Text-implied uncertainty in 10-K filings: Do investors get the message?

Abstract

We investigate the role of text-implied uncertainty in the cross-section of stock returns. We employ word embeddings to derive an uncertainty dictionary from 10-K annual company filings. Stocks in the highest uncertainty quintile generate a 0.281% higher monthly return compared to stocks in the lowest quintile. This outperformance survives risk-adjustment. Furthermore, uncertainty is negatively related to filing period abnormal returns. This information is priced within one day after filing. High-uncertainty companies have lower return-on-assets, unexpected earnings, and are less likely to increase dividends in the subsequent fiscal year. Uncertainty averse investors demand a premium for holding high uncertainty stocks.

Keywords: Uncertainty, GloVe embeddings, text mining, uncertainty aversion

JEL: G10, G12, G40.

1 Introduction

We investigate the asset pricing implications of text-based uncertainty in 10-K annual company filings in the cross-section of the US stock market. We derive an uncertainty dictionary from pre-trained GloVe embeddings (see Pennington et al., 2014) to capture the perception of uncertainty from the management's perspective, as expressed in the annual reports. Our uncertainty dictionary covers a variety of natural expressions for uncertainty, e.g. "confusion" or "anxiety", and is thus a meaningful extension of simple word-count approaches based on relatives and synonyms of the word uncertainty. In order to capture the potentially related concepts of risk and sentiment, we construct two additional dictionaries which contain risk- and sentiment-related terms. By assigning each word to the dictionary with the highest similarity to the seed word and further orthogonalizing our uncertainty score with risk and sentiment, we ensure that our final score primarily reflects uncertainty.

Turning to the asset pricing implications, we find that stocks with high uncertainty scores earn higher average returns during our sample period from March 1994 to December 2020. This result holds true in portfolio sorts, where stocks in the highest quintile outperform stocks in the lowest quintile with up to 0.281%, even after risk-adjustment with most recent factor models. Factor model alphas on a monthly basis vary between 0.251% (t-statistic = 2.37) for the six-factor model of Fama and French (2018) and 0.449% (t-statistic = 3.61) for the q-4 factor model of Hou et al. (2015).

This result extends to cross-sectional Fama and MacBeth (1973) regressions where we find a positive relationship between our uncertainty score and stock returns in the subsequent month. This relationship remains highly significant above and beyond well-known predictors in the cross-section of stock returns, for example size, book-to-market and short-term reversal. In stark contrast, sentiment- and risk-related terms in 10-K reports have no significant impact on average returns.

To understand which characteristics are possibly related to uncertainty, we regress our uncertainty score on various stock characteristics and risk measures, for example illiquidity and idiosyncratic volatility. High-uncertainty stocks are more likely to be small growth stocks, have a higher market beta and higher idiosyncratic volatility. We furthermore find a negative relationship between our uncertainty score and the exposure to aggregate uncertainty, as documented by Bali et al. (2017), when controlling for all stock characteristics under consideration. However, even taken together, the well-established stock characteristics only explain roughly 4.2% of the cross-sectional variation in uncertainty.

Estimating uncertainty from 10-K filings allows us to study the instantaneous impact of text-based uncertainty on stock prices around the filing dates. We find a price drop within the first trading day after publication, suggesting that investors immediately sell stocks with high uncertainty in annual reports. The one-day lag is consistent with the fact that EDGAR filings are available to the general public after one to two business days (see Griffin, 2003). If we consider pre- and post-filing abnormal returns, i.e. the month before and the week after the filing date, there is no significant impact of uncertainty on abnormal returns. We draw two main conclusions from this finding. First, the uncertainty-related terms in 10-K filings represent new information for the public, since there is no differential impact in the month *before* the filing date. Second, there is no over- or underreaction to the new information in 10-K reports and investors promptly sell high-uncertainty stocks within one business day based on the new information. As a consequence of this price drop, high-uncertainty stocks earn higher average returns in the future. Once more, sentiment- and risk-related terms in 10-K filings have no significant impact on stock prices.

Finally, we study the predictive power of uncertainty for firm fundamentals in the subsequent fiscal year. Our uncertainty score negatively predicts future returns-on-assets, suggesting that uncertainty is negatively related to future profitability. Furthermore, high-uncertainty firms are more likely to report negative earnings surprises over the next fiscal year, since the relationship between our uncertainty score and standardized unexpected earnings is also negative. Additionally, a dividend increase in the next year is less likely for high-uncertainty firms. These findings are in line with the price drop in response to uncertainty around the filing date, suggesting that uncertainty-related terms in the 10-K raise concerns about deteriorating fundamentals of the corresponding company.

Our study contributes to a large and growing body of literature on the asset pricing implications of uncertainty.[1] From a theoretical point of view, Leippold et al. (2008) and Epstein and Schneider (2010) introduce an unobservable or ambiguous expected growth rate of dividends to show that ambiguity-averse investors require a higher equity premium. In Anderson et al. (2009), the positive ambiguity premium is driven by the uncertainty from forecasters' disagreement and is more important for the expected market return than risk, measured in terms of return volatility. This finding is in line with Brenner and Izhakian (2018) who account for ambiguity in order to restore a positive risk-return relation and document a positive ambiguity premium when investors are ambiguity-

[1] Our literature review is far from conclusive and we omit studies on the aggregate impact of uncertainty on future market returns for the sake of brevity (Bahmani-Oskooee and Saha, 2019; Chiang, 2020; Gupta et al., 2020)

averse. Similarly, Pastor and Veronesi (2013) document a positive uncertainty premium due to political uncertainty. In general, our finding of a positive uncertainty premium is consistent with the predictions of the theoretical models above and reveals that investors strongly dislike uncertainty-related terms in 10-K filings. The weak explanatory power of risk-related terms highlights the distinction between risk and uncertainty and the importance to account for both factors, as stressed out by Brenner and Izhakian (2018).

In stark contrast to the theoretical literature, empirical studies on the pricing impact of uncertainty reveal no such consensus regarding the sign of the uncertainty premium in the cross-section of stock returns. The majority of studies employs the exposure to aggregate uncertainty indices, for example the macroeconomic uncertainty index proposed by Jurado et al. (2015) or the Baker et al. (2016) economic policy uncertainty index (EPU). Here, Brogaard and Detzel (2015) and Gao et al. (2019) show that stocks with a high exposure to innovations in the EPU earn lower average returns. Bali et al. (2017) estimate betas with respect to the Jurado et al. (2015) macroeconomic uncertainty index and also document a negative cross-sectional premium. Both studies argue that stocks with high exposures to uncertainty hedge against deteriorating future investment opportunities.

A second strand on the cross-sectional pricing implications of uncertainty focuses on firm-level proxies for uncertainty. Diether et al. (2002) use the dispersion in analysts' earnings forecasts and find a negative relationship between their measure and expected stock returns. Baltussen et al. (2018) measure uncertainty on the firm level as the volatility of the implied volatility (vol-of-vol) of stock options, showing that this uncertainty proxy is negatively priced in the cross-section of stock returns. Somewhat interestingly, both studies leave the economic mechanism behind the negative pricing impact up to future research.

Our text-based uncertainty score provides a clean measure for firm-level uncertainty. Textual analysis is a promising approach to capture uncertainty as an expression of the opinion holder, which is, by construction, hard to measure with macroeconomic or market data. In contrast to analysts' forecasts or volatility, the uncertainty expressed by the management in 10-K reports reveals otherwise private information to the public from an insider's perspective rather than market expectations. Our text-based approach circumvents the measurement error when estimating the exposure to aggregate uncertainty indices and allows a novel distinction between uncertainty, sentiment and risk. The exact filing date of our 10-K reports furthermore provides us with a unique opportunity to study the immediate pricing impact of uncertainty-related terms in annual reports.

The study closest to ours is Zhang (2021), who derives uncertainty from 10-K filings, 10-Q filings and conference calls by using a word counting approach of six uncertainty-related terms (i.e. "uncertainty", "uncertain", "uncertainties", "ambiguity", "ambiguities", and "ambiguous") and apparently rests on investors who - contrary to common assumption in financial economics - love uncertainty. To control for risk-related terms, he scales his uncertainty word-count with risk-related words in the denominator (i.e. "risk", "risks", "risked", "riskier", "riskiest", "risking", "risky"). He finds a negative relationship between manager uncertainty and average stock returns and argues that stocks with a high manager uncertainty have higher real options value. Our approach extends his analysis in various directions, both in terms of uncertainty measurement as well as the empirical application of uncertainty. Our dictionary is based on GloVe embeddings and thus captures a higher number of relevant terms which might be no direct synonyms to uncertainty, but are equally important in our natural language to express uncertainty, e.g. "confusion", "doubts" or "anxiety". If managers deliberately avoid the phrase "uncertainty", the simple word-count approach is likely to miss key messages in the 10-K report. Furthermore, our multi-dimensional approach to distinguish uncertainty from risk and sentiment is more likely to separate between the three concepts, providing us with a clean measure for uncertainty. Finally, with respect to the empirical application of our measure, our matching procedure based on the publication dates of the reports as well as the focus on one-month ahead returns truly emphasizes the investor's perspective on manager uncertainty. In contrast, he aggregates text-based uncertainty throughout the entire calendar year, inducing a look-ahead bias which makes the strategy not investable. At first glance, our results appear contradictory to the negative risk-premium on high-uncertainty stocks documented by him. However, since he employs contemporaneous Fama and MacBeth (1973) regressions, this result is well in line with our event study findings which document a significant price drop one day after filing.[2]

The paper is organized as follows. Section 2 provides an overview about the data and explains the construction of our uncertainty measure from 10-K filings. Section 3 studies the asset pricing implications of our text-based uncertainty measure and Section 4 takes the relationship between uncertainty and firm fundamentals into account. Section 5 provides robustness checks and 6 concludes the paper.

[2] It is worth noting that this negative contemporary relation is hard to reconcile with the interpretation of manager uncertainty as real options value.

2 Data and Uncertainty Measurement

2.1 Data Cleaning

Our main data source to measure uncertainty are Form 10-K filings on the Security and Exchange Commission's (SEC) Electronic Data Gathering and Retrieval (EDGAR) website. We use stage one parsed files from 1994 to 2020 provided on the Notre Dame Software Repository for Accounting and Finance (SRAF).[3] In this preliminary parsing process, the 10-Ks are cleaned from original HTML, XBRL and XML markup language tags, ASCII-encoded graphics and tables embedded in the filings.

We process the filings further by first converting them into vectors of tokens, i.e. unigrams. Since we are only interested in words, we further remove numbers, punctuation and specific symbols. Furthermore, we clean the files of common English stopwords since they usually don't provide any useful content. Finally, we remove duplicates and 10-K/A files, leaving us with a total amount of 193,923 10-K filings and an average number of words of about 25,568 per filing.

The monthly return data is from CRSP and ranges from March 1994 to December 2020. We consider only common stocks with share code 10 or 11 with a price greater than $5 to avoid biased results due to the influence of small and micro cap stocks. We merge the text and return data with the relevant link tables from CRSP/Compustat. After merging, we are left with 8,372 distinct CIK-PERMNO combinations in the sample.

[3] We thank Bill McDonald for providing the parsed company filings data. For further information about the Parsing procedure see: https://sraf.nd.edu/sec-edgar-data/cleaned-10x-files/10x-stage-one-parsing-documentation/

In the next step, we remove documents with a reporting lag of 24 months and a total number of words per document below 250 (see Theil et al., 2018), which leaves us with 92,539 distinct 10-K files with an average number of about 33,904 words. The total number of companies in the dataset amounts to 8,197.

For our asset pricing analysis we use data on common risk factors and cross-sectional predictors. An overview of the variables is provided in Table 12. The data is obtained from the data library of Kenneth French, the q-factor website and openassetpricing.com by Chen and Zimmermann (2021).[4] The data we use for the analysis of the relationship between uncertainty and firm fundamentals is obtained from CRSP/Compustat and IBES.

