Supplemental Material for "Ensemble Methods for Survival Function Estimation with Time-Varying Covariates"

Weichi Yao[1*], Halina Frydman[1], Denis Larocque[2], and Jeffrey S. Simonoff[1]

[*]Correspondence: wyao@stern.nyu.edu
[1]Department of Technology, Operations, and Statistics, Stern School of Business, New York University, New York City, USA
[2]Department of Decision Sciences, HEC Montréal, Montréal, CA

Contents

S1 Proposed forests for time-varying covariate survival data **3**

S1.1 Generating the survival times . . . 3

S1.2 Changing pattern of the time-varying covariates in the simulations . . . 3

S1.3 Histogram of simulated survival times . . . 5

S1.4 Parameters set in the simulation study . . . 6

S1.5 Bootstrapping subjects vs. bootstrapping pseudo-subjects . . . 7

S1.6 Regulating the construction of trees in forests . . . 9

S1.7 Performance comparison for Weibull-Decreasing distribution . . . 14

S1.8 Analysis of variance . . . 15

S2 Proposed forests for time-invariant covariate survival data **15**

S2.1 Model setup . . . 15

S2.2 Regulating the construction of trees in forests . . . 16

S2.3 Performance comparison and IBS-based CV model selection . . . 22

S1 Proposed forests for time-varying covariate survival data

S1.1 Generating the survival times

Suppose the covariates $\boldsymbol{X}$ are initially observed at $t_0 = 0$, and change values at $t_1, \ldots, t_{m-1}$, before the subject is censored or the event occurs at t_m. That means there are in total m intervals, within which the values of all covariates remain the same, $\boldsymbol{X}(t) = \boldsymbol{x}_i$, for $t \in [t_i, t_{i+1})$, $i = 0, 1, \ldots, m-1$. Consider the Weibull distribution with coefficient (λ, v) and a survival relationship $\vartheta(t) = \vartheta(\boldsymbol{X}(t))$. Denote $\vartheta_i = \vartheta(\boldsymbol{x}_i)$, $i = 0, \ldots, m-1$.

The survival times are generated as follows:

1. Compute the values of cumulative hazard function $H(t)$ at $t_1, t_2, \ldots, t_{m-1}$ from

$$H(t) = \begin{cases} \lambda \exp(\vartheta_0) t^v & \text{if } 0 \le t \le t_1; \\ \lambda \exp(\vartheta_i)\left(t^v - t_i^v\right) + H(t_i, \vartheta_{i-1}) & \text{if } t_i < t \le t_{i+1},\ i = 1, \ldots, m-2; \\ \lambda \exp(\vartheta_{m-1})\left(t^v - t_{m-1}^v\right) + H(t_{m-1}, \vartheta_{m-2}) & \text{if } t_{m-1} < t. \end{cases}$$

2. Divide time lines into the following regions

$$R_i = \begin{cases} [H(t_{i-1}), H(t_i)] & i = 1, \ldots, m-1; \\ [H(t_{i-1}), \infty) & i = m. \end{cases}$$

3. Randomly generate $u \in \mathcal{U}(0,1)$, the survival time is determined by which region $-\log(u)$ falls in:

$$T = \begin{cases} \left[\frac{-\log(u)}{\lambda \exp(\vartheta_0)}\right]^{1/v} & \text{if } -\log(u) \in R_1; \\ \left[\frac{-\log(u) - H(t_{i-1})}{\lambda \exp(\vartheta_{i-1})} + t_{i-1}^v\right]^{1/v} & \text{if } -\log(u) \in R_i,\ i = 2, \ldots, m. \end{cases}$$

4. The corresponding survival probability function is

$$S(t) = \exp(-H(t)),$$

where t falls into mutually exclusive intervals $D_1 = [0, t_1), D_2 = [t_1, t_2), \ldots, D_m = [t_{m-1}, \infty)$.

S1.2 Changing pattern of the time-varying covariates in the simulations

This section provide further details on the changing pattern of the time-varying covariates, specifically, on how to generate the observed values of time-varying covariates X_6, X_{13}, X_{16}, X_{18}, X_{20} at $t_0, \ldots, t_{m-1}$: $x_{j,k}$, for $j = 0, \ldots, m-1$ and $k = 6, 13, 16, 18, 20$.

- X_6, whose initial value is randomly generated from $\{0, 1, 2\}$ with equal probability:
  - if the initial value is 2 then it stays at 2 for all the following time points: $x_{0,6} = \cdots = x_{m-1,6} = 2$;

- if the initial value is less than 2, randomly sample the number of time points from $\{1, \ldots, m\}$ that it will stay at the initial value, $\tilde{n}_0$,
  - if the initial value is 1,
    - if $\tilde{n}_0 < m$, then the values at the rest $m - \tilde{n}_0$ time points are all $2's$: $x_{0,6} = \cdots = x_{\tilde{n}_0-1,6} = 1$ and $x_{\tilde{n}_0,6} = \cdots = x_{m-1,6} = 2$;
    - otherwise, it stays at 1 for all time points: $x_{0,6} = x_{1,6} = \cdots = x_{m-1,6} = 1$;
  - if the initial value is 0, randomly sample the number of time points from $\{1, \ldots, m - \tilde{n}_0\}$ that its value becomes 1, $\tilde{n}_1$,
    - if $\tilde{n}_1 < m - \tilde{n}_0$, then its values at the rest $m - \tilde{n}_0 - \tilde{n}_1$ time points are $2's$: $x_{0,6} = \cdots = x_{\tilde{n}_0-1,6} = 0$, $x_{\tilde{n}_0,6} = \cdots = x_{\tilde{n}_0+\tilde{n}_1-1,6} = 1$ and $x_{\tilde{n}_0+\tilde{n}_1,6} = \cdots = x_{m-1,6} = 2$;
    - otherwise its values at the rest $m - \tilde{n}_0$ time points are 1: $x_{0,6} = \cdots = x_{\tilde{n}_0-1,6} = 0$ and $x_{\tilde{n}_0,6} = \cdots = x_{m-1,6} = 1$.

- X_{13}, whose changing pattern is always $0 \rightarrow 1$: randomly sample the number of time points from $\{1, \ldots, m-1\}$ that it stays at the initial value 0, $\tilde{n}_0$, and it then changes value to 1 for the rest of the time points: $x_{0,13} = \cdots = x_{\tilde{n}_0-1,13} = 0$ and $x_{\tilde{n}_0,13} = \cdots = x_{m-1,13} = 1$;

- X_{16}, whose changing pattern is either $0 \rightarrow 1$ or $1 \rightarrow 2$: first randomly generated its initial value from Bern(0.5), then randomly sample the number of time points from $\{1, \ldots, m-1\}$ that it stays at the initial value, $\tilde{n}_0$, and it then increases its value by 1 for the rest of the time points: $x_{0,16} = \cdots = x_{\tilde{n}_0-1,16}$ and $x_{\tilde{n}_0,16} = \cdots = x_{m-1,16} = x_{0,16} + 1$;

- X_{18}, whose changing pattern is $0 \rightarrow 1 \rightarrow 2$: randomly sample two numbers of time points from $\{1, \ldots, m-1\}$, $\tilde{n}_0$ and $\tilde{n}_1$, then $x_{0,18} = \cdots = x_{\tilde{n}_0-1,18} = 1$, $x_{\tilde{n}_0,18} = \cdots = x_{\tilde{n}_1-1,18} = 1$ and $x_{\tilde{n}_1,18} = \cdots = x_{m-1,18} = 2$;

- X_{20}, which is a linear function of the left-truncated time point of the interval with slope and intercept follows Unif(0, 1): first generate $c_1, c_2 \sim$ Unif(0, 1), then $x_{j,20} = c_1 t_j + c_2$, $j = 0, \ldots, m-1$.

S1.3 Histogram of simulated survival times

Histograms of survival times for typical samples with the number of subjects $N = 500$ in each scenario are provided in Figure S1.1.

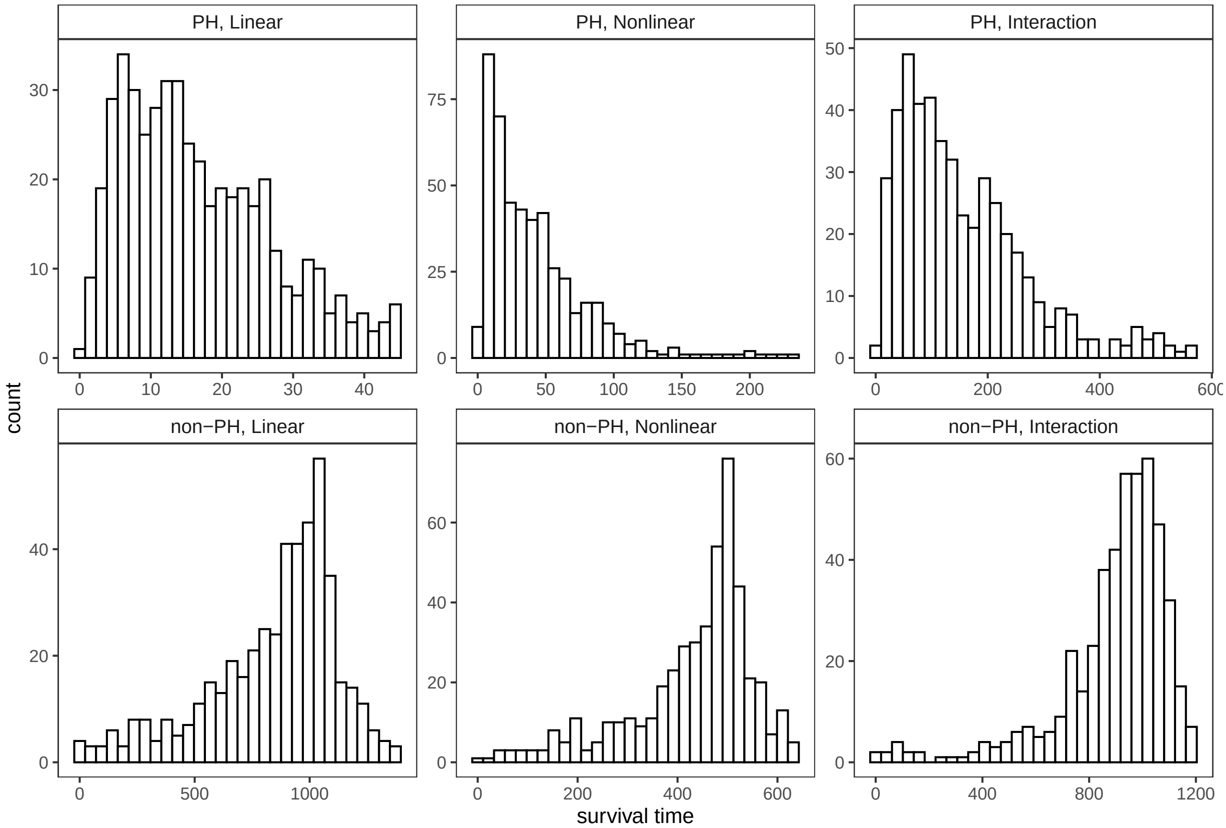

Figure S1.1: Histograms of the lower 95% of the survival times generated in typical samples with no right-censoring, with the number of subjects $N = 500$. The first row gives results under the PH setting, the second under the non-PH setting. The first column gives results for the linear survival relationship, second column for the nonlinear survival relationship, and the last column for the interaction survival relationship.

S1.4 Parameters set in the simulation study

In the simulation study, we set the parameters for the basic scenario "2TI + 4TV" as follows.

