Internet appendix

Balancing external vs. internal validity: An application of causal forest in finance

Huseyin Gulen Candace E. Jens T. Beau Page

Internet Appendix

Table of Contents

A Summary of related literature

Table IA.1. Summary of topics of study and key empirical differences between papers in the literature following Chava and Roberts (2008) in treating bond covenant default as an exogenous event. There are two definitions of violation used in the literature. "Any" means any quarter or year in which a firm is in technical default is deemed a violation. "New" means the authors focus on technical defaults that follow at least four quarters without a technical default. The bandwidth column indicates whether or not the authors tested within a smaller sample around the default threshold. Measure is the distance to default measure used to define the threshold given in the paper, when available. McCrary indicates whether the paper included some formal test of bunching or manipulation around the threshold. The measure and McCrary columns are N/A for papers in the Roberts and Sufi (2009) setting where a distance to default cannot be calculated.

No.	topic	paper	violation	frequency	bandwidth	measure	McCrary
Panel A: Summary of literature studying causal effect of bond covenant violations on...							
1.	CEO compensation contracts	Akins, Bitting, DeAngelis, and Gaulin (2019)	any	annual	yes	unscaled distance to default	yes
2.	CEO compensation and package	Balsam, Gu, and Mao (2018)	any	annual	yes	Chava and Roberts (2008)	yes
3.	acquisitions	Becher et al. (2022)	new	annual	no	N/A	N/A
4.	audit fees; auditor actions	Bhaskar, Krishnan, and Yu (2016)	new	annual	no	N/A	N/A
5.	rival firms	Billett, Esmer, and Yu (2018)	new	quarterly	no	N/A	N/A
6.	value of lender relationships	Bird, Hertzel, Karolyi, and Ruchti (2023)	any	quarterly	yes	distance scaled by SD(8)*	no**
7.	dividend policy	Bulan and Hull (2013)	any	quarterly	yes	unscaled distance to default	no
8.	employee safety	Chatterjee, Hass, Hribar, and Kalogirou (2021)	any	quarterly	yes	unscaled distance to default	no
9.	R&D and innovation	Chava et al. (2017)	any	quarterly	yes	Chava and Roberts (2008)	no
10.	CDS spreads	Chen, Kim, and Zhu (2017)	any	quarterly	yes	distance scaled by SD(20)*	no
11.	non-GAAP reporting practices	Christensen, Pei, Pierce, and Tan (2019)	any	quarterly	no	N/A	N/A
12.	corporate tax avoidance	Cook, Ma, and Zhao (2020)	any	annual	no	N/A	N/A
13.	merger premium, CARs, deal type, payment type	Daher and Ismail (2018)	any	quarterly	no	Chava and Roberts (2008)	N/A
14.	employment; sales and closures of establishments	Ersahin, Irani, and Le (2021)	new	annual	no	N/A	N/A
15.	employment	Falato and Liang (2016)	any	annual	yes	distance scaled by threshold	yes
16.	independent directors	Ferreira, Ferreira, and Mariano (2018)	any	annual	yes	distance scaled by threshold	no
17.	tax avoidance	Francis, Shen, and Wu (2023)	new	annual	no	N/A	N/A
18.	future loan contracts	Freudenberg, Imbierowicz, Saunders, and Steffen (2017)	any	both	yes	difference scaled by SD(12)*	no
19.	information asymmetry	Gao, Khan, and Tan (2017)	new	quarterly	no	N/A	N/A
20.	innovation	Gu et al. (2017)	new	annual	yes	Chava and Roberts (2008)	yes
21.	CSR activities	He, Zhang, and Zhong (2021)	any	annual	no	N/A	N/A
22.	audit verification	Jiang and Zhou (2016)	any	annual	no	N/A	N/A
23.	banking relationships	Keil (2023)	any	quarterly	yes	unscaled distance to default	no
24.	monitoring; savings behavior	Lin, Xin, Zhang, and Zhang (2017)	new	annual	no	N/A	N/A
25.	investment (capital expenditures and acquisitions); leverage; payout; turnover	Nini et al. (2012)	new	quarterly	yes	unclear	no
26.	competitors' investment and financing	Nordlund (2018)	new	quarterly	no	N/A	N/A
27.	conservatism	Tan (2013)	any	quarterly	yes	Chava and Roberts (2008)	no
28.	disclosure	Vashishtha (2014)	any	quarterly	no	N/A	N/A
29.	trade credit	Zhang (2019)	any	quarterly	yes	Chava and Roberts (2008)	yes
Panel B: Summary of literature studying how ... affects causal effect of bond covenant violations on...							
30.	credit conditions; enforcement	Bird, Ertan, Karolyi, and Ruchti (2022a)	any	quarterly	yes	distance scaled by SD(8)*	no
31.	short-termism; enforcement	Bird, Ertan, Karolyi, and Ruchti (2022b)	any	quarterly	no	distance scaled by SD(8)*	no
32.	whether a firm has CDSs; investment and bankruptcy	Chakraborty, Chava, and Ganduri (2022)	any	quarterly	yes	unclear	no
33.	dual ownership; investment	Chava, Wang, and Zou (2019)	any	quarterly	no	N/A	N/A
34.	distracted loan officers; investment and default rates	Gao, Karolyi, and Pacelli (2018)	any	quarterly	no	N/A	N/A
Panel C: Other studies							
35.	violation shocks monitoring intensity	Colonnello, Koetter, and Stieglitz (2020)	new	quarterly	yes	distance scaled by threshold	yes

*Note: SD(n) = standard deviation of the accounting variable over the previous *n* quarters.
**Note: McCrary (2008) test is mentioned but no results are presented in the paper.

B Variable definitions

- Altman's Z-score: (3.3 × pre-tax income + sales + 1.4 × retained earnings + 1.2 × net working capital) / total assets + 0.6 × market value / market value of assets.
- Bond rating: a binary variable equal to one if a firm has a bond rating in a quarter.
- Capital / Assets: Net PPE / total assets.
- Cash / Assets: current cash / total assets.
- Cash flow: (income before extraordinary items + depreciation and amortization) / start-of-period PPE.
- Current ratio: current assets / current liabilities.
- Firm size: natural logarithm of total assets. Total assets are deflated to December 2000 by the all-urban CPI. CPI data are from the Federal Reserve Bank of St. Louis website (https://fred.stlouisfed.org/).
- Investment: capital expenditures / start-of-period PPE.
- Leverage: total debt / total assets.
- Macro Q: (total book debt + market equity − total inventories) / start-of-period PPE.
- Market-to-book: (market equity + total debt + preferred stock liquidation − deferred taxes and investment tax credits) / total assets.
- Net worth: total assets − total liabilities.
- ROA: operating assets before depreciation / total assets.
- Syndicate size: number of banks in a loan's syndicate.
- Tangible net worth: current assets + net PPE + other assets − total liabilities.

C Supplementary calculations for Monte Carlo experiments

C.1 Additional details on simulated data

For all Monte Carlo specifications, our simulated data have a second order polynomial as the functional form:

$$Y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025(x_1 - \bar{x_1})^2 - 0.01(x_2 - \bar{x_2})^2 + 0.015(x_3 - \bar{x_3})^2 + 0.02W + \varepsilon, \tag{IA.1}$$

in which Y is the dependent variable, the three x variables are additional covariates, $\bar{x}$ is the expectation of x (i.e., $E(x)$), W is a binary variable equal to one if a continuous forcing variable w is greater than zero, and ε is the mean zero error term with a standard deviation of 0.065. We simulate the x and w variables using a multivariate normal distribution. For the vector (x_1, x_2, x_3, w), we set the mean to $(1.40, 5.50, 0.10, 0.20)$, standard deviation to $(1.00, 1.50, 0.30, 0.30)$, and use the correlation matrix:

$$\left\{ \begin{array}{cccc} 1.00 & -0.05 & -0.30 & 0.15 \\ -0.05 & 1.00 & 0.20 & 0.35 \\ -0.30 & 0.20 & 1.00 & 0.50 \\ 0.15 & 0.35 & 0.50 & 1.00 \end{array} \right\}. \tag{IA.2}$$

Latent variable x_4 is joint-normally distributed with w, Y, and the observable covariates and has a mean of 0.3 and a standard deviation of 0.8.

C.2 Relation between Γ and p_{man}

Manipulation of treatment status effectively creates an omitted variable that induces a stronger relation between W and Y that is unrelated to the values of the observable covariates. In our baseline setting, the probability an observation is treated is the probability that w is greater than zero. Once manipulation is introduced, the probability an observation is treated depends on both the probability that w is greater than zero and the probability that manipulation occurs. Manipulation occurs if two conditions are met: without manipulation w would be below zero but above some point $\underline{w}$ below which manipulation is too costly to perform, and the firm is randomly chosen (according to p_{man}) to be allowed to manipulate. So,

$$P\left(\tilde{w} > 0 | \mathbf{x} = x, p_{man}\right) = P\left(w > 0 | \mathbf{x} = x\right) + p_{man} P\left(\underline{w} \leq w \leq 0 | \mathbf{x} = x\right), \tag{IA.3}$$

in which $\tilde{w}$ is the observed value of w after manipulation occurs. The true value of w is unobservable.

