SUPPLEMENT TO "DO FORECASTS OF BANKRUPTCY CAUSE BANKRUPTCY? A MACHINE LEARNING SENSITIVITY ANALYSIS."

BY DEMETRIOS PAPAKOSTAS[1,b] P. RICHARD HAHN[1,a], JARED MURRAY[2,c] AND FRANK ZHOU[3,d] JOSEPH GERAKOS[4,e]

[1]*School of Mathematical and Statistical Sciences, Arizona State University,* [a]*prhahn@asu.edu;* [b]*dpapakos@asu.edu*

[2]*Department of Information, Risk and Operations Management, The University of Texas at Austin,* [c]*jared.murray@mccombs.utexas.edu*

[3]*The Wharton School, University of Pennsylvania,* [d]*szho@wharton.upenn.edu*

[4]*Tuck School of Business, Dartmouth College,* [e]*Joseph.J.Gerakos@tuck.dartmouth.edu*

1. Estimating the risk difference as estimand of interest. Rather than looking at the ratio of potential outcomes, it is often the case we want to investigate the difference in the expected value of each, i.e. we can look at *risk differences*:

$$\text{Risk Difference} \rightarrow \Delta \equiv \mathrm{E}(B^1) - E(B^0)$$

In our framework, following similar reasoning as in section *Modular sensitivity analysis with machine learning* in the main file, risk differences can be defined as

$$\Delta(\mathbf{x}_i) = \int_{\mathbb{R}} \Phi\big(b_1(\mathbf{x}) + u\big) f(u)\mathrm{d}u - \int_{\mathbb{R}} \Phi\big(b_0(\mathbf{x}) + u\big) f(u)\mathrm{d}u$$

The sample average risk difference (ARD) is therefore $\frac{1}{n}\sum_{i=1}^{n}\Delta(\mathbf{x}_i)$. In the case of the audit data, the average risk difference refers to percentage point difference in going bankrupt after receiving a going concern opinion. We esimate the risk difference using our methodology as well as the bivariate probit with endogenous regressor model, (described in equation(4) in the main file), to the audit data. Specifically, we used the same covariates as we used when fitting monotone bart, used the bankruptcy indicator as the binary outcome, and whether or not a going concern was issued as the "treatment" indicator. Using our methodology, results for estimating risk differences on the audit data are presented in Table 1.

Analogs of tables 5-8 of the main file but with risk differences as the estimands of interest are presented in Table 2, Table 3, Table 4, and Table 5.

2. Bivariate probit simulation study. Table 6 shows the results when fitting the bivariate probit regression with a maximum likelihood estimate to the simulated bivariate probit data. Unsurprisingly, this performs well, with the caveat that we require large N ($\mathrm{N} = 100,000$) to get these impressive results. We simulated the samples from the bivariate probit model of the main document, where we sum 5 uniform(-1,1) $\mathbf{x}_i$ covariates each with

Distribution of $f(u)$	ARD post (%)	mean B_1 (%)	95% Credible interval for ARD (%)
$\mathrm{N}(0, \sigma = 0.1)$	9.97	11.5	$(7.08, 12.9)$
$\mathrm{N}(0, \sigma = 0.5)$	4.12	5.96	$(2.74, 5.60)$
$\mathrm{N}(0, \sigma = 1)$	0.70	2.90	$(0.40, 1.01)$
Shark $q = 0.25$, $s = 0.5$; $\sigma = 1.05$	0.28	2.60	$(0.14, 0.45)$
Shark $q = 0.75$, $s = 1.25$; $\sigma = 0.89$	2.84	4.87	$(1.81, 3.97)$
Symmetric Mixture ($\sigma = 0.64$)	2.30	4.52	$(2.23, 2.92)$
Asymmetric Mixture ($\sigma = 0.49$)	2.47	4.62	$(2.43, 3.14)$

TABLE 1

The reduced form probabilities (equation (15) in the main file) were estimated using BART with a monotonicity constraint on the going concern variable. We further require $b_1(\mathbf{x}) > b_0(\mathbf{x})$ in the projection step. Posterior summaries based on 500 Monte Carlo samples. σ refers to the implied standard deviations of the different distributions. Values listed as the percentage increase in probability of bankruptcy.

