Contextualizing Tax Outcomes:
The Informational Role of Financial Statement Narratives

Andrew Belnap
University of Texas at Austin
andrew.belnap@mccombs.utexas.edu

Till Muenster
University of Mannheim
till.muenster@uni-mannheim.de

February 2026

Preliminary. Please do not distribute.

Abstract. We use large language models to examine how narrative text in financial statements contextualizes accounting numbers. We extend prior research, which broadly finds that text is informative, by explicitly modeling *context* as an interaction between text and numbers that improves the mapping to firm outcomes. Tax outcomes are an ideal setting for this analysis because the link between an accounting number (pre-tax income) and a firm outcome (tax expense) is both conceptually grounded and systematically distorted by reporting and regulatory rules that narrative disclosures can help explain. Using embeddings derived from management discussion and analysis (MD&A) sections of 10-K filings, we train deep neural networks that learn how textual context alters the relation between pre-tax income and tax expense. We show that context from the MD&A has significant explanatory power, improving the mapping between book income and tax outcomes by 17.6% to 23.9%. In contrast, income tax footnote narratives often obscure rather than clarify this mapping. Moreover, we show that MD&A context is useful to financial statement users—e.g., by improving analysts' effective tax rate forecasts. Finally, we provide insights into the relation between context informativeness and the narratives' underlying disclosure topics,. Collectively, our findings demonstrate the value of contextual information in understanding distortions between book and tax numbers.

Keywords: 10-K filings; accruals; book-tax differences; cash; contextual information; deep neural networks; disclosure; embeddings; machine learning; tax outcomes.
JEL Classification: C45, C58, H25, H26, M41, M42, K42.

We thank Kay Blaufus, Matthias Breuer, Ulf Brüggemann, Shuping Chen, Holger Daske, Lisa De Simone, Philipp Dörrenberg, Adam Du Pon (discussant), Alex Edwards, Joachim Gassen, Jeff Hales, Nick Hallman, Jeff Hoopes, Adrian Kubata, Harm Schütt, Barbara Stage, Maximilian Todtenhaupt, Jaron Wilde, and Brady Williams, Ricky Yao (discussant), as well as other conference and workshop participants at the 5U Accounting Conference 2025, Hawai'i Accounting Research Conference 2025, NTA Annual Conference 2025, Leibniz University of Hanover, University of Mannheim, University of Texas at Austin, and Humboldt University of Berlin for their helpful comments and suggestions. We thank Junwei Xia for providing tax rate reconciliation data on her website. We gratefully acknowledge financial support from our respective institutions. Data are available from the sources cited in the text.

1 INTRODUCTION

We use large language models to examine how narrative text in financial statements contextualizes accounting numbers. Financial statements are not meant to be interpreted solely through numbers. Regulatory bodies including the U.S. Securities and Exchange Commission (SEC) explicitly mandate narrative disclosures, such as the Management Discussion and Analysis (MD&A), precisely to provide context for financial statement numbers (SEC, 1989). Indeed, textual analysis studies have long recognized the value of financial statement text and identified correlations between overall features of the text and firm outcomes (e.g., Li, 2008; Loughran and McDonald, 2011). Yet, despite the well-established value of narrative text, few studies actually examine *context*, which entails an interaction between financial statement numbers and disclosure content.

We leverage recent advances in LLMs and deep learning to explicitly model context as an interaction between text and numbers that improves the mapping to firm outcomes. Specifically, we create high-dimensional representations of text, which we use to quantify disclosure content and examine the explanatory power of context in mapping the relationship between income and tax expense. We choose the setting of corporate tax outcomes for two reasons. First, there is a strong mechanical link between pre-tax income and tax expense, as income taxes are generally calculated on some form of profit. In other words, after-tax outcomes are a function of pre-tax performance, usually represented via the statutory tax rate applicable to a given firm-year. Empirical research often assumes a proportional or linear tax function in this relationship (Edwards et al., 2021). In reality, however, the link between pre-tax income and tax expense is highly non-linear, making it notoriously difficult to assess corporate tax avoidance using the relationship between pre-tax income and tax expense (e.g., Belnap et al., 2025; Schwab et al., 2022). Non-linearity arises mainly from (1) differential treatment of income components for GAAP versus tax purposes, and (2) blended statutory tax rates when income components fall under multiple tax regimes. Therefore, the tax setting provides a promising scope for examining the degree to which narrative disclosure context can explain these non-linearities.

Second, the amount of tax expense for a given level of pre-tax income depends inherently on *how* that income is generated. Tax laws are complex as they apply differential treatments to a long array of firm-specific operating facts (Amberger et al., 2025; Hoppe et al., 2023). Thus, two firms with identical book income can face very different tax outcomes, for example due to differing cost structures, investments, geographic footprints, and past profitability. Capturing these and other important features of firms' operations with numeric variables requires a nuanced hand-collection, which is prone to fragmentation and omitted interactions. To the extent that these operating facts are embedded in firms' narrative disclosures, disclosure context offers a holistic representation of firms' operations.

Our aim is to measure the extent to which contextual information about firm operations, as embedded in financial statement text, can explain the mapping between firms' pre-tax income and tax expense and ultimately provide information to financial statement users. To achieve this, we employ a two-step methodology. First, we quantify the narrative context surrounding tax outcomes using LLMs. We extract the MD&A section from firms' annual 10-K filings, as this section is intended to provide management's perspective on performance and relevant context. We process this text using SBERT, a modification of the seminal transformer model "Bidirectional Encoder Representations from Transformers" (BERT) that is fine-tuned towards the meaningful quantification of sentences, to generate high-dimensional vector representations (or "embeddings") for each firm-year's MD&A. These embeddings capture the rich semantic and contextual information contained in the narrative and allow us to interact narrative information with financial data.[1]

Second, we train a deep neural network designed to learn how this narrative information modifies the relation between pre-tax income and tax expense. Following Farrell et al., (2021) and Kim and Nikolaev (2024a, 2024b), instead of feeding both text embeddings and pre-tax income as concurrent inputs into a feedforward neural network, we structure the model such that the text embeddings (the "input") determine the parameters (the "output") of a simple linear relation linking pre-tax income (*PI*) to tax expense—GAAP tax expense (*TXT*) or cash taxes paid (*TXPD*). Specifically, during training, the neural network is given data on pre-tax income, tax expense, and narrative disclosure embeddings, and uses embeddings as only input to the neural network to minimize the squared error between tax expense and pre-tax income.

In doing so, the neural network outputs two parameters, θ_0 and θ_1, for each firm-year, representing the context-adjusted intercept and slope in a linear tax function ($Tax\ Expense_{i,t} = \theta_{0i,t} + \theta_{1i,t} \cdot PI_{i,t}$). An important feature of this approach is that the slope, θ_1, reflects "context" as an interaction between the text and pre-tax income, thus explicitly modeling how well narrative disclosures can explain *PI*'s mapping into tax expense. Intuitively, the neural network "learns" to use narrative disclosure embeddings to adjust pre-tax income such that it aligns as well as possible with tax expense. Therefore, context-adjusted pre-tax income reflects both the original pre-tax income and the portion of narrative information that is informative about differences between pre-tax income and tax expense.

After training the neural network on four-year rolling windows (i.e., *t-4* to *t-1*), we apply the resulting model to the data in year *t* to ensure out-of-sample validity. Specifically, we feed the resulting model with narrative disclosure embeddings as input and receive as output two

[1] We conduct a battery of robustness checks using alternative LLMs (including BERT and FinBERT, a fine-tuned alternative to BERT that is trained on a corpus of financial text) to assess the sensitivity of our results towards the use of SBERT as the primary embedding model.

parameters that reflect the extent to which the disclosures would suggest adjustments to *PI* in order to better resemble tax expense. Importantly, these parameters are firm-year-specific as they build on individual narrative disclosure inputs, consistent with the idea that two firms with identical pre-tax income are to be assessed differently based on their respective operations.

Our final measure of context-adjusted pre-tax income is then calculated using the output parameters in interaction with the reported pre-tax income in the same year: $PI^{Context}_{i,t} = \theta_{0i,t} + \theta_{1i,t} \cdot PI_{i,t}$. Intuitively, $PI^{Context}$ is a firm-year-specific derivative of *PI* that adjusts the original value towards tax expense to the extent that contextual narrative information explains deviations between both variables. We also separately train a model that learns to resemble tax expense without information on pre-tax income, allowing us to derive a narrative-only measure ($PI^{Text}_{i,t} = \theta_{0i,t}$) to compare the influence of text without the contextualization to pre-tax income.

Our empirical results strongly support the value of contextualization. When examining *TXT* as the dependent variable, we find that $PI^{Context}$ significantly outperforms both pre-tax income and the narrative-only measure, PI^{Text}, in predicting tax expense. Specifically, $PI^{Context}$ has a coefficient close to one and substantially increases explanatory power. *PI* itself exhibits high explanatory power, with an adjusted R^2 of 0.8930, making the bar for incremental contribution exceptionally high. PI^{Text} has an adjusted R^2 of 0.6654, indicating that text alone does a much worse job of predicting tax expense than pre-tax income. However, $PI^{Context}$ increases the adjusted R^2 to 0.9118, reducing the unexplained variation by 17.6% (from 0.1070 to 0.0882). The increase is statistically significant and indicates that context can meaningfully improve the mapping between *PI* and *TXT*. Moreover, we find similar improvements for cash taxes paid, where our contextualized measure continues to offer a reduction in unexplained variation relative to regular pre-tax income by 23.9% (from 0.1936 to 0.1474). However, we also find that $PI^{Context}$ does not fully subsume the informational value in *PI*, suggesting that

MD&A narrative disclosures, with their focus on strategic and operating context, align more closely with accrual-based measures like tax expense rather than cash-based tax payments.

Next, we compare the contextual value derived from MD&A disclosures with that from income tax footnotes. Ex ante, it is possible that both disclosure sections can be incrementally informative in explaining the relation between pre-tax income and tax expense because income tax footnotes specifically discuss firms' tax outcomes, whereas MD&As provide a broader overview of firms' operations. Surprisingly, however, our results suggest that there is very limited contextual value from income tax footnotes. In contrast to the MD&A results, we show that $PI^{Context}$ variables from income tax footnotes yield coefficients that are well below one and often *decrease* explanatory power, relative to *PI* alone. These results hold when we examine income tax footnote-based contextualized variables alone as well as together with MD&A-based contextualized variables. In additional analyses we also show that while context from the MD&A consistently provides incremental explanatory power across a wide range of model specifications, context from the tax footnote consistently provides less explanatory power than *PI* itself. This finding provides empirical evidence to support recent survey evidence that tax executives prefer vague income tax footnote disclosures to comply with regulation but not provide informative information (Richter et al., 2024).

After showing that contextual information from MD&A narrative disclosures informs variation in the relation between pre-tax book income and tax outcomes, we shift toward deepening our understanding of context. Specifically, we examine how certain disclosure attributes and firms' underlying characteristics impact the relation. To identify what makes narrative context informative, we examine which specific MD&A disclosure topics are associated with more or less accurate contextualization of tax expense. We use Latent Dirichlet Allocation (LDA) to model topics and follow Dyer et al., (2017) to categories the topics. We find that narratives focusing on "Employees and Executives," "Geographic", and "Contracts and Legal" information, among others, are linked to more informative disclosures. Conversely, discussions concerning "Property and Leasing," "Stock and Options," and "Investments, Securities, and Derivatives" tend to provide less useful context. This suggests that the contextual utility of disclosures is higher when they pertain to core operational aspects and strategic firm activities rather than complex transactional or potentially boilerplate information.

In terms of firm characteristics, we focus on four cross-sectional tests where we expect context to exhibit an important moderating effect: multinational status, product market fluidity, size, and tax avoidance. Our results indicate that incorporating narrative context provides a relatively larger improvement in explaining tax outcomes for large firms and firms with low effective tax rates (ETRs), suggesting that context is particularly informative about tax outcomes of firms with less complex business operation and when firms are more likely to engage in tax avoidance. However, we find lower improvements among multinational firms compared to domestic firms, which may identify a potential limitation to the informativeness of narrative disclosures (i.e., when tax outcomes are complicated by cross-border operations).

Lastly, we show that the contextual information captured by our models is relevant to financial statement users. Using analysts' effective tax rate forecasts, we find that narrative context from the MD&A (but not tax footnotes) is associated with analyst forecast accuracy. While larger contextual adjustments are associated with greater forecast errors in cross-sectional tests, within-firm analyses reveal that increases in contextual informativeness improve analysts' forecast accuracy. That is, between firms, greater context may be associated with greater complexity that makes forecasting less precise. However, within a firm (and controlling for unobservables between firms, such as complexity), greater context appears to provide value to analysts. These findings suggest that MD&A narratives convey tax-relevant context that sophisticated users can incorporate.

Our study builds upon and contributes to several streams of literature. First, we quantify and provide strong evidence for the value of narrative context for understanding tax outcomes. Moreover, we shed light on where annual reports provide contextual value for corporate tax outcomes and where they do not. We show that the broad economic context provided in the MD&A is effective for explaining a substantial portion of both accrual-based and cash-based tax outcomes, whereas the specific, technical disclosures in the income tax footnote are generally uninformative. Further, we find descriptive evidence consistent with the notion that the value of contextualization is due to disclosures about firms' operating environments. Our findings offer a nuanced perspective on the value of different sources of narrative disclosures and highlight the value of broader economic disclosures (as per MD&A disclosures) over more specific explanations (as per footnote disclosures). To this end, our results provide evidence for standard-setters and regulators on the value of reporting narratives.

Second, our research adds to the rapidly evolving field of textual analysis and machine learning in accounting and finance. Early work in this area often used dictionary-based methods or "bag-of-words" approaches to quantify sentiment, tone, or specific topics (e.g., Li, 2008; Kothari et al., 2009; Feldman et al., 2010; Li, 2010; Loughran and McDonald, 2011; Davis et al., 2012; Law and Mills, 2015; Henry and Leone, 2016; Allen et al., 2021). While valuable, these methods often struggle with context and domain-specific nuances. Methodologically, we adapt and apply an innovative deep learning framework (Farrell et al., 2021; Kim and Nikolaev, 2024a, 2024b) to the specific challenge of mapping pre-tax income to taxes. This provides a concrete use case and a potentially generalizable approach for researchers seeking to "contextualize" other key financial statement variables where narrative discussion provides crucial interpretive information.

2 BACKGROUND

Recent years have witnessed a significant shift in accounting research, with scholars increasingly employing sophisticated textual analysis and machine learning techniques to extract insights from the considerable narrative information accompanying financial statements. Traditional methods often relied on the characteristics of narrative disclosure or dictionary-based "bag-of-words" approaches (e.g., Li, 2008; Li, 2010; Loughran and McDonald, 2016; Loughran and McDonald, 2020; Rogers et al., 2011; Law and Mills, 2015), treating text as an unordered collection of words and sometimes struggling to capture the nuances of domain-specific language. However, the limitations of these earlier techniques, particularly their inability to fully grasp context and semantic relationships, are being overcome by modern LLMs like BERT and GPT (generative pre-trained transformer). These advanced models can process language bidirectionally—meaning that the model reads text both to the left and to the right of a word simultaneously—and thus understand the context in which words are used, positioning them as the new standard for analyzing complex financial and non-financial disclosures. That is, researchers can now explore how narrative context interacts with and alters the interpretation of reported financial numbers.

A key theme emerging from this research is the critical role of narrative context in interpreting quantitative financial data. Kim and Nikolaev (2024a, 2024b) use deep learning to model the interactions between text and numbers, showing that accounting numbers are substantially more informative in predicting future performance when combined with narrative disclosures about the same topic, particularly when numeric data alone might be less reliable or persistent. Their findings suggest the value derived from these interactions often dominates the direct informational content of the narrative itself, and that both markets and analysts incorporate this contextual information in their decision-making. This highlights that the "story" surrounding the numbers significantly shapes their meaning and predictive value. Our study adds to this understanding by shifting from prediction tasks, where relations between two variables are inherently uncertain due to their future-orientation, to the mapping between two contemporaneous variables that should theoretically be closely aligned and are only distorted by complex non-linearities that are difficult to map using standard methods and variables. Moreover, it is unclear ex ante whether and to what extent narrative disclosures are informative in explaining tax outcomes since the operating facts driving them do not necessarily overlap with the operating facts being discussed in firms' narrative disclosures. Also, instead of using a narrow set of narrative disclosures that are most likely to be explanatory about the variables we seek to explain (here: tax expense), we consider the informative value of the full MD&A section to allow a broad number of contents to provide meaningful context. Given the strong mechanical link between pre-tax income and tax expense and the ambiguity about whether narrative disclosures capture the complex non-linearities beyond this relationship, it remains an open empirical question whether narrative context can incrementally explain this relation.