2.2 Measuring Uncertainty

Our main focus is the quantification of uncertainty in 10-K filings using Natural Language Processing (NLP) techniques in terms of a single measure applicable for asset pricing tests. In the empirical finance context, the three concepts of uncertainty, risk, and sentiment are well-defined and clearly separated from another, but may be used interchangeably in natural language expressions. Therefore, we also derive dictionaries for the concepts of risk and sentiment, which enables us to investigate whether terms related to the three concepts are distinct from each other or tend to mix up in the vector space. In this respect, the methodology for quantifying risk, sentiment and uncertainty is the same. We compute scores regarding the three concepts by employing a dictionary-based word counting procedure. The advantage of such an unsupervised approach lies in the fact that no labelled data is needed to compute measures regarding the three concepts.

[4] For more details see: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, http://global-q.org/index.html and https://www.openassetpricing.com/

To derive a dictionary, usually one of two approaches is chosen. First, a dictionary can be derived manually by expert judgement. Examples of such dictionaries are the psychologial dictionary Harvard-IV within the General Inquirer Augmented Spreadsheet[5], the finance-specific word list of Henry (2008) or the Loughran and Mcdonald (2011) dictionary derived from 10-K filings. The second approach involves the derivation of the dictionary semi-automatically by specifying seed words from which dictionary terms are derived with NLP techniques. To avoid a potential subjectivity bias which possibly arises by using the former approach, we rely on word embeddings to derive semantically similar words to each of the three seed words "uncertainty", "risk" and "sentiment".

Word embeddings represent each word of a given vocabulary as a real-valued vector in the d-dimensional vector space. The numbers within the vector encode word-specific semantic information. Most of the word embedding techniques encode the word vectors in a way that semantically similar words are positioned close to each other in the vector space. Semantic similarity itself is usually defined by the context of the words such that words which often appear in the same context are highly semantically similar. Famous techniques to derive word embeddings include word2vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), fasttext (Bojanowski et al., 2017) and BERT (Devlin et al., 2019).

To derive dictionary terms for each concept, we use pre-trained GloVe embeddings from Pennington et al. (2014). More precisely, we use embeddings with a dimensionality of 100, derived from a six billion token corpus.[6] For each seed word of the three concepts we identify the most similar words in terms of the cosine similarity. To derive a choice for the

[5] http://www.wjh.harvard.edu/~inquirer/Home.html.

[6] For more information see: https://nlp.stanford.edu/projects/glove/

number of words to incorporate into each dictionary, we implement a K-Means clustering algorithm to the word vectors of the respective dictionary words for a varying number of words. This method avoids choosing an arbitrary number of words to incorporate into each dictionary. Since we want to disentangle words regarding the concepts of uncertainty, risk and sentiment as clearly as possible, we investigate how well the K-Means algorithm does in clustering the dictionary words into their true clusters (uncertainty, risk and sentiment). Figure 1 shows the results of this procedure with the number of words per dictionary ranging from 10 to 100. The figure shows the frequency of correctly clustered words on the y-axis with 20 being the number where the K-Means algorithm does best in assigning the correct clusters to the respective dictionary words. However, it should be noted that with 20 words the algorithm does not cluster all words correctly, providing a first hint that some words regarding the three concepts tend to mix up in the vector space. The final dictionary incorporates the top 20 words including the seed word. If a word appears in multiple dictionaries, we assign it to the dictionary where the word has the highest cosine similarity with the seed word. This leaves us with three mutually exclusive dictionaries.

Table 1 shows the words belonging to each dictionary and the respective values of the cosine similarity to the target word. The seed words themselves have a cosine similarity of one. All three dictionaries contain words which seem to be closely related to the seed concepts. For instance the uncertainty dictionary contains its relatives ("uncertainties", "uncertain") as well as other uncertainty-related terms like "worries" or "concerns". Similarly, the risk and sentiment dictionary contain risk- and sentiment-related terms like "danger" or "likelihood" in the former and "optimism" or "bullish" in the latter case. However, the dictionaries also contain words which seem to be more related to one of the other concepts like "pessimism" in the uncertainty dictionary which might be more related to sentiment or "volatility" in the uncertainty dictionary which might be more related to the concept of risk. Moreover, some words do not seem to be directly related to the seed concept, but seem to be indirectly related to the latter like "reduce" in the case of risk, which is not a synonym to risk, but might show up with risk in terms of "reducing risk". These findings show that word embeddings are not just able to capture word relationships in terms of synonymy, but are also capable of encoding the semantic similarity between words with respect to the context in which they are usually used.

The K-Means clustering exercise provided a first hint to the fact that not all words can be correctly assigned to their true clusters, indicating that some words tend to mix up in the vector space. To investigate more thoroughly whether the words of the three dictionaries are well separated from each other in the vector space, we perform a Principal Component Analysis (PCA) on the word vectors. This enables us to reduce the dimensionality of the matrix of word vectors and visualize them in the two-dimensional vector space. Figure 2 shows a plot of the first two Principal Components (PCs). The blue dots show uncertainty-related terms, whereas the green and red dots present sentiment- and risk-related terms, respectively. The figure shows a clustering of risk-related terms, indicating that the concept of risk seems to be clearly separable from the other two concepts. In contrast, the terms of the other two concepts are less isolated from each other but tend to mix up in the vector space, illustrating the fact that the concepts of uncertainty and sentiment are not clearly separable from each other. However, it should be noted that risk-related words are only distinct from words of the other concepts in the two-dimensional space. Since the original word vectors are vectors of 100 dimensions, it is very likely that risk-related words mix up with the words of the other concepts in higher dimensions. For this reason, as well as the fact that the uncertainty dictionary contains terms which might be more related to the other concepts, an uncertainty measure derived from the respective uncertainty dictionary might not solely reflect uncertainty, but also risk or sentiment. Therefore, we compute one raw score for each concept as well as an orthogonalized uncertainty score.[7]

Raw uncertainty, risk and sentiment are measured in the following way:

$$UNC_{i,t} = \frac{N_{i,t}^{UNC}}{N_{i,t}^{doc}} \qquad RISK_{i,t} = \frac{N_{i,t}^{RISK}}{N_{i,t}^{doc}} \qquad SENT_{i,t} = \frac{N_{i,t}^{SENT}}{N_{i,t}^{doc}} \tag{1}$$

where $N_{i,t}^{UNC}$, $N_{i,t}^{RISK}$ and $N_{i,t}^{SENT}$ are measures of the counts of uncertainty-, risk- and sentiment-related terms in the 10-K filing of company i at time t and $N_{i,t}^{doc}$ measures the total number of words in the filing of company i at time t.[8]

To orthogonalize uncertainty, uncertainty scores are regressed on the risk and sentiment scores using cross-sectional regressions. The econometric specification is the following:

$$UNC_{i,t} = \alpha + \beta RISK_{i,t} + \gamma SENT_{i,t} + \epsilon_{i,t} \tag{2}$$

[7] As an alternative way to derive a clean uncertainty dictionary, we investigated the words with the highest cosine similarity to the word vector: "uncertainty" - "risk" - "sentiment". However, this approach does not lead to meaningful results as words which previously had a high positive cosine similarity with the word "uncertainty" now have a high negative cosine similarity to the latter, suggesting that the derived word vector points in the opposite direction to the vector of the word "uncertainty".

[8] Due to the size of the dictonaries, the value of the nominator in Equation (1) is usually relatively small. To investigate whether the results reported in the following sections are driven by the denominator which might be related to measures of readability or complexity like the Fog-Index (see Loughran and Mcdonald, 2014), we run Placebo tests with $1/N_{i,t}^{doc}$. In unreported results, we show that this measure does not have any explanatory power for expected returns and does not subsume the relationship between uncertainty and stock returns.

where $UNC_{i,t}$, $SENT_{i,t}$ and $RISK_{i,t}$ correspond to the raw uncertainty, sentiment and risk score of company i at time t from Equation (1). We then use the residual $\epsilon_{i,t}$ of the regression in Equation (2) as the orthogonalized uncertainty score denoted as $UNC_{i,t}^{Orth}$.

As a result of this data preparation process, we receive text-implied scores for uncertainty-, risk-, and sentiment-related terms for each filing company on the respective filing date. Since we are interested in the impact of text-implied uncertainty at the time of the publication, we match this score with return and stock data based on the filing date instead of the report date. Filing dates strongly cluster in the first quarter of each calendar year and roughly 78% of all 10-K reports are filed in February and March. In order to avoid a strongly unbalanced sample throughout the year, we hold each of the text-implied measures constant, until a new 10-K is filed or drop the stock from the sample if the information is stale for 24 months. A constant number of stocks throughout the year is particularly important for the portfolio sorts in Section 3.1 in order to avoid unbalanced and very small portfolios in months with a low filing activity. In case of the cross-sectional regressions in Section 3.2, holding values constant in between filing dates allows us to to follow the standard and account for well-known predictors on a monthly basis. However, robustness checks in Section 5 show that or findings also hold without filling our text-based measures throughout the year.

2.3 Desciptive Statistics

Table 2 presents descriptive statistics for variables of interest, including the one-month-ahead return given in percent (R_{t+1}), the uncertainty score (UNC), risk score (RISK), sentiment score (SENT) and orthogonalized uncertainty score (UNC^{Orth}) as well as cross-sectional predictors defined in Table 12. The average one-month-ahead return is about 1.53%, exhibits a positive skewness and a relatively high excess kurtosis. The uncertainty, risk and sentiment measures all have a positive mean, positive skewness and a positive excess kurtosis. The uncertainty measure has the highest excess kurtosis among the three raw measures. Orthogonalized uncertainty has a zero mean and a standard deviation close to the one of the raw uncertainty measure. It has positive skewness and excess kurtosis as well, indicating that its distribution is slightly asymmetric around the mean and has fat tails.

Table 3 presents the time series averages of the cross-sectional correlations between the variables of interest. The one-month-ahead return has a relatively weak correlation with all of the considered variables. The uncertainty measure exhibits a weakly positive correlation with the risk and sentiment measures which further supports the idea of orthogonalizing uncertainty with respect to risk and sentiment. After orthogonalization, this correlation is close to zero. The minor positive correlation with sentiment might arise because of the fact that the orthogonalization is conducted in a multivariate setting. The orthogonalized measure is weakly correlated with several cross-sectional predictors. For example, it is negatively correlated with SIZE, indicating that stocks with higher uncertainty tend to be smaller stocks, and has a positive correlation with the market beta (β^{MKT}), indicating that stocks with higher uncertainty have higher exposure to market returns. Furthermore, stocks with higher uncertainty scores have a higher idiosyncratic volatility and are more likely to be lottery stocks given the positive correlation to IVOL and MAX.

3 Uncertainty and Stock Returns

3.1 Univariate Portfolio Analysis

We now turn to the asset pricing implications of our text-based uncertainty measure and start with univariate sorts. Each month we sort stocks into quintiles from stocks with the lowest to stocks with the highest orthogonalized uncertainty score. We hold each portfolio for one month and then compute the value-weighted returns of the quintile portfolios as well as the returns of a long-short portfolio which invests in the high-uncertainty quintile, while shorting stocks in the lowest quintile. In addition to monthly excess returns, we compute risk-adjusted returns (alphas) from six factor models, namely i) the CAPM, ii) the Fama and French (1993) three-factor model ("FF3"), iii) the Fama and French (2015) five-factor model ("FF5"), iv) the Fama and French (2018) six-factor model ("FF6"), v) the q-4 model according to Hou et al. (2015), and vi) the q-5 model introduced by Hou et al. (2021) ("q-5").