- Under the PH setting:
  - When the survival relationship is linear: set $\lambda = 0.012$, $\nu = 0.8$ if the underlying survival distribution is Weibull-Decreasing, and $\lambda = 0.001$, $\nu = 2$ for Weibull-Increasing.
    * When the signal-to-noise ratio is "Low," $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = -1$, $\beta_4 = -1$, $\beta_5 = 0.5$, $\beta_6 = -0.5$;
    * When the signal-to-noise ratio is "High," $\beta_0 = -4$, $\beta_1 = 5$, $\beta_2 = 5$, $\beta_3 = -5$, $\beta_4 = -5$, $\beta_5 = 2.5$, $\beta_6 = -2.5$;
  - When the survival relationship is nonlinear: set $\lambda = 0.15$, $\nu = 0.8$ if the underlying survival distribution is Weibull-Decreasing, and $\lambda = 0.0025$, $\nu = 1.8$ for Weibull-Increasing.
    * When the signal-to-noise ratio is "Low," $\phi_1 = \phi_4 = 0$, $\phi_2 = \phi_3 = -1$, $\psi_0 = 0.2$, $\psi_3 = 0.1$, $\psi_5 = 0.05$, $\psi_6 = 1$;
    * When the signal-to-noise ratio is "High," $\phi_1 = 0$, $\phi_2 = \phi_3 = -5$, $\phi_4 = 5$, $\psi_0 = 0.2$, $\psi_3 = 0.1$, $\psi_5 = 0.05$, $\psi_6 = 1$;
  - When the survival relationship is interaction, set $A = [0.7, 1]$, $B = \{4, 5\}$, and $\lambda = 0.14$, $\nu = 0.5$ if the underlying survival distribution is Weibull-Decreasing, $\lambda = 0.0001$, $\nu = 1.8$ for Weibull-Increasing.
    * When the signal-to-noise ratio is "Low," $\eta_1 = \eta_3 = 1$, $\eta_2 = \eta_4 = 0$, $\gamma_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = -1$, $\gamma_4 = -1$, $\gamma_5 = 0.5$, $\gamma_6 = -0.5$, $\alpha_0 = 0$, $\alpha_1 = -1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\alpha_4 = -1$, $\alpha_5 = -0.5$, $\alpha_6 = -0.5$;
    * When the signal-to-noise ratio is "High," $\eta_1 = \eta_3 = 5$, $\eta_2 = 1.5$, $\eta_4 = 1.101$, $\gamma_0 = -7$, $\gamma_1 = 5$, $\gamma_2 = 5$, $\gamma_3 = -5$, $\gamma_4 = -5$, $\gamma_5 = 2.5$, $\gamma_6 = -2.5$, $\alpha_0 = -10$, $\alpha_1 = -5$, $\alpha_2 = 5$, $\alpha_3 = 5$, $\alpha_4 = -5$, $\alpha_5 = -2.5$, $\alpha_6 = -2.5$;
- Under the non-PH setting:
  - When the survival relationship is linear, set $\lambda = 0.001$ for Weibull-Increasing.
    * When the signal-to-noise ratio is "Low," $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$, $\beta_4 = 1$, $\beta_5 = 10$, $\beta_6 = 1$;
    * When the signal-to-noise ratio is "High," $\beta_0 = -64.5$, $\beta_1 = 3$, $\beta_2 = 3$, $\beta_3 = 3$, $\beta_4 = 3$, $\beta_5 = 30$, $\beta_6 = 3$;
  - When the survival relationship is nonlinear, set $\lambda = 0.002$ for Weibull-Increasing.
    * When the signal-to-noise ratio is "Low," $\phi_1 = 1$, $\phi_2 = \phi_3 = \phi_4 = 0$;
    * When the signal-to-noise ratio is "High," $\phi_1 = 5$, $\phi_2 = \phi_3 = 0$, $\phi_4 = -2.835$;
  - When the survival relationship is interaction, set $A = [0.7, 1]$, $B = \{5\}$, and $\lambda = 0.001$ for Weibull-Increasing as the underlying survival function.
    * When the signal-to-noise ratio is "Low," $\eta_1 = \eta_3 = 1$, $\eta_2 = \eta_4 = 0$, $\gamma_0 = 0$, $\gamma_1 = -1$, $\gamma_2 = -1$, $\gamma_3 = -1$, $\gamma_4 = -1$, $\gamma_5 = -10$, $\gamma_6 = -1$, $\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\alpha_4 = 1$, $\alpha_5 = 10$, $\alpha_6 = 1$;
    * When the signal-to-noise ratio is "High," $\eta_1 = \eta_3 = 5$, $\eta_2 = 0.6$, $\eta_4 = 1.101$, $\gamma_0 = 64.5$, $\gamma_1 = -3$, $\gamma_2 = -3$, $\gamma_3 = -3$, $\gamma_4 = -3$, $\gamma_5 = -30$, $\gamma_6 = -3$, $\alpha_0 = -64.5$, $\alpha_1 = 3$, $\alpha_2 = 3$, $\alpha_3 = 3$, $\alpha_4 = 3$, $\alpha_5 = 30$, $\alpha_6 = 3$.

S1.5 Bootstrapping subjects vs. bootstrapping pseudo-subjects

Figures S1.2 and S1.3 give side-by-side boxplots of integrated L_2 difference, showing the performance comparison between bootstrapping pseudo-subjects and bootstrapping subjects for each type of the forests under the PH setting, and under the non-PH setting, respectively. The results are provided for data generated under the true model 2TI + 4TV.

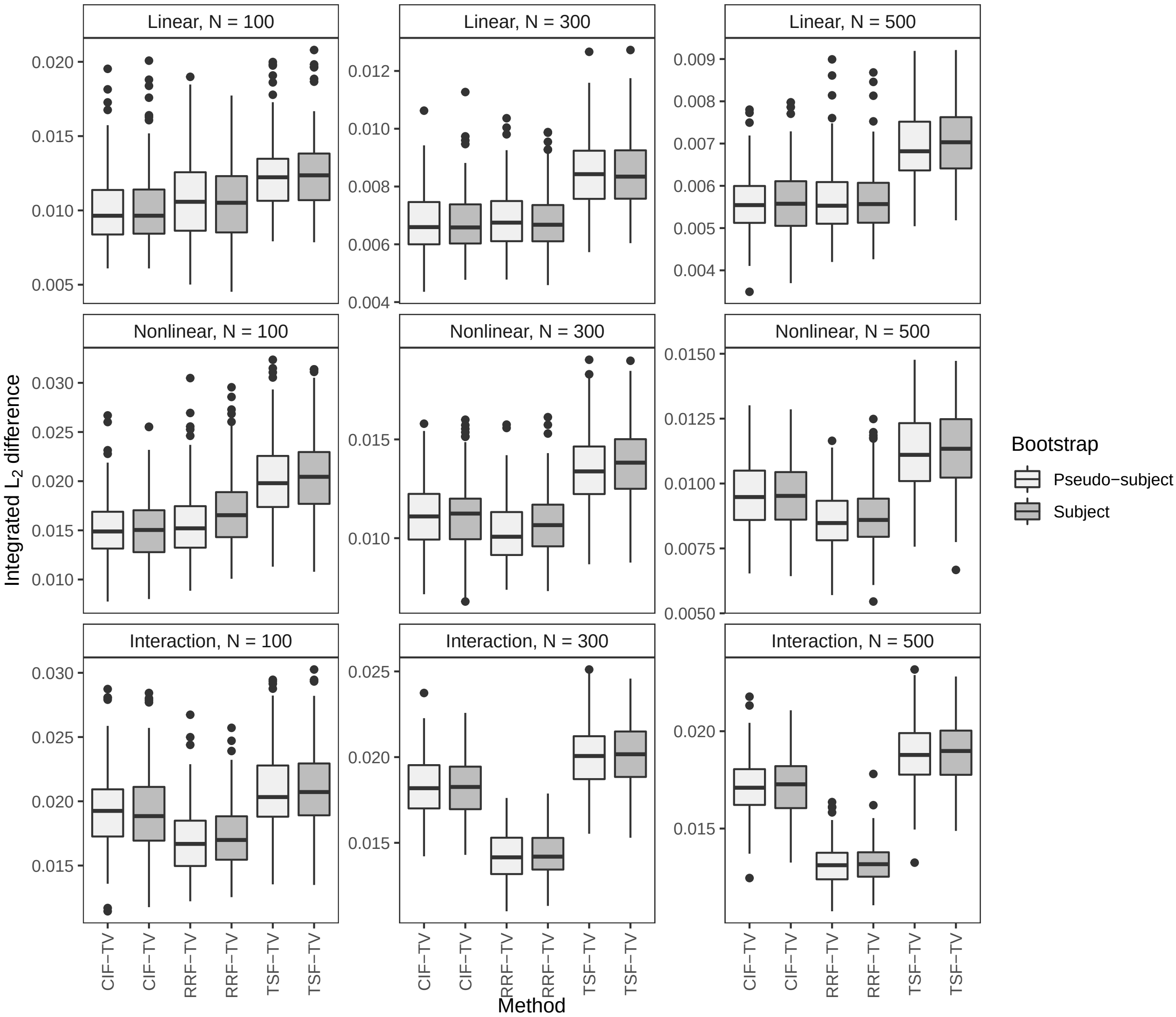

Figure S1.2: Integrated L_2 difference of three forest methods with different bootstrap mechanisms, on datasets having 2TI + 4TV in the true model, survival times generated from a Weibull-Increasing distribution under the PH setting, light right-censoring rate (20%). All forest methods are trained with $mtry = 5$ by default, and tuning parameters set to be $\sqrt{n}$. From the top row to the bottom, it gives results for the linear, nonlinear and the interaction survival relationship. From the first column to the third, it gives results for the number of subjects $N = 100, 300, 500$.

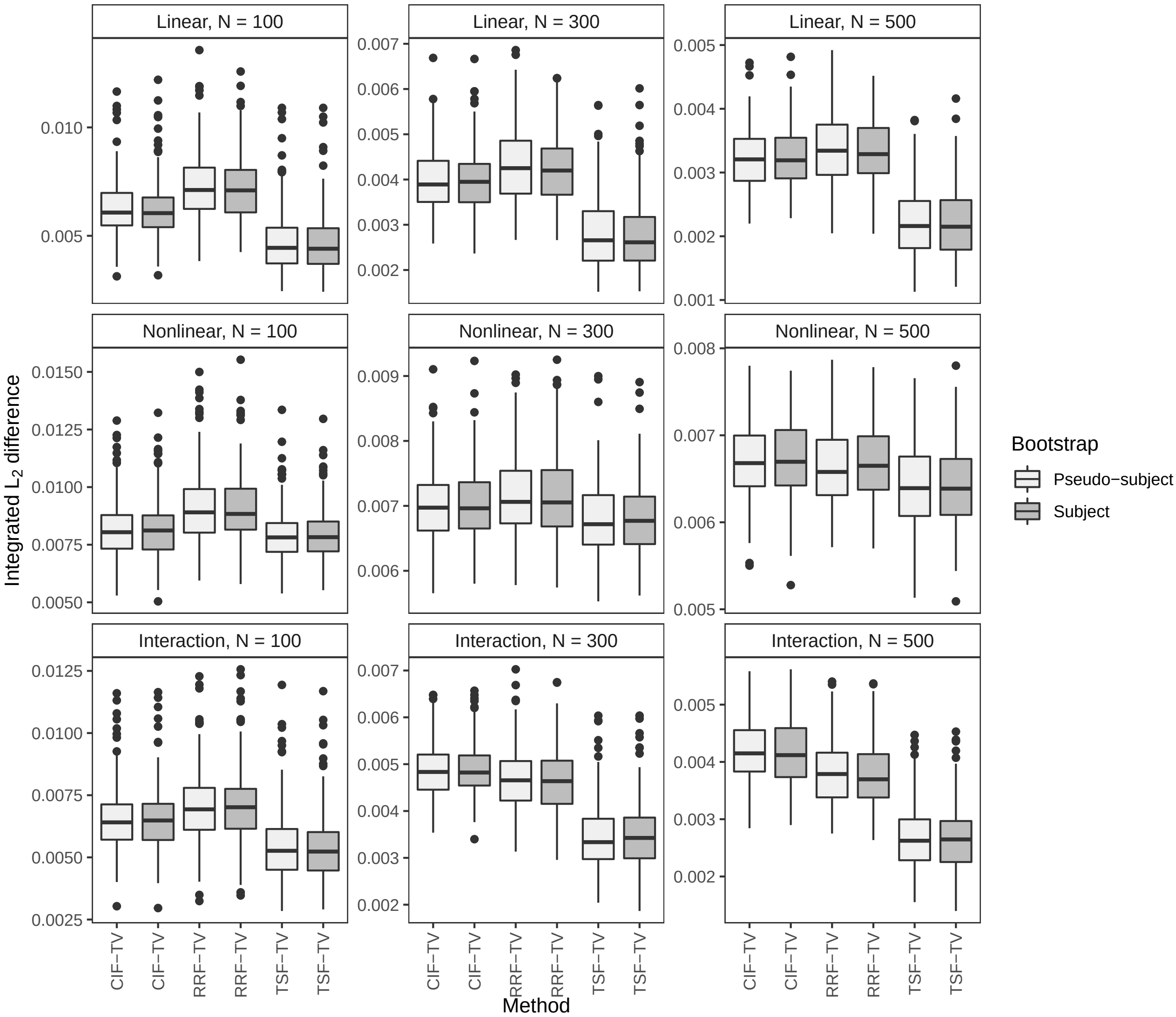

Figure S1.3: Integrated L_2 difference of three forest methods with different bootstrap mechanisms, on datasets having 2TI + 4TV in the true model, survival times generated from a Weibull-Increasing distribution under the non-PH setting, light right-censoring rate (20%). All forest methods are trained with $mtry = 5$ by default, and tuning parameters set to be $\sqrt{n}$. From the top row to the bottom, it gives results for the linear, nonlinear and the interaction survival relationship. From the first column to the third, it gives results for the number of subjects $N = 100, 300, 500$.