Tax-specific research has undergone a similar evolution. Early studies used bag-of-words approaches to predict ETRs (Allen et al., 2021) or show that text-based measures of financial constraints are associated with tax avoidance (Law and Mills, 2015). More recent studies apply machine learning to prediction and analysis. Jennings et al., (2024) and Bogachek et al., (2024) use Latent Dirichlet Allocation and other techniques to predict ETRs and other tax outcomes—finding, among other things, that textual data improve predictions.[2]

Researchers are also pushing the boundaries by using generative artificial intelligence (GAI) and developing specialized models for tax. Choi and Kim (2024) employ GAI to interpret narrative disclosures and measure firm-level tax audit periods, uncovering effects on tax avoidance and firm risk. Addressing the need for domain expertise, Hechtner et al., (2025)

[2] Another related paper is Guenther et al., (2023) which uses machine learning to predict future ETRs, but that study uses only quantitative financial data.

guide the creation of specialized LLMs like TaxBERT, pretrained on tax-specific corpora, which demonstrably outperform generic language models in analyzing tax disclosures. These advancements collectively showcase the power of sophisticated machine learning approaches in extracting nuanced insights from complex accounting and tax data.

Burd et al., (2025), is perhaps the most closely related study to ours. It does not utilize machine learning but creates a measure of "discussed numbers" that is the number of times the text in a 10-K mentions the same words of a tax-related XBRL tag. While innovative in linking XBRL to text, this approach relies on frequency counts and direct word matching, which potentially misses deeper context and meaning, and excludes any text not directly associated with specific tax tags. As we describe in detail in the next section, our study exploits the recent advances in LLMs to provide a much deeper context for tax numbers, including both tax- and non-tax related narrative context.

3 METHODOLOGY

3.1 Conceptual Framework

Our aim is to show how context can improve the prediction of a firm's tax expense given the firm's income. From a mathematical perspective, the tax liability of any firm *i* in year *t* can be determined by applying a specific formula, such as the applicable statutory corporate income tax rate, to its taxable pre-tax income, as expressed in the following general form:

$$\textit{Tax Expense}_{it} = f(\textit{Pre-tax Taxable Income}_{it}) \tag{1}$$

where $f(\cdot)$ denotes the applicable average statutory corporate income tax rate for firm *i* in year *t*. For example, in the United States, a constant statutory tax rate of 35% (21%) applied prior to (following) the enactment of the U.S. Tax Cuts and Jobs Act (TCJA). In a simplified setting, pre-tax taxable income may therefore map linearly into tax expense as such:

$$\textit{Tax Expense}_{it} = \textit{Pre-tax Taxable Income}_{it} \cdot \textit{Statutory Corporate Income Tax Rate}_{t} \tag{2}$$

In practice, however, this function is often non-linear due to factors such as preferential tax rates for specific types of income and varying statutory tax rates for income earned in different jurisdictions. Additionally, we must rely on *Pre-tax Book Income*—as determined by Generally Accepted Accounting Principles (GAAP)—because *Pre-tax Taxable Income* is generally not observable. The differential treatment of income components under book and tax, which is defined by the U.S. Internal Revenue Code and the laws of other states and jurisdictions where firms earn their income, further complicates the relation between pre-tax income and tax expense. A more representative formulation of Eq. (1), which accounts for both the non-linear, firm-specific nature of $f(\cdot)$ and differences between book and taxable income would be as follows:

$$Tax\ Expense_{it} = f_{it}(Pre\text{-}tax\ Book\ Income_{it} \pm Book\text{-}Tax\ Differences_{it}) \quad (3)$$

Given that prior studies have found value in the text of financial statements, we expect that when this text is used to explicitly interact with financial statement numbers—that is, to provide context—it will improve predictions by capturing both the book-tax differences and the other non-linearity $f(\cdot)$ components in this equation.

Note that when we use the GAAP tax expense (*TXT*) as the left-hand side variable in Eq. (3), we capture only permanent book-tax differences; whereas when we use cash taxes paid (*TXPD*), we can also capture temporary differences. This approach is consistent with the tax avoidance literature, where GAAP ETRs capture only permanent book-tax differences and cash ETRs reflect deferral strategies.

3.2 Linking Tax Expense, Book Income, and Narrative Context

Our approach to capture these components leverages recent advancements in machine learning, specifically the methodology developed by Farrell et al., (2021) and applied to narrative disclosures by Kim and Nikolaev (2024a, 2024b). This methodology employs deep neural networks to process unstructured data (i.e., text) and minimize the difference between a target outcome variable (i.e., tax expense) and explanatory financial data (i.e., pre-tax book income).

Following Kim and Nikolaev (2024a), we adopt a two-step approach. First, we generate numeric representations of narrative disclosures to capture the contextual information embedded in firms' narrative disclosures. Specifically, we employ SBERT, a fine-tuned version of Bidirectional Encoder Representations from Transformers (BERT), which is a transformer-based large language model that encodes text into vector representations that explicitly capture linguistic and contextual nuances (Devlin et al., 2019; Reimers and Gurevych, 2019). We choose SBERT over BERT or similar alternatives as our primary model because the pre-training of SBERT emphasized the embedding of the entire text section over the embedding of individual words, making it likely that SBERT produces more accurate embedding vectors. We apply SBERT to convert each firm-year's narrative disclosures into 768-dimensional vector embeddings. For each document, we extract the classification token (CLS), which summarizes the text's overall context, to obtain a document-specific vector representing a firm-year's narrative disclosure. If a disclosure exceeds SBERT's 512-token context window, we split it into chunks of up to 512 tokens with a 50 token overlap, embed each chunk separately, and average all CLS tokens. This procedure yields a single 768-dimensional vector representation per document.[3]

In the second step, we train a deep neural network to use the SBERT embeddings to explain variation in tax expense beyond what is captured by pre-tax book income alone. While one option is to use SBERT embeddings and pre-tax income jointly as inputs to predict tax expense, this strategy would likely require substantial data to model the complex interactions between textual embeddings and performance numbers effectively (Kim and Nikolaev, 2024a).

[3] Before running documents through SBERT, we pre-process all documents by deleting special characters, tables, visuals, HTML markup and other fragments contained in raw EDGAR downloads. Moreover, we mask numeric information in narrative disclosures by replacing any numbers with a placeholder, allowing SBERT to embed words based on the existence of surrounding numbers but not their meaning.

Therefore, we adopt a more data-efficient design that uses only the embeddings as inputs while incorporating pre-tax book income into the loss function. Specifically, we train the neural network to minimize the *difference* between tax expense and pre-tax book income.

We structure the neural network to learn how the interaction between narrative disclosures and pre-tax book income explains variation in total tax expense. Formally, we use the 768-dimensional SBERT embedding as the input layer and define a two-dimensional output layer that yields parameters θ_0 and θ_1 for each observation. The model minimizes the following loss function, equivalent to the root mean squared error (RMSE):

$$\mathrm{L}(\theta_0, \theta_1) = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} \left(Tax\ Expense_n - \theta_{0,n} - \theta_{1,n} \cdot PI_n \right)^2} \tag{4}$$

where *PI* is pre-tax book income and *Tax Expense* is either *TXT* or *TXPD*, both divided by the appliable U.S. statutory corporate income tax rate. This normalization aligns both quantities in the same unit, mitigating distortions in training performance related to the TCJA. Thus, the loss function effectively minimizes book-tax differences. Intuitively, unlike a non-contextual model, which minimizes the error between tax expense and one or more standalone parameters θ_0, θ_1, …, θ_n, our architecture constraints the neural network to predict tax expense both directly (θ_0) and via interaction with pre-tax book income (θ_1). In essence, since we adjust tax expense by the applicable statutory tax rate, the values of tax expense the model predicts serve as a measure of "expected taxable income" based on pre-tax book income and embedded narrative disclosures.

Figure 1 illustrates the model architecture. The 768-dimensional SBERT embeddings vector is the input layer. The network consists of three hidden layers with 512, 128, and 32 nodes, respectively. Each hidden layer applies a 20% dropout rate to reduce overfitting by randomly deactivating nodes during training. The output is a two-dimensional parameter layer producing θ_0 and θ_1. We train the model using a four-year rolling window (years *t–4* to *t–1*) and evaluate performance using Eq. (4) on a validation set comprising 20% of randomly sampled training observations. The model trains for up to 250 epochs, with early stopping if the value of the loss function for the validation set fails to improve by at least 0.001 for ten consecutive periods. We retain the best-performing model state using the Adam optimizer with a learning rate of 5×10^{-4} and a batch size of 64.

After identifying the optimal model state, we apply the final model to the out-of-sample year *t* that follows the training window to obtain a context-adjusted measure of pre-tax book income, denoted $PI^{Context}$. Specifically, for each four-year rolling window comprising all observations in years *t–4* through *t–1*, we calculate $PI^{Context}$ for year *t* as follows:

$$PI_{it}^{Context} = \theta_{0,it} + \theta_{1,it} \cdot PI_{it} \tag{5}$$

This structure ensures that model training and validation relies only on data from the rolling window and does not learn from any information from year *t*, preserving a strict out-of-sample evaluation.

As a benchmark, we also construct a text-only approximation of tax expense, which uses the same steps as above but omits interactions with *PI*. We re-train the same architecture but reduce the parameter layer to have only one dimension, θ_0, optimized via a standard RMSE loss function. This produces a purely narrative-driven proxy for tax expense:

$$PI_{i,t}^{Text} = \theta_{0,i,t} \tag{6}$$

Including this measure allows us to compare our contextually informed version of pre-tax book income—which interacts reported pre-tax book income with narrative disclosures—against both the raw pre-tax income and the narrative-only prediction. Note that we repeat this process two times to obtain contextually informed versions of pre-tax book income that approximate *TXT* and *TXPD,* respectively.

4 DATA AND SAMPLE SELECTION

We obtain financial accounting data from Compustat and narrative disclosures from firms' 10-K filings, which we collect from the SEC's EDGAR database. To isolate the relevant textual data, we search all filings for their MD&A sections as the primary source of narrative disclosure. We exclude observations with narrative disclosures that are insufficiently long to be informative.[4] We focus on MD&A sections because they predominantly contain qualitative discussions of firm operations and are designed to be highly firm-specific and managerially tailored sources of information for readers (Clarkson et al., 1999).

Our sample period spans 1996 to 2024, starting with the first year of comprehensive availability of 10-K filings in EDGAR. We exclude all observations with negative values for *PI*, *TXT, or TXPD* because pre-tax book income maps less clearly to taxable income for these observations. Moreover, since the initial four years of our sample period are reserved for the initial training window, our evaluation period effectively begins in 2000, yielding a final sample of 52,113 firm-year observations. In additional tests involving income tax footnotes, we further restrict our sample period to effectively begin in 2015 due to inconsistent availability of income tax footnote text in CalcBench before 2011, yielding a restricted footnote sample comprising 16,109 observations. Table 1 summarizes the sample selection procedure.

Table 2 presents descriptive statistics for all variables used in our main analysis. Panel A provides summary statistics, with all variables presented in millions. Recall that *TXT* and *TXPD* are divided by the applicable statutory corporate income tax rate in year *t*. *TXT* has a mean of $338 million and median of $46 million. *TXPD* is comparable, but with a lower mean ($301 million) and median ($33 million). Both *TXT* and *TXPD* are lower than *PI*, which has a mean of $390 million and a median of $59 million. Our first contextualized measure, $PI^{Context}$

[4] Specifically, we exclude all texts with fewer than 500 words because these texts are likely fragmented and do not contain valid information that can be used to train our models.

(TXT), more closely resembles *TXT*, with a mean of $310 million and median of $47 million. The narrative-only measure, PI^{Text} *(TXT)*, has a smaller mean but a larger median than $PI^{Context}$ *(TXT)*, suggesting weaker performance in the tails of the distribution. Moreover, consistent with *TXPD* being on average lower than *TXT*, both PI^{Text} *(TXPD)* and $PI^{Context}$ *(TXPD)* are lower than their accruals-based counterparts. As before, the contextualized measure resembles *TXPD* more closely than the text-only measure.

Panel B reports pairwise correlations among the main variables. *PI*, PI^{Text} *(TXT)*, and $PI^{Context}$ *(TXT)* are all significantly correlated with *TXT*. The correlation between *TXT* and $PI^{Context}$ *(TXT)* of 0.9549 exceeds that between *TXT* and either *PI* (0.9450) or PI^{Text} *(TXT)* (0.8157), indicating that the contextualized measure captures additional variation in tax expense. As expected, *PI* and $PI^{Context}$ *(TXT)* are highly correlated (0.9741), consistent with the idea that $PI^{Context}$ *(TXT)* builds upon *PI* by incorporating contextual information from narrative disclosures. In contrast, PI^{Text} *(TXT)* is less correlated with both *PI* (0.8131) and $PI^{Context}$ *(TXT)* (0.8286), consistent with its construction relying on textual disclosures without explicit interaction with pre-tax book income.

5 RESULTS

5.1 Informational Value of Context

To evaluate the informational value of context conveyed through MD&A disclosures, we test whether contextually adjusted pre-tax book income provides a better mapping into tax expense than raw pre-tax book income. We estimate linear ordinary least squares (OLS) regressions with varying versions of pre-tax income as independent variables. The full model specifying all three "types" of pre-tax income takes the following form:

$$Tax\ Expense_{it} = \beta_0 + \beta_1 \cdot PI_{it} + \beta_2 \cdot PI_{it}^{Text} + \beta_3 \cdot PI_{it}^{Context} + \varepsilon_{\mathrm{it}} \tag{7}$$

where the *Tax Expense*, PI^{Text}, and $PI^{Context}$ variables relate to either *TXT* or *TXPD*. In this specification, β_1 captures the mapping between standard pre-tax book income and tax expense, β_2 captures the mapping with the narrative-only income measure, and β_3, our primary coefficient of interest, captures the mapping between contextualized pre-tax book income and tax expense. Coefficients closer to zero indicate weaker alignment with tax expense, whereas values closer to one suggest stronger co-movements.

In Table 3, Panel A, we report the results of the *TXT* regressions. In Column (1), we establish the association between *PI* and *TXT*. The coefficient estimate of 0.8218 indicates a strong positive relation, consistent with the premise that tax expense is a function of pre-tax income. Recall that *TXT* is defined as total tax expense scaled by the statutory corporate income tax rate, meaning that the coefficient indicates that, on average, an additional dollar of book income is associated with an additional 82 cents of taxable income. The adjusted R^2 of 0.8930 suggests that standard pre-tax book income explains the majority of the variation in *TXT*, consistent with our conceptual framework in Section 3.1. Further, this baseline level of explanatory power sets an extremely high bar for other variables to incrementally contribute. In subsequent columns, we test whether the incremental explanatory power is statistically significant using a Vuong (1989) test.

In Column (2), our narrative-only measure variant, PI^{Text}, yields a coefficient estimate near unity, consistent with the neural network's design to predict average tax values. However, the adjusted R^2 is substantially smaller than in the baseline regression, indicating a lack of explanatory power in capturing cross-sectional variation in tax expense. The coefficient estimate for our contextualized measure, $PI^{Context}$, in Column (3) is similar to the one of the narrative-only measure. Moreover, the adjusted R^2 is higher than that reported in the baseline regression. Specifically, while *PI* alone leaves about 0.1070 of variation in *TXT* unexplained, the contextualized measure reduces this to about 0.0882, indicating an approximate 17.57% improvement in explanatory power. When including both *PI* and $PI^{Context}$ (Column (4)) or even all three measures of pre-tax book income simultaneously (Column (5)), all coefficients remain statistically significant, but the magnitude of the coefficient on $PI^{Context}$ far exceeds the others. Moreover, the marginal increase in adjusted R^2 relative to Column (3) is small. These results suggest that the contextually adjusted measure subsumes most of the explanatory power of both standard and narrative-only pre-tax income measures. Collectively, these findings provide strong support for the incremental value of context in explaining variation in tax expense.

In Panel B of Table 3, we present the results when using *TXPD* as the dependent variable. The baseline model in Column (1) yields a coefficient estimate of 0.7452, implying that, on average, cash tax payments for every generated dollar of book income are about 75 cents. Notably, adjusted R^2 is lower than in Panel A (0.8064), indicating that *PI* explains less of the variation in *TXPD* than of the variation in accruals-based *TXT*. Compared to standard pre-tax book income, both PI^{Text} and $PI^{Context}$ produce larger coefficient estimates, slightly overshooting actual tax payments. Moreover, the contextualized model offers a substantial increase in explanatory power in Column (3) relative to the baseline (the unexplained portion of the variation in *TXPD* decreases by 23.86%). The adjusted R^2 in Column (4) is similar, indicating that contextualizing *PI* again subsumes most variation in standard pre-tax income. However, including all three measures jointly in Column (5), the adjusted R^2 increases further (0.8635), indicating that the text-only measure provides a modest amount of additional informative value. Collectively, we conclude that our contextually adjusted measure of pre-tax book income continues to be superior to standard pre-tax book income in explaining variation in cash taxes paid but that contextualization subsumes the information contained in pre-tax income less clearly than for tax expense. These findings are consistent with the nature of the MD&A, which reflects a firm's strategic narrative and operating context over the prior fiscal year. While such disclosures are well aligned with accruals-based measures such as tax expense—whose purpose is to match expenses to economic performance—they may be less predictive of actual tax payments.

