SUPPLEMENTAL MATERIALS

Appendix A: Implementation

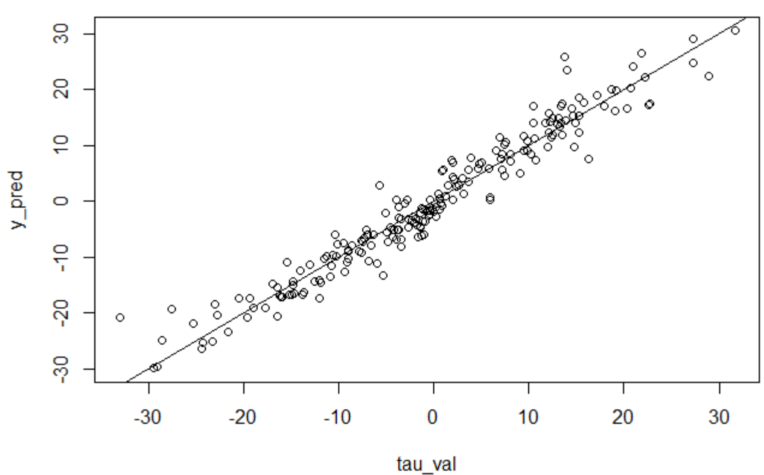
Title: RCATE package

R-package for robust estimation of CATE: R package **RCATE** containing code for 9 robust estimation algorithms of CATE described in the article and also the methods based on additive B-spline LAD regression in R.Li. The package also contains the dataset used as example in the article.

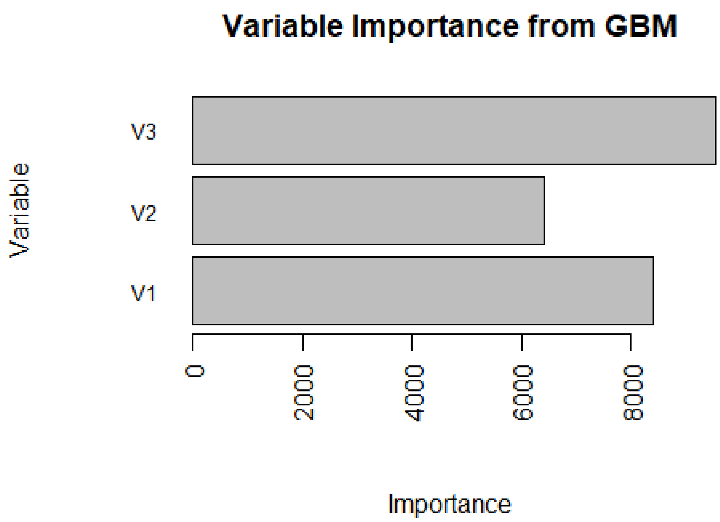
Hypertension dataset: Data set used in the illustration of robust estimation of CATE algorithms in Section 4.

Example of usage:

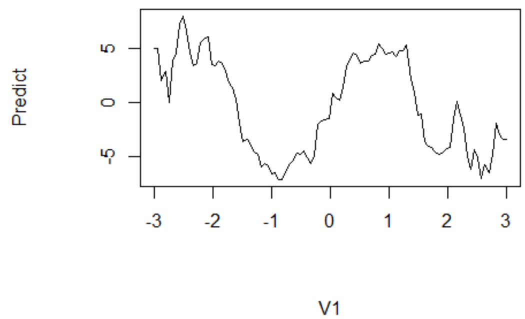
```
## Install package
require(devtools)
devtools::install_github("rhli-Hannah/RCATE")
library(RCATE)

## Data generation
n <- 1000; p <- 3; set.seed(2223)
X <- as.data.frame(matrix(runif(n*p,-3,3),nrow=n,ncol=p))
tau = 6*sin(2*X[,1])+3*(X[,2]+3)*X[,3]
p = 1/(1+exp(-X[,1]+X[,2]))
d = rbinom(n,1,p)
t = 2*d-1
y = 100+4*X[,1]+X[,2]-3*X[,3]+tau*t/2 + rnorm(n,0,1)
set.seed(2223)
x_val = as.data.frame(matrix(rnorm(200*3,0,1),nrow=200,ncol=3))
tau_val = 6*sin(2*x_val[,1])+3*(x_val[,2]+3)*x_val[,3]

## Use robust GBM + R-learning to estimate CATE
fit <- rcate.ml(X,y,d,method='RL',algorithm='GBM')
y_pred <- predict(fit,x_val)$predict
plot(tau_val,y_pred);abline(0,1)
```

