

Identifying M&A Targets from Textual Disclosures: A Transformer Neural Network Approach

Nils Lohmeier Lennart Stitz Paul Drecker*

This Version: October 9, 2026

Abstract: Can textual information from firm disclosures help to identify M&A targets? We employ two state-of-the-art transformer neural networks, RoBERTa and Longformer, based on 102,987 annual reports of publicly listed U.S. firms to estimate takeover likelihoods. We show that incorporating publicly available, standardized textual information can improve the predictability of corporate takeovers significantly in out-of-sample tests and that this information is not fully incorporated in stock prices. We use explainable artificial intelligence methods to examine the reasons for the improved predictions. Our analyses indicate that the machine learning algorithms are able to identify product offerings and firm-specific capabilities sought by acquirers.

Keywords: Mergers & Acquisitions, Takeover Probability, Machine Learning, Natural Language Processing, Deep Learning.

JEL: C45, G34

*Lohmeier (corresponding author), nils.lohmeier@wiwi.uni-muenster.de, University of Münster Finance Center Münster; Stitz, lennart.stitz@wiwi.uni-muenster.de, University of Münster Finance Center Münster; Drecker, paul.drecker@uni-muenster.de, University of Münster Department of Political Science. The authors thank Jarrad Harford (the editor), an anonymous referee, Mong Shan Ee (discussant), Christopher Felder (discussant), Paul Hark, Stefan Klein, Alejandro Lopez Lira, Johannes Kriebel, Thomas Langer, Nathalie Mitev, Tapas Mohapatra (discussant), Hannes Mohrschladt, Andreas Pfingsten, Daniel Rabetti (discussant), Edona Selimaj, Christoph Schneider, Susanne Siedhoff, Yannick Wilner, Tao Yuan (discussant), and participants at numerous seminars and conferences for their helpful comments and suggestions. Calculations for this project were performed on the HPC cluster PALMA II of the University of Münster, subsidized by the DFG (INST 211/667-1). We are responsible for any remaining errors or omissions.

I. INTRODUCTION

Mergers and acquisitions (M&A) are integral elements of modern economies, affecting involved parties and shaping broader industrial landscapes. One central finding of the M&A literature is that the targets of corporate acquisitions experience large abnormal returns around the offer announcement (see, e.g., Jensen and Ruback (1983), Jarrell, Brickley, and Netter (1988), Andrade, Mitchell, and Stafford (2001)), making the prediction of targets a potentially highly profitable task (Cremers, Nair, and John (2009)). Identifying M&A targets based on financial data, however, has proven to be an elusive challenge (Palepu (1986)). Recent contributions in the finance literature have begun to tap into a large, previously untouched source of firm-specific information, namely unstructured textual data, yielding significant improvements in predictions such as corporate directors' performance (Erel, Stern, Tan, and Weisbach (2021)) and stock performance (Azimi and Agrawal (2021), Brogaard and Zareei (2023)).

In this paper, we employ RoBERTa and Longformer, two state-of-the-art deep learning algorithms, to predict targets of corporate acquisitions based on 102,987 annual financial reports. Our empirical analyses show that incorporating textual disclosures yields a statistically and economically significant improvement of predictions beyond standard financial variables. A simple trading strategy based on these predictions yields large positive excess returns of up to 0.85% per month. We employ novel advances in explainable artificial intelligence to shed further light on our improved predictions. These analyses suggest that predictability stems from fundamental information about a target's product offerings, operations, and capabilities, which trigger a change in corporate control.

To test the relevance of financial disclosures to takeover predictions, we access the Securities and Exchange Commission's EDGAR website and obtain 102,987 Form 10-K filings, the

mandatory annual financial report, from 1993 to 2020. In our analyses, we focus on the Business Description (BD, Item 1) and the Management Discussion and Analysis (MDA, Item 7) as the firm’s business model and managers’ analyses of past results are arguably most relevant to identifying takeover targets (Hoberg and Phillips (2010), Routledge, Sacchetto, and Smith (2017), Frésard, Hoberg, and Phillips (2020)). We merge textual information with M&A data from SDC to identify takeover targets and add financial controls and stock returns from Compustat and CRSP. Then, we employ RoBERTa and Longformer to harness the textual information for takeover predictions.

RoBERTa is an intensively pre-trained transformer neural network that is able to process text cohesively rather than sequentially by using a context-specific word embedding (Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer, and Stoyanov (2019)). The exceptional text comprehension ability of transformer neural networks has been the focus of public attention since ChatGPT (Chat Generative Pre-trained Transformer) was launched as a prototype in 2022. RoBERTa’s architecture is based on the BERT model (Devlin, Chang, Lee, and Toutanova (2019)), i.e., it uses a transformer-encoder architecture with self-attention. Encoders translate the text used as input into a numeric format. The resulting encoding of each word is context-specific since it considers all words used in the sentence via self-attention (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin (2017)). RoBERTa’s improvement in performance compared to BERT originates from a changed pre-training approach that focuses on masked language modeling (i.e., predicting a missing word in a sentence) using a dynamic masking approach and more training data. RoBERTa outperforms other modern deep-learning models for natural language processing in relevant benchmark tasks such as the Stanford Question Answering Dataset (SQuAD, Rajpurkar, Zhang, Lopyrev, and Liang (2016), Rajpurkar, Jia, and Liang (2018)) and the General

Language Understanding Evaluation (GLUE, Wang, Singh, Michael, Hill, Levy, and Bowman (2018)). We use the pre-trained RoBERTa model with a binary classifier head and fine-tune it to predict takeover likelihoods using three-fold cross-validation to optimize a set of relevant hyperparameters.

A central drawback of RoBERTa is that it can handle only 512 tokens (Liu et al. (2019)) such that using RoBERTa implicitly limits the processed texts approximately to their first 400 words. One deep learning algorithm that explicitly seeks to alleviate this limitation is Longformer (Beltagy, Peters, and Cohan (2020)). Longformer is also a transformer model built on BERT and is designed to process sequences of up to 4,096 tokens, which translates to approximately 3,200 words per text. Consequently, applying Longformer for prediction tasks has shown a lot of promise, in particular in the context of long legal documents (Xiao, Hu, Liu, Tu, and Sun (2021), Acikalin, Caskurlu, Hoberg, and Phillips (2026)). Therefore, we fine-tune Longformer to the task of predicting takeover likelihoods alongside RoBERTa and compare the performance of both deep learning models.

First, we examine the impact of the text-based predictions on the overall prediction quality. We include text-based predictions in logistic regressions on the likelihood to be targeted in a corporate acquisition beside standard financial controls. We find that text-based predictions based on the BD and the MDA sections of the 10-K filings derived by both RoBERTa and Longformer are highly statistically significant when considered separately, indicating that the financial disclosures contain relevant information for M&A target prediction beyond previously considered variables. In a regression, which considers predictions from BD and MDA jointly, we find that there is substantial information overlap between the BD and the MDA.

Second, we test the predictive ability of our textual measures by comparing the out-of-sample area under the curve (AUC), which is a common measure for prediction quality when dealing with a binary dependent variable. The empirical investigations confirm that including information based on both sections significantly raises the AUC for unseen data by 2.96 (RoBERTa) and 3.26 (Longformer) percentage points when compared to a model based purely on financial controls.

Third, we examine firms' stock performance based on estimated takeover probabilities. Based on the findings from the prior literature that targets in corporate acquisitions receive large announcement returns (see, e.g., Bhagat, Dong, Hirshleifer, and Noah (2005), Betton, Eckbo, and Thorburn (2008), Eckbo (2014)) and that takeover sensitivity is positively associated with stock returns (Cremers et al. (2009), Easterwood, Netter, Paye, and Stegemoller (2024)), we hypothesize that improved predictions should allow us to form a trading strategy that yields positive excess returns. Our empirical analyses based on non-parametric portfolio sorts support this conjecture. Specifically, we find that firms with a high estimated takeover probability outperform low-takeover-probability firms by a monthly alpha of 0.85% (RoBERTa) and 0.60% (Longformer) after controlling for relevant risk factors, a non-text-based takeover vulnerability factor (Easterwood et al. (2024)), and accounting for transaction costs. The monthly outperformance is statistically significant for RoBERTa and marginally significant for Longformer. Our findings suggest that markets fail to incorporate takeover sensitivities fully into stock prices, enabling a trading strategy that exploits financial disclosures to yield positive excess returns.

Fourth, we study the extent to which the inclusion of text-based takeover predictions enables the ex-ante identification of individual takeover targets in the out-of-sample period. To do so, we impose a strict binary takeover classification based on an expanding window of past takeover

probabilities. Our analyses confirm that the share of correctly identified targets is significantly higher when textual information is incorporated in takeover predictions.

Then, we explore how and why the inclusion of textual information can contribute significantly and strongly to the predictability of M&A targets. Specifically, we employ two text-based analyses to shed light on the workings of our text-based predictions. First, we turn to topic modeling, a statistical approach used to uncover latent topics in text, to identify whether there are systematic differences in the underlying topics between firms with high and low estimated takeover probability. Notably, information on firms' operations and their finances varies strongly with RoBERTa's takeover predictions. Second, we employ novel methods in explainable artificial intelligence. Specifically, we implement effect attribution based on layer integrated gradients to quantify the link between tokenized texts and predictions by analyzing how different permutations of the input affect the output (Sundararajan, Taly, and Yan (2017), Janizek, Sturmfels, and Lee (2021)), which enables the identification of context-specific phrases with the strongest impact on our text-based predictions. Aggregating the most impactful phrases documents that product- and operations-specific information are most relevant in the BD section, whereas firm-specific capabilities, in particular with respect to information technology, dominate in the MDA section. We also find some evidence that discussions of firms' operations and finances affect the takeover exposure.

Our contribution to the literature is threefold. First, we add to the emerging literature that applies advances in machine learning in conjunction with unstructured textual data to answer open questions in finance. For example, researchers have employed machine learning to extract readability (Loughran and McDonald (2014)) and sentiment (Frankel, Jennings, and Lee (2022)) from financial disclosures, to measure corporate culture (Li, Mai, Shen, and Yan (2021)), firm

complexity (Loughran and McDonald (2024)), and financial constraints (Hoberg and Maksimovic (2015)), and to predict credit defaults (Kriebel and Stitz (2022)) and director performance (Erel et al. (2021)). More specifically, Routledge et al. (2017) and Katsafados, Androutsopoulos, Chalkidis, Fergadiotis, Leledakis, and Pyrgiotakis (2021) have incorporated information from 10-K filings into M&A target prediction by using word dictionaries and simple phrases, respectively, providing important evidence of an improved prediction performance. We build on these findings and implement two significantly more advanced machine learning algorithms, RoBERTa and Longformer, which are able to grasp the contextual meaning of words and are thus well-suited for the task of natural language processing. Moreover, we address the lack of interpretability of the baseline predictions from transformer models by analyzing RoBERTa's predictions and implementing state-of-the-art advances in explainable artificial intelligence.

Second, we provide novel evidence on the identification of M&A targets. Importantly, employing such transformer neural networks represents a methodological contribution as the machine learning algorithms enable a sophisticated incorporation of takeover-related information from Form 10-K filings. Moreover, based on the improved prediction through RoBERTa and Longformer, we are also able to contribute to our understanding of the identification of M&A targets. While Palepu (1986) has established that the prediction accuracy of target identification is generally low, a strand of literature has identified factors that influence the probability to be targeted. For example, managerial voting rights (Mikkelsen and Partch (1989)), asset tangibility (Ambrose and Megginson (1992)), board composition (Shivdasani (1993)), product-market synergies (Hoberg and Phillips (2010)), corporate pension plans (Cocco and Volpin (2013)), firm risk (Schneider and Spalt (2017)), stock prices (Edmans, Goldstein, and Jiang (2012)), and anti-takeover provisions (Cuñat, Giné, and Guadalupe (2020), John, Kadyrzhanova, and Lee

(2024)) have been shown to determine the likelihood of receiving a takeover offer. Employing a novel methodology in the M&A context, we show that business operations and firm-specific capabilities, identified from Form 10-K filings, drive takeover likelihoods. Incorporating text-based predictions significantly uplifts our ability to identify M&A targets.

Third, our findings contribute to the overarching question of which function the market for corporate control serves. The evidence based on textual information complements several existing theories on target selection in M&A deals. (1) The importance of industry information for takeover probabilities is consistent with the notion that technological shocks cause industry-clustered merger waves (see, e.g., Mitchell and Mulherin (1996), Harford (2005), Ahern and Harford (2014)). (2) The relevance of firms' operations and capabilities is in line with acquirers seeking specific assets (Rhodes-Kropf and Robinson (2008)). (3) Textual information relating to governance and potential financial struggles points towards the disciplinary role of the market for corporate control as assets are reallocated to a more efficient management (see, e.g., Gort (1969), Jovanovic and Rousseau (2002)).

The remainder of this paper is structured as follows. Section II details our data sources, while Section III discusses the empirical design and methodology. Section IV presents results for the prediction of M&A targets. Section V documents our explainable artificial intelligence analyses. Section VI concludes.

II. DATA

Sources. The basis of our sample is the universe of Compustat firm-years between 1993 and 2020. We require the availability of Form 10-K's BD and MDA sections within the SEC's EDGAR system to analyze the information content of textual disclosures with respect to takeovers. To obtain

the 10-K filings, EDGAR is automatically scraped for all filings titled "10-K", "10-KSB", "10-K405", and "10-KSB40". We extract the BD and the MDA sections from the annual financial reports and delete the section titles. Moreover, we limit our sample to filings with existing text data by dropping all observations with fewer than 50 words. Since the models we use are trained on natural text, we refrain from any text preparation apart from deleting line breaks and superfluous additional spaces. We supplement the textual data with corresponding accounting information obtained from Compustat. Data on M&A deals is obtained from LSEG SDC. We consider deals with publicly listed targets in the United States and require that the bidder seeks more than 50% of the target's shares. Recapitalizations, repurchases, and rumored deals are excluded. Additional data on horizontal and vertical 10-K Text-based Network Industry Classifications (TNIC) by Hoberg and Phillips (2010), Hoberg and Phillips (2016), and Frésard et al. (2020) are drawn from the [Hoberg-Phillips Data Library](#). We use data on industrial production from the Federal Reserve Bank of St. Louis. Lastly, for our return analyses, we use the Fama and French (1993) factors from [Kenneth R. French's website](#), profitability and investment factors from [the Hou, Xue, and Zhang \(2015\) q-factor data repository](#), the Pástor and Stambaugh (2003) liquidity factor from [Robert Stambaugh's website](#), the rolling window takeover vulnerability factor from [Sara Easterwood's website](#), and stock returns obtained from CRSP.

Main Variables. Our central dependent variable is the *Takeover Dummy*, which is equal to 1 if a firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise. If a firm files a new annual financial report within one year, we assign the takeover to the subsequent 10-K filing, such that the most recent, publicly available information is always utilized. The main explanatory variables for our analyses are the estimated takeover likelihoods derived by RoBERTa

(Longformer) from the BD and the MDA: *RoBERTa BD* and *RoBERTa MDA* (*Longformer BD* and *Longformer MDA*). Section III details how these takeover likelihoods are estimated.

Financial Controls. We define further control variables in line with the prior M&A literature (see, e.g., Baker, Pan, and Wurgler (2012), Schneider and Spalt (2017)). We briefly introduce all variables in the following and report additional detail in Table A1 in the Online Appendix.

Importantly, we use accounting characteristics from the last fiscal-year end prior to the M&A announcement to ensure that no forward-looking information is included (see, e.g., Masulis, Wang, and Xie (2007), Golubov, Yawson, and Zhang (2015)) and all explanatory variables are lagged relative to our dependent variable.¹ We calculate return on equity (*RoE*) as the three-year average of the ratio of net income (Compustat: NI) to shareholders' equity (Compustat: SEQ), *Market Cap* as the natural logarithm of the product of the stock price at the fiscal year-end (Compustat: PRCC_F) and shares outstanding (Compustat: CSHO), and the market-to-book (*MB*) ratio by deflating the (non-logarithmized) *Market Cap* with shareholders' equity (Compustat: SEQ). Similarly, the price-earnings (*PE*) ratio is defined as the ratio of the market capitalization to the firm's net income (Compustat: NI), while *Leverage* is long-term debt (Compustat: DLTT) over shareholders' equity (Compustat: SEQ). *Tangible Assets* is calculated as the ratio of property, plant, and equipment (Compustat: PPENT) to total assets (Compustat: AT) and *Sales Growth* as the three-year average of the percentage change in revenue (Compustat: SALE) since the prior fiscal year. We use the three-year average of the sum of cash and cash equivalents (Compustat: CHE) and total receivables (Compustat: RECT) over total assets (Compustat: AT) to define *Liquidity*. A

Growth-Resource-Mismatch dummy variable is set to 1 if a firm is classified as low

¹Assume for instance that a firm files its annual financial report for the fiscal year end of December 31, 2000 on March 30, 2001. Then, we use the textual and financial data as of December 31, 2000 to examine whether the firm receives a takeover offer within 365 days after March 30, 2001. Thereby, we ensure that explanatory variables are available for public dissemination and lagged relative to potential takeover offers.

Sales Growth, low *Leverage*, and high *Liquidity* or, conversely, high *Sales Growth*, high *Leverage*, and low *Liquidity*, based on the respective annual median. Moreover, we define *Takeover Activity* as a dummy variable equal to 1 if we record at least one merger in the TNIC-3 peer group in the previous year. We include industry dummies according to the Fama and French (1997) 12 industry portfolio classification.

Industry & Demand Controls. Motivated by the evidence that firms' productivity and demand shocks affect the market for corporate assets (Maksimovic and Phillips (2001)), we employ a number of industry and demand variables in further tests. *Industry Profit* is calculated as the mean profitability of a firm's industry peers, defined as the ratio of net income (Compustat: NI) to sales (Compustat: SALE), based on the TNIC-3 database by Hoberg and Phillips (2010) and Hoberg and Phillips (2016). *Low Profit* is a dummy variable equal to 1 if the firm's profitability is below that of its industry peers, and 0 otherwise. We also include *TNIC HHI*, which is the Herfindahl-Hirschman index provided by Hoberg and Phillips (2016). *Downstream Sales Growth* is defined as the mean growth rate of the average sales growth downstream of the firm according to the vertical TNIC database by Frésard et al. (2020) in the previous year. Lastly, we compute *Industrial Production* as the growth rate of the United States' industrial production over the quarter prior to the respective fiscal year end. We winsorize all continuous control variables at the 1% and the 99% levels.

Data Split. We split our data into a training set and a test set, such that the first 19 fiscal years (1993 – 2011) are used for training the transformer models RoBERTa and Longformer.² Thus, the most recent years (2012 – 2020) are assigned to the test set and used to evaluate out-of-sample

²Coverage of 10-K filings in EDGAR in the first years of our training sample is incomplete (see Table A2 in the Online Appendix). Larger and more established firms are more likely to have such early filings available. Since such larger firms are less likely to be takeover targets, the sample in the early period may bias takeover predictions downward. To test the influence of these early years, we exclude the first four years of the training sample due to lower coverage and start the training (test) sample in 1997 (2016). As shown in Table A3 in the Online Appendix, prediction quality remains qualitatively unchanged such that we use all available observations for our training sample.

model performance.³ We opt for a sample split along the time axis instead of a random assignment into the training and test samples for two main reasons. First, this procedure ensures that we do not incorporate any forward-looking information that could potentially distort the predictive nature of our text-based takeover forecasts. Specifically, since we also use our forecasts to predict subsequent returns, any inclusion of subsequent 10-K filings might lead to information leakage, which in turn might taint our results. Second, by relying strictly on past textual disclosures, we provide a conservative estimate of the 10-K's takeover-relevant information content. While structural breaks in the textual determinants of takeover likelihoods after the training set will thus not be picked up, putting a strain on our model's predictive performance, our text-based predictions strictly exclude any forward-looking information. Consequently, our AUC estimates might also be interpreted as lower bound estimates. We further limit the test dataset to observations without missing financial control variables, resulting in 19,665 observations in the test set.

[Insert Table 1 approximately here]

Summary Statistics. Table 1 reports descriptive statistics for both the dependent variable and the control variables used in our main analysis. On average, 4.06% of all firm-year observations are targeted by a corporate acquisition.⁴ The median firm within our sample has a market capitalization of \$560 million and exhibits *MB* and *PE* ratios of 1.90 and 9.41, respectively. The mean *Leverage* equals 57.77%, while 22.10% of the firms' assets are *Tangible Assets* and 39.00% are classified as *Liquidity*. The median *Sales Growth* equals 12.47% with a substantially larger mean driven by a very small number of extreme observations. We provide additional summary statistics on the temporal distribution of 10-K filings in Table A2 in the Online Appendix.

³While we limit our sample time period to the end of fiscal year 2020, the corresponding last filing date is in March 2021. Since we consider merger offers within 365 days after the filing date, the last considered takeover attempt is announced in March 2022.

