Do Forecasts of Bankruptcies Cause Bankruptcies?

11/10/2022

Demetrios Papakostas

Arizona State University

Fall 2022

Do Forecasts of Bankruptcy Cause Bankruptcy?

The Annals of Applied Statistics
2023, Vol. 17, No. 1, 711–739
https://doi.org/10.1214/22-AOAS1648

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

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

[1]*School of Mathematical and Statistical Sciences, Arizona State University,* [a]*dpapakos@asu.edu,* [b]*prhahn@asu.edu*
[2]*Department of Information, Risk and Operations Management, The University of Texas at Austin,*
[c]*jared.murray@mccombs.utexas.edu*
[3]*Accounting Department, 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*

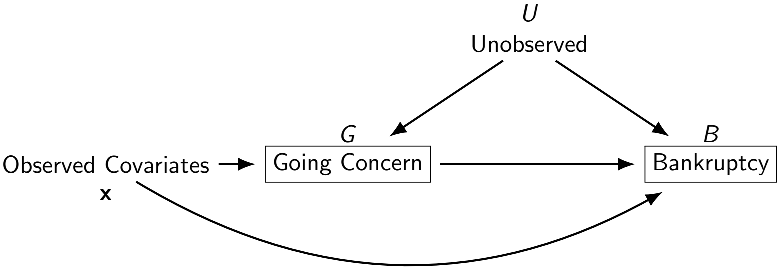
Forecasts…Forecause?

Introduction

What is the problem?

- Auditors' have "public data" and "hidden info" and use these to determine whether a company is at risk for bankruptcy.
- Does issuing going concern opinion for a company cause their bankruptcy, via a self-fulfilling prophecy mechanism?
- What do auditors know that we (researchers, "public") do not?

Schematically



Terminology

Potential Outcomes

- B_i^1 and B_i^0 are the potential outcomes. "Counter-factual" scenarios

  $B_i^1 \rightarrow$ Outcome of company i were they to receive going concern

  $B_i^0 \rightarrow$ Outcome of company i were they not to get going concern

- Fundamental problem: Can only **observe** one or other of each potential outcome for each company.

What should we do?

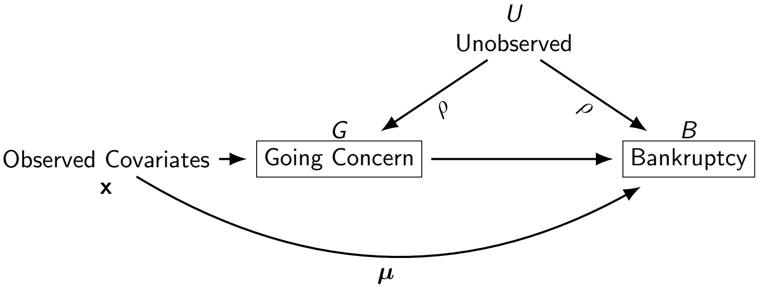
- Common tools, like instrumental variable analysis, didn't fit.
- E-Values: Quantify how much unexplained confounding we need to explain away a causal effect [PV16].
- Bivariate Probit Regression: A nice model that gives us an identified causal effect, but has its fair share of problems. We propose a method that "mixes" these.

Bivariate probit with endogenous regressor

Model introduction

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

- ρ can be thought of as arrow from confounding to outcome, $\boldsymbol{x}_i$ represents the controlled data.



Bivariate probit with endogenous regressor

Model formulation

- Define treatment, G, and outcome, B, as:

$$G = \mathbb{1}\{Z_{g,i} \geq 0\} \tag{2}$$
$$B = \mathbb{1}\{Z_{b,i} \geq -\gamma G\} \tag{3}$$

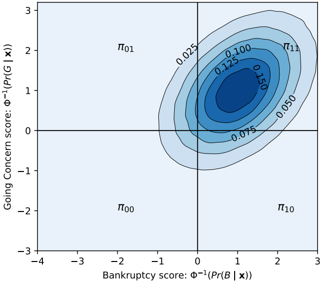
- The factor γ is introduced as a way to measure the average risk difference or the risk ratio.