$f(u)$	**true** ARD	Correct est.	Correct RMSE	Wrong $f(u)$	Wrong est.	Wrong RMSE	LBP est.	LBP RMSE	SBP est.	SBP RMSE
$N(0,1)$	0.30	0.31	0.05	Lap(0,1.2^2)	0.17	0.15	0.14	0.16	0.16	0.16
$N(0,1.5^2)$	0.26	0.26	0.05	Lap(0, 1.75^2)	0.16	0.12	0.13	0.13	0.06	0.19
$N(0,2^2)$	0.23	0.23	0.04	Lap(0, 2.5^2)	0.12	0.12	0.23	0.06	0.25	0.07
$N(0,2.5^2)$	0.20	0.20	0.04	Lap(0, 2^2)	0.25	0.	0.07	0.07	0.23	0.06
$N(-1,1)$	0.17	0.18	0.04	Lap(-1, 1.3^2)	0.05	0.14	0.10	0.16	0.11	0.16
$N(1,2^2)$	0.25	0.22	0.06	Lap(1, 2.4^2)	0.11	0.16	0.17	0.06	0.16	0.08
$N(-2,2^2)$	0.11	0.03	0.09	Lap$(-2, 2.3^2)$	0.00	0.12	0.07	0.08	0.07	0.09
$N(2,1)$	0.30	0.29	0.05	Lap(2, 1.3^2)	0.10	0.22	0.28	0.13	0.24	0.13

TABLE 2

Different $f(u)$ as described in equation (28) of the main file. $N = 25000$. Wrong $f(u)$ indicates the distribution of U we used to solve the system of equations in equation (15) , i.e. how we mis-specified. True indicates true ARD, and correct est indicates our estimate of the ARD when correctly *specifying $f(u)$. Lap refers to the Laplacian distribution. Smooth refers to bivariate probit regression with smoothing covariates.*

$f(u)$ sharkfin with parameters q, s	**true** ARD	true est. ARD	true RMSE	wrong q	wrong q ARD est.	wrong q RMSE
(0.25, 0.82; 3)	0.28	0.28	0.04	(0.40,1.37;3.00)	0.22	0.05
(0.40, 1.37; 3)	0.26	0.27	0.05	(0.70,2.34;3.00)	0.29	0.05
(0.60, 1.06; 3)	0.23	0.22	0.04	(0.30,1.00;3.00)	0.15	0.07
(0.75, 2.46; 3)	0.18	0.17	0.04	(0.92,2.77;3.00)	0.21	0.08
(0.25, 0.34; 0.5)	0.38	0.39	0.05	(0.10,0.12;0.50)	0.39	0.05
(0.40, 0.56; 0.5)	0.35	0.36	0.06	(0.20,0.26;0.50)	0.35	0.06
(0.60, 0.84; 0.5)	0.29	0.30	0.05	(0.80,1.05;0.50)	0.31	0.06
(0.75, 1.00; 0.5)	0.24	0.25	0.05	(0.45,1.63;0.50)	0.21	0.06

TABLE 3

Different $f(u)$ as described in equation (28), all of the "sharkfin" family. $N = 25000$. Wrong q indicates that we purposely mis-specified q when solving our system of equations, whereas the true est. AR and true RMSE columns indicate where we correctly specified $f(u)$, both the q and s parameters, when solving our system. ; indicates the variance, whereas the first 2 entries in shark are the q and s parameters. Here we vary the skewness while keeping variance constant.

the same β_1 and α_1 coefficients respectively. We set $\beta_0 = 0, \beta_1 = -0.2, \alpha_0 = -0.5, \alpha_1 = 0.7$ to generate reasonable number of going concerns and bankruptcies.

$f(u)$ sharkfin with parameters q, s	**true** ARD	true est. ARD	true RMSE	wrong σ^2	wrong σ^2 ARD est.	wrong σ^2 RMSE
(0.25, 0.82; 3)	0.28	0.28	0.06	(0.25,0.47;1.0)	0.50	0.23
(0.40, 1.37; 3)	0.26	0.27	0.06	(0.40,1.12;2.0)	0.35	0.10
(0.60, 1.06; 3)	0.23	0.22	0.06	(0.60,0.92;0.6)	0.48	0.27
(0.75, 2.46; 3)	0.18	0.17	0.05	(0.75,1.74;1.5)	0.26	0.09
(0.25, 0.34; 0.5)	0.38	0.39	0.07	(0.25,0.67;2.0)	0.15	0.24
(0.40, 0.56; 0.5)	0.35	0.36	0.07	(0.40,1.12;2.0)	0.13	0.23
(0.60, 0.84; 0.5)	0.29	0.30	0.07	(0.60,2.38;4.0)	0.04	0.28
(0.75, 1.00; 0.5)	0.24	0.25	0.06	(0.75,3.18;5.0)	0.04	0.24

TABLE 4

Different $f(u)$ as described in equation (28) of the main document, all of the "sharkfin" family. $N = 25000$. Wrong σ^2 indicates that we purposely mis-specified our variance (by varying the s parameter) when solving our system of equations, whereas the true est. ARD and true RMSE columns indicate where we correctly specified $f(u)$, both the q and s parameters, when solving our system. ; indicates the variance, whereas the first 2 entries in shark are the q and s parameters. Here we vary the variance keeping skewness constant.