Table 4 presents the results. The first row of the table shows that when moving from quintile 1 to 5, the value-weighted return increases monotonically from 0.697% to 0.978%. The average return difference between quintile 5 and quintile 1 stocks is 0.281% with a Newey and West (1987) (NW) t-statistic of 2.02. This implies that stocks in the highest quintile generate on average 3.372% higher annual returns than stocks in the lowest quintile. In addition to the the value-weighted raw returns, Table 4 provides an overview about the magnitude and statistical significance of the risk-adjusted returns from six factor models. Four out of the six models, namely the CAPM, FF3, q-4 and q-5 model, exhibit monotonically increasing risk-adjusted returns from quintile 1 to 5. The risk-adjusted returns of the FF5 and FF6 model decrease from quintile 1 to quintile 2 and monotonically increase thereafter. The difference in alphas between quintile 5 and 1 of the six factor models ranges from 0.251% for the FF6 model to 0.449% for the q-4 model, tantamount to an annual risk-adjusted return difference from 3.012% to 5.388% between stocks of the highest and lowest quintile. The NW t-statistics indicate that the differences in alphas are statistically significant for all factor models. Furthermore the results seem to be driven by an outperformance of high-uncertainty stocks, i.e. the long leg of the long-short portfolio. The portfolio sorts provide a first indication that stocks with high uncertainty tend to earn higher risk-adjusted returns.

3.2 Fama-Macbeth Regressions

In this section, we analyze the cross-sectional relation between our orthogonalized measure of uncertainty and expected stock returns in Fama and MacBeth (1973) regressions (FMB). We present time-series averages of the slope coefficients from the regressions of one-month-ahead stock returns on the uncertainty measure. We furthermore control for the risk and sentiment measures as well as for various cross-sectional predictors. The final regression has the following form:

$$R_{i,t+1} = \lambda_{0,t} + \lambda_{1,t} UNC_{i,t}^{Orth} + \lambda_{2,t} RISK_{i,t} + \lambda_{3,t} SENT_{i,t} + \lambda_{4,t} X_{i,t} + \epsilon_{i,t+1} \tag{3}$$

where $R_{i,t+1}$ is the realized excess return on stock i in month $t+1$, $UNC_{i,t}^{Orth}$, $RISK_{i,t}$ and $SENT_{i,t}$ are the orthogonalized uncertainty, risk and sentiment scores of company i in month t. $X_{i,t}$ is a set of stock-specific control variables for stock i in month t as defined in Table 12. Control variables are: SIZE, the market beta (β^{MKT}) (Fama and MacBeth, 1973), the ratio of book value of equity to market value of equity (BM) (Fama and French, 1992), Momentum (MOM) (Jegadeesh and Titman, 1993), idiosyncratic volatility (IVOL) (Ang et al., 2006), illiquidity (ILLIQ) (Amihud, 2002), short-term reversal (REV) (Jegadeesh, 1990), Coskewness (COSKEW) (Harvey and Siddique, 2000), the maximum return (Bali et al., 2011), the VIX beta (β^{VIX}) (Ang et al., 2006), and the uncertainty beta (β^{Unc}) according to (Bali et al., 2017). Table 5 presents the time-series averages of slope coefficients and the NW t-statistics in parentheses.

The first column indicates a significantly positive relationship between the orthogonalized uncertainty measure and the one-month-ahead stock returns. The average slope on $UNC_{i,t}^{Orth}$ in the univariate regression is 8.67 with a NW t-statistic of 4.80. Columns (2) and (3) present the average slopes of risk and sentiment, showing that both of the measures exhibit insignificant coefficients in the univariate framework. Column (4) controls for all text-based measures simultaneously, revealing that only the orthogonalized uncertainty measure has a significant coefficient. Columns (5) to (7) add cross-sectional predictors to each of the text-implied measures, presenting a significantly positive average slope for orthogonalized uncertainty while risk and sentiment display insignificant coefficients, in line with the results from the univariate regressions. Column (8) controls for all text-based measures and cross-sectional predictors simultaneously. Again, among the text-derived measures only orthogonalized uncertainty has a significantly positive coefficient of 4.45 with a NW t-statistic of 3.61 while risk and sentiment exhibit insignificant coefficients.

The signs of the coefficients of the cross-sectional predictors are, in the most cases, in line with prior empirical evidence. The effect on BM is positive and significant in all specifications while the SIZE effect is significantly negative. Furthermore, there is evidence for a statistically significant short-term reversal effect. Moreover, ILLIQ and COSKEW exhibit positive coefficients, statistically significant at the 10%-level. Contrary to prior empirical research, the stocks in our sample do not exhibit any momentum effects and the coefficient on IVOL is significantly positive. MAX, β^{VIX} and β^{UNC} exhibit insignificant coefficients in all regression specifications under consideration, probably due to the short sample period. Finally, somewhat surprisingly, the coefficient of β^{MKT} is significantly positive in all of the regression specifications.[9]

3.3 Average Stock Characteristics

To understand the relationship between our text-based uncertainty measure and other well-known firm characteristics, we examine the average stock characteristics of low and high uncertainty stocks. We perform additional monthly FMB regressions with the orthogononalized uncertainty score as the dependent variable, i.e.

$$UNC_{i,t}^{Orth} = \lambda_{0,t} + \lambda_{1,t} X_{i,t} + \epsilon_{i,t}, \tag{4}$$

where $UNC_{i,t}^{Orth}$ is the orthogonalized uncertainty score of stock i in month t and $X_{i,t}$ is a collection of stock-level predictors as defined in Table 12.[10] Table 6 reports the time series averages of the slope coefficients from the regression in Equation (4).

[9] In unreported results, we furthermore control for an uncertainty measure derived from the Loughran and Mcdonald (2011) uncertain sentiment dictionary in order to investigate whether controlling for the latter subsumes the relationship between our uncertainty measure and one-month ahead returns. However, we find no empirical evidence for this conjecture.

[10] We don't use a price filter in this framework, because this potentially biases the relationship between SIZE and uncertainty.

In univariate settings in Columns (1) - (8), we find a significant relationship between UNC^{Orth} and the majority of firm-level characteristics. High-uncertainty stocks tend to be smaller, exhibit a higher β^{MKT}, but lower momentum and a lower book-to-market ratio. High-uncertainty thus tend to be growth stocks. Conversely, stocks with high UNC^{Orth} tend to have higher IVOL, be more lottery-like, exhibit a positive COSKEW and higher ILLIQ.

All the effects above are significant at the one percent level. When we consider the exposure to systematic factors, Columns (9) and (10) reveal that high-uncertainty stocks have a stronger exposure to the VIX as well as the macroeconomic uncertainty index UNC. Both relationships are somewhat unexpected as both factors tend to earn negative risk premia in the cross-section, but may be driven by omitted characteristics of high-uncertainty stocks in the univariate regressions.

Column (11) shows the results of the regression specification including all regressors simultaneously. With four interesting exceptions, the univariate results also hold in a multivariate setting. The relationship between SIZE and UNC^{Orth} is significantly positive (t-statistic = 4.82) once we control for all variables at once. The sign of MAX and ILLIQ switches from positive to negative, where the former is most likely driven by the high correlation between MAX and IVOL ($\rho = 0.87$). While the relationship between UNC^{Orth} and β^{VIX} is now insignificant, we find that high-uncertainty stocks tend to have a lower β^{UNC}. This finding is consistent with Bali et al. (2017) and suggests that stocks with a higher UNC^{Orth} are unlikely hedges against increases in aggregate uncertainty.

3.4 Event Study Analysis

Our previous results suggest that stocks with a high uncertainty score earn higher average returns. To investigate the exact pricing mechanism of this relation, we follow Jegadeesh and Wu (2013) and Loughran and Mcdonald (2011) and perform an event study. Since we are measuring uncertainty from 10-K filings with an exact filing date, we can study the instantaneous impact of uncertainty around the publication of the new information in the 10-K.

As a first step, we zoom-in on the performance of the quintile portfolios in Section 3.1 around the filing date. Instead of monthly returns, we compute daily market-adjusted abnormal returns for each stock from

$$ret_{i,t}^{exc} = ret_{i,t} - ret_{MKT,t} \tag{5}$$

where $ret_{i,t}$ defines the return on stock i at day t and $ret_{MKT,t}$ defines the market return on day t. Again, we sort stocks into quintile portfolios based on their risk, sentiment or orthogonalized uncertainty score at the end of the previous month. For each portfolio, we then compute the equal-weighted average daily abnormal return. We focus on an event window from day -5 to day +5.

Figure 3 presents the average abnormal filing period returns for quintile sorts on the different scores. Panel A presents average abnormal returns for the five UNC^{Orth} quintiles and reveals an interesting pattern. While the returns from quintiles one to four co-move throughout the entire event window, the highest quintile exhibits a strong price drop one day after the filing with an abnormal return of -0.27%. The one-day delay of the pricing effect is consistent with the availability of EDGAR filings for average investors without a paid subscription, who receive the filings one day after publication (see Griffin, 2003). We repeat this analysis in Panels B and C for sentiment and risk, but find no such pattern. Here, the abnormal returns of portfolios formed on the sentiment and risk scores co-move before and after the filing date except for quintile 2 of the risk-sorted quintiles where we see a relatively large spike in abnormal returns two days prior to the filing date.

Next, we take a closer look at this finding and perform cross-sectional regressions of the abnormal returns on the orthogonalized uncertainty score as well as the risk and sentiment scores. For this purpose, we follow Jegadeesh and Wu (2013) and Loughran and Mcdonald (2011) and compute filing period abnormal returns in a four-day event window ranging from event day 0 through +3 as follows:

$$\tilde{r}_i = \prod_{t=0}^{3} ret_{i,t} - \prod_{t=0}^{3} ret_{MKT,t} \tag{6}$$

$\tilde{r}_i$ gives us the buy-and-hold excess return for each stock over the filing period. We consider two additional time windows to emphasize that the impact of uncertainty concentrates on the filing date. First, to account for the performance prior to the publication date, we use an pre-event window of -20 to -1 days relative to the filing dates. Second, in order to investigate whether the investors initially over- or underreact to the information regarding uncertainty in the 10-K files, we also consider a post-event window from day +5 through +9. For both event windows, we compute abnormal returns according to Equation (6).

Next, we follow Loughran and Mcdonald (2011) and Jegadeesh and Wu (2013) by performing quarterly FMB regressions of the following form:

$$\tilde{r}_{i,t} = \lambda_{0,t} + \lambda_{1,t} UNC^{Orth}_{i,t} + \lambda_{2,t} Risk_{i,t} + \lambda_{3,t} Sent_{i,t} + \lambda_{4,t} X_{i,t} + \epsilon_{i,t} \tag{7}$$

where $\tilde{r}_{i,t}$ is the abnormal filing period return of company i in quarter t, $UNC^{Orth}_{i,t}$, $Risk_{i,t}$ and $Sent_{i,t}$ are the orthogonalized uncertainty, risk and sentiment scores of company i in quarter t and $X_{i,t}$ is a set of control variables which follows Jegadeesh and Wu (2013). Control variables are SIZE, BM, Accruals as defined by Sloan (1996) and VOLATILITY as well as TURNOVER defined by Jegadeesh and Wu (2013). Table 12 provides further details regarding the construction of these controls.

Table 7 presents the results of the cross-sectional regressions of the abnormal filing period returns on the orthogonalized uncertainty, risk and sentiment scores as well as of the multivariate framework where we control for various firm-level characteristics. Column (1) reveals a negative and statistically significant relationship between orthogonalized uncertainty and filing period abnormal returns. This finding also holds true when we control for sentiment- and risk-related terms in Column (4). The highly negative coefficient on $UNC^{Orth}_{i,t}$ is also robust to the inclusion of firm-level control variable in Columns (5) and (8), independently whether or not we include sentiment- and risk-related terms. SENT and RISK have no consistent impact on abnormal daily returns, as expected from Figure 3.

