Online Appendix

B Data Sources, Cleaning, Pre-Processing and Matching

B.1 SEC 10-K Filings

The SEC deposits firms' 10-K filings in the SEC Edgar Database, which we download via the Notre Dame Software Repository for Accounting and Finance (SRAF) maintained by Loughran and McDonald (2016). The SRAF provides annual cleaned 10-X filings for the financial years 1993 to 2023, where extraneous characters have been removed. We focus on the filings relevant to our analysis, namely the filing types: 10-K, 10-K405, 10-KSB and 10-KSB40, henceforth simply referred to as 10-K.[46] We supplement this with available 10-K documents (up until 2Q 2025) for the financial year 2024, downloaded directly from the SEC website.

From each 10-K document, we extract identifying information: the Central Index Key (CIK) number, Filing Type and Filing Date. We further extract from the 10-K filing header, the Company Conformed Name, the Conformed Period of Report, the Address, the SIC Classification Code, Former Company Name, and the Date of Name Change, using regular expression techniques.[47] This information is used to match the 10-K document with firm-year observations in the Compustat Database. From each 10-K filing, we extract the Management's Discussion and Analysis (MD&A) section. The MD&A section generally appears as Item 7 in 10-K documents and as Item 6 in 10-KSB documents and its variants. Some MD&As are incorporated in the reference sections and we make use of regular expression techniques to detect these MD&A sections. We

[46] 10-K denotes the main filing type. The other three filing types we consider are variants of annual reports that companies file based on their size, reporting status, and compliance with specific rules. These variants of 10-K forms have now been phased out. Since December 2008, all firms are required to file a 10-K.

[47] There are a total of 1,163 reports which did not have a 'Conformed Period of Report' in their headers. Most of these reports include a statement 'for the fiscal/financial year ended (*date*)'. We extract this date as the 'Conformed Period of Report' for that particular document. The very few remaining reports without the fiscal/financial year statement, have a standalone date in the header of the report. We extract this date as the 'Conformed Period of Report' for that particular report.

disregard all 10-K documents for which we cannot extract a valid MD&A section.[48] we remove non-narrative content such as tables, figures, and headings from the MD&A text using the open-source python library, *unstructured*.

B.2 Compustat Data

We download Annual Compustat North America for the years 1993 to 2024. Variables used in our analysis are defined and cleaned as follows and deflated using the GDP deflator, except for Capital Expenditure (CAPX), which is deflated using the implicit price deflator for private, fixed non-residential investment.

B.2.1 Compustat Variable Construction

- Total Assets is the variable AT.
- Firm age is the number of years since the firm first appears in the Compustat database.
- Tobin's Q is is defined as (AT+(PRCC_F*CSHO)-CEQ)/AT, where PRCC_F is the Annual Price Close (fiscal year end), CSHO is Common Shares Outstanding, AT is Total Assets and CEQ is Common Equity.
- R&D Expenditure is the variable XRD.
- The R&D to lagged Total Asset ratio is constructed by taking R&D Expenditure (XRD), divided by lagged Total Assets (AT).
- Total Debt is the sum of Debt in Current Liabilities (DLC) and Total Long-Term Debt (DLTT).
- The Total Debt to lagged Total Assets ratio is calculated by taking the ratio of Total Debt to lagged Total Assets (AT).
- Book Leverage is the ratio of Total Liabilities (LT) to Total Assets (AT).

[48]In a few instances, MD&A sections cannot be extracted as they do not appear under the conventional headers or, particularly for older reports, poor formatting impedes text extraction.

- The Cash to (lagged) Total Assets ratio is the ratio of Cash and Short Term Investments (CHE) to (lagged) Total Assets (AT).
- Cashflow is constructed by adding Depreciation and Amortization (DP) to Income Before Extraordinary Items (IB).
- Cashflow to lagged Total Assets ratio is Cashflow over lagged AT.
- Cashflow to Total Assets ratio is Cashflow over AT.
- Working Capital is the difference between Current Assets (ACT) and Current Liabilities (LCT).
- Net (Non-Cash) Working Capital is defined as ACT-LCT-CHE.
- Cash dividends is the variable Cash Dividends per share (DVR).
- Cash dividends to lagged Total Assets is the ratio of Cash Dividends to lagged Total Assets (AT).
- Total Dividends is the variable DVT.
- Capital expenditure is the variable CAPX.
- The investment rate is defined as the ratio of CAPX to lagged Net Property, Plant and Equipment (PPENT).
- Return on Assets (ROA) is defined as the ratio of Income Before Extraordinary Items (IB) to Total Assets (AT).
- Acquisitions is the variable AQC.
- Share repurchases is the variable PRSTKC.
- Tangibility is defined as the ratio of Net Property, Plant and Equipment (PPENT) to (current) Total Assets (AT)
- Investment Opportunities is defined as (PRCC_F*CSHPRI + PSTKL + Total Debt - TXDITC)/AT, where PRCC_F is the Annual Price Close (fiscal year end), CSHPRI is Common Shares Outstanding used to calculate earnings per share, Total Debt is the sum of long term and short term debt, and TXDITC is Deferred Taxes and Investment Tax Credit.

- We use variables from Compustat Quarterly for the check proposed by Farre-Mensa and Ljungqvist (2016). These are SSTKY (Sale of Common and Preferred Stock), CSHOQ (Common Shares Outstanding (Quarterly) and PRCCQ (Closing Share Price). As outlined in Farre-Mensa and Ljungqvist (2016), these variables were used to create Equity Issuance Proceeds, as it appears in Table 11.
- Other Sources of Funds is defined as the sum of DLTIS (Long Term Debt Issuance), IBC (Income before extraordinary items), DPC (Depreciation and Amortization), SPPE (Sale of Property, Plant and Equipment), SIV (Sale of Investments) net of Equity Proceeds, scaled by Total Assets (AT).
- Bankrupt and liquidated firms are identified by the values of the variable DLRSN ('02' for bankruptcy and '03' for liquidation).

B.2.2 Variable Cleaning

We apply the general data selection criteria outlined in Section 3. These are:

- We include only firms which use the US Dollar as the reporting currency. Additionally, we drop firms which are not incorporated in the US
- We drop firm-years with zero or missing total assets
- We drop firm-years if sales or the sum of common and preferred stock is zero or negative.
- We drop firms whose reporting sector is SIC 4900-4999 or SIC 6000-6999. Additionally, we drop firms that report using the financial services format, which is used by banks, insurance companies, and other financial institutions.
- We exclude firm-year observations as soon as a firm is in financial distress. Specifically, if the variable DLRSN takes the value '02' or '03'.