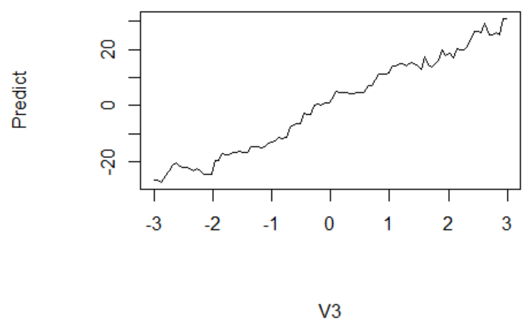


```
## Variable importance level
importance <- importance.rcate(fit)
```




```
## Marginal treatment effect plot
marginal.rcate(fit, 'V1')
marginal.rcate(fit, 'V3')
```





Appendix B: Supplemental Simulation Study Results

Supplemental Simulation Study (S): We compared the algorithm-based robust estimators against the model-based ones when the true treatment effect models were correctly specified. Here we assumed that the true effect effect τ was an additive function of $\mathbf{X}$. In such a situation, robust methods based on generalized additive models (GAM) should provide correct estimates. We also included in the simulation an L_1-based Q-learner (robust QL) for comparison.

Specifically, we defined the last two methods as follows:

$$\text{Robust GAM: } \hat{\beta} = argmin_{\beta} \frac{1}{n} \sum_{i=1}^{n} w_i^*(X_i, T_i) |Y_i^* - B(X_i)^T \beta| + \Lambda_n(\beta),$$

$$\text{Robust QL: } \hat{\gamma}, \hat{\beta} = argmin_{\gamma,\beta} \frac{1}{n} \sum_{i=1}^{n} |Y_i - B(X_i)^T \gamma - \frac{T_i}{2} B(X_i)^T \beta| + \Lambda_n(\gamma, \beta),$$

where Λ is a smoothness-sparsity penalty for group-wise variable selection and for smoothness of the regression line.

Model-based estimators can be more efficient when they depict the treatment effect with the right function. We simulated a situation where the true treatment effect τ is an additive function of $\mathbf{X}$. Since the model-based estimators used GAM to depict $\tau(\mathbf{X})$, we expect them to perform well. Algorithm-based estimators, on the other hand, may have reduced efficiency while offering a greater protection against model misspecification. Here we used model-based methods as a benchmark, and compared the performance of the algorithm-based estimators as sample size increased.

We compared all methods indicated by "(S)" in Table 4. We considered two scenarios: (1) For the robust GAMs, we fixed the sample size at $n_0 = 200$, and for robust GBMs, robust RFs, and robust ANNs, we increased the sample size from 200 to 1000 by an increment of 200; (2) For the robust GAMs, we fixed the sample size at $n_0 = 1000$; for the proposed robust algorithms, we increased the sample size from 1000 to 7000 by an increment of 2000. Specifically, we used two different error distributions $P = N(0, 100)$ and $P = Laplace(0, \sqrt{50})$, while fixing the proportion of outliers at $p_o = 0.15$. The covariates were continuous variables ($\mathbf{X}_i \sim N_{10}(0, 1)$).

Functions $b_0(\mathbf{X}_i)$ and $\tau_0(\mathbf{X}_i)$ in the response surface were

$$b_0(\mathbf{X}_i) = 100 + 4X_{i1} + X_{i2} - 3X_{i3},$$

$$\tau_0(\mathbf{X}_i) = 6sin(2X_{i1}) + 3X_{i2} + X_{i3} + 9tanh(0.5X_{i4}) + 3X_{i5},$$

where the true treatment effect function was an additive model of covariates. We reported the MSE of the CATE estimates graphically in Figure B.1.

