Update Spatial Random Effects & Variance:

W

τ^2

ρ

Impute Censored Data:

ij

$$\tilde{y}_{ij} = \{ \ TruncNormal\Big(\mu_{ij}, \sigma^2; \log y_{ij}\Big), if \delta_{ij} = 0, \log y_{ij}, if \delta_{ij} = 1.$$

Update BART:

$y_{ij} -$

W_i

$(Z_{ij}, \mathbf{X}_{ij})$

$\{\mathcal{T}_h, \mathcal{M}_h\}$

σ^2

Essay Defense: Analysis of spatially clustered survival data with unobserved covariates using SBART and Estimating Treatment effects

Durbadal Ghosh
Florida State University

Methods

Important Features / Novelties in our method

- We implement the **Two-stage method:** (i) estimate PS $\hat{e}_{ij}$, (ii) plugin $\hat{e}_{ij}$ into the outcome model; Doubly robust
- Non-parametric(mBART) outcome modelling
- Frailty/ Random county effects
- Spatial Association among clusters.

Notation and Definitions

- Study with K clusters; cluster i has n_i subjects; total $N = \sum_{i=1}^{K} n_i$.
- Binary Treatment: For each subject j in cluster i, $Z_{ij} \in \{0, 1\}$.
- For subject j in cluster i:
  - Individual covariates: $\mathbf{X}_{ij}$.
  - Failure time: T_{ij} (possibly right-censored at C_{ij}).
  - Observed outcome: $y_{ij} = \min(T_{ij}, C_{ij})$ with censoring indicator $\delta_{ij} = I(T_{ij} < C_{ij})$.
- Cluster-level covariates: $\mathbf{V}_i$.
- Counterfactual outcomes: $T_{ij}(1)$ and $T_{ij}(0)$ for treatment $Z_{ij} = 1$ and $Z_{ij} = 0$, respectively.

Potential outcomes

- $\{T_{ij}(1), T_{ij}(0)\}$: potential times
- $\{C_{ij}(1), C_{ij}(0)\}$: potential censoring times
- **Under Consistency, the relation between counterfactual and factual data:**

$$
\begin{aligned}
T_{ij} &= Z_{ij}T_{ij}(1) + (1 - Z_{ij})T_{ij}(0) \\
C_{ij} &= Z_{ij}C_{ij}(1) + (1 - Z_{ij})C_{ij}(0)
\end{aligned}
$$

Causal Assumptions

(A1) SUTVA: For any two subjects j and j' in clusters i and i', with treatment assignments Z_{ij} and $Z_{i'j'}$:

$$T_{ij}(Z_{ij}, Z_{i'j'}) = T_{ij}(Z_{ij}, Z'_{i'j'})$$

(A2) Consistency:

$$T_{ij} = T_{ij}(1)I(Z_{ij} = 1) + T_{ij}(0)I(Z_{ij} = 0)$$

(A3) Weak Unconfoundedness:

$$T_{ij}(z) \perp\!\!\!\perp Z_{ij} \mid \mathbf{X}_{ij}, \mathbf{V}_i \quad \text{for } z = 0, 1$$

(A4) Positivity: The propensity score

$$e(\mathbf{X}_{ij}, \mathbf{V}_i) = P(Z_{ij} = 1 \mid \mathbf{X}_{ij}, \mathbf{V}_i)$$

is bounded away from 0 and 1.

