approaches.

- **MixGPD** (proposed): We evaluate three bulk specifications — **MixGammaGPD** (Gamma kernel), **MixInvNormGPD** (inverse Gaussian kernel), and **MixLogNormGPD** (lognormal kernel) — each combined with a GPD tail. All variants use Bayesian DPM inference for the bulk and a shared GPD component for exceedances above a covariate-dependent threshold, allowing assessment of robustness to bulk misspecification.
- **IPW**: an inverse-probability-weighted quantile regression estimator (Rosenbaum and Rubin, 1983), where arm-specific conditional quantiles are estimated using stabilized propensity score weights and combined to form CQTEs.
- **Extreme QR**: a weighted quantile regression estimator with inference adapted for extreme quantiles using extremal subsampling methods (Chernozhukov and Fernández-Val, 2011).
- **Skew-t**: a parametric log-skew-t regression model fitted separately by treatment arm, with CQTEs obtained as differences between fitted conditional quantiles (Morán-Vásquez et al., 2025).

For MixGPD, each arm is initialized with 5 mixture components, with weights below $\epsilon = 0.01$ removed. Posterior inference is performed via MCMC using `nimble`, with three chains and standard convergence diagnostics. To assess extreme behavior, we use

$$\tau \in \left\{0.25, 0.50, 0.75, 0.95, 1 - \frac{10}{n}, 1 - \frac{5}{n}, 1 - \frac{1}{n^{1.25}}, 1 - \frac{1}{n \log n}, 1 - \frac{1}{n^{3/2}}\right\}.$$

We construct evaluation points $\tilde{D}_x$ using coordinate-wise deciles of each covariate, yielding 9 representative covariate profiles. For each $(\tau, \tilde{\boldsymbol{x}})$, we compute the following:

$$\text{Bias} = q(\tau \mid \tilde{\boldsymbol{x}}) - \frac{1}{R}\sum_{r=1}^{R} \widehat{q}_r(\tau \mid \tilde{\boldsymbol{x}}),$$

$$\text{SD} = \left[\frac{1}{R-1}\sum_{r=1}^{R}\left(\widehat{q}_r(\tau \mid \tilde{\boldsymbol{x}}) - \overline{q}(\tau \mid \tilde{\boldsymbol{x}})\right)^2\right]^{1/2},$$

$$\text{CP} = \frac{1}{R}\sum_{r=1}^{R} \mathbb{I}\left\{\widehat{q}_r^{L}(\tau \mid \tilde{\boldsymbol{x}}) \le q(\tau \mid \tilde{\boldsymbol{x}}) \le \widehat{q}_r^{U}(\tau \mid \tilde{\boldsymbol{x}})\right\},$$

$$\text{AIW} = \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{q}_r^{U}(\tau \mid \tilde{\boldsymbol{x}}) - \widehat{q}_r^{L}(\tau \mid \tilde{\boldsymbol{x}})\right).$$

4.3. Simulation Results

The simulation results are summarized using heatmaps of bias and coverage probabilities (Figures 2 and 3). The figures provide insights not only into the stability of the estimator but also into the interaction between the bulk and the tail component of the estimator. Across all configurations, the MixGPD estimators show consistent behavior as the quantile level increases toward 1. The bias is small across central and moderately high quantiles and increases gradually towards the extreme quantiles. The Monte Carlo standard deviation also increases with the quantile level, reflecting the reduced amount of data available in the tail, but this increase is gradual rather than abrupt. The figures do not show any evidence of discontinuities or irregular patterns near the threshold, indicating that the transition between the bulk mixture component and the GPD tail is captured well by the model.

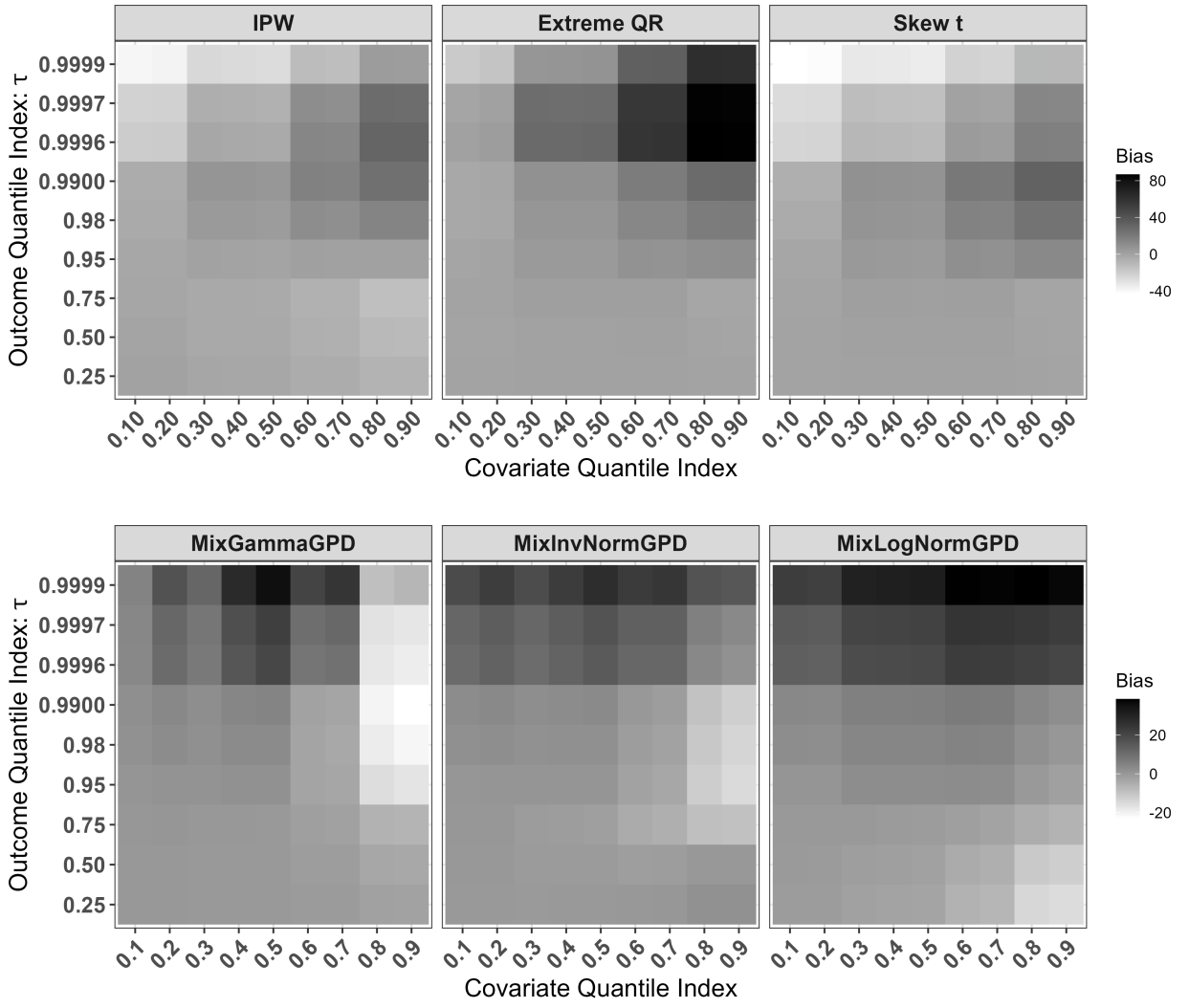

Figure 2: Bias of conditional quantile treatment effect estimates for $n = 500$. **Top panel:** Results for IPW, Extreme QR, and Skew t. **Bottom panel:** Results for MixGammaGPD, MixInvNormalGPD and MixLogNormGPD. Bias is evaluated across nine quantile indices of the outcome variable and nine coordinate-wise quantile based group of predictor variables. High intensity of red indicates high positive bias while intense blue indicates high negative bias. Light intensity colors indicate bias close to zero.

On the other hand, the Lognormal kernel reveals a more abrupt pattern of behavior. The bias increases steeply as the quantile index reaches unity, and the Monte Carlo standard deviations increase at a rate that exceeds the other specifications. The plots for the interval widths reveal a general tendency to expand the intervals, which are often disproportionately large. The smaller coverages at the extreme quantiles suggests that the Lognormal kernel does not mix well with the semiparametric structure of the tail. This may be due to the fact that the Lognormal kernel puts too much probability mass into the high-end bulk region. This led to unstable estimates of the parameters of the tail, a fact which can be readily observed in the simulation plots. This behavior is similar to the effects observed in earlier studies on semiparametric extreme value models. Excessive flexibility of the bulk component can impair the reliability of the extrapolation to the extremes.