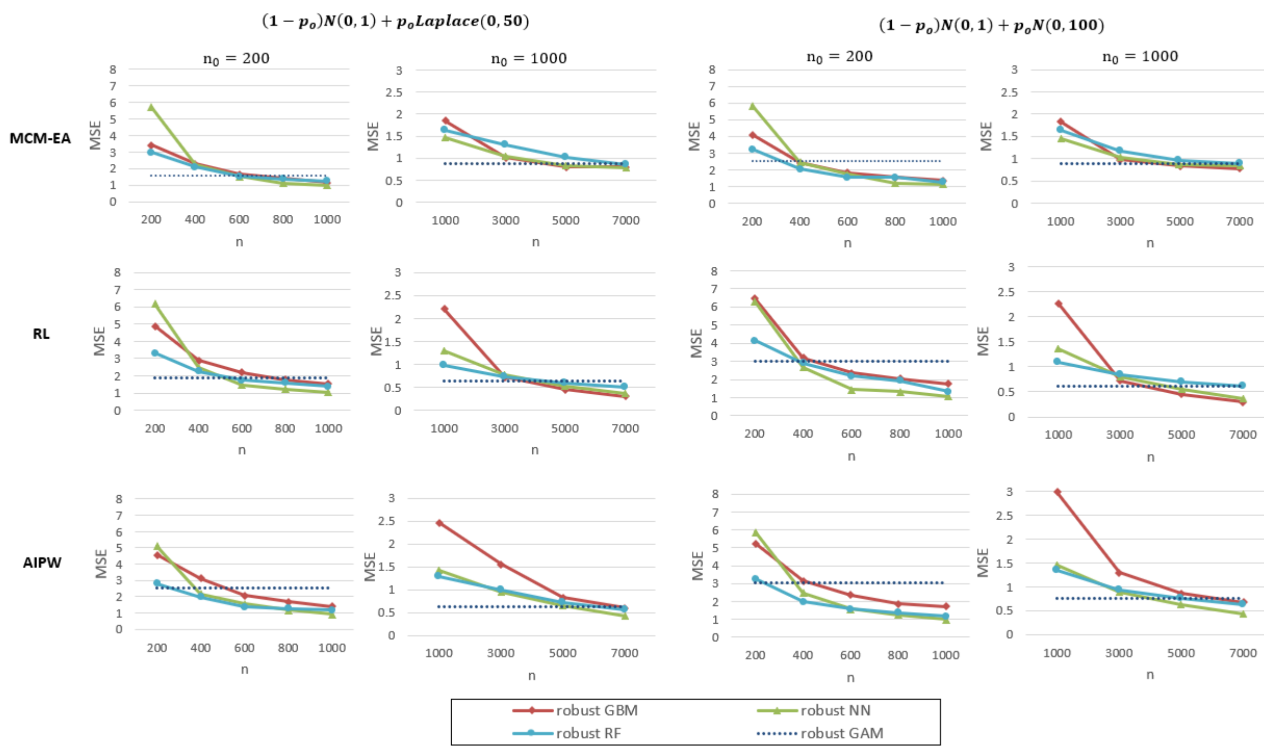


Figure B.1: Simulation results of Simulation S - MSE of different methods under different sample sizes. The robust GBMs were indicated by red solid line, the robust RFs were indicated by blue solid line, the robust ANNs were indicated by green solid line. The robust GAMs were indicated by blue dotted line. In the first and third columns of figures, the sample size of robust GAMs methods was $n_0 = 200$; in the second the fourth columns of figures, the sample size of robust GAMs methods was $n_0 = 1000$.

Figure B.1 showed that machine-learning algorithms' performance improved with the sample size.

Table B.1: Simulation Results of Simulation S ($n_0 = 200$)

		MCM-EA									
		$Laplace(0, \sqrt{50})$					$N(0, 100)$				
	n	200	400	600	800	1000	200	400	600	800	1000
MSE	robust GBM	3.43	2.28	1.66	1.39	1.14	4.08	2.43	1.80	1.55	1.37
	GBM	19.96	13.87	10.95	8.99	7.69	21.64	14.18	11.33	9.36	7.33
	robust NN	5.70	2.25	1.54	1.12	1.00	5.84	2.45	1.74	1.21	1.14
	NN	6.50	3.95	3.01	2.60	2.17	6.79	4.09	3.01	2.69	2.23
	robust RF	2.96	2.10	1.53	1.37	1.24	3.21	2.05	1.55	1.54	1.25
	RF	11.12	9.94	8.77	7.94	7.75	10.05	9.37	8.33	8.38	7.23
	robust GAM	1.54					2.54				
	n	200	400	600	800	1000	200	400	600	800	1000
MAE	robust GBM	1.44	1.32	0.99	0.90	0.80	1.57	1.21	1.03	0.94	0.87
	GBM	3.40	3.56	2.49	2.23	2.05	3.62	2.91	2.58	2.34	2.06
	robust NN	1.86	1.14	0.93	0.80	0.75	1.89	1.19	0.98	0.83	0.79
	NN	1.99	1.53	1.33	1.23	1.14	2.04	1.57	1.34	1.26	1.14
	robust RF	1.27	0.99	0.90	0.84	0.79	1.32	1.03	0.89	0.89	0.80
	RF	2.02	3.56	1.80	1.73	1.70	2.08	4.08	1.89	1.83	1.75
	robust GAM	0.83					1.13				

