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

BY DEMETRIOS PAPAKOSTAS[1,†] P. RICHARD HAHN[1,*], JARED MURRAY[2,‡] AND FRANK ZHOU[3,§] JOSEPH GERAKOS[4,¶]

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

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

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

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

It is widely speculated that auditors' public forecasts of bankruptcy are, at least in part, self-fulfilling prophecies in the sense that they might actually cause bankruptcies that would not have otherwise occurred. This conjecture is hard to prove, however, because the strong association between bankruptcies and bankruptcy forecasts could simply indicate that auditors are skillful forecasters with unique access to highly predictive covariates. In this paper, we investigate the causal effect of bankruptcy forecasts on bankruptcy using nonparametric sensitivity analysis. We contrast our analysis with two alternative approaches: a linear bivariate probit model with an endogenous regressor, and a recently developed bound on risk ratios called E-values. Additionally, our machine learning approach incorporates a monotonicity constraint corresponding to the assumption that bankruptcy forecasts do not make bankruptcies less likely. Finally, a tree-based posterior summary of the treatment effect estimates allows us to explore which observable firm characteristics moderate the inducement effect.

1. Introduction. A "going concern opinion" is an assessment by an auditor that a firm is at risk of going out of business in the coming year. Here, a "concern" refers to a firm, and "going" refers to staying, as opposed to going out of, business. According to U.S. securities regulations, a public company that receives an adverse going concern opinion must disclose it in the firm's annual filings with the Securities and Exchange Commission. Once issued and disclosed, a going concern opinion may directly contribute to a firm's bankruptcy risk, for example, by inducing lenders to pull lines of credit or increase borrowing costs.[1] As reported in **?**:

Keywords and phrases: BART, Causal Inference, heterogeneous treatment effects, self-fulfilling prophecy, sensitivity analysis.

[1]See **?** for a recent discussion of going concern opinions in the news. See **?** for a discussion of how adverse going concern opinions can adversely affect borrowing costs.

> Companies that receive a going-concern audit opinion may be subjected to more rigorous covenant terms or downgrades in their credit ratings, said Anna Pinedo, a partner at law firm Mayer Brown. Fractured relationships with customers could also strengthen a business's competitors, she said.

Estimating the magnitude of such an "inducement effect" is complicated by the unavailability of the auditors' private information to the analyst. That is, in addition to publicly available firm information, auditors have access to "private information" gleaned from confidential documents and via firsthand knowledge of undocumented attributes such as the firm's corporate culture. This paper considers the question: do going-concern opinions help to predict bankruptcy because they incorporate the auditor's private information or because of an inducement effect? This is a textbook example of causal inference where the potential unobserved confounders are particularly pictureseque: what do auditors know that we (the analysts) do not? We introduce methodology to quantify the impact of private information on the probability that a firm files for bankruptcy in the fiscal year following the issuance of a going concern opinion. We conduct a sensitivity analysis rooted in nonlinear, semiparametric regression techniques and a generalization of the bivariate probit model with an endogenous regressor. We conclude that there is evidence for inducement under plausible assumptions on the distribution of the auditors' private information.

1.1. *Methodological background.* Denote the treatment variable by G_i for "going concern" so that $G_i = 1$ for the ith firm in our sample if that firm disclosed a going concern opinion in the prior year.

Denote the outcome variable B_i for "bankrupt" so that $B_i = 1$ filed for bankruptcy, In terms of potential outcomes [**?**], we are interested in two scenarios: B_i^1 and B_i^0, which are respectively the outcome of a firm i if it had received the treatment and if it had not received the treatment, respectively; only one of these potential outcomes is observed.

The primary estimand of interest will be the causal risk ratio (CRR):

$$\tau \equiv \mathrm{E}(B^1)/\mathrm{E}(B^0) \tag{1}$$

which we will often refer to as simply the "inducement effect." Alternatively, the inducement effect in terms of the "do"-operator of **?** as

$$\tau \equiv \mathrm{E}(B = 1 \mid \mathrm{do}(G = 1))/\mathrm{E}(B = 1 \mid \mathrm{do}(G = 0)) \tag{2}$$

where $\mathrm{do}(G = g)$ refers to an exogeneous intervention, in contradistinction to probabilistic conditioning. We will also consider the risk difference

$$\Delta \equiv \mathrm{E}(B^1) - \mathrm{E}(B^0) \tag{3}$$

and consider how these two estimands differ as a function of observable firm characteristics.

The fundamental problem of causal inference [**?**] is that (B^1, B^0) are never observed simultaneously, rather only one or the other is observed. Consequently, the conditions under which the CRR can be estimated must be carefully assessed and their plausibility debated. There are three widely used methods for estimating average treatment effects: randomization, regression adjustment (broadly construed to include matching and propensity score based methods), and instrumental variables analysis. To briefly review:

- In a randomized controlled trial the treatment variable — G in the present context — is independent of the potential outcomes B^1 and B^0; in this case $\mathrm{E}(B^1) = \mathrm{E}(B \mid G = 1)$, the right hand side of which is readily estimable from observed data (and likewise for the $G = 0$ case).
- When a randomized experiment is not possible (such as in the present example) one instead may hope to find a set of control variables $\mathbf{x}$ for which $\mathrm{E}(B^1 \mid \mathbf{x}) = \mathrm{E}(B \mid G = 1, \mathbf{x})$ and $\mathrm{E}(B^0 \mid \mathbf{x}) = \mathrm{E}(B \mid G = 0, \mathbf{x})$, in which case treatment effects can be estimated by estimating these conditional expectations via regression modeling. This condition in called *conditional ignorability* or, alternatively, $\mathbf{x}$ are said to satisfy the *back-door criterion*.
- A third possibility is that a sufficient set of controls is unavailable, but an *instrument* for the treatment assignment is available. An instrument is a variable that is causally related to the treatment but not otherwise associated with the response variable. In the current context, an instrument variable (IV) would be a one that affects the probability that an auditor issues a going concern opinion without directly affecting bankruptcy probabilities or sharing common causes with bankruptcies. Here we do not elaborate on the details of instrumental variable regression, but see **?** for a recent survey and **?** for a discussion of the use of IV specifically in accounting research.

In the present context, none of these three approaches are available. A sufficient set of controls is certainly not readily available and the existence of a valid instrument is doubtful because firms choose their own auditor, rendering auditor attributes endogenous. Although there are other approaches — such as regression continuity design [**??**], difference-in-differences [**?**], and the synthetic control method [**??**] — they apply in idiosyncratic settings that do not apply to the bankruptcy inducement problem.

With none of the usual tools available to us, it may be possible to make additional modeling assumptions that yield identification of the treatment effect. One such model for bivariate binary observations is the bivariate probit model with an endogeneous regressor [**?**, Section 15.7.3]. Such model-based identification is generally undesirable because the identifying form of the likelihood typically lacks plausible justification [**?**]. Accordingly, it is prudent to consider a range of different assumptions (model specifications) and observe how the estimated treatment effects vary as a result. In this paper, we propose a method for modeling the strength of unobserved confounding in a machine learning framework which permits convenient sensitivity analysis without constraining the observed data distribution unrealistically.

1.2. *Methodological contribution of this paper.* This paper brings together three lines of methodological research. First, we develop a generalization of the bivariate probit with endogeneous regressor and use this unidentified model to conduct a sensitivity analysis. Second, we use modern Bayesian tree-based classification models to estimate the identified parameters in our model and describe a numerical procedure to map these parameters back to the causal estimands of interest. This approach represents both a novel use of Bayesian machine learning as well as a novel application of machine learning to the applied problem of whether going concern opinions induce bankruptcy. Additionally, this model incorporates the assumption that going concern opinions cannot make bankruptcies less likely, a plausible assumption that potentially improves estimation accuracy. Finally, we apply a tree-based posterior summarization strategy to our estimates of the individual treatment effects to identify interesting subgroups for further scrutiny, a method first described in **?**, building on a framework laid out in **?** for linear models.

1.3. *Paper structure.* Because this work touches on many disparate areas, an overview organizing the contents may be helpful.

- First, we review the traditional parametric model used for the binary-treatment-binary-response setting with unmeasured confounding, which is the bivariate probit model with endogenous regressor. We provide a novel justification of this model in terms of Pearl's causal calculus using a latent factor representation of the bivariate probit likelihood.
- Next, we generalize this model by relaxing the linearity and distributional assumptions, making it robust to misspecification.
- The generalized bivariate probit model is not point identified, making a sensitivity analysis necessary. A computationally efficient method for conducting the sensitivity analysis is developed, which uses a single Bayesian model fit of the reduced form parameters.
- We then introduce monotone Bayesian additive regression trees, which is a custom modification of the popular BART model [**?**], and describe the Markov chain updates for enforcing monotonicity in the treatment variable.
- Putting these pieces together, the new machine learning sensitivity analysis is applied to over 25,000 data points from publicly traded U.S. firms. Results are compared to a model-free sensitivity analysis approach called E-values [**?**], which generalize the well known Cornfield bounds [**?**]. Decision trees are used as a posterior summarization tool to discover variables that moderate the inducement effect.
- Additionally, the new approach is investigated via several simulation studies to evaluate its behavior relative to alternative approaches when the data generating process is known.

2. The bivariate probit model with endogenous predictor. A well-known model that has been used for problems similar to the one described here is the bivariate probit with endogenous predictor [**?**, Section 15.7.3]. This model can be expressed in terms of bivariate Gaussian latent utilities Z_g and Z_b that relate to going concern opinions and bankruptcy:

$$\begin{pmatrix} Z_{g,i} \\ Z_{b,i} \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma) \qquad \mu = \begin{pmatrix} \beta_0 + \beta_1 \mathbf{x}_i \\ \alpha_0 + \alpha_1 \mathbf{x}_i \end{pmatrix} \qquad \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{4}$$

The premise of this model is that ρ reflects the influence of private information available to the auditor but not the researcher, and $\mathbf{x}_i$ represents covariates of a company that is available to both the auditor and to the researcher. The observed binary indicators, G and B, relate to these latent utilities via

$$G = \mathbb{1}\left\{Z_{g,i} \geq 0\right\} \tag{5}$$

$$B = \mathbb{1}\left\{Z_{b,i} \geq -\gamma G\right\} \tag{6}$$

The coefficient γ governs the strength of the inducement effect.

The basic identification strategy can be motivated geometrically. Let

$$\mathbf{\Pi} = \begin{pmatrix} \pi_{01} & \pi_{11} \\ \pi_{00} & \pi_{10} \end{pmatrix}$$

where $\pi_{jk} = \Pr(B = j, G = k)$, which describes the four scenarios resulting from our equations for G and B. Figure 1 gives a visual representation of the $\mathbf{\Pi}$ matrix.

Note in Figure 1 that μ determines the location (center of ellipse) and the correlation ρ determines the tilt and concentration of the probability contours. Inducement introduces an extra parameter which lowers the threshold for bankruptcy by γ.

