JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 59, No. 6, Sep. 2024, pp. 2582–2617

doi:10.1017/S0022109023000650

Uncovering Financial Constraints

Matthew Linn
University of Massachusetts Amherst Isenberg School of Management
mlinn@isenberg.umass.edu (corresponding author)

Daniel Weagley
Georgia Institute of Technology Scheller College of Business
daniel.weagley@scheller.gatech.edu

Abstract

We use a random forest model to classify firms' financial constraints using only financial variables. Our methodology expands the range of classified firms compared to text-based measures while maintaining similar levels of informativeness. We construct two versions of our constraint measures, one using many firm characteristics and the other using a small set of more primitive characteristics. Using our measures, we find that institutional investors hold a lower percentage of shares in equity-focused constrained firms, while retail investors show a preference for them. Equity issuance and investment of constrained firms also increases during periods of high investor sentiment.

I. Introduction

Understanding the effects of firms' external financing constraints is one of the most studied topics in economics. Proper empirical analysis in this area requires an accurate estimation of firm-level constraints. We introduce a new methodology that uses random decision forests, a machine learning algorithm, to estimate a multi-dimensional mapping between firm-level accounting variables and financial constraints. We estimate constraints both with a large set of accounting variables as predictors and, alternatively, with a small set of primitive, less endogenous variables, producing two different versions of our measures. Our methodology provides informative constraint estimates and allows us to estimate constraints for

We thank Jarrad Harford (the editor) and Gerard Hoberg (the referee) for constructive comments in the review process. We are grateful to Taylor Begley, Donald Bowen, András Danis, Joan Farra-Mensa, Anisha Ghosh (discussant), Sudip Gupta (discussant), Charles Hadlock (discussant), Armen Hovakimian (discussant), Alex Hsu, Russell Jame, Nikunj Kapadia, Soohun Kim, Bradley Paye, Manpreet Singh, Jason Sockin (discussant), Huan Yang, and seminar and conference participants at Georgia Tech, University of Massachusetts, Virginia Tech, the 2018 Northern Finance Association Conference, the 2019 Dolomites Winter Finance Conference, the 2019 Financial Management Association Conference, the 2020 Midwest Finance Association Conference, and the 2022 University of Florida Research Conference on Machine Learning in Finance for helpful comments on the article. Any remaining errors are our own. Some of the ideas in this article were spun out of a previous working article entitled "Seeing the Forest Through The Trees: Do Investors Underreact to Systemic Events?" This article was briefly circulated under the title "Estimating Financial Constraints with Machine Learning."

https://doi.org/10.1017/S0022109023000650 Published online by Cambridge University Press

most publicly traded companies back to 1972.[1] We apply our measures to provide new insights into how equity constraints are negatively (positively) related to institutional (retail) investor ownership and, additionally, how the timing of financing and investment of constrained equity-focused firms is positively related to investor sentiment.

Producing estimates of firms' financial constraints that are both informative and cover most publicly traded firms has proven to be an empirical challenge. A common method for estimating constraints is to use indices constructed from accounting variables. However, Bodnaruk, Loughran, and McDonald (2015) and Farre-Mensa and Ljungqvist (2016) call into question the reliability of popular firm-level constraint indices. Recent constraint classification methods based on algorithmic textual analysis of firms' 10-K filings have been shown to be informative, yet they lack coverage both in the cross section and the time series. The lack of coverage is a result of analyzing specific sections of firms' 10-K filings, which are missing or cannot be parsed for a large number of firms.

To extend the coverage of the text-based constraint measures without losing informativeness, we train a random forest on the textual analysis-based financial constraint measures of Hoberg and Maksimovic (HM) (2015).[2] This allows us to expand coverage of equity and debt-focused constraint measures both in the cross section and time series, classifying over 165,000 more firm-years than the HM measures.[3] During the years 1997–2015 (the time period covered by HM), we are able to increase the number of classified firms by an average of 43% per year. Our classifications extend to time periods pre-1997 and post-2015, greatly expanding the potential set of analyses that can be conducted.

Our methodology can be thought of as a modern, more algorithmic, and larger-scale version of the Kaplan and Zingales (1997) and Hadlock and Pierce (2010) indices: A "training" sample of firms is classified as more or less constrained based on the text in their 10-Ks, a statistical model then creates a mapping between accounting variables and constraints.[4] Our methodology improves on previous indices by using a much broader training set and applying modern statistical learning which captures important nonlinearities and interactions that simple linear models do not.

[1]The data corresponding to the constraint measures described in this article are available at: http://www.danielweagley.com.

[2]The use of large-scale textual analysis of firms' 10-K filings to assess financial constraints represented a major methodological breakthrough (see Bodnaruk et al. (2015), Hoberg and Maksimovic (HM) (2015), and Buehlmaier and Whited (2018)). These methods allow for a more direct identification of firms' financial constraints by parsing through the firms' correspondence with investors. While the text-based measures of Bodnaruk et al. (2015) and Buehlmaier and Whited (2018) are similar, we use HM (2015) as our training measure because the authors have made their data publicly available allowing for easy replication of our results.

[3]While HM's method classifies firms as being constrained and equity- or debt-focused, they show that firms with equity (debt) focused constraints behave as if they face more binding equity (debt) constraints. We therefore use their equity (debt) focused constraints as a measure of equity (debt) constraints and provide evidence that these firms indeed face constraints in the relevant source of financing.

[4]The use of the regression coefficients for the construction of an index and for out-of-sample analysis was not necessarily the goal of the authors of these papers, but is widely practiced in the subsequent literature.

We estimate two versions of the random forest model that differ in the set of predictors used. The first version uses the union of accounting variables used in creating three leading constraint indices from the literature: those of Kaplan and Zingales (1997), Whited and Wu (2006), and Hadlock and Pierce (2010). The second version uses a small subset of these variables that are the most primitive to the firm and are less likely to be endogenously determined by firm managers. While the full model has better out-of-sample predictive power, we find both models produce informative measures of constraints. Tests using each of our measures tend to exhibit greater statistical precision than the HM measures and we find that the firms we classify as constrained behave similarly in and out of sample.

We use HM constraints as our training sample for several reasons. First, the algorithmic-nature of HM's classifications ensures consistency and objectivity. Second, the HM measures have been shown to be informative, which we confirm in our analysis. Third, the set of HM classifications is much larger than could be achieved by manually reading 10-K text as in Kaplan and Zingales (1997) and Hadlock and Pierce (2010). Fourth, the use of HM measures addresses concerns that previous methods extrapolate from a small and, in some cases, nonrepresentative training set. Finally, the HM measure is constructed to maximize its ability to identify constraints in the cross section. This involves removing common "boilerplate" information each period to sharpen focus on informative cross-sectional content. It therefore does not reflect time variation in aggregate constraints. Our measures inherit these features, making them very informative as cross-sectional measures as opposed to capturing time-variation in average constraint levels.

We conduct tests to confirm the performance of the random forest model both in and out of sample. In addition, we show that the random forest is superior to an analogously constructed linear (OLS) model in terms of out-of-sample fit. The flexibility of random forests allows it to capture nonlinearities and interactions that simple linear models cannot.

We perform a number of validation tests from the literature to address the Bodnaruk et al. (2015) and Farre-Mensa and Ljungqvist (2016) critique, in which they provide evidence suggesting existing accounting-based constraint indices do not correctly identify financially constrained firms. The tests we choose are those used by Bodnaruk et al. (2015). These tests include one originally introduced by Farre-Mensa and Ljungqvist (2016). The remaining tests were first proposed by Bodnaruk et al. (2015). In all of the tests, where previous accounting variable-based indices have been shown to fall short, our estimators (both the full-model and "Primitive"-model) perform well, suggesting their efficacy as constraint measures.

After establishing the effectiveness of our constraint estimators, we examine how equity constraints are related to firms' raising and investing capital. First, we examine how equity-related constraints are related to investor type. Institutional ownership has been linked to the reduction in asymmetric information through better governance (Carleton, Nelson, and Weisbach (1998), Admati and Pfleiderer (2009), Appel, Gormley, and Keim (2016), and McCahery, Sautner, and Starks (2016)) and more efficient pricing (Boehmer and Kelley (2009)). It follows that institutional ownership should be negatively associated with equity-related constraints holding all else equal. We find a strong monotonically, negative relationship between equity-related constraints and institutional ownership, even after controlling for firm size. The disparity in ownership between the most constrained and least constrained stocks has consistently been between 10 and 20 percentage points over the past 40 years.

We find a starkly different pattern for retail investors. Using brokerage firm data from 1991:Q2 to 1996:Q3 and Robinhood investor positions from May 2018 to Aug. 2020, we show that retail investors are relatively more likely to hold the equity of the most equity-constrained firms. Although consistent with some of the behavior observed during the meme stock trading era, the patterns we document are true even in the early 1990s and 2018–2019, suggesting retail investors – at least these subsets – are more likely to target firms that are facing equity constraints, potentially due to lack of interest from institutional investors.

Second, we test whether equity-constrained firms' financing and investment decisions are more sensitive to shifts in investor sentiment than less constrained firms.[5] We hypothesize that because they face constraints, equity-constrained firms are more likely to take advantage of the lower equity costs leading to greater equity issuance and investment compared to unconstrained firms. As discussed by Baker and Wurgler (2006), the equity prices of firms that are difficult to value are likely to be most sensitive to investor sentiment. Equity-constrained firms, who are characterized by higher levels of information asymmetry, are more difficult to value and should, therefore, be more sensitive to investor sentiment. Our empirical results are consistent with the hypothesis: equity-constrained firms' real outcomes are more positively correlated with proxies for market sentiment.

In Section II, we discuss our methodology and how it relates to existing constraint indices. In Section III, we discuss our methodology's predictive performance and the estimated relationships between firm characteristics and constraints. In Section IV, we assess whether firms classified as constrained by the random forest model behave as though they are constrained. In Section V, we examine the relationship between constraints and institutional and retail investor stock ownership and examine how firms' constraints and their financing and investment are related to market sentiment. Section VI concludes.

II. Estimating Financial Constraints

A. Existing Financial Constraint Measures

There are broadly three approaches to estimating financial constraints: i) use proxies such as cash holdings, dividend payouts, and firm size (among many others) to infer whether firms are financially constrained, ii) use a financial constraint index of accounting variables motivated by a model or patterns uncovered from reading firms' 10-Ks (mainly the "Kaplan–Zingales" Index, "Whited–Wu" Index, and the "Hadlock–Pierce" Index), or iii) classify firms based on large-scale textual analysis of the firms' 10-Ks. There is significant debate about the ability of these measures, especially the first two categories, to capture financial constraints. Recent work by Bodnaruk et al. (2015) and Farre-Mensa and Ljungqvist (2016) provides a number

[5]We focus on equity-focused constrained firms because constraints tend to affect these firms more than those that rely more on debt financing (HM (2015)).

TABLE 1

Rank Correlations of Constraint Indices

Table 1 reports rank correlations for constraint indices. HM_EQUITY and HM_DEBT denote equity and debt-focused constraints from Hoberg and Maksimovic (HM) (2015). HP denotes the constraint index of Hadlock and Pierce (2010), KZ denotes the Kaplan and Zingales (1997) index, and WW denotes the Whited and Wu (2006) index. For the HP, KZ, and WW indices, we include all observations (excluding financials and utilities) between 1972 and 2021 for which we are able to construct the given index using annual Compustat data. The HM_EQUITY and HM_DEBT indices include the observations provided in the data from HM (2015) which range from 1997 to 2015. Each pairwise rank correlation includes only firms-years for which we have observations in both indices. All rank correlations are reported as percentages.

	HM_EQUITY	HM_DEBT	HP	KZ	WW
HM_EQUITY	100.00				
HM_DEBT	−17.43	100.00			
HP	23.38	−20.90	100.00		
KZ	0.78	21.37	9.89	100.00	
WW	21.20	−17.01	73.09	27.31	100.00

of empirical results suggesting existing accounting-based measures do a poor job of capturing constraints. The text-based measures of Bodnaruk et al. (2015), HM (2015), and Buehlmaier and Whited (2018), on the other hand, show significant promise in identifying financially constrained firms though they lack coverage in the cross section and time-series. We discuss this lack of coverage more in the Section II.C.

The different approaches can lead to significant variation across measures in the firms that are classified as constrained. Table 1 reports within-year rank correlations between the HM debt-focused and equity-focused constraints (which we use as our training measures) and other well-known indices used throughout the financial constraints literature (measures based on results in Kaplan and Zingales (1997), Whited and Wu (2006), and Hadlock and Pierce (2010)). The relatively low correlations between the various indices are perhaps surprising. Only the Hadlock and Pierce (HP) index and the Whited and Wu (WW) index show a high correlation with each other. Furthermore, because these indices measure general constraint levels, it is interesting to see whether they coincide more with the HM debt or equity-focused constraints. The table shows that the HP and WW indices are moderately (positively) correlated with the equity-focused constraint measure (HM_EQUITY), while the KZ index is moderately correlated with the debt-focused constraint measure (HM_DEBT). There is very little correlation between HM debt and HM equity. Similarly, the indices that are positively correlated with HM equity are not strongly correlated with HM debt and vice versa. These results suggest that the information in HM equity and HM debt measures is sufficiently different from other constraint indices.