True $f(u)$	true ATT	ATT est.	ATC true	ATC est.	Wrong q $f(u)$	ATT est. wrong	ATC est. wrong
shark(0.1, 0.30; 3)	0.28	0.29	0.29	0.29	shark(0.9, 2.74;3)	0.14	0.20
shark(0.1, 0.12; 0.5)	0.40	0.40	0.38	0.39	shark(0.9, 1.12; 0.5)	0.39	0.40
shark(0.1, 0.18; 1)	0.37	0.37	0.37	0.36	shark(0.9, 1.58; 1)	0.32	0.37
shark(0.1, 0.18; 1)	0.37	0.37	0.37	0.36	shark(0.5, 1; 1)	0.34	0.36
shark(0.5, 1; 1)	0.32	0.31	0.28	0.29	shark(0.1, 0.18; 1)	0.29	0.24
shark(0.5, 1; 1)	0.32	0.31	0.28	0.29	shark(0.9, 1.58; 1)	0.35	0.34
shark(0.9, 1.58; 1)	0.21	0.20	0.16	0.16	shark(0.1, 0.18; 1)	0.11	0.06
shark(0.9, 1.58; 1)	0.21	0.20	0.16	0.16	shark(0.5, 1; 1)	0.15	0.10

TABLE 5

Comparing estimates of average causal risk difference on treated (ACRDT) and average causal risk difference on controls (ACRDC) when we more aggressively misspecify the q parameter, which controls the skewness.

ACRD true	ACRD est	ICRD cor	ICRD RMSE	ACRR true	ACRR est	ICRR cor	ICRR rmse	γ true	γ est.	ρ	ρ est.
0.23	0.24	0.97	0.02	2.24	2.07	1.00	0.24	1.00	0.77	0.25	0.37
0.46	0.46	0.99	0.02	4.32	3.89	1.00	0.87	1.75	1.62	0.25	0.31
0.58	0.57	1.00	0.02	6.24	5.40	0.99	2.34	2.50	2.38	0.25	0.30
0.26	0.26	0.99	0.01	2.42	2.27	1.00	0.22	1.00	0.85	0.40	0.47
0.46	0.46	0.99	0.02	4.33	3.89	1.00	0.90	1.75	1.63	0.40	0.45
0.57	0.56	0.99	0.02	6.14	5.13	1.00	2.83	2.50	2.34	0.40	0.46
0.28	0.28	0.99	0.01	2.57	2.46	1.00	0.17	1.00	0.92	0.60	0.63
0.47	0.47	1.00	0.01	4.51	4.24	1.00	0.63	1.75	1.70	0.60	0.61
0.59	0.58	1.00	0.01	6.41	5.89	0.99	1.67	2.50	2.45	0.60	0.61
0.31	0.31	1.00	0.00	2.79	2.80	1.00	0.02	1.00	1.02	0.80	0.80
0.47	0.47	1.00	0.01	4.51	4.26	1.00	0.56	1.75	1.70	0.80	0.81
0.58	0.58	1.00	0.01	6.31	5.56	0.99	2.25	2.50	2.41	0.80	0.82

TABLE 6

N=100,000. Fit the simulated bivariate probit with the bivariate probit regression. Validates the MLE of the bivariate probit regression performs well, as well as the validity of our data generation process, however required a large N to get accurate results. ACRD refers to average risk difference, ICRD individual risk difference. ACRR refers to average causal risk ratio, whereas ICRR refers to individual causal risk ratio.

Table 7 shows the results when we simulated from the bivariate probit and fit with our methodology, with $f(u)$ assigned appropriately, only this time we are interested in the treatment effect. Our method does well here, with N $= 25,000$ and $p = 5$.

γ	ρ	ACRD true	ACRD est	ICRD cor	ICRD RMSE
1.00	0.25	0.29	0.29	0.89	0.05
1.75	0.25	0.47	0.47	0.96	0.04
2.50	0.25	0.58	0.57	0.97	0.05
1.00	0.40	0.29	0.29	0.90	0.05
1.75	0.40	0.47	0.46	0.94	0.06
2.50	0.40	0.58	0.55	0.96	0.09
1.00	0.60	0.29	0.29	0.90	0.05
1.75	0.60	0.47	0.46	0.95	0.05
2.50	0.60	0.58	0.57	0.98	0.04
1.00	0.80	0.29	0.27	0.91	0.05
1.75	0.80	0.47	0.44	0.95	0.05
2.50	0.80	0.58	0.55	0.98	0.05

TABLE 7

We simulated from the bivariate probit with 25,000 observations and deploy our methodology. cor refers to the correlation between predicted and true for the average causal risk difference (ACRD), and the rmse is the root mean square error. Fit using the 'GJRM' package of Marra and Radice [2011].