The results above indicate that investors react to the publication of the 10-K filings by selling stocks considered as highly uncertain, possibly due to uncertainty aversion.

Furthermore, none of the cross-sectional control variables has a significant coefficient in the regressions.

Table 8 presents the results of the cross-sectional regressions for the pre- and post-event windows. In Columns (1) - (4), the dependent variable is the abnormal return from day -20 to day -1 relative to the filing date. The coefficient on the orthogonalized uncertainty measure is insignificant in all of the specifications, suggesting that uncertainty has no significant effect on abnormal returns before the filing dates. The same holds true for RISK and SENT, where all coefficients are insignificant as well. Columns (5) to (8) present results for the abnormal post-event returns from event day +5 through +9. Again, orthogonalized uncertainty is not statistically significant, suggesting that investors neither overreact, nor underreact to the information on uncertainty in 10-K filings, but that this information is promptly incorporated into stock prices. As Figure 3 suggests, this information is priced into stocks with a delay of one day. Column (6) indicates a positive relationship between the SENT score and the post filing period returns in the univariate setting. However, this relationship vanishes in the multivariate setting in Column (8).

4 Uncertainty and Firm Fundamentals

In this section, we investigate the relationship between orthogonalized uncertainty and various firm fundamentals in order to identify whether uncertainty affects future firm cash flows and thus signals deteriorating firm prospects. For this purpose we consider the following panel regression framework:

$$F_{i,t+1} = \lambda_{0,t} + \lambda_{1,t} UNC_{i,t}^{Orth} + \lambda_{2,t} RISK_{i,t} + \lambda_{3,t} SENT_{i,t} + \lambda_{4,t} X_{i,t} + \epsilon_{i,t+1} \tag{8}$$

where $F_{i,t+1}$ is a firm fundamentals variable of company i in the fiscal year $t+1$. The firm fundamentals variable is either one of three variables: the return on assets (ROA), the standardized unexpected earnings (SUE) or a dividend increase dummy (DIV), set to the value of one if a company increases its dividends during the fiscal year t. $UNC_{i,t}^{Orth}$, $RISK_{i,t}$ and $SENT_{i,t}$ are the orthogonalized uncertainty, risk and sentiment score of company i in year t. $X_{i,t}$ is the same set of control variables used in the event study in Table 7, i.e. ACCRUALS, VOLATILITY, BM, SIZE, TURNOVER, and VOLATILITY. For the purpose of this panel regression, we report t-statistics from heteroscedasticity-consistent standard errors clustered both by firm and year in parentheses. We estimate the models in Columns (1) to (4) with OLS and employ a logit model in the case of the dividend dummy in Columns (5) and (6). In the latter case, the reported R^2 is the McFadden Pseudo R^2.

Table 9 presents the results of the panel regressions. Columns (1) and (2) displays the regression results with ROA as dependent variable, revealing a significantly negative relationship between uncertainty and ROA in the next fiscal year. Thus, firms with a high orthogonalized uncertainty score tend to be less profitable in the subsequent fiscal year. In contrast to the previous results, RISK is also negatively related to ROA while SENT is significantly positive once we account for control variables. Among the control variables, BM is the only insignificant variable. ACCRUALS and Size exhibit significant positive coefficients while TURNOVER and VOLATILITY are significantly negatively related to ROA.

Columns (3) and (4) present the results with SUE as the dependent variable. Column (3) reveals that among the text-implied measures, only UNC^{Orth} exhibits a significant coefficient, which is negative and thus in line with the results in the Columns (1) and (2). After accounting for control variables in Column (4), UNC^{Orth} remains significantly negative at the 10% level. Among the control variables ACCRUALS and SIZE exhibit significantly positive coefficients, the latter being significant at the 10%-level. TURNOVER and VOLATILITY both have negative coefficients with t-statistics of -2.13 and -1.83, respectively.

Columns (5) and (6) present the results from a logistic regression of the dividend increase dummy on the text-implied measures and firm-level controls. Again, uncertainty is significantly negative in both specifications. RISK is once more insignificant, SENT is significantly positive in the full model in Column (6). Regarding the control variables, VOLATILITY and TURNOVER are again significantly negative. BM is significantly negative with a t-statistic of -1.84, while the coefficient on SIZE is significantly positive with a t-statistic of 14.78.

The results reveal several interesting relationships between uncertainty information from 10-K filings and future firm fundamentals, indicating that high-uncertainty firms tend to have a lower return on assets, lower standardized unexpected earnings and are less likely to increase dividends in the fiscal year following the annual report. Thus, firms with a higher level of uncertainty in the annual report are more likely to face a deterioration in future cash flows, raising concerns among investors, which possibly explains the price drop after the publication of the 10-K through a cash flow channel.

5 Robustness Checks

5.1 Conditional Double Sorts

The cross-sectional regressions in Table 6 reveal several significant relationships between $UNC_{i,t}^{Orth}$ and our control variables. To ensure that $UNC_{i,t}^{Orth}$ is important above and beyond those alternative predictors, we perform conditional bivariate sorts and re-examine average excess returns as well as the alphas of the FF6 and q-5 models. We choose both models based on their good performance in the univariate sorts, where the FF6 model produces the lowest alphas among the factor models of Fama and French (1993, 2015, 2018) and the q-5 model outperforms the q-4 model of Hou et al. (2015).

Each month, we sort stocks into quintiles based on our main control variables from Table 5: The natural logarithm of market capitalization (SIZE), book-to-market ratio (BM), momentum (MOM), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), short-term reversal (REV), coskewness (COSKEW), maximum return (MAX), market beta (β^{MKT}), VIX beta (β^{VIX}), and uncertainty beta (β^{UNC}). Within each quintile, we further form quintiles based on $UNC_{i,t}^{Orth}$, hold the portfolio for one month and compute value-weighted excess returns as well as factor model alphas. Our main focus are the average 5-1 long-short portfolios for each quintile. Table 10 presents the results.

For each variable under consideration, the average excess return of the 5-1 portfolio is statistically significant at the 5% level. Average returns range from 0.207% for BM to 0.553% per month for Size with NW t-statistics of 1.96 and 4.49, respectively. For the largest part, this finding extends to the factor model alphas with two minor exceptions. The FF6 model explains the 5-1 returns in case of the double sorts based on β^{MKT} and BM, since the FF6 α turns insignificant with t-statistics of 1.61 and 1.41, respectively. In stark contrast, however, the q-5 model leaves a significant alpha in both cases. More precisely, the double sorts which control for β^{MKT} create a q-5 alpha of 0.343% per month with a t-statistic of 1.89. In the case of BM, the q-5 alpha of 0.350% per month is significant at any conventional level (t-statistic = 2.93). Thus, controlling for β^{MKT} and BM weakens the impact of $UNC_{i,t}^{Orth}$ on average returns, but does not fully capture the asset pricing implications of $UNC_{i,t}^{Orth}$.

5.2 Fama-Macbeth Regressions, Filing Months Only

As outlined in Section 2.1, we assume that our text-based measures UNC^{Orth}, RISK, and SENT remain constant in between filing dates in order to handle the filing clusters in February and March and avoid unbalanced portfolios in the univariate sorts. Since this approach is technically not necessary for the FMB regressions in Section 3.2, we repeat the analysis only for the filing months in our sample. Due the smaller number of firms in each month, the average number of observations is considerably lower. Other than that, the setup is identical to Table 5. For the sake of brevity, we do not report control variables and focus on our main measures, namely UNC^{Orth}, $RISK$, and $SENT$. Table 11 presents the results.

The results in Table 11 strongly support the explanatory power of text-implied uncertainty. UNC^{Orth} remains significant at the 1% level throughout all specifications, independently of the included control variables. The largest difference to the baseline results is the remarkably good performance of SENT in Columns (3) and (4) without the inclusion of control variables. Here, SENT is significantly positive, indicating that 10-K reports with a large number of sentiment-related terms are associated with higher returns in the next month. However, including controls in Columns (7) and (8) fully wipes out this effect and SENT turns, once more, insignificant. In general, we conclude that the positive relationship between UNC^{Orth} and average returns is robust to this adjustment of our data preparation process.

6 Concluding remarks

This paper investigates the role of text-implied uncertainty in the cross-sectional pricing of individual stocks. We measure uncertainty on the firm level from 10-K filings using a dictionary-based approach. We utilize pre-trained GloVe embeddings (Pennington et al., 2014) to derive uncertainty, sentiment and risk dictionaries and use those dictionaries to compute scores regarding the respective concepts.

Univariate sorts indicate that stocks in the highest uncertainty quintile generate on average 3.372% higher annual returns than stocks in the lowest quintile. This result extends to risk-adjusted returns. The positive cross-sectional relationship between orthogonalized uncertainty and expected returns is confirmed by FMB regressions. Those results indicate uncertainty-averse investor behavior, whereas investors require a positive uncertainty premium for holding stocks with a high level of uncertainty. This is in line with theoretical studies on this subject (e.g. Anderson et al., 2009; Brenner and Izhakian, 2018; Leippold et al., 2008) as well as with the behavior of individuals observed in experimental studies (e.g. Ellsberg, 1961).

The event study analysis furthermore confirms those results by showing that orthogonalized uncertainty is negatively related to filing period abnormal returns. Moreover, the information on uncertainty seems to be priced efficiently with a lag of one day after the initial filing date as the orthogonalized uncertainty measure is not significantly related to post filing period abnormal returns.

Studying the impact of orthogonalized uncertainty on firm fundamentals, we find that high uncertainty firms are more likely to face a deterioration of future cash flows, possibly explaining the negative pricing effect of uncertainty through a cash flow channel.

To the extent that our measure of uncertainty proxies for the degree of a manager's uncertainty regarding the future, this degree of uncertainty conveys important information to investors. The results suggest that investors require a higher premium for holding stocks when the manager is highly uncertain about the future as expressed by specific words in the 10-K filings. This indicates that managers should choose their words wisely when writing the company report.

References

Amihud, Yakov (2002), 'Illiquidity and Stock Returns: Cross-Section and Time-Series Effects', *Journal of Financial Markets* **5**(1), 31–56.

Anderson, Evan W., Eric Ghysels and Jennifer L. Juergens (2009), 'The impact of risk and uncertainty on expected returns', *Journal of Financial Economics* **94**(2), 233–263.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing and Xiaoyan Zhang (2006), 'The Cross-Section of Volatility and Expected Returns', *The Journal of Finance* **61**(1), 259–299.

Bahmani-Oskooee, Mohsen and Sujata Saha (2019), 'On the effects of policy uncertainty on stock prices: an asymmetric analysis', *Quantitative Finance and Economics* **3**(2), 412–424.

Baker, Scott R., Nicholas Bloom and Steven J. Davis (2016), 'Measuring Economic Policy Uncertainty*', *The Quarterly Journal of Economics* **131**(4), 1593–1636.

Bali, Turan G., Nusret Cakici and Robert F. Whitelaw (2011), 'Maxing out: Stocks as Lotteries and the Cross-Section of Expected Returns', *Journal of Financial Economics* **99**(2), 427–446.

Bali, Turan G, Stephen J Brown and Yi Tang (2017), 'Is Economic Uncertainty Priced in the Cross-Section of Stock Returns?', *Journal of Financial Economics* **126**(3), 471–489.

Baltussen, Guido, Sjoerd van Bekkum and Bart van der Grient (2018), 'Unknown Unknowns: Uncertainty About Risk and Stock Returns', *Journal of Financial and Quantitative Analysis* **53**(4), 1615–1651.

Bojanowski, Piotr, Edouard Grave, Armand Joulin and Tomas Mikolov (2017), 'Enriching Word Vectors with Subword Information', *Transactions of the Association for Computational Linguistics* **5**, 135–146.

Brenner, Menachem and Yehuda Izhakian (2018), 'Asset Pricing and Ambiguity: Empirical evidence', *Journal of Financial Economics* **130**(3), 503–531.