B. Our Training Measures

We choose to use the HM measures as our training sample because the data is readily available and have been shown to be informative. HM analyze the text of firms' 10-K filings to estimate annual measures of firms' financial constraints. Their algorithm identifies firms that report potential delays in investment due to being unable to raise capital. Importantly, they are able to separate delays in investment for firms focusing on different methods of raising capital. They provide four measures of financial constraints based upon firms' discussions of delaying investment due to lack of liquidity: a general measure of financial constraints, a debt-focused financing constraint measure, an equity-focused financing constraint measure, and a private placement financing constraint measure. In their analysis, HM show that firms with high debt-focused constraint measures tend to differ from the types of firms with high equity-focused constraint measures.[6] Through a number of validation tests, the authors show that firms classified as having higher values of the equity (debt) focused constraint measure behave as though they face relatively tighter equity (debt) constraints. This serves as external validation that their algorithm deciphers tightness of equity and debt constraints separately.

C. Coverage Limitations of Text-Based Measures

Text-based measures of financial constraints, such as the HM measures, offer the advantage of directly capturing firms' reported inability to raise capital. However, there are some limitations to using text-based measures of constraints. First, the HM text-based constraint measure data goes back only as far as 1997, which makes their data unsuitable for analysis requiring long time series or for the study of phenomena pre-1997. Second, HM examine the text in the Liquidity and Capital Resources subsection of the "Management Discussion and Analysis" (MD&A) section of firms' 10-Ks. This limits the coverage of their measure. They are only able to classify 42%–68% of U.S. domestic firms in COMPUSTAT each year (excluding firms in the finance and utilities industries). Expanding this coverage has the potential to significantly increase the power of most tests even within the HM sample period.[7]

Another concern that arises from the use of 10-K filings to construct the HM measure is potential reporting bias. Although the direction of the bias is unclear. It is possible that constrained firms will strategically under-report their financing issues since this revelation may hurt firm value. HM argue that the likely bias is in the other direction: Less constrained firms will not fill out the relevant subsection and, therefore, will not be in their sample. If this is the case, then the most constrained firms will still be classified by their method. As discussed below, we use a random forest model to create a mapping between accounting variables and financial constraints as measured by HM. A benefit of the random forest methodology we employ is that we are able to classify firms even if firms' selectively report their constraints in the text. Importantly, even if there is a slight reporting bias in the HM classifications they should still provide relevant classifications as long as there are enough constrained firms discussing their constraints in the relevant subsection.

D. Our Methodology

Our approach is to model the text-based measure as a function of a broad set of measurable accounting variables that are common to the literature studying financial constraints. Specifically, we estimate the following equation:

[6]The two types of constraints are not mutually exclusive. For example, equity-constrained firms may also face constraints in raising debt financing.

[7]Buehlmaier and Whited (2018) propose an alternative method to expand the cross section of firms with constraint estimates. Their method involves further analyzing the text of firms' 10-Ks and cannot expand to firm years without 10-Ks.

$$\text{CONSTRAINT}_{i,t} = f(X_{i,t}) + \varepsilon_{i,t}, \quad (1)$$

where $\text{CONSTRAINT}_{i,t}$ is the level of debt or equity constraint for firm i in year t and $X_{i,t}$ is a set of predictor variables. We discuss the set of predictor variables in Section II.E. As discussed in Section II.B, we train the model using the equity and debt financing constraint measures of HM (2015). We focus on the debt and equity financing constraint measures, but our method can be applied to the other two measures as well. After training the model, we extend the model to all firm-years in the COMPUSTAT data file between 1972 and 2021 that have the requisite predictor variables.

We use random forests to estimate firm-level financial constraints because they offer a very flexible alternative to linear models like ordinary least squares regression or an ordered probit.[8] Random forests allow for, and are very effective in detecting and incorporating nonlinear relationships and interaction effects between dependent and explanatory variables. We find that both nonlinearities and interaction effects are prevalent in the relationship between the text-based HM measures and the set of accounting variables typically associated with financial constraints. Importantly, these nonlinearities and interactions do not need to be specified ex ante, they are instead uncovered by the algorithm. The researcher only selects the set of predictor variables. This removes a layer of subjectivity from the classification process, which should minimize concern of omissions of important relationships between accounting variables and financial constraints. Furthermore, random forests are able to incorporate a large number of predictors without the need to invert covariance matrices of predictors thus circumventing the multicollinearity problems commonly associated with linear regressions.

An important reason for creating our measure is to extend the coverage of the HM data both in the cross section and the time series. For this reason, we take care to ensure that we minimize the potential for problems associated with possible time effects or structural effects. These types of effects could cause difficulty in extrapolating from the training set to out-of-sample sets of data. This is important in practice because a number of previously constructed financial constraint indices are used by economists who apply in-sample parameter estimates to out-of-sample data.

We transform the data in two ways to mitigate problems arising due to extrapolation. First, within each year we normalize the HM constraint measures among firms that are included in our sample. The HM measures themselves are constructed by taking residuals of regressions on common, "boilerplate" content and are, therefore, approximately standardized themselves. Our normalization just ensures standardization within our sample. This standardization creates a distribution of measures that is approximately the same from year to year which helps us avoid the problem of time variation in the distribution of constraints over

[8]One could employ any of a number of statistical algorithms in place of random forests. We find constraint estimates using gradient boosting and support vector regressions are highly correlated with random forest-based estimates. We use random forest regression because it has the highest in-sample and out-of-sample predictive performance and due to its more intuitive nature.

time. Since our goal is to classify firms as most to least constrained relative to other firms within a given year, this normalization is natural and has the benefit that our nonparametric model does not have to contend with time variation in the data-generating process.

Second, for each predictor variable, each year, we transform the variable into its percentile rank for the year. That is, for each predictor, each year, we transform numerical values to percentile values. This mitigates problems related to extrapolation that may arise if there is a trend or other temporal effects across firms where levels of predictor variables vary at different times within the training sample. This transformation of predictors has become a common practice in the literature relating firm characteristics to unconditional stock returns (e.g., Kelly, Pruitt, and Su (2019), Freyberger, Neuhierl, and Weber (2020)) in a manner similar to our goal of mapping firm characteristics to constraints. By transforming the values of all predictors to percentiles within a year, the value relative to other firms in the year is all that matters for our model estimation. These steps optimize our model's ability to measure constraints cross-sectionally. However, the choice of HM as our training set, while providing strong cross-sectional constraint information, prohibits the ability of our measures to describe time-series variation in aggregate constraint levels.

E. Predictor Variables

1. Model with Full Set of Predictors

An important consideration for our methodology is the selection of accounting-based predictor variables. One approach would be to include a large number of predictors and let the random forest algorithm select the relevant variables. A drawback to this approach is that if any variable value is missing for a firm-year, the firm-year cannot be classified without forcing the algorithm to impute values of the missing variables. Therefore, we face a trade-off between coverage and accuracy in selecting the set of variables. The set of accounting variables we include in our predictor set was selected with this trade-off in mind.

We limit our set of accounting variables to those used in three prior accounting-based measures of financial constraints. Specifically, we include the underlying variables used to construct the Kaplan and Zingales (1997), Whited and Wu (2006), and Hadlock and Pierce (2010) measures. Those three measures use the following characteristics: the ratio of cash flow to k (where $k =$ previous year property plant and equipment), the ratio of cash to k, dividends to k, Tobin's q, the ratio of debt to total capital, sales growth (both firm-level and industry-level), age, and size.

Most of the previously used characteristics are combinations of multiple Compustat-level variables (e.g., sales growth is a combination of sales (SALE) and lagged sales (LAG_SALE)). To give the random forest as much flexibility as possible, we use the 19 subcomponents as our main set of predictors (e.g., SALE and LAG_SALE, instead of sales growth). We use variable definitions described by Hadlock and Pierce (2010) in reconstructing the Kaplan and Zingales (1997), and Whited and Wu (2006) indices and creating their size and age index. In Table 2, we provide the list of variables and their definitions. We also report which prior constraint indices use the variable in their construction. While the results of most

TABLE 2
Model Variables

Table 2 gives descriptions of the Compustat variables used in the random forest model. The first column of each row lists the symbol of the variable as presented in the article. The second column gives the definition of the variable. The third column lists the transformed variables used in prior indices that the given variable is used to construct. The fourth column lists the existing financial constraints indices that use the variable in their construction. The indices are proposed or derived from results by: Kaplan and Zingales (1997) (KZ), Whited and Wu (2006) (WW), and Hadlock and Pierce (2010) (HP). * indicates the inclusion of the contemporaneous and the lagged version of the variable in our "Full" random forest model.

Symbol	Definition	Use	Index
AGE	Number of years in Compustat with nonmissing stock price (beginning in 1950)	Age	HP
AT	Total assets	Size, Tobin's Q	HP, KZ
CEQ	Common ordinary equity	TOBIN'S Q	KZ
CHE	Cash and short term investments	Cash/K	KZ
DLC	Debt current liabilities – Total	Debt-to-total capital	KZ, WW
DLTT	Long-term debt – Total	Debt-to-total capital	KZ, WW
DP	Depreciation and amortization	Cash-flow/K	KZ, WW
DVC	Dividends common/Ordinary	Dividends/K, dividend dummy	KZ, WW
DVP	Dividends – Preferred/Preference	Dividends/K, dividend dummy	KZ, WW
IB	Income before extraordinary items	Cash-flow/K	KZ, WW
SALE*	Sales/Turnover (Net)	Sales growth	WW
SEQ	Stockholders' equity – Total	Debt-to-total capital	KZ, WW
SIC3_SALES*	Total sales in 3-digit SIC	Industry sales growth	WW
TXDB	Deferred taxes (Balance sheet)	Tobin's Q	KZ
CSHO	Common shares outstanding	Tobin's Q	KZ
PRCC_F	Price close – Annual – Fiscal	Tobin's Q	KZ
PPEGT	Property, plant, and equipment	Cash-flow/K, cash/K, Dividends/K	KZ

tests in the article are broadly similar whether we use the 9 characteristics of the 3 prior constraint indices or the 19 subcomponents, we use the 19 subcomponents as they provide superior out-of-sample fit.

2. Model with Primitive Predictors

Our inclusion of 19 predictors in our main model stands in contrast to the only two predictors used in the index of Hadlock and Pierce (2010). A strength of the Hadlock and Pierce (2010) model is that it only relies on firm size and age, neither of which are typically considered to be endogenously determined by firms' managers. The cost of using only two less endogenous predictors is the potential neglect of important information in other accounting variables. Researchers, therefore, face a trade-off: including more information has the potential to make a model more informative, however, potential endogeneity in predictors can result in misleading inference.

We address potential endogeneity concerns by constructing an alternative version of our constraint measures that uses only primitive firm characteristics as predictors. Specifically, we include only total assets (AT), age (AGE), lagged industry sales (LAG_SIC3_SALES), and contemporaneous industry sales (SIC3_SALES) in the "Primitive"-version of the model. Firm size and age are used by the Hadlock and Pierce (2010) measure, and the industry sales variables are the components of industry sales growth, which is unlikely to be a choice variable of the firm. We conduct most of our analysis using estimates of constraints both from the more comprehensive model with the full set of predictors (labeled "Full"), and the model with the more primitive predictors (labeled "Primitive").

F. Data and Sample Construction

To build our sample, we start with the entire Compustat annual file of firms and keep data from 1972 to 2021, exclude financial firms and utilities, and remove any firm-year for which at least one of the predictor variables is missing.[9] We detail the list of Compustat variables we use as predictors in Table 2. We do not inflation-adjust any of the predictors because they are all transformed to percentiles within a given year before they enter our algorithm. We start the sample in 1972 to balance the desire to increase the time-series coverage of our measure against the potential for greater extrapolation error as we extend further back in time. Further, the NASDAQ opened in 1971, meaning 1972 is the first full year of trading on the NASDAQ. We merge this set of firms with the HM data. The most recent update of data from HM reports annual measures beginning in 1997, running through 2015. Therefore, 1997–2015 is the time period of our training sample.

-Summary statistics for the predictor variables and the constraint measures are provided in Table 3. We provide separate summary statistics for the subsample in which there are HM(2015) financial constraint data in the Supplementary Material. We construct a number of additional variables using Compustat data. These variables are used in tests to assess the informativeness of our financial constraint classifications in Section IV. The yearly change in payouts to shareholders (ΔPAYOUT), the yearly change in equity issuance proceeds (ΔEQUITY_ISSUANCE), the change in other funding sources (ΔOTHER_ FUNDING), and change in firm size (ΔSIZE) are all defined as in Farre-Mensa and Ljungqvist (2016). The indicator for a firm omitting its dividend (DIVIDEND_OMISSION_DUMMY), the indicator for the firm increasing its dividend (DIVIDEND_INCREASE_DUMMY), and an indicator for a firm under-funding its pension (UNDERFUND_PENSION_DUMMY) are defined as in Bodnaruk et al. (2015). The dividend-related indicators require the firm to have paid a dividend in the prior year and the underfunded pension dummy requires the firm to have a pension, hence the lower observation numbers.

In Section III.B, we examine how our estimated constraints are related to certain firm-level characteristics that have previously been associated with constraints. We construct those characteristics as follows: AGE is defined by the number of years a given company has been listed in the Compustat database at the time of each annual observation, CASH_FLOW is defined as operating income (IB) plus depreciation (DP), cash is defined as cash plus marketable securities (CHE), dividends are total annual dividend payments (DVC + DVP), TOBINS_Q is defined as book assets (AT) minus book common equity (CEQ) minus deferred taxes (TXDB) plus market equity (PRCC_F × CSHO), all divided by book assets (AT), LEVERAGE is defined as the ratio of short-term (DLC) plus long-term debt (DLTT) to short term (DLC) plus long-term debt (DLTT) plus total stockholders' equity (SEQ), SALES_GROWTH is defined as the ratio of sales (SALE) in year t, minus sales in year $t-1$ (LAG_SALE) to sales in year $t-1$, SIZE is the log of book assets. SIC3_SALES_GROWTH is defined as the aggregate sum of all sales within a given 3-digit SIC code (SIC3_SALES) minus the previous year's total sales

[9]We limit our COMPUSTAT sample to domestic firms and use the consolidated financial statements.