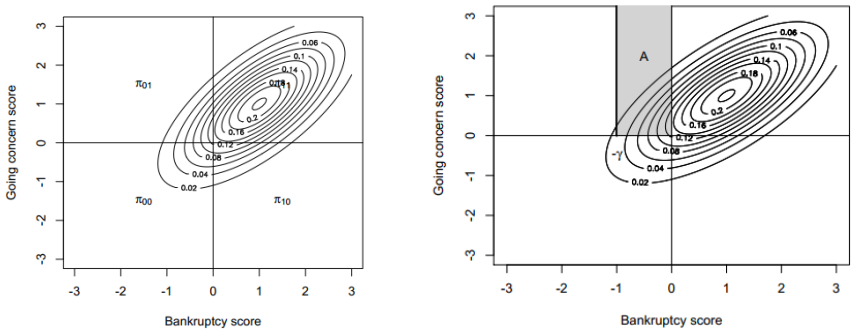


Fig 1: The bivariate probit entails ellipse shaped probability contours, where (when $\gamma = 0$) the probability mass associated to each quadrant represents the four combinations of the bivariate binary observed variables (B, G). The shaded region in the right panel, labeled "A", is subtracted from the upper left quadrant and added to the upper right quadrant when a going concern is issued, thus reflecting the endogeneity of the going concern variable. The parameters ρ and γ are estimable because changes in the shape of the ellipses, governed by ρ, lead to distinct apportioning of probability than do changes in the width of the A region, governed by γ.

2.1. *A causal interpretation of* γ. Having presumed a particular parametric model for the distribution of the data (G, B) (conditional on covariates $\mathbf{x}$), we would like additional license for the interpretation that ρ captures the contribution of auditor's additional information on bankruptcy likelihood while γ captures the contribution of inducement effects on bankruptcy likelihood. To justify this interpretation, we turn to the causal analysis framework of **?**. Recall that in Pearl's framework, the inducement effect would be written as

$$\Pr(B = 1 \mid \mathbf{x}, \text{do}(G = 1))/\Pr(B = 1 \mid \mathbf{x}, \text{do}(G = 0)), \tag{7}$$

where $\text{do}(G = 1)$ denotes the intervention of issuing a going concern, irrespective of the stochastic data generating process. Denote by U the auditor's additional information. Suppressing the covariates $\mathbf{x}$, the relationship between G and B can be expressed using the causal diagram depicted in Figure 2.

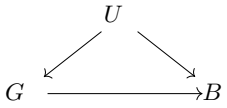


Fig 2: Conditional on observable attributes $\mathbf{x}$ (not shown), the causal diagram above stipulates the temporal ordering among the firm's private information U, the auditor risk assessment G, and the firm's bankruptcy outcome, B.

This diagram asserts several causal assumptions. First, the issuance of a going concern does not cause the existence of auditor's additional information: there is no arrow running from G to U. Second, bankruptcies cannot cause going concerns: there is no arrow running from B to G. Similarly, bankruptcies do not cause the creation of auditor's additional information for predicting bankruptcy: there is no arrow from B to U. All of these assumptions follow straightforwardly from a temporal ordering—auditor's first procure information concerning bankruptcy propensity (U), they then issue going concern opinions (G) and then firms either go bankrupt or not (B). Because U disconnects alternative routes from B to G

and no directed path exists from G to U, U is said to satisfy the back-door criterion **?**, and we can compute $\Pr(B = 1 \mid \mathbf{x}, \text{do}(G = 1))$ via the expression:

$$\Pr(B = 1 \mid \mathbf{x}, \text{do}(G = 1)) = \int \Pr(B = 1 \mid \mathbf{x}, U = u, G = 1) f(u) \mathrm{d}u, \tag{8}$$

where $f(u)$ is the marginal density of the random variable U.

The difficulty, of course, is that U is unobserved in our problem so $f(u)$ can never be estimated from data. However, we can re-express the bivariate probit model directly in terms of U in order to derive expression (8) in terms of parameters ρ, γ, and β. This demonstrates how the functional form of the model dictates the causal estimand in (7), which in turn establishes the causal interpretation of the γ parameter.

In detail, re-writing (4) conditional on U gives a model with diagonal error covariance:

$$\begin{pmatrix} Z_{g,i} \\ Z_{b,i} \end{pmatrix} \sim \text{N}(\mu, \Sigma), \quad \mu = \begin{pmatrix} \beta_0 + \beta_1 \mathbf{x}_i + \eta_g U \\ \alpha_0 + \alpha_1 \mathbf{x}_i + \eta_b U \end{pmatrix}, \quad \Sigma = \begin{pmatrix} v_g & 0 \\ 0 & v_b \end{pmatrix}, \tag{9}$$

where $U \sim \text{N}(0, 1)$, $v_g = 1 - \eta_g^2$, $v_b = 1 - \eta_b^2$ and $\rho = \eta_g \eta_b$. Although this representation is non-unique in $(\eta_g, \eta_b, v_g, v_b)$, it turns out that expression (8) will not depend on these values. This representation allows us to apply the causal assumptions depicted in the causal diagram above, which in turn allows us to derive the counterfactual probability of bankruptcy as:

$$\begin{aligned}
\Pr(B = 1 \mid \mathbf{x}, \text{do}(G = 1)) &= \int \Pr(B = 1 \mid \mathbf{x}, U = u, G = 1) \text{N}_u(0, 1) \mathrm{d}u, \\
&= \int 1 - \Phi(0; \gamma + \alpha_0 + \alpha_1 \mathbf{x} + \eta_b u) \text{N}_u(0, 1) \mathrm{d}u, \\
&= \int 1 - \int_{-\infty}^{0} \text{N}_w(\gamma + \alpha_0 + \alpha_1 \mathbf{x} + \eta_b u, v_b) \mathrm{d}w \text{N}_u(0, 1) \mathrm{d}u, \\
&= 1 - \int_{-\infty}^{0} \int \text{N}_w(\gamma + \alpha_0 + \alpha_1 \mathbf{x} + \eta_b a, v_b) N_u(0, 1) \mathrm{d}a \mathrm{d}w, \\
&= 1 - \int_{-\infty}^{0} \text{N}_w(\gamma + \alpha_0 + \alpha_1 \mathbf{x}, 1) \mathrm{d}w, \\
&= 1 - \Phi(0; \gamma + \alpha_0 + \alpha_1 \mathbf{x}), \\
&= \Phi(\gamma + \alpha_0 + \alpha_1 \mathbf{x}).
\end{aligned} \tag{10}$$

Here $\Phi(0; \mu)$ denotes the CDF of a normal distribution with mean μ and variance 1, evaluated at 0. A similar calculation can be done for $\Pr(B = 1 \mid \mathbf{x}, \text{do}(G = 0))$, allowing us to recover the causal risk ratio as

$$\tau(\mathbf{x}_i) = \Phi(\gamma + \alpha_0 + \mathbf{x}\alpha_1) / \Phi(\alpha_0 + \mathbf{x}\alpha_1).$$

In other words, fitting a bivariate probit model to the data $(G, B, \mathbf{x})$, coupled with the causal assumptions encoded in the causal diagram (Figure 2), implies a causal inducement effect that can be written in terms of α and γ. Although γ is a shared constant parameter, its impact on the risk ratio for a given firm will depend on both $\mathbf{x}$ and α.

2.2. *Identification and estimation for bivariate probit models.* The previous section related the parameters of the bivariate probit model with endogenous regressor to the causal risk ratio. However, identifiability is a distinct concern. Identification of parameters in bivariate probit models is subtle and deserves a careful discussion. The treatment in **?** derives the bivariate probit model from a system of simultaneous equations. Section 3 of **?**, page 949, provides a proof that the associated reduced form parameters of the model are identified without any exclusion restrictions, i.e., the going concern and bankruptcy equations do not share all of their covariates in common. Identification follows from the functional form of the probit likelihood, and indeed **?** contains a section devoted to maximum likelihood estimation. **?** also treats the continuous (non-binary response) version of the same structural system; in that case, exclusion restrictions are necessary for identification, and in that case estimation can proceed by a two-stage least squares procedure without specifying a likelihood function.

? study an applied problem using the binary response formulation of the **?** model, but do not assume the probit formulation and rather proceed to estimate parameters using an OLS based procedure. In this context, the role of an exclusion restriction is an open question as **?** point out; however, the two step procedure applied to the binary response setting gives inconsistent estimates.

In summary, textbook treatments of the bivariate probit model equivocate on the necessity of an exclusion restriction [**?**, Chapter 15]. To be clear, if one assumes the bivariate probit formulation, then an exclusion restriction is not necessary. If fitting a generalized linear model to a bivariate binary response *without* specifying a link function, it is unknown if (but plausible that) an exclusion restriction is necessary. Here, these concerns are secondary, as we do not demand identification, but proceed instead via a sensitivity analysis.

3. Modular sensitivity analysis with machine learning. In this section we propose our new approach for machine learning-based sensitivity analysis by generalizing the bivariate probit model. We begin by defining the joint probability of treatment and outcome as

$$\Pr\left(B, G \mid \mathbf{x}\right) = \int_{\mathbb{R}} \Pr\left(B \mid \mathbf{x}, U = u, G\right) \Pr\left(G \mid \mathbf{x}, U = u\right) f(u) \mathrm{d}u \tag{11}$$

for latent variable U. In this formulation, U has two special properties. First, it is assumed to be the *orthogonal* component of the private information in the sense that $U \perp\!\!\!\perp X$, hence $\mathbf{x}$ does not appear in $f(u)$. Second, U is assumed to be *complete*, in the sense that $\Pr\big(B \mid \mathbf{x}, u, G\big)$ can be interpreted causally in G, because U is a sufficient control variable. That is, $\Pr\big(B^1 \mid \mathbf{x}, u\big) = \Pr\big(B \mid \mathbf{x}, \mathrm{do}(G = 1), u\big) = \Pr\big(B \mid \mathbf{x}, G = 1, u\big)$ and similarly for $G = 0$; accordingly, the inducement effect for firm i is

$$\tau(\mathbf{x}_i) \equiv \frac{\int_{\mathbb{R}} \Pr\left(B = 1 \mid \mathbf{x}, G = 1, u\right) f(u) \mathrm{d}u}{\int_{\mathbb{R}} \Pr\left(B = 1 \mid \mathbf{x}, G = 0, u\right) f(u) \mathrm{d}u}. \tag{12}$$

Because the outcome and treatment are both binary, we can expand this probability into its four constituent parts. For convenience, we specify a probit link, yielding

$$\begin{aligned} \Pr\left(B = 1 \mid \mathbf{x}, U = u, G = 1\right) &= \Phi\left(b_1(\mathbf{x}) + u\right), \\ \Pr\left(B = 1 \mid \mathbf{x}, U = u, G = 0\right) &= \Phi\left(b_0(\mathbf{x}) + u\right), \\ \Pr\left(G = 1 \mid \mathbf{x}, U = u\right) &= \Phi\left(g(\mathbf{x}) + u\right). \end{aligned} \tag{13}$$

Therefore, in terms of f, b_1, b_0 and g, the individual inducement effect for firm i is

$$\tau(\mathbf{x}_i) = \frac{\int_{\mathbb{R}} \Phi\big(b_1(\mathbf{x}) + u\big) f(u) \mathrm{d}u}{\int_{\mathbb{R}} \Phi\big(b_0(\mathbf{x}) + u\big) f(u) \mathrm{d}u} \tag{14}$$

and we denote the sample average inducement effect (or average causal risk ratio: ACRR) as $\bar{\tau} = \frac{1}{n}\sum_{i=1}^{n} \tau(\mathbf{x}_i)$. Importantly, the orthogonality and completeness of U, as well as the choice of the probit link, are not substantive assumptions, as U is unobserved and b_1, b_0 and g are nonparametric functions of $\mathbf{x}$. Rather, these assumptions *define* U and give the specification of $f(\cdot)$ meaning; the choice of f, therefore, *is* a substantive assumption (as it is in the bivariate probit model as well).