Because p_{man} has no effect on the observable xs, the expected value of w given $\mathbf{x}$ and the standard deviation are as given in Subsection 4.3. In this context, we can re-cast the above equation in terms of the standard normal CDF:

$$P\left(\tilde{w}>0|\mathbf{x}=x,p_{man}\right)=1-\left(1-p_{man}\right)\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)-p_{man}\Phi\left(\frac{\underline{w}-\overline{w}_x}{\sigma_{w|x}}\right). \tag{IA.4}$$

Eq. IA.4 shows that the treatment probability given $\mathbf{x}$ is weakly increasing in p_{man} because the density $P\left(\underline{w}\leq w\leq 0|\mathbf{x}=x\right)\geq 0$.

Using Eq. IA.4, we can also see how increasing p_{man} can impair overlap, as the probability of observing values between $\underline{w}$ and 0 approaches to 0 as p_{man} approaches 1. If we combine Eq. IA.4 with the equation for Γ from our main text (Eq. 11):

$$\Gamma\left(x,p_{man}\right)=\frac{\left[1-\left(1-p_{man}\right)\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)-p_{man}\Phi\left(\frac{\underline{w}-\overline{w}_x}{\sigma_{w|x}}\right)\right]/\left[\left(1-p_{man}\right)\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)+p_{man}\Phi\left(\frac{\underline{w}-\overline{w}_x}{\sigma_{w|x}}\right)\right]}{\left[1-\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)\right]/\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)}. \tag{IA.5}$$

Note that the odds ratio based on the observable values of $\mathbf{x}$ is unchanged from our base setting.

Finally, Γ as given in Eq. IA.5 is increasing in p_{man}. The partial derivative is:

$$\frac{\partial\Gamma\left(x,p_{man}\right)}{\partial p_{man}}=\frac{\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)-\Phi\left(\frac{\underline{w}-\overline{w}_x}{\sigma_{w|x}}\right)}{\frac{1-\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)}{\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)}\times\left[\left(1-p_{man}\right)\Phi\left(\frac{-\overline{w}_x}{\sigma_{w|x}}\right)+p_{man}\Phi\left(\frac{\underline{w}-\overline{w}_x}{\sigma_{w|x}}\right)\right]^2}>0. \tag{IA.6}$$

Figure IA.1. **Additional Monte Carlo results for baseline specification, varying sample sizes.** Auxiliary results for Figure 3. See the caption of Figure 3 for specification details.

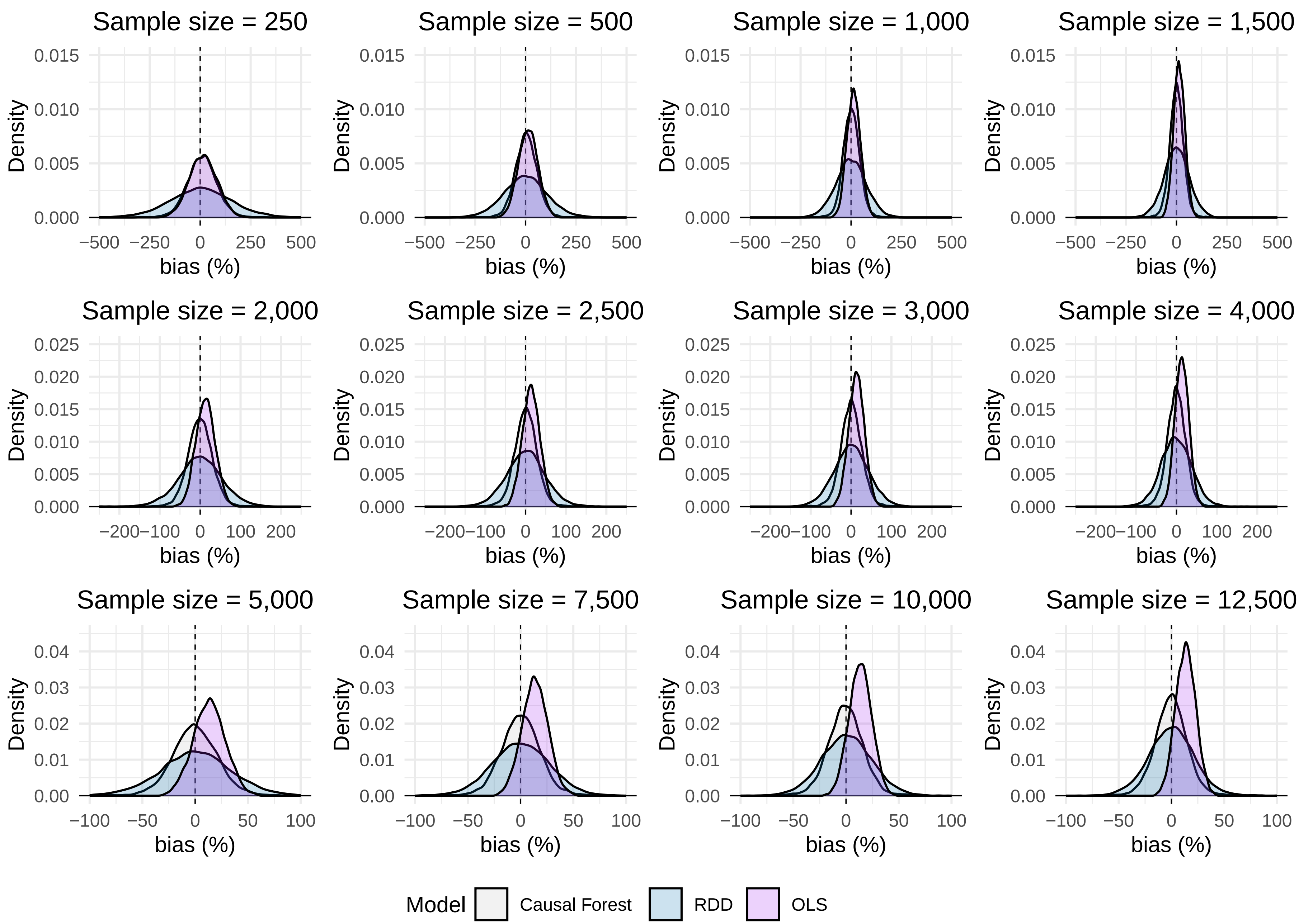

Figure IA.2. **Additional Monte Carlo results for latent variable equal to interaction of already-included covariates.** Each panel shows sensitivities of bias—the difference between the estimated and true treatment effect, scaled by the treatment effect, and scaled by 100—for 10,000 sets of simulated data with a sample size of 10,000, omitting the twenty most extreme estimations for each estimator. We augment the specification given in the caption of Figure 3 with an interaction of x_2 and x_3. This interaction is a latent variable and so is used in the data generating process but omitted from the causal forest, OLS, and RDD estimations. The interaction strength given in the header of each panel indicates the proportion of outstanding variation in the outcome variable that the interaction explains. Interaction strengths equal to $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0\}$ correspond with slopes on the interaction term equal to $\{0.0149, 0.0298, 0.0447, 0.0596, 0.0745, 0.0894, 0.1043, 0.1192, 0.1341, 0.1490\}$.

