The heterogeneous effects of default on investment: An application of causal forest in corporate finance*

Huseyin Gulen[†] Candace E. Jens[‡] T. Beau Page[§]

December 26, 2021

Abstract

Answering causal questions with extendable results is challenging. Regression discontinuity design (RDD) recovers selection-bias-free estimates that are uninformative outside of the threshold sample. Using Monte Carlo experiments, we compare the performance of RDD against causal forest, a non-parametric, machine-learning-based matching estimator, at recovering estimates in panel data. Even in simulations with selection bias, causal forest recovers estimates that are low-bias and much more precise than RDD estimates. Consequently, causal forest commonly outperforms RDD at recovering "true" treatment effects. We re-visit a popular RDD design, debt covenant defaults, to show in practice how extendable and heterogeneous causal forest estimates enhance inferences.

Keywords: causal forest, investment, financing, RDD, machine learning.

JEL Classification: G32, G31, C50.

*We thank Anat Admati, Josh Angrist, Shai Bernstein, Alex Butler, Zhaojing Chen, Sergey Chernenko, David Denis, Isil Erel, Miguel Ferreira, Oleg Gredil, Selin Gulen, Manasa Gopal, Camille Hebert, Steven Jens, Nishad Kapadia, Weiling Liu, Michelle Lowry, Anya Nakhmurina, Ryan Peters, Mitchell Petersen (discussant), Robert Prilmeier, Matthew Ringgenberg (discussant), James Reeder, Amir Sufi, Jin Xu, Irene Yi, Morad Zekhnini, and seminar participants at Tulane University, the Federal Reserve Board of Governors, the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC), University of Massachusetts (Boston), Wayne State University, the New Technologies in Finance Conference (Columbia Business School), the University of Iowa, the Northeastern University Finance Conference 2020, and the 2020 Texas A&M Young Scholars in Finance Conference for helpful comments and suggestions. This paper was previously circulated under the title, "An application of causal forest in corporate finance: How does financing affect investment?" The views in this paper are those of the authors and do not necessarily reflect those of the Office of the Comptroller of the Currency or the US Department of the Treasury.

[†]Krannert School of Management, Purdue University, 403 West State Street, West Lafayette, IN 47907. Tel: (765) 496-2689, e-mail: hgulen@purdue.edu.

[‡]A.B. Freeman School of Business Tulane University Goldring/Woldenberg Business Complex 7 McAlister Dr. New Orleans, LA 70118-5645 phone: (504) 865-5042 e-mail: cjens@tulane.edu.

[§]Office of the Comptroller of the Currency, e-mail: trenton.page@occ.treas.gov.

1 Introduction

Establishing causality is a challenge in corporate finance. As researchers rarely have the opportunity to test theories in field experiments, they increasingly rely on econometric techniques that allow for causal inference by mimicking randomized control trials ("pseudo-randomized techniques"), including: difference-in-difference (DD), regression discontinuity design (RDD), and pseudo-randomized experiments.[1] Such studies tend to focus on small samples wherein the strict assumptions required by these techniques are upheld. While these studies have stronger claims to causality than previous literature, this new line of literature can be criticized because such narrow settings produce results with limited extendibility, raising the question of what we are learning about the general behavior of firms from these studies.[2] Despite their limitations, pseudo-randomized techniques are now widely used because only a few econometric techniques can address sources of endogeneity common in corporate finance.

The trade-off between establishing causality and extendibility of results is particularly salient for RDD studies. RDD can recover treatment effects free of selection bias, a common endogeneity concern, but these estimates have limited external validity. For example, a number of studies examine consequences of violating debt covenants (i.e., technical default) using an RDD setup comparing the investment of firms in and not in default (Chava and Roberts, 2008; Nini, Smith, and Sufi, 2009; Demiroglu and James, 2010; Falato and Liang, 2016). If firms at default thresholds randomly default, this framework recovers a treatment effect net of underlying differences between firms in and not in default (i.e., selection bias). Unfortunately, however, we are unable to learn from this setup about the costs of default for firms far below default thresholds or the hypothetical costs to firms well out of default.

We introduce into the corporate finance literature a new machine learning technique, causal forest from Wager and Athey (2018), that can help balance preserving extendibility while establishing causality of results. Causal forest estimates heterogeneous, observation-level treatment effects by using trees to group like treatment and control firms. Existing literature uses causal forest to recover heterogeneity in effects in randomized studies.[3] We

[1] Bowen, Frésard, and Taillard (2017) provide statistics documenting the rise in such techniques in corporate finance publications.

[2] See, for example, discussion in Welch (2015) and Kahn and Whited (2018).

[3] See Davis and Heller (2017), Athey and Wager (2019), Knittel and Stolper (2019), and O'Neill and Weeks (2019). The use of machine learning in finance outside of creating or improving the measure of explanatory variables is rare but its popularity is increasing. One exception is Erel, Stern, Tan, and Weisbach (2021), who use a variety of machine learning algorithms for prediction, but these do not include causal forest. Also, Easley, de Prado, O'Hara, and Zhang (2021) use a regression forest in a microstructure setting. Both regression forests and causal forests are generalized random forests. Causal forest predicts an estimated treatment effect for an unobservable "ground truth", whereas a regression forest tries to predict an outcome.

recognize that, while not explicitly designed to control for selection bias, causal forest is a data-driven, non-parametric estimator that provides better "matching" in observational data than traditional techniques in the literature. Because selection bias is driven by underlying differences between treatment and control firms, this improvement in matching reduces bias in estimates.

The goal of our study is to evaluate whether causal forest performs well enough at ameliorating selection bias that it offers a viable alternative to techniques like RDD. If causal forest can recover treatment effects across a broad sample with sufficiently low bias, causal forest bridges the middle ground between OLS and traditional matching methods and pseudo-randomized techniques. In this case, depending on the specifics of an empirical setting, causal forest may provide benefits over pseudo-randomized techniques at answering a question, proving a valuable addition to our econometrics toolkit.

Our paper has three parts. First, we discuss the theoretical benefits of causal forest relative to matching estimators that corporate finance researchers traditionally use. Understanding causal forest's benefits over traditional matching estimators is fairly straightforward because causal forest relies on similar intuition and assumptions as matching estimators. Second, we use Monte Carlo experiments to empirically compare causal forest and RDD's recovery of treatment effects across a variety of scenarios. Comparing causal forest to RDD is more complex, as both techniques rely on different underlying assumptions and have different strengths and weaknesses. Our simulations are designed to pit the techniques against each other so we can directly compare their disparate advantages and disadvantages. Third, we provide an example application of causal forest to give concreteness to its benefits. In this application, we use causal forest to re-examine the effects of technical default on investment, recover smaller estimates than RDD studies in the literature, and use the heterogeneity in our causal forest estimations to reconcile between causal forest and RDD estimates.

Correspondingly, our study makes three major contributions. First, we provide a complete and intuitive discussion of causal forest estimation and results. We draw on this discussion to motivate theoretical reasons causal forest is relatively robust to omitted variable bias. Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019) provide technical treatment of causal trees and generalized random forests, of which causal forest is a special case. We focus on building intuition for causal forest through comparison with popular matching estimators.

Causal forests comprise trees, which "grow" to create comparison sets of treatment and control observations. Each tree is grown on a subset of data, which creates variation in comparison sets across trees. (For brevity, we gloss over this sampling process and other technical details here, but provide a full discussion of trees in Section 2.2.) To grow a tree, an algorithm repeatedly partitions data by identifying splitting points in the data that maximize heterogeneity in estimated effects. For example, if technical default greatly affects firms with negative cash flow but not firms with positive cash flow, the first partition splits the sample at zero cash flow. This split creates the first level in the tree and results in two subsamples of observations with very dissimilar treatment effects. Each subsample is then split in the same fashion—for examine, on firm size or again on cash flow, depending on which variables drive heterogeneity in treatment effects—creating a second level in the tree. This partitioning process continues until reaching final partitions, within which firms are alike on both treatment effects and covariates. (See the bottom panel of Figure 1 for an example of partitioned data.)

To detail the benefits of causal forest, we compare its tree-based matching against several other matching estimators, including coarsened exact matching. In coarsened exact matching, a researcher "coarsens" or discretizes variables, effectively creating a grid. If treatment and control observations fall within the same cell in the grid, they are matched (middle panel of Figure 1). We discuss four important differences between coarsened exact matching and causal forest's tree-based matching:

1. The splitting procedure in causal forest is designed to maximize heterogeneity, but simultaneously creates partitions alike on covariates. In coarsened exact matching, the goal is to match on covariates. Causal forest achieves a similar outcome via a different process.
2. Causal forest is data-driven. In coarsened exact matching, the researcher determines the covariates and coarsening process. In causal forest, an algorithm identifies comparison sets.
3. Causal forest estimates a series of trees, each of which provides "good" matches. The final step in a forest estimation aggregates across trees to calculate for each observation how likely that observation is to "match" with each other observation. Observation-level treatment effects are calculated as weighted averages, using these matching weights, of the differences between the outcome of each observation and its matches. In coarsened exact matching, the researcher attempts to identify one "best" set of matches.
4. Causal trees do not discard data. In coarsened exact matching, un-matched treated and control observations are discarded. In contrast, trees are grown such that all data is retained, so treatment effects are calculated for all observations within a sample.

Several of these differences drive causal forest's ability to better match data and thereby minimize selection bias. Data-driven matching is more likely to result in similar treatment and control groups than researcher-determined matching, unless the researcher is omniscient. Calculating averages of good weights allows for a match between an observation and a control comprising of weighted observations. Intuitively, this process finds a best match that can consist of an observation that does not physically exist in the dataset.

The second contribution of our study is to use Monte Carlo simulations to empirically compare causal forest and RDD. We analyze five different sets of simulations, each for three different finite sample sizes (1,000, 5,000, and 10,000). Although Wager and Athey (2018) show that observation-level causal forest estimates are asymptotically unbiased, causal forest's finite sample performance is largely unexplored. Thus, in addition to a comparison against RDD, these simulations also provide an idea of how causal forest performs in relatively small datasets.

First, we compare treatment effects RDD and causal forest recover in a simulated dataset with an outcome variable, a binary treatment with a homogeneous effect, and three covariates. We do not incorporate any source of latent variable bias or heterogeneity. In this simulation, both causal forest and RDD recover treatment effects with near-zero average bias—the difference between the simulated "true" treatment effect and the recovered treatment effect, scaled by the true effect. However, causal forest's estimates have much lower variance because causal forest analyzes a whole sample whereas RDD focuses only on the threshold. Although causal forest performs better than RDD, this setup is not a "fair fight"—no researcher would pay the costs associated with RDD if selection bias were not a concern—but provides a useful baseline against which to compare later simulation results.

Next, we incorporate into our data a heterogeneous treatment effect. Whereas causal forest continues to recover low-bias treatment effects, incorporating heterogeneity into treatment effects biases RDD estimates. We find that the higher the heterogeneity, the greater is the bias in RDD treatment effects. These simulations show that causal forest recovers heterogeneous treatment effects with much lower variance and average bias than RDD.

Our third set of simulations incorporates latent variables with varying levels of correlation with included covariates. In these simulations, we purposefully break an assumption required for unbiased causal forest treatment effects: the potential outcome within each final partition must be uncorrelated with treatment status (i.e., unconfoundedness). Unsurprisingly, RDD continues to recover treatment effects with zero bias, as RDD is designed to control for selection bias created by latent variables. Also unsurprisingly, the average bias in causal forest estimates is higher in these simulations than in the baseline simulations. However, across these specifications, causal forest's average bias is relatively small. Because the variance of RDD's estimates remains high, causal forest's performance *relative* to RDD is effectively unchanged from the baseline simulations. Approximately one-third of RDD estimates are more biased than 99% of causal forest estimates in both the baseline and in all latency specifications.

These results showing that causal forest bests RDD as an estimator even when a latent variable is present may be surprising. However, they highlight an important drawback of RDD that is commonly overlooked. While researchers use RDD to reduce selection bias, the benefit of eliminating bias only comes if averaging over treatment effects from *repeated estimations*. Frequently, in empirical corporate finance, we do not design RDDs that can be repeated across groups or cohorts, like an RDD using a test score as a cutoff. Rather, we find RDDs "in the wild" and study one accompanying sample. For example, our application analysis is limited to the available sample of technical defaults. As such, we are at the mercy of RDD's high variance. Furthermore, unlike in these simulations, as researchers we never see the "true" treatment effect and so never know whether we fall in the one-third of RDDs that perform worse than almost all causal forest estimations. Given the "one draw" nature of RDDs in corporate finance, causal forest's high precision is a strong point in its favor.

Keeping in mind that, in practice, many RDDs are not designed by researchers, we next incorporate into our data non-random assignment of treatment status at the threshold. A valid RDD setup requires crossing a threshold to have an economically meaningful benefit (or consequence) for firms but that firms are either unwilling or unable to cross (or avoid crossing) that threshold. These simulations are important because we should understand how robust RDD estimates are to violations of this unprovable assumption, particularly for "wild" RDDs. We find that even modest levels of non-random assortment in as small as 5% of the sample induce large bias in RDD estimates. Increasing the percentage of non-random observations greatly increases this bias. Thus, whereas causal forest is relatively robust to latent variables, in comparison, RDD is very sensitive to even minor violations of its assumption regarding unobservables.

These simulations, in which firms can manipulate treatment status, highlight a benefit of trees that make causal forest fairly robust to omitted variables. The progressive partitioning of trees creates a hierarchy in covariates, in which some covariates matter more to tree growth because they are better predictors of treatment effects; trees will partition on these covariates first. Other covariates predict treatment effects less well, so the algorithm partitions on these variables towards the bottom of the tree, if at all. This hierarchy matters because a latent variable cannot bias treatment effects if, were the variable included in the estimation, it would not change assortment of observations to final partitions. Thus, in a causal forest estimation, latent variables that do not bias treatment effects can exist. If a latent variable is not sufficiently correlated with treatment status and its inclusion would not change the calculated weights and, thus, treatment effects, the variable does not bias treatment effects.

If a latent variable is strongly correlated with already-included covariates, the variable is redundant in the causal forest estimation, as information from the latent variable is already captured by the already-included covariates. Likewise, the variable does not bias treatment effects. In these simulations, non-random assortment at the threshold can be thought of as a latent variable (e.g., "ability to avoid treatment") that is not a "relevant" enough omitted variable to bias causal forest estimates.

In our final set of Monte Carlos, we combine conditions from earlier simulations, including both a latent variable and non-random assignment of treatment. A common concern regarding evidence from simulations is that the data generating process in simulations is simpler than for real data. How, then, can we be certain that conclusions from our simulations extend to real-world conditions? These simulations show that, as data complexity increases, causal forest's performance *improves* because causal forest is a non-parametric, machine-learning based estimator. In contrast, RDD's performance remains the same or worsens as data complexity increases because RDD is so dependent on rigid assumptions.

The third and final major contribution of our study is an example application of causal forest. In this application, we re-visit the effects of technical default on firm investment, which a number of studies recover using RDD.[4] There are several *ex ante* reasons causal forest provides benefits over RDD in this setting. First, there likely is substantial heterogeneity in the effects of default on investment; related literature finds a variety of creditor reactions to technical default (Gopalakrishnan and Parkash, 1995; Chen and Wei, 1993; Beneish and Press, 1993; Griffin, Nini, and Smith, 2019). Second, at least some firms likely can avoid default. Dichev and Skinner (2002) show that firms bunch just above default thresholds, and a survey of managers report that they would forgo profitable investment to avoid default (Graham, Harvey, and Rajgopal, 2005). Our Monte Carlos show that both heterogeneity and non-random assortment bias RDD estimates. Thus, causal forest estimates are likely less biased than RDD estimates in this setting, and causal forest can recover heterogeneity in effects that previous studies have overlooked.

In a sample of quarterly observations from 1994 to 2017, we find only a modest effect of technical default on investment (capital expenditures scaled by lagged PPE). Our average observation-level treatment effect is only -0.24% (t-statistic of -1.32). Covariates most important to treatment effects include: firm size (total assets), cash flow, and Macro Q. The heterogeneity in our results shows that technical default affects firms when default leads to a binding constraint in small firms with few outside borrowing options but profitable investment opportunities. However, as these firms are relatively sparse in the sample, we

[4]These studies include: Chava and Roberts (2008), Nini et al. (2009), Demiroglu and James (2010), and Falato and Liang (2016), among others.

find a small average treatment effect of default on investment.

Our causal forest results suggest a lower effect of default on investment than previous literature documents. RDD recovers local average treatment effects of default on investment as high as -1.5% to -2% (Chava and Roberts, 2008; Nini, Smith, and Sufi, 2009; Demiroglu and James, 2010; Falato and Liang, 2016). We consider a number of possible explanations for this difference in Section 7.6, but note here that an important covariate that determines treatment effects is cash flow. Firms with negative cash flow, on average, are more affected by default. Because firms are more likely to default if they are in more precarious financial positions, an RDD setup that focuses on firms at the threshold also over-samples firms that we find to have greater-than-average treatment effects. The difference between our results and RDD results highlights the consequences of relying on techniques that recover results with limited extendibility and the importance of techniques like causal forest that can disentangle such complicated correlations within the data.

Our causal forest results add to the literature on investment, financing, and the role of debt covenants. A key question in this literature is whether there are meaningful consequences to technical default. The implications of this question extend beyond the finance literature, as a series of accounting studies examine whether firms manipulate the accounting variables on which debt covenant thresholds are based because manipulation is only expected if defaults affect firms.[5] Early studies find, on average, no measurable effects of technical default on firms.[6] These findings seem at odds with those in Graham, Harvey, and Rajgopal (2005) and Dichev and Skinner (2002), who suggest that at least some firms successfully avoid default. The RDD results from Chava and Roberts (2008), among others, that show large real effects of default on firms can be taken to resolve this tension in the literature.

In contrast, our results show heterogeneity in treatment effects is the key to resolving disparate findings in this literature. We find that a relatively small group of firms are adversely affected by default: firms with investment opportunities but insufficient internal funding. Additionally, only a small portion of a sample clustering above the default threshold is sufficient to conclude manipulation exists. Our results show, simultaneously, bunching at the threshold and, on average, an insignificant effect of technical default on investment. Thus, we conclude that there is no conflict between early results in this literature if heterogeneity of effects is recognized. Our study highlights the importance of a technique like causal forest designed to recover heterogeneous effects to this area of study.

We join a growing literature re-examining the role of pseudo-randomized settings in em-

[5] This literature includes Watts and Zimmerman (1986), DeFond and Jiambalvo (1994), Sweeney (1994), Bartov (1993), Beneish and Press (1993), and Haw, Jung, and Lilien (1991). More recent literature in the area includes Franz, HassabElnaby, and Lobo (2014), among others.

[6] See, for example, discussion in Beneish and Press (1995) and Ertan and Karolyi (2016).

pirical research. Baker, Larcker, and Wang (2021) look broadly at DD analysis in corporate finance and accounting and shows how common implementations of the technique can cause biased results. Karpoff and Wittry (2018) examine and identify pitfalls for a specific DD setup, state-level anti-takeover law changes. In this paper, we both highlight general weaknesses with RDD and in a specific RDD setup that treats technical default as exogenous. Most importantly, however, we show that causal forest is a valuable addition to our econometrics toolkit that can ameliorate selection bias, a common source of endogeneity, and recover heterogeneous, plausibly unbiased, and extendable treatment effects that enhance our understanding of the behavior of firms.

2 Causal forest primer

For many economic questions, we are interested in whether a significant effect exists *and* what drives variation in that effect. Unfortunately, however, exploring treatment effect heterogeneity using traditional matching or regression methods is challenging. In Section 2.1, we detail the limitations of these estimators in recovering treatment effect heterogeneity. These limitations arise from the inability of traditional techniques to handle problems with very high dimensionality and, consequently, to determine which of a number of correlated covariates drives variation in treatment effects.

Causal forest, a machine-learning technique, is designed to overcome these limitations and recover heterogeneity in effects. Causal forests comprise trees, which handle higher dimensionality well. In Section 2.2, we describe the intuition underlying trees. In Section 2.3, we detail how causal forest combines results across many trees to measure observation-level, heterogeneous treatment effects. Section 2.4 provides one additional technical detail on forest calculations. In Section 2.5, we discuss how causal forest results provide several types of conditional average treatment effects (CATEs), or average treatment effects conditional on the characteristics of observations, that show how individual covariates drive treatment effects. Through these CATEs, causal forests provide more complete and nuanced answers to economic questions than are available using our traditional econometrics toolkit.

2.1 Measuring heterogeneity without machine learning

If our goal is to recover the treatment effect of binary W_i on outcome Y_i using OLS, we estimate:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 X_i + \epsilon, \tag{1}$$

where X_i is an explanatory variable and ϵ is i.i.d. error. To estimate heterogeneity in the treatment effect, or the effect of W_i on Y_i conditional on X_i, we add an interaction term between W_i and X_i:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 X_i + \beta_3 W_i X_i + \epsilon. \tag{2}$$

With only one covariate, measuring conditional effect $\beta_1 + \beta_3 X_i$ is fairly straightforward.

However, as the number of covariates rises, capturing and understanding effect heterogeneity becomes increasingly challenging. To estimate each additional covariate's conditional effect, we must interact that covariate with the treatment and each already-included covariate. This process gets out of hand quickly. For example, estimating conditional effects for five covariates involves estimating coefficients for 64 terms. This problem is compounded if we recognize that relations between at least some covariates are likely nonlinear and add polynomials and accompanying interactions into the mix. While it is likely that only some of these interactions are important to explaining Y_i, a traditional OLS regression is unable to automatically put more weight on more important covariates. At its crux, challenges measuring conditional effects using OLS are driven by OLS' inability to handle high dimensionality.

Constrained by this weakness of OLS, corporate finance researchers typically examine heterogeneity in two fairly limited ways. First, we add one interaction to an estimation, for example between a continuous explanatory variable (e.g., firm size) and the treatment variable. Second, we discretize a continuous variable, for example creating binary variables for firm size terciles, and estimate an OLS specification separately for firms in the smallest and largest terciles. These approaches are similar; estimating an effect for firms in the smallest size tercile subsample is equivalent to an estimation with the whole sample in which the tercile indicator is interacted with treatment and all other covariates.

Although these techniques provide some information on how effects vary across firms, they have two major flaws. First, these techniques are unable to control for correlations between the interacted variable or the variable defining the sample split and other covariates. Consequently, we may find higher average treatment effects for small firms but are unable to disentangle whether this heterogeneity is driven by firm size or the myriad of covariates correlated with size (e.g., financial constraints or investment opportunities). Second, these techniques require subjective identification of the interacted covariate or partitioning covariate and subsample partitions. In consequence, we only find heterogeneity in effects for which we are explicitly looking. These flaws mean that using OLS to understand heterogeneity can result in important sources of heterogeneity being misidentified or overlooked entirely.

2.2 Introducing trees

Trees provide benefits over traditional regression and matching methods in high-dimensional estimations. To understand the intuition underlying trees, it is helpful to recast Equation 1 in a matching framework:

$$\hat{\beta}_1 = \frac{1}{n}\sum_{n=1}^{n}[Y(W_i = 1|X_i) - Y(W_i = 0|X_i)]. \tag{3}$$

Using OLS to recover $\hat{\beta}_1$ in Equation 1 is effectively the same as using all observations for which $W_i = 0$ as control or comparison observations for all observations for which $W_i = 1$ (treated observations). If desired, researchers can use matching techniques to define alternate treatment and control sets. Using any matching estimator, $\hat{\beta}_1$ is an average of treatment effects estimated as the difference in Y_i between treated and control observations within comparisons sets. Following Wager and Athey (2018), we use τ to indicate treatment effects calculated as $Y_i(1) - Y_i(0)$ within comparison sets.

To build intuition for causal forest as a matching estimator, Figure 1 compares causal forest to two commonly-used matching estimators in corporate finance: Barber and Lyon (1996) matching (top panel) and coarsened exact matching (Iacus, King, and Porro, 2011) (middle panel). In addition to Barber and Lyon (1996), which historically has been widely used in corporate finance research, we discuss coarsened exact matching because it is a more technically advanced method on which an increasing number of studies rely.[7] However, the intuition we provide here extends to a number of matching estimators.

Figure 1 provides a visual comparison of the subsamples created by each technique using simulated data. In Barber and Lyon (1996) style matching, treated observations (triangles) are first matched with control observations (open circles) within the same industry. The figure shows observations plotted along five discrete industry categories on the y-axis. Then, control observations are limited to those within bands, typically 90% to 110% or 70% to 130%, of the treated firm's size or ROA, depending on the application. The figure plots size on the x-axis and these bands in parentheses. The middle panel of Figure 1 shows an example of coarsened exact matching (CEM) in which researchers "coarsen" or discretize data into bins. This process creates a grid wherein treated and control observations are matched if they are located within a shared cell of the grid. In both techniques, the researcher then has discretion on how to weight and compare observations within these bands or bins. Additionally, data outside of these comparison sets are discarded, as are unmatched treated observations.

[7] For finance applications, see: Balsmeier, Fleming, and Manso (2017), Galasso and Simcoe (2011), Eaton, Howell, and Yannelis (2020), and Heath and Mace (2020), among others.

Like Barber and Lyon (1996) and coarsened exact matching, causal forest "bins" and weights data to recover treatment effects. However, causal forest differs from these techniques in at least four ways. First, causal forest calculates weights to maximize heterogeneity in treatment effects. In contrast, the goal of Barber and Lyon (1996) and coarsened exact matching is to weight data to create treatment and control sets that are similar along selected covariates. Comparison sets within a causal forest are alike on covariates, but this similarity is achieved via a different process.

Second, causal forest is data driven. In Barber and Lyon (1996) and coarsened exact matching, researchers specify bins and do not ex ante know whether the covariates defining the bins are important to treatment. Within a forest, trees are grown using an algorithm that bins similar observations together (bottom panel of Figure 1), creating balance only for those variables important to treatment effects. Additionally, trees allow researchers to control for far more variables than Barber and Lyon (1996) two-by-two style matching.

Third, trees do not discard data. Because of the way in which trees are grown, all observations are included in a partition. Additionally, trees do not create partitions with fewer than a set number of treated and control observations, so treatment effects are calculated for all observations within a sample.

Finally, causal forests comprise a series of trees, each of which identifies reasonably "good" matches for each observation. The forest then aggregates over these matches to calculate weights. Consequently, the comparison set for observation i is a weighted set of observations across many estimations rather than just one set of observations. In contrast, the goal of Barber and Lyon (1996) and coarsened exact matching is to calculate one set of "best" matches (and weights) once. We discuss how differences between trees come to be and this aggregation procedure in Section 2.3. We now describe in greater detail how individual trees are grown.

Trees group like observations by progressively ("recursively") splitting samples into finer and finer subsamples. In the bottom panel of Figure 1, the first split in the data is on age. The two resulting portions of data, above and below the horizontal dashed line that intersects the y-axis, are known as nodes. Because these nodes are created by a split in the data, they are known as "children nodes". Each child node can then be split again into additional, smaller children nodes (and thus become a "parent" node, or a node that contains smaller nodes within it). If a node does not have any children, it is known as a "leaf", or a final partition of the data. The tree in Figure 1 has seven leaves. In a causal tree, treatment effects are calculated by comparing treatment and control observations within leaves.