This formulation entails that as $u \to -\infty$, the probability of bankruptcy approaches 0, regardless of whether the treatment is administered or not. As $u \to \infty$, the probability of bankruptcy approaches 1. The special case $u = 0$ corresponds to no unobserved confounding and the inducement effect can be computed directly from the observed joint probabilities. Finally, because G and B must have a valid joint distribution at each $\mathbf{x}$ value, we have the following system of equations defining our data generating process:

$$\begin{aligned}
\Pr\big(B = 1, G = 1 \mid \mathbf{x}\big) &= \int_{\mathbb{R}} \Phi\big(g(\mathbf{x}) + u\big) \Phi\big(b_1(\mathbf{x}) + u\big) f(u)\, \mathrm{d}u\,, \\
\Pr\big(B = 1, G = 0 \mid \mathbf{x}\big) &= \int_{\mathbb{R}} \Big(1 - \Phi\big(g(\mathbf{x}) + u\big)\Big) \Phi\big(b_0(\mathbf{x}) + u\big) f(u)\, \mathrm{d}u\,, \\
\Pr\big(B = 0, G = 1 \mid \mathbf{x}\big) &= \int_{\mathbb{R}} \Phi\big(g(\mathbf{x}) + u\big) \Big(1 - \Phi\big(b_1(\mathbf{x}) + u\big)\Big) f(u) \mathrm{d}u.
\end{aligned} \tag{15}$$

Observe that this generalizes the bivariate probit model with endogenous regressor: when $U \sim \mathrm{N}(0, \rho/(1-\rho))$, $b_0(\mathbf{x}) = \alpha_0 + \alpha_1 \mathbf{x}$, $b_1(\mathbf{x}) = \alpha_0 + \alpha_1 \mathbf{x} + \gamma$, and $g(\mathbf{x}) = \beta_0 + \beta_1 \mathbf{x}$ we recovers that model exactly. Our formulation is quite a lot more flexible: we relax the Gaussian assumption on the marginal distribution of U, drop the parallel relationship between $b_0(\cdot)$ and $b_1(\cdot)$, and allow b_1, b_0 and g to be nonlinear[2]. The price of the extra flexibility of our relaxed specification is that $f(u)$ is now unidentified, whereas in the bivariate probit case it is assumed to be Gaussian but with an identified correlation parameter ρ.

The left hand side of the system in (15) — the *reduced form* parameters — can be estimated from the observed data. Any of a host of machine learning classification methods, such as random forest [**?**], xgboost [**?**], Bayesian additive regression trees (BART) [**?**], among others, can be used to obtain estimates of these probabilities. Here, we focus our attention on BART for two reasons: one, we can impose monotonicity so that going concerns can only increase the probability of bankruptcy, and two, we obtain a Bayesian measure of uncertainty based on Markov chain Monte Carlo sampling methods.

3.1. *Projecting the reduced form probabilities onto the causal parameters.* What remains is to solve for $b_1(\cdot), b_0(\cdot), g(\cdot)$, the *structural*, or causal, parameters. To do so, we

[2]Observe that when the form of b_1, b_0 and g are constrained, as in the linear probit model, the choice of the probit link becomes a substantive modeling assumption, while in our more flexible formulation it is merely a convenience.

take a numerical approach, by minimizing the sum of the squared distance between the three left-hand right-hand pairs in (15):

$$\left[\Phi^{-1}\Big(\Pr\big(B=1,G=1\mid\mathbf{x}\big)\Big)-\Phi^{-1}\left(\int_{\mathbb{R}}\Phi\big(g(\mathbf{x})+u\big)\Phi\big(b_1(\mathbf{x})+u\big)f(u)\,\mathrm{d}u\right)\right]^2+$$

$$\left[\Phi^{-1}\Big(\Pr\big(B=1,G=0\mid\mathbf{x}\big)\Big)-\Phi^{-1}\left(\int_{\mathbb{R}}\Big(1-\Phi\big(g(\mathbf{x})+u\big)\Big)\Phi\big(b_0(\mathbf{x}+u)\big)f(u)\,\mathrm{d}u\right)\right]^2+$$

$$\left[\Phi^{-1}\Big(\Pr\big(B=0,G=1\mid\mathbf{x}\big)\Big)-\Phi^{-1}\left(\int_{\mathbb{R}}\Phi\big(g(\mathbf{x})+u\big)\Big(1-\Phi\big(b_1(\mathbf{x})+u\big)\Big)f(u)\mathrm{d}u\right)\right]^2.$$

Although it is unclear that (15) has a unique solution in b_1, b_0, g, numerical solvers converge readily in our experience. Heuristically, as a convex combination of monotone functions, each of the individual integrals in (15) is likely to be nearly linear over much of its domain. Note that the use of the normal inverse CDF simply ensures that the range of our objective function is unbounded; we observe that this improves numerical stability of our solver.

We refer to this process as *modular* because it requires fitting the reduced form model just one time. Sensitivity of the causal estimates to different choices of f can be assessed independently using the same estimates (or posterior samples) from a single reduced form model fit.

4. Monotone BART for reduced form inference.

4.1. *Probit BART Overview.* BART, Bayesian additive regression trees, is at its core a sum-of-trees model. For a $p-$dimensional vector of covariates $\mathbf{x}$ and a continuous response variable y, the BART model is

$$Y=t(\mathbf{x})+\varepsilon,\qquad \varepsilon\sim\mathrm{N}(0,\sigma^2) \tag{16}$$

where $t(\mathbf{x})=\mathrm{E}(Y\mid\mathbf{x})$ denotes a sum of L regression trees, i.e. $t(\mathbf{x})=\sum_{l=1}^{L}q_l(\mathbf{x})$. Figure 3 for presents an example regression tree. In addition to this additive tree representation, BART uses a stochastic process tree prior that favors smaller trees; the prior probability of splitting at depth d is $\eta(1+d)^{-\zeta}$, $\eta\in(0,1)$, $\zeta\in[0,\infty)$ [**?**].

At each leaf of the tree parameters are assigned independent regularization priors, $m_{lb}\sim\mathrm{N}(0,\sigma_\mu^2)$, where $\sigma_\mu=0.5/(k\sqrt{L})$, where L is the number of trees.

To handle binary outcomes, BART may be extended through a latent probit formulation, using the data augmentation approach of **?**. For binary outcome B

$$\begin{aligned}B^*&=t(\mathbf{x})+\varepsilon,\qquad \varepsilon\sim\mathrm{N}(0,1),\\ B&=\mathbb{1}(B^*>0),\end{aligned}$$

which implies

$$\Pr\big(B=1\mid\mathbf{x}\big)=\Phi(t(\mathbf{x})) \tag{17}$$

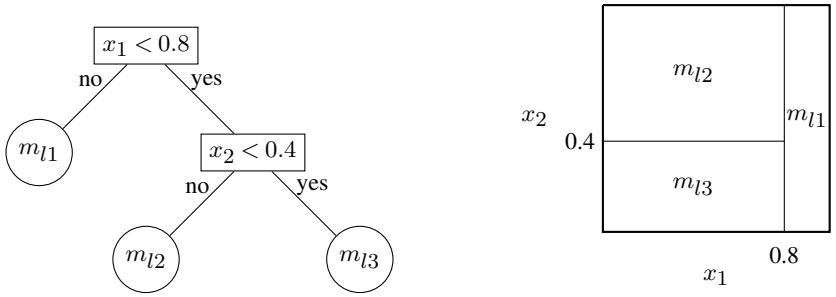


Fig 3: (Left) An example binary tree, with internal nodes labelled by their splitting rules and terminal nodes labelled with the corresponding parameters m_{lb}. (Right) The corresponding partition of the sample space and the step function. Figure from [**?**].

where Φ is the standard normal CDF.

The B^* variables may be imputed from their truncated normal full conditional distributions; conditional on B^* the BART fitting algorithm can be applied as usual.

4.2. *Monotone probit BART.* We turn now to a modification of the BART probit model for the bankruptcy and going concern data. We model the left-hand side of the system in (15) using a compositional representation, using two "chained" regression models, one for $\Pr\big(G \mid \mathbf{x}\big)$ and another for $\Pr\big(B \mid \mathbf{x}, G\big)$. This formulation permits us to insist that $\Pr\big(B = 1 \mid G = 1, \mathbf{x}\big) \geq \Pr\big(B = 1 \mid G = 0, \mathbf{x}\big)$ for all $\mathbf{x}$, encoding the uncontroversial belief that going concern opinions never mitigate bankruptcy risk. To enforce this constraint, we parameterize $\Pr\big(B = 1 \mid G, \mathbf{x}\big)$ as follows:

$$
\begin{aligned}
\Pr\big(B = 1 \mid G = 1, \mathbf{x}\big) &= \Phi[h_1(\mathbf{x})], \\
\Pr\big(B = 1 \mid G = 0, \mathbf{x}\big) &= \Phi[h_0(\mathbf{x})] \Pr\big(B = 1 \mid G = 1, \mathbf{x}\big), \\
&= \Phi[h_0(\mathbf{x})] \Phi[h_1 \mathbf{x})], \\
\Pr\big(G = 1 \mid \mathbf{x}\big) &= \Phi[w(\mathbf{x})].
\end{aligned} \tag{18}
$$

For each function h_0, h_1, and w we specify independent BART priors which allows us to fit the treatment and outcome models separately.

The likelihood for the bankruptcy model is

$$
\begin{aligned}
L(h_0, h_1; B, G, \mathbf{X}) = &\prod_{i:G_i=1} \Phi(h_1(\mathbf{x}_i))^{B_i} (1 - \Phi(h_1(\mathbf{x}_i)))^{1-B_i} \times \\
&\prod_{i:G_i=0} [\Phi(h_0(\mathbf{x}_i)) \Phi(h_1(\mathbf{x}_i))]^{B_i} (1 - \Phi(h_0(\mathbf{x}_i)) \Phi(h_1(\mathbf{x}_i)))^{1-B_i}.
\end{aligned} \tag{19}
$$

This likelihood is challenging: The expression $1 - \Phi(h_0(\mathbf{x}_i))\Phi(h_1(\mathbf{x}_i))$ does not factor into separate terms involving the unknown functions h_0 and h_1, making it difficult to adapt the BART MCMC sampler for posterior inference. To overcome this, we introduce a data-augmented representation that permits updating h_0 and h_1 independently using standard MCMC for probit BART.