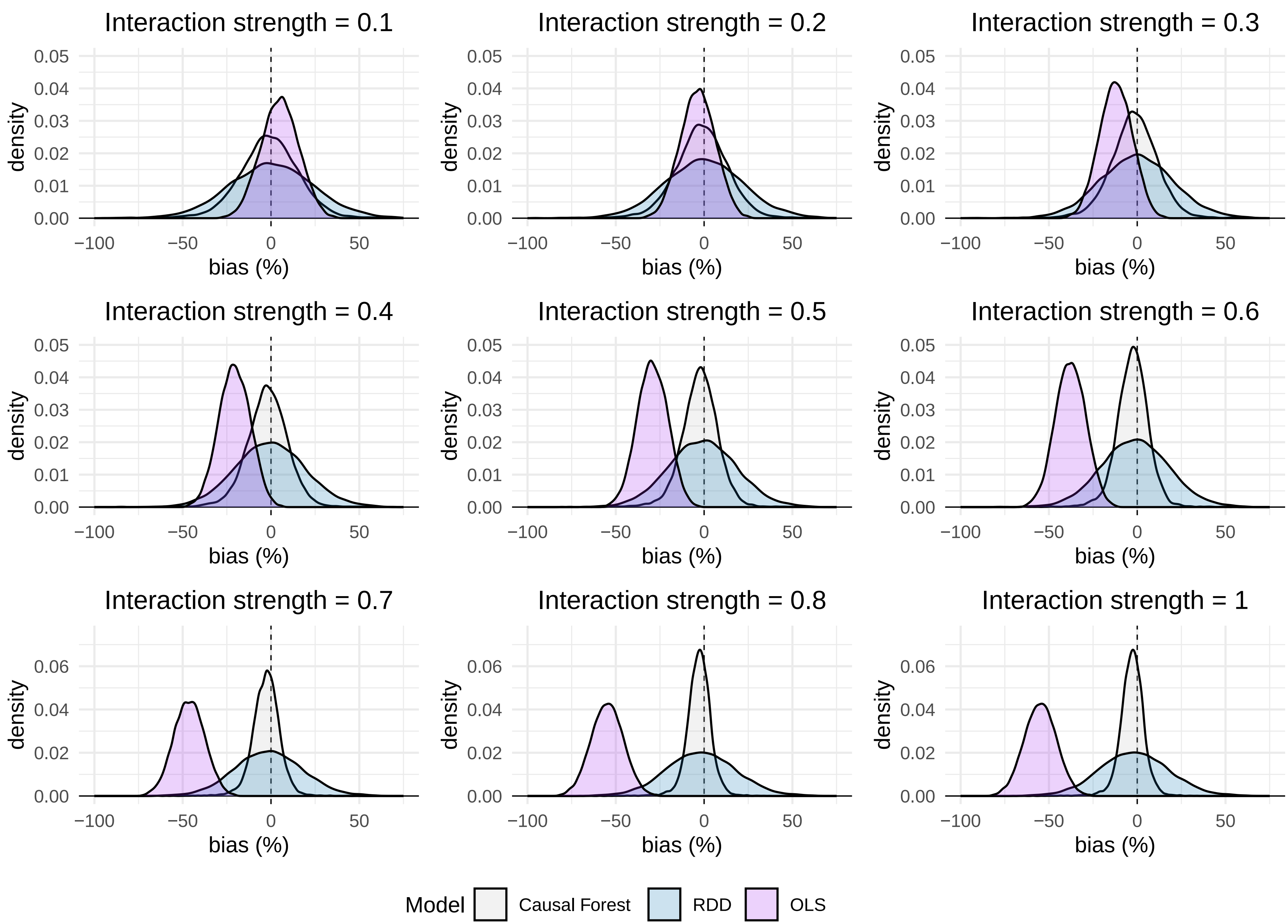

Figure IA.3. **Additional Monte Carlo results for alternate functional form.** Each panel shows sensitivities of bias—the difference between the estimated and true treatment effect, scaled by the treatment effect, and scaled by 100—for 10,000 sets of simulated data with a sample size of 10,000, omitting the twenty most extreme estimations for each estimator. The alternative functional form specification is: $Y = 0.05\,(x1 < \overline{x_1} - \sigma_{x_1}) - 0.005\,(x1 < \overline{x_1} + 0.5\sigma_{x_1}) + 0.01\,(x2 < \overline{x_2}) + 0.02\,(x2 < \overline{x_2} + 0.5\sigma_{x_2}) + 0.025\,(x2 < \overline{x_2} + \sigma_{x_2}) - 0.01\,(x2 \geq \overline{x_2} + \sigma_{x_2}) + 0.015x_3 + 0.02W + \varepsilon$ where Y is the dependent variable, the three x variables are additional covariates (with $\overline{x}$ indicating the mean of the variable and σ_x its standard deviation), inequalities in parentheses being indicator functions that are equal to one when the inequality is true, W is a binary variable equal to one if the forcing variable w is greater than zero, and ε is the mean zero error term.

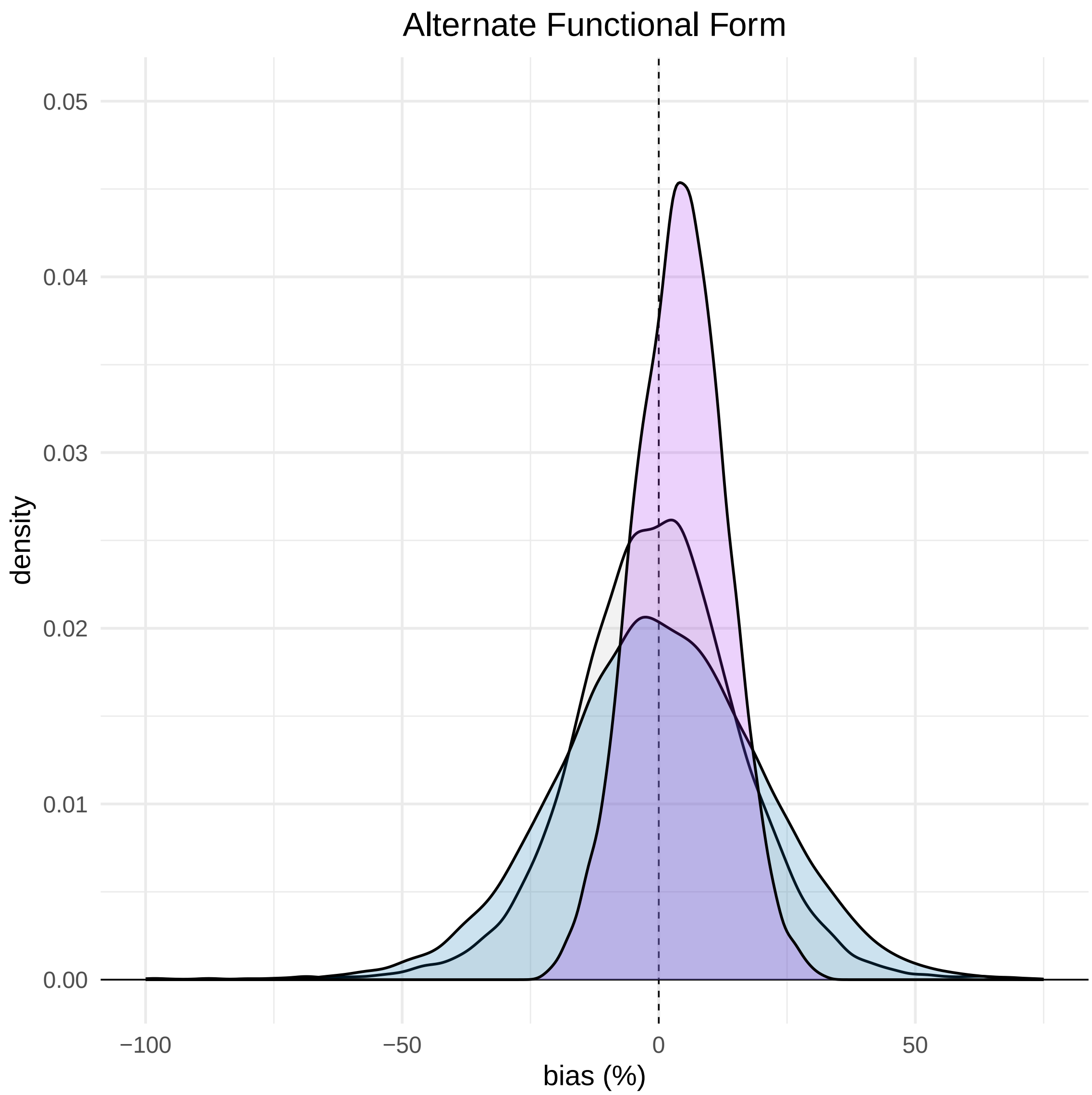

Table IA.2. Comparison of bias and test coverage for causal forest, OLS, and RDD for baseline specification with differing sample sizes. Each panel shows mean bias (as a percent of the true average treatment effect scaled by 100), root mean squared error (also as a percent of the true effect scaled by 100), and coverage (the percent of Monte Carlo simulations for which a 5% t-test of the difference between the estimate and true effect rejects the null hypothesis) for 10,000 sets of simulated data with the sample size given in the panel title, omitting the twenty most extreme estimations for each estimator. The baseline specification is: $Y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025(x_1 - \bar{x_1})^2 - 0.01(x_2 - \bar{x_2})^2 + 0.015(x_3 - \bar{x_3})^2 + 0.02W + \varepsilon$ where Y is the dependent variable, the three x variables are additional covariates, W is a binary variable equal to one if the forcing variable w is greater than zero, and ε is the mean zero error term.

	Bias	RMSE	Coverage		Bias	RMSE	Coverage
Sample size = 500				Sample size = 1,000			
RDD	−1.06	106.19	6.10	RDD	−0.46	74.57	5.38
OLS	13.10	49.36	5.56	OLS	13.08	36.06	6.52
CF	2.91	55.29	6.73	CF	1.79	41.96	6.39
Sample size = 2,000				Sample size = 3,000			
RDD	0.15	53.15	5.79	RDD	0.61	42.87	5.44
OLS	13.23	27.02	8.13	OLS	13.59	23.48	9.86
CF	−0.07	31.47	6.26	CF	0.63	26.50	5.95
Sample size = 5,000				Sample size = 10,000			
RDD	0.74	33.68	5.65	RDD	0.51	23.75	5.22
OLS	13.42	20.13	13.89	OLS	13.51	17.17	23.69
CF	−0.19	22.25	5.86	CF	−0.49	18.06	4.98

Table IA.3. Comparison of bias and test coverage for causal forest, OLS, and RDD for specifications with latent variable that adds information to the forest. Each panel shows mean bias (as a percent of the true average treatment effect scaled by 100), root mean squared error (also as a percent of the true effect scaled by 100), and coverage (the percent of Monte Carlo simulations for which a 5% t-test of the difference between the estimate and true effect rejects the null hypothesis) for 10,000 sets of simulated data with the sample size given in the panel title, omitting the twenty most extreme estimations for each estimator. We augment the specification given in the caption of Figure 3 with a latent variable, x_4, that is included in the data generating process but not in the causal forest, OLS, and RDD estimations. The slope on x_4 (κ) is given in the panel titles and varies by column. The correlation between x_4 and the forcing variable, w_i, is given in the panel titles and varies by row. Table 1 gives the corresponding λ and $\widetilde{\Gamma}$ and for each column and row, respectively.