Although all three mix-GPD estimators vary in terms of accuracy metrics, they all exhibit a similar qualitative behavior. In Figure 8, we observe that all three estimators behave as expected in terms of graudal increased variability as we approach a quantile index value of one. The most conservative intervals

are obtained with the Gamma kernel, whereas the Inverse Normal kernel provides narrower intervals for a small sacrifice in terms of coverage. The Lognormal kernel is the least reliable one, which is consistent with the tendency of the lognormal distribution to put substantial probability mass near the upper end of the bulk region, which interferes with the estimation of the tail component and leads to instability in the GPD fit (Bopp et al., 2021).

The benchmark estimators provide a useful comparison. The IPW estimator performs reasonably well for central quantiles but exhibits substantial variability in the tail, driven by instability in the estimated weights and a reduction in effective sample size. The Extreme QR approach is designed for tail estimation, but its performance becomes highly variable at the most extreme quantiles, with both point estimates and interval widths showing large fluctuations. The skew-t model captures moderate skewness in the data but does not adequately represent the tail behavior induced by the data-generating process, resulting in unreliable estimates at the extreme quantile indices.

Overall, the results demonstrate that accurate estimation of conditional quantile treatment effects in the tail requires explicit modeling of both the bulk and tail of the outcome distribution. The MixGPD approach achieves this by combining flexible modeling in the central region with a parametric tail specification, leading to lower bias, more controlled variability, and better-calibrated uncertainty across the full range of quantiles, including the extremes.

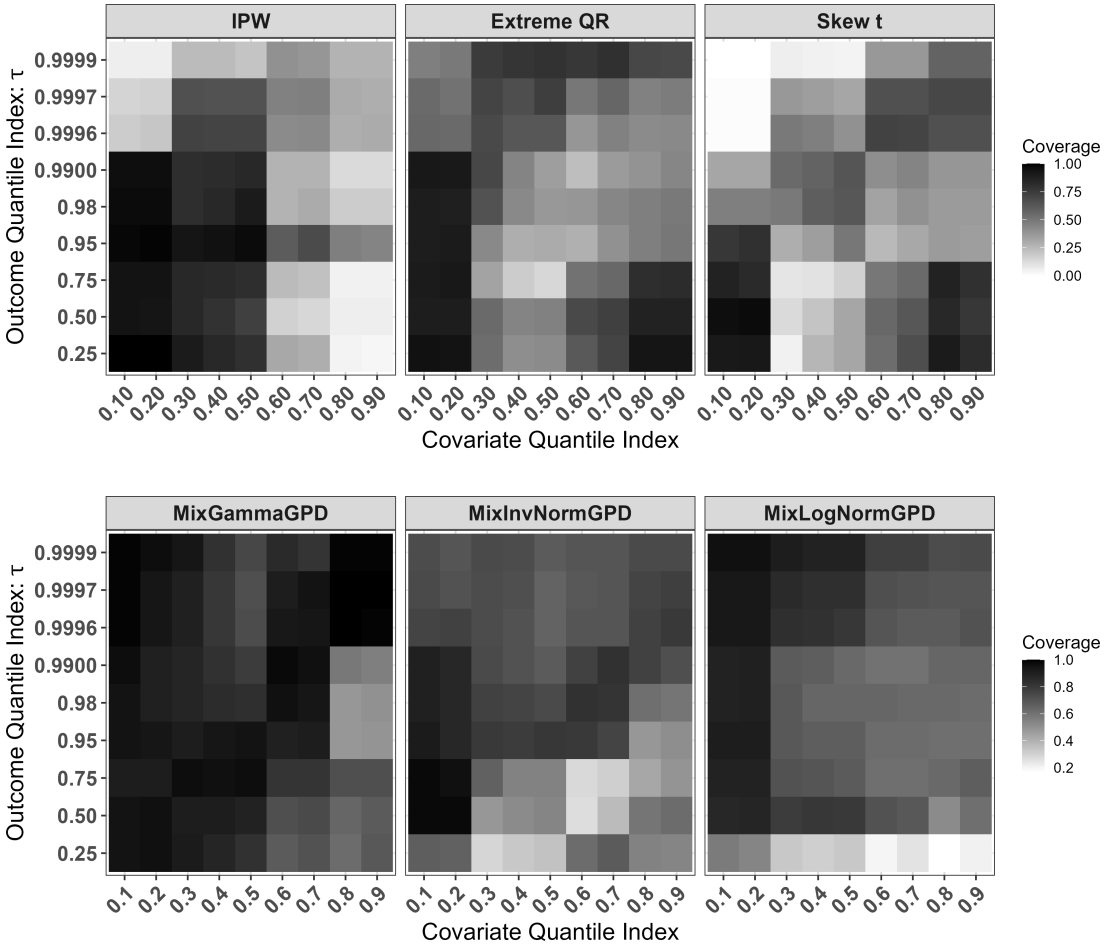

Figure 3: Empirical coverage of 90% posterior credible intervals or confidence intervals for $n = 500$. **Top panel:** Competing estimators including IPW, Extreme QR, and Skew t. **Bottom panel:** Comparison of MixGammaGPD, MixInvNormalGPD and MixLogNormGPD. Coverage is evaluated across nine quantile indices of the outcome variable and nine quantile-based groups of predictor variables. Red indicates a high coverage value and blue indicates a low coverage.

5. Real Data Analysis

To demonstrate the practical utility of the proposed method, we analyze an observational dataset in which estimating conditional quantile treatment effects (CQTEs) is of direct clinical interest. The goal is to examine how the effect of obesity varies across the distribution of glycemic outcomes and across covariate profiles. Obesity is a well-established risk factor for Type 2 Diabetes Mellitus (T2DM). Excess adiposity, particularly visceral fat, contributes to insulin resistance, which plays a central role in the development of T2DM. This relationship is well documented in both epidemiological and biomedical studies (Hu et al. (2001), Ng et al. (2021)). Projections indicate that by 2050, more than half of the global adult population may be overweight or obese, with a corresponding increase in the burden of T2DM (The Lancet, 2023). In the United States, current estimates suggest that approximately 11% of adults have diabetes, with projections rising to nearly 33% by 2050.

We study the relationship between obesity and long-term glycemic control, measured by glycosylated hemoglobin (HbA1c). This biomarker reflects average blood glucose levels over the preceding two to three months, and values above 7.0 are commonly used to indicate poor glycemic control. We use the `Diabetes` dataset available from the Vanderbilt Biostatistics Datasets, which contains data on 403 individuals and 19 covariates collected to study obesity, diabetes, and cardiovascular risk factors among African Americans in central Virginia. Variables with more than 50% missingness (repeat systolic and diastolic blood pressure) were excluded. Remaining missing values were imputed using multiple imputation by chained equations (MICE).

We define obesity status as the treatment variable and HbA1c as the outcome. Treating obesity as an exposure, we estimate conditional quantile treatment effects to characterize how its impact varies across the outcome distribution. In this dataset, approximately 30% of individuals have HbA1c levels exceeding 7.0, indicating substantial mass in the upper tail of the distribution.

The distribution of HbA1c is right-skewed with significant mass in the upper tail, and the difference between treatment groups is most pronounced at higher quantile levels. The proposed MixGPD approach is particularly suited for this setting, as it leverages the generalized Pareto distribution for the extreme tail, as supported by extreme value theory.

5.1. Analysis Framework

To evaluate heterogeneity in the effect of obesity across clinically relevant subpopulations, we construct a prediction grid varying four key covariates: age, BMI, gender, and stabilized glucose (`stab.glu`). The selected levels, summarized in Table 1, produce 162 unique covariate profiles. All remaining covariates are fixed at representative values, defined as the median for continuous variables and the mode for categorical variables. For each covariate profile, we estimate CQTE at nine quantile levels:

$$\tau \in \left\{0.25,\, 0.50,\, 0.75,\, 0.95,\, 1-\tfrac{10}{n},\, 1-\tfrac{5}{n},\, 1-\tfrac{1}{n^{1.25}},\, 1-\tfrac{1}{n\log n},\, 1-\tfrac{1}{n^{3/2}}\right\},$$

where $n = 403$. This grid spans central, upper, and extreme regions of the outcome distribution, allowing us to examine how treatment effects evolve toward the tail.

