Ambiguity vs. Risk: Decomposing Uncertainty in Asset Pricing

Yingzhen Jiang, *University of Macau*

Charles Kang, *University of Hong Kong*

January 2026

Abstract

This paper decomposes uncertainty in financial markets into its two fundamental dimensions—ambiguity (epistemic uncertainty) and risk (aleatory uncertainty)—and examines their distinct effects on asset pricing. Using a deep learning approach to analyze corporate disclosures, we find that epistemic uncertainty is associated with negative contemporaneous stock returns, consistent with ambiguity aversion, while aleatory uncertainty is positively linked to future realized returns, reflecting compensation for measurable risk. We further show that idiosyncratic risk, a component of aleatory uncertainty, is positively priced in the cross-section of returns, and that ambiguity differentially affects the pricing of idiosyncratic and systemic risk. These findings underscore the dual role of uncertainty in asset pricing and provide new insights into investor behavior and the pricing of ambiguity and risk.

JEL: G12, G14, G32, G41, D81, D84

Keywords: Uncertainty, risk, ambiguity aversion, idiosyncratic volatility puzzle

1. Introduction

Uncertainty is a fundamental feature of financial markets, shaping investor behavior, corporate decision-making, and asset pricing. Yet, despite its pervasive influence, the literature often treats uncertainty as a monolithic concept, overlooking its distinct dimensions. This paper addresses this gap by decomposing uncertainty into its two fundamental components: ambiguity (epistemic uncertainty) and risk (aleatory uncertainty). Drawing on foundational work in probability theory (Savage, 1954; Hacking, 1975) and Knight's (1921) distinction between ambiguity and risk, we provide a refined framework for understanding how these distinct dimensions influence financial markets.

The distinction between these dimensions has deep roots in classical philosophy and probability theory. Aristotle, in his *Physics* and *Nicomachean Ethics*, distinguishes between chance (*tyche*), which "happens unpredictably and without a deliberate cause," and ignorance (*agnoia*), "a lack of knowledge, whether from not having learned or from not being able to understand." This philosophical groundwork laid the foundation for later developments in probability theory. Knight (1921) formalizes the distinction by differentiating between measurable risk, which can be quantified due to known probabilities, and ambiguity, which arises when probabilities are "unknown, indeterminate, or not capable of being calculated." Hacking (1975) further refines this by differentiating uncertainty stemming from randomness (aleatory) from that arising from incomplete knowledge (epistemic). These insights highlight that uncertainty is multifaceted, requiring disaggregation to fully understand its effects on financial markets.

Building on this foundation, we define epistemic uncertainty as arising from incomplete information, subjective judgments, or a lack of knowledge, reflecting situations where the underlying probabilities of outcomes are unknown or difficult to assess (Fox and Ülkümen, 2011). This aligns with Knight's (1921) concept of ambiguity, where uncertainty stems from

the inability to assign precise probabilities to future events. In contrast, aleatory uncertainty captures inherent randomness or variability in outcomes, where probabilities are well-defined and can be estimated with confidence.

Despite the theoretical importance of this distinction, the finance literature lacks robust empirical measures to isolate the effects of epistemic and aleatory uncertainty. Traditional proxies, such as analyst forecast dispersion, stock volatility, and the VIX, conflate ambiguity with measurable risk, obscuring their unique influences on investor behavior and asset pricing. While theoretical models of ambiguity aversion (e.g., Ellsberg, 1961; Gilboa and Schmeidler, 1989) provide a conceptual framework, their empirical application remains limited, as existing measures fail to disentangle the two dimensions. This underscores the need for more precise tools to analyze how different types of uncertainty shape financial markets.

Corporate disclosures offer a rich source of information about uncertainty. However, traditional methods, such as dictionary-based approaches, struggle to differentiate meaningful uncertainty from boilerplate language. To address this, we employ a convolutional neural network (CNN), a state-of-the-art deep learning model, to identify and classify uncertainty statements in the Management Discussion and Analysis (MD&A) section of 10-K filings. Our CNN is trained on a manually annotated dataset of approximately 10,000 sentences. The model achieves high accuracy, significantly outperforming traditional methods and simpler machine learning algorithms.

Using the CNN model, we decompose uncertainty-related sentences in the MD&A sections into meaningful uncertainty statements, boilerplate statements, and further into aleatory and epistemic dimensions. While 28% of MD&A sentences contain uncertainty triggers identified by the Loughran and McDonald (2011) dictionary, only 28.6% of these are classified as meaningful uncertainty statements by the CNN. Among these, two-thirds pertain to aleatory uncertainty and one-third to epistemic uncertainty. Our final sample consists of

56,480 firm-year observations from 1994 to 2024, constructed from 10-K filings, Compustat, CRSP, and I/B/E/S data.

Epstein and Schneider (2008, 2010) provide a theoretical framework for understanding how investors respond to ambiguity. Their models predict that when ambiguity-averse investors face ambiguity, they price assets by assuming the worst-case interpretation of signals. This behavior leads to asymmetric market reactions: bad news is treated as highly reliable, causing sharp price drops, while good news is discounted as noisy, resulting in muted price increases. This asymmetry generates a persistent discount in asset prices, known as the ambiguity premium, as investors demand compensation for enduring ambiguity. Building on this framework, we predict that epistemic uncertainty triggers ambiguity aversion, leading investors to discount assets with ambiguous prospects. In contrast, aleatory uncertainty, which captures measurable risk, is expected to command a risk premium, consistent with classical asset pricing theory (Sharpe, 1964).

We test these predictions by examining the impact of uncertainty on contemporaneous and future realized returns. We find that epistemic uncertainty exhibits a negative and statistically significant effect on contemporaneous returns, consistent with ambiguity aversion. This suggests that investors discount assets when faced with incomplete information or subjective uncertainty. In contrast, aleatory uncertainty, which reflects inherent randomness, elicits no significant market reaction.

For future realized returns, uncertainty becomes highly significant in predicting one-quarter-ahead stock returns. Decomposing uncertainty into its components reveals that aleatory uncertainty is a significant positive predictor, supporting the idea that aleatory uncertainty conveys quantifiable risk for which investors demand compensation. In contrast, epistemic uncertainty does not systematically affect future realized returns. This decomposition shows

that the predictive power of uncertainty on future realized returns is primarily driven by aleatory uncertainty, which accounts for two-thirds of CNN-identified uncertainty.

To validate our findings, we conduct portfolio analyses by sorting stocks into quintiles based on aleatory and epistemic uncertainty. For three-day returns, epistemic uncertainty yields significant negative hedge portfolio returns of –0.42% to –0.48% across risk-factor adjustments, whereas aleatory uncertainty exhibits no significant contemporaneous effect. This evidence is consistent with ambiguity aversion, suggesting that investors discount assets when uncertainty stems from limited knowledge or information. In contrast, for one-quarter-ahead returns, aleatory uncertainty produces economically meaningful hedge portfolio returns of 0.86% to 1.08% across risk-factor adjustments—annualizing to between 3.48% and 4.37%—with returns increasing monotonically across quintiles. These results support a risk premium interpretation of aleatory uncertainty, indicating that investors require higher expected returns as compensation for bearing such risk.

The Fama-MacBeth regressions further reinforce these findings. Epistemic uncertainty has a significant negative effect on three-day contemporaneous returns, consistent with ambiguity aversion. In contrast, aleatory uncertainty has no significant contemporaneous impact but exhibits a positive and significant effect on one-quarter-ahead returns, in line with a risk premium interpretation. Together, these results highlight the distinct roles of epistemic and aleatory uncertainty: epistemic uncertainty drives immediate price discounts due to ambiguity aversion, while aleatory uncertainty commands a risk premium.

Furthermore, we test how ambiguity and measurable risk interact in shaping asset prices. Drawing on the models of Epstein and Schneider (2008, 2010), Epstein and Wang (1994), and Illeditsch (2011), we focus on two channels: (i) the ambiguity-premium channel, in which the price discount from ambiguity is amplified when risk is high; and (ii) the portfolio inertia channel, in which elevated ambiguity weakens the risk premium associated with measurable

risk. We implement these tests by partitioning the sample along one dimension of uncertainty and re-estimating return regressions for the other. Consistent with the ambiguity-premium channel, the negative contemporaneous effect of ambiguity is concentrated in the high-aleatory risk subsample. Consistent with the portfolio inertia channel, the positive future realized returns to aleatory risk are stronger when ambiguity is low. These interaction results show that the asset-pricing impact of uncertainty depends not only on its type, but also on the joint distribution of its ambiguity and risk components.

We also decompose aleatory uncertainty into systemic and idiosyncratic components to address the longstanding "idiosyncratic risk puzzle." Classical asset pricing theory, rooted in the Capital Asset Pricing Model (CAPM), posits that only systematic risk should be priced, as idiosyncratic risk can be diversified away in a well-constructed portfolio. However, empirical evidence on the relationship between idiosyncratic risk and expected returns has been mixed. Early studies, such as those by Ang et al. (2006, 2009), document a negative relationship between idiosyncratic volatility and future returns, suggesting that stocks with higher idiosyncratic risk tend to underperform. In contrast, other studies, such as Fu (2009) and Bégin et al. (2019), find a positive relationship, attributing the negative results to measurement issues or focusing on tail risk. Our decomposition of aleatory uncertainty reveals that both its systemic and idiosyncratic components are positively and significantly related to future realized returns. This suggests the aggregate positive risk premium for aleatory risk compensates for both systematic and idiosyncratic exposures, consistent with Fu (2009) and Bégin et al. (2019).

Building on these findings, we further examine how ambiguity influences the pricing of systemic and idiosyncratic risk. We find that ambiguity fundamentally conditions the premiums investors require for bearing these risks. In low-ambiguity environments, where firm-specific risk can be assessed with confidence, idiosyncratic risk commands a significant and robust positive premium across all horizons, while the premium for systemic risk is weaker

and reliably significant only over longer horizons. This relationship reverses markedly in high-ambiguity environments: the idiosyncratic risk premium largely disappears as ambiguity-averse investors diversify away or avoid such exposures, while a greater premium for systemic risk emerges at short and intermediate horizons, compensating investors for bearing ambiguous and non-diversifiable exposures. This conditional interaction framework offers more insight into the conflicting evidence surrounding the idiosyncratic risk puzzle and clarifies the circumstances under which both idiosyncratic and systemic risk are priced.

This paper makes several key contributions to the literature on uncertainty, investor behavior, and asset pricing. First, we introduce a novel framework for decomposing uncertainty into its two fundamental dimensions: epistemic uncertainty (ambiguity) and aleatory uncertainty (risk). While prior research has often treated uncertainty as a monolithic concept, our approach builds on foundational work in probability theory (Savage, 1954; Hacking, 1975) and Knight's (1921) distinction between risk and ambiguity to provide a more nuanced understanding of how different types of uncertainty influence financial markets. Our central contribution is to clarify the conceptual relationship between risk and uncertainty, demonstrating that risk *derives* from uncertainty, rather than an independent construct arising from a different source. Ambiguity, often equated with "uncertainty" in empirical work, is likewise only one component of overall uncertainty, alongside risk. By operationalizing these dimensions using a deep learning model applied to corporate disclosures, we offer a robust methodology for isolating and measuring different dimensions of uncertainty in financial texts.

Second, we provide empirical evidence that these two dimensions of uncertainty elicit distinct market reactions. Epistemic uncertainty, which reflects ambiguity about the precision or reliability of information, is associated with negative contemporaneous stock returns, consistent with theories of ambiguity aversion (Ellsberg, 1961; Gilboa and Schmeidler, 1989). In contrast, aleatory uncertainty, which captures measurable risk, is positively associated with

future realized returns, as investors demand compensation for bearing measurable risk. These findings underscore the importance of distinguishing between ambiguity and risk in understanding investor behavior and asset pricing, as they convey fundamentally different types of uncertainty and elicit qualitatively different market responses. Our findings also highlight the need for careful interpretation of commonly used uncertainty proxies in finance and accounting research, such as analyst forecast dispersion, the VIX, or dictionary-based uncertainty scores, which do not fully disentangle these distinct dimensions. Researchers and practitioners should consider the specific nature of uncertainty when employing these measures to avoid conflating risk and ambiguity.

Our study also addresses the longstanding "idiosyncratic risk puzzle." Classical asset pricing theory posits that only systematic risk should be priced, while idiosyncratic risk, being diversifiable, should not command a premium. However, empirical evidence has been mixed: Ang et al. (2006, 2009) find a negative relationship between idiosyncratic volatility and expected returns, while Fu (2009) and Goyal and Santa-Clara (2003) document a positive relationship when using alternative volatility measures. By decomposing aleatory uncertainty into its systematic and idiosyncratic components, we provide a unifying perspective on these conflicting findings. We show that the risk premium associated with aleatory uncertainty persists even after controlling for standard risk factors and that compensation for idiosyncratic risk is particularly strong in low-ambiguity environments, when firm-specific risk can be assessed with greater precision (Merton, 1987; Bégin et al., 2019). In contrast, in high-ambiguity environments, the idiosyncratic risk premium largely disappears as ambiguity-averse investors diversify away from or avoid such exposures, while the premium for systemic risk becomes significantly stronger at short and intermediate horizons, compensating investors for bearing ambiguous and non-diversifiable macro risk. This conditional pricing framework clarifies the circumstances under which idiosyncratic and systemic risks are priced and

demonstrates that ambiguity is a critical determinant of how both types of risk are reflected in asset prices.

Finally, we contribute to the growing literature on textual analysis in finance by demonstrating the value of deep learning techniques for analyzing corporate disclosures. Our convolutional neural network (CNN) model outperforms traditional methods in identifying and decomposing uncertainty, providing a more precise framework for understanding its role in financial markets. This approach opens new avenues for research on the role of textual information in asset pricing. By leveraging CNNs, we introduce a significant methodological innovation that enhances classification accuracy and enables the nuanced measurement of complex constructs such as uncertainty, risk, and ambiguity.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature and develops our hypotheses. Section 3 describes our research design. Section 4 presents our empirical results. Section 5 concludes with a discussion of the implications of our findings for future research.

2. Literature Review and Hypotheses

Uncertainty is a central concept in finance and accounting, yet its interpretation often conflates distinct dimensions that may have qualitatively different effects on investor behavior and asset pricing. While prior research has treated uncertainty as a single dimension, the literature on probability theory highlights its dual nature: aleatory uncertainty (related to randomness or chance) and epistemic uncertainty (related to knowledge or lack thereof) (Savage, 1954; Hacking, 1975). Building on Knight's (1921) seminal work on ambiguity, we decompose uncertainty into these two distinct components: epistemic uncertainty (ambiguity) and aleatory uncertainty (risk). This decomposition is critical for understanding how investors perceive and price uncertainty, as these components elicit different behavioral and economic responses.

Epistemic uncertainty arises from a lack of knowledge or ambiguity about the probability distribution of outcomes. It reflects situations where information is incomplete, imprecise, or subject to interpretation, making it difficult for investors to assign precise probabilities to future events (Fox and Ülkümen, 2011). This conceptualization aligns closely with Knight's (1921) notion of ambiguity, which distinguishes it from measurable risk. In contrast, aleatory uncertainty reflects inherent variability or randomness in outcomes, where probabilities are well-defined and can be estimated with confidence. It stems from probabilistic processes rather than informational limitations. Recent experimental studies provide robust evidence for this distinction (Tannenbaum et al., 2017). For example, Walters et al. (2023) show that investors facing greater epistemic uncertainty are more likely to seek expert guidance, while those confronting higher aleatory uncertainty prioritize diversification strategies and risk preferences.

The distinction between epistemic and aleatory uncertainty is particularly relevant in the context of ambiguity aversion—the preference for known risks over unknown uncertainties (Ellsberg, 1961; Gilboa and Schmeidler, 1989). Theoretical models of portfolio choice suggest that ambiguity aversion can lead to nonparticipation in equity markets (Dow and Werlang, 1992; Garlappi et al., 2007; Easley and O'Hara, 2009; Bossaerts et al., 2010; Epstein and Schneider, 2010; Cao et al., 2011). Empirical evidence supports these predictions: Antoniou et al. (2015) and Dimmock et al. (2016) find that ambiguity aversion reduces stock market participation, while Guidolin and Rinaldi (2013) and Peijnenburg (2018) argue that it explains phenomena such as the home equity bias and the equity premium puzzle.

Theoretical asset pricing models further highlight the distinct effects of ambiguity aversion on market outcomes (Epstein and Schneider, 2007, 2008; Caskey, 2009; Ilut and Schneider, 2014; Dicks and Fulghieri, 2019, 2021; Kostopoulos, Meyer, and Uhr, 2022). For example, Wang (2005) shows that uncertainty-averse investors may hold portfolios that deviate from mean-variance efficiency, challenging traditional asset pricing models. Similarly, Epstein

and Schneider (2008) argue that ambiguity-averse investors, uncertain about the quality of information they receive, price assets by assuming the worst-case interpretation of signals. This leads to asymmetric reactions: bad news is treated as highly reliable, causing sharp price drops, while good news is discounted as noisy, resulting in muted price increases. This asymmetry generates negative skewness in returns and introduces a persistent discount in asset prices—an ambiguity premium—that compensate investors for enduring ambiguity (Epstein and Schneider, 2010). Crucially, the model highlights anticipatory effects: prices drop preemptively when investors expect future ambiguous signals, even if fundamentals remain unchanged. In a related vein, Williams (2015) finds that shocks to macro-uncertainty lead investors to respond asymmetrically to firm-specific earnings news, placing greater weight on bad news than on good news.

Despite these advances, prior literature often equates uncertainty with Knightian ambiguity, and existing measures, such as dispersion in analyst forecasts and the VIX, fail to reliably isolate ambiguity from aleatory uncertainty. For example, analyst forecast dispersion reflects disagreement among analysts, but this disagreement can arise from differences in ability, expertise, and access to information, rather than ambiguity per se. Similarly, the VIX captures uncertainty related to macroeconomic fundamentals, making it difficult to isolate the effects of ambiguity (Bekaert, Hoerova, and Duca, 2013). This conflation limits the ability to examine the distinct roles of ambiguity and aleatory uncertainty in financial markets. Our study bridges this gap by operationalizing epistemic uncertainty as a proxy for ambiguity and explicitly separating it from aleatory uncertainty, providing a clearer framework to examine their effects.