To begin, note that the first term above (corresponding to $G = 1$) involves only h_1 so we only need to augment data in the $G = 0$ "arm". When $G = 0$, we relate B to two independent binary latent variables R_0 and R_1 as follows:

$$\Pr\big(R_0 = 1 \mid \mathbf{x}, G = 0\big) = \Phi(h_0(\mathbf{x})),$$
$$\Pr\big(R_1 = 1 \mid \mathbf{x}, G = 0\big) = \Phi(h_1(\mathbf{x}))$$

and $B = R_0 R_1$. Integrating out the latent variables gives $\Pr\big(B = 1 \mid \mathbf{x}, G = 0\big) = \Phi(h_0(\mathbf{x}))\Phi(h_1(\mathbf{x}))$ and $\Pr\big(B = 0 \mid \mathbf{x}, G = 0\big) = 1 - \Phi(h_0(\mathbf{x}))\Phi(h_1(\mathbf{x}))$ as required[3]. The augmented likelihood function (including R_0, R_1) is

$$\begin{aligned} L(h_0, h_1; R, B, G, \mathbf{X}) = &\prod_{i:G_i=1} \Phi(h_1(\mathbf{x}_i))^{B_i}(1 - \Phi(h_1(\mathbf{x}_i)))^{1-B_i} \times \\ &\prod_{i:G_i=0} \Phi(h_1(\mathbf{x}_i))^{R_{1i}}(1 - \Phi(h_1(\mathbf{x}_i)))^{1-R_{1i}} \times \\ &\prod_{i:G_i=0} \Phi(h_0(\mathbf{x}_i))^{R_{0i}}(1 - \Phi(h_0(\mathbf{x}_i)))^{1-R_{0i}} \times \\ &\prod_{i:G_i=0} \mathbb{1}(B_i = 1 \text{ if } R_{0i} = R_{1i} = 1) \end{aligned} \tag{20}$$

After rearranging terms, we have two separate probit likelihoods in h_0 and h_1 (and the domain restriction in the last term). Conditional on R_0, R_1 we can update h_0, h_1 using standard probit BART MCMC steps. To update the latent variables R_{0i} and R_{1i}, first note that they are fixed at 1 when $B_i = 1, G_i = 0$. When $B_i = 0, G_i = 0$, $R_i \equiv (R_{0i}, R_{1i})$ is sampled from:

$$\begin{aligned} \Pr\big(R_i = r \mid h_0, h_1, B_i = 0, G_i = 0\big) \propto &\Phi(h_0(\mathbf{x}_i))^{R_{0i}}(1 - \Phi(h_0(\mathbf{x}_i)))^{1-R_{0i}} \times \\ &\Phi(h_1(\mathbf{x}_i))^{R_{1i}}(1 - \Phi(h_1(\mathbf{x}_i)))^{1-R_{1i}} \times \\ &\mathbb{1}(r \neq (1, 1)), \end{aligned} \tag{21}$$

which is the joint probability distribution of the latent variables from Eq. (4.2), truncated away from the $R_{0i} = R_{1i} = 1$ region.[4]

5. Empirical analysis of bankruptcy data. In this section, we study the question of whether unfavorable going concern opinions cause bankruptcy. We conduct a modular sensitivity analysis based on a monotone BART model fit. This combination allows us to use machine learning methods to learn potentially complex functional forms for the observable data distribution – while reaping the estimation benefits of imposing monotonicity – and obtain valid measures of uncertainty for average and subgroup average effects under different assumptions about the distribution of private information.

Data collection is described in section 5.1. Results are presented in section 5.2, specifically posterior summaries of firm-year estimated inducement effects as $f(u)$ is varied. For illustration, several individual firms are investigated in section 5.4. Finally, firm characteristics which moderate the inducement effect are investigated in section 5.5.

[3]Observe that $\Pr\big(B = 1 \mid \mathbf{x}, G = 1\big) = \Pr\big(R_1 = 1 \mid \mathbf{x}, G = 0\big)$, so thinking about this as a generative model we can interpret R_1 as a simulated outcome if we had observed $G = 1$ and R_0 as an indicator that this outcome is "thinned" to enforce monotonicity, since in reality $G = 0$.

[4]Formally, this MCMC sampler effects joint updates for R_i and the latent variables in the two probit BART models

5.1. *Data.* Data was collected and merged from Audit Analytics, Compustat, and BankruptcyData.com for the sample period of 2000–2014 leading to 25,350 firm-year observations. The bankruptcy indicator was assigned value of 1 if it occurred within a year of the audit report. This was done because Statement of Auditing Standards No. 59 requires audit firms to opine whether there is substantial doubt regarding a client's ability to continue operating as a "going concern" over the twelve months following the financial statement audit.

The following are the control covariates that constitute $\mathbf{x}$:

1. `Log(Assets)`: Natural log of total assets
2. `Leverage`: Ratio of total liabilities to total assets
3. `Investment`: Ratio of short-term investments to total assets
4. `Cash`: Ratio of cash and cash equivalents to total assets
5. `ROA`: Ratio of income before extraordinary items to total assets
6. `Log(Price)`: Natural log of stock price
7. `Intangible assets`: Ratio of intangible assets to total assets
8. `R&D`: Ratio of research and development expenditures to sales
9. `R&D missing`: Indicator for missing R&D expenditures
10. `No S&P rating`: Indicator for the existence of a S&P credit rating
11. `Rating below CCC+`: Indicator for S&P credit rating below CCC+
12. `Rating downgrade`: Indicator for an S&P credit rating downgrade from above CCC+ to CCC+ or below
13. `Non-audit fees`: Ratio of non-audit fees to total audit fees
14. `Non-audit fees missing`: Indicator for missing non-audit fees
15. `Years client`: Number of years of client used auditor

These variables are similar to those used in **?**, which were inspired by **?**, and were chosen due to their potential relevance to a companies' upcoming bankruptcy risk as well as their relevance to the issuance of a going concern opinion.

5.2. *Sensitivity to the distribution of private information.* For fixed conditional probabilities on (B, G) outcomes (15), different choices of $f(u)$ will yield different causal estimates based on solutions to (b_0, b_1, g). Specifically, the right tail of the density $f(u)$ governs how likely an auditor is to observe information that would make a bankruptcy much more likely than suggested by the available covariates, while the left tail governs how likely an auditor is to observe information that would make bankruptcy much less likely than indicated by the available covariates. For reference, in a bivariate probit analysis, $f(u)$ is assumed to have a $\mathrm{N}(0, \sigma)$ distribution, where $\sigma = \sqrt{\rho/(1-\rho)}$; larger σ means the available covariates are a more incomplete guide to actual bankruptcy risk. Table 1, reports estimated inducement effects for various specifications of the standard deviation of the private information, $\sigma = \sqrt{V(U)}$.

In addition to varying σ for a Gaussian distribution over U, we also consider a unimodal asymmetric specifications, reflecting the belief that the unreported information is more likely to inflate (or deflate) bankruptcy probabilities even though it is most likely that there is no private information. Specifically, we consider a skewed unimodal (at zero) density with Gaus-

sian tails called the "sharkfin" [**?**], which has the following expression:

$$\pi(\beta) = \begin{cases} 2qf(\beta) & \beta \leq 0 \\ 2f\left(\frac{\beta}{1-q} \cdot q\right) \cdot q & \beta > 0 \end{cases} \tag{22}$$

where $f(\cdot)$ is the pdf of the normal distribution with standard deviation s, and $q = \Pr(U < 0)$ controls the skewness. The right panel of Figure 4 depicts two sharkfin densities with $q = 0.1$ and $q = 0.9$ for illustration.

Additionally, we consider two three-component Gaussian mixtures, one symmetric about zero and the other asymmetric with a high weight on the component with the positive mean parameter:

$$f(u) = 0.05\phi(u; -2, s^2) + 0.90\phi(u; 0, s^2) + 0.05\phi(u; 2, s^2)$$

and

$$f(u) = 0.01\phi(u; -2, s^2) + 0.94\phi(u; 0, s^2) + 0.05\phi(u; 2, s^2)$$

respectively, with $s = 0.05$. Each of these models reflects the case of a small possibility of quite strong positive or negative private information regarding a firm's bankruptcy risk.

Table 1 reports posterior estimates of the average inducement effect across the firms in our study for various choices of $f(u)$. The left panel of Figure 4 shows the sample average inducement effect (causal risk ratio) as a fraction of the observed risk ratio plotted against σ (the standard deviation of U) for various specifications of $f(u)$; consistent with intuition, it shows that greater dispersion of $f(u)$ drives the estimated inducement effect to zero, while the skewness dictates the rate of decay.

Distribution of $f(u)$	inducement posterior mean	95% Credible interval for mean inducement
$\mathrm{N}(0, \sigma = 0.1)$	48.7	$(2.14, 337)$
$\mathrm{N}(0, \sigma = 0.5)$	15.5	$(1.18, 97.4)$
$\mathrm{N}(0, \sigma = 1)$	2.33	$(1.00, 9.51)$
Shark $q = 0.25$, $s = 0.5$ ($\sigma = 1.05$)	1.22	$(1.00, 2.48)$
Shark $q = 0.75$, $s = 1.25$ ($\sigma = 0.88$)	12.6	$(1.05, 80.9)$
Symmetric mixture ($\sigma = 0.64$)	9.30	$(1.00, 72.1)$
Asymmetric mixture ($\sigma = 0.48$)	9.73	$(1.00, 76.0)$

TABLE 1

The reduced form probabilities (15) 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.

5.3. *Comparison with the E-value.* Rather than modeling the distribution of unobserved information $f(u)$, an alternative approach is to consider the strength of unobserved confounding that would be necessary to entirely explain the observed association. This approach can be found as early as **?**, and has recently been generalized in **?** and **?**, who prove that

$$\max(\mathrm{RR}_{GU}, \mathrm{RR}_{UB}) \geq \left\{ \mathrm{RR}_{GB}^{\text{obs}} + \sqrt{\mathrm{RR}_{GB}^{\text{obs}}\left(\mathrm{RR}_{GB}^{\text{obs}} - \mathrm{RR}_{GB}^{\text{true}}\right)} \right\} \Big/ \mathrm{RR}_{GB}^{\text{true}} \tag{23}$$

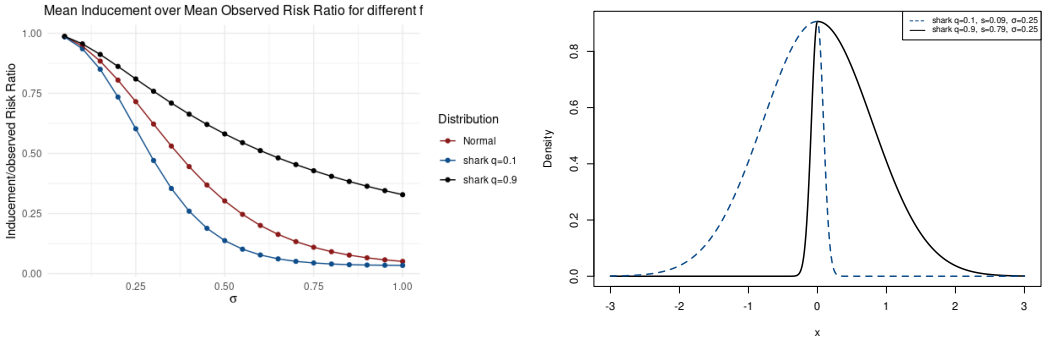


Fig 4: Left: Plot of inducement effect over observed risk ratio for different standard deviations σ for f normally distributed (red), sharkfin ($q = 0.9$) with right skew (black), and sharkfin ($q = 0.1$) with left skew (blue). The mean observed risk ratio was 30.80. On right is a plot of the shark fin with $q = 0.1$ and $q = 0.9$, for visual purposes.

where

$$\text{RR}_{GU|\mathbf{x}} = \max_{k} \frac{\Pr\left(U = k \mid G = 1, \mathbf{x}\right)}{\Pr\left(U = k \mid G = 0, \mathbf{x}\right)},$$

$$\text{RR}_{UB|\mathbf{x}} = \max_{k,k',g} \frac{\Pr\left(B = 1 \mid G = g, \mathbf{x}, U = k\right)}{\Pr\left(B = 1 \mid G = g, \mathbf{x}, U = k'\right)}$$

for $g \in \{0, 1\}$ and

$$\text{RR}_{GB}^{\text{true}} = \frac{\int \Pr\big(B = 1 \mid G = 1, \mathbf{x}, U\big) \Pr\big(U \mid \mathbf{x}\big) \mathrm{d}u}{\int \Pr\big(B = 1 \mid G = 0, \mathbf{x}, U\big) \Pr\big(U \mid \mathbf{x}\big) \mathrm{d}u}. \tag{24}$$

Figure 5 provides a visualization of these terms.