TABLE 3

Firm-Level Summary Statistics

Table 3 presents summary statistics of the financial variables used as predictors in our model of financial constraints, constraint estimates, and additional variables used in our analysis. Each variable is measured at the annual frequency. Variable descriptions of the predictors used in our model (AGE–PPEGT) are provided in Table 2. HM_DEBT (EQUITY) are the standardized Hoberg and Maksimovic (2015) constraint measures. RF_DEBT (EQUITY) are the full-model random forest constraint estimates. RF_PRIMITIVE_DEBT (EQUITY) are the primitives-model random forest constraint estimates. The yearly change in payouts to shareholders (ΔPAYOUT), the yearly change in equity issuance proceeds (ΔEQUITY_ISSUANCE), the change in other funding sources (ΔOTHER_FUNDING), and change in firm size (ΔSIZE) are all defined as in Farre-Mensa and Ljungqvist (2016). The indicator for a firm omitting its dividend (DIVIDEND_OMISSION_DUMMY), the indicator for the firm increasing its dividend (DIVIDEND_INCREASE_DUMMY), and an indicator for a firm under-funding its pension (UNDERFUND_PENSION_DUMMY) are defined as in Bodnaruk et al. (2015). The dividend-related indicators require the firm to have paid a dividend in the prior year and the underfunded pension dummy requires the firm to have a pension.

Variable	No. of Obs.	Mean	Std. Dev.	P25	P50	P75
AGE	251,202	15.034	12.054	6	11	21
AT	251,202	1,675.395	5,882.795	18.21	89.777	538.237
CEQ	251,202	614.705	2,134.623	6.73	40.03	224.151
CHE	251,202	155.038	563.669	1.124	7.733	51.727
DLC	251,202	72.759	317.059	0.048	1.43	11.138
DLTT	251,202	394.031	1,415.225	0.032	5.353	91.183
DP	251,202	76.041	279.773	0.516	3.154	22.226
DVC	251,202	28.479	129.557	0	0	1.272
DVP	251,202	0.374	2.028	0	0	0
IB	251,202	68.046	324.734	−2.715	1.038	16.514
SALE	251,202	1,369.534	4,654.778	12.589	82.808	506.189
SEQ	251,202	621.584	2,149.419	7.124	40.964	228.697
SIC3_SALES	251,202	195,707.3	293,365.9	14,182.95	56,287.73	259,698.8
TXDB	251,202	56.454	245.433	0	0	5.845
CSHO	251,202	70.287	183.86	5.205	16.28	50.308
PRCC_F	251,202	15.34	20.778	2	7.75	20.125
PPEGT	251,202	1,058.773	4,058.585	5.524	34.151	256.422
HM_DEBT	69,982	0	1	−0.722	−0.118	0.62
HM_EQUITY	69,982	0	1	−0.732	−0.124	0.606
RF_DEBT	251,202	0.011	0.573	−0.421	−0.024	0.403
RF_EQUITY	251,202	0.048	0.64	−0.401	−0.088	0.432
RF_PRIMITIVE_DEBT	251,202	0.013	0.540	−0.365	−0.026	0.354
RF_PRIMITIVE_EQUITY	251,202	0.020	0.605	−0.399	−0.067	0.361
ΔPAYOUT	178,750	0.002	0.048	0	0	0.001
ΔEQUITY_ISSUANCE	194,743	0.039	0.463	0	0	0
ΔOTHER_FUNDING	114,276	0.077	1.126	−0.115	0.013	0.162
ΔSIZE	232,770	0.013	0.807	−0.08	0.051	0.197
DIVIDEND_OMISSION_DUMMY	99,446	0.084	0.277	0	0	0
DIVIDEND_INCREASE_DUMMY	99,504	0.615	0.487	0	1	1
UNDERFUND_PENSION_DUMMY	38,868	0.655	0.475	0	1	1

(LAG_SIC3_SALES) within that SIC code, all divided by the previous year's total sales within the SIC code (LAG_SIC3_SALES).

Lastly, in our application, we use the Baker and Wurgler (2006) quarterly investor sentiment data for the 1972–2021 time period obtained from Jeffrey Wurgler's website. We use institutional investor holdings (13F) data obtained from the Thomson Reuters Institutional Holdings (s34) database for the time period of 1980:Q1 to 2021:Q4. We use data on Robinhood investors' positions obtained from Robintrack.net, which is available from May 2, 2018, to Aug. 13, 2020. For each stock, each day, Robintrack provides the total number of account holders invested in the stock. For more details on the Robintrack data see Welch (2022). Finally, we use holdings data for accounts at a large discount brokerage firm over the time period of Jan. 1991 to Nov. 1996. This is the data used by Barber and Odean (2000).

III. Performance of the Random Forest Model

In this section, we characterize the financial constraint estimates from our random forest model. We examine the additional coverage provided by our model, the relationship between firm characteristics and financial constraints, and discuss the predictive performance of the random forest model compared to an alternative OLS model.

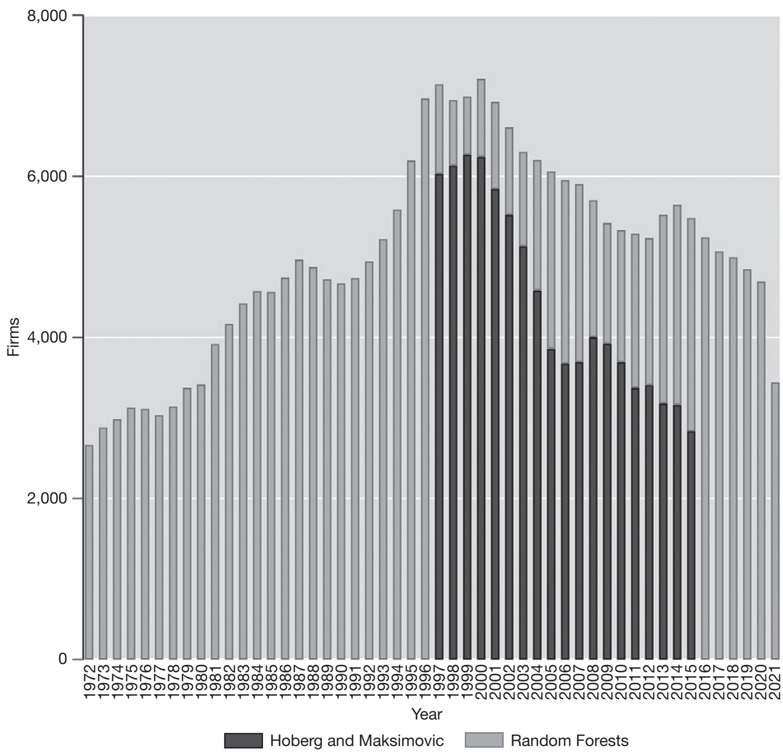
A. Expanded Coverage

Figure 1 exhibits the additional coverage gained by our random forest model. We plot the number of firms with a HM (2015) constraint estimate and the number of firms with a random forest estimate each year from 1972 to2021. The HM measures cover the 1997–2015 time period. We are able to expand coverage to 31 additional years (1972–1996 and 2016–2021) adding an average of 4,365 firms per year in these years. Within the HM sample period, the random forest methodology increases coverage by 12% to 93% per year with an average of 43% per year. The years post-2004 exhibit the greatest increase in coverage. In total, we are able to estimate constraints for 251,202 firm-years, compared to 84,552 firm-years in the original HM sample. This is a 197% increase in the number of classified firm-years relative to the HM sample. This increased coverage should increase the power of

FIGURE 1

Financial Constraint Classification Coverage

Figure 1 depicts the number of firms classified by our random forest model and the number of firms in the Hoberg and Maksimovic (2015) annual financial constraint classifications each year.



https://doi.org/10.1017/S0022109023000650 Published online by Cambridge University Press

most tests and allow for analyses of time periods outside of the period covered by HM.

B. Relationship Between Predictors and Financial Constraints

Since random forests are a more complicated technique than linear regression, understanding the relationship between individual predictors and financial constraints is not as straight forward as simply looking at a regression coefficient and its standard error. However, we are able to exploit the richer estimation technique in different ways to understand which firm characteristics appear to be related to financial constraints.

We first calculate the variable importance for each predictor variable. Variable importance is a measure of how the variable decreases "impurity" (here, variance). The measure is the average reduction across each node of each tree of the random forest. The greater the average reduction in variance the more important the variable. We normalize the measures of variable importance so that the largest variable importance measure is equal to 1. Interpreting the variable importance for the individual predictors can only bring so much insight as the individual variables are not as economically relevant as their underlying combinations (e.g., price and shares outstanding combine for market value, or the combination of predictors into Tobin's q).

Graph A of Figure 2 shows the variable importance of each predictor variable for the equity constraint model. Lagged sales (LAG_SALE) is by far the most important predictor. Next, is income before extraordinary items (IB), followed by shares outstanding (CSHO), age (AGE), and cash and short-term investments (CHE) which all have similar importance.

In Graph B of Figure 2, we report the variable importance of each predictor for the debt constraint model. Long-term debt (DLTT) is the most important predictor, followed by cash and short-term investments (CHE) with lagged sales (LAG_SALE), and debt in current liabilities (DLC). It is perhaps unsurprising that debt, cash, and sales are important predictors for debt-focused constrained firms.

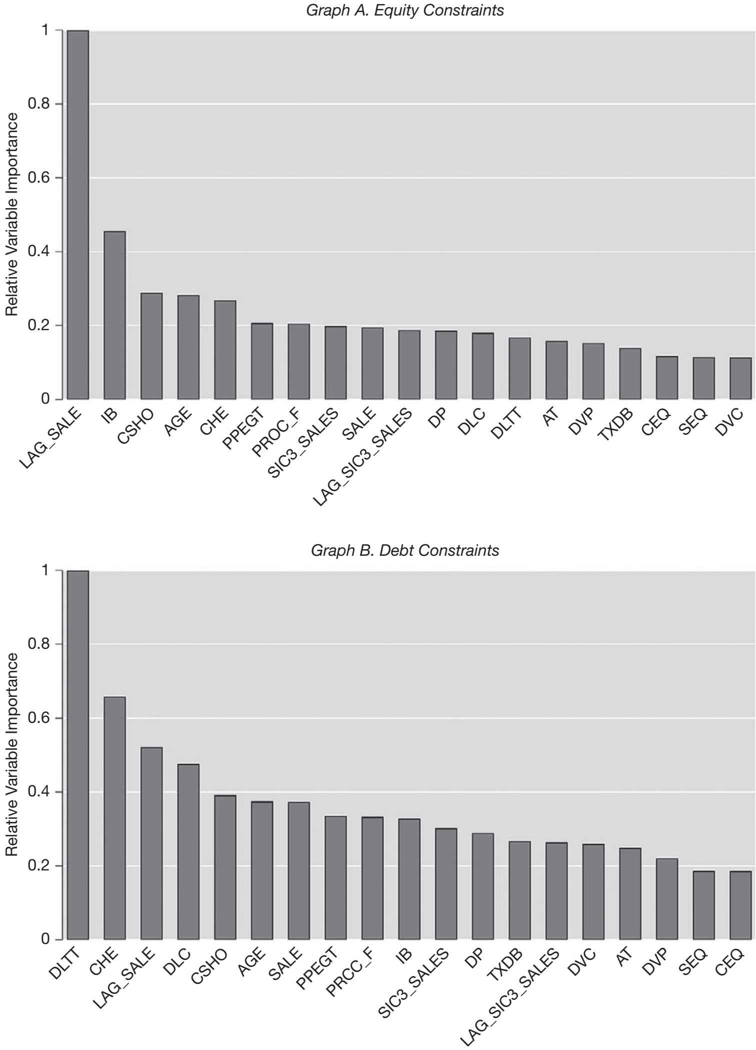
Our set of predictor variables is the subcomponents of 9 accounting variables (or measures) that have been previously related to financial constraints. In Figures 3 and 4, we examine how the HM constraint measures and our measures are related to these 9 variables. We find the random forest does a very good job matching the underlying relationships between the HM constraint measures and these accounting variables. Comparing Figures 3 and 4, it is clear that debt and equity constraints befall very different types of firms. Equity-constrained firms tend to be younger, smaller firms, with lower cash flow, and higher Tobin's q. Debt-constrained firms tend to be slightly older and relatively larger, with greater leverage, and less cash. The differences across equity and debt-constrained firms highlight the importance of modeling the two types of constraints separately (see HM (2015) for further discussion).

As is clearly seen in the figures, significant nonlinearities exist between constraints and the accounting variables. These nonlinearities are present in the underlying data as is demonstrated by the blue curves depicting the projections

FIGURE 2

Predictor Variable Importance

Figure 2 depicts the relative importance of each predictor used in our random forest model for equity constraints (Graph A) and debt constraints (Graph B). Variable importance is based on the average reduction in variance for each predictor variable. We normalize the variable importance of each predictor by the importance measure of the predictor with the highest importance.