5.2 Income Tax Footnotes

In this section, we compare the informational value of narrative context derived from the MD&A with that provided by income tax footnotes. Our main specification relies on the full MD&A section, which includes both tax and non-tax information. In contrast, the income tax footnote focuses specifically on a firm's tax behavior, tax compliance, and reconciliations between expected and actual tax outcomes. As such, it may offer a more targeted summary of information relevant to understanding the relation between pre-tax income and tax expense. However, this narrower focus may limit its informational value by omitting broader context about a firm's economic performance with relevance to tax outcomes. Moreover, prior research suggests that tax-specific disclosures tend to be more boilerplate than other contents, potentially reducing their informativeness (Bilicka et al., 2024; Richter et al., 2024).

To formally assess the relative value of income tax footnote disclosures, we repeat our empirical procedure for the subset of firm-years with available income tax footnotes on CalcBench. After applying the same preprocessing steps used for MD&A text, embedding all disclosures using SBERT, and feeding the resulting vectors into separate neural networks, we construct four new measures: $PI^{Text\_Footnote}$ and $PI^{Context\_Footnote}$ (each with a *TXT* and *TXPD* version), which incorporate income tax footnotes to minimize the difference between *PI* and *TXT* or *TXPD*, respectively. These variants allow us to estimate extended versions of Eq. (7) and compare the relative informativeness of income tax footnotes versus the MD&A.

Table 4, Panel A presents results for the informational value of income tax footnotes in explaining the relation between pre-tax income and tax expense, following the same structure as our main specification. The baseline relation between standard pre-tax income and tax expense is comparable to the full sample reported before, with a coefficient estimate of 0.8303 and an adjusted R^2 of 0.8869 in Column (1). As before, we find that the text-only measure explains less variation in the dependent variable than *PI*. Informing pre-tax income with context from income tax footnotes yields a coefficient of 0.7812 and an adjusted R^2 of 0.8549, which is significantly lower than the *PI*-only benchmark. Consistently, in Columns (4) and (5), the *PI* coefficients dominate over both those of $PI^{Text\_Footnote}$ and $PI^{Context\_Footnote}$, and the incremental improvement in adjusted R^2 compared to Column (1) is modest. These findings suggest that the income tax footnote provides limited incremental context for understanding the mapping between pre-tax book income and tax expense.

Panel B presents analogous results using cash taxes paid as the dependent variable. We find in the baseline regression that a one dollar increase in *PI* is associated with a 0.7757 dollar increase in *TXPD*, indicating underpayment relative to the scaled statutory base. Moreover, we continue to find that both measures built on the income tax footnote shift this relation closer to one. As before, we continue to find that the adjusted R^2 for *PI* (0.8064) exceeds both the explanatory power of PI^{Text} (0.3873) and $PI^{Context}$ (0.8060), with a moderate and statistically significant increase in the joint regressions in Column (4) to 0.8290 and in Column (5) to 0.8334. These findings suggest similar inferences about the relevance of income tax footnotes as our analysis of *TXT*, though our measure of contextualized *PI* does not result in a substantial decline in explanatory power. Thus far, our findings suggest that income tax footnotes add little explanatory power to understanding the relation between pre-tax income and tax outcomes.

To further assess this point, we directly compare the informational value of context from the income tax footnote to that of the MD&A by adding the original variables, PI^{Text} and $PI^{Context}$, to our regressions. We present the results in Table 5, Panel A (B) for analyses aiming to explain *TXT* (*TXPD*). Comparing both the text-only (Column (1)) and the contextually informed measures (Column (2)), we find that the MD&A-based measures are much stronger associated with tax outcomes than the tax footnote-based measures, based on the coefficients. In Column (3), we include *PI* and only the MD&A measures to examine whether the reduced explanatory power compared to our previous analysis of MD&A measures is due to differences in sample composition. Comparing Table 5, Column (3) with Table 4, Column (5) reveals contextualization through MD&As yields larger values of adjusted R^2 than through the contextual interaction from income tax footnotes, and the Vuong test shows that the difference is statistically significant. In the full model (Column (4)), $PI^{Context}$ has by far the largest coefficient estimate among all five *PI* measures in both Panel A and Panel B. More importantly, adding variables that are contextually informed with tax footnote information adds almost no additional explanatory power to the regression compared to Column (3), suggesting that virtually all the added information stems from MD&A context.

Together, our findings suggest that income tax footnotes offer little contextual value for explaining variation in both GAAP tax expense and cash tax payments, with a slightly better ability to explain *TXPD* than *TXT*. This highlights a nuanced role for narrative disclosures in the context of taxation: MD&A narratives are generally effective in explaining a substantial portion of variation in tax outcomes, especially when it comes to accrual-based outcomes, whereas income tax footnotes are somewhat insightful only when explaining cash tax outflows. Our findings are also consistent with concurrent survey evidence from Richter et al., (2024) that tax executives aim for the narrative text in tax footnotes to be vague.

5.3 Relative Importance of Narrative Disclosure Topics

The previous sections establish that narrative context can add incremental information helping to explain the mapping between pre-tax book income and tax outcomes. In this section, we examine which specific MD&A disclosure topics contribute most meaningfully to this informational value, i.e., their ability to contribute to an accurate mapping between pre-tax book income and tax expense.

Specifically, we examine which narrative disclosure topics are associated with more or less accurate contextualization. We begin by constructing a binary measure of disclosure informativeness. We define observations as informative based on the absolute value of the residuals of the regression reported in Table 3, Column (3). Specifically, we set *Informative* to one if the residual is less than the value of the dependent variable, and zero otherwise. This process flags 65.69% of all observations as being "informative" and 34.31% as "uninformative," providing a tractable classification for subsequent analysis.

To examine the topical drivers of disclosure informativeness, we apply Latent Dirichlet Allocation (LDA) to the MD&A corpus. LDA is a Bayesian statistical model that uncovers hidden thematic structures in a collection of documents by representing each document as a mixture of topics and each topic as a distribution over words (Blei et al., 2003). We iteratively explore the appropriate number of topics to be extracted from the disclosures and use a combination of human topic interpretability and topic coherence scores for evaluation. This process leaves us with an optimal number of 100 disclosure topics. We then aggregate the 100 original topics to the 13 topic categories proposed by Dyer et al. (2017). Finally, we apply the LDA model to compute the relative frequency of each topic category, yielding 13 metrics that jointly describe the distribution of topic emphasis in each MD&A. Figure 2 visualizes the distribution of topic categories over time.

We analyze the association between the prevalence of each topic category in an MD&A and its status as being "informative." That is, we run separate regressions of *Informative* on each respective variable describing the relative frequency of any given topic category. We run individual regressions for two reasons: (1) to isolate the effect of each topic's prevalence on the likelihood of a disclosure being informative; and (2) since topic frequencies sum to one by construction, running all metrics jointly would require us to omit one "baseline" topic category.

Figure 3, Panel A presents the coefficient estimates from these regressions. We find that topic categories like "Employees and Executives", "Contracts and Legal", "Business Structure and M&A", "Geographic", "Industry-Specifics", and "Operations and Strategy" are associated with a higher likelihood of informative disclosures. In contrast, several other topic categories—

including “Stock and Options”, “Property and Leasing”, “SEC, Accounting Standards”, “Investments, Securities, Derivatives”, and “Intellectual Property and R&D”—are associated with a higher likelihood of having uninformative disclosures or do not contribute to disclosure informativeness at all.

These findings provide initial evidence of the relation between the prevalence of certain topic categories and the informativeness of context for tax outcomes. However, one alternative explanation is that the corresponding circumstances themselves drive this finding, and not their disclosure. For example, more extensive disclosures of M&A deals are likely to be strongly correlated with the propensity of having recent M&A activities, which could then affect the relation between pre-tax book income and tax outcomes. To mitigate this concern, we re-run the same set of analyses but include an array of control variables that capture the underlying economic circumstances. Specifically, we control for size, intangible asset ratio, market-to-book ratio, leverage, firm age, return on assets, number of employees, special items, R&D expenditures, and the number of geographic segments and business segments. Moreover, we include indicator variables denoting the existence of non-zero R&D expenditure, the presence of a Big N auditor, recent losses, recent M&A deals, and non-zero foreign currency translations. Finally, we add industry- and year-fixed effects to control for industry-specific conditions and macroeconomic trends over time.

We report the results after controlling for firms’ real economic conditions in Figure 3, Panel B. We continue to find statistically significant relations between *Informative* and seven topic categories. Specifically, topic categories such as “Property and Leasing”, “Business Structure and M&A”, and “Investments, Securities, Derivatives” are associated with a lower propensity of disclosures being informative, whereas “Contracts and Legal”, “Employees and Executives”, and (weakly) “Geographic” and “Performance, Revenues, Customers” are associated with a higher propensity of disclosures being informative. The direction of most topic categories remains unaltered by adding control variables, suggesting the robustness of our inferences against the inclusion of control variables. Notably, the coefficient on "SEC, Accounting Standards" switches from weak negative significance to moderately positive significance, suggesting an ambivalent contribution of disclosures regarding SEC compliance and accounting standards to explaining the relation between pre-tax income and tax expense. Moreover, categories related to complex business transactions (such as "Business Structure and M&A" and "Investments, Securities, Derivatives") are consistently related to a lower likelihood of disclosures being informative. These findings suggest that boilerplate-intensive topics, such as discussions regarding business structure, and complex transactions, such as discussions regarding stock options and financial instruments, do not contribute positively to the informativeness of a disclosure.

5.4 Firm Characteristics

To better understand when context is most informative, we examine how the incremental value of context varies across four key firm characteristics. We compare domestic versus multinational entities (MNEs) because MNEs face more complex tax environments, leading to differences in firms' ability and willingness to report transparently about tax strategies. Similarly, we split firms by product market fluidity (Hoberg et al., 2014), as firms facing rapid competitive change may have evolving or novel tax strategies that are less easily captured in standard financial metrics but potentially described in narrative text. We also explore firm size, as smaller firms may have a less stable relation between pre-tax income and tax outcomes. Lastly, we examine firms with low versus high effective tax rates, since low-ETR firms are more likely to engage in aggressive tax planning that creates a complex mapping between pre-tax book income and tax expense.

We employ commonality analysis (Newton and Spurrell, 1967; Mood, 1969) to compare the incremental contributions of standard *PI* ("baseline") and $PI^{Context}$ to the variation explained by a combined model that includes both variables.[5] We begin by repeating our main analysis separately for both cross-sectional groups. Specifically, we run three models: One "full model" that includes both *PI* and the contextually informed $PI^{Context}$, and two models containing only one of the two variables, respectively (equivalent to Table 3, Columns (1), (3), and (4)). We then compute the incremental adjusted R^2 (labelled $U_{Context}$) added by $PI^{Context}$ as the difference between the adjusted R^2 of the full model and the baseline, and vice versa for the incremental adjusted R^2 of *PI* (labelled $U_{Baseline}$). We then compute the common variation that is shared by both variables as the difference between the adjusted R^2 of the full model, $U_{Context}$, and $U_{Baseline}$. We then create a measure of the relative dominance of $PI^{Context}$ over *PI* by scaling $U_{Context}$ by the difference between the adjusted R^2 of the full model and the common variation (which reflects the expectable range of possible values). Values closer to one reflect a strong dominance of context, whereas values close to zero suggest that context adds no value.

We present the results in Table 6, Panel A (B) for analyses using *TXT* (*TXPD*) as dependent variable. For GAAP tax expense, we find that contextualized pre-tax income dominates standard pre-tax income more clearly in domestic firms, large firms, and low-ETR firms. We also find a modestly higher dominance for the subsample of firms having high product-market fluidity. The incremental adjusted R^2 added by contextualization is far below that of standard pre-tax income in the subsamples of small firms and firms with high ETRs. However, we also note that the general level of adjusted R^2 in the subsample of small firms is small, irrespective of the independent variable. These findings suggest that the main results are concentrated in (1) domestic firms, suggesting that contextualization performs less reliable when the relation between tax expense and book income is complicated by cross-border

[5] The advantage of using commonality analysis over Shapley values is that it specifically allows for "shared variance", which likely plays a prominent role in our setting because $PI^{Context}$ builds on *PI*. Since Shapley values average marginal contributions across all possible permutations of variables, it effectively splits common explanatory power across all variables. While Shapley values are generally better suited when the number of predictors is large and the independent variables are independent of each other, commonality analysis is appropriate here given our small set of predictors and the fact that contextualized measures build on *PI*.

operations; (2) large firms, consistent with smaller firms having more straightforward operations and less complex tax functions; and (3) low-ETR firms, which are more likely to engage in corporate tax avoidance, thus having a less clear mapping between pre-tax book income and tax expense absent the consideration of context.

For cash taxes paid, we observe similar mechanisms, although the dominance of contextualization in firms with low product market fluidity slightly exceeds that of their high-fluidity counterparts. The dominance of contextualized pre-tax income over standard *PI* is at a similar level for both small and large firms, although the overall level of adjusted R^2 is still substantially lower in small firms. Collectively, these tests validate that context adds the most value where tax outcomes are least transparent and most contingent on firm-specific nuances.

6 ADDITIONAL ANALYSES

6.1 Analyst Responses to Disclosure Informativeness

The main results suggest that narratives in the MD&A are incrementally informative for understanding differences between expected and actual tax outcomes, whereas the narrative component of the income tax footnote seems to be less informative. In this section, we study whether these findings extend to the utilization of corporate narrative disclosures by financial statements users. Specifically, we examine whether analysts consider the narrative information conveyed through MD&As and income tax footnotes, respectively, when forming assessments about firms' ETRs. Prior research provides mixed evidence on analysts' ability to incorporate tax information into their forecasting. For example, Plumlee (2003) and Hoopes (2018) find that analysts struggle to incorporate news about complex tax law changes into their forecasts. However, Hutchens (2017) finds that analyst ETR forecasts are more accurate when income tax footnotes are more readable and use less complex jargon. Similarly, Burd et al. (2025) show improved analyst ETR forecast accuracy when tax numbers are discussed in firms' narrative disclosures. While these studies link analysts' ETR forecasting abilities to broader disclosure styles, none of them explicitly examine whether analysts effectively consider the informativeness of the context provided in these disclosures.

Our methodology allows us to explicitly measure to what extent information about the relation between pre-tax income and tax expense is incorporated in narrative disclosures. Specifically, since the neural networks described in Section 3.2 learn to interpret what portions of the narrative disclosures adjust book-tax differences, the parameters θ_0 and θ_1 effectively reflect the extent to which narrative disclosures explain these differences. In other words, the deviation between actual pre-tax income and the contextualized values suggest by our methodology reflects the informative value of context in explaining book-tax differences. Thus, we quantify the amount of contextual information conveyed through disclosures as the adjustments suggested by our model:

$$Adjustment_{it} = \frac{\left|PI_{it}^{Context} - PI_{it}\right|}{PI_{it}} \tag{8}$$

Intuitively, *Adjustment* indicates the absolute percentage value of the adjustment suggested by narrative context. Using percentages instead of raw values allows us to produce comparable magnitudes of *Adjustment* for firms with different profitability. However, while conceptually strongly related to the amount of contextual information given in the narrative disclosures, it does not explicitly reflect corporate tax behavior. To be able to examine analyst responses to corporate tax outcomes in conjunction with its contextually suggested shift, we specify the following alternative variable:

$$Wedge_{it} = |ETR_{it} - xETR_{it}| \tag{9}$$

where the "explainable ETR", $xETR_{it}$ is a variant of ETR_{it} that builds on $PI^{Context}$ instead of *PI* as the denominator, effectively containing a firm's GAAP ETR adjusted for all deviations from expected levels that are explained by the narrative disclosure.

We measure analysts' ETR forecasting ability using the mean absolute error between analysts' implied ETR forecasts for year *t+1* and the actual ETR in year *t+1*, including all analyst forecasts issued within 90 days before the fiscal year-end (Bratten et al., 2017). To examine how analysts incorporate narrative information from year *t* into their implied ETR forecasts for year *t+1*, we estimate the following linear regressions:

$$\mathit{ETR\ Forecast\ Error}_{it+1} = \beta_0 + \beta_1 \cdot \mathit{Adjustment}_{it} + \sum_{k=2}^{K+1} \beta_\mathrm{k} \cdot \mathit{Controls}_{it} + \varepsilon_\mathrm{it} \qquad (10)$$

$$\mathit{ETR\ Forecast\ Error}_{it+1} = \beta_0 + \beta_1 \cdot \mathit{Wedge}_{it} + \beta_2 \cdot \mathit{ETR}_{it} + \sum_{k=3}^{K+2} \beta_\mathrm{k} \cdot \mathit{Controls}_{it} + \varepsilon_\mathrm{it} \qquad (11)$$

If analysts are able to incorporate the adjustments suggested by narrative disclosures, we expect a positive coefficient on β_1. However, it is also possible that analysts do not incorporate this information into their forecasting, in which case we could expect an insignificant or negative coefficient. We include an array of control variables consistent with prior literature to mitigate concerns that our inferences are confounded by firm characteristics such as size, profitability, analyst following, and business complexity, or disclosure characteristics such as readability, specificity, and boilerplate language. We further use year fixed effects to control for macroeconomic trends and industry-fixed effects to remove cross-sectional differences shared by firms in the same industries. We define all control variables in the Appendix and provide descriptive statistics in Table 7, Panel A.