In addition, we apply the cleaning criteria below to the following variables:

- We set negative values of Stockholder's Equity (SEQ) to missing. For firm-year observations where Total Debt is less than DLC, we set DLC, DLTT and Total Debt to missing.

- We set values of AT, DLC, DLTT, ACT, LCT, CHE and SALE to missing for extreme changes in these variables. Specifically, values for these variables are replaced with missing in the top 1 percentile of their growth rates.
- We set negative values of CHE, SEQ, DVC, DVT and DVPSP_F to missing. Additionally, the top 0.1 percentile of DVPSP_F is set to missing. For Total Debt/Lagged Assets, Cashflow/Lagged Total Assets, R&D Spending to Lagged Total Assets, R&D Spending to Total Assets, Cash Dividends to lagged Total Assets, Investment Rate, Long Term Debt to Total Assets, Short Term Debt to Total Assets, Total Inventories to Lagged Sales, we set the top 0.1 percentile to missing.
- We set the top and bottom 0.1 percentile to missing of Lagged Change in Tangibility, Lagged Change in Investment Opportunities, Lagged Change in Return on Assets, Cashflow to Total Assets, Cashflow to Lagged Total Assets, Return on Assets, Change in Cash to Total Assets, Change in Current Liabilities to Total Assets, Change in Short Term Debt to Total Assets, Change in Long Term Debt to Total Assets, Change in Investment Rate, Change in Total Dividends to Total Assets, Log Change in Total Liabilities and Change in R&D Spending to Total Assets.
- We set the top and bottom percentile of Working Capital, Tobin's Q and Leverage to missing as well as the same percentiles of the change in the ratios ratio of Working Capital, Accounts Payable and Accrued Expenses to Total Assets. Additionally, we set to missing the bottom percentile of Working Capital over Total Assets and the top percentile of Accounts Payable over Total Assets and Accrued Expenses over Total Assets.

B.3 Matching Compustat Data with the MD&A Information

This section describes how we match MD&A sections from 10-K filings to the corresponding firm-year observations in Compustat. We begin by matching using Central Index Keys (CIK) and the exact report dates. The CIK is a unique identifier assigned by the US Securities and Exchange Commission. It is provided in Compustat and the SEC

Edgar Database. The report date is the Fiscal Period End Date in Compustat, and the Period of Report end date which we extract from the 10-K files. For matches by CIK, we classify the report dates as a match if these are no more than 14 days apart.

Occasionally, firms' CIK codes change, for example, due to mergers or organizational restructuring. A firm may thus have multiple historical CIK codes that differ from its most recent CIK code. By default, Compustat only provides the most recent CIK. We use historical CIK codes provided by Leo Lui and repeat the matching process described above for all remaining unmatched Compustat firms. The steps described above allow us to match the vast majority of firm-year observations (95%) of our final sample. The remaining matched observations in our final sample come from other matching approaches, described below.

For all remaining unmatched Compustat firm-year observations, we perform fuzzy matching on the company name. Prior to using this algorithm, we clean and simplify all company names. We remove non-essential terms and only keep alphanumeric characters, which removes, for example, special characters that sometimes appear in company names. We remove empty spaces and convert all letters to lower case. For the fuzzy matching procedure, we adopt a cosine similarity threshold of 0.75 and proceed to fuzzy match Company Names in Compustat with the 'Company Conformed Name' and 'Former Company Conformed Name' from the 10-K documents. We discard the matches where the matched string is less than 6 characters long. Among observations for which company names were matched, we further match based on the report dates, that is, the Compustat Fiscal Period End Date and the 10-K Period of Report end date. Again, we retain matches if these dates fall within a 14 day interval. This fuzzy matching procedure on corporation name allows us to add another 4,405 firm-year observations to the final data set, around 3.7% of our matches.

Similar to the case of CIKs, companies may change their name over time, yet Compustat by default only records the most recent company name. We access historical company names for every corporate financial year from the CRSP database and link this information to the respective Compustat firm-year observations. This allows us to account for company name changes over time. We repeat the fuzzy company name matching and date matching step as documented in the previous paragraph which allows us to match an additional 0.5% of our final data set.

We further collect historical company names using the DISCERN 2.0 database provided by Arora et al. (2024a,b). The DISCERN 2.0 data set matches patent ownership to a particular company, and captures company name changes over time. Name changes in this data set are already linked to the Compustat firm and fiscal year end variables, 'gvkey' and 'fyear'. We repeat the fuzzy company name matching and date matching step as documented previously. We match an additional 0.2% of our final sample in this way.

Our matching procedures achieve a 83.65% matching rate for cleaned Compustat firm-year observations. Figure D2 shows the matching rate over time. Since 1996, the matching rate is consistently above 87% (on average 91.2%) which is in line with previous studies.[49] Prior to the SEC's full implementation of the Edgar system in the early 1990s, filings were often submitted in paper format. For this reason, filings at the beginning of our sample may not be electronically available on the Edgar database which results in the substantially lower matching rates during the first three years. The decline in publicly listed firms evident from Figure D2 is well documented in the literature, see e.g. Kahle and Stulz (2017) or Doidge et al. (2017).

C Training and Performance of BERT Models

C.1 Details on Generating Training Data Using Active Learning

Figure B1 provides a graphical overview of the model training steps for the Constraint Classifier model, including the Active Learning algorithm. Further details of these steps are described below.

[49] Chu et al. (2021) for example achieve a match rate of 63.8% for their sample.

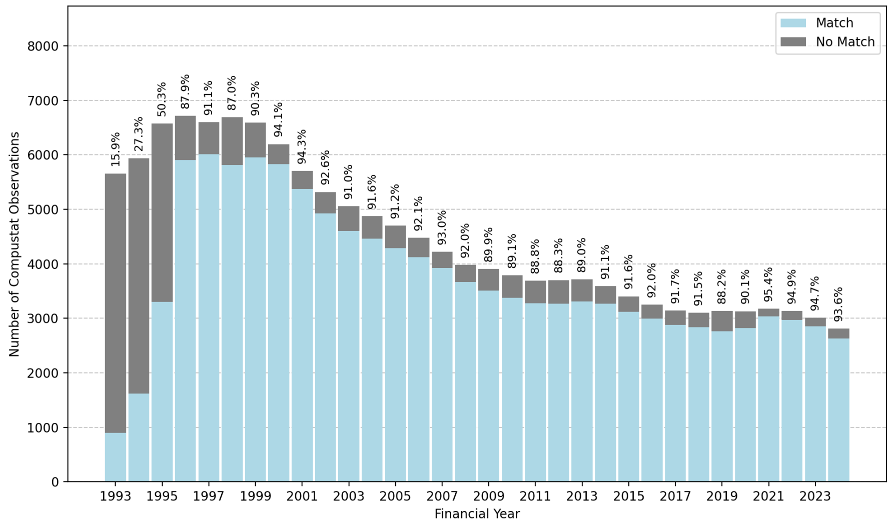

Figure D2: Compustat Firm-Year Observations and Matching Rates with MD&A Sections

Step 1: Initial Data Collection

Our first step is to collect an initial seed data set from the SEC filings. For this purpose we manually read MD&A sections and extract and label text segments as constrained or unconstrained. These text segments typically have the length of a paragraph (see Section 3 for details). Our seed data set contains 251 text segments labeled 'constrained' and 2,000 segments labeled 'unconstrained'.