Covariate	Levels used in the grid
Age (years)	{34, 45, 60}
BMI (kg/m^2)	{24.09, 27.79, 32.26}
Gender (coded)	{male = 1, female = 2}
Stabilized glucose (`stab.glu`)	{81, 89, 106}

Table 1
Covariate levels used to construct the prediction grid. Levels reflect representative values across the range of each covariate.

We evaluate three MixGPD specifications: MixGammaGPD (Gamma kernel), MixInvNormGPD (inverse Gaussian kernel), and MixLogNormGPD (lognormal kernel). Each combines a Bayesian DPM bulk with a GPD tail, allowing us to assess robustness to bulk misspecification across the full outcome distribution, including extreme quantiles. Posterior inference is performed via MCMC using `nimble` with three chains;

Table 2
Point estimate and 90% credible interval for CQTE by age and quantile level, for MixInvNormGPD and MixGammaGPD. Each cell reports the posterior mean followed by the interval (lower, upper), averaged over the prediction grid sharing the same age, in units of HbA1c.

	MixInvNormGPD			MixGammaGPD		
Quantile (τ)	Age 34	Age 45	Age 60	Age 34	Age 45	Age 60
0.25	0.13 (-0.48, 0.79)	0.19 (-0.42, 0.85)	0.29 (-0.41, 1.01)	0.16 (-0.45, 0.77)	0.19 (-0.40, 0.79)	0.24 (-0.41, 0.91)
0.50	0.13 (-0.45, 0.76)	0.20 (-0.37, 0.81)	0.29 (-0.34, 0.97)	0.19 (-0.39, 0.78)	0.22 (-0.35, 0.79)	0.27 (-0.34, 0.89)
0.75	0.13 (-0.52, 0.84)	0.19 (-0.44, 0.87)	0.29 (-0.38, 1.00)	0.26 (-0.41, 0.95)	0.29 (-0.35, 0.94)	0.33 (-0.32, 1.00)
0.95	0.09 (-1.21, 1.52)	0.16 (-1.13, 1.55)	0.25 (-1.05, 1.62)	0.45 (-0.84, 1.87)	0.48 (-0.77, 1.86)	0.52 (-0.70, 1.88)
0.975	0.06 (-1.74, 2.01)	0.12 (-1.66, 2.05)	0.21 (-1.58, 2.11)	0.55 (-1.20, 2.47)	0.58 (-1.14, 2.46)	0.62 (-1.05, 2.48)
0.988	0.01 (-2.42, 2.66)	0.07 (-2.35, 2.69)	0.16 (-2.26, 2.74)	0.66 (-1.69, 3.24)	0.69 (-1.63, 3.24)	0.74 (-1.54, 3.24)
0.9994	-0.57 (-9.05, 8.51)	-0.51 (-8.94, 8.53)	-0.43 (-8.82, 8.57)	1.35 (-6.88, 10.68)	1.39 (-6.84, 10.70)	1.44 (-6.77, 10.74)
0.9996	-0.67 (-10.14, 9.52)	-0.62 (-10.05, 9.55)	-0.53 (-9.90, 9.58)	1.43 (-7.76, 11.93)	1.47 (-7.72, 11.97)	1.52 (-7.66, 12.01)
0.9999	-1.26 (-16.29, 14.97)	-1.21 (-16.27, 15.03)	-1.13 (-16.16, 15.03)	1.82 (-13.13, 18.89)	1.86 (-13.11, 18.93)	1.92 (-13.09, 18.97)

Table 3
Point estimate and 90% credible interval for CQTE by stabilized glucose (`stab.glu`) and quantile level, for MixInvNormGPD and MixGammaGPD. Each cell reports the posterior mean followed by the interval (lower, upper), averaged over the prediction grid sharing the same stab.glu value, in units of HbA1c.

	MixInvNormGPD			MixGammaGPD		
Quantile (τ)	stab.glu 81	stab.glu 89	stab.glu 106	stab.glu 81	stab.glu 89	stab.glu 106
0.25	0.18 (-0.41, 0.82)	0.20 (-0.42, 0.86)	0.23 (-0.47, 0.97)	0.19 (-0.39, 0.77)	0.19 (-0.41, 0.80)	0.21 (-0.47, 0.89)
0.50	0.18 (-0.38, 0.79)	0.20 (-0.38, 0.82)	0.24 (-0.41, 0.93)	0.22 (-0.34, 0.78)	0.23 (-0.35, 0.81)	0.24 (-0.39, 0.87)
0.75	0.18 (-0.45, 0.86)	0.20 (-0.44, 0.88)	0.23 (-0.44, 0.97)	0.29 (-0.35, 0.93)	0.29 (-0.35, 0.95)	0.31 (-0.38, 1.01)
0.95	0.14 (-1.15, 1.55)	0.16 (-1.13, 1.56)	0.20 (-1.10, 1.58)	0.47 (-0.78, 1.85)	0.48 (-0.77, 1.86)	0.49 (-0.77, 1.89)
0.975	0.11 (-1.69, 2.05)	0.13 (-1.66, 2.05)	0.16 (-1.61, 2.06)	0.57 (-1.14, 2.46)	0.58 (-1.13, 2.47)	0.59 (-1.12, 2.48)
0.988	0.06 (-2.38, 2.70)	0.08 (-2.35, 2.70)	0.11 (-2.30, 2.69)	0.69 (-1.63, 3.23)	0.69 (-1.62, 3.24)	0.71 (-1.61, 3.25)
0.9994	-0.52 (-9.00, 8.57)	-0.51 (-8.96, 8.55)	-0.48 (-8.85, 8.49)	1.38 (-6.84, 10.69)	1.39 (-6.83, 10.70)	1.40 (-6.81, 10.71)
0.9996	-0.63 (-10.10, 9.59)	-0.61 (-10.05, 9.56)	-0.58 (-9.95, 9.49)	1.46 (-7.73, 11.96)	1.47 (-7.72, 11.97)	1.49 (-7.70, 11.98)
0.9999	-1.22 (-16.32, 15.07)	-1.21 (-16.26, 15.03)	-1.18 (-16.14, 14.93)	1.85 (-13.14, 18.93)	1.86 (-13.12, 18.93)	1.88 (-13.07, 18.92)

MCMC convergence diagnostics (traceplots and effective sample sizes) for the diabetes application are provided in Appendix A.2 (Figures 4–6).

5.2. Analysis Results

We assess how obesity influences the conditional distribution of HbA1c by estimating CQTEs across a sampled grid of quantiles, extending into the extreme upper tail. We focus on three clinically relevant covariates: age, stabilized glucose, and gender. The estimated CQTEs and 90% credible intervals are summarized in Tables 2–4.

Before examining individual covariates, several general patterns emerge. Uncertainty increases sharply for $\tau > 0.99$, reflecting limited information in the extreme tail. The MixGPD specifications respond differently to this sparsity. The Gamma kernel produces progressively larger tail effects, with interval widths increasing steadily across quantiles. The Inverse-Normal kernel yields smooth and gradual growth in both estimates and uncertainty. In contrast, the Lognormal kernel exhibits pronounced variability at high quantiles, leading to inflated and less stable intervals.

Table 2 shows that the effect of obesity on HbA1c increases with both age and quantile level. Differences across age groups are modest in the central region but widen substantially for τ greater than 0.95, where older individuals exhibit markedly larger CQTEs. Interval widths increase with τ for all methods, though the rate of increase varies by model. The Inverse-Normal kernel produces smooth and stable growth, whereas the Gamma kernel yields larger tail estimates and wider intervals. The Lognormal kernel shows increased variability at high quantiles.

Table 3 reports the corresponding 90% credible intervals stratified by stabilized glucose. The stabilized-glucose stratification reveals a strong monotonic relationship between baseline glucose levels and CQTE magnitude. Individuals with higher stabilized glucose exhibit larger treatment effects across almost all quantiles, with differences becoming pronounced in the upper tail. This indicates that obesity has a higher

Table 4
Point estimate and 90% credible interval for CQTE by gender and quantile level, for MixInvNormGPD and MixGammaGPD. Each cell reports the posterior mean followed by the interval (lower, upper), averaged over the prediction grid sharing the same gender, in units of HbA1c.