⁴We refrain from balancing the dependent variable via oversampling, since both our transformer models are based on text and the logistic regressions estimate probabilities across the complete range as indicated in Figures 2 and 3. We provide additional evidence based on a balanced sample in the Online Appendix.

III. METHODOLOGY

Since textual data is very complex and diverse, using machine learning to automatically extract relevant information is no easy task. Many basic approaches focus on counting the occurrences of specific words or word combinations, thereby losing important information such as word-order and context. With the rise in data availability and computing power, deep neural networks have surfaced as a strong contender in automatically analyzing and understanding text. In general, an artificial neural network consists of layers of neurons that transmit information from the input layer through a set of hidden layers to the output layer, thereby gradually aggregating multi-dimensional information (LeCun, Bengio, and Hinton (2015)). Such a model can represent the semantic meaning of words in an n-dimensional vector space (i.e., create a word embedding that translates text into numbers). High-dimensional word embeddings are thus a complex representation of used language that can be automatically processed by neural networks. For our application, the BD and the MDA sections from 10-K filings serve as inputs with the *Takeover Dummy* as the relevant output to predict.

We use the state-of-the-art transformer models RoBERTa (“a robustly optimized BERT pretraining approach”; Liu et al. (2019)) and Longformer (Beltagy et al. (2020)) as binary classifiers to predict takeovers based on text from the 10-K filings (separately for the BD and the MDA sections). Transformer models are a special class of deep neural networks that use a transformer-encoder architecture with self-attention (Vaswani et al. (2017)). The self-attention concept implies that the model is able to process text cohesively rather than sequentially to optimize the embedding of each word by considering the complete sentence. In contrast to more traditional embeddings, the precise vector depends on the surrounding words used in the sentence. This enables RoBERTa and Longformer to capture context-specific differences in meaning within the

embedding. Since transformer models require training on an exceptionally large number of parameters, their usefulness depends on transfer learning (i.e., we employ a pre-trained model and use the text data from 10-K filings to fine-tune the model). Both RoBERTa and Longformer are adjusted variants of the very successful BERT (“bidirectional encoder representations from transformers”; Devlin et al. (2019)) model that achieved breakthrough results in benchmark tasks for processing and understanding text. In contrast to BERT, RoBERTa is built on an adjusted and more intensive training approach (i.e., focusing on masked language modeling, using a dynamic masking approach, and training on more data and longer sentences), resulting in an even better performance concerning the benchmark tasks SQuAD (Rajpurkar et al. (2016), Rajpurkar et al. (2018)) and GLUE (Wang et al. (2018)). Longformer adjusts BERT’s attention mechanism such that it scales linearly with sequence length (instead of quadratically as BERT and RoBERTa) and is thus able to process longer texts (Beltagy et al. (2020)).

We use the Adam optimizer — an extended variant of stochastic gradient descent⁵ — when fine-tuning our RoBERTa and Longformer models and additionally consider the set of hyperparameters as stated in Table 2.⁶ We conduct a complete search of the respective hyperparameter space using three-fold cross-validation on our training dataset.⁷ To account for the rare occurrence of takeovers in our sample, we use stratified sampling within the three-fold cross-validation, i.e., we ensure that each class is proportionally represented in training and validation samples.⁸ The optimal hyperparameter combination is chosen based on the lowest

⁵Stochastic gradient descent is an iterative method of optimization. Compared to original gradient descent, it uses only an estimate of the actual gradient to reduce computational effort.

⁶Due to computational limitations, we only consider a reduced hyperparameter space when fine-tuning Longformer.

⁷Three-fold cross-validation describes a process used to train robust models and estimate their out-of-sample performance. Specifically, the training data is subdivided into three sets. In turn, each of the sets is used to validate the performance of the model trained on the other two sets.

⁸Additionally, we implement an alternative fine-tuning approach based on cost-sensitive training such that we penalize misclassifications of actual takeover targets disproportionately. We report the results, which are qualitatively equivalent to our main analyses, in Table A4 and Figure A1 in the Online Appendix.

arithmetic mean of binary cross-entropy (log-loss) across the three folds. We subsequently use the arithmetic mean of the three models' predictions to evaluate the performance on unseen data from the test dataset.

[Insert Table 2 approximately here]

We use logistic regressions to combine the text-based predictions with the defined control variables to analyze the improvement in prediction performance when also considering structured data. This allows us to compare the relation between text-based predictions and actual takeovers to those of the financial variables and takeovers. The logistic regressions are conducted using a rolling-regression approach based on the test dataset (i.e., we separately run a logistic regression for out-of-sample prediction of each year in the test sample starting from 2013). These logistic regressions are estimated on all preceding data from the test set (e.g., the logistic regression used to predict 2017 is estimated on data from 2012 to 2016).⁹ Importantly, we do not retrain the machine learning models in the out-of-sample periods and keep all their parameters fixed. The predictions resulting from these annual logistic regressions are used together to calculate the AUC for the test sample to evaluate out-of-sample performance. AUC uses the receiver operating characteristic (ROC) curve to compare true-positive rates against false-positive rates for all decision thresholds. A good classifier is characterized by a favorable trade-off between accepting additional false positives to achieve more true positives. The AUC of a model that randomly produces predictions is equal to 0.5, while a model that perfectly predicts every observation scores 1.0.

[Insert Figure 1 approximately here]

Figure 1 summarizes our approach. For our textual predictions, we partition the observations along the time axis into a training and a test dataset. RoBERTa and Longformer are fine-tuned solely on the training dataset such that we derive text-based predictions in the test dataset

⁹Since we use year-dummies in a rolling regression approach, we assume that the coefficient for the year-dummy for which the model is required to make out-of-sample predictions is equal to the coefficient of the final year-dummy in the training data.

on unseen data (*RoBERTa BD*, *RoBERTa MDA*, *Longformer BD*, *Longformer MDA*). Then, we use these text-based predictions as input in logistic regressions to examine their incremental predictive power relative to the considered financial variables. In particular, we conduct eight separate logistic regressions for each year of the test dataset and successively expand the estimation window of the logistic regressions. For instance, predictions for 2013 are based on logistic regressions that incorporate observations from 2012 only, while predictions for 2014 are based on observations from 2012 and 2013 (see also Figure 1).

IV. EVALUATING TEXTUAL PREDICTIONS

In this section, we test whether the information extracted from 10-K filings via deep learning algorithms helps to identify takeover targets. First, we present multivariate regression evidence that the contribution of textual information to takeover predictions is statistically significant and economically large. Second, we present an out-of-sample AUC-based model comparison. Third, we demonstrate that a trading strategy based on our complete model’s predictions earns excess returns. Fourth, we evaluate the practical predictive performance of our models in the out-of-sample period by analyzing classification accuracy, precision, recall, and F_1 -Scores based on estimated probability thresholds.

A. In-Sample Regression Evidence

To test whether the 10-K filings contain relevant information for the prediction of subsequent M&A targets, we present marginal effects from logistic regressions with the *Takeover Dummy* as the dependent variable in Table 3. We report results separately for predictions derived from RoBERTa and Longformer in Panels A and B, respectively. While specifications (1) and (2) test the impact of the BD and the MDA without any controls separately, specification (3) displays

their joint impact. Specifications (4) to (7) (additionally) include financial controls as well as year- and industry-fixed effects. The reported regression results correspond to the final rolling regression (i.e., it was conducted using the years 2012 to 2019).¹⁰ Thus, while the regression is conducted on the previously unused test sample, the reported marginal effects correspond to the data on which the regression was run (i.e., the regression evidence is in-sample).

[Insert Table 3 approximately here]

The results displayed in Table 3 highlight that text-based predictions are a statistically significant and economically large predictor of takeovers across all specifications. Focusing on Panel A, *RoBERTa BD* displays an average marginal effect between 0.38 and 0.42. Incorporating additional control variables and fixed effects barely affects the average marginal effect of *RoBERTa BD*, leaving it statistically significant at the 1% level. The predictive effect is also economically large as a one-standard-deviation (i.e., 0.0249) increase in *RoBERTa BD* is associated, on average, with an increase in the estimated takeover probability of 0.95 ($= 0.3808 \times 0.0249 \times 100$) percentage points in the most rigorous specification, which corresponds to 4.81% of the dependent variable's standard deviation.

In contrast, the average marginal effect of *RoBERTa MDA* is highly sensitive to the simultaneous inclusion of *RoBERTa BD*. Specifically, *RoBERTa MDA* is statistically significant and economically large when it is the sole predictor of *Takeover Dummy*, and remains so with an average marginal effect of 0.29 when control variables and fixed effects are included. Specification (6) indicates that a one-standard-deviation increase in *RoBERTa MDA* is, on average, associated with an increase in the takeover likelihood of 0.74 ($= 0.2875 \times 0.0258 \times 100$) percentage points, which equals 3.75% of the standard deviation of the *Takeover Dummy*. Including *RoBERTa BD*, however, turns the average marginal effect of *RoBERTa MDA* statistically insignificant.

¹⁰17,967 observations from 2012 to 2019 serve as input for the logistic regressions, which we evaluate based on the takeover offers across the 1,698 observations in 2020.

Consequently, the results of the logistic regression model indicate that RoBERTa is able to extract information that is relevant to the prediction of takeover likelihoods from both the BD and the MDA. The evidence indicates that the two 10-K sections may contain partially overlapping information. This interpretation is consistent with the strongly positive correlation of *RoBERTa BD* and *RoBERTa MDA* of 0.66 reported in Table A5 in the Online Appendix.¹¹

Panel B of Table 3 reports results from equivalent specifications with *Longformer BD* and *Longformer MDA*, respectively, as the main explanatory variables. While the marginal effect of *Longformer MDA* is significantly positive and varies between 0.54 and 0.75 across all specifications, the effect size and statistical significance of *Longformer BD* drop upon the inclusion of *Longformer MDA*. Again, the correlation between *Longformer BD* and *Longformer MDA* of 0.68 reported in Table A5 in the Online Appendix points toward substantial overlap in the takeover-relevant information extracted from the BD and the MDA. Nonetheless, both *Longformer BD* and *Longformer MDA* remain statistically significant in the most rigorous specification at the 5% and 1% level, respectively. Here, a one-standard-deviation increase in *Longformer BD* (*Longformer MDA*) is, on average, associated with an increase in the takeover likelihood of 0.44 (0.86) percentage points, which corresponds to 2.23% (4.36%) of the dependent variable's sample standard deviation.

[Insert Table 4 approximately here]

We provide further evidence on the economic size of text-based takeover predictions in the Online Appendix: Tables A6 and A7 report regression coefficients from standardized linear probability models with predictions derived from RoBERTa and Longformer, respectively. The magnitude of *RoBERTa BD* in Table A6 exceeds all financial control variables. *RoBERTa MDA* also

¹¹Several of the control variables significantly affect the takeover probability in line with our expectations in at least some of the specifications. The negative marginal effect of *Market Cap* points toward a preference for smaller targets. Moreover, bidders more commonly target firms with higher *Leverage*, higher *Tangible Assets*, and (in one specification) higher *Liquidity*. *Sales Growth* is significantly negative in all specifications.

exhibits a standardized coefficient that exceeds all other variables as long as *RoBERTa BD* is not included in the regression specification. Table A7 confirms that textual predictions strongly contribute to the estimation of takeover likelihoods as both *Longformer BD* and *Longformer MDA* exceed the economic magnitude of all control variables whenever they are included separately. Our regression evidence emphasizes the economic significance of the utilized text data.

In addition to the Compustat-based variables that comprise our financial controls, we also examine whether RoBERTa’s and Longformer’s predictions explain subsequent takeover offers beyond additional industry and demand variables in Table 4. Importantly, several of these variables build on (takeover-relevant) information on horizontal and vertical industrial relatedness extracted by Hoberg and Phillips (2010) and Frésard et al. (2020) from 10-K filings. The results in Table 4 confirm the robustness of our previous insights: takeover predictions by RoBERTa and Longformer remain statistically significant and economically important predictors of subsequent takeovers.¹²

B. Out-of-Sample Model Comparison

In addition to the regression evidence in Subsection A, we also conduct an out-of-sample model comparison. Specifically, we calculate the AUC for takeover predictions on unseen data for different sets of predictors akin to the preceding subsection. While a model that solely incorporates fixed effects and financial controls serves as a benchmark, we employ different combinations including *RoBERTa BD*, *RoBERTa MDA*, *Longformer BD*, and *Longformer MDA* to investigate the contribution of textual disclosures to the prediction of M&A targets.

[Insert Table 5 approximately here]

Table 5 compares the prediction performance on unseen data of different logit model specifications that all include financial controls as well as industry- and year-fixed effects. The

¹²We document in Table A8 in the Online Appendix that our results are robust to alternative definitions of industry-level productivity.

resulting AUC values are compared to a benchmark model using only financial control variables and fixed effects, which we report in the first row. Including RoBERTa’s text-based predictions using the BD (MDA) sections results in an out-of-sample AUC that is 3.03 (1.43) percentage points higher than that of the benchmark model. This difference in AUC is highly statistically significant with a z-value of 4.63 (3.19). When combining both text-based predictions from RoBERTa with control variables and industry- and year-fixed effects in a single logit model, the resulting AUC uplift is 2.96 percentage points with a corresponding z-value of 4.54. Figure 2 depicts the AUC uplift by showing the underlying difference in ROC curves.

[Insert Figure 2 approximately here]

Additionally, Table 5 also reports the AUC values based on Longformer’s text-based predictions. Incorporating *Longformer BD (Longformer MDA)* alongside the financial controls and fixed effects raises the AUC by 1.68 (3.47) percentage points, which is statistically significant with a z-value of 2.69 (4.50). Combining Longformer’s predictions from both the BD and the MDA with the control variables and fixed effects raises the AUC relative to the benchmark model by 3.26 percentage points with a z-value of 4.08. Figure 3 plots the corresponding ROC curve.

[Insert Figure 3 approximately here]

The results displayed in Table 5, Figure 2, and Figure 3 yield three main insights. First, and most importantly, the inclusion of textual predictions raises the AUC in all specifications such that the predictability of M&A targets in out-of-sample tests improves significantly. Consequently, the machine learning algorithms are clearly able to extract takeover-relevant information from the BD and the MDA. We provide additional evidence that these AUC uplifts of the combined text specifications remain statistically significant and economically large when using (1) a more granular industry classification in Table A9 in the Online Appendix and when using (2) an alternative text-based industry classification as industry-fixed effects in Table A10.

Second, there are notable differences in the AUC generated by RoBERTa and Longformer based on the BD and the MDA, respectively. We test these differences for significance and report the results in Table A11 in the Online Appendix. While RoBERTa is able to achieve a 1.35 percentage point higher AUC based on the BD (z-value of 2.60), Longformer reaches an AUC based on the MDA that is 2.04 percentage points higher (z-value of 3.50). The difference in the complete models based on RoBERTa and Longformer of 0.30 percentage points, however, is statistically insignificant. Thus, Longformer seems to be particularly good at extracting takeover-relevant information from the MDA compared to RoBERTa. This could be ascribed to the boilerplate preface (e.g., highlighting uncertainty with respect to forward-looking statements) commonly included in MDAs before delving into economically relevant discussions. BDs, conversely, typically do not have such extensive preambles. Truncation occurs for nearly all documents under both transformer architectures,¹³ yet the relative placement of takeover-relevant information within each 10-K section could plausibly contribute to the differences in performance between RoBERTa and Longformer. In particular, the ability of Longformer to retain a substantially larger portion of the MDA is likely to be important. Conversely, RoBERTa's full self-attention may be advantageous for the BD, where relevant information tends to be concentrated up-front, such that richer token-to-token interactions may better capture contextual nuances compared to Longformer's predominantly local attention mechanism.

Third, the fact that combining information from the BD and the MDA does not increase the prediction quality substantially (either with or without financial controls, industry- and year-fixed effects), indicates that there is an overlap between the information uncovered by the machine learning algorithms from the BD and the MDA. This interpretation is supported by the large

¹³In our test sample, 98.03% and 98.74% of BDs and MDAs, respectively, exceed the 512-token input limit of RoBERTa. Likewise, 91.03% and 87.19% exceed the 4,096-token input limit of Longformer.

correlations between the prediction probabilities generated from the two 10-K sections across both transformer architectures reported in Table A5 in the Online Appendix.

We provide six additional pieces of evidence in the Online Appendix. (1) We report the AUC of text-based only predictions by year in Table A12. The results confirm that the predictions derived from the 10-K filings by RoBERTa and Longformer consistently improve M&A target predictions (all AUCs > 0.5). Moreover, a trend of diminishing prediction performance from 2012 (AUCs between 0.59 and 0.64) to 2020 (AUCs between 0.54 and 0.56) emerges as the time between prediction and training data increases. Comparing these annual AUCs with the annual AUCs based on a shifted test window starting in 2016 (Table A3) suggests that including more recent observations in the training dataset is insufficient to eliminate this performance drift over time. One potential explanation for this pattern could be that 10-K filings have become longer and less readable over time (Dyer, Lang, and Stice-Lawrence (2017)), such that it might become comparably harder to extract takeover-relevant information from more recent annual reports.¹⁴ More broadly, differences in disclosure norms and the regulatory environment between the training and test sample periods – for instance due to the implementation of Regulation G in 2003 and the JOBS Act in 2012 – may contribute to the observed performance drift. Alternatively, given the ample evidence of time-variant determinants of M&A activity (see, e.g., Mitchell and Mulherin (1996), Harford (2005)), the underlying economic drivers of takeover activity that the machine learning models seek to extract may themselves evolve over time, for instance due to shifts in target-industry composition.

(2) We provide further evidence on the out-of-sample predictability of M&A targets based on simple portfolio sorts in Table A13. Specifically, we sort observations annually into quintiles

¹⁴Descriptive analyses of the test sample confirm a clear upward trend in the BD word count from 2012 to 2020. Similarly, the Fog Index confirms that BD texts are significantly harder to read in 2020 than in 2012.

based on the predicted likelihood of being targeted and report the average *Takeover Dummy*, i.e., the likelihood of being targeted in the subsequent year. Across all models, the share of subsequent targets is significantly higher among high-takeover-probability than low-takeover-probability firms.

(3) Figures A2 and A3 in the Online Appendix document that models incorporating textual predictions continue to outperform when we incorporate industry and demand variables in the reduced sample. (4) Excluding fixed effects from the control set widens the gap between the logit models incorporating text-based predictions and the one using financial controls only (see Table A14 and Figures A4 and A5).

(5) We report AUCs of logistic regressions based on text-based predictions only in Table A15 and examine whether adding financial controls yields a better predictive performance beyond RoBERTa's and Longformer's estimates of takeover likelihood. Our results document that adding financial controls does not significantly change the prediction quality. One explanation of this pattern could be that the predictions extracted by the transformer models already contain the takeover-relevant information captured by the financial controls. Alternatively, it is also possible that part of the predictive power extracted from textual disclosures reflects firms' underlying financial condition, which is also captured by financial controls such as *Leverage* and *Liquidity*. If textual disclosures primarily reflected information already captured by conventional financial controls, the economic contribution of text-based predictions would naturally be more limited. However, the significant incremental predictive power of text-based predictions documented in Table 5 suggests that textual disclosures contain takeover-relevant information beyond conventional financial controls.

(6) In an alternative approach discussed in detail in the Online Appendix, we conduct a random sample split into a train and test set instead of a split along the time axis. Our results

confirm that RoBERTa is able to extract information from the BD of the 10-K filing that helps to identify takeover targets - even against a matched sample within the same industry-year and of similar size (see Table A20). To better understand what type of information is uncovered from both the BD and the MDA, we apply an in-depth analysis using explainable artificial intelligence methods in Section V.

C. Portfolio Evidence

We further seek to explore the economic implications of the improved predictions of subsequent M&A targets. Specifically, we examine the relationship between estimated takeover probabilities and subsequent stock returns. Our motivation to examine the effect of takeover exposure on asset prices is twofold. First, large abnormal returns of targets around takeover announcements are well-documented as targets extract large offer premiums over current market valuations from bidders (Andrade et al. (2001)). Second, both Cremers et al. (2009) and Easterwood et al. (2024) provide evidence that the likelihood of being targeted in an M&A deal is positively associated with expected returns, even in the absence of realized takeovers. Hence, a positive association between our model’s takeover predictions and subsequent stock returns would serve as proof-of-concept for our predictions and illustrate the associated economic relevance.

[Insert Table 6 approximately here]

Therefore, we conduct a portfolio sort based on our complete model’s takeover predictions. We start with observations within our test period from March 2013 to February 2022 and limit our sample to firms with common ordinary shares listed on the NYSE, AMEX, or NASDAQ. Additionally, we exclude stocks with share prices lower than \$5 (Jegadeesh and Titman (2001)). Then, we define monthly, cross-sectional portfolio breakpoints based on estimated takeover probabilities for quintile sorts.