We report the results in Table 7, Panel B. When examining the contextual information from MD&As to form *Adjustment* and *Wedge*, we find that greater adjustments (as suggested by the narratives) are associated with less accurate analyst ETR forecasts (Columns (1) and (2)). This finding suggests that adjustments indicated in the MD&A reflect an incremental component of reporting complexity (beyond the metrics we control for) that considers the contents of a text rather than broader features. However, we find no significant association between the amount of contextualization and analyst ETR forecast accuracy when considering context from the narrative portion of the income tax footnote (Columns (3) and (4)), consistent with our previous finding that these narratives do not help explain variation in tax outcomes.

In a separate specification, we replace industry-fixed effects by firm-fixed effects to examine within-firm variation in the informational value of narrative disclosures. This specification allows us to examine how analysts respond to more informative narrative context conditional on the broader complexity of a firm's reporting environment. We report the results in Table 7, Panel C. We find that greater levels of *Adjustment* or *Wedge* in specifications using MD&A context are associated with lower analyst ETR forecast errors. Thus, after controlling for any unobservable firm-level attributes that determine the complexity of a firm's narrative disclosures, our findings suggest that analysts can use the context conveyed by the MD&A to produce better forecasts. Again, we do not find a significant coefficient on either variable when considering context from the income tax footnote. Overall, our findings indicate that the contextual information embedded in MD&A disclosures is relevant for financial statement users.

6.2 Sensitivity to Hyperparameter Selection

We assess the sensitivity of our results to alternative hyperparameter choices and the use of alternative embedding models. Specifically, we examine how the model's estimated parameters θ_0 and θ_1 vary when training under alternative configurations of batch size, learning rate, and dropout rate. We systematically vary each of these three hyperparameters across three levels: batch sizes of 32, 64, and 128; learning rates of 10^{-2}, 10^{-3}, and 5×10^{-4}; and dropout rates of 0%, 10%, and 20%. In sum, we evaluate $3^3 = 27$ unique model configurations. For each combination, we re-train the neural network following the same procedure outlined in Section 3 and calculate a corresponding text-only and contextualized version *PI*.

Figure 4 summarizes the performance of all model variants. Specifically, the graph shows trends in the adjusted R^2 across all 27 model configurations for three regressions of tax expense on versions of *PI*: standard pre-tax income, which is the same across all configurations (equivalent to Table 3, Column (1); solid line); contextualized pre-tax income (equivalent to

Table 3, Column (3); short-dashed line); and both (equivalent to Table 3, Column (4); long-dashed line). We report results for *TXT* (*TXPD*) as tax expense variable in Panel A (*B*). For both tax variables and across all hyperparameter configurations, we find that $PI^{Context}$ consistently outperforms traditional pre-tax book income. In Panel A, the average unexplained portion of the full model variation is 13.48% lower when adding context, compared to standard pre-tax income, confirming the robustness of our main finding. Moreover, the average additional adjusted R^2 in the combined model is just 0.48% compared to that of $PI^{Context}$, suggesting that the contextualized variables subsume most of the signal provided in *PI*. Similarly, for *TXPD* as dependent variable, we find an average reduction in unexplained variation of 17.86%, and a 0.55% improvement of adding both variables instead of just contextualized pre-tax income. However, we also observe meaningful differences in model performance across hyperparameter settings. Specifically, models tend to perform better when learning rates are smaller, consistent with the intuition that larger learning rates tend to "overshoot" optimal results. Moreover, we note that the combination of high learning rates and high dropout rates leads to the smallest improvements relative to standard *PI*, whereas higher dropout rates slightly improve estimation in in other specifications, suggesting that dropout helps prevent overfitting by deactivating a fraction of the neural network during training. We observe little sensitivity to our choice of batch size.

In Panels C and D of Figure 4, we repeat the hyperparameter sensitivity analysis using narrative disclosures from income tax footnotes rather than MD&A sections. As with the MD&A-based models, we vary batch size, learning rate, and dropout rate across the 27 possible configurations and plot the resulting adjusted R^2 values for regressions of tax expense (Panel C) and cash taxes paid (Panel D). Strikingly, the contextualized pre-tax income measures derived from footnotes provide far weaker incremental explanatory power than those based on MD&A disclosures. For tax expense in Panel C, $PI^{Context\_Footnote}$ consistently underperforms relative to the baseline, reinforcing our earlier regression findings (Table 4) that footnote disclosures add little incremental context for accrual-based tax expense, and, in fact, appear to reduce explanatory power. In Panel D, which examines cash taxes paid, we find a similar though slightly more favorable role for footnote-based context. $PI^{Context\_Footnote}$ is still consistently below but is much closer to the baseline adjusted R^2 of *PI*. Overall, Panels C and D provide further evidence that footnote-based context provides limited incremental information beyond standard pre-tax income.

Next, we examine whether our findings are sensitive to our choice of embedding model. We use SBERT for our main analysis because it is designed to be a sentence embedding model as opposed to other models, including BERT, which are word embedding models that produce the CLS token as a "byproduct". However, we note that SBERT possibly faces other disadvantages that other embedding models could potentially overcome. For example, FinBERT is a fine-tuned version of BERT that is trained on a corpus of financial texts, allowing it to produce more accurate embeddings in financial contexts (Huang et al., 2023). Other models, such as E5 and BAAI General Embedding (BGE) models are trained on larger corpora and with more recent algorithms, possibly allowing for superior general-purpose performance. We compare the performance of SBERT to that of BERT, FinBERT, E5, and BGE using the same 27 hyperparameter specifications as above, and evaluate models based on the performance of the versions of $PI^{Context}$ they produce, relative to standard *PI*.[6]

We report the results in Figure 5. We find that all five models consistently outperform standard *PI* in explaining variation in tax expense across all 27 hyperparameter configurations. Out of all five models, we find that BERT generally performs the weakest across most configurations. E5 and BGE generally perform similarly well, although both produce smaller

[6] Specifically, we employ BERT's base-uncased model, SBERT's base model, E5 large v2, and BAAI's General Embedder large English v1.5, and the FinBERT model developed by Huang et al. (2023).

improvements in adjusted R^2 than SBERT or FinBERT. This is consistent with expectations because SBERT is pre-trained to embed sentences over words and FinBERT is pre-trained to embed context in the financial domain. The average unexplained portion of the full model variation is 13.53% (19.17%) lower when adding context to regressions explaining *TXT* (*TXPD*). Thus, FinBERT's performance is on par with SBERT's in our accrual-based analyses and slightly exceeds it in our cash-based analysis (though this is mostly driven by higher performance in configurations with smaller learning rates). We conclude that our findings are robust to a large number of alternative model specifications. The findings reinforce our conclusion that context improves the mapping between tax expense and pre-tax book income.

7 CONCLUSION

A large stream of empirical research on corporate tax avoidance focuses on explaining variation in the relation between pre-tax book income and tax outcomes (e.g., tax expense, cash taxes paid), often in the form of ETRs or book-tax differences (Hanlon and Heitzman, 2010; Wilde and Wilson, 2018). This study examines to what extent narrative disclosures can inform readers about unexplained variation in the relation between income and tax expense. Specifically, we leverage computational advancements in LLMs and deep learning to quantify the information in corporate narrative disclosures. We use these "embedded" disclosures to train neural networks with the goal of minimizing the difference between pre-tax book income and tax outcomes, enabling us to evaluate the relative value of contextually informed pre-tax book income over their standard version in explaining variation in tax outcomes.

We find that adding context from MD&A sections to the relation between pre-tax book income and tax expense explains a substantial portion of unexplained variation in both GAAP tax expense and cash taxes paid, highlighting the informational value of narrative disclosures. Importantly, we do not find an equivalent informational value of context conveyed through income tax footnotes, suggesting that MD&A disclosures provide more informative value.

Moreover, we use LDA to analyze which disclosure topics drive the informational value of context. We find that topics directly linked to firms' tax outcomes (e.g., geographic disclosures, discussions of intellectual property) and broader economic activity (e.g., performance, revenues, and customers) contribute to a better mapping between pre-tax book income and tax expense, whereas possibly complex topics (e.g., M&A disclosures, technical discussions on accounting rules) do not. These findings are consistent with the idea that information on firms' operating environments drive the informational value of context. Finally, we identify several firm characteristics that are associated with the performance increase we gain by using contextually informed measures instead of standard pre-tax book income.

Collectively, our results suggest that adding context to the relation between pre-tax book income and tax outcomes using publicly available narrative disclosures explains about 17.6% to 23.9% of the unexplained variation in said relation. Our study adds to the literatures on tax avoidance and corporate disclosures by showing the relevance of interactions between numeric tax information and narrative disclosures. Moreover, we add to the growing research stream that leverages machine learning techniques for accounting applications by showcasing how such techniques allow researchers and others to overcome limitations in measurement quality.

REFERENCES

Allen, E., D. O'Leary, H. Qu, & C. Swenson. (2021). Tax Specific versus Generic Accounting-Based Textual Analysis and the Relationship with Effective Tax Rates: Building Context. *Journal of Information Systems 35*(2): 115–147.

Amberger, H., J. Gallemore, & J. Wilde. (2024). Corporate Tax System Complexity and Investment Sensitivity to Tax Policy Changes. *Working Paper*.

Belnap, A., K. Kroeger, & J. Thornock. (2025). Explaining Corporate Tax Avoidance. *Working Paper*.

Bilicka, K., E. Casi, C. Seregni, & B. Stage. (2025). Tax Strategy Disclosure: A Greenwashing Mandate? *Journal of Accounting Research*, forthcoming.

Blei, D., A. Ng, & M. Jordan. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*(1): 993–1022.

Bogachek, O., A. De Vito, P. Demere, & F. Grossetti. (2024). Using Narrative Disclosures to Predict Tax Outcomes. *Working Paper*.

Burd, C., E. Casi, F. Grossetti, & P. Lisowsky. (2025). FinTag2Text: Mapping Financial Statement Disclosures with XBRL Data. *Working Paper*.

Bratten, B., C. Gleason, S. Larocque, & L. Mills. (2017). Forecasting Taxes: New Evidence from Analysts. *The Accounting Review 92*(3): 1–29.

Choi, G.Y., & A. Kim. (2024). Firm-Level Tax Audits: A Generative AI-Based Measurement. *Working Paper*.

Clarkson, P., J. Kao, & G. Richardson. (1999). Evidence That Management Discussion and Analysis (MD&A) is a Part of a Firm's Overall Disclosure Package. *Contemporary Accounting Research 16*(1): 111–134.

Davis, A., J. Piger, & L. Sedor. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research 29*(3): 845–868.

Devlin, J., M.-W. Chang, K. Lee, & K. Toutanova. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Working Paper*.

Dechow, P., R. Sloan, & A. Sweeney. (1995). Detecting Earnings Management. *The Accounting Review 70*(2): 193–225.

Dyer, T., M. Lang, & L. Stice-Lawrence. (2017). The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics 64*(2–3): 221–245.

Edwards, A., A. Kubata, & T. Shevlin. (2021). The Decreasing Trend in U.S. Cash Effective Tax Rates: The Role of Growth in Pre-Tax Income. *The Accounting Review 96*(5): 231–261.

Farrell, M., T. Liang, & S. Misra. (2021). Deep Learning for Individual Heterogeneity: An Automatic Inference Framework. *Working Paper*.

Feldman, R., S. Govindaraj, J. Livnat, & B. Segal. (2010). Management's Tone Change, Post Earnings Announcement Drift and Accruals. *Review of Accounting Studies 15*(4): 915–953.

Guenther, D., K. Peterson, J. Searcy, & B. Williams. (2023). How Useful Are Tax Disclosures in Predicting Effective Tax Rates? A Machine Learning Approach. *The Accounting Review 98*(5): 297–322.

Hanlon, M., & S. Heitzman. (2010). A Review of Tax Research. *Journal of Accounting and Economics 50*(2–3): 127–178.

Hechtner, F., L. Schmidt, A. Seebeck, & M. Weiß. (2025). How to Design and Employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT. *Working Paper*.

Henry, E., & A. Leone. (2016). Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *The Accounting Review 91*(1): 153–178.

Hoberg, G., G. Phillips, & N. Prabhala. (2014). Product Market Threats, Payouts, and Financial Flexibility. *The Journal of Finance 69*(1): 293–324.

Hoopes, J. (2018). The Effect of Temporary Tax Laws on Understanding and Predicting Corporate Earnings. *Working Paper*.

Hope, O.-K., D. Hu, & H. Lu. (2016). The Benefits of Specific Risk-Factor Disclosures. *Review of Accounting Studies 21*: 1005–1045.

Hoppe, T., D. Schanz, S. Sturm, & C. Sureth-Sloane. (2023). The Tax Complexity Index – A Survey-Based Country Measure of Tax Code and Framework Complexity. *European Accounting Review 32*(2): 239–273.

Huang, A., H. Wang, & Y. Yang. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research 40*(2): 806–841.

Hutchens, M. (2017). Can Disclosure Characteristics Improve Analyst Forecast Accuracy? *Working Paper*.

Jennings, J., J. Lee, & E. Towery. (2024). Estimating the Likelihood of Future Tax Settlements. *Working Paper*.

Jones, J. (1991). Earnings Management During Import Relief Investigations. *Journal of Accounting Research 29*(2): 193–228.

Kim, A., & V. Nikolaev. (2024a). Contextualizing Profitability. *Working Paper*.

Kim, A., & V. Nikolaev. (2024b). Context-Based Interpretation of Financial Information. *Journal of Accounting Research*, forthcoming.

Kothari, S., X. Li, & J. Short. (2009). The Effect of Disclosures by Management, Analysts, and Financial Press on the Equity Cost of Capital: A Study Using Content Analysis. *The Accounting Review 84*(5): 1639–1670.

Lang, M., & L. Stice-Lawrence. (2015). Textual Analysis and International Financial Reporting: Large Sample Evidence. *Journal of Accounting and Economics 60*: 110–135.

Law, K., & L. Mills. (2015). Taxes and Financial Constraints: Evidence from Linguistic Cues. *Journal of Accounting Research 53*(4): 777–819.

Li, F. (2008). Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics 45*(2–3): 221–247.

Li, F. (2010). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature 29*: 143–165.

Loughran, T., & B. McDonald. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance 66*(1): 35–65.

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

Loughran, T., & B. McDonald. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics 12*: 357–375.

Mood, A. (1969). Macro-Analysis of the American Educational System. *Operations Research 17*(5): 770–784.

Newton, R., & D. Spurrell. (1967). A Development of Multiple Regression for the Analysis of Routine Data. *Applied Statistics 16*(1): 51–64.

Plumlee, M. (2003). The Effect of Information Complexity on Analysts' Use of That Information. *The Accounting Review 78*(1): 275–296.

Reimers, N., & I. Gurevych. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Working Paper*.

Richter, S., J. Seidman, R. Sinha, & B. Stomberg (2024). How Tax Executives Craft Income Tax Disclosures in Response to Tax-Based Proprietary Costs. *Working Paper*.

Rogers, J., A. Van Buskirk, & S. Zechman. (2011). Disclosure Tone and Shareholder Litigation. *The Accounting Review 86*(6): 2155–2183.

Schwab, C., B. Stomberg, & J. Xia. (2022). What Determines Effective Tax Rates? The Relative Influence of Tax and Other Factors. *Contemporary Accounting Research 39*(1): 459–497.

SEC. (1989). Release No. 33-6835.

Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica 57*(2) 307–333.

Wilde, J., & R. Wilson. (2018). Perspectives on Corporate Tax Planning: Observations from the Past Decade. *Journal of the American Taxation Association 40*(2): 63–81.