In the context of corporate disclosures, uncertainty is often embedded in qualitative narratives, such as the Management Discussion and Analysis (MD&A) section of financial reports. Epistemic uncertainty in these disclosures reflects managerial ambiguity stemming

from incomplete or imprecise information, which can trigger ambiguity aversion and result in negative market reactions. In contrast, aleatory uncertainty, which reflects measurable risk, is likely to be priced into expected returns as compensation for risk (Fama and French, 1996; Anderson, Ghysels, and Juergens, 2009; Andrei and Hasler, 2015; Brenner and Izhakian, 2018).

Building on this framework, we hypothesize that epistemic uncertainty, capturing ambiguity, elicits negative market reactions due to ambiguity aversion, while aleatory uncertainty, capturing risk, is positively associated with future realized returns as investors price in risk premiums. Formally stated:

H1a: Epistemic uncertainty is associated with contemporaneous negative market reactions due to ambiguity aversion.

H1b: Aleatory uncertainty is associated with higher future realized returns.

3. Research Design

3.1. Identifying Uncertainty in MD&A

We employ the convolutional neural network (CNN), a state-of-the-art deep learning model, to identify uncertainty statements in financial text. Over the past decade, deep learning techniques have achieved remarkable progress, with recent research demonstrating their superiority over traditional methods such as domain-specific dictionaries (e.g., Loughran and McDonald, 2011) and simpler machine learning algorithms (e.g., Naïve Bayes) in analyzing financial disclosures (Azimi and Agrawal, 2021; Huang et al., 2022; Brown et al., 2024). Among the most widely adopted deep learning architectures, CNNs—originally developed for image processing—have proven highly effective in sentence classification tasks, including applications in financial contexts (Kim, 2014; Zhang and Wallace, 2015).[1]

[1] Although the Google BERT model can accommodate tasks such as natural language processing, it is pre-trained on general text and may not perform optimically for financial texts, which can differ significantly in vocabulary and writing style (Huang et al., 2022).

The application of CNN models to sentence classification mirrors their use in computer vision, with minor adaptations. The core principle of CNN involves dimensionality reduction: condensing all possible information in a sentence into features and associating these features with sentence classification. While the width and height of images can be fixed, the length of sentences varies significantly. To address this, we follow Kim (2014) and Zhang and Wallace (2015) by using *padding* to standardize the sentence length (i.e., a customized sentence length of $n = 80$).[2] Each word is represented by a word vector, and the dimensionality of the word vectors is d (where d represents the size of the vocabulary). The entire sentence is then transformed into a matrix, which can be treated as an 'image' by the CNN, and the dimensionality of the sentence matrix is $n \times d$.

Next, we condense the information in a sentence into features by performing convolution on the sentence matrix via *filters*, which are the sub-matrixes that slide over the entire sentence matrix. In the text application of CNN, we need to consider the inherent sequential structure of the words. We employ *filters* with widths equal to the dimensionality of the word vectors d and heights equal to 3, 4, and 5, which is the number of adjacent rows in the sentence matrix. To determine the appropriate number of *filters* for optimal computational performance, we balance prediction accuracy and computational cost. Following Huang et al. (2022), we conduct a grid search over the hyperparameter set [32, 64, 128, 256, 512] and identify 128 as the optimal choice for the number of *filters*.[3] The input sentence matrix is then convolved with these *filters* to produce multiple feature maps, each corresponding to a specific *filter*. These feature maps

[2] We set the customized sentence length to $n = 80$ because the mean length of the sentences in our training sample is 80 words. For sentences longer than 80 words, we select the first 80 words. For sentences shorter than 80 words, we append append 'zero' tokens to the end to standardize their length. Initially, we set the customized sentence length to the maximum value in the hope of increasing the capacity to capture all the information in a sentence. Nonetheless, we acknowledge that CNN becomes inefficient in classifying short sentences when the customized sentence length is too large, due to the excessive padding with zero tokens. We also note that our results are robust to using alternative parameter settings for the CNN models.

[3] Increasing the number of feature maps can help the machine learn more complex and abstract features, but it also increases the computational cost.

capture linguistic patterns in the input sentence, such as n-grams, parts of speech, and syntactic structures.[4]

Since the dimensionality of the feature maps varies due to the different heights of the *filters* (i.e., 3, 4, and 5), we apply *1-max pooling* (Boureau et al., 2010) to each feature map to induce a scalar. For the 128 feature maps, this process generates 128 scalars, which are later concatenated to form a single feature vector for the penultimate layer. Finally, we map this feature vector to the corresponding sentence classification by passing it through a *softmax function*. Figure 1, adapted from Zhang and Wallace (2015), provides a simplified illustration of these procedures in the CNN models.

[Insert Figure 1]

Before training the model to recognize uncertainty statements, we construct a training sample of approximately 10,000 sentences. To isolate sentences most likely to convey uncertainty, we apply a keyword-based screening process to sentences extracted from the MD&A sections of 10-K filings.[5] Our keyword selection begins with terms classified under the "uncertain" category in Loughran and McDonald (2011). We supplement this list with additional uncertainty-related keywords not included in their dictionary, such as "potential*, susceptible, influenc*, whether, necessarily, primarily" (where * denotes the inclusion of suffixes). These supplementary keywords are drawn from prior literature (Kravet and Muslu, 2013) or identified through our manual review of relevant texts.

We first extract the Management Discussion and Analysis (MD&A) sections from 10-K filings between 1995 and 2024. After identifying the set of MD&A sentences that include "uncertain" triggers (hence most likely to be uncertainty statements), we generate a training sample by randomly selecting sentences from this full set. To account for variations in

[4] The feature maps for image recognition represent edges, lines, corners, shapes, textures, and other visual patterns.
[5] This screening process is necessary to increase the efficiency of our method, by excluding sentences that do not pertain to uncertainty.

uncertainty descriptions across years and firms of different sizes, we divide all firms in a given year into quintiles based on their total assets and randomly select one firm from each quintile. This process results in 150 randomly selected firms over the 30-year period. Additionally, because firms in different industries face industry-specific uncertainties in their operations and business activities, we randomly select one firm from each two-digit SIC industry. We then extract sentences containing uncertain triggers from the MD&A sections of these randomly selected firm-years, yielding a training sample of over 10,000 sentences.

Next, we manually classify each sentence in the training sample into three categories: (i) boilerplate statements unrelated to uncertainty,[6] (ii) statements related to aleatory uncertainty, and (iii) statements related to epistemic uncertainty. To ensure consistency and accuracy in classification, we identify specific linguistic patterns that distinguish aleatory and epistemic uncertainty.[7] Appendix A provides illustrative examples of these patterns, which guide our manual classification process. Broadly, aleatory uncertainty encompasses external, distributional factors that are irreducible and and relate to variability or unpredictability in outcomes. Examples include: "The prices for certain commodities have fluctuated (cue) in recent years due to changes in global supply and demand" and "Business cycles and other changes in demand could (cue) adversely impact our revenue and profits." In contrast, epistemic uncertainty generally corresponds to internal uncertainty, which stems from a lack of knowledge, incomplete information, or subjective judgments. This type of uncertainty can, in principle, be reduced through targeted efforts, such as acquiring additional or higher-quality

[6] This is because the Loughran and McDonald (2011) uncertainty dictionary includes words such as "may," "might," "could," and "can," which often function as meaningless modals or boilerplate disclosures. Additionally, the patterns of "approximately/about + #number" are considered boilerplate because only the approximate value of the numerical information is mentioned. Furthermore, certain accounting terminologies—such as variable costs, credit exposure, doubtful accounts, intangible assets, and risk adjustments in accounting practices— can be boilerplate when they do not refer to the firm's future operations.

[7] During this process, we also utilized ChatGPT to enhance our understanding of certain sentences. Specifically, we first fed the definitions and some definitive sample sentences of boilerplate, aleatory and epistemic uncertainty into ChatGPT. Once ChatGPT had developed a similar understanding of the topic, we used it to label sentences for which our human coders were uncertain about the appropriate category. Our team also discussed the complex sentences and consistently reinforced our understanding of the concepts throughout the labelling process.

information, consulting external experts, or implementing structured decision-making frameworks. A typical example of epistemic uncertainty is: "Management believes (cue) that any losses resulting from such proceedings would not have a material adverse effect on the corporation's results of operations or financial position." After manual classification, we input the labeled sentences into the CNN and train the models to recognize these patterns out of the sample.

We validate the CNN models using an evaluation sample of an additional 1,000 sentences (randomly selected) by comparing machine and manual classifications and calculating model performance metrics. Table 1 reports the classification performance of the CNN in the validation sample. In Panel A, we present the confusion matrix, which shows that the CNN is highly unlikely to misclassify epistemic sentences as aleatory, or vice versa. In panel B, we report several performance metrics for each uncertainty category as labeled by the researchers. These include precision (the number of sentences correctly classified into a category relative to the total number of sentences classified into that category), recall (the number of sentences correctly classified into a category relative to the number of sentences researchers label as that category), accuracy (the number of sentences correctly classified relative to the total number of sentences), and F1 score (the harmonic mean of precision and recall). The results demonstrate that the CNN model achieves desirable performance in classifying uncertainty categories.[8] After validation, we use the model to identify all the sentences in our full sample that exhibit patterns of uncertainty statements.

[Insert Table 1]

3.2. Sample Selection

[8] Previous studies have used the CNN method for sentiment classification (Huang et al., 2022) and forward-looking statement recognition (Brown et al., 2024). Our overall model performance is comparable to and slightly better than the results reported in these studies. For instance, our mean recall rate of 86.1% exceeds the 71.4% reported by Huang et al. (2022) for sentiment classification and the 75.3% reported by Brown et al. (2024) for forward-looking statement recognition. We also find better model performance when broken down by categories.

We obtain 10-K reports filed with the SEC from EDGAR, financial variables from Compustat, stock prices from CRSP, and analyst forecasts from I/B/E/S. The time period spans from 1995 through 2024. Following D'Augusta and DeAngelis (2020), we remove observations with financial statement data in non-U.S. currencies and those with stock prices below $1. We also exclude observations for which we cannot obtain the MD&A section or where the MD&A word count is fewer than 100 words. Additionally, we remove observations missing the necessary data items required to calculate the variables. This procedure yields a final sample of 56,480 firm-year observations (covering from years 1995 through 2024) for empirical testing. All continuous variables are winsorized at the 1st and 99th percentiles of their distributions.

3.3. Regression Models

We examine whether uncertainty measures derived from the Loughran and McDonald (LM) dictionary or convolutional neural networks (CNN) can explain contemporaneous stock returns surrounding the 10-K filing date stock returns around 10-K filing dates and predict future realized returns . Following Loughran and McDonald (2011) and Frankel et al. (2022), we estimate the following regression model:

$$BHR[-1,1] \text{ (or } BHR[2, n]) = a_0 + a_1 \times UncertainyMeasures_{i,t} + TextualControls_{i,t} + FirmControls_{i,t} + Industry\ FE + Year\ FE + e_{i,t}, \quad (1)$$

The dependent variable *BHR [-1,1]*, measures contemporaneous stock returns, computed as the buy-and-hold raw returns over the three-day window surrounding the 10-K filing date. The dependent variable *BHR[2,n]*, captures future realized returns, calculated as buy-and-hold raw returns over one-month (n=20), one-quarter (n=60) and half-year (n=125) windows starting two days after the 10-K filing dates. To ensure robustness, we adjust buy-and-hold returns by risk factors. We compute market-adjusted returns, denoted as *BHMAR[-1,1]* and *BHMAR[2,n]* respectively, using CRSP value-weighted index return. We compute size and

book-to-market (BM) adjusted returns, denoted as *BHSAR[-1,1]* and *BHSAR[2,n]* respectively, using portfolios of firms matched on size and BM ratios. We also compute Carhart 4-factor-adjusted returns, where the factor loadings are based on a regression of a four-factor model using trading days [−252, −6] relative to the 10-K filing date (Carhart, 1997). For ease of interpretation, all stock returns are expressed in percentage terms by multiplying by 100.

The explanatory variable, *UncertainyMeasures*, is the proportion of uncertainty-related sentences in the MD&As, as identified by the LM dictionary or CNN.[9] To highlight the limitations of traditional dictionary-based approaches, we first estimate Eq. (1) using the LM-based measure (*LM*), which represents the proportion of sentences containing uncertainty triggers as defined by Loughran and McDonald (2011). However, our main focus is on the deep learning approach (CNN). As described in Section 3.1, we employ CNN to differentiate meaningful uncertainty statements (*CNN*) from boilerplate statements (*B*). The *LM* is then decomposed into the sum of *CNN* and *B*, demonstrating that a significant portion of LM-identified uncertainty is boilerplate and not meaningful. Furthermore, we categorize uncertainty statements (*CNN*) into those reflecting aleatory uncertainty (*A*) and epistemic uncertainty (*E*), consistent with our theoretical framework. Our results are robust when using the counts of uncertainty-related sentences in the MD&As.

We control for textual attributes commonly used in the literature (Li, 2008; Loughran and McDonald, 2011), including disclosure amount, sentiment, and readability. The disclosure amount is measured as the natural logarithm of the number of sentences (*Length*) in the MD&A section. Prior literature suggests that 10-K filing returns are associated with sentiment (e.g., Loughran and McDonald, 2011). To control for sentiment, we use the finance-oriented sentiment dictionary developed by Loughran and McDonald (2011). Specifically, *Tone* is calculated as the difference between the number of positive and negative words, scaled by the

[9] Our results are robust when using the raw number of sentences count.

total number of sentiment words. Since linguistic complexity is linked to operational uncertainty (Bushee et al., 2018), we use the Fog index (*Fog*) to control for readability (Li, 2008).

Following prior studies (e.g., Loughran and McDonald, 2011; Frankel et al., 2022), we control for firm-level factors (*FirmControls*) associated with stock returns. These include current earnings performance (*ROA*), the natural logarithm of market value (*LnMV*), the book-to-market ratio (*BTM*), the natural logarithm of the number of business and geographic segments (*LnSeg*), the natural logarithm of the sum of one and the number of analysts following(*LnAF*), stock turnover (*Turnover*), Carhart alpha (*PreAlpha*), and buy and hold raw return (*PastReturn*) calculated over trading days [−252, −6] relative to the 10-K filing date. Additionally, we incorporate industry fixed effects and year fixed effects. Standard errors are clustered by firm in all regressions.

3.4. Descriptive Statistics

Table 2 presents the summary statistics for the base sample of 56,480 observations. The mean value of *LM* is 0.28, indicating that 28% of the sentences in the MD&A contain "uncertainty" triggers identified by Loughran and McDonald (2011). The mean of *CNN* is 0.07, suggesting that only 25% (0.07/0.28) of the sentences classified as uncertainty-related by the LM dictionary are indeed uncertainty-related (as indicated by our deep learning approach *CNN*). Among the *CNN*-identified uncertainty statements, about two-thirds pertain to aleatory uncertainty, while the remaining one-third relate to epistemic uncertainty. The mean of *Length* is 4.86, which coresponds to approximately 129 sentences in the MD&A. The average sentiment (*Tone*) is negative at -0.19. The mean Fog index (*Fog*) for the MD&A is 22.81, indicating that the readability is classified as "unreadable" based on Li (2008). The summary statistics for other control variables are comparable to those reported in prior studies (e.g., Frankel et al., 2021).

[Insert Table 2]

Table 3 reports the pairwise correlations between the variables. As showed in the Pearson correlations, *CNN* is significantly positively correlated with future realized returns *BHSAR[2,60]*, *BHFAR[2,60]* and *BHFAR[2,125]*. While *CNN* is significantly negatively correlated with sentiment (*Tone*), it is significantly positively correlated with complexity (*Fog*). These suggests that uncertainty-related statements tend to be more negative and less readable.

Both *CNN* and aleatory uncertainty (*A*) are positively and significantly correlated with firm size (*LnMV*), indicating that larger firms are more likely to exhibit higher levels of aleatory uncertainty. Epistemic uncertainty (*E*) also shows a significantly negative correlation with firm size, suggesting that smaller firms would have higher level of ambiguity. In terms of profitability, both *CNN* and aleatory uncertainty (*A*) exhibit a positive and statistically significant correlation with current profitability (*ROA*), suggesting that riskier firms tend to exhibit stronger current performance. Epistemic uncertainty (*E*), however, shows a negative correlation with *ROA*. The correlations between *CNN* and the other control variables are generally low, alleviating concerns about multicollinearity.

[Insert Table 3]

Figure 2 plots the equal-weighted aggregate uncertainty (using different measures) over time. Panel A presents the results for *LM* and *B*, while Panel B shows the results for *CNN*, *A*, and *E*. While boilerplate statements (*B*) exhibit a steady trend (except for the earliest period), aggregate uncertainty statements (*CNN*) capture time-varying aggregate uncertainty information associated with the macro environment. Three distinct peaks are observed: in 1999 (potentially driven by the speculative environment of the dot-com bubble), 2009 (during the financial crisis), and 2021 (amid the COVID-19 pandemic). In Panel B, the comparison between aleatory and epistemic uncertainty reveals that variations in *CNN* are primarily driven by changes in statements related to aleatory uncertainty.