We report the central results from portfolio sorts based on takeover probability estimates in Table 6, with Panels A and B reporting the extreme legs of a quintile sort based on RoBERTa and Longformer, respectively.¹⁵ Since takeover targets tend to be relatively smaller stocks, we focus on equal-weighted returns. We account for transaction costs by estimating monthly portfolio turnover from changes in portfolio weights. Transaction costs are based on bid-ask spreads estimated following Abdi and Rinaldo (2017) and are computed as half the bid-ask spread multiplied by portfolio weight changes. Net returns are obtained by subtracting these costs from gross returns. Gross returns are reported in Table A16 in the Online Appendix. To evaluate whether the return spread between high- and low-takeover probability portfolios reflects compensation for common sources of systematic risk, we estimate portfolio excess returns (α) relative to a multi-factor benchmark model including the Fama and French (1993) three-factor model, momentum (Jegadeesh and Titman (1993), Carhart (1997)), profitability, investment (Hou et al. (2015)), liquidity (Pástor and Stambaugh (2003)), and takeover vulnerability (Cremers et al. (2009), Easterwood et al. (2024)).¹⁶

According to Panel A of Table 6, high-takeover probability stocks earn 0.84% higher returns per month than low-takeover probability stocks on an equal-weighted basis. Examining excess returns reveals that high-takeover-probability stocks outperform low-takeover-probability stocks by 0.85% per month. The difference portfolio's outperformance is statistically significant with a t-value of 2.69 and is economically large as it implies annual excess returns of 10.69%.

¹⁵We only report the extreme legs of the portfolio sort since we do not claim that takeover likelihood is necessarily a return factor, which would imply a (linear) relationship across all portfolio buckets. Rather, we focus primarily on the high-takeover probability portfolios and use the low-takeover probability portfolios as natural counterfactual.

¹⁶This specification accounts for the possibility that firms that are assigned high takeover likelihoods by RoBERTa and Longformer differ systematically from other firms along several dimensions. For instance, takeover targets may be less liquid and exhibit different profitability and investment characteristics that are associated with expected returns. Moreover, adjusting for the takeover-vulnerability factor of Easterwood et al. (2024) allows us to examine whether transformer-extracted takeover predictions generate excess returns beyond established models of takeover likelihood.

Panel B of Table 6 reports the corresponding results of a sort based on the logistic regressions including Longformer predictions, financial controls, and fixed effects. The results confirm that high-takeover probability stocks outperform low-takeover probability stocks with a monthly raw return difference of 0.53%, which translates to an excess return of 0.60%. The monthly outperformance is marginally significant with a t-value of 1.88.¹⁷ Overall, the portfolio evidence, in particular based on RoBERTa’s predictions, documents outperformance of the difference portfolio constructed based on transformer-extracted takeover likelihoods. Importantly, the corresponding portfolio alphas remain positive and at least marginally significant after accounting for transaction costs and common sources of systematic risk such as liquidity and takeover vulnerability. These findings underline the economic relevance of takeover predictions that incorporate both financial and textual information.

D. Identifying Individual Takeover Targets

Our prior analyses consistently document that the incorporation of textual information via machine learning algorithms significantly and meaningfully improves the prediction of takeover targets out-of-sample. The predicted takeover likelihoods by firm, however, are low, reflecting the inherent difficulty of predicting corporate takeovers (Palepu (1986)) and the low base rate of takeover offers in our sample (see average takeover rate of 4.06% in our test sample). While we are able to extract meaningful variation in text-based predicted takeover likelihoods based on this naturalistic prediction set-up (see Tables 3, 5, and 6), it is unclear to which extent the model is able to correctly identify individual takeover targets.

¹⁷Motivated by the declining AUC over time (see Table A12), we repeat the analyses of Table 6 separately for the first and the second part of the return sample period. The results reported in Table A17 do not suggest that the documented return predictability is confined to the earliest years of the sample period.

To test the ability of our model to improve the prediction of individual takeover targets, we extend our approach. Specifically, we start with the takeover estimates derived by the logistic regressions based on different models as depicted in Figure 1. Then, we estimate thresholds for binary takeover classifications (i.e., forcing the model to label takeover targets in the subsequent year rather than estimating takeover likelihoods) such that the estimated threshold optimizes the F_1 -Score separately by model. We report these thresholds by year in Table A18 in the Online Appendix. As in our logistic regression approach, we ensure that the analysis is purely out-of-sample by estimating thresholds based on an expanding window of all prior years of the test sample. For example, the threshold in 2013 is estimated on 2012 data and the threshold in 2020 is estimated on data from 2012 to 2019. This enables us to compute sensible performance metrics for prediction quality beyond AUC, based on true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (see, e.g., Acikalin et al. (2026)). In particular, we define the following additional performance metrics:

$$(1) \quad Precision = \frac{TP}{TP + FP},$$

$$(2) \quad Recall = \frac{TP}{TP + FN},$$

$$(3) \quad F_1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall},$$

$$(4) \quad Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

[Insert Table 7 approximately here]

Table 7 reports F_1 -scores, accuracy, precision, and recall for a model based on fixed effects and financial controls, models based on textual data only, and models that combine financial controls and textual data separately for RoBERTa and Longformer, respectively. Three important patterns emerge in the data. First, F_1 -scores are low overall, varying between 8.70% and 10.50%. Second, using takeover predictions derived from textual data *only* yields F_1 -scores (insignificantly) higher than those of the model combining several established financial control variables and fixed effects. The corresponding uplifts in F_1 -scores of 0.75 to 0.90 percentage points are primarily driven by sizable improvements in recall values by 6.46 to 7.97 percentage points, which represent increases of 44.37% to 54.74% in the share of subsequent takeovers that are correctly identified when using textual predictions only. At the same time, the level of precision based on textual predictions is slightly below (-0.10 percentage points) that of the model based on fixed effects and financial controls. Third, combining textual predictions with financial controls leads to statistically significant increases in F_1 -scores of 1.80 percentage points for RoBERTa (significant at the 5% level based on firm-clustered bootstrapping; see Petersen (2009)) and of 1.74 percentage points for Longformer (significant at the 5% level) relative to the model based on financial variables only. Interestingly, the higher F_1 -scores of the combined models for RoBERTa and Longformer reflect two distinct mechanisms. The improvement from including RoBERTa's predictions seems to be driven more strongly by an increase in recall (+7.56 percentage points) than precision (+0.68 percentage points), while the inclusion of Longformer's predictions is driven more strongly by an increase in precision (+1.67 percentage points) than recall (+0.96 percentage points).

[Insert Table 8 approximately here]

Table 8 complements the aggregate classification results reported in Table 7 by reporting prediction performance metrics across deciles of estimated takeover probability. We separately sort the test sample's firm-year observations into deciles of estimated takeover likelihoods based on the

Controls model (Panel A), the *RoBERTa BD*, *RoBERTa MDA*, & *Controls* model (Panel B), and the *Longformer BD*, *Longformer MDA*, & *Controls* model (Panel C).¹⁸ For each decile, we report the number of realized takeovers, the realized takeover rate, the number of true positives, the F_1 -score, and the classification accuracy.

The results in Table 8 enable us to study whether the incremental value of including textual information is primarily concentrated in the highest predicted-probability observations. Compared to the *Controls* model, both RoBERTa and Longformer assign more realized takeover targets to the highest predicted-probability decile such that takeover rates are slightly higher. Moreover, both transformer-based models identify more true positive predictions in the highest decile compared to the benchmark model. Consequently, the F_1 -scores in the highest decile based on RoBERTa (14.12%) and Longformer (14.37%) exceed the *Controls* model (11.46%). While this improvement is most pronounced in the highest decile for both transformer architectures, RoBERTa continues to identify additional true positives (and thus non-zero F_1 -scores) in deciles 9 through 7, whereas Longformer’s gains are largely confined to the top decile. Hence, textual information improves prediction quality primarily at the upper end of the takeover-likelihood distribution. Overall, the analyses presented in Tables 7 and 8 suggest that text-based predictions are an important complement to established takeover determinants when identifying firms that are more likely to be targeted subsequently.

V. EXPLAINING TEXTUAL PREDICTIONS

The improved takeover predictability based on textual disclosures shown in Section IV raises the question of where the predictive power regarding future takeovers in 10-K filings stems

¹⁸We additionally report corresponding results based on the *RoBERTa BD* & *RoBERTa MDA* and the *Longformer BD* & *Longformer MDA* models in Table A19 in the Online Appendix.

from. In this section, we examine which takeover-relevant information RoBERTa¹⁹ is able to extract from both the BD and the MDA using two different approaches. First, we use a latent Dirichlet allocation (LDA) topic model to identify topics that are associated with low vs. high predicted takeover probabilities. Second, we use an explainable artificial intelligence approach to calculate feature attributions with integrated gradients to identify which words have a strong impact on the predictions made by the model.

A. Topic Modeling

Topic models such as LDA treat documents as a combination of topics with each topic having a certain probability of producing a certain combination of words (Blei, Ng, and Jordan (2003)). We treat each BD and MDA text as an individual document and analyze the predominance of specific topics according to LDA within quintiles based on the predicted takeover probability. Therefore, we calculate the arithmetic mean of the estimated topic proportions across documents within each quintile. We remove the most common 500 words for both the BD and the MDA to exclude generic terms across documents. The main issue when using topic models is that the resulting topics are only described by the most dominant words (i.e., the topics are latent and not specifically named). Thus, we manually assign names to the topics found by the model. Figure 4 (6) shows the resulting named topics and the top 10 words that make up the topics for the BD (MDA). Figure 5 (7) depicts how these topics are distributed across BD (MDA) quintiles.

[Insert Figure 4 approximately here]

Figure 4 lists the four topics uncovered by LDA from the BD, while Figure 5 depicts how the topic prevalence varies across the five portfolios built on the takeover probabilities estimated by RoBERTa. Topic 1 is mainly characterized by words related to firms' physical assets and operations

¹⁹In the following section, we focus on RoBERTa only. We made this decision in the interest of brevity and due to excessive computational effort required to duplicate this analysis using Longformer.

("vehicle", "reserves", "unique", "parts", and "cells"). Its prevalence is generally higher in the upper takeover-probability quintiles, rising from 17.13% in the lowest quintile to 27.42% in the fourth quintile before declining slightly to 25.81% in the highest quintile. Topic 2 touches upon industrial themes as it includes words pertaining to (heavy) industry ("coal", "site", "air", "contracts", and "center"). Its prevalence varies across the quintiles but does not exhibit a clear systematic relation with the estimated takeover probabilities. Topic 3 consists of several words that relate to a firm's finances and governance ("disclosures", "stockholder", "owners", and "expectations"). The prevalence of Topic 3 monotonically decreases from the lowest takeover-probability quintile (33.08%) to the highest takeover-probability quintile (19.92%). The fourth topic contains a heterogeneous set of terms and does not exhibit a clear systematic or monotonic association with the estimated takeover probabilities.

[Insert Figure 5 approximately here]

Figure 6 presents the four latent topics within the MDA, whereas Figure 7 depicts the topic importance across the five takeover portfolios analogous to Figure 5. The first topic features accounting-related terms such as "impairment", "fair", and "estimated". Topic 2 contains several terms related to banking and finance ("deposits," "banking," "shareholders," and "lease"), while Topic 3 captures industry- and operations-related language ("crude," "pipeline," "joint," "managed," and "China"). Topic 4 comprises a more heterogeneous set of terms. Although the prevalence of these topics varies across the MDA quintiles, none exhibits a clear and consistently interpretable association with the estimated takeover probabilities.

[Insert Figure 6 approximately here]

While 10-K filings are by construction standardized and thus vary relatively little, the topic modeling uncovers three patterns in the BD in particular. First, estimated takeover exposures depend on operations-specific information picked up by the machine learning algorithm as, for

example, discussions of physical assets are associated with higher predicted takeover probabilities. Second, takeover predictions vary with the prevalence of a firm's focus on its industry. Third, discussions of finances and governance affect estimated takeover exposures.

[Insert Figure 7 approximately here]

B. Integrated Gradient Feature Attribution

Since transformer neural networks use complex interactions within their multi-layer structure, estimating the strength of the link between model inputs (i.e., tokenized text) and output is not straightforward. Thus, we use a recent advancement in explainable artificial intelligence called (layer) integrated gradients to estimate the feature attribution when considering interactions between the inputs (Sundararajan et al. (2017), Janizek et al. (2021)). This approach is based on a comparison with a so-called baseline input (i.e., the zero embedding token corresponding to an empty text input). Tokens of the actual input sequence are added to the baseline input (individually or in groups) and the corresponding effects on the prediction output are analyzed. The changes in the predictions corresponding to different interpolations between the baseline input and the actual input are aggregated into attribution values that quantify the complex link between inputs and outputs.

Figure 8 depicts one example each drawn from the BDs and MDAs of our test sample. As expected, the BD describes the company's overall business model. Although not every attribution is easily interpretable, an interesting pattern emerges. In particular, technologically advanced product characteristics (e.g., "speed", "interconnect", "transform", "cloud") load positively. By contrast, terms related to the particular industrial application ("communications") and operational effects for customers ("performance", "costs") seem to decrease the takeover exposure. The second exhibit,

taken from an MDA, discusses a firm's capabilities. Here, the reliance on information technology ("data assets", "AI technology", "digital advertisements") leads to increased takeover estimates.

[Insert Figure 8 approximately here]

Figure 9 reports the 20 words with the strongest average positive impact on RoBERTa predictions in the BD and the MDA, respectively, since we are especially interested in the words that strongly increase the predicted takeover likelihood across observations.²⁰ Unsurprisingly, words that pertain to operations and business models receive the strongest attributions in the BD as depicted in the blue bars in Figure 9. Descriptions of the firm's market position such as "leader", "premier", and "leading" lead to higher takeover estimates. Similarly, words which may relate to the firm's operations and products such as "personalized", "cloud", "focused", "platform", "targeted", and "deploy" help to identify likely M&A targets. Notably, "leverage" also appears to be particularly relevant for RoBERTa's takeover estimates. By construction, it is hard to aggregate the context-specific attributions of RoBERTa. However, it is clear that RoBERTa's predictions based on the BD incorporate operations- and product-specific characteristics that cannot be picked up by standard financial controls and industry-fixed effects.

[Insert Figure 9 approximately here]

The corresponding 20 words with the strongest positive impact within the MDA are depicted in green in Figure 9. In the MDA, the single most important driver of takeover estimates in our test sample seems to be the field of information technologies as terms such as "cloud", "internet", "online", "websites", and "software" indicate. In addition, potentially related terms such as "platform", "customized", "analytics", "advertising", and "targeted" suggest that related offerings and capabilities are associated with higher takeover probabilities. Moreover, finance-related terms, such as "mortgage", "banking", "bank", and "lending" are associated with increased estimates of

²⁰We limit the feature attribution analysis to a random sample of 2,000 observations (400 per quintile) due to the computational effort. We focus on words that appear in at least 30 texts.

takeover likelihoods.²¹ Lastly, industry information with respect to the healthcare sector ("hospitals", "healthcare", "patients") is clearly relevant to the prediction of takeover targets. Overall, RoBERTa seems to assign higher takeover likelihoods to firms with MDAs which describe specific capabilities, and there is some evidence that struggling firms might also be more likely to be taken over.

We manually assign the words included in Figure 9 to three classes of takeover motives (see Figure A6 in the Online Appendix for an illustration of this subjective classification). First, the *synergy* view of M&As is that firms seek to generate synergies by purchasing specific assets, obtaining desired capabilities, and entering (complementary) product markets (see, e.g., Rhodes-Kropf and Robinson (2008), Hoberg and Phillips (2010)). Words assigned to the “synergy” class of takeover motives dominate Figure 9: terms relating to firms’ business models (e.g., "platform", "personalized") are central drivers of takeover predictions in the BD, while specific (information technology) capabilities (e.g., "cloud", "analytics") discussed in the MDA similarly drive estimated takeover exposures. Second, *industrial reorganization* is argued to be a central driver of clustered M&A activity as industry-specific shocks cause firms to reorganize (see, e.g., Schlingemann, Stulz, and Walkling (2002), Harford (2005)). Consistent with the notion of industry-specific reorganization, several impactful words in the BD describe firms’ competitive role within the industry (e.g., "leader"). In the MDA, words relating to the industry (e.g., "healthcare") drive takeover estimates. Third, the market for corporate control is argued to play a *disciplinary* role as control over struggling firms is reallocated (see, e.g., Gort (1969)). Here, several words are consistent with the notion that financial struggles drive M&As as words in both the BD and the MDA are related to firms’ financial situation (see, e.g., "leverage", "mortgage"). Taken together, the

²¹This finding is also consistent with the notion that part of the predictive power extracted from textual disclosures reflects firms’ underlying financial condition (see the discussion of Table A15).

transformer neural networks are able to uncover valuable fundamental information related to existing takeover theories from unstructured data.

VI. CONCLUSION

In this paper, we examine how textual information can supplement financials in identifying subsequent targets in M&A deals. Therefore, we compile a sample of 102,987 annual 10-K filings, extract two central sections, and train RoBERTa and Longformer, two state-of-the-art deep learning algorithms, to predict takeover targets from textual data. Our results show that incorporating information from publicly available and highly standardized annual disclosures improves the predictability of takeovers significantly compared to models solely relying on structured data. Further analyses reveal that both the Business Description and the Management Discussion and Analysis contribute to the improved prediction. A simple trading strategy based on exposure to takeovers yields significantly positive excess returns out-of-sample.

We employ topic modeling and feature attribution from explainable artificial intelligence methods to shed light on the contribution of textual disclosures to the improved predictability of M&A targets. These analyses suggest that narrative disclosures in financial reports possess relevant information on firms' industrial environment, operations, and governance with regard to takeover likelihoods. Overall, these findings indicate that textual information contributes to the identification of takeover targets by extracting fundamental information from unstructured data.

Our paper presents novel evidence on how recent advances in machine learning can be employed to enhance our understanding of open questions in finance by accessing previously unused sources of unstructured data. We demonstrate how explainable artificial intelligence can improve our comprehension of the market for corporate control.

Figure 1. Sample Split and Regression Procedure

This figure illustrates our training and regression approach. We separate our sample along the time axis into a training (fiscal years 1993 to 2011) and a test (fiscal years 2012 to 2020) dataset. We fine-tune our two transformer neural networks (TNNs) – RoBERTa and Longformer – in the training dataset and use the resulting predictions for the test dataset as input to the logistic regressions. Within the test dataset, we run logistic regressions separately for each year with an expanding estimation window.