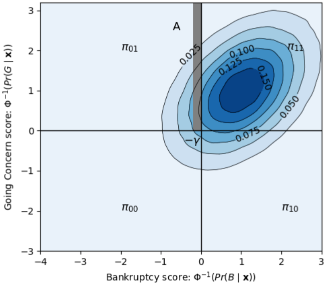
$$\Delta_i = \Phi(\gamma + \alpha_0 + \alpha_1 \boldsymbol{x}_i) - \Phi(\alpha_0 + \alpha_1 \boldsymbol{x}_i)$$
$$\tau_i = \Phi(\gamma + \alpha_0 + \alpha_1 \boldsymbol{x}_i)/\Phi(\alpha_0 + \alpha_1 \boldsymbol{x}_i)$$

Bivariate probit with endogenous regressor

Visual representation of identification

- Let $\Pi = \begin{pmatrix} \pi_{01} & \pi_{11} \\ \pi_{00} & \pi_{10} \end{pmatrix}$ where $\pi_{ij} = \Pr(B = i, G = j)$.





- $\boldsymbol{\mu}$ dictates the location, ρ (confounding) gives the tilt and concentration.

A relaxed bivariate probit

Model inspired from the bivariate probit

- From law of total probability:

$$\Pr(B, G|\boldsymbol{x}) = \int_{\mathbb{R}} \Pr(B = 1|\boldsymbol{x}, U = u, G)\Pr(G|\boldsymbol{x}, U = u)f(u)\,\mathrm{d}u \quad (4)$$

- This configuration allows us to account for hidden variable U.

Define model: Specify U

Main assumptions

- We stipulate:

$$\begin{aligned} \Pr\left(B = 1 | \mathbf{x}, U = u, G = 1\right) &= \Phi\left(b_1(\mathbf{x}) + u\right) \\ \Pr\left(B = 1 | \mathbf{x}, U = u, G = 0\right) &= \Phi\left(b_0(\mathbf{x}) + u\right) \qquad (5) \\ \Pr\left(G = 1 | \mathbf{x}, U = u\right) &= \Phi\left(g(\mathbf{x}) + u\right) \end{aligned}$$

- We solve for $b_1(\cdot), b_0(\cdot), g(\cdot)$, which are not restricted to be linear. Since Φ is a standard normal cdf

$$\begin{aligned} u \to -\infty &\Longrightarrow \Pr(B) \to 0 \\ u \to \infty &\Longrightarrow \Pr(B) \to 1 \\ u \to 0 &\Longrightarrow \text{hidden info plays no role} \end{aligned}$$

System of equations

Setting up sensitivity analysis

$$
\begin{aligned}
\Pr(B=1, G=1|\mathbf{x}) &= \int_{\mathbb{R}} \Phi(g(\mathbf{x})+u)\Phi(b_1(\mathbf{x})+u)f(u)\mathrm{d}u \\
\Pr(B=1, G=0|\mathbf{x}) &= \int_{\mathbb{R}} (1-\Phi(g(\mathbf{x})+u))\Phi(b_0(\mathbf{x})+u)f(u)\mathrm{d}u \\
\underbrace{\Pr(B=0, G=1|\mathbf{x})}_{\text{reduced form}} &= \int_{\mathbb{R}} \Phi(g(\mathbf{x})+u)(1-\Phi(b_1(\mathbf{x})+u))f(u)\mathrm{d}u
\end{aligned}
\tag{6}
$$

- **Specify** $f(u)$, then estimate the functions $b_1(\cdot), b_0(\cdot), g(\cdot)$ by minimizing distance between lhs and rhs.

(Non-linear) Optimization

- To find $b_1(\cdot), b_0(\cdot), g(\cdot)$, we minimize distance:

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

- Sum over all 3 "pairs" from 6, minimize this value. This gives us our estimates of $b_1(\cdot), b_0(\cdot), g(\cdot)$.

Inducement effect

Use ratios to match E-value approach

- Define **inducement** across individuals as:

$$\tau(\mathbf{x}) \equiv \int_{\mathbb{R}} \Pr(B=1|\mathbf{x}, G=1, u) f(u) \mathrm{d}u) \,/\, \int_{\mathbb{R}} \Pr(B=1|\mathbf{x}, G=0, u)\, f(u) \mathrm{d}u$$

In our framework, for each firm we have:

$$\tau(\mathbf{x}_i) = \int_{\mathbb{R}} \Phi(b_1(\mathbf{x}_i)+u) f(u) \mathrm{d}u / \int_{\mathbb{R}} \Phi(b_0(\mathbf{x}_i)+u) f(u) \mathrm{d}u \qquad (7)$$

- The average inducement is: $\frac{1}{n}\sum_{i=1}^{n} \tau(\mathbf{x}_i)$.