It was stressed in the main document how using a BART model with a monotonicity constraint improves our estimation of the ICRR, but the improvement is even more pronounced when studying risk differences. In figure Figure 1, we look at the comparison of IRD (individual risk difference) estimates from data generated by the bivariate probit, with the left hand side of our system of equation probabilites estimated with BART and monotone BART. We display in the main file, but with inducements as the estimand of interest.

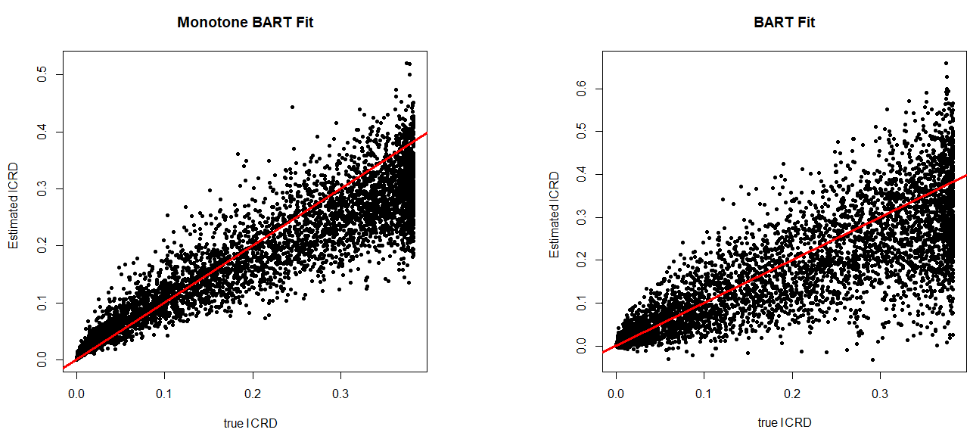
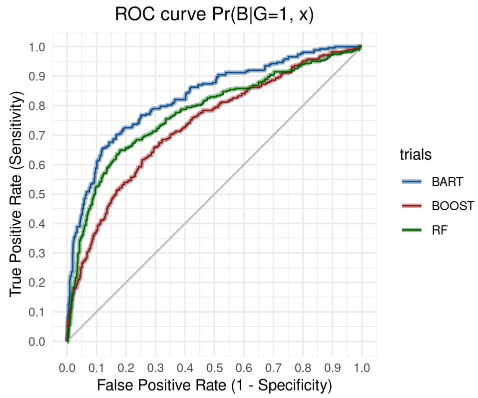


Fig 1: Plots of expected individual causal risk differences (ICRD) vs our estimates, i.e. a plot comparing the difference in potential outcomes from the bivariate probit model $(\Phi(\alpha_0 + \alpha_1 \mathbf{x}_i + \gamma) - \Phi(\alpha_0 + \alpha_1 \mathbf{x}_i))$ versus our estimate. In the DGP, $\rho = 0.25, \gamma = 1$. The monotone BART correlation between τ and $\hat{\tau}$ is 0.928 and for BART is 0.826

3. Comparing machine learning methods for the observational data. Here we present our results from fitting the left hand side of equation(15) in the main file. In this section, we compare the performance in predicting the left hand side of equation(15) using various non-parametric "machine learning" tools. In particular, we compare using monotone



	XGBoost	RanFor	BART
Case	AUC	AUC	AUC
B_1, G_0	0.80	0.83	0.87
B_1, G_1	0.73	0.78	0.83
G_1	0.88	0.90	0.90

Fig 2: Left: Area under curve of ROC plot, balanced 5-fold CV. Plot of ROC performance for predicting $\Pr(B \mid G = 1, \mathbf{x})$. Corresponds to shaded row in table on right. Right: case 1 is predicting bankruptcy when no going concern is issued, case 2 is when no concern issued, and case 3 predicts if concern is issued. All the methods are *similar*, with monotone BART seeming to be the top performer.

BART [1], random forests [Breiman, 2001], and xgboost [Chen and Guestrin, 2016]. Referencing Figure 2 seems to indicate our methodology outperforms competitors in a cross-validation assessment[2].

REFERENCES

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. 5

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *Arxiv Preprint*, 2016. 5

H.A. Chipman, E.I. George, and R.E. McCulloch. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, 2010. 5

G. Marra and R Radice. Estimation of a semiparametric recursive bivariate probit model in the presence of endogeneity. *Canadian Journal of Statistics*, 2011. 4

[1] We use a BART [Chipman et al., 2010] model for the $\Pr(G \mid \mathbf{x})$ scenario, consistent with the main text.

[2] In our main text, we do not do a cross validation to obtain our probabilities, but rather get the probabilities from deploying the monotone BART models on the entire dataset. In this case, our B_1, G_0 auc was 0.88, B_1, G_1 was 0.83, and G_1 was 0.92.