(A5) Covariate-dependent Censoring:

$$T_{ij}(z) \perp\!\!\!\perp C_{ij}(z) \mid \mathbf{X}_{ij}, \mathbf{V}_i, Z_{ij} \quad \text{for } z = 0, 1.$$

Estimands

- **Counterfactual survival functions for $z = 0, 1$**
  - $S^{(z)}(t|\mathbf{X}, \mathbf{V}) = P(T(z) \geq t|\mathbf{X}, \mathbf{V})$
  - $S^{(z)}(t) = P(T(z) \geq t) = \mathbb{E}_{\mathbf{X},\mathbf{V}}[S^{(z)}(t|\mathbf{X}, \mathbf{V})]$
- **Survival probability causal effect (SPCE) at t (Mao et al. 2018):**

$$\Delta^{SPCE}(t) := S^{(1)}(t) - S^{(0)}(t)$$

- **Average causal effect (ACE) :**

$$\Delta^{ACE} = \mathbb{E}[T(1)] - \mathbb{E}[T(0)]$$

- **Restricted average causal effect (RACE) :**

$$\Delta^{RACE}(t^*) = \mathbb{E}[\min(T(1), t^*)] - \mathbb{E}[\min(T(0), t^*)]$$

Estimands

- **Conditional survival probability causal effect (CSPCE) at t :**

$$\Delta^{CSPCE}(t, \mathbf{X}, \mathbf{V}) := S^{(1)}(t, \mathbf{X}, \mathbf{V}) - S^{(0)}(t, \mathbf{X}, \mathbf{V})$$

- **Conditional average causal effect (CACE) :**

$$\Delta^{CACE}(\mathbf{X}, \mathbf{V}) = \mathbb{E}[T(1) \mid \mathbf{X}, \mathbf{V}] - \mathbb{E}[T(0) \mid \mathbf{X}, \mathbf{V}]$$

- **Conditional restricted average causal effect (CRACE) :**

$$\Delta^{CRACE}(t^*, \mathbf{X}, \mathbf{V}) = \mathbb{E}[\min(T(1), t^*) \mid \mathbf{X}, \mathbf{V}] - \mathbb{E}[\min(T(0), t^*) \mid \mathbf{X},$$

Method

Model for Propensity Score

$P[Z_{ij} = 1 \mid \mathbf{X}_{ij}, \mathbf{V_i}] = e_{ij}, \quad logit(e_{ij}) = g(\mathbf{X}_{ij})$, where $g(\cdot) \sim BART$

Two-stage implementation: (i) estimate PS $\hat{e}_{ij}$, (ii) plug in $\hat{e}_{ij}$ into the survival model. (Doubly Robust)

The AFT-BART Model with Spatial CAR Prior

- **Model:**

$$\log T_{ij} = f\Big(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)\Big) + W_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2).$$

- **Spatial Random Effects:**

$$p(W \mid \tau^2, \rho) \propto \exp\Big\{-\frac{1}{2\tau^2} W^\top (D - \rho A) W\Big\},$$

where A is the spatial adjacency matrix, D is diagonal with $d_{ii} = \sum_{i'} A_{ii'}$, and ρ is the spatial parameter.

- **BART:**

$$f(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)) = \sum_{h=1}^{H} g(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i); \mathcal{T}_h, \mathcal{M}_h)$$

Prior Specification

- Error variance: $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$.
- Spatial variance: $\sigma_W^2 \sim \mathrm{IG}(a_W, b_W)$.
- Spatial correlations: $\rho \sim Uniform(\frac{1}{\alpha_{(1)}}, \frac{1}{\alpha_{(K)}})$. $\alpha_{(1)}, \alpha_{(K)}$ are the minimum and maximum eigenvalues of A, respectively.
- Function $f(\cdot) \sim \mathrm{SBART}$

Observed Data Likelihood

- For each subject j in cluster i, we observe

$$y_{ij} = \min(T_{ij}, C_{ij}), \quad \delta_{ij} = 1(T_{ij} < C_{ij}).$$

- Define the model mean (on the log-scale) as

$$\mu_{ij} = f(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)) + W_i.$$

- Then the individual likelihood contribution is

$$L_{ij}^{\text{obs}}(\theta) = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log y_{ij} - \mu_{ij})^2}{2\sigma^2} \right] \right\}^{\delta_{ij}} \left\{ 1 - \Phi\left(\frac{\log y_{ij} - \mu}{\sigma} \right.\right.$$

where $\Phi(\cdot)$ is the standard normal CDF.