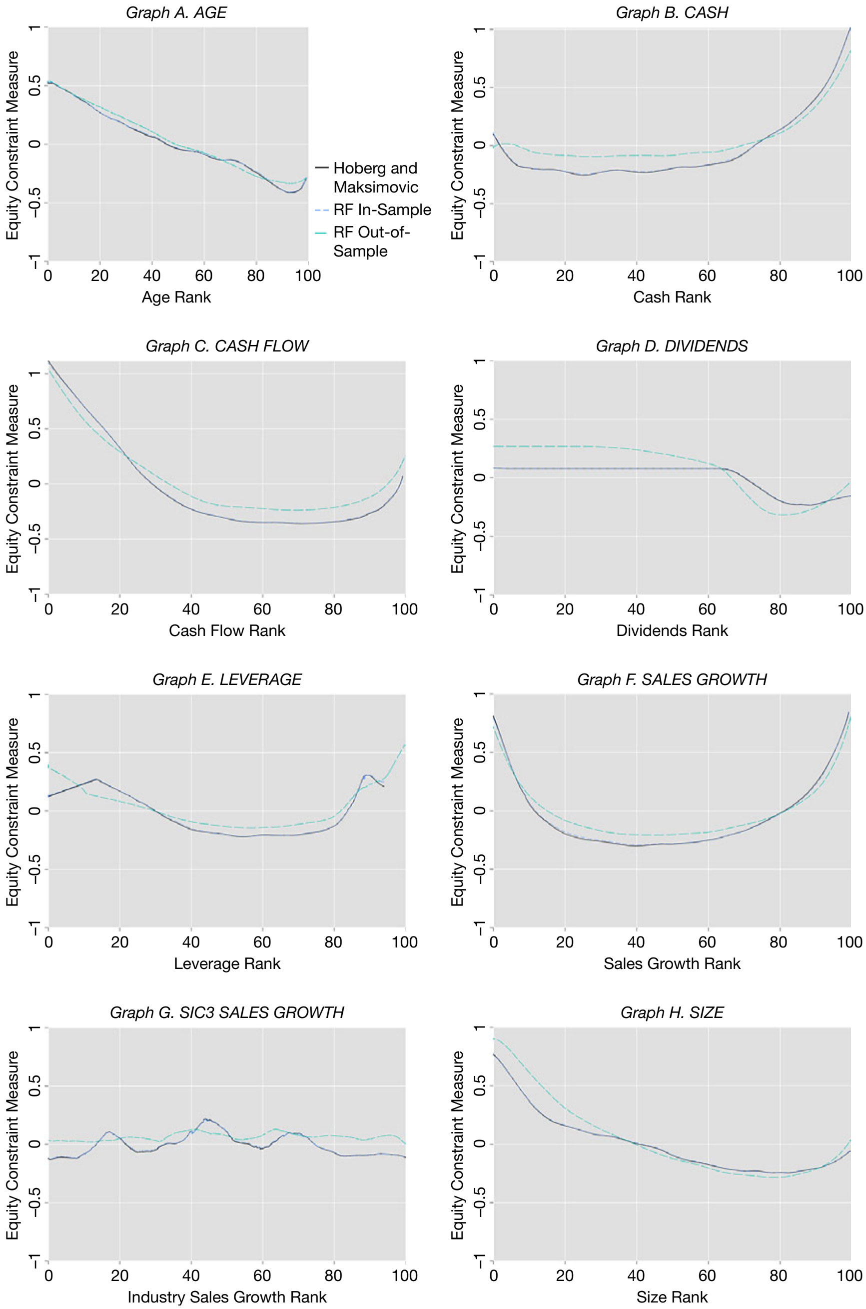


of the HM constraints onto each accounting variable. The random forest model appears to capture these nonlinearities quite well. While it is possible to capture these nonlinearities using a least squares regression model (e.g., by including additional higher order terms in the regression specification), this would require prespecification of a functional form and, likely, significant trial and error to find the specification that best fits the data or performs the best out of sample. The random forest is a much more efficient method for capturing these nonlinearities.

FIGURE 3

Equity Constraint Predictor Variables Lowess Plots

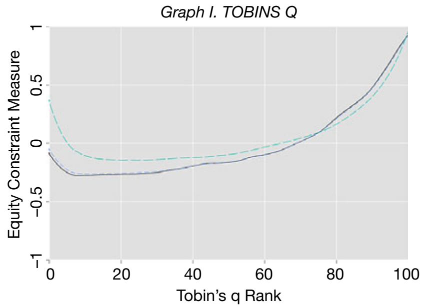
Figure 3 depicts the relationship between various firm characteristics that have been previously associated with constraints and the equity constraint measures. The characteristic of interest for each graph is stated on the *x*-axis and in the graph title. We plot the lowess-smoothed equity constraint classifications as a function of the percentile rank of the characteristic of interest. We present the relationship for the standardized Hoberg and Maksimovic (HM) (2015) constraint measure, the random forest (RF) constraint measure for in-sample firm-years also classified by HM (2015), and the random forest constraint measure for out-of-sample firm-years not classified by HM (2015).



(continued on next page)

FIGURE 3 (continued)

Equity Constraint Predictor Variables Lowess Plots



C. Model Performance

In this section, we examine the performance of the random forest. We examine the in- and out-of-sample fits, and we compare the performance of the random forest model to an ordinary least squares model using Monte Carlo cross-validation.

To examine the in-sample performance of our model, we present cross-tabulations of the number of firms from each HM quintile that are in each random forest quintile in Table 4. We present results both for the full model (Panels A and B) and the "Primitive" model (Panels C and D). Quintiles are defined by ranking HM and model predicted HM values each year. The goal is to have a large number of observations on the diagonal of the cross-tabulation table as this is indicative of a correct classification. For the purposes of this article, we aim particularly to properly classify the least constrained (quintile 1) and most constrained (quintile 5) firms. The random forest tends to correctly classify both the debt and equity constraints across all levels of constraints (quintiles 1–5), for both the full and "Primitive" constraint estimators. This suggests the model does a very good job of estimating firms' constraint levels within the set of firm years that are included in the HM sample.

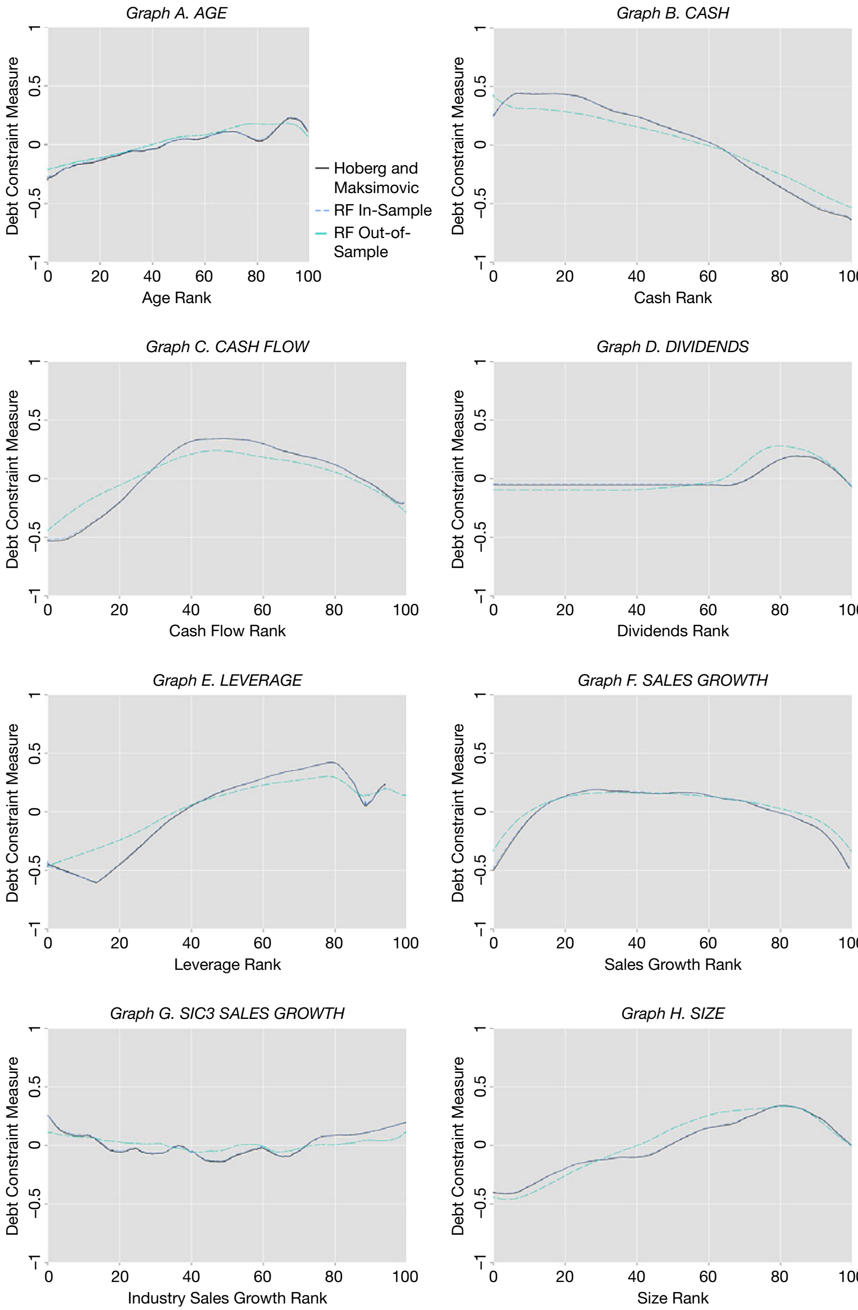
While in-sample fit is important, we are especially interested in the ability of the random forest to predict constraints out of sample. Table 5 shows the out-of-sample analogue of Table 4. We estimate the out-of-sample constraints in the following manner. We first divide the HM sample into 20 equal-sized subsamples. For each subsample, we train the random forest on all observations except for the (left out) subsample of interest. We then fit the model for the subset of interest to obtain the out-of-sample estimates for the left-out subset. We examine the resultant cross-tabulations in Table 5. In Panels A and B, the $(1,1)$ and $(5,5)$ entries are heavily populated indicating the random forest using the full set of predictors performs reasonably well out of sample. The "Primitive" model also does fairly well at estimating constraints out of sample, though the performance is somewhat poorer than that of the full model.

Finally, we perform Monte Carlo cross-validation (MCCV) analyses to ensure that the random forest exhibits superior out-of-sample performance compared to a similarly defined OLS estimator. First proposed by Picard and Cook (1984), and shown to be highly effective in model selection by Shao (1993), MCCV provides a

FIGURE 4

Debt Constraint Predictor Variables Lowess Plots

Figure 4 depicts the relationship between various firm characteristics that have been previously associated with constraints and the debt constraint measures. The characteristic of interest for each graph is stated on the *x*-axis and in the graph title. We plot the lowess-smoothed debt constraint measures as a function of the percentile rank of the characteristic of interest. We present the relationship for the standardized Hoberg and Maksimovic (HM) (2015) constraint measure, the random forest (RF) constraint measure for in-sample firm-years also classified by HM (2015), and the random forest constraint measure for out-of-sample firm-years not classified by HM (2015).



(continued on next page)

FIGURE 4 (continued)

Debt Constraint Predictor Variables Lowess Plots

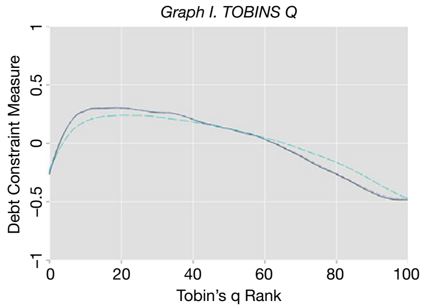


TABLE 4

In-Sample Random Forest Classifications Cross-Tabulation

Table 4 depicts in-sample cross-tabulations of the random forest constraint measures and Hoberg and Maksimovic (2015) (HM) constraint measures. Each column represents a quintile of the HM constraint measures. Each row represents a quintile of the random forest model. We present results both for the full model and the model with only arguably "primitive" predictors. The set of predictors used and the constraint measure estimated is denoted above each table.

	1	2	3	4	5	Total
Panel A. Equity Constraints (Full Model)						
1	12,272	1,732	1	–	–	14,005
2	1,624	10,109	2,263	–	–	13,996
3	107	2,030	9,690	2,171	–	13,998
4	2	125	2,017	10,256	1,596	13,996
5	–	–	27	1,569	12,391	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982
Panel B. Debt Constraints (Full Model)						
1	11,999	2,001	5	–	–	14,005
2	1,921	9,285	2,773	17	–	13,996
3	85	2,640	8,799	2,468	6	13,998
4	–	70	2,418	10,032	1,476	13,996
5	–	–	3	1,479	12,505	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982
Panel C. Equity Constraints ("Primitive" Model)						
1	11,762	2,230	13	–	–	14,005
2	2,019	9,177	2,777	23	–	13,996
3	200	2,360	8,876	2,552	10	13,998
4	23	225	2,244	9,634	1,870	13,996
5	1	4	88	1,787	12,107	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982
Panel D. Debt Constraints ("Primitive" Model)						
1	11,616	2,362	26	1	–	14,005
2	2,255	8,801	2,887	53	–	13,996
3	134	2,729	8,612	2,504	19	13,998
4	0	104	2,455	9,811	1,626	13,996
5	–	–	18	1,627	12,342	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982

comprehensive way of examining out-of-sample predictive performance. In the OLS model, we include the same set of percentile ranked predictors as the random forest. The OLS analogue of equation (1) is thus given by: $\text{CONSTRAINT}_{i,t} = a + \sum_{i=1}^{19} \beta_i X_{i,t} + \varepsilon_{i,t}$.