Step 2: Training a FinBERT model

We use the seed dataset of text segments from step one to train a BERT model based on the FinBERT pre-trained model. Training involves replacing the output layer of the pre-trained FinBERT model with a binary classification layer that can be trained to predict for a given text segment whether the text is suggestive of financial constraints or not.

For each input text segment, the trained model outputs a probability value for the estimated likelihood that the input sample indicates financial constraints. Throughout this training procedure, we refer to a segment as being labeled constrained by our model if the predicted probability is greater than 0.5 and unconstrained if it is below 0.5. This cut-off is standard in the literature (e.g. Rosa (2010) and Pang et al. (2002)). We train

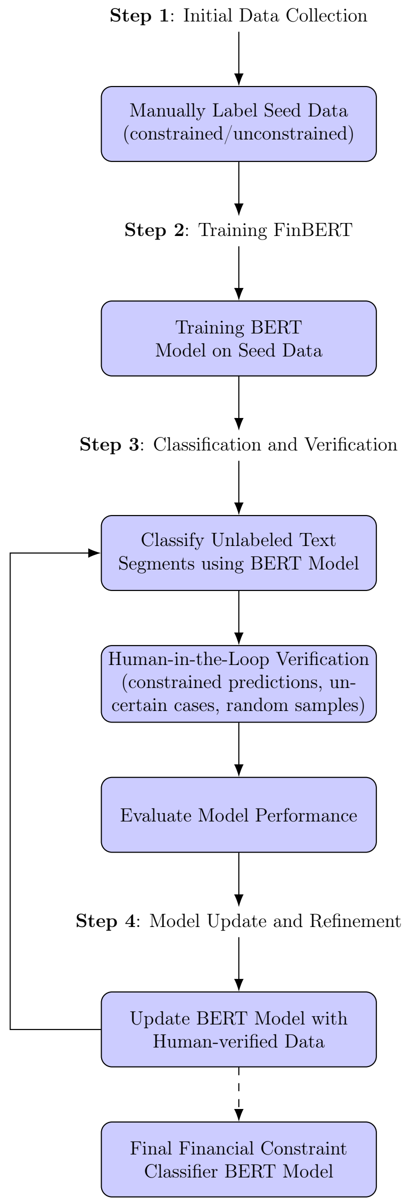

Figure B1: Active Learning Process for Financial Constraint Classification

our BERT models using all hand-labeled data available in each iteration. For more details on the specifics of the training of a BERT model see Appendix C.2 which shows that the predicted probabilities tend to be concentrated either near zero or near one, far away from the 0.5 cut-off threshold.

Step 3: Automated Classification and Human-in-the-Loop Verification

After completing the initial training phase, we employ our model to categorize a broader corpus of text segments outside our initial hand-labeled training sample. Its primary objective is to identify text segments and automatically classify them based on the learned patterns. This process results in a pre-classified dataset, with predictions generated by our latest BERT model.

After the automated classification, expert readers begin a detailed verification process. This involves a strategic review of the pre-classified text segments. We select three separate samples of text segments: a) text segments classified as indicating financial constraints. This is to ensure we specifically monitor the examples that the algorithm indicates as containing financial constraints, in order to minimize the risk of false positives. b) text segments that were classified not to indicate financial constraints, but where the predicted probability of financial constraints exceeds 10%. These are text segments, where the model is uncertain about the correctness of its classification, and therefore these cases are highly valuable for improving the model's performance in subsequent iterations. This also helps to reduce the risk of false negatives. c) a random subset of text segments where the model assigns a probability value of less than 10%. These are text segments likely to be correctly classified as not containing financial constraints, but we include a number of these segments in order to monitor overall model accuracy and ensure an adequate balance of constrained and unconstrained samples in future training runs.

This tripartite selection strategy is aimed at enhancing the robustness of our training data set, by not only capturing clear instances of financial constraints but also by considering edge cases which are essential for improving the model's predictive capabilities. In each iteration of the human-in-the-loop process, we collect around 500 sample text segments of type a) and 500 type b) segments and add to this an equal number of type c) examples, resulting in a total addition to the training data set of roughly 2,000 text segments per iteration which are subsequently read and classified by the expert readers.

Step 4: Model Update and Refinement Loop

Using the approximately 2,000 text segments verified by the expert readers in step three, we update the model through training on the expanded dataset. This training is essential in enabling the model to correct its biases and improve performance over time.

By letting the model pre-label examples and flagging those associated with high levels of uncertainty, we ensure that human attention is focused on high value examples. This allows us to collect a large number of diverse example text segments within a reasonable period of time. We perform several iterations of this refinement cycle, utilizing the progressively updated model to analyze fresh batches of text segments.

During each iteration, we monitor the model's performance metrics. One of these is the agreement rate between the model's predictions and the human readers' judgments, which is defined as:

$$\text{Agreement Rate}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{1}(D_{i,j}^h = D_{i,j}^m),$$

where N_i is the total number of text segments evaluated in one particular iteration $i = 1, \ldots, I$. $D_{i,j}^h$ and $D_{i,j}^m$ are the labels assigned to a particular text segment $j = 1, \ldots, N_i$ by the human expert and the model, respectively. $\mathbf{1}$ is the indicator function that equals 1 when $D_{i,j}^h = D_{i,j}^m$, showing agreement between the human and machine-based classification, and 0 otherwise. This metric serves as an indicator of the alignment between the model's assessments and the experts' judgments.

We further monitor the model's precision in identifying those text segments indicating financial constraints. This metric, referred to as the 'Precision Rate Constrained', is calculated for each iteration as the share of correctly identified text segments indicating financial constraints (where both the model and the human experts agree that a constraint exists) of the total number of instances classified as constrained by the model. It can be expressed as

$$\text{Precision Rate Constrained}_i = \frac{\sum_{j=1}^{N_i} \mathbf{1}(D_{i,j}^m = 1 \text{ and } D_{i,j}^h = 1)}{\sum_{j=1}^{N_i} \mathbf{1}(D_{i,j}^m = 1)},$$

and helps to ensure that the model reliably detects actual constraints and avoids false positives (cases in which a text segment is incorrectly classified as indicative of financial constraints).