One can see that for each type of forest, the results when bootstrapping pseudo-subjects are very similar to those when bootstrapping the subjects. That is, the two different bootstrapping mechanisms do not result in fundamentally different levels of performance.

S1.6 Regulating the construction of trees in forests

Figures S1.4 to S1.7 give examples of how RRF-TV and TSF-TV perform with different values of *mtry* under the PH setting and the non-PH setting, respectively. The *mtry* values are chosen by the "out-of-bag" tuning procedure. The results are very similar compared to those for CIF-TV given in the manuscript.

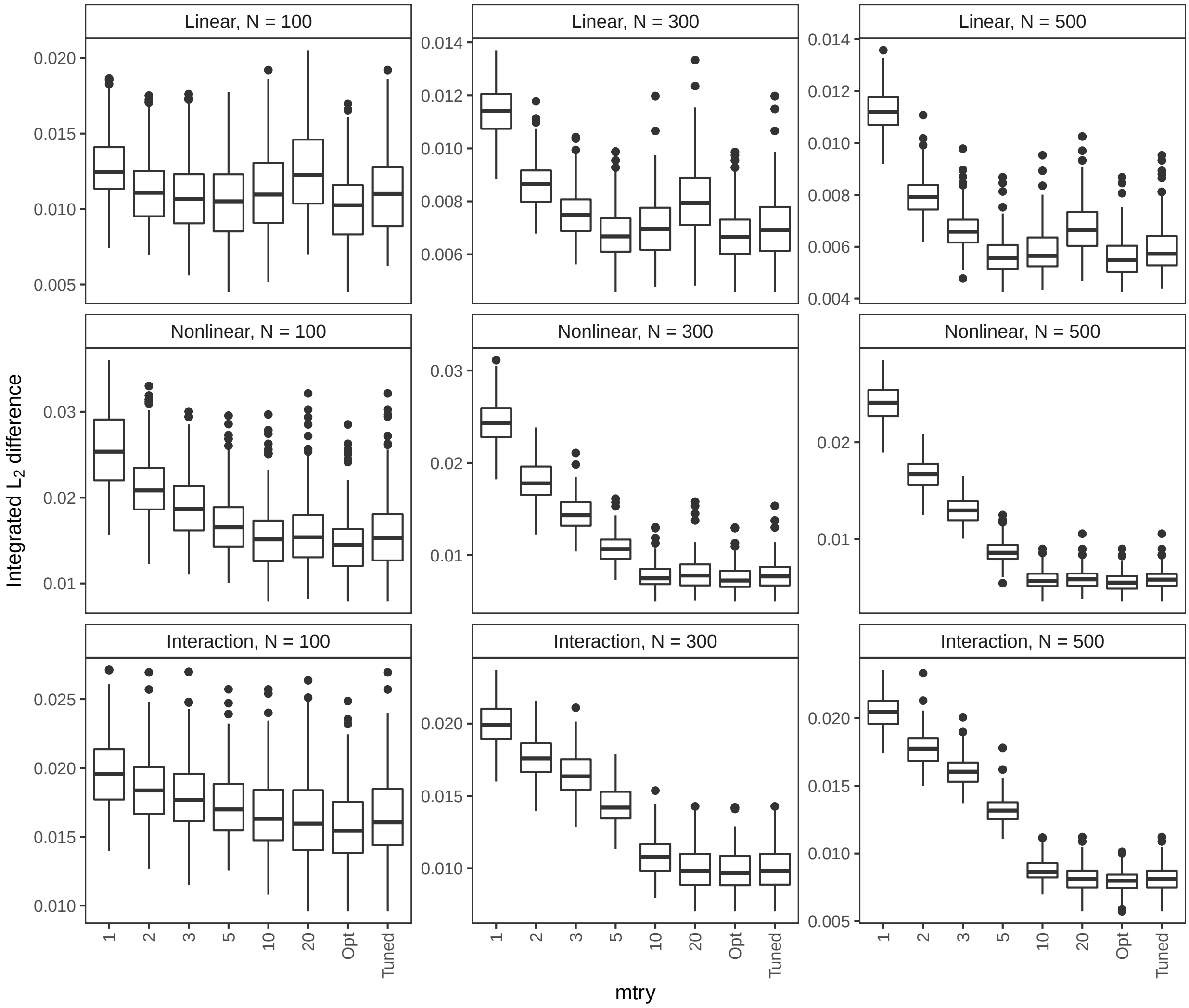

Figure S1.4: Integrated L_2 difference of RRF-TV with different *mtry* values under the PH setting. Datasets are generated with light right-censoring rate (20%), survival times generated from a Weibull-Increasing distribution. From the top row to the bottom, it gives results for the linear, nonlinear and the interaction survival relationship. From the first column to the last, it gives results for the number of subjects $N = 100, 300, 500$. In each plot, 1–RRF-TV with $mtry = 1$; 2–RRF-TV with $mtry = 2$; 3–RRF-TV with $mtry = 3$; 5–RRF-TV with $mtry = 5$; 10–RRF-TV with $mtry = 10$; 20–RRF-TV with $mtry = 20$; Opt–RRF-TV with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–RRF-TV with the value of *mtry* tuned by the "out-of-bag" tuning procedure.

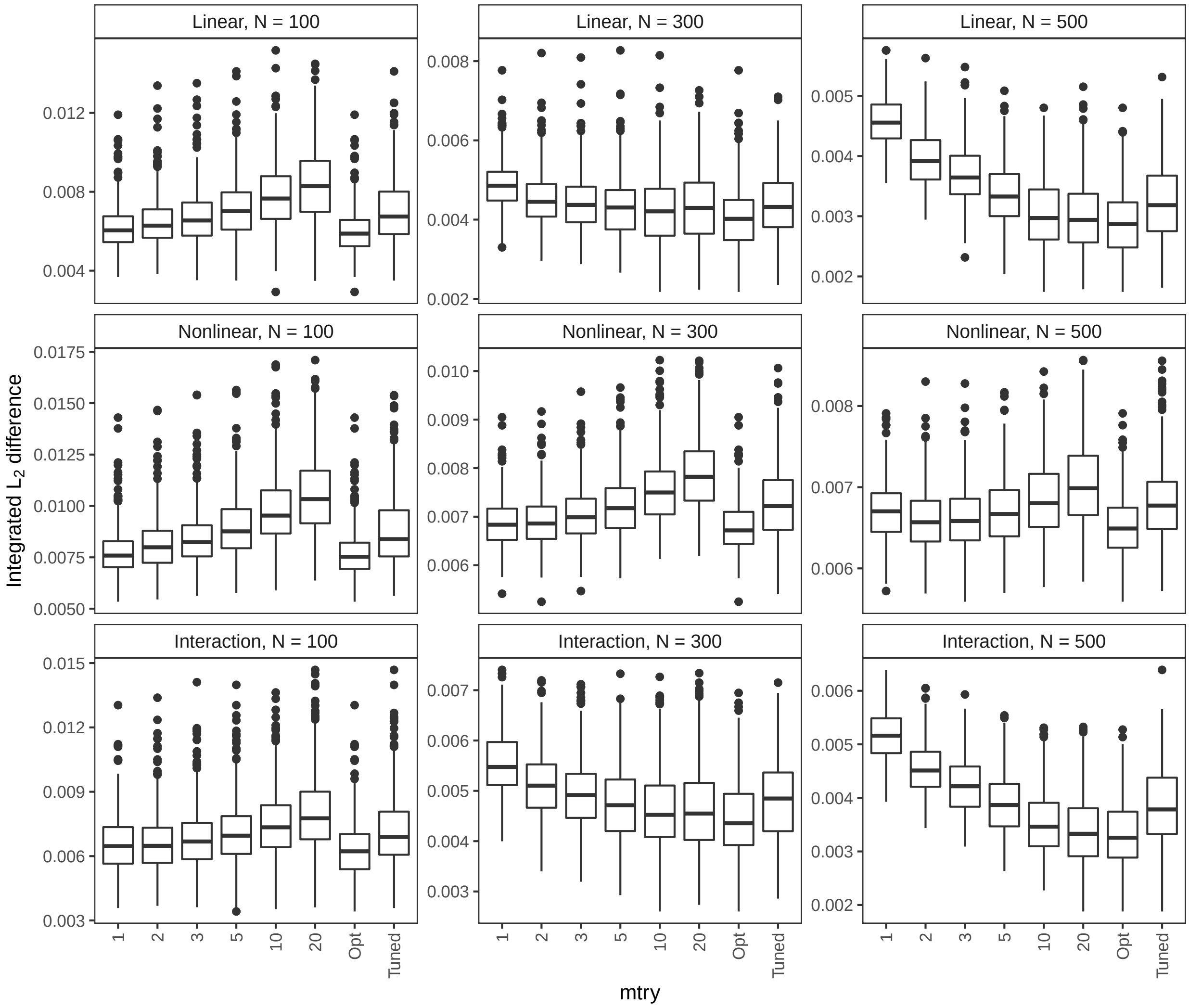

Figure S1.5: Integrated L_2 difference of RRF-TV with different *mtry* values under the non-PH setting. Datasets are generated with light right-censoring rate (20%), survival times generated from a Weibull-Increasing distribution. From the top row to the bottom, it gives results for the linear, nonlinear and the interaction survival relationship. From the first column to the last, it gives results for the number of subjects $N = 100, 300, 500$. In each plot, 1–RRF-TV with $mtry = 1$; 2–RRF-TV with $mtry = 2$; 3–RRF-TV with $mtry = 3$; 5–RRF-TV with $mtry = 5$; 10–RRF-TV with $mtry = 10$; 20–RRF-TV with $mtry = 20$; Opt–RRF-TV with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–RRF-TV with the value of *mtry* tuned by the "out-of-bag" tuning procedure.

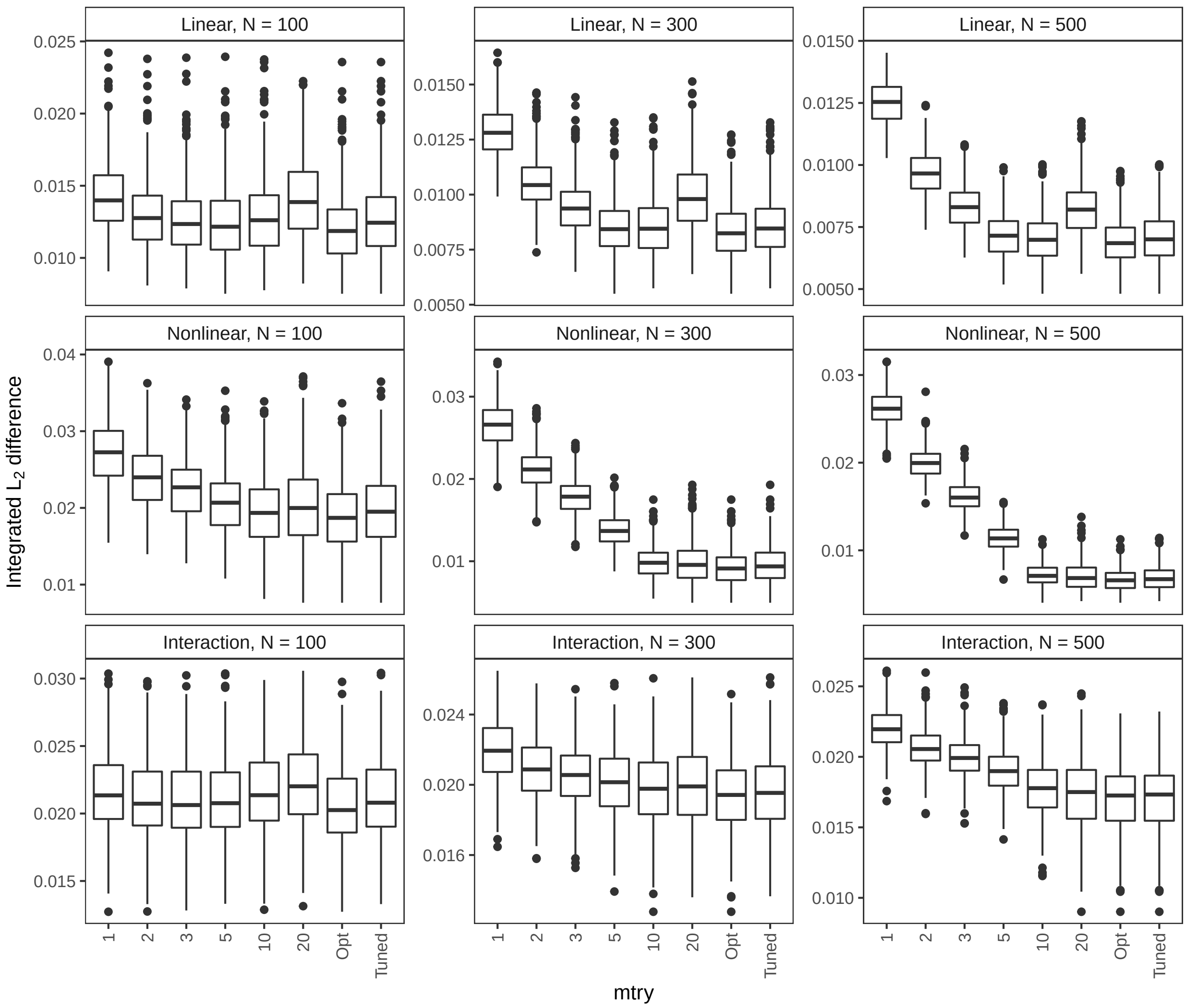

Figure S1.6: Integrated L_2 difference of TSF-TV with different *mtry* values under the PH setting. Datasets are generated with light right-censoring rate (20%), survival times generated from a Weibull-Increasing distribution. From the top row to the bottom, it gives results for the linear, nonlinear and the interaction survival relationship. From the first column to the last, it gives results for the number of subjects $N = 100, 300, 500$. In each plot, 1–TSF-TV with $mtry = 1$; 2–TSF-TV with $mtry = 2$; 3–TSF-TV with $mtry = 3$; 5–TSF-TV with $mtry = 5$; 10–TSF-TV with $mtry = 10$; 20–TSF-TV with $mtry = 20$; Opt–TSF-TV with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–TSF-TV with the value of *mtry* tuned by the "out-of-bag" tuning procedure.