Brogaard, Jonathan and Andrew Detzel (2015), 'The Asset-Pricing Implications of Government Economic Policy Uncertainty', *Management Science* **61**(1), 3–18.

Carhart, Mark M. (1997), 'On Persistence in Mutual Fund Performance', *The Journal of Finance* **52**(1), 57–82.

Chen, Andrew Y and Tom Zimmermann (2021), 'Open Source Cross-Sectional Asset Pricing', *Critical Finance Review, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3604626 or http://dx.doi.org/10.2139/ssrn.3604626* .

Chiang, Thomas C. (2020), 'Economic policy uncertainty and stock returns—evidence from the Japanese market', *Quantitative Finance and Economics* **4**(3), 430–458.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2019), 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* **1**, 4171–4168.

Diether, Karl B., Christopher J. Malloy and Anna Scherbina (2002), 'Differences of Opinion and the Cross Section of Stock Returns', *The Journal of Finance* **57**(5), 2113–2141.

Ellsberg, Daniel (1961), 'Risk, Ambiguity, and the Savage Axioms', *The Quarterly Journal of Economics* **75**(4), 643–669.

Epstein, Larry G. and Martin Schneider (2010), 'Ambiguity and Asset Markets', *Annual Review of Financial Economics* **2**(1), 315–346.

Fama, Eugene F. and James D. MacBeth (1973), 'Risk, Return, and Equilibrium: Empirical Tests', *Journal of Political Economy* **81**(3), 607–636.

Fama, Eugene F. and Kenneth R. French (1992), 'The Cross-Section of Expected Stock Returns', *The Journal of Finance* **47**(2), 427–465.

Fama, Eugene F. and Kenneth R. French (1993), 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, Eugene F. and Kenneth R. French (2015), 'A Five-Factor Asset Pricing Model', *Journal of Financial Economics* **116**(1), 1–22.

Fama, Eugene F. and Kenneth R. French (2018), 'Choosing Factors', *Journal of Financial Economics* **128**(2), 234–252.

Gao, Jun, Sheng Zhu, Niall O'Sullivan and Meadhbh Sherman (2019), 'The Role of Economic Uncertainty in UK Stock Returns', *Journal of Risk and Financial Management* **12**(1), 5.

Griffin, Paul A. (2003), 'Got Information? Investor Response to Form 10-K and Form 10-Q EDGAR Filings', *Review of Accounting Studies* **8**, 433–460.

Gupta, Rangan, Hardik A. Marfatia and Eric Olson (2020), 'Effect of uncertainty on U.S. stock returns and volatility: evidence from over eighty years of high-frequency data', *Applied Economics Letters* **27**(16), 1305–1311.

Harvey, Campbell R. and Akhtar Siddique (2000), 'Conditional Skewness in Asset Pricing Tests', *The Journal of Finance* **55**(3), 1263–1295.

Henry, E. (2008), 'Are Investors Influenced By How Earnings Press Releases Are Written?', *Journal of Business Communication* **45**(4), 363–407.

Hou, Kewei, Chen Xue and Lu Zhang (2015), 'Digesting Anomalies: An Investment Approach', *Review of Financial Studies* **28**(3), 650–705.

Hou, Kewei, Haitao Mo, Chen Xue and Lu Zhang (2021), 'An Augmented q-factor Model with Expected Growth', *Review of Finance* **25**(1), 1–41.

Jegadeesh, Narasimhan (1990), 'Evidence of Predictable Behavior of Security Returns', *The Journal of Finance* **45**(3), 881–898.

Jegadeesh, Narasimhan and Di Wu (2013), 'Word power: A new approach for content analysis', *Journal of Financial Economics* **110**(3), 712–729.

Jegadeesh, Narasimhan and Sheridan Titman (1993), 'Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency', *The Journal of Finance* **48**(1), 65–91.

Jurado, Kyle, Sydney C. Ludvigson and Serena Ng (2015), 'Measuring Uncertainty', *American Economic Review* **105**(3), 1177–1216.

Leippold, Markus, Fabio Trojani and Paolo Vanini (2008), 'Learning and Asset Prices Under Ambiguous Information', *Review of Financial Studies* **21**(6), 2565–2597.

Loughran, Tim and Bill Mcdonald (2011), 'When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks', *The Journal of Finance* **66**(1), 35–65.

Loughran, Tim and Bill Mcdonald (2014), 'Measuring Readability in Financial Disclosures: Measuring Readability in Financial Disclosures', *The Journal of Finance* **69**(4), 1643–1671.

Mikolov, Tomas, Kai Chen, Greg Corrado and Jeffrey Dean (2013), 'Efficient Estimation of Word Representations in Vector Space', *Proceedings of the 26th International Conference on Neural Information Processing Systems* **2**, 3111–3119.

Newey, Whitney K. and Kenneth D. West (1987), 'A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica* **55**(3), 703–708.

Pastor, Lubos and Pietro Veronesi (2013), 'Political uncertainty and risk premia', *Journal of Financial Economics* **110**(3), 520–545.

Pennington, Jeffrey, Richard Socher and Christopher Manning (2014), Glove: Global Vectors for Word Representation, *in* 'Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)', Association for Computational Linguistics, Doha, Qatar, pp. 1532–1543.

Sloan, Richard G (1996), 'Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?', *The Accounting Review* **71**(3), 289–315.

Theil, Christoph Kilian, Sanja Štajner and Heiner Stuckenschmidt (2018), Word Embeddings-Based Uncertainty Detection in Financial Disclosures, *in* 'Proceedings of the First Workshop on Economics and Natural Language Processing', Association for Computational Linguistics, Melbourne, Australia, pp. 32–37.

Zhang, Tengfei (2021), 'Manager Uncertainty and the Cross-Section of Stock Returns', *Available at SSRN: https://ssrn.com/abstract=3654376 or http://dx.doi.org/10.2139/ssrn.3654376* p. 67.

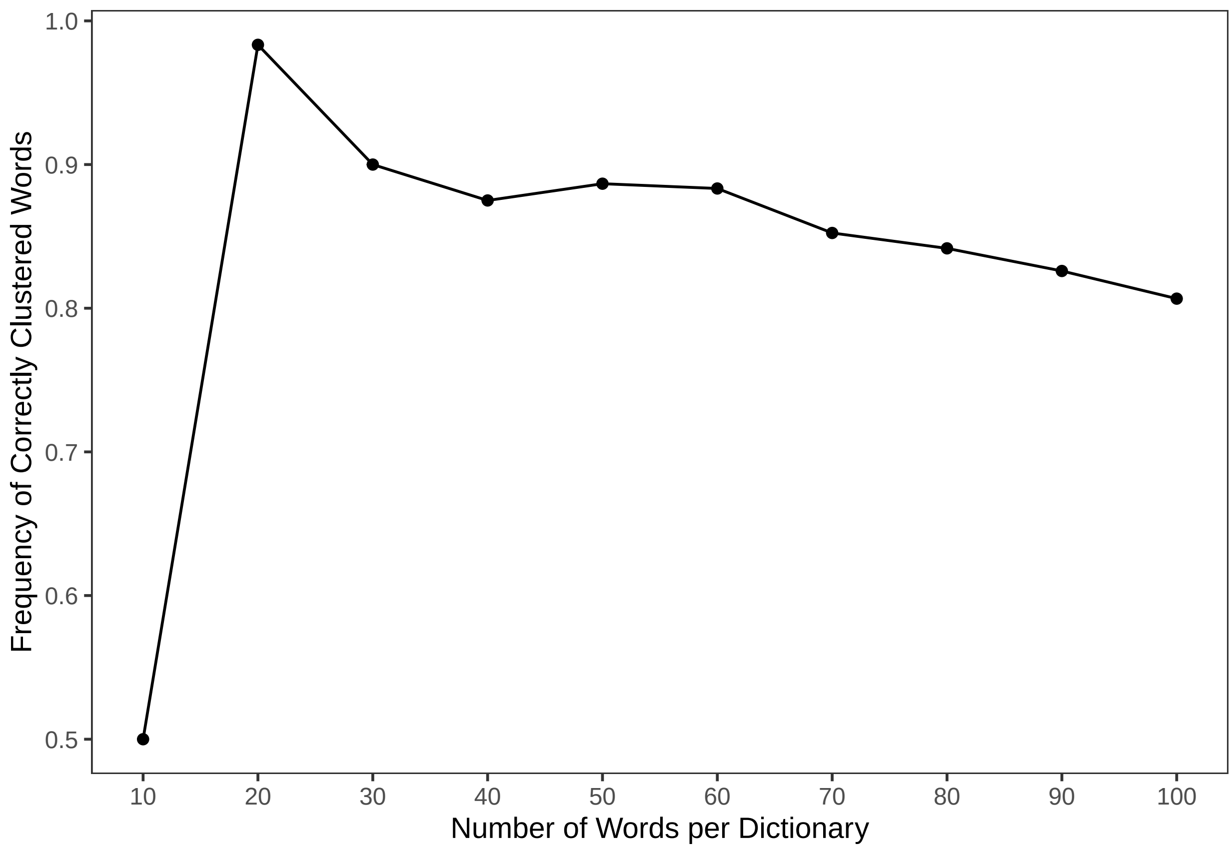

Figure 1: Frequency of Correctly Clustered Words by K-Means Algorithm
This figure shows the frequency of dictionary words which were correctly clustered by a K-Means algorithm applied to the respective word vectors. The K-Means algorithm is implemented with three clusters (uncertainty, risk and sentiment) and a varying number of words per dictionary, which we increase from 10 to 100 in steps of 10. For each word we compare the true cluster based on our derived dictionaries with the cluster assigned by the K-Means algortihm.

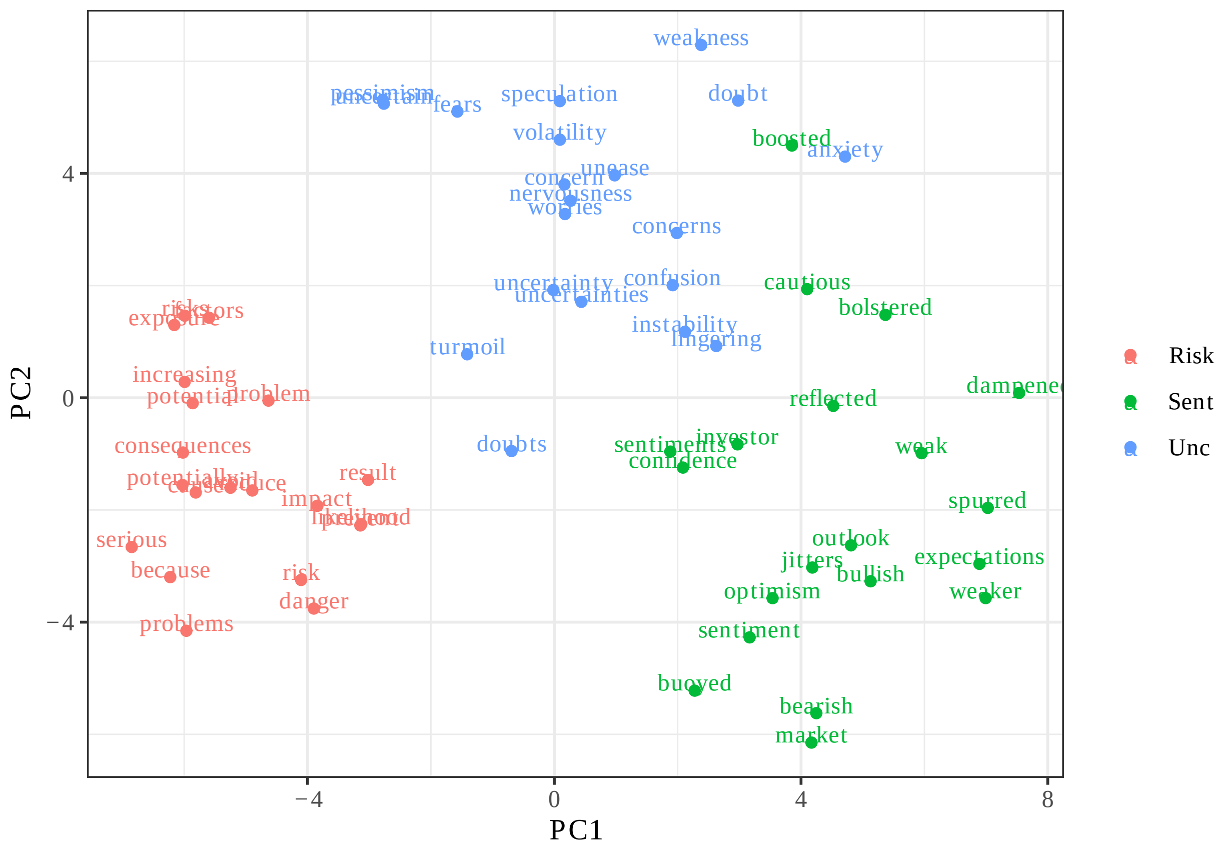

Figure 2: First two PCs of Dictionary Terms.
This figure shows the first two Prinicipal Components (PC) of a Principal Component Analysis (PCA) on the 100-dimensional word vectors of the risk-, sentiment- and uncertainty-related words.