Analogously, we monitor the 'Precision Rate Unconstrained' which measures the model's success in identifying text segments indicating no notion of financial constraints. For each iteration, this metric is given by

$$\text{Precision Rate Unconstrained}_i = \frac{\sum_{j=1}^{N_i} \mathbf{1}(D_{i,j}^m = 0 \text{ and } D_{i,j}^h = 0)}{\sum_{j=1}^{N_i} \mathbf{1}(D_{i,j}^m = 0)},$$

tracking the share of text segments indicating no notion of financial constraints (identified in agreement by the human readers and the model) in the total number of such instances classified by the model.

Figure B2 illustrates the performance improvements of our BERT models over each iteration. The left panel shows the agreement rate, i.e. the share of BERT-classified text segments that are judged to be correct by the expert readers. The rate starts at around 84% in iteration one, which is encouraging given the small initial training data set, and increases to approximately 95% in iteration 10.

The right panel of Figure B2 shows the precision rate constrained and the unconstrained counterpart over the ten iterations. In iteration one, the model's ability to correctly identify cases of financial constraints is rather limited. Only about 50% of BERT-classified constrained examples were also judged to be constrained by the expert readers. This is indicative of the initial model's inability to generalize well, and a result of the limited representativeness of the language used in the step-one training data for expressing financial constraints in a textual setting. Since the initial training data (step one) came from a non-random subset of all MD&A text segments, the model has learned to associate certain phrases and words with constraints without realizing that sometimes these can occur in unconstrained contexts. For example, the phrase 'raising additional funds', is likely to occur frequently in the initial examples, mostly in a negative context. However, since there are likely many cases, where the raising of funds is successful, the model initially mistakes these for incidences of constraints. As we provide feedback, the model learns from these mistakes and steadily improves its performance to be correct in over 90% of its own predictions of constraints. We stop our Active Learning procedure after ten iterations when performance plateaus at a high level.

The right panel of Figure B2 shows that the model is successful throughout all iterations at identifying those text segments without any notion of financial constraints. The precision measure for recognizing unconstrained text segments is well above 90% across all iterations.

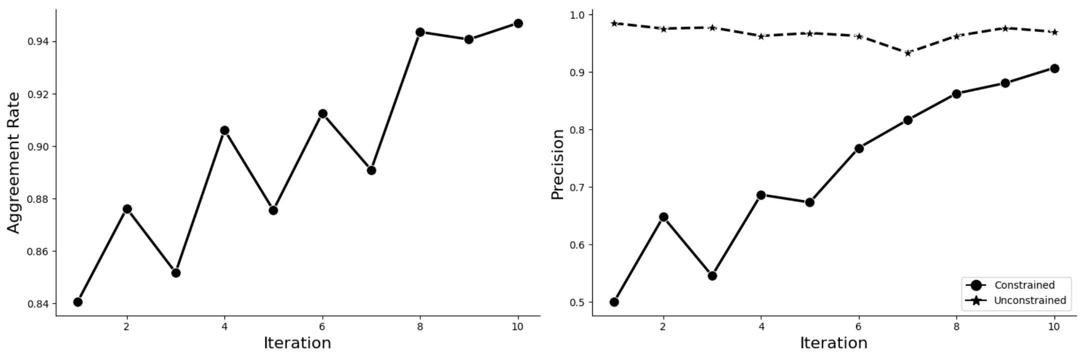

Figure B2: Trained BERT Model Performance over Active Learning Iterations

Table B1 provides additional details on the model's performance throughout the active learning iterations. All metrics are calculated based on the fully human-annotated prelabeled dataset from Step 3. The false positive rate in the first iteration is almost 50%, resulting in a correspondingly low precision of 0.50. We take this as evidence of a biased initial seed sample – the model has learned to associate certain patterns with financially constrained text segments in the seed sample, but when extrapolating these to the wider universe of MD&A documents, it turns up a large number of false positives. Applying our Active Learning strategy alleviates the initial bias over time, although the improvements are not always monotonic. With respect to the rate of false positives, the model takes around 5 iterations to get to a point where the rate improves steadily between every iteration. This process mimics a form of explore-exploit pattern, where the model adapts its understanding of financial constraints based on human feedback. Once the rate of improvements begins to level off, we terminate the Active Learning process.

Table B1: Active Learning Algorithm Performance Metrics

Iteration	Accuracy	Precision	Recall	F1 Score	True Positives	False Positives	True Negatives	False Negatives	Checked Samples
1	0.84	0.50	0.93	0.65	313	312	1,451	22	2,098
2	0.88	0.65	0.92	0.76	418	227	1,445	36	2,126
3	0.85	0.55	0.91	0.68	341	284	1,487	34	2,146
4	0.91	0.69	0.83	0.75	287	131	1,558	60	2,036
5	0.88	0.67	0.91	0.77	441	214	1,389	46	2,090
6	0.91	0.77	0.88	0.82	404	122	1,454	56	2,036
7	0.89	0.82	0.88	0.85	628	141	1,252	89	2,110
8	0.94	0.86	0.85	0.86	352	56	1,637	63	2,108
9	0.94	0.88	0.96	0.92	688	93	1,263	30	2,074
10	0.95	0.91	0.95	0.93	695	71	1,266	39	2,071

Each iteration corresponds to the evaluation of the model's performance on a new set of human-labeled data acquired through the Active Learning process described in Appendix C. Accuracy measures the overall proportion of correctly classified samples. Precision is the proportion of correctly classified constrained samples. Recall is the proportion of actual constrained samples correctly identified by the model. The F1 Score is the harmonic mean of precision and recall. True Positives and True Negatives represent correctly classified samples, while False Positives and False Negatives represent misclassifications.

C.2 Training the Constraints Classifier Model

We train the model over 10 epochs, using a learning rate of $2 \times e^{-5}$ and a weight decay of 0.01.[50] The training process minimizes the cross-entropy loss function, updating the model parameters using gradient descent.[51] Cross-entropy loss is commonly used in classification tasks to measure the difference between the true distribution of labels and the predicted probabilities output by a model. For a multi-class classification problem, it is defined as:[52]

$$L = -\sum_{c=1}^{C} y_{o,c} \log(p_{o,c}), \tag{4}$$

where C is the number of classes, $y_{o,c}$ is a binary indicator (0 or 1) that is unity if class label c is the correct classification for observation o, and $p_{o,c}$ is the predicted probability of observation o being of class c. The loss function penalizes the model more the greater the discrepancy between the actual and the predicted probabilities, effectively guiding the model to increase the probability of the true class label while decreasing the probabilities of the incorrect labels.