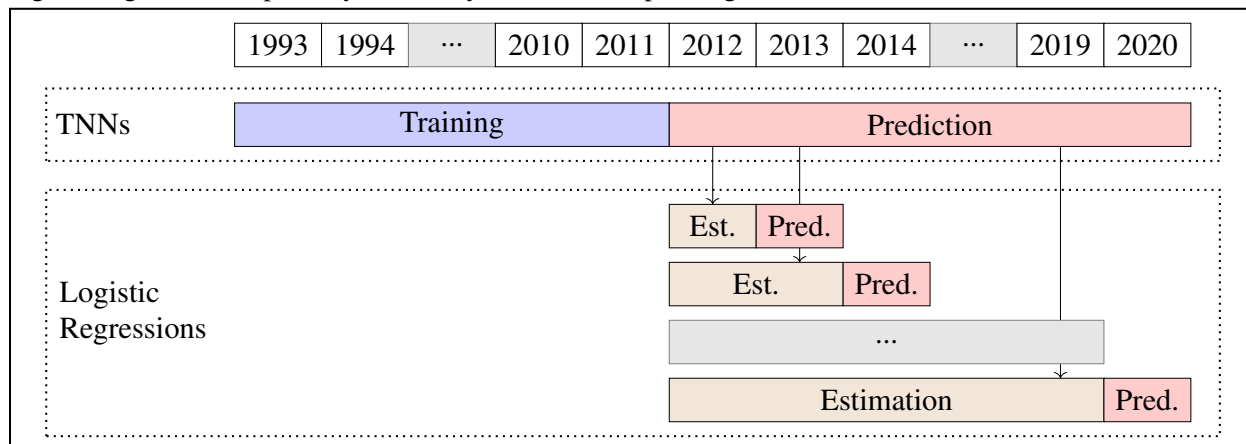


Figure 2. ROC Curves of Benchmark Model and RoBERTa

This figure compares the ROC curve of the logit model based on financial controls, year fixed effects, and industry fixed effects (light red) with the model also including text-based predictions derived by RoBERTa from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

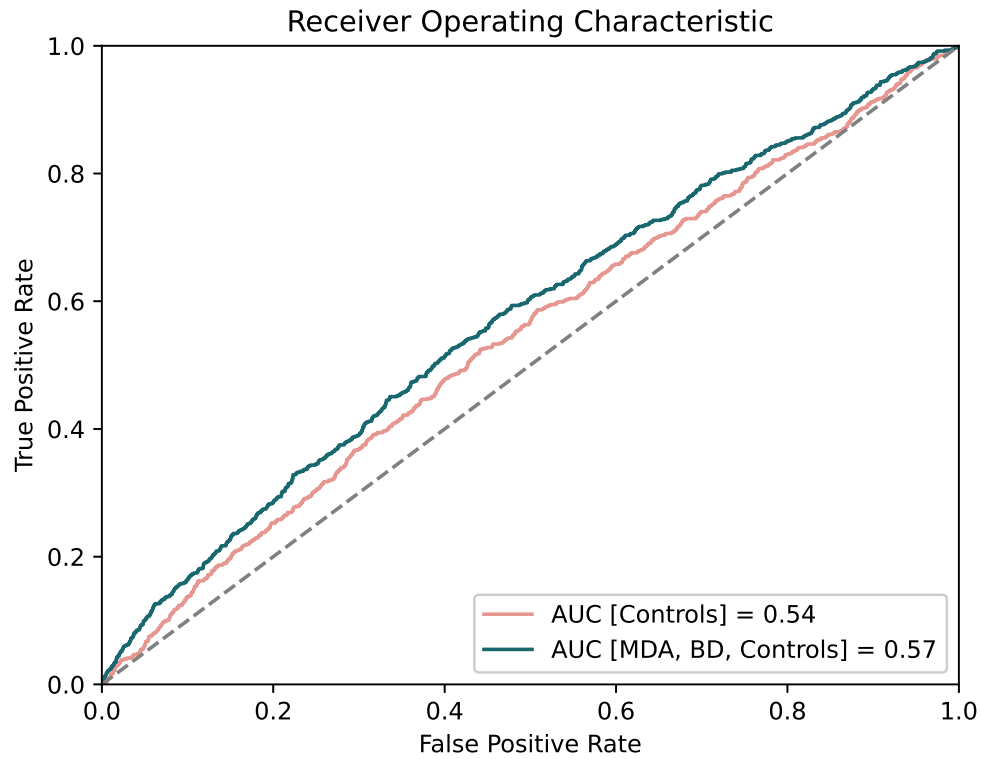


Figure 3. ROC Curves of Benchmark Model and Longformer

This figure compares the ROC curve of the logit model based on financial controls, year-fixed effects, and industry-fixed effects (light red) with the model also including text-based predictions derived by Longformer from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

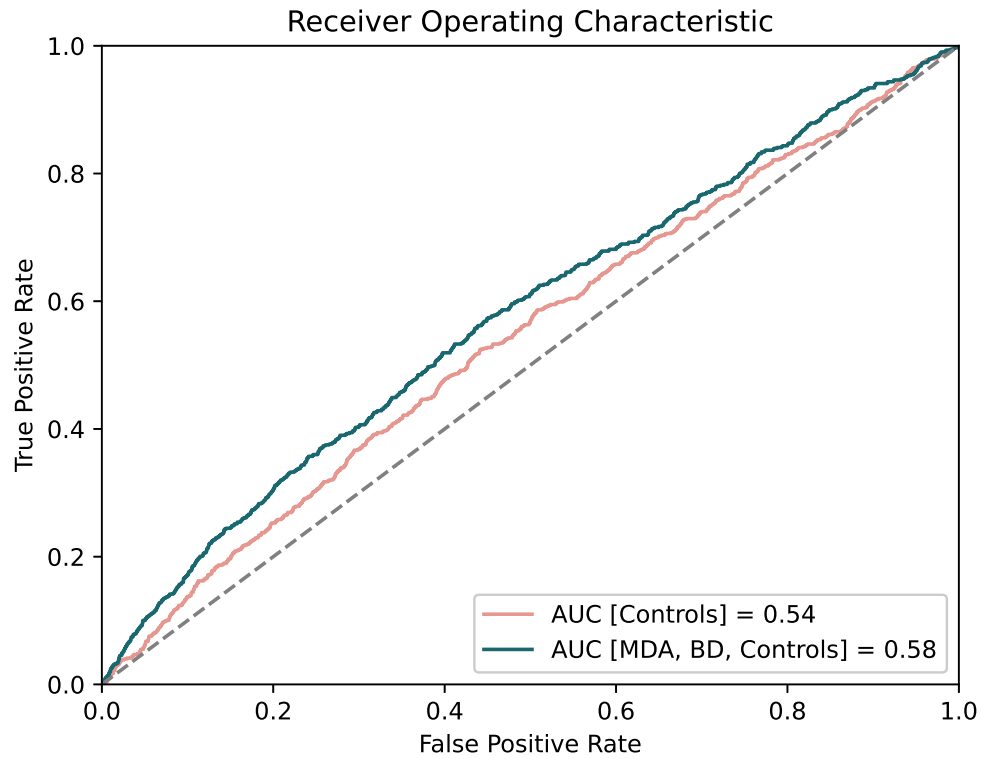


Figure 4. Latent Dirichlet Allocation Topics: BD

This figure shows the 10 highest-weighted words in each topic found for the BD based on a Latent Dirichlet Allocation analysis of test sample observations. Manually added titles for each topic are reported above each word cloud.

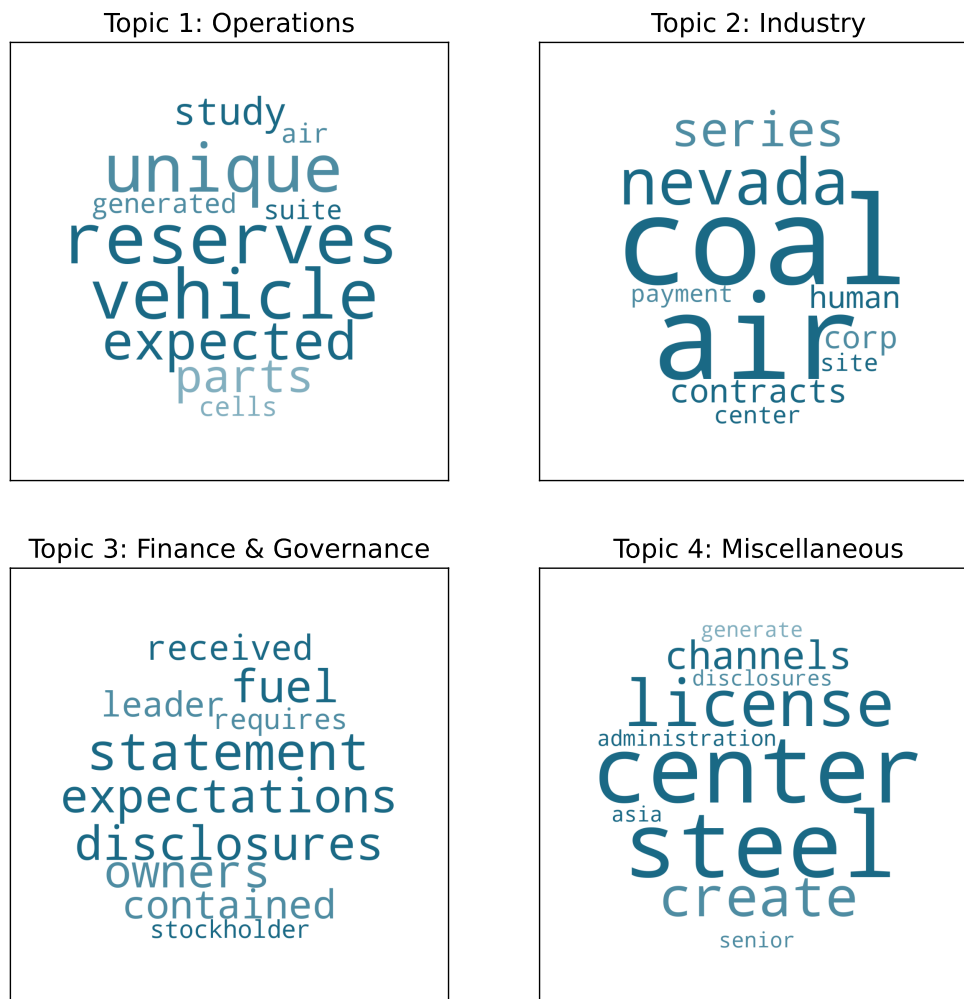


Figure 5. Latent Dirichlet Allocation Topic Distribution Across BD Quintiles

This figure displays the prevalence of Latent Dirichlet Allocation topics across quintiles based on the predicted takeover likelihood using BD only. The height of a column refers to the average topic association of the BD across all observations in the test sample. A greater height of a column indicates a higher relevance of the topic within texts in the corresponding quintile. The darker the column, the higher the associated takeover likelihood quintile.

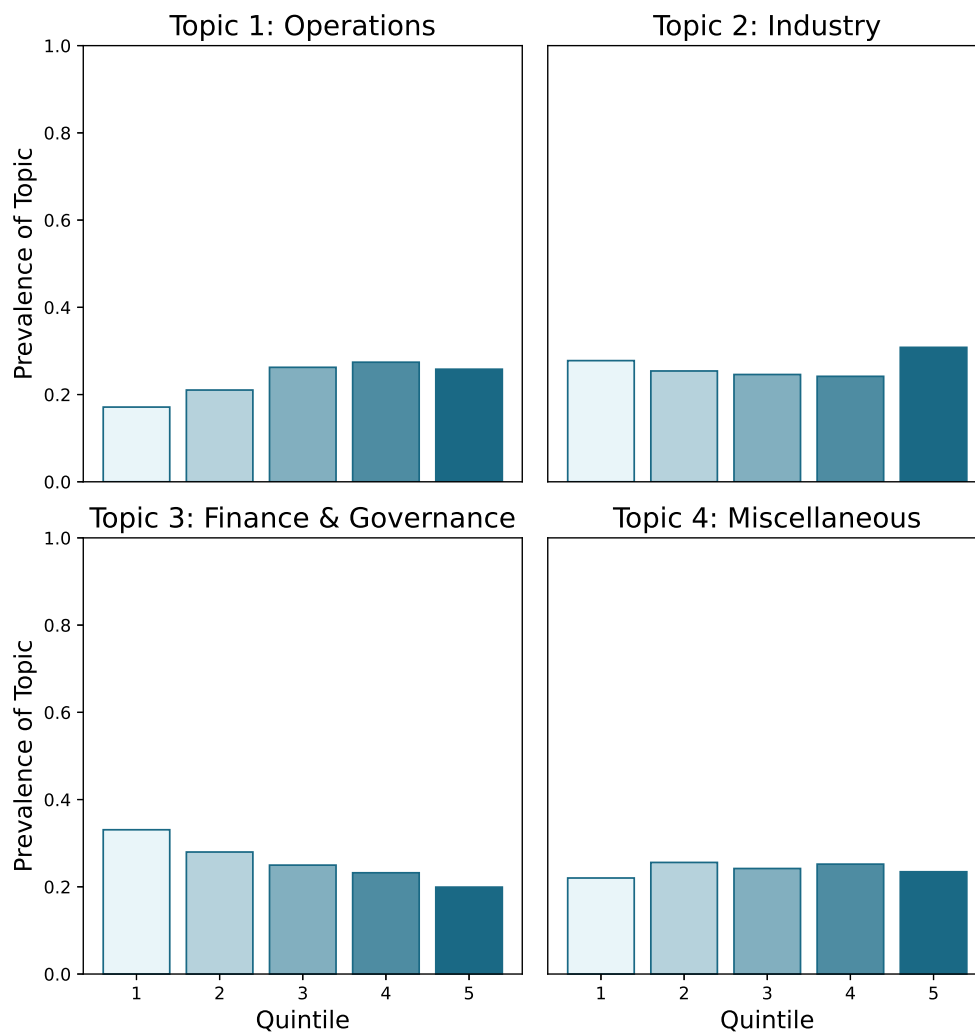


Figure 6. Latent Dirichlet Allocation Topics: MDA

This figure shows the 10 highest-weighted words in each topic found for the MDA based on a Latent Dirichlet Allocation analysis of test sample observations. Manually added titles for each topic are reported above each word cloud.

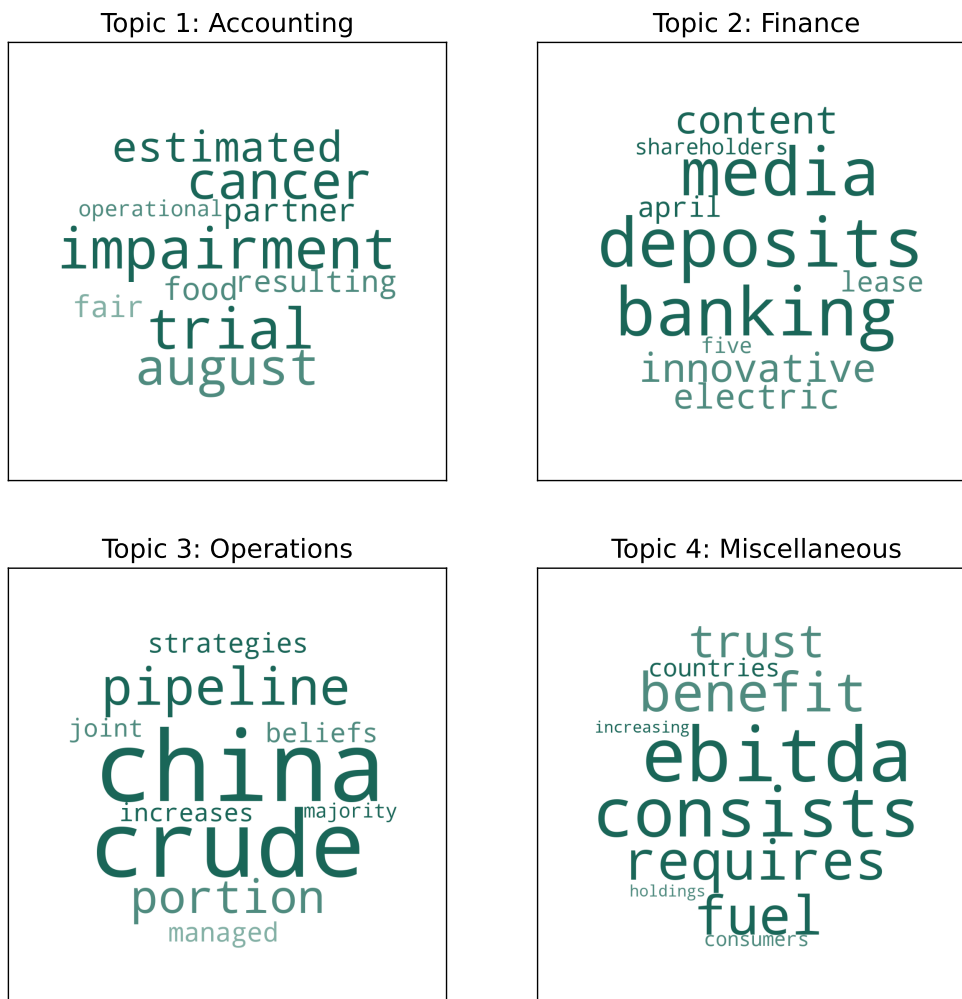


Figure 7. Latent Dirichlet Allocation Topic Distribution Across MDA Quintiles

This figure displays the prevalence of Latent Dirichlet Allocation topics across quintiles based on the predicted takeover likelihood using MDA only. The height of a column refers to the average topic association of the MDA across all observations in the test sample. A greater column height indicates a higher topic relevance within texts in the corresponding quintile. The darker a column, the higher the associated takeover likelihood quintile.

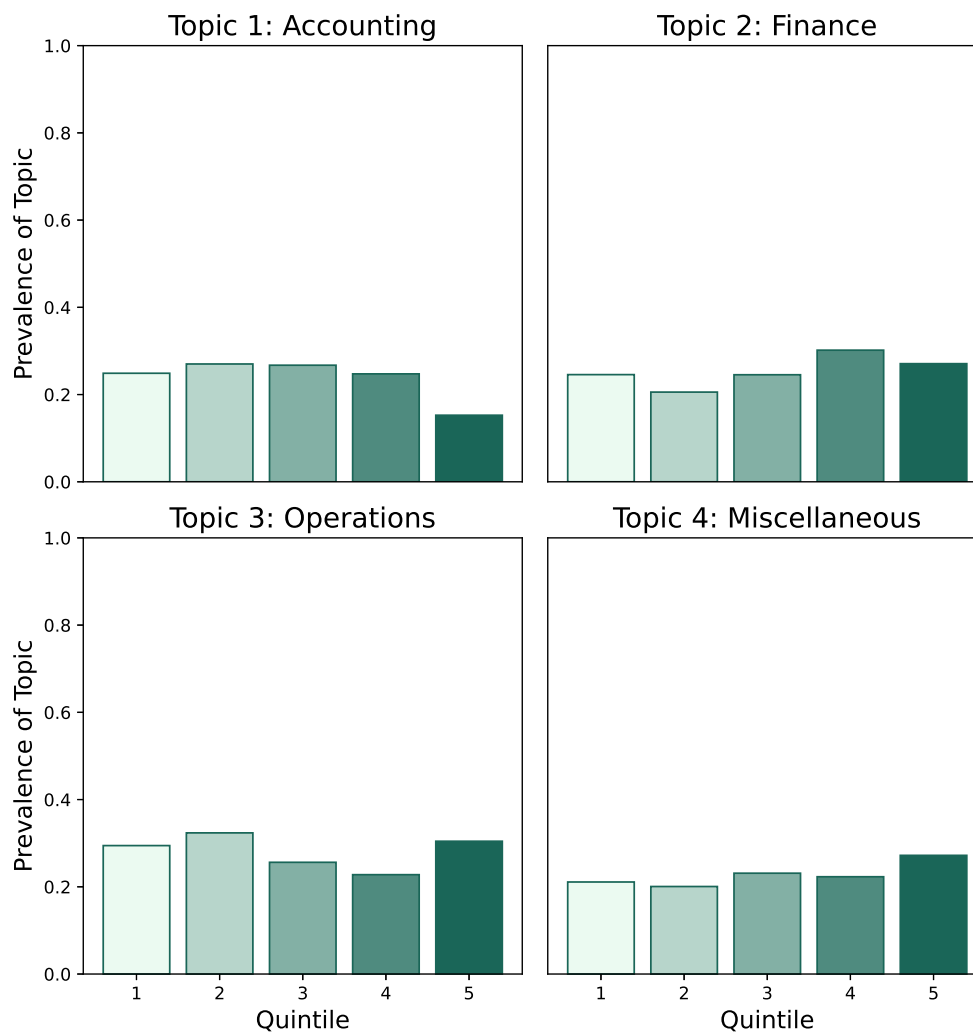


Figure 8. Feature Attribution from (Layer) Integrated Gradient Analysis: Examples

This figure displays two examples of applied (layer) integrated gradients to text excerpts from our test set. For illustrative purposes, positive attributions larger than 0.015 are highlighted in green; negative attributions smaller than -0.015 are highlighted in red.

Our mission is to deliver high speed coherent optical interconnect products that transform communications networks relied upon by cloud infrastructure operators and content and communication service providers through improvements in performance and capacity and reductions in associated costs.

To drive measurable results for clients, we activate our data assets through proprietary AI technology to engage consumers in real time through the pricing and delivery of highly relevant digital advertisements across devices and environments.

Figure 9. Aggregate Feature Attribution from (Layer) Integrated Gradient Analysis

This figure displays feature attributions based on (layer) integrated gradient analysis for both the BD and the MDA. The top 20 words are displayed with the strongest average positive impact on the predicted takeover likelihood. Reported attributions are based on a random sample of 2,000 observations (400 per quintile) from our test dataset. For this figure, we only consider words that appear in at least 30 observations.