		RL									
		$Laplace(0, \sqrt{50})$					$N(0, 100)$				
	n	200	400	600	800	1000	200	400	600	800	1000
MSE	robust GBM	4.87	2.88	2.19	1.77	1.52	6.46	3.19	2.36	2.05	1.77
	GBM	36.93	22.98	17.28	14.28	11.88	38.67	24.42	17.34	16.77	11.55
	robust NN	6.19	2.48	1.48	1.23	1.04	6.28	2.67	1.44	1.33	1.07
	NN	6.78	4.28	3.45	2.87	2.60	7.30	4.39	3.51	3.10	2.62
	robust RF	3.32	2.24	1.76	1.59	1.39	4.14	2.92	2.20	1.92	1.34
	RF	49.57	49.48	9.36	43.16	47.60	133.48	92.12	40.54	68.37	50.74
	robust GAM	1.87					3.03				
	n	200	400	600	800	1000	200	400	600	800	1000
MAE	robust GBM	1.71	1.32	1.07	1.01	0.93	1.95	1.38	1.18	1.08	1.00
	GBM	4.56	3.56	2.50	2.73	2.48	4.78	3.71	3.16	2.97	2.53
	robust NN	1.94	1.19	0.91	0.83	0.77	1.96	1.23	0.92	0.87	0.78
	NN	2.03	1.60	1.43	1.31	1.24	2.12	1.63	1.44	1.36	1.26
	robust RF	1.27	0.99	0.86	0.81	0.72	1.36	1.03	0.87	0.85	0.77
	RF	3.61	3.56	1.81	3.57	3.67	4.61	4.08	3.64	4.35	3.89
	robust GAM	0.79					1.10				

		AIPW									
		$Laplace(0, \sqrt{50})$					$N(0, 100)$				
	n	200	400	600	800	1000	200	400	600	800	1000
MSE	robust GBM	4.55	3.13	2.09	1.70	1.41	5.23	3.15	2.36	1.87	1.73
	GBM	19.10	14.01	11.75	8.61	7.32	20.36	13.40	12.07	9.90	7.52
	robust NN	5.12	2.17	1.57	1.18	0.94	5.88	2.48	1.58	1.24	0.99
	NN	1.97	3.77	2.91	2.46	2.23	6.86	4.24	3.14	2.57	2.34
	robust RF	2.80	1.98	1.36	1.29	1.18	3.25	1.98	1.56	1.37	1.18
	RF	10.57	9.66	9.36	8.58	8.33	9.98	8.81	9.21	9.86	8.66
	robust GAM	2.52					3.04				
	n	200	400	600	800	1000	200	400	600	800	1000
MAE	robust GBM	1.65	1.33	1.07	0.94	0.84	1.77	1.33	1.13	0.99	0.92
	GBM	3.35	2.78	2.50	2.13	1.98	3.50	2.81	2.59	2.31	2.05
	robust NN	1.76	1.11	0.93	0.81	0.73	1.89	1.19	0.96	0.84	0.76
	NN	1.97	1.50	1.31	1.20	1.15	2.05	1.59	1.37	1.23	1.17
	robust RF	1.21	0.99	0.86	0.82	0.78	1.32	1.01	0.89	0.84	0.77
	RF	1.98	1.89	1.81	1.73	1.68	2.09	1.94	1.92	1.87	1.76
	robust GAM	0.99					1.11				

Table B.2: Simulation Results of Simulation S ($n_0 = 1000$)

		MCM-EA							
		$Laplace(0, \sqrt{50})$				$N(0, 100)$			
	n	1000	3000	5000	7000	1000	3000	5000	7000
MSE	robust GBM	1.85	1.01	0.80	0.80	1.83	0.97	0.83	0.77
	GBM	11.22	2.71	1.99	1.55	11.10	2.83	2.03	1.59
	robust NN	1.48	1.04	0.84	0.78	1.46	1.03	0.86	0.84
	NN	2.76	2.61	2.18	1.88	2.74	2.76	2.20	1.95
	robust RF	1.63	1.31	1.01	0.84	1.62	1.17	0.96	0.88
	RF	5.12	5.00	4.28	4.13	5.26	4.85	4.11	3.81
	robust GAM	0.87				0.88			
	n	1000	3000	5000	7000	1000	3000	5000	7000
MAE	robust GBM	1.06	0.75	0.67	0.61	1.06	0.74	0.68	0.68
	GBM	2.56	1.23	1.05	0.93	2.58	1.28	1.07	0.94
	robust NN	0.94	0.77	0.69	0.62	0.93	0.77	0.70	0.72
	NN	1.28	1.23	1.13	1.04	1.28	1.27	1.13	1.07
	robust RF	0.94	0.84	0.72	0.67	0.92	0.79	0.71	0.68
	RF	1.51	1.48	1.40	1.36	1.58	1.53	1.44	1.37
	robust GAM	0.60				0.63			