Causal trees are built to maximize the heterogeneity in treatment effect predictions. At each parent node (P), the causal forest algorithm partitions the data into two children nodes

(C_1 and C_2) such that the difference in treatment effect estimates between the children nodes is maximized.

As performing this computation within each node can be expensive, the algorithm uses a gradient descent approach to split parent nodes into children and estimates the treatment effect in the parent node as:

$$\hat{\tau}_P = \frac{1}{n_P} A_P^{-1} \sum_{\{i:X_i \in P\}} \left[\left(W_i - \bar{W}_P \right) \left(Y_i - \bar{Y}_P \right) \right], \tag{4}$$

in which $\bar{Y}_P$ and $\bar{W}_P$ are the average values of Y_i and W_i in node P, respectively, and

$$A_P = \frac{1}{|\{i : X_i \in P\}|} \sum_{\{i:X_i \in P\}} \left(W_i - \bar{W}_P \right)^2. \tag{5}$$

The influence function for observation i in node P is:

$$\rho_{i,P} = A_P^{-1} \left(W_i - \bar{W}_P \right) \left(Y_i - \bar{Y}_P - \left(W_i - \bar{W}_P \right) \hat{\tau}_P \right). \tag{6}$$

In the gradient descent approach, we approximate the treatment effect in the child nodes of a potential split as:

$$\hat{\tau}_{C_j} = \hat{\tau}_P - \frac{1}{|\{i : X_i \in C_j\}|} \sum_{\{i:X_i \in C_j\}} \rho_{i,P}. \tag{7}$$

Intuitively, the average influence function for observations assigned to node C_j captures how much the treatment effect in the child node is likely to differ from the treatment effect estimate in the parent node. As $\hat{\tau}_P$ is constant no matter how the data are split to create the children nodes, the squared difference in treatment effects between C_1 and C_2 for a potential split is:

$$\widetilde{\Delta} \left(C_1, C_2 \right) = \sum_{j=1}^{2} \frac{1}{|\{i : X_i \in C_j\}|} \left(\sum_{\{i:X_i \in C_j\}} \rho_{i,P} \right)^2. \tag{8}$$

Causal forest chooses the split that maximizes (8), which is computationally less expensive than inverting a matrix to evaluate each potential split.[8]

This approach is fairly "greedy" in how quickly it maximizes heterogeneity. As a result, a potential problem that can arise is that a split creates a child node with too few treatment or control observations for the estimate to be informative (i.e., an "unbalanced" split). The causal forest algorithm relies on several parameters to control balance:[9] the minimum number

[8]For readers more familiar with regression trees, Equation 8 is the equivalent of minimizing mean squared error (MSE) in a regression tree. For more discussion on this, see Wager and Athey (2018).

[9]We discuss two here. There is a third parameter, the imbalance penalty, that is available in the `grf` $\mathbb{R}$

of both treatment and control observations within a node (k; note that the minimum number of observations within a leaf is then $2k$) and the minimum proportion of the parent node that must be contained within each child node. Together, these parameters control imbalance created by splitting and restrict how "deep" a tree can grow. Parameterization of a causal forest estimation is sample-specific, so in internet appendix Section C.1, we list parameters and discuss parameterization in the context of our application.

Panel C of Figure 1 shows how this splitting creates comparison groups of observations. Each subsequent split results in subsets that are more and more alike in treatment effects. (This statement must be true because the first split creates subsamples that are the most *dissimilar* in treatment effects.) Because partitions are defined using covariates, subsamples are also more and more alike on covariates (specifically, covariates that determine heterogeneity in effects). This partitioning process allows trees to capture heterogeneity without encountering the dimensionality limitations we discuss in Section 2.1.

2.3 Combining trees into forests

A causal forest contains $1, \ldots, B$ trees, each grown according to the following procedure:

1. Draw a subset of data, S_b.
2. Divide S_b into two equal halves, S^{tr} and S^{est}.
3. Using only a subset of covariates, outcomes (Y_i) from S^{tr}, and treatment (W_i) and covariates (X_i) from both S^{tr} and S^{est}, grow tree b.
4. Use the partitioning rules that define tree b to partition S^{est}.

This procedure grows "honest" trees. An honest tree does not use the outcomes Y_i from S^{est} to determine partitioning. Honesty is key to ensuring trees do not overfit the data (and eliminates the need for any pruning) but requires "wasting" half of the data when growing each tree. However, because this sampling procedure is repeated for each tree within the forest, observations cycle in and out of S^{str} and S^{tr} across trees and mean squared error of estimates is not compromised via this honesty procedure.

Once B trees are grown, an aggregation across these trees determines how frequently the i^{th} observation falls within the same leaf as an observation characterized by $L_b(X)$:

$$\alpha_i = \frac{1}{B} \sum_{i=1}^{B} \frac{1(X_i \in L_b(X), i \in S_b)}{i : X_i \in L_b(X), i \in S_b}. \tag{9}$$

package (Tibshirani, Athey, Friedberg, Hadad, Hirshberg, Miner, Sverdrup, Wager, and Wright, 2020). It has not proven important in our testing, so, for brevity, we omit a discussion of this parameter. The package authors note that it is an experimental parameter.

The bottom panel of Figure 2 shows these weights for an example forest grown with simulated data.[10] To create this figure, we simulate a binary outcome variable and two continuous covariates (*variable 1* and *variable 2* shown on the x- and y-axes, respectively) and estimate a forest with 1,000 trees. The top left panel plots this data, including: treated observations (triangles), control observations (open circles), and an example point (blue solid square) for which we calculate α_i.

The top right and middle two plots show three example trees. The splitting rules for each tree are indicated with dashed lines. Points that are contained within the same leaf as the example point are larger and darker.

The bottom plot in Figure 2 shows α_i. Larger, darker points are contained within more leaves with the example point and have a larger α_i. For example, the example point and the dark triangle to its bottom left share many leaves with the example point, including in all three trees shown in Figure 2. In contrast, the pair of triangles to the upper left of the example point are lighter than that single triangle and so are contained in fewer leaves with the example point.

Observation-level treatment effects are calculated with α_i:

$$\hat{\tau}(x) = \left(\sum_{i=1}^{n} \alpha_i(x)(W_i - \overline{W}_\alpha)^2\right)^{-1} \sum_{i=1}^{n} \alpha_i(x)(W_i - \overline{W}_\alpha)(Y_i - \overline{Y}_\alpha), \tag{10}$$

where $\overline{W}_\alpha = \sum \alpha_i(x) W_i$ and $\overline{Y}_\alpha = \sum \alpha_i(x) Y_i$. Intuitively, $\hat{\tau}(x)$ is a weighted least squares estimator, with weights equal to observation i's α_i. In Figure 2, the observation-level heterogeneous treatment effect (HTE) for the example point will be more similar to the HTE for the darker, closer triangle than the HTEs for the lighter, farther pair of triangles.

Wager and Athey (2018) show that HTEs from a causal forest containing honest trees are asymptotically unbiased under two assumptions: unconfoundedness, $Y_i(1), Y_i(0) \perp W_i | X_i$, and overlap, $0 < Pr(W_i = 1 | X_i = x) < 1 \; \forall \; x$. Unconfoundedness requires that, once the set of covariates is controlled for, the potential outcome of treated and control observations is orthogonal to treatment status. Stated otherwise, within each leaf, potential outcomes for treated and control observations must be uncorrelated with the treatment status. Overlap means that there exist sufficiently similar control observations against which to compare treated observations.

[10] We are indebted to Athey, Tibshirani, and Wager (2019) for inspiring this figure and related discussion.

2.4 A note on centering

Our discussion of the forest procedure in Section 2.3 omits, for simplicity, one step that occurs *before* the forest estimation in which the data are "centered". Locally centering the data involves differencing out the effect of the covariates, X_i, on both the outcome (Y_i) and treatment (W_i). We estimate regression trees for both the outcome and treatment separately and calculate centered $\widetilde{Y}_i = Y_i - \hat{y}_i^{(-i)}(X_i)$ and $\widetilde{W}_i = W_i - \hat{w}_i^{(-i)}(X_i)$ where $\hat{y}_i^{(-i)}(X_i)$ and $\hat{w}_i^{(-i)}(X_i)$ are leave-one-out estimates of marginal expectations, omitting the i^{th} observation. We estimate the forest according to the procedure outlined in Section 2.3 using centered $\widetilde{Y}_i$ and $\widetilde{W}_i$ rather than Y_i and W_i. Athey, Tibshirani, and Wager (2019) note that this procedure follows Robinson (1988) and improves the performance of forests by making the estimator more robust to confounding effects. For additional details, see Section 6.1.1 of Athey, Tibshirani, and Wager (2019).

2.5 Conditional average treatment effects from a causal forest

A causal forest estimation provides two types of results capturing heterogeneity in treatment effects. The first set of results are observation-level, heterogeneous treatment effects, which can be summarized in density plots or sample and subsample averages. In our application, we estimate the average treatment effect over all observations in the sample (ATE), subsample averages for treated and control firms separately (ATT and ATC, respectively), and an average over the whole sample with more weight put on leaves with better overlap (ATO).[11] We also calculate ATE, ATT, ATC, and ATO within subsamples of interest, for example small firms and firms with negative cash flows. These averages are conditional average treatment effects. If treated and control samples have different characteristics that determine treatment effects, ATT and ATC differ because averages of treatment effects reflect subsample characteristics. All averages are calculated using augmented inverse propensity weighting (AIPW), as is standard in the literature (Robinson, 1988). For additional details on AIPW, please see our technical internet appendix.

The second set of results from a causal forest estimation is a mapping between covariates and treatment effects. Trees grow by partitioning data on covariates that predict variation in treatment effects, so a forest effectively provides a model of how covariates determine

[11] In this section, we discuss summaries and analyses of causal forest results that we find informative in our application. The reader may expect a discussion of result presentations "common in the literature", rather than a focus on our analysis. However, because we are the first to use causal forest in observational data and corporate finance, we decide how best to summarize and describe our results. Thus, this section does not discuss common practices in the literature because this literature does not yet exist. Subsequent studies may choose to summarize or analyze causal forest results differently, and this choice may be driven in part by the objectives of the study and research question.

treatment effects. The bottom panel of Figure 2 shows an example of this mapping. Because larger, darker points are weighted more heavily in determining the treatment effect for the example point (blue square), this area of the data has similar treatment effects *and* is similar on the two covariates on the axes. We can use this mapping to explore heterogeneity in effects with counterfactuals and ask, for a firm with these characteristics (these covariate values), what is the predicted treatment effect from our forest?

In our application, we first use this map to plot conditional average treatment effects (CATEs) for each covariate. To estimate these CATEs, we hold all covariates equal to their values for the median firm and vary one covariate (e.g., cash flows) over its support. We then use forest mapping to estimate hypothetical treatment effects across these possible values for cash flows. The estimated treatment effects show which firms are more affected by treatment *and* important inflection points in treatment effects. For example, we find an important inflection point at zero cash flows; firms with positive cash flows have near-zero treatment effects of default on investment, while firms with negative cash flows have increasingly large, negative treatment effects. In contrast, if a covariate is not important to determining treatment effects, the CATE is flat over the covariate's support. Thus, these CATEs provide a great deal more information on treatment effect heterogeneity than sample splits via OLS we discuss in Section 2.1.

In a second set of tests, we use a twist on this procedure to understand which covariates are the most important to predicting firms' treatment effects. Intuitively, trees split on covariates that are better predictors of variation in treatment effects closer to the top of trees. These covariates are the most important to determining into which leaves firms fall. Because treatment effects are calculated based on assortment to leaves, if these covariates are altered or removed from the sample, calculated treatment effects would vary greatly. Accordingly, we calculate relative covariates' importance to treatment effects using this procedure:

1. Take one covariate and randomly distribute the covariate across firms in the sample (i.e., permute the covariate) while holding all other covariates constant.
2. Use the forest mapping with the permuted data created in step 1 to re-estimate treatment effects.
3. Calculate the difference between these new treatment effects and the treatment effects calculated with un-permuted data.
4. Repeat steps 1 through 3 to determine, on average, how much permuting the covariate changes treatment effects.

Repeating this procedure for each covariate in turn, we can rank covariates by these averages. Covariates with higher average differences between treatment effects calculated with permuted and un-permuted data are relatively more important to tree growth within the forest and to the determination of treatment effects. Relative differences between the sizes of these averages also provides information. Covariates with very small averages have almost no role in determining treatment effects.

This permutation process is a powerful tool in understanding treatment effect heterogeneity because it effectively severs correlations between covariates. In Section 2.1, we discuss how OLS estimations on subsamples of data are limited in their ability to recover heterogeneity because these tests are unable to disentangle between the effects of multiple correlated covariates. In contrast, this permutation process can show that only one of several correlated covariates is important to determining treatment effects. In this scenario, significant differences between treatment effects in subsamples defined using the other correlated covariates are likely driven by this one important covariate. Alternatively, the permutation process may show that all the correlated covariates matter to treatment effects. This disentangling between covariates is not possible without a technique like causal forest that provides a model of each covariates' individual effect on treatment effects.

3 Causal forest, selection bias, and identification

3.1 Relative benefits of causal forest

Causal forest controls for selection bias well because of the flexible, data-driven way in which trees calculate matched comparison sets and the repeated sampling process that provides best weights rather than one set of best matches. Returning to Equation 3 (with X_i dropped for simplicity), an average treatment effect (ATE) calculated as:

$$\begin{aligned} E(Y_i|W_i = 1) - E(Y_i|W_i = 0) = \{E[Y_i(1)|W_i = 1] - E[Y_i(0)|W_i = 1]\} \\ + \{E[Y_i(0)|W_i = 1] - E[Y_i(0)|W_i = 0]\}, \end{aligned} \tag{11}$$

which is the average treatment effect of the treated (ATT) plus selection bias. ATE estimated via matching is biased if there are any underlying differences between the treatment and control groups (i.e., $E[Y_i(0)|W_i = 1] - E[Y_i(0)|W_i = 0] \neq 0$). However, selection bias can be reduced if better matches, or matches with more similar treatment and control observations, are obtained.

There are four theoretical reasons why causal forest is more robust to selection bias than OLS or traditional matching methods:

1. Because of the progressive (recursive) way in which trees split data, trees automati-

cally consider as covariates all interactions and polynomials of included covariates. In contrast, including a covariate in an OLS regression does not rule out the square of that covariate as a latent variable.

2. Causal forest matching is data-driven, which likely provides better matches of treatment and control observations than researcher-led matching.

3. The centering procedure we detail in Section 2.4 renders forest estimates more robust to confounders (Athey, Tibshirani, and Wager, 2019). Trees create leaves by predicting treatment effects, but causal forest assumes random treatment within leaves. Potentially, then there can exist a covariate correlated with treatment and potential outcome that is unrelated to predicted treatment effects and so not controlled for in the partitioning process. In this scenario, unconfoundedness could be violated. However, considering the centering procedure, unconfoundedness assumes that: $\widetilde{Y}_i(1), \widetilde{Y}_i(0) \perp \widetilde{W}_i | X_i$. This assumption means that *unpredicted* potential outcome is uncorrelated with *unpredicted* treatment, where predicted potential outcome and treatment are predicted with regression trees including X_i as covariates. Thus, even though leaves are grown on predicted treatment effects, any possible correlations between potential outcome and covariates and treatment status and covariates are previously controlled for with the centering procedure.

4. The inherent hierarchy within trees creates a "hurdle" for latent variables; if a latent variable is not sufficiently "relevant" to an estimation, it does not bias estimates. Say we omit a variable from our set of covariates in a causal forest estimation. This latent variable only biases treatment effects if, were it included in the estimation, it changes assortment of observations to leaves, within which treatment effects are calculated. Thus, in a causal forest estimation, latent variables that do not bias treatment effects can exist. If a latent variable is not sufficiently correlated with treatment status and its inclusion would not change treatment effect calculation, the variable does not bias treatment effects. Alternatively, if a latent variable is strongly correlated with already-included covariates, the variable is redundant in the causal forest estimation, as information from the latent variable is already captured by the included covariates. Likewise, the variable does not bias treatment effects. (This is the same intuition underlying the discussion in Section 2.5 regarding ranking covariates according to their importance in growing forests.) This property of causal forest means that causal forest requires weaker assumptions about unobservables than OLS, as *any* correlated latent variable biases treatment effects recovered with OLS.

3.2 Monte Carlo simulations

Handling omitted variable bias better than OLS or traditional matching methods does not mean, however, that causal forest estimates are bias-free. Furthermore, these theoretical benefits do not mean that causal forest can recover treatment effects with sufficiently low bias to rival RDD. To demonstrate causal forest's success at matching on observables to control for bias from unobservables, and its relative performance against RDD, we use a series of Monte Carlo experiments. (The only way to directly compare estimators on bias is using simulated data, as bias is the difference between recovered and "true" treatment effect. Outside of simulations, we can never observe the true treatment effect and so cannot calculate bias.)

To test how well causal forest and RDD perform in the kind of panel data common in corporate finance, we simulate a dataset similar to our application data, in which the outcome (dependent) variable is investment, treatment variable is a binary indicator for default status, and covariates include: cash flow, firm size (total assets), and Macro Q. The structure of the simulated data mimics the average and variance of each covariate, as well as their pairwise correlations. The covariates have both linear and quadratic relations with the outcome variable. Later in this section, we discuss alternate functional forms and the importance of functional form to the performance of each estimator. Further details on the data simulation process are in our technical internet appendix (Section D).

Because causal forest is a relatively new estimator, research has not yet explored its finite sample properties. While Wager and Athey (2018) show that causal forest estimates are asymptotically unbiased, we do not know how large of a sample is required for causal forest to perform well. Thus, we estimate simulations in finite samples, 1,000, 5,000, and 10,000 observations. For comparison, the sample size for our application is 32,530.

The top panel of Figure 3 shows a baseline comparison of causal forest (left, grey distributions) and RDD (right, blue distributions) across three sample sizes on the x-axis. For each sample size, we simulate 10,000 sets of data with no source of endogeneity and a homogeneous treatment effect. For each of the 10,000 simulated datasets, we estimate a causal forest and RDD and calculate bias as the difference between the estimated and true treatment effect scaled by the true treatment effect. Figure 3 plots the densities of bias from these estimations. The boxes within the densities are interquartile ranges; the horizontal bar within the interquartile range is the average bias. We prefer estimators with average bias closer to zero (indicated with the horizontal solid line across all sample sizes) and smaller variance (i.e., a smaller vertical spread of the bias density).

The densities in the top panel of Figure 3 show that, in the baseline specification, both causal forest and RDD recover treatment effects with near-zero bias. However, RDD estimates have large variance. As sample size increases, variance in bias for both causal forest and RDD declines, but the variance of RDD's estimates is consistently higher than the variance of causal forest's estimates. The difference in variance between the estimators occurs because each causal forest analysis uses the whole available sample, while each RDD is limited to the Imbens and Kalyanaraman (2012) optimal bandwidth (i.e., a sample with fewer observations).[12] Consequently, in each set of simulations, causal forest has better precision than RDD.

To compare the performance of RDD and causal forest, we consider both the average and variance of bias. Researchers use RDD because of concerns about selection bias. On average, RDD estimates are bias-free. If a researcher uses, for example, a test score as a cutoff for an RDD, she can repeat the experiment across schools or cohorts, average treatment effects across estimations, and recover an average effect close to the true treatment effect. However, in corporate finance, researchers generally identify RDD setups "in the wild" rather than design an RDD experiment. This distinction matters because we typically are not afforded the opportunity to re-run "experiments". For example, in our application, our analysis is limited to the current sample of defaults available. If we get only one draw of data, we are unable to truly take advantage of the lower bias RDD provides. The variance of RDD treatment effects gives an idea of how far estimates from one draw can be from the "true" effect.

We present statistics in Table 1 to help quantify the implications of the "one draw" of data to the comparison between RDD and causal forest as estimators. Table 1 shows the percentage of RDD estimates that have greater bias than all, 99%, 98%, or 95% of causal forest estimates for each sample size. These percentages stay relatively consistent across sample sizes, echoing that both causal forest and RDD recover increasingly precise estimates as sample size increases. In our baseline simulation, approximately 45% of RDD estimates have greater bias than 95% of causal forest estimates. This statistic means that one RDD estimation in this setting has a 45% chance of performing worse than 95% of causal forest estimations. Given that we tend to only get one draw of a sample for an RDD estimation and *never* learn whether we are in that 45%, these results are not comforting.

Even more concerning, the variance of RDD estimates is so high that a portion of estimated treatment effects are *of the wrong sign*. The "true" treatment effect in our simulated data is positive, so any estimates with bias less than -100% are negative. In our baseline specification, 9.42% of RDD estimates for the 1,000 sample size are incorrectly signed. Although the RDD estimates have almost zero bias on average, the high variance means that a substantial proportion of the estimates are not particularly close to the true effect.

[12]Other bandwidths are available, but this point stands regardless of the bandwidth algorithm used.

Notably, causal forest's bias is near-zero even in sample sizes as small as 1,000 observations. Although causal forest is designed to handle large, complex datasets, the first major takeaway from these simulations is that causal forest also recovers low-bias estimates in relatively small samples.

In the middle and bottom panels of Figure 3, we introduce into our data heterogeneous treatment effects. In data for the middle panel, the treatment effect for each observation is a function of the observation's covariate values. For the bottom panel, we use the same functional form for treatment effect calculation as in the middle panel but use a multiplier to increase the variance of the treatment effect. This multiplier increases the heterogeneity in the bottom panel relative to the middle panel.

The introduction of heterogeneity into treatment effects does not affect causal forest's performance but badly biases RDD estimates. The causal forest bias densities in Panels B and C are very similar to those in Panel A. Given that causal forest is designed to recover heterogeneous treatment effects, it is unsurprising but reassuring that the introduction of heterogeneity does not affect causal forest's estimates. In contrast, average bias in RDD estimates—the vertical difference between the zero line and horizontal bar in the interquartile range box—increases moving from Panel A to Panel B to Panel C. These simulations show that RDD is not well-equipped to handle heterogeneous treatment effect estimation and that, as heterogeneity increases, bias in RDD estimates increase.

Next, we introduce into our dataset a correlated omitted variable, with low, medium, and high levels of correlation with already-included covariates. Figure 4 shows that average RDD bias is unaffected by the inclusion of a latent variable. Densities for RDD in Figure 4 are unchanged from our baseline results in Figure 3. In contrast, the latent variable increases average bias in our causal forest estimates. This result is unsurprising, as Wager and Athey (2018) show that unconfoundedness is required for the recovery of zero-bias treatment effects with causal forest. However, the average bias in these estimates is small, so small it can be best seen in the 10,000 sample size estimations. (Note that, although the average bias appears to increase as sample size increases, this visual is driven by an increase in precision tightening the distribution around the average bias, not a meaningful increase in average bias.)

There are three takeaways from this set of simulations. First, introducing a latent variable does, on average, bias causal forest estimates. However, the bias is small, approximately 4% to 6% on average across all specifications. To put that in perspective, if the true treatment effect in our estimations is 10%, these simulations recover an average treatment effect of 9.4% to 9.6%. For sample sizes greater than 10,000, many causal forest estimations recover treatment effects that are very close to the true treatment effect. Second, a highly-correlated latent variable (Panel A) induces only slightly higher bias in causal forest estimates than less-correlated latent variables (Panels B and C). These simulation results support our earlier theoretical discussion from Section 3.1 that causal forest is relatively robust to omitted variable bias. Finally, most causal forest estimations still recover higher quality (lower bias) estimates than many RDD estimations. In Table 1, there are no meaningful differences between the proportions of RDD estimates that are more biased than causal forest estimates between the baseline and latent specifications. These results mean that, even in the setting for which RDD is designed and has the greatest benefits, causal forest bests RDD as an estimation technique as strongly as in the baseline specification. Intuitively, the increased difference in bias between RDD and causal forest estimates is not strong enough to overcome the continuing difference in variance.

In our third set of simulations, we introduce the ability of firms to avoid treatment. Each firm has a randomly chosen probability to manipulate treatment status but only does so if the benefit (i.e., avoiding the treatment effect) exceeds a cost. In the low, medium, and high specifications in Figure 5, no more than 5%, 10%, and 20% of firms in the bandwidth sample avoid treatment, respectively. These probabilities are kept low to ensure we do not compromise overlap by creating treatment and control samples that are too dissimilar.

Before discussing the results from these simulations, we offer an important caveat. A researcher should not use RDD in a setting in which firms can plausibly avoid treatment. We expect that introducing the possibility of manipulation causes differences between our treatment and control groups and biases our RDD estimates. Additionally, although we continue to use the Imbens and Kalyanaraman (2012) "optimal" bandwidth calculation for our RDDs, this calculation is based on a trade-off between bias and precision that does not make sense in a setting where manipulation is present. Despite these concerns, we present these results because it is important to recognize the consequences of even small levels of manipulation to RDD's ability to recover treatment effects. Most RDDs in corporate finance rely on a researcher identifying a threshold that meaningfully affects firms while simultaneously assuming that firms do not or cannot avoid (or seek out) crossing the threshold (i.e., treatment). This assumption is unprovable, so we should understand the consequences of (presumably unwittingly) breaking this assumption.

The bias densities in Figure 5 show that the introduction of manipulation into the simulations biases RDD estimates, with increasing bias as the percentage of sample avoiding treatment increases. Even modest levels of manipulation, less than 5% of the sample in the top panel, badly bias RDD estimates because RDD focuses entirely on the sample in which manipulation occurs. These simulations show how susceptible RDD estimates are to only a relatively small proportion of observations breaking its key assumption regarding unobservables.

In contrast, causal forest estimates are not biased by the presence of manipulation. Whereas RDD focuses on a bandwidth sample, causal forest analyzes the entire sample in which the threshold comprises a small portion. Causal forest estimates are robust to this source of omitted variable bias because this latent variable affects so few observations within the forest. Consequently, "ability to manipulate" is a latent variable that is irrelevant in a causal forest estimation. Intuitively, if "ability to manipulate" were included in our forest, its inclusion would not change assortment of firms to leaves in any significant way, so its exclusion does not bias estimates. For further discussion of causal forest's relative robustness to omitted variable bias, see Section 3.1.

We use a final set of simulations to address a common complaint in studies using simulated data. One natural concern regarding evidence from simulations is how sensitive the results are to alternative data generating processes. Generally, we expect that the data generating process for simulations is too simple relative to real data. A key strength of causal forest is that causal forest performs *better* in more complex datasets because the technique is nonparametric, data-driven, and tree-based.

To demonstrate causal forest's enhanced performance in more complex data, in Figure 6, we compare causal forest and RDD in nine additional datasets combining low, medium, and high manipulation probability and correlation between a latent variable and included covariates for a sample size of 2,500. Whereas RDD estimates are badly biased in all specifications, causal forest bias declines with the introduction of higher manipulation probability into each latency specification. The improvements are small because the starting bias is small, but these estimations show how causal forest performs better in more complex data. This final set of simulations is important because we compare RDD and causal forest's recovery of treatment effects from relatively simple polynomial specifications. Thus, particularly in the simulations in Figure 4, we have stacked the deck in RDD's favor, and RDD still comes up short. If we were to use a more complex data generating process and re-examine these simulations, causal forest would be even more poised to shine.