- The full observed-data likelihood is

$$L_{\text{obs}}(\theta) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} L_{ij}^{\text{obs}}(\theta).$$

Data Augmentation

- Define the latent log survival time $\tilde{y}_{ij}$ as

$$\tilde{y}_{ij} = \begin{cases} \text{TruncNormal}\Big(\mu_{ij}, \sigma^2; \log y_{ij}\Big), & \text{if } \delta_{ij} = 0, \\ \log y_{ij}, & \text{if } \delta_{ij} = 1. \end{cases}$$

  Here, $\text{TruncNormal}(\mu, \sigma^2; a)$ denotes a $N(\mu, \sigma^2)$ distribution truncated to the interval (a, ∞). The imputed values are used in the complete-data likelihood.

Complete Data Likelihood

- Introduce the latent (complete) log survival times:

$$\tilde{y}_{ij} = \begin{cases} \log y_{ij}, & \delta_{ij} = 1, \\ \text{draw from } N(\mu_{ij}, \sigma^2) \text{ truncated to } [\log y_{ij}, \infty), & \delta_{ij} = 0. \end{cases}$$

- With μ_{ij} defined as before, the complete-data likelihood is

$$L_{\text{complete}}(\theta) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\tilde{y}_{ij} - \mu_{ij})^2}{2\sigma^2} \right].$$

Algorithm 1: A Single Iteration

1. **Update Spatial Random Effects & Variance:** Update W, τ^2, and ρ from their full conditionals based on the CAR prior.
2. **Impute Censored Data:** For subjects ij, sample the latent log survival time as

$$\tilde{y}_{ij} = \begin{cases} \text{TruncNormal}\Big(\mu_{ij}, \sigma^2; \log y_{ij}\Big), & \text{if } \Delta_{ij} = 0, \\ \log y_{ij}, & \text{if } \Delta_{ij} = 1. \end{cases}$$

3. **Update BART:** With responses $\tilde{y}_{ij} - W_i$ and covariates $(z_{ij}, \mathbf{x}_{ij})$, update the BART parameters $\{\mathcal{T}_h, \mathcal{M}_h\}$ and the error variance σ^2 via Bayesian backfitting.

Full Conditionals

Latent Log Survival Times $\tilde{y}_{ij}$

- **For observed events ($\delta_{ij} = 1$):**

$$\tilde{y}_{ij} = \log y_{ij} \quad \text{(fixed)}.$$

- **For censored observations ($\delta_{ij} = 0$):**

$$\tilde{y}_{ij} \mid \text{Rest} \sim \text{TruncNormal}\Big(\mu_{ij}, \sigma^2; [\log y_{ij}, \infty)\Big).$$

Full Conditional for $\mathbf{W}$

- Let $\mathbf{W} = (W_1, \ldots, W_K)^\top$ be the vector of spatial random effects. For each cluster i with n_i observations, define

$$s_i = \sum_{j \in \text{cluster } i} \Big(\tilde{y}_{ij} - f_{ij} \Big),$$

  and set $\mathbf{s} = (s_1, \ldots, s_K)^\top$.

- The likelihood contribution for cluster i is

$$\exp\left\{ -\frac{n_i}{2\sigma^2} W_i^2 + \frac{s_i}{\sigma^2} W_i \right\},$$

  so that for all clusters,

$$L(\mathbf{W}) \propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{K} \left[\frac{n_i}{\sigma^2} W_i^2 - 2\frac{s_i}{\sigma^2} W_i \right] \right\}.$$

- Multiplying likelihood and prior, the joint full conditional is proportional to

$$p(\mathbf{W} \mid \text{Rest}) \propto \exp\left\{-\frac{1}{2}\left[\mathbf{W}^\top \left(\frac{1}{\sigma^2}\operatorname{diag}(n_i) + \frac{1}{\tau^2}(D - \rho A)\right)\mathbf{W} - 2\right.\right.$$