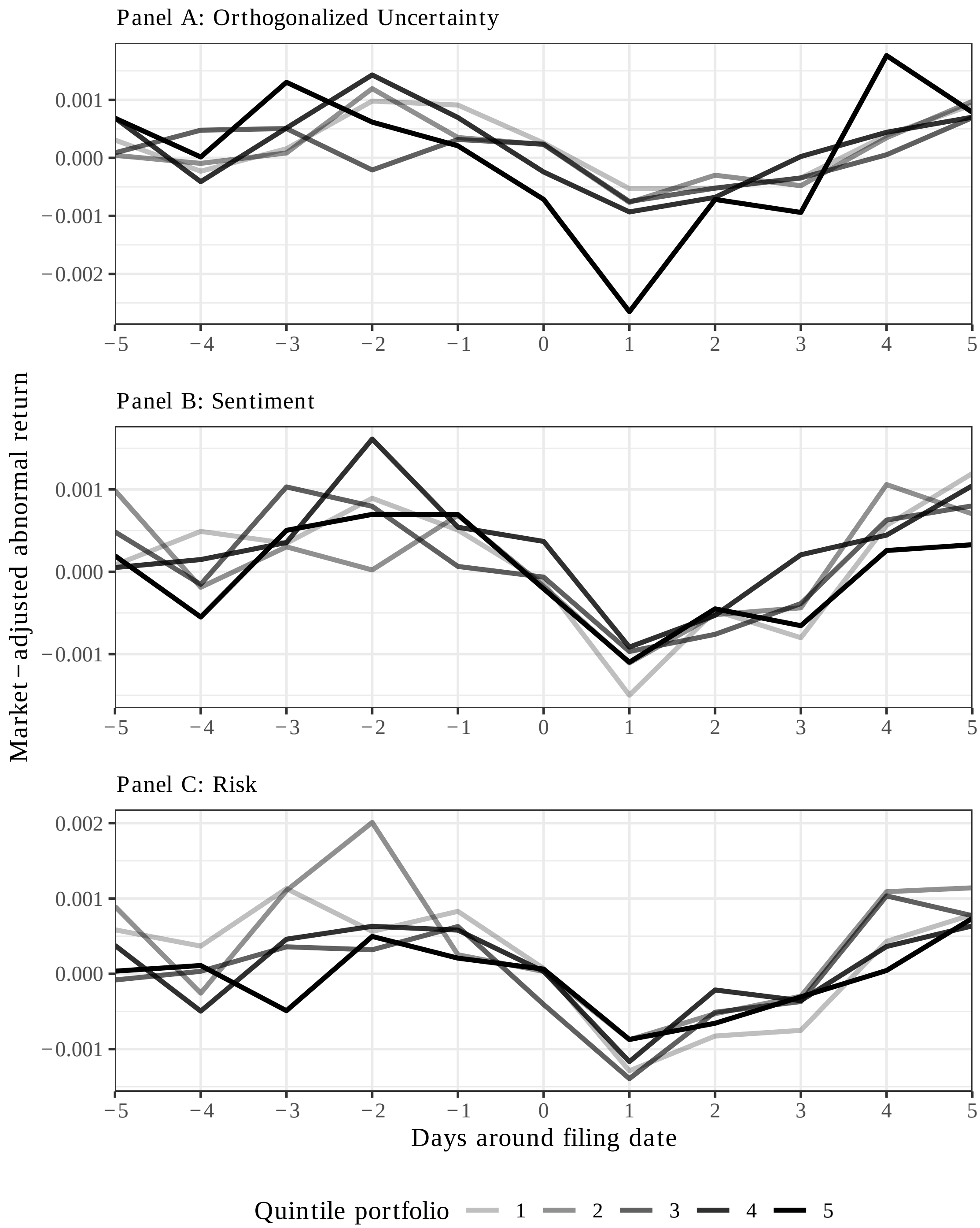

Figure 3: Price reaction to orthogonalized uncertainty, sentiment and risk on the filing day.
The figure presents the daily market-adjusted abnormal returns +-5 trading days around the filing date of our sample 10-K reports. We sort stocks into quintile portfolios given the value for UNC^{Orth}, Sentiment and Risk-related GloVe scores as of the end of the previous month and compute the equal-weighted average for each day in our 10-day event window.

Table 1: Uncertainty, Risk and Sentiment Dictionaries.
This table shows the words assigned to the uncertainty, risk and sentiment dictionaries based on pre-trained GloVe embeddings. Each single subtable consists of two columns reporting the word and the cosine similarity of the respective word to the seed word.

Uncertainty		Risk		Sentiment	
Word	Value	Word	Value	Word	Value
uncertainty	1.000	risk	1.000	sentiment	1.000
uncertainties	0.811	risks	0.897	sentiments	0.732
worries	0.761	danger	0.745	optimism	0.732
turmoil	0.758	exposure	0.741	buoyed	0.719
confusion	0.742	potential	0.728	bullish	0.713
concerns	0.727	cause	0.720	reflected	0.698
volatility	0.725	avoid	0.716	dampened	0.688
concern	0.721	consequences	0.710	confidence	0.681
nervousness	0.716	likelihood	0.709	bearish	0.676
doubts	0.716	serious	0.708	bolstered	0.669
lingering	0.713	potentially	0.707	outlook	0.665
pessimism	0.712	problems	0.698	market	0.662
fears	0.703	prevent	0.697	investor	0.657
uncertain	0.702	factors	0.697	cautious	0.657
speculation	0.701	problem	0.693	weaker	0.657
instability	0.695	impact	0.693	jitters	0.657
unease	0.693	increasing	0.691	expectations	0.655
anxiety	0.685	because	0.691	boosted	0.649
weakness	0.669	reduce	0.691	weak	0.648
doubt	0.667	result	0.690	spurred	0.648

Table 2: Descriptive Statistics.
This table shows the descriptive statistics of various variables. Variables include the one-month-ahead return (R_{t+1}), the uncertainty score (UNC), the risk score (RISK), the sentiment score (SENT), the orthogonalized uncertainty score (UNC^{Orth}), the natural logarithm of market capitalization (SIZE), the book-to-market rato (BM), momentum (MOM), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), short-term reversal (REV), coskewness (COSKEW), maximum return (MAX), market beta (β^{MKT}), VIX beta (β^{VIX}) and the uncertainty beta (β^{UNC}). The one-month ahead return as well as the text-implied measures are given in percentage points. The sample period is March 1994 to December 2020.

Variable	Mean	SD	Skew	Kurt	Min	P05	P25	Median	P75	P95	Max	N
R_{t+1}	1.53	13.17	5.14	155.10	-52.94	-14.89	-4.68	0.62	6.37	20.13	241.38	2844
UNC	0.07	0.04	1.72	16.22	0.00	0.02	0.04	0.06	0.09	0.14	0.42	2857
$RISK$	0.75	0.22	0.58	4.15	0.08	0.43	0.60	0.74	0.88	1.13	1.91	2857
$SENT$	0.33	0.13	1.30	8.54	0.04	0.15	0.23	0.31	0.40	0.56	1.46	2857
UNC^{Orth}	0.00	0.03	1.19	7.01	-0.10	-0.05	-0.02	-0.01	0.01	0.05	0.22	2857
SIZE	13.56	1.77	0.32	3.01	8.13	10.83	12.29	13.46	14.69	16.67	19.75	2857
BM	0.56	0.49	4.97	81.38	0.01	0.09	0.26	0.46	0.73	1.32	8.95	2453
MOM	0.19	0.58	5.24	83.80	-0.82	-0.37	-0.09	0.10	0.33	1.03	10.23	2545
$IVOL$	0.02	0.01	6.79	163.44	0.00	0.01	0.01	0.02	0.02	0.04	0.31	2857
$ILLIQ$	0.98	8.14	17.35	510.01	0.00	0.00	0.00	0.01	0.09	3.04	287.58	2783
REV	0.02	0.13	3.61	91.65	-0.53	-0.15	-0.05	0.01	0.07	0.20	2.01	2856
$COSKEW$	-0.22	0.24	0.16	3.99	-1.23	-0.59	-0.37	-0.22	-0.07	0.18	0.93	2777
MAX	0.06	0.06	9.86	267.30	0.00	0.02	0.03	0.04	0.07	0.13	1.44	2857
β^{MKT}	0.92	0.69	1.55	24.20	-2.09	0.11	0.46	0.79	1.24	2.21	6.64	2682
β^{VIX}	0.00	0.01	0.18	77.54	-0.11	-0.01	0.00	0.00	0.00	0.02	0.11	2857
β^{UNC}	0.02	0.52	0.39	37.04	-4.22	-0.71	-0.21	0.01	0.24	0.78	5.11	2629

Table 3: Correlation of Orthogonalized Uncertainty with Text-Based Measures and Common Firm Characteristics.
This table shows the time series average cross-sectional correlations between various variables including the one-month-ahead return (R_{t+1}), the uncertainty score (UNC), the risk score (RISK), the sentiment score (SENT), the orthogonalized uncertainty score (UNC^{Orth})), the natural logarithm of market capitalization (SIZE), the book-to-market ratio (BM), momentum (MOM), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), short-term reversal (REV), coskewness (COSKEW), maximum return (MAX), market beta (β^{MKT}), VIX beta (β^{VIX}) and the uncertainty beta (β^{UNC}). The sample period is March 1994 to December 2020.