In sum, our Monte Carlo simulations show that RDD is more susceptible to violations in its assumption regarding unobservables than is causal forest and that causal forest recovers higher-quality estimates than RDD across many samples and settings. We also show that RDD estimates are biased by heterogeneity. However, these comparisons focus only on comparing bias in average treatment effects. There are other benefits to causal forest as an estimator, including the recovery of heterogeneous, observation-level and conditional average treatment effects that we detail in Section 2.5. Because of the structure of trees, the intuition from the Monte Carlos for average treatment effects holds for conditional average treatment effects as well, since conditional average treatment effects are essentially average treatment effects in a subset of trees. Additionally, RDD estimates tend to be very sensitive to subjective researcher decisions, including functional form, bandwidth, and sample selections. In Section 7.6, we show how sensitive RDD results in our application are to these decisions and, in Section 7.3, how causal forest estimates are robust to such decisions. Considering these additional factors tips the scale even more in causal forest's favor.

4 Application setting

In this section, we detail the setting of our causal forest application: examining the heterogeneous effects of default on investment. We begin with a discussion of how covenant design leads to multiple sources of selection bias, each of which complicates the recovery of unbiased treatment effects in Section 4.1. Then, in Section 4.2, we draw on our Monte Carlo results from Section 3.2 to explain why causal forest can plausibly control for each source of selection bias in our setting.

4.1 Several sources of selection bias

Three aspects of the role and design of covenants can introduce bias into the recovery of treatment effects of default. First, covenants, which exist to protect lenders (Smith and Warner, 1979), are set via a two-way negotiation between firms and lenders (Jensen and Meckling, 1976). These negotiations result in covenant types and strictness that are correlated with firm and lender characteristics (i.e, endogenous). Demiroglu and James (2010) empirically confirm non-random assortment of covenants and strictness at origination by showing that riskier firms and firms with fewer investment opportunities receive tighter covenants.

Second, covenants are intended to transfer control rights to lenders in "bad" states. Consequently, firms in and near default are, on average, in more precarious financial positions than firms well out of default. If a researcher were to estimate the treatment effect of default on investment by merely differencing the average investment of firms in and not in default, underlying differences between firms in and not in default (i.e., selection bias) would confound the estimate.

Finally, covenant violations are costly. Following a technical default, firms reduce: capital investment (Roberts and Sufi, 2009; Nini, Smith, and Sufi, 2009, 2012), acquisition activity (Becher, Griffin, and Nini, 2020), employment (Falato and Liang, 2016), and payout (Nini, Smith, and Sufi, 2012). R&D expenditures and innovation quality are lower following default (Chava, Nanda, and Xiao, 2017; Gu, Mao, and Tian, 2017). Additionally, Roberts and Sufi

(2009) show leverage decreases after default. These costs create incentives for firms to avoid or delay violation, if possible (Dichev and Skinner, 2002). If some firms are more successful at avoiding violation than others, this creates another source of selection bias primarily located in firms at the default threshold.

Given the costs associated with default, it is unsurprising that empirical evidence suggests that firms act to avoid default. Two early studies show that firms alter real activities to avoid default: Bartov (1993), who examines the timing of income recognition, and Haw, Jung, and Lilien (1991), who study the timing of pension plan settlement. Both studies proxy for distance to covenant default with leverage and compare real activities of high leverage to low leverage firms to draw conclusions. Additionally, survey results from Graham, Harvey, and Rajgopal (2005) suggest that firms manipulate investment to ensure making earnings targets and show that managers report a benefit of making earnings targets is avoiding covenant violations. Concerningly, Graham, Harvey, and Rajgopal (2005) find the strongest evidence of managers reporting they would forgo profitable investment to avoid default in managers of firms that are weaker financially, or in firms that are more likely to be at the default threshold. Taken together, these studies show that avoidance of default via altering real activities is a real concern to identification in this setting. Any such avoidance of default, whether via accruals or real activities like investment, can cause differences in treatment and control firms at the threshold.

Additional empirical evidence from the accounting literature suggests that firms manipulate accruals and accounting choices to avoid default. Consistent with the Watts and Zimmerman (1986) hypothesis that firms should make income-increasing decisions to avoid covenant violations, DeFond and Jiambalvo (1994) and Jha (2013) show that abnormal accruals in firms preceding default. Additionally, Sweeney (1994) finds that firms make income-increasing accounting changes before default.

It is important to note that the presence of several sources of selection bias in this setting poses a challenge for researchers, but does not imply that covenants are not "working" as intended. We raise several issues in this section that are related to econometrics. While these econometrics issues are driven by the role and design of covenants, this discussion should not be taken as a critique of the role and design of covenants. Take, for example, the possibility that firms reduce investment to avoid violating a covenant. In this scenario, arguably, covenants are working as intended to ensure lenders are repaid. However, firms reducing investment in this way has serious implications for a researcher's ability to recover treatment effects of default, particularly if the researcher intends to use the default threshold as a source of "as-good-as-random" variation, like in an RDD estimation. In consequence, our job, as researchers, to recover the treatment effects of default on investment is made more difficult. However, these econometric arguments do not mean that covenants are not important or are in any way flawed.

4.2 Causal forest in this setting

Combining these institutional details with our Monte Carlo simulation results shows that: (i.) RDD is not well-suited to answer this question, and, (ii.) that causal forest can plausibly recover low- to no-bias treatment effects in this setting. The post-default waiver process introduces heterogeneity into the effects of default on investment. Additionally, empirical literature suggests that at least some firms are able to avoid default. Our simulations show that both of these features of the setting bias RDD estimates. Given RDD's relative weakness in this setting, we focus here on discussing causal forest's assumptions of overlap and unconfoundedness with regards to each source of selection bias that we list in the previous section.

Variation in firms, lenders, and covenant thresholds provides overlap in our sample. Even among firms with covenants defined using the same accounting ratio, for example, net worth covenants, there is variation in the specific ratio used as the threshold. Additionally, even if all similar firms receive the same exact default threshold, the passage of time between origination and observations in our sample provides additional variation. Consequently, there is no one covariate value, even for net worth, that wholly determines default for all firms, or even some firms most of the time. Thus, there is sufficient overlap in the sample. To alleviate concerns about overlap, we include in our results ATO estimates, which support our conclusions throughout our results.

The passage of time between origination and observations in our sample is important to both our discussion of overlap and our ability to control for the first source of selection bias in our sample. Demiroglu and James (2010) show that, at origination, firm characteristics are correlated with covenant strictness. This correlation weakens as time passes and firms deviate from their characteristics at origination. In our data, there is variation in the length of time between observations and origination, so this correlation is weaker for some observations than for others. Although we argue that the passage of time weakens this correlation such that it is not a major source of endogeneity bias in our results, we also provide empirical reports supporting this argument. We include both contemporaneous and initial distance to default (slack) in our causal forest estimates and show that these covariates are relatively unimportant to our estimation of treatment effects (see Table 4). These results, which we discuss further in Section 7.1, suggest that these latent variables are not relevant enough to our estimation to bias estimates.

The most concerning source of selection bias in our setting comes from differences between firms in and not in default. On average, firms in default are weaker financially and engage in lower investment than firms not in default. However, lower investment of firms in default can be driven by their financial positions rather than default itself.

Ex ante, there are two reasons to believe that, despite the existence of this source of selection bias, unconfoundedness is a reasonable assumption in our setting. First, the same covariates determine both propensity to receive treatment and magnitude of treatment. Firms in weaker financial positions are more likely to default. Also, according to Chen and Wei (1993), firms that are in weaker financial positions experience the greatest consequences for default. We confirm their conclusions in our sample. To bias our treatment effects, a latent variable must be sufficiently correlated with treatment propensity *and* insufficiently correlated with treatment effects. Because the set of covariates that predict both propensity to receive treatment and treatment effects are very similar, the existence of such a latent variable is improbable.

In a second argument for why unconfoundedness exists in our setting, we again can lever some degree of randomness that comes from covenant thresholds having been set in the past. Effectively, causal forest provides really good matching between firms based on current covariates. Covenant thresholds were set at some time in the past. While there is correlation between initial covenant slack and firm characteristics at the time of origination, there is still variation in covenant thresholds. Additionally, despite correlations that exist at the time of origination, firms do not grow and change identically following origination. On top of variation in initial covenant thresholds, some covenant thresholds adjust over the life of the loan. These different sources of variation in covenant thresholds—differences at origination, differences in adjustment, and differences due to the passage of time since origination—create randomness in covenant thresholds for similar firms partitioned together with causal forest. We argue that, within a partition created with very good matching based on contemporaneous firm characteristics, whether a firm is slightly in or out of default is random because of these sources of variation in the contemporaneous default threshold.

The final source of selection bias in our setting, driven by some firms at the default threshold avoiding default, is relatively less concerning. In an RDD where the focus is on the bandwidth sample, this bias is catastrophic. In contrast, a causal forest estimation uses all available data, so threshold observations comprise a very small portion of the sample. As we show in our Monte Carlo results in Figure 4 and discuss in Section 3.2, bias from so few observations is unlikely to affect our estimations.

Finally, we expect that causal forest outperforms RDD in this setting because the effects of default on investment are heterogeneous. Following a covenant violation, lenders can force the firm into bankruptcy, but rarely do so (Gopalakrishnan and Parkash, 1995). Rather, the most common outcome of a technical default is a renegotiation process in which lenders "waive" the violation and may adjust the terms of the lending agreement going forward. For example, lenders can reduce lines of credit, accelerate payments, or increase interest rates (Chava and Roberts, 2008). Renegotiation can also result in tighter or additional covenants (Nini, Smith, and Sufi, 2009, 2012; Becher, Griffin, and Nini, 2020). Lenders' discretion in this renegotiation process results in heterogeneity in when and how firms are affected by default. Chen and Wei (1993) show that 45% of violations are waived with no observable changes to loan terms and that these waivers happen nonrandomly—firms with lower leverage ratios and stronger financial positions are more likely to receive a waiver. Griffin, Nini, and Smith (2019) describe violations in which lenders take no consequential actions as "foot faults". Griffin, Nini, and Smith (2019) show that up to one-third to two-thirds of violations constitute foot faults and substantial variation in these incidences over time. Results from these studies show heterogeneity in the effects of default, which our Monte Carlo results show bias RDD effects.

5 Data, variables, and sampling procedure

To examine the effects of default on investment, we focus on tangible net worth and net worth covenants. These covenants have unambiguous definitions that are standardized across lenders. Other covenants, particularly those involving debt, are not clearly defined and specified in the Dealscan covenant data (long-term vs. short term debt, for example, is not specified in debt-to-EBITDA covenants), which makes measurement of a distance to default difficult. Demerjian and Owens (2016) confirm this statement by collecting covenant calculation details from Tearsheets and providing statistics on how frequently Tearsheet definitions match standard definitions. Additionally, while current ratio covenant definitions are also standardized, we leave analysis of current ratio covenants to our internet appendix due to concerns about overlap in the sample (Section F). We discuss these concerns and results in Section 7.4. We begin our sample in 1994 due to low coverage of covenant data before 1994.

We combine loan covenant data from Loan Pricing Corporation's (LPC) Dealscan database and firm data from Compustat to create a dataset with firm-quarter observations of firm investment, distance to covenant violation, and controls. Dealscan data are matched to Compustat quarterly data using the linking file on Michael Roberts' webpage.[13] Our sample ends in 2017 because the linking file runs through the end of 2017. We elect not to continue

[13] Accessed January, 2019: http://finance.wharton.upenn.edu/ mrrobert/.

our sample further to ensure no differences in matching procedure affect our later sample. Generally, covenants apply to all loans in a package, so we identify firms as being bound by covenants from the earliest start date to the final maturity date of any loan in the package.

We determine whether a firm is in default in a quarter based on the calculation of net worth from Compustat and the violation threshold given in Dealscan data. To calculate distance to default, we first take the difference between the firm's net worth and the default threshold. For covenants with adjusting thresholds, we linearly interpolate quarterly thresholds between the starting and ending thresholds. This difference is then divided by the threshold itself, providing a percentage distance to default. Following Chava and Roberts (2008), this variable is trimmed at the upper and lower 2.5 percentiles.

We define the treatment variable, $bind_{it}$:

$$bind_{it} = \begin{cases} 1, & z_{it} - z_{it}^0 < 0 \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

for firm i in year-quarter t, where z_{it} is the firm's net worth and z_{it}^0 is the default threshold given in the covenant. For any firm bound by multiple covenants, $bind_{it} = 1$ if any one covenant is in default in a quarter. Covenant violations are common. Approximately 42% of firms violate a covenant at some point in the sample period.

We include in the sample for our main analyses firms that do not default in the sample and omitting loans with negative initial distance to default (slack). We include firms that do not default to increase sample size and overlap. We follow Chava and Roberts (2008) in excluding loans with negative slack at origination. In robustness work, we examine causal forest estimates on samples omitting firms that do not default at least once in the sample period and samples that omit or include loans with negative initial slack (Section 7.3). Results from these estimations allow us to demonstrate how causal forest can be used to reconcile results across such dissimilar samples and, simultaneously, the robustness of our conclusions.

Table 2 shows summary statistics for our sample. To remain in the sample, we require each firm to have the Compustat accounting variables necessary to calculate net worth, non-missing control variables, and positive leverage. We omit financial firms (SIC codes 6000–6999). Following the literature, all Compustat control variables are trimmed at the upper and lower first percentiles. In our causal forest estimation, our outcome variable is investment (capital expenditures scaled by beginning of period net PPE). A firm-quarter is treated if the firm is in technical default in the quarter; otherwise, the firm-quarter is a control observation. We include as covariates the set of controls common in this literature, including one-period lags of Macro Q, log(assets), and Altman Z-score (Chava and Roberts, 2008). Additionally, we include contemporaneous and a one-quarter lag of cash flow, initial distance to default (slack),[14] or the difference between the firm's net worth and the net worth covenant threshold at the loan start date, and contemporaneous distance to default (slack), year, binary quarter indicators, and firm fixed effects.[15] Finally, we include measures of information asymmetry Chava and Roberts (2008) show are important to the sensitivity of firms' investment to default: a binary variable for whether the firm has a credit rating, the loan's syndicate size, and current cash scaled by assets. Formal definitions for all Compustat control variables are available in the caption of Table 2 and internet appendix B.

6 Application analysis, results, and discussion

6.1 Main results: HTEs and CATEs

Figure 7 presents a density plot of estimated observation-level heterogeneous treatment effects (HTEs). The peak of the distribution is between 0% and -0.5% and a large percentage of estimates fall between -0.5% and 0.5%. However, there is substantial heterogeneity in estimated treatment effects. A sizable portion of the sample have HTEs less than -0.25%, including observations with HTEs between -1% and -2%, which are economically meaningful estimates. Although on average the treatment effect of default on investment is small, there are firms that are relatively more affected by default.

Table 3 summarizes observation-level treatment effect estimates as ATE, ATT, ATC, and ATO. The ATE is -0.24% and is statistically insignificant (t-statistic of -1.323). These results suggest that the effect of default on firm investment is, on average, negative and small. The ATT of -0.55% is economically and statistically significant (t-statistic of -5.845). In comparison, the ATC is only -0.19%. Were the control observations treated, on average, the effect of technical default would be small and insignificant. There are more control than treated observations in the sample, so the ATE is closer to the ATC than the ATT.

Our causal forest estimation recovers ATE, ATT, and ATC that are numerically different, which highlights the importance of several aspects of our empirical design to fully capturing the effect of default on investment. In this setting, it would be inappropriate to assume that ATT is equal to ATC. Differences between ATT and ATC show that the treatment and control groups are dissimilar along a covariate that is important to heterogeneity. In subsequent sections (Sections 6.2 and 6.3), we identify key covariates driving treatment effects

[14]Contemporaneous slack cannot be included in the forest as it perfectly predicts default status.

[15]We use the fixed effects estimator Jens, Page, and Reeder, III (2021) describe.

as measures of firm financial conditions. Thus, the dissimilarity between ATT and ATC is not surprising; because of how financial covenants are specified (Section 4.1), we expect that firms in default are in more precarious financial positions than firms not in default. The correlation between covariates that both predict treatment and drive treatment effects results in differences between ATT and ATC.

The calculation of ATE, ATT, and ATC separately is a benefit of using causal forest over other estimators, including both RDD and OLS. In an RDD, only one local average treatment effect, or LATE, equal to both ATT and ATC is estimated. Similarly, in OLS and the matching methods we describe in Section 2.2, the focus is on calculating ATE by isolating ATT from selection bias. The inferences Table 3 provides are not possible with RDD and OLS.

In Table 3, we also estimate an ATO of −0.54%. This estimate is close to the ATT estimate, −0.55%, and greater in magnitude than the ATE estimate, −0.24%. The ATO measures average treatment effect with higher weight placed on observations where there is particularly good overlap in the covariate space. Mechanically, ATO is calculated with more weight on observations with estimated treatment propensities near 50%. We find a stronger effect of default on investment using ATO than ATE for the same reason we find a smaller ATT than ATC. Firms more likely to be in default are also more likely to have stronger treatment effects of default on investment.

6.2 Comparative statics identifying key covariates

In Figure 8, we provide comparative statics that show how predicted treatment effects vary over each covariate. For each panel of Figure 8, we set all covariate values to the sample geometric median vector except for the covariate on the x-axis.[16] So, Figure 8 shows the predicted treatment effect for the range of each covariate, while all other covariates are set to represent a median firm in the sample. By varying only one covariate at a time, while holding the others constant, we isolate that covariate's effect on the treatment effect. In effect, we use the forest's mapping between covariates and treatment effects to run experiments on how each variable affects firms' response to default (see Section 2.5).

The first takeaway from Figure 8 is that, for most of the support of most covariates, pre-

[16] Rather than set each variable to its median, we use the geometric median, which is the vector of covariates that best approximates the "median" firm in the sample. We do this to avoid creating a "Frankenfirm" with the median value for each covariate that does not resemble any actual firms in the sample. Firms with close to median size, for example, have very low cash holdings, while firms with median cash holdings are more likely to be small, so it would be inappropriate to create a representative firm with median size and cash holdings. The geometric median looks for a combination of size and cash holdings that is most typical, so the geometric median observation is a better fit for a "typical" observation in our sample.

dicted treatment effects are close to zero, which is consistent with our small and statistically insignificant ATE estimate from Table 3. For most firms in the sample, characterized by a wide set of covariate ranges, we find no effect of default on firm investment.

Evidence of negative treatment effects is concentrated in firms that are less well off financially and in smaller firms. For firms over approximately $40 million in size, predicted treatment effects are close to zero. However, for firms below $40 million, we estimate negative treatment effects, or a decline in investment with default. We predict approximately a 0.33% drop in investment with default for firms between $29 and $40 million in size. There is a steep decline in treatment effects below approximately $29 million in size. Microcaps between $10 and $20 million in size have predicted treatment effects between approximately −2% and −3%.[17] Additionally, predicted treatment effects for contemporaneous and lagged cash flows are relatively stable and approximately equal to zero while cash flows are positive. However, a decline in investment of approximately −1% is predicted for firms with negative contemporaneous and lagged cash flows. Predicted treatment effects are closer to zero for firms with very low Macro Q and negative for firms with higher Macro Q. Figure 8 shows the key covariates to determining the effect of treatment status are: size, Macro Q, cash flow, and lagged cash flow.

6.3 Additional analysis of key covariates and discussion

Table 4 presents the ranking of our covariates in importance to predicting treatment effects calculated using the permutation procedure we detail in Section 2.5. We provide three statistics from this analysis. First, to calculate Absolute Mean Difference (column 1), we find the average of the absolute value of the difference between treatment effects calculated using permuted data and un-permuted data. All else equal, a higher absolute mean difference indicates that a covariate is more important to the determination of treatment effects. Standard Deviation of Differences (column 2) is the standard deviation of these differences. Finally, Ratio of Difference Variance to HTE Variance (column 3) scales the variance of these differences by the variance of the original treatment effect estimates. This ratio shows how large of an effect a hypothetical change in a variable's distribution would have on observation-level treatment effects in comparison to the actual variation in those treatment effects. All else equal, a higher ratio of variances means a covariate is more important to the analysis.

The results in Table 4 show firm size is the most important variable to determining treatment effects, followed by contemporaneous cash flow and Macro Q. These results are consistent with our comparative statics, which show the largest range of treatment effects

[17] Note that assets are deflated to December 2000 by the all-urban CPI.

occurs over the support of firm size. Firm fixed effects are also relatively important to our causal forest analysis, followed by lagged cash flow, Altman Z, and cash-over-assets. Many of the most highly-ranked covariates are measures of firm financial strength and stability. Variables further down the list that, if omitted from our analysis would have little to no effect on calculated HTEs, include: initial and contemporaneous distance to default (slack), year and quarter, syndicate size, and a binary credit rating variable.

Using our comparative statics from Figure 8 and covariate rankings from Table 4, we identify subsamples of firms with large predicted declines in investment with default. In Table 5, we split our sample into firms with negative and non-negative cash flows and firms in the highest and lowest terciles of Macro Q and average our estimated HTEs for each subsample. Table 6 shows averages of HTEs for firms with negative and non-negative cash flows and firms in largest and smallest size terciles. Because of overlap between groups of firms with high predicted treatment effects, for example, driven by correlation between lagged and contemporaneous cash flows, these samples provide a comprehensive look at treatment effects in samples with high treatment effects. Sample splits on lagged and contemporaneous cash flows, for example, are redundant.

The results in Table 5 show large declines in investment with default are concentrated in firms with negative cash flow and high Macro Q. We estimate these firms experience, on average, a 1.4% to 1.5% decline in investment when in default. In contrast, we estimate an ATE near zero for firms with high Macro Q and positive cash flows. However, firms with negative cash flow and high Macro Q comprise only 4.3% of our sample, suggesting that while there are firms with investment that is greatly affected by covenant violations, these firms are not typical in the sample. In Table 5 we show firms with negative cash flow have larger declines in investment while in default than firms with positive cash flow, but that this effect is concentrated in firms with many potential investment projects (high Macro Q).

In Table 6, we split our sample on size and negative vs. non-negative cash flow and show the largest declines in investment in default occur for small firms with negative cash flow. For these firms, which comprise about 7% of our sample, we estimate treatment effects of approximately -0.9% to -1%. For comparison, small firms with positive cash flow have an ATE of 0.33% that is statistically indistinguishable from zero (t-statistic of 0.58). Additionally, whereas large firms with negative cash flow experience, on average, a 0.66% decline in investment in default, large firms with positive cash flows have a smaller estimated treatment effect, -0.16%. The ATE for large firms with negative cash flow has a t-statistic of -1.402 and so is not statistically significant at the 10% level, but the sample size of 742 is relatively small. Comparing the sample sizes of subsamples in this table highlights the non-random nature of these covariates—there are over three times as many small firms as large firms with negative cash flow. Our inferences throughout this section are possible because causal forest is so effectively able to disentangle between such correlations.

Our results across covariates and subsamples make sense economically and are consistent with results in related studies. Covenants are intended to be the canary in the coal mine, alerting lenders to potentially concerning conditions within a firm. However, although firms in default are, on average, distressed, covenant violation does not necessarily mean a firm is in distress, and lenders have discretion in their actions when a firm violates a covenant. Related literature shows heterogeneity in lender actions consistent with the heterogeneity in treatment effects we recover. Chen and Wei (1993) hand collect a small sample of technical defaults and show 57 of 128 violations were waived (45%) and that firms with high leverage and bankruptcy risk are the least likely to receive waivers. Likewise, Beneish and Press (1993) show heterogeneity in outcomes of technical default. Although these early studies are limited to small samples, their descriptive statistics suggest that the average firm may not be greatly affected by technical default, and the average cost of technical default on firm investment may be low and limited to subsamples of firms in financial distress. More recently, Griffin, Nini, and Smith (2019) find that, after up to two-thirds of covenant violations, lenders take no meaningful actions and borrowers face no discernible consequences of violation. Evidence in Griffin, Nini, and Smith (2019) is also consistent with a low and insignificant average cost of technical default.

We find an important cutoff in determining negative treatment effects is zero cash flows, which echoes the importance of the banking regulations Chernenko, Erel, and Prilmeier (2020) use for identification. Chernenko, Erel, and Prilmeier (2020) study nonbank lending and find a jump in the probability of nonbank lending at zero EBITDA, suggesting that this cutoff is important to banks and nonbanks in lending. They argue this jump is driven by guidance from The Comptroller of the Currency (OCC) Handbook on Rating Credit Risk (2001), which recommends that lenders focus on the "strength of the primary repayment source" in evaluating credit risk.[18] Chernenko, Erel, and Prilmeier (2020) note that the primary repayment source for most loans is operating cash flow. Thus, we find strong, negative treatment effects in firms that, according to Chernenko, Erel, and Prilmeier (2020), are more likely to receive a *non-pass* rating in banks' internal credit risk rating systems and "attract special regulatory scrutiny". The consistency between our results and those in related studies, including Chen and Wei (1993), Beneish and Press (1993), Griffin, Nini, and Smith (2019), and Chernenko, Erel, and Prilmeier (2020), suggests our causal forest estimation recovers important heterogeneity in the effect of default on investment.

[18] https://www.occ.treas.gov/publications/publications-by-type/comptrollers-handbook/rating-credit-risk/pub-chrating- credit-risk.pdf.

7 Supporting analyses and robustness

7.1 Empirical evidence supporting unconfoundedness

To aid in our discussion of the appropriateness of assuming unconfoundedness in our setting, we estimate a classification forest predicting the propensity of each firm-quarter to be in default. We rank covariates by importance in predicting these propensities in Panel A of Table 7. The results in Panel A of Table 7 show that the most important covariate to predicting default status is Altman Z. This is unsurprising because Altman Z was designed to predict whether a firm will default. Additional covariates important to predicting whether a firm-quarter is in default are: initial slack, or how tightly a covenant is set originally, firm size, and contemporaneous cash flows. Other covariates, including syndicate size and lagged cash flow, are less important.

We compare the ranking in Panel A of Table 7 with the ranking of covariates that are important in predicting treatment effects of default in Table 4. A comparison of the covariates important to predicting default status (Table 7) with the covariates important to predicting the treatment effect of default on investment (Table 4) shows that similar covariates are important in both forests.

The similarity between the covariates that predict treatment status (default) and treatment effects of default on investment is important to understanding whether assuming unconfoundedness in our main forest estimation is reasonable. Unconfoundedness requires that treatment status is uncorrelated with the potential outcome, controlling for covariates that predict treatment effects. Generally, it is possible a variable exists that: predicts treatment status, is correlated with the outcome variable, and is not a particularly good predictor of treatment effects. In this hypothetical setting, assortment of observations to final partitions by predicting treatment effects would not control for this variable, so assuming unconfoundedness would be difficult. However, in our setting, overlap between covariates that strongly predict treatment and treatment status means that our main forest estimation creates final partitions of firm-quarters that are similar in the effect of default on investment *and* the likelihood of default. This setting-specific detail, combined with the benefits of the structure of trees that make forest analyses relatively robust to concerns about omitted variable, mean that the assumption of unconfoundedness is plausible in our setting.