	Bias	RMSE	Coverage		Bias	RMSE	Coverage		Bias	RMSE	Coverage		Bias	RMSE	Coverage
Slope 0.01, Corr 0.05				Slope 0.03, Corr 0.05				Slope 0.05, Corr 0.05				Slope 0.07, Corr 0.05			
RDD	0.22	22.83	5.22	RDD	0.29	22.93	5.54	RDD	0.42	23.07	5.62	RDD	0.51	23.23	5.61
OLS	17.96	20.73	39.76	OLS	26.69	28.63	72.13	OLS	35.42	36.90	92.23	OLS	44.15	45.34	98.97
CF	4.74	18.01	6.48	CF	14.37	22.57	18.95	CF	24.17	29.97	41.61	CF	33.82	38.22	65.59
Slope 0.01, Corr 0.10				Slope 0.03, Corr 0.10				Slope 0.05, Corr 0.10				Slope 0.07, Corr 0.10			
RDD	0.29	22.89	5.19	RDD	0.18	22.88	5.09	RDD	0.13	22.88	4.91	RDD	0.06	22.94	4.93
OLS	22.29	24.59	56.54	OLS	39.71	41.04	96.55	OLS	57.14	58.06	100.00	OLS	74.55	75.26	100.00
CF	9.48	19.57	11.40	CF	28.85	33.50	53.87	CF	48.39	51.34	87.59	CF	67.88	70.06	96.20
Slope 0.01, Corr 0.20				Slope 0.03, Corr 0.20				Slope 0.05, Corr 0.20				Slope 0.07, Corr 0.20			
RDD	0.21	23.28	5.45	RDD	0.22	23.33	5.72	RDD	0.15	23.12	5.59	RDD	0.33	22.94	5.48
OLS	30.96	32.66	84.21	OLS	65.80	66.62	100.00	OLS	100.66	101.18	100.00	OLS	135.52	135.90	100.00
CF	19.36	26.11	29.65	CF	58.36	60.83	93.71	CF	97.30	98.81	98.84	CF	136.36	137.45	99.62
Slope 0.01, Corr 0.40				Slope 0.03, Corr 0.40				Slope 0.05, Corr 0.40				Slope 0.07, Corr 0.40			
RDD	0.08	22.88	5.09	RDD	0.09	22.55	4.99	RDD	0.17	22.07	5.26	RDD	0.21	21.18	5.03
OLS	48.32	49.41	99.65	OLS	118.02	118.46	100.00	OLS	187.70	187.97	100.00	OLS	257.41	257.60	100.00
CF	38.62	42.61	74.34	CF	116.55	117.77	99.40	CF	194.39	195.13	99.87	CF	272.35	272.85	99.94
Slope 0.01, Corr 0.55				Slope 0.03, Corr 0.55				Slope 0.05, Corr 0.55				Slope 0.07, Corr 0.55			
RDD	0.64	23.00	5.30	RDD	0.61	22.55	5.45	RDD	0.62	21.49	5.67	RDD	0.65	19.78	5.60
OLS	61.51	62.37	100.00	OLS	157.34	157.66	100.00	OLS	253.14	253.33	100.00	OLS	348.92	349.05	100.00
CF	53.44	56.43	90.71	CF	160.58	161.52	99.81	CF	267.86	268.42	100.00	CF	374.98	375.33	100.00

Table IA.4. Comparison of bias and test coverage for causal forest, OLS, and RDD for alternative functional form. Each panel shows mean bias (as a percent of the true average treatment effect scaled by 100), root mean squared error (also as a percent of the true effect scaled by 100), and coverage (the percent of Monte Carlo simulations for which a 5% t-test of the difference between the estimate and true effect rejects the null hypothesis) for 10,000 sets of simulated data with the sample size given in the panel title, omitting the twenty most extreme estimations for each estimator. The alternative functional form specification is: $Y = 0.050.05\,(x1 < \overline{x_1} - \sigma_{x_1}) - 0.005\,(x1 < \overline{x_1} + 0.5\sigma_{x_1}) + 0.01\,(x2 < \overline{x_2}) + 0.02\,(x2 < \overline{x_2} + 0.5\sigma_{x_2}) + 0.025\,(x2 < \overline{x_2} + \sigma_{x_2}) - 0.01\,(x2 \geq \overline{x_2} + \sigma_{x_2}) + 0.015x_3 + 0.02W + \varepsilon$ where Y is the dependent variable, the three x variables are additional covariates (with $\overline{x}$ indicating the mean of the variable and σ_x its standard deviation), inequalities in parentheses being indicator functions that are equal to one when the inequality is true, W is a binary variable equal to one if the forcing variable w is greater than zero, and ε is the mean zero error term.

	Bias	RMSE	Coverage
RDD	0.28	19.83	4.96
OLS	4.27	9.75	7.19
CF	−0.11	17.18	5.20

D Details of our application causal forest parameterization

The goal of this section is to fully describe our parameterization and the reasoning behind our choice of parameters. Parameters are listed in the table below.

List of causal forest parameters and definitions	
B	number of trees within a forest
sample fraction; $\frac{S_b}{B}$	proportion of data used in each tree
honesty fraction; $\frac{S_b}{2}$	proportion of data in S^{tr} and S^{est}
–	number of covariates selected for randomized splitting
k	minimum number of treatment and control observations in each child node
α	minimum fraction of the sample in each child node after split

Following Wager and Athey (2018), we employ the double-sample tree specification. In a double-sample tree, the algorithm selects a percentage of data for each estimation, called the sample fraction, and then splits that selected data into S^{tr} and S^{est}. We estimate a causal forest of 1,000 trees with honesty and sample fractions of 50%. Reducing the amount of data in each S^{tr} and S^{est} introduces variance into estimates but is important to reduce bias caused by overfitting. A forest of 1,000 trees is large enough that any remaining variation in the HTEs is due to the data and not the randomized tree growing process. We select 1,000 trees and honesty and sample fractions of 50% because this is a large enough number of trees and small enough honesty and sample fractions such that reasonable variations in these choices do not affect our estimates.

Like all forests (Breiman, 2001), causal forest employs randomized splitting to ensure trees do not overfit. Our causal forest estimation includes eleven covariates, and we set the number of covariates available at each node for splitting to nine. Thus, our algorithm selects the optimal partition from among nine randomly selected covariates at each node. We set this parameter to nine because parameters less than nine result in trees being dropped from our estimation. A tree drops from the sample if no split can be made on the included covariates. A covariate must be sufficiently relevant to the estimation in order for a split to be made. If a randomly selected set of covariates does not include enough relevant covariates, the tree will not grow. In our causal forest estimation, we find that a small number of covariates dominate in importance. If at least some of these important covariates are not included in the randomly selected subset of covariates, trees cannot be estimated and are dropped from the sample. Thus, we cannot set the number of included covariates any less than nine because doing so results in random draws of covariates that are not sufficiently relevant to the estimation and drops trees from the sample.

We also select a minimum node size (k) that ensures our estimation does not drop trees. Node size is the minimum number of treated observations that must be included in a node after a split. If a split results in fewer treated observations than this minimum, the resulting node is a terminal node (i.e., is not split any further). When an honest tree is grown, the splitting rules are calculated on one portion of data (S^{tr}) and treatment effects are calculated on a second portion of data (S^{est}) using these splitting rules. If the minimum node size is set too small, it is possible for terminal nodes in S^{est} to have no treated observations. If a terminal node has no treated observations, treatment effects cannot be calculated and the tree drops from the forest. We set minimum node size to ten. Minimum node size is determined similarly to number of covariates available on which to split. We set minimum node size large enough to ensure trees will not drop from our estimation but small enough such that our final results are not sensitive to increasing this number. In addition to ensuring that results from our forest estimation are not sensitive to any one parameter selection, we also ensure that the combination of selected parameters is appropriate by estimating a series of mini-forests with slightly different parameter selections and combinations.[1]

We allow a tuning function to select our final parameters for our estimation and use the `grf` *R* package for our causal forest estimation. This function will estimate a series of small forests (500 forests of 200 trees each) with various parameter sets and estimate a set of parameters that minimize error in the estimation defined by Nie and Wager (2021). This estimation yields an α, or the minimum fraction of the sample that must be contained in each daughter node after each split, of just over 8% and an imbalance penalty of approximately 0.06%. The imbalance penalty biases each partition towards a 50-50 split, requiring that the reduction in MSE obtained at each partition is greater than what would be obtained splitting at the median. These parameterizations are important to ensure that each estimated tree does not grow too deep and overfit the data.