TABLE 5

Out-of-Sample Random Forest Classifications Cross-Tabulation

Table 5 depicts cross-tabulations for out-of-sample random forest constraint measures and Hoberg and Maksimovic (2015) (HM) constraint measures. We calculate out-of-sample constraints by randomly splitting the data into 20 subsamples. For each subsample, the model is trained on the rest of the data, then evaluated using the left-out subsample. This gives an out-of-sample estimate for all observations. Each column represents a quintile of the HM constraint measures. Each row represents a quintile of the random forest model constraint measure. We present results both for the full model (Panels A and B) and the model with only "primitive" predictors (Panels C and D). The set of predictors used and the constraint measure estimated is denoted above each table.

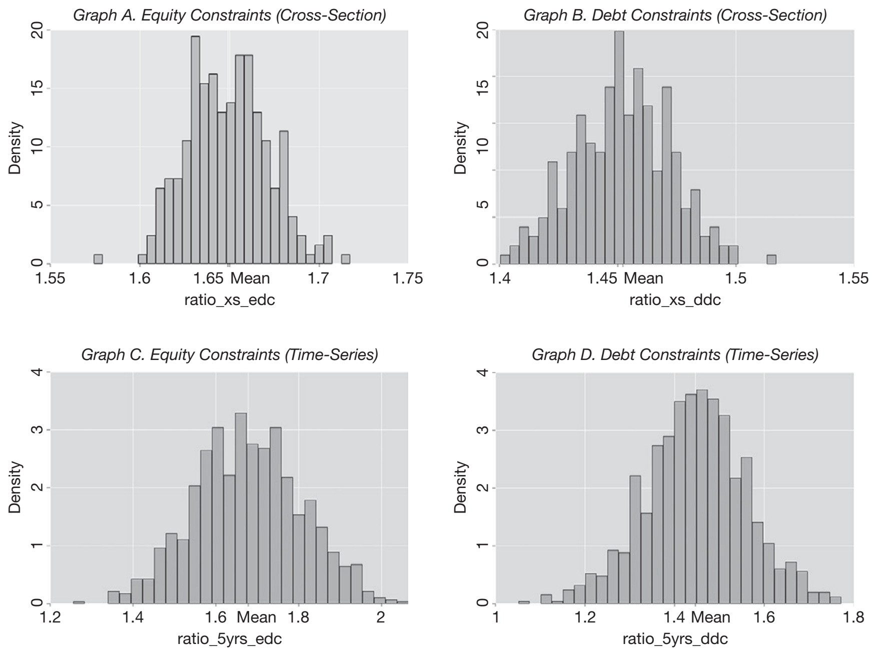
	1	2	3	4	5	Total
Panel A. Equity Constraints (Full Model)						
1	7,290	3,977	1,793	757	188	14,005
2	3,310	4,285	3,673	2,104	624	13,996
3	1,932	2,893	3,976	3,662	1,535	13,998
4	1,075	1,997	3,080	4,343	3,501	13,996
5	398	844	1,476	3,130	8,139	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982
Panel B. Debt Constraints (Full Model)						
1	6,503	3,771	2,331	1,105	295	14,005
2	3,760	4,002	3,282	2,163	789	13,996
3	2,187	3,173	3,509	3,202	1,927	13,998
4	1,111	2,047	3,042	4,048	3,748	13,996
5	444	1,003	1,834	3,478	7,228	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982
Panel C. Equity Constraints ("Primitive" Model)						
1	5,172	3,690	2,627	1,709	807	14,005
2	3,359	3,540	3,304	2,443	1,350	13,996
3	2,610	2,939	3,224	3,180	2,045	13,998
4	1,848	2,416	2,934	3,549	3,249	13,996
5	1,016	1,411	1,909	3,115	6,536	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982
Panel D. Debt Constraints ("Primitive" Model)						
1	4,758	3,598	2,809	1,852	988	14,005
2	3,488	3,379	3,026	2,442	1,661	13,996
3	2,621	2,907	2,944	2,940	2,586	13,998
4	1,917	2,350	2,772	3,304	3,653	13,996
5	1,221	1,762	2,447	3,458	5,099	13,987
Total	14,005	13,996	13,998	13,996	13,987	69,982

We first examine the out-of-sample performance in the cross section. For this, we hold out a random selection of 25% of the HM data (firm-years) as our test data. We then train the random forest and the OLS models on the remaining 75% of the data. After training the models, we examine the fit using the test data (i.e., predict constraint classifications) and calculating the ratio of the out-of-sample R^2 for the random forest model (RF) to the out-of-sample R^2 for the ordinary least squares model (OLS) $\left(\frac{\text{OOS}R^2 \text{ of RF}}{\text{OOS}R^2 \text{ of OLS}}\right)$. We repeat this 1,000 times and examine the distribution of the resulting out-of-sample R^2 ratios. Values above 1 correspond to a larger out-of-sample R^2 for the random forest than for the OLS projection. Graphs A and B of Figure 5 plot the distribution of these ratios for models fitting equity and debt constraints, respectively. For equity-focused constraints, realized values range between 1.57 and 1.72 with a mean of approximately 1.65. Debt-focused constraints have an out-of-sample ratio that averages around 1.45. In other words, using the random forest delivers an out-of-sample R^2 that is 65% and 45% larger on average than the OLS out-of-sample R^2 for equity and debt focused constraints, respectively.

FIGURE 5

Random Forest Out-of-Sample Fits Compared to OLS

Figure 5 depicts histograms of the ratio of out-of-sample R^2s for the full random forest model (RF) compared to an OLS model with the same predictors ((RF OOS R^2)/(OLS OOS R^2)). In Graphs A and B, we hold out a random sample of 25% of the firm-year observations for testing, train each model (RF and OLS) on the remaining data, then calculate the out-of-sample R^2 from regressing the standardized Hoberg and Maksimovic(2015) measures on the model-implied measures using the hold-out data. We repeat this process 1,000 times. For Graphs C and D, we run an analogous procedure, but randomly select 5 years of data as our hold-out sample each iteration. The constraint type (debt or equity) is indicated below each histogram.



To assess the ability of the random forest to fit the data when expanding coverage to a sample of dates outside the original HM sample, we perform a similar MCCV. The only difference is that we randomly select 5 calendar years (which need not be contiguous) to leave out of the training sample at each of 1,000 iteration. Resulting R^2 ratios are presented in Graphs C and D of Figure 5. On average, the random forest out-of-sample R^2 is 68% and 45% larger than the OLS out-of-sample R^2 for equity and debt constraints, respectively. These results highlight the significant improvement in out-of-sample predictability achieved by the random forest model. They also suggest that when extrapolating outside of the training sample, the random forest will deliver significantly superior performance than the standard linear model.

IV. Do Constraint Classifications Capture Financial Constraints?

In Section IV, we examine whether firms classified as constrained by the random forest actually behave as if they are constrained. While a number of tests have been proposed in the literature to validate constraint estimates, we focus on the set of tests used by Bodnaruk et al. (2015), one of which was originally proposed by Farre-Mensa and Ljungqvist (2016). Bodnaruk et al. (2015) show that the commonly used accounting indices (Kaplan and Zingales (1997), Whited and Wu (2006), and Hadlock and Pierce, (2010)) are unable to pass more than half of the proposed tests. This set of tests, while certainly not exhaustive, presents an established and reasonably high bar with which to assess our measures.

The first test examines the relative equity recycling behavior of constrained versus unconstrained firms following the methodology first proposed by Farre-Mensa and Ljungqvist (2016). Equity recycling entails raising equity financing and increasing payouts to equity holders in the same period. The motivation for this test is that constrained firms should recycle equity less than unconstrained firms if more constrained firms face a greater wedge in the cost of financing between external equity funds and internal funds.[10]

In our tests, the regression is specified exactly as in Farre-Mensa and Ljungqvist (2016). We regress the yearly change in payouts to shareholders on the yearly change in equity issuance proceeds (ΔEQUITY_ISSUANCE). We control for the change in other funding sources (ΔOTHER_FUNDING) and change in firm size (ΔSIZE). We include industry-by-year fixed effects. We run separate regressions for firms classified as most constrained and firms classified as least constrained and compare the coefficient on the change in equity issuance across the two regressions. For this analysis, we consider the top 20% of firms as most constrained and bottom 20% as the least constrained. We focus on equity constraints for this analysis as the test is focused on equity recycling. In the Supplementary Material, we provide the results for the debt constraint measures. If our classifications are capturing constraints, we expect a significantly smaller ΔEQUITY_ISSUANCE coefficient for the most constrained firms relative to the least constrained firms.

The results are presented in Table 6. We classify firms using the current period constraint classification (the current period classification is based on the previous year's accounting variables). We present results for regressions examining the full random forest model classification using the entire sample period, the period pre-1997 (i.e., out-of-sample), and the period covering the years 1997 and later (mostly in-sample). Across all three sample periods, we find large and significant differences in the equity recycling behavior of the most constrained (labeled "Constrained") and least constrained (labeled "Unconstrained") firms. Examining the full sample period results, we see the estimated coefficient for the most constrained firms is 0.002, which is much smaller than the coefficient for the least constrained firms of 0.024. This difference is economically significant with the least constrained firms increasing payouts to equity holders for each dollar of equity raised over 10 times as much as the most constrained firms. A Wald Test indicates the difference in coefficients is statistically significant at the 1% level. We see a similar pattern for both the pre-1997 subperiod and the post-1997 subperiod. Based

[10]Fazzari, Hubbard, and Petersen (1988) first proposed that more constrained firms face a greater wedge between external and internal costs of capital than less constrained firms. This wedge may arise due to information asymmetry between the firm and financial capital suppliers (Tirole (2006)). This definition is different from the curvature of the capital supply curve definition (e.g., Stiglitz and Weiss (1981), Almeida and Campello (2001), and Whited and Wu (2006)). Under the latter definition, constrained firms should not conduct any equity recycling. See Farre-Mensa and Ljungqvist (2016) for more discussion.

TABLE 6

Equity Recycling by Financial Constraint Classification

Table 6 examines the difference in equity recycling behavior between the most equity-constrained firms (top 20%) and the least equity-constrained firms (bottom 20%). We follow the procedure of Farre-Mensa and Ljungqvist (2016). We regress the yearly change ($t-1 \rightarrow t$) in payouts to shareholders on the yearly change in equity issuance proceeds (ΔEQUITY_ISSUANCE). We control for the change in other funding sources (ΔOTHER_FUNDING) and change in firm size (ΔSIZE). All variables are scaled by the beginning-of-year ($t-1$) total assets except size. We include industry-by-year fixed effects. In the first 6 columns, we present results for the main random forest measure in different time periods. In columns 7 and 8, we present the results for the Hoberg and Maksimovic (2015) equity constraint measure. In the last 2 columns, we present results for the "Primitive" model for the full sample period. Standard errors are clustered at the firm level. We report the results of a Wald test comparing the coefficient of interest for constrained and unconstrained firms. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	1	2	3	4	5	6	7	8	9	10
ΔEQUITY_ISSUANCE	0.002***	0.024***	0.005***	0.039***	0.001***	0.022***	0.003***	0.012***	0.001***	0.023***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ΔOTHER_FUNDING	0.000	0.013***	0.003**	0.015***	−0.000	0.012***	0.000	0.006***	0.000	0.012***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ΔSIZE	0.001*	−0.018***	0.001	−0.014***	0.001	−0.020***	0.000	−0.008***	0.000	−0.013***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constraint group:	Top 20%	Bottom 20%	Top 20%	Bottom 20%	Top 20%	Bottom 20%	Top 20%	Bottom 20%	Top 20%	Bottom 20%
Time period	1972–2021		1972–1996		1997–2021		1997–2015		1972–2021	
Constraint measure:	RF full		RF full		RF full		HM		RF primitive	
Wald test:	34.37***		14.08***		25.11***		5.54**		39.36***	
No. of obs.	19,823	21,749	4,650	5,790	15,173	15,959	6,673	7,539	19,443	20,569
R^2	0.028	0.042	0.034	0.041	0.027	0.043	0.041	0.049	0.024	0.042

on this test, our equity constraint classifications seem to capture variation in equity constraints fairly well.

We provide the results for the HM (2015) classifications in the next 2 columns. Using their classifications, we find there is a difference in equity recycling behavior across constrained and unconstrained firms. The difference in coefficients is not as large as using our measure in the post-1997 time period. There are two potential explanations for why there is greater variation in equity recycling behavior using the random forest classifications than the HM (2015) classifications: expanded coverage and noise reduction. By mapping back to firm fundamentals, we may be providing a closer approximation to true constraints.[11] These results suggest that econometricians may be better able to understand the behavior of constrained firms when using our measures than when using existing constraint indices.

Finally, in the last 2 columns, we present results using the "Primitive" constraint measures for the full sample-period. The results are consistent with the full-model findings. All coefficient estimates are very similar and achieve similarly high levels of significance. This suggests that constraint classifications using only primitive predictors align with firm behavior similarly to classifications that include potentially endogenously determined predictors. This helps alleviate concerns that endogenously determined predictors are driving the results of our analysis by capturing something other than firm constraints.

[11]We find greater variation in equity recycling behavior using the random forest classifications on only the firm-years in the HM sample (Wald of 12.57 using RF measures vs. Wald of 5.44 using the HM measures, and a greater difference in coefficients between constrained and unconstrained). This suggests that both noise reduction and expanded coverage contribute to the larger variation in equity recycling behavior when using random forest classifications.