[Insert Figure 2]

4. Results

4.1. Evidence on the Stock Return

We employ the regression model specified in Equation (1) to examine whether uncertainty measures explain contemporaneous and future realized returns. Panel A of Table 4 presents results where the dependent variables are contemporaneous returns. Columns 1–3 report results for three-day buy-and-hold raw returns (*BHR*[*-1,1*]). Columns 4–6 report results for three-day market-adjusted buy-and-hold returns (*BHMAR*[*-1,1*]), and columns 7–9 report results for size-and-value-adjusted buy-and-hold returns (*BHSAR*[*-1,1*]) surrounding 10-K filings. Columns 10–12 report results for three-day buy-and-hold returns adjusted by Carhart 4 factor model. In columns (1), (4), (7), and (10), the coefficients on *LM* are statistically insignificant, which could reflect either the limited validity of the dictionary-based approach or the conflation of distinct uncertainty attributes. Similarly, the coefficients on *CNN*—a deep learning-based measure capturing the proportion of uncertainty statements after removing boilerplate language—are also insignificant in columns (2), (5), (8), and (11). These results indicate that overall uncertainty is not associated with contemporaneous stock returns.

However, when we disaggregate uncertainty into its epistemic and aleatory components, we uncover economically meaningful distinctions. In columns (3), (6), (9), and (12), epistemic uncertainty (*E*) exhibits a negative and statistically significant effect ($p < 0.10$), suggesting that the market reacts adversely to statements reflecting epistemic uncertainty. This finding aligns with theories of ambiguity aversion, which predict that investors discount assets when faced with uncertainty about the underlying probabilities or incomplete information. In contrast, the market reaction to aleatory uncertainty (*A*) is statistically insignificant, indicating that investors do not exhibit the same aversion to uncertainty arising from inherent randomness or variability in outcomes. This divergence underscores the importance of distinguishing between these two

dimensions of uncertainty, as they convey fundamentally different types of uncertainty. The results provide strong support for *H1a*, which posits that epistemic uncertainty elicits a negative market reaction due to ambiguity aversion.

[Insert Table 4]

Sentiment (*Tone*) displays positive but statistically insignificant coefficients across all specifications. This result is consistent with Frankel et al. (2022), who find that sentiment measures based on Loughran and McDonald's (2011) dictionary no longer explain stock returns in recent 10-K filings.

Panel B (Panel C and Panel D) of Table 4 reports results for future realized returns, measured as buy-and-hold abnormal returns over the one-month (one-quarter and half-year) window beginning two days after the 10-K filings. The coefficients on *LM*, the dictionary-based measure, are statistically insignificant in both Panels B, C, and D. In contrast, the coefficients on *CNN*, our deep learning-based measure, are marginally significant in Panel B but become highly significant ($p < 0.01$) in explaining one-quarter-ahead stock returns in Panel C. This result underscores the superior ability of our deep learning approach to identify uncertainty statements in the MD&A that predict future stock returns, compared to traditional dictionary-based methods.

To isolate the drivers of this result, we decompose total uncertainty (*CNN*) into its epistemic and aleatory components in column (3) of Panels C. The coefficient on aleatory uncertainty (*A*) is positive and highly significant (*coeff.* = 5.62, t = 2.89), indicating that aleatory uncertainty statements are a significant positive predictor of one-quarter-ahead stock returns. This supports *H1b*, which posits that aleatory uncertainty, reflecting variability due to inherent stochasticity, conveys quantifiable risk for which the market demands compensation through a risk premium. In contrast, the coefficient on epistemic uncertainty (*E*) is insignificant. While epistemic uncertainty induces ambiguity aversion and contemporaneous

price discounts, it does not systematically affect future realized returns. Importantly, the decomposition reveals that the results for total uncertainty (*CNN*) are primarily attributable to aleatory uncertainty, which accounts for two-thirds of the *CNN*-identified uncertainty statements.

4.2. Portfolio Analyses

To further examine the relationship between uncertainty measures and stock returns, we conduct portfolio analyses by sorting stocks into quintiles based on their levels of three measures: *CNN* (total uncertainty), *A* (aleatory uncertainty), and *E* (epistemic uncertainty). We calculate raw returns and market-adjusted, size-and-value-adjusted, and Carhart 4-factor-adjusted returns for each quintile and construct hedge portfolios to assess the differential returns between the highest and lowest uncertainty quintiles (Carhart, 1997). The results are presented in Table 5.

[Insert Table 5]

In Panel A, we report equal-weighted three-day portfolio returns surrounding the 10-K filing dates.[10] For epistemic uncertainty (*E*), the hedge portfolio return is -0.44% (t = -3.30) for raw returns, -0.48% (t = -3.72) for market-adjusted returns, -0.47% (t = -4.03) for size-and-value-adjusted returns, and -0.42% (t = -3.97) for Carhart 4-factor-adjusted returns. The quintile returns for epistemic uncertainty (*E*) also exhibit a monotonic pattern from the bottom (lowest uncertainty) to the top quintile (highest uncertainty). These findings are consistent with ambiguity aversion, suggesting that investors discount assets when facing uncertainty stemming from a lack of knowledge or information.

We report one-month portfolio returns in Panel B, one-quarter portfolio returns in Panel C and half-year portfolio returns in Panel D. For aleatory uncertainty (*A*), the one-quarter hedge portfolio return is 1.08% (t = 3.05) for raw returns, 0.92% (t = 2.44) for market-adjusted returns,

[10] The results are robust when equal-weighted portfolios are replaced by value-weighted portfolios.

0.92% (t = 2.73) for size-and-value-adjusted returns, and 0.86% (t = 2.43) for Carhart 4-factor-adjusted returns, annualizing to 4.37%, 3.73%, 3.73%, and 3.48%, respectively. Notably, the returns increase monotonically from the bottom (lowest uncertainty) to the top quintile (highest uncertainty). These results support a risk premium interpretation for aleatory uncertainty, indicating that investors demand higher returns for bearing risk. In Panel D, we find less significant half-year hedge portfolio returns for aleatory uncertainty.

4.3. Fama-Macbeth Regressions

In this section, we conduct Fama-MacBeth regressions, applying a Newey-West correction with six lags to account for cross-sectional and serial correlation in the panel data. We show the results on three-day contemporaneous stock returns, one-month-ahead, one-quarter-ahead, and half-year ahead stock returns in Panels A, B, C, and D, respectively. To facilitate the comparison of economic magnitudes, we transform our independent variables into decile ranks for each year (denoted as *D_CNN*, *D_A*, and *D_E*). These ranks are initially numbered 1 through 10 and then scaled to range from -0.5 to +0.5, with a mean of zero and a range of one. This scaling ensures that the regression coefficients represent the abnormal returns from a zero-investment strategy of going long on the highest decile and short on the lowest decile.

The results are presented in Table 6. Panel A shows that the coefficient on *D_E*, our measure of epistemic uncertainty, is significantly negative across all specifications: -0.18, -0.21, -0.22, and -0.19 in columns (2), (4), (6), and (8), respectively. These estimates imply a three-day raw, market-adjusted, size-and-value-adjusted, or Carhart four-factor-adjusted return of -0.18%, -0.21%, -0.22%, or -0.19%, respectively, for a hedge portfolio that is long the highest decile of epistemic uncertainty and short the lowest decile. This finding aligns with the notion that epistemic uncertainty elicits negative contemporaneous returns due to ambiguity aversion.

[Insert Table 6]

Panel C shows that quarterly returns for a hedge portfolio based on aleatory uncertainty (*D_A*) are 1.21%, 1.09%, 1.05%, and 1.10% for raw, market-adjusted, size-and-value-adjusted, and Carhart four-factor-adjusted returns, respectively, translating to annualized returns exceeding 4.89%, 4.43%, 4.24%, and 4.49%. In columns (2), (4), (6), and (8), the coefficients on total uncertainty (*D_CNN*) are also significantly positive. Untabulated analyses further show that the economic magnitudes of *D_A* and *D_CNN* are substantially larger than those of other standardized textual attributes in predicting one-quarter-ahead abnormal returns.

4.4. Interaction between Ambiguity and Risk

In Epstein and Schneider's (2008) framework, ambiguity aversion generates a persistent "ambiguity premium" — a contemporaneous discount in asset prices — that becomes more pronounced when fundamentals are volatile. This premium arises from the interaction between ambiguity and idiosyncratic risk, such that the price impact of epistemic uncertainty is amplified in high-aleatory environments. This mechanism implies that the negative contemporaneous return associated with ambiguity should be strongest when uncertainty about fundamentals is elevated.

A related implication from prior ambiguity models is that high ambiguity can result in portfolio inertia and departures from equilibrium prices, thereby preventing the efficient pricing of measurable (aleatory) risk (Epstein and Wang, 1994; Epstein and Schneider, 2010; Illeditsch, 2011). For example, Epstein and Schneider (2010) show that portfolio inertia can arise in risky portfolios when investors use an ambiguous asset to hedge. Under this mechanism, the interaction of risk and ambiguity can lead to portfolio inertia in risky portfolios when investors process ambiguous news, resulting in a continuum of equilibrium stock prices for some fundamentals.

We test these two implications by partitioning the sample and re-estimating return regressions for epistemic and aleatory uncertainty. Panel A of Table 7 reports the contemporaneous (three-day) return results when the sample is partitioned by aleatory uncertainty (risk). Consistent with the model of Epstein and Schneider (2008), the negative three-day contemporaneous stock price discount associated with epistemic uncertainty (ambiguity) is concentrated in the high-risk subsample. Across the four return measures, the coefficients on epistemic uncertainty (ambiguity) in the high-risk subsample are negative and statistically significant, whereas in the low-risk subsample the coefficients are statistically insignificant. These results indicate that ambiguity-premium-driven discounts are materially larger when fundamental risk is elevated.

[Insert Table 7]

Panels B–D present complementary tests of the portfolio inertia mechanism by partitioning the sample on epistemic uncertainty (ambiguity) and estimating future realized return regressions at one-month, one-quarter, and half-year horizons. The positive future returns associated with aleatory uncertainty (risk) are large and highly significant in the subsample with low ambiguity but insignificant in the high-ambiguity subsample. Across all return horizons, measurable risk yields a positive risk premium only in low-ambiguity environments; when ambiguity is elevated, the compensation for aleatory risk is attenuated or absent.

Taken together, these results provide clear empirical support for both theoretical channels. First, ambiguity aversion produces a contemporaneous price discount that is amplified when fundamentals are volatile, consistent with Epstein and Schneider (2008). Second, high ambiguity induces portfolio inertia and a persistent discount in asset prices, which prevents the efficient pricing of measurable risk, as predicted by Epstein and Wang (1994) and Illeditsch (2011). These findings highlight that the asset-pricing consequences of uncertainty depend not

only on its nature, but also on the joint distribution of its ambiguity and aleatory risk components.

4.5. Idiosyncratic Risk Puzzle

A longstanding debate in asset pricing concerns whether idiosyncratic risk is compensated in expected returns. Classical portfolio theory predicts that only non-diversifiable (systematic) risk should be priced, while idiosyncratic risk, being diversifiable, should not command a premium. Merton (1987), however, shows that with incomplete information and limited diversification, investors may require compensation for bearing firm-specific risk, implying a positive relation between idiosyncratic risk and expected returns.

Empirical evidence on this prediction is mixed. Ang et al. (2006, 2009) document a striking negative relation between lagged idiosyncratic volatility and future returns, giving rise to the "idiosyncratic risk puzzle." In contrast, Fu (2009) attributes the negative relationship observed by Ang et al. (2006) to return reversals among small, high-volatility stocks. By using conditional volatility estimates, Fu (2009) finds a positive relationship between idiosyncratic risk and expected returns. Similarly, Goyal and Santa-Clara (2003) find a positive relationship between average stock variance (largely idiosyncratic) and market returns, though Bali et al. (2005) argue that this result is driven by small stocks and liquidity premiums. Chen and Petkova (2012) and Bégin et al. (2019) emphasize the role of tail risk and variance innovations and find contrasting results.

Our setting provides a new angle on this debate. Our earlier results in Tables 4-7 suggest that the risk premium linked to aleatory uncertainty remains significant even after controlling for standard risk factors. This evidence is consistent with the notion that idiosyncratic risk commands a positive expected return, consistent with the positive associations identified by Fu (2009) and Bégin et al. (2019).

To further investigate the "idiosyncratic risk puzzle," we decompose aleatory risk into systemic and idiosyncratic components by projecting aleatory risk on industry-by-year fixed effects:

$$A_{i,t} = \alpha + \text{Industry-Year FEs} + e_{i,t} \quad (2),$$

where $A_{i,t}$ is our measure of aleatory uncertainty for firm i in year t; The fitted values from (2) represent the component of aleatory risk explained by common industry–year conditions (the systemic component *A_Sys*), while the residuals capture the remaining firm-specific variation (the idiosyncratic component *A_Idio*).

Having decomposed aleatory uncertainty (risk) into systemic and idiosyncratic components, we next examine how each is priced. To mitigate the influence of outliers, we transform both measures into decile ranks. Panel A of Table 8 reports regressions of future buy-and-hold returns over one-, three-, and six-month horizons on the systemic component (*D_A_Sys*) and the idiosyncratic component (*D_A_Idio*), controlling for the same covariates as in previous analyses. We find that both components of aleatory risk are positively and significantly related to future realized returns, with the magnitude of this relationship increasing with the investment horizon. Specifically, for the six-month horizon (column (3)), the coefficient on *D_A_Sys* is 1.79 (t = 3.33) and the coefficient on *D_A_Idio* is 1.02 (t = 2.09), indicating that firms with higher systemic and idiosyncratic risk, respectively, are associated with materially higher six-month-ahead returns. Collectively, these results suggest that the positive risk premium for aleatory risk documented in Tables 4–7 reflects compensation for both systemic and idiosyncratic risk.

[Insert Table 8]

4.6. Ambiguity and the Pricing of Systemic and Idiosyncratic Risk

The evidence in Section 4.4 shows that the return implications of aleatory risk depend critically on the prevailing level of ambiguity. To examine this further, Panel B of Table 8

reports return regressions for the systematic (*D_A_Sys*) and idiosyncratic (*D_A_Idio*) components, partitioned into low- and high-ambiguity states. Columns (1), (3), and (5) present results for the low-ambiguity subsample; Columns (2), (4), and (6) present results for the high-ambiguity subsample. Specifications mirror those in Panel A, with identical controls and fixed effects.

In low-ambiguity states, idiosyncratic risk earns a sizable and robust premium across all horizons. The coefficients on *D_A_Idio* are positive and highly significant: 0.75 ($t = 2.79$) for one-month returns, 1.58 ($t = 3.28$) for three-month returns, and 2.23 ($t = 3.41$) for six-month returns. Low ambiguity enables investors to assess firm-specific risk with greater precision, allowing accurate estimation of required compensation. When compensation adequately reflects the risk–return tradeoff, investors rationally forgo costly diversification, bearing diversifiable risk until marginal compensation equals marginal diversification costs. At this equilibrium, residual idiosyncratic risk is effectively non-diversifiable at the margin and commands a risk premium, consistent with an environment in which clearer information improves risk assessment so that idiosyncratic risk is both willingly borne and systematically priced. In contrast, under high ambiguity, the idiosyncratic risk premium effectively disappears. The coefficients for *D_A_Idio* are statistically insignificant at all horizons. This pattern is consistent with ambiguity aversion. When both idiosyncratic risk and ambiguity are high, investors diversify away such exposures, eliminating any systematic compensation.

Systemic risk exhibits the opposite conditioning pattern. In low-ambiguity states, the premium associated with *D_A_Sys* is weak at shorter horizons and becomes statistically significant only at the six-month horizon, consistent with exposure to persistent macroeconomic state variables whose premia are small relative to transitory noise in the short run and require longer horizons for cumulative effects to be detected. Under high ambiguity, the distribution of systemic risk is harder to assess. Because macroeconomic shocks and the

uncertainty surrounding them are inherently undiversifiable, investors must bear these exposures and demand higher expected returns—an ambiguity premium. Consistent with this, *D_A_Sys* is positive and significant at both the three-month (1.24, t = 2.19) and six-month (1.43, t = 1.76) horizons.

Taken together, these findings substantially refine the interaction documented in Section 4.4, showing that ambiguity plays a pivotal role in conditioning the pricing of both components of aleatory risk. In low-ambiguity environments, firm-specific risk earns a substantial premium, whereas in high-ambiguity environments, ambiguity-averse investors avoid such exposures and the premium vanishes. For systemic risk, the premium under low ambiguity emerges only over longer horizons, whereas under high ambiguity it is present even at shorter horizons, reflecting heightened compensation for ambiguous and undiversifiable macroeconomic exposure. This conditional interaction provides a fresh lens on the idiosyncratic risk puzzle and offers a coherent explanation for the mixed empirical evidence on compensation for idiosyncratic risk.

4.7. Comparison with Existing Measures of Ambiguity or Information Uncertainty

Finally, as a validation test, we benchmark our ambiguity measure, *D_A*, and aleatory risk measure, *D_E*, against existing proxies for "ambiguity" and "information uncertainty" used in the literature. Specifically, we examine their correlations with analyst forecast dispersion (*D_Disp*), the VIX index (*D_VIX*), firm-level implied volatility (*D_ImplVol*), the standard deviation of operating cash flows (*D_StdCFO*), firm size (*D_LnMV*), and firm age (*D_Age*). All variables are converted into decile ranks, and the coefficients of the regressions are reported in Table 9.

[Insert Table 9]

The evidence in Table 9 shows that the existing proxies are not "pure" measures of either ambiguity or risk; instead, they load on both *D_A* and *D_E*. Analyst forecast dispersion,

VIX, implied volatility, and cash flow volatility are all positively and significantly related to both measures. For example, *D_ImplVol* and *D_StdCFO* have sizeable positive correlations with *D_E* (0.18 and 0.19, respectively), but they are also positively correlated with *D_A* (0.03 and 0.02). Likewise, *D_Disp* and *D_VIX* are positively related to both *D_A* and *D_E*. Thus, when these variables are used as stand-alone proxies for "ambiguity" or "information uncertainty," they combine ambiguity and risk in a single scalar, making it impossible to isolate the separate effects of each component.