APPENDIX

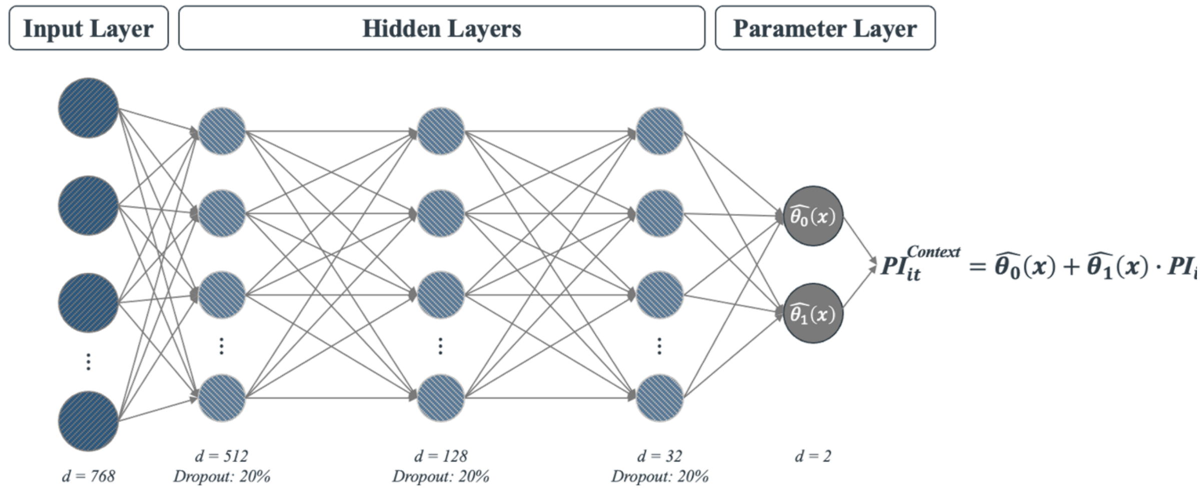
Variable Name	Variable Definition
Main Variables	
PI	Pre-tax book income.
TXT	Total tax expense divided by the U.S. corporate income tax rate in the associated year.
TXPD	Cash taxes paid divided by the U.S. corporate income tax rate in the associated year.
PI^{Text} *(TXT)*	Neural network-based estimation of pre-tax book income, trained on past values of *TXT* and SBERT embeddings representing firm-years' narrative disclosures (i.e., MD&As), calculated using current-year SBERT embeddings as per Eq. (6). For parsimony, we include the "*(TXT)*" qualifier only in descriptive statistics.
$PI^{Context}$ *(TXT)*	Neural network-based variation of *PI*, trained on past values of *PI*, *TXT*, and SBERT embeddings representing firm-years' narrative disclosures (i.e., MD&As), and calculated using current-year SBERT embeddings and *PI* as per Eq. (5). For parsimony, we include the "*(TXT)*" qualifier only in descriptive statistics.
PI^{Text} *(TXPD)*	Neural network-based estimation of pre-tax book income, trained on past values of *TXPD* and SBERT embeddings representing firm-years' narrative disclosures (i.e., MD&As), calculated using current-year SBERT embeddings as per Eq. (6). For parsimony, we include the "*(TXPD)*" qualifier only in descriptive statistics.
$PI^{Context}$ *(TXPD)*	Neural network-based variation of *PI*, trained on past values of *PI*, *TXPD*, and SBERT embeddings representing firm-years' narrative disclosures (i.e., MD&As), and calculated using current-year SBERT embeddings and *PI* as per Eq. (5). For parsimony, we include the "*(TXPD)*" qualifier only in descriptive statistics.
$PI^{Text\_Footnote}$ *(TXT)*	Neural network-based estimation of pre-tax book income, trained on past values of *TXT* and SBERT embeddings representing firm-years' narrative disclosures (i.e., income tax footnotes), calculated using current-year SBERT embeddings as per Eq. (6). For parsimony, we include the "*(TXT)*" qualifier only in descriptive statistics.
$PI^{Context\_Footnote}$ *(TXT)*	Neural network-based variation of *PI*, trained on past values of *PI*, *TXT*, and SBERT embeddings representing firm-years' narrative disclosures (i.e., income tax footnotes), and calculated using current-year SBERT embeddings and *PI* as per Eq. (5).

	For parsimony, we include the "*(TXT)*" qualifier only in descriptive statistics.
$PI^{Text\_Footnote}$ *(TXPD)*	Neural network-based estimation of pre-tax book income, trained on past values of *TXPD* and SBERT embeddings representing firm-years' narrative disclosures (i.e., income tax footnotes), calculated using current-year SBERT embeddings as per Eq. (6). For parsimony, we include the "*(TXPD)*" qualifier only in descriptive statistics.
$PI^{Context\_Footnote}$ *(TXPD)*	Neural network-based variation of *PI*, trained on past values of *PI*, *TXPD*, and SBERT embeddings representing firm-years' narrative disclosures (i.e., income tax footnotes), and calculated using current-year SBERT embeddings and *PI* as per Eq. (5). For parsimony, we include the "*(TXPD)*" qualifier only in descriptive statistics.
Additional Variables	
Informative	Indicator variable that equals one if the absolute value of the residual in a linear regression of *TXT* on $PI^{Context\_TXT}$ (i.e., Table 3, Column 3) is smaller than the respective value of *TXT*, and zero otherwise.
ETR Forecast Error	Mean absolute difference between a firm's actual ETR and the implied ETR forecasted by analysts within 90 days prior to the fiscal year-end. Implied ETR forecasts are the difference between the estimates for pre-tax income and net income, divided by the estimate for pre-tax income.
Adjustment	Absolute value of the difference between contextualized pre-tax income and pre-tax book income, as a percentage of pre-tax book income.
Wedge	Absolute value of the difference between *ETR* and *xETR*, where *xETR* is tax expense divided by contextualized pre-tax income.
Accruals Quality	Absolute value of discretionary accruals as estimated using the Jones (1991) model in its modified version as per Dechow et al. (1995).
Age	Natural logarithm of one plus the number of years between the current firm-year and the firm's first year with data availability in Compustat.
Analyst Following	Number of analysts following a firm in a given year.
Big N Auditor	Indicator variable that equals one if the auditor is a Big Four accounting firm (or any of its predecessors), and zero otherwise.
Boilerplate	Percentage of sentences in the MD&A that contain at least one boilerplate tetragram. Boilerplate tetragrams are defined following Lang & Stice-Lawrence (2015).

Business Segments	Number of business segments.
ETR	Tax expense divided by pre-tax book income.
ETR Indicator	Indicator variable that equals one if the GAAP ETR is above the sample median, and zero otherwise.
Fluidity Indicator	Indicator variable that equals one if product market fluidity as per Hoberg et al. (2014) is above the sample median, and zero otherwise.
Fog	Fog index of the MD&A, calculated as 0.4 × (Number of words / Number of sentences + Complex words / Total words). Complex words are words with three or more syllables.
Geographic Segments	Number of geographic segments.
Intangible	Intangible assets, divided by total assets.
Length	Natural logarithm of the total number of words in the MD&A.
Leverage	Sum of current and long-term total debt, scaled by total assets.
M&A Indicator	Indicator variable that equals one if special items before taxes that correspond to merger and acquisition are non-missing and non-zero, and zero otherwise.
MNC Status	Indicator variable that equals one if foreign pre-tax income is non-missing and foreign pre-tax income or foreign tax expense are non-zero, and zero otherwise.
Market-to-Book	Closing stock price times common shared outstanding on balance sheet date, divided by common equity.
Negativity	Percentage of negative words in the MD&A, as per Loughran & McDonald (2011).
R&D	Research and development expenditures, divided by total assets. Set to zero if missing.
ROA	Pre-tax income, divided by total assets.
Size	Natural logarithm of one plus total assets.
Size Indicator	Indicator variable that equals one if *Size* is above the sample median, and zero otherwise.
Specificity	Number of named entities reported in the MD&A, divided by the number of words in the MD&A, after removing stopwords, following Hope et al. (2016). Named entities are identified using the Natural Language Toolkit.

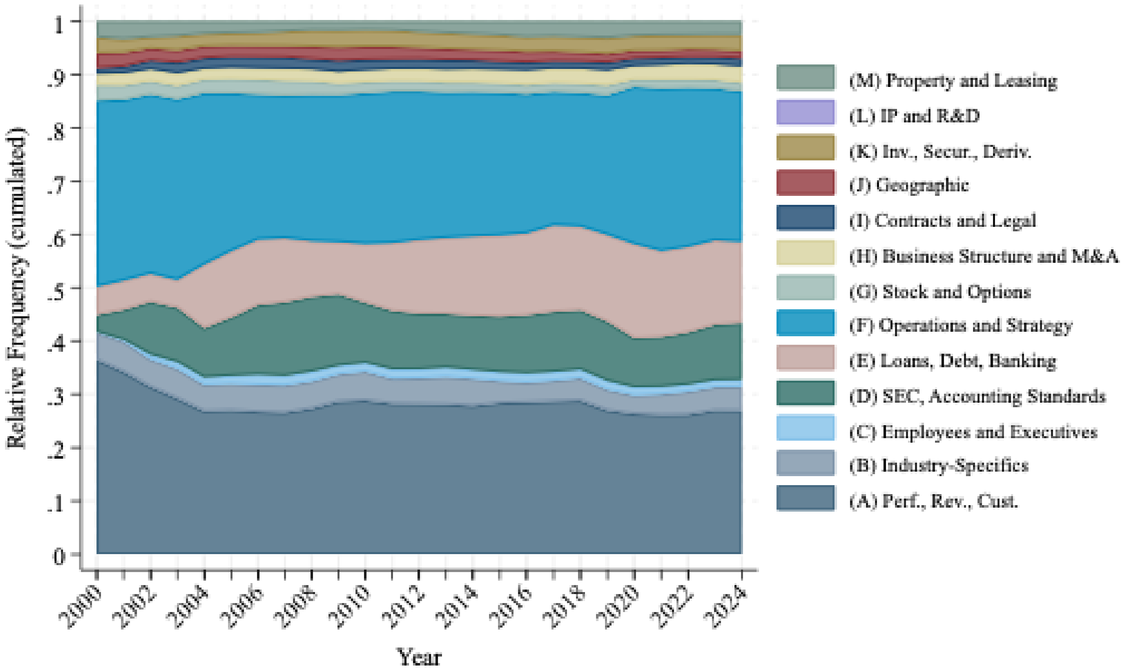
FIGURES AND TABLES

FIGURE 1 Visualization of Deep Neural Network



Notes: This figure describes the second step of the contextualization of pre-tax book income. Specifically, we use a deep neural network to transform SBERT-based sentence embeddings (represented by one 768-dimensional vector per document) into two parameters, θ_0 and θ_1. The deep neural network comprises three hidden layers with 512, 128, and 32 dimensions, respectively. Throughout training, we establish a dropout rate to randomly a fraction of all nodes in each layer to avoid overfitting. At the end of each epoch, we use the loss function described in Eq. (4) to evaluate the accuracy of the deep neural network with respect to the validation set (20% from observations randomly selected from the training set). The deep neural network trains for up to 250 epochs; training ends early if the loss function does not improve by at least 0.001 for ten consecutive periods. In this case, we retain the model weights from the epoch with the lowest loss value with respect to the validation set. We apply the final network to our test set (i.e., all observations in the period immediately following the respective training period) to obtain $PI^{Context}$.

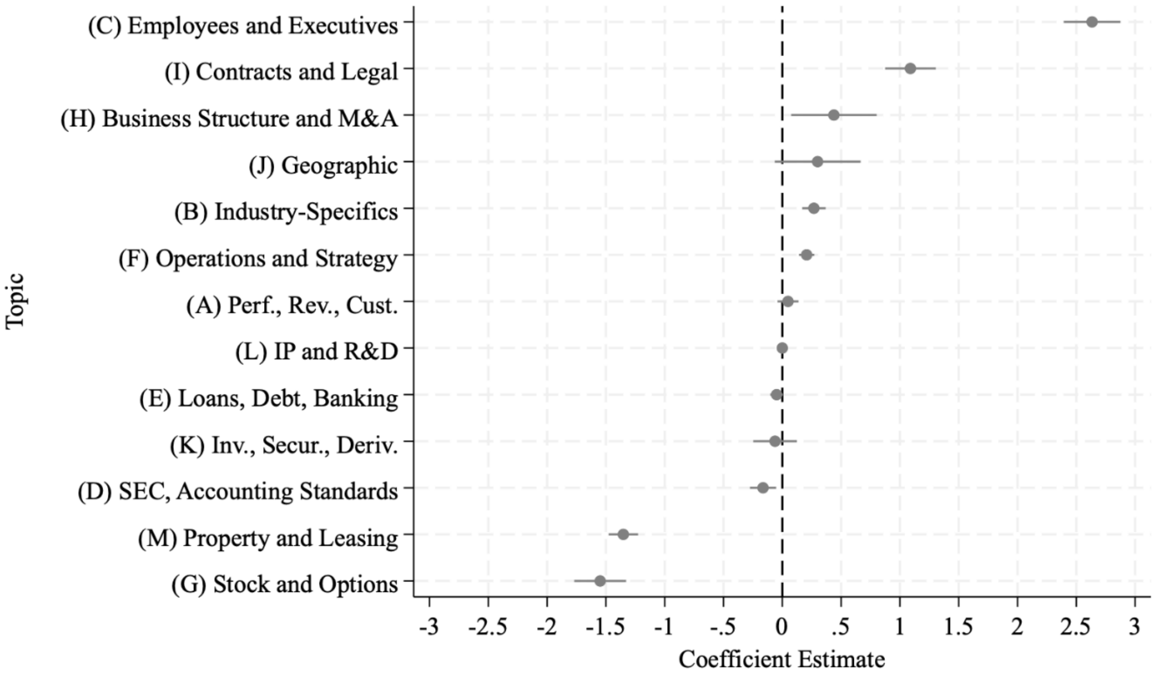
FIGURE 2 Time Trends in Topic Models



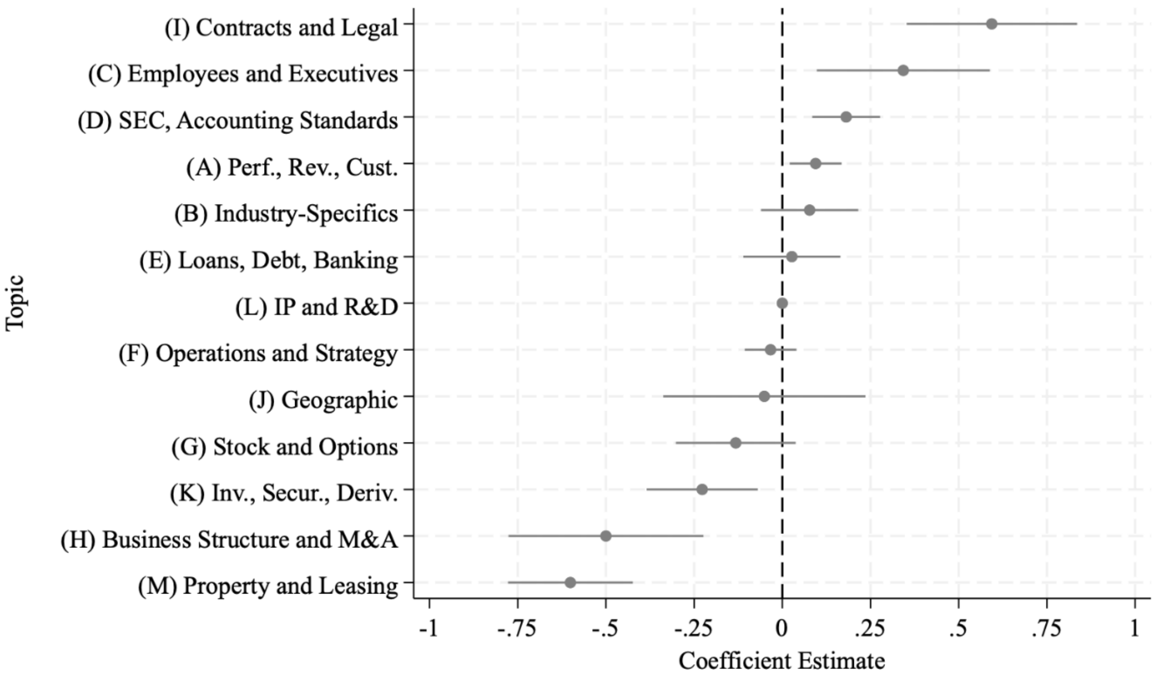
Notes: This figure presents the relative frequency of different disclosure topics over time. Topics are identified using LDA.