Connection to the bivariate probit

Model reformulated

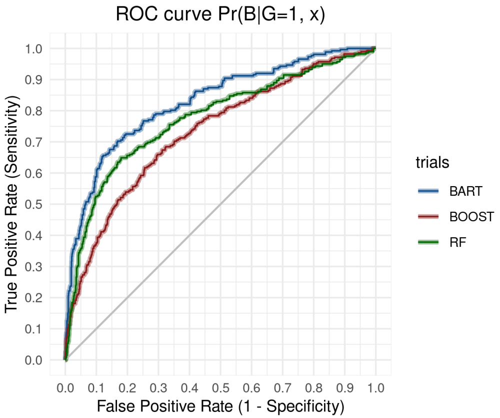
- This model is actually a reformulation of the bivariate probit model, with the distribution of U corresponding to specific covariance matrix
- If we generate data from the bivariate probit and deploy our methodology with $U \sim N\left(0, \sigma^2 = \frac{\rho}{1-\rho}\right)$, we should return the true estimates of the inducement (or risk difference) effects.
- We relax the conditions on the marginal distribution of u, drop the linear requirement of $b_1(\cdot), b_0(\cdot), g(\cdot)$, and allow the impact of the inducement to be non-additive.

How to fit Reduced Form

- The reduced form equations can be fit using observed data
- Can choose any method, but would like something flexible
- Our data is quite imbalanced, making this a hard classification problem

Different Methods

Balanced 5-fold CV



Note, we do not use cross-validation in our analysis. This just shows BART (with monotonicity modification) performs well in out of sample prediction compared to peers

Monotone BART details

Going concern issued cannot lower probability of bankruptcy

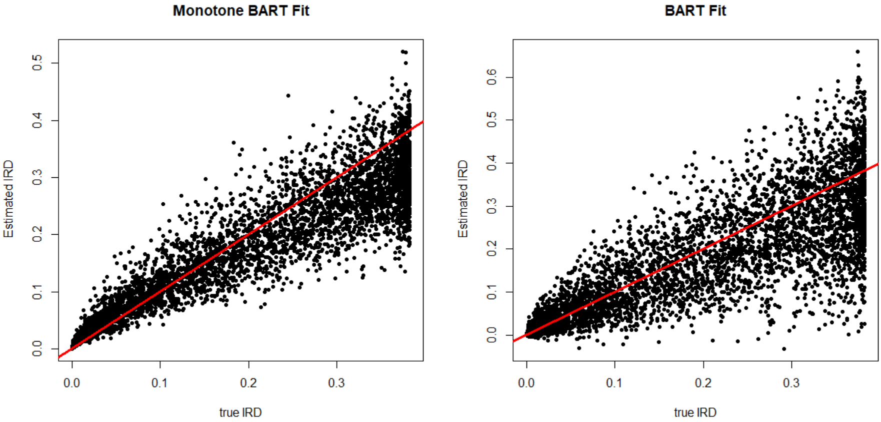
- Want $\Pr(B|G=1,\boldsymbol{x}) \geq \Pr(B|G=0,\boldsymbol{x})$, so we parameterize $\Pr(B=1 \mid G, \mathbf{x})$ as follows:

$$\begin{aligned} \Pr(B=1 \mid G=1,\mathbf{x}) &= \Phi[h_1(\mathbf{x})], \\ \Pr(B=1 \mid G=0,\mathbf{x}) &= \Phi[h_0(\mathbf{x})]\Pr(B=1 \mid G=1,\mathbf{x}), \\ &= \Phi[h_0(\mathbf{x})]\Phi[h_1(\mathbf{x})], \\ \Pr(G=1 \mid \mathbf{x}) &= \Phi[w(\mathbf{x})]. \end{aligned} \tag{8}$$

- Independent BART ([CGM10]) priors on each. Use data-augmented representation that permits updating h_0 and h_1 independently using standard MCMC for probit BART.