At the same time, the patterns underscore that *D_A* and *D_E* capture distinct dimensions of uncertainty. The sharpest contrast arises for firm size and firm age. *D_A* is positively related to both *D_LnMV* and *D_Age* (0.04 and 0.08), whereas *D_E* is negatively related to these same variables (–0.14 and –0.17). Thus, risk, as captured by *D_A*, tends to be higher for larger and more mature firms, while ambiguity, as captured by *D_E*, is stronger for smaller and younger firms. The negative association of *D_E* with size and age is consistent with the view that smaller and younger firms are more informationally opaque and harder to understand. In contrast, the positive association of *D_A* with size and age suggests that aleatory risk is not confined to small, young, or "risky" firms in the conventional sense; large and mature firms can also exhibit substantial inherent risk, for example because of complex operations, extensive scale, or exposure to aggregate shocks. The opposing signs on size and age make clear that ambiguity and risk do not move in lockstep and would be conflated if one relied on these characteristics, or on standard "uncertainty" proxies more generally.

Moreover, the associations, while statistically strong, are modest in magnitude, indicating that *D_A* and *D_E* are not simple relabelings of existing measures. Rather, they extract two distinct components—aleatory risk (*D_A*) and ambiguity (*D_E*)—from the broader construct that traditional proxies mix together. This decomposition helps reinterpret prior work that uses forecast dispersion, volatility, or firm characteristics as generic uncertainty proxies

and provides a cleaner framework for studying how different facets of uncertainty are priced and shape corporate decisions.

5. Conclusion

This paper advances the understanding of uncertainty in financial markets by decomposing it into its two fundamental dimensions: ambiguity (epistemic uncertainty) and risk (aleatory uncertainty). Using a deep learning approach, we operationalize these dimensions in corporate disclosures, providing a novel framework to analyze their distinct effects on investor behavior and asset pricing. Our findings reveal that epistemic uncertainty, reflecting ambiguity about the precision or reliability of information, elicits negative contemporaneous stock returns, consistent with theories of ambiguity aversion. In contrast, aleatory uncertainty, capturing measurable risk, is positively associated with future realized returns, as investors demand compensation for bearing quantifiable risk.

Our study also provides new insights into the longstanding idiosyncratic risk puzzle. While traditional asset pricing theory suggests that idiosyncratic risk should not command a premium, we find that the risk premium persists for both the systematic and idiosyncratic components of aleatory uncertainty. Importantly, compensation for idiosyncratic risk is significantly higher in low-ambiguity environments, when investors can more confidently assess the nature of the risk. In high-ambiguity states, this premium largely disappears, consistent with ambiguity-averse investors diversifying away such exposures, while a substantial premium for systemic risk emerges to compensate for bearing ambiguous and undiversifiable systemic exposures. This conditional pricing mechanism clarifies the circumstances under which idiosyncratic and systemic risks are priced and demonstrates that ambiguity is a critical moderating factor for both. Our findings complement Merton's (1987) investor recognition hypothesis and offer a unified framework for understanding how the interplay between risk type and ambiguity shapes asset prices.

Methodologically, we contribute to the literature on textual analysis in finance by demonstrating the value of deep learning techniques for analyzing corporate disclosures. Our convolutional neural network (CNN) model outperforms traditional methods in identifying and decomposing uncertainty, providing a more precise framework for understanding its role in financial markets. This approach opens new avenues for research on the role of textual information in asset pricing and offers a robust tool for measuring complex constructs such as uncertainty, risk, and ambiguity.

Our framework has important implications for corporate disclosure policy. While regulators have emphasized the importance of transparency in corporate reporting, our results suggest that the nature of disclosed uncertainty—whether epistemic or aleatory—matters greatly for investor response. Firms and regulators should consider not only the quantity of uncertainty disclosed but also its type and source. Ambiguity is likely to trigger adverse market reactions, while risk-related disclosures may be priced as part of expected returns. This insight can guide firms in crafting more effective disclosures that reduce ambiguity and better communicate measurable risks.

References

Anderson, E. W., Ghysels, E., & Juergens, J. L. (2009). The impact of risk and uncertainty on expected returns. *Journal of Financial Economics*, *94*(2), 233-263.

Andrei, D., & Hasler, M. (2015). Investor attention and stock market volatility. *The Review of Financial Studies*, *28*(1), 33-72.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics*, *91*(1), 1-23.

Akbas, F., Jiang, C., & Koch, P. D. (2017). The trend in firm profitability and the cross-section of stock returns. *The Accounting Review*, *92*(5), 1–32.

Antoniou, C., Harris, R. D., & Zhang, R. (2015). Ambiguity Aversion and Market Participation: Evidence from Fund Flows. *Journal of Banking and Finance, 58,* 57-70.

Aristotle. (1984). *Physics* (R. P. Hardie & R. K. Gaye, Trans.). In J. Barnes (Ed.), *The complete works of Aristotle: The revised Oxford translation* (Vol. 1, pp. 315–446). Princeton University Press. (Original work published ca. 350 B.C.E.)

Aristotle. (1984). *Nicomachean Ethics* (W. D. Ross, Trans.). In J. Barnes (Ed.), *The complete works of Aristotle: The revised Oxford translation* (Vol. 2, pp. 1729–1867). Princeton University Press. (Original work published ca. 350 B.C.E.)

Azimi, M., & Agrawal, A. (2021). Is positive sentiment in corporate annual reports informative? Evidence from deep learning. *Review of Asset Pricing Studies*, *11*(4), 762–805.

Bae, J. W., Belo, F., Li, J., Lin, X., & Zhao, X. (2023). The opposing effects of complexity and information content on uncertainty dynamics: Evidence from 10-K filings. *Management Science*, *69*(10), 6313-6332.

Bali, T. G., Cakici, N., Yan, X., & Zhang, Z. (2005). Does idiosyncratic risk really matter?. *The Journal of Finance*, *60*(2), 905-929.

Bégin, J. F., Dorion, C., & Gauthier, G. (2019). Idiosyncratic jump risk matters: Evidence from equity returns and options. *The Review of Financial Studies*, *33*(1), 155-211.

Bekaert, G., Hoerova, M., & Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, *60*(7), 771-788.

Bossaerts, P., Ghirardato, P., Guarnaschelli, S., & Zame, W. R. (2010). Ambiguity in asset markets: Theory and experiment. *The Review of Financial Studies*, *23*(4), 1325-1359.

Boureau, Y. L., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 111–118).

Brenner, M., & Izhakian, Y. (2018). Asset pricing and ambiguity: Empirical evidence. *Journal of Financial Economics*, *130*(3), 503-531.

Brown, S. V., Hinson, L. A., & Tucker, J. W. (2024). Financial statement adequacy and firms' MD&A disclosures. *Contemporary Accounting Research*, *41*(1), 126-162.

Cao, H. H., Han, B., Hirshleifer, D., & Zhang, H. H. (2011). Fear of the unknown: Familiarity and economic decisions. *Review of Finance*, *15*(1), 173-206.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57-82.

Caskey, J. A. (2009). Information in equity markets with ambiguity-averse investors. *The Review of Financial Studies*, *22*(9), 3595-3627.

Chen, Z., & Petkova, R. (2012). Does idiosyncratic volatility proxy for risk exposure? *The Review of Financial Studies*, *25*(9), 2745-2787.

D'Augusta, C., & DeAngelis, M. D. (2020). Tone concavity around expected earnings. *The Accounting Review*, *95*(1), 133–164.

Dicks, D. L., & Fulghieri, P. (2019). Uncertainty aversion and systemic risk. *Journal of Political Economy*, *127*(3), 1118-1155.

Dicks, D., & Fulghieri, P. (2021). Uncertainty, investor sentiment, and innovation. *The Review of Financial Studies*, *34*(3), 1236-1279.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics*, *119*(3), 559-577.

Dow, J., & da Costa Werlang, S. R. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica: Journal of the Econometric Society*, *60*(1), 197-204.

Easley, D., & O'Hara, M. (2009). Ambiguity and nonparticipation: The role of regulation. *The Review of Financial Studies*, *22*(5), 1817-1843.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, *75*(4), 643-669.

Epstein, L., & Wang, T. (1994). Intertemporal Asset Pricing Under Knightian Uncertainty. *Econometrica*, *62*(2), 283-322.

Epstein, L. G., & Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies*, *74*(4), 1275-1303.

Epstein, L. G., & Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *The Journal of Finance*, *63*(1), 197-228.

Epstein, L. G., & Schneider, M. (2010). Ambiguity and asset markets. *The Annual Review of Financial Economics*, *2*(1), 315-346.

Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, *20*(1), 41-81.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141-153.

Gleason, C. A., & Lee, C. M. (2003). Analyst forecast revisions and market price discovery. *The Accounting Review*, *78*(1), 193-225.

Guidolin, M., & Rinaldi, F. (2013). Ambiguity in asset pricing and portfolio choice: A review of the literature. *Theory and Decision*, *74*(2), 183-217.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, *51*(1), 55-84.

Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In Brun, W., Keren, G., Kirkebøen, G., and Montgomery, H., editors, *Perspectives on Thinking, Judging, and Decision Making: A tribute to Karl Halvor Teigen*, pages 21–35. Universitetsforlaget, Oslo.

Frankel, R., Jennings, J., & Lee, J. (2022). Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science*, *68*(7), 5514–5532.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics*, *91*(1), 24-37.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141-153.

Goyal, A., & Santa-Clara, P. (2003). Idiosyncratic risk matters!. *The Journal of Finance*, *58*(3), 975-1007.

Hacking, I. (1975). *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*. Cambridge University Press

Huang, A., Wang, H., & Yang, Y. (2022). FinBERT—A deep learning approach to extracting textual information. *Contemporary Accounting Research*, *40*(2), 806-841

Illeditsch, P. K. (2011). Ambiguous information, portfolio inertia, and excess volatility. *The Journal of Finance*, *66*(6), 2213-2247.

Ilut, C. L., & Schneider, M. (2014). Ambiguous business cycles. *American Economic Review*, *104*(8), 2368-2399.

Kostopoulos, D., Meyer, S., & Uhr, C. (2022). Ambiguity about volatility and investor behavior. *Journal of Financial Economics*, *145*(1), 277-296.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2–3), 221–247.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35–65.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Knight, F. H. (1921). *Risk, Nncertainty, and Profit*. Chicago, IL: University of Chicago Press.

Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, *18*, 1088–1122.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance, 42*, 483-510

Peijnenburg, K. (2018). Life-cycle asset allocation with ambiguity aversion and learning. *Journal of Financial and Quantitative Analysis*, *53*(5), 1963-1994.

Savage, L. J. (1954). *The Foundations of Statistics*. New York, Wiley.

Securities and Exchange Commission (SEC). (2005). Securities and Exchange Commission final rule, release no. 33–8591 (FR-75). http://www.sec.gov/rules/final/33-8591.pdf

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425-442.

Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment extremity and accuracy under epistemic vs. aleatory uncertainty. *Management Science*, *63*(2), 497–518.

Walters, D. J., Ülkümen, G., Tannenbaum, D., Erner, C., & Fox, C. R. (2023). Investor behavior under epistemic vs. aleatory uncertainty. *Management Science*, *69*(5), 2761–2777.

Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, *18*(2), 673-705.

Williams, C. D. (2015). Asymmetric responses to earnings news: A case for ambiguity. *The Accounting Review*, *90*(2), 785-817.

Appendix A Examples of Boilerplate, Aleatory, and Epistemic Uncertainty Statements

In this appendix, we illustrate several patterns and list sample sentences of the three uncertainty categories.

Boilerplate statements (B):

The boilerplate information refers to sentences that are not associated with any uncertainty arguments. This is because the LM dictionary contains words like "may", "might", "could" or "can", which are meaningless modals and boilerplate disclosures. Also, the patterns of "approximately/about + #Number" are considered to be boilerplate because only the approximate value of the numerical information is mentioned. Some accounting terminologies can be boilerplate when they are not referring to firms' operations, e.g., variable costs, credit exposure, doubtful accounts, intangible assets, and risk adjustment in accounting practices.

Sentences with "uncertain" triggers (in bold) that are boilerplate:

*B1) Operating expenses operation and maintenance expenses decreased **approximately** $2.3 million in 1994 when compared to 1993.*

*B2) The write-off of goodwill and certain other **intangible** assets, primarily tradenames and franchise agreements, represent noncash charges of $1,267.7 million and $207.1 million, respectively.*

*B3) If yields are materially different than projected, work-in-process inventory **may** need to be revalued.*

*B4) If actual future after-tax cash flows differ significantly from their estimates, we **may** be required to record an impairment charge to write down the identifiable **intangible** assets to their realizable values.*

Aleatory uncertainty (A):

Aleatory uncertainty (1) is represented in relation to a class of possible outcomes, (2) is focused on assessing an event's propensity, (3) is naturally measured by relative frequency, and (4) is attributed to stochastic behavior. A convenient judgment is whether it mentions any increase of randomness. Our "aleatory" category generally corresponds to external-distributional uncertainty, which cannot be reduced. Our manual reading reveals that in most cases aleatory uncertainty is attached to negative sentiment.

The pattern "…**could** have an adverse impact on" & "**may** negatively impact" should be considered to be aleatory because the uncertainty increases due to external factors (see Sample A2).

Sentences with "uncertain" triggers that are aleatory:

*A1) Our cash balances **fluctuate** from quarter to quarter due to slight seasonality in our business and the timing of certain cash disbursements.*

*A2) Business cycles and other changes in demand **could** adversely impact our revenue and profits.*

Epistemic uncertainty (E):

Epistemic uncertainty: (1) is represented in terms of a single case, (2) is focused on the extent to which an event is or will be true or false, (3) is naturally measured by confidence in one's knowledge or model of the causal system determining the outcome, and (4) is attributed to missing information or expertise. Our "epistemic" category generally corresponds to internal uncertainty, which can be reduced by managers' efforts. In most cases, epistemic uncertainty is attached to positive sentiment.

The pattern "we **anticipate** that … will increase" or "we **believe** that … will" should be considered to be epistemic (in most cases), as it reflects managers' confidence in future outcomes (See Sample E1). In some cases, when "believe" has the same meaning as "think" and the sentences do not refer to future operations, they are boilerplate. For example, "*We*

believe *the following critical accounting policies contain the more significant judgments and estimates used in the preparation of our financial statements.*"

The pattern "We **may**" or "Our company **may**" can be the epistemic category (because the uncertainty is attributed to their information or expertise, see Sample E2) or boilerplate (Sample B2), depending on the context.

Sentences with "uncertain" triggers that are epistemic:

E1) We ***believe*** *our future operating cash flows will be more than sufficient to cover debt repayments, other contractual obligations, capital expenditures and dividends.*

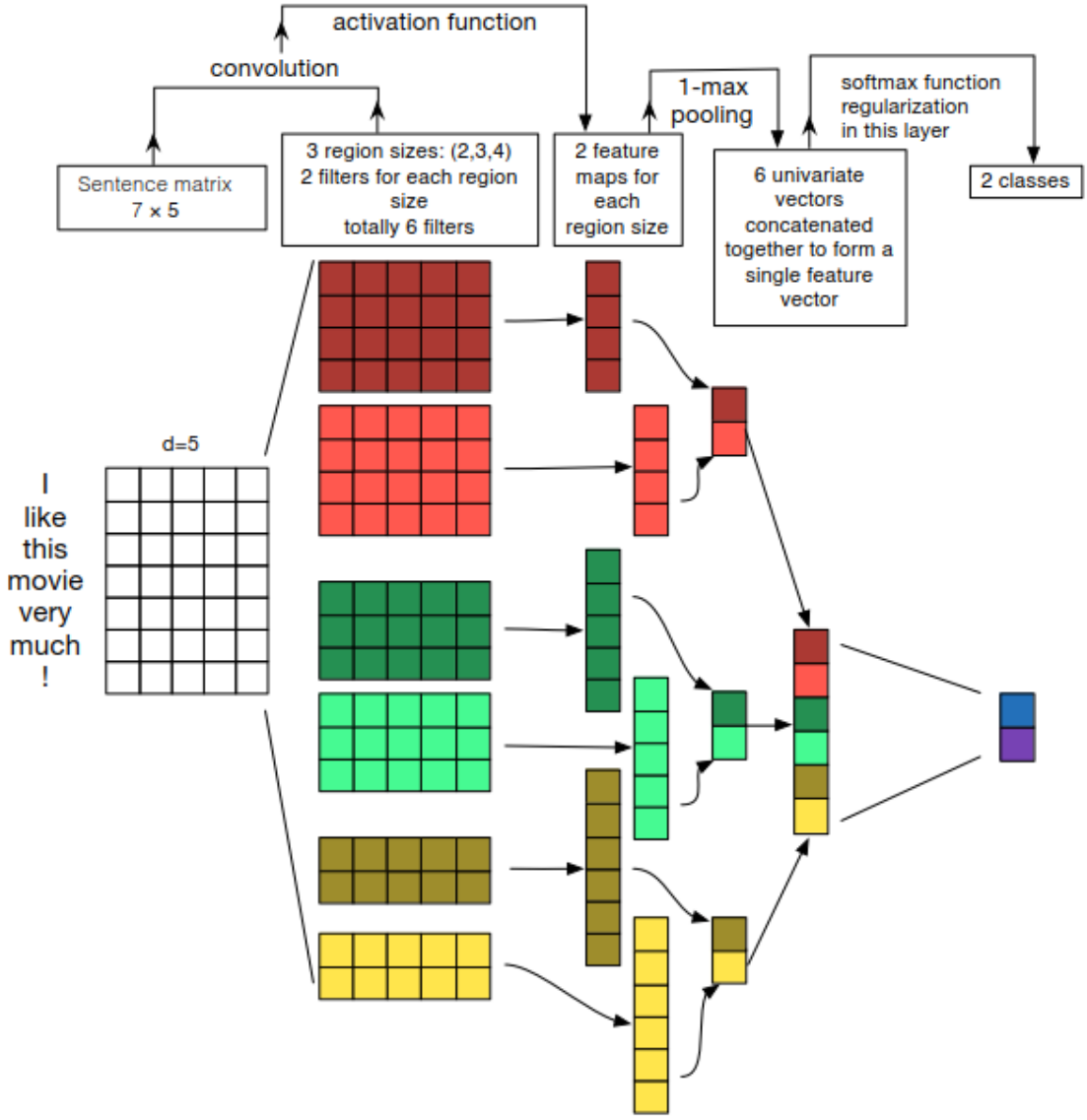
E2) Under some circumstances, we ***may*** *be able to redesign or reformulate our relationships or arrangements to address these challenges.*