[1]We discuss two parameters here: (i) the minimum number of both treatment and control observations within a node (k; note that the minimum number of observations within a leaf is then $2k$) and (ii) the minimum proportion of the parent node that must be contained within each child node (α). There is a third parameter, the imbalance penalty, that is available in the `grf` $\mathbb{R}$ package (Tibshirani, Athey, Friedberg, Hadad, Hirshberg, Miner, Sverdrup, Wager, and Wright, 2020). The package authors note that it is an experimental parameter. It has not proven important in our testing, so, for brevity, we omit a discussion of this parameter.

E Auxiliary tables and figures from our application

List of auxiliary tables and figures:

- Figure IA.4 plots HTEs against covariates. We estimate HTEs in the causal forest estimation described in the caption of Figure 5.
- Figure IA.5 shows results from a McCrary (2008) test for our whole sample. We reject the null of no bunching with a p-value of 0.0013, also reported in Panel A of Table 6.
- Figure IA.6 shows McCrary (2008) tests for subsamples defined in the caption of Table 6. We fail to reject the null of no bunching for the top panel (p-value of 0.4184). We reject the null of no bunching in the middle and bottom panels with p-values of 0.0053 and 0.0581, respectively, at the 1% and 10% levels.
- In Figure IA.7, we use the classification forest predicting default propensities from our forest estimation and plot these predicted propensities against the distance to default expressed as a ratio. There is a concerning pattern in default propensities around the default threshold, a jump in default propensities moving from just above the default threshold to just below the default threshold. This jump means that firms just above and just below the threshold differ from each other in such a way that, even without the slack variable included in the classification forest, the estimation is able to distinguish between these two groups of firms. The difference between these two average predicted default propensities is statistically significant (t-statistic of 4.31). This analysis provides additional empirical evidence that firms just above and just below the threshold are not random and loan covenant defaults do not generally provide "as good as random" assortment of firms around the default threshold.
- Table IA.5 presents summary statistics for the sample for our causal forest estimation.
- Table IA.6 provides ATC and ATO estimates, in addition to the ATE and ATT estimates given in Figure 5, summarizing HTEs for the causal forest estimation.
- Table IA.7 provides CATE, CATT, CATC, and CATO estimates for subsamples defined on firm size and cash flow.
- Table IA.8 provides CATC and CATO estimates, in addition to the CATE and CATT estimates in Table 4, for subsamples defined on firm macro Q and cash flow.
- Table IA.9 shows estimated treatment effects for ad hoc bandwidths around the default threshold defined using slack ratio.
- Table IA.10 provides variable importance ranking for the centering step in which we predict both default propensity (Panel A) and investment (Panel B).
- Table IA.11 provides re-estimations of ATE from causal forest estimations in which we leave each covariate out of the estimation in turn. Only removing Altman Z-score from the estimation results in an ATE that is statistically different from the ATE in Table IA.6. Removing Altman Z results in

a more negative ATE because the forest is less able to control for differences between firms in and not in default, which biases treatment effects downward.

- Table IA.12 provides results from a statistical test of balance. To gauge covariate overlap (i.e, balance), we follow McCaffrey et al. (2013) and calculate "population" standardized bias (PSB). For each covariate k, we calculate:

$$PSB_k = \frac{\overline{k} - \overline{k_p}}{\widehat{\sigma_{k_p}}}, \tag{IA.7}$$

where $\overline{k_p}$ and $\widehat{\sigma_{k_p}}$ are the mean and standard deviation for covariate k for the pooled sample (denoted p), and $\overline{k}$ is the propensity-score-weighted mean of covariate k, calculated as:

$$\overline{k} = \frac{\sum_{i=1}^{n} \frac{W_i \times k_i}{\widehat{\pi}(X)}}{\sum_{i=1}^{n} \frac{W_i}{\widehat{\pi}(X)}}, \tag{IA.8}$$

where $\widehat{\pi}(X)$, the estimated propensity score for observation X with covariates X_i and treatment status W_i, is calculated with the pooled sample. Intuitively, PSB captures the similarity between covariates of observations that are treated and the entire sample; if the propensity-score-weighted mean and actual mean of covariate k are closer, PSB is smaller.

Table IA.12 reports PSB statistics for our key covariates that suggest that we have relatively good overlap. According to the guidelines McCaffrey et al. (2013) provide, a PSB score for a covariate less than 0.20 is considered low, around 0.40 is considered moderate, and greater than 0.60 is considered high. All of our covariates have PSB statistics between 0.22 and 0.35. The highest PSB score, 0.35, is for Altman Z score. Intuitively, we expect Altman Z, a measure of firm financial health, to differ between firms in and not in technical default. However, the PSB statistic for Altman Z is still within the "moderate" range. Thus, PSB statistics suggest that our setting has reasonable overlap for our key covariates.

- Table IA.13 presents RDD results using triangular kernels and local linear specifications for the subsamples that we examine in Figure 7. Figure 7 presents estimates that are differences in average investment between firms above and below the treatment threshold for varying bandwidths. Results in Figure 7 show that, for firms in Panel A, LATE is negative close to the threshold and, for firms in Panel B, LATE is positive close to the threshold. Similarly, Table IA.13 shows negative and positive LATE estimates for firms in Panels A and B, respectively, albeit using more sophisticated RDD specifications. These results are also similar to those in Panels B and C of Table 6. All three sets of results are different ways of using causal forest's centering step to model treatment status and use RDD to estimate LATE for firms that plausibly can and cannot avoid default.

Figure IA.4. **HTE plots.** Heterogeneous treatment effects plotted against covariates. Details of the causal forest estimation are in the caption of Figure 5.

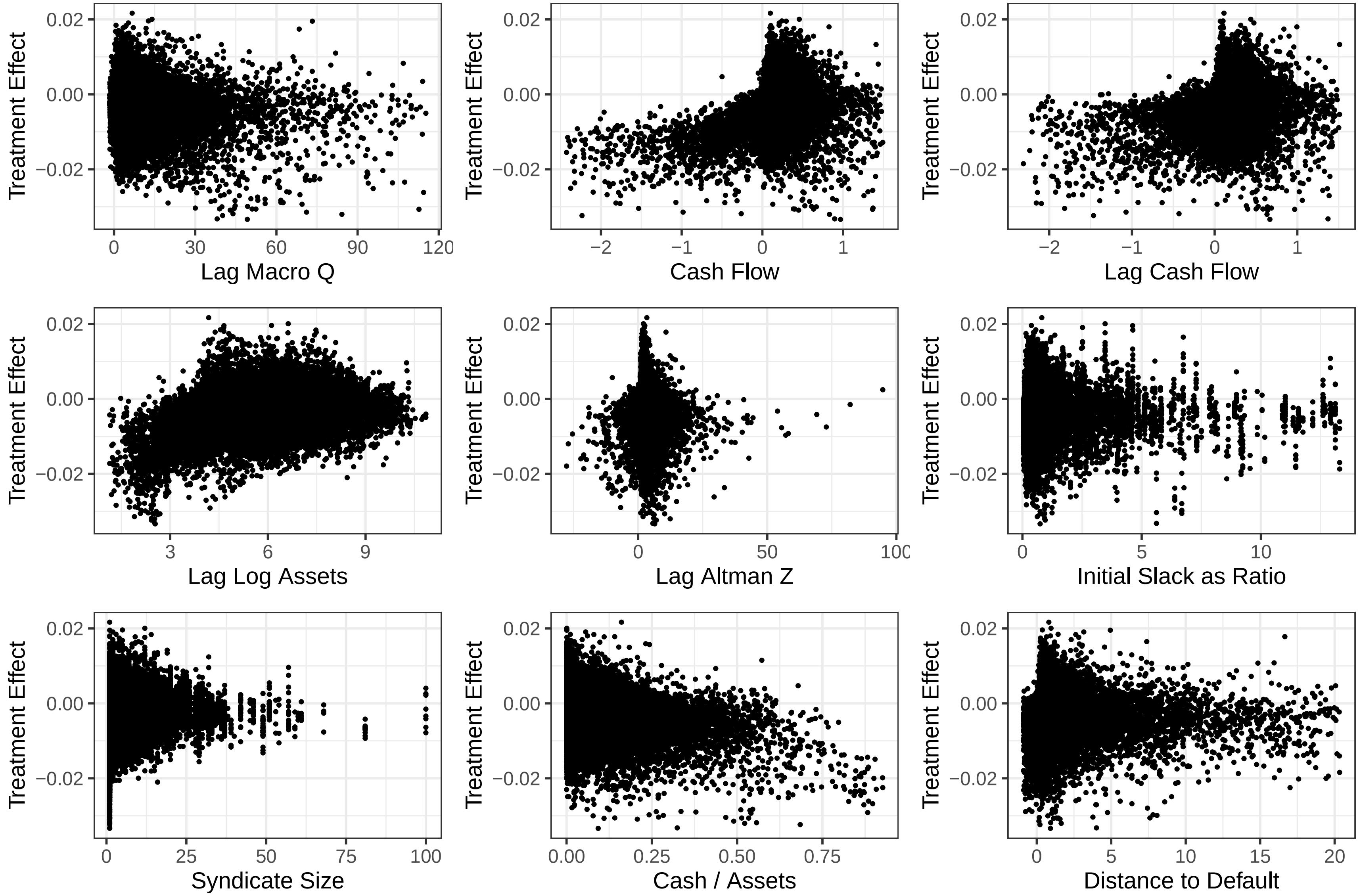

Figure IA.5. **McCrary test results showing a discontinuity in the forcing variable at the threshold for the whole sample.** McCrary (2008) tests of manipulation at the default with the forcing variable expressed as a ratio (percentage of threshold), so that the default threshold equals zero, with the sample limited to observations with ratios between −1 and 1. The sample is described in the caption of Table IA.5. The null of no manipulation is rejected (p-value of 0.0096).