Furthermore, we can use the variable importance ranking in Table 4 to show how important a latent variable would need to be to our analysis to bias our estimated treatment effects. We compare the "standard deviation of differences" column from Table 4 and see that variables ranked below our firm fixed effect have approximately one-third the standard deviation of the ATE estimate. Thus, were we to omit any of these covariates—lagged cash flow, lagged Altman Z, cash-over-assets, initial and contemporaneous slack, year, syndicate size, quarter, or credit rating—from our analysis, the treatment effect estimates would change, but still be well within one standard deviation of our initial estimates. This comparison gives an idea of how important an omitted covariate would need to be to greatly affect our estimates. Such a covariate would need to dominate these covariates in importance to the estimation, while being sufficiently uncorrelated with covariates already included, to be a relevant latent variable. Existence of such an omitted variable, while possible, is unlikely.

7.2 Re-examining drivers of variation in treatment effects

While we show that firm financial health is the most important covariate in determining treatment effects, Chava and Roberts (2008) theorize that a major goal of covenants is to mitigate agency problems, so covenant violations should be more severe for firms with higher information asymmetry, as creditors will be more likely to intervene if these firms default. Using sample splits on information asymmetry proxies, Chava and Roberts (2008) conclude that information asymmetry drives important variation in the treatment effect of default on investment.

In our permutation analysis, in which we sever correlations between covariates to determine how each individual covariate drives variation in treatment effects, we find that the most important sources of heterogeneity in our results—firm size, Macro Q, and cash flow—do not include the Chava and Roberts (2008) proxies for information asymmetry (Table 4). Indeed, two proxies, a binary credit rating variable and syndicate size, are among the least important covariates in our analysis. Likely, then, information asymmetry subsample results in Chava and Roberts (2008) are driven by a covariate important to HTE calculation and correlated with the Chava and Roberts (2008) information asymmetry proxies, such as firm size (log of assets). Consistent with this explanation, rating and syndicate size are highly correlated with the most important covariate in our forest, firm size (correlations of 0.60 and 0.62, respectively). For more discussion on the strengths of the permutation analysis in isolating the effects of individual covariates, see Section 2.5.

7.3 Causal forest results across samples

In internet appendix Table IA.7, we show our causal forest results are robust across several sample constructions. Table IA.7 presents the effects of default on investment for samples: keeping all firms and all loans (Panel A), keeping all firms and dropping loans with negative initial slack (Panel B), keeping only firms that default at least once in the sample and keeping all loans (Panel C), and keeping only firms that default at least once in the sample and dropping loans with negative initial slack (Panel D). Results in Panel B are also given in Table 3, as this is the sample we use for our main analysis.

Our results in Table IA.7 are consistent with ATE, ATT, ATC, and ATO in Table 3. Whether we keep or drop loans with negative initial slack does not greatly affect our estimates; results in Panels A and B and Panels C and D are very similar. However, ATEs are lower when firms that never default are dropped from the sample, −0.31% and −0.24% in Panels A and B versus −0.43% in both Panels C and D. Intuitively, these results make sense. Firms that default at least once in the sample period are, on average, in worse shape financially than firms that never default. We show that firms that are in worse shape financially are relatively more affected by default. Thus, the ATE is more negative if a sample includes more firms that should have more negative treatment effects.

When interpreting the results in Table IA.7, it is important to remember that results are conditional average treatment effects that should differ across these samples. The different samples we examine in Table IA.7 are selected on observables. Because observables determine the degree to which a firm is affected by default, estimations differ across these samples. While typically in a robustness section, a researcher provides results from a series of tests that more or less find the same average treatment effect (and, thus, argues her results are "robust"), this is not the goal of our robustness analyses. Causal forest should recover similar observation-level treatment effects across estimations. However, conditional average treatment effects of these observation-level results depend on sample composition. Thus, our goal here is not to show similar average effects across samples, but rather understandable differences between average effects across samples.

7.4 Current ratio covenant results and discussion

In internet appendix Section F, we fully discuss our decision to leave current ratio covenants out of our main analysis sample. Here, we detail one important concern that has implications for both causal forest and RDD's performance in recovering treatment effects in this sample. In Figure 9, we present results from a McCrary (2008) test. A McCrary (2008) test statistically compares the density of observations occurring above and below a threshold. Bunching above the default threshold is concerning as it suggests that firms can avoid default or select their default status. In Figure 9, we reject the null of no bunching at the 0.01% level in both the current ratio and net worth samples (McCrary, 2008). In our Monte Carlo experiments, we show that RDD treatment effects are badly biased by this kind of non-random assortment. Additionally, bunching in the current ratio sample is so strong, it raises the concern that there is a lack of overlap between the treatment and control groups. In Section F, we present and discuss causal forest results in the current ratio sample consistent with a lack of overlap.

7.5 Further discussion of threshold bunching

In Section 7.1, we describe a classification forest estimation predicting default propensity, results from which can be used to further demonstrate the bunching of firms that exists at the default threshold. In Figure IA.4, we plot these predicted propensities against the slack 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.

Taken with the McCrary (2008) results from Section 7.4, these results have important implications for the over thirty studies that follow Chava and Roberts (2008) and treat debt covenant defaults as an exogenous event. We list and summarize these studies in internet appendix Section A, Table IA.1. All studies using this RDD framework should be concerned about non-random assortment of firms at the default threshold introducing bias into their estimates and proceed cautiously. Additionally, we, along with other studies including Chen and Wei (1993) and Griffin, Nini, and Smith (2019), show heterogeneity in the effects of default on investment. Our Monte Carlos show that heterogeneity also biases RDD estimates; this heterogeneity is a second reason to proceed with caution when using RDD to examine the effects of technical default on any outcome variable.

7.6 Reconciling with previous literature

Having demonstrated how default heterogeneously affects firm investment, we reconcile our causal forest HTE estimates with average treatment effects existing literature documents.

We begin with three important caveats to this discussion. First, the most recent literature on the costs of default on investment answer a different question than our research question. We are interested in heterogeneity in the costs of default, or determining which and how firms are affected by default. Previous studies ask whether there is statistical evidence that default affects investment. Accordingly, these studies primarily rely on various RDD specifications, which are designed to directly address their question. In using an RDD, researchers focus narrowly around a plausibly exogenous threshold to recover an ATE that is only applicable to this narrow sample. Only so much variation in threshold observations exists, so RDDs are limited in their ability to recover heterogeneity in effects.

Second, causal forest relies on totally different assumptions than does RDD. The validity of our results depend on the credibility of our assumptions in our setting, which we detail in Section 7.1. RDD's ability to recover an unbiased effect of default on investment depends on the credibility of RDD assumptions in a setting, including the assumption of local continuity. Any concerns regarding the plausibility of RDD assumptions have no bearing on our causal forest setup.

Finally, the goal of our paper is to introduce causal forest into the literature. We argue that causal forest is a valuable new econometric technique. As a new technology in the literature, it is unsurprising that causal forest can provide different answers to some empirical questions than researchers have obtained using older technologies. However, an improvement in technology that provides a new answer to a question does not constitute evidence that older studies within a literature are "good" or "bad". The relative contributions and merits of studies should be weighed based on the standard and methods of the time in which they were written.

To aid in our reconciliation between existing results in the literature and our causal forest results, we analyze the debt covenant default RDD setting and show that results from this setting are sensitive to two competing sources of endogeneity bias. Tables 8 and 9 provide estimations for net worth and current samples separately, and Tables IA.13 and IA.14 provide the same analysis for a combined sample of firms with one of these two covenants.[19] These results show that RDD does not accurately capture the costs of default on investment even for the small number of firms at the threshold. Across a variety of RDD specifications with a forcing variable of distance to covenant violation, we show that treatment effects of default on investment range from -1.22% to 0.42%. These results show a push-and-pull between two competing sources of selection bias.

We find the largest negative treatment effects in specifications in which differences between treated and control firms are controlled for poorly (i.e., specifications with fewer explanatory variables and in broader samples around the threshold). Intuitively, it makes sense that treatment effects that reflect this source of selection bias are more negative because firms in default have, on average, lower investment than firms not in default.

In RDD specifications emphasizing observations closer to the threshold, in which the

[19]For the calculations in Table 9, we use the RDD $\mathbb{R}$ package (Dimmery, 2016) and the RDrobust $\mathbb{R}$ package (Calonico, Cattaneo, Farrell, and Titiunik, 2021).

negative effect should be more pronounced, we instead find a positive treatment effect. These specifications include estimations with narrow bandwidth samples and placing relatively more weight on observations closer to the threshold (i.e., triangular kernels). In contrast with studies in the literature that document large negative effects of default on investment (Chava and Roberts, 2008; Demiroglu and James, 2010; Falato and Liang, 2016), we find that firms in technical default, just below the threshold, actually have *higher* investment than firms just above the threshold. This result suggests that at least some firms can avoid default, which is consistent with Dichev and Skinner (2002), who show that firms bunch just above default thresholds, and a survey of managers who report that they would forgo profitable investment to avoid default (Graham, Harvey, and Rajgopal, 2005).

Using our results from Tables 8 and 9, we can compare our causal forest results to those in two streams of literature on default and investment. Studies in this literature can be categorized into two groups by how the study identifies covenant violations. Some studies follow Chava and Roberts (2008) and link covenant thresholds from Dealscan to Compustat firm variables and calculate a distance to default (slack); if slack is negative, the firm is in technical default. Calculating a distance to default allows the researcher to identify a "narrow miss" sample to use as a control group. However, these studies are limited to focus only on covenants defined using unambiguous accounting definitions, typically, (tangible) net worth and current ratio. In our sample, this limits us to examining only three of 22 possible covenants. Others, including Roberts and Sufi (2009), identify violations using 10-K and 10-Q filings. This method identifies relatively more violations, but robs the researcher of the ability to use distance to default to define a control group. Several studies, for example, Demiroglu and James (2010), use both methods for robustness.

We begin with a discussion of studies using Dealscan data and distance to default as a control variable. These studies include Chava and Roberts (2008), Demiroglu and James (2010), and Falato and Liang (2016), who find large, negative effects of default on investment using an RDD framework, approximately -0.8% to -1.5%. In RDD specifications most similar to theirs, we find comparable treatment effects of approximately -0.7% to -1.2%. These estimates are economically large, suggesting default causes a 12% to 20% decline in investment rate, relative to the sample average. In contrast, our causal forest estimates an average treatment effect of -0.24%, or an average decline in investment rate of 4%. Our causal forest estimates suggest default has an economically significant effect on investment but that the effect is more modest than previous studies find.

Our causal forest results differ from those obtained using RDD because causal forest is better able to control for differences between firms in and not in default than RDD. In theory, polynomial RDD specifications should be able to fully control for differences between firms in and not in default.[20] However, using causal forest, we find nonlinearities in heterogeneous treatment effects that are difficult to control for in such an RDD setup. If a specification does not control for underlying differences in investment between firms in and not in default, we expect to recover overly large, negative effects of default on investment. In comparison, our causal forest treatment effects are closer to zero than these RDD estimates because trees are better able to control for setting-specific nonlinearities in covariates than "off the shelf" polynomial specifications.

In our discussion this far, we focus on why our conclusions differ from those from other studies using Dealscan data to identify defaults. We now shift to discussing our results relative to those from papers using SEC data to identify defaults. These studies include Roberts and Sufi (2009) and Nini, Smith, and Sufi (2012), who report finding similar drops in investment around covenant violations as do Chava and Roberts (2008).[21] In a related study, Nini, Smith, and Sufi (2009) find lower investment following firms obtaining a debt agreement with a capital expenditure restriction. Because we do not examine defaults from SEC data in this paper, we are more limited in our discussion of our results relative to those in Roberts and Sufi (2009) and Nini, Smith, and Sufi (2009). However, we can discuss how our results should differ from those in these studies because of the advantages of causal forest relative to commonly-used empirical methods in this literature.

Roberts and Sufi (2009) and Nini, Smith, and Sufi (2012) face the same endogeneity issue as Chava and Roberts (2008) and us—non-random default status. However, without the ability to measure distance to default, Roberts and Sufi (2009) and Nini, Smith, and Sufi (2012) rely on an identification strategy that is effectively a difference-in-difference framework with an endogenous "treatment" variable (default status). To identify an unbiased treatment effect, these studies must fully control for differences between their treated and control samples, which is a difficult task. Otherwise, the coefficient on the treatment variable captures differences between the treated and control samples (i.e., selection bias). To this end, Nini, Smith, and Sufi (2012) report that linear, quadratic, and cubic contemporaneous and lagged ratios on which covenants are commonly defined and market-to-book ratio are included as control variables in their regressions. Additionally, Nini, Smith, and Sufi (2012) use a first-difference specification to control for time-invariant effects at the firm level.

Unfortunately, this specification does not capture the heterogeneity in effects our causal forest estimation shows to be important. We find that size, particularly in small firms, is a key source of heterogeneity. This specification can control for heterogeneity like this, but only

[20] Gelman and Imbens (2018) discuss general weaknesses in such specifications. We focus on weaknesses specific to this setting.

[21] This statement pertains to the main results in these papers. For robustness, Roberts and Sufi (2009) also examine Dealscan data and the RDD framework from Chava and Roberts (2008).

with interactions between the treatment variable (default) and measures of size. A squared or cubed size variable would be hard-pressed to capture the nonlinearities we document in size, flat for medium-sized and large firms but rapidly decreasing as firm size drops below approximately $40 million. Additionally, while first-differencing can control for persistence in variables and size may be persistent for some firms, we do not expect persistence in size in these small firms. Thus, our causal forest analysis provides results that suggest serious concerns about omitted variables in this empirical framework that can result in differences between our results and conclusions and those in Roberts and Sufi (2009) and Nini, Smith, and Sufi (2012).

It is important to recognize that causal forest is an improvement over OLS only if the researcher does not perfectly know the functional form. As the BLUE, OLS bests causal forest as an estimator, so long as the appropriate covariates and interactions between these covariates are included. However, in practice, the researcher likely does not know the true functional form. Even the "kitchen sink models", as Roberts and Sufi (2009) describe their estimations, are insufficient to capture the heterogeneity our causal forest estimation uncovers. A fully saturated OLS can capture all interactions a forest captures, but cannot be estimated in finite data. Section 2.1 provides more supporting discussion regarding OLS' inability to capture heterogeneity.

Finally, we note that differences between samples can drive differences between results in our study and those in related studies. For example, Nini, Smith, and Sufi (2009) study covenants that restrict firms' capital expenditures and the conditions under which these covenants are imposed. They find that, within this non-random sample of firms, covenants impede investment. It is outside of the scope of our study to examine such covenants in detail, but the results in Nini, Smith, and Sufi (2009) do not necessarily contradict ours. If, for example, there is overlap between firms we find are particularly affected by default and those Nini, Smith, and Sufi (2009) study, we agree that the investment of these firms is affected by covenants and covenant violation. Unfortunately, summary statistics in Nini, Smith, and Sufi (2009) do not provide enough information to make a direct comparison possible. Future work directly comparing such studies using causal forest can highlight the ability of causal forest to reconcile between studies with different samples selected on observables.

8 Conclusion

We show that causal forest is a powerful new econometric method that is well-suited to addressing a common endogeneity complaint in corporate finance, selection bias. We use Monte Carlo experiments to demonstrate causal forest's recovery of low-variance treatment effects across a variety of settings. Causal forest also recovers low- to no-bias treatment effects when we incorporate heterogeneity and some forms of latency into our simulations. We also show that, even when a correlated omitted variable is introduced, causal forest performs well, recovering precise treatment effects with only 4% to 6% bias, on average. Thus, causal forest controls for selection bias well, although not explicitly designed to do so.

We also provide a critical analysis of RDD's role in empirical corporate finance research. While, theoretically, RDD controls for selection bias well, we highlight three concerns with how RDD works in practice. First, RDD estimates are high variance. Second, heterogeneity biases RDD estimates. Third, even modest levels of manipulation badly bias RDD estimates; RDD estimates are very sensitive to unprovable assumptions regarding random assortment. All of these concerns are exacerbated by the fact that we typically do not have the opportunity to design an RDD but find them "in the wild", and so are limited to analyzing one sample. Another limitation of RDD we have not explicitly mentioned that is also relevant to applied research is that RDDs in the wild are rare. If a researcher is dependent on finding a valid RDD setup to control for selection bias, more likely than not, that research question is going unanswered.

To showcase causal forest's practical usefulness, we re-visit a popular RDD setting in the literature to examine heterogeneity in the effects of default on investment and demonstrate how causal forest can provide new economic intuition and reconcile between conflicting results in existing studies. Early literature on the consequences of default show evidence that firms manipulate accounting ratios to avoid default but no evidence that defaulting adversely affects firms. Taken together these results are puzzling—why would firms try to avoid default if there are no consequences to default? We use heterogeneous treatment effects to show that, while most firms decrease investment by a relatively small amount, if at all, there is a small subset of firms that react strongly to technical default. Firms lacking internal cash flows, especially if they are small or have good investment opportunities, decrease investment strongly, showing that technical default creates a binding financial constraint for such firms. Simultaneously, we show evidence that some firms are able to avoid default. However, because firms only avoid default at the threshold, these firms comprise a very small portion of our overall sample. Recognizing heterogeneity in effects across the whole sample is key to understanding how these results in the literature fit together.

Causal forest can be thought of as a flexible, non-parametric estimator that bridges the gap between OLS and traditional matching methods and pseudo-randomized techniques. Estimates from OLS and simpler matching methods cannot control for bias as effectively as causal forest. Pseudo-randomized techniques tend to require strong assumptions in relatively narrow settings and recover treatment effects with limited extendibility. In contrast, causal forest recovers precise and low- to no-bias treatment effects across a whole sample, striking balance between recovering quality and extendible treatment effects. As such, causal forest proves a valuable addition to our econometric toolkit.

References

Akins, B., Bitting, J., DeAngelis, D., Gaulin, M., 2019. Do CEO compensation policies respond to debt contracting?, Unpublished working paper.

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. The Annals of Statistics 47, 1148–1178.

Athey, S., Wager, S., 2019. Estimating treatment effects with causal forests: an application, unpublished working paper.

Baker, A., Larcker, D. F., Wang, C. C. Y., 2021. How much should we trust staggered difference-in-difference estimates?, European Corporate Governance Institute—Finance Working Paper No. 731/2021; Rock Center for Corporate Governance at Stanford University Working Paper No. 246.

Balsam, S., Gu, Y., Mao, C. X., 2018. Credit influence and ceo compensation: Evidence from debt covenant violations. Accounting Review 93, 23–50.

Balsmeier, B., Fleming, L., Manso, G., 2017. Independent boards and innovation. Journal of Financial Economics 123, 536–557.

Barber, B. M., Lyon, J. D., 1996. Detecting abnormal operating performance: the empirical power and specification of test statistics. Journal of Financial Economics 41, 359–399.

Bartov, E., 1993. The timing of asset sales and earnings manipulation. Accounting Review 68, 840–855.

Becher, D. A., Griffin, T. P., Nini, G., 2020. Congruence in governance: Evidence from credit monitoring of corporate acquisitions, Unpublished working paper.

Beneish, M. D., Press, E., 1993. Costs of technical violation of accounting-based debt covenants. Accounting Review 68, 233–257.

Beneish, M. D., Press, E., 1995. The resolution of technical default. Accounting Review 70, 337–353.

Bhaskar, L. S., Krishnan, G. V., Yu, W., 2016. Debt covenant violations, firm financial distress, and auditor actions. Contemporary Accounting Research 34, 186–215.

Billett, M. T., Esmer, B., Yu, M., 2018. Creditor control and product-market competition. Journal of Banking & Finance 86, 87–100.

Bird, A., Ertan, A., Karolyi, S. A., Ruchti, T. G., 2017. Lender forbearance, Unpublished working paper.

Bird, A., Ertan, A., Karolyi, S. A., Ruchti, T. G., 2019. Short-termism spillovers from the financial industry, Unpublished working paper.

Bowen, D. E., Frésard, L., Taillard, J. P., 2017. What's your identification strategy? Innovation in corporate finance research. Management Science 63, 2529–2548.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Bulan, L., Hull, T., 2013. The impact of technical defaults on dividend policy. Journal of Banking & Finance 37, 814–823.

Calonico, S., Cattaneo, M. D., Farrell, M. H., 2020. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. Econometrics Journal 23, 192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., Titiunik, R., 2021. rdrobust: Robust data-driven statistical inference in regression-discontinuity designs. R package version 0.99.9.

Chakraborty, I., Chava, S., Ganduri, R., 2015. Credit default swaps and moral hazard in bank lending, Unpublished working paper.

Chava, S., Nanda, V., Xiao, S. C., 2017. Lending to innovative firms. Review of Corporate Finance Studies 6, 234–289.

Chava, S., Roberts, M. R., 2008. How does financing impact investment? The role of debt covenants. The Journal of Finance 63, 2085–2121.

Chava, S., Wang, R., Zou, H., 2019. Covenants, creditors' simultaneous equity holdings, and firm investment policies. Journal of Financial and Quantitative Analysis 54, 481–512.

Chen, C., Kim, J., Zhu, C., 2017. The impact of debt covenant violation on credit default swap spreads, Unpublished working paper.

Chen, K. C. W., Wei, K. C. J., 1993. Creditors' decisions to waive violations of accounting-based debt covenants. Accounting Review 68, 218–232.

Chernenko, S., Erel, I., Prilmeier, R., 2020. Why do firms borrow directly from nonbanks?, working paper.

Christensen, T. E., Pei, H., Pierce, S. R., Tan, L., 2019. Non-gaap reporting following debt covenant violations. Review of Accounting Studies .

Colonnello, S., Koetter, M., Stieglitz, M., 2019. Benign neglect of covenant violations: Blissful banking or ignorant monitoring?, Unpublished working paper.

Cook, K., Ma, T., Zhao, Y. E., 2019. Do creditors influence corporate tax planning? Evidence from loan covenants, Unpublished working paper.

Davis, J. M. V., Heller, S. B., 2017. Using causal forests to predict heterogeneity: An application to summer jobs. American Economic Review 107, 546–550.

DeFond, M. L., Jiambalvo, J., 1994. Debt covenant violations and manipulation of accruals. Journal of Accounting and Economics 17, 145–176.

Demerjian, P. R., Owens, E. L., 2016. Measuring the probability of financial covenant violation in private debt contracts. Journal of Accounting and Economics 61, 433–447.

Demiroglu, C., James, C. M., 2010. The information content of bank loan covenants. Review of Financial Studies 23, 3700–3737.

Dichev, I. D., Skinner, D. J., 2002. Large-sample evidence on the debt covenant hypothesis. Journal of Accounting Research 40, 1091–1123.

Dimmery, D., 2016. rdd: Regression discontinuity estimation. R package version 0.57.

Easley, D., de Prado, M. L., O'Hara, M., Zhang, Z., 2021. Microstructure in the machine age. Review of Financial Studies 34, 3316–3363.

Eaton, C., Howell, S. T., Yannelis, C., 2020. When investor incentives and consumer interests diverge: Private equity in higher education. Review of Financial Studies 33, 4024–4060.

Erel, I., Stern, L. H., Tan, C., Weisbach, M. S., 2021. Selecting directors using machine learning, forthcoming, Review of Financial Studies.

Ersahin, N., Irani, R. M., Le, H., 2020. Creditor control rights and resource allocation within firms, forthcoming, Journal of Financial Economics.

Ertan, A., Karolyi, S. A., 2016. Debt covenants and the expected cost of technical default, Unpublished working paper.

Falato, A., Liang, N., 2016. Do credit rights increase employment risk? Evidence from loan covenants. Journal of Finance 71, 2545–2590.

Ferreira, D., Ferreira, M. A., Mariano, B., 2018. Creditor control rights and board independence. Journal of Finance 73, 2385–2423.

Francis, B., Shen, Y. V., Wu, Q., 2017. Do creditor control rights impact corporate tax avoidance? Evidence from debt covenant violations, Unpublished working paper.

Franz, D. R., HassabElnaby, H. R., Lobo, G. J., 2014. Impact of proximity to debt covenant violation on earning management. Review of Accounting Studies 19, 473–505.

Freudenberg, F., Imbierowicz, B., Saunders, A., Steffen, S., 2017. Covenant violations and dynamic loan contracting. Journal of Corporate Finance 45, 540–565.

Galasso, A., Simcoe, T. S., 2011. Ceo overconfidence and innovation. Management Science 57, 1469–1484.

Gao, J., Karolyi, S. A., Pacelli, J., 2018. Screening and monitoring by inattentive corporate loan officers, Unpublished working paper.

Gao, Y., Khan, M., Tan, L., 2017. Further evidence on the consequences of debt covenant violations. Contemporary Accounting Research 34, 1489–1521.

Gelman, A., Imbens, G., 2018. Why high-order polynomials should not be used in regression discontinuity designs. Journal of Business & Economic Statistics 37, 1–10.

Gopalakrishnan, V., Parkash, M., 1995. Borrower and lender perceptions of accounting information in corporate lending agreements. Accounting horizons 9, 13–26.

Graham, J. R., Harvey, C. R., Rajgopal, S., 2005. The economic implications of corporate financial reporting. Journal of Accounting and Economics 40, 3–73.

Griffin, T. P., Nini, G., Smith, D. C., 2019. Losing control? The 20-year decline in loan covenant restrictions, working paper.

Gu, Y., Mao, C. X., Tian, X., 2017. Banks' interventions and firms' innovation: Evidence from debt covenant violations. Journal of Law and Economics 60, 637–671.

Haw, I.-M., Jung, K., Lilien, S. B., 1991. Overfunded defined benefit pension plan settlements without asset reversions. Journal of Accounting and Economics 14, 295–320.

He, L., Zhang, J., Zhong, L., 2018. The real effects of financing on corporate social responsibility: Evidence from covenant violations, Unpublished working paper.

Heath, D., Mace, C., 2020. The strategic effects of trademark protection. Review of Financial Studies 33, 1848–1877.

Iacus, S. M., King, G., Porro, G., 2011. Causal inference without balance checking: Coarsened exact matching. Political analysis 20, 1–24.

Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. Review of Economic Studies 79, 933–959.

Jens, C. E., Page, T. B., Reeder, III, J. C., 2021. Controlling for group-level heterogeneity in causal forest, unpublished working paper.

Jensen, M. C., Meckling, W. H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3, 305–360.

Jha, A., 2013. Earnings management around debt-covenant violations—an empirical investigation using a large sample of quarterly data. Journal of Accounting, Auditing, and Finance 28, 369–396.

Jiang, L., Zhou, H., 2016. The role of audit verification in debt contracting: Evidence from covenant violations. Review of Accounting Studies 29, 1911–1942.

Kahn, R. J., Whited, T. M., 2018. Identification is not causality, and vice versa. Review of Corporate Finance Studies 7, 1–21.

Karpoff, J. M., Wittry, M. D., 2018. Institutional and legal context in natural experiments: The case of state antitakeover laws. Journal of Finance 73, 657–714.

Keil, J., 2018. The value of lending relationships when creditors are in control, Unpublished working paper.

Knittel, C. R., Stolper, S., 2019. Using machine learning to target treatment: The case of household energy use, NBER working paper.

Li, F., Morgan, K. L., Zaslavsky, A. M., 2018. Balancing covariates via propensity score weighting. Journal of the American Statistical Association 113, 390–400.

Lin, Y., Xin, X., Zhang, L., Zhang, Z., 2017. Bank monitoring and corporate tax planning: Evidence from loan covenants, Unpublished working paper.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142, 698–714.

Nie, X., Wager, S., 2017. Quasi-oracle estimation of heterogeneous treatment effects, Unpublished working paper.