We further assess the robustness of these results by using forward-looking equity constraint classifications and by examining only share repurchases as the measure of equity holder payouts to gauge robustness (following Farre-Mensa and Ljungqvist (2016)). Results are presented in the Supplementary Material. We find a very similar pattern as in the baseline tests: more constrained firms conduct significantly less equity recycling than less constrained firms. The differences in coefficients are economically and statistically significant in all specifications using the random forest classifications. We also show in the Supplementary Material that similar results obtain for the equity recycling tests if we use the modified version of the random forest model that does not include dividends as a predictor. These results help ease any concern that our tests are mechanical due to dividends being used to construct the dependent variable and constraint measure. The performance of the random forest-based measures on the equity recycling tests gives further confidence that our equity constraint classifications are capturing important variation in equity financing constraints. This stands in stark contrast to the poor performance of the well-known accounting-based constraint measures on similar equity recycling tests (Bodnaruk et al. (2015), Farre-Mensa and Ljungqvist (2016)).

Next, we examine the performance of the equity constraint measures on two dividend-policy tests from Bodnaruk et al. (2015). These tests examine dividend omissions and dividend increases across constrained and unconstrained firms. We continue to focus on equity constraints as these tests are related to behavior surrounding equity payouts.[12] Since it is especially costly for constrained firms to access external capital, they are more likely to omit paying a dividend to shareholders than less constrained firms with better access to external markets. Similarly, unconstrained firms are more likely to increase dividend payments than are constrained firms.

The tests are performed by regressing a dummy variable for a dividend omission or a dividend increase on a constrained firm dummy variable. A firm-year observation has a dividend omission dummy equal to 1 if the firm paid dividends in the previous year, but elected not to pay a dividend in the current year. The dummy takes a value of 0 if the firm paid dividends in the previous year and still pays dividends in the current year. Thus, only firms which paid dividends in the previous year are included in the sample. Similarly, the dividend increase dummy is equal to 1 if a firm paid positive dividends in the previous year and increased dividend payouts in the current year. The dummy takes a value of 0 if the firm paid dividends in the previous year, but does not increase its dividends in the current year. The explanatory variable of interest is a dummy equal to 1 if the firm is in the top 20% of constraints in a given year (i.e., most constrained) and 0 if in the bottom 20% of constraints (i.e., least constrained). We control for lagged market capitalization (log(MKT_CAP)), lagged book to market (log(BM)), a negative earnings dummy (NEG_EARNINGS_DUMMY), and lagged returns in excess of the CRSP value-weighted market return (PAST_EXCESS_RETURN). In all specifications, we include year and industry fixed effects where industries are defined according to

[12]For completeness, we conduct similar tests (equity recycling, dividend omissions, and dividend increases) for the debt constraint classifications even though the predicted relationship between debt constraints and equity payout behavior is unclear. Results are presented in the Supplementary Material.

TABLE 7

Dividend Tests

Table 7 examines the difference in dividend payment behavior between the most equity-constrained firms (top 20%) and the least equity-constrained firms (bottom 20%). In Panel A, the dependent variable is a dividend omission dummy equal to 1 if the firm did not pay a dividend during the year and paid a dividend the previous year ($t-1 \rightarrow t$). In Panel B, the dependent variable is a dividend increase dummy equal to 1 if a firm increased its dividend between the previous and current year ($t-1 \rightarrow t$). The main independent variable of interest (CONSTRAINED_DUMMY) is a dummy variable equal to 1 (0) if the firm is in the most (least) constrained quintile in year $t-1$. We control for the logarithm of market capitalization (year $t-1$), logarithm of book-to-market (year $t-1$) winsorized at the 1% level, a negative earnings dummy (year $t-1$) and the firm's equity return in excess of the market in the previous year ($t-2 \rightarrow t-1$). We only include firm-year observations in which the firm paid a dividend in year $t-1$. We include industry and year fixed effects. We identify the model and time period at the top of each column. Standard errors are clustered at the industry and year levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	1	2	3	4	5
Panel A. Dividend Omissions					
CONSTRAINED_DUMMY	0.044***	0.036***	0.060***	0.045***	0.032***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log(MKT_CAP)	−0.014***	−0.014***	−0.015***	−0.014***	−0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(BM)	−0.002	−0.009	0.004	0.006	−0.003
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
NEG_EARNINGS_DUMMY	0.146***	0.155***	0.130***	0.117***	0.132***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
PAST_EXCESS_RETURN	0.002	−0.005	0.008	0.014**	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Time period	1972–2021	1972–1996	1997–2021	1997–2015	1972–2021
Constraint measure	RF full	RF full	RF full	HM	RF primitive
No. of obs.	26,901	15,171	11,730	5,504	27,790
R^2	0.108	0.101	0.112	0.101	0.106
Panel B. Dividend Increases					
CONSTRAINED_DUMMY	−0.116***	−0.143***	−0.056*	−0.048	−0.054***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)
log(MKT_CAP)	0.048***	0.058***	0.039***	0.037***	0.046***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(BM)	−0.047***	−0.076***	−0.021**	−0.010	−0.040***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
NEG_EARNINGS_DUMMY	−0.222***	−0.272***	−0.198***	−0.167***	−0.236***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
PAST_EXCESS_RETURN	0.047***	0.079***	0.021	0.011	0.042***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Time period	1972–2021	1972–1996	1997–2021	1997–2015	1972–2021
Constraint measure	RF full	RF full	RF full	HM	RF primitive
No. of obs.	26,904	15,171	11,733	5,506	27,801
R^2	0.159	0.192	0.140	0.140	0.157

the 48 Fama–French industry classifications. Standard errors are clustered by year and industry.[13]

Panel A of Table 7 presents the results for the dividend omission test. Column 1 shows the results using the entire sample of our full-model random forest classification. As predicted, there is a strong positive relation between the constraint dummy and the dividend omission dummy indicating constrained firms are more likely to cut dividends to 0. The coefficient is highly significant (p-value < 0.01) and economically meaningful with a point estimate of 0.044. In columns 2 and 3, we

[13]The industry classifications and the fixed effects are different for the equity recycling tests and the remaining tests to be perfectly consistent with the methodologies of Bodnaruk et al. (2015) and Farre-Mensa and Ljungqvist (2016), respectively.

show that the coefficient is positive and significant in both the pre-1997 and post-1997 periods. Using the HM measures, we also find a positive and significant coefficient. Column 5 shows the result using the "Primitive" random forest model estimates for constraints. The coefficient for the constraint dummy of 0.032 is similar, though slightly smaller in magnitude, to the coefficient estimate when using the full model and is again highly significant. These results strongly suggest our equity constraint classifications are capturing firm-level financial constraints.

Panel B of Table 7 shows results for the dividend increase test. We find constrained firms are less likely to increase their dividends with a negative coefficient on the constrained dummy across all specifications. The coefficients are relatively less consistent across time periods with a stronger relationship between constraints and dividend increases in the pre-1997 period than in the post-1997 period. The full-model random forest constraint dummy for the post-1997 sample and the HM constraint dummy have similar point estimates although only the random forest constraint dummy is significant at the 10%-level. This highlights the additional power provided by expanding coverage using our methodology. The coefficient on the constrained dummy for the "Primitive" model is smaller in magnitude than for the full model but is still significant at the 1% level.

One potential concern with the dividend-related tests is that the results may be mechanical given that dividends are a predictor in our main random forest model. The fact that we find a similar pattern using the more primitive model helps to address that concern. In addition, we show in the Supplementary Material that we obtain similar results for the dividend increase and dividend omission tests when using a modified version of the constraints model that does not include dividends (DVC and DVP) as predictors. Overall, the results of the dividend increase and dividend omission tests indicate that our constraint measure is likely to identify constrained firms well.

The final test of the financial constraint classifications from Bodnaruk et al. (2015) examines the funding of pensions by constrained and unconstrained firms. Motivation for this test comes from Rauh (2006), who shows mandatory pension obligations are negatively related to firm investment. Bodnaruk et al. (2015) show that financially constrained firms are more likely to underfund pensions. We use the same method as Bodnaruk et al. (2015) to examine whether firms we classify as debt-constrained are more likely to underfund their pensions. We focus on debt constraints for two reasons. First, as discussed in Matsa (2010) and Chava et al. (2020), pension plans are similar to leverage in a firm's capital structure since pension plans specify a fixed set of payments firms commit to make periodically over time. Second, very few of our firms classified as equity-constrained have pensions at all. Given that a number of firms do not have pensions, we classify the top 30% and bottom 30% of firms as the most and least constrained, respectively to increase the sample size of our tests. This is especially important in the pre-1997 period, where even with the broader definition there are only 4,629 firm-year observations.

Results testing for the relationship between debt constraints and pension underfunding are presented in Table 8. We find that using the full model and the entire sample period, there is a positive and significant (p-value < 0.01) relationship between debt constraints and pension underfunding. Examining the two subperiods

TABLE 8

Pension Underfunding Tests

Table 8 examines the difference in pension funding behavior between more debt-constrained firms (top 30%) and less debt-constrained firms (bottom 30%). The dependent variable is a pension underfunded dummy equal to 1 if a firm's pension is underfunded in year t. The main independent variable of interest (CONSTRAINED_DUMMY) is a dummy variable equal to 1 (0) if the firm is in the most (least) constrained 30% of firms in year $t-1$. We control for the lagged dependent variable (year $t-1$). We control for the logarithm of market capitalization (year $t-1$), logarithm of book-to-market (year $t-1$) winsorized at the 1% level, a negative earnings dummy (year $t-1$) and the firm's equity return in excess of the market in the previous year ($t-2 \rightarrow t-1$). We only include firm-year observations in which the firm had pension obligations in the year $t-1$. We include industry and year fixed effects. We identify the model and time period at the top of each column. Standard errors are clustered at the industry and year levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	1	2	3	4	5
CONSTRAINED_DUMMY	0.020***	0.025	0.020**	0.003	0.011**
	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
LAG_UNDERFUND_DUMMY	0.591***	0.596***	0.579***	0.530***	0.590***
	(0.03)	(0.02)	(0.04)	(0.05)	(0.03)
log(MKT_CAP)	−0.003*	−0.010*	−0.001	−0.001	−0.004**
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
log(BM)	−0.009**	−0.042***	0.001	−0.002	−0.012**
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
NEG_EARNINGS_DUMMY	−0.007	−0.020	−0.002	−0.002	−0.002
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
PAST_EXCESS_RETURN	−0.001	0.003	−0.002	−0.003	0.001
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Time period	1972–2021	1972–1996	1997–2021	1997–2015	1972–2021
Constraint measure	RF full	RF full	RF full	HM	RF primitive
No. of obs.	16,820	4,629	12,191	6,870	18,419
R^2	0.620	0.401	0.493	0.483	0.625

we see the coefficients are similar. The constrained dummy coefficient is 0.025 in the pre-1997 period and 0.020 in the post-1997 period. Only the post-1997 period estimate is significant at the 5% significance level, however. The coefficient on the HM constraint dummy is positive and insignificant, suggesting the increased sample size is important for increasing statistical precision. The constrained dummy coefficient when using the "Primitive" model is smaller (point estimate of 0.011) but is significant at the 5% level. The results provide suggestive evidence the random forest debt constraint classifications properly classify firms' debt constraints.

Overall, the performance of our constraint measures on this battery of tests provides evidence that firms we classify as more constrained actually are more constrained.

V. Equity Constraints, Sentiment, and the Presence of Institutional and Retail Investors

In Section V, we use our new equity constraint measures to examine how equity constraints are related to institutional and retail ownership, and to examine how changes in investor sentiment are related to relative financing and investment activity of equity-constrained firms.

A. Institutional Ownership of Constrained Firms

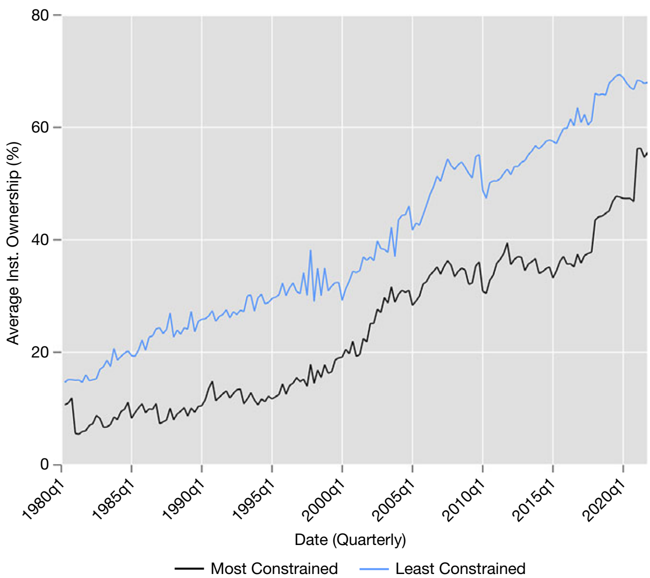
Firm managers are known to go to great lengths to court and sustain investment by institutions who control large sums of capital. The presence of institutional owners can help ease financial constraints by reducing asymmetries in information between equity investors and firm managers. This occurs through two main channels. First, as shown in Boehmer and Kelley (2009), stocks with greater institutional ownership are priced more efficiently. Investors can have some confidence there is a lower degree of mispricing and companies have more accurately valued equity with more institutional ownership. Second, institutional shareholders improve corporate governance as discussed in Carleton et al. (1998), Appel et al. (2016), and McCahery et al. (2016). They govern through direct intervention when large shareholders directly engage with firm managers to make changes or demands. They may also exert influence through the threat of selling their shares when firm management performs poorly, which serves as an indirect form of governance (Admati and Pfleiderer (2009)). The improved governance and reduction in asymmetric information associated with institutional ownership should relax firm's financing constraints. For these reasons, we hypothesize that institutional ownership is negatively related to the constraints of equity-focused firms.