Comparing BART vs. monotone BART

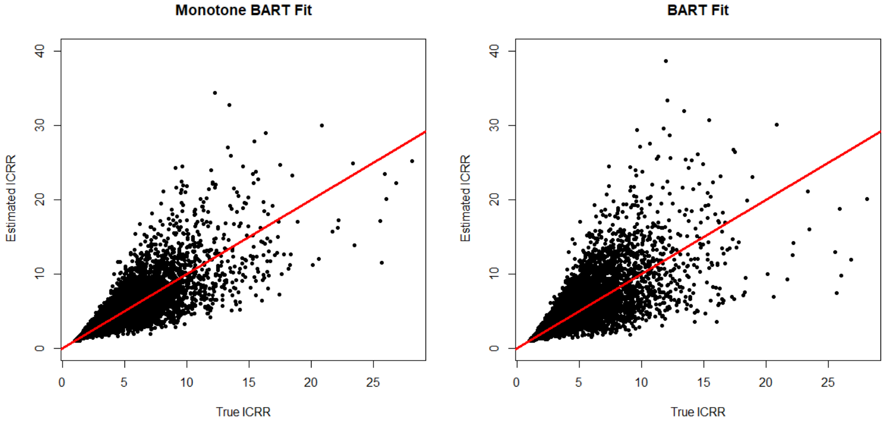
How we estimate observed probabilities



- BART predicts struggles with difference estimand. Monotone BART helps.

Comparing BART vs. monotone BART

How we estimate observed probabilities

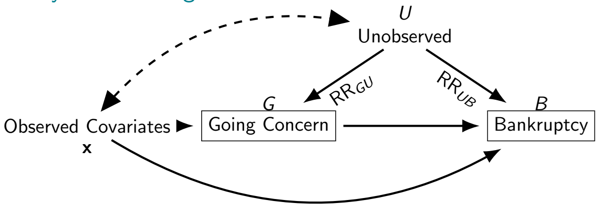


- The disparity is not as drastic when looking at ratios (simulated from bivariate probit)

Comparison with E-values

E-Values

A bound on necessary confounding



$$\mathrm{RR}_{GU|\boldsymbol{x}} = \max_k \frac{\Pr(U = k \mid G = 1, \boldsymbol{x})}{\Pr(U = k \mid G = 0, \boldsymbol{x})},$$

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

- RR_{UB} is maximum risk ratio for outcome comparing any two categories of confounding.
- RR_{GU} is the maximum risk ratio for any specific level of the U comparing those w/& w/o treatment.

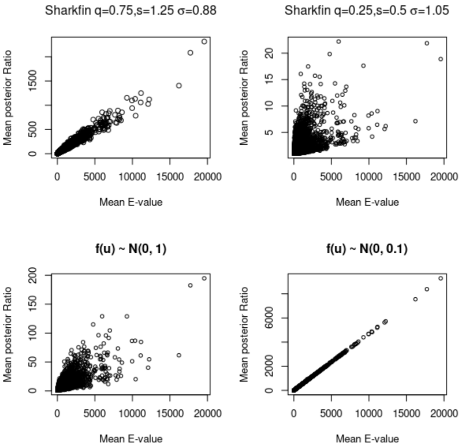
E-Value

- [PV16] define the *E-value* (for evidence value) as

$$\text{E-value} = \mathrm{RR}_{GB}^{\mathrm{obs}} + \sqrt{\mathrm{RR}_{GB}^{\mathrm{obs}}\left(\mathrm{RR}_{GB}^{\mathrm{obs}} - 1\right)}, \tag{9}$$

- Minimum strength of association that an unmeasured confounder would need to have with both the G and B (conditional on $\boldsymbol{x}$) to fully explain the observed treatment-outcome association.

Comparison with E-value



Further Details

- Code Repository: Code to reproduce the tables/figures in the paper
- Monbart: An R-package to implement the monotone BART (written by Jared Murray)
- SFPSA: An R-package to implement the methods presented

Other Applications

- Really any binary treatment/binary outcome[1].
- Does being labeled an underdog in sports *cause* teams to win more often than they should??
- Highly heterogeneous with likely lots of moderation occurring

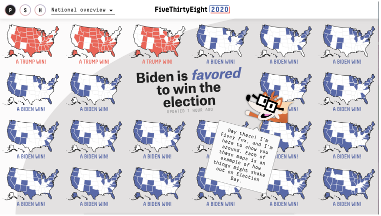
Run Line Performance

OVERALL	OVERALL
55-36	**46-43**
UNDERDOG	FAVORITE
27-8	**29-50**

[1] Could modify system for categorical outcomes

Other Applications

- Do polls (or even election models) have a causal effect on winner of elections?
- A way to model the endogeneity present in those models after the fact.