Table B2 shows that the Constraint Classifier Model is subject to low uncertainty when

[50] An epoch in machine learning is a full pass through the entire training data set by the model. The learning rate controls the size of the parameter updates during gradient descent – smaller values lead to more gradual, stable convergence. Weight decay is a regularization method that penalizes large parameter values to help prevent overfitting.

[51] Gradient descent is an iterative optimization technique that adjusts the model parameters in the direction that most rapidly decreases the loss function.

[52] Alternatively, since there are only two classes in this case, this could be considered a binary classification problem.

labeling text segments as either constrained or unconstrained. It assigns high probabilities (mostly >90%) to text segments classified as constrained and low probabilities (mostly $< 1\%$) to those classified as unconstrained. This indicates that the applied threshold of 0.5 as a cut-off between constrained and unconstrained is not very binding.

Table B2: Distribution of Probability Scores by Financial Constraints Classifier

	All Text Segments	Constrained Text Segments	Unconstrained Text Segments
Mean	0.0161	0.9506	0.0014
Variance	0.0143	0.0130	0.0002
Minimum	0.0001	0.5000	0.0001
Maximum	0.9990	0.9990	0.4999

The table reports statistics of the distribution of BERT probabilities (confidence scores) for the Financial Constraints Classifier Model.

C.3 Training of the Time Horizon and Severity Classifier BERT Models

We train two additional BERT models for the classification tasks related to Time Horizon and Severity. Both models use as a starting point the off-the-shelf FinBERT model, trained on about 1,000 hand-labeled text segments for the respective categories. The models are trained over 10 epochs with a learning rate of $2e^{-5}$ and a weight decay of 0.01. The training process employs specific loss functions tailored to each classification task, detailed below.

The Time Horizon Classifier relies on multi-label models, meaning they are designed to predict multiple labels simultaneously for a single instance. For example, a text segment might simultaneously reference both financial constraints at present and in future. The training process minimizes the binary cross-entropy loss, which treats the prediction for each label as an independent binary classification problem (i.e. "Does this label apply?"). Specifically, the loss function for the multi-label classification task is defined as:

$$L = -\frac{1}{M} \sum_{m=1}^{M} \left[y_{o,m} \log(p_{o,m}) + (1 - y_{o,m}) \log(1 - p_{o,m}) \right],$$

where M is the number of labels, $y_{o,m}$ is a binary indicator (0 or 1) indicating whether label

m is true for observation o, and $p_{o,m}$ is the predicted probability for label m. For multi-label classification, the loss is computed for each label separately and then averaged across all labels. This ensures that the model learns to predict each label effectively without interference from others. By handling each label independently, the model is able to account for the complexity of our data where multiple aspects of financial constraints (e.g. frictions at present and anticipated in future) may be present simultaneously in a single text segment.

The Severity Classifier, in contrast, is a multi-class model, where each observation belongs to a single class. This model minimizes the categorical cross-entropy loss as given in equation (4). This loss penalizes discrepancies between the true class and predicted probabilities, guiding the model to maximize the probability of the correct class.

Table B3 shows that the Time Horizon and the Severity Classifier Model are subject to low uncertainty when labeling text segments. It assigns high probabilities (on average >95% with a low variance) to text segments. In case of these classifiers, there are four possible options, so for a category to be assigned, at least a probability score of 25% is required and the reported minimum is substantially above this value. This indicates that the model is highly certain about its classifications.

Table B3: Average Probability Scores for Severity and Time Horizon Classifiers

Panel A: Severity Classifier				
	Mean	Variance	Min	Max
Mild	0.9676	0.0080	0.3793	0.9999
Moderate	0.9838	0.0040	0.3539	0.9999
Severe	0.9890	0.0029	0.3969	0.9999
Panel B: Time Horizon Classifier				
	Mean	Variance	Min	Max
Current	0.9976	0.0003	0.5278	0.9997
Future	0.9981	0.0003	0.5002	0.9998
Current & Future	0.9975	0.0001	0.5889	0.9997

The above statistics show the average BERT probabilities (confidence) on the categories assigned to each text segment.

D Classifying Financially Constrained Text Segments

D.1 Categorizing Financially Constrained Text Segments

This section provides further details on the classification of financial constraints in the time horizon and severity dimensions.

Time Horizon of Constraints. The time horizon dimension distinguishes between constraints that affect the firm in the current financial year and those expected to impact in the future. We allow text segments to be labeled as 'Current', 'Future', or 'Current & Future'. In addition, we allow segments to be labeled as 'unclear' if it cannot clearly be mapped into one of the former categories. Table C1 summarizes the detailed definitions for each category.

Table C1: Definitions for Classifying the Time Horizon of Constraints

Time Horizon	**Definition**
Current	When financial constraints occur in the current financial year.
Future	When financial constraints are expected to occur in future (after the current financial year).
Current & Future	When financial constraints occur in the current financial year and are expected to occur in future financial years.
Unclear	Time horizon cannot be evaluated from the text.

Severity of Constraints. Severity measures the intensity of the financial constraints, classified as 'mild', 'moderate', or 'severe'. Mild constraints may cause minor operational constraints, while severe constraints threaten a firm's survival. These classifications are mutually exclusive. If a text segment discusses different constraints of varying severity, we select the most severe classification. Further, we allow for an additional label, 'unclear', if the segment cannot clearly be mapped to one of the three former categories. Table C2 summarizes the detailed definitions for each category.

Using the training data set, we then train two BERT models, each specialized in predicting either the time horizon or severity dimension. Details about the training and performance of these models are provided in Appendices C.2 and C.3.

Table C2: Definitions for Classifying the Severity of Constraints

Severity	Definition
Mild	When constraints may affect a firm's ability to finance investment projects. The lack of funds implies no direct risk to a firm's survival and continuation of ongoing operations.
Moderate	When constraints do affect a firm's ability to finance investment projects or to finance at least some ongoing operations. The inability to raise funds is no direct concern to a firm's survival.
Severe	When constraints affect a firm's ability to finance investment projects or ongoing operations, and this is a threat to firm survival.
Unclear	When severity of constraints cannot clearly be mapped to the above categories.