Before running our statistical tests, we plot the time series of average institutional ownership of the most (top 20%) and least (bottom 20%) equity-constrained firms. In Figure 6, we plot the time series of the average percent of shares held by institutions among the most constrained equity-focused stocks and least constrained stocks each year-quarter from 1980, the beginning of the Thomson Reuters Institutional Holdings database, to 2021. Throughout the entire time period, the least constrained stocks have significantly higher average institutional ownership than the most equity-constrained stocks. In the Supplementary Material, we show a

FIGURE 6

Institutional Ownership and Equity Constraints Over Time

Figure 6 shows the average institutional ownership (percentage of shares held by institutions) for firms in the top and bottom quintiles of equity constraints. Each year, firms are sorted by their previous year equity constraint (RF_EQUITY) quintile.



https://doi.org/10.1017/S0022109023000650 Published online by Cambridge University Press

TABLE 9

Institutional Ownership and Constraints of Equity-Focused Firms

Table 9 examines the relationship between institutional ownership and equity constraints. We regress institutional ownership (INSTITUTIONAL_OWNERSHIP), defined as the percentage of shares held by institutions, on dummies for the lagged quintiles of equity constraints (LAG_CONSTRAINTS_Q). We include the logarithm of market capitalization (log(MKT_CAP)) as a control variable in columns 2 and 3. In columns 1 and 2, we use constraint estimates from the full random forest model for equity-focused constraints. In column 3, we use constraint estimates from the "Primitive" random forest model for equity-focused constraints. We include year-quarter fixed effects in all specifications. Standard errors are clustered at the year and firm-level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	1	2	3
LAG_CONSTRAINTS_Q2	0.906*	−2.411***	−1.721***
	(0.51)	(0.44)	(0.42)
LAG_CONSTRAINTS_Q3	−1.163*	−5.320***	−3.014***
	(0.62)	(0.55)	(0.49)
LAG_CONSTRAINTS_Q4	−6.643***	−8.092***	−4.787***
	(0.71)	(0.60)	(0.54)
LAG_CONSTRAINTS_Q5	−15.986***	−12.347***	−8.485***
	(0.75)	(0.62)	(0.63)
log(MKT_CAP)		7.178***	7.229***
		(0.22)	(0.22)
Constraint measure	RF full	RF full	RF primitive
No. of obs.	456,825	456,825	456,825
R^2	0.193	0.338	0.330

similar pattern using the "Primitive"-model or HM constraint measures. These time-series results are consistent with institutional ownership being negatively related with firm's equity constraints.

Next, we regress the institutional ownership of each firm on lagged equity constraint quintile dummies. We define institutional ownership as the percent of shares held by institutional investors for a given firm in a given quarter. Column 1 in Table 9 reports the results of regressing institutional ownership on indicators for each quintile of equity-focused constraints from the fully specified random forest. Indicators for the two most constrained quintiles are significantly negative at the 1% level suggesting that institutional ownership is significantly lower among these firms than in the least constrained quintile. Institutional ownership is 15.986 percentage points lower in the most constrained quintile than the least constrained quintile on average. In column 2, we control for firm size as institutions are known to strongly prefer investing in large firms and we want to ensure that we are not just capturing the preference for larger companies. Even after controlling for size, we see a strong negative relationship between constraints and institutional ownership. In column 3, we regress institutional holdings on equity constraint quintile dummies as defined by the "Primitive" random forest model, controlling for firm size. The results are similar to those using the full random forest model. Controlling for firm size, the institutional ownership of firms in the most constrained quintile is 8.485 percentage points lower than that of firms in the least constrained quintile. Overall, the results show that institutions tend to invest significantly less in constrained firms even after controlling for firm size.

B. Retail Investors and Equity Constraints

The results above indicate that institutional investors are a smaller proportion of the investor base for the relatively more constrained equity-focused stocks.

We next examine whether this is also true for retail investors.[14] For this analysis, we use the retail brokerage database of Barber and Odean (2000) with data from 1991:Q2 to 1996:Q3 and the Robinhood investor position data from Robintrack.net from May 2, 2018 to Aug. 13, 2020. Each of these data sets represents a different, small subset of retail investors over different, relatively short, time periods.

Using the brokerage data from the early 1990s, we create a retail stock ownership measure that is the total dollar amount invested in each stock across all accounts at the end of each quarter divided by the stock's market capitalization. This is the retail equivalent of the institutional investor ownership variable. Due to the small scale of the brokerage data, we standardize the ownership measure each quarter to ease interpretation.

We plot the average retail ownership measure for the most equity-constrained and least equity-constrained quintiles each quarter in Graph A of Figure 7. In contrast to institutional investors, we see that throughout the relatively short sample period, retail investors have a greater presence in the most equity-constrained firms relative to the least equity-constrained firms.

We next examine whether a similar pattern of behavior is present among Robinhood investors. We cannot calculate an ownership percentage variable within this sample as Robintrack does not provide the dollar amount invested in each stock. Instead, Robintrack reports the number of Robinhood users invested in each stock, each day. We calculate an ownership measure as follows: $\frac{\sum_{i \in j} N_i}{\sum N_i}$, where N_i is the number of account holders invested in stock i and j is a constraint quintile grouping. The measure is a transformation of breadth and is based on the ARH measure of Welch (2022).

Graph B of Figure 7 presents the average of the Robinhood ownership measure across the most and least equity-constrained quintiles as defined by the full random forest model. Similar to the 1990s discount brokerage investors, Robinhood investors display a relative preference for the *most* equity-constrained firms compared to the least constrained firms throughout the entire sample period with total ownership about three to four times as large for the most constrained as the least constrained firms. Overall, the results suggest retail investors are more willing to invest in the more equity-constrained firms that are out of favor with institutional investors.

Robinhood investors do shift out of the more equity-constrained stocks in the months following the start of the COVID-19 pandemic (we mark the WHO declaration date of Mar. 13, 2020 on the figure) and invest slightly more in the least constrained stocks. Though, a large gap in holdings remains between the most and least equity-constrained firms.

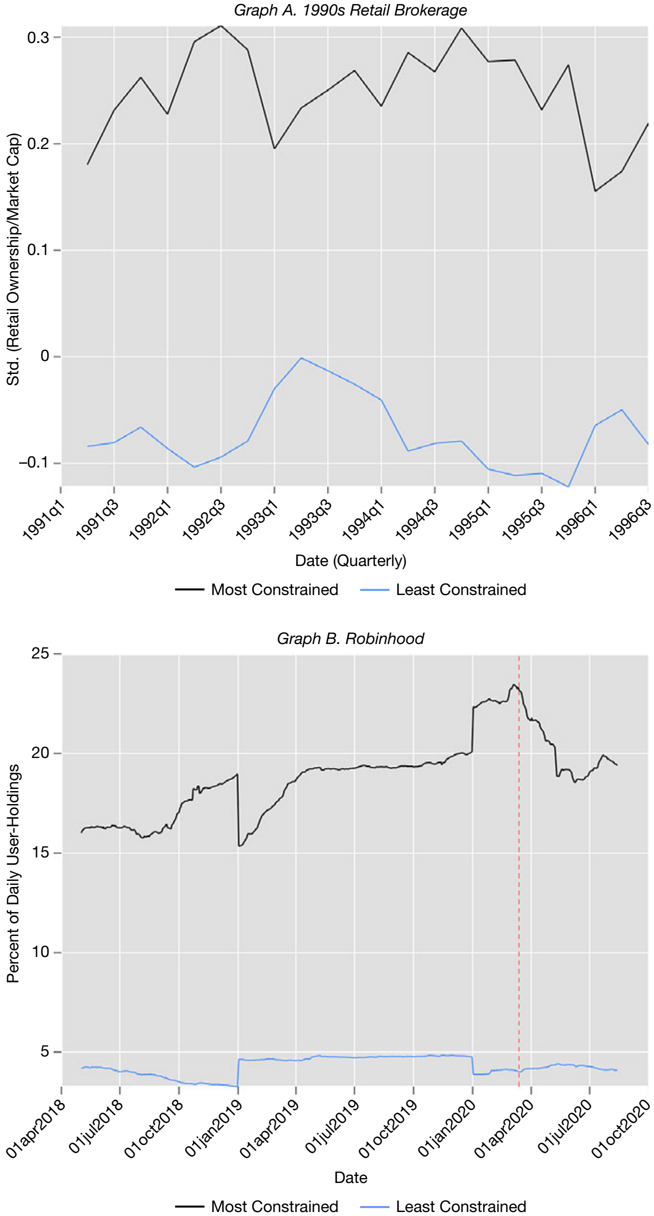
During the pandemic, there was significant media coverage of retail investors' trading behavior. One of the main themes was how Robinhood investors were

[14]We cannot just assume that retail investors are holding the remaining shares not held by institutions and, therefore, should exhibit an opposing relationship between constraints and investor presence. The institutional ownership data we use comes from 13F filings which are only required of institutions with more than $100 million in assets under management, leaving small institutions unrepresented in the data. Furthermore, some foreign institutions with US stock holdings are not required to report holdings to the Securities and Exchange Commission.

FIGURE 7

Retail Investor Holdings and Equity Constraints Over Time

Figure 7 displays the relationship between retail investor stock ownership and equity constraints. Graph A shows the percent of stock ownership by investors who are clients of a retail brokerage in the period of 1991:Q2 to 1996:Q3. Ownership is calculated as total dollar amount invested across all accounts divided by the market capitalization of the stock. The measure is standardized within each quarter. Graph B shows the average percent of daily Robinhood investor holdings that are in the top quintile of equity constraints (most constrained) or bottom quintile of equity constraints (least constrained). Constraints are based on the previous year equity constraint (RF_EQUITY) quintile.



https://doi.org/10.1017/S0022109023000650 Published online by Cambridge University Press

purchasing stocks that were out of favor with institutional investors (e.g., Hertz stock during bankruptcy proceedings or Gamestop stock). Our results suggest that this is a more widespread phenomenon – retail investors tend to target these out-of-favor companies that are facing equity constraints. This finding is broadly consistent with the findings of Farrell, Green, Jame, and Markov (2022), who find that finance social media coverage, which caters to retail investors, tends to focus on stocks with lower institutional holdings.

C. Sentiment, Constraints, and Firm Investment

In our final set of tests, we examine whether financially constrained firms' equity issuance and investment are more sensitive to investor sentiment. Financially constrained firms face external costs of capital that are high enough to prevent them from undertaking positive net present value projects and this stifling of investment can have a significant impact on the economy (Bernanke, Gertler, and Gilchrist (1996), Kiyotaki and Moore (1997)). While most of the macroeconomic literature has focused on the effects of macro shocks on firms with borrowing constraints, we examine whether variation in market sentiment is related to the real outcomes of constrained, equity-focused firms.

We test whether the most constrained equity-focused firms increase their investment and equity offering activities *more* than unconstrained firms during periods of heightened investor sentiment. Investor sentiment is likely to have a greater effect on the equity prices of firms that are difficult to value (Baker and Wurgler (2006)). Equity-constrained firms, who are characterized by higher levels of information asymmetry, are more difficult to value and should, therefore, be more sensitive to investor sentiment. The loosening of constraints for the more constrained firms should lead these firms to issue relatively more equity and invest more as their cost of external capital decreases. The equity issuance and investment of unconstrained firms, on the other hand, should be less sensitive to shifts in investor sentiment.

We begin our analysis by regressing firm equity issuance scaled by lagged total assets on market sentiment and indicators for each quintile of constraints from the full random forest model. Firms are sorted into quintiles within each year and industry. The coefficients of interest are interactions of sentiment and the equity constraint quintiles. A positive coefficient on the interaction term for a particular quintile means firms within that quintile tend to increase their equity issuance (increase investment) more during periods of high sentiment than firms estimated to be in the least-constrained quintile. We include firm fixed effects to absorb time-invariant differences across firms, and industry × year-quarter fixed effects to account for differential shocks across industries over time. We also control for lagged debt to assets, lagged cash to assets, lagged firm size (log of market capitalization), and lagged book-to-market, all of which are also interacted with sentiment. For brevity, the point estimates and standard errors for these controls are not reported. Reported standard errors are clustered at the year and firm level.