FIGURE 3 Topic Analysis

Panel A: Without Controls



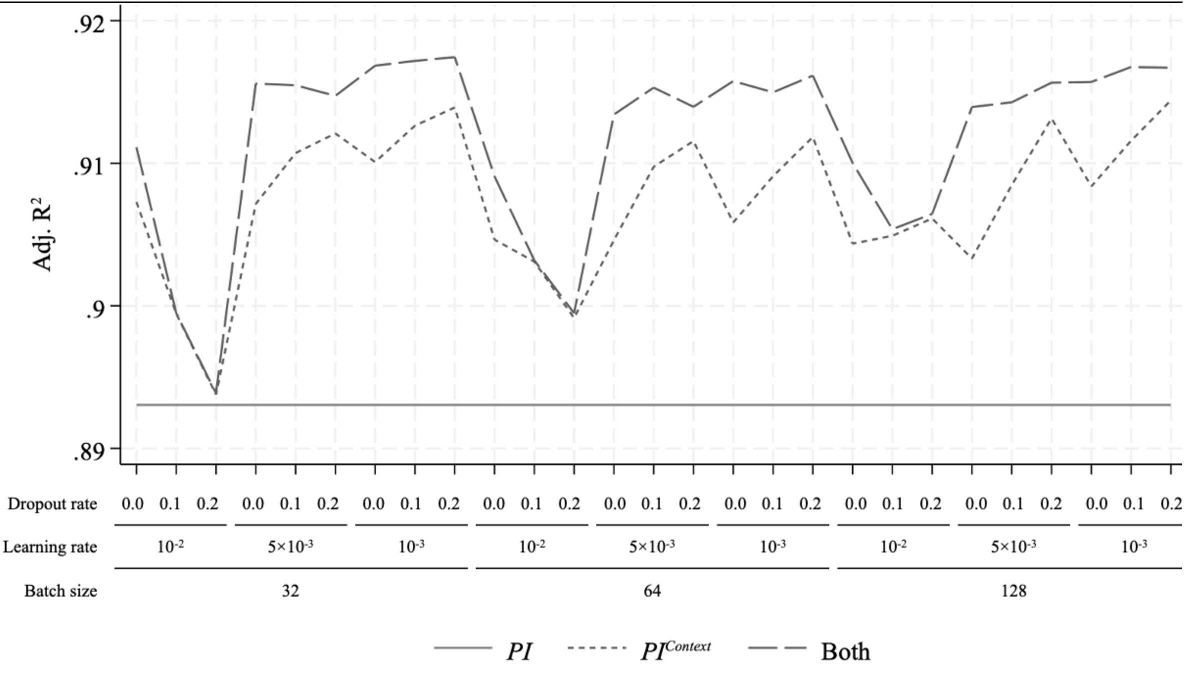
Panel B: With Controls



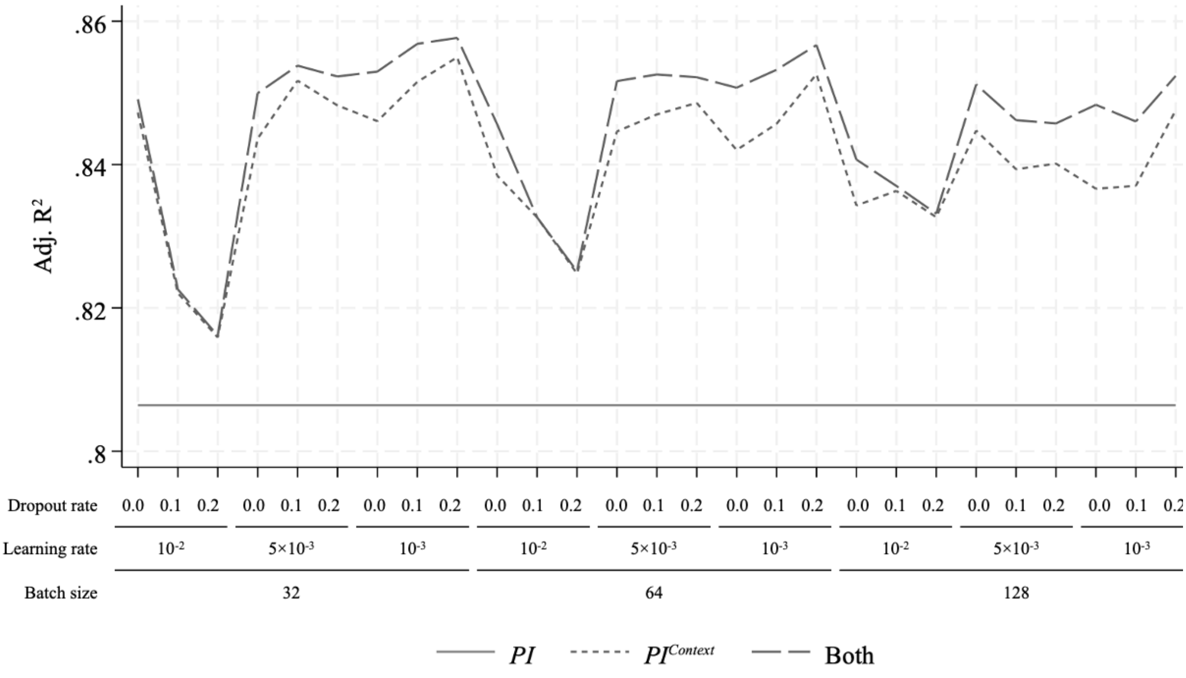
Notes: This figure presents the coefficient estimates from 13 separate linear OLS regressions of *Informative* on the relative frequency of one of 13 topic categories ((A) through (M)). The horizontal lines reflect 95% confidence intervals. Panel A presents the results without control variables or fixed effects, while Panel B includes an array of control variables and industry- and year-fixed effects. We cluster standard errors at the firm level.

FIGURE 4 Robustness: Hyperparameter Selection

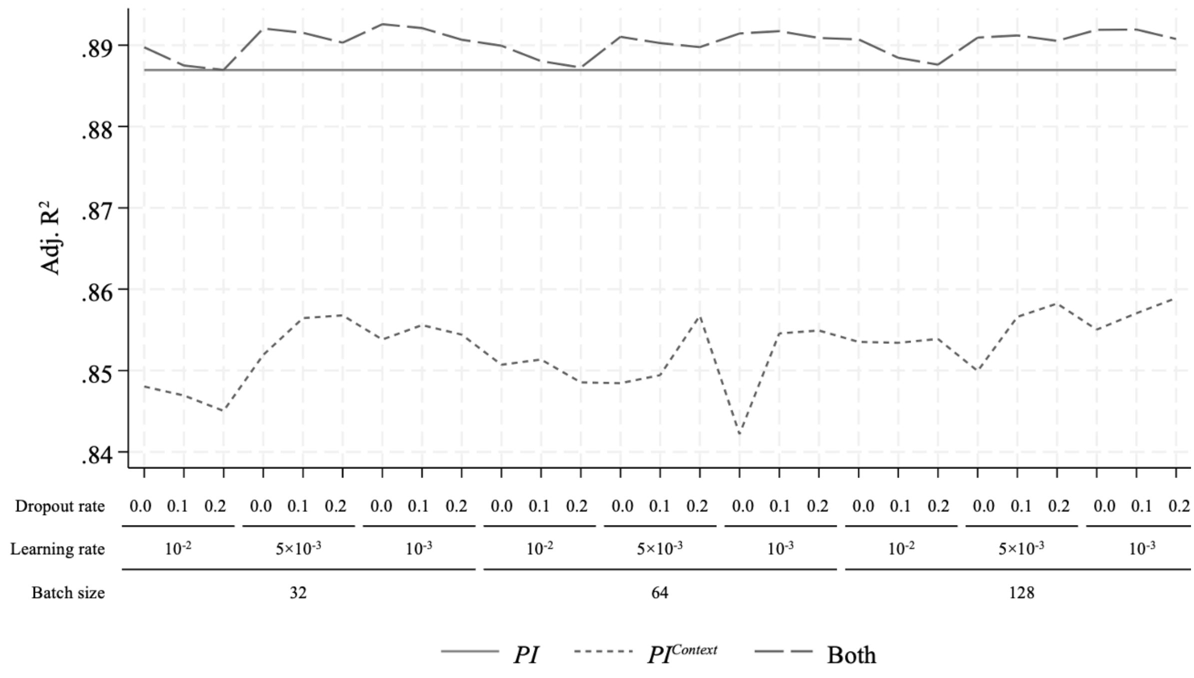
Panel A: Tax expense; context from MD&A sections



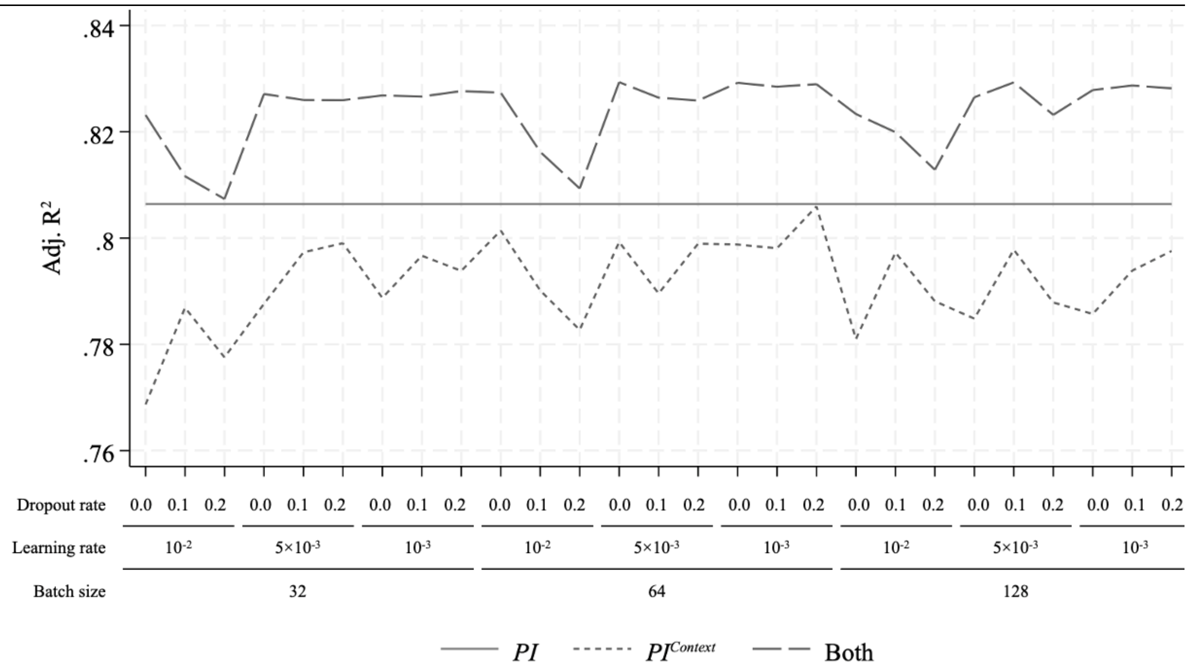
Panel B: Cash taxes paid; context from MD&A sections



Panel C: Tax expense; context from income tax footnotes



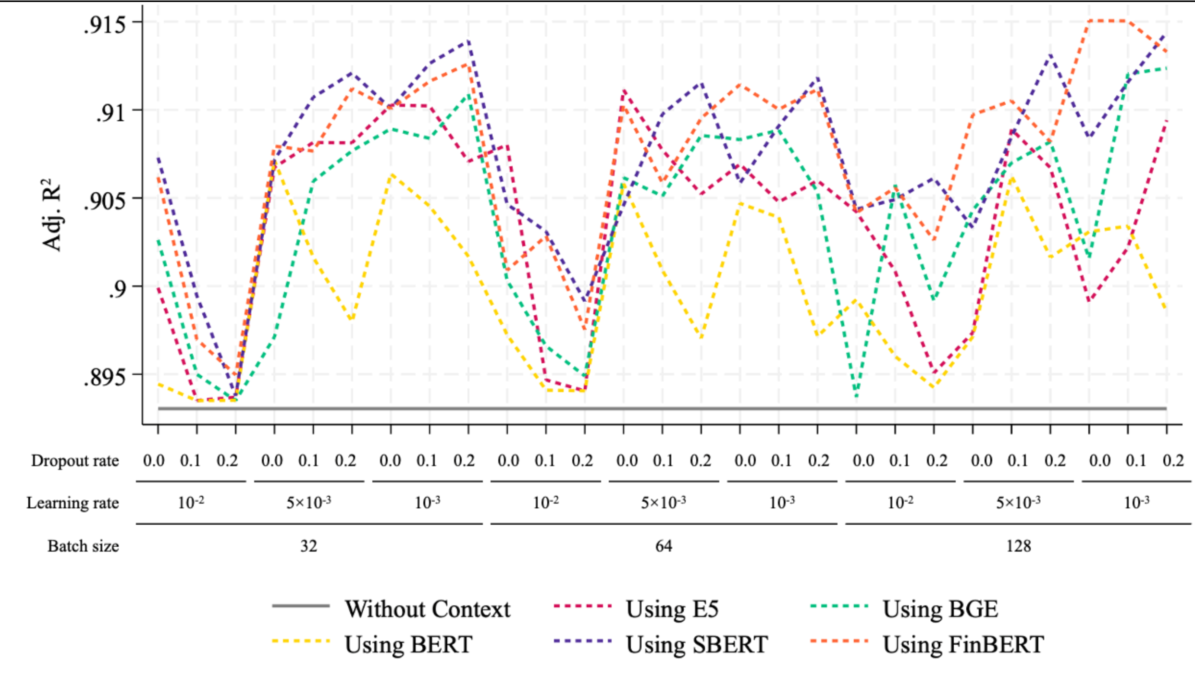
Panel D: Cash taxes paid; context from income tax footnotes



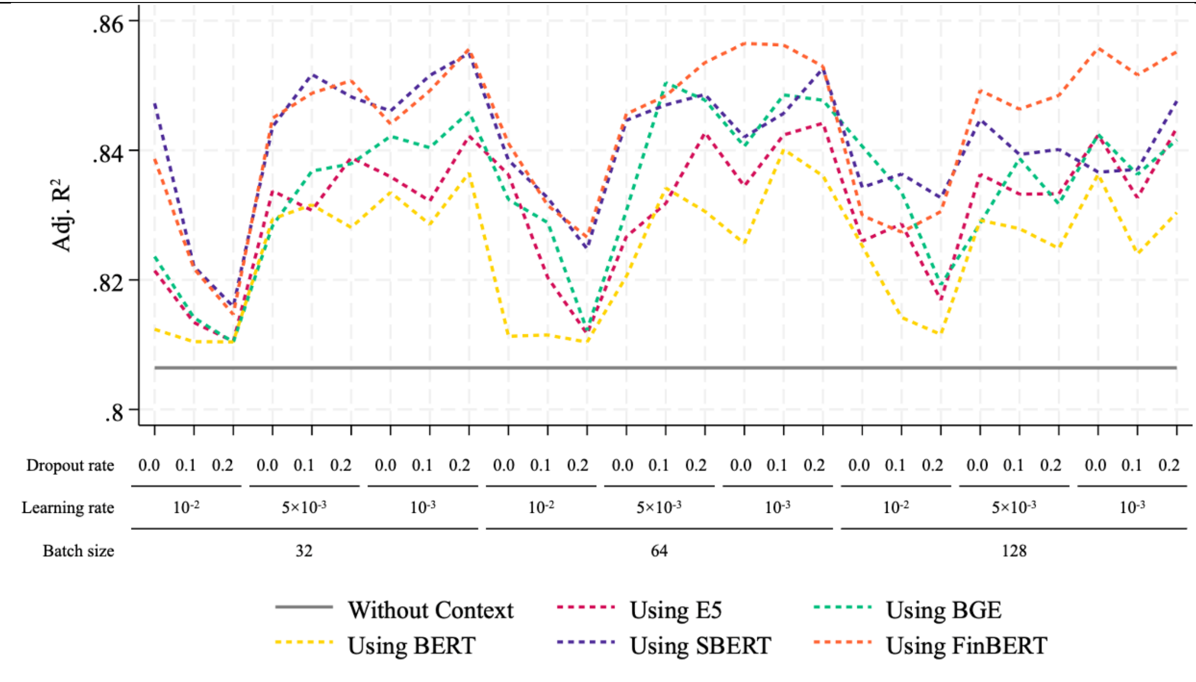
Notes: This figure presents the sensitivity of our main results to alternative hyperparameter choices. Specifically, we train the same neural networks in 27 different specifications by varying batch size (32, 64, or 128), learning rate (10^{-2}, 5×10^{-2}, or 10^{-3}), and dropout rate (0, 0.1, or 0.2). We run the resulting versions of *PI*^{Context} both in a stand-alone version and in a joint model with *PI*, i.e., in linear OLS regressions akin to those presented in Tables 3 and 4, Columns (3) and (4). We run the baseline model containing only *PI* as a benchmark. The lines represent the adjusted R^2 of the three resulting regressions per hyperparameter configuration. In Panel A (B), we report the results for tests using *TXT* (*TXPD*) as the dependent variable and narrative disclosures from MD&As. In Panel C (D), we report the results for tests using *TXT* (*TXPD*) as the dependent variable and narrative disclosures from income tax footnotes.