Appendix B Definition of variables

Variable	Definition
BHR[-1,1]	Buy and hold raw return during the three-day window surrounding the 10-K filing dates, multiplied by 100.
BHMAR[-1,1]	Buy and hold market-adjusted return during the three-day window surrounding the 10-K filing dates, defined as the raw return adjusted for the same period CRSP value-weighted index return and multiplied by 100.
BHSAR[-1,1]	Buy and hold size-and-value adjusted return during the three-day window surrounding the 10-K filing dates, defined as the raw return minus the return on the portfolio of firms with approximately the same size and book-to-market (BM) ratio and multiplied by 100.
BHFAR[-1,1]	Buy and hold Carhart 4-factor-adjusted return during the three-day window surrounding the 10-K filing dates, multiplied by 100. The factor loadings are based on a regression of a four-factor model using trading days [−252, −6] relative to the 10-K filing date.
BHR[2,20]	Buy and hold raw return during the one-month window, multiplied by 100.
BHMAR[2,20]	Buy and hold market-adjusted return during the one-month window, defined as the raw return adjusted for the same period CRSP value-weighted index return and multiplied by 100.
BHSAR[2,20]	Buy and hold size-and-value adjusted return during the one-month window, defined as the raw return minus the return on the portfolio of firms with approximately the same size and book-to-market (BM) ratio and multiplied by 100.
BHFAR[2,20]	Buy and hold Carhart 4-factor-adjusted return during the one-month window, multiplied by 100. The factor loadings are based on a regression of a four-factor model using trading days [−252, −6] relative to the 10-K filing date.
BHR[2,60]	Buy and hold raw return during the one-quarter window, multiplied by 100.
BHMAR[2,60]	Buy and hold market-adjusted return during the one-quarter window, defined as the raw return adjusted for the same period CRSP value-weighted index return and multiplied by 100
BHSAR[2,60]	Buy and hold size-and-value adjusted return during the one-quarter window, defined as the raw return minus the return on the portfolio of firms with approximately the same size and book-to-market (BM) ratio and multiplied by 100.
BHFAR[2,60]	Buy and hold Carhart 4-factor-adjusted return during the one-quarter window, multiplied by 100. The factor loadings are based on a regression of a four-factor model using trading days [−252, −6] relative to the 10-K filing date.
BHR[2,125]	Buy and hold raw return during the half-year window, multiplied by 100.
BHMAR[2,125]	Buy and hold market-adjusted return during the half-year window, defined as the raw return adjusted for the same period CRSP value-weighted index return and multiplied by 100
BHSAR[2, 125]	Buy and hold size-and-value adjusted return during the half-year window, defined as the raw return minus the return on the portfolio of firms with approximately the same size and book-to-market (BM) ratio and multiplied by 100.
BHFAR[2,125]	Buy and hold Carhart 4-factor-adjusted return during the half-year window, multiplied by 100. The factor loadings are based on a regression of a four-factor model using trading days [−252, −6] relative to the 10-K filing date.

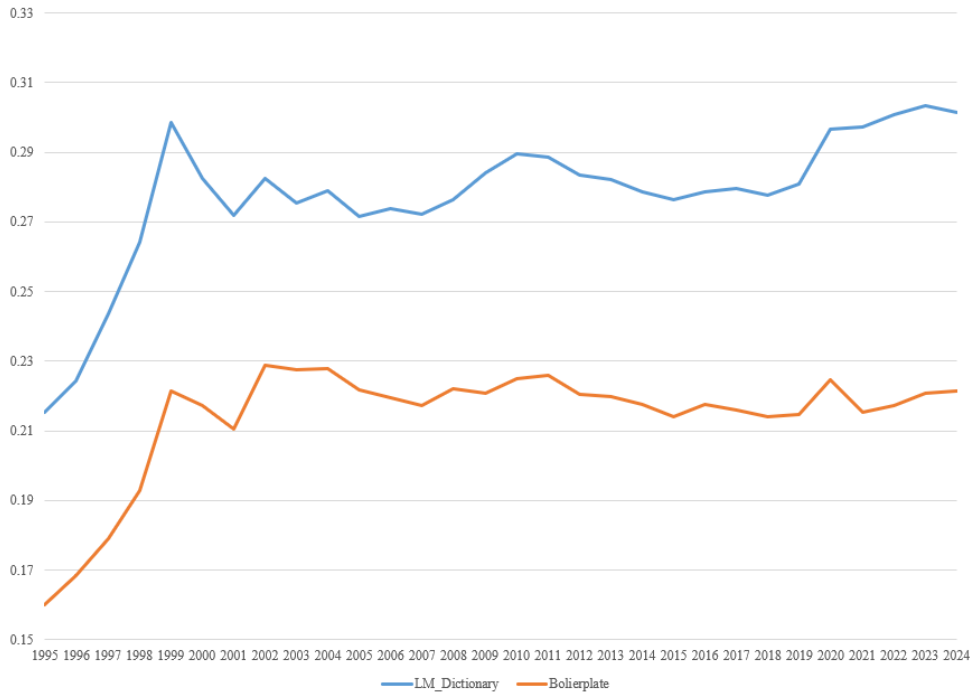
LM	The proportion of sentences in the MD&A that include the triggers which are summarized by Loughran and McDonald (2011) under the "uncertain" category.
CNN	The proportion of sentences that are classified as aleatory or epistemic uncertainty by CNN.
A	The proportion of sentences that are classified as aleatory uncertainty by CNN.
E	The proportion of sentences that are classified as epistemic uncertainty by CNN.
B	The proportion of sentences that are classified as boilerplate by CNN.
Length	The log of the number of sentence counts in the MD&A Section.
Tone	The difference between positive words and negative words, scaled by total sentiment words. We use the commonly used finance-oriented sentiment dictionary proposed by Loughran and McDonald (2011).
Fog	The Fog index.
LnMV	The natural logarithm of the market value of equity.
BTM	The ratio of the book value of equity to the market value of equity.
LnSeg	The natural logarithm of the number of business and geographic segments.
LnAF	The natural logarithm of the sum of one and the number of analysts following.
ROA	Net income (NI) over total assets (AT) at the beginning of current year.
Turnover	The number of shares traded for the firm in the trading days [−252, −6] relative to the 10-K filing date divided by the firm's shares outstanding at the 10-K filing date
PreAlpha	Carhart alpha based on a regression of their four-factor model using trading days [−252, −6] relative to the 10-K filing date. At least 60 observations of daily returns must be available to be included in the sample.
PastReturn	Buy and hold return during the period of [−252, −6] relative to the 10-K filing date. At least 60 observations of daily returns must be available to be included in the sample.
A_Sys	Systemic component of aleatory uncertainty, captureing industry–year–level fluctuations that affect all firms in the same industry and year. Formally, we run the following regressions: $A_{i,t} = \alpha_0$ +Industry-Year FEs+ $e_{i,t}$, where the fitted value is an estimate of the latent systemic component (*A_Sys*) and the residual is an estimate of the latent idiosyncratic component (*A_Idio*). Industry–year fixed effects are included to capture systemic fluctuations that affect firms in the same industry-year simultaneously. The fitted values from this regression capture the *systemic* component that co-moves across firms, whereas the residuals measure firm-level *idiosyncratic* deviations.
A_Idio	Idiosyncratic component of aleatory uncertainty.
Disp	Analyst forecast dispersion, measured using the last forecast from each analyst within the 30-day window preceding the 10-K filing date.
VIX	VIX index at the 10-K filing date.
ImplVol	Implied volatility at the 10 K filing date, constructed following Bae et al. (2023) at the firm level.
StdCFO	Standard deviation of the annual operating cash flow of the past seven years
Age	Firm age.

Figure 1 Illustration of CNN Models for Sentence Classification from Zhang and Wallace (2015)

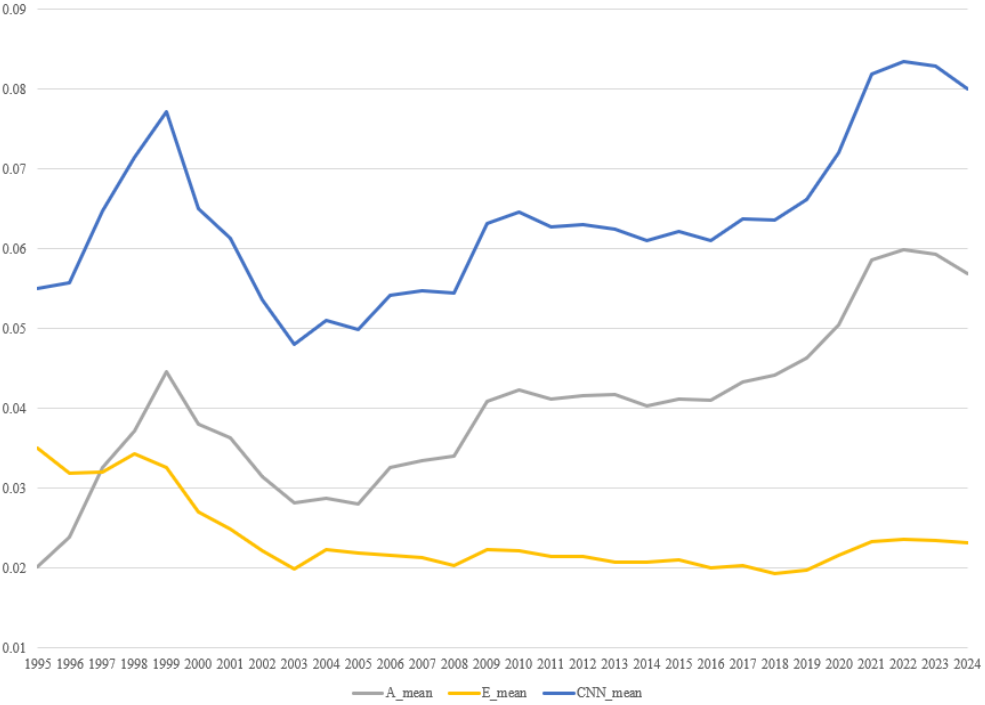


Notes. Figure 1 provides a simple example from Zhang and Wallace (2015) to illustrate the procedures in the CNN models. In this simple example, the customized sentence length n=7 and the dimensionality of the word vectors d=5. Next, 6 *filters* are used, with their widths equal to 5 and heights equal to 2, 3, and 4. Then the sentence matrix is convolved with 6 *filters* to produce 6 feature maps. As the dimensionality of the feature maps varies due to different heights of *filters*, we further apply *1-max pooling* (Boureau et al., 2010) to each feature map to induce a scalar. 6 scalars are generated and then concatenated together to form a single feature vector for the penultimate layer. Finally, we feed the feature vector through a *softmax function* to link the final classification, which has two classes.

Figure 2 Aggregate Uncertainty Measures over Time
Panel A: Results of *LM* and *B*



Panel B: Results of *CNN*, *A*, and *E*



Notes. Figure 2 plots the equal-weighted aggregate uncertainty (using different measures) within a year over time. Panel A shows the results of *LM* and *B* while panel B shows the results of *CNN*, *A*, and *E*. This figure shows that while boilerplate statements (*B*) exhibit a steady trend, the aggregate uncertainty statements (*CNN*) can capture time-varying aggregate uncertainty information associated with the macro environment.

Table 1 Uncertainty Classification Performance of CNN in the Validation Sample

Panel A: Confusion Matrix

		Researcher Label			
		A	**B**	**E**	**Sum**
CNN Predict	**A**	133	14	3	150
	B	28	712	17	757
	E	2	11	80	93
	Sum	163	737	100	

Panel B: Performance indices for each uncertainty category

	Precision	**Recall**	**Accuracy**	**F1 score**
A	88.67%	81.60%	95.30%	84.98%
B	94.06%	96.61%	93.00%	95.31%
E	86.02%	80.00%	96.70%	82.90%

Notes. This table presents our classification performance of the CNN in the validation sample. In panel A, we tabulate the confusion matrix. Each cell reports the number of sentences with the corresponding label of the CNN and of the researchers. In panel B, we report precision, recall, accuracy, and F1 score for each uncertainty category. Precision is the number of sentences correctly classified into a category scaled by the total number of sentences classified into that category. Recall is the number of sentences correctly classified into a category scaled by the number of sentences researchers label as that category. Accuracy is the number of sentences correctly classified scaled by the total number of sentences. F1 score is the harmonic mean of precision and recall, that is 2×Precision×Recall / (Precision+Recall).

Table 2 Variable Descriptive Statistics

Panel A: Summary statistics of variables in the regression models

Variable	N	Mean	SD	P25	Median	P75
BHR[-1,1]	56480	-0.17	7.50	-3.12	0.00	2.81
BHMAR[-1,1]	56480	-0.13	7.16	-2.93	-0.13	2.64
BHSAR[-1,1]	56480	-0.17	7.09	-2.95	-0.19	2.53
BHFAR[-1,1]	56480	-0.09	7.07	-2.80	-0.10	2.50
BHR[2,20]	56480	1.78	15.35	-5.27	1.15	7.92
BHMAR[2,20]	56480	0.60	13.98	-6.14	-0.02	6.25
BHSAR[2,20]	56480	0.60	13.58	-5.98	-0.06	6.08
BHFAR[2,20]	56480	0.77	13.53	-5.54	0.21	6.14
BHR[2,60]	56480	5.06	27.65	-8.69	3.21	16.43
BHMAR[2,60]	56480	1.97	25.52	-10.71	0.42	12.65
BHSAR[2,60]	56480	1.68	24.80	-10.91	0.44	12.16
BHFAR[2,60]	56480	2.34	24.73	-9.86	1.25	12.88
BHR[2,125]	56480	7.77	38.73	-11.59	5.62	23.96
BHMAR[2,125]	56480	2.89	35.84	-15.21	1.06	18.17
BHSAR[2,125]	56480	2.54	34.77	-15.15	1.02	17.64
BHFAR[2,125]	56480	3.89	35.57	-13.99	2.46	19.04
LM	56480	0.28	0.12	0.20	0.26	0.33
CNN	56480	0.07	0.06	0.03	0.05	0.08
A	56480	0.04	0.05	0.01	0.02	0.05
E	56480	0.02	0.02	0.01	0.02	0.03
Length	56480	4.86	0.57	4.85	5.02	5.14
Tone	56480	-0.19	0.27	-0.38	-0.22	-0.02
Fog	56480	22.81	2.02	21.45	22.62	23.97
ROA	56480	-0.05	0.27	-0.06	0.02	0.07
LnMV	56480	5.98	2.11	4.40	5.89	7.43
BTM	56480	0.72	0.34	0.45	0.73	0.97
LnSeg	56480	2.05	0.85	1.39	2.20	2.71
LnAF	56480	1.12	1.12	0.00	1.10	2.08
Turnover	56480	2.01	2.44	0.62	1.33	2.45
PreAlpha	56480	0.00	0.00	-0.00	0.00	0.00
PastReturn	56480	0.14	0.68	-0.25	0.04	0.35

Notes. This table presents the summary statistics of variables in the regression models. To calculate the summary statistics, we use the sample with non-missing variables for the OLS regressions. Variable definitions are provided in Appendix B.