D.2 Details on the Performance Metrics for Multi-Label Classifiers

Owing to the non-exclusive multi-label nature of the Time Horizon Classifier, we calculate the accuracy, precision, recall, and F1-score using micro-averaging methods suitable for multi-label classification. Specifically, accuracy is computed as the Hamming Accuracy, which measures the proportion of correctly predicted labels over the total number of labels. This is calculated by comparing the predicted and true labels for each class independently and averaging the correctness across all labels and samples.

For precision, recall, and F1-score, we aggregate the true positives (TP), false positives (FP), and false negatives (FN) across all labels and samples before computing the metrics. Using micro-averaging, these metrics are defined as:

$$\begin{aligned}
\text{Precision}_{\text{micro}} &= \frac{\sum_{\text{labels}} \text{TP}}{\sum_{\text{labels}} (\text{TP} + \text{FP})}, \\
\text{Recall}_{\text{micro}} &= \frac{\sum_{\text{labels}} \text{TP}}{\sum_{\text{labels}} (\text{TP} + \text{FN})}, \\
\text{F1}_{\text{micro}} &= 2 \times \frac{\text{Precision}_{\text{micro}} \times \text{Recall}_{\text{micro}}}{\text{Precision}_{\text{micro}} + \text{Recall}_{\text{micro}}}.
\end{aligned}$$

This approach treats each individual label prediction equally, effectively turning the multi-label problem into multiple binary classification problems, and provides an overall performance metric that accounts for all labels.

For the Severity Classifier, which operates as an exclusive multi-class classifier with mutually exclusive labels (mild, moderate, severe, unclear), accuracy, precision, recall, and F1-score are computed differently compared to a non-exclusive multi-label setup. Accuracy is defined as the proportion of correctly predicted labels to the total number of segments and provides an overall measure of the classifier's performance. Precision for a given class is calculated as the ratio of true positives to the sum of true positives and false positives, while recall is calculated as the ratio of true positives to the sum of true positives and false negatives. The F1-score combines precision and recall into a single metric using the harmonic mean.

D.3 Descriptive Statistics on Text Segments' Classifications

Table C3 shows distributions of text segments, conditional on the number of constrained text segments within an MD&A. Panel A concerns the severity distribution while Panel B focuses on the time horizon.

Table C3: Severity and Time Horizon Distribution of Constrained Text Segments, Conditional on the Number of Constrained Segments within an MD&A Document

Panel A: Severity	1	2	3–5	6–10	>10	≥1
Mild	40.86%	26.85%	14.93%	9.73%	7.02%	21.73%
Moderate	50.69%	58.54%	63.41%	65.39%	68.50%	60.27%
Severe	7.25%	13.30%	19.82%	22.38%	20.77%	16.20%
Unclear	1.20%	1.31%	1.84%	2.51%	3.72%	1.81%
Total by Column	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Panel B: Time Horizon	1	2	3–5	6–10	>10	≥1
Current	7.97%	12.59%	18.16%	21.46%	21.33%	15.50%
Future	82.01%	74.46%	64.27%	55.27%	50.32%	67.98%
Current & Future	9.14%	11.59%	15.21%	19.11%	21.47%	14.13%
Unclear	0.87%	1.37%	2.37%	4.16%	6.88%	2.39%
Total by Column	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

The column headers of each Panel indicate the number of constrained segments within an MD&A section. Each cell reports the share of segments classified as mild, moderate, severe, or unclear conditional on the number of constrained segments within each MD&A section.

Table C4 displays the distribution of MD&A documents by the number of decisive text segments. The vast majority of constrained firms have between one and three decisive text segments, although a few firms have more than ten. This relatively low number of decisive text segments per MD&A document helps in identifying the time horizon of financial constraints.

Table C4: Distributions of Decisive Text Segments and Time Horizon Classifications

	Firm-Year Obs. Conditional on Number of Decisive Text Segments Per MD&A						
Number of Decisive Segments	1	2	3	4–5	6–7	8–9	≥10
Mildly Constrained MD&As	10,689	1,521	130	18	0	0	0
Moderately Constrained MD&As	17,709	6,861	2,650	1,954	478	156	111
Severely Constrained MD&As	6,435	2,895	1,115	683	114	33	6
All Constrained MD&As	34,833	11,277	3,895	2,655	592	189	117

D.4 Comparing BERT to Large Language Models

In this section, we benchmark our trained BERT model against a decoder-based large language model (LLM). While BERT models are optimized for classification tasks, recent advances in decoder-based LLMs, which have primary strength in reasoning, have substantially improved their classification ability.

Advantages of using BERT models in an academic context are that results are fully reproducible and stable across time, and they can be trained to optimize performance on specific tasks. Importantly, BERT models learn exactly the example definitions one provides. LLMs may inject prior world knowledge or interpret labels creatively, which can blur fine distinctions (e.g. current vs. expected constraints). Classification with BERT yields explicit probability scores. This enables clear thresholding and computation of standard performance metrics (e.g. ROC AUC), while LLM prompt-based classification is harder to rigorously evaluate. LLMs excel when labels are fuzzy, examples are scarce, or reasoning across long contexts is required. Based on these arguments, it appears that BERT-style models are the better choice for our for a high-volume, well-defined classification task. Nonetheless, it is interesting to compare their performance against LLMs.

For the LLM-based evaluation, we use `gpt-5-nano`, a small-scale OpenAI model chosen to balance quality with speed and cost control.[53] We evaluate both the BERT and the LLM models using our testing data set from Section 4.1.2, corresponding to 15% of

[53]Specifically, we use the model-snapshot `gpt-5-nano-2025-08-07`. The `gpt-5-nano` model is the fastest and smallest current version of GPT-5, and is recommended for classification tasks.

the labeled segment-level training corpus.

We explore how LLM performance varies along two key dimensions. First, we vary the number of in-context examples provided to the model (zero-shot and N-shot prompting), allowing us to study the marginal contribution of labeled examples. These examples are selected using a similarity-based retrieval strategy from the training set, ensuring that the examples are semantically close to the target text segment.[54] Our retrieval approach for N-shot prompting provides a strong advantage to the LLM. Second, we vary the level of explicit reasoning effort that the model is asked to deploy, ranging from minimal instructions to more deliberate, step-by-step reasoning, in order to assess how structured reasoning interacts with retrieved examples in shaping the LLM's classification performance. The LLM prompt is shown in Table C6.

Table C5 reports the performance statistics of the trained BERT classifier and the LLM across different prompting setups. Without any examples (zero-shot) and varying reasoning effort, the LLM's performance is worse that BERT's in all regards. Its accuracy, precision, recall and F1 Score are all substantially inferior. This holds in particular for precision where BERT achieves 86.13% and the zero-shot LLM ranges from 63.78%-66.63% for varying reasoning effort. Providing examples to the LLM, based on our large training data set, can improve performance, but only in conjunction with higher reasoning effort, and it still falls below BERT's performance.