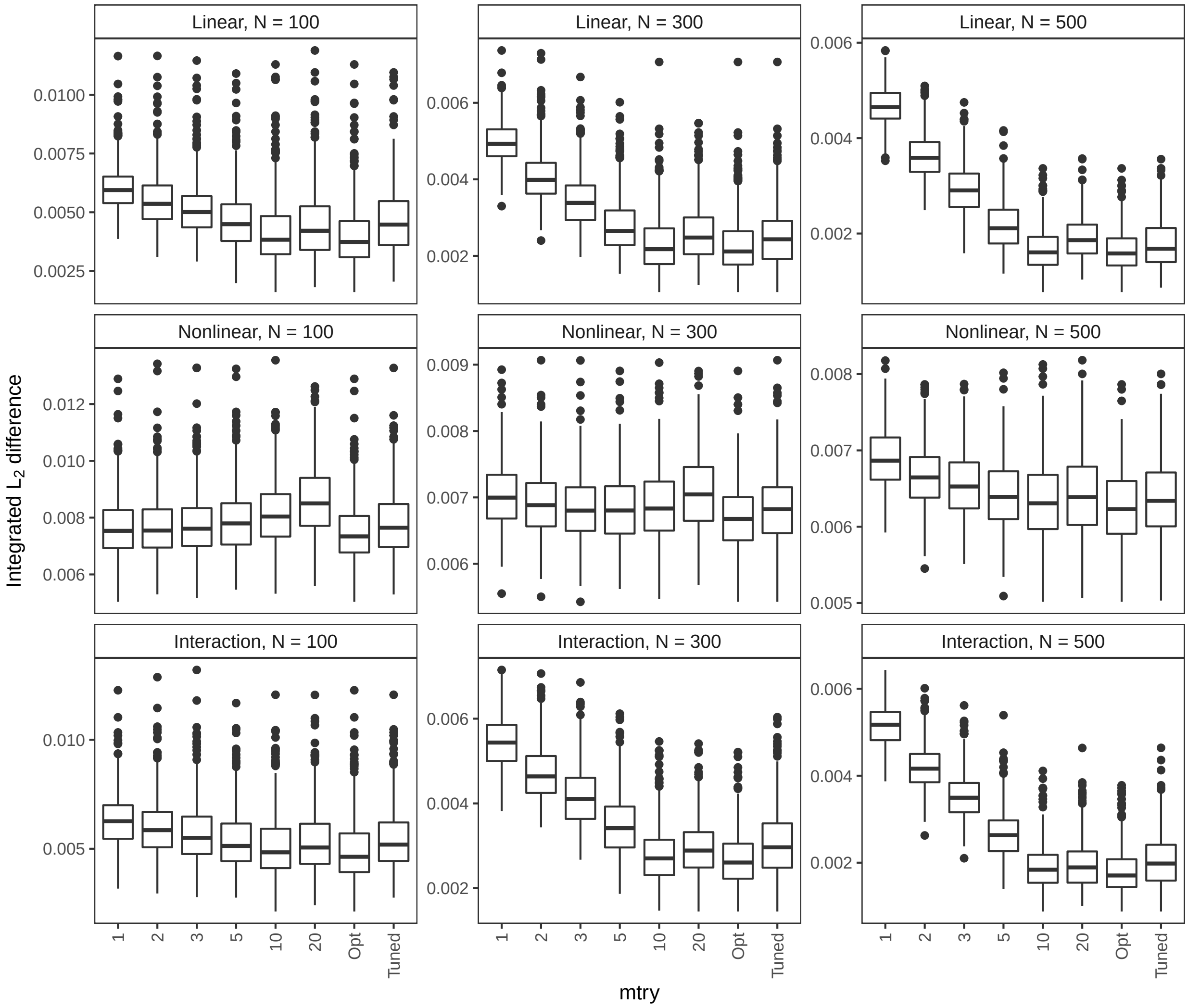

Figure S1.7: Integrated L_2 difference of TSF-TV with different *mtry* values under the non-PH setting. Datasets are generated with light right-censoring rate (20%), survival times generated from a Weibull-Increasing distribution. From the top row to the bottom, it gives results for the linear, nonlinear and the interaction survival relationship. From the first column to the last, it gives results for the number of subjects $N = 100, 300, 500$. In each plot, 1–TSF-TV with $mtry = 1$; 2–TSF-TV with $mtry = 2$; 3–TSF-TV with $mtry = 3$; 5–TSF-TV with $mtry = 5$; 10–TSF-TV with $mtry = 10$; 20–TSF-TV with $mtry = 20$; Opt–TSF-TV with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–TSF-TV with the value of *mtry* tuned by the "out-of-bag" tuning procedure.

Table S1.1 gives the performance comparison between each forest method with its default parameter settings and with the proposed parameter settings for the nonlinear survival relationship under the PH setting and the non-PH setting, respectively. Table S1.2 gives the performance comparison between each forest method with its default parameter settings and with the proposed parameter settings for the interaction survival relationship under the PH setting and the non-PH setting, respectively.

Figure S1.8 shows the effect of different numbers of trees for bootstrap samples on CIF-TV, RRF-TV and TSF-TV. The number of trees increases from 100 to 500. The results show that the improvement in integrated L_2 difference from more trees is negligible at 100.

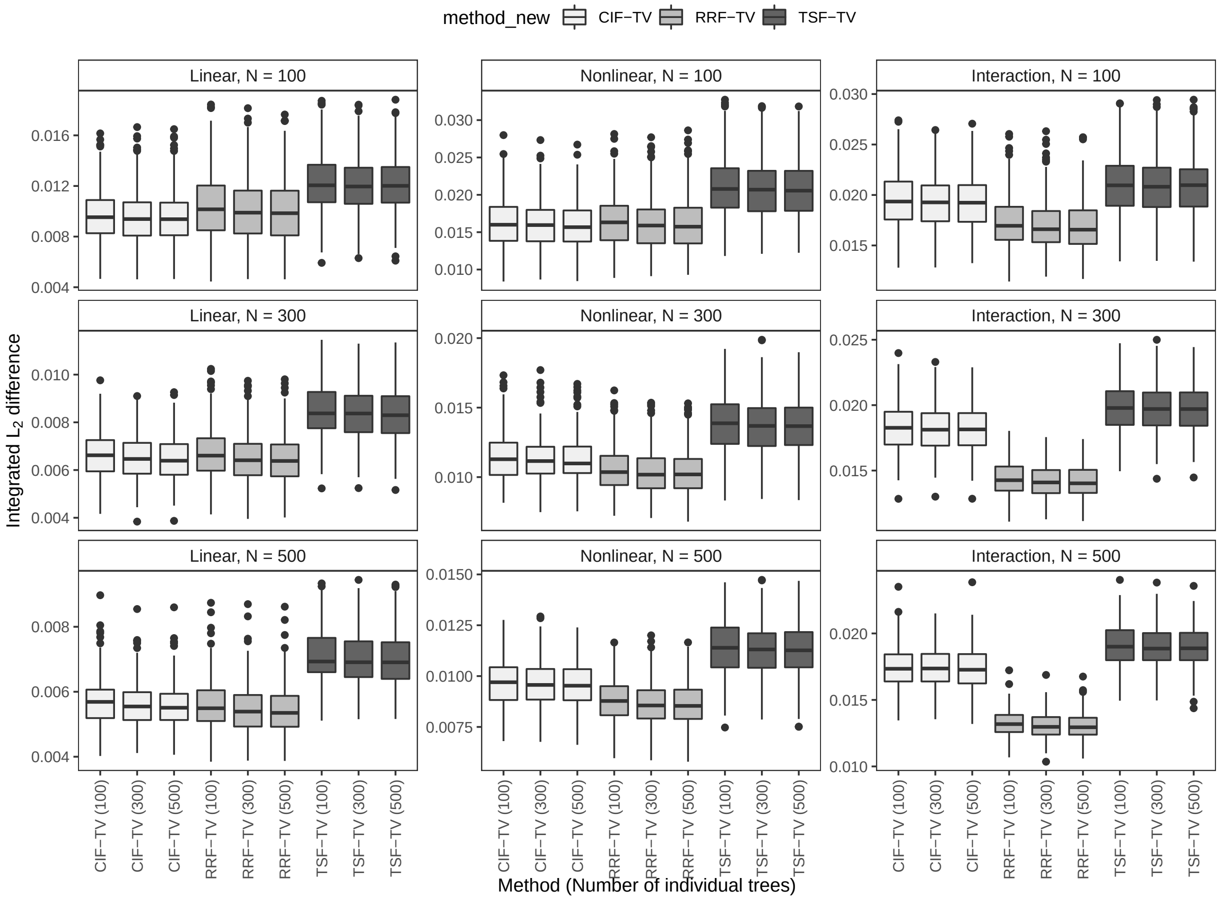

Figure S1.8: Effects of different numbers of trees for bootstrap samples on CIF-TV, RRF-TV and TSF-TV. Datasets are generated with survival times following a Weibull-Increasing distribution, light right-censoring rate 20% under the PH setting. The first column shows results for the number of subjects $N = 100$, second column for $N = 300$, bottom column for $N = 500$; the top row shows results for linear survival relationship, middle row for nonlinear, the bottom for interaction. In each of the plots, the set of boxplots lightly shaded shows the performance of CIF-TV with number of trees $100, 300, 500$ from left to right; the set moderately shaded shows the performance of RRF-TV with number of trees $100, 300, 500$ from left to right; the set heavily shaded shows the performance of TSF-TV with number of trees $100, 300, 500$ from left to right.

S1.7 Performance comparison for Weibull-Decreasing distribution

Figure S1.9 gives side-by-side integrated L_2 difference boxplots on datasets with survival times generated following a Weibull-Decreasing distribution.