- Thus, we have

$$\mathbf{W} \mid \text{Rest} \sim N\Big(\boldsymbol{\mu}_W,\ \Sigma_W\Big),$$

where

$$\Sigma_W = \left(\frac{1}{\sigma^2}\operatorname{diag}(n_i) + \frac{1}{\tau^2}(D - \rho A)\right)^{-1}$$

and

$$\boldsymbol{\mu}_W = \Sigma_W \left(\frac{\mathbf{s}}{\sigma^2}\right).$$

Error Variance σ^2

- The residual sum-of-squares is

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\tilde{y}_{ij} - f_{ij} - W_i)^2.$$

- With the prior $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$, the full conditional is

$$\sigma^2 \mid \mathrm{Rest} \sim \mathrm{IG}\left(a_\sigma + \frac{N}{2},\ b_\sigma + \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\tilde{y}_{ij} - f_{ij} - W_i)^2\right),$$

where $N = \sum_{i=1}^{K} n_i$.

Spatial Variance τ^2 (or σ_W^2)

- With the CAR prior and $\tau^2 \sim \mathrm{IG}(a_W, b_W)$, the full conditional is

$$\tau^2 \mid \mathrm{Rest} \sim \mathrm{IG}\left(a_W + \frac{K}{2},\ b_W + \frac{1}{2}\, W^\top (D - \rho A) W\right).$$

Spatial Correlation Parameter ρ

- Prior:

$$\rho \sim \text{Uniform}\Big(\frac{1}{\alpha_{(1)}}, \frac{1}{\alpha_{(K)}}\Big),$$

  where $\alpha_{(1)}$ and $\alpha_{(K)}$ are the minimum and maximum eigenvalues of A, respectively.

- The full conditional (up to normalizing constant) is:

$$p(\rho \mid \text{Rest}) \propto \exp\left\{-\frac{1}{2\tau^2} W^\top (D - \rho A) W\right\} \mathbb{I}\Big(\rho \in \Big[\frac{1}{\alpha_{(1)}}, \frac{1}{\alpha_{(K)}}\Big]\Big).$$

- This is typically updated via a Metropolis–Hastings step.

BART Function Parameters

- The outcome model (after accounting for W_i) is

$$\tilde{y}_{ij} - W_i = \sum_{h=1}^{H} g(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i); \mathcal{T}_h, \mathcal{M}_h) + \epsilon_{ij},$$

with $\epsilon_{ij} \sim N(0, \sigma^2)$.

- For tree h, define the partial residuals:

$$R_{ij}^{(h)} = \tilde{y}_{ij} - W_i - \sum_{l \neq h} g(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i); \mathcal{T}_l, \mathcal{M}_l).$$

- The full conditional for the h^{th} tree is then

$$p(\mathcal{T}_h, \mathcal{M}_h \mid \text{Rest}) \propto \prod_{i=1}^{K} \prod_{j=1}^{n_i} \exp\left\{ - \frac{\left(R_{ij}^{(h)} - g(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i\right.}{2\sigma^2}$$

- Updates are performed via the Bayesian backfitting algorithm.

Posterior Inference of Causal Effects

Linking Potential Outcomes and Observed Data

- **Consistency:** For subjects with $Z = 1$,

$$T = T(1).$$

- **Weak Unconfoundedness:** $T(1) \perp Z \mid \mathbf{X}, \mathbf{V}$ implies that

$$P\big(T(1) \geq t \mid \mathbf{X}, \mathbf{V}\big) = P\big(T(1) \geq t \mid \mathbf{X}, \mathbf{V}, Z = 1\big).$$

- **By Consistency:** For treated subjects,

$$P\big(T(1) \geq t \mid \mathbf{X}, \mathbf{V}, Z = 1\big) = P\big(T \geq t \mid \mathbf{X}, \mathbf{V}, Z = 1\big).$$