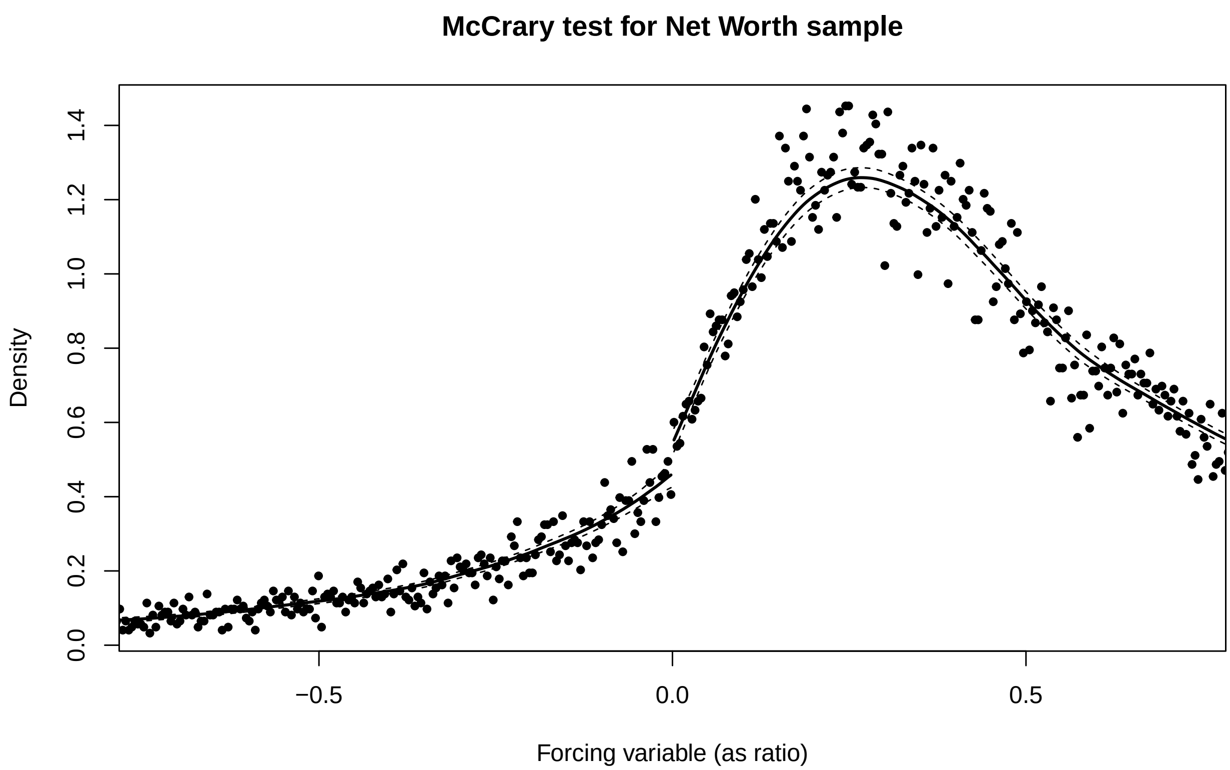

Figure IA.6. McCrary tests for a discontinuity in the forcing variable at the threshold in subsamples. McCrary (2008) tests of manipulation at the default with the forcing variable expressed as a ratio (percentage of threshold), so that the default threshold equals zero, with the sample limited to observations with ratios between −1 and 1 for the three subsamples defined in the caption of Table 6. See Table 6 for RDD estimates and McCrary (2008) p-values.

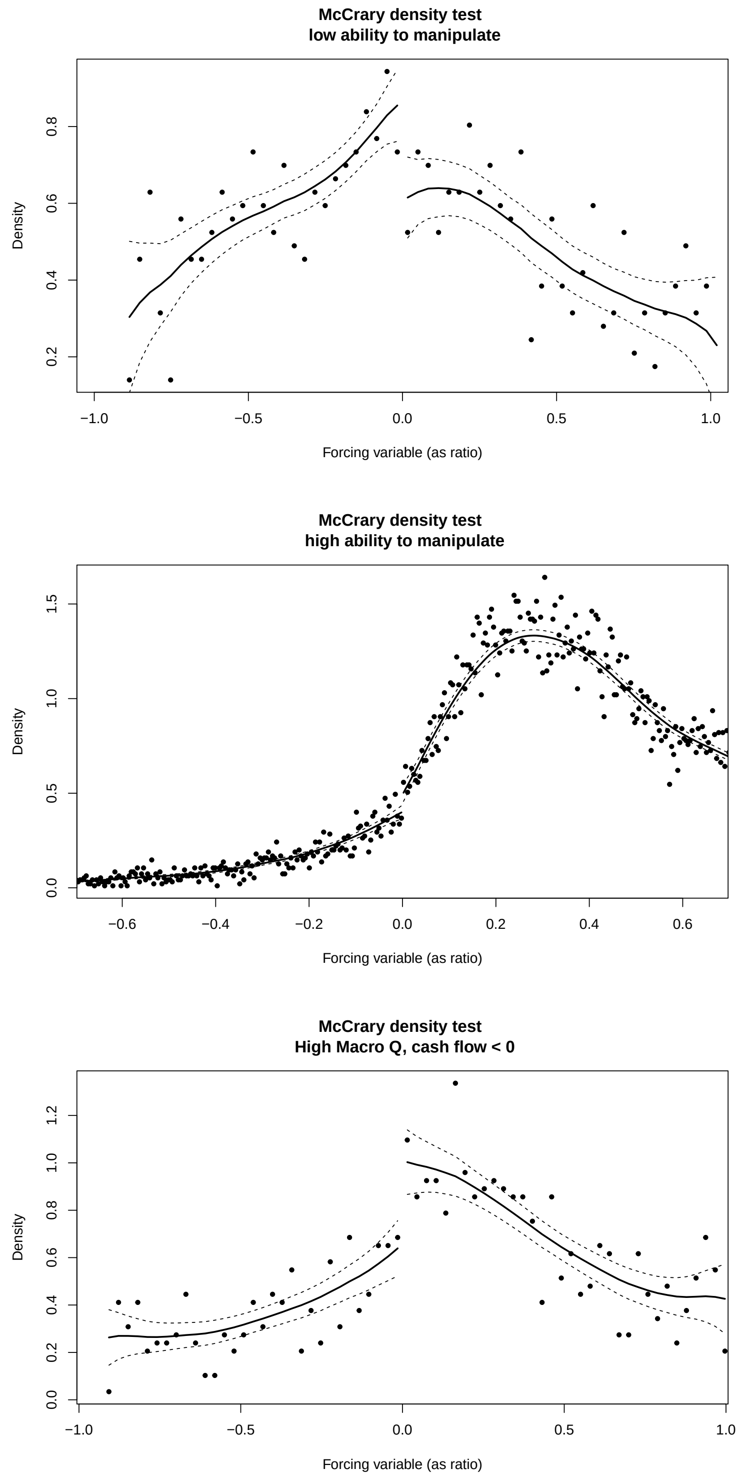

Figure IA.7. Predicted default propensities. Predicted defaults estimated using a classification forest estimation described in the caption of Table IA.10. We plot estimated default propensities, each with a light blue x, against slack ratio, or the distance to default scaled by the default threshold. Averages for 5% bands of slack ratio are given with horizontal blue bars. The height of the blue boxes indicates 95% confidence intervals around these averages. The average predicted default propensity just above and just below the default threshold (slack ratio of zero) are statistically different from each other (t-statistic of 4.31).