	MixInvNormGPD		MixGammaGPD	
Quantile (τ)	Female	Male	Female	Male
0.25	0.33 (-0.17, 0.86)	0.07 (-0.70, 0.91)	0.30 (-0.18, 0.81)	0.09 (-0.66, 0.84)
0.50	0.33 (-0.14, 0.82)	0.08 (-0.64, 0.88)	0.33 (-0.13, 0.80)	0.13 (-0.59, 0.85)
0.75	0.33 (-0.23, 0.90)	0.08 (-0.66, 0.90)	0.40 (-0.17, 0.97)	0.19 (-0.55, 0.95)
0.95	0.29 (-0.99, 1.66)	0.04 (-1.27, 1.47)	0.59 (-0.66, 1.93)	0.38 (-0.89, 1.80)
0.975	0.26 (-1.53, 2.17)	0.01 (-1.78, 1.94)	0.69 (-1.03, 2.55)	0.47 (-1.23, 2.38)
0.988	0.21 (-2.23, 2.82)	-0.04 (-2.46, 2.57)	0.81 (-1.52, 3.34)	0.58 (-1.72, 3.14)
0.9994	-0.37 (-8.83, 8.67)	-0.63 (-9.04, 8.41)	1.52 (-6.75, 10.81)	1.26 (-6.91, 10.59)
0.9996	-0.48 (-9.93, 9.69)	-0.74 (-10.14, 9.41)	1.61 (-7.62, 12.09)	1.34 (-7.81, 11.85)
0.9999	-1.07 (-16.13, 15.17)	-1.33 (-16.35, 14.85)	2.01 (-12.96, 19.01)	1.72 (-13.26, 18.84)

impact on glycemic control among individuals with poorer baseline metabolic status. Uncertainty also increases with glucose level, especially at extreme quantiles. The Inverse-Normal kernel shows stable increase in uncertainty, the Gamma kernel amplifies tail effects, and the Lognormal kernel introduces additional variability.

Table 4 reports the corresponding 90% credible intervals stratified by gender. The gender-stratified analysis shows consistently larger obesity effects among women, particularly for τ greater than 0.95, while men exhibit more moderate increases. These differences indicate meaningful gender heterogeneity in upper-tail outcomes. Interval widths increase sharply across quantile levels, but the mixture models provide smooth and interpretable scaling. The Inverse-Normal kernel again yields the most stable behavior, the Gamma kernel emphasizes tail amplification, and the Lognormal kernel exhibits increased volatility.

Across all stratifications, a consistent pattern emerges: the effect of obesity on HbA1c strengthens toward the upper tail and is most pronounced among higher-risk groups, including older individuals, those with elevated glucose levels, and women. The MixGPD estimators, particularly with the Inverse-Normal specification, deliver stable and interpretable estimates across the full distribution. These results highlight the importance of flexible bulk–tail modeling for capturing heterogeneous treatment effects in extreme outcome regions.

The statistical significance of the CQTEs is evaluated using the 90% credible intervals in Table 2, Table 3, and Table 4. As shown in Table 2, credible intervals for both intermediate ($\tau = 0.25, 0.50, 0.75$) and higher ($\tau \geq 0.95$) quantiles all contain zero across all age groups, indicating that no quantile treatment effect reaches statistical significance at the 90% credible interval level. Nonetheless, the point estimates at higher quantiles tend to be positive, particularly for older individuals, suggesting a trend toward increased HbA1c with obesity in the upper tail. A similar pattern is observed in Table 3: while point estimates grow with baseline glucose level and quantile level, all credible intervals contain zero, so these effects remain statistically inconclusive. Table 4 likewise shows larger positive point estimates among females at high quantiles, but all credible intervals again straddle zero.

6. Discussion

This paper develops a semiparametric framework for estimating CQTEs across both central and extreme regions of the outcome distribution. By combining a Dirichlet process mixture (DPM) model for the bulk with a generalized Pareto distribution (GPD) for the tail, the proposed MixGPD approach models the entire conditional distribution while allowing a smooth transition into regions where data are sparse. This is particularly important in settings where upper quantiles are of primary interest, such as clinical risk assessment, and where standard methods either impose restrictive tail assumptions or fail to provide reliable uncertainty quantification near the boundary of the observed support.

The simulation study demonstrates consistent performance gains over existing methods. Across a range of sample sizes, distributional settings, and quantile levels, the MixGPD estimators exhibit lower bias, reduced variability, and higher interval coverage relative to IPW, Extreme QR, and Skew-t approaches. These benchmark methods do not explicitly model exceedances or tail behavior, which leads to unreliable estimates as the quantile level approaches the tail. In contrast, the MixGPD framework jointly models the bulk and tail, with tail parameters estimated directly from exceedances. This enables stable extrapolation at high quantiles without imposing parametric assumptions on the full distribution.

The comparison of mixture kernels highlights the role of the bulk model in tail inference. The Gamma mixture performs well across a range of settings but can allocate excess mass near the upper boundary, inflating uncertainty when the true distribution is light-tailed. The Lognormal mixture exhibits greater variability in the upper tail, resulting in wider credible intervals. The Inverse–Normal kernel produces smoother transitions into the tail and more controlled growth in uncertainty. These findings indicate that tail estimation depends not only on the GPD component but also on how well the bulk model captures the distribution below the threshold.

The real data analysis of HbA1c reinforces these findings. Across stratifications by age, stabilized glucose, and gender, the effect of obesity increases toward the upper tail, indicating stronger effects among higher-risk individuals. Standard methods either attenuate these patterns or become unstable in the tail due to limited data support. The MixGPD approach yields smooth and interpretable CQTE estimates, with uncertainty increasing in line with the decreasing number of effective observations. Among the mixture specifications, the Inverse–Normal kernel provides the most stable performance, while the Gamma and Lognormal kernels introduce varying degrees of tail amplification and variability.

Several limitations should be noted. Although the threshold is estimated within the model, inference at extreme quantiles is sensitive to threshold placement, prior specification, and the number of exceedances. Computational cost also increases with the complexity of the mixture representation and the number of components required to capture skewness or multimodality.

In addition, The framework can be naturally extended to lower-tail inference. By applying the same methodology to a reflected outcome, extreme lower-tail quantiles can be estimated using the upper-tail construction and mapped back to the original scale.

Overall, this work provides a novel and flexible approach for estimating heterogeneous treatment effects across the full outcome distribution. By combining Bayesian nonparametric modeling with extreme value theory, the proposed framework yields interpretable inference in regions where standard methods are unreliable. This work highlights that accurate estimation of extreme quantile treatment effects requires careful alignment between bulk and tail components, particularly in applications where tail behavior has direct scientific or clinical importance.

Data availability

The `Diabetes` dataset used in Section 5 is publicly available from the Vanderbilt Biostatistics Datasets (https://hbiostat.org/data/), as cited in the main text.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

References

Blette, B.S., Granholm, A., Li, F., Shankar-Hari, M., Lange, T., Munch, M.W., Møller, M.H., Perner, A., Harhay, M.O., 2023. Causal Bayesian machine learning to assess treatment effect heterogeneity by dexamethasone dose for patients with COVID-19 and severe hypoxemia. Scientific Reports 13, 1–10. doi:10.1038/s41598-023-33425-3.

Bopp, A., Shaby, B.A., Huser, R., 2021. Semiparametric modelling of covariate-dependent extremes. Extremes 24, 373–404.

Cabras, S., Castellanos, M.E., 2011. A Bayesian approach for estimating extreme quantiles under a semiparametric mixture model. ASTIN Bulletin: The Journal of the IAA 41, 87–106. doi:10.2143/AST.41.1.2084387.

Chernozhukov, V., Fernández-Val, I., 2011. Inference for extremal conditional quantile models, with an application to market and birthweight risks. The Review of Economic Studies 78, 559–589. URL: https://doi.org/10.1093/restud/rdq020, doi:10.1093/restud/rdq020.

Chernozhukov, V., Hansen, C., 2005. An iv model of quantile treatment effects. Econometrica 73, 245–261.

Deuber, M., et al., 2022. Estimating extreme quantiles in the presence of measurement error. Electronic Journal of Statistics 16, 2074–2101.

Firpo, S., 2007. Efficient semiparametric estimation of quantile treatment effects. Econometrica 75, 259–276. doi:10.1111/j.1468-0262.2007.00738.x.

Folks, J.L., Chhikara, R.S., 1978. The inverse gaussian distribution and its statistical application—a review. Journal of the Royal Statistical Society: Series B (Methodological) 40, 263–275. URL: https://doi.org/10.1111/j.2517-6161.1978.tb01039.x, doi:10.1111/j.2517-6161.1978.tb01039.x.