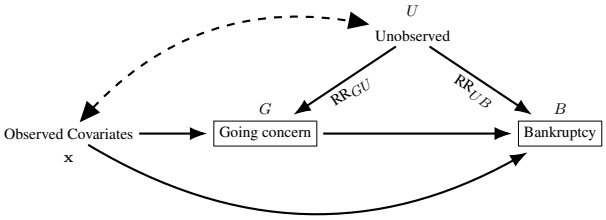


Fig 5: RR_{GU} is the maximum risk ratio comparing any two categories of confounding and RR_{UB} is the maximum risk ratio for any specific level of the unmeasured confounders comparing those with and without treatment, controlling for $\mathbf{x}$.

Setting $\text{RR}_{GB}^{\text{true}} = 1$ in expression 23, **?** define the *E-value* (for evidence value) as

$$\text{E-value} = \text{RR}_{GB}^{\text{obs}} + \sqrt{\text{RR}_{GB}^{\text{obs}}\left(\text{RR}_{GB}^{\text{obs}} - 1\right)}, \tag{25}$$

which can be interpreted as the minimum strength of association that an unmeasured confounder would need to have with both the G and B (conditional on $\mathbf{x}$) to fully explain the observed treatment-outcome association. Note that for large observed risk ratios (that is,

$RR^{\text{obs}} \approx RR^{\text{obs}} - 1$), the E-value is essentially proportional to the observed risk ratio itself. Accordingly, if we compare our model-based sensitivity analysis estimates to the E-value, we find that when $f(u)$ concentrates around zero, the associated causal risk ratio becomes the observed risk ratio, which is effectively the E-value. However, for different choices of $f(u)$, the associated causal risk ratio at different $\mathbf{x}$ values can differ from the observed risk ratio in interesting ways, which we explore in the following sections. Figure 6 plots posterior means of τ against the posterior mean of the E-value for the auditing data for the distributions of U reported in 11. Essentially, E-values are simply a scale multiple of the observed risk ratio, which is precisely the causal risk ratio when there is assumed to be no private information (lower right panel of Figure 6). However, less dogmatic choices of $f(u)$ also yield substantial inducement effect estimates for some firms (first three panels of Figure 6).

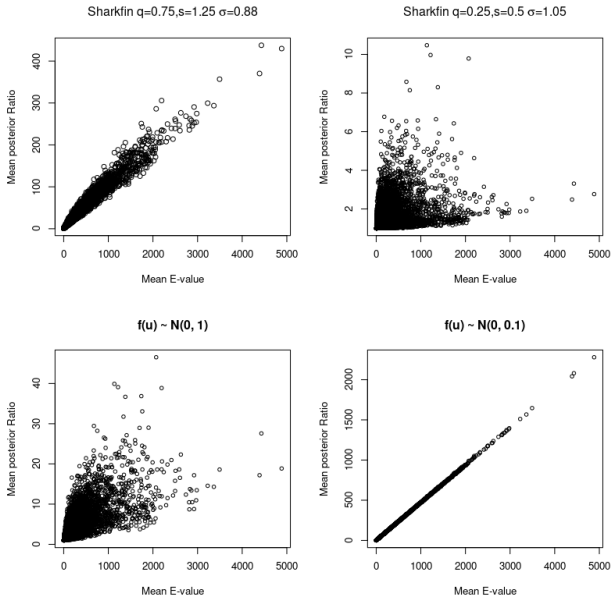


Fig 6: Posterior means of τ across 500 draws for different distributions of $f(u)$ vs the E-value per firm calculated from the posterior mean of the risk ratio from $\text{RR}^{\text{obs}}_{GB|\mathbf{x}} = \Pr\left(B = 1 \mid G = 1, \mathbf{x}\right) / \Pr\left(B = 1 \mid G = 0, \mathbf{x}\right)$.

5.4. *Posterior Individual Inducement Effects for Specific Firms.* By numerically solving 15 for (b_0, b_1, g) at each posterior draw, for a given firm-year observation and a given choice of $f(u)$, a full posterior distribution over causal estimands for that observation can be obtain. Scrutinizing these posteriors for specific firms provides an intuitive approach to investigating the results of the sensitivity analysis that is more granular than simply reporting sample averages across all observations. To this end, the posterior mean inducement effect, as well as a 95% credible interval, are presented in Table 2 for a selection of illustrative firms. Figure 7 depicts a histogram of posterior draws of the inducement effect for Apple (from year 2001) and Jetblue (from 2007) using asymmetric Gaussian mixture.

We find that the inducement effect varies both across posterior draws as well as across firms as a function of the density $f(u)$. Differences between firms are illuminating: For example, Apple in 2001 had a significantly higher inducement effect than Blockbuster in 2009,

Firm	Going Concern	Bankruptcy	Auditor	mean RR_{obs}	mean B_0	mean B_1	mean τ post	95% Credible interval for τ (%)	mean B_0	mean B_1	mean τ post	95% Credible interval for τ (%)
JetBlue (2007)	No	No	E&Y	45.0	0.005	0.062	16.8	$(2.32, 50.0)$	0.008	0.045	7.45	$(1.07, 25.3)$
JetBlue (2009)	No	No	E&Y	10.9	0.015	0.050	3.94	$(1.00, 12.6)$	0.022	0.038	2.32	$(1.05, 7.52)$
Apple (2001)	No	No	KPMG	403	0.001	0.062	103.7	$(6.62, 461)$	0.003	0.042	43.6	$(1.02, 408)$
Build a Bear (2010)	No	No	KPMG	23.3	0.004	0.021	7.23	$(1.13, 24.3)$	0.007	0.027	5.31	$(1.08, 11.1)$
Build a Bear (2014)	No	No	E&Y	49.1	0.004	0.039	15.1	$(2.62, 55.7)$	0.006	0.048	9.49	$(1.08, 25.3)$
Radioshack (2013)	No	No	PWC	4.66	0.122	0.297	2.99	$(1.02, 7.74)$	0.132	0.262	2.26	$(1.00, 5.79)$
Radioshack (2014)	No	Yes	PWC	3.03	0.143	0.251	1.98	$(1.00, 4.64)$	0.151	0.231	1.63	$(1.00, 3.99)$
Blockbuster (2004)	No	No	PWC	42.9	0.003	0.024	13.6	$(1.48, 54.8)$	0.007	0.016	4.10	$(1.07, 11.2)$
Blockbuster (2009)	Yes	No	PWC	4.89	0.082	0.222	3.02	$(1.15, 6.25)$	0.093	0.156	1.75	$(1.00.4.13)$
Six Flags (2006)	No	No	KPMG	14.4	0.017	0.084	6.13	$(1.36, 15.3)$	0.021	0.032	1.78	$(1.00, 4.96)$
Six Flags (2009)	Yes	Yes	KPMG	3.78	0.136	0.307	2.49	$(1.03, 5.66)$	0.144	0.307	2.32	$(1.00, 5.089)$

TABLE 2

Left: Posterior estimates of the inducement effect given $f(u) \sim N(0, \sigma = 0.5)$ for select firms. Right: Posterior estimates of the inducement effect given $f(u)$ is the asymmetric Gaussian mixture with an upweighted right component for the same firms.

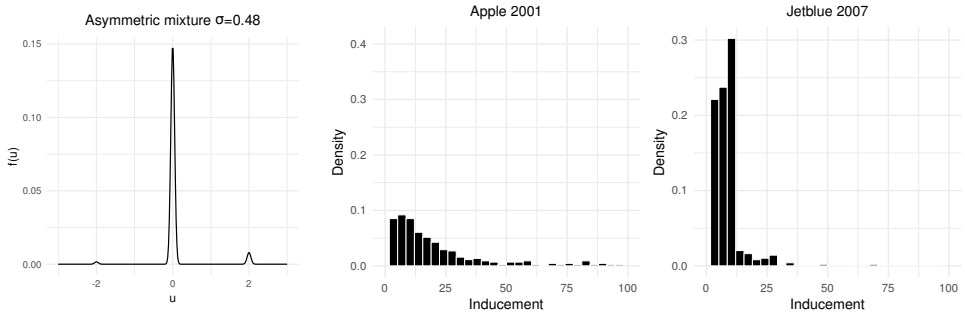


Fig 7: Histogram of posterior estimates of the individual inducement effects given $f(u)$ with the distribution on the left (moderate confounding) for Apple 2001 and JetBlue 2007. Neither received a going concern, nor did either go bankrupt. JetBlue was audited by Ernst and Young, and Apple audited by KPMG.

but this is at least in part an artifact of Apple 2001 having extremely low probability of bankruptcy. This points to a general phenomenon with risk ratios, which is that they can be dramatically impacted by the denominator; we explore this fact further in the following section.

5.5. *Exploratory subgroup analysis.* With firm-year specific treatment effects in hand, one can conduct an ex post regression tree analysis to isolate subgroups of firms with subgroup average treatment effects that depart from the overall average. Specifically, we identify moderating subgroup of variables by fitting a single regression tree using the individual inducement effect estimates (posterior means) as the response variable and observable firm (and auditor) features as predictors (as detailed in **?**). For predictors we use the same covariates reported in 5.1, all of which are plausible moderators of the inducement effect.

The subgroup analysis presented here is based on $U \sim \mathrm{N}(0, \sigma = 0.5)$ to the left hand side of equation (15). The left panel of Figure 8 shows the resulting tree fit. Using this tree, we can identify subgroups based on the corresponding partition implied by terminal node (leaf) membership. However, the resulting point estimates only tell part of the story. For a fuller picture, we can consider the posterior distribution of subgroup *differences*, even for different choices of $f(u)$ than the one used to produce the tree. We compute the subgroup difference of mean inducement effects for each posterior draw between the subgroups with the largest and

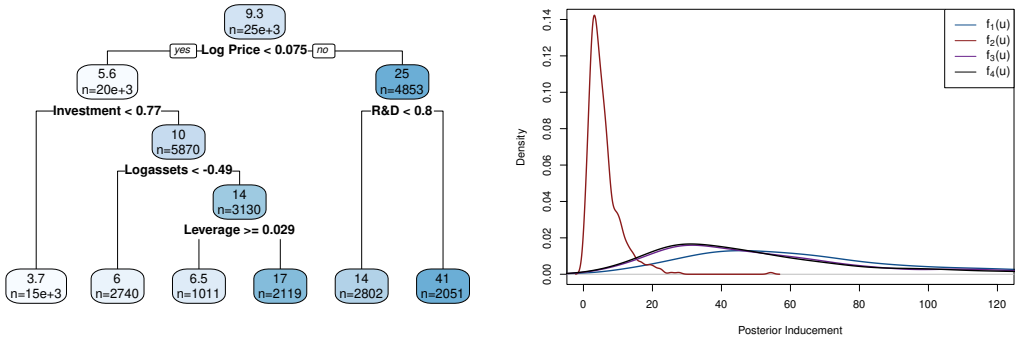


Fig 8: Left: A small tree fit to inducement effects (risk ratios). This is also the group of variables we investigate as moderators. Follow down tree to identify subgroup. Right: Plot of difference of inducement effects across the posterior draws between the largest and smallest inducement effect subgroups (bottom right and bottom left respectively on the tree).

smallest subgroup effects as determined by the regression tree. This analysis is repeated for four different distributions of $f(u)$: $f_1(u) \sim \mathrm{N}(0, \sigma = 0.5)$, $f_2 \sim \mathrm{N}(0, 1)$, $f_3(u)$ which is a a mixture model with more weight on a far bump to the right (see Figure 7), and $f_4(u)$ which is a three component Gaussian mixture with 90% of the area centered around 0, and 5% around $u = -2$ and $u = 2$. The right panel of Figure 8 shows posteriors of subgroup differences in inducement effects (causal risk ratios); the sign of the differences is preserved across various choice, while the magnitude varies (as one might anticipate).