- **Thus,**

$$P\big(T(1) \geq t \mid \mathbf{X}, \mathbf{V}\big) = P\big(T \geq t \mid \mathbf{X}, \mathbf{V}, Z = 1\big).$$

Deriving the Survival Function for $T(1)$

- For treated subjects ($Z = 1$), our outcome model specifies

$$\log T \mid \mathbf{X}, \mathbf{V}, Z = 1 \sim N\Big(f(1, \mathbf{X}, \mathbf{V}, \hat{e}(\mathbf{X}, \mathbf{V})) + W,\ \sigma^2\Big).$$

- Define

$$\mu(1, \mathbf{X}, \mathbf{V}) = f\Big(1, \mathbf{X}, \mathbf{V}, \hat{e}(\mathbf{X}, \mathbf{V})\Big) + W.$$

- Then, the survival probability (on the log-scale) is

$$P\big(\log T \geq \log t \mid \mathbf{X}, \mathbf{V}, Z = 1\big) = 1 - \Phi\left(\frac{\log t - \mu(1, \mathbf{X}, \mathbf{V})}{\sigma}\right).$$

- By the causal assumptions, this equals the survival function for the potential outcome:

$$S^{(1)}(t \mid \mathbf{X}, \mathbf{V}) = P\big(T(1) \geq t \mid \mathbf{X}, \mathbf{V}\big) = 1 - \Phi\left(\frac{\log t - \mu(1, \mathbf{X}, \mathbf{V})}{\sigma}\right).$$

Estimating $S^{(1)}(t \mid \mathbf{X}, \mathbf{V})$ from MCMC Samples

- For each MCMC iteration $m = 1, \ldots, M$, obtain posterior draws:

$$f^{(m)}(1, \mathbf{X}, \mathbf{V}, \hat{e}(\mathbf{X}, \mathbf{V})), \quad W^{(m)}, \quad \sigma^{(m)}.$$

- Compute the linear predictor:

$$\mu^{(m)}(1, \mathbf{X}, \mathbf{V}) = f^{(m)}(1, \mathbf{X}, \mathbf{V}, \hat{e}(\mathbf{X}, \mathbf{V})) + W^{(m)}.$$

- For a given time t, the iteration-specific survival estimate is:

$$S_m^{(1)}(t \mid \mathbf{X}, \mathbf{V}) = 1 - \Phi\left(\frac{\log t - \mu^{(m)}(1, \mathbf{X}, \mathbf{V})}{\sigma^{(m)}}\right).$$

- The overall posterior estimate is the Monte Carlo average:

$$\widehat{S^{(1)}}(t \mid \mathbf{X}, \mathbf{V}) = \frac{1}{M} \sum_{m=1}^{M} S_m^{(1)}(t \mid \mathbf{X}, \mathbf{V}).$$

Estimating the Marginal SPCE from MCMC Samples

- Recall the conditional survival functions for treatment and control:

$$S^{(1)}(t \mid \mathbf{X}, \mathbf{V}) = 1-\Phi\left(\frac{\log t - \mu(1, \mathbf{X}, \mathbf{V})}{\sigma}\right), \quad S^{(0)}(t \mid \mathbf{X}, \mathbf{V}) = 1-$$

- For each MCMC iteration m, we have draws:

$$\mu^{(z),(m)}(\mathbf{X}, \mathbf{V}) = f^{(m)}(z, \mathbf{X}, \mathbf{V}, \hat{e}(\mathbf{X}, \mathbf{V}))+W^{(m)}, \quad \sigma^{(m)}, \quad z = 0, 1.$$

- Compute the iteration-specific conditional survival functions:

$$S_m^{(z)}(t \mid \mathbf{X}, \mathbf{V}) = 1 - \Phi\left(\frac{\log t - \mu^{(z),(m)}(\mathbf{X}, \mathbf{V})}{\sigma^{(m)}}\right), \quad z = 0, 1.$$