Fúquene Patiño, J.A., 2015. A semi-parametric Bayesian extreme value model using a Dirichlet process mixture of gamma densities. Journal of Applied Statistics 42, 267–280. URL: https://doi.org/10.1080/02664763.2014.947357, doi:10.1080/02664763.2014.947357.

de Haan, L., Ferreira, A., 2007. Extreme Value Theory: An Introduction. Springer Science & Business Media.

Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00442, doi:10.1111/1468-0262.00442.

Hu, F.B., Manson, J.E., Stampfer, M.J., Colditz, G., Liu, S., Solomon, C.G., Willett, W.C., 2001. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. New England Journal of Medicine 345, 790–797.

Imbens, G.W., 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. The Review of Economics and Statistics 86, 4–29. doi:10.1162/003465304323023651.

Li, F., Ding, P., Mealli, F., 2022. Bayesian causal inference: a critical review. arXiv preprint arXiv:2206.15460 URL: https://arxiv.org/abs/2206.15460.

Lopez, M.J., Gutman, R., 2017. Estimation of causal effects with multiple treatments: a review and new ideas. Statistical Science 32, 432–454. doi:10.1214/17-STS612.

Morán-Vásquez, R.A., Giraldo-Melo, A.D., Mazo-Lopera, M.A., 2025. Quantile estimation based on the Log-Skew-t linear regression model: Statistical aspects, simulations, and applications. Stats 8, 58. URL: https://doi.org/10.3390/stats8030058, doi:10.3390/stats8030058.

do Nascimento, F.F., Gamerman, D., Lopes, H.F., 2011. Peak over threshold modeling for estimating extreme quantiles. Journal of Statistical Computation and Simulation 81, 401–415.

Ng, A.C., Delgado, V., Borlaug, B.A., Bax, J.J., 2021. Diabesity: the combined burden of obesity and diabetes on heart disease and the role of imaging. Nature Reviews Cardiology 18, 291–304.

Oganisian, A., Roy, J.A., 2021. A practical introduction to Bayesian estimation of causal effects: parametric and nonparametric approaches. Statistics in Medicine 40, 518–551.

Pickands, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3, 119–131.

Powell, D., 2020. Quantile treatment effects in the presence of covariates. Review of Economics and Statistics 102, 994–1005.

Ray, K., Szabó, B., 2019. Debiased Bayesian inference for average treatment effects. URL: https://arxiv.org/abs/1909.12078, arXiv:1909.12078.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

The Lancet, 2023. Diabetes: a defining disease of the 21st century. The Lancet 401, 2087. URL: https://doi.org/10.1016/S0140-6736(23)01296-5, doi:10.1016/S0140-6736(23)01296-5.

Venturini, S., Alfo, M., et al., 2015. Quantile regression for longitudinal data based on a latent factor. Statistics in Medicine 34, 1646–1657.

Xu, D., 2018. A Dirichlet process mixture model for causal inference with many covariates. Journal of Causal Inference 6, 20180006.

Zhang, Y., et al., 2018. Estimating quantile treatment effects in the presence of competing risks. Journal of the American Statistical Association 113, 137–149.

Zigler, C.M., 2016. The central role of Bayes' theorem for joint estimation of causal effects and propensity scores. The American Statistician 70, 47–54. doi:10.1080/00031305.2015.1111260.

A. Appendices

A.1. MCMC Sampling Algorithm for Model Parameters

To explore the joint posterior distributions defined in equations (5) and (8), we employ a blocked Metropolis–Hastings (MH) within–Gibbs sampler. Each iteration begins with an update of the propensity score regression coefficients $\boldsymbol{\beta}_{ps}$ from the design stage, followed by a sequence of updates for the outcome-stage parameter blocks

$$\Theta^{(a)} = \left\{ \boldsymbol{\beta}_{\mu}^{(a)}, \boldsymbol{\eta}^{(a)}, \boldsymbol{\beta}_{u}^{(a)}, \sigma^{(a)}, \xi^{(a)} \right\}, \tag{10}$$

as defined above, together with the associated bulk location parameters $\{\mu_{ij}^{(a)}\}$. These updates are applied separately for each treatment arm $a \in \{0, 1\}$. All likelihood contributions correspond to the expressions in equations (4) and (6). We adopt the prior distributions detailed in Section 3.3 and the hierarchical structure specified in (3). We will denote samples from proposal distributions using asterisk superscript (*) over original parameter symbol.

A.1.1. Design-stage update

Given the data $(\boldsymbol{X}, \boldsymbol{A})$, we update the vector $\boldsymbol{\beta}_{ps} \in \mathbb{R}^p$ using a standard random-walk MH step.

1. We first propose

$$\boldsymbol{\beta}_{ps}^{*} \sim \mathcal{N}_p\left(\boldsymbol{\beta}_{ps}^{(m-1)}, \boldsymbol{S}_{ps}\right), \tag{11}$$

   where $\boldsymbol{S}_{ps}$ is a user-specified $p \times p$ tuning matrix, and $\boldsymbol{\beta}_{ps}^{(m-1)}$ represents the current (i.e., previous) state of the Markov chain.

2. Using the log-posterior kernel from equation (5), we compute the acceptance probability

$$a_{\text{ps}} = \min\left\{1, \exp\left(\log p\big(\boldsymbol{\beta}_{ps}^{*} \mid \boldsymbol{X}, \boldsymbol{A}\big) - \log p\Big(\boldsymbol{\beta}_{ps}^{(m-1)} \mid \boldsymbol{X}, \boldsymbol{A}\Big)\right)\right\}. \tag{12}$$

3. We accept the proposed value with probability a_{ps}. Specifically, we set $\boldsymbol{\beta}_{ps}^{(m)} = \boldsymbol{\beta}_{ps}^{*}$ if the proposal is accepted and $\boldsymbol{\beta}_{ps}^{(m)} = \boldsymbol{\beta}_{ps}^{(m-1)}$ otherwise. The fitted propensity scores $\widehat{\rho}_{x_i}$ and the corresponding augmented predictors $\boldsymbol{r(x_i)}$ are then recomputed according to equations (2) and (3).

A.1.2. Outcome-stage updates

For each treatment arm $a \in \{0, 1\}$, we update the parameter blocks in $\Theta^{(a)}$ and the bulk locations $\{\mu_{ij}^{(a)}\}$, conditioning on the current value of $\boldsymbol{\beta}_{ps}$ and keeping the parameters for the other arm fixed. Let

$$\log p\Big(\Theta^{(1)}, \Theta^{(0)} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{A}\Big)$$

denote the posterior kernel in equation (8). A generic MH update for any scalar or vector sub-block $\theta \subset \Theta^{(a)}$ uses the following acceptance probability:

$$a\Big(\theta^{*}, \theta^{(m-1)}\Big) = \min\left\{1, \exp\left(\log p\Big(\Theta^{(1)*}, \Theta^{(0)*} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{A}\Big) - \log p\Big(\Theta^{(1)}, \Theta^{(0)} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{A}\Big)\right)\right\}, \tag{13}$$

where only the block θ is replaced with θ^{*} in $\Theta^{(a)*}$, and all other components remain unchanged at their current values.

Each outcome-stage sweep for arm a proceeds through the following updates.

1. *Dirichlet process concentration and stick-breaking weights.* Working with the truncated representation in equation (3), we fix a truncation level of J mixture components. Let $\alpha^{(a)}$ denote the concentration parameter, and denote the stick-breaking components and corresponding weights as $\{v^{(a,j)}, w^{(a,j)}\}_{j=1}^{J}$.

   (a) Each $v^{(a,j)}$ is updated from its full conditional distribution, which combines the Beta prior (Section 3.3) with the mixture allocations implied by equation (6). The resulting conditional

distribution remains in the Beta family, with shape parameters adjusted according to the current component-specific counts. The weights $w^{(a,j)}$ are then reconstructed using the standard stick-breaking formula in equation (3).

(b) The parameter $\alpha^{(a)}$ is updated via a one-dimensional MH step. We adopt the proposal

$$\alpha^{(a)*} \sim \mathcal{N}\Big(\alpha^{(a)(m-1)}, s_\alpha^2\Big)\, \mathcal{I}\Big\{\alpha^{(a)*} > 0\Big\}, \tag{14}$$

with acceptance probability computed using equation (13).