With respect to economic interpretation, the tree presented Figure 8 shows that firms with higher stock prices [`Log(Price)`], greater assets [`Log(Assets)`], lower leverage [`Leverage`], and greater investments [`Investments` and `R&D`] have higher risk ratios. In general, these characteristics are consistent with such firms being "safe" (i.e., having a lower risk of bankruptcy). This interpretation is consistent with the risk ratios for these firms being driven by small denominators.

At this point it is instructive to consider if different estimands may be moderated by different covariates. In particular, risk ratios may be dominated by the denominator, which may be affected by different variables than those which affect the numerator. Accordingly, in Figure 9, we fit a regression tree to point estimates of the $\Pr\big(B = 1 \mid \mathbf{x}, \mathrm{do}(G = 0)\big)$. For this tree, we find that firms with lower leverage and higher stock prices have lower probabilities of bankruptcy absent a going concern opinion.[5]

We next consider the risk difference $\Pr\big(B = 1 \mid \mathbf{x}, \mathrm{do}(G = 1)\big) - \Pr\big(B = 1 \mid \mathbf{x}, \mathrm{do}(G = 0)\big)$. While risk ratios can be unappealingly large for firms with very small bankruptcy risk, risk differences (necessarily) have the opposite complication, which is that a difference of 0.1 "means" something quite different for a firm with control probability of 0.5 than it does for one with control probability 0.9. Fortunately, the risk difference has another interpretation in contexts like the present one where treatment effects are assumed to be monotonic: the risk difference is equivalent to the probability that a firm went bankrupt *because of* the going concern opinion. This interpretation is derived as follows. Consider the four possible potential outcomes, depicted in Table 3, which gives each configuration a suggestive name.

[5] **?** find similarly that the probability of bankruptcy increases in leverage and decreases in size and share price.

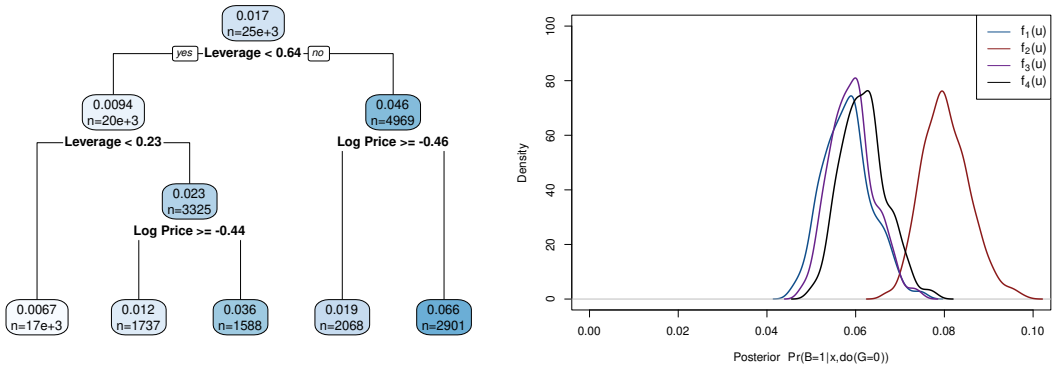


Fig 9: Left: A small tree fit to the $\Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=0)\big)$, B_0 for short. Right: Plot of differences of B_0 across the posterior draws between the largest and smallest B_0 effect subgroups (bottom right and bottom left respectively on the tree).

Name	B^1	B^0
No Inducement	0	0
Prevention	0	1
Induced bankruptcy	1	0
No prevention	1	1

TABLE 3

Because we are operating in the binary treatment/binary response world, we have just four outcomes. The first row refers a firm that, regardless of a receiving going concern opinion, does not go bankrupt ("No Inducement"). "Prevention" refers to the situation in which, without the treatment, the firm would have gone bankrupt, but with the going concern opinion it does not. We do not allow for this situation given our monotonicity assumption $\Pr\big(B=1 \mid \mathbf{x}, G=1\big) \geq \Pr\big(B=1 \mid \mathbf{x}, G=0\big)$. *"Induced bankruptcy" refers to the situation in which the firm goes bankrupt because of the going concern opinion. "No prevention" means, regardless of going concern opinion being issued, the company goes bankrupt.*

The marginal probabilities are then simply the sum of rows where "1" appears in the corresponding column of Table 3:

$$\begin{aligned} \Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=0)\big) &= \Pr(\text{Prevention}) + \Pr(\text{No prevention}) \\ \Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=1)\big) &= \Pr(\text{Induced bankruptcy}) + \Pr(\text{No prevention}) \end{aligned} \tag{26}$$

But, under the monotonicity assumption, $\Pr(\text{Prevention}) = 0$, in which case

$$\Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=1)\big) - \Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=0)\big) = \Pr(\text{Induced bankruptcy}). \tag{27}$$

Accordingly, in Figure 10, we fit a regression tree to point estimates of the (causal) risk difference $\Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=1)\big) - \Pr\big(B=1 \mid \mathbf{x}, \text{do}(G=0)\big)$. At the top of the tree, we find that firms with greater leverage are more likely to have an inducement effect. This result is consistent with **?**, who find that debt contracts often include covenants that mechanically increase interest rates when the borrow receives an adverse going concern opinion. The second level of the tree shows that larger firms are more likely to have an inducement effect. This result could be due to firms' information environments varying with firm size. At the third level, inducement is likely to occur when the firm has an S&P credit rating. Consistent with this result, **?** find that S&P tends downgrade credit ratings after the issuance of a going concern opinion.

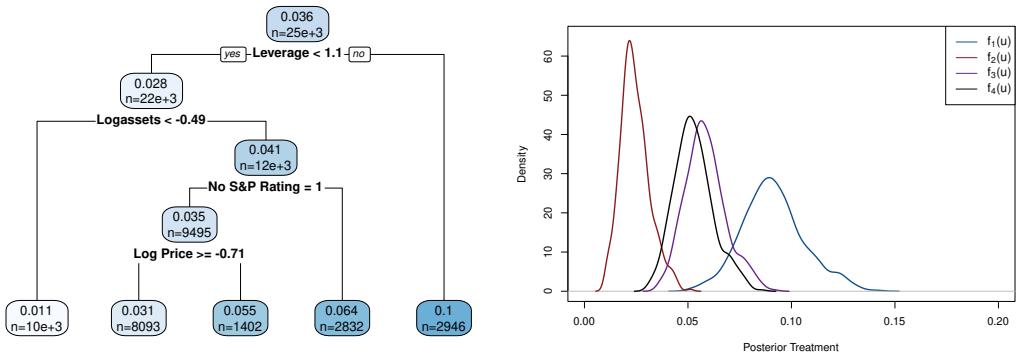


Fig 10: Left: A small tree fit to the risk difference, $\Pr\big(B=1 \mid \mathbf{x}, \mathrm{do}(G=1)\big) - \Pr\big(B=1 \mid \mathbf{x}, \mathrm{do}(G=0)\big)$, which is under monotonicity of going concerns is equivalent to the probability that the bankruptcy was induced. Right: Posterior subgroup differences between the largest and smallest treatment effect subgroups (bottom right and bottom left respectively on the tree).

It bears emphasis that the tree-based posterior subgroup analysis presented above is simply an exploration of the posterior distribution. Consequently, the posterior difference shown in the right panels of Figures 8, 9, and 10 require no further adjustment. Similarly, the CART fits presented in the left panels cannot be "over-fit". The posterior distribution is where the inferences are performed, CART is being used merely as a way to navigate a high dimensional posterior. Trees are restricted to be small to ease interpretation and to focus on subgroups with relatively large sample sizes. Ideally, these summaries would not be endpoints of an analysis, but the starting point for further investigation into the moderating role of particular attributes.

6. Simulation studies. In this section we investigate how the new method performs under a variety of different simulated settings, in an effort to build confidence in the empirical analysis above. We begin by comparing our method to the bivariate probit regression model in 6.1, and show that we perform comparably well when the data is generated according to the bivariate probit, and in 6.2 we show how better than the bivariate probit under more complicated non-linear data generating processes. 6.2 also explores how mis-specification of f affects our modeling. Specifically, we explore the effects of mis-specifying the scale, location, skew, and tail weights of different distributions and report on those effects. Section 6.3 repeats the results of Figure 6 but with simulated data. In 6.4, we show the benefits of using the monotonicity constraint in BART.

6.1. *Comparison to bivariate probit.* To verify that the proposed machine learning sensitivity analysis yields sensible answers, we take advantage of the relationship between our model and the bivariate probit model with endogenous binary regressor: if we generate the data from the bivariate probit model with $U \sim \mathrm{N}(0, \sigma = \sqrt{\rho/(1-\rho)})$, the true causal risk ratios should be recoverable[6]. Table 4 reports the results of fitting our sensitivity analysis

[6]Note, the success of our sensitivity analysis is predicated upon minimizing the squared distance between the three left hand-side pairs is equation(15). We use Nelder-Mead to do so, a commonly used numerical method for minimization of loss functions [**?**] (although we also employed a simulated annealing approach and the Broyden–Fletcher–Goldfarb–Shanno algorithm, both giving similar results as Nelder-Mead)

model data generated from the bivariate probit

$$\begin{pmatrix} Z_{g,i} \\ Z_{b,i} \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \mathbf{\Sigma}) \qquad \mu = \begin{pmatrix} \beta_0 + \beta_1 \mathbf{x}_i \\ \alpha_0 + \alpha_1 \mathbf{x}_i \end{pmatrix} \qquad \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

We simulated 25,000 samples, where we sum 5 uniform(-1,1) $\mathbf{x}_i$ covariates each with the same β_1 and α_1 coefficients respectively. We set $\beta_0 = 0, \beta_1 = -0.2, \alpha_0 = -0.5, \alpha_1 = -0.5$ to generate reasonable number of going concerns and bankruptcies. We fit the left hand side of equation(15) using BART with the monotonicity constraint, whose benefit is shown in section[6.4]. Note, in our simulations, we do not solve our systems for every BART posterior estimate of equation 15 due to computational constraints. Instead, we take the mean of the posterior BART probability estimates in the fitting stage and then solving for our causal parameters $b_0(\cdot), b_1(\cdot), g(\cdot)$ once for each observation[7]. We impose the constraint that $b_1(\mathbf{x}) \geq b_0(\mathbf{x})$ when solving for the causal parameters.