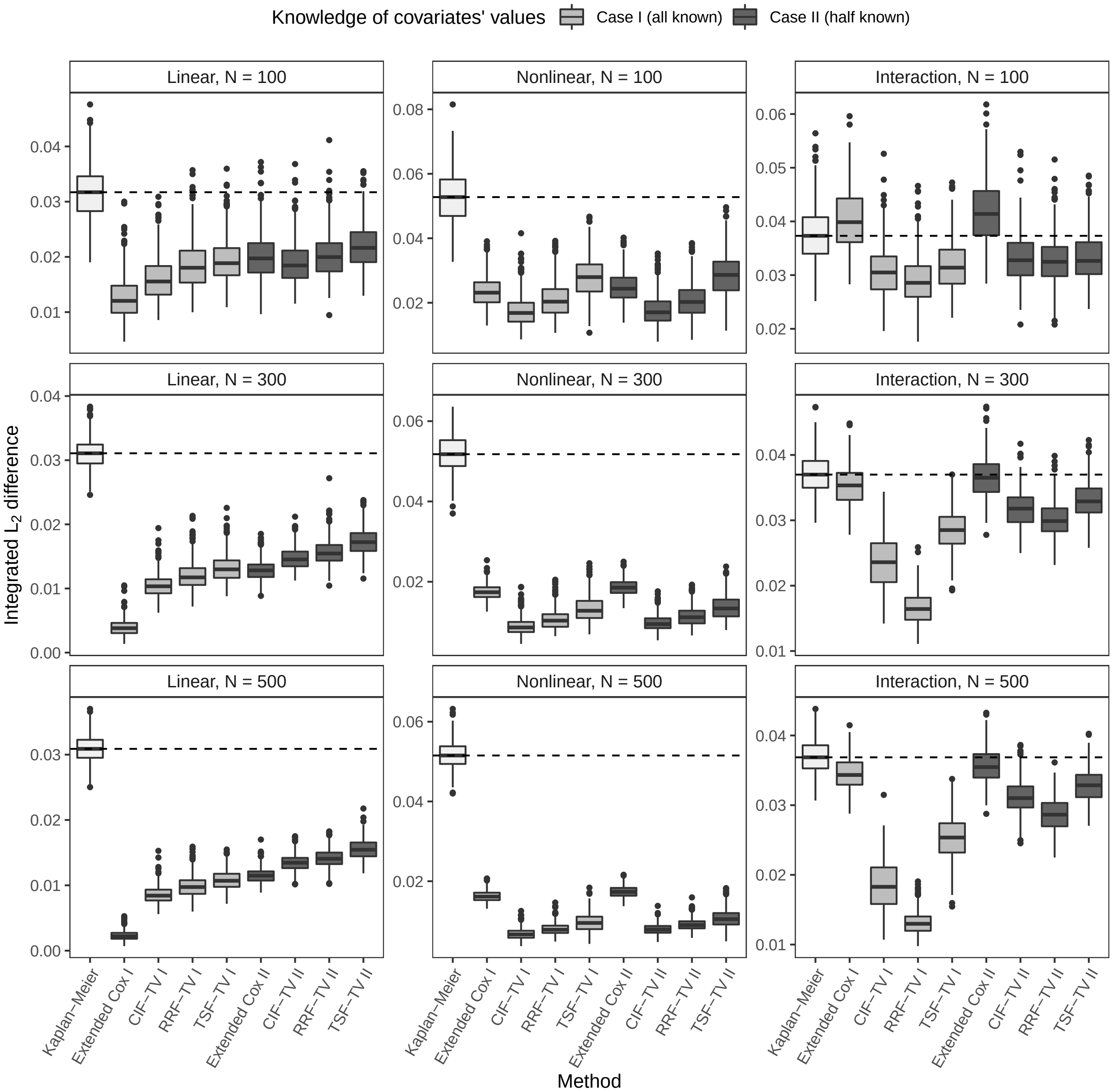

Figure S1.9: Boxplots of integrated L_2 difference for performance comparison under the PH setting. Datasets are generated with survival times following a Weibull-Decreasing distribution, light right-censoring rate 20%. The first row shows results for the number of subjects $N = 100$, second row for $N = 300$, bottom row for $N = 500$; the first column shows results for linear survival relationship, second column for nonlinear, the third column for interaction. The horizontal dashed line shows the median integrated L_2 difference of a Kaplan-Meier fit on the datasets. In each of the plots, the set of boxplots lightly shaded shows the performance of different methods on datasets with history of changes in covariates' values known; the set heavily shaded shows the performance on datasets with part of history of changes in covariates' values unknown.

S1.8 Analysis of variance

Table S1.3: ANOVA table for comparing group means of CIF-TV/RRF-TV improvement over a simple Kaplan-Meier fit.

Response: $(L_2(\text{Method}) - L_2(\text{KM}))/L_2(\text{KM})$						
Method:	CIF-TV			RRF-TV		
	Sum Sq	Df	Pr(>F)	Sum Sq	Df	Pr(>F)
Censoring rate	0.1849	1	0.0046	0.2443	1	0.0015
Knowledge	1.0512	1	< 0.0001	0.8215	1	< 0.0001
Relationship	1.6353	2	< 0.0001	0.3697	2	0.0005
Sample size	1.5899	2	< 0.0001	3.1521	2	< 0.0001
Scenario	0.4215	1	< 0.0001	0.4597	1	< 0.0001
Setting	10.6751	1	< 0.0001	15.9043	1	< 0.0001
SNR	0.0000	1	0.9729	0.0633	1	0.1039

S2 Proposed forests for time-invariant covariate survival data

The section is organized as follows. Section S2.1 describes the model setup, Section S2.2 provides the performance evaluation of the "out-of-bag" tuning procedure for *mtry* and gives comparative performance between the proposed parameter settings and the default settings. The overall performance comparison among the Cox model, the three forest methods with the proposed parameter settings, as well as the best method and the method selected by the IBS-based 10-fold CV rule are given in Section S2.3. All results with integrated L_2 difference are computed with $\tau_i = \widetilde{T}_i$. The comparative performance of the different methods was broadly similar when using $\tau_i = \widetilde{\tau} = \max_j \widetilde{T}_j$.

Here we use the proposed forest methods for left-truncated right-censored survival data with time-invariant covariates, the proposed forest methods and the transformation forest method are referred to as LTRC CIF, LTRC RRF and LTRC TSF, respectively.

S2.1 Model setup

We generate left-truncated right-censored survival time data with time-invariant covariates as follows. The left-truncation time T_0 is generated as a $\mathcal{U}(0, L)$ random variable with some constant $L > 0$. The event time T is randomly generated with a Weibull distribution. If the generated $T < T_0$, i.e. the event time is less than the left-truncation time, then this observation is discarded. Otherwise, the observation is retained, with censoring time $C = T_0 + D$, where $D \sim \textit{weibull}(\text{shape} = 2, \text{scale} = \lambda_D)$ has an weibull distribution. The parameter λ_D is selected to ensure 20% censoring rate. If $C < T$, then this observation is censored $(\Delta = 0)$, otherwise the survival time T is observed $(\Delta = 1)$. Note that D and T_0 are both independently generated from T and from each other. The observed response for each observation is a triplet $(T_0, \widetilde{T}, \Delta)$, where $Y = \min(T, C)$.

There are 20 covariates in total, with the first six determining the survival times. $X_1, X_3 \sim$ Bern(0.5), $X_2, X_4 \sim$ Unif(0, 1), X_5 follows a categorical distribution with possible outcomes $\{1, 2, 3, 4, 5\}$ with equal probability, X_6 follows a categorical distribution with possible outcomes $\{0, 1, 2\}$ with equal probability. Among the rest of the 14 noise covariates, X_7, X_{10}, X_{15}, X_{17}, $X_{20} \sim$ Unif(0, 1), $X_8 \sim$ Unif(1, 2), X_{11}, X_{13}, X_{16}, $X_{19} \sim$ Bern(0.5), X_{12} and X_{18} both follow a categorical distribution with possible outcomes $\{0, 1, 2\}$ with equal probability, X_9 and X_{14} both follow a categorical distribution with possible outcomes $\{1, 2, 3, 4, 5\}$ with equal probability. The survival times generating schemes with linear, nonlinear and interaction survival relationships are the same as in the time-varying cases, described in Section 3.2 in the manuscript.

S2.2 Regulating the construction of trees in forests

In the simulations the value of *mtry* is tuned based on the "out-of-bag observations," and the values of *minprob* and *minbucket* are set to be the maximum of the default value and the square root of the number of pseudo-subject observations n.

Figures S2.1 and S2.2 show how LTRC CIF performs with different values of *mtry* under the PH setting and non-PH setting, respectively. The datasets are generated with survival times following a Weibull-Increasing distribution, light (right-)censoring rate. The results for LTRC RRF and LTRC TSF are similar, as given in Figures S2.3 to S2.6.

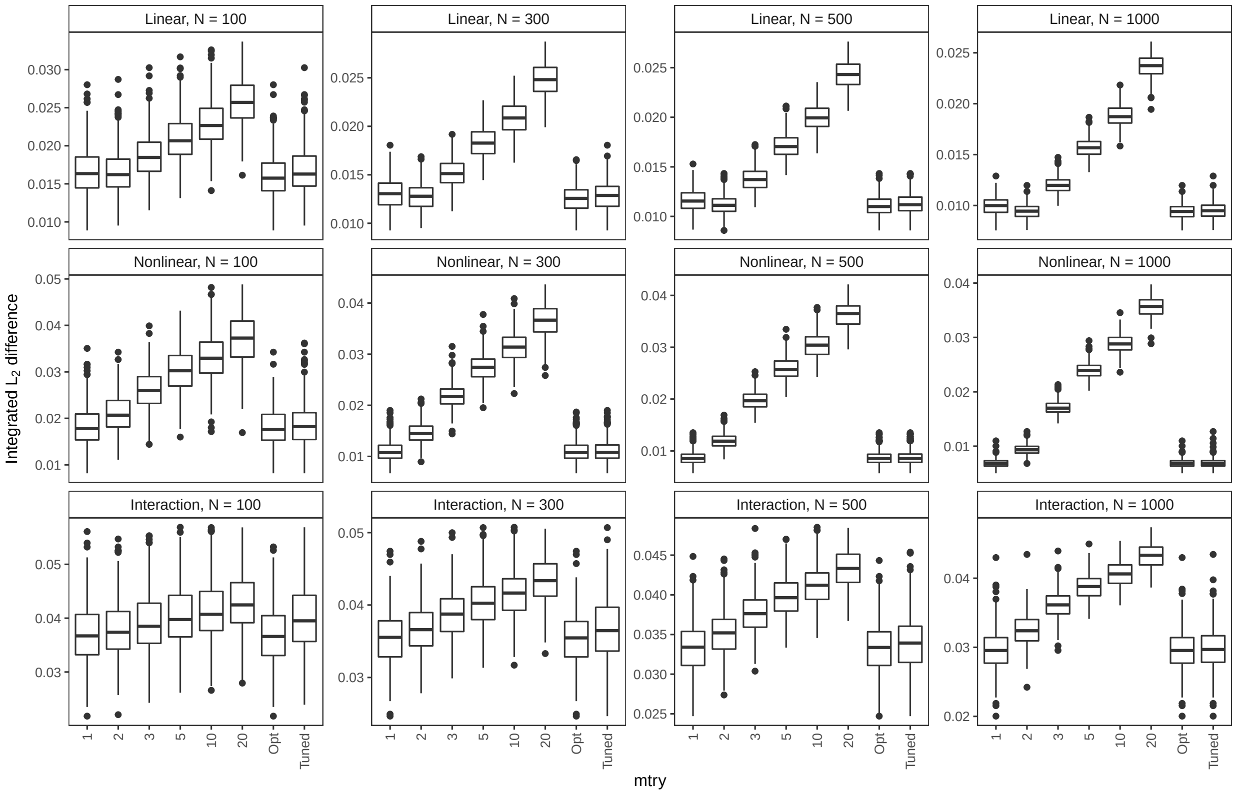

Figure S2.1: Integrated L_2 difference of LTRC CIF with different $mtry$ values distribution under the PH setting. Datasets are generated with time-invariant covariates, light right-censoring rate (20%), left-truncated and right-censored survival times following a Weibull-Increasing. From the top row to the bottom, are given results for the linear, nonlinear and interaction survival relationship. From the first column to the last, are given results for the number of subjects $N = 100, 300, 500, 1000$. In each plot, 1–LTRC CIF with $mtry = 1$; 2–LTRC CIF with $mtry = 2$; 3–LTRC CIF with $mtry = 3$; 5–LTRC CIF with $mtry = 5$; 10–LTRC CIF with $mtry = 10$; 20–LTRC CIF with $mtry = 20$; Opt–LTRC CIF with value of $mtry$ that gives the smallest Integrated L_2 difference in each round; Tuned–LTRC CIF with the value of $mtry$ tuned by the "out-of-bag" tuning procedure. The default value in LTRC CIF is $mtry = 5$.