		RL							
		$Laplace(0, \sqrt{50})$				$N(0, 100)$			
	n	1000	3000	5000	7000	1000	3000	5000	7000
MSE	robust GBM	2.21	0.72	0.45	0.30	2.27	0.72	0.45	0.30
	GBM	29.31	4.55	3.01	2.14	27.49	5.02	3.04	1.80
	robust NN	1.29	0.77	0.53	0.36	1.36	0.80	0.55	0.37
	NN	3.58	3.26	2.58	1.90	3.66	3.51	2.55	1.93
	robust RF	0.98	0.73	0.60	0.50	1.09	0.85	0.70	0.62
	RF	107.18	91.18	91.12	72.19	127.63	91.68	81.65	70.37
	robust GAM	0.64				0.61			
	n	1000	3000	5000	7000	1000	3000	5000	7000
MAE	robust GBM	1.14	0.64	0.50	0.40	1.15	0.63	0.49	0.40
	GBM	3.80	1.51	1.20	0.97	3.75	1.59	1.25	0.96
	robust NN	0.88	0.67	0.55	0.46	0.90	0.68	0.56	0.46
	NN	1.45	1.37	1.22	1.04	1.47	1.42	1.22	1.06
	robust RF	0.71	0.62	0.54	0.46	0.70	0.60	0.51	0.47
	RF	5.44	4.89	4.61	4.04	5.73	4.96	4.65	4.19
	robust GAM	0.56				0.53			

		AIPW							
		$Laplace(0, \sqrt{50})$				$N(0, 100)$			
	n	1000	3000	5000	7000	1000	3000	5000	7000
MSE	robust GBM	2.46	1.55	0.82	0.60	2.99	1.29	1.07	0.98
	GBM	14.43	3.98	2.47	1.54	15.25	3.51	4.17	1.88
	robust NN	1.43	0.94	0.66	0.43	1.46	0.89	0.63	0.43
	NN	3.83	3.27	2.54	2.04	3.61	3.15	2.67	2.06
	robust RF	1.28	1.00	0.73	0.57	1.35	0.93	0.75	0.63
	RF	7.80	7.76	7.51	6.20	7.97	7.39	7.09	6.72
	robust GAM	0.63				0.75			
	n	1000	3000	5000	7000	1000	3000	5000	7000
MAE	robust GBM	1.48	0.77	0.59	0.48	1.15	0.76	0.61	0.48
	GBM	2.82	1.37	1.10	0.85	2.87	1.37	1.11	0.85
	robust NN	0.92	0.73	0.61	0.49	0.93	0.72	0.60	0.49
	NN	1.46	1.35	1.20	1.07	1.46	1.34	1.23	1.07
	robust RF	0.81	0.71	0.59	0.47	0.83	0.69	0.60	0.53
	RF	1.65	1.64	1.56	1.45	1.69	1.69	1.59	1.55
	robust GAM	0.52				0.53			