	R_{t+1}	UNC	$RISK$	$SENT$	UNC^{Orth}	$SIZE$	BM	MOM	$IVOL$	$ILLIQ$	REV	$COSKEW$	MAX	β^{MKT}	β^{VIX}
UNC	0.00														
$RISK$	0.00	0.38													
$SENT$	0.00	0.28	0.42												
UNC^{Orth}	0.00	0.89	0.00	0.01											
$SIZE$	0.00	-0.11	0.04	-0.10	-0.10										
BM	0.02	0.00	0.00	0.01	-0.01	-0.35									
MOM	0.01	-0.03	0.00	-0.01	-0.03	0.16	-0.25								
$IVOL$	0.00	0.14	-0.03	0.02	0.15	-0.43	0.23	-0.09							
$ILLIQ$	0.02	0.03	-0.04	0.02	0.04	-0.25	0.17	-0.03	0.26						
REV	0.03	0.00	0.00	0.00	0.00	0.06	-0.11	0.00	0.14	0.02					
$COSKEW$	-0.01	0.02	-0.01	0.01	0.02	-0.08	0.02	-0.01	0.06	0.03	0.01				
MAX	0.01	0.11	-0.02	0.01	0.12	-0.30	0.15	-0.08	0.87	0.20	0.35	0.04			
β^{MKT}	-0.01	0.13	0.02	0.02	0.14	-0.13	0.04	-0.02	0.30	-0.02	0.00	-0.08	0.25		
β^{VIX}	-0.01	0.00	0.01	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	-0.02	0.00	0.01	0.02	
β^{Unc}	-0.01	0.02	0.01	0.01	0.02	0.04	-0.06	0.05	0.02	0.00	0.05	0.09	0.03	0.07	0.00

Table 4: Univariate Sorts of Stocks by Orthogonalized Uncertainty.
For each month, quintile portfolios are formed by sorting individual stocks based on their orthogonalized uncertainty score (UNC^{Orth}), where quintile 1 (5) contains stocks with lowest (highest) UNC^{Orth}. For each portfolio value-weighted one-month-ahead returns are computed. The first row reports the value-weighted excess returns on the quintile portfolios as well as of the high-low portfolio (5-1). The following rows report alphas for various factor models. CAPM is the alpha relative to the market factor; FF3 is the alpha relative to the market, size and book-to-market factor. FF5 ist the alpha relative to the market, size, book-to-market, profitability and investment factor; FF6 is the alpha relative to the market, size, book-to-market, investment, profitability and momentum factor; q-4 is the alpha relative to the market, size, profitability and investment factor; q-5 is the alpha relative to the market, size, profitability, investment and expected growth factor. Newey-West adjusted t-statistics are reported in parentheses. The sample period is March 1994 to December 2020.

Model	1	2	3	4	5	5-1
Excess Return	0.697	0.747	0.815	0.855	0.978	0.281
	(2.32)	(2.88)	(3.33)	(3.25)	(3.93)	(2.02)
CAPM	-0.139	-0.048	0.041	0.079	0.233	0.372
	(-1.16)	(-0.60)	(0.58)	(1.07)	(3.33)	(2.59)
FF3	-0.175	-0.063	0.036	0.091	0.239	0.414
	(-1.79)	(-0.86)	(0.53)	(1.37)	(3.42)	(3.28)
FF5	-0.056	-0.085	-0.051	0.020	0.204	0.260
	(-0.62)	(-1.23)	(-0.79)	(0.29)	(3.17)	(2.36)
FF6	-0.047	-0.093	-0.046	0.017	0.204	0.251
	(-0.54)	(-1.34)	(-0.71)	(0.25)	(3.10)	(2.38)
q-4	-0.219	-0.094	-0.032	0.075	0.229	0.449
	(-2.12)	(-1.19)	(-0.45)	(0.95)	(3.39)	(3.61)
q-5	-0.176	-0.062	-0.048	0.025	0.155	0.331
	(-1.76)	(-0.85)	(-0.61)	(0.32)	(2.73)	(2.59)

Table 5: Fama-MacBeth Cross-Sectional Regressions.
This table reports time-series averages of the slope coefficients obtained from regressing monthly excess returns on the orthogonalized uncertainty (UNC^{Orth}), risk (RISK) and sentiment (SENT) measure and a set of predictor variables using the Fama-MacBeth methodology. The control variables are the natural logarithm of market capitalization (SIZE), book-to-market ratio (BM), momentum (MOM), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), short-term reversal (REV), coskewness (COSKEW), maximum return (MAX), market beta (β^{MKT}), VIX beta (β^{VIX}), uncertainty beta (β^{UNC}). The cross-sectional regressions are run at a monthly frequency from March 1994 to December 2020. Newey-West adjusted t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	1.56	1.56	1.51	1.56	2.95	2.92	3.09	3.09
	(6.59)	(5.99)	(5.85)	(5.72)	(4.94)	(4.90)	(5.09)	(5.12)
UNC^{Orth}	8.67			8.73	4.30			4.45
	(4.80)			(4.81)	(3.44)			(3.61)
RISK		0.03		0.01		0.17		0.21
		(0.16)		(0.05)		(0.89)		(1.29)
SENT			0.09	0.16			-0.38	-0.34
			(0.32)	(0.70)			(-1.57)	(-1.45)
SIZE					-0.22	-0.23	-0.23	-0.23
					(-5.44)	(-5.54)	(-5.52)	(-5.61)
BM					0.47	0.43	0.44	0.46
					(4.04)	(3.70)	(3.69)	(4.05)
MOM					-0.00	-0.01	-0.01	-0.01
					(-0.01)	(-0.03)	(-0.05)	(-0.06)
IVOL					36.08	37.23	36.96	35.67
					(5.52)	(5.60)	(5.55)	(5.47)
ILLIQ					0.04	0.04	0.04	0.04
					(1.90)	(1.76)	(1.84)	(1.82)
REV					-3.75	-3.79	-3.79	-3.83
					(-7.00)	(-7.08)	(-7.03)	(-7.19)
COSKEW					0.34	0.37	0.38	0.37
					(1.79)	(1.90)	(1.93)	(1.90)
MAX					1.76	1.70	1.67	1.85
					(1.18)	(1.15)	(1.10)	(1.23)
β^{MKT}					0.61	0.64	0.64	0.60
					(2.93)	(3.00)	(3.03)	(2.90)
β^{VIX}					-5.17	-5.35	-5.23	-5.18
					(-1.13)	(-1.16)	(-1.14)	(-1.14)
β^{UNC}					-0.22	-0.23	-0.23	-0.20
					(-1.53)	(-1.56)	(-1.52)	(-1.39)
N	2885	2885	2885	2885	2312	2312	2312	2312
R^2	0.19	0.12	0.12	0.39	5.96	5.96	5.97	6.09

Table 6: Average Stock Characteristics.
This table reports the time series averages of the slope coefficients from the regressions of the orthogonalized uncertainty measure on various firm-level characteristics using the Fama-MacBeth methodology. The considered stock-specific variables are the book-to-market rato (BM), the natural logarithm of market capitalization (SIZE), momentum (MOM), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), short-term reversal (REV), coskewness (COSKEW), maximum return (MAX), market beta (β^{MKT}), VIX beta (β^{VIX}) and the uncertainty beta (β^{UNC}). The cross-sectional regressions are run at a monthly frequency from March 1994 to December 2020. Newey-West adjusted t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Intercept	2.14	-0.72	0.11	0.07	-0.69	-0.43	0.03	-0.07	-0.36	-0.34	-1.59
	(11.58)	(-15.88)	(5.52)	(3.91)	(-23.45)	(-26.32)	(1.43)	(-11.39)	(-14.11)	(-13.27)	(-7.57)
SIZE	-0.17										0.07
	(-12.28)										(4.82)
β^{MKT}		0.64									0.52
		(16.58)									(12.59)
BM			-0.07								-0.55
			(-3.22)								(-8.51)
MOM				-0.30							-0.15
				(-4.80)							(-3.79)
IVOL					27.83						28.00
					(17.27)						(7.70)
MAX						5.83					-2.80
						(19.70)					(-6.54)
COSKEW							0.34				0.33
							(4.35)				(3.88)
ILLIQ								0.01			-0.01
								(4.48)			(-2.03)
β^{VIX}									2.70		0.81
									(1.93)		(0.84)
β^{UNC}										0.07	-0.10
										(1.82)	(-2.24)
N	3593	3380	3074	3238	3592	3593	3496	3500	2898	2652	2323
R^2	1.15	2.38	0.27	0.47	2.81	1.68	0.18	0.29	0.12	0.15	4.21

Table 7: Filing Period Abnormal Return Regression.
This table reports the estimates of the regression of filing period abnormal returns,defined as a firm's buy-and-hold return minus the market return over the four-day window of [filingdate, filing date +3] against the orthogonalized uncertainty, risk and sentiment score and various control variables, including Accruals (ACCRUALS), book-to-market ratio (BM), the natural logarithm of the stock's market capitalization (SIZE), turnover (TURNOVER) and volatility (VOLATILITY). The coefficients are based on quarterly Fama-MacBeth regressions in a sample period from March 1994 to December 2020. Newey-West adjusted t-statistics are reported in parentheses

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	-0.06	0.11	-0.21	-0.14	0.55	0.75	0.57	0.83
	(-0.95)	(0.69)	(-1.10)	(-0.68)	(0.61)	(0.79)	(0.62)	(0.92)
UNC^{Orth}	-7.27			-7.31	-4.89			-4.72
	(-3.92)			(-3.89)	(-2.39)			(-2.44)
SENT		-47.00		-78.05		-45.19		-67.03
		(-1.15)		(-1.70)		(-1.06)		(-1.31)
RISK			10.29	36.33			-15.28	6.13
			(0.44)	(1.36)			(-0.66)	(0.21)
ACCRUALS					1.18	1.35	1.33	1.32
					(1.36)	(1.59)	(1.55)	(1.53)
BM					0.13	0.16	0.17	0.13
					(1.23)	(1.43)	(1.46)	(1.16)
SIZE					0.04	0.05	0.05	0.03
					(1.00)	(1.19)	(1.18)	(0.70)
TURNOVER					-0.07	-0.09	-0.08	-0.07
					(-0.86)	(-1.05)	(-0.93)	(-0.91)
VOLATILITY					-1.34	-1.19	-1.38	-1.25
					(-1.06)	(-0.91)	(-1.09)	(-0.97)
N	906	906	906	906	696	696	696	696
R^2	0.25	0.19	-0.02	0.46	2.67	2.75	2.44	2.94

Table 8: Pre/Post-Filing Period Abnormal Return Regression.
This table reports the estimates of the regression of pre- and post- filing period abnormal returns, defined as a firm's buy-and-hold return minus the market return over the windows of [filing date -20, filing date -1] and [filing date +5, filing date +9] against the orthogonalized uncertainty, risk and sentiment score and various control variables, including Accruals (ACCRUALS), book-to-market ratio (BM), the natural logarithm of the stocks market capitalization (SIZE), turnover (TURNOVER) and volatility (VOLATILITY). The coefficients are based on quarterly Fama-MacBeth regressions in a sample period from March 1994 to December 2020. Newey-West adjusted t-statistics are reported in parentheses

	Pre-filing date [-20, -1]				Post-filing date [+5, +9]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.60	1.09	-4.28	-3.87	0.25	-0.09	-0.38	-0.55
	(2.92)	(2.62)	(-2.12)	(-1.83)	(2.79)	(-0.54)	(-0.41)	(-0.57)
UNC^{Orth}	1.44	1.36	-1.80	-1.76	-0.27	-0.35	0.80	0.43
	(0.39)	(0.37)	(-0.47)	(-0.45)	(-0.14)	(-0.18)	(0.45)	(0.23)
SENT		-21.64		-42.68		43.31		41.99
		(-0.45)		(-0.88)		(1.71)		(1.34)
RISK		-79.62		-21.45		32.50		-1.86
		(-1.01)		(-0.23)		(0.78)		(-0.04)
ACCRUALS			-2.83	-2.78			0.86	0.89
			(-1.61)	(-1.59)			(1.07)	(1.05)
BM			-0.46	-0.46			0.28	0.25
			(-2.49)	(-2.56)			(2.29)	(2.00)
SIZE			-0.05	-0.04			-0.07	-0.07
			(-0.46)	(-0.40)			(-1.54)	(-1.58)
TURNOVER			0.38	0.38			0.09	0.09
			(1.99)	(2.00)			(1.16)	(1.16)
VOLATILITY			4.46	4.14			0.40	0.01
			(1.38)	(1.29)			(0.30)	(0.01)
N	906	906	696	696	906	906	696	696
R^2	0.31	0.36	3.4	3.35	0.23	0.31	3.18	3.36