FIGURE 5 Robustness: Alternative Embedding Models

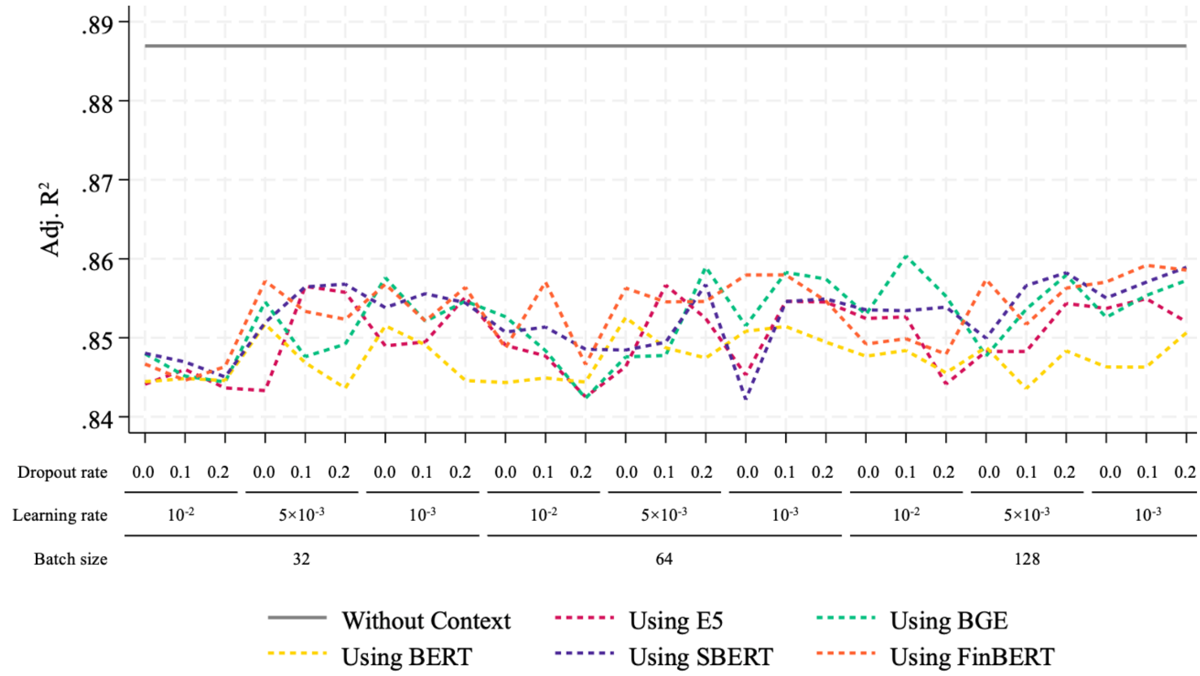
Panel A: Tax expense; context from MD&A sections



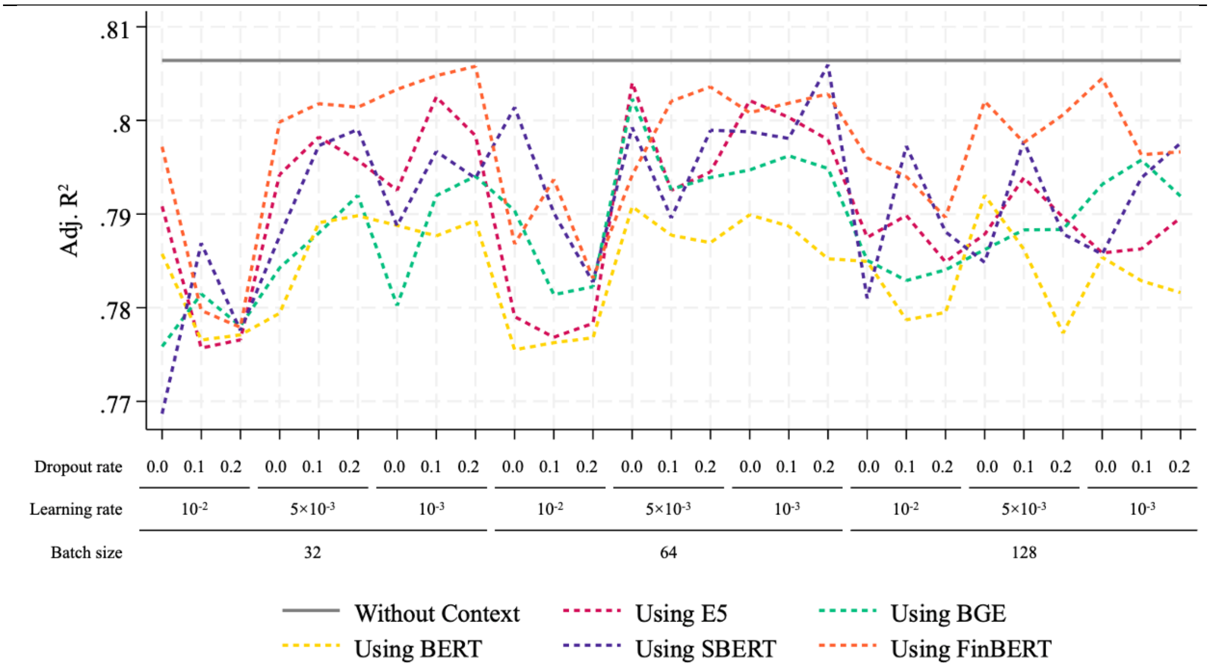
Panel B: Cash taxes paid; context from MD&A sections



Panel C: Tax expense; context from income tax footnotes



Panel D: Cash taxes paid; context from income tax footnotes



Notes: This figure presents the sensitivity of our main results to alternative hyperparameter choices and embedding models. Specifically, we train the same neural networks in 27 different specifications by varying batch size (32, 64, or 128), learning rate (10^{-2}, $5{\times}10^{-2}$, or 10^{-3}), dropout rate (0, 0.1, or 0.2), and embedding models (E5, BGE, BERT, SBERT, or FinBERT). We run the resulting versions of $PI^{Context}$ in simple regressions, i.e., in linear OLS regressions akin to those presented in Tables 3 and 4, Column (3). We run the baseline model containing only *PI* as a benchmark. The lines represent the adjusted R^2 of the resulting regression per hyperparameter configuration and embedding model. In Panel A (B), we report the results for tests using *TXT* (*TXPD*) as the dependent variable and narrative disclosures from MD&As. In Panel C (D), we report the results for tests using *TXT* (*TXPD*) as the dependent variable and narrative disclosures from income tax footnotes.

TABLE 1 Sample Selection

	Requirement	Less	# Obs.
	Firm-years in Compustat with fiscal years ending between 1996 and 2024		245,014
./.	No identifiable 10-K filing from SEC EDGAR	(75,684)	169,330
./.	No identifiable MD&A with at least 500 words	(16,848)	152,482
./.	Non-missing and non-negative data on pre-tax income and tax expense	(84,237)	68,245
./.	Non-missing data on control variables	(8,241)	60,004
./.	Initial four-year training window	(7,891)	52,113
=	**MD&A Sample**		**52,113**
./.	Firm-years ending earlier than 2011	(23,454)	28,659
./.	No identifiable income tax footnote with at least 500 words	(5,787)	22,872
./.	Initial four-year training window	(6,763)	16,109
=	**Income Tax Footnote Sample**	**14,163**	**14,163**

Notes: This table summarizes the sample selection procedure. We obtain firm-year level data on tax outcomes from Compustat. We obtain MD&As from SEC EDGAR, while income tax footnotes are from CalcBench. We reserve the first four years of our sample periods for the initial training window, reducing the effective number of observations available for evaluation of our deep learning models. The sample sizes for tests involving additional variables are smaller because they require non-missing information on additional variables. If data requirements for the income tax footnote sample differ from those for the MD&A sample, we report them in square brackets.

TABLE 2 Descriptive Statistics

Panel A: Summary Statistics

Dependent variable	N	Mean	SD	P25	Median	P75
PI	52,113	390.444	1042.761	13.080	58.674	244.200
TXT	52,113	337.940	906.798	8.089	45.914	208.740
TXPD	52,113	301.465	865.317	4.586	32.520	162.857
PI^{Text} *(TXT)*	52,113	281.849	676.278	30.597	78.818	204.853
$PI^{Context}$ *(TXT)*	52,113	309.868	817.363	11.600	46.892	193.777
PI^{Text}*(TXPD)*	52,113	237.330	627.887	20.738	53.784	152.624
$PI^{Context}$ *(TXPD)*	52,113	247.178	685.622	8.024	32.605	140.681
$PI^{Text\_Footnote}$*(TXT)*	16,109	561.636	987.885	124.949	226.669	475.810
$PI^{Context\_Footnote}$*(TXT)*	16,109	632.470	1475.577	33.372	122.875	471.068
$PI^{Text\_Footnote}$ *(TXPD)*	16,109	550.701	1114.022	99.691	195.868	405.207
$PI^{Context\_Footnote}$ *(TXPD)*	16,109	545.097	1325.569	24.469	95.373	389.125

Panel B: Pairwise Correlations

	PI	*TXT*	*TXPD*	PI^{Text} *(TXT)*	$PI^{Context}$ *(TXT)*	PI^{Text} *(TXPD)*	$PI^{Context}$ *(TXPD)*	$PI^{Text\_Footnote}$ *(TXT)*	$PI^{Context\_Footnote}$ *(TXT)*	$PI^{Text\_Footnote}$ *(TXPD)*	$PI^{Context\_Footnote}$ *(TXPD)*
PI	1										
TXT	0.9450	1									
TXPD	0.8980	0.9115	1								
PI^{Text} *(TXT)*	0.8131	0.8157	0.8023	1							
$PI^{Context}$ *(TXT)*	0.9741	0.9549	0.9072	0.8286	1						
PI^{Text} *(TXPD)*	0.7979	0.7993	0.8208	0.9291	0.8034	1					
$PI^{Context}$ *(TXPD)*	0.9513	0.9332	0.9234	0.8129	0.9609	0.8412	1				
$PI^{Text\_Footnote}$ *(TXT)*	0.5943	0.5934	0.6048	0.5818	0.6045	0.5799	0.6098	1			
$PI^{Context\_Footnote}$ *(TXT)*	0.9640	0.9246	0.8874	0.8062	0.9510	0.7958	0.9309	0.6096	1		
$PI^{Text\_Footnote}$ *(TXPD)*	0.5914	0.5890	0.6224	0.5716	0.6018	0.5928	0.6246	0.9078	0.6101	1	
$PI^{Context\_Footnote}$ *(TXPD)*	0.9450	0.9079	0.8978	0.7969	0.9339	0.8129	0.9380	0.6194	0.9733	0.6473	1

Notes: This table reports summary statistics. Panel A (B) presents descriptive statistics (pairwise Pearson correlations) for all dependent and independent variables included in our main analysis reported in Tables 3 and 4. All continuous variables are winsorized at the 1st and 99th percentile. All variables are as defined in Appendix A.

TABLE 3 Informational Value of Context

Panel A: Tax expense

Dependent variable	*TXT* (1)	*TXT* (2)	*TXT* (3)	*TXT* (4)	*TXT* (5)
PI	0.8218***			0.2524***	0.2450***
	(75.442)			(5.658)	(5.529)
PI^{Text}		1.0938***			0.0975***
		(39.417)			(5.779)
$PI^{Context}$			1.0594***	0.7457***	0.6880***
			(88.374)	(14.093)	(13.888)
Intercept	17.0754***	29.6548***	9.6692***	8.3146***	1.5852
	(4.837)	(3.624)	(2.868)	(2.962)	(0.579)
p-value$_{PI=1}$	0.000***	0.002***	0.000***	0.000***	0.000***
Vuong's Test		-7.421***	2.981***		
Clustered SE	Firm + Year	Firm + Year	Firm + Year	Firm + Year	Firm + Year
Adjusted R^2	0.8930	0.6654	0.9118	0.9161	0.9178
N	52,113	52,113	52,113	52,113	52,113

Panel B: Cash taxes paid

Dependent variable	*TXPD* (1)	*TXPD* (2)	*TXPD* (3)	*TXPD* (4)	*TXPD* (5)
PI	0.7452***			0.1713***	0.1744***
	(39.205)			(5.601)	(5.956)
PI^{Text}		1.1312***			0.2100***
		(34.546)			(8.007)
$PI^{Context}$			1.1654***	0.9176***	0.7513***
			(68.148)	(18.551)	(17.454)
Intercept	10.5080**	33.0005***	13.4066***	7.7825**	-2.1716
	(2.123)	(4.037)	(3.710)	(2.354)	(-0.754)
p-value$_{PI=1}$	0.000***	0.001***	0.000***	0.109	0.000***
Vuong's Test		-5.960***	3.668***		
Clustered SE	Firm + Year	Firm + Year	Firm + Year	Firm + Year	Firm + Year
Adjusted R^2	0.8064	0.6737	0.8526	0.8567	0.8635
N	52,113	52,113	52,113	52,113	52,113

Notes: This table presents the results of estimating versions of Eq. (7). Panel A (B) presents the results using *TXT* (*TXPD*) as the dependent variable. In Column (1), we regress the tax outcome variable on pre-tax book income. In Column (2), we regress the tax outcome variable on the text-only variable that approximates the tax outcome using only MD&A narrative disclosures. In Column (3), we regress the tax outcome variable on contextually informed pre-tax book income using information from MD&As. We report a regression including both pre-tax book income and contextually informed pre-tax book income in Column (4). We report the full model in Column (5), including all three measures jointly. We present t-statistics in parentheses. In a separate row, we present the p-value corresponding to t-tests testing whether the respective variable of interest (as highlighted through grey shading) is statistically significantly different from one. Finally, we report the z-statistics of Vuong's tests comparing the explanatory power of non-nested models with the baseline model in Column (1). Standard errors are clustered by firm and year. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively, based on two-tailed tests. All continuous variables are winsorized at the 1st and 99th percentiles. All variables are as defined in Appendix A.

TABLE 4 Informational Value of Context from Income Tax Footnotes

Panel A: Tax expense

Dependent variable	*TXT* (1)	*TXT* (2)	*TXT* (3)	*TXT* (4)	*TXT* (5)
PI	0.8303***			0.6292***	0.6259***
	(78.931)			(12.476)	(12.632)
$PI^{Text\_Footnote}$		0.7489***			0.0503***
		(21.049)			(4.593)
$PI^{Context\_Footnote}$			0.7812***	0.1999***	0.1825***
			(38.467)	(4.556)	(4.387)
Intercept	45.1168***	194.3662***	120.8740***	56.6990***	41.7752***
	(5.084)	(7.308)	(10.276)	(5.842)	(5.807)
p-value$_{PI\,=\,1}$	0.000***	0.000***	0.000***	0.000***	0.000***
Vuong's Test		-9.433***	-4.383***		
Clustered SE	Firm + Year	Firm + Year	Firm + Year	Firm + Year	Firm + Year
Adjusted R^2	0.8869	0.3521	0.8549	0.8909	0.8919
N	16,109	16,109	16,109	16,109	16,109

Panel B: Cash taxes paid

Dependent variable	*TXPD* (1)	*TXPD* (2)	*TXPD* (3)	*TXPD* (4)	*TXPD* (5)
PI	0.7757***			0.4007***	0.4151***
	(40.829)			(9.349)	(9.834)
$PI^{Text\_Footnote}$		0.6825***			0.0964***
		(19.648)			(6.850)
$PI^{Context\_Footnote}$			0.8273***	0.4234***	0.3564***
			(56.360)	(10.159)	(8.743)
Intercept	44.3052**	200.8561***	125.7342***	70.9293***	44.4671***
	(2.893)	(7.895)	(6.303)	(4.194)	(3.360)
p-value$_{PI\,=\,1}$	0.000***	0.000***	0.000***	0.000***	0.000***
Vuong's Test		-8.154***	-0.048		
Clustered SE	Firm + Year	Firm + Year	Firm + Year	Firm + Year	Firm + Year
Adjusted R^2	0.8064	0.3873	0.8060	0.8290	0.8334
N	16,109	16,109	16,109	16,109	16,109

Notes: This table presents the results of estimating versions of Eq. (7). Panel A (B) presents the results using *TXT* (*TXPD*) as the dependent variable. In Column (1), we regress the tax outcome variable on pre-tax book income. In Column (2), we regress the tax outcome variable on the text-only variable that approximates the tax outcome using only income tax footnote narrative disclosures. In Column (3), we regress the tax outcome variable on contextually informed pre-tax book income using information from income tax footnotes. We report a regression including both pre-tax book income and contextually informed pre-tax book income in Column (4). We report the full model in Column (5), including all three measures jointly. We present t-statistics in parentheses. In a separate row, we present the p-value corresponding to t-tests testing whether the respective variable of interest (as highlighted through grey shading) is statistically significantly different from one. Finally, we report the z-statistics of Vuong's tests comparing the explanatory power of non-nested models with the baseline model in Column (1). Standard errors are clustered by firm and year. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively, based on two-tailed tests. All continuous variables are winsorized at the 1st and 99th percentiles. All variables are as defined in Appendix A.