Table B.3: Simulation Results (Coverage Probabilities) of Simulation 1

	MCM-EA													
	$Laplace(0, \sqrt{50})$							$N(0, 100)$						
p_o	0.00	0.05	0.10	0.15	0.20	0.30	0.50	0.00	0.05	0.10	0.15	0.20	0.30	0.50
Robust GBM	0.95	0.94	0.91	0.93	0.93	0.95	0.97	0.95	0.94	0.95	0.94	0.94	0.98	0.97
GBM	0.95	0.93	0.96	0.96	0.94	0.97	0.96	0.95	0.96	0.97	0.99	0.99	0.91	0.95
Robust NN	0.95	0.94	0.96	0.92	0.94	0.99	0.98	0.95	0.92	0.98	0.92	0.95	0.98	0.93
NN	0.96	0.76	0.87	0.47	0.83	0.79	0.89	0.96	0.39	0.43	0.27	0.32	0.35	0.29
Robust RF	0.94	0.96	0.97	0.94	0.92	0.93	0.98	0.94	0.95	0.94	0.93	0.92	0.96	0.98
RF	0.96	0.96	0.98	0.97	0.97	0.89	0.96	0.96	0.90	0.92	0.92	0.93	0.89	0.87
Robust GAM	0.23	0.38	0.49	0.54	0.57	0.66	0.74	0.23	0.39	0.48	0.54	0.59	0.66	0.76
Robust QL	0.47	0.49	0.49	0.52	0.53	0.53	0.00	0.47	0.49	0.50	0.51	0.53	0.33	0.00
	RL													
	$Laplace(0, \sqrt{50})$							$N(0, 100)$						
p_o	0.00	0.05	0.10	0.15	0.20	0.30	0.50	0.00	0.05	0.10	0.15	0.20	0.30	0.50
Robust GBM	0.95	0.94	0.95	0.96	0.95	0.97	0.97	0.95	0.95	0.96	0.94	0.96	0.91	0.98
GBM	0.93	0.97	0.98	0.98	0.97	0.96	0.95	0.93	0.99	0.98	0.99	1.00	1.00	0.99
Robust NN	0.94	0.95	0.93	0.96	0.92	0.93	0.96	0.94	0.95	0.96	0.93	0.97	0.91	0.98
NN	0.96	0.82	0.94	0.95	0.88	0.85	0.91	0.96	0.94	0.93	0.94	0.85	0.82	0.36
Robust RF	0.95	0.96	0.97	0.94	0.98	0.95	0.94	0.95	0.92	0.91	0.90	0.87	0.82	0.72
RF	0.94	0.89	0.98	0.97	0.87	0.93	0.94	0.94	0.82	0.93	0.89	0.95	0.89	0.86
Robust GAM	0.27	0.48	0.57	0.61	0.65	0.71	0.80	0.27	0.49	0.58	0.64	0.65	0.72	0.81
Robust QL	0.47	0.49	0.49	0.52	0.53	0.53	0.00	0.47	0.49	0.50	0.51	0.53	0.33	0.00
	AIPW													
	$Laplace(0, \sqrt{50})$							$N(0, 100)$						
p_o	0.00	0.05	0.10	0.15	0.20	0.30	0.50	0.00	0.05	0.10	0.15	0.20	0.30	0.50
Robust GBM	0.96	0.93	0.92	0.91	0.90	0.89	0.90	0.96	0.94	0.93	0.92	0.90	0.98	0.93
GBM	0.96	0.92	0.98	0.97	0.95	0.95	0.98	0.96	0.97	0.96	0.99	0.97	0.90	0.91
Robust NN	0.94	0.92	0.95	0.93	0.92	0.96	0.98	0.94	0.98	0.96	0.93	0.95	0.93	0.98
NN	0.95	0.74	0.89	0.96	0.80	0.73	0.88	0.95	0.44	0.88	0.84	0.87	0.83	0.78
Robust RF	0.93	0.94	0.98	0.91	0.97	0.96	0.98	0.93	0.96	0.97	0.94	0.95	0.96	0.95
RF	0.94	0.98	0.97	0.93	0.93	0.82	0.96	0.94	0.94	0.91	0.87	0.95	0.86	0.82
Robust GAM	0.54	0.57	0.59	0.61	0.62	0.64	0.70	0.54	0.58	0.59	0.61	0.64	0.66	0.73
Robust QL	0.47	0.49	0.49	0.52	0.53	0.53	0.00	0.47	0.49	0.50	0.51	0.53	0.33	0.00

Table B.4: Tuning parameters of considered methods in simulation

Method	Parameter	Value
RF-based algorithms	Number of trees	50
	Fraction of feathers used in splitting	0.8
	Minimum node size	3
Boosting-based algorithms	Number of trees	1000
	Depth of trees	2
	Learning rate	0.1
Robust ANN	Number of hidden layers	2
	Number of neurons in hidden layers	p and $p/2$
	Adam optimization	$\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$
	L_1 regularization ($p = 100, 2000$)	0.1, if $p = 100$; 0.02, if $p = 2000$.
	Number of neurons in hidden layers ($p = 2000$)	$p/10$ and $p/40$
Robust GAM and QL	Number of knots	$\sqrt{n}/2$
	Number of degree	3
	γ in SCAD	3.7