2. *Bulk locations $\mu_i^{(a,j)}$, regression coefficients $\boldsymbol{\beta}_\mu^{(a,j)}$, and kernel parameters $\eta^{(a,j)}$.* For each mixture component $j = 1, \ldots, J$ and each observation with $a_i = a$, the bulk kernel induces a location parameter $\mu_{ij}^{(a)}$, whose distribution depends on the chosen bulk family. Conditionally on the linear predictor $\boldsymbol{r}(\boldsymbol{x_i})^\top \boldsymbol{\beta}_\mu^{(a,j)}$, we specify the posterior sample form bulk location parameter as

$$\mu_i^{(a,j)} \mid \boldsymbol{r}(\boldsymbol{x_i}), \boldsymbol{\beta}_\mu^{(a,j)} = \begin{cases} \exp\left(\boldsymbol{r}(\boldsymbol{x_i})^\top \boldsymbol{\beta}_\mu^{(a,j)}\right), & \text{Gamma or inverse-normal bulk,} \\ \boldsymbol{r}(\boldsymbol{x_i})^\top \boldsymbol{\beta}_\mu^{(a,j)}, & \text{lognormal bulk.} \end{cases}$$

Each $\mu_i^{(a,j)}$ is deterministic and doesn't need MCMC steps.

The regression coefficients $\boldsymbol{\beta}_\mu^{(a,j)} \in \mathbb{R}^{p+1}$ are updated using a multivariate random-walk proposal,

$$\boldsymbol{\beta}_\mu^{(a,j)*} \sim \mathcal{N}_{p+1}\Big(\boldsymbol{m}_{\beta_\mu}^{(a,j)(m-1)}, \boldsymbol{S}_\mu\Big), \tag{15}$$

with acceptance probability computed via equation (13), where the likelihood contribution is mediated through the variables $\mu_i^{(a,j)}$ and the kernel-specific bulk density.

The additional kernel parameter $\eta^{(a,j)}$ (shape or scale, depending on the bulk family) is updated using the scalar random-walk MH proposal

$$\eta^{(a,j)*} \sim \mathcal{G}(\alpha_\eta, \beta_\eta), \tag{16}$$

with acceptance computed using equation (13) and the Gamma or inverse-Gamma prior specified in Section 3.3, depending on whether the bulk kernel is Gamma, inverse-normal, or lognormal.

3. *Threshold regression coefficients $\boldsymbol{\beta}_u^{(a)}$.* The regression vector $\boldsymbol{\beta}_u^{(a)}$, which determines the covariate-dependent threshold surface, is updated using a multivariate random-walk proposal,

$$\boldsymbol{\beta}_u^{(a)*} \sim \mathcal{N}_{p+1}\Big(\boldsymbol{m}_{\beta_u}^{(a)(m-1)}, \boldsymbol{S}_u\Big), \tag{17}$$

where the linear predictor $\boldsymbol{r}(\boldsymbol{x_i})^\top \boldsymbol{\beta}_u^{(a)}$ enters the lognormal model for the threshold as specified in equation (3).

4. *Covariate-dependent thresholds $u^{(a)}(\boldsymbol{x}_i)$.* For observations in arm a (i.e., $A_i = a$), the threshold is modeled directly through a lognormal regression surface, with log-mean given by the linear predictor and variance parameter $s_u^{(a)}$:

$$u_i^{(a)} \equiv u^{(a)}(\boldsymbol{x}_i) \sim \mathcal{LN}\Big(\boldsymbol{r}(\boldsymbol{x}_i)^\top \boldsymbol{\beta}_u^{(a)},\ (s_u^{(a)})^2\Big),$$

To enforce the empirical lower bound defined in Section (3.3), we truncate below at $l_{u^{(a)}}(x_i)$ (the arm-specific empirical percentile), yielding the prior

$$u_i^{(a)} \sim \mathcal{LN}\Big(\boldsymbol{r}(\boldsymbol{x}_i)^\top \boldsymbol{\beta}_u^{(a)}),\ (s_u^{(a)})^2\Big)\, \mathcal{I}\Big\{u_i^{(a)} \geq l_{u^{(a)}}(\boldsymbol{x}_i)\Big\}.$$

We update $u_i^{(a)}$ using a random-walk Metropolis step on the natural scale with a lognormal proposal centered at the current value:

$$u_i^{(a)*} \sim \mathcal{LN}\Big(\log u_i^{(a)(m-1)},\, s^2_{\text{prop}}\Big)\, \mathcal{I}\Big\{u_i^{(a)*} \geq l_{u^{(a)}}(\boldsymbol{x}_i)\Big\}. \tag{18}$$

The acceptance probability is calculated using (13).

5. *GPD scale parameter* $\sigma^{(a)}$. The scale parameter $\sigma^{(a)}$ is updated via a positive-constrained random-walk proposal,

$$\sigma^{(a)*} \sim \mathcal{G}(\alpha_\sigma, \beta_\sigma), \tag{19}$$

with acceptance determined by the contribution of the generalized Pareto likelihood for observations exceeding the threshold and the prior on $\sigma^{(a)}$ given in Section 3.3.

6. *GPD shape parameter* $\xi^{(a)}$. Finally, the tail index $\xi^{(a)}$ is updated via a truncated normal random-walk proposal:

$$\xi^{(a)*} \sim \mathcal{N}\Big(\xi^{(a)(m-1)}, s^2_\xi\Big)\, \mathcal{I}\Big\{\xi^{(a)*} > -0.5\Big\}, \tag{20}$$

which enforces the regularity constraint implied by equation (3.2) and the prior specification in Section 3.3. The acceptance probability is again given by equation (13).

Each complete MCMC iteration therefore consists of the design-stage update described in equations (11)–(12), followed by the outcome-stage updates for $a = 0, 1$, using the generic acceptance rule in equation (13) and the posterior kernel in equation (8).

A.2. MCMC Convergence Diagnostics for the Diabetes Application

We assess MCMC convergence for the diabetes application using traceplots and effective sample sizes (ESS) for three parameters: the GPD shape index ξ, the GPD scale σ, and the threshold $\hat{u}$ evaluated at the arm-specific median covariate profiles. MCMC setup were described in Section 5. Traceplots for the MixGammaGPD and MixLogNormGPD fits are shown in Figures 4 and 5, respectively. In both cases, all three parameters exhibit good mixing across the three chains, with no evidence of divergence or multimodality. The GPD shape parameter ξ shows stable behavior around values close to zero, indicating light-tailed behavior, while the scale parameter σ and threshold $\hat{u}$ also show consistent mixing patterns. Effective sample sizes for the same parameters are reported in Figure 6.