Nini, G., Smith, D. C., Sufi, A., 2009. Creditor control rights and firm investment policy. Journal of Financial Economics 92, 400–420.

Nini, G., Smith, D. C., Sufi, A., 2012. Creditor control rights, corporate governance, and firm value. Review of Financial Studies 25, 1713–1761.

Nordlund, J., 2018. Spillovers from creditor control, Unpublished working paper.

O'Neill, E., Weeks, M., 2019. Causal forest estimation of heterogeneous household response to time-of-use electricity pricing schemes, unpublished working paper.

Roberts, M. R., Sufi, A., 2009. Control rights and capital structure: an empirical investigation. Journal of Finance 64, 1657–1695.

Robins, J. M., Rotnitzky, A., Zhao, L. P., 1994. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association 89, 846–866.

Robinson, P. M., 1988. Root-n-consistent semiparametric regression. Econometrica 56, 931–954.

Smith, C. W., Warner, J. B., 1979. On financial contracting: An analysis of bond covenants. Journal of Financial Economics 7, 117–161.

Sweeney, A. P., 1994. Debt-covenant violations and managers' accounting responses. Journal of Accounting and Economics 17, 281–308.

Tan, L., 2013. Creditor control rights, state of nature verification, and financial reporting conservatism. Journal of Accounting and Economics 55, 1–22.

Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Hirshberg, D., Miner, L., Sverdrup, E., Wager, S., Wright, M., 2020. grf: Generalized Random Forests. R package version 1.2.0.

Vashishtha, R., 2014. The role of bank monitoring in borrowers' discretionary disclosure: Evidence from covenant violations. Journal of Accounting and Economics 57, 176–195.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association 113, 1228–1242.

Watts, R. L., Zimmerman, J. L., 1986. Positive accounting theory. Prentice-Hall, Englewood Cliffs, N.J.

Welch, I., 2015. Plausibility, Unpublished working paper.

Zhang, Z., 2019. Bank interventions and trade credit: Evidence from debt covenant violations. Journal of Financial and Quantitative Analysis 54, 2179–2207.

Figure 1. Example matched samples comparison. Figure illustrating how Barber and Lyon (1996) style matching, coarsened exact matching (CEM), and a tree create subsamples of treated (triangles) and control (open circles) observations. In Barber and Lyon (1996) matching, treated firms are matched to control firms in the same industry and within bands (parentheses) of 90% to 110% of the treated firm's size. In CEM, the researcher "coarsens" continuous variables, creating a grid within which treated and control firms are matched within shared cells. Trees use an algorithm to partition the covariate space.

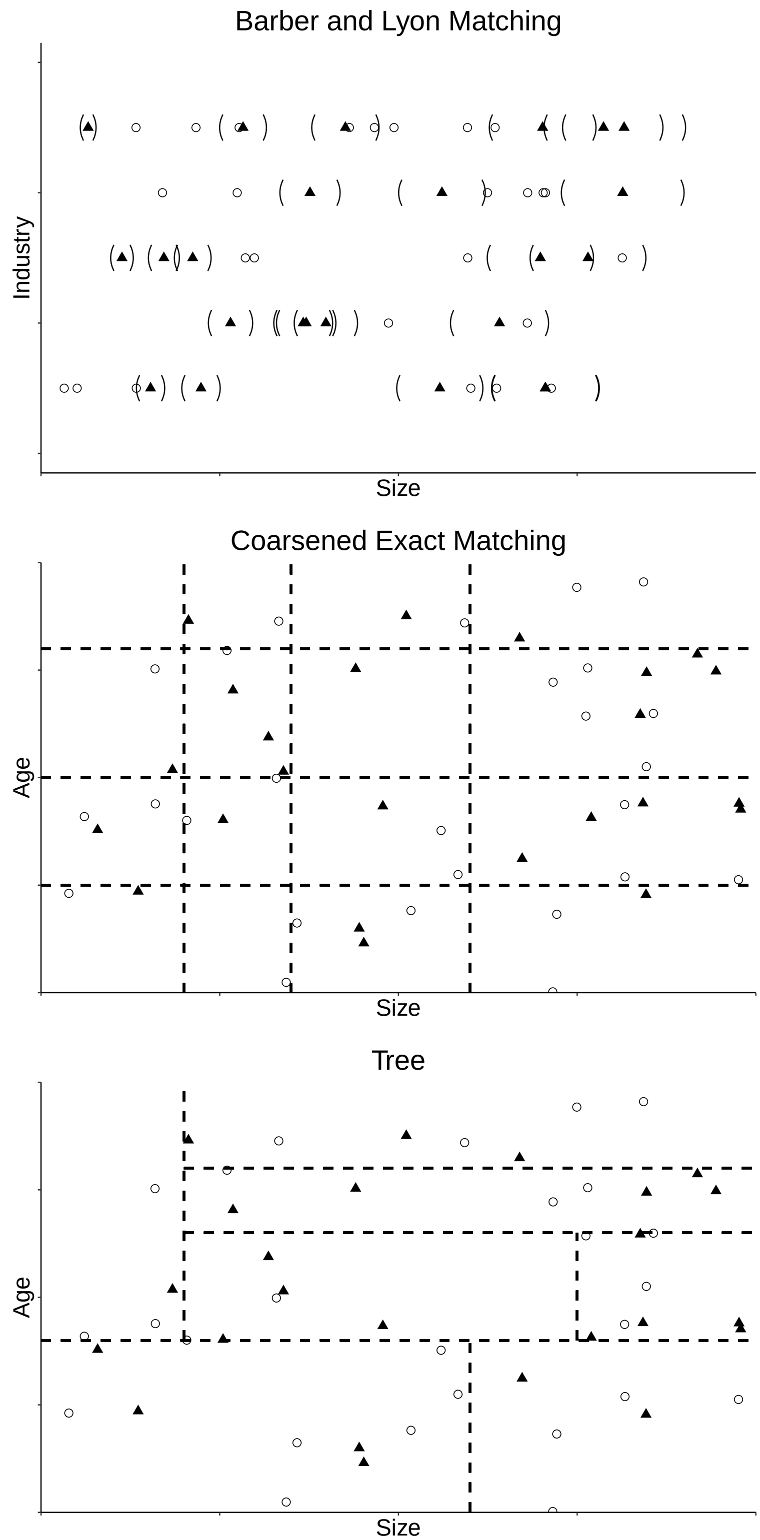


Figure 2. Example forest aggregation. Figure of an example forest weighting in covariate space. The top left plot shows simulated data with a binary outcome variable and two continuous covariates, variable 1 and variable 2. Treated and control observations are indicated with triangles and open circles, respectively. An example data point is plotted in a solid blue square. The top right and middle left and right plots (tree 1, tree 2, and tree 3, respectively) show example partitions from trees in a 1,000 tree forest. In each tree plot, the dashed lines indicate partitions. Points contained in the final partition, or terminal node, as the data point of interest (square) are larger and shaded more heavily. The bottom plot demonstrates how likely each point is to be in the same terminal node as the example point (square) by shading and size. Larger, darker points are more likely to be contained in the same terminal node as the example point. This figure was inspired by a figure in Athey et al. (2019).

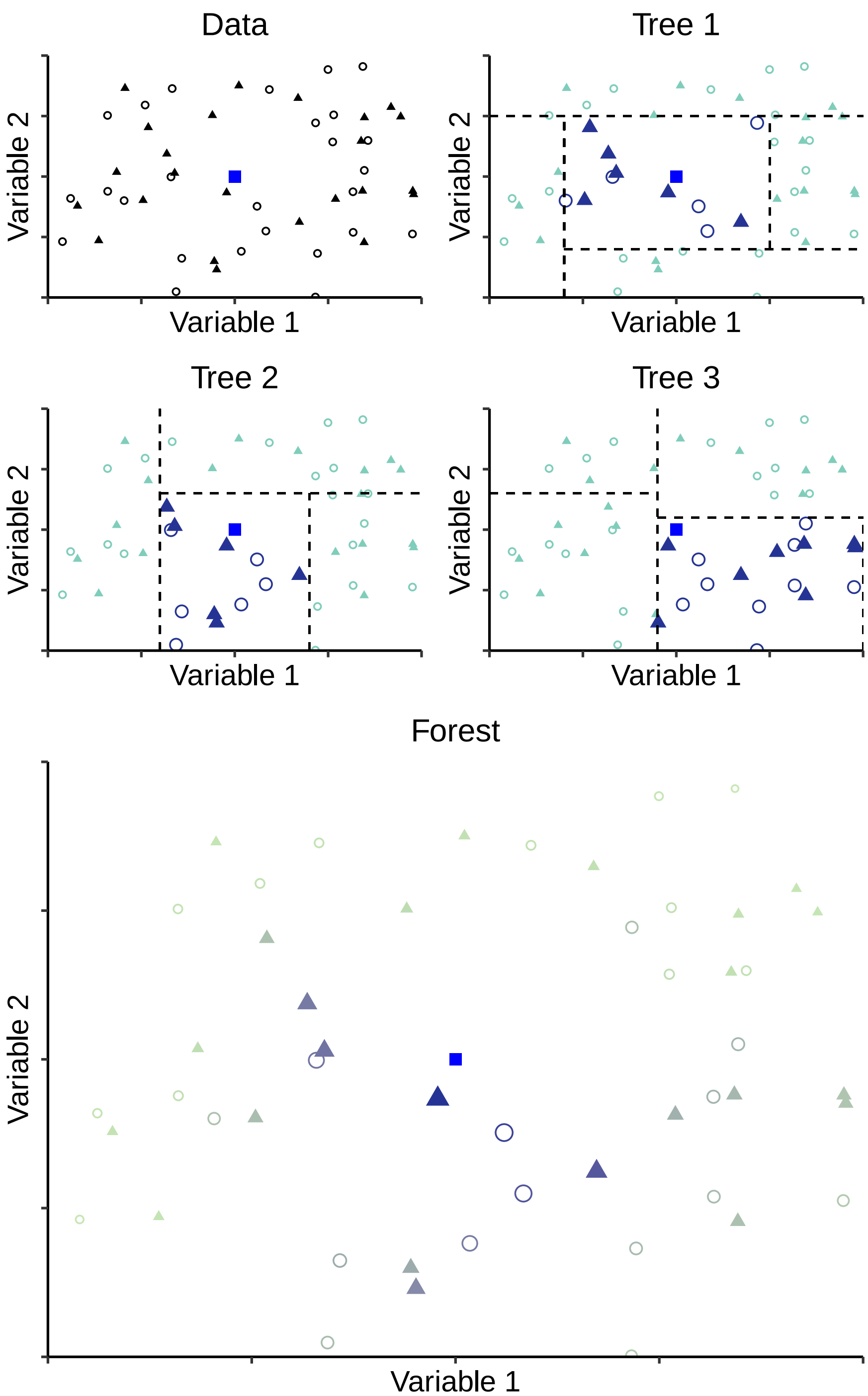


Figure 3. Densities of bias in estimates from causal forest and RDD in finite samples. Bias densities and interquartile ranges from simulations for the specifications detailed in the panel titles. Details on data simulation are in internet appendix Section D.

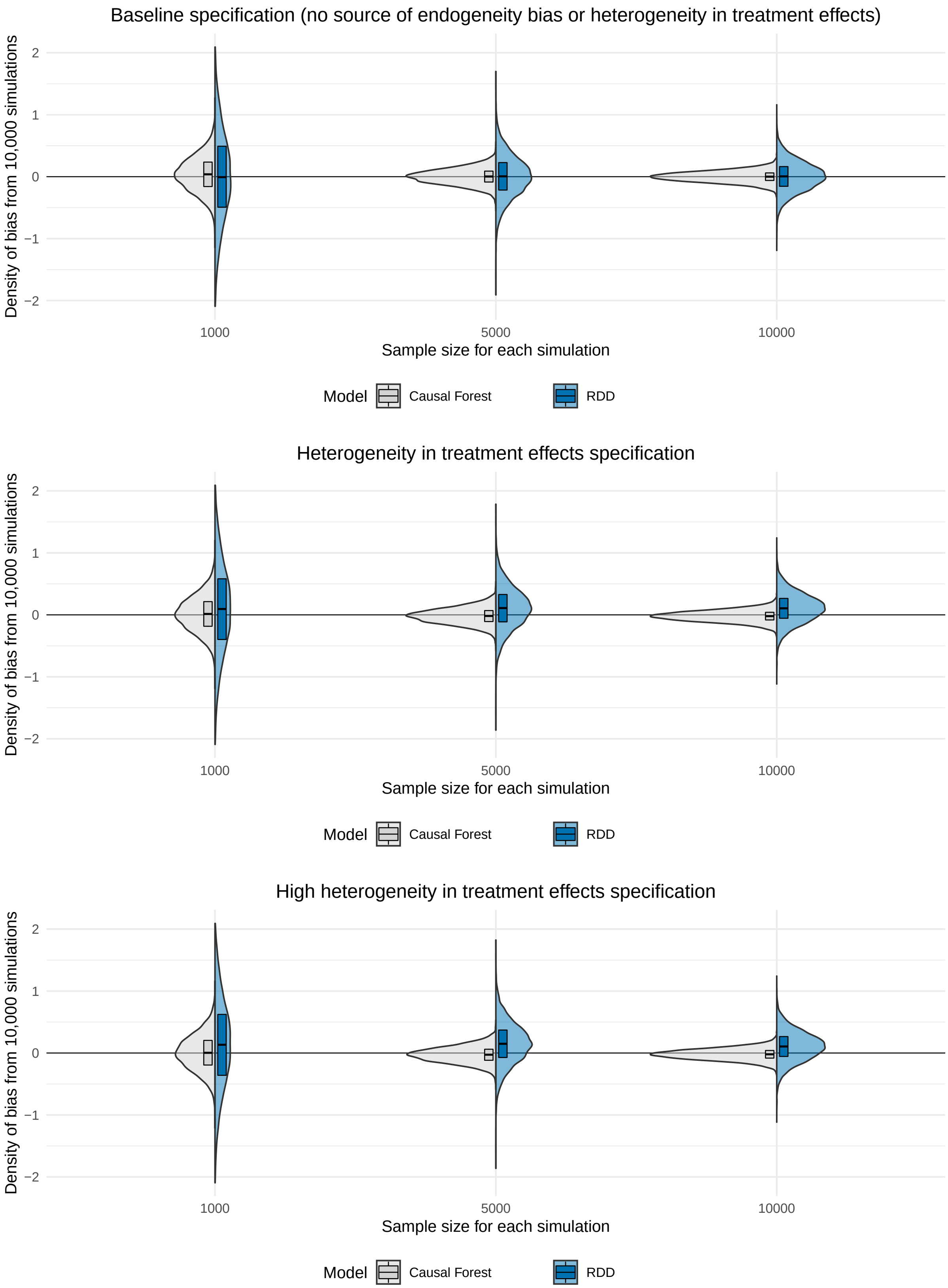


Figure 4. Densities of bias in estimates from causal forest and RDD in finite samples: latent variable specifications. Bias densities and interquartile ranges from simulations for the specifications detailed in the panel titles. Details on data simulation are in internet appendix Section D.

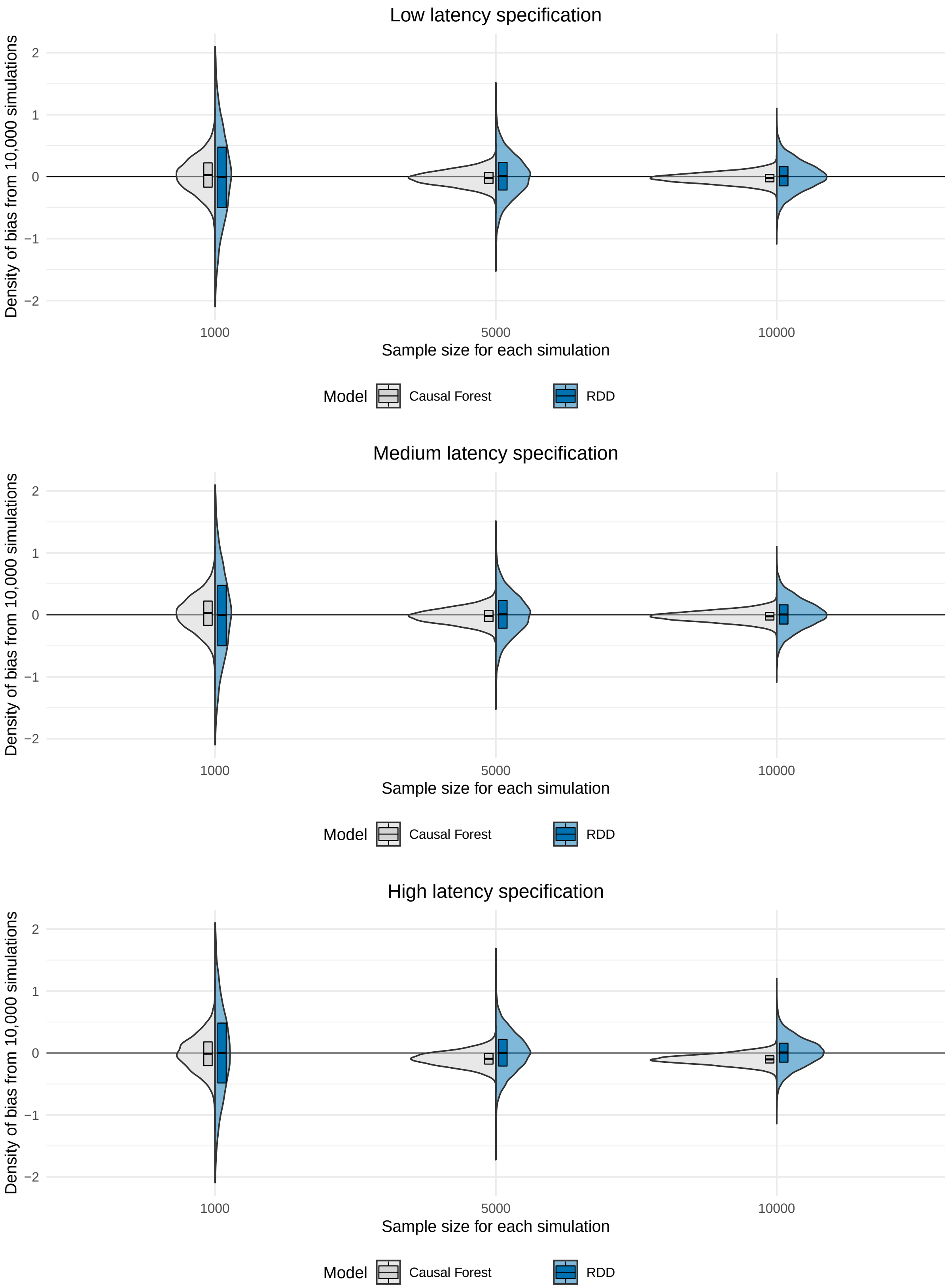


Figure 5. Densities of bias in estimates from causal forest and RDD in finite samples: manipulation specifications. Bias densities and interquartile ranges from simulations for the specifications detailed in the panel titles. Details on data simulation are in internet appendix Section D.

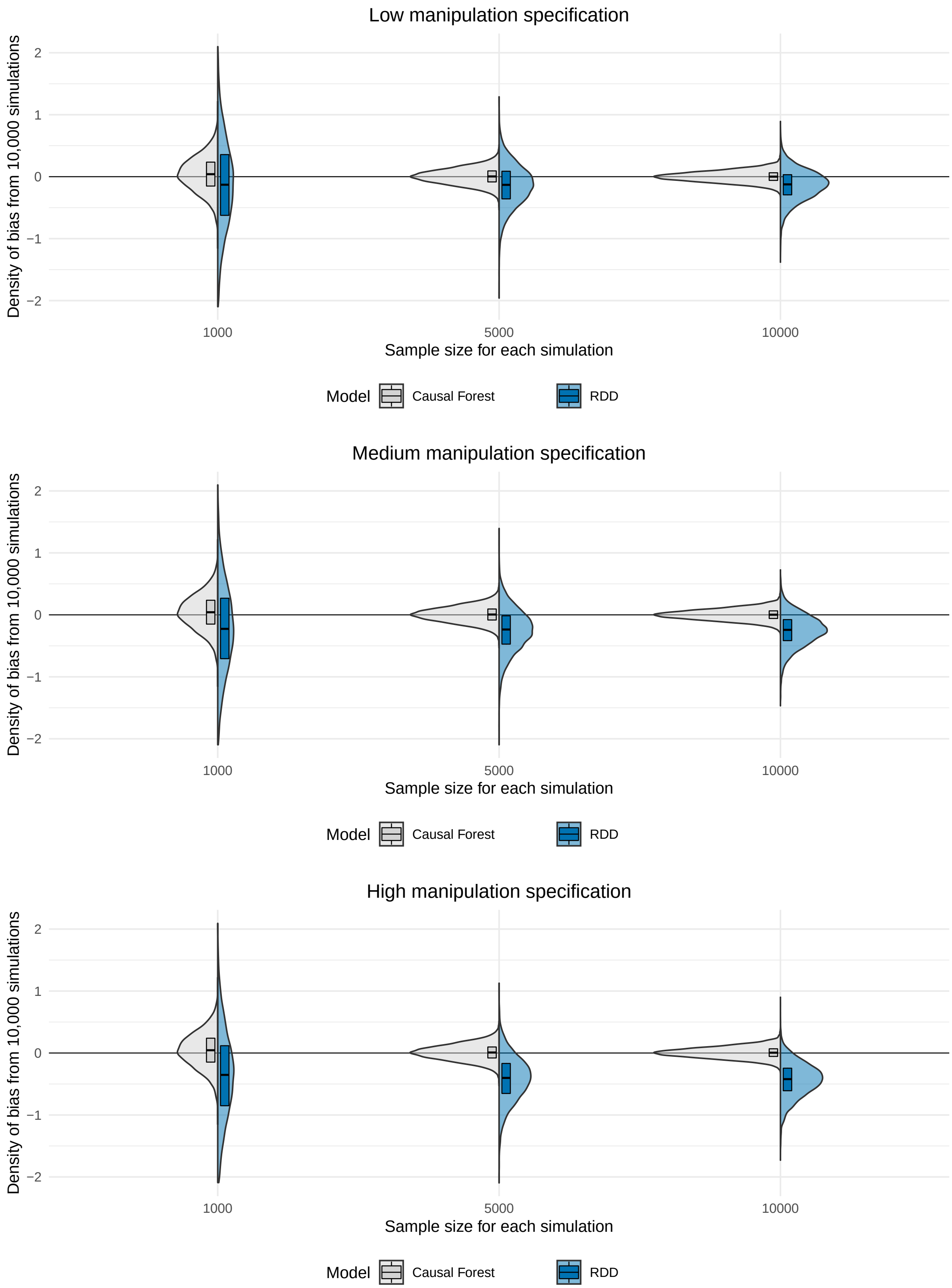


Figure 6. Densities of bias in estimates from causal forest and RDD: latent variable and manipulation specifications, 2,500 obs. Bias densities and interquartile ranges from simulations for the specifications detailed in the panel titles. Details on data simulation are in internet appendix Section D.

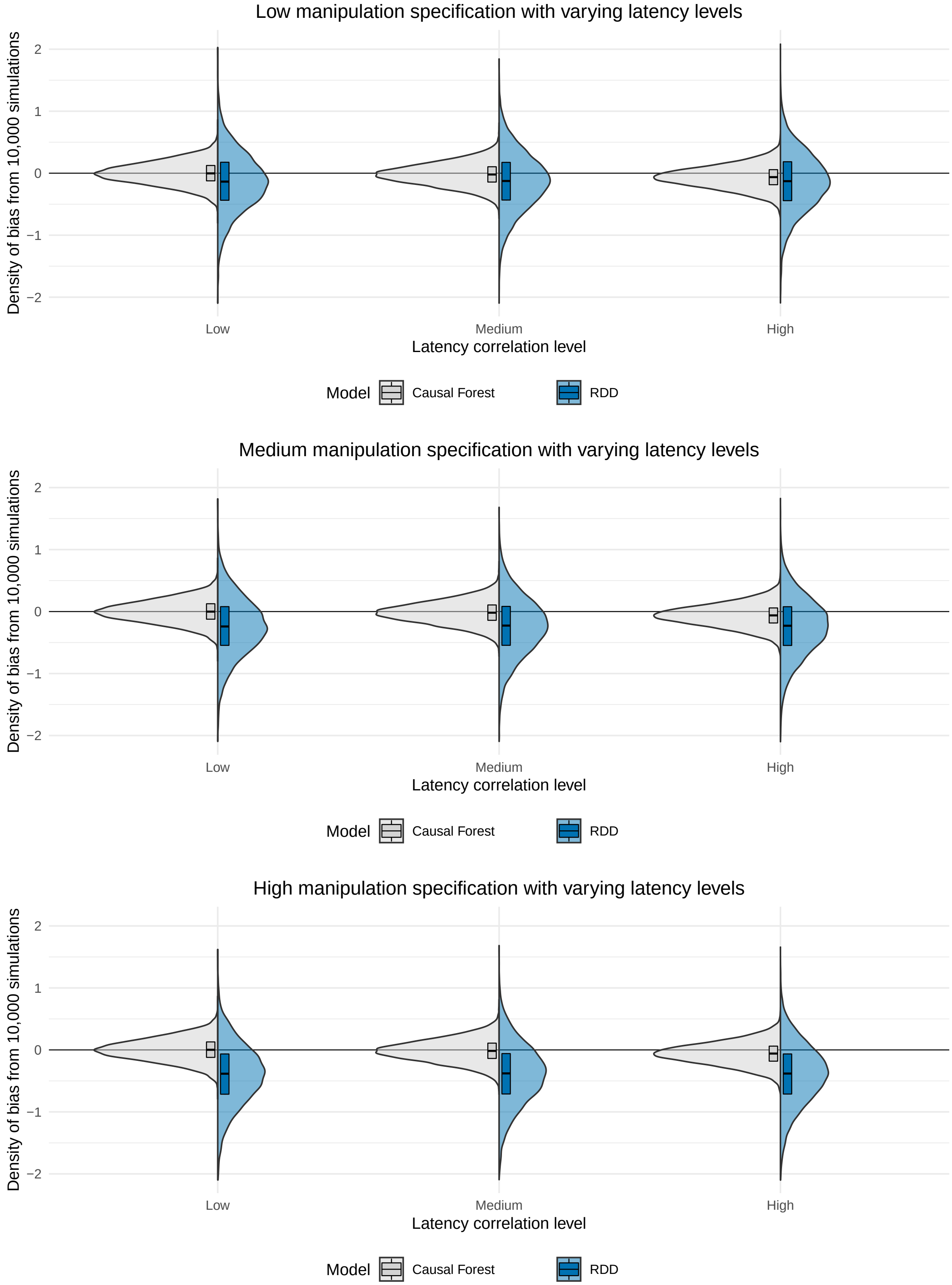


Figure 7. HTE density plot. Density plot of heterogeneous treatment effects estimated in the causal forest estimation described in the caption of Table 3.