In simulations (not reported here) we observe that at $\mathrm{N} = 100,000$ the bivariate probit regression (unsurprisingly) works remarkably well when the data generation process is the bivariate probit[8]. Table 4 shows that our method works well even with $\mathrm{N} = 25,000$ and $p = 5$. The bivariate probit regression should perform better when correctly specified, but the broadly similar estimates in this case is reassuring that the projection approach is trustworthy.

γ	ρ	ACRR true	ACRR est	ICRR cor	ICRR rmse
1.00	0.25	2.90	2.94	0.88	1.12
1.75	0.25	5.35	5.08	0.88	3.74
2.50	0.25	8.34	7.37	0.89	11.03
1.00	0.40	2.90	2.82	0.86	1.06
1.75	0.40	5.35	4.99	0.90	3.57
2.50	0.40	8.34	6.74	0.89	12.83
1.00	0.60	2.90	2.77	0.83	1.18
1.75	0.60	5.35	4.68	0.86	4.45
2.50	0.60	8.34	6.75	0.85	13.53
1.00	0.80	2.90	2.23	0.61	1.82
1.75	0.80	5.35	3.35	0.67	6.97
2.50	0.80	8.34	4.53	0.67	18.99

TABLE 4

We simulate from the bivariate probit with 25,000 observations and deploy our methodology. ACRR = average causal risk ratio. ICRR = individual causal risk ratio, cor refers to the correlation between predicted and true for the individual causal risk ratios, and the rmse is the root mean square error.

[7] This methodology held for all the simulated data; when analyzing the real data we repeated the integrals for random samples of the posterior BART estimates

[8] It is well-known that maximum likelihood estimates of the bivariate probit model can be unstable (i.e., many local modes), especially when there are a large number of predictor variables (see **?** and **?**). Our simulations bear this out; with thousands of observations, estimates of ρ were quite inaccurate. Therefore, to verify that we obtain consistent parameter estimates with maximum likelihood estimation (and to cross-check our data generating process) we generate and train our models on 100,000 observations.

6.2. *Sensitivity to f.* We do much better with our methodology when the data were generated from a non-linear data generating process, as described below:

$$
\begin{aligned}
b_0(\mathbf{x}) &= \mathbf{x}_5 + \mathbf{x}_1 \sin(2\mathbf{x}_6) - 1.75 \\
b_1(\mathbf{x}) &= b_0(\mathbf{x}) + 1.5 \\
g(\mathbf{x}) &= 0.5 b_0(\mathbf{x}) + \mathbf{x}_2 + 0.25 \\
U &\sim \mathrm{N}(\mu, \sigma^2) \\
G &\sim \mathrm{Bin}\big(\Phi(g(\mathbf{x}) + u)\big) \\
B \mid G = 1 &\sim \mathrm{Bin}\big(\Phi(b_1(\mathbf{x}) + u)\big) \\
B \mid G = 0 &\sim \mathrm{Bin}\big(\Phi(b_0(\mathbf{x}) + u)\big)
\end{aligned}
\tag{28}
$$

where we draw u and $b_i(\cdot)$, G conditional on those values, and subsequently the values of B are drawn conditional on our values of G. The $\mathbf{x}_i$ are drawn uniform(-1,1), with some $\mathbf{x}_i$ passed as covariates in our monotone BART fitting stage that do not appear in the DGP; these extraneous variables serve as "noise" to complicate the problem and make it more realistic. Table 5 demonstrates how in this setting our model performs much better than the bivariate probit. Additionally, we mis-specify $f(u)$ to see if we can still return true individual treatment effects, and, if we fail, what type of distributions cause problems. In Table 5, we mis-specify with Laplacian distributions, as the fatter tail weight could be problematic, and the table confirms this does appear to be an issue. Additionally, we compare our methodology with the bivariate probit model, fit with regression spline smoothing and without. Our methodology does comparatively much better in this setting, as the DGP is highly non-linear.

In Table 6, we generate $f(u)$ according to the shark fin but with σ varied to attain certain variances. The choice of q affects the skewness of the distribution. The shark fin provides us insight into whether or not skewness or large variances affect our models estimates; as the previous table showed mean offsets do not seem to impact our estimates too badly. In Table 7, we see getting q wrong (skewness) seems less impactful, meanwhile downwardly estimating variance seems to bias the estimates of the average causal risk ratio (ACRR) up, while guessing variance too high downwardly biases the average causal risk ratio. Table 8 investigates more drastically mis-specifying q or σ.

$f(u)$	**true** ACRR	true est. ACRR	RMSE	Wrong $f(u)$	Wrong est.	Wrong RMSE	LBP est.	LBP RMSE	SBP est.	SBP RMSE
$\mathrm{N}(0,1)$	4.43	4.71	1.66	Lap(0,1.2)	2.02	3.09	4.19	1.98	4.46	2.00
$\mathrm{N}(0,1.5)$	2.80	2.81	0.70	Lap(0, 1.75)	1.68	1.36	3.22	0.85	3.38	0.94
$\mathrm{N}(0,2)$	2.14	2.11	0.36	Lap(0, 2.5)	1.42	0.83	0.44	1.81	0.37	0.73
$\mathrm{N}(0,2.5)$	1.81	1.80	0.25	Lap(0, 2)	2.04	0.37	1.81	0.46	0.37	1.46
$\mathrm{N}(-1,1)$	8.18	9.38	5.07	Lap$(-1, 1.3)$	1.53	7.83	2.96	6.51	2.83	6.61
$\mathrm{N}(1,2)$	1.74	1.45	0.34	Lap(1, 2.4)	1.20	0.57	1.89	0.30	0.62	1.15
$\mathrm{N}(-2,2)$	3.43	5.88	4.32	Lap$(-2, 2.3)$	1.02	2.49	2.77	0.92	3.03	0.76
$\mathrm{N}(2,1)$	1.68	1.62	0.23	Lap(2, 1.3)	1.18	1.39	0.61	0.45	1.75	1.42

TABLE 5

Different $f(u)$ as described in DGP(28). $N = 25,000$. Wrong $f(u)$ indicates the distribution of U we used to solve the system of equations in 13, i.e. how we mis-specified. True indicates true average causal risk ratio (ACRR), and correct est. indicates our estimate of the ACRR when correctly *specifying $f(u)$. We use standard deviation instead of variance for our spread parameter. Lap refers to the Laplacian distribution. LBP refers to bivariate probit regression without smoothing, and SBP refers to bivariate probit regression with smoothing covariates, i.e. where the smooth term for each covariate is made of basis functions.*

$f(u)$ sharkfin with parameters q, s	**true** ACRR	true est. ACRR	true RMSE	wrong q	wrong q est.	wrong q RMSE
(0.25, 0.82; 3)	1.79	1.81	0.21	(0.40,1.37;3.00)	1.80	0.23
(0.40, 1.37; 3)	2.07	2.08	0.31	(0.70,2.34;3.00)	2.55	0.94
(0.60, 1.06; 3)	3.10	2.97	0.80	(0.30,1.00;3.00)	1.97	1.43
(0.75, 2.46; 3)	5.86	5.37	2.59	(0.92,2.77;3.00)	8.41	5.13
(0.25, 0.34; 0.5)	4.11	4.27	1.54	(0.10,0.12;0.50)	4.13	1.44
(0.40, 0.56; 0.5)	5.28	5.95	2.71	(0.20,0.26;0.50)	5.25	2.01
(0.60, 0.84; 0.5)	8.63	8.80	5.31	(0.80,1.05;0.50)	10.9	7.60
(0.75, 1.00; 0.5)	13.4	12.3	8.25	(0.45,1.63;0.50)	7.74	10.6

TABLE 6

Different $f(u)$ as described in DGP(28), all of the "sharkfin" family. ACRR = average causal risk ratio. $N = 25,000$. Wrong q indicates that we purposely mis-specified q when solving our system of equations, whereas the true est.columns indicate where we correctly specified $f(u)$, both the q and s parameters, when solving our system. ; indicates the variance, whereas the first two entries in shark are the q and s parameters. Here we vary the skewness while keeping variance constant.

$f(u)$ sharkfin with parameters q, s	**true** ACRR	true est. ACRR	true RMSE	wrong σ^2	wrong σ^2 est.	wrong σ^2 RMSE
(0.25, 0.82; 3)	1.79	1.76	0.38	(0.25,0.47;1.0)	3.55	2.12
(0.40, 1.37; 3)	2.07	2.09	0.61	(0.40,1.12;2.0)	2.76	0.90
(0.60, 1.06; 3)	3.10	3.27	1.54	(0.60,0.92;0.6)	11.3	10.4
(0.75, 2.46; 3)	5.86	7.40	7.94	(0.75,1.74;1.5)	9.79	6.20
(0.25, 0.34; 0.5)	4.11	5.34	6.25	(0.25,0.67;2.0)	1.56	3.16
(0.40, 0.56; 0.5)	5.28	8.91	10.7	(0.40,1.12;2.0)	1.83	4.28
(0.60, 0.84; 0.5)	8.63	9.65	22.5	(0.60,2.38;4.0)	1.40	9.29
(0.75, 1.00; 0.5)	13.4	16.6	57.2	(0.75,3.18;5.0)	1.90	15.5

TABLE 7

Different $f(u)$ as described in DGP(28), all of the "sharkfin" family. $N = 25,000$. Wrong σ^2 indicates that we purposely mis-specified our variance (by varying the σ parameter) when solving our system of equations, whereas the true est. columns indicate where we correctly specified $f(u)$, both the q and a parameters, when solving our system. ; indicates the variance, whereas the first two entries in shark are the q and s parameters. Here we vary the variance keeping skewness constant.

True $f(u)$	true ACRRT	ACRRT est.	ACRRC true	ACRRC est.	Wrong q $f(u)$	ACRRT est. wrong	ACRRC est. wrong
shark(0.1, 0.30; 3)	1.60	1.62	1.67	1.69	shark(0.9, 2.74;3)	1.36	1.60
shark(0.1, 0.12; 0.5)	3.18	0.40	3.79	3.75	shark(0.9, 1.12; 0.5)	3.82	5.19
shark(0.1, 0.18; 1)	2.32	2.39	2.58	2.69	shark(0.9, 1.58; 1)	2.75	3.72
shark(0.1, 0.18; 1)	2.32	2.39	2.58	2.69	shark(0.5, 1; 1)	2.43	2.97
shark(0.5, 1; 1)	4.00	4.18	4.66	5.18	shark(0.1, 0.18; 1)	3.16	3.47
shark(0.5, 1; 1)	4.00	4.18	4.66	5.18	shark(0.9, 1.58; 1)	6.78	9.32
shark(0.9, 1.58; 1)	13.1	12.0	19.2	17.6	shark(0.1, 0.18; 1)	2.75	2.56
shark(0.9, 1.58; 1)	13.1	12.0	19.2	17.6	shark(0.5, 1; 1)	4.96	5.66

TABLE 8

Comparing estimates of average causal risk ratio on treated (ACRRT) and average causal risk ratio on controls (ACRRC) when we more aggressively mis-specify the q parameter, which controls the skewness.