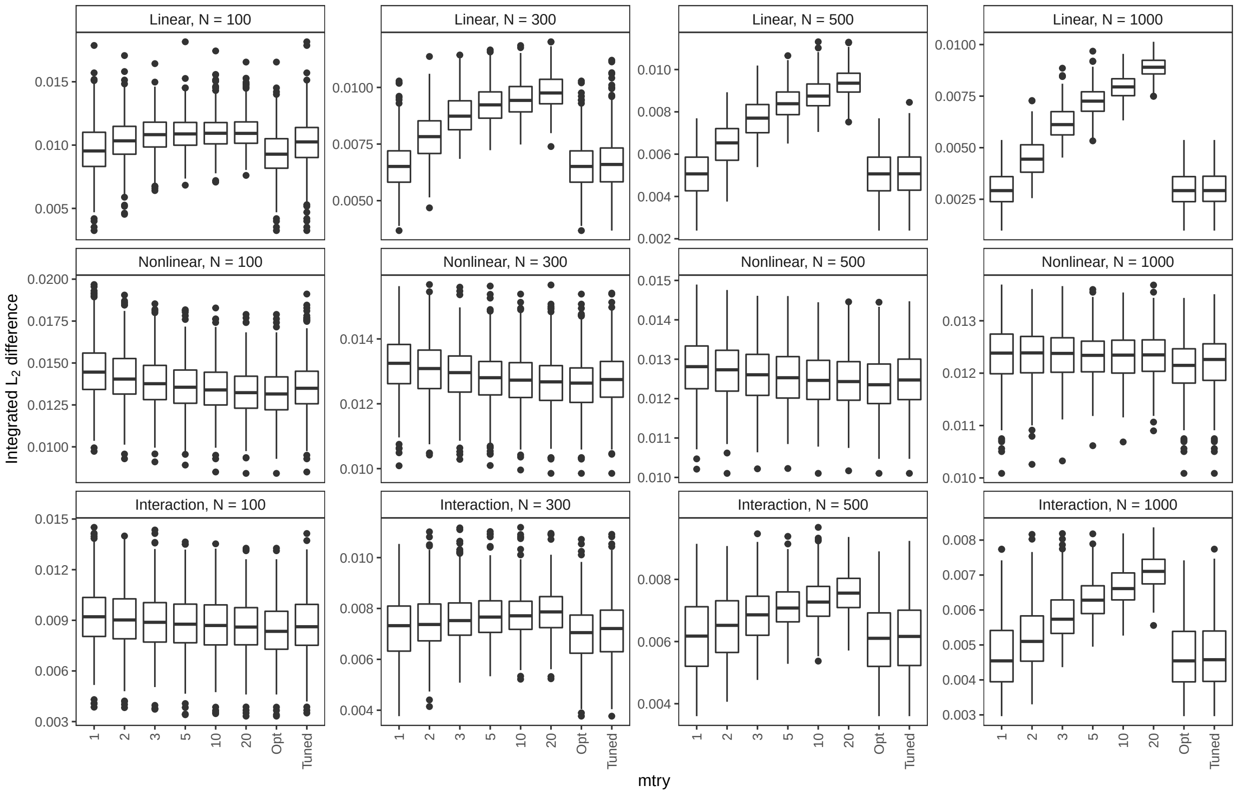

Figure S2.2: Integrated L_2 difference of LTRC CIF with different *mtry* values under the non-PH setting. Datasets are generated with time-invariant covariates, light right-censoring rate (20%), left-truncated and right-censored survival times following a Weibull-Increasing distribution. From the top row to the bottom, are given results for the linear, nonlinear and interaction survival relationship. From the first column to the last, are given results for the number of subjects $N = 100, 300, 500, 1000$. In each plot, 1–LTRC CIF with $mtry = 1$; 2–LTRC CIF with $mtry = 2$; 3–LTRC CIF with $mtry = 3$; 5–LTRC CIF with $mtry = 5$; 10–LTRC CIF with $mtry = 10$; 20–LTRC CIF with $mtry = 20$; Opt–LTRC CIF with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–LTRC CIF with the value of *mtry* tuned by the "out-of-bag" tuning procedure. The default value in LTRC CIF is $mtry = 5$.

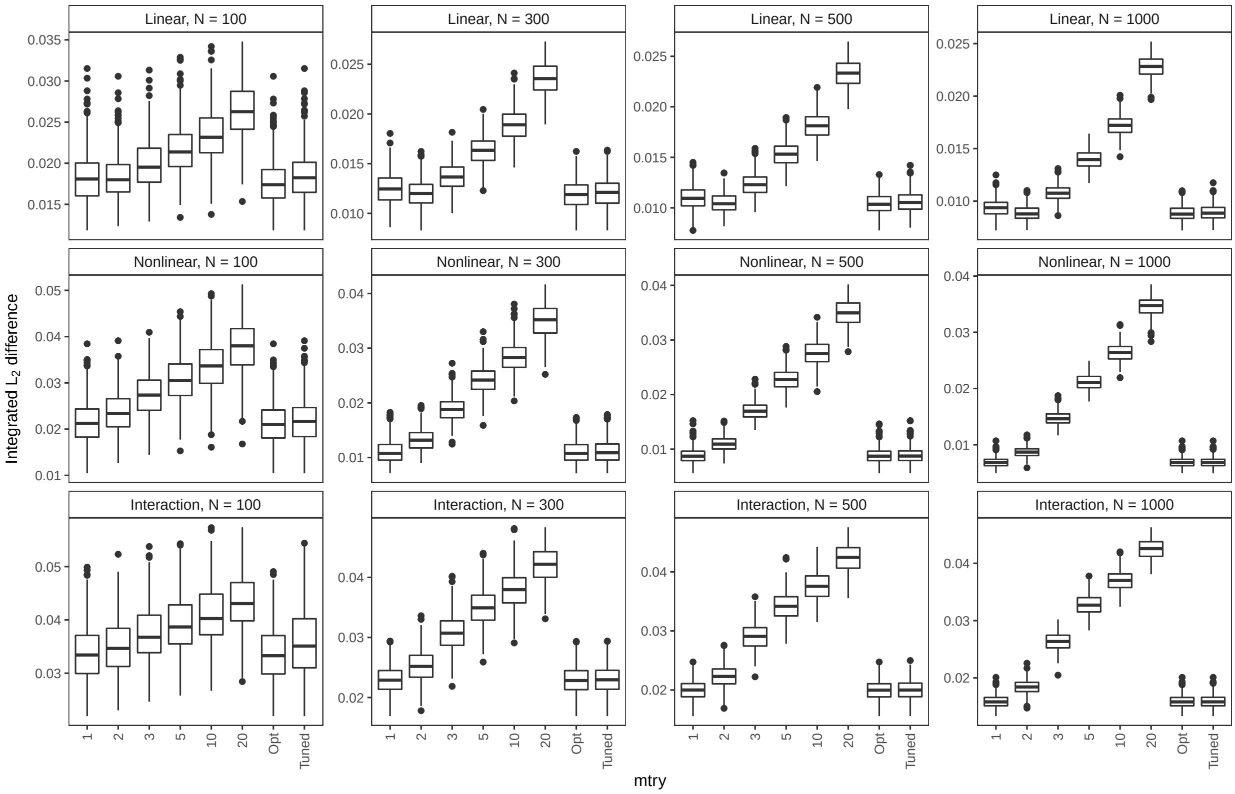

Figure S2.3: Integrated L_2 difference of LTRC RRF with different *mtry* values under the PH setting. Datasets are generated with time-invariant covariates, light right-censoring rate (20%), left-truncated and right-censored survival times following a Weibull-Increasing distribution. From the top row to the bottom, are given results for the linear, nonlinear and interaction survival relationship. From the first column to the last, are given results for the number of subjects $N = 100, 300, 500, 1000$. In each plot, 1–LTRC RRF with $mtry = 1$; 2–LTRC RRF with $mtry = 2$; 3–LTRC RRF with $mtry = 3$; 5–LTRC RRF with $mtry = 5$; 10–LTRC RRF with $mtry = 10$; 20–LTRC RRF with $mtry = 20$; Opt–LTRC RRF with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–LTRC RRF with the value of *mtry* tuned by the "out-of-bag" tuning procedure. The default value in LTRC RRF is $mtry = 5$.

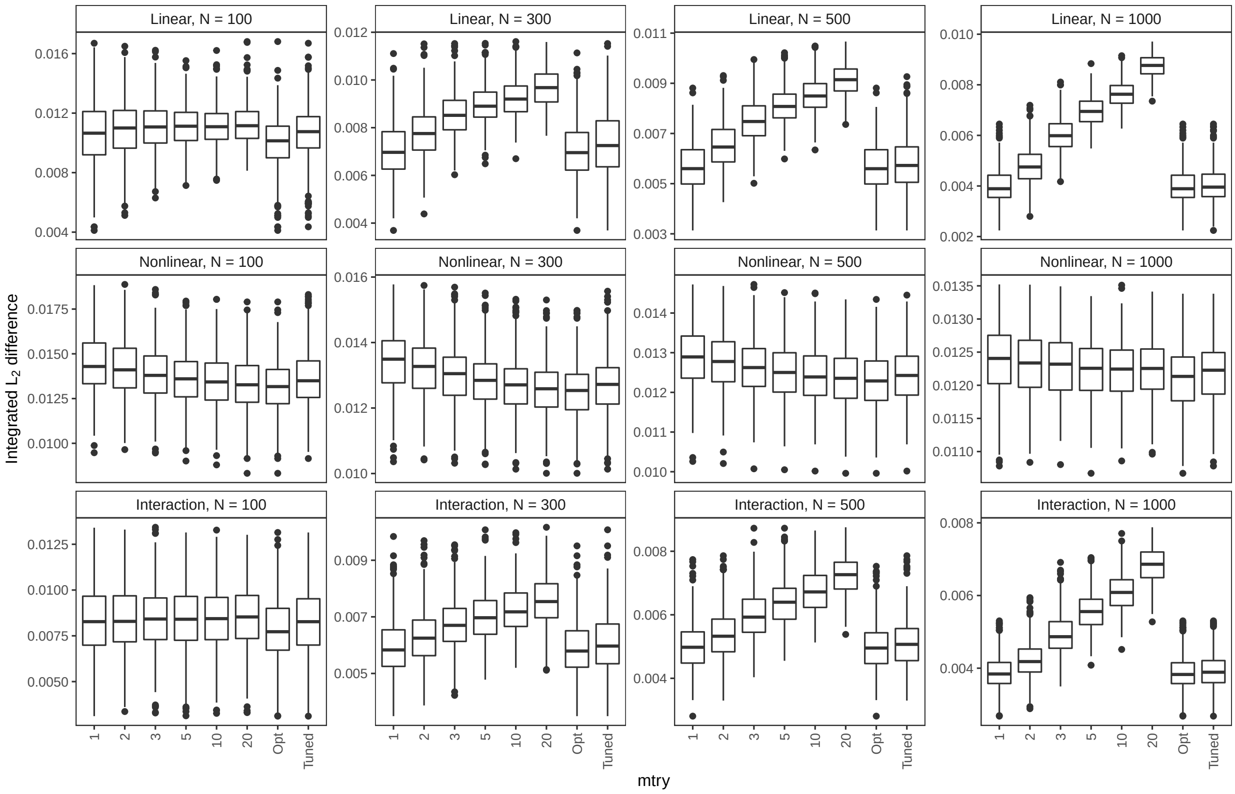

Figure S2.4: Integrated L_2 difference of LTRC RRF with different *mtry* values under the non-PH setting. Datasets are generated with time-invariant covariates, light right-censoring rate (20%), left-truncated and right-censored survival times following a Weibull-Increasing distribution. From the top row to the bottom, are given results for the linear, nonlinear and interaction survival relationship. From the first column to the last, are given results for the number of subjects $N = 100, 300, 500, 1000$. In each plot, 1–LTRC RRF with $mtry = 1$; 2–LTRC RRF with $mtry = 2$; 3–LTRC RRF with $mtry = 3$; 5–LTRC RRF with $mtry = 5$; 10–LTRC RRF with $mtry = 10$; 20–LTRC RRF with $mtry = 20$; Opt–LTRC RRF with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–LTRC RRF with the value of *mtry* tuned by the "out-of-bag" tuning procedure. The default value in LTRC RRF is $mtry = 5$.