Table 3 Correlation Matrix

Panel A: Correlations between the variables (1-16) in the OLS model

	1.	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*	*15.*	*16.*
1.BHR[-1,1]		**0.92**	**0.89**	**0.87**	**-0.01**	0.00	-0.01	**-0.01**	**-0.02**	0.00	0.00	-0.01	**-0.01**	0.00	0.01	0.00
2.BHMAR[-1,1]	**0.96**		**0.97**	**0.92**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	0.00	-0.01	-0.01	0.00	**0.01**	0.01	0.01	0.01
3.BHSAR[-1,1]	**0.94**	**0.99**		**0.93**	**-0.02**	**-0.02**	**-0.02**	**-0.02**	-0.01	**-0.01**	**-0.01**	-0.01	**0.01**	0.00	0.00	0.00
4.BHFAR[-1,1]	**0.92**	**0.96**	**0.96**		-0.01	**-0.01**	**-0.01**	**-0.01**	0.00	0.00	0.00	0.00	**0.01**	0.01	0.01	**0.01**
5.BHR[2,20]	**-0.02**	**-0.02**	**-0.03**	**-0.02**		**0.91**	**0.85**	**0.81**	**0.53**	**0.50**	**0.46**	**0.43**	**0.37**	**0.36**	**0.34**	**0.31**
6.BHMAR[2,20]	**-0.02**	**-0.02**	**-0.03**	**-0.02**	**0.94**		**0.96**	**0.88**	**0.47**	**0.52**	**0.50**	**0.45**	**0.34**	**0.38**	**0.36**	**0.32**
7.BHSAR[2,20]	**-0.02**	**-0.03**	**-0.03**	**-0.03**	**0.90**	**0.98**		**0.89**	**0.45**	**0.49**	**0.51**	**0.45**	**0.33**	**0.36**	**0.37**	**0.32**
8.BHFAR[2,20]	**-0.02**	**-0.03**	**-0.03**	**-0.02**	**0.85**	**0.91**	**0.92**		**0.43**	**0.46**	**0.46**	**0.51**	**0.32**	**0.34**	**0.34**	**0.37**
9.BHR[2,60]	**-0.04**	**-0.02**	**-0.02**	**-0.01**	**0.58**	**0.54**	**0.51**	**0.48**		**0.93**	**0.88**	**0.84**	**0.71**	**0.66**	**0.62**	**0.59**
10.BHMAR[2,60]	**-0.02**	**-0.02**	**-0.02**	**-0.01**	**0.56**	**0.57**	**0.55**	**0.51**	**0.96**		**0.96**	**0.89**	**0.65**	**0.70**	**0.68**	**0.62**
11.BHSAR[2,60]	**-0.02**	**-0.02**	**-0.02**	**-0.01**	**0.53**	**0.55**	**0.56**	**0.51**	**0.93**	**0.98**		**0.90**	**0.62**	**0.67**	**0.70**	**0.62**
12.BHFAR[2,60]	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**0.49**	**0.50**	**0.51**	**0.56**	**0.87**	**0.92**	**0.93**		**0.60**	**0.63**	**0.64**	**0.69**
13.BHR[2,125]	**-0.02**	0.00	-0.01	0.00	**0.42**	**0.39**	**0.38**	**0.36**	**0.75**	**0.71**	**0.69**	**0.65**		**0.93**	**0.87**	**0.83**
14.BHMAR[2,125]	**-0.01**	0.00	-0.01	0.00	**0.41**	**0.42**	**0.40**	**0.38**	**0.71**	**0.74**	**0.72**	**0.68**	**0.96**		**0.96**	**0.88**
15.BHSAR[2,125]	**-0.01**	-0.01	-0.01	0.00	**0.39**	**0.41**	**0.41**	**0.38**	**0.69**	**0.72**	**0.73**	**0.69**	**0.92**	**0.98**		**0.90**
16.BHFAR[2,125]	-0.01	0.00	-0.01	0.00	**0.36**	**0.37**	**0.37**	**0.40**	**0.64**	**0.67**	**0.67**	**0.73**	**0.87**	**0.92**	**0.93**	
17.LM	0.00	0.00	0.00	0.00	**-0.01**	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18.CNN	0.00	0.00	0.00	0.00	-0.01	0.00	0.01	0.01	0.00	0.01	**0.01**	**0.01**	0.00	0.00	0.01	**0.01**
19.A	0.00	0.00	0.01	0.01	**-0.01**	-0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.00	-0.01	0.00	0.00
20.E	**-0.01**	**-0.02**	**-0.02**	**-0.01**	0.00	0.00	**0.01**	**0.01**	**0.01**	**0.01**	**0.02**	**0.02**	**0.01**	**0.01**	**0.01**	**0.02**
21.Length	-0.01	-0.01	-0.01	**-0.01**	0.00	**-0.01**	**-0.01**	**-0.01**	0.00	0.00	0.00	**-0.01**	-0.01	-0.01	**-0.01**	**-0.02**
22.Tone	0.00	0.00	0.01	0.00	**-0.03**	**-0.02**	**-0.01**	**-0.01**	**-0.02**	**-0.02**	**-0.02**	**-0.01**	**-0.03**	**-0.03**	**-0.02**	**-0.02**
23.Fog	0.00	0.00	0.01	0.01	**-0.03**	**-0.02**	**-0.01**	**-0.01**	**-0.03**	**-0.03**	**-0.02**	**-0.03**	**-0.03**	**-0.04**	**-0.02**	**-0.03**
24.ROA	**0.07**	**0.07**	**0.07**	**0.07**	**0.01**	**0.02**	**0.02**	**0.02**	0.00	0.01	0.01	**0.02**	0.00	0.00	0.00	**0.01**
25.LnMV	**0.02**	**0.03**	**0.04**	**0.03**	**-0.08**	**-0.05**	**-0.04**	**-0.05**	**-0.09**	**-0.08**	**-0.07**	**-0.09**	**-0.10**	**-0.11**	**-0.09**	**-0.12**
26.BTM	**0.01**	**0.01**	**0.01**	**0.01**	**0.08**	**0.05**	**0.04**	**0.05**	**0.10**	**0.09**	**0.07**	**0.08**	**0.10**	**0.10**	**0.08**	**0.09**
27. LnSeg	0.01	0.01	**0.01**	0.01	0.01	**0.01**	0.01	0.00	0.00	**0.01**	0.00	**-0.01**	0.00	0.00	-0.01	**-0.02**
28. LnAF	**0.02**	**0.03**	**0.03**	**0.03**	**-0.04**	**-0.02**	**-0.02**	**-0.02**	**-0.04**	**-0.03**	**-0.02**	**-0.04**	**-0.03**	**-0.04**	**-0.03**	**-0.05**
29.Turnover	**-0.05**	**-0.05**	**-0.05**	**-0.05**	**-0.05**	**-0.05**	**-0.04**	**-0.05**	**-0.06**	**-0.06**	**-0.05**	**-0.07**	**-0.08**	**-0.09**	**-0.08**	**-0.09**
30.PreAlpha	0.00	0.00	0.00	0.00	0.01	**0.01**	0.00	**0.01**	**-0.01**	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00
31.PastReturn	**0.01**	0.00	0.00	0.00	**-0.07**	**-0.04**	**-0.04**	**-0.01**	**-0.11**	**-0.07**	**-0.06**	**-0.04**	**-0.12**	**-0.09**	**-0.06**	**-0.05**

Panel B: Correlations between the variables (17-31) in the return model, continued from Panel A

	17.	*18.*	*19.*	*20.*	*21.*	*22.*	*23.*	*24.*	*25.*	*26.*	*27.*	*28.*	*29.*	*30.*	*31.*
1.BHR[-1,1]	**-0.01**	**-0.01**	**-0.01**	**-0.02**	0.00	0.00	0.00	**0.06**	**0.03**	**0.01**	**0.01**	**0.03**	**-0.04**	**0.02**	**0.04**
2.BHMAR[-1,1]	**-0.01**	**-0.02**	**-0.01**	**-0.02**	0.00	0.00	0.00	**0.07**	**0.05**	**0.01**	**0.02**	**0.03**	**-0.03**	**0.02**	**0.04**
3.BHSAR[-1,1]	**-0.01**	**-0.01**	-0.01	**-0.02**	0.00	0.01	0.01	**0.07**	**0.06**	0.01	**0.02**	**0.04**	**-0.02**	**0.01**	**0.03**
4.BHFAR[-1,1]	-0.01	**-0.01**	-0.01	**-0.02**	0.00	0.00	0.01	**0.07**	**0.05**	**0.01**	**0.01**	**0.04**	**-0.03**	**0.02**	**0.02**
5.BHR[2,20]	**-0.01**	-0.01	0.00	-0.01	**0.01**	**-0.02**	**-0.03**	**0.02**	**-0.03**	**0.06**	**0.02**	**-0.01**	**-0.03**	**0.02**	**-0.05**
6.BHMAR[2,20]	**-0.01**	-0.01	-0.01	-0.01	0.01	**-0.01**	**-0.03**	**0.04**	0.00	**0.03**	**0.03**	0.00	**-0.03**	**0.02**	-0.01
7.BHSAR[2,20]	-0.01	0.00	0.00	0.00	0.00	-0.01	**-0.02**	**0.04**	0.00	**0.02**	**0.02**	0.01	**-0.02**	**0.02**	0.00
8.BHFAR[2,20]	0.00	0.01	0.01	0.00	**-0.01**	-0.01	**-0.02**	**0.04**	**-0.02**	**0.03**	0.01	0.00	**-0.05**	**0.02**	**0.01**
9.BHR[2,60]	-0.01	0.00	0.00	0.00	0.00	**-0.02**	**-0.03**	**0.01**	**-0.05**	**0.08**	0.01	**-0.02**	**-0.06**	0.01	**-0.10**
10.BHMAR[2,60]	0.00	0.01	0.00	0.00	0.01	**-0.02**	**-0.03**	**0.02**	**-0.04**	**0.06**	**0.03**	-0.01	**-0.05**	0.00	**-0.06**
11.BHSAR[2,60]	0.00	**0.01**	0.01	0.01	0.00	**-0.01**	**-0.02**	**0.02**	**-0.02**	**0.05**	**0.01**	0.00	**-0.04**	0.00	**-0.03**
12.BHFAR[2,60]	0.00	**0.02**	0.01	**0.01**	**-0.01**	**-0.01**	**-0.02**	**0.02**	**-0.05**	**0.06**	-0.01	**-0.02**	**-0.08**	0.01	**-0.02**
13.BHR[2,125]	-0.01	0.00	0.00	0.01	0.00	**-0.02**	**-0.03**	0.01	**-0.06**	**0.08**	0.00	**-0.01**	**-0.08**	**0.02**	**-0.11**
14.BHMAR[2,125]	-0.01	0.00	0.00	0.00	0.01	**-0.02**	**-0.04**	0.00	**-0.07**	**0.07**	**0.01**	**-0.02**	**-0.09**	**0.02**	**-0.07**
15.BHSAR[2,125]	0.00	**0.01**	0.00	0.01	0.00	**-0.02**	**-0.02**	0.00	**-0.05**	**0.06**	0.00	**-0.01**	**-0.07**	**0.02**	**-0.03**
16.BHFAR[2,125]	0.00	**0.01**	0.00	**0.01**	**-0.01**	**-0.02**	**-0.03**	**0.01**	**-0.08**	**0.08**	**-0.02**	**-0.04**	**-0.12**	0.00	**-0.03**
17.LM		**0.53**	**0.54**	**0.15**	**-0.10**	**-0.24**	**0.18**	**-0.06**	**-0.04**	**-0.02**	**-0.03**	0.00	**0.03**	-0.01	**-0.03**
18.CNN	**0.68**		**0.81**	**0.60**	**-0.02**	**0.02**	**0.04**	**0.01**	**-0.03**	**-0.07**	**-0.02**	**0.01**	**0.07**	0.00	**-0.03**
19.A	**0.69**	**0.93**		**0.13**	0.00	**-0.14**	**0.08**	0.01	**0.02**	**-0.02**	**0.02**	**0.05**	**0.06**	0.00	**-0.02**
20.E	**0.10**	**0.40**	**0.04**		0.00	**0.23**	**-0.04**	**-0.01**	**-0.14**	**-0.11**	**-0.08**	**-0.07**	**0.05**	**0.01**	**-0.03**
21.Length	**-0.18**	**0.02**	**0.01**	**0.04**		**-0.04**	**-0.59**	0.01	**0.09**	**0.02**	**0.15**	**0.07**	**0.05**	**-0.01**	0.01
22.Tone	**-0.24**	**-0.07**	**-0.18**	**0.25**	**-0.10**		**-0.05**	**0.14**	**0.11**	**-0.20**	0.01	**0.08**	**0.08**	**0.02**	**0.04**
23.Fog	**0.17**	**0.04**	**0.05**	**-0.03**	**-0.28**	**-0.04**		**-0.10**	**0.09**	**-0.01**	**-0.04**	**0.09**	**0.09**	**-0.04**	**-0.02**
24.ROA	**-0.04**	**0.02**	**0.03**	**-0.04**	**-0.03**	**0.05**	**-0.09**		**0.38**	**-0.16**	**0.21**	**0.18**	0.00	**0.14**	**0.28**
25.LnMV	0.00	**0.03**	**0.08**	**-0.12**	**0.03**	**0.12**	**0.09**	**0.29**		**-0.32**	**0.32**	**0.62**	**0.40**	**0.01**	**0.24**
26.BTM	0.00	**-0.03**	0.00	**-0.08**	-0.01	**-0.18**	**-0.02**	**0.11**	**-0.34**		**-0.02**	**-0.20**	**-0.28**	**-0.19**	**-0.24**
27. LnSeg	-0.01	**0.03**	**0.06**	**-0.06**	**0.08**	**0.02**	**-0.03**	**0.20**	**0.30**	**-0.03**		**0.21**	**0.11**	**-0.02**	**0.05**
28. LnAF	**0.02**	**0.05**	**0.07**	**-0.06**	**0.04**	**0.09**	**0.09**	**0.14**	**0.64**	**-0.21**	**0.20**		**0.35**	**-0.04**	**0.08**
29.Turnover	**0.02**	**0.04**	**0.02**	**0.04**	**0.05**	**0.03**	**0.06**	**-0.21**	**0.10**	**-0.12**	0.00	**0.15**		**-0.03**	**-0.02**
30.PreAlpha	-0.01	0.00	-0.01	**0.03**	**-0.01**	**0.01**	**-0.04**	**0.07**	**-0.01**	**-0.20**	**-0.02**	**-0.06**	**0.03**		**0.72**
31.PastReturn	**-0.01**	0.00	0.00	0.01	0.00	**0.03**	**-0.02**	**0.11**	**0.13**	**-0.25**	**0.02**	**0.02**	**0.04**	**0.73**	

Notes. This table presents Pearson (Spearman) correlations in the lower (upper) diagonal of the panels between the variables used in the OLS model. Correlations in bold are statistically significant at the 5 percent or 1 percent level. Variable definitions are provided in Appendix B.

Table 4 Uncertainty and Stock Returns
Panel A: Three-day abnormal returns

	BHR[-1,1]			*BHMAR[-1,1]*			*BHSAR[-1,1]*			*BHFAR[-1,1]*		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LM	-0.18			-0.12			-0.17			-0.01		
	(-0.65)			(-0.45)			(-0.63)			(-0.03)		
CNN		-0.68			-0.52			-0.51			-0.16	
		(-1.29)			(-1.03)			(-1.02)			(-0.33)	
A			-0.22			-0.04			-0.03			0.22
			(-0.37)			(-0.07)			(-0.06)			(0.40)
E			-3.64**			-3.55**			-3.52**			-2.88*
			(-2.11)			(-2.13)			(-2.14)			(-1.77)
Length	-0.06	-0.05	-0.04	-0.04	-0.04	-0.03	-0.04	-0.03	-0.02	-0.05	-0.05	-0.04
	(-1.16)	(-1.02)	(-0.79)	(-0.85)	(-0.75)	(-0.52)	(-0.84)	(-0.70)	(-0.46)	(-1.00)	(-1.02)	(-0.90)
Tone	0.08	0.08	0.15	0.02	0.02	0.09	-0.03	-0.02	0.05	0.00	-0.00	0.10
	(0.60)	(0.65)	(1.19)	(0.14)	(0.15)	(0.75)	(-0.23)	(-0.18)	(0.43)	(0.01)	(-0.02)	(0.82)
Fog	0.00	-0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
	(0.02)	(-0.01)	(0.04)	(0.37)	(0.35)	(0.41)	(0.41)	(0.37)	(0.43)	(0.58)	(0.58)	(1.00)
ROA	1.68***	1.69***	1.68***	1.72***	1.72***	1.72***	1.66***	1.66***	1.66***	1.55***	1.55***	1.50***
	(9.32)	(9.34)	(9.33)	(9.78)	(9.80)	(9.78)	(9.53)	(9.55)	(9.53)	(8.95)	(8.95)	(8.86)
LnMV	-0.03	-0.03	-0.03	0.04*	0.04*	0.03	0.04*	0.04*	0.04*	0.03	0.03	0.01
	(-1.09)	(-1.08)	(-1.29)	(1.67)	(1.67)	(1.42)	(1.93)	(1.93)	(1.67)	(1.42)	(1.42)	(0.59)
BTM	0.17	0.17	0.16	0.24*	0.24*	0.23*	0.15	0.15	0.14	0.25*	0.25*	0.20*
	(1.29)	(1.29)	(1.21)	(1.85)	(1.84)	(1.75)	(1.17)	(1.17)	(1.09)	(1.92)	(1.92)	(1.66)
LnSeg	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	-0.06
	(-0.67)	(-0.65)	(-0.73)	(-0.77)	(-0.76)	(-0.85)	(-0.61)	(-0.59)	(-0.68)	(-0.60)	(-0.60)	(-1.37)
LnAF	0.09**	0.09**	0.09**	0.09***	0.10***	0.10***	0.09***	0.09***	0.09***	0.10***	0.10***	0.11***
	(2.35)	(2.36)	(2.40)	(2.77)	(2.78)	(2.81)	(2.71)	(2.72)	(2.76)	(3.09)	(3.09)	(3.41)
Turnover	-0.15***	-0.15***	-0.15***	-0.13***	-0.13***	-0.13***	-0.13***	-0.13***	-0.12***	-0.12***	-0.12***	-0.11***
	(-7.58)	(-7.58)	(-7.56)	(-7.20)	(-7.20)	(-7.18)	(-6.82)	(-6.82)	(-6.80)	(-6.29)	(-6.29)	(-6.25)
PreAlpha	-8.51	-8.31	-8.17	6.52	6.67	6.81	-10.06	-9.92	-9.78	-3.17	-3.11	-3.75
	(-0.25)	(-0.24)	(-0.24)	(0.20)	(0.20)	(0.21)	(-0.31)	(-0.30)	(-0.30)	(-0.10)	(-0.09)	(-0.11)
PastReturn	0.07	0.07	0.07	0.01	0.01	0.01	0.04	0.04	0.04	0.05	0.05	0.05
	(0.70)	(0.69)	(0.69)	(0.13)	(0.13)	(0.12)	(0.36)	(0.36)	(0.35)	(0.51)	(0.51)	(0.53)
Observation	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.0205	0.0205	0.0206	0.0108	0.0108	0.0108	0.0090	0.0090	0.0090	0.0080	0.0080	0.0080