[54]To implement the similarity based retrieval strategy, we create the embeddings for all training and test segments using the `text-embedding-3-small` model. For each text segment we then select the N most similar text segments from the training data using cosine similarity. These are then provided into the LLM's context window with the associated labels.

Table C5: Classifier performance on held-out test set

Panel A: Trained BERT		Accuracy	Recall	Precision	F1 Score
BERT		94.14%	86.81%	86.13%	86.47%
Panel B: LLM via API					
N-shot	Reasoning effort	Accuracy	Recall	Precision	F1 Score
0	minimal	86.21%	72.19%	66.63%	69.30%
0	medium	86.27%	76.37%	65.58%	70.57%
0	high	85.79%	78.85%	63.78%	70.52%
1	minimal	84.67%	76.63%	61.60%	68.30%
1	medium	88.89%	82.25%	70.87%	76.13%
1	high	88.83%	81.85%	70.85%	75.95%
3	minimal	84.67%	80.81%	60.87%	69.43%
3	medium	89.11%	83.16%	71.17%	76.70%
3	high	89.87%	85.90%	72.31%	78.52%
10	minimal	84.64%	79.76%	60.98%	69.12%
10	medium	88.38%	83.68%	69.00%	75.63%
10	high	89.56%	84.86%	71.82%	77.80%

The LLM results use dynamic similarity-based retrieval of in-context examples from the training set. The LLM outputs are evaluated as binary predictions.

Table C6: LLM Prompt Template (Few-shot)

LLM Prompt Template (Few-shot)

```
You are an expert in corporate finance and credit risk.
Task:  Classify short text segments from firms' disclosures into one of two
categories.
• 1 (constrained):  The text indicates that the firm is, or is very likely
  to be, financially constrained.  Signals may include:  difficulty
  accessing external finance, binding borrowing limits or covenants,
  inability to fund profitable projects, stressed liquidity, urgent need
  to deleverage, or strong dependence on internal cash flow.
• 0 (unconstrained):  The text does NOT clearly indicate that the firm is
  financially constrained.  It may discuss generic risks, macro conditions,
  or market volatility, but without strong evidence that the firm itself
  currently faces binding financing frictions.
Your job:  Given a new text segment, decide whether it should be labeled 0
or 1.
Here are some labeled examples (for your guidance):
Example 1
Label:  [0 or 1]

[EXAMPLE TEXT 1]

Example 2
Label:  [0 or 1]

[EXAMPLE TEXT 2]

...

Now classify the following new text segment:

[TEXT]

Instructions:
- Respond with EXACTLY ONE CHARACTER:
- "1" if the firm appears financially constrained.
- "0" otherwise.
- Do NOT include any explanation, spaces, or extra characters.

Answer (one character only:  0 or 1):
```

D.5 Examples of AI-Classified Text Segments

Table C7 shows examples of text segments classified using our trained AI models.

Table C7: Example Text Segments

Text Segment	Labels
ACURA PHARMACEUTICALS INC (2/4/1997)	
From late December 1996 through March 31, 1997, the Company borrowed an aggregate of approximately 1,100,000 from certain of the holders of the Company's convertible subordinated debentures. These borrowings, which are evidenced by unsecured promissory notes, are due and payable as to principal on demand. Interest, at the rate of 10 per annum, is payable on a quarterly basis. The Company utilized the proceeds of these borrowings for working capital. As previously indicated, the Company has continued to actively pursue financing. At the current time, the Company is discussing with several parties obtaining financing which will replace the Company's banks and provide additional working capital. There can be no assurance that the Company will be able to obtain any such financing on commercially acceptable terms.	current; mild
ALLIED HEALTHCARE INTL INC (27/1/1997)	
Further expansion of the Company's business (particularly through acquisitions) may require the Company to incur additional debt or offer additional equity if cash generated from operations, cash on hand and amounts available under the Credit Facility are inadequate or not available to meet its needs. There can be no assurance that any such additional debt or equity will be available to the Company, or if available, will be on terms acceptable to the Company. In addition, covenants contained in the Credit Facility restrict the Company from entering into transactions not in the ordinary course of business, including making acquisitions and issuing capital stock, without consent.	future; mild
AMERICAN AIRLINES GROUP INC (25/2/2005)	
Credit Ratings Since the Terrorist Attacks, AMR s and American s credit ratings have been lowered to significantly below investment grade. These reductions have increased borrowing costs and otherwise adversely affected borrowing terms, and limited borrowing options. Additional reductions in the credit ratings could further increase borrowing or other costs and further restrict the availability of future financing.	current & future; mild
UNIVERSAL DETECTION TECHNOL (15/4/2003)	
On June 11, 2002 we were notified by The Nasdaq Stock Market that we did not meet the continued listing requirements of The Nasdaq Small Capital Market and our common shares were delisted on the close of business on June 19, 2002. Our common stock currently is trading on the OTC Bulletin Board. It may be more difficult to raise additional debt or equity financing while trading on the OTC Bulletin Board. If we are unable to raise additional financing, we will not be able to accomplish our business objectives and may consider steps to protect our assets against creditors.	current; moderate