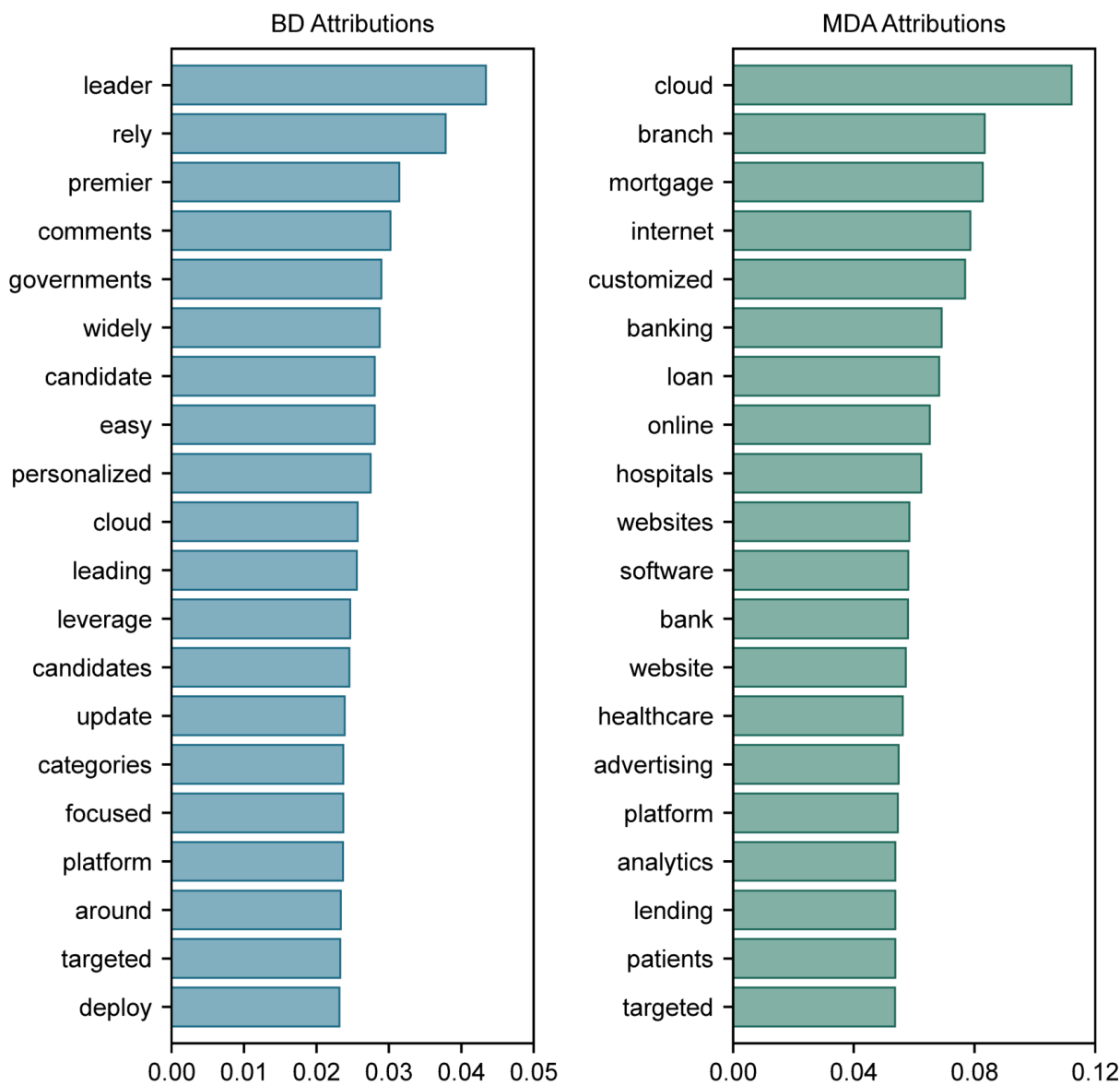


Table 1. Descriptive Statistics of Structured Data

This table reports summary statistics for both the dependent variable (*Takeover Dummy*) and the set of financial control variables for the test sample. Listed are the arithmetic mean (Mean), the standard deviation (St. Dev.), three quartiles ($X_{0.25}$, $X_{0.50}$, $X_{0.75}$), and the number of observations (N).

	Mean	St. Dev.	$X_{0.25}$	$X_{0.50}$	$X_{0.75}$	N
<i>Takeover Dummy</i>	0.0406	0.1974	0.0000	0.0000	0.0000	19,665
<i>RoE</i>	-0.0036	1.5866	-0.1233	0.0667	0.1554	19,665
<i>Market Cap</i>	6.1329	2.4610	4.4111	6.3282	7.8861	19,665
<i>MB</i>	3.0626	10.9093	0.9719	1.8952	3.9010	19,665
<i>PE</i>	8.1893	70.4342	-4.7727	9.4110	21.9821	19,665
<i>Leverage</i>	0.5777	2.3568	0.0000	0.2518	0.8196	19,665
<i>Tangible Assets</i>	0.2210	0.2706	0.0250	0.1056	0.3266	19,665
<i>Sales Growth</i>	0.3928	1.1298	0.0201	0.1247	0.3268	19,665
<i>Liquidity</i>	0.3900	0.2706	0.1620	0.3235	0.6037	19,665
<i>Growth-Resource-Mismatch</i>	0.2745	0.4463	0.0000	0.0000	1.0000	19,665
<i>Takeover Activity</i>	0.4462	0.4971	0.0000	0.0000	1.0000	19,665

Table 2. Hyperparameters for RoBERTa and Longformer Models

This table reports the considered set of hyperparameters for fine-tuning the RoBERTa (Panel A) and Longformer (Panel B) models and their optimal values for both BD and MDA sections. The batch size describes the number of observations used for each step when updating the model weights. The learning rate determines the size of the adjustment to model weights conducted in each updating step.

Hyperparameter	possible values	optimal value BD	optimal value MDA
<i>Panel A: RoBERTa</i>			
global batch size	[16, 32, 64]	32	64
learning rate	[.000001, .000005, .00001]	.000005	.000005
<i>Panel B: Longformer</i>			
global batch size	[16]	16	16
learning rate	[.000001, .000005, .00001]	.000001	.000001

Identifying M&A Targets from Textual Disclosures

Table 3. Textual Disclosures and Takeover Likelihood

This table reports the marginal effects calculated across all actual observations from logit regressions. The conducted regressions correspond to the last rolling-window (i.e., using data from 2012 to 2019). The dependent variable is the *Takeover Dummy*, which is equal to 1 if a firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise. The main explanatory variables are takeover predictions derived by RoBERTa (Longformer) from BD and MDA in Panel A (Panel B). A constant term is included but not reported. z-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: <i>Takeover Dummy</i>						
	1	2	3	4	5	6	7
<i>Panel A: RoBERTa Predictions</i>							
<i>RoBERTa BD</i>	0.4226*** (8.81)		0.4051*** (6.01)		0.4033*** (7.39)		0.3808*** (5.60)
<i>RoBERTa MDA</i>		0.3169*** (6.38)	0.0260 (0.37)			0.2875*** (4.72)	0.0421 (0.56)
<i>RoE</i>				-0.0007 (-0.77)	-0.0006 (-0.67)	-0.0007 (-0.74)	-0.0006 (-0.56)
<i>Market Cap</i>				-0.0022*** (-3.34)	-0.0025*** (-3.59)	-0.0025*** (-3.70)	-0.0025*** (-3.62)
<i>MB</i>				-0.0001 (-0.76)	-0.0002 (-1.01)	-0.0002 (-0.99)	-0.0002 (-1.03)
<i>PE</i>				0.0000 (0.78)	0.0000 (0.72)	0.0000 (0.74)	0.0000 (0.72)
<i>Leverage</i>				0.0015** (2.44)	0.0015** (2.41)	0.0015** (2.43)	0.0015** (2.41)
<i>Tangible Assets</i>				0.0171** (2.07)	0.0207** (2.52)	0.0215*** (2.60)	0.0212** (2.56)
<i>Sales Growth</i>				-0.0035** (-2.03)	-0.0035** (-2.04)	-0.0035** (-2.02)	-0.0035** (-2.04)
<i>Liquidity</i>				0.0134* (1.76)	0.0109 (1.42)	0.0110 (1.43)	0.0107 (1.40)
<i>Growth-Resource-Mismatch</i>				-0.0056 (-1.48)	-0.0051 (-1.36)	-0.0054 (-1.42)	-0.0051 (-1.36)
<i>Takeover Activity</i>				0.0181*** (5.31)	0.0146*** (4.29)	0.0160*** (4.68)	0.0145*** (4.25)
Year & Industry FE	No	No	No	Yes	Yes	Yes	Yes
N	17,967	17,967	17,967	17,967	17,967	17,967	17,967
Pseudo R2	0.0120	0.0061	0.0120	0.0161	0.0248	0.0197	0.0249
<i>Panel B: Longformer Predictions</i>							
<i>Longformer BD</i>	0.5194*** (7.62)		0.2127** (2.22)		0.4888*** (6.19)		0.2406** (2.46)
<i>Longformer MDA</i>		0.7186*** (8.42)	0.5377*** (4.55)			0.7479*** (7.00)	0.5640*** (4.32)
Financial Controls	No	No	No	Yes	Yes	Yes	Yes
Year & Industry FE	No	No	No	Yes	Yes	Yes	Yes
N	17,967	17,967	17,967	17,967	17,967	17,967	17,967
Pseudo R2	0.0091	0.0116	0.0124	0.0161	0.0223	0.0244	0.0254

Identifying M&A Targets from Textual Disclosures

Table 4. Textual Disclosures and Takeover Likelihood: Industry & Demand Controls

This table reports the marginal effects calculated across all actual observations from logit regressions. The conducted regressions correspond to the last rolling-window (i.e., using data from 2012 to 2019). The dependent variable is the *Takeover Dummy*, which is equal to 1 if a firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise. The main explanatory variables are takeover predictions derived by RoBERTa and Longformer from BD and MDA, respectively. In addition to the “Financial Controls” included in Table 3, the reported regressions include additional industry and demand variables as control variables. A constant term is included but not reported. z-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: <i>Takeover Dummy</i>					
	1	2	3	4	5	6
<i>RoBERTa BD</i>	0.3557*** (4.83)		0.2940*** (3.23)			
<i>RoBERTa MDA</i>		0.3086*** (3.82)	0.1206 (1.21)			
<i>Longformer BD</i>				0.3608*** (3.48)		0.1648 (1.26)
<i>Longformer MDA</i>					0.5445*** (3.93)	0.4223*** (2.44)
<i>Industry Profit</i>	0.0001 (0.59)	0.0001 (0.58)	0.0001 (0.53)	0.0001 (0.63)	0.0001 (0.59)	0.0001 (0.51)
<i>Low Profit</i>	0.0026 (0.64)	0.0024 (0.59)	0.0025 (0.61)	0.0026 (0.63)	0.0027 (0.65)	0.0027 (0.65)
<i>TNIC HHI</i>	-0.0223** (-1.97)	-0.0244** (-2.14)	-0.0229** (-2.02)	-0.0235** (-2.06)	-0.0244** (-2.14)	-0.0235** (-2.07)
<i>Downstream Sales Growth</i>	-0.0059 (-0.38)	-0.0051 (-0.31)	-0.0046 (-0.29)	-0.0074 (-0.46)	-0.0045 (-0.29)	-0.0042 (-0.27)
<i>Industrial Production</i>	-0.0114 (-0.02)	-0.0132 (-0.03)	-0.0422 (-0.09)	-0.0241 (-0.05)	-0.0392 (-0.08)	-0.1003 (-0.21)
Financial Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year & Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
N	11,860	11,860	11,860	11,860	11,860	11,860
Pseudo R2	0.0349	0.0329	0.0353	0.0323	0.0332	0.0337

Table 5. AUC of Different Model Specifications

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. Our *Controls* model, which includes financial controls as well as year- and industry-fixed effects, serves as a benchmark. The uplift in AUC reported in the last column is denoted relative to the *Controls* model. The significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift compared to controls
<i>Controls</i>	0.5433	-
<i>RoBERTa BD & Controls</i>	0.5736	0.0303*** (4.63)
<i>RoBERTa MDA & Controls</i>	0.5576	0.0143*** (3.19)
<i>RoBERTa BD, RoBERTa MDA, & Controls</i>	0.5729	0.0296*** (4.54)
<i>Longformer BD & Controls</i>	0.5601	0.0168*** (2.69)
<i>Longformer MDA & Controls</i>	0.5780	0.0347*** (4.50)
<i>Longformer BD, Longformer MDA & Controls</i>	0.5759	0.0326*** (4.08)

Table 6. Stock Market Performance of Predicted Targets

This table reports the extreme legs of monthly quintile portfolio sorts based on the estimated probability to become an M&A target derived from the logistic regressions including RoBERTa (Panel A) and Longformer (Panel B) as well as the aforementioned control variables and fixed effects for the period from March 2013 to February 2022. The table presents subsequent equal-weighted monthly returns R after accounting for transaction costs. For the difference portfolio (5-1), transaction costs are calculated and deducted separately for the long and short legs. The portfolio excess returns (α) and the corresponding factor loadings for the market (β_{MKT}), size (β_{SMB}), value (β_{HML}), momentum (β_{WML}), profitability (β_{ROE}), investment (β_{INV}), liquidity (β_{LIQ}), and takeover vulnerability factors (β_{TMA}) are also provided. The t-statistics in parentheses refer to the difference portfolio and are based on standard errors following Newey and West (1987) using twelve lags. Returns and alphas are stated in %.

	R	α	β_{MKT}	β_{SMB}	β_{HML}	β_{WML}	β_{ROE}	β_{INV}	β_{LIQ}	β_{TMA}
<i>Panel A: RoBERTa Predictions</i>										
1	1.49	0.36	1.01	0.55	0.20	-0.09	0.01	-0.02	0.00	-0.04
5	2.53	1.40	0.99	0.57	0.16	0.11	-0.37	-0.01	0.02	0.16
5-1	0.84	0.85	-0.01	0.02	0.01	0.22	-0.39	-0.03	0.04	0.18
t(5-1)	(3.10)	(2.69)	(-0.27)	(0.09)	(0.17)	(2.71)	(-1.88)	(-0.34)	(0.42)	(1.71)
<i>Panel B: Longformer Predictions</i>										
1	1.70	0.56	1.04	0.53	0.24	-0.08	-0.01	-0.10	-0.04	-0.05
5	2.44	1.35	0.95	0.52	0.09	0.07	-0.34	0.06	0.10	0.09
5-1	0.53	0.60	-0.08	-0.01	-0.10	0.18	-0.34	0.12	0.16	0.12
t(5-1)	(1.71)	(1.88)	(-1.34)	(-0.03)	(-1.54)	(2.54)	(-2.58)	(0.81)	(1.30)	(1.42)

Table 7. Picking Takeover Targets Out-of-Sample

This table reports the out-of-sample accuracy, precision, recall, and F_1 -score for logit models based on different sets of predictors for the test sample. We report year- and model-specific thresholds for takeover classifications in Table A18 in the Online Appendix. The uplift in the F_1 -Score in the last column is denoted relative to the *Controls* model. The significances of differences in the F_1 -Score are calculated using firm-clustered bootstrapping. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor Set	Accuracy	Precision	Recall	F_1 -Score	Uplift to Controls
<i>Controls</i>	0.8735	0.0620	0.1456	0.0870	-
<i>RoBERTa BD & RoBERTa MDA</i>	0.8334	0.0610	0.2102	0.0945	0.0075
<i>RoBERTa BD, RoBERTa MDA, & Controls</i>	0.8439	0.0688	0.2212	0.1050	0.0180**
<i>Longformer BD & Longformer MDA</i>	0.8244	0.0610	0.2253	0.0960	0.0090
<i>Longformer BD, Longformer MDA, & Controls</i>	0.8898	0.0787	0.1552	0.1044	0.0174**

Table 8. Picking Takeover Targets Out-of-Sample – Decile-Level Performance

This table reports out-of-sample prediction performance across deciles of predicted probabilities separately for the *Controls* model (Panel A), the *RoBERTa BD*, *RoBERTa MDA*, & *Controls* model (Panel B), and the *Longformer BD*, *Longformer MDA*, & *Controls* model (Panel C). For each model, firm-years are sorted into deciles based on predicted takeover likelihoods, where Decile 10 contains observations with the highest predicted probabilities and Decile 1 contains observations with the lowest predicted probabilities. “True Takeovers” reports the number of realized takeover targets in a given decile, “Takeover Rate” the fraction of firm-years that become takeover targets, and “TP” the number of true positive predictions. Takeover predictions are assigned using the model-specific classification thresholds reported in Table A18 in the Online Appendix. All metrics are computed separately within each decile using only observations assigned to that decile.

Decile	True Takeovers	Takeover Rate	TP	F_1 -Score	Accuracy
<i>Panel A: Controls</i>					
10	111	0.0630	84	0.1146	0.2633
9	90	0.0512	22	0.0991	0.7723
8	74	0.0420	0	0.0000	0.9580
7	71	0.0404	0	0.0000	0.9596
6	73	0.0415	0	0.0000	0.9585
5	77	0.0438	0	0.0000	0.9562
4	65	0.0370	0	0.0000	0.9630
3	67	0.0381	0	0.0000	0.9619
2	47	0.0267	0	0.0000	0.9733
1	53	0.0301	0	0.0000	0.9699
<i>Panel B: RoBERTa BD, RoBERTa MDA, & Controls</i>					
10	129	0.0732	103	0.1412	0.2889
9	92	0.0524	40	0.0979	0.5805
8	90	0.0511	15	0.0912	0.8301
7	78	0.0444	3	0.0484	0.9328
6	78	0.0444	0	0.0000	0.9556
5	55	0.0313	0	0.0000	0.9687
4	59	0.0336	0	0.0000	0.9664
3	49	0.0279	0	0.0000	0.9721
2	44	0.0250	0	0.0000	0.9750
1	54	0.0306	0	0.0000	0.9694
<i>Panel C: Longformer BD, Longformer MDA, & Controls</i>					
10	130	0.0738	106	0.1437	0.2832
9	99	0.0563	7	0.0737	0.8998
8	79	0.0449	0	0.0000	0.9551
7	68	0.0387	0	0.0000	0.9613
6	72	0.0410	0	0.0000	0.9590
5	72	0.0410	0	0.0000	0.9590
4	61	0.0347	0	0.0000	0.9653
3	47	0.0267	0	0.0000	0.9733
2	55	0.0313	0	0.0000	0.9687
1	45	0.0255	0	0.0000	0.9745

REFERENCES

- Abdi, F., and A. Rinaldo. “A simple estimation of bid-ask spreads from daily close, high, and low prices.” *Review of Financial Studies*, 30 (2017), 4437–4480.
- Acikalin, U.; T. Caskurlu; G. Hoberg; and G. Phillips. “Intellectual property protection lost and competition: An examination using large language models.” *Journal of Financial Economics*, 182 (2026), 104306.
- Ahern, K. R., and J. Harford. “The importance of industry links in merger waves.” *Journal of Finance*, 69 (2014), 527–576.
- Ambrose, B. W., and W. L. Megginson. “The role of asset structure, ownership structure, and takeover defenses in determining acquisition likelihood.” *Journal of Financial and Quantitative Analysis*, 27 (1992), 575–589.
- Andrade, G.; M. Mitchell; and E. Stafford. “New evidence and perspectives on mergers.” *Journal of Economic Perspectives*, 15 (2001), 103–120.
- Azimi, M., and A. Agrawal. “Is positive sentiment in corporate annual reports informative? Evidence from deep learning.” *Review of Asset Pricing Studies*, 11 (2021), 762–805.
- Baker, M.; X. Pan; and J. Wurgler. “The effect of reference point prices on mergers and acquisitions.” *Journal of Financial Economics*, 106 (2012), 49–71.
- Beltagy, I.; M. E. Peters; and A. Cohan. “Longformer: The long-document transformer.” arXiv preprint arXiv:2004.05150 (2020).
- Betton, S.; B. E. Eckbo; and K. S. Thorburn. “Corporate takeovers.” In *Handbook of Empirical Corporate Finance*, Vol. 2, B. E. Eckbo, ed. Elsevier (2008), 291–429.
- Bhagat, S.; M. Dong; D. Hirshleifer; and R. Noah. “Do tender offers create value? New methods and evidence.” *Journal of Financial Economics*, 76 (2005), 3–60.

- Blei, D. M.; A. Y. Ng; and M. I. Jordan. “Latent Dirichlet allocation.” *Journal of Machine Learning Research*, 3 (2003), 993–1022.
- Brogaard, J., and A. Zareei. “Machine learning and the stock market.” *Journal of Financial and Quantitative Analysis*, 58 (2023), 1431–1472.
- Carhart, M. M. “On persistence in mutual fund performance.” *Journal of Finance*, 52 (1997), 57–82.
- Cocco, J. F., and P. F. Volpin. “Corporate pension plans as takeover deterrents.” *Journal of Financial and Quantitative Analysis*, 48 (2013), 1119–1144.
- Cremers, K. M.; V. B. Nair; and K. John. “Takeovers and the cross-section of returns.” *Review of Financial Studies*, 22 (2009), 1409–1445.
- Cuñat, V.; M. Giné; and M. Guadalupe. “Price and probability: Decomposing the takeover effects of anti-takeover provisions.” *Journal of Finance*, 75 (2020), 2591–2629.
- Devlin, J.; M.-W. Chang; K. Lee; and K. Toutanova. In *Proceedings of the 2019 Conference of the NAACL-HLT*. “BERT: Pre-training of deep bidirectional transformers for language understanding.”, 4171–4186. Association for Computational Linguistics (2019).
- Dyer, T.; M. Lang; and L. Stice-Lawrence. “The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation.” *Journal of Accounting and Economics*, 64 (2017), 221–245.
- Easterwood, S.; J. Netter; B. Paye; and M. Stegemoller. “Taking over the size effect: Asset pricing implications of merger activity.” *Journal of Financial and Quantitative Analysis*, 59 (2024), 690–726.
- Eckbo, B. E. “Corporate takeovers and economic efficiency.” *Annual Review of Financial Economics*, 6 (2014), 51–74.