TABLE 5 MD&A Sections versus Income Tax Footnotes

Panel A: Tax expense

Dependent variable	*TXT* (1)	*TXT* (2)	*TXT* (3)	*TXT* (4)
PI			0.3155^{***}	0.2424^{***}
			(5.408)	(3.622)
$PI^{Text\_Footnote}$	0.2292^{***}			0.0196^{***}
	(10.133)			(3.656)
$PI^{Context\_Footnote}$		0.2020^{***}		0.0953^{***}
		(6.180)		(3.449)
PI^{Text}	0.9390^{***}		0.0897^{***}	0.0796^{**}
	(26.638)		(3.880)	(3.243)
$PI^{Context}$		0.8163^{***}	0.6158^{***}	0.5821^{***}
		(19.540)	(9.724)	(10.171)
Intercept	13.8765	47.1366^{***}	21.3426^{***}	23.4149^{***}
	(0.665)	(6.268)	(3.464)	(3.859)
p-value$_{PI^{Footnote}=1}$	0.000^{***}	0.000^{***}		0.000^{***}
p-value$_{PI=1}$	0.118	0.002^{***}	0.000^{***}	0.000^{***}
Vuong's Test (vs. Table 4, Panel A, Col. (5))			3.071^{**}	
Clustered SE	Firm + Year	Firm + Year	Firm + Year	Firm + Year
Adjusted R^2	0.6835	0.9046	0.9077	0.9088
N	16,109	16,109	16,109	16,109

Panel B: Cash taxes paid

Dependent variable	*TXPD* (1)	*TXPD* (2)	*TXPD* (3)	*TXPD* (4)
PI			0.1762^{***}	0.0986^{*}
			(3.931)	(1.907)
$PI^{Text\_Footnote}$	0.2267^{***}			0.0490^{***}
	(10.038)			(4.781)
$PI^{Context\_Footnote}$		0.2358^{***}		0.1502^{***}
		(10.313)		(5.233)
PI^{Text}	0.9652^{***}		0.2039^{***}	0.1737^{***}
	(25.179)		(5.216)	(4.392)
$PI^{Context}$		0.8711^{***}	0.7701^{***}	0.6722^{***}
		(20.906)	(13.275)	(12.654)
Intercept	22.7554	46.0294^{***}	9.6188	12.5680
	(1.200)	(4.124)	(1.073)	(1.488)
p-value$_{PI^{Footnote}=1}$	0.000^{***}	0.000^{***}		0.000^{***}
p-value$_{PI=1}$	0.388	0.013^{**}	0.003^{***}	0.000^{***}
Vuong's Test (vs. Table 4, Panel B, Col. 5)			4.608^{***}	
Clustered SE	Firm + Year	Firm + Year	Firm + Year	Firm + Year
Adjusted R^2	0.7063	0.8622	0.8643	0.8687
N	16,109	16,109	16,109	16,109

Notes: This table presents the results of estimating extended versions of Eq. (7) that compare contextual enhancements from MD&A sections to those from income tax footnotes. Panel A (B) presents the results using *TXT* (*TXPD*) as the dependent variable. In Column (1), we compare the both text-only measures with each other. In Column (2), we compare both contextualized measures of pre-tax income with each other. In Column (3), we re-run Eq. (7) using only standard pre-tax income and contextually informed measures using narrative disclosures from the MD&A section on the subsample of observations with non-missing income tax footnotes. We report the full model in Column (4), including all five measures jointly. We present t-statistics in parentheses. In a separate row, we present the *p*-value corresponding to t-tests testing whether the respective variable of interest (as highlighted through grey shading) is statistically significantly different from one. Finally, we report the *z*-statistics of Vuong's tests comparing the explanatory power of the full model that builds only on income tax footnote context with the model that builds only on MD&A context. Standard errors are clustered by firm and year. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively, based on two-tailed tests. All continuous variables are winsorized at the 1st and 99th percentiles. All variables are as defined in Appendix A.

TABLE 6 Cross-Sectional Analyses

Panel A: Tax expense; context from MD&A sections

			Adjusted R^2			Evaluation				
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			$PI + PI^{Context}$	PI	$PI^{Context}$	$U_{Context}$ (1) – (2)	$U_{Baseline}$ (1) – (3)	Common Var. (1) – (4) – (5)	$Dominance_{Context}$ (4) / ((1) – (6))	$(7)_{(A)}/(7)_{(B)}$
MNC Status	*Domestic*	(A)	0.9258	0.8810	0.9251	0.0448	0.0007	0.8803	0.9846	1.4229
	Multinational	(B)	0.9088	0.8924	0.9015	0.0164	0.0073	0.8851	0.6920	
Fluidity Indicator	*Low*	(A)	0.9218	0.9040	0.9169	0.0178	0.0049	0.8991	0.7841	0.9002
	High	(B)	0.9107	0.8830	0.9066	0.0277	0.0041	0.8789	0.8711	
Size Indicator	*Small*	(A)	0.5232	0.5206	0.5087	0.0026	0.0145	0.5061	0.1520	0.1795
	Large	(B)	0.9063	0.8803	0.9016	0.0260	0.0047	0.8756	0.8469	
ETR Indicator	*Low*	(A)	0.9186	0.8916	0.9136	0.0270	0.0050	0.8866	0.8438	5.7857
	High	(B)	0.9317	0.9296	0.9194	0.0021	0.0123	0.9173	0.1458	
Tax Avoidance	*Unfavorable*	(A)	0.928	0.9206	0.9236	0.0074	0.0044	0.9162	0.6271	0.8687
	Favorable	(B)	0.9535	0.9413	0.9488	0.0122	0.0047	0.9366	0.7219	

Panel B: Cash taxes paid; context from MD&A sections

		Adjusted R^2			Evaluation				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		$PI + PI^{Context}$	PI	$PI^{Context}$	$U_{Context}$ (1) – (2)	$U_{Baseline}$ (1) – (3)	Common Var. (1) – (4) – (5)	$Dominance_{Context}$ (4) / ((1) – (6))	$(7)_{(A)}/(7)_{(B)}$
MNC Status	*Domestic* (A)	0.8337	0.7562	0.8302	0.0775	0.0035	0.7527	0.9568	1.0714
	Multinational (B)	0.8571	0.8162	0.8522	0.0409	0.0049	0.8113	0.8930	
Fluidity Indicator	*Low* (A)	0.8678	0.8247	0.8647	0.0431	0.0031	0.8216	0.9329	1.0429
	High (B)	0.8439	0.7930	0.8379	0.0509	0.0060	0.7870	0.8946	
Size Indicator	*Small* (A)	0.3236	0.3013	0.3218	0.0223	0.0018	0.2995	0.9253	0.9965
	Large (B)	0.8418	0.7859	0.8375	0.0559	0.0043	0.7816	0.9286	
ETR Indicator	*Low* (A)	0.8614	0.7981	0.8588	0.0633	0.0026	0.7955	0.9605	1.4452
	High (B)	0.8470	0.8252	0.8360	0.0218	0.0110	0.8142	0.6646	
Tax Avoidance	*Unfavorable* (A)	0.8531	0.828	0.8459	0.0251	0.0072	0.8208	0.7771	1.2730
	Favorable (B)	0.8738	0.8539	0.8611	0.0199	0.0127	0.8412	0.6104	

Panel C: Tax expense; context from income tax footnotes

		Adjusted R^2			Evaluation				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		PI + *PI*^{Context}	*PI*	*PI*^{Context}	$U_{Context}$ (1) – (2)	$U_{Baseline}$ (1) – (3)	Common Var. (1) – (4) – (5)	$Dominance_{Context}$ (4) / ((1) – (6))	$(7)_{(A)}/(7)_{(B)}$
MNC Status	*Domestic* (A)	0.9768	0.9753	0.9006	0.0015	0.0762	0.8991	0.0193	0.1218
	Multinational (B)	0.9301	0.9214	0.8839	0.0087	0.0462	0.8752	0.1585	
Fluidity Indicator	*Low* (A)	0.9291	0.9273	0.8758	0.0018	0.0533	0.8740	0.0327	0.1405
	High (B)	0.9461	0.9338	0.9055	0.0123	0.0406	0.8932	0.2325	
Size Indicator	*Small* (A)	0.7233	0.7235	0.6852	-0.0002	0.0381	0.6854	- .0053	-0.0420
	Large (B)	0.9283	0.9211	0.8782	0.0072	0.0501	0.8710	0.1257	
ETR Indicator	*Low* (A)	0.9561	0.9373	0.9220	0.0188	0.0341	0.9032	0.3554	12.9463
	High (B)	0.9625	0.9597	0.8633	0.0028	0.0992	0.8605	0.0275	
Tax Avoidance	*Unfavorable* (A)	0.9405	0.9336	0.8258	0.0069	0.1147	0.8189	0.0567	0.3151
	Favorable (B)	0.9450	0.9356	0.9022	0.0094	0.0428	0.8928	0.1801	

Panel D: Cash taxes paid; context from income tax footnotes

			Adjusted R^2			Evaluation				
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			PI + *PI*^{Context}	*PI*	*PI*^{Context}	$U_{Context}$ (1) – (2)	$U_{Baseline}$ (1) – (3)	Common Var. (1) – (4) – (5)	$Dominance_{Context}$ (4) / ((1) – (6))	$(7)_{(A)}/(7)_{(B)}$
MNC Status	*Domestic*	(A)	0.8988	0.8975	0.8149	0.0013	0.0839	0.8136	0.0153	0.0880
	Multinational	(B)	0.8851	0.8742	0.8331	0.0109	0.0520	0.8222	0.1733	
Fluidity Indicator	*Low*	(A)	0.8811	0.8738	0.8288	0.0073	0.0523	0.8215	0.1225	0.8549
	High	(B)	0.8923	0.8823	0.8325	0.0100	0.0598	0.8225	0.1433	
Size Indicator	*Small*	(A)	0.5781	0.5752	0.5216	0.0029	0.0565	0.5187	0.0488	0.3325
	Large	(B)	0.8737	0.8637	0.8156	0.0100	0.0581	0.8056	0.1468	
ETR Indicator	*Low*	(A)	0.9013	0.8886	0.8498	0.0127	0.0515	0.8371	0.1978	3.7332
	High	(B)	0.8832	0.8793	0.8135	0.0039	0.0697	0.8096	0.0530	
Tax Avoidance	*Unfavorable*	(A)	0.8681	0.8662	0.8076	0.0019	0.0605	0.8057	0.0304	0.2220
	Favorable	(B)	0.8926	0.8833	0.8341	0.0093	0.0585	0.8248	0.1372	

Notes: This table presents the adjusted R^2 plus additional calculations for several linear OLS regressions. Panel A (B) presents the results using *TXT* (*TXPD*) as the dependent variable and narrative disclosures from MD&As. In Panel C (D). we report the results for tests using *TXT* (*TXPD*) as the dependent variable and narrative disclosures from income tax footnotes. We split the main sample cross-sectionally into two groups using four categories (MNC status. level of product market fluidity. size, and level of ETR), respectively, and repeat three regressions: One using only *PI* as the independent variable, one using only *PI*^{Context} as the independent variable, and one using both. The dependent variable in each regression is *TXT*. We evaluate the respective models' performances in Columns (4) through (8) using commonality analysis.

TABLE 7 Analyst ETR Forecasts

Panel A: Summary Statistics

Dependent variable	N	Mean	SD	P25	Median	P75
ETR Forecast Error$_{t+1}$	14,244	0.045	0.066	0.007	0.019	0.051
Adjustment	14,244	0.215	0.306	0.078	0.159	0.265
ETR	14,244	0.309	0.214	0.210	0.306	0.374
Wedge	14,244	0.088	0.149	0.019	0.046	0.092
Adjustment (Footnote)	5,201	0.216	0.178	0.090	0.186	0.298
Wedge (Footnote)	5,201	0.094	0.152	0.018	0.044	0.102
Size	14,244	7.660	1.788	6.382	7.584	8.862
ROA	14,244	0.122	0.110	0.051	0.094	0.157
Leverage	14,244	0.303	0.278	0.072	0.258	0.440
Market-to-Book	14,244	3.684	5.164	1.609	2.607	4.381
Intangible	14,244	0.266	0.257	0.040	0.192	0.435
Accruals Quality	14,244	0.854	2.242	0.044	0.147	0.595
MNC Status	14,244	0.569	0.495	0	1	1
Big N Auditor	14,244	0.891	0.312	1	1	1
R&D	14,244	0.025	0.049	0.000	0.000	0.026
Age	14,244	2.784	0.769	2.197	2.708	3.332
Fog	14,244	23.187	1.687	22.012	23.135	24.296
Length	14,244	9.343	0.452	9.068	9.348	9.627
Negativity	14,244	0.011	0.004	0.008	0.010	0.013
Specificity	14,244	0.066	0.014	0.056	0.065	0.075
Boilerplate	14,244	0.063	0.056	0.023	0.040	0.085
Analyst Following	14,244	6.906	5.460	3	5	10

Panel B: Regression Across Firms

Var. of interest build on:	*MD&A*		*Tax Footnotes*	
Dependent variable	*ETR Forecast Error$_{t+1}$*	*ETR Forecast Error$_{t+1}$*	*ETR Forecast Error$_{t+1}$*	*ETR Forecast Error$_{t+1}$*
	(1)	(2)	(3)	(4)
Adjustment	**0.0036***		**0.0121***	
	(1.666)		**(1.705)**	
Wedge		**0.0148*****		**0.0120**
		(3.035)		**(1.419)**
ETR		**-0.0239*****		**-0.0089**
		(-4.362)		**(-1.268)**
Size	-0.0005	-0.0010	-0.0017	-0.0018
	(-0.739)	(-1.407)	(-1.433)	(-1.545)
ROA	-0.0777***	-0.0809***	-0.0588***	-0.0610***
	(-11.498)	(-12.016)	(-4.150)	(-4.310)
Leverage	0.0122***	0.0116***	0.0092**	0.0092**
	(4.144)	(3.959)	(2.034)	(2.039)
Market-to-Book	-0.0000	-0.0000	0.0001	0.0001
	(-0.235)	(-0.213)	(0.395)	(0.401)
Intangible	-0.0123***	-0.0111***	-0.0165***	-0.0165***
	(-3.820)	(-3.460)	(-3.400)	(-3.407)
Accruals Quality	0.0010***	0.0010***	0.0013**	0.0013**
	(2.933)	(2.955)	(2.180)	(2.189)
MNC Status	0.0019	0.0020	0.0014	0.0015
	(0.986)	(1.059)	(0.451)	(0.504)
Big N Auditor	-0.0052**	-0.0047*	0.0011	0.0010
	(-2.042)	(-1.861)	(0.287)	(0.255)
R&D	0.0627***	0.0577***	0.0245	0.0244
	(3.536)	(3.270)	(0.914)	(0.907)
Age	-0.0057***	-0.0055***	-0.0069***	-0.0069***
	(-5.142)	(-4.996)	(-4.030)	(-4.062)
Fog	0.0001	0.0000	-0.0010	-0.0009
	(0.192)	(0.080)	(-1.164)	(-1.109)
Length	0.0050**	0.0049**	0.0068**	0.0068**
	(2.375)	(2.344)	(2.132)	(2.132)
Negativity	0.5420***	0.5394***	0.4717	0.4752
	(2.734)	(2.734)	(1.369)	(1.386)
Specificity	0.0275	0.0246	0.0054	0.0088
	(0.478)	(0.429)	(0.062)	(0.100)
Boilerplate	0.0360***	0.0355***	0.0374**	0.0372**
	(2.669)	(2.649)	(2.086)	(2.073)
Analyst Following	-0.0017***	-0.0017***	-0.0016***	-0.0016***
	(-10.899)	(-10.817)	(-6.570)	(-6.500)
Clustered SE	Firm	Firm	Firm	Firm
Industry FE	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.0836	0.0854	0.1072	0.1066
N	14,816	14,816	5,386	5,386

Panel C: Regression Within Firms

Var. of interest build on:	*MD&A*		*Tax Footnotes*	
Dependent variable	*ETR Forecast Error$_{t+1}$*	*ETR Forecast Error$_{t+1}$*	*ETR Forecast Error$_{t+1}$*	*ETR Forecast Error$_{t+1}$*
	(1)	(2)	(3)	(4)
Adjustment	**-0.0069****		**-0.0026**	
	(-2.176)		**(-0.335)**	
Wedge		**-0.0090***		**-0.0079**
		(-1.650)		**(-0.872)**
ETR		**-0.0044**		**0.0006**
		(-0.745)		**(0.087)**
Size	0.0056***	0.0057***	0.0095**	0.0095**
	(3.023)	(3.070)	(2.425)	(2.433)
ROA	-0.0811***	-0.0807***	-0.0656***	-0.0653***
	(-9.368)	(-9.287)	(-3.580)	(-3.548)
Leverage	0.0059	0.0060	-0.0026	-0.0025
	(1.333)	(1.360)	(-0.371)	(-0.358)
Market-to-Book	-0.0001	-0.0001	-0.0001	-0.0001
	(-0.949)	(-0.936)	(-0.778)	(-0.797)
Intangible	-0.0129**	-0.0131**	-0.0144	-0.0145
	(-2.512)	(-2.557)	(-1.494)	(-1.504)
Accruals Quality	0.0005	0.0005	0.0014***	0.0014***
	(1.645)	(1.605)	(2.638)	(2.653)
MNC Status	0.0023	0.0024	0.0028	0.0027
	(0.732)	(0.762)	(0.433)	(0.425)
Big N Auditor	0.0020	0.0020	-0.0019	-0.0019
	(0.427)	(0.414)	(-0.168)	(-0.168)
R&D	0.0136	0.0138	-0.0815	-0.0832
	(0.339)	(0.344)	(-1.286)	(-1.313)
Age	-0.0122***	-0.0122***	-0.0308***	-0.0310***
	(-3.108)	(-3.099)	(-3.361)	(-3.394)
Fog	-0.0019**	-0.0019**	-0.0015	-0.0015
	(-2.383)	(-2.359)	(-1.131)	(-1.152)
Length	0.0022	0.0023	-0.0027	-0.0026
	(0.841)	(0.843)	(-0.532)	(-0.513)
Negativity	0.4067*	0.4123*	-0.0573	-0.0555
	(1.656)	(1.678)	(-0.113)	(-0.110)
Specificity	0.1714*	0.1712*	0.0883	0.0859
	(1.911)	(1.908)	(0.525)	(0.513)
Boilerplate	-0.0094	-0.0094	-0.0559	-0.0554
	(-0.401)	(-0.400)	(-1.429)	(-1.417)
Analyst Following	-0.0019***	-0.0019***	-0.0019***	-0.0019***
	(-8.166)	(-8.160)	(-4.682)	(-4.618)
Clustered SE	Firm	Firm	Firm	Firm
Industry FE	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.2477	0.2474	0.3199	0.3198
N	14,244	14,244	5,049	5,049