Conclusions

- We generalize the bivariate probit model, giving a flexible framework for binary treatment, binary outcome scenario
- We show our method is analogous, but more useful, then the "E-value"
- Apply our method to a cool dataset

Bibliography I

[CGM10] H.A. Chipman, E.I. George, and R.E. McCulloch, *Bart: Bayesian additive regression trees*, The Annals of Applied Statistics **4** (2010), no. 1, 266–298.

[PV16] D. Peng and T.J. VanderWeele, *Sensitivity analysis without assumptions*, Epidemiology (Cambridge, Mass) **27** (2016), no. 3, 368–377.

Application Results

- `Log(Assets)`: Natural log of total assets
- `Leverage`: Ratio of total liabilities to total assets
- `Investment`: Ratio of short-term investments to total assets
- `Cash`: Ratio of cash and cash equivalents to total assets
- `ROA`: Ratio of income before extraordinary items to total assets
- `Log(Price)`: Natural log of stock price
- `Intangible assets`: Ratio of intangible assets to total assets
- `R&D`: Ratio of research and development expenditures to sales
- `R&D missing`: Indicator for missing R&D expenditures
- `No S&P rating`: Indicator for the existence of a S&P credit rating
- `Rating below CCC+`: Indicator for S&P credit rating below CCC+
- `Rating downgrade`: Indicator for an S&P credit rating downgrade from above CCC+ to CCC+ or below

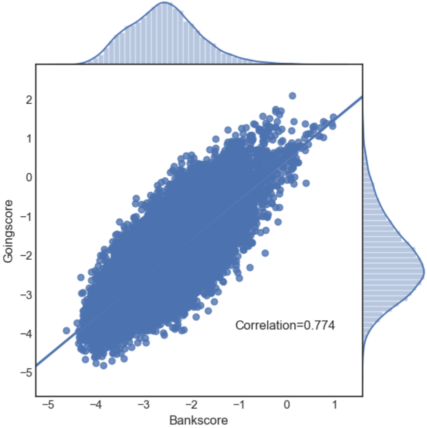
List of Covariates Part II

- `Non-audit fees`: Ratio of non-audit fees to total audit fees
- `Non-audit fees missing`: Indicator for missing non-audit fees
- `Years client`: Number of years of client used auditor
- `Average short interest`: Interest expense/total assets
- `Short interest ratio`: Average short interest (measured in number of shares)/total shares outstanding three months prior to the auditor signature date
- `Sum of log returns`: The sum of log daily return in year t
- `Return Volatility`: The standard deviation of daily returns in year t
- `Time fixed effect`: A dummy variable for the years 2000–2014

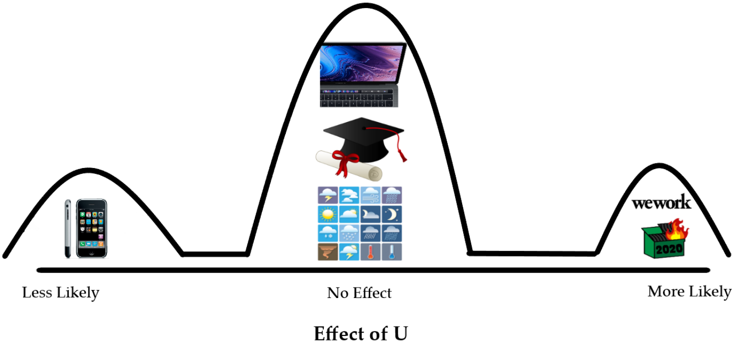
Going Concerns vs. Bankruptcy Graphically

Visualization of our Data

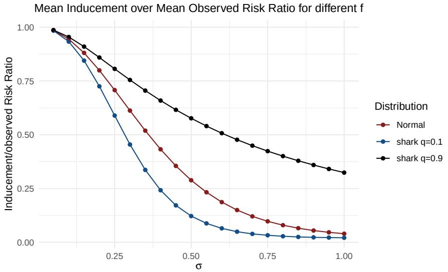
Graph of the z-scores for the going concern and bankruptcy (fit with same covariates)