- Edmans, A.; I. Goldstein; and W. Jiang. “The real effects of financial markets: The impact of prices on takeovers.” *Journal of Finance*, 67 (2012), 933–971.
- Erel, I.; L. H. Stern; C. Tan; and M. S. Weisbach. “Selecting directors using machine learning.” *Review of Financial Studies*, 34 (2021), 3226–3264.
- Fama, E. F., and K. R. French. “Common risk factors in the returns on stocks and bonds.” *Journal of Financial Economics*, 33 (1993), 3–56.
- Fama, E. F., and K. R. French. “Industry costs of equity.” *Journal of Financial Economics*, 43 (1997), 153–193.
- Frankel, R.; J. Jennings; and J. Lee. “Disclosure sentiment: Machine learning vs. dictionary methods.” *Management Science*, 68 (2022), 5514–5532.
- Frésard, L.; G. Hoberg; and G. Phillips. “Innovation activities and integration through vertical acquisitions.” *Review of Financial Studies*, 33 (2020), 2937–2976.
- Golubov, A.; A. Yawson; and H. Zhang. “Extraordinary acquirers.” *Journal of Financial Economics*, 116 (2015), 314–330.
- Gort, M. “An economic disturbance theory of mergers.” *Quarterly Journal of Economics*, 83 (1969), 624–642.
- Harford, J. “What drives merger waves?” *Journal of Financial Economics*, 77 (2005), 529–560.
- Hoberg, G., and V. Maksimovic. “Redefining financial constraints: A text-based analysis.” *Review of Financial Studies*, 28 (2015), 1312–1352.
- Hoberg, G., and G. Phillips. “Product market synergies and competition in mergers and acquisitions: A text-based analysis.” *Review of Financial Studies*, 23 (2010), 3773–3811.
- Hoberg, G., and G. Phillips. “Text-based network industries and endogenous product differentiation.” *Journal of Political Economy*, 124 (2016), 1423–1465.

- Hou, K.; C. Xue; and L. Zhang. “Digesting anomalies: An investment approach.” *Review of Financial Studies*, 28 (2015), 650–705.
- Janizek, J. D.; P. Sturmfels; and S.-I. Lee. “Explaining explanations: Axiomatic feature interactions for deep networks.” *Journal of Machine Learning Research*, 22 (2021), 1–54.
- Jarrell, G. A.; J. A. Brickley; and J. M. Netter. “The market for corporate control: The empirical evidence since 1980.” *Journal of Economic Perspectives*, 2 (1988), 49–68.
- Jegadeesh, N., and S. Titman. “Returns to buying winners and selling losers: Implications for stock market efficiency.” *Journal of Finance*, 48 (1993), 65–91.
- Jegadeesh, N., and S. Titman. “Profitability of momentum strategies: An evaluation of alternative explanations.” *Journal of Finance*, 56 (2001), 699–720.
- Jensen, M. C., and R. S. Ruback. “The market for corporate control: The scientific evidence.” *Journal of Financial Economics*, 11 (1983), 5–50.
- John, K.; D. Kadyrzhanova; and S. Lee. “Do classified boards deter takeovers? Evidence from merger waves.” *Journal of Financial and Quantitative Analysis*, 59 (2024), 759–795.
- Jovanovic, B., and P. L. Rousseau. “The Q-theory of mergers.” *American Economic Review*, 92 (2002), 198–204.
- Katsafados, A. G.; I. Androutsopoulos; I. Chalkidis; E. Fergadiotis; G. N. Leledakis; and E. G. Pyrgiotakis. “Using textual analysis to identify merger participants: Evidence from the U.S. banking industry.” *Finance Research Letters*, 42 (2021), 101949.
- Kriebel, J., and L. Stitz. “Credit default prediction from user-generated text in peer-to-peer lending using deep learning.” *European Journal of Operational Research*, 302 (2022), 309–323.
- LeCun, Y.; Y. Bengio; and G. Hinton. “Deep learning.” *Nature*, 521 (2015), 436–444.

- Li, K.; F. Mai; R. Shen; and X. Yan. “Measuring corporate culture using machine learning.” *Review of Financial Studies*, 34 (2021), 3265–3315.
- Liu, Y.; M. Ott; N. Goyal; J. Du; M. Joshi; D. Chen; O. Levy; M. Lewis; L. Zettlemoyer; and V. Stoyanov. “RoBERTa: A robustly optimized BERT pretraining approach.” arXiv preprint arXiv:1907.11692 (2019).
- Loughran, T., and B. McDonald. “Measuring readability in financial disclosures.” *Journal of Finance*, 69 (2014), 1643–1671.
- Loughran, T., and B. McDonald. “Measuring firm complexity.” *Journal of Financial and Quantitative Analysis*, 59 (2024), 2487–2514.
- Maksimovic, V., and G. Phillips. “The market for corporate assets: Who engages in mergers and asset sales and are there efficiency gains?” *Journal of Finance*, 56 (2001), 2019–2065.
- Masulis, R. W.; C. Wang; and F. Xie. “Corporate governance and acquirer returns.” *Journal of Finance*, 62 (2007), 1851–1889.
- Mikkelson, W. H., and M. M. Partch. “Managers’ voting rights and corporate control.” *Journal of Financial Economics*, 25 (1989), 263–290.
- Mitchell, M. L., and J. H. Mulherin. “The impact of industry shocks on takeover and restructuring activity.” *Journal of Financial Economics*, 41 (1996), 193–229.
- Newey, W. K., and K. D. West. “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix.” *Econometrica*, 55 (1987), 703–708.
- Palepu, K. G. “Predicting takeover targets: A methodological and empirical analysis.” *Journal of Accounting and Economics*, 8 (1986), 3–35.
- Pástor, L., and R. F. Stambaugh. “Liquidity risk and expected stock returns.” *Journal of Political Economy*, 111 (2003), 642–685.

- Petersen, M. A. “Estimating standard errors in finance panel data sets: Comparing approaches.” *Review of Financial Studies*, 22 (2009), 435–480.
- Rajpurkar, P.; R. Jia; and P. Liang. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers*. “Know what you don’t know: Unanswerable questions for SQuAD.”, 784–789. Association for Computational Linguistics (2018).
- Rajpurkar, P.; J. Zhang; K. Lopyrev; and P. Liang. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. “SQuAD: 100,000+ questions for machine comprehension of text.”, 2383–2392. Association for Computational Linguistics (2016).
- Rhodes-Kropf, M., and D. T. Robinson. “The market for mergers and the boundaries of the firm.” *Journal of Finance*, 63 (2008), 1169–1211.
- Routledge, B. R.; S. Sacchetto; and N. A. Smith. “Predicting merger targets and acquirers from text.” Working Paper, Carnegie Mellon University (2017).
- Schlingemann, F. P.; R. M. Stulz; and R. A. Walkling. “Divestitures and the liquidity of the market for corporate assets.” *Journal of Financial Economics*, 64 (2002), 117–144.
- Schneider, C., and O. Spalt. “Acquisitions as lotteries: The selection of target-firm risk and its impact on merger outcomes.” *Critical Finance Review*, 6 (2017), 77–132.
- Shivdasani, A. “Board composition, ownership structure, and hostile takeovers.” *Journal of Accounting and Economics*, 16 (1993), 167–198.
- Sundararajan, M.; A. Taly; and Q. Yan. In *Proceedings of the 34th International Conference on Machine Learning*. “Axiomatic attribution for deep networks.”, 3319–3328. PMLR (2017).
- Vaswani, A.; N. Shazeer; N. Parmar; J. Uszkoreit; L. Jones; A. N. Gomez; Ł. Kaiser; and I. Polosukhin. “Attention is all you need.” *Advances in Neural Information Processing Systems*, 30 (2017), 5998–6008.

Wang, A.; A. Singh; J. Michael; F. Hill; O. Levy; and S. R. Bowman. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*.

“GLUE: A multi-task benchmark and analysis platform for natural language understanding.”, 353–355. Association for Computational Linguistics (2018).

Xiao, C.; X. Hu; Z. Liu; C. Tu; and M. Sun. “Lawformer: A pre-trained language model for Chinese legal long documents.” *AI Open*, 2 (2021), 79–84.

Online Appendix for

**“Identifying M&A Targets from Textual Disclosures:
A Transformer Neural Network Approach”**

A. VARIABLE DEFINITIONS

Table A1. Variable Definitions

In this table, we define all variables used in the main paper and in additional analyses in our Online Appendix.

Variable	Variable Definition
<i>Takeover Dummy</i>	<i>Takeover Dummy</i> is a dummy variable equal to 1 if the firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise.
<i>RoBERTa BD</i>	<i>RoBERTa BD</i> is the takeover likelihood prediction derived by RoBERTa based on the "Business Description" section of the firm's Form 10-K filing. Section III of the main paper details how these predictions are constructed.
<i>RoBERTa MDA</i>	<i>RoBERTa MDA</i> is the takeover likelihood prediction derived by RoBERTa based on the "Management Discussion and Analysis" section of the firm's Form 10-K filing. Section III of the main paper details how these predictions are constructed.
<i>Longformer BD</i>	<i>Longformer BD</i> is the takeover likelihood prediction derived by Longformer based on the "Business Description" section of the firm's Form 10-K filing. Section III of the main paper details how these predictions are constructed.
<i>Longformer MDA</i>	<i>Longformer MDA</i> is the takeover likelihood prediction derived by Longformer based on the "Management Discussion and Analysis" section of the firm's Form 10-K filing. Section III of the main paper details how these predictions are constructed.
<i>RoE</i>	<i>RoE</i> is the firm's three-year average return on equity, calculated as the ratio of net income (Compustat: NI) to shareholders' equity (Compustat: SEQ).
<i>Market Cap</i>	<i>Market Cap</i> is the natural logarithm of the product of the firm's stock price (Compustat: PRCC_F) and shares outstanding (Compustat: CSHO).
<i>MB</i>	<i>MB</i> is calculated as the ratio of the firm's market capitalization (Compustat: PRCC_F $\times$ CSHO) to shareholders' equity (Compustat: SEQ).
<i>PE</i>	<i>PE</i> is defined as the ratio of the firm's market capitalization (Compustat: PRCC_F $\times$ CSHO) to its net income (Compustat: NI).

Identifying M&A Targets from Textual Disclosures

Variable	Variable Definition
<i>Leverage</i>	<i>Leverage</i> is the ratio of long-term debt (Compustat: DLTT) to shareholders' equity (Compustat: SEQ).
<i>Tangible Assets</i>	<i>Tangible Assets</i> is defined as the ratio of net property, plant, and equipment (Compustat: PPENT) to a firm's total assets (Compustat: AT).
<i>Sales Growth</i>	<i>Sales Growth</i> is calculated as the three-year average of the percentage change in revenue (Compustat: SALE) since the prior fiscal year.
<i>Liquidity</i>	<i>Liquidity</i> is defined as the three-year average of the sum of cash and cash equivalents (Compustat: CHE) and total receivables (Compustat: RECT) relative to the firm's total assets (Compustat: AT).
<i>Growth-Resource-Mismatch</i>	<i>Growth-Resource-Mismatch</i> is a dummy variable equal to 1 if a firm is classified as low <i>Sales Growth</i> , low <i>Leverage</i> , and high <i>Liquidity</i> or, conversely, high <i>Sales Growth</i> , high <i>Leverage</i> , and low <i>Liquidity</i> , based on the respective annual median, and 0 otherwise.
<i>Takeover Activity</i>	<i>Takeover Activity</i> is a dummy variable equal to 1 if we record at least one takeover offer for one of the firm's industry peers based on the 10-K Text-based Network Industry Classification within the previous year and 0 otherwise.
<i>Industry Profit</i>	<i>Industry Profit</i> is the mean profit of a firm's industry peers based on the 10-K Text-based Network Industry Classification (Hoberg and Phillips (2010), Hoberg and Phillips (2016)) at the TNIC-3 granularity level. Profit is defined as the ratio of net income (Compustat: NI) to sales (Compustat: SALE).
<i>Low Profit</i>	<i>Low Profit</i> is a dummy variable equal to 1 if the firm's profit is below its respective <i>Industry Profit</i> , and 0 otherwise. Profit is defined as the ratio of net income (Compustat: NI) to sales (Compustat: SALE).

Identifying M&A Targets from Textual Disclosures

Variable	Variable Definition
<i>Industry Productivity</i>	<i>Industry Productivity</i> is the mean productivity of a firm's industry peers based on the 10-K Text-based Network Industry Classification (Hoberg and Phillips (2010), Hoberg and Phillips (2016)) at the TNIC-3 granularity level. Productivity is defined as the difference of sales (Compustat: SALE) and cost of goods sold (Compustat: COGS), relative to the number of employees (Compustat: EMP).
<i>Low Productivity</i>	<i>Low Productivity</i> is a dummy variable equal to 1 if the firm's productivity is below its respective <i>Industry Productivity</i> , and 0 otherwise. Productivity is defined as the difference of sales (Compustat: SALE) and cost of goods sold (Compustat: COGS), relative to the number of employees (Compustat: EMP).
<i>Industry Cash Flow</i>	<i>Industry Cash Flow</i> is the mean cash flow generation of a firm's industry peers based on the 10-K Text-based Network Industry Classification (Hoberg and Phillips (2010), Hoberg and Phillips (2016)) at the TNIC-3 granularity level. Cash flow generation is defined as the difference of sales (Compustat: SALE) and the sum of cost of goods sold (Compustat: COGS) and capital expenditures (Compustat: CAPX), relative to the firm's sales (Compustat: SALE).
<i>Low Cash Flow</i>	<i>Low Cash Flow</i> is a dummy variable equal to 1 if the firm's cash flow generation is below its respective <i>Industry Cash Flow</i> , and 0 otherwise. Cash flow generation is defined as the difference of sales (Compustat: SALE) and the sum of cost of goods sold (Compustat: COGS) and capital expenditures (Compustat: CAPX), relative to the firm's sales (Compustat: SALE).
<i>TNIC HHI</i>	<i>TNIC HHI</i> is the Herfindahl-Hirschman index based on the 10-K Text-based Network Industry Classification (Hoberg and Phillips (2016)).
<i>Downstream Sales Growth</i>	<i>Downstream Sales Growth</i> is the mean of the sales growth of firms downstream of the firm according to the 10% granularity version of the vertical textual network industry relatedness classification (Frésard et al. (2020)) in the past year.

Identifying M&A Targets from Textual Disclosures

Variable	Variable Definition
<i>Industrial Production</i>	<i>Industrial Production</i> is the growth rate of the United States' industrial production over the quarter prior to the fiscal year end according to the Federal Reserve Bank of St. Louis.

B. ADDITIONAL FIGURES

Figure A1. ROC Curves of Benchmark Model and RoBERTa: Cost-Sensitive Training

This figure compares the ROC curve of the logit model based on financial controls, year-fixed effects, and industry-fixed effects (light red) with the model also including text-based predictions derived by RoBERTa under cost-sensitive training from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

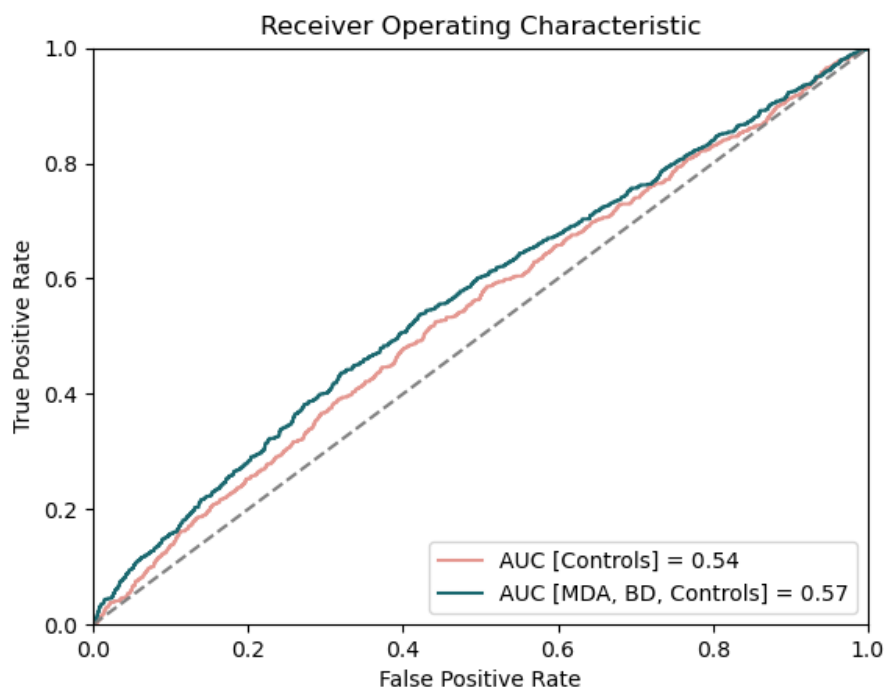


Figure A2. ROC Curves of Benchmark Model and RoBERTa: Additional Industry & Demand Controls

This figure compares the ROC curve of the logit model based on financial controls, industry- and demand controls, year-fixed effects, and industry-fixed effects (light red) with the model also including text-based predictions derived by RoBERTa from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

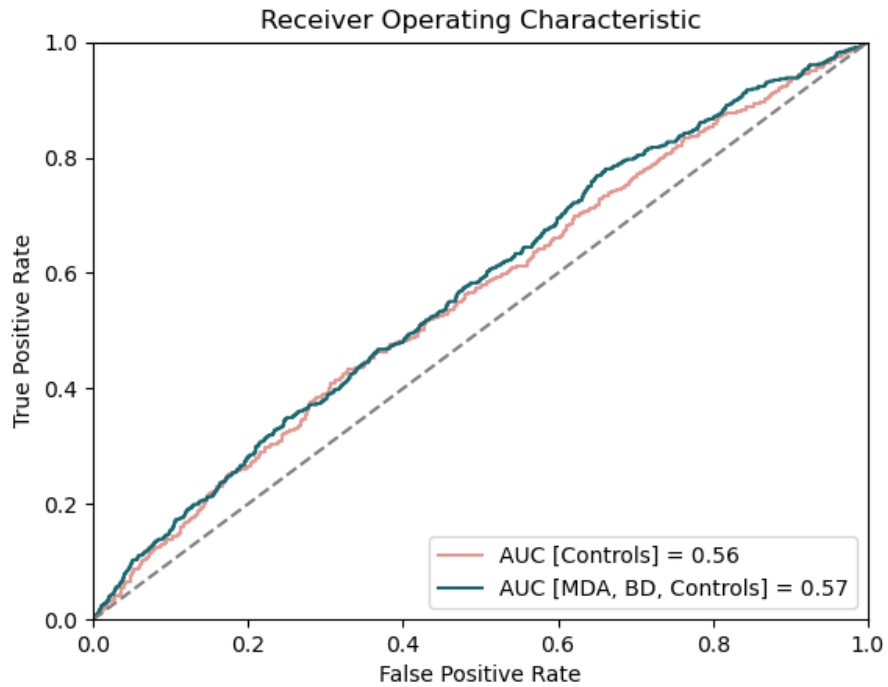


Figure A3. ROC Curves of Benchmark Model and Longformer: Additional Industry & Demand Controls

This figure compares the ROC curve of the logit model based on financial controls, industry- and demand controls, year fixed effects, and industry fixed effects (light red) with the model also including text-based predictions derived by Longformer from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

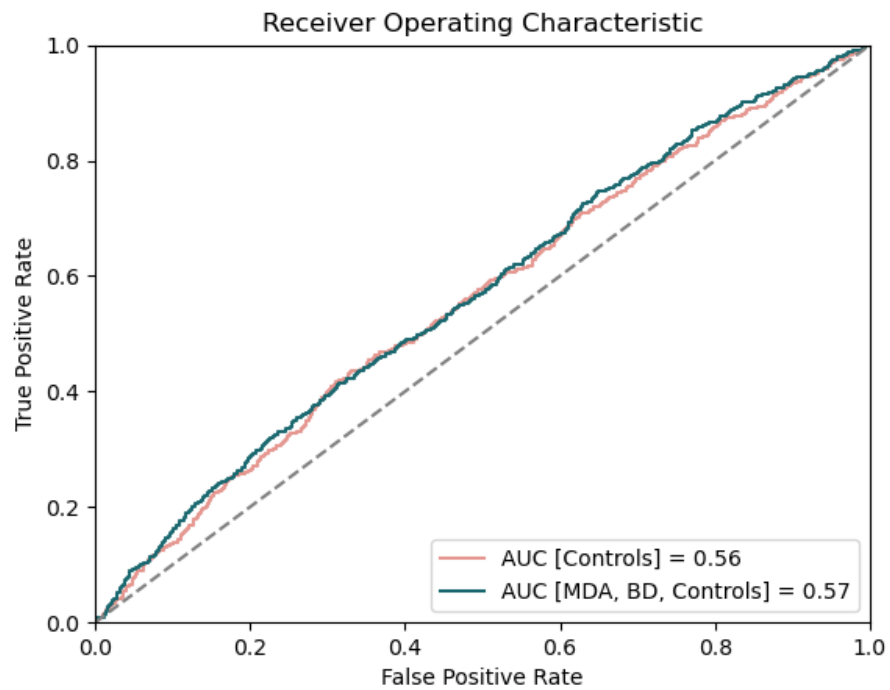


Figure A4. ROC Curves of Benchmark Model and RoBERTa: Financial Controls Only

This figure compares the ROC curve of the logit model based on financial controls without any fixed effects (light red) with the model also including text-based predictions derived by RoBERTa from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

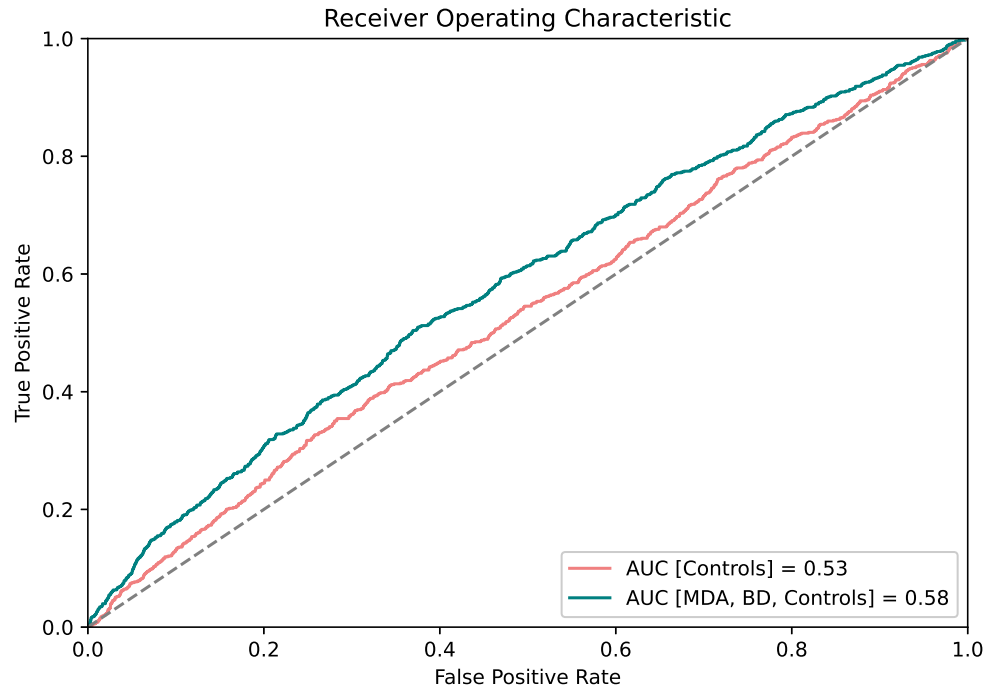


Figure A5. ROC Curves of Benchmark Model and Longformer: Financial Controls Only

This figure compares the ROC curve of the logit model based on financial controls without any fixed effects (light red) with the model also including text-based predictions derived by Longformer from the BD and the MDA (dark green). The dotted grey line corresponds to the performance of a random prediction.