Table 9: Panel Regressions with Firm Fundamentals.
This table reports the estimates of regressions of firm fundamentals against the orthogonalized uncertainty, risk and sentiment score and various control variables. Considered firm fundamentals are the return on assets (ROA), standardized unexpected earnings (SUE) and a dividend dummy (DIV), which is set to one if a firm increases its dividend in the fiscal year t. The control variables include Accruals (ACCRUALS), book-to-market ratio (BM), the natural logarithm of the stocks market capitalization (SIZE), turnover (TURNOVER) and volatility (VOLATILITY). The coefficients are based on panel regressions with a yearly frequency, regressing firm fundamentals in fiscal year t+1 on text-implied measures and controls in year t with a sample period from 1993 to 2020. z-statistics based on double clustered and heteroscedasticity-consistent standard errors are reported in parentheses

	ROA		SUE		DIV	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.085	-0.044	-0.005	-0.036	-0.421	3.819
	(1.93)	(-0.74)	(-0.60)	(-0.40)	(-3.31)	(5.80)
UNC^{Orth}	-1.616	-0.839	-0.952	-0.684	-11.505	-5.959
	(-5.81)	(-5.48)	(-2.15)	(-1.73)	(-10.71)	(-7.52)
RISK	-0.079	-0.083	-0.015	-0.033	0.093	-0.193
	(-4.25)	(-5.70)	(-0.90)	(-1.18)	(0.64)	(-1.25)
SENT	-0.117	0.068	-0.031	0.018	-0.278	0.971
	(-1.44)	(3.09)	(-1.21)	(0.62)	(-1.34)	(4.41)
ACCRUALS		0.093		0.105		-0.224
		(2.77)		(2.01)		(-0.84)
BM		-0.002		-0.000		-0.083
		(-0.88)		(-0.10)		(-1.84)
SIZE		0.026		0.018		0.359
		(6.66)		(1.69)		(14.78)
TURNOVER		-0.014		-0.014		-0.541
		(-2.18)		(-2.13)		(-11.88)
VOLATILITY		-0.694		-0.131		-14.112
		(-5.87)		(-1.83)		(-10.95)
R^2	0.245	9.447	0.177	0.353	2.556	23.491
N	50498	40616	53048	42531	48536	39146

Table 10: Bivariate Portfolio Sorts.
This table reports average monthly value-weighted returns of portfolios sorted by control variables and the orthogonalized uncertainty score (UNC^{Orth}). Within each characteristic quintile, we sort stocks into five additional portfolios based on UNC^{Orth} and then average each of the orthogonalized uncertainty portfolios across the five quintiles from the first sort. We compute the returns of the resuling portfolios over the subsequent month, as well as the return on the 5-1 portfolio. We furthermore compute the risk-adjusted return of the 5-1 portfolio between relative to the FF6 and q-5 model. Newey-West adjusted t-statistics are reported in parentheses. The sample period is March 1994 to December 2020.

Variable	1	2	3	4	5	5-1	5-1 (FF6 α)	5-1 (q-5 α)
Size	1.181	1.334	1.585	1.720	1.761	0.553	0.511	0.537
	(9.46)	(9.60)	(9.47)	(9.95)	(10.36)	(4.83)	(4.49)	(4.47)
β^{MKT}	0.848	1.186	1.143	1.100	1.202	0.351	0.271	0.343
	(4.27)	(3.85)	(6.33)	(6.25)	(7.31)	(2.18)	(1.61)	(1.89)
BM	0.849	0.921	1.336	1.269	1.081	0.207	0.158	0.350
	(5.88)	(6.27)	(6.13)	(6.69)	(7.66)	(1.96)	(1.41)	(2.93)
MOM	0.681	0.925	1.108	1.031	1.114	0.422	0.332	0.375
	(4.52)	(4.42)	(6.54)	(6.22)	(7.32)	(3.81)	(2.89)	(3.04)
IVOL	0.726	1.049	1.023	1.036	1.109	0.385	0.350	0.365
	(4.33)	(5.62)	(5.94)	(5.87)	(6.40)	(3.03)	(2.62)	(2.57)
MAX	0.701	0.964	1.149	1.016	1.150	0.393	0.385	0.435
	(4.49)	(5.89)	(6.78)	(5.86)	(6.92)	(3.25)	(3.08)	(3.23)
COSKEW	0.717	0.894	0.897	1.444	1.081	0.338	0.295	0.421
	(5.22)	(6.17)	(6.25)	(4.33)	(8.03)	(3.07)	(2.60)	(3.41)
ILLIQ	0.942	1.234	1.406	1.604	1.462	0.525	0.417	0.431
	(7.70)	(8.36)	(8.87)	(8.43)	(9.08)	(4.48)	(3.63)	(3.45)
β^{VIX}	0.736	0.846	0.987	1.155	1.093	0.328	0.266	0.236
	(5.11)	(5.70)	(6.44)	(6.96)	(7.32)	(2.73)	(2.12)	(1.76)
β^{Unc}	0.776	0.940	1.140	1.048	1.113	0.330	0.337	0.379
	(5.20)	(6.44)	(6.68)	(6.08)	(7.64)	(2.77)	(2.76)	(2.89)

Table 11: Fama-MacBeth Cross-Sectional Regressions, Filings Months only.
This table reports time-series averages of the slope coefficients obtained from regressing monthly excess returns on the orthogonalized uncertainty (UNC^{Orth}), risk (RISK) and sentiment (SENT) measure and a set of predictor variables using the Fama-MacBeth methodology. The control variables are the natural logarithm of market capitalization (SIZE), book-to-market ratio (BM), momentum (MOM), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), short-term reversal (REV), coskewness (COSKEW), maximum return (MAX), market beta (β^{MKT}), VIX beta (β^{VIX}), uncertainty beta (β^{UNC}). The cross-sectional regressions are run at a monthly frequency from March 1994 to December 2020. Newey-West adjusted t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	4.01	3.77	2.70	3.25	7.79	8.12	7.54	7.63
	(10.23)	(6.31)	(6.51)	(5.42)	(3.67)	(3.90)	(3.63)	(3.55)
UNC^{Orth}	28.36			28.39	11.83			15.18
	(5.26)			(4.91)	(3.10)			(3.71)
RISK		0.63		-1.01		-0.34		-1.00
		(0.86)		(-1.15)		(-0.37)		(-0.98)
SENT			4.08	4.95			1.81	2.39
			(4.01)	(3.93)			(1.61)	(1.74)
SIZE					-0.62	-0.63	-0.63	-0.61
					(-4.74)	(-4.82)	(-4.93)	(-4.53)
BM					0.70	0.48	0.45	0.63
					(0.98)	(0.67)	(0.64)	(0.86)
MOM					-0.04	-0.10	-0.09	0.01
					(-0.09)	(-0.26)	(-0.22)	(0.02)
IVOL					143.73	148.70	143.02	142.61
					(4.74)	(4.52)	(4.67)	(4.24)
ILLIQ					-0.40	-0.42	-0.38	-0.37
					(-1.67)	(-1.75)	(-1.74)	(-1.68)
REV					1.96	2.41	1.92	1.99
					(1.45)	(1.83)	(1.42)	(1.47)
COSKEW					-0.74	-0.72	-0.56	-0.73
					(-0.86)	(-0.93)	(-0.67)	(-0.90)
MAX					-16.43	-16.73	-15.33	-16.18
					(-2.27)	(-2.19)	(-2.07)	(-2.11)
β^{MKT}					1.36	1.32	1.25	1.37
					(5.10)	(4.94)	(4.47)	(4.77)
β^{VIX}					10.15	13.13	8.08	7.18
					(0.57)	(0.77)	(0.51)	(0.38)
β^{UNC}					0.29	0.24	0.24	0.21
					(0.70)	(0.59)	(0.62)	(0.55)
N	266	266	266	266	215	215	215	215
R^2	0.77	0.35	0.16	1.08	16.9	16.35	16.34	16.73

Table 12: Variable Definitions for Control Variables.
This table contains an overview of the control variables we use in our empirical analysis. The first column shows the variable name. The second column provides a description of the respective variable. The last column shows the data source and a link to the public source (if available).

Variable	Description	Source
MKTRF	Excess return of the CRSP value-weighted market portfolio over the one-month Treasury bill rate	Kenneth R. French
SMB	Small minus big factor return as defined in Fama and French (1993)	Kenneth R. French
HML	High minus low factor return as defined in Fama and French (1993)	Kenneth R. French
MOM	Momentum factor as defined by Carhart (1997)	Kenneth R. French
RMW	Robust minus weak factor as defined by Fama and French (2015)	Kenneth R. French
CMA	Conservative minus aggressive factor as defined by Fama and French (2015)	Kenneth R. French
R_{MKT}	Market factor as defined by Hou et al. (2015)	Hou, Xue, Zhang
R_{ME}	Size factor returns as defined by Hou et al. (2015)	Hou, Xue, Zhang
$R_{I/A}$	Investment factor as defined by Hou et al. (2015)	Hou, Xue, Zhang
R_{ROE}	Profitability factor as defined by Hou et al. (2015)	Hou, Xue, Zhang
R_{EG}	Expected growth factor as defined by Hou et al. (2021)	Hou, Xue, Zhang
β^{MKT}	Market beta estimated by a 60-month rolling window regression defined by Fama and MacBeth (1973)	Chen, Zimmermann
SIZE	Size defined as the natural logarithm of the stocks market capitalization	CRSP
BM	Book value of equity to market value of equity ratio as defined by Fama and French (1992)	Chen, Zimmermann
MOM	Momentum over 12 months as defined by Jegadeesh and Titman (1993)	Chen, Zimmermann
IVOL	Idiosyncratiy volatility measured as the standard deviation of residuals from Fama-French three factor regressions with daily data defined by Ang et al. (2006)	Chen, Zimmermann
ILLIQ	Illiquidity computed as the 12-month average of the daily return divided by the turnover defined by Amihud (2002)	Chen, Zimmermann
REV	Short-term reversal specified as the stock return in month t as defined by Jegadeesh (1990)	Chen, Zimmermann
β^{VIX}	The VIX beta estimated by a 1-month rolling window regression of daily excess stock returns on the daily change in the VIX as defined by Ang et al. (2006)	Chen, Zimmermann
COSKEW	Coskewness estimated as $E[\tilde{r}_{it}\tilde{r}_{mt}^2]/(SD[\tilde{r}_{it}]SD[\tilde{r}_{mt}]^2$ where $\tilde{r}_{it}$ is the de-meaned stock excess return and $\tilde{r}_{mt}$ is the de-meaned market excess return as defined by Harvey and Siddique (2000)	Chen, Zimmermann
MAX	Maximum daily return over the previous month as defined by Bali et al. (2011)	Chen, Zimmermann
β^{UNC}	Uncertainty beta estimated by a 60-month rolling window regression of excess stock returns on the macroeconomic uncertainty index by Jurado et al. (2015) and risk factors defined by Bali et al. (2017)	CRSP
ACCRUALS	Accruals measured as the annual change in current total assets minus the annual change in cash and short-term investments, minus the annual change in current liabilities and debt in current liabilities minus the change in income taxes all divided by average total assets defined by Sloan (1996)	Chen, Zimmermann
VOLATILITY	Volatility measured as the standard deviation of the firm-specific component of returns using 60 months of data as of the end of the month prior to filing defined by Jegadeesh and Wu (2013)	CRSP
TURNOVER	Turnover computed as the natural logarithm of the number of shares traded during the period from six to 252 trading days before the filing date divided by the number of shares on the filing date defined by Jegadeesh and Wu (2013)	CRSP
ROA	Income before extraordinary items scaled by average total assets	Compustat
SUE	Standardized unexpected earnings defined as absolute value of (Actual EPS minus median IBES EPS estimate) scaled by stock price.	Compustat, IBES (Datastream)
DIV	Dividend increase dummy set to one if a firm increases ist dividends in year t	Compustat