Results are reported in Table 10. In Panel A, we use the standardized, orthogonalized measure from Baker and Wurgler (2006) as our sentiment proxy. Each row shows the interaction effects for a given equity constraint quintile with sentiment.

https://doi.org/10.1017/S0022109023000650 Published online by Cambridge University Press

TABLE 10

Sentiment and Real Outcomes for Equity-Focused Constrained Firms

Table 10 examines the relationship between firm real outcomes, sentiment, and constraints of equity-focused firms. We regress the firm outcome of interest on an interaction between the investor sentiment measure (SENTIMENT) and dummies for the lagged quintiles of equity constraints (LAG_CONSTRAINTS_Q). Quintiles are formed within industry-year. In Panel A, the sentiment measure is from Baker and Wurgler (2006). In Panel B, the sentiment measure is an adjusted version of the Baker and Wurgler (2006) constructed from three components: closed-end fund discount, dividend premium, and equity share in new issues. We standardize all sentiment measures to mean of 0 and standard deviation of 1. We include the quintile dummies independently as well (coefficients not reported for brevity). The outcomes of interest are equity issuance to lagged assets (EQUITY_TO_ASSETS) in columns 1–3, and capital expenditures to lagged property, plant, and equipment (CAPX_TO_K) in columns 4–6. In columns 1 and 4, we do not include controls. In the remaining regressions, we include the lagged logarithm of the firm's market capitalization, the lagged book-to-market, the lagged cash-to-assets ratio, and the lagged debt-to-assets ratio and each control variables' interaction with sentiment (coefficients not reported). We denote at the top of each column the model used to estimate equity constraints (either the full model or the model using "primitive" predictors). Both firm fixed effects and industry × year-quarter fixed effects are included in the regressions. The main effect of sentiment is absorbed by the fixed effects. Standard errors are clustered at the year and firm levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	EQUITY_TO_ASSETS			CAPX_TO_K		
	1	2	3	4	5	6
Panel A. Sentiment						
LAG_CONSTRAINTS_Q2 × SENTIMENT	0.003**	0.003**	0.002	0.006**	0.001	0.004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAG_CONSTRAINTS_Q3 × SENTIMENT	0.008**	0.007***	0.016**	0.013***	0.001	0.006***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
LAG_CONSTRAINTS_Q4 × SENTIMENT	0.025**	0.022***	0.025***	0.023***	0.008**	0.022***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
LAG_CONSTRAINTS_Q5 × SENTIMENT	0.052***	0.044***	0.031***	0.054***	0.028***	0.029***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Constraint measure	RF full	RF full	RF primitive	RF full	RF full	RF primitive
Controls?	No	Yes	Yes	No	Yes	Yes
No. of obs.	201,889	134,337	134,337	202,994	135,830	135,830
R^2	0.389	0.369	0.368	0.324	0.403	0.403
Panel B. Adjusted Sentiment						
LAG_CONSTRAINTS_Q2 × SENTIMENT	0.002	0.002	0.000	−0.000	−0.002	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAG_CONSTRAINTS_Q3 × SENTIMENT	0.005**	0.003*	0.008**	0.004	−0.002	0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAG_CONSTRAINTS_Q4 × SENTIMENT	0.018***	0.007**	0.009**	0.005	−0.003	0.006*
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAG_CONSTRAINTS_Q5 × SENTIMENT	0.038***	0.021***	0.015***	0.022***	0.011**	0.014***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
Constraint model	RF full	RF full	RF primitive	RF full	RF full	RF primitive
Controls?	No	Yes	Yes	No	Yes	Yes
No. of obs.	201,889	134,337	134,337	202,994	135,830	135,830
R^2	0.389	0.368	0.367	0.323	0.399	0.400

Columns 1–3 show the results when equity issuance is the dependent variable. We find a clear, monotonically increasing relationship between equity issuance and constraints interacted with sentiment. We find similar results whether we do not include controls (column 1), include controls (column 2) or use the "Primitive" model for estimating constraints (column 3). Across all three specifications, the interaction with the most constrained firms (quintile 5) is significant at the 1% level. These results indicate that constrained firms issue significantly more new equity than the least constrained firms as sentiment increases.

In columns 4–6, we run analogous regressions examining how firm-level investment is related to investor sentiment. The dependent variable is capital expenditures deflated by lagged property, plant, and equipment. We find a similar pattern as with equity issuance. Across all three specifications, the more equity-constrained a firm, the more sensitive its investment activity is to changes in sentiment. The pattern is monotonically increasing and is highly significant in each column. The results of Panel A suggest that as sentiment increases, equity-focused constrained firms issue more equity and invest more than less constrained firms.

Since the Baker and Wurgler (2006) sentiment index is arguably the most widely used proxy for market-wide sentiment in the literature, we use it as our first proxy. However, measuring sentiment is an imprecise science and it is difficult to attribute all variation in any index to sentiment as opposed to other economic factors. One potential concern is that sentiment is correlated with aggregate investment opportunities or growth options. When investment opportunities are most favorable, equity-focused, constrained firms may be more likely to see their equity values increase, with more demand for growth and innovation, and this may allow them to issue equity at more favorable prices and invest more.

We take a number of steps to address this concern. First, our controls in Table 10 include the ratio of the firms' book-to-market equity values and its interaction with sentiment in order to control for firm investment opportunities. Second, we use industry-time fixed effects and sort firms into quintiles of constraints within-industry, which will absorb time-variation in investment opportunities across industries and ensure we are identifying off within-industry differences in constraints. Third, we use the version of the Baker and Wurgler (2006) index that is orthogonalized with respect to six macroeconomic variables. This should help to purge the index of rational optimism based upon economic information.

However, there may still remain concerns that two of the components of the sentiment index are IPO first-day returns and IPO volume (among other things), which should be related to aggregate investment opportunities. In order to address this potential concern, we reconstruct the Baker and Wurgler index without including the aggregate measures of IPO volume and first-day IPO returns. The index, which we call "Adjusted Sentiment," uses factor analysis to extract the first principal component of the remaining proxies used in the Baker and Wurgler (2006) index: value-weighted dividend premium (pdnd), closed-end fund discount (cefd), and the equity share in new issues (s). We then orthogonalize this with respect to the same macroeconomic variables as the original index uses.

Panel B of Table 10 shows results when using the Adjusted Sentiment proxy. The main coefficients of interest, interactions between constraint quintiles and sentiment, remain mostly monotonic in constraint levels. We find 5 of the 6 tests show the most constrained quintile has a coefficient that is significant at the 1% level. For the full-model, capital expenditure column with controls, the coefficient is significant at the 5% level. The results suggest that periods of high sentiment are associated with a relative easing of constraints for the most constrained firms.

In the Supplementary Material, we conduct further tests to address this concern. We run the same set of tests with each of the individual proxies for sentiment used in constructing the Baker and Wurgler (2006) index to see which components are driving the relationships observed in Table 10. We find all of the individual components of the Baker and Wurgler index tend to exhibit a monotonic and significant relationship *except* IPO returns – the component most likely to be correlated with aggregate investment opportunities. We also run the tests using the University of Michigan Consumer Sentiment Index, which is derived from survey data as opposed to market-based measures of mispricing from which the Baker and Wurgler index is derived. Qiu and Welch (2004) find that the University of Michigan Index is a reasonable measure of investor sentiment, however, for our particular application, mispricing realizations may be a better measure of the sentiment firm managers base their decisions on. Using the University of Michigan Index, we find similar monotonic relationships between the constraints and sentiment interactions for both equity issuance and capital expenditures, yet the statistical significance is less pronounced. Interestingly, when using the University of Michigan index, the "Primitive" constraint quintiles show higher levels of statistical significance than those formed using the full set of predictors. While we cannot completely rule out alternative hypotheses, the body of the evidence suggests that the financing and investment activity of equity-focused constrained firms are sensitive to changes in investor sentiment.

These results are related to the results of Baker and Wurgler (2006), who show that in periods of high sentiment, small stocks, young stocks, high volatility stocks, unprofitable stocks, nondividend-paying stocks, and extreme growth stocks earn lower returns (i.e., lower costs of equity capital). While their results are focused on stock returns and many characteristics that are often associated with constrained companies, our focus is on real firm outcomes.[15] The results are also related to HM (2015), who show that constrained firms do relatively poorly during times of extreme stress in financial markets. Our results show that equity-constrained firms' financing activities are generally more sensitive to investor sentiment, which does not necessarily derive from macroeconomic fundamentals.

VI. Conclusion

We propose a novel method for estimating firms' financial constraints over a large cross section and time series using regression random forests and accounting variables. Our estimates capture similar information to the text-based estimates of financial constraints of Hoberg and Maksimovic (2015), yet we are able to classify over 165,000 additional firm years. The improvement in coverage is greatest in the time series in which we more than double the number of years in which firms can be classified according to their financial constraints. We cover the vast majority of firm years between 1972 and 2021. This expanded coverage offers the potential to significantly improve empirical research on financial constraints.

We provide two separate estimates of constraints for both debt- and equity-focused firm types. The first uses a large set of predictors aggregated from well-known constraint estimators from the literature. The second relies only on the subset of these primitive predictors that are unlikely to be endogenously determined by the firm.

[15]It has been shown that financial constraints are not always reflected in the stock returns associated with constrained firms. For example, Lamont, Polk, and Saaá-Requejo (2001) show that constrained firms do not earn a premium even though theory and empirical evidence suggests that they are sensitive to aggregate economy wide risk.

We show that both versions of our model of financial constraints perform well out of sample. Additionally, we provide significant evidence that firms we classify as constrained behave in ways that indicate the constraint classifications are actually capturing firms' financial constraints.

We use our constraint classifications to uncover novel details about the relationship between financial constraints, investor behavior, and firm outcomes. We find firms classified as equity-focused and constrained have a much lower presence of institutional investors compared to unconstrained firms. In contrast, we find a large subset of retail investors, investors using a discount brokerage in the 1990s and those using the Robinhood platform in 2018–2020, appear to favor constrained equity-focused firms relative to unconstrained firms. This is consistent with the notion that retail investors target stocks which are out of favor with institutional investors. Finally, we show equity-focused, constrained firms' financing and investment timing are more sensitive to market-wide measures of sentiment than unconstrained firms. These results provide novel insights into the relationship between investors, equity-related constraints, and real outcomes.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109023000650.

References

Admati, A. R., and P. Pfleiderer. "The "Wall Street Walk" and Shareholder Activism: Exit as a Form of Voice." *Review of Financial Studies*, 22 (2009), 2645–2685.

Almeida, H., and M. Campello. "Financial Constraints and Investment-Cash Flow Sensitivities: New Research Directions." Twelfth Annual Utah Winter Finance Conference (2001).

Appel, I. R.; T. A. Gormley; and D. B. Keim. "Passive Investors, Not Passive Owners." *Journal of Financial Economics*, 121 (2016), 111–141.

Baker, M., and J. Wurgler. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance*, 61 (2006), 1645–1680.

Barber, B. M., and T. Odean. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance*, 55 (2000), 773–806.

Bernanke, B. S.; M. Gertler; and S. Gilchrist. "The Flight to Quality and the Financial Accelerator." *Review of Economics and Statistics*, 78 (1996), 1–15.

Bodnaruk, A.; T. Loughran; and B. McDonald. "Using 10-K Text to Gauge Financial Constraints." *Journal of Financial and Quantitative Analysis*, 50 (2015), 623–646.

Boehmer, E., and E. K. Kelley. "Institutional Investors and the Informational Efficiency of Prices." *Review of Financial Studies*, 22 (2009), 3563–3594.

Buehlmaier, M. M., and T. M. Whited. "Are Financial Constraints Priced? Evidence from Textual Analysis." *Review of Financial Studies*, 31 (2018), 2693–2728.

Carleton, W. T.; J. M. Nelson; and M. S. Weisbach. "The Influence of Institutions on Corporate Governance through Private Negotiations: Evidence from TIAA-CREF." *Journal of Finance*, 53 (1998), 1335–1362.

Chava, S.; A. Danis; and A. Hsu. "The Economic Impact of Right-to-Work Laws: Evidence from Collective Bargaining Agreements and Corporate Policies." *Journal of Financial Economics*, 137 (2020), 451–469.

Farrell, M.; T. C. Green; R. Jame; and S. Markov. "The Democratization of Investment Research and the Informativeness of Retail Investor Trading." *Journal of Financial Economics*, 145 (2022), 616–641.

Farre-Mensa, J., and A. Ljungqvist. "Do Measures of Financial Constraints Measure Financial Constraints?" *Review of Financial Studies*, 29 (2016), 271–308.

Fazzari, S. M.; R. G. Hubbard; and B. C. Petersen. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity*, 1988 (1988), 141–195.

Freyberger, J.; A. Neuhierl; and M. Weber. "Dissecting Characteristics Nonparametrically." *Review of Financial Studies*, 33 (2020), 2326–2377.
Hadlock, C. J., and J. R. Pierce. "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index." *Review of Financial Studies*, 23 (2010), 1909–1940.
Hoberg, G., and V. Maksimovic. "Redefining Financial Constraints: A Text-Based Analysis." *Review of Financial Studies*, 28 (2015), 1312–1352.
Kaplan, S. N., and L. Zingales. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *Quarterly Journal of Economics*, 112 (1997), 169–215.
Kelly, B. T.; S. Pruitt; and Y. Su. "Characteristics are Covariances: A Unified Model of Risk and Return." *Journal of Financial Economics*, 134 (2019), 501–524.
Kiyotaki, N., and J. Moore. "Credit Cycles." *Journal of Political Economy*, 105 (1997), 211–248.
Lamont, O.; C. Polk; and J. Saaá-Requejo. "Financial Constraints and Stock Returns." *Review of Financial Studies*, 14 (2001), 529–554.
Matsa, D. A. "Capital Structure as a Strategic Variable: Evidence from Collective Bargaining." *Journal of Finance*, 65 (2010), 1197–1232.
McCahery, J. A.; Z. Sautner; and L. T. Starks. "Behind the Scenes: The Corporate Governance Preferences of Institutional Investors." *Journal of Finance*, 71 (2016), 2905–2932.
Picard, R. R., and R. D. Cook. "Cross-Validation of Regression Models." *Journal of the American Statistical Association*, 79 (1984), 575–583.
Qiu, L., and I. Welch. "Investor Sentiment Measures." NBER Working Paper Series 10794 (2004).
Rauh, J. D. "Investment and Financing Constraints: Evidence from the Funding of Corporate Pension Plans." *Journal of Finance*, 61 (2006), 33–71.
Shao, J. "Linear Model Selection by Cross-Validation." *Journal of the American Statistical Association*, 88 (1993), 486–494.
Stiglitz, J. E., and A. Weiss. "Credit Rationing in Markets with Imperfect Information." *American Economic Review*, 71 (1981), 393–410.
Tirole, J. *The Theory of Corporate Finance*. Princeton: Princeton University Press (2006).
Welch, I. "The Wisdom of the Robinhood Crowd." *Journal of Finance*, 77 (2022), 1489–1527.
Whited, T. M., and G. Wu. "Financial Constraints Risk." *Review of Financial Studies*, 19 (2006), 531–559.