- **Marginalize over the covariate distribution:** For a sample of n subjects with covariates $\{(\mathbf{X}_{ij}, \mathbf{V}_i)\}_{i=1}^{n}$, define

$$\widehat{S}_m^{(z)}(t) = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} S_m^{(z)}(t \mid \mathbf{X}_{ij}, \mathbf{V}_i), \quad z = 0, 1.$$

- The iteration-specific marginal survival probability causal effect is:

$$\Delta_m^{SPCE}(t) = \widehat{S}_m^{(1)}(t) - \widehat{S}_m^{(0)}(t).$$

- **Posterior Estimate:** Average over all M MCMC samples to obtain

$$\widehat{\Delta}^{SPCE}(t) = \frac{1}{M} \sum_{m=1}^{M} \Delta_m^{SPCE}(t).$$

Simulations

Simulation 4: Data Generation

- **Subjects:** $n = 5000$ observations.
- **Covariates:** Two confounders:

$$X_1, X_2 \sim U(0, 1).$$

- **Clusters:** $K = 10$ clusters, with subjects randomly assigned.

Treatment Assignment

- **Propensity Score:**

$$p = \text{expit}\Big(0.3X_1 - 0.2X_2 + 0.5X_1X_2\Big),$$

  where $\text{expit}(x) = \frac{1}{1+e^{-x}}$.

- **Treatment:** $Z \sim \text{Bernoulli}(p)$.

- **Note:** Treatment assignment depends on the confounders.

Spatial Structure and CAR Random Effects

- **Adjacency:** A structure for $K = 10$ clusters: Based on 10 Florida counties in the Western region.
- **CAR Prior:**

$$p(\mathbf{W} \mid \sigma_W^2, \rho) \propto |Q(\rho)|^{1/2} \exp\Big\{-\frac{1}{2\sigma_W^2}\mathbf{W}^\top Q(\rho)\mathbf{W}\Big\},$$

where

$$Q(\rho) = D - \rho A, \quad D_{ii} = \sum_j A_{ij}.$$

- **Generation:** Cluster effects are drawn from a multivariate normal with covariance $\sigma_W^2 Q(\rho)^{-1}$.

Outcome Model and Censoring

- **True Outcome Model (log-scale):**

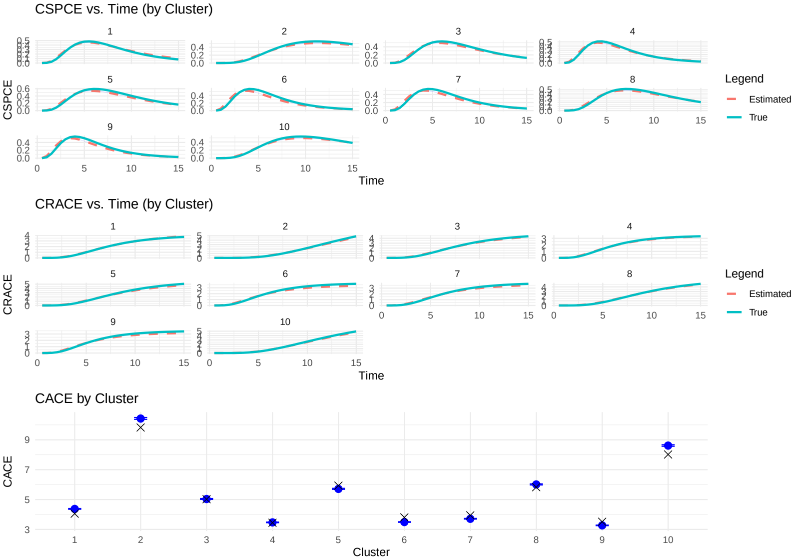
$$\log T = f(X, Z) + W \cdot \Big(1 + 0.5\, Z\, X_1\Big) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