Continued on next page

Text Segment	Labels
US AIRWAYS GROUP INC (27/3/2003)	
US Airways currently has an agreement with NPC which expires March 31, 2003, but which provides for a 45-day extension if a series of milestones are met. Specifically, the 45-day extension with NPC requires that the Company have a confirmed plan of reorganization and a new credit card processing agreement in place by March 31, 2003. The Company then has to substantially complete the plan and close on the ATSB Loan by April 15, 2003, allowing for an extension of the NPC agreement until May 15, 2003. Assuming these conditions are met, US Airways would expect to transition to its new processor in late April or early May. However, there can be no assurance that the current contract can be extended beyond March 31, 2003, if US Airways does not execute an agreement with a new processor by this date. The termination of credit card processing agreements could have a material adverse effect on the Company s liquidity, financial condition and results of operations. On August 12, 2002, Standard and Poor's (S&P) lowered its corporate credit ratings on US Airways Group and US Airways to D (Default) from SD (Selective Default). Ratings on most issues of US Airways that had not already defaulted were lowered as well, and remain on CreditWatch with developing implications. On August 12, 2002, Moody s Investor Service (Moody s) downgraded its Senior Implied Ratings of the Company and US Airways to Caa3 from Caa2 and the rating outlook is negative.	current; moderate
OTELCO INC (31/12/2008)	
There can be no assurance that we will have sufficient cash in the future to pay dividends on our Class A common stock at the historical rate or at all. If we do not generate sufficient cash from our operating activities in the future to pay dividends, we may have to reduce or eliminate dividends or rely on cash provided by financing activities in order to fund dividend payments, if such financing is available. However, if we use working capital or borrowings under our credit facility to fund dividends, we would have less cash available for future dividends and we may not have sufficient cash to pursue growth opportunities such as the introduction of new services and the acquisition of other telephone companies, or to respond to unanticipated events such as the failure of a portion of our switching or network facilities. If we do not have sufficient cash to finance growth opportunities or capital expenditures that would otherwise be necessary or desirable, and cannot find alternative sources of financing, our financial condition and our business will suffer. Our current dividend policy, our high indebtedness levels and related debt service requirements and our capital expenditure requirements will significantly limit any cash available from operations for other uses for the foreseeable future.	future; moderate
PLUS THERAPEUTICS INC (5/3/2024)	
To date, our operating losses have been funded primarily from outside sources of invested capital from issuance of our common and preferred stocks, proceeds from our Term Loan with Oxford and grant funding. We have had, and will continue to have, an ongoing need to raise additional cash from outside sources to fund our future clinical development programs and other operations. There can be no assurance that we will be able to continue to raise additional capital in the future. Our inability to raise additional cash would have a material and adverse impact on our operations and would cause us to default on our Term Loan.	current & future; moderate

Continued on next page

Text Segment	Labels
SALON MEDIA GROUP INC (31/12/2004)	
PricewaterhouseCoopers LLP, Salon's independent accountants through November 13, 2003 have included a paragraph in their report indicating that substantial doubt exists as to Salon's ability to continue as a going concern because of Salon's recurring operating losses, negative cash flow and an accumulated deficit for the for the years ended March 31, 2002 and March 31, 2003. Salon's current independent accountants, Burr, Pilger & Mayer LLP make the same assertions in their report for Salon's year ended March 31, 2004.	current; severe
GLOBALNET CORP (31/12/2003)	
We anticipate that the full amount of the convertible debentures, together with accrued interest, will be converted into shares of our common stock, in accordance with the terms of the convertible debentures. If we are required to repay the convertible debentures, we would be required to use our limited working capital and raise additional funds. If we were unable to repay the debentures when required, the debenture holders could commence legal action against us and foreclose on all of our assets to recover the amounts due. Any such action would require us to curtail or cease operations.	future; severe
ENERGY FOCUS INC (31/12/2018)	
If we fail to obtain required additional financing to sustain our business before we are able to produce levels of revenue to meet our financial needs, we will need to delay, scale back or eliminate our business plan and further reduce our operating costs and headcount, each of which would have a material adverse effect on our business, future prospects, and financial condition. A lack of additional financing could also result in our inability to continue as a going concern and force us to sell certain assets or discontinue or curtain our operations and, as a result, investors in the Company could lose their entire investment.	future; severe
RXBAZAAR INC (31/12/2003)	
We are not profitable. In the year ended December 31, 2003 we incurred a net loss of 4.7 million, and as of December 31, 2003, we had a stockholders' deficit of 5.7 million. We anticipate future losses and negative cash flow to continue for the foreseeable future, even though monthly sales have recently increased. Because of our history of operating losses and working capital deficit, our independent auditors have expressed doubt as to our ability to continue as a going concern. We can give no assurance that we will be able to operate profitably or produce a positive cash flow. In general, our ability to become profitable depends on our ability to generate and sustain substantially higher revenues while maintaining reasonable expense levels. If we were to achieve profitability, we cannot give any assurance that we would be able to sustain or increase profitability on a quarterly or annual basis in the future.	current & future; severe
WORLDS INC (24/3/2008)	
We have had to severely diminish our operations due to a lack of liquidity. We will attempt to continue to operate in this manner until we find an additional source of capital. We have no current arrangements with respect to, or sources of, additional financing and there can be no assurance that any such financing would become available. We may need to completely halt all operations.	current & future; severe

References in the Appendix

Almeida, H., Campello, M., and Weisbach, M. S. (2004). The Cash Flow Sensitivity of Cash. *Journal of Finance*, 59(4):1777–1804.

Almeida, H., Campello, M., Weisbach, M. S., et al. (2024). The Cash Flow Sensitivity of Cash: Replication, Extension, and Robustness. *Critical Finance Review*, 13(3-4):351–365.

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., Shin, H. M., and Shvadron, D. (2024a). DISCERN 2.0: Duke Innovation & Scientific Enterprises Research Network [Dataset].

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., and Shvadron, D. (2024b). Back to the Future: Are Big Firms Regaining their Scientific and Technological Dominance? Evidence from DISCERN 2.0. (available soon).

Chu, K. K., Chen, S., and Leung, T. (2021). A novel algorithm for generating a gvkey-cik link table. *Journal of Information Systems*, 35(1):27–46.

Doidge, C., Karolyi, G. A., and Stulz, R. M. (2017). The US Listing Gap. *Journal of Financial Economics*, 123(3):464–487.

Farre-Mensa, J. and Ljungqvist, A. (2016). Do Measures of Financial Constraints Measure Financial Constraints? *Review of Financial Studies*, 29(2):271–308.

Grieser, W. D. and Hadlock, C. J. (2019). Panel-Data Estimation in Finance: Testable Assumptions and Parameter (In)Consistency. *Journal of Financial and Quantitative Analysis*, 54(1):1–29.

Hadlock, C. J. and Pierce, J. R. (2010). New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index. *Review of Financial Studies*, 23(5):1909–1940.

Hoberg, G. and Maksimovic, V. (2015). Redefining Financial Constraints: A Text-Based Analysis. *Review of Financial Studies*, 28(5):1312–1352.

Kahle, K. M. and Stulz, R. M. (2017). Is the US Public Corporation in Trouble? *Journal of Economic Perspectives*, 31(3):67–88.

Loughran, T. and McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4):1187–1230.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 79–86. Association for Computational Linguistics.

Riddick, L. A. and Whited, T. M. (2009). The Corporate Propensity to Save. *Journal of Finance*, 64(4):1729–1766.

Rosa, G. J. (2010). The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Hastie, T., Tibshirani, R., and Friedman, J. *Biometrics*, 66(4):1315.

Welch, I. (2021). Ratios of Changes: How Real Estate Shocks Did Not Affect Corporate Investment. *American Economic Review*, 111(10):3351–3387.

Whited, T. and Wu, G. (2006). Financial Constraints Risk. *Review of Financial Studies*, 19(2):531–559.