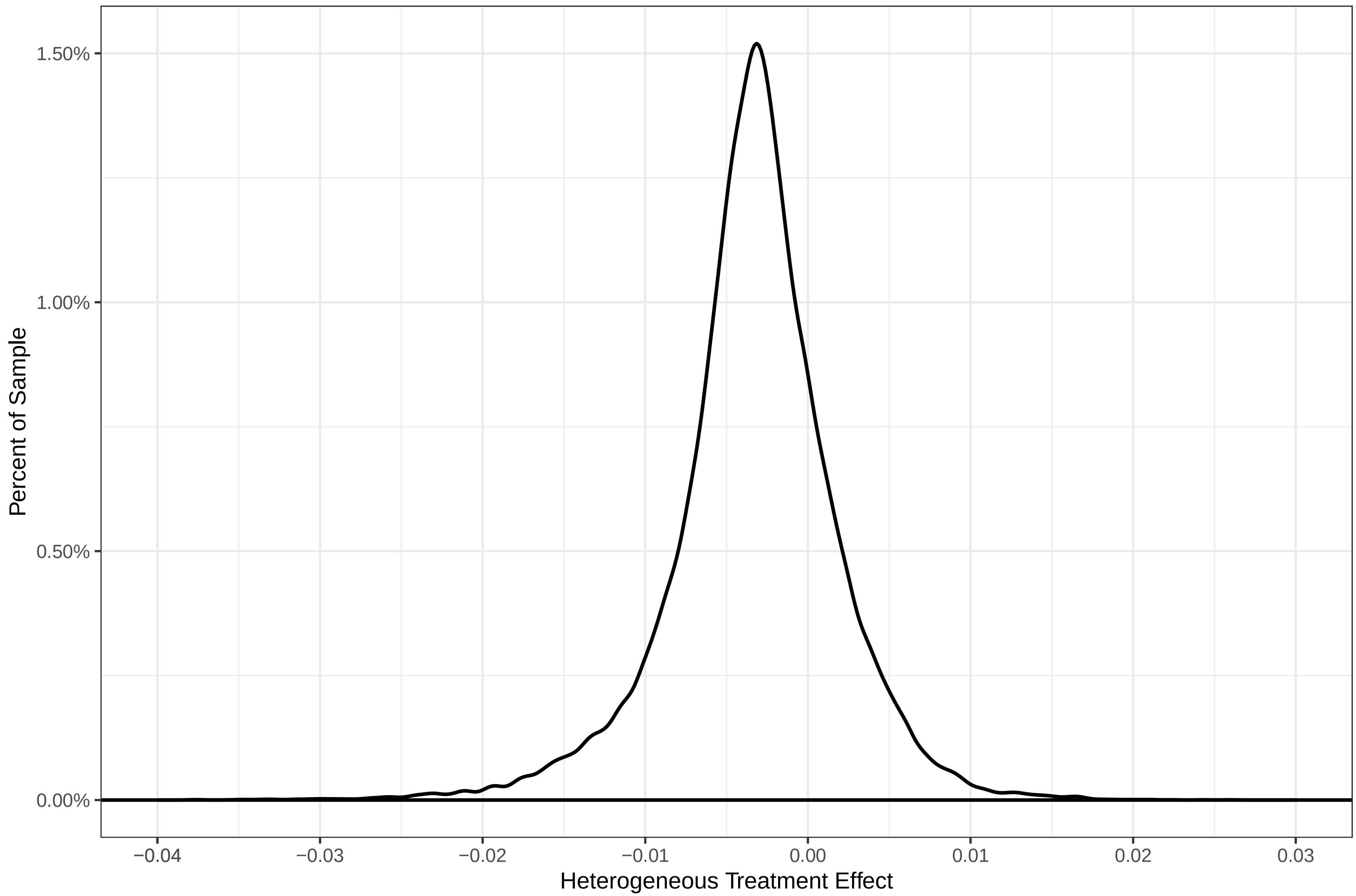


Figure 8. Treatment effect comparative statics. Comparative statics for a set of covariates for the causal forest estimate detailed in the caption of Table 3. In each panel, we set covariate values to the sample geometric median vector, then generate treatment effect predictions while varying the covariate on the x-axis over its support.

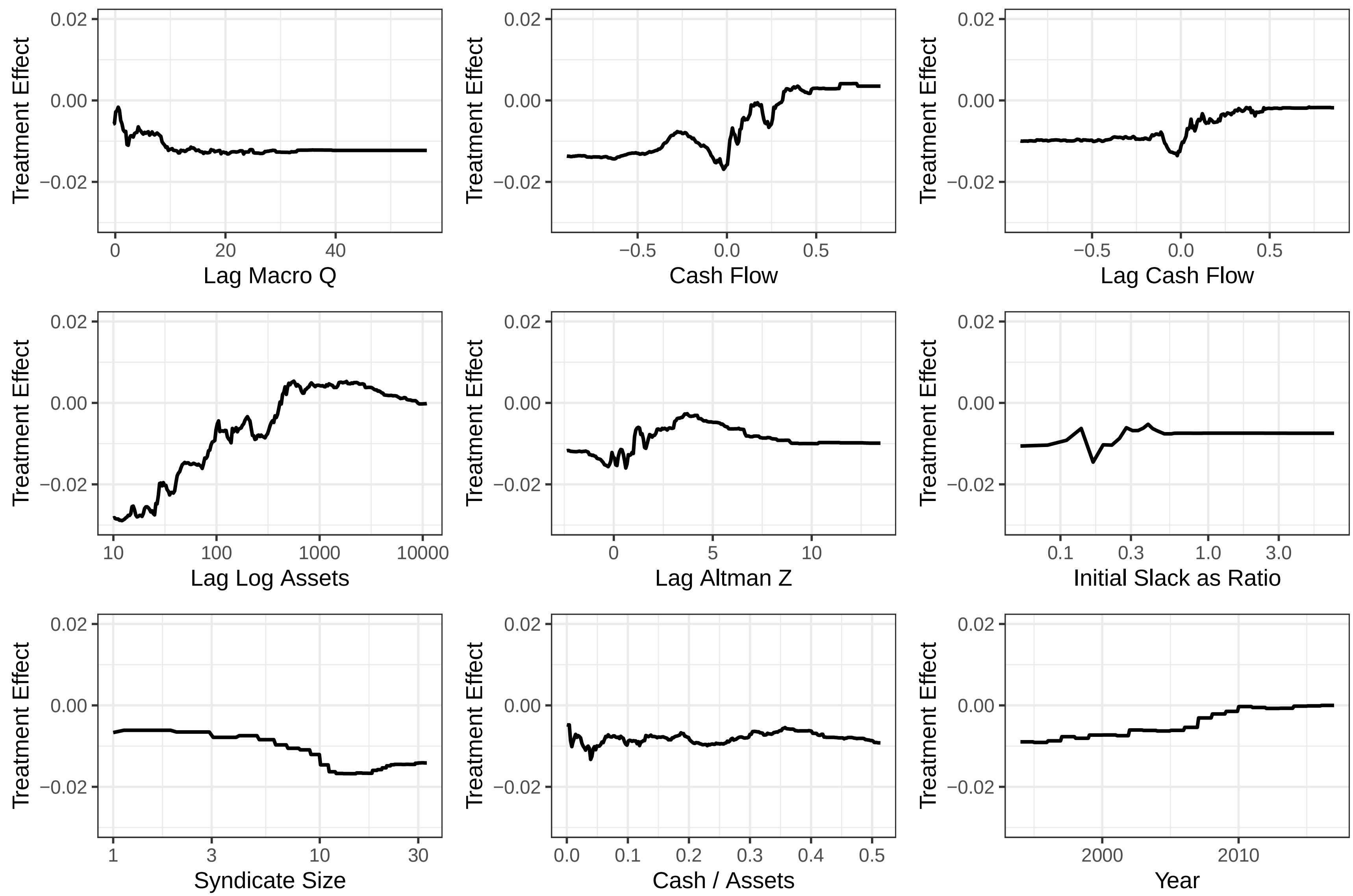
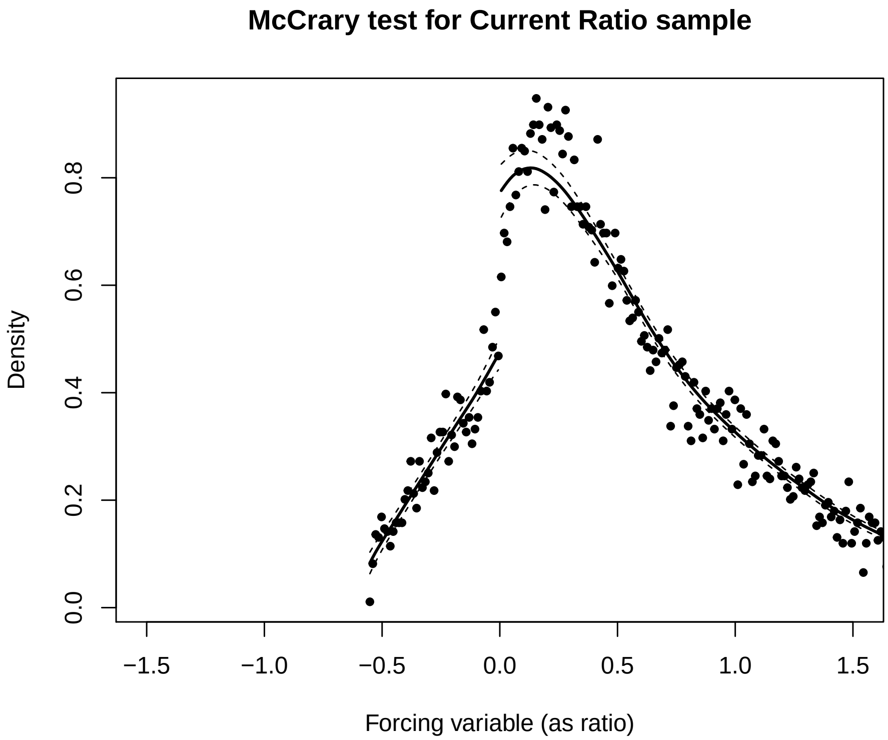


Figure 9. McCrary test results. McCrary (2008) tests of manipulation at the threshold for current ratio (top panel) and net worth (bottom panel) forcing variables expressed as ratios (percentages of threshold) minus one, so that the default threshold equals zero. In both panels, the null of no manipulation is rejected (p-values of 6.2E-5 and 2.53E-23 for the net worth and current ratio samples, respectively).



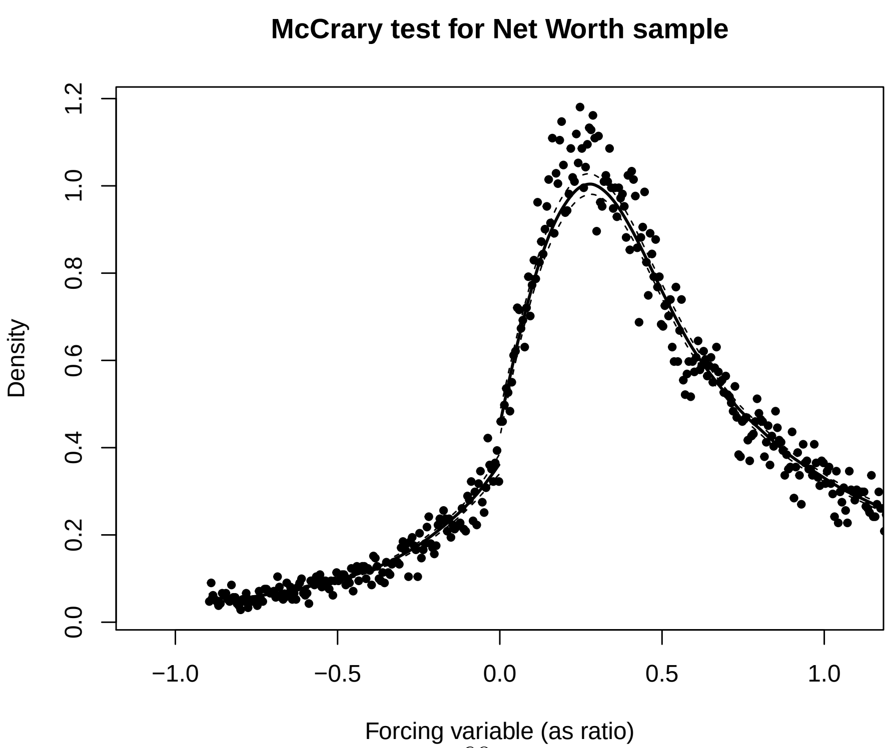


Table 1. Percent of RDD estimates with greater bias than causal forest estimates. Statistics are reported in Panels A, B, and C for simulations in Figure 3, Panel D for simulations in Figure 4, and Panel E for simulations in Figure 5.

		RDD estimates with greater bias than [column header] of causal forest estimates (%)				McCrary test fail (%)
	n	100%	99%	98%	95%	
Panel A: Baseline						
–	1,000	15.30	31.47	35.92	44.32	6.59
–	5,000	13.82	31.75	35.80	44.04	6.11
–	10,000	14.83	32.96	37.82	45.05	5.62
Panel B: Heterogeneity in treatment effects						
–	1,000	15.70	31.41	35.93	43.94	6.59
–	5,000	16.39	35.11	40.16	47.38	6.11
–	10,000	19.96	38.29	43.81	51.45	5.62
Panel C: High heterogeneity in treatment effects						
–	1,000	17.01	31.91	36.42	44.39	6.59
–	5,000	18.44	38.48	42.60	50.44	6.11
–	10,000	24.48	44.33	49.39	56.49	5.62
Panel D: Latency specifications						
low	1,000	16.06	30.60	36.02	44.10	6.38
low	5,000	13.35	30.70	36.77	44.29	6.06
low	10,000	14.46	32.60	37.06	44.32	5.55
medium	1,000	15.99	30.81	36.29	44.36	6.33
medium	5,000	13.18	32.13	37.19	44.48	6.20
medium	10,000	15.02	32.81	38.59	45.85	5.77
high	1,000	13.96	31.38	36.18	44.37	6.11
high	5,000	15.82	33.37	38.07	45.76	6.28
high	10,000	22.07	38.16	43.00	49.71	6.36
Panel E: Manipulation specifications						
low	1,000	16.15	32.61	37.74	45.85	28.30
low	5,000	18.31	35.71	40.31	48.25	83.17
low	10,000	24.92	39.91	44.14	50.58	98.12
medium	1,000	17.46	34.33	38.96	47.17	67.80
medium	5,000	26.13	44.07	48.75	55.41	99.97
medium	10,000	40.36	55.70	59.51	65.32	100.00
high	1,000	22.00	39.00	43.69	51.80	99.27
high	5,000	42.64	60.66	64.63	70.45	100.00
high	10,000	67.66	78.41	80.92	84.27	100.00

Table 2. **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 3. **Causal forest results.** 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	-1.323	0.186	32,530
ATT	-0.55^{***}	0.09	-5.845	<0.001	32,530
ATC	-0.19	0.21	-0.877	0.381	32,530
ATO	-0.54^{***}	0.10	-5.475	<0.001	32,530

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

Table 4. **Ranking of covariates by importance to tree growth, predicting magnitude of treatment effect.** Variable importance ranking in the net worth sample (quarterly observations, 1994 – 2017) for the causal forest estimation detailed in the caption of Table 3. To calculate variable importance, we select a variable, say log(assets). We hold all other variables constant and randomly assign log(assets) to observations. Then, we predict treatment effects for the permuted observations using our estimated causal forest 100 times. The Absolute Mean Difference column is the average absolute value of differences between treatment effects calculated using permuted data and our original treatment effect. The Standard Deviation of Differences column is the standard deviation of the differences between treatment effects calculated using our permuted data and our original treatment effect. The Ratio of Difference Variance to HTE Variance is the variance of differences between permuted treatment effects and our original treatment effects scaled by the variance of the heterogeneous treatment effects originally calculated in our causal forest estimation. We repeat this procedure for each covariate in turn to calculate variable importance for each covariate. Covariates are ranked by their relative importance to treatment effect calculation.

	Absolute Mean Difference	Standard Deviation of Differences	Ratio of Difference Variance to HTE Variance
Lag assets	0.26	0.46	0.810
Cash flow	0.24	0.38	0.562
Lag Macro Q	0.21	0.34	0.437
Firm fixed effect	0.20	0.32	0.393
Lag cash flow	0.17	0.26	0.256
Lag Altman Z	0.16	0.26	0.255
Cash / assets	0.14	0.24	0.227
Initial distance to default (slack)	0.13	0.22	0.189
Distance to default (slack)	0.09	0.16	0.100
Year	0.10	0.15	0.091
Syndicate size	0.08	0.13	0.066
Quarter	0.05	0.08	0.026
Credit rating (binary)	0.03	0.08	0.024

Table 5. Average HTEs in subsamples based on cash flow and Macro Q. 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 3.

	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					
ATE	-0.49^{**}	0.20	-2.391	0.017	1,631	ATE	-0.34	0.22	-1.585	0.113	9,212
ATT	-0.32	0.21	-1.533	0.125	1,631	ATT	-0.30^{**}	0.13	-2.244	0.025	9,212
ATC	-0.54^{**}	0.23	-2.399	0.016	1,631	ATC	-0.34	0.25	-1.405	0.160	9,212
ATO	-0.35	0.22	-1.581	0.114	1,631	ATO	-0.28^{*}	0.16	-1.740	0.082	9,212
Panel B: High Macro Q, Cash flow < 0						Panel D: High Macro Q, Cash flow ≥ 0					
ATE	-1.44^{***}	0.43	-3.324	0.001	1,407	ATE	-0.07	0.53	-0.125	0.901	9,436
ATT	-1.49^{***}	0.47	-3.191	0.001	1,407	ATT	-0.59^{**}	0.29	-2.001	0.045	9,436
ATC	-1.40^{***}	0.48	-2.899	0.004	1,407	ATC	-0.01	0.58	-0.019	0.985	9,436
ATO	-1.43^{***}	0.48	-2.946	0.003	1,407	ATO	-0.65^{**}	0.29	-2.246	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 6. Average HTEs in subsamples based on cash flow and size. 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 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 3.

	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					
ATE	-0.98^{***}	0.24	-4.071	<0.001	2,290	ATE	0.33	0.57	0.580	0.562	8,553
ATT	-0.91^{***}	0.27	-3.389	0.001	2,290	ATT	-0.34^{*}	0.19	-1.843	0.065	8,553
ATC	-0.98^{***}	0.26	-3.752	<0.001	2,290	ATC	0.50	0.67	0.754	0.451	8,553
ATO	-0.91^{***}	0.27	-3.306	0.001	2,290	ATO	-0.38^{*}	0.21	-1.833	0.067	8,553
Panel B: Large firms, Cash flow < 0						Panel D: Large firms, Cash flow ≥ 0					
ATE	-0.66	0.47	-1.402	0.161	742	ATE	-0.16	0.28	-0.572	0.567	10,101
ATT	-0.64^{*}	0.33	-1.954	0.051	742	ATT	-0.20	0.20	-0.999	0.318	10,101
ATC	-0.66	0.52	-1.256	0.209	742	ATC	-0.15	0.29	-0.535	0.592	10,101
ATO	-0.67^{*}	0.35	-1.930	0.054	742	ATO	-0.16	0.22	-0.760	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 7. Ranking of variable importance to tree growth, predicting treatment status and investment. Variable importance 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.

	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 8. RDD regression results. Results for a regression of investment (quarterly capital expenditures divided by beginning-of-period PPE and scaled by 100 for ease in interpretation) on a binary variable (*bind*) equal to one if a firm is in default in a quarter. In the columns indicated, control variables included are: firm size, Macro Q and cash flow. Variable descriptions are given in the caption of Table 2. Year-quarter and firm fixed effects are included in each specification. Where indicated by the column header, all powers of slack are included up to the highest polynomial (i.e., if the highest polynomial indicated in the column header is "third", slack-squared and slack are also included). If interactions between bind and slack are included, all slack variables are interacted with bind. Bandwidth sizes for regressions in Panel C are Calonico et al. (2020) optimal bandwidths. Standard errors are in parentheses.

			Highest polynomial of slack included			
			first	second	third	fourth
Panel A: Baseline regressions						
bind	-1.22^{***}	-0.94^{***}	-0.82^{***}	-0.70^{***}	-0.55^{***}	-0.60^{***}
	(0.09)	(0.10)	(0.10)	(0.11)	(0.11)	(0.12)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	42,439	36,246	36,246	36,246	36,246	36,246
R^2	0.42	0.46	0.46	0.46	0.46	0.46
Panel B: Including *bind* and slack interactions						
bind	–	–	-0.47^{***}	-0.24	-0.06	-0.27
			(0.14)	(0.18)	(0.22)	(0.26)
controls	–	–	yes	yes	yes	yes
slack	–	–	yes	yes	yes	yes
interactions	–	-	yes	yes	yes	yes
Obs.	–	–	36,246	36,246	36,246	36,246
R^2	–	–	0.46	0.46	0.46	0.46
Panel C: Threshold sample						
bind	-0.44^{***}	-0.40^{***}	0.20	0.27	0.42	0.35
	(0.13)	(0.14)	(0.21)	(0.22)	(0.27)	(0.29)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	9,841	8,394	8,394	8,394	8,394	8,394
R^2	0.53	0.57	0.57	0.57	0.57	0.57

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

Table 9. RDD results estimated with triangular kernels. 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) for sample described in the caption of Table 2. 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
bind	0.10	0.24	0.31	0.21	0.13
	(0.23)	(0.26)	(0.34)	(0.42)	(0.47)
Bandwidth	0.34	0.25	0.32	0.36	0.43
Obs. below	2,769	2,288	2,657	2,846	3,168
Obs. above	10,694	7,553	9,878	11,410	13,944

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. (2020)	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.	dividend policy	Bulan and Hull (2013)	any	quarterly	yes	unscaled distance to default	no
7.	R&D and innovation	Chava et al. (2017)	any	quarterly	yes	Chava and Roberts (2008)	no
8.	CDS spreads	Chen, Kim, and Zhu (2017)	any	quarterly	yes	distance scaled by SD(20)*	no
9.	non-GAAP reporting practices	Christensen, Pei, Pierce, and Tan (2019)	any	quarterly	no	N/A	N/A
10.	corporate tax avoidance	Cook, Ma, and Zhao (2019)	any	annual	no	N/A	N/A
11.	employment; sales and closures of establishments	Ersahin, Irani, and Le (2020)	new	annual	no	N/A	N/A
12.	employment	Falato and Liang (2016)	any	annual	yes	distance scaled by threshold	yes
13.	independent directors	Ferreira, Ferreira, and Mariano (2018)	any	annual	yes	distance scaled by threshold	no
14.	tax avoidance	Francis, Shen, and Wu (2017)	new	annual	no	N/A	N/A
15.	future loan contracts	Freudenberg, Imbierowicz, Saunders, and Steffen (2017)	any	both	yes	difference scaled by SD(12)*	no
16.	information asymmetry	Gao, Khan, and Tan (2017)	new	quarterly	no	N/A	N/A
17.	innovation	Gu et al. (2017)	new	annual	yes	Chava and Roberts (2008)	yes
18.	CSR activities	He, Zhang, and Zhong (2018)	any	annual	no	N/A	N/A
19.	audit verification	Jiang and Zhou (2016)	any	annual	no	N/A	N/A
20.	banking relationships	Keil (2018)	any	quarterly	yes	unscaled distance to default	no
21.	monitoring; savings behavior	Lin, Xin, Zhang, and Zhang (2017)	new	annual	no	N/A	N/A
22.	investment (capital expenditures and acquisitions); leverage; payout; turnover	Nini et al. (2012)	new	quarterly	yes	unclear	no
23.	competitors' investment and financing	Nordlund (2018)	new	quarterly	no	N/A	N/A
24.	conservatism	Tan (2013)	any	quarterly	yes	Chava and Roberts (2008)	no
25.	disclosure	Vashishtha (2014)	any	quarterly	no	N/A	N/A
26.	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...							
27.	credit conditions; enforcement	Bird, Ertan, Karolyi, and Ruchti (2017)	any	quarterly	yes	distance scaled by SD(8)*	no
28.	short-termism; enforcement	Bird, Ertan, Karolyi, and Ruchti (2019)	any	quarterly	no	distance scaled by SD(8)*	no
29.	whether a firm has CDSs; investment and bankruptcy	Chakraborty, Chava, and Ganduri (2015)	any	quarterly	yes	unclear	no
30.	dual ownership; investment	Chava, Wang, and Zou (2019)	any	quarterly	no	N/A	N/A
31.	distracted loan officers; investment and default rates	Gao, Karolyi, and Pacelli (2018)	any	quarterly	no	N/A	N/A
Panel C: Other studies							
32.	violation shocks monitoring intensity	Colonnello, Koetter, and Stieglitz (2019)	new	quarterly	yes	distance scaled by threshold	yes

*Note: SD(n) = standard deviation of the accounting variable over the previous n quarters.

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.

Table IA.2. List of parameters for causal forest

parameter	description
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

C Causal Forest technical appendix

C.1 Details of causal forest parameterization

The goal of this section is to fully describe our parameterization and the reasoning behind our choice of parameters. 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 a few of these important covariates are not included in the randomly selected subset of covariates, trees cannot be estimated and will be 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.

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 (2017). This estimation yields an *alpha*, 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.

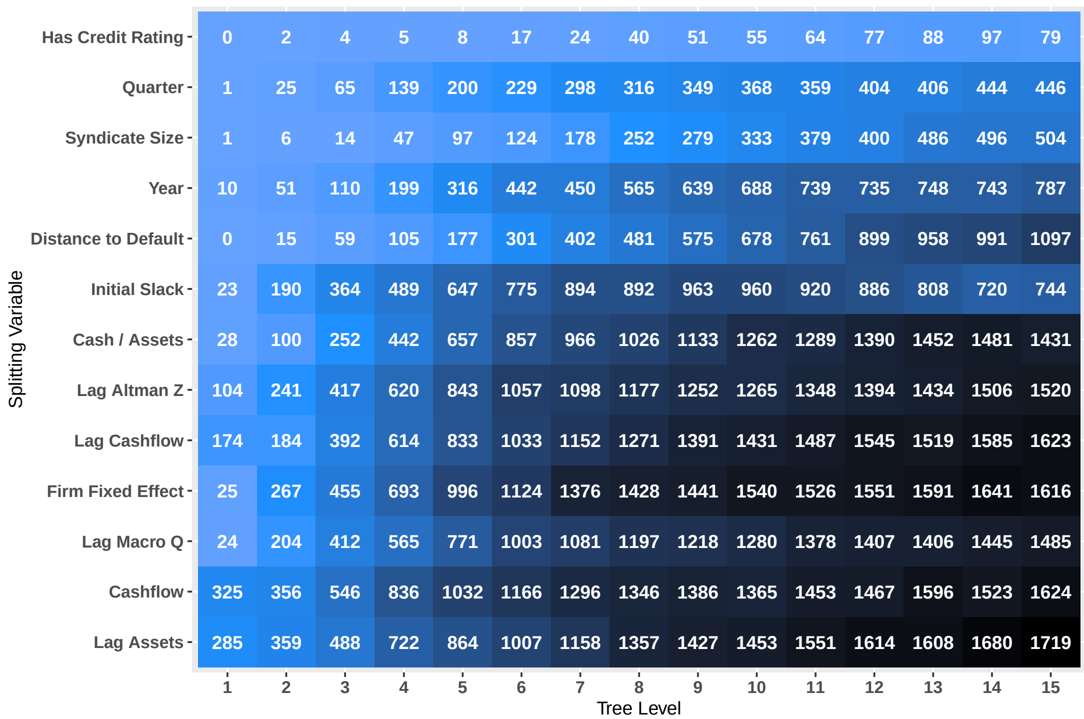
C.2 Randomized splitting

Figure IA.1 demonstrates how the randomized splitting rule affects our estimation by showing the count (top panel) and percentage of observations (bottom panel) of each covariate (vertical axis) for each level of tree (horizontal axis). Covariates with higher counts and percentages are shaded darker, while covariates with lower counts and percentages are lighter. With a randomized splitting rule, a set of covariates is randomly selected at each node, and the tree splits along the covariate more important to treatment status. In comparison, a "greedy" tree always splits along the most important variable determining treatment and can overfit the data.