- **True Function:**

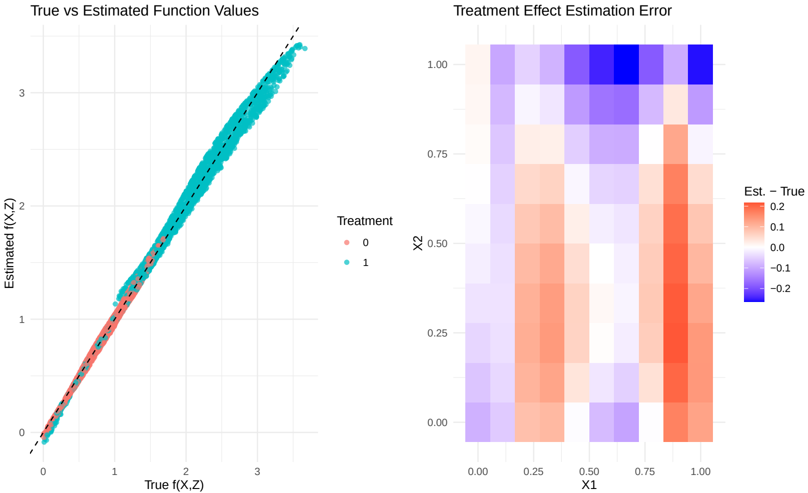
$$f(X, Z) = \sin(\pi X_1) + \ln(1 + X_2^2) + 2Z\,(X_1 X_2) + (X_1^2)Z.$$

- **Survival Times:** $T = \exp\{\log T\}$.
- **Censoring:**
  - Censoring times $C \sim \text{Exponential}(\lambda_c)$ (e.g., $\lambda_c = 0.05$).
  - Observed time: $y = \min(T, C)$ and indicator $\delta = I(T \leq C)$.

Results (Rep1)

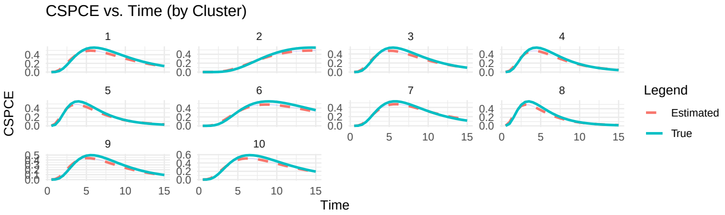


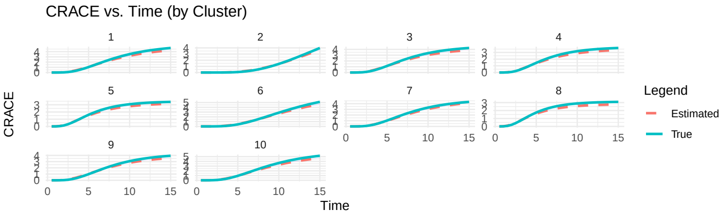
Results (Rep1)