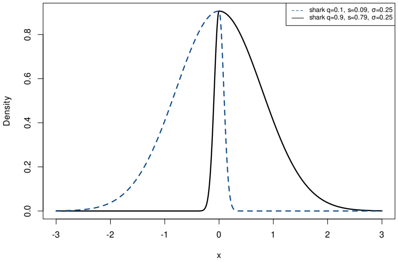


How to choose distribution of U



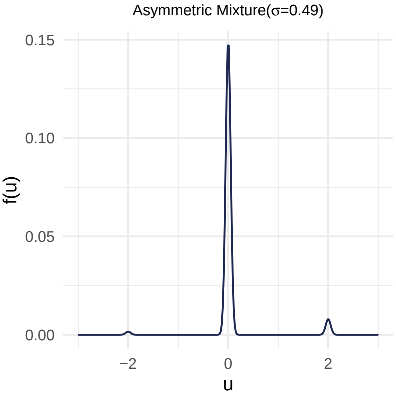
How much does $f(u)$ matter?

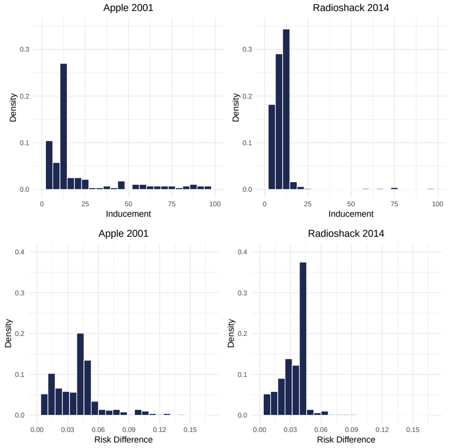




- Left: Plot of inducement effect over observed risk ratio. The mean observed risk ratio was 30.80. On right is a plot of the shark fin with $q = 0.1$ and $q = 0.9$, for visual purposes.

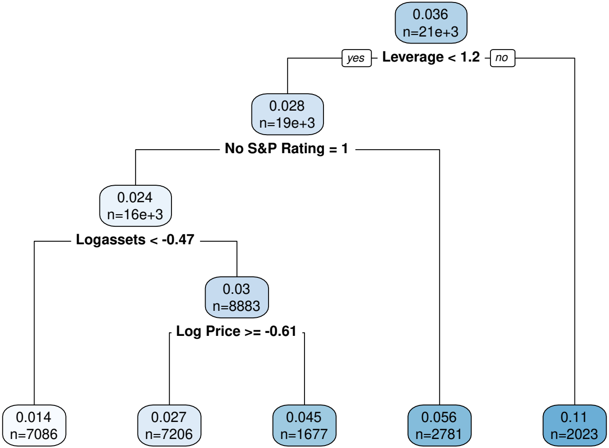
Individual Firms





A closer look

Moderation of Risk Differences



Empirical Results

Distribution of $f(u)$	inducement posterior mean	95% CI for mean inducement	RD Posterior Mean	95 % CI for mean RD
$\mathrm{N}(0, \sigma = 0.1)$	111	$(39.6, 279)$	0.100	$(0.071, 0.129)$
$\mathrm{N}(0, \sigma = 0.5)$	33.9	$(11.8, 91.7)$	0.041	$(0.027, 0.056)$
$\mathrm{N}(0, \sigma = 1)$	4.08	$(1.82, 9.79)$	0.007	$(0.004, 0.011)$
Shark $q = 0.25$, $s = 0.5$ ($\sigma = 1.05$)	1.51	$(1.14, 2.46)$	0.003	$(0.001, 0.004)$
Shark $q = 0.75$, $s = 1.25$ ($\sigma = 0.88$)	27.8	$(9.69, 74.9)$	0.028	$(0.018, 0.040)$
Symmetric mixture ($\sigma = 0.64$)	24.4	$(8.09, 64.2)$	0.025	$(0.018, 0.029)$
Asymmetric mixture ($\sigma = 0.49$)	25.6	$(8.40, 72.0)$	0.023	$(0.019, 0.031)$

RD is risk difference, CI is credible interval, Further restrain $b_1(\mathbf{x}) > b_0(\mathbf{x})$ in optimization step

Thank you