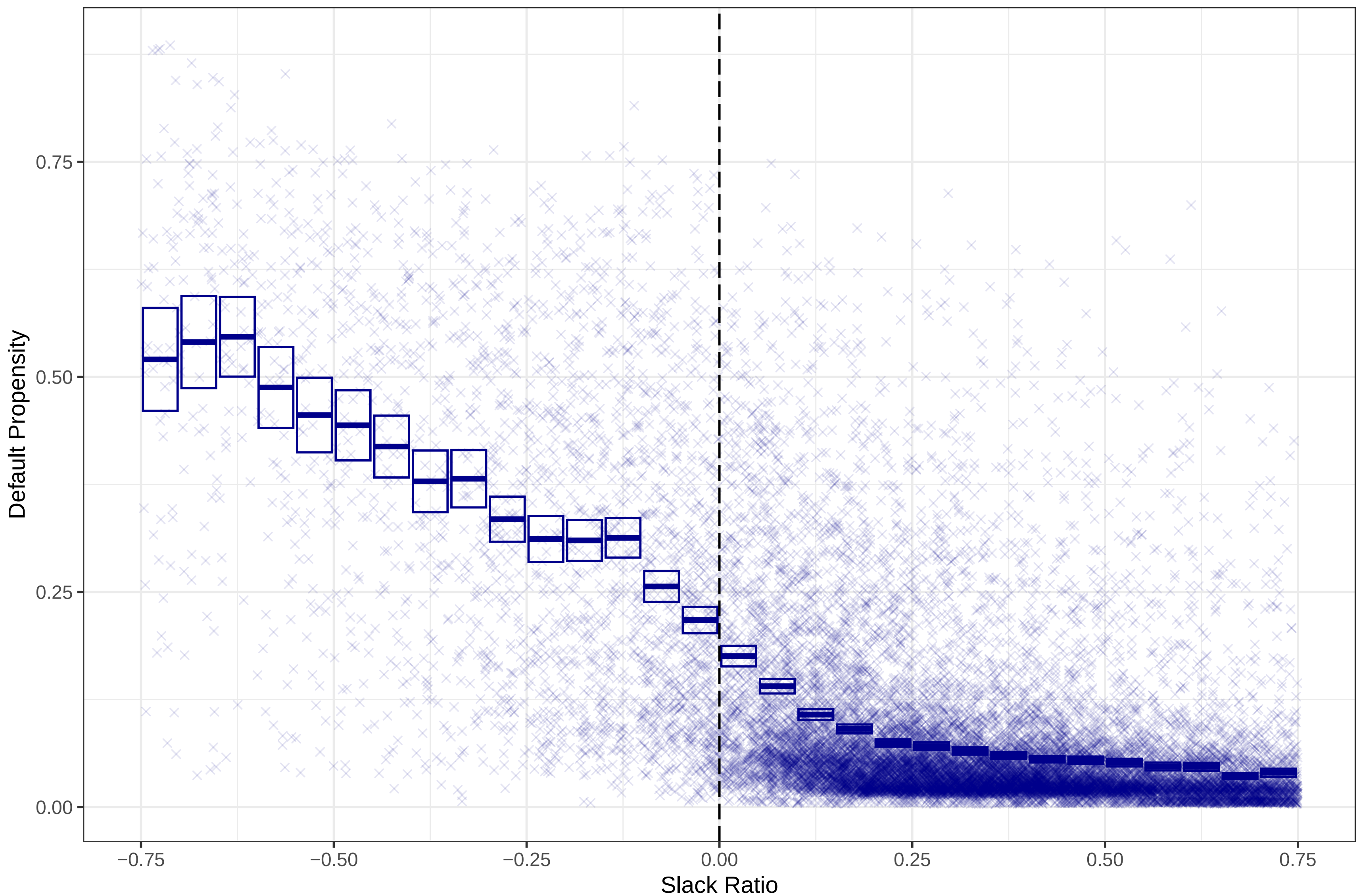

Table IA.5. Summary statistics. Averages, [medians], and (standard errors) of firm characteristics for firm-quarter observations for deals and loans. The sample includes all firm-quarter observations for firms with a covenant restricting minimum (tangible) net worth listed in Dealscan between 1994 and 2017. Variable definitions are as follows: Current ratio is current assets divided by current liabilities; net worth is total assets minus total liabilities; tangible net worth is the sum of current assets, net PPE, and other assets, minus total liabilities; log(assets), or firm size, is calculated as the natural logarithm of total assets, deflated to December 2000 by the all-urban CPI; market-to-book is the sum of market equity, total debt, and preferred stock liquidation minus deferred taxes and investment tax credits, divided by total assets; Macro Q is total book debt plus market equity minus total inventories, divided by start-of-period PPE; ROA is operating income before depreciation divided by total assets; capital/assets is net PPE divided by total assets; investment/capital is capital expenditures divided by start-of-period PPE; cash flow is the sum of income before extraordinary items and depreciation and amortization, divided by start-of-period PPE; and leverage is total debt divided by total assets. All variables are winsorized at the top and bottom 1%. Firms must have positive debt and non-missing current ratio or net worth to remain in the sample.

Variable	Mean [Median]	(SE)
Net Worth	789.19	(14.86)
	[186.31]	
Tangible Net Worth	773.19	(14.01)
	[183.27]	
Current Ratio	2.11	(0.01)
	[1.79]	
Log(Assets)	6.06	(0.01)
	[6.09]	
Market-to-Book	1.29	(0.01)
	[0.97]	
Macro Q	7.17	(0.06)
	[3.13]	
ROA	0.03	(0.00)
	[0.03]	
Capital/Assets	0.33	(0.00)
	[0.26]	
Investment/Capital	0.06	(0.00)
	[0.04]	
Cash Flow	0.08	(0.00)
	[0.07]	
Leverage	0.24	(0.00)
	[0.22]	
Firm-Quarter Obs	46,306	
Firms	2,628	

Table IA.6. **Causal forest results including ATC and ATO.** Average treatment effect (ATE), average treatment effect on the treated (ATT), average treatment effect on the control (ATC), and average treatment effect with an overlap correction of Li et al. (2018) (ATO) of default on firm investment (quarterly capital expenditures divided by beginning-of-period PPE and scaled by 100 for ease in interpretation) for the net worth sample (quarterly observations, 1994 – 2017). Causal forest is estimated following Wager and Athey (2018) and contains 1,000 trees. Covariates included in the estimation that are lagged one quarter are: Macro Q, log(assets), and Altman Z-score. Also included are current quarter and lagged one quarter cash flow, a binary variable for whether the firm has a credit rating, the loan syndicate's size, current cash-over-assets, initial and contemporaneous distance to default, and fixed effects for firm, year, and quarter.

	est	se	t-stat	p-val	N
ATE	-0.24	0.18	-1.326	0.185	32,530
ATT	-0.54^{***}	0.09	-5.670	<0.001	32,530
ATC	-0.18	0.21	-0.873	0.383	32,530
ATO	-0.53^{***}	0.10	-5.358	<0.001	32,530

Note: *p<0.1; **p<0.05; ***p<0.01

Table IA.7. **Averages of causal forest HTEs: Subsamples conditional on cash flow and size including CATC and CATO.** Conditional average treatment effect (CATE), conditional average treatment effect on the treated (CATT), conditional average treatment effect on the control (CATC), and conditional average treatment effect with an overlap correction of Li et al. (2018) (CATO) of default on investment for subsamples of small (top panels) and large (bottom panels) firms and firms with negative (results to left) and non-negative (results to right) cash flows. HTEs are estimated with the causal forest estimation described in the caption of Table IA.6.

	est	se	t-stat	p-val	N		est	se	t-stat	p-val	N
Panel A: Small firms, Cash flow < 0						Panel C: Small firms, Cash flow ≥ 0					
CATE	-0.96^{***}	0.24	-3.938	<0.001	2,290	CATE	0.34	0.54	0.625	0.532	8,553
CATT	-0.89^{***}	0.27	-3.291	0.001	2,290	CATT	-0.31^{*}	0.19	-1.667	0.096	8,553
CATC	-0.96^{***}	0.27	-3.623	<0.001	2,290	CATC	0.50	0.63	0.797	0.426	8,553
CATO	-0.89^{***}	0.28	-3.226	0.001	2,290	CATO	-0.36^{*}	0.21	-1.714	0.087	8,553
Panel B: Large firms, Cash flow < 0						Panel D: Large firms, Cash flow ≥ 0					
CATE	-0.69	0.50	-1.389	0.165	742	CATE	-0.18	0.27	-0.670	0.503	10,101
CATT	-0.68^{**}	0.33	-2.064	0.039	742	CATT	-0.21	0.20	-1.012	0.312	10,101
CATC	-0.68	0.55	-1.233	0.218	742	CATC	-0.18	0.28	-0.631	0.528	10,101
CATO	-0.72^{**}	0.35	-2.046	0.041	742	CATO	-0.16	0.22	-0.758	0.448	10,101

Note: *p<0.1; **p<0.05; ***p<0.01

Note: Small (large) firms are observations in the lowest (highest) size tercile.