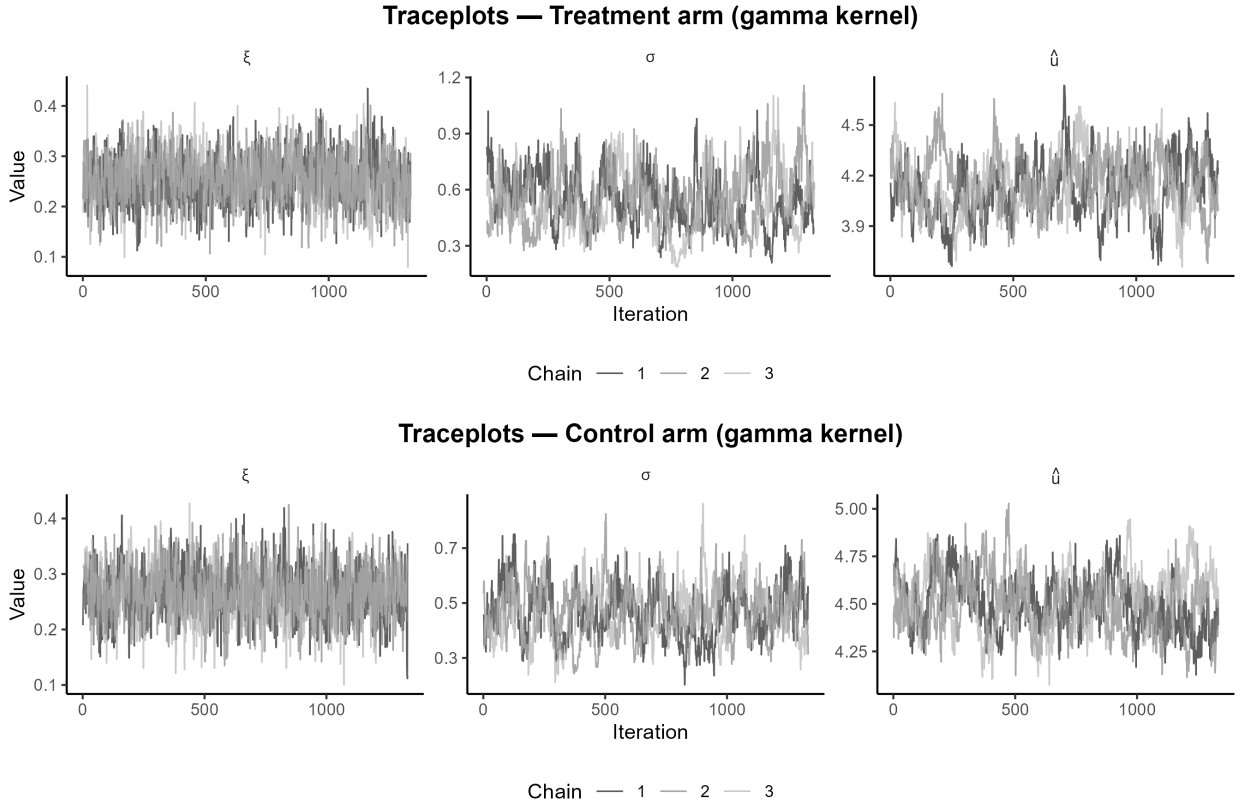

Figure 4: Traceplots for the MixGammaGPD fit to the diabetes data. Top: treatment arm. Bottom: control arm. Columns correspond to ξ (GPD shape), σ (GPD scale), and $\hat{u}$ (threshold at the arm-specific median covariate profile).

A.3. Supplementary Simulation Figures

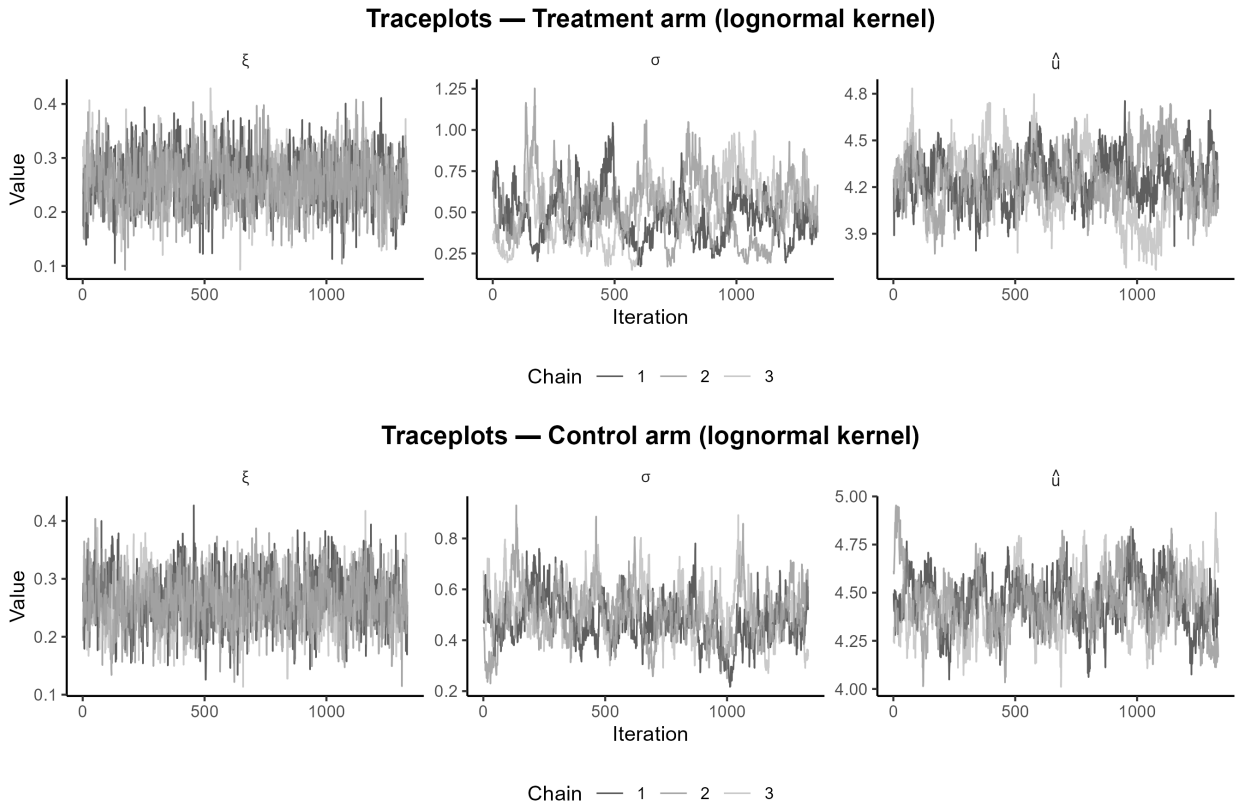

Figure 5: Traceplots for the MixLogNormGPD fit to the diabetes data. Top: treatment arm. Bottom: control arm. Columns correspond to ξ (GPD shape), σ (GPD scale), and $\hat{u}$ (threshold at the arm-specific median covariate profile).

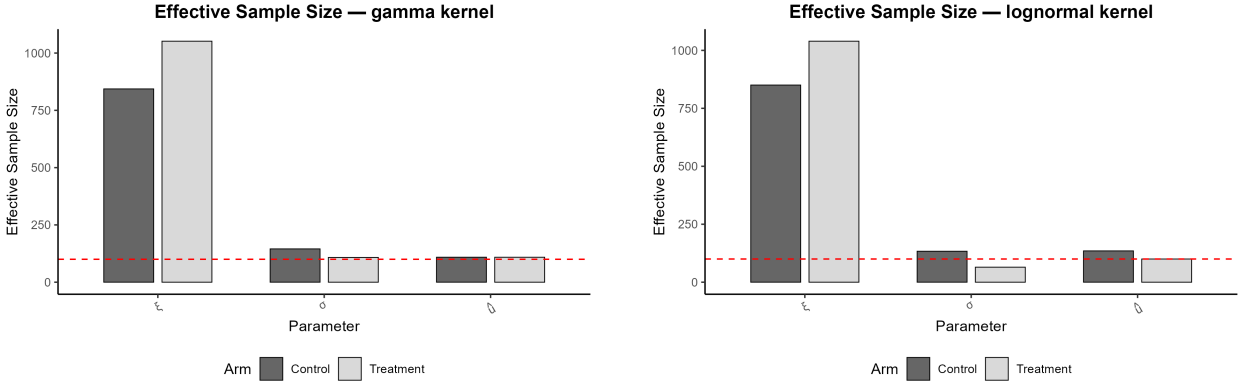

Figure 6: Effective sample sizes (ESS) for the MixGammaGPD (left) and MixLogNormGPD (right) fits to the diabetes data. Bars are shown separately for the treatment and control arms. The dashed red line marks ESS $= 100$. All three parameters (ξ, σ, and $\hat{u}$) achieve ESS well above this threshold for both arms.