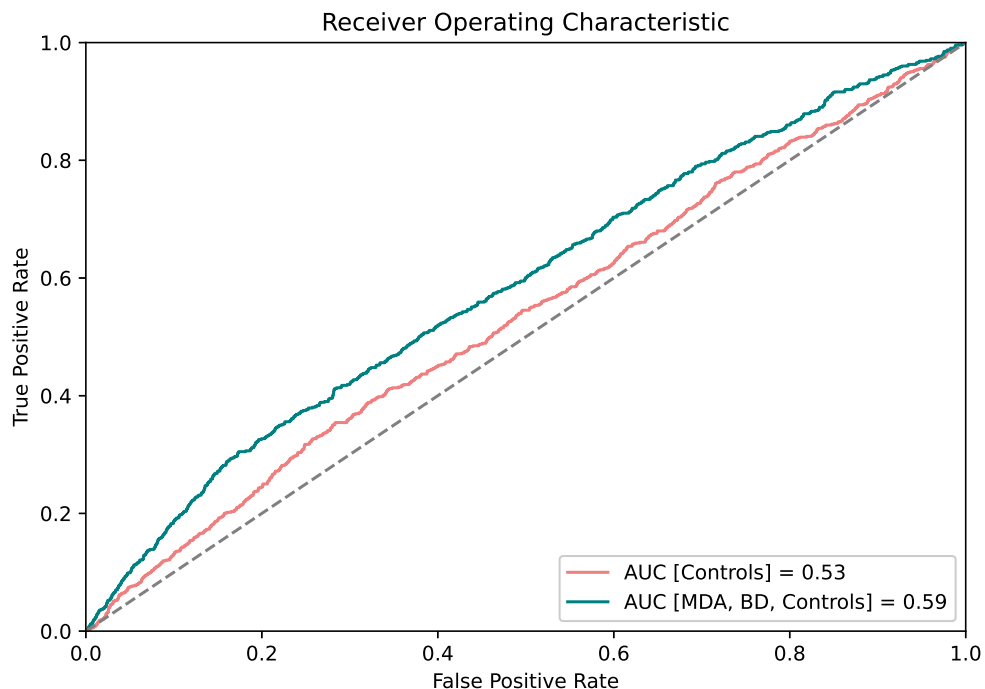
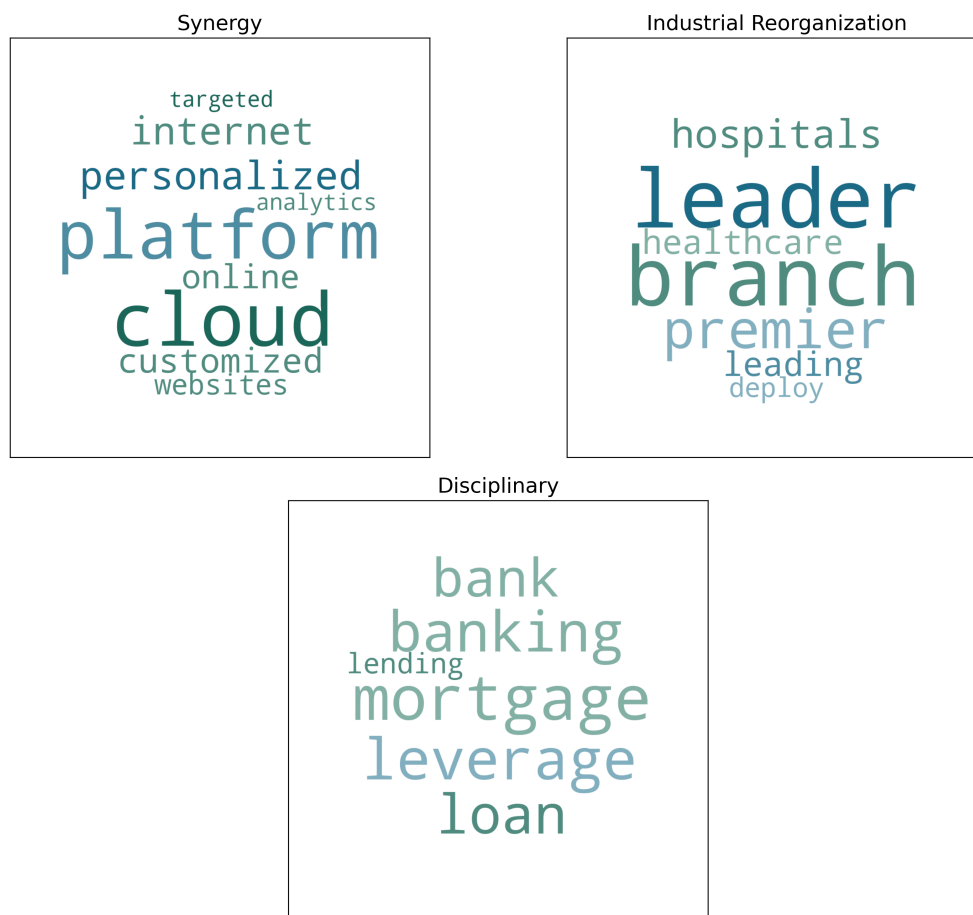


Figure A6. Illustration: Feature Attribution and Takeover Motives

This figure displays the most impactful words for takeover predictions, based on the integrated gradient feature attribution (see Figure 9 in the main paper) by topic. Words are color-coded by document source: blue for BD words and green for MDA words. We *manually* assign words to three canonical classes of takeover motives: “Synergy”, “Industrial Reorganization”, and “Disciplinary”. Please note that this assignment of words to takeover motives is subjective (i.e., it is not the output of a formal classification) and serves illustrative purposes only. Within each motive category, word size reflects the relative rank of its feature attribution importance among BD and MDA words, respectively, assigned to that motive (displayed in Figure 9 in the main paper).



C. ADDITIONAL SUMMARY STATISTICS

Table A2. Annual Filings and M&A Offers by Year – Textual Data

This table reports additional detail on our sample. In particular, we report the number of textual observations based on the training sample (Panel A) and the test sample (Panel B). In this table, we report textual observations *before* adding control variables needed for our analyses on predictions (i.e., reported numbers exceed the final observation count used in the logistic regressions). We assign observations to a year based on the last day of its fiscal year such that “Textual Observations” captures the 10-K filings by year. “Number of M&A Offers” is the number of firms that receive a takeover offer in the subsequent 365 days. “Average Takeover Dummy” is defined as the ratio of “Number of M&A Offers” to the “Textual Observations”.

Year	Textual Observations	Number of M&A Offers	Average Takeover Dummy
<i>Panel A: Training Data</i>			
1993	599	14	0.0234
1994	1,082	53	0.0490
1995	2,198	121	0.0551
1996	4,503	331	0.0735
1997	4,822	385	0.0798
1998	5,072	441	0.0869
1999	5,178	340	0.0657
2000	5,266	310	0.0589
2001	4,873	203	0.0417
2002	4,290	231	0.0538
2003	4,322	192	0.0444
2004	4,461	210	0.0471
2005	3,789	235	0.0620
2006	3,384	206	0.0609
2007	2,945	135	0.0458
2008	2,854	128	0.0448
2009	2,891	143	0.0495
2010	3,706	127	0.0343
2011	3,632	136	0.0374
<i>Panel B: Test Data</i>			
2012	3,793	104	0.0274
2013	4,014	119	0.0296
2014	4,004	134	0.0335
2015	3,929	149	0.0379
2016	3,844	124	0.0323
2017	3,771	139	0.0369
2018	3,694	118	0.0319
2019	3,384	82	0.0242
2020	2,687	119	0.0443

D. ADDITIONAL EMPIRICAL EVIDENCE: ROBUSTNESS

Table A3. AUC by Year – Post-2015 Text-only Predictions

This table reports the out-of-sample AUC values based on text-based predictions only for each year of the test sample. The AUC is computed separately for predictions based on the BD and the MDA, which are derived by RoBERTa. For this table, we retrain RoBERTa based on textual data starting from 1997 through 2015.

Year	AUC	
	RoBERTa BD	RoBERTa MDA
2016	0.6140	0.5937
2017	0.5811	0.5539
2018	0.5775	0.5001
2019	0.5511	0.5390
2020	0.5889	0.5467

Table A4. AUC of Different Model Specifications - Cost-Sensitive Training

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. Compared to Table 5 in the main paper, the training of the textual prediction model was explicitly cost-sensitive, penalizing the failure to identify a target through a weighted loss function. Our *Controls* model, which includes financial controls as well as year- and industry-fixed effects, serves as a benchmark. The uplift in AUC reported in the last column is denoted relative to the *Controls* model. The significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift compared to controls
<i>Controls</i>	0.5433	-
<i>RoBERTa BD & Controls</i>	0.5697	-
		0.0264***
		(4.52)
<i>RoBERTa MDA & Controls</i>	0.5508	0.0075***
		(2.66)
<i>RoBERTa BD, RoBERTa MDA, & Controls</i>	0.5665	0.0232***
		(3.65)

Table A5. Correlations of Text-based Predictions

This table reports firm-year-level pairwise correlations between takeover prediction probabilities generated from *RoBERTa BD*, *RoBERTa MDA*, *Longformer BD*, and *Longformer MDA* in the test sample. All variables are described in Table A1 in the Online Appendix. *p*-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	<i>RoBERTa BD</i>	<i>RoBERTa MDA</i>	<i>Longformer BD</i>	<i>Longformer MDA</i>
<i>RoBERTa BD</i>	1.0000			
<i>RoBERTa MDA</i>	0.6610*** (0.0000)	1.0000		
<i>Longformer BD</i>	0.6799*** (0.0000)	0.6404*** (0.0000)	1.0000	
<i>Longformer MDA</i>	0.6261*** (0.0000)	0.7645*** (0.0000)	0.6826*** (0.0000)	1.0000

Identifying M&A Targets from Textual Disclosures

Table A6. Textual Disclosures and Takeover Likelihood – RoBERTa Standardized

This table reports the coefficients from OLS regressions. The conducted regressions correspond to the last rolling-window (i.e., using data from 2012 to 2019). The dependent variable is the *Takeover Dummy*, which is equal to 1 if a firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise. The main explanatory variables are takeover predictions derived by RoBERTa from the BD and the MDA. All independent variables and the dependent variable have been standardized to zero mean and unit variance. A constant term is included but not reported. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: <i>Takeover Dummy</i>						
	1	2	3	4	5	6	7
<i>RoBERTa BD</i>	0.0688*** (9.25)		0.0664*** (6.71)		0.0687*** (8.26)		0.0645*** (6.42)
<i>RoBERTa MDA</i>		0.0474*** (6.36)	0.0037 (0.38)			0.0470*** (5.26)	0.0077 (0.71)
<i>RoE</i>				-0.0066 (-0.86)	-0.0054 (-0.70)	-0.0060 (-0.79)	-0.0052 (-0.68)
<i>Market Cap</i>				-0.0274*** (-3.32)	-0.0303*** (-3.67)	-0.0312*** (-3.77)	-0.0307*** (-3.71)
<i>MB</i>				-0.0066 (-0.81)	-0.0082 (-1.02)	-0.0080 (-0.99)	-0.0084 (-1.04)
<i>PE</i>				0.0059 (0.78)	0.0044 (0.58)	0.0052 (0.69)	0.0044 (0.59)
<i>Leverage</i>				0.0195** (2.45)	0.0189** (2.36)	0.0196** (2.46)	0.0191** (2.40)
<i>Tangible Assets</i>				0.0220** (2.02)	0.0278** (2.54)	0.0286*** (2.60)	0.0284*** (2.59)
<i>Sales Growth</i>				-0.0153** (-2.06)	-0.0154** (-2.07)	-0.0151** (-2.03)	-0.0153** (-2.06)
<i>Liquidity</i>				0.0185* (1.74)	0.0150 (1.41)	0.0153 (1.44)	0.0146 (1.37)
<i>Growth-Resource-Mismatch</i>				-0.0132 (-1.51)	-0.0124 (-1.41)	-0.0129 (-1.47)	-0.0123 (-1.41)
<i>Takeover Activity</i>				0.0465*** (5.44)	0.0376*** (4.36)	0.0414*** (4.81)	0.0373*** (4.33)
Year Fixed Effects	No	No	No	Yes	Yes	Yes	Yes
Industry Fixed Effects	No	No	No	Yes	Yes	Yes	Yes
N	17,967	17,967	17,967	17,967	17,967	17,967	17,967
Adjusted R2	0.0047	0.0022	0.0046	0.0038	0.0075	0.0053	0.0075

Identifying M&A Targets from Textual Disclosures

Table A7. Textual Disclosures and Takeover Likelihood – Longformer Standardized

This table reports the coefficients from OLS regressions. The conducted regressions correspond to the last rolling-window (i.e., using data from 2012 to 2019). The dependent variable is the *Takeover Dummy*, which is equal to 1 if a firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise. The main explanatory variables are takeover predictions derived by Longformer from the BD and the MDA. All independent variables and the dependent variable have been standardized to zero mean and unit variance. A constant term is included but not reported. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: <i>Takeover Dummy</i>						
	1	2	3	4	5	6	7
<i>Longformer BD</i>	0.0588*** (7.90)		0.0265*** (2.61)		0.0588*** (6.90)		0.0308*** (2.95)
<i>Longformer MDA</i>		0.0655*** (8.80)	0.0474*** (4.67)			0.0700*** (7.76)	0.0511*** (4.61)
<i>RoE</i>				-0.0066 (-0.86)	-0.0058 (-0.76)	-0.0054 (-0.70)	-0.0053 (-0.69)
<i>Market Cap</i>				-0.0274*** (-3.32)	-0.0372*** (-4.45)	-0.0340*** (-4.11)	-0.0374*** (-4.47)
<i>MB</i>				-0.0066 (-0.81)	-0.0079 (-0.97)	-0.0080 (-0.99)	-0.0083 (-1.03)
<i>PE</i>				0.0059 (0.78)	0.0047 (0.62)	0.0038 (0.51)	0.0037 (0.50)
<i>Leverage</i>				0.0195** (2.45)	0.0195** (2.44)	0.0194** (2.44)	0.0194** (2.44)
<i>Tangible Assets</i>				0.0220** (2.02)	0.0242** (2.22)	0.0294*** (2.68)	0.0286*** (2.61)
<i>Sales Growth</i>				-0.0153** (-2.06)	-0.0152** (-2.04)	-0.0146** (-1.96)	-0.0148** (-1.98)
<i>Liquidity</i>				0.0185* (1.74)	0.0101 (0.94)	0.0082 (0.77)	0.0066 (0.62)
<i>Growth-Resource-Mismatch</i>				-0.0132 (-1.51)	-0.0108 (-1.24)	-0.0114 (-1.30)	-0.0106 (-1.21)
<i>Takeover Activity</i>				0.0465*** (5.44)	0.0369*** (4.26)	0.0372*** (4.32)	0.0347*** (4.00)
Year Fixed Effects	No	No	No	Yes	Yes	Yes	Yes
Industry Fixed Effects	No	No	No	Yes	Yes	Yes	Yes
N	17,967	17,967	17,967	17,967	17,967	17,967	17,967
Adjusted R2	0.0034	0.0042	0.0046	0.0038	0.0064	0.0071	0.0075

Identifying M&A Targets from Textual Disclosures

Table A8. Textual Disclosures and Takeover Likelihood – Alt. Industry & Demand Controls

This table reports the marginal effects calculated across all actual observations from logit regressions. The conducted regressions correspond to the last rolling-window (i.e., using data from 2012 to 2019). The dependent variable is the *Takeover Dummy*, which is equal to 1 if a firm receives a takeover offer within 365 days after the 10-K filing date and 0 otherwise. The main explanatory variables are takeover predictions derived by RoBERTa and Longformer from BD and MDA, respectively. Panel A (Panel B) includes *Industry Productivity* and *Low Productivity* (*Industry Cash Flow* and *Low Cash Flow*) as additional control variables for the firm's productivity (cash flow generation) relative to its industry peers. A constant term is included but not reported. z-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Dependent Variable: <i>Takeover Dummy</i>					
	1	2	3	4	5	6
Panel A: Industry Productivity						
<i>RoBERTa BD</i>	0.3620*** (4.89)		0.2935*** (3.22)			
<i>RoBERTa MDA</i>		0.3228*** (3.98)	0.1344 (1.35)			
<i>Longformer BD</i>				0.3764*** (3.61)		0.1797 (1.38)
<i>Longformer MDA</i>					0.5598*** (4.01)	0.4282** (2.48)
<i>Industry Productivity</i>	0.0000 (0.16)	0.0000 (0.10)	0.0000 (0.13)	0.0000 (0.14)	0.0000 (0.15)	0.0000 (0.13)
<i>Low Productivity</i>	-0.0020 (-0.49)	-0.0015 (-0.37)	-0.0019 (-0.45)	-0.0020 (-0.47)	-0.0017 (-0.41)	-0.0018 (-0.43)
<i>TNIC HHI</i>	-0.0219* (-1.94)	-0.0240** (-2.12)	-0.0225** (-1.99)	-0.0223** (-1.97)	-0.0237** (-2.09)	-0.0229** (-2.02)
<i>Downstream Sales Growth</i>	-0.0033 (-0.20)	-0.0023 (-0.14)	-0.0010 (-0.06)	-0.0045 (-0.27)	-0.0013 (-0.08)	-0.0009 (-0.05)
<i>Industrial Production</i>	-0.0121 (-0.02)	-0.0140 (-0.03)	-0.0484 (-0.10)	-0.0231 (-0.05)	-0.0433 (-0.09)	-0.1095 (-0.22)
Financial Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year & Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
N	11,311	11,311	11,311	11,311	11,311	11,311
Pseudo R2	0.0366	0.0348	0.0372	0.0340	0.0349	0.0355
Panel B: Industry Cash Flow Generation						
<i>RoBERTa BD</i>	0.3511*** (4.75)		0.2958*** (3.24)			
<i>RoBERTa MDA</i>		0.3116*** (3.84)	0.1129 (1.13)			
<i>Longformer BD</i>				0.3618*** (3.47)		0.1819 (1.39)
<i>Longformer MDA</i>					0.5155*** (3.70)	0.3712** (2.14)
<i>Industry Cash Flow</i>	0.1591 (1.09)	0.0831 (0.57)	0.1253 (0.86)	0.1136 (0.78)	0.1715 (1.16)	0.1689 (1.15)
<i>Low Cash Flow</i>	-0.0012 (-0.29)	-0.0013 (-0.32)	-0.0013 (-0.31)	-0.0009 (-0.22)	-0.0008 (-0.19)	-0.0009 (-0.22)
<i>TNIC HHI</i>	-0.0232** (-2.04)	-0.0250** (-2.19)	-0.0233** (-2.05)	-0.0236** (-2.08)	-0.0251** (-2.20)	-0.0243** (-2.14)
<i>Downstream Sales Growth</i>	-0.0066 (-0.42)	-0.0055 (-0.35)	-0.0049 (-0.31)	-0.0074 (-0.47)	-0.0053 (-0.34)	-0.0049 (-0.31)
<i>Industrial Production</i>	-0.0555 (-0.12)	-0.0191 (-0.04)	-0.0436 (-0.09)	-0.0243 (-0.05)	-0.0370 (-0.08)	-0.0832 (-0.18)
Financial Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year & Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
N	11,818	11,818	11,818	11,818	11,818	11,818
Pseudo R2	0.0352	0.0332	0.0355	0.0326	0.0333	0.0338