6.3. *Relationship with E-values: Simulations.* Here, we replicate the analysis presented in Figure 6 with simulated data. Rather than using all the posterior draws given by the BART model in the simulated data setting, we instead take the mean of the posterior BART probability estimates in the fitting stage and then solving for our causal parameters $b_0(\cdot), b_1(\cdot), g(\cdot)$ once for each observation. We impose the constraint that $b_1(\mathbf{x}) \geq b_0(\mathbf{x})$ when solving for the

causal parameters. We do this for different distributions of f with the data generated in accordance with 28. In Figure 11, we compare our estimate of the inducement effect vs the E-value, for different distributions of U.

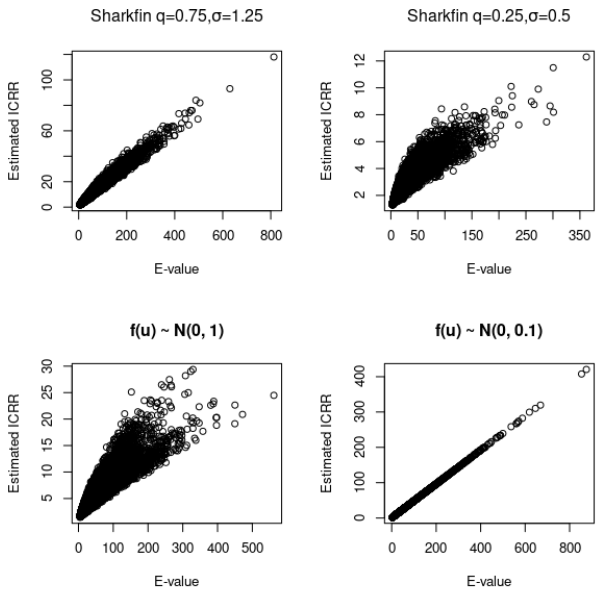


Fig 11: Comparison of our individual causal risk ratio estimates vs individual E-value estimates for 25,000 simulations drawn from the same dgp as specified in 28. Shown are different distributions of $f(u)$, with the bottom right "low u" setting mimicking the E-value as expected.

6.4. *Value of monotonicity.* The monotonicity constraint proved valuable, as Figure 12 shows the improvement we see in estimating the individual causal risk ratios (ICRR) by using monotone BART instead of BART. The reason for this is that even though BART estimates each potential outcome well, it does not estimate the coupling of the two potential outcomes[9]. Monotone BART helps remedy this by not allowing for going concerns to lower the probability of bankruptcy.

7. Discussion. Compared to the popular bivariate probit model, the machine learning sensitivity analysis introduced here is more flexible and hence, more credible in empirical analyses. This increased flexibility comes at the price of identification, but this should not be a barrier to empirical investigation: a thorough sensitivity analysis can still yield evidence and insight, especially when coupled with posterior subgroup analysis.

Specifically, we conclude that at least some firms appear to experience induced bankruptcies; the degree of private information would have to be extreme to rule this out entirely. Moreover, it appears that induced bankruptcies are more likely to occur for firms that have high levels of leverage and that have an S&P credit rating. These results are reassuring given

[9]For more on issues with fitting BART estimates in causal inference settings, see **?**.

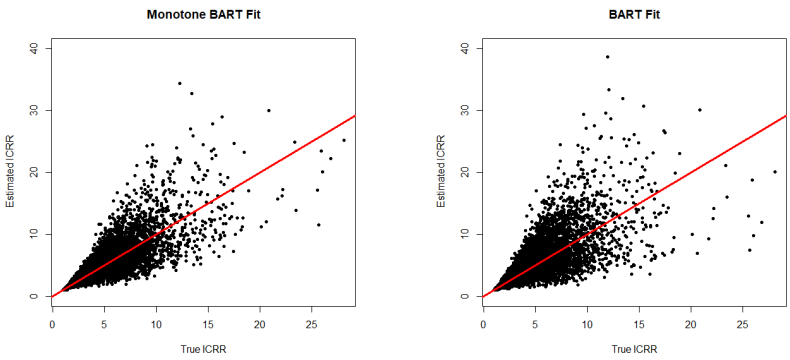


Fig 12: Plots of expected individual causal risk ratios vs our estimates, i.e. a plot comparing the ratio of potential outcomes from model 4 ($\Phi(\alpha_0 + \alpha_1 \mathbf{x}_i + \gamma)/\Phi(\alpha_0 + \alpha_1 \mathbf{x}_i)$) versus our estimate within the integral of equation(14). In the DGP, $\rho = 0.25, \gamma = 1$. The monotone BART correlation between the truth and estimate is 0.88 and BART is 0.83.

that adverse going concern opinions can mechanically lead to higher borrowing costs and credit rating downgrades. The fact that these moderating variables were uncovered by the model without explicit instruction lends credence to the inducement hypothesis.

Data analyses which mirror the "self-fulfilling prophecy" of the bankruptcy inducement problem have the potential to benefit from the modular machine learning sensitivity analysis developed here. For example, the question of whether catholic high schools lead to higher college enrollment [**?**] would be of particular interest, as that analysis employed the bivariate probit with endogenous regressor approach that we have generalized.

Acknowledgements. The authors would like to acknowledge support from NSF grant #1502640 .

APPENDIX A: 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[3] 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. 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 10.

γ	ρ	ARD true	ARD est	IRD cor	IRD 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 9

We simulate from the bivariate probit with 25,000 observations and deploy our methodology. cor refers to the correlation between predicted and true for the average risk difference (ARD), and the rmse is the root mean square error.

When fitting the bivariate probit regression, parameters are estimated using maximum likelihood. Our ARD from this regression was 2.90 percentage points and a mean risk ratio of 2.82 with 95% CI (1.33, 6.04). Note, we used a regression spline approach to smooth our numeric covariates, where the smooth term for each covariate is made of basis functions, see **?** for details. This gives us additional flexibility in fitting the model.

It was stressed in the main document how using BART with a monotonicity constraint improves our estimation of the ICRR, but the improvement is even more pronounced when studying risk differences. In figure 13, 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. This is a similar plot to figure 14 in the main file, but with a different estimand of interest.

Distribution of $f(u)$	ARD (%)	ARD post (%)	mean B_1 (%)	95% Credible interval for ARD (%)
$N(0, \sigma = 0.1)$	8.89	8.91	10.3	$(6.95, 11.2)$
$N(0, \sigma = 0.5)$	3.62	3.76	5.29	$(2.78, 4.97)$
$N(0, \sigma = 1)$	0.41	0.63	2.57	$(0.37, 0.97)$
Shark $q = 0.25$, $s = 0.5$; $\sigma = 1.05$	0.11	0.24	2.44	$(0.13, 0.40)$
Shark $q = 0.75$, $s = 1.25$; $\sigma = 0.88$	2.39	2.57	4.30	$(1.82, 3.52)$
"Right Bump" $\sigma = 0.48$	1.98	2.17	4.10	$(1.78, 2.69)$
98% peak $\sigma = 0.29$	6.85	7.08	8.62	$(5.41, 9.06)$
90% peak $\sigma = 0.64$	1.85	2.03	4.02	$(1.68, 2.51)$

TABLE 10

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.

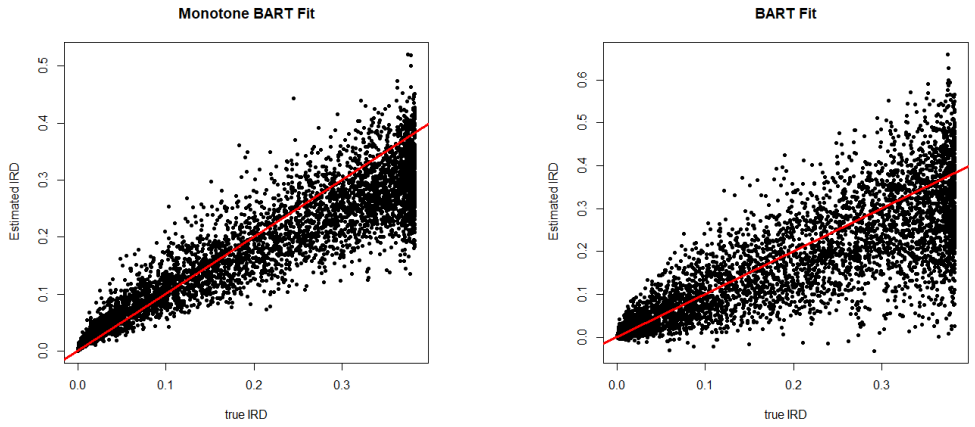


Fig 13: Plots of expected individual risk differences (IRD) vs our estimates, i.e. a plot comparing the difference in potential outcomes from the bivariate probit model $(\Phi(\alpha_0 + \alpha_1 \mathbf{x}_i + \gamma) - \Phi(\alpha_0 + \alpha_1 \mathbf{x}_i))$ versus our estimate. In the DGP, $\rho = 0.25, \gamma = 1$. The monotone BART correlation between τ and $\hat{\tau}$ is 0.928 and for BART is 0.826

APPENDIX B: BIVARIATE PROBIT SIMULATION STUDY

Table 11 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 ($N = 100,000$) 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 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.

Table 9 shows the results when we simulate 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 only $N = 25,000$ and $p = 5$.

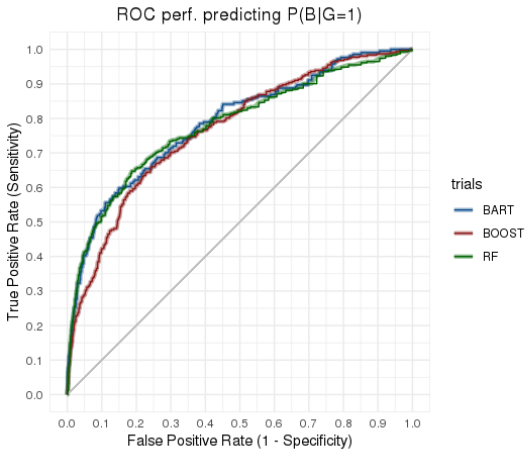
ARD true	ARD est	IRD cor	IRD RMSE	ACRR true	ACRR est	ICRR cor	ICRR rmse	γ true	γ est.	ρ	ρ est.
0.23	0.24	0.97	0.02	2.24	2.07	1.00	0.24	1.00	0.77	0.25	0.37
0.46	0.46	0.99	0.02	4.32	3.89	1.00	0.87	1.75	1.62	0.25	0.31
0.58	0.57	1.00	0.02	6.24	5.40	0.99	2.34	2.50	2.38	0.25	0.30
0.26	0.26	0.99	0.01	2.42	2.27	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 11

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. ARD refers to average risk difference, IRD individual risk difference. ACRR refers to average causal risk ratio, whereas ICRR is individual causal risk ratio.

APPENDIX C: 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 BART (described earlier), random forests, and XGBoost. Referencing figure 14 actually tells two different tales. According to the table, all the methods perform relatively similarly by the auc metric, but the roc curve shows the methods provide different probability estimates.



	XGBoost	RanFor	BART
Case	AUC	AUC	AUC
B_1, G_0	0.76	0.77	0.78
B_1, G_1	0.79	0.79	0.79
G_1	0.88	0.89	0.89

Fig 14: 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.

REFERENCES