Panel B: One-month abnormal returns

	BHR[2,20]			*BHMAR[2,20]*			*BHSAR[2,20]*			*BHFAR[2,20]*		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LM	0.13			-0.03			0.12			0.29		
	(0.24)			(-0.06)			(0.25)			(0.60)		
CNN		1.46			1.67*			1.62*			1.98**	
		(1.41)			(1.74)			(1.71)			(2.12)	
A			1.85*			1.85*			1.99*			2.20**
			(1.65)			(1.77)			(1.93)			(2.16)
E			-0.70			0.85			-0.49			1.03
			(-0.22)			(0.28)			(-0.16)			(0.34)
Length	0.02	0.01	0.02	-0.01	-0.02	-0.02	0.00	-0.01	0.00	-0.07	-0.09	-0.09
	(0.17)	(0.09)	(0.16)	(-0.14)	(-0.18)	(-0.15)	(0.03)	(-0.06)	(0.01)	(-0.74)	(-0.95)	(-0.92)
Tone	-0.31	-0.29	-0.23	-0.28	-0.24	-0.22	-0.47**	-0.44**	-0.39	-0.30	-0.28	-0.26
	(-1.24)	(-1.21)	(-0.91)	(-1.22)	(-1.08)	(-0.92)	(-2.01)	(-1.97)	(-1.63)	(-1.29)	(-1.27)	(-1.09)
Fog	0.06*	0.06*	0.06*	0.04	0.04	0.04	0.05*	0.05*	0.05*	0.01	0.01	0.01
	(1.80)	(1.81)	(1.83)	(1.43)	(1.42)	(1.43)	(1.65)	(1.66)	(1.68)	(0.43)	(0.46)	(0.47)
ROA	0.47	0.45	0.45	0.72**	0.70**	0.70**	0.61*	0.59*	0.59*	1.10***	1.08***	1.08***
	(1.28)	(1.24)	(1.24)	(2.11)	(2.06)	(2.06)	(1.73)	(1.70)	(1.69)	(3.11)	(3.07)	(3.06)
LnMV	-0.28***	-0.28***	-0.28***	-0.19***	-0.19***	-0.19***	-0.23***	-0.23***	-0.24***	-0.37***	-0.37***	-0.37***
	(-6.20)	(-6.18)	(-6.21)	(-4.65)	(-4.62)	(-4.60)	(-5.30)	(-5.31)	(-5.34)	(-8.48)	(-8.49)	(-8.45)
BTM	1.04***	1.05***	1.04***	1.03***	1.05***	1.05***	0.92***	0.93***	0.92***	0.96***	0.96***	0.96***
	(4.39)	(4.44)	(4.37)	(4.62)	(4.69)	(4.65)	(3.65)	(3.67)	(3.63)	(3.73)	(3.75)	(3.73)
LnSeg	0.40***	0.40***	0.40***	0.41***	0.41***	0.41***	0.26***	0.26***	0.25***	0.20**	0.20**	0.20**
	(5.15)	(5.08)	(5.08)	(5.54)	(5.46)	(5.46)	(3.11)	(3.10)	(3.06)	(2.51)	(2.48)	(2.47)
LnAF	0.42***	0.42***	0.42***	0.46***	0.45***	0.45***	0.42***	0.42***	0.42***	0.40***	0.40***	0.40***
	(5.75)	(5.73)	(5.75)	(6.67)	(6.64)	(6.65)	(6.26)	(6.25)	(6.26)	(5.96)	(5.94)	(5.95)
Turnover	-0.13***	-0.13***	-0.13***	-0.15***	-0.15***	-0.15***	-0.17***	-0.17***	-0.17***	-0.22***	-0.22***	-0.22***
	(-3.57)	(-3.58)	(-3.55)	(-4.34)	(-4.36)	(-4.34)	(-4.71)	(-4.70)	(-4.69)	(-6.24)	(-6.23)	(-6.22)
PreAlpha	540.45***	540.23***	540.14***	441.90***	441.64***	441.58***	349.96***	349.44***	349.51***	123.42*	122.80*	122.81*
	(7.56)	(7.56)	(7.56)	(6.53)	(6.53)	(6.53)	(5.24)	(5.23)	(5.23)	(1.84)	(1.83)	(1.83)
PastReturn	-2.32***	-2.32***	-2.32***	-1.77***	-1.77***	-1.77***	-1.41***	-1.41***	-1.41***	-0.17	-0.17	-0.17
	(-10.74)	(-10.74)	(-10.74)	(-8.68)	(-8.67)	(-8.67)	(-7.03)	(-7.02)	(-7.03)	(-0.83)	(-0.82)	(-0.82)
Observation	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.0874	0.0875	0.0875	0.0294	0.0294	0.0294	0.0140	0.0140	0.0140	0.0123	0.0123	0.0123

Panel C: One-quarter abnormal returns

	BHR[2,60]			*BHMAR[2,60]*			*BHSAR[2,60]*			*BHFAR[2,60]*		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LM	0.78			0.46			0.83			1.04		
	(0.83)			(0.50)			(0.94)			(1.18)		
CNN		5.59***			5.20***			5.94***			6.45***	
		(3.13)			(3.06)			(3.57)			(3.88)	
A			5.62***			5.30***			5.87***			6.05***
			(2.89)			(2.87)			(3.23)			(3.36)
E			4.54			3.75			5.58			8.18
			(0.77)			(0.66)			(0.99)			(1.46)
Length	0.09	0.04	0.05	0.00	-0.03	-0.03	-0.01	-0.06	-0.06	-0.21	-0.27	-0.28
	(0.47)	(0.24)	(0.24)	(0.03)	(-0.16)	(-0.14)	(-0.03)	(-0.32)	(-0.33)	(-1.19)	(-1.59)	(-1.63)
Tone	-0.10	-0.04	-0.02	-0.03	0.05	0.08	-0.23	-0.17	-0.16	-0.02	0.03	-0.02
	(-0.22)	(-0.10)	(-0.04)	(-0.07)	(0.12)	(0.19)	(-0.53)	(-0.41)	(-0.38)	(-0.04)	(0.07)	(-0.04)
Fog	0.01	0.02	0.02	0.01	0.01	0.01	-0.02	-0.01	-0.01	-0.10*	-0.09*	-0.10*
	(0.23)	(0.26)	(0.27)	(0.24)	(0.25)	(0.26)	(-0.27)	(-0.23)	(-0.23)	(-1.74)	(-1.69)	(-1.70)
ROA	0.70	0.66	0.66	0.73	0.69	0.69	0.67	0.63	0.63	1.77***	1.73***	1.73***
	(1.02)	(0.96)	(0.96)	(1.09)	(1.03)	(1.03)	(1.03)	(0.96)	(0.96)	(2.72)	(2.65)	(2.66)
LnMV	-0.79***	-0.79***	-0.79***	-0.80***	-0.80***	-0.80***	-0.77***	-0.78***	-0.78***	-1.01***	-1.01***	-1.01***
	(-9.15)	(-9.18)	(-9.11)	(-9.66)	(-9.68)	(-9.62)	(-9.44)	(-9.46)	(-9.37)	(-12.49)	(-12.51)	(-12.36)
BTM	3.40***	3.42***	3.41***	3.56***	3.58***	3.57***	3.16***	3.18***	3.17***	3.07***	3.09***	3.09***
	(7.17)	(7.20)	(7.18)	(7.74)	(7.78)	(7.75)	(6.95)	(6.98)	(6.97)	(6.67)	(6.71)	(6.71)
LnSeg	0.29*	0.28*	0.28*	0.28*	0.27*	0.27*	0.28*	0.27*	0.27*	0.08	0.08	0.08
	(1.84)	(1.81)	(1.80)	(1.82)	(1.79)	(1.78)	(1.85)	(1.82)	(1.81)	(0.56)	(0.52)	(0.54)
LnAF	1.20***	1.20***	1.20***	1.21***	1.21***	1.21***	1.18***	1.17***	1.17***	1.03***	1.03***	1.03***
	(9.07)	(9.05)	(9.05)	(9.56)	(9.54)	(9.54)	(9.35)	(9.33)	(9.33)	(8.33)	(8.31)	(8.30)
Turnover	-0.51***	-0.51***	-0.51***	-0.55***	-0.55***	-0.55***	-0.49***	-0.49***	-0.49***	-0.60***	-0.60***	-0.60***
	(-7.17)	(-7.16)	(-7.16)	(-8.05)	(-8.04)	(-8.03)	(-7.31)	(-7.30)	(-7.29)	(-8.99)	(-8.98)	(-8.98)
PreAlpha	835.77***	834.03***	834.07***	589.15***	587.49***	587.55***	477.24***	475.39***	475.39***	173.55	171.57	171.48
	(6.65)	(6.64)	(6.64)	(4.84)	(4.83)	(4.83)	(3.98)	(3.96)	(3.96)	(1.41)	(1.39)	(1.39)
PastReturn	-4.00***	-3.99***	-3.99***	-2.71***	-2.71***	-2.71***	-2.20***	-2.19***	-2.19***	-0.47	-0.46	-0.46
	(-10.44)	(-10.43)	(-10.43)	(-7.36)	(-7.35)	(-7.35)	(-6.07)	(-6.06)	(-6.06)	(-1.27)	(-1.25)	(-1.25)
Observation	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.1393	0.1395	0.1394	0.0611	0.0613	0.0612	0.0327	0.0328	0.0328	0.0314	0.0316	0.0316

Panel D: Half-year abnormal returns

	BHR[2,125]			*BHMAR[2,125]*			*BHSAR[2,125]*			*BHFAR[2,125]*		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LM	0.76			-0.16			-0.06			0.27		
	(0.59)			(-0.13)			(-0.05)			(0.22)		
CNN		7.36***			5.72**			6.68***			7.88***	
		(3.05)			(2.43)			(2.89)			(3.35)	
A			7.48***			6.09**			6.96***			7.65***
			(2.85)			(2.38)			(2.78)			(3.00)
E			5.72			2.77			4.36			8.39
			(0.70)			(0.35)			(0.55)			(1.04)
Length	-0.23	-0.29	-0.28	-0.31	-0.32	-0.31	-0.35	-0.37	-0.36	-0.52**	-0.55**	-0.55**
	(-0.89)	(-1.13)	(-1.11)	(-1.21)	(-1.29)	(-1.24)	(-1.36)	(-1.47)	(-1.43)	(-2.02)	(-2.22)	(-2.22)
Tone	-0.35	-0.25	-0.21	-0.42	-0.25	-0.17	-0.84	-0.66	-0.60	-0.41	-0.24	-0.25
	(-0.58)	(-0.43)	(-0.34)	(-0.69)	(-0.43)	(-0.29)	(-1.43)	(-1.17)	(-1.01)	(-0.68)	(-0.41)	(-0.41)
Fog	-0.13	-0.13	-0.13	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.20**	-0.20**	-0.20**
	(-1.60)	(-1.57)	(-1.56)	(-1.07)	(-1.10)	(-1.08)	(-1.09)	(-1.11)	(-1.10)	(-2.52)	(-2.52)	(-2.52)
ROA	1.93**	1.88**	1.88**	1.63*	1.58*	1.58*	1.11	1.06	1.06	2.46***	2.41***	2.41***
	(2.06)	(2.01)	(2.00)	(1.78)	(1.73)	(1.73)	(1.23)	(1.17)	(1.17)	(2.68)	(2.62)	(2.62)
LnMV	-1.45***	-1.45***	-1.45***	-1.51***	-1.51***	-1.52***	-1.48***	-1.48***	-1.49***	-1.85***	-1.85***	-1.85***
	(-12.27)	(-12.29)	(-12.21)	(-13.18)	(-13.20)	(-13.14)	(-13.00)	(-13.02)	(-12.95)	(-15.78)	(-15.80)	(-15.68)
BTM	5.55***	5.57***	5.56***	5.63***	5.66***	5.65***	4.93***	4.96***	4.95***	5.62***	5.65***	5.65***
	(8.29)	(8.33)	(8.31)	(8.64)	(8.68)	(8.65)	(7.69)	(7.74)	(7.71)	(8.51)	(8.56)	(8.55)
LnSeg	0.22	0.22	0.22	0.09	0.09	0.09	0.21	0.21	0.20	0.22	0.22	0.22
	(1.06)	(1.03)	(1.02)	(0.44)	(0.43)	(0.41)	(1.03)	(1.01)	(1.00)	(1.08)	(1.06)	(1.06)
LnAF	2.12***	2.12***	2.12***	2.11***	2.11***	2.11***	2.03***	2.02***	2.03***	1.79***	1.79***	1.79***
	(11.68)	(11.66)	(11.66)	(11.95)	(11.93)	(11.94)	(11.53)	(11.51)	(11.51)	(10.01)	(9.98)	(9.98)
Turnover	-1.07***	-1.06***	-1.06***	-1.06***	-1.06***	-1.06***	-0.98***	-0.97***	-0.97***	-1.11***	-1.11***	-1.11***
	(-10.62)	(-10.62)	(-10.61)	(-10.94)	(-10.94)	(-10.93)	(-10.21)	(-10.20)	(-10.19)	(-11.45)	(-11.44)	(-11.44)
PreAlpha	1,695.39***	1,693.07***	1,693.11***	1,226.88***	1,224.95***	1,225.06***	1,007.38***	1,005.14***	1,005.21***	501.80***	499.22***	499.18***
	(9.80)	(9.79)	(9.79)	(7.26)	(7.25)	(7.25)	(6.01)	(6.00)	(6.00)	(2.86)	(2.84)	(2.84)
PastReturn	-6.99***	-6.98***	-6.98***	-4.81***	-4.80***	-4.81***	-3.51***	-3.51***	-3.51***	-1.75***	-1.74***	-1.74***
	(-13.11)	(-13.10)	(-13.10)	(-9.30)	(-9.29)	(-9.29)	(-6.88)	(-6.87)	(-6.87)	(-3.26)	(-3.25)	(-3.25)
Observation	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478	56,478
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.1675	0.1676	0.1676	0.0802	0.0803	0.0802	0.0386	0.0387	0.0387	0.0434	0.0435	0.0435

Notes. This table reports the regression results testing the relation between uncertainty measures and stock returns. Panel A shows the regression results for three-day buy and hold return. Panel B shows the regression results for one-month buy and hold return. Panel C shows the regression results for one-quarter buy and hold return. Panel D shows the regression results for half-year

buy and hold return. See Appendix B for variable definitions. We control for industry and year fixed effects. The *t*-statistics are based on standard errors that are clustered by firm. ***, **, * indicate significantly different from zero at the 1%, 5%, 10% level, respectively.

Table 5 Portfolios Formed Based on Uncertainty Measures.

Panel A: Equal-weighted portfolios three-day abnormal returns

	BHR[-1,1]			*BHMAR[-1,1]*			*BHSAR[-1,1]*			*BHFAR[-1,1]*		
Deciles	CNN	A	E	CNN	A	E	CNN	A	E	CNN	A	E
Low	0.06	-0.09	0.05	0.11	-0.09	0.11	0.06	-0.13**	0.06	0.09	-0.06	0.11
	(0.34)	(-0.55)	(0.34)	(1.38)	(-1.19)	(1.55)	(0.98)	(-2.64)	(0.94)	(1.40)	(-1.11)	(1.63)
2	-0.15	-0.10	-0.08	-0.13	-0.03	-0.02	-0.16**	-0.08	-0.06	-0.09	-0.03	0.00
	(-0.87)	(-0.57)	(-0.46)	(-1.38)	(-0.33)	(-0.24)	(-2.37)	(-1.02)	(-0.85)	(-1.10)	(-0.33)	(0.01)
3	-0.20	-0.25	-0.20	-0.14	-0.18	-0.16	-0.17*	-0.21**	-0.18**	-0.12	-0.15	-0.11
	(-1.11)	(-1.34)	(-1.13)	(-1.32)	(-1.49)	(-1.62)	(-1.83)	(-2.08)	(-2.41)	(-1.29)	(-1.44)	(-1.28)
4	-0.28	-0.17	-0.21	-0.25*	-0.14	-0.19*	-0.29**	-0.18	-0.21**	-0.18*	-0.11	-0.14
	(-1.47)	(-0.83)	(-1.09)	(-1.97)	(-1.10)	(-1.74)	(-2.73)	(-1.62)	(-2.19)	(-1.73)	(-0.97)	(-1.68)
High	-0.26	-0.22	-0.39*	-0.21*	-0.17*	-0.37***	-0.24**	-0.19*	-0.41***	-0.16*	-0.09	-0.31***
	(-1.31)	(-1.20)	(-2.03)	(-1.97)	(-1.71)	(-3.14)	(-2.37)	(-2.04)	(-4.25)	(-1.91)	(-1.13)	(-3.64)
Hight-Low	-0.32**	-0.13	-0.44***	-0.32**	-0.08	-0.48***	-0.29***	-0.06	-0.47***	-0.25***	-0.03	-0.42***
	(-2.41)	(-1.12)	(-3.30)	(-2.47)	(-0.69)	(-3.72)	(-2.61)	(-0.62)	(-4.03)	(-2.32)	(-0.26)	(-3.97)

Panel B: Equal-weighted portfolios one-month abnormal returns

	BHR[2,20]			*BHMAR[2,20]*			*BHSAR[2,20]*			*BHFAR[2,20]*		
Deciles	CNN	A	E	CNN	A	E	CNN	A	E	CNN	A	E
Low	1.36	1.46*	1.53*	0.32	0.43	0.49	0.34	0.48*	0.48*	0.53**	0.65***	0.68***
	(1.70)	(1.83)	(1.90)	(0.72)	(1.05)	(1.13)	(1.38)	(2.01)	(1.97)	(2.71)	(3.19)	(3.22)
2	1.44*	1.45*	1.56*	0.37	0.37	0.46	0.44*	0.38	0.52**	0.59**	0.60***	0.75***
	(1.86)	(1.86)	(1.90)	(0.94)	(0.92)	(1.08)	(1.77)	(1.64)	(2.15)	(2.49)	(2.77)	(3.50)
3	1.78**	1.60*	1.75**	0.68	0.48	0.64	0.74**	0.56*	0.68**	1.02***	0.76***	0.92***
	(2.24)	(1.91)	(2.16)	(1.66)	(1.12)	(1.55)	(2.62)	(1.96)	(2.46)	(3.81)	(2.90)	(3.34)
4	1.79**	1.80**	1.85**	0.59	0.62	0.72	0.65*	0.68**	0.82**	0.88***	0.93***	0.94***
	(2.16)	(2.31)	(2.25)	(1.32)	(1.45)	(1.65)	(2.03)	(2.42)	(2.61)	(3.08)	(3.39)	(3.27)
High	1.91**	1.95**	1.58*	0.83*	0.86*	0.48	0.95***	0.94***	0.60*	1.12***	1.13***	0.84***
	(2.15)	(2.19)	(1.91)	(1.73)	(1.91)	(1.04)	(2.96)	(3.08)	(2.00)	(4.03)	(4.20)	(3.30)
Hight-Low	0.54**	0.48*	0.05	0.51*	0.42	-0.00	0.61**	0.45*	0.12	0.58***	0.47**	0.16
	(1.97)	(1.84)	(0.14)	(1.74)	(1.60)	(-0.02)	(2.39)	(1.88)	(0.52)	(3.39)	(2.42)	(0.95)