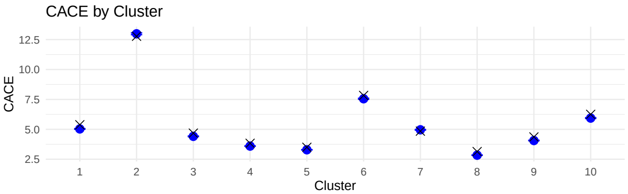


Performance Metrics:
RMSE: 0.0530
MAE: 0.0379
Correlation: 0.9977

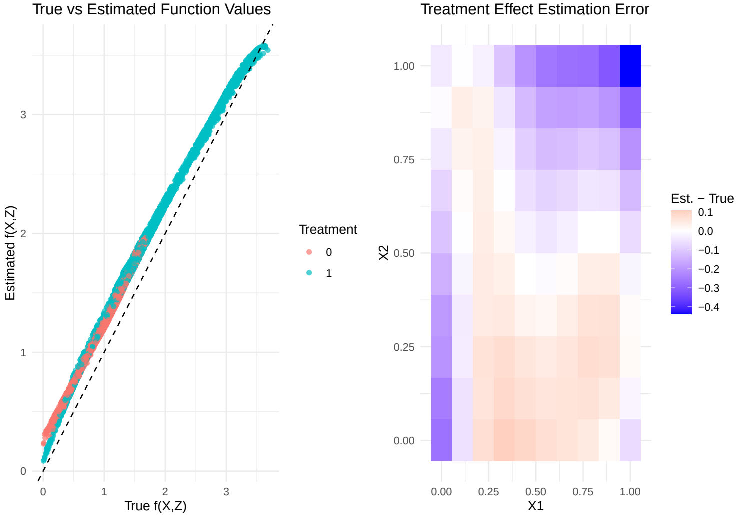
Results (Rep2)





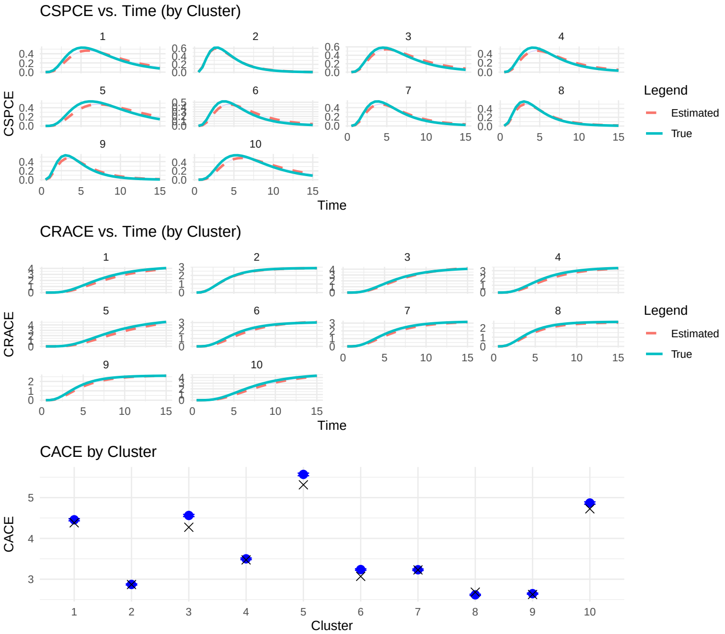


Results (Rep2)

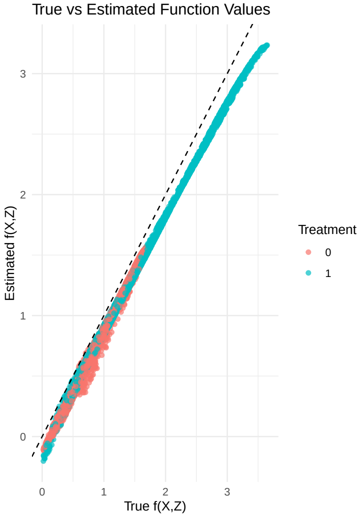


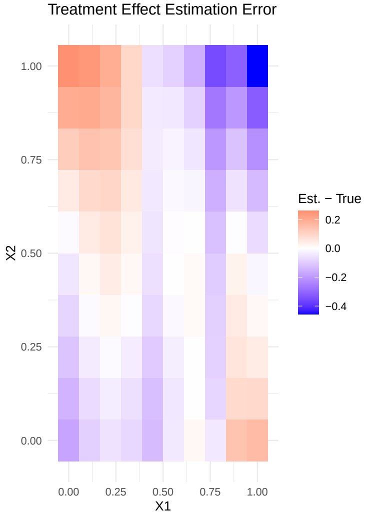
Performance Metrics:
RMSE: 0.2449
MAE: 0.2391
Correlation: 0.9981

Results (Rep3)



Results (Rep3)





Performance Metrics:
RMSE: 0.1709
MAE: 0.1624
Correlation: 0.9987

Thank you!