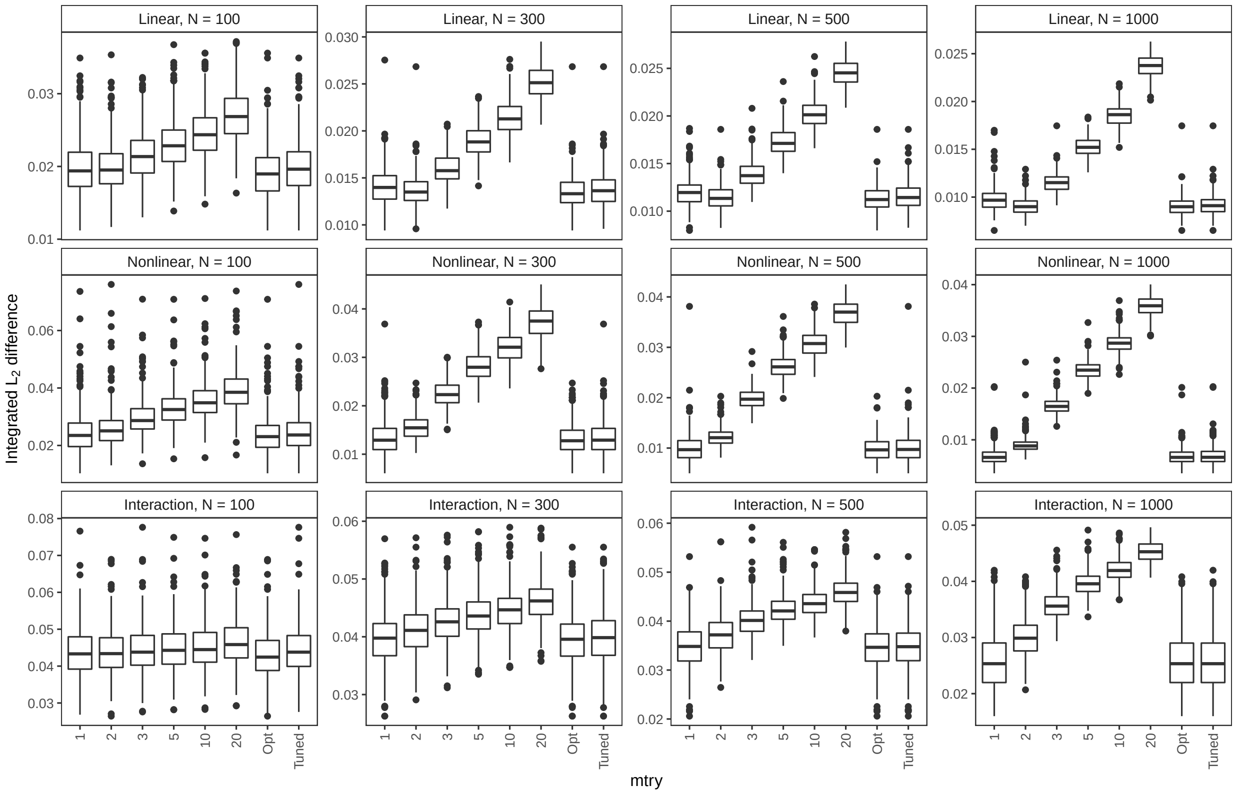

Figure S2.5: Integrated L_2 difference of LTRC TSF with different *mtry* values under the PH setting. Datasets are generated with time-invariant covariates, light right-censoring rate (20%), left-truncated and right-censored survival times following a Weibull-Increasing distribution. From the top row to the bottom, are given results for the linear, nonlinear and interaction survival relationship. From the first column to the last, are given results for the number of subjects $N = 100, 300, 500, 1000$. In each plot, 1–LTRC TSF with $mtry = 1$; 2–LTRC TSF with $mtry = 2$; 3–LTRC TSF with $mtry = 3$; 5–LTRC TSF with $mtry = 5$; 10–LTRC TSF with $mtry = 10$; 20–LTRC TSF with $mtry = 20$; Opt–LTRC TSF with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–LTRC TSF with the value of *mtry* tuned by the "out-of-bag" tuning procedure. The default value in LTRC TSF is $mtry = 5$.

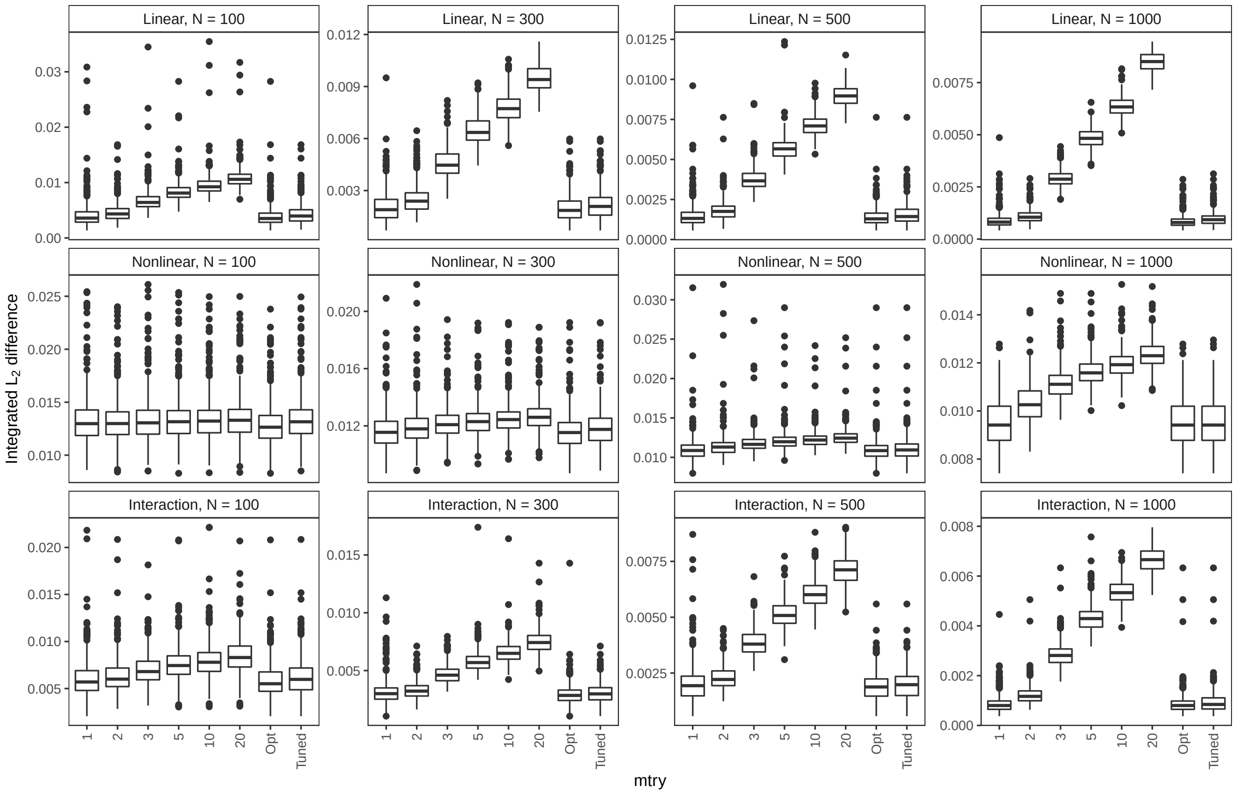

Figure S2.6: Integrated L_2 difference of LTRC TSF with different *mtry* values under the non-PH setting. Datasets are generated with time-invariant covariates, light right-censoring rate (20%), left-truncated and right-censored survival times following a Weibull-Increasing distribution. From the top row to the bottom, are given results for the linear, nonlinear and interaction survival relationship. From the first column to the last, are given results for the number of subjects $N = 100, 300, 500, 1000$. In each plot, 1–LTRC TSF with $mtry = 1$; 2–LTRC TSF with $mtry = 2$; 3–LTRC TSF with $mtry = 3$; 5–LTRC TSF with $mtry = 5$; 10–LTRC TSF with $mtry = 10$; 20–LTRC TSF with $mtry = 20$; Opt–LTRC TSF with value of *mtry* that gives the smallest Integrated L_2 difference in each round; Tuned–LTRC TSF with the value of *mtry* tuned by the "out-of-bag" tuning procedure. The default value in LTRC TSF is $mtry = 5$.

Table S2.1 gives performance comparison between each forest method with its default parameter settings and with the proposed parameter settings, using datasets generated with survival times following a Weibull-Increasing distribution, nonlinear relationship, with 20% right-censoring rate, as an example.

In Table S2.1, positive numbers indicate a decrease in integrated L_2 difference compared to a Cox model on the dataset, while negative numbers indicate an increase. The absolute value of the numbers represents the size of the difference between the integrated L_2 difference of the candidate and that of a Cox model. The two tables show that forests with the proposed parameter settings can provide improved performance over those with default parameter settings across almost all different numbers of subjects N by a substantial amount, under both PH and non-PH settings.

S2.3 Performance comparison and IBS-based CV model selection

Figures S2.7 and S2.8 give side-by-side integrated L_2 difference boxplots on datasets with light (right-)censoring rate, survival times generated following a Weibull-I distribution under the

PH setting and the non-PH setting, respectively. These figures compare the performance of the four methods, the best method and the method chosen by IBS-based 10-fold CV rule with $\tau_i = 1.5\widetilde{T}_i$.

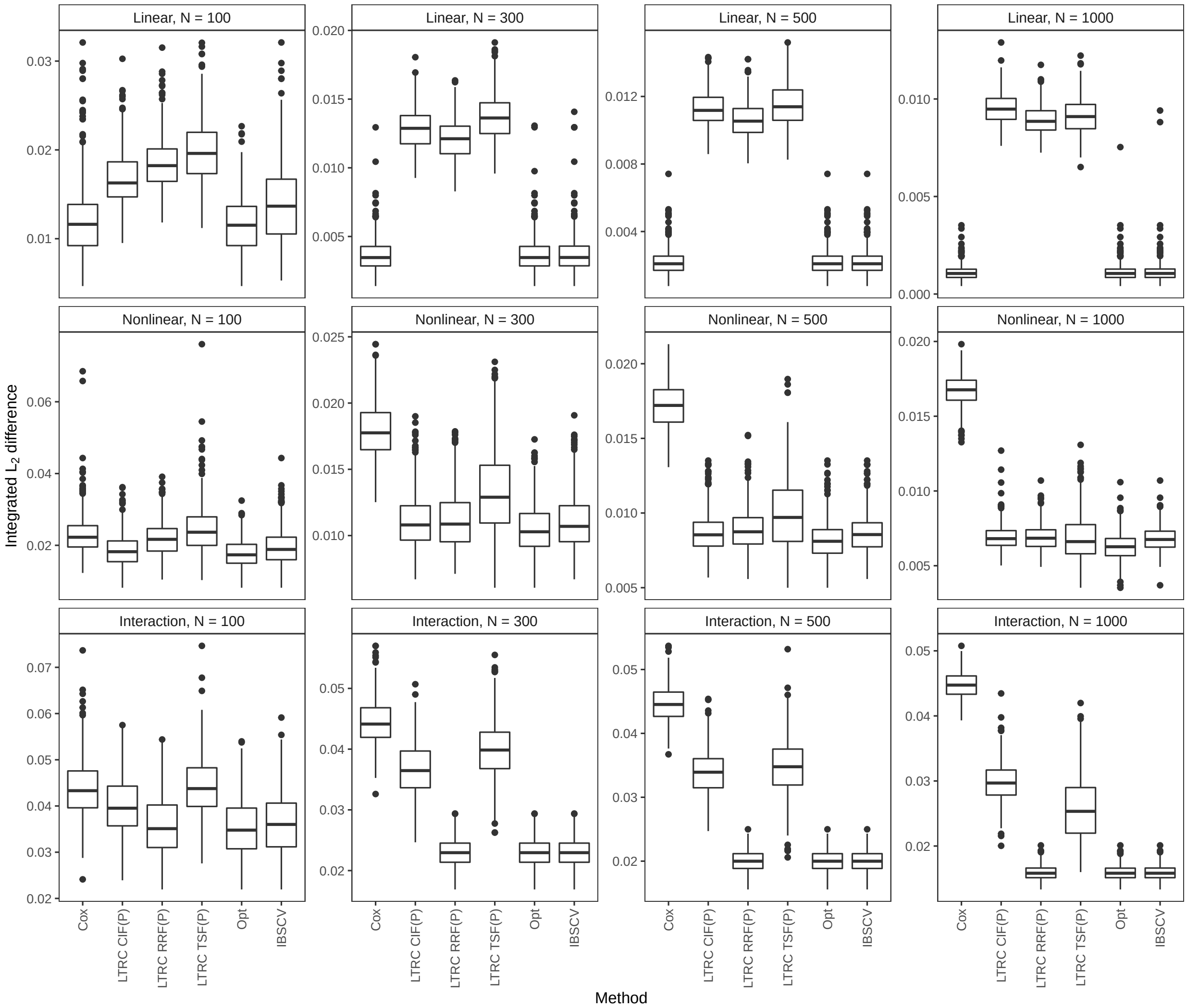

Figure S2.7: Boxplots of integrated L_2 difference for performance comparison across different survival relationships and different numbers of subjects N, under the PH setting. Datasets are generated with time-invariant covariates, left-truncated right-censored survival times following a Weibull-Increasing distribution. The first row shows results for the number of subjects $N = 100$, second row for $N = 300$, third row for $N = 500$, bottom row for $N = 1000$; the first column shows results for linear survival relationship, second column for nonlinear, the third column for interaction. In each plot, LTRC CIF(P)–LTRC CIF with proposed parameter settings; LTRC RRF(P)–LTRC RRF with proposed parameter settings; LTRC TSF(P)–LTRC TSF with proposed parameter settings; Opt–Best method; IBSCV–Method chosen by IBS-based 10-fold CV.