As Figure IA.1 shows, the most frequent covariate at the highest level of trees is size measured by cash flow, but this covariate only appears at level one in 325 of 1,000 trees. Even though cash flow appears to be a key variable in determining treatment status, cash flow is not always included in the first set of variables on which the tree splits. Thus, other variables, including size, lagged cash flow, and Altman Z score, also rank highly in level one. In contrast, a binary credit rating variable is never the first splitting variable; syndicate size, quarter, and year are very infrequently the first splitting variable. As additional levels of trees are grown, cash flow remains a key variable on which to split, but other variables grow in importance. Note that the total count of covariates in level one is 1,000, which matches the number of trees in our analysis, but the covariate count in subsequent levels is determined by how each individual tree develops.

Figure IA.1 shows only the top 15 levels of trees, but trees can grow deeper than 15 levels. Figure IA.2 expands Figure IA.1 to 40 levels. In Figure IA.2, covariates with higher counts and percentages are again shaded darker, and covariates with lower counts and percentages are lighter. Expanding to 40 levels shows that distance to default, syndicate size, and year become more important to tree growth in deeper levels. In contrast, the binary credit rating variable remains unimportant even when 40 levels of tree growth are examined.

Figure IA.1. Covariates by level in trees. Count and frequency of each covariate appearing at each level of tree in the causal forest estimation described in the caption of Table 3. For example, of 1,000 trees, contemporaneous cash flow appears 325 in the first level (top panel), which comprises 32.5% of covariates at that level (bottom panel). Lighter squares indicate lower counts (top panel) or percentages (bottom panel) at each level; darker squares indicate higher counts (top panel) or percentages (bottom panel) at each level. Number of observations in each level depends on previous splitting and the depth of trees. This figure shows up to 15 levels; Figure IA.2 shows up to 40 levels.



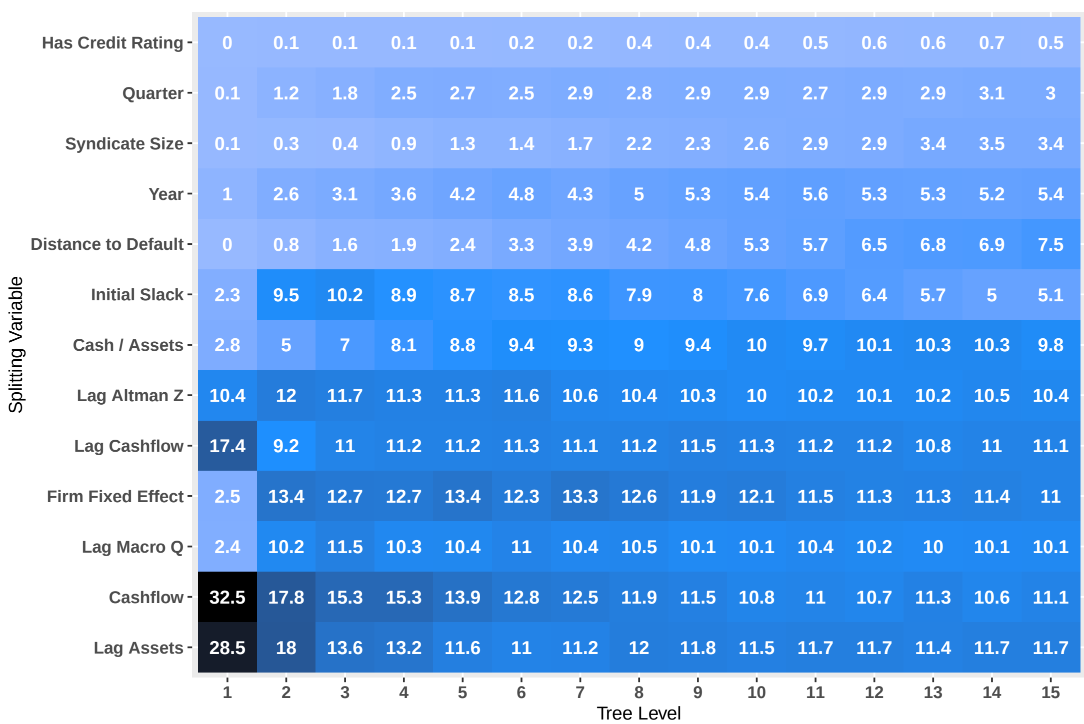
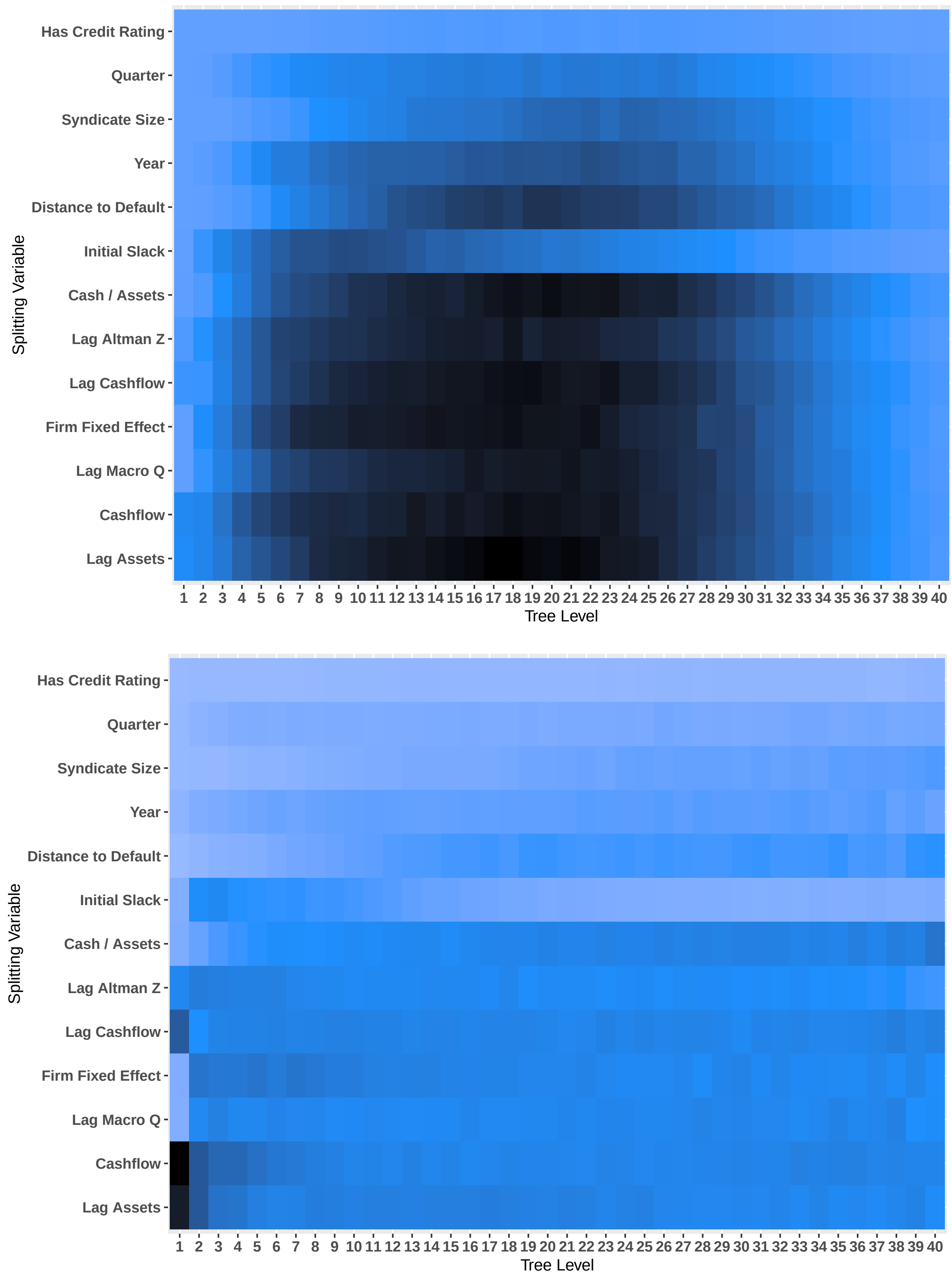


Figure IA.2. Covariates by level in trees up to 40 levels. This figure expands Figure IA.1 to 40 levels of trees. Lighter squares indicate lower counts (top panel) or percentages (bottom panel) at each level; darker squares indicate higher counts (top panel) or percentages (bottom panel) at each level.



C.3 Augmented Inverse Propensity Weighting (AIPW)

Because simple averages of treatment effects can have poor finite sample properties, we use augmented inverse propensity weighting (AIPW) to calculate these averages (Robins, Rotnitzky, and Zhao, 1994). The AIPW estimator for ATE is a simple average across treatment effects ($\hat{\tau}_i$) plus an adjustment factor (δ):

$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}_i + \delta, \tag{13}$$

where the adjustment factor,

$$\delta = \frac{1}{n}\sum_{i=1}^{n}\gamma_i\left[W_i(Y_i - \hat{Y}(X_i, W_i = 1)) - (1 - W_i)(Y_i - \hat{Y}(X_i, W_i = 0))\right]. \tag{14}$$

In Equation (14), $\hat{Y}(X_i, W_i)$ is the predicted outcome given a firm's covariates and treatment status, and γ_i is the normalized inverse probability weights.

$$\gamma_i = W_i\frac{(\hat{W}(X_i))^{-1}}{\frac{1}{|i:W_i=1|}\sum_{i:W_i=1}(\hat{W}(X_i))^{-1}} + (1 - W_i)\frac{(1 - \hat{W}(X_i))^{-1}}{\frac{1}{|i:W_i=0|}\sum_{i:W_i=0}(1 - \hat{W}(X_i))^{-1}}. \tag{15}$$

In Equation 15, $\hat{W}(X_i)$ is the estimated treatment propensity for firm i given covariates X_i.

The AIPW estimators for ATT and ATC are similarly calculated, but replace γ_i in δ for ATT with:

$$\gamma_i = W_i + (1 - W_i)\frac{\hat{W}(X_i)(1 - \hat{W}(X_i))^{-1}}{\frac{1}{|i:W_i=0|}\sum_{i:W_i=0}\hat{W}(X_i)(1 - \hat{W}(X_i))^{-1}} \tag{16}$$

and, for ATC with:

$$\gamma_i = W_i\frac{(1 - \hat{W}(X_i))(\hat{W}(X_i))^{-1}}{\frac{1}{|i:W_i=1|}\sum_{i:W_i=1}(1 - \hat{W}(X_i))(\hat{W}(X_i))^{-1}} + (1 - W_i). \tag{17}$$

In our application, we also calculate an overlap-weighted estimator (ATO) from Li, Morgan, and Zaslavsky (2018), which uses an alternative finite-sample adjustment:

$$\widehat{ATO} = \frac{\sum_{i=1}^{n}(W_i - \hat{W}_i)(Y_i - \hat{Y}_i)}{\sum_{i=1}^{n}(W_i - \hat{W}_i)^2}. \tag{18}$$

Essentially, this ATO calculation balances individual treatment effect estimates according to the estimated treatment propensity of the individual, with more weight being given to observations with treatment propensities near 0.5. Intuitively, this estimator puts more weight on portions of the sample in which there is more overlap between treated and control observations. Estimates from these areas of the covariate space are theoretically more informative. We calculate ATE, ATT, ATE, and ATO using the `grf` package.

D Simulated data for Monte Carlos

We simulate data with a second order polynomial as the functional form:

$$y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025x_1^2 - 0.01x_2^2 + 0.015x_3^2 + 0.02D + \varepsilon,$$

in which y is the dependent variable, the three x variables are additional covariates, D is a binary variable equal to one if the forcing variable d is greater than zero, and ε is the mean zero error term. The coefficients on x_1, x_2, and x_3 are similar to the data we use for our application estimation. Macro q, firm size, and cash flow are approximately represented as x_1, x_2, and x_3, respectively, and d is slack, or distance to technical default. The coefficients on the second order terms are chosen so that each has an important effect, but less than the first order terms.

We simulate the x and d variables using a multivariate normal distribution. For the vector (x_1, x_2, x_3, d), 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:

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

This correlation structure is selected to mimic correlation between the covariates in our simulation data. The ε term has a mean of zero and a standard deviation of 0.065. We estimate simulated samples of sizes 1,000, 2,500, 5,000, and 10,000.

In our baseline simulations, treatment effect is a constant (0.02), but we also consider heterogeneous treatment effects. In these simulations, an observation's treatment effect (τ_i) is a function of its covariates:

$$\tau_i = 0.02 + [0.01(x_{i1} - \bar{x_1}) - 0.01(x_{i3} - \bar{x_3}) - 0.01(D_i - \bar{D})]\gamma$$

where γ is a multiplier that magnifies the heterogeneity in treatment effects. In our initial heterogeneity specifications, we set γ equal to one. In our "high heterogeneity" specifications, $\gamma = \sqrt{2}$ to double the variance of the treatment effect.

We introduce manipulation into our simulations by setting a manipulation cost, c_{man}, and a manipulation opportunity probability, p_{man}. The manipulation cost determines how much of the outcome variable, y, is exchanged for an increase in the forcing variable, d. As an example, if $c_{man} = 0.2$, as it is in our simulations, then an increase in the forcing variable of 0.4 comes from a decrease in the outcome variable of $0.4 \times 0.2 = 0.08$. We limit the firms that manipulate based on the treatment effect; a firm can only manipulate if the cost of manipulation (in the form of a decrease in y) is less than the treatment effect. In our simulations, in which the true treatment effect is 0.02, firms only manipulate if the value of the forcing variable is between -0.10 and 0 (because $0.02/0.2 = 0.10$). Not all firms that are close to the threshold engage in manipulation. Instead, firms are randomly chosen to manipulate, with each firm having a probability of p_{man} of manipulating. We perform Monte Carlo experiments for $p_{man} = \{0.05, 0.1, 0.2\}$.

We again use a second-order polynomial functional form:

$$y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025x_1^2 - 0.01x_2^2 + 0.015x_3^2 + 0.02D + \Phi c_{man} d + \varepsilon.$$

in which y is the dependent variable, the three x variables are additional covariates, D is a binary variable equal to one if the forcing variable d is greater than zero (after manipulation), ε is the mean zero error term, and $\Phi = 1$ if there is manipulation, in which case the pre-manipulation value of d times the manipulation cost is added. We simulate data as previously described.

In our final set of Monte Carlo experiments, we include an omitted variable (z):

$$y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025x_1^2 - 0.01x_2^2 + 0.015x_3^2 + 0.02D + 0.02z - 0.01z^2 + \varepsilon.$$

We simulate the x, z and d variables using a multivariate normal distribution. For the vector (x_1, x_2, x_3, d, z), we set the mean to $(1.40, 5.50, 0.10, 0.20, 0.30)$ and standard deviation to $(1.00, 1.50, 0.30, 0.30, 0.80)$. We use three different correlation matrices to simulate x, z and d, to vary the degree of endogeneity in the specification. Using the first matrix, z is uncorrelated with the other regressors:

$$\left\{ \begin{array}{ccccc} 1.00 & -0.05 & -0.30 & 0.15 & 0.00 \\ -0.05 & 1.00 & 0.20 & 0.35 & 0.00 \\ -0.30 & 0.20 & 1.00 & 0.50 & 0.00 \\ 0.15 & 0.35 & 0.50 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{array} \right\}.$$

With the second matrix, z has low levels of correlation with the other regressors:

$$\left\{\begin{array}{ccccc} 1.00 & -0.05 & -0.30 & 0.15 & 0.10 \\ -0.05 & 1.00 & 0.20 & 0.35 & 0.10 \\ -0.30 & 0.20 & 1.00 & 0.50 & 0.10 \\ 0.15 & 0.35 & 0.50 & 1.00 & 0.10 \\ 0.10 & 0.10 & 0.10 & 0.10 & 1.00 \end{array}\right\}.$$

With the third matrix, z has medium levels of correlation with the other regressions:

$$\left\{\begin{array}{ccccc} 1.00 & -0.05 & -0.30 & 0.15 & 0.20 \\ -0.05 & 1.00 & 0.20 & 0.35 & 0.20 \\ -0.30 & 0.20 & 1.00 & 0.50 & 0.20 \\ 0.15 & 0.35 & 0.50 & 1.00 & 0.20 \\ 0.20 & 0.20 & 0.20 & 0.20 & 1.00 \end{array}\right\}.$$

Finally, with the fourth matrix, z has high levels of correlation with the other regressors:

$$\left\{\begin{array}{ccccc} 1.00 & -0.05 & -0.30 & 0.15 & 0.40 \\ -0.05 & 1.00 & 0.20 & 0.35 & 0.40 \\ -0.30 & 0.20 & 1.00 & 0.50 & 0.40 \\ 0.15 & 0.35 & 0.50 & 1.00 & 0.40 \\ 0.40 & 0.40 & 0.40 & 0.40 & 1.00 \end{array}\right\}.$$

Table IA.3. Monte Carlo baseline results. Results from a Monte Carlo experiment comparing RDD, OLS, and causal forest estimators of the treatment effects of a treatment status randomly assigned by a forcing threshold. The functional form is a second order polynomial: $y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025x_1^2 - 0.01x_2^2 + 0.015x_3^2 + 0.02D + \epsilon$. The table reports bias, precision, and coverage for each estimator. Bias is the mean difference between the estimate and true value of 0.02, reported as a percentage of the true value. Precision is the root mean squared error (RMSE), measured as the standard deviation of the difference between the estimate and the true value and reported as a percentage of the true value. Coverage is the fraction of trials in which a t-test of the estimate compared to the true value rejects at the 5% level. The top three panels present these statistics for an RDD using the optimal bandwidth of Imbens and Kalyanaraman (2012), half this bandwidth, and twice this bandwidth, respectively. The bottom two panels present these statistics for OLS and causal forest estimators, respectively. Each Monte Carlo has 10,000 iterations of simulated sample sizes 1,000, 5,000, and 10,000 (indicated by column header). Data simulation procedure is detailed in Section D.

	Sample size		
Estimator	1,000	5,000	10,000
RDD (Optimal bandwidth)			
Bias	0.09	0.51	−0.10
RMSE	73.94	32.72	23.41
P(t)	5.86	4.96	5.48
RDD (Half optimal bandwidth)			
Bias	−0.12	0.39	−0.19
RMSE	98.91	44.21	31.06
P(t)	5.27	5.12	4.67
RDD (Twice optimal bandwidth)			
Bias	1.36	1.75	1.41
RMSE	55.78	24.61	17.74
P(t)	5.60	4.81	5.36
OLS			
Bias	12.92	13.53	13.28
RMSE	32.88	14.78	10.55
P(t)	6.51	14.86	24.25
Causal forest (ATO)			
Bias	5.65	0.50	0.22
RMSE	27.80	12.32	8.80
P(t)	5.35	5.13	5.12

Table IA.4. Monte Carlo heterogeneity results. Results from a Monte Carlo experiment comparing RDD, OLS, and causal forest estimators of the treatment effects of a treatment status randomly assigned by a forcing threshold. The functional form is the same as is described in Table IA.3, but treatment effect is calculated as: $\tau_i = 0.02 + [0.01(x_{i1} - \bar{x_1}) - 0.01(x_{i3} - \bar{x_3}) - 0.01(D_i - \bar{D})]\gamma$. The multiplier, γ, is 1 for the panel to the left and $\sqrt{2}$ for the panel to the right. The table reports bias, precision, and coverage for each estimator. Bias is the mean difference between the estimate and true value of 0.02, reported as a percentage of the true value. Precision is the root mean squared error (RMSE), measured as the standard deviation of the difference between the estimate and the true value and reported as a percentage of the true value. Coverage is the fraction of trials in which a t-test of the estimate compared to the true value rejects at the 5% level. The top three panels present these statistics for an RDD using the optimal bandwidth of Imbens and Kalyanaraman (2012), half this bandwidth, and twice this bandwidth, respectively. The bottom two panels present these statistics for OLS and causal forest estimators, respectively. Each Monte Carlo has 10,000 iterations of simulated sample sizes 1,000, 5,000, and 10,000 (indicated by column header). Data simulation procedure is detailed in Section D.

	Heterogeneity			2x Heterogeneity		
	Sample size			Sample size		
Estimator	1,000	5,000	10,000	1,000	5,000	10,000
Panel A: Treatment effect results						
RDD (Optimal bandwidth)						
Bias	9.67	10.83	10.55	13.83	15.01	14.70
RMSE	76.38	35.94	26.37	77.27	37.50	28.36
P(t)	5.89	7.28	8.15	6.21	8.60	10.71
RDD (Half optimal bandwidth)						
Bias	8.88	10.48	10.11	13.00	14.66	14.27
RMSE	102.38	47.24	33.77	103.18	48.45	35.39
P(t)	5.69	5.94	6.09	5.86	6.57	7.33
RDD (Twice optimal bandwidth)						
Bias	11.22	12.02	11.93	15.32	16.18	16.07
RMSE	58.89	28.48	21.65	60.05	30.56	24.24
P(t)	5.96	8.23	10.93	6.40	10.15	14.88
OLS						
Bias	12.07	12.38	12.45	11.65	11.94	12.01
RMSE	36.20	19.63	16.48	36.20	19.42	16.19
P(t)	6.10	12.44	20.67	6.15	11.99	19.51
Causal forest (ATO)						
Bias	1.68	−1.78	−2.03	0.68	−2.66	−2.91
RMSE	29.11	13.11	9.30	29.23	13.32	9.59
P(t)	4.74	5.35	5.66	4.66	5.58	6.31

Table IA.5. Monte Carlo with manipulation results. Results from a Monte Carlo experiment comparing RDD, OLS, and causal forest estimators of the treatment effects of a treatment status with some manipulation of treatment status. This experiment uses the polynomial specification detailed in the caption of Table IA.3, but introduces the probability of manipulation given in the column header. Details of the manipulation specification are given in Section D. Panel A provides results of the experiment, and Panel B shows for which percentage of simulations the McCrary (2008) null of no manipulation is rejected.

	5% manipulation prob.			10% manipulation prob.			20% manipulation prob.		
	Sample size			Sample size			Sample size		
Estimator	1,000	5,000	10,000	1,000	5,000	10,000	1,000	5,000	10,000
Panel A: Treatment effect results									
RDD (Optimal bandwidth)									
Bias	−11.55	−11.27	−11.46	−21.74	−21.65	−22.62	−37.72	−39.23	−42.63
RMSE	74.20	33.07	23.67	74.38	33.96	25.22	76.33	35.86	27.97
P(t)	6.17	6.83	8.30	6.80	10.57	16.68	8.78	21.27	40.14
RDD (Half optimal bandwidth)									
Bias	−20.39	−20.85	−21.23	−36.48	−37.14	−38.90	−58.57	−61.21	−65.96
RMSE	98.51	44.18	32.06	98.00	45.52	34.08	99.27	47.40	37.08
P(t)	5.95	6.83	10.41	7.19	13.71	23.33	9.96	26.81	50.54
RDD (Twice optimal bandwidth)									
Bias	−4.90	−4.52	−4.68	−10.92	−10.78	−11.37	−21.44	−22.43	−24.72
RMSE	56.22	24.90	17.90	56.59	25.51	18.82	58.55	26.87	20.52
P(t)	5.66	5.90	5.88	5.81	7.40	10.24	6.90	14.43	25.56
OLS									
Bias	14.53	14.68	14.55	15.49	15.64	15.51	17.42	17.57	17.44
RMSE	32.90	14.65	10.51	32.90	14.65	10.51	32.91	14.65	10.51
P(t)	6.74	16.07	27.98	7.01	17.76	31.38	7.62	21.10	38.13
Causal forest (ATO)									
Bias	7.70	1.68	1.24	8.66	2.61	2.22	10.60	4.61	4.19
RMSE	28.17	12.32	8.73	28.18	12.30	8.73	28.19	12.32	8.73
P(t)	5.41	5.15	5.41	5.61	5.56	5.91	6.15	6.59	7.85
Panel B: Percent rejections at 5% level using McCrary (2008) test									
	18.45	58.14	84.16	43.62	97.69	99.95	90.42	100.00	100.00

Table IA.6. Monte Carlo with latent variable results. Results from a Monte Carlo experiment comparing RDD, OLS, and causal forest estimators of the treatment effects of a treatment status randomly assigned by a forcing threshold. The functional form adds an endogenous regressor, z, to the linear specification detailed in the caption of Table IA.3: $y = 0.05x_1 - 0.005x_2 + 0.01x_3 + 0.025x_1^2 - 0.01x_2^2 + 0.015x_3^2 + 0.02D + 0.02z - 0.01z^2 + \epsilon$ The level of correlation between z and the other regressors is given in the column header and detailed further in the text.

	No correlation			Low correlation			High correlation		
	Sample size			Sample size			Sample size		
Estimator	1,000	5,000	10,000	1,000	5,000	10,000	1,000	5,000	10,000
RDD (Optimal bandwidth)									
Bias	0.05	0.17	0.61	0.56	−0.12	0.43	0.49	0.26	0.74
RMSE	74.61	33.65	23.98	74.39	33.79	23.85	74.55	33.14	23.68
P(t)	5.60	5.19	5.45	5.58	5.42	5.48	5.55	5.50	5.43
RDD (Half optimal bandwidth)									
Bias	−0.21	−0.14	0.20	0.19	−0.46	−0.09	−0.29	0.22	0.27
RMSE	100.91	45.53	31.88	100.69	45.72	31.68	100.30	44.64	31.61
P(t)	5.12	4.86	4.75	5.23	5.20	4.96	5.16	4.77	4.59
RDD (Twice optimal bandwidth)									
Bias	1.44	1.62	1.98	1.78	1.56	1.94	2.62	2.65	2.81
RMSE	56.73	25.48	18.15	56.61	25.61	18.05	56.93	25.19	18.23
P(t)	5.34	5.26	5.40	5.23	5.30	5.49	5.63	5.28	5.69
OLS									
Bias	13.06	13.39	13.52	16.10	16.31	16.43	33.48	33.42	33.41
RMSE	36.42	20.24	17.26	37.65	22.33	19.62	47.43	36.64	35.07
P(t)	6.41	13.99	23.86	7.35	18.34	32.79	16.54	59.59	87.64
Causal forest (ATO)									
Bias	5.90	0.96	0.43	9.53	3.86	3.19	26.15	16.99	16.46
RMSE	29.30	12.84	9.01	30.36	13.40	9.60	38.83	21.25	18.75
P(t)	4.98	5.11	5.33	5.84	5.97	6.46	14.36	26.46	44.83

E Auxiliary tables and figures

Notes on tables and figures:

- Figure IA.3 plots HTEs against covariates. We estimate HTEs in the causal forest estimation described in the caption of Table 3.
- Figure IA.4 plot predicted default propensities from a classification forest against slack ratio, or our forcing variable expressed as the distance to default scaled by the default threshold. Each predicted default propensity is plotted in a light blue x. The dark blue horizontal lines present average predicted default propensities at each level of slack ratio, and the height of the blue boxes around the dark blue horizontal lines are 95% confidence intervals. The forest estimation includes all covariates included in our main causal forest analysis of investment except slack (distance to default) because slack perfectly predicts default status.
- Table IA.7 shows averages of HTEs across samples constructed on observables.