Panel C: Equal-weighted portfolios one-quarter abnormal returns

	BHR[2,60]			*BHMAR[2,60]*			*BHSAR[2,60]*			*BHFAR[2,60]*		
Deciles	CNN	A	E	CNN	A	E	CNN	A	E	CNN	A	E
Low	4.16**	4.41**	4.44***	1.10	1.37	1.31	0.87	1.20*	1.09**	1.66***	1.97***	1.93***
	(2.67)	(2.69)	(3.00)	(1.25)	(1.47)	(1.60)	(1.62)	(2.03)	(2.17)	(3.19)	(3.46)	(3.89)
2	4.66***	4.57**	4.77***	1.57	1.53	1.63*	1.38**	1.20*	1.44**	2.18***	2.04***	2.30***
	(2.84)	(2.67)	(2.92)	(1.65)	(1.51)	(1.77)	(2.29)	(1.85)	(2.61)	(3.79)	(3.11)	(4.48)
3	4.93***	5.07***	5.17***	1.75*	1.90*	2.04*	1.57**	1.72**	1.79**	2.43***	2.48***	2.60***
	(2.83)	(2.85)	(2.89)	(1.75)	(1.88)	(1.91)	(2.47)	(2.62)	(2.61)	(3.93)	(3.97)	(3.82)
4	5.65***	5.27***	5.30***	2.40**	2.08**	2.15*	2.15***	1.87***	1.99**	2.88***	2.59***	2.58***
	(3.01)	(3.03)	(2.85)	(2.11)	(2.06)	(1.87)	(2.77)	(2.84)	(2.50)	(3.85)	(4.00)	(3.49)
High	5.56***	5.49***	5.26***	2.41**	2.29**	2.09*	2.27***	2.12***	1.93**	2.96***	2.84***	2.71***
	(3.05)	(3.06)	(2.77)	(2.18)	(2.16)	(1.81)	(2.92)	(3.02)	(2.36)	(4.19)	(4.43)	(3.65)
Hight-Low	1.40***	1.08***	0.82	1.31**	0.92**	0.78	1.40***	0.92***	0.84	1.30***	0.86**	0.77*
	(2.98)	(3.05)	(1.20)	(2.55)	(2.44)	(1.11)	(2.96)	(2.73)	(1.46)	(3.27)	(2.43)	(1.68)

Panel D: Equal-weighted portfolios half-year abnormal returns

	BHR[2,125]			*BHMAR[2, 125]*			*BHSAR[2, 125]*			*BHFAR[2, 125]*		
Deciles	CNN	A	E	CNN	A	E	CNN	A	E	CNN	A	E
Low	6.57**	7.27***	6.79***	1.70	2.33	1.84	1.58*	2.23**	1.74**	3.15***	3.80***	3.33***
	(2.74)	(2.86)	(3.00)	(1.19)	(1.55)	(1.41)	(1.87)	(2.41)	(2.26)	(3.66)	(3.93)	(4.14)
2	7.42***	7.02**	7.70***	2.39	2.21	2.65*	2.28**	1.92*	2.58***	4.03***	3.65***	4.14***
	(2.91)	(2.65)	(3.00)	(1.62)	(1.39)	(1.73)	(2.65)	(2.05)	(2.96)	(4.23)	(3.57)	(4.59)
3	8.00***	7.72***	8.04***	2.83*	2.58	2.94*	2.65***	2.45***	2.75**	4.14***	3.92***	4.41***
	(3.03)	(2.94)	(2.91)	(1.78)	(1.67)	(1.75)	(2.91)	(2.77)	(2.75)	(4.23)	(4.00)	(4.02)
4	8.57***	8.21***	8.27***	3.30*	3.01*	3.03*	3.06**	2.88***	2.87**	4.55***	4.28***	4.25***
	(3.01)	(3.15)	(2.96)	(1.88)	(1.90)	(1.73)	(2.75)	(3.18)	(2.61)	(3.72)	(4.24)	(3.56)
High	8.43***	8.35***	8.18***	3.24*	3.14*	3.00*	3.20***	3.06***	2.83**	4.58***	4.42***	4.31***
	(3.05)	(2.97)	(2.90)	(1.90)	(1.84)	(1.72)	(2.91)	(2.81)	(2.52)	(4.05)	(3.93)	(3.68)
Hight-Low	1.86**	1.08*	1.39	1.54*	0.81	1.16	1.62**	0.82	1.09	1.43**	0.62	0.98
	(2.44)	(1.66)	(1.38)	(1.92)	(1.19)	(1.12)	(2.38)	(1.39)	(1.29)	(2.05)	(0.92)	(1.21)

Notes. This table reports the average three-day, one month, one-quarter and half-year (raw and risk-adjusted) buy and hold returns for portfolios formed based on uncertainty measures. The t-statistics are reported in parentheses, and are calculated based on the time-series of the portfolio (raw and risk-adjusted) stock returns.

Table 6 Uncertainty and Stock Returns: Fama-Macbeth Regression

Panel A: Three-day abnormal returns

	BHR[-1,1]		*BHMAR[-1,1]*		*BHSAR[-1,1]*		*BHFAR[-1,1]*	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_CNN	-0.13		-0.11		-0.11		-0.08	
	(-1.08)		(-0.98)		(-1.13)		(-0.73)	
D_A		-0.02		0.03		0.04		0.05
		(-0.14)		(0.33)		(0.40)		(0.48)
D_E		-0.18*		-0.21**		-0.22**		-0.19**
		(-2.01)		(-2.52)		(-2.72)		(-2.52)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	56,480	56,480	56,480	56,480	56,480	56,480	56,480	56,480
# of Groups	30	30	30	30	30	30	30	30
Adj. R^2	0.0147	0.0147	0.0129	0.0129	0.0105	0.0106	0.0087	0.0087

Panel B: One month buy and hold adjusted returns

	BHR[2,20]		*BHMAR[2,20]*		*BHSAR[2,20]*		*BHFAR[2,20]*	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_CNN	0.35**		0.34**		0.41**		0.52***	
	(2.77)		(2.43)		(2.62)		(4.07)	
D_A		0.49**		0.37**		0.39***		0.47***
		(2.43)		(2.53)		(2.95)		(4.07)
D_E		-0.21*		-0.09		-0.02		0.08
		(-1.95)		(-0.73)		(-0.13)		(0.74)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	56,480	56,480	56,480	56,480	56,480	56,480	56,480	56,480
# of Groups	30	30	30	30	30	30	30	30
Adj. R^2	0.0483	0.0482	0.0433	0.0432	0.0330	0.0328	0.0236	0.0235

Panel C: One quarter buy and hold adjusted returns

	BHR[2,60]		*BHMAR[2,60]*		*BHSAR[2,60]*		*BHFAR[2,60]*	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_CNN	1.34***		1.27***		1.28***		1.42***	
	(5.05)		(4.74)		(4.64)		(4.86)	
D_A		1.21***		1.09***		1.05***		1.10***
		(5.02)		(4.55)		(4.44)		(4.58)
D_E		0.31		0.37		0.43		0.52
		(1.15)		(1.48)		(1.60)		(1.60)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	56,480	56,480	56,480	56,480	56,480	56,480	56,480	56,480
# of Groups	30	30	30	30	30	30	30	30
Adj. R^2	0.0585	0.0583	0.0542	0.0540	0.0438	0.0435	0.0391	0.0388

Panel D: Half year buy and hold adjusted returns

	BHR[2,125]		*BHMAR[2,125]*		*BHSAR[2,125]*		*BHFAR[2,125]*	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_CNN	1.68***		1.43**		1.38**		1.55**	
	(3.05)		(2.50)		(2.43)		(2.68)	
D_A		1.58***		1.34***		1.28***		1.23**
		(3.34)		(3.05)		(2.96)		(2.56)
D_E		0.27		0.18		0.14		0.39
		(0.47)		(0.29)		(0.22)		(0.64)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	56,480	56,480	56,480	56,480	56,480	56,480	56,480	56,480
# of Groups	30	30	30	30	30	30	30	30
Adj. R^2	0.0709	0.0708	0.0646	0.0645	0.0529	0.0528	0.0599	0.0597

Notes. This table reports the regression results testing the relation between uncertainty disclosures and stock returns using Fama-Macbeth regressions. Standard errors are from a Fama-MacBeth estimation with Newey-West correction for up to six lags. The t-statistics are reported in parentheses. ***, **, * indicate significantly different from zero at the 1%, 5%, 10% level, respectively.

Table 7 Interaction between Ambiguity and Risk

Panel A: Three-day abnormal returns

	BHR[-1,1]		*BHMAR[-1,1]*		*BHSAR[-1,1]*		*BHFAR[-1,1]*	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low A.	High A.	Low A.	High A.	Low A.	High A.	Low A.	High A.
A	-1.10	-0.08	2.11	0.18	1.15	0.11	1.46	0.21
	(-0.18)	(-0.10)	(0.36)	(0.25)	(0.20)	(0.17)	(0.25)	(0.30)
E	-2.12	-5.02**	-1.81	-5.19**	-1.71	-5.26**	-1.13	-4.22*
	(-0.91)	(-2.02)	(-0.80)	(-2.16)	(-0.77)	(-2.22)	(-0.51)	(-1.78)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	28,286	28,191	28,286	28,191	28,286	28,191	28,286	28,191
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.0228	0.0202	0.0134	0.0102	0.0109	0.0089	0.0103	0.0072

Panel B: One month buy and hold adjusted returns

	BHR[2,20]		*BHMAR[2,20]*		*BHSAR[2,20]*		*BHFAR[2,20]*	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low E.	High E.	Low E.	High E.	Low E.	High E.	Low E.	High E.
A	3.95***	-1.58	3.35**	-1.05	3.73***	0.12	3.92***	0.46
	(2.65)	(-0.94)	(2.43)	(-0.67)	(2.74)	(0.08)	(2.84)	(0.31)
E	-13.12	-0.78	-7.89	1.61	-1.43	2.26	3.94	2.87
	(-0.89)	(-0.15)	(-0.57)	(0.32)	(-0.11)	(0.46)	(0.29)	(0.58)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	28,311	28,167	28,311	28,167	28,311	28,167	28,311	28,167
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.0959	0.0834	0.0323	0.0302	0.0137	0.0148	0.0127	0.0123

Panel C: One quarter buy and hold adjusted returns

VARIABLES	*BHR[2,60]*		*BHMAR[2,60]*		*BHSAR[2,60]*		*BHFAR[2,60]*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low E.	High E.	Low E.	High E.	Low E.	High E.	Low E.	High E.
A	9.28***	1.80	8.47***	1.90	8.94***	2.71	9.34***	2.90
	(3.60)	(0.61)	(3.45)	(0.68)	(3.69)	(0.99)	(3.87)	(1.08)
E	-1.93	6.80	5.59	8.22	13.59	7.83	22.98	12.08
	(-0.08)	(0.73)	(0.23)	(0.92)	(0.56)	(0.88)	(0.94)	(1.36)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	28,311	28,167	28,311	28,167	28,311	28,167	28,311	28,167
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.1394	0.1410	0.0589	0.0653	0.0307	0.0360	0.0296	0.0344

Panel D: Half year buy and hold adjusted returns

VARIABLES	*BHR[2,125]*		*BHMAR[2,125]*		*BHSAR[2,125]*		*BHFAR[2,125]*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low E.	High E.	Low E.	High E.	Low E.	High E.	Low E.	High E.
A	13.46***	1.20	10.83***	1.10	11.71***	2.13	12.64***	2.59
	(3.83)	(0.30)	(3.15)	(0.29)	(3.49)	(0.57)	(3.65)	(0.69)
E	-11.13	5.68	-13.82	7.67	-2.68	7.56	31.50	6.89
	(-0.32)	(0.44)	(-0.40)	(0.61)	(-0.08)	(0.60)	(0.91)	(0.53)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	28,311	28,167	28,311	28,167	28,311	28,167	28,311	28,167
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.1737	0.1648	0.0811	0.0824	0.0398	0.0401	0.0416	0.0471

Notes. This table reports the regression results testing the interactions between aleatory and epistemic uncertainty. Panel A shows the regression results for three-day buy and hold return. Panel B shows the regression results for one-month buy and hold return. Panel C shows the regression results for one-quarter buy and hold return. Panel D shows the regression results for half-year buy and hold return. See Appendix B for variable definitions. We control for industry and year fixed effects. The *t*-statistics are based on standard errors that are clustered by firm. ***, **, * indicate significantly different from zero at the 1%, 5%, 10% level, respectively.

Table 8 Decomposition of Risk

Panel A: Full Sample

	BHR[2,20]	*BHR[2,60]*	*BHR[2,125]*
VARIABLES	(1)	(2)	(3)
D_A_Sys	0.08	0.86**	1.79***
	(0.38)	(2.21)	(3.33)
D_A_Idio	0.38*	0.81**	1.02**
	(1.88)	(2.25)	(2.09)
D_E	-0.28	0.04	-0.09
	(-1.35)	(0.11)	(-0.18)
Controls	Yes	Yes	Yes
Observation	56,478	56,478	56,478
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm
Adj. R^2	0.0884	0.1394	0.1676

Panel B: Partition by Ambiguity

	BHR[2,20]	*BHR[2,20]*	*BHR[2,60]*	*BHR[2,60]*	*BHR[2,125]*	*BHR[2,125]*
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Low E.	High E.	Low E.	High E.	Low E.	High E.
D_A_Sys	-0.19	0.32	0.30	1.24**	1.90***	1.43*
	(-0.64)	(0.98)	(0.56)	(2.19)	(2.64)	(1.76)
D_A_Idio	0.75***	-0.03	1.58***	0.01	2.23***	-0.27
	(2.79)	(-0.10)	(3.28)	(0.02)	(3.41)	(-0.38)
D_E	-0.45	-0.54	-0.30	0.03	-0.89	-1.00
	(-0.80)	(-0.87)	(-0.30)	(0.02)	(-0.66)	(-0.68)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observation	28,311	28,167	28,311	28,167	28,311	28,167
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Standard Errors	Firm	Firm	Firm	Firm	Firm	Firm
Adj. R^2	0.0960	0.0834	0.1394	0.1410	0.1738	0.1648

Notes. This table reports the regression results on the decomposition of aleatory uncertainty into systemic and idiosyncratic components. Specifically, we decompose aleatory uncertainty into systemic and idiosyncratic parts. We expect that the systemic components are reflected in the industry-year fixed effects. Formally, we run the following regressions:

$$A_{i,t} = \alpha + \text{Industry-Year FEs} + e_{i,t},$$

where the fitted value is an estimate of the latent systemic component (*A_Sys*) and the residual is an estimate of the latent idiosyncratic component (*A_Idio*). Industry–year fixed effects are included to capture systemic fluctuations that affect multiple firms simultaneously. To minimize the effects of outliers, we transform our independent variables of factor loadings into decile ranks. The fitted values from this regression capture the *systemic* component that co-moves across firms (by time and industry), whereas the residuals measure firm-level *idiosyncratic* deviations.

Panel A reports the regression results testing the relation between aleatory uncertainty components and stock returns. To facilitate the comparison of economic magnitudes, we transform our independent variables *A_Sys* and *A_Idio* into decile ranks for each year (denoted as *D_A_Sys* and *D_A_Idio*). These ranks are initially numbered 1 through 10 and then scaled to range from -0.5 to +0.5, with a mean of zero and a range of one. Column 1 shows the regression results for one-month buy and hold raw return. Column 2 shows the regression results for one-quarter buy and hold raw return. Column 3 shows the regression results for half-year buy and hold raw return. Panel B reports the partition results testing the interactions with epistemic uncertainty. See Appendix B for variable definitions. We control for industry and year fixed effects. The *t*-statistics are based on standard errors that are clustered by firm. ***, **, * indicate significantly different from zero at the 1%, 5%, 10% level, respectively.

Table 9 Associations with Existing Proxies of Ambiguity or Information Uncertainty

	D_Disp	*D_VIX*	*D_ImplVol*	*D_StdCFO*	*D_LnMV*	*D_Age*
VARIABLES	(1)	(2)	(3)	(4)	(6)	(7)
D_A	0.03***	0.05***	0.03***	0.02***	0.04***	0.08***
	(4.30)	(11.65)	(6.05)	(4.72)	(11.06)	(19.49)
D_E	0.02***	0.04***	0.18***	0.19***	-0.14***	-0.17***
	(2.64)	(10.51)	(31.59)	(48.28)	(-37.19)	(-43.60)
Controls	No	No	No	No	No	No
Observation	20,390	64,617	29,708	58,669	64,579	64,413
Industry FE	No	No	No	No	No	No
Year FE	No	No	No	No	No	No
Standard Errors	No	No	No	No	No	No
Adj. R^2	0.0014	0.0045	0.0367	0.0409	0.0213	0.0309

Notes. This table reports the regression results on the relationship between our our measures and proxies for ambiguity or information uncertainty from the prior literature. These proxies include 1) analyst forecast dispersion, measured using the last forecast from each analyst within the 30-day window preceding the 10-K filing date; 2) the VIX index at the 10-K filing date; 3) implied volatility at the 10-K filing date, constructed following Bae et al. (2023) at the firm level; 4) the standard deviation of operating cash flow; 5) firm size; and 6) firm age. To minimize the effects of outliers, we transform our variables into decile ranks. Consequently, the reported regression results reflect Spearman's rank correlation coefficients between variable pairs. See Appendix B for variable definitions. ***, **, * indicate significantly different from zero at the 1%, 5%, 10% level, respectively.