Table IA.8. Average HTEs in subsamples based on cash flow and Macro *Q* including CATC and CATO. Average treatment effect (ATE), average treatment effect on the treated (ATT), average treatment effect on the control (ATC), and average treatment effect with an overlap correction of Li et al. (2018) (ATO) of default on firm investment for subsamples of firms with low (top panels) and high (bottom panels) Macro *Q* and firms with negative (results to left) and non-negative (results to right) cash flows. HTEs are estimated with the causal forest estimation described in the caption of Table IA.6.

	est	se	t-stat	p-val	N		est	se	t-stat	p-val	N
Panel A: Low Macro Q, Cash flow < 0						Panel C: Low Macro Q, Cash flow ≥ 0					
CATE	-0.48^{**}	0.21	-2.315	0.021	1,631	CATE	-0.31	0.21	-1.436	0.151	9,212
CATT	-0.28	0.21	-1.365	0.172	1,631	CATT	-0.28^{**}	0.13	-2.112	0.035	9,212
CATC	-0.55^{**}	0.23	-2.360	0.018	1,631	CATC	-0.30	0.24	-1.253	0.210	9,212
CATO	-0.33	0.22	-1.495	0.135	1,631	CATO	-0.26	0.16	-1.637	0.102	9,212
Panel B: High Macro Q, Cash flow < 0						Panel D: High Macro Q, Cash flow ≥ 0					
CATE	-1.40^{***}	0.44	-3.160	0.002	1,407	CATE	-0.14	0.51	-0.280	0.780	9,436
CATT	-1.45^{***}	0.47	-3.102	0.002	1,407	CATT	-0.59^{**}	0.29	-1.989	0.047	9,436
CATC	-1.36^{***}	0.50	-2.737	0.006	1,407	CATC	-0.09	0.54	-0.116	0.868	9,436
CATO	-1.39^{***}	0.48	-2.867	0.004	1,407	CATO	-0.65^{**}	0.29	-2.242	0.025	9,436

Note: *p<0.1; **p<0.05; ***p<0.01
Note: High (low) Macro *Q* sample is observations in the highest (lowest) Macro *Q* tercile.

Table IA.9. Average HTEs in subsamples based on distance to default including CATC and CATO. Conditional average treatment effect (CATE), conditional average treatment effect on the treated (CATT), conditional average treatment effect on the control (CATC), and conditional average treatment effect with an overlap correction of Li et al. (2018) (CATO) of default on investment for subsamples of firms within the distance to default given in the panel header. HTEs are estimated with the causal forest estimation described in the caption of Table IA.6.

	est	se	t-stat	p-val	N
Panel A: Slack ratio $\leq$ 0.50					
CATE	-0.14	0.20	-0.716	0.474	14,727
CATT	-0.39	0.11	-3.572	0.000	14,727
CATC	-0.09	0.23	-0.411	0.681	14,727
CATO	-0.42	0.11	-3.930	0.000	14,727
Panel B: Slack ratio $\leq$ 0.25					
CATE	-0.12	0.24	-0.489	0.625	7,264
CATT	-0.25	0.15	-1.691	0.091	7,264
CATC	-0.10	0.27	-0.349	0.727	7,264
CATO	-0.32	0.13	-2.528	0.011	7,264
Panel C: Slack ratio $\leq$ 0.10					
CATE	-0.0002	0.0037	-0.052	0.958	2,372
CATT	-0.0016	0.0023	-0.687	0.492	2,372
CATC	-0.0001	0.0040	-0.014	0.988	2,372
CATO	-0.0014	0.0018	-0.780	0.436	2,372
Panel D: Slack ratio $\leq$ 0.05					
CATE	0.20	0.56	0.362	0.717	1,043
CATT	-0.16	0.36	-0.448	0.654	1,043
CATC	0.24	0.60	0.409	0.683	1,043
CATO	-0.15	0.24	-0.627	0.530	1,043

Note: *p<0.1; **p<0.05; ***p<0.01

Table IA.10. Ranking of variable importance to tree growth, predicting treatment status and investment. Variable importance for the centering step of the forest estimation. Ranking using net worth sample (quarterly observations, 1994 – 2017) for a classification forest estimation predicting propensity to default (Panel A) and a regression forest estimation for investment (Panel B). Covariates are ranked by their relative importance to investment and propensity estimation. See the caption of Table 3 for details on ranking procedure.

	Absolute Mean Difference	Standard Deviation of Differences	Ratio of Difference Variance to HTE Variance
Panel A: Variable importance for a classification forest predicting default propensity			
lag Altman Z	6.38	12.80	1.0204
initial slack	4.90	9.75	0.5922
lag assets	2.85	6.73	0.2820
cash flow	2.40	6.04	0.2272
syndicate size	0.94	2.82	0.0494
lag cash flow	0.95	2.27	0.0322
firm fixed effect	0.78	2.02	0.0254
lag Macro Q	0.80	1.90	0.0226
year	0.64	1.65	0.0170
cash / assets	0.57	1.56	0.0151
credit rating (binary)	0.07	0.30	0.0006
quarter	0.08	0.22	0.0003
Panel B: Variable importance for a regression forest predicting investment			
firm fixed effect	2.77	3.88	1.1988
lag Macro Q	1.71	2.50	0.4969
lag Altman Z	0.63	0.90	0.0640
year	0.60	0.89	0.0634
lag assets	0.57	0.84	0.0557
lag cash flow	0.33	0.49	0.0188
cash flow	0.21	0.31	0.0079
cash / assets	0.11	0.19	0.0029
distance to default (slack)	0.11	0.18	0.0025
quarter	0.09	0.16	0.0021
initial slack	0.08	0.13	0.0013
syndicate size	0.07	0.12	0.0011
credit rating (binary)	0.02	0.04	0.0001

Table IA.11. Averages of causal forest HTEs: ATE re-estimations leaving out one covariate. Re-estimations of ATE from forest estimations leaving out the covariate listed in each row. The last five rows test whether this ATE statistically differs from the ATE in Table IA.6, our main specification in which all covariates are included.

	ATE from forest without covariate					Difference between estimate and ATE in Table IA.6				
dropped covariate	est	se	t-stat	p-val	N	diff	se	t-stat	p-val	N
lag Altman Z	-0.75^{***}	0.12	-6.439	0.000	32,530	-0.51^{**}	0.22	-2.260	0.024	32,530
lag cash flow	-0.04	0.30	-0.131	0.895	32,530	0.20	0.36	0.566	0.571	32,530
firm fixed effect	-0.39^{*}	0.24	-1.653	0.098	32,530	-0.15	0.30	-0.482	0.630	32,530
initial slack	-0.38^{***}	0.11	-3.310	0.001	32,530	-0.13	0.22	-0.593	0.553	32,530
lag assets	-0.12	0.22	-0.528	0.598	32,530	0.13	0.29	0.432	0.666	32,530
slack ratio	-0.37^{*}	0.21	-1.767	0.077	32,530	-0.12	0.28	-0.441	0.659	32,530
cash flow	-0.36^{**}	0.16	-2.317	0.021	32,530	-0.12	0.25	-0.477	0.633	32,530
year	-0.34^{*}	0.20	-1.682	0.093	32,530	-0.09	0.28	-0.335	0.738	32,530
syndicate size	-0.33	0.20	-1.615	0.106	32,530	-0.09	0.28	-0.311	0.756	32,530
quarter	-0.18	0.22	-0.817	0.414	32,530	0.06	0.29	0.209	0.834	32,530
credit rating (binary)	-0.20	0.21	-0.953	0.341	32,530	0.04	0.28	0.151	0.880	32,530
lag(Macro Q)	-0.24	0.20	-1.184	0.236	32,530	0.01	0.28	0.031	0.975	32,530
cash / assets	-0.25	0.29	-0.841	0.400	32,530	0.00	0.35	-0.011	0.991	32,530

Note: *p<0.1; **p<0.05; ***p<0.01

Table IA.12. Test for balance in covariates. Population standardized bias (PSB) statistics from McCaffrey et al. (2013) for key covariates from the forest estimation described in the caption of Figure 5.

covariate	PSB
lag Macro Q	0.23
lag assets	0.27
initial slack	0.31
ROA	0.26
cash flow	0.30
lag cash flow	0.30
Altman Z-score	0.35

Table IA.13. RDD results estimated with triangular kernels using subsamples based on technical default propensities. Estimations of the effect of default on investment (quarterly capital expenditures divided by beginning-of-period PPE and scaled by 100 for ease in interpretation). Each panel presents a different subsample of the sample described in the caption of Table IA.5. In Panel A the sample is restricted to observations likely to be in default, defined as observations having a default propensity (estimated using a regression forest) between 22.99% and 75.48% (the range within which there are an equal number of defaults and non-defaults). In Panel B the sample is restricted to observations with below 22.99% default propensity. The "IK Bandwidth" result is calculated using a local linear specification in the Imbens and Kalyanaraman (2012) optimal bandwidth (from the RDD $\mathbb{R}$ package). Results in the remaining four columns are for local polynomial specifications calculated in Calonico et al. (2020) optimal bandwidths with the estimation constrained to have a maximum polynomial given in the column header (from the RDrobust $\mathbb{R}$ package).

	Model				
	IK Bandwidth	Linear	Quadratic	Cubic	Quartic
Panel A: High propensity to be in default					
bind	−0.10	−0.09	−0.45	−0.91	−1.24
	(0.55)	(0.56)	(0.73)	(0.92)	(1.05)
Bandwidth	0.29	0.30	0.36	0.34	0.36
Obs. below	849	871	985	953	989
Obs. above	837	853	943	919	944
Panel B: Low propensity to be in default					
bind	0.04	0.31	0.44	0.54	0.57
	(0.32)	(0.41)	(0.55)	(0.62)	(0.82)
Bandwidth	0.38	0.21	0.25	0.37	0.33
Obs. below	898	709	786	888	862
Obs. above	8,543	3,799	5,110	8,268	7,172