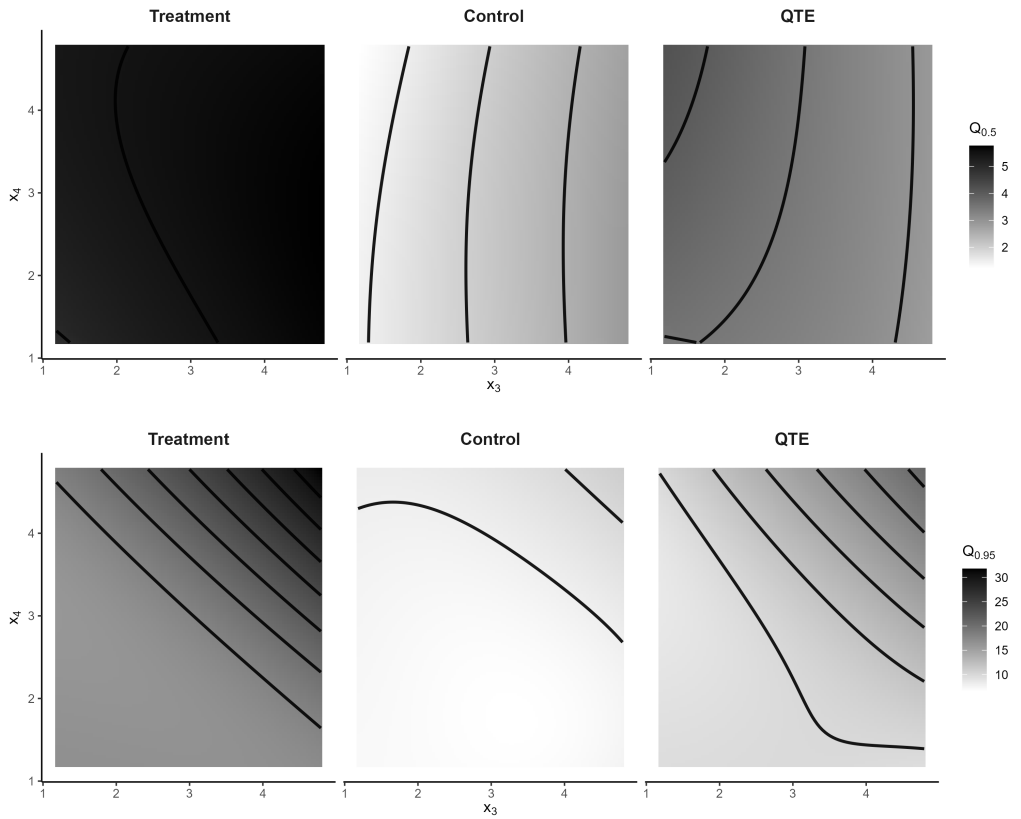

Figure 7: Conditional quantile surfaces over the covariate pair (x_3, x_4) with the remaining covariates fixed at $x_1 = 2$ and $x_2 = 3$. The top display corresponds to $\tau = 0.50$ and the bottom display corresponds to $\tau = 0.95$. Within each display, the treatment surface, control surface, and resulting QTE surface are shown side by side. Color shading represents the magnitude of the conditional quantile, with lighter regions corresponding to larger values and darker regions corresponding to smaller values. White contour curves connect locations with equal quantile or QTE values and therefore document the strength and direction of local curvature across the covariate space.

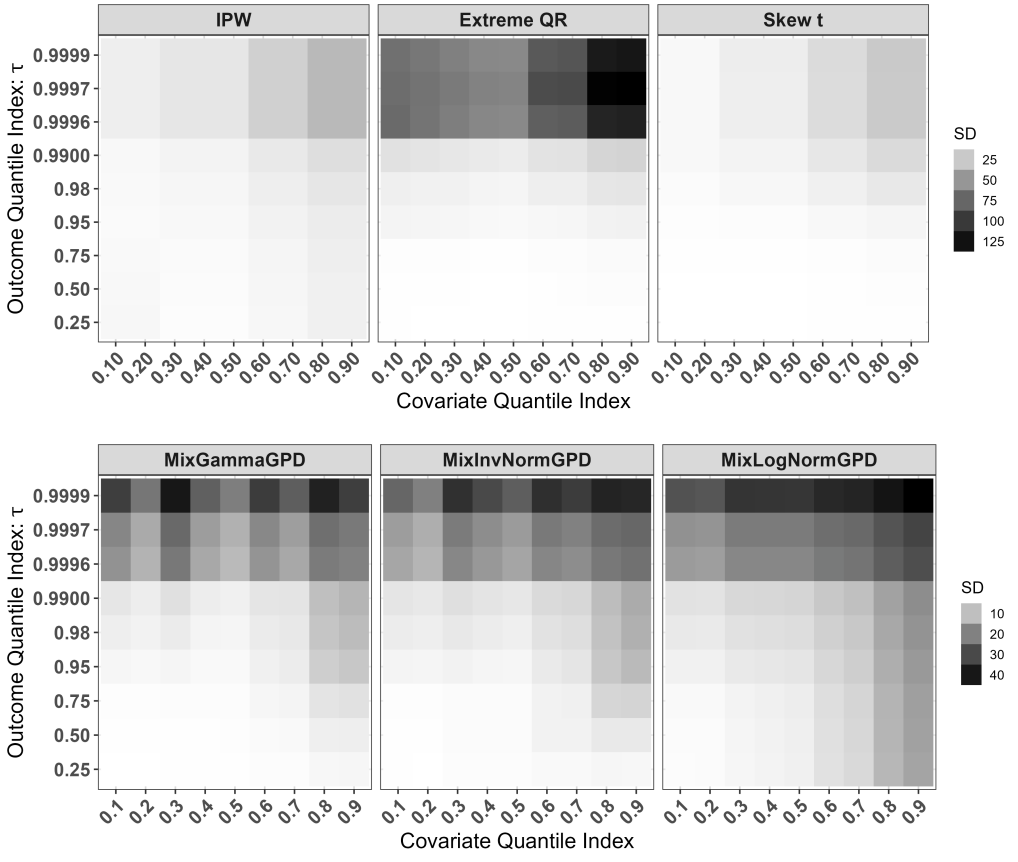

Figure 8: Standard-deviation heatmaps for conditional quantile treatment effect estimates when $n = 500$. The top panel reports the three competing methods, namely IPW, Extreme QR, and Skew t. The bottom panel reports the three proposed bulk specifications, MixGammaGPD, MixInvNormGPD, and MixLogNormGPD. In each heatmap, columns index the nine covariate-grid profiles and rows index the nine quantile levels. Color intensity increases from blue to red as Monte Carlo variability increases, so warm colors identify regions where repeated-sample CQTE estimates fluctuate more strongly.

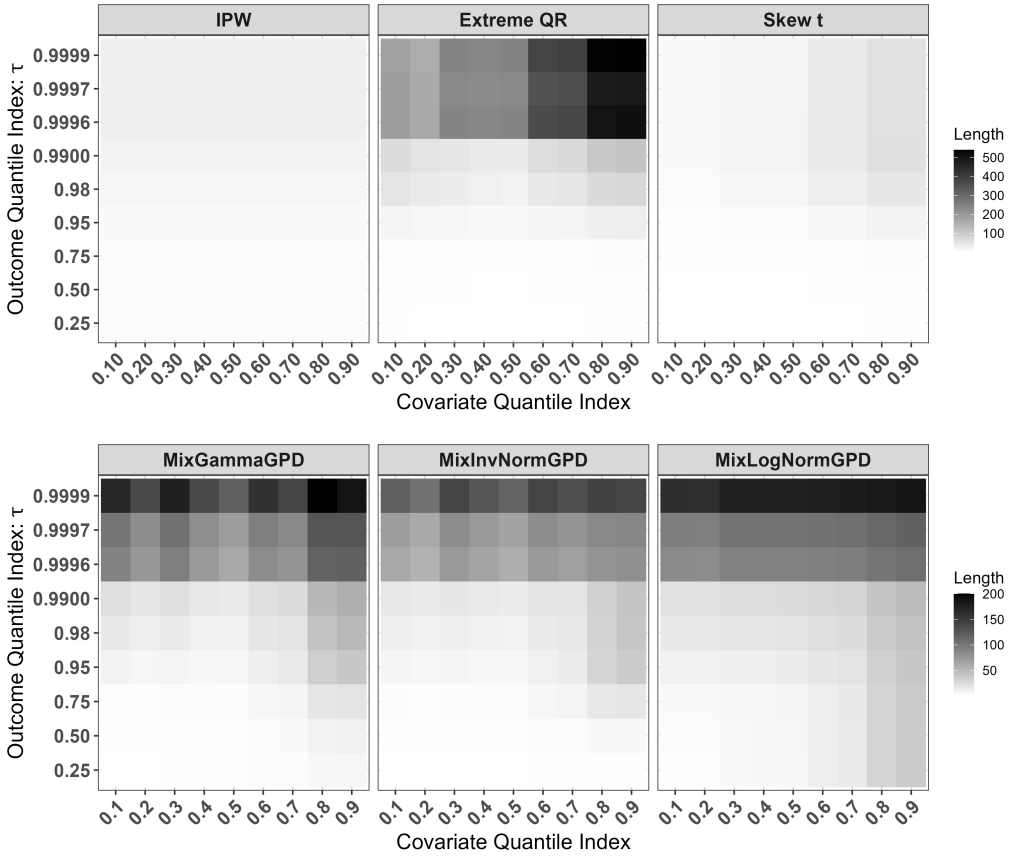

Figure 9: Average interval-width heatmaps for conditional quantile treatment effect estimates when $n = 500$. The top panel reports the three competing methods, namely IPW, Extreme QR, and Skew t. The bottom panel reports the three proposed bulk specifications, MixGammaGPD, MixInvNormGPD, and MixLogNormGPD. In each heatmap, columns index the nine covariate-grid profiles and rows index the nine quantile levels. Blue cells indicate shorter uncertainty intervals, whereas red cells indicate longer intervals, thereby documenting how inferential width changes across the covariate–quantile grid.