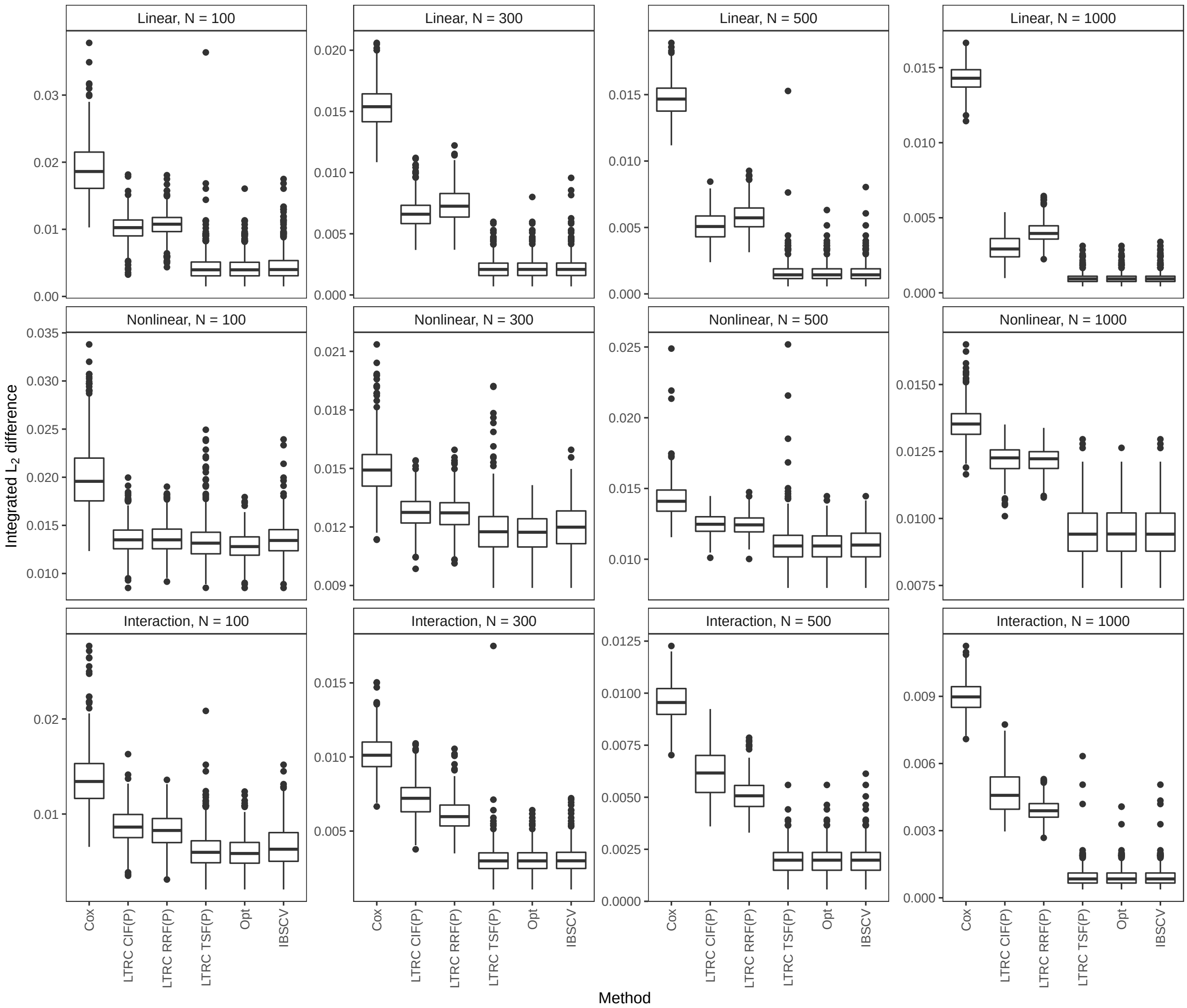

Figure S2.8: Boxplots of integrated L_2 difference for performance comparison across different survival relationships and different numbers of subjects N, under the non-PH setting. Datasets are generated with time-invariant covariates, left-truncated right-censored survival times following a Weibull-Increasing distribution. The first row shows results for the number of subjects $N = 100$, second row for $N = 300$, third row for $N = 500$, bottom row for $N = 1000$; the first column shows results for linear survival relationship, second column for nonlinear, the third column for interaction. In each plot, LTRC CIF(P)–LTRC CIF with proposed parameter settings; LTRC RRF(P)–LTRC RRF with proposed parameter settings; LTRC TSF(P)–LTRC TSF with proposed parameter settings; Opt–Best method; IBSCV–Method chosen by IBS-based 10-fold CV.

References

Table S1.1: Comparison between forests with default (D) and proposed parameter settings (P) across different numbers of subjects N given a nonlinear survival relationship. Given a method A, each cell value are given as mean $\pm$ one standard deviation of $\big(L_2(\text{KM}) - L_2(\text{A})\big)/L_2(\text{KM})$ based on all simulations. The mean value is the average % decrease in integrated L_2 difference compared to the Kaplan-Meier fit.

Proportional hazards setting							
N	Case I. All changes in covariates' values are known						
	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	0.55 ± 0.10	0.53 ± 0.14	0.66 ± 0.12	0.52 ± 0.14	0.56 ± 0.15	0.41 ± 0.17	0.46 ± 0.13
300	0.65 ± 0.04	0.64 ± 0.07	0.82 ± 0.04	0.68 ± 0.06	0.79 ± 0.05	0.55 ± 0.09	0.74 ± 0.06
500	0.66 ± 0.03	0.68 ± 0.05	0.87 ± 0.03	0.72 ± 0.05	0.84 ± 0.03	0.61 ± 0.07	0.81 ± 0.04
N	Case II. Half of changes in covariates' values are unknown						
	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	0.52 ± 0.10	0.59 ± 0.10	0.65 ± 0.10	0.54 ± 0.13	0.57 ± 0.13	0.47 ± 0.12	0.45 ± 0.12
300	0.63 ± 0.04	0.71 ± 0.05	0.80 ± 0.04	0.71 ± 0.05	0.77 ± 0.05	0.66 ± 0.06	0.72 ± 0.06
500	0.64 ± 0.03	0.75 ± 0.04	0.84 ± 0.03	0.75 ± 0.04	0.81 ± 0.03	0.72 ± 0.04	0.78 ± 0.04
Non-proportional hazards setting							
N	Case I. All changes in covariates' values are known						
	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	-0.42 ± 0.22	-0.33 ± 0.30	0.02 ± 0.14	-0.40 ± 0.29	0.02 ± 0.15	-0.20 ± 0.27	0.06 ± 0.11
300	-0.13 ± 0.06	-0.31 ± 0.20	0.08 ± 0.07	-0.35 ± 0.20	0.10 ± 0.07	-0.26 ± 0.19	0.11 ± 0.06
500	-0.07 ± 0.04	-0.25 ± 0.16	0.12 ± 0.05	-0.29 ± 0.15	0.12 ± 0.07	-0.18 ± 0.17	0.16 ± 0.06
N	Case II. Half of changes in covariates' values are unknown						
	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	-0.45 ± 0.23	-0.16 ± 0.23	0.02 ± 0.12	-0.22 ± 0.23	0.02 ± 0.13	-0.06 ± 0.19	0.03 ± 0.10
300	-0.13 ± 0.07	-0.15 ± 0.14	0.04 ± 0.05	-0.20 ± 0.16	0.07 ± 0.05	-0.08 ± 0.14	0.04 ± 0.04
500	-0.08 ± 0.04	-0.11 ± 0.12	0.05 ± 0.03	-0.15 ± 0.12	0.07 ± 0.07	-0.04 ± 0.11	0.05 ± 0.03

Table S1.2: Comparison between forests with default (D) and proposed parameter settings (P) across different numbers of subjects N given an interaction survival relationship. Given a method A, each cell value are given as mean $\pm$ one standard deviation of $\big(L_2(\mathrm{KM}) - L_2(\mathrm{A})\big)/L_2(\mathrm{KM})$ based on all simulations. The mean value is the average % decrease in integrated L_2 difference compared to the Kaplan-Meier fit.

Proportional hazards setting							
	Case I. All changes in covariates' values are known						
N	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	-0.02 ± 0.12	0.21 ± 0.19	0.26 ± 0.11	0.29 ± 0.18	0.35 ± 0.16	0.15 ± 0.20	0.19 ± 0.08
300	0.05 ± 0.05	0.29 ± 0.11	0.41 ± 0.11	0.40 ± 0.10	0.61 ± 0.07	0.17 ± 0.13	0.23 ± 0.06
500	0.06 ± 0.03	0.33 ± 0.09	0.55 ± 0.11	0.43 ± 0.07	0.68 ± 0.04	0.22 ± 0.11	0.33 ± 0.09
	Case II. Half of changes in covariates' values are unknown						
N	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	-0.05 ± 0.13	0.24 ± 0.14	0.19 ± 0.08	0.24 ± 0.17	0.21 ± 0.16	0.19 ± 0.12	0.14 ± 0.06
300	0.02 ± 0.05	0.28 ± 0.08	0.16 ± 0.05	0.30 ± 0.08	0.24 ± 0.10	0.23 ± 0.08	0.12 ± 0.03
500	0.04 ± 0.03	0.30 ± 0.06	0.16 ± 0.04	0.33 ± 0.06	0.26 ± 0.07	0.24 ± 0.06	0.11 ± 0.03
Non-proportional hazards setting							
	Case I. All changes in covariates' values are known						
N	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	-0.34 ± 0.21	-0.27 ± 0.31	0.09 ± 0.18	-0.20 ± 0.27	0.09 ± 0.14	-0.11 ± 0.30	0.25 ± 0.18
300	-0.09 ± 0.06	-0.14 ± 0.19	0.32 ± 0.12	-0.06 ± 0.18	0.27 ± 0.10	-0.11 ± 0.22	0.55 ± 0.11
500	-0.05 ± 0.03	-0.09 ± 0.15	0.44 ± 0.11	0.02 ± 0.13	0.38 ± 0.11	0.00 ± 0.20	0.70 ± 0.09
	Case II. Half of changes in covariates' values are unknown						
N	Extended Cox	CIF-TV(D)	CIF-TV(P)	RRF-TV(D)	RRF-TV(P)	TSF-TV(D)	TSF-TV(P)
100	-0.35 ± 0.22	-0.07 ± 0.22	0.05 ± 0.13	-0.08 ± 0.22	0.07 ± 0.14	0.05 ± 0.22	0.13 ± 0.13
300	-0.10 ± 0.05	0.01 ± 0.14	0.13 ± 0.07	0.02 ± 0.14	0.15 ± 0.06	0.10 ± 0.16	0.25 ± 0.11
500	-0.06 ± 0.03	0.06 ± 0.12	0.17 ± 0.08	0.06 ± 0.11	0.18 ± 0.06	0.17 ± 0.14	0.31 ± 0.11

Table S2.1: Comparison between forests with default (D) and proposed parameter settings (P) across different numbers of subjects N for left-truncated right-censored survival data. Datasets are generated with time-invariant covariates, under the nonlinear survival relationship. Given a method A, each cell value are given as mean $\pm$ one standard deviation of $\big(L_2(\text{Cox}) - L_2(\text{A})\big)/L_2(\text{KM})$ based on all simulations. The mean value is the average % decrease in integrated L_2 difference compared to the Cox model.

Proportional hazards setting	LTRC CIF(D)	LTRC CIF(P)	LTRC RRF(D)	LTRC RRF(P)	LTRC TSF(D)	LTRC TSF(P)
$N = 100$	-0.10 ± 0.22	0.16 ± 0.22	-0.22 ± 0.25	0.02 ± 0.25	-0.27 ± 0.26	-0.09 ± 0.29
$N = 300$	0.07 ± 0.12	0.37 ± 0.13	0.02 ± 0.12	0.37 ± 0.14	-0.00 ± 0.13	0.25 ± 0.18
$N = 500$	0.19 ± 0.09	0.49 ± 0.09	0.16 ± 0.09	0.48 ± 0.10	0.17 ± 0.09	0.42 ± 0.15
$N = 1000$	0.34 ± 0.06	0.59 ± 0.06	0.33 ± 0.06	0.59 ± 0.06	0.37 ± 0.09	0.58 ± 0.10
Non-proportional hazards setting	LTRC CIF(D)	LTRC CIF(P)	LTRC RRF(D)	LTRC RRF(P)	LTRC TSF(D)	LTRC TSF(P)
$N = 100$	0.29 ± 0.10	0.31 ± 0.10	0.30 ± 0.10	0.31 ± 0.10	0.33 ± 0.09	0.33 ± 0.09
$N = 300$	0.13 ± 0.07	0.15 ± 0.06	0.13 ± 0.07	0.15 ± 0.06	0.23 ± 0.06	0.21 ± 0.07
$N = 500$	0.10 ± 0.07	0.12 ± 0.06	0.10 ± 0.07	0.12 ± 0.06	0.22 ± 0.09	0.22 ± 0.09
$N = 1000$	0.10 ± 0.05	0.10 ± 0.04	0.09 ± 0.05	0.10 ± 0.04	0.27 ± 0.05	0.30 ± 0.07