Figure IA.3. HTE plots. Heterogeneous treatment effects plotted against covariates. Details of the causal forest estimation are in the caption of Table 3.

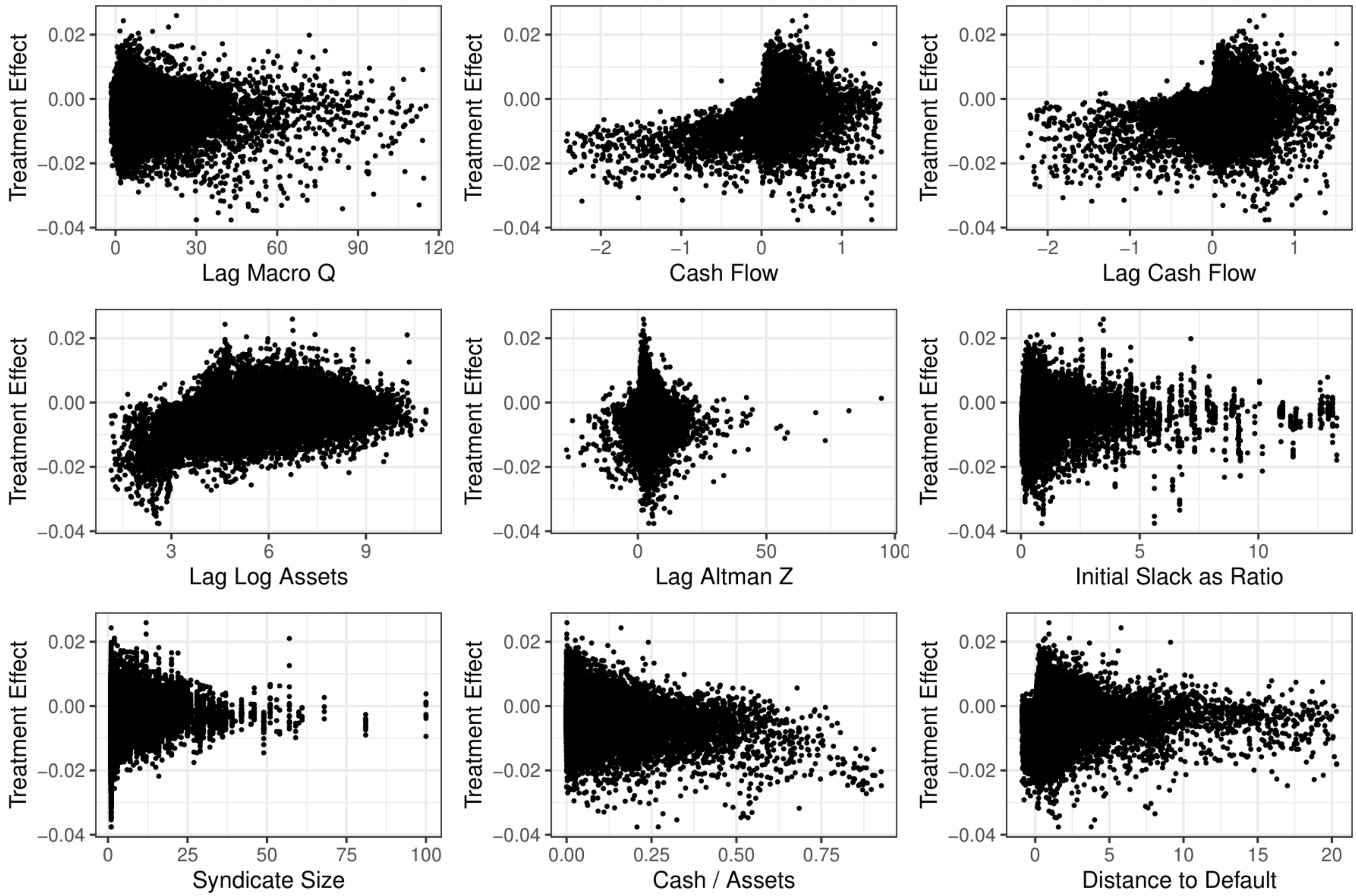


Figure IA.4. Predicted default propensities. Predicted defaults estimated using a classification forest estimation described in the caption of Table 7. We plot estimated default propensities, each with a light blue x, against slack ratio, or the distance to default scaled by the default threshold. Averages 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).

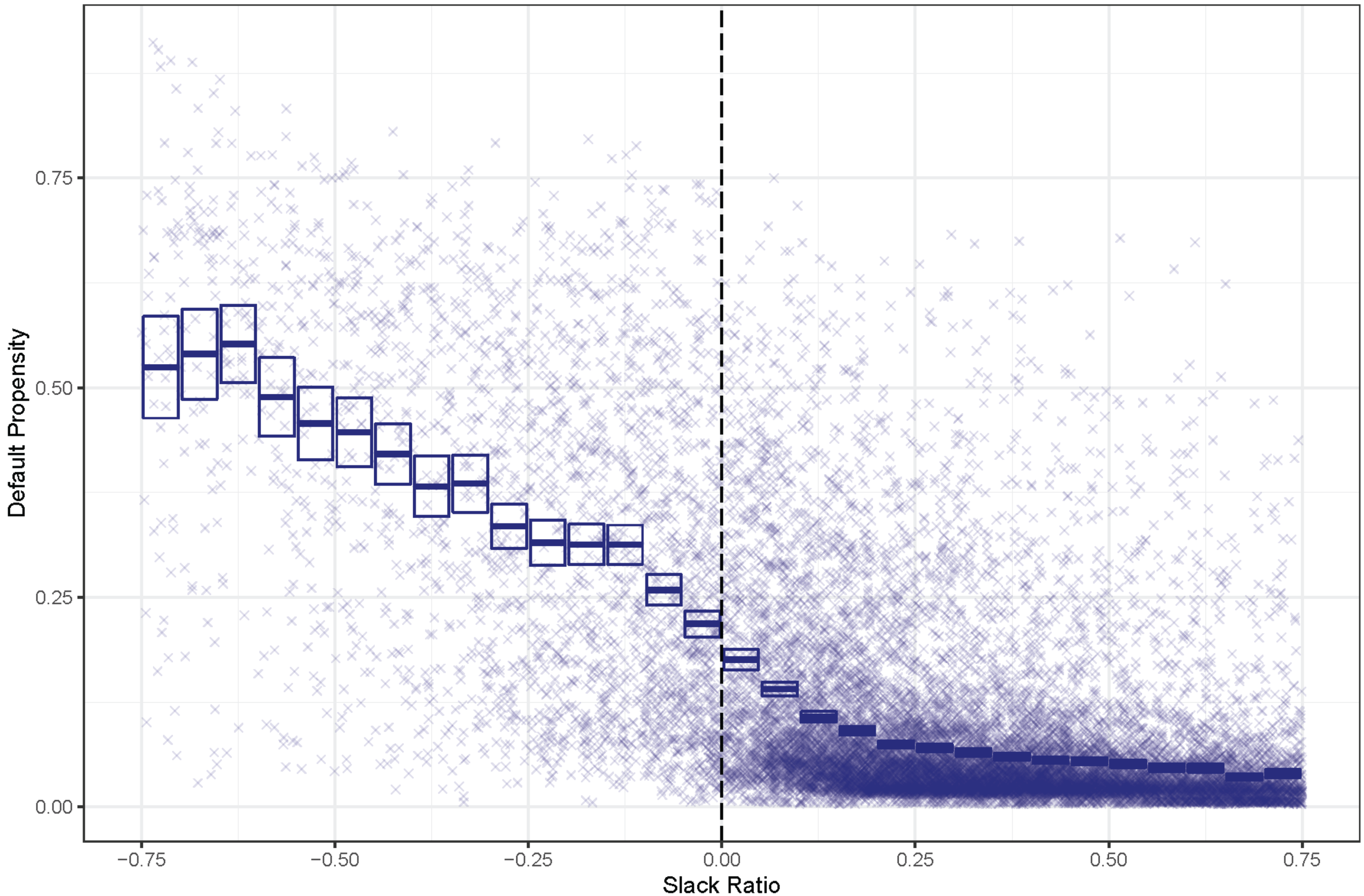


Table IA.7. Causal forest results across samples defined with observables. Re-estimations of average treatment effect (ATE), average treatment effect on the treated (ATT), average treatment effect on the control (ATC), and average treatment effect with an overlap correction of Li et al. (2018) (ATO) of default on firm investment. Samples are variations on the net worth extended sample (quarterly observations, 1994 – 2017) defined in the italicized headers. We follow the causal forest estimation described in the caption of Table 3. For more information on sampling procedures, see the caption of Table IA.9.

	estimate	standard error	t-statistic	p-value	N
Panel A: Keep all firms and all loans					
ATE	-0.31^{**}	0.14	-2.193	0.028	35,502
ATT	-0.50^{***}	0.09	-5.281	<0.001	35,502
ATC	-0.25	0.18	-1.365	0.172	35,502
ATO	-0.50^{***}	0.09	-5.418	<0.001	35,502
Panel B: Keep all firms; drop loans with negative initial slack					
ATE	-0.24	0.18	-1.323	0.186	32,530
ATT	-0.55^{***}	0.09	-5.845	<0.001	32,530
ATC	-0.19	0.21	-0.877	0.381	32,530
ATO	-0.54^{***}	0.10	-5.475	<0.001	32,530
Panel C: Drop firms that never default; keep all loans					
ATE	-0.43^{***}	0.10	-4.520	<0.001	15,120
ATT	-0.36^{***}	0.10	-3.398	0.001	15,120
ATC	-0.43^{***}	0.12	-3.528	<0.001	15,120
ATO	-0.42^{***}	0.10	-4.070	<0.001	15,120
Panel D: Drop firms that never default and loans with negative initial slack					
ATE	-0.43^{***}	0.10	-4.182	<0.001	11,256
ATT	-0.44^{***}	0.11	-4.064	<0.001	11,256
ATC	-0.41^{***}	0.12	-3.477	0.001	11,256
ATO	-0.45^{***}	0.11	-4.043	<0.001	11,256

F Current Ratio covenant sample

We analyze samples of firms with current ratio and net worth covenants separately for two reasons. First, there are empirical differences between the two samples (Table 2). These differences are consistent with firms not receiving covenant types randomly (Demiroglu and James, 2010). So, in limiting our sample to firms with the same kind of covenant we are removing one source of selection bias from our setting.

Second, there is a compelling theoretical reason to separate the current ratio and net worth samples. A secondary source of endogeneity in our setting is firms' ability to avoid default. Because current ratio and net worth are defined using different accounting variables, firms' strategies to avoid default by altering these variables will differ. For example, in our setting firms can decrease investment to increase their current ratio and avoid default. Current ratio is calculated as current assets divided by current liabilities. A decrease in investment that would have been financed, even in part, with current assets, will result in a higher current ratio. Firms avoiding default by altering investment is of particular concern in our setting because it introduces a clear potential source of reverse causality. In separating these two samples, we can observe and address these sources of endogeneity more clearly and directly.

Additionally, as we discuss in the main text of the paper, there is considerable evidence of manipulation in our current ratio sample. Likely, this manipulation threatens overlap. Thus, separating the samples allows us to examine only the net worth sample, in which the assumption of overlap is more plausible.

In Table IA.10, we report causal forest results for the current ratio sample that are consistent with a lack of overlap. One concerning way in which firms can avoid default is to cut investment. Graham et al. (2005) survey managers who report a willingness to cut investment to avoid violating covenants. Cutting investment that would be paid for, even partly, with current assets (i.e., cash), would boost current ratio, which is calculated as current assets scaled by current liabilities. If this manipulation is present, we would expect *higher* investment in firms above the default threshold.

Panel B of Table IA.10 shows positive treatment effect estimates in the current ratio sample. However, the ATE and ATC are positive and significant, whereas the ATT is insignificant. This is the reverse of the significance we see in the net worth sample (Panel A). A lack of significance in the ATT sample means that the ATE is driven by the control observations, rather than the treated observations. An estimated ATC of 0.54% means that, were these firms to default, we expect an *increase* in investment. This positive effect reflects the *higher* investment we expect in firms that do not avoid default by cutting investment.

The results in Panel B of Table IA.10 point to a lack of overlap, or insufficiently similar treatment and control firms. An important piece of evidence here is that the ATO is small and insignificant. Thus, firms with closer to 50% propensities to default, with which the ATO is calculated, are dissimilar from the control firms in the ATC calculation. Thus, the ATE is driven by firms that are not in default and do not have much of a likelihood to default. Because these firms are not likely to default, they are not appropriate control observations for many of the treated observations within the same. Treatment effects here are likely driven by firms' avoidance of default, so this setup does not capture the effects of default on investment.

Table IA.10 also includes causal forest estimates for a combined sample of net worth and current ratio observations that highlight how different are the current ratio and net worth samples. The ATE for this sample is 0.01%, which makes sense; we expect the negative and positive effect of default on investment we estimate in the net worth and current ratio samples, respectively, separately should net out close to zero. Additionally, when we split results of our combined analysis into subsamples of firms with current ratio or net worth covenants, we find similar average treatment effects as when we analyze these samples separately. For example, Table 3 reports an ATE, ATT, ATC, and ATO for our net worth sample of: −0.24%, −0.55%, −0.19%, and −0.54%. These estimates for the net worth subsample of our combined analysis are: −0.28%, −0.56%, −0.23%, and −0.54%. These similarities suggest that the causal forest can distinguish between net worth and current ratio observations, so these observations differ in measurable ways. These results provide more evidence that our decision to separate the current ratio and net worth samples for analysis is appropriate.

Notes on tables and figures:

- Table IA.8 provides summary statistics for both samples. The net worth sample summary statistics are also presented in the main paper.
- Table IA.9 provides sample sizes and McCrary (2008) test results for current ratio, net worth, and combined samples.
- Table IA.10 shows averages of HTEs for forests estimated on current ratio, net worth, and combined samples.
- Table IA.11 shows polynomial RDD results for both the net worth and current ratio samples. This table adds the current ratio results to the net worth results, which are in the main text of the paper, side-by-side for comparison.
- Table IA.12 shows kernel RDD results for both net worth and current ratio samples. Again, net worth results are also in the paper, but we present both here to facilitate

side-by-side comparison.

- Table IA.13 shows polynomial RDD estimations in a combined net worth and current ratio sample.
- Table IA.14 shows kernel RDD estimations in a combined net worth and current ratio sample.

Table IA.8. Summary statistics. Averages, [medians], and (standard errors) of firm characteristics for firm-quarter observations for deals and loans from 1994 to 2017. The current ratio sample is all firm-quarter observations for firms with a covenant restricting the minimum current ratio listed in Dealscan between 1994 and 2017. The net worth sample is 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.

	Current Ratio		Net Worth	
Variable	Mean [Median]	(SE)	Mean [Median]	(SE)
Current Ratio	2.22	(0.01)	2.11	(0.01)
	[1.90]		[1.79]	
Net Worth	517.55	(34.81)	789.19	(14.86)
	[97.14]		[186.31]	
Tangible Net Worth	517.82	(34.85)	773.19	(14.01)
	[96.95]		[183.27]	
Log(Assets)	5.44	(0.01)	6.06	(0.01)
	[5.43]		[6.09]	
Market-to-Book	1.32	(0.01)	1.29	(0.01)
	[1.00]		[0.97]	
Macro q	6.98	(0.10)	7.17	(0.06)
	[2.77]		[3.13]	
ROA	0.03	(0.00)	0.03	(0.00)
	[0.03]		[0.03]	
Capital/Assets	0.37	(0.00)	0.33	(0.00)
	[0.29]		[0.26]	
Investment/Capital	0.07	(0.00)	0.06	(0.00)
	[0.05]		[0.04]	
Cash Flow	0.08	(0.00)	0.08	(0.00)
	[0.07]		[0.07]	
Leverage	0.25	(0.00)	0.24	(0.00)
	[0.22]		[0.22]	
Firm-Quarter Obs	15,347		46,306	
Firms	992		2,628	

Table IA.9. Sample sizes and McCrary test results across samples. This table shows how sampling decisions affect sample sizes and McCrary (2008) p-values for a null hypothesis of no manipulation of treatment status. Non-defaulters are firms that do not default in at least one firm-quarter during the sample period. We either keep all loans with negative initial slack, drop these loans from our sample entirely, or drop only the firm-quarters in which a default occurs because a loan originated with in default (with negative initial slack). All samples require non-missing investment, Macro Q, and cash flow.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Current ratio sample						
Non-defaulters:						
keep	✓	✓	✓	–	–	–
drop	–	–	–	✓	✓	✓
Loans with negative initial slack:						
keep all	✓	–	–	✓	–	–
drop loans	–	✓	–	–	✓	–
drop firm-quarters	–	–	✓	–	–	✓
obs.	13,826	11,518	13,067	8,500	6,086	7,717
McCrary p-value	<0.001	<0.001	<0.001	0.118	<0.001	0.024
	(7)	(8)	(9)	(10)	(11)	(12)
Panel B: Net worth sample						
Defaulters:						
keep	✓	✓	✓	–	–	–
drop	–	–	–	✓	✓	✓
Loans with negative initial slack:						
keep all	✓	–	–	✓	–	–
drop loans	–	✓	–	–	✓	–
drop firm-quarters	–	–	✓	–	–	✓
obs.	38,046	34,589	36,889	16,557	12,444	15,602
McCrary p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	(13)	(14)	(15)	(16)	(17)	(18)
Panel C: Combined current ratio and net worth sample						
Defaulters:						
keep	✓	✓	✓	–	–	–
drop	–	–	–	✓	✓	✓
Loans with negative initial slack:						
keep all	✓	–	–	✓	–	–
drop loans	–	✓	–	–	✓	–
drop firm-quarters	–	–	✓	–	–	✓
obs.	43,822	39,409	42,157	23,184	17,576	21,533
McCrary p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table IA.10. **Causal forest results across covenant samples.** 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 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
Panel A: Net worth covenant sample					
ATE	-0.24	0.18	-1.323	0.186	32,530
ATT	-0.55^{***}	0.09	-5.845	<0.001	32,530
ATC	-0.19	0.21	-0.877	0.381	32,530
ATO	-0.54^{***}	0.10	-5.475	<0.001	32,530
Panel B: Current ratio covenant sample					
ATE	0.46^{**}	0.22	2.111	0.035	10,891
ATT	0.20	0.17	1.185	0.236	10,891
ATC	0.54^{**}	0.25	2.177	0.030	10,891
ATO	0.19	0.18	1.022	0.307	10,891
Panel C: Net worth and current ratio covenant sample					
Both covenants (whole sample) ATE	0.01	0.12	0.049	0.961	36,856
Both covenants (whole sample) ATT	-0.22^{**}	0.09	-2.586	0.010	36,856
Both covenants (whole sample) ATC	0.06	0.14	0.467	0.640	36,856
Both covenants (whole sample) ATO	-0.20^{**}	0.09	-2.166	0.030	36,856
Current ratio covenants subsample ATE	0.51^{**}	0.22	2.314	0.021	10,569
Current ratio covenants subsample ATT	0.35^{**}	0.16	2.225	0.026	10,569
Current ratio covenants subsample ATC	0.58^{**}	0.25	2.278	0.023	10,569
Current ratio covenants subsample ATO	0.36^{**}	0.17	2.131	0.033	10,569
Net worth covenants subsample ATE	-0.28^{**}	0.14	-2.037	0.042	26,287
Net worth covenants subsample ATT	-0.56^{***}	0.10	-5.738	<0.001	26,287
Net worth covenants subsample ATC	-0.23	0.16	-1.453	0.146	26,287
Net worth covenants subsample ATO	-0.54^{***}	0.11	-5.165	<0.001	26,287

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

Table IA.11. RDD regression results. Results for a regression of investment (quarterly capital expenditures divided by beginning-of-period PPE and scaled by 100 for ease in interpretation) on a binary variable (*bind*) equal to one if a firm is in default in a quarter. In the columns indicated, control variables included are: firm size, Macro Q and cash flow. Variable descriptions are given in the caption of Table 2. Samples for Panels A.1 through A.3 and B.1 through B.3 are quarterly firm observations from 1994 to 2017 with net worth and current ratio covenants, respectively. Year-quarter and firm fixed effects are included in each specification. Where indicated by the column header, all powers of slack are included up to the highest polynomial (i.e., if the highest polynomial indicated in the column header is "third", slack-squared and slack are also included). If interactions between bind and slack are included, all slack variables are interacted with bind. Bandwidth sizes for regressions in Panels A.3 and B.3 are Calonico et al. (2020) optimal bandwidths. Standard errors are in parentheses.

			Highest polynomial of slack included			
			first	second	third	fourth
Panel A.1: Net worth sample baseline regressions						
bind	-1.22^{***}	-0.94^{***}	-0.82^{***}	-0.70^{***}	-0.55^{***}	-0.60^{***}
	(0.09)	(0.10)	(0.10)	(0.11)	(0.11)	(0.12)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	42,439	36,246	36,246	36,246	36,246	36,246
R^2	0.42	0.46	0.46	0.46	0.46	0.46
Panel A.2: Net worth sample with *bind* and slack interactions						
bind	–	–	-0.47^{***}	-0.24	-0.06	-0.27
			(0.14)	(0.18)	(0.22)	(0.26)
controls	–	–	yes	yes	yes	yes
slack	–	–	yes	yes	yes	yes
interactions	–	-	yes	yes	yes	yes
Obs.	–	–	36,246	36,246	36,246	36,246
R^2	–	–	0.46	0.46	0.46	0.46
Panel A.3: Net worth threshold sample						
bind	-0.44^{***}	-0.40^{***}	0.20	0.27	0.42	0.35
	(0.13)	(0.14)	(0.21)	(0.22)	(0.27)	(0.29)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	9,841	8,394	8,394	8,394	8,394	8,394
R^2	0.53	0.57	0.57	0.57	0.57	0.57

			Highest polynomial of slack included			
			first	second	third	fourth
Panel B.1: Current ratio sample baseline regressions						
bind	-0.67^{***}	-0.44^{**}	-0.48^{**}	-0.53^{**}	-0.56^{**}	-0.44
	(0.17)	(0.19)	(0.20)	(0.23)	(0.27)	(0.29)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	14,119	11,969	11,969	11,969	11,969	11,969
R^2	0.39	0.43	0.43	0.43	0.43	0.43
Panel B.2: Current ratio sample with *bind* and slack interactions						
bind	–	–	-0.19	-0.02	-0.32	-0.47
	–	–	(0.27)	(0.36)	(0.46)	(0.55)
controls	–	–	yes	yes	yes	yes
slack	–	–	yes	yes	yes	yes
interactions	–	-	yes	yes	yes	yes
Obs.	–	–	11,969	11,969	11,969	11,969
R^2	–	–	0.43	0.43	0.43	0.43
Panel B.3: Current ratio threshold sample						
bind	-0.42	-0.13	-0.05	0.04	0.17	0.17
	(0.27)	(0.29)	(0.49)	(0.50)	(0.64)	(0.65)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	3,078	2,535	2,535	2,535	2,535	2,535
R^2	0.50	0.56	0.56	0.56	0.56	0.56

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

Table IA.12. RDD results estimated with triangular kernels. 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) for the current ratio (Panel A) and net worth (Panel B) samples described in the caption of Table 2. 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: Current ratio sample					
bind	-0.74^{**}	-0.42	-0.23	-0.25	0.22
	(0.35)	(0.60)	(0.76)	(0.95)	(1.21)
Bandwidth	0.54	0.18	0.25	0.28	0.25
Obs. below	2,171	1,027	1,350	1,454	1,373
Obs. above	5,848	2,051	2,859	3,185	2,928
Panel B: Net worth sample					
bind	0.10	0.24	0.31	0.21	0.13
	(0.23)	(0.26)	(0.34)	(0.42)	(0.47)
Bandwidth	0.34	0.25	0.32	0.36	0.43
Obs. below	2,769	2,288	2,657	2,846	3,168
Obs. above	10,694	7,553	9,878	11,410	13,944

Table IA.13. RDD regression results for combined covenants sample. Results for a regression of investment (quarterly capital expenditures divided by beginning-of-period PPE and scaled by 100 for ease in interpretation) on a binary variable (*bind*) equal to one if a firm is in default in a quarter. In the columns indicated, control variables included are: firm size, Macro Q and cash flow. Variable descriptions are given in the caption of Table 2. Samples for Panels A through C are quarterly firm observations from 1994 to 2017 with net worth or current ratio covenants. Year-quarter and firm fixed effects are included in each specification. Where indicated by the column header, all powers of slack are included up to the highest polynomial (i.e., if the highest polynomial indicated in the column header is "third", slack-squared and slack are also included). If interactions between bind and slack are included, all slack variables are interacted with bind. Bandwidth sizes for regressions in Panels C are Calonico et al. (2020) optimal bandwidths. Standard errors are in parentheses.

			Highest polynomial of slack included			
			first	second	third	fourth
Panel A: Baseline regressions						
bind	−1.08***	−0.82***	−0.68***	−0.45***	−0.57***	−0.52***
	(0.08)	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	48,407	41,232	41,232	41,232	41,232	41,232
R^2	0.41	0.45	0.45	0.45	0.45	0.45
Panel B: Including *bind* and slack interactions						
bind	–	–	−0.35***	−0.13	−0.18	−0.40*
			(0.12)	(0.16)	(0.19)	(0.23)
controls	–	–	yes	yes	yes	yes
slack	–	–	yes	yes	yes	yes
interactions	–	-	yes	yes	yes	yes
Obs.	–	–	41,232	41,232	41,232	41,232
R^2	–	–	0.45	0.45	0.45	0.45
Panel C: Threshold sample						
bind	−0.50***	−0.44***	−0.23	−0.09	0.03	−0.05
	(0.11)	(0.12)	(0.19)	(0.19)	(0.24)	(0.25)
controls	no	yes	yes	yes	yes	yes
slack	no	no	yes	yes	yes	yes
interactions	no	no	no	no	no	no
Obs.	13,632	11,548	11,548	11,548	11,548	11,548
R^2	0.50	0.54	0.54	0.54	0.54	0.54

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

Table IA.14. **RDD results estimated with triangular kernels for combined covenants sample.** 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) for quarterly firm observations from 1994 to 2017 with net worth or current ratio covenants. The "IK Bandwidth" result is calculated using a local linear specification in the Imbens and Kalyanaraman (2012) optimal bandwidth (from the RDD 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 ℝ package).

	Model				
	IK Bandwidth	Linear	Quadratic	Cubic	Quartic
bind	−0.26	−0.18	−0.15	0.01	0.06
	(0.18)	(0.24)	(0.29)	(0.37)	(0.44)
Bandwidth	0.51	0.28	0.41	0.43	0.47
Obs. below	5,246	3,698	4,694	4,839	5,063
Obs. above	18,422	9,934	15,134	15,982	17,377