Table A9. AUC of Different Model Specifications - Increased Number of Industry Controls

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. We show AUCs for models also including financial controls, year-fixed effects, and industry-fixed effects. Compared to Table 5 in the main paper, the regressions underlying the AUC reported here use industry dummies according to the Fama and French (1997) 48 industry portfolio classification. The uplift in AUC compared to the model only based on these controls is listed in the third column. Significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift compared to controls
<i>Controls</i>	0.5419	-
<i>RoBERTa BD & Controls</i>	0.5650	0.0232*** (4.21)
<i>RoBERTa MDA & Controls</i>	0.5491	0.0073** (2.21)
<i>RoBERTa BD, RoBERTa MDA, & Controls</i>	0.5641	0.0223*** (4.03)
<i>Longformer BD & Controls</i>	0.5519	0.0101** (2.08)
<i>Longformer MDA & Controls</i>	0.5656	0.0238*** (3.98)
<i>Longformer BD, Longformer MDA & Controls</i>	0.5643	0.0224*** (3.54)

Table A10. AUC of Different Model Specifications - Alternative Industry Classification

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. We show AUCs for models also including financial controls, industry- and demand controls, year-fixed effects, and industry-fixed effects. Compared to Table 5 in the main paper, the regressions underlying the AUC reported here use industry dummies according to the Hoberg and Phillips (2016) 100 fixed industry classification and the industry- and demand controls. Due to the availability of the industry classification, the sample period for this analysis ends in 2015. The uplift in AUC compared to the model only based on these controls is listed in the third column. Significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift compared to controls
<i>Controls</i>	0.5311	-
<i>RoBERTa BD & Controls</i>	0.5511	0.0200*** (2.85)
<i>RoBERTa MDA & Controls</i>	0.5349	0.0039 (1.08)
<i>RoBERTa BD, RoBERTa MDA, & Controls</i>	0.5503	0.0192** (2.51)
<i>Longformer BD & Controls</i>	0.5454	0.0143** (2.57)
<i>Longformer MDA & Controls</i>	0.5474	0.0163* (1.96)
<i>Longformer BD, Longformer MDA & Controls</i>	0.5478	0.0167** (2.13)

Table A11. AUC of RoBERTa vs. Longformer

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. The uplift in AUC reported in the last column is denoted relative to the corresponding RoBERTa model, i.e., we compare predictions based on the same textual data. The significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift compared to RoBERTa
<i>RoBERTa BD & Controls</i>	0.5736	-
<i>RoBERTa MDA & Controls</i>	0.5576	-
<i>RoBERTa BD, RoBERTa MDA, & Controls</i>	0.5729	-
<i>Longformer BD & Controls</i>	0.5601	-0.0135*** (-2.60)
<i>Longformer MDA & Controls</i>	0.5780	0.0204*** (3.50)
<i>Longformer BD, Longformer MDA & Controls</i>	0.5759	0.0030 (0.47)

Table A12. AUC by Year – Text-only Predictions

This table reports the out-of-sample AUC values based on text-based predictions only for each year of the test sample. The AUC is computed separately for BD and MDA-based predictions generated by each of RoBERTa and Longformer.

Year	AUC			
	RoBERTa BD	RoBERTa MDA	Longformer BD	Longformer MDA
2012	0.6020	0.5871	0.6214	0.6361
2013	0.6179	0.5958	0.5596	0.6259
2014	0.5923	0.5514	0.5654	0.5751
2015	0.6110	0.5879	0.6391	0.6029
2016	0.6236	0.5886	0.5834	0.6121
2017	0.5465	0.5600	0.5436	0.5762
2018	0.5524	0.5129	0.5316	0.5302
2019	0.5525	0.5266	0.5245	0.5308
2020	0.5644	0.5494	0.5355	0.5553

Table A13. Out-of-Sample Performance of Takeover Predictions – Portfolio Evidence

This table reports the out-of-sample takeover rates of annual quintile portfolios formed on estimated takeover probabilities in the test sample. *Controls* refers to takeover predictions based on the financial variables and fixed effects, *BD* and *MDA* to the predictions of the respective Form 10-K filing, *Text* to the combination of *BD* and *MDA*, and *Complete* to the combination of *Controls* and *Text*. t-statistics, which are reported in parentheses, refer to the difference between the extreme portfolios formed based on predicted takeover probability and are based on standard errors corrected for heteroskedasticity (White (1980)).

	Controls	BD	MDA	Text	Complete
<i>Panel A: RoBERTa Predictions</i>					
1	0.0284	0.0250	0.0298	0.0261	0.0290
2	0.0372	0.0324	0.0389	0.0324	0.0276
3	0.0424	0.0460	0.0370	0.0441	0.0418
4	0.0424	0.0461	0.0432	0.0472	0.0492
5	0.0566	0.0575	0.0581	0.0572	0.0595
5-1	0.0282	0.0325	0.0282	0.0311	0.0305
t(5-1)	(5.88)	(6.87)	(5.79)	(6.54)	(6.24)
<i>Panel B: Longformer Predictions</i>					
1	0.0284	0.0293	0.0275	0.0281	0.0290
2	0.0372	0.0395	0.0364	0.0358	0.0287
3	0.0424	0.0384	0.0392	0.0404	0.0458
4	0.0424	0.0435	0.0449	0.0381	0.0378
5	0.0566	0.0563	0.0589	0.0646	0.0657
5-1	0.0282	0.0271	0.0314	0.0365	0.0368
t(5-1)	(5.88)	(5.62)	(6.48)	(7.30)	(7.28)

Table A14. AUC of Different Model Specifications – Financial Controls Only

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. Our *Financial Controls* model includes financial controls only, i.e., we exclude year- and industry fixed effects, and serves as a benchmark. The uplift in AUC reported in the last column is denoted relative to the *Financial Controls* model. The significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift compared to financial controls
<i>Financial Controls</i>	0.5340	-
<i>RoBERTa BD & Financial Controls</i>	0.5838	0.0499*** (5.55)
<i>RoBERTa MDA & Financial Controls</i>	0.5675	0.0336*** (4.44)
<i>RoBERTa BD, RoBERTa MDA, & Financial Controls</i>	0.5838	0.0498*** (5.58)
<i>Longformer BD & Financial Controls</i>	0.5682	0.0342*** (3.86)
<i>Longformer MDA & Financial Controls</i>	0.5892	0.0552*** (5.37)
<i>Longformer BD, Longformer MDA, & Financial Controls</i>	0.5860	0.0520*** (5.04)

Table A15. AUC of Different Model Specifications – Reverse

This table reports the out-of-sample AUC values for logit models based on different sets of predictors. We report the AUC uplifts which we obtain upon adding *Controls*, which includes financial controls as well as year- and industry-fixed effects, to the respective text-based predictions. The significances of differences in AUC are calculated using the DeLong test. Corresponding z-values are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Predictor set	AUC	Uplift through controls
<i>RoBERTa BD</i>	0.5778	-
<i>RoBERTa MDA</i>	0.5527	-
<i>RoBERTa BD & Controls</i>	0.5736	-0.0042 (-0.43)
<i>RoBERTa MDA & Controls</i>	0.5576	0.0049 (0.44)
<i>Longformer BD</i>	0.5580	-
<i>Longformer MDA</i>	0.5800	-
<i>Longformer BD & Controls</i>	0.5601	0.0021 (0.21)
<i>Longformer MDA & Controls</i>	0.5780	-0.0020 (-0.22)

Table A16. Stock Market Performance of Predicted Targets — Gross Returns

This table reports the extreme legs of monthly quintile portfolio sorts based on the estimated probability to become an M&A target derived from the logistic regressions including RoBERTa (Panel A) and Longformer (Panel B) as well as the aforementioned control variables and fixed effects for the period from March 2013 to February 2022. The table presents subsequent equal-weighted monthly gross returns R . The portfolio excess gross returns (α) and the corresponding factor loadings for the market (β_{MKT}), size (β_{SMB}), value (β_{HML}), momentum (β_{WML}), profitability (β_{ROE}), investment (β_{INV}), liquidity (β_{LIQ}), and takeover vulnerability factors (β_{TMA}) are also provided. The t-statistics in parentheses refer to the difference portfolio and are based on standard errors following Newey and West (1987) using twelve lags. Returns and alphas are stated in %.

	R	α	β_{MKT}	β_{SMB}	β_{HML}	β_{WML}	β_{ROE}	β_{INV}	β_{LIQ}	β_{TMA}
<i>Panel A: RoBERTa</i>										
1	1.59	0.45	1.01	0.55	0.17	-0.10	0.01	0.00	-0.01	-0.03
5	2.65	1.50	0.99	0.56	0.14	0.10	-0.37	0.02	0.01	0.16
5-1	1.06	1.05	-0.01	0.01	-0.03	0.20	-0.38	0.02	0.02	0.20
t(5-1)	(3.74)	(3.29)	(-0.27)	(0.07)	(-0.66)	(2.49)	(-1.81)	(0.18)	(0.22)	(1.82)
<i>Panel B: Longformer</i>										
1	1.81	0.66	1.03	0.53	0.21	-0.09	-0.01	-0.08	-0.06	-0.04
5	2.55	1.46	0.96	0.52	0.07	0.06	-0.34	0.08	0.09	0.10
5-1	0.74	0.80	-0.08	-0.01	-0.15	0.16	-0.33	0.16	0.15	0.14
t(5-1)	(2.41)	(2.61)	(-1.37)	(-0.06)	(-2.60)	(2.18)	(-2.52)	(1.15)	(1.15)	(1.52)

Table A17. Stock Market Performance of Predicted Targets — Time Split

This table reports the extreme legs of monthly quintile portfolio sorts based on the estimated probability to become an M&A target derived from the logistic regressions including RoBERTa (Panels A and B) and Longformer (Panels C and D) as well as the aforementioned control variables and fixed effects separately for the first and second half of the sample period from March 2013 to February 2022, respectively. The table presents subsequent equal-weighted monthly returns R after accounting for transaction costs. The portfolio excess returns (α) and the corresponding factor loadings for the market (β_{MKT}), size (β_{SMB}), value (β_{HML}), momentum (β_{WML}), profitability (β_{ROE}), investment (β_{INV}), liquidity (β_{LIQ}), and takeover vulnerability (β_{TMA}) are also provided. The t-statistics in parentheses refer to the difference portfolio and are based on standard errors following Newey and West (1987) using twelve lags. Returns and alphas are stated in %.

	R	α	β_{MKT}	β_{SMB}	β_{HML}	β_{WML}	β_{ROE}	β_{INV}	β_{LIQ}	β_{TMA}
<i>Panel A: RoBERTa – First Half of the Sample</i>										
1	1.40	0.22	0.97	0.63	0.30	0.05	0.12	-0.26	-0.10	0.00
5	2.62	1.81	0.98	0.44	0.09	0.25	-0.75	0.09	-0.05	-0.08
5-1	1.01	1.40	0.02	-0.18	-0.24	0.21	-0.85	0.37	0.06	-0.10
t(5-1)	(2.22)	(1.51)	(0.16)	(-0.46)	(-1.66)	(0.97)	(-1.62)	(0.95)	(0.30)	(-0.32)
<i>Panel B: RoBERTa – Second Half of the Sample</i>										
1	1.58	0.40	1.01	0.49	0.16	-0.19	-0.08	0.00	-0.01	-0.11
5	2.44	1.29	0.95	0.74	0.18	0.07	-0.12	0.08	0.09	0.27
5-1	0.67	0.72	-0.06	0.26	0.10	0.29	-0.06	0.01	0.13	0.36
t(5-1)	(2.24)	(2.33)	(-0.92)	(3.00)	(2.29)	(7.93)	(-0.81)	(0.09)	(0.90)	(2.90)
<i>Panel C: Longformer – First Half of the Sample</i>										
1	1.97	0.94	0.99	0.54	0.25	-0.01	-0.05	-0.13	-0.08	-0.08
5	2.27	1.23	1.00	0.28	0.11	0.28	-0.54	0.07	-0.04	0.07
5-1	0.08	0.10	0.02	-0.27	-0.17	0.30	-0.49	0.21	0.06	0.13
t(5-1)	(0.23)	(0.14)	(0.10)	(-0.80)	(-0.86)	(1.30)	(-1.24)	(0.43)	(0.32)	(0.69)
<i>Panel D: Longformer – Second Half of the Sample</i>										
1	1.44	0.21	1.06	0.54	0.21	-0.14	-0.02	-0.05	0.00	-0.15
5	2.60	1.51	0.90	0.78	0.12	0.00	-0.16	0.08	0.15	0.11
5-1	0.98	1.14	-0.16	0.26	-0.02	0.17	-0.15	0.06	0.19	0.24
t(5-1)	(2.64)	(4.75)	(-2.94)	(2.32)	(-0.26)	(3.09)	(-2.23)	(0.43)	(1.04)	(1.50)

Table A18. Picking Takeover Targets Out-of-Sample – Thresholds

This table reports the annual classification thresholds used to convert predicted takeover probabilities into binary takeover predictions in the out-of-sample evaluation. For each model and year, the threshold is estimated to optimize the F_1 -score using all preceding years of the test period. The final row reports the average threshold across years.

Year	Controls	RoBERTa Text	RoBERTa Complete	Longformer Text	Longformer Complete
2013	0.0600	0.0443	0.1918	0.1165	0.0791
2014	0.0645	0.0447	0.0603	0.0430	0.0832
2015	0.0621	0.0472	0.0490	0.0435	0.0650
2016	0.0629	0.0510	0.0556	0.0639	0.0709
2017	0.0618	0.0617	0.0611	0.0624	0.0662
2018	0.0672	0.0563	0.0591	0.0593	0.0687
2019	0.0646	0.0546	0.0693	0.0567	0.0668
2020	0.0594	0.0554	0.0591	0.0530	0.0684
Average	0.0628	0.0519	0.0757	0.0623	0.0710

Table A19. Picking Takeover Targets Out-of-Sample – Decile-Level Performance – Alternative Models

This table reports out-of-sample prediction performance across deciles of predicted probabilities separately for the *RoBERTa BD & RoBERTa MDA* model (Panel A) and the *Longformer BD & Longformer MDA* model (Panel B). For each model, firm-years are sorted into deciles based on predicted takeover likelihoods, where Decile 10 contains observations with the highest predicted probabilities and Decile 1 contains observations with the lowest predicted probabilities. “True Takeovers” reports the number of realized takeover targets in a given decile, “Takeover Rate” the fraction of firm-years that become takeover targets, and “TP” the number of true positive predictions. Takeover predictions are assigned using the model-specific classification thresholds reported in Table A18 in the Online Appendix. F_1 -scores and accuracy are computed separately within each decile using only observations assigned to that decile.

Decile	True Takeovers	Takeover Rate	TP	F_1 -Score	Accuracy
<i>Panel A: RoBERTa BD & RoBERTa MDA</i>					
10	119	0.0675	116	0.1258	0.0851
9	83	0.0472	36	0.0860	0.5646
8	83	0.0472	1	0.0177	0.9369
7	82	0.0467	0	0.0000	0.9533
6	94	0.0535	0	0.0000	0.9465
5	62	0.0353	0	0.0000	0.9647
4	52	0.0296	0	0.0000	0.9704
3	61	0.0347	0	0.0000	0.9653
2	37	0.0210	0	0.0000	0.9790
1	55	0.0312	0	0.0000	0.9688
<i>Panel B: Longformer BD & Longformer MDA</i>					
10	125	0.0709	102	0.1248	0.1884
9	103	0.0586	28	0.0926	0.6875
8	55	0.0312	21	0.0802	0.7261
7	78	0.0444	13	0.0906	0.8514
6	77	0.0438	0	0.0000	0.9562
5	65	0.0370	0	0.0000	0.9630
4	68	0.0387	0	0.0000	0.9613
3	58	0.0330	0	0.0000	0.9670
2	51	0.0290	0	0.0000	0.9710
1	48	0.0272	0	0.0000	0.9728

E. ADDITIONAL EMPIRICAL EVIDENCE: RANDOM SPLIT

All previously reported results throughout the main paper and this Online Appendix are built on a sample split along the time axis, i.e., sorting observations through 2011 into the training sample and observations after 2011 into the test sample (as described in Section III of the paper). This approach is essential to ensure that there is no leakage of forward-looking information. At the same time, this conservative approach drags on the predictive ability of our deep learning algorithms: given that time-varying industry shocks (and many other time-varying factors) are key determinants of M&A activity, RoBERTa and Longformer will fail to pick up any systematic changes in determinants of M&A targets after 2011. The impact of temporal changes in prediction quality is also apparent in Table A12 as the AUC deteriorates over time.

One alternative approach when employing machine learning for prediction tasks is a random split, i.e., sorting observations randomly into training and test samples (irrespective of the time). We believe that this approach is less suitable for M&A target prediction as it entails the implicit use of forward-looking information. Assume for instance that a manager discusses increased merger activity in her firm’s industry within its 10-K filing in the previous year. If this observation is sorted into the training sample and there are any industry peers from the previous year in the test sample, this will clearly inflate RoBERTa’s and Longformer’s prediction quality; a time split strictly avoids such issues.

Nonetheless, a random split is potentially helpful as it also enables us to address the fact that the *Takeover Dummy* is a highly unbalanced variable (only approximately 4% of firm-year observations are associated with a takeover offer). While we abstain from balancing the *Takeover Dummy* in all previous tests (see, e.g., Palepu (1986) for a methodological criticism of oversampling in the context of M&A prediction), we do so for illustrative purposes in the following. In particular, we select all

firms that receive a takeover offer and then match each firm with a firm-year observation, where no takeover offer is received. Specifically, we match exactly on the same year and two-digit SIC industry and then use the nearest neighbor based on the firm's market capitalization (where available). We randomly sort 70% of the observations of this matched sample into the training sample. Then, we train RoBERTa to identify takeover targets in this training sample and evaluate the predictive performance based on the remaining 30% in the holdout sample (after conducting a hyperparameter optimization). Since this procedure (artificially) increases the average *Takeover Dummy* in the training sample, RoBERTa is more likely to directly label observations as "targets".

Table A20. Text-based Predictions - Random Split

This table reports the F_1 Score, Accuracy, AUC, as well as the elements of the confusion matrix (TP , TN , FP , FN) for RoBERTa predictions based on the BD and the MDA. These predictions are based on a random split of a sample that contains all takeovers in our main sample as well as matching observations.

Predictor set	F_1 Score	Accuracy	AUC	TP	TN	FP	FN
<i>BD Prediction</i>	0.5271	0.5308	0.5561	759	782	675	687
<i>MDA Prediction</i>	0.4744	0.5260	0.5449	621	906	551	825

Table [A20](#) reports the performance metrics for RoBERTa's predictions based on the BD and the MDA in the holdout sample. The results confirm three key insights from the main paper. First, the information extracted by RoBERTa from the 10-K filings is relevant to the prediction of M&A targets (five out of six metrics $> 50\%$) *even* when using firms from the same industry-year and of a similar size as counterfactual. Second, while RoBERTa is able to uncover takeover-relevant information, the market for corporate control is clearly also driven by other factors (e.g., bidder-specific motives), which cannot be uncovered based on the target's annual financial report. Third, in line with all prior analyses, the information contained in the BD seems to be more relevant than the MDA when relying on RoBERTa's predictions.

REFERENCES

- Fama, E. F., and K. R. French. “Industry costs of equity.” *Journal of Financial Economics*, 43 (1997), 153–193.
- Frésard, L.; G. Hoberg; and G. Phillips. “Innovation activities and integration through vertical acquisitions.” *Review of Financial Studies*, 33 (2020), 2937–2976.
- Hoberg, G., and G. Phillips. “Product market synergies and competition in mergers and acquisitions: A text-based analysis.” *Review of Financial Studies*, 23 (2010), 3773–3811.
- Hoberg, G., and G. Phillips. “Text-based network industries and endogenous product differentiation.” *Journal of Political Economy*, 124 (2016), 1423–1465.
- Newey, W. K., and K. D. West. “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix.” *Econometrica*, 55 (1987), 703–708.
- Palepu, K. G. “Predicting takeover targets: A methodological and empirical analysis.” *Journal of Accounting and Economics*, 8 (1986), 3–35.
- White, H. “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity.” *Econometrica*, 48 (1980), 817–838.