Asymmetric News Repetition and the Cross-section of Stock Returns*

Ruicong Li†

this draft: November, 2020
check for the latest version

Abstract

This paper studies the implications of disclosure repetitiveness on firm performance, information processing costs, and future stock returns. I propose an entropy-based measure of disclosure repetitiveness – the information redundancy ratio – which is assumption-free with respect to the underlying language models. I then decompose my measure into news redundancy ratio (NRR) and stale-news redundancy ratio (SNR), which allow me to separately analyze managers' disclosure behaviors regarding new and old contents compared to the previous year. In contrast with previous studies, I find that operating performance and filing announcement returns are positively (negatively) correlated with NRR (SRR) and that investors under-react to information in NRR. A portfolio with long (short) positions on high (low) NRR stocks generates value-weighted alphas of 5%-11% per annum. These results are consistent with the notion that managers present good (bad) news with more (less) repetition and repeat stale-news to obfuscate unfavorable information, while investors have difficulty extracting news due to limited attention.

JEL Classification: D8, G12, G14, M41
Keywords: Disclosure redundancy, Processing cost, Shannon's Entropy, Stale news, Return prediction, Textual analysis

*I would like to thank Emilio Bisetti, Zhanhui Chen, Redouane Elkamhi, Paul Gao, Yan Ji, Kai Li, Peter MacKay, Abhiroop Mukherjee, John Nash, Yoshio Nozawa, George Panayotov, Yan Xiong, Haifeng You, Jialin Yu, Alminas Zaldokas, Chu Zhang, Zilong Zhang, and the seminar participants at HKUST for helpful comments and suggestions. All remaining errors are my own.

†Ph.D. Candidate, Department of Finance, Hong Kong University of Science and Technology (HKUST); Email: rliah@connect.ust.hk.

1 Introduction

In many communication scenarios, the way how we present information affects its processing cost. In the financial market, disclosures in annual and quarterly reports are among the major communication channels from managers to outside investors, and the disclosure processing cost is key to market efficiency (Grossman and Stiglitz, 1980). It has been documented that managers are more likely to reduce the processing costs of favorable information and increase costs for unfavorable information. Manipulating disclosure redundancy[1] (repetition) is believed to serve this purpose, but the underlying mechanism is not clearly understood yet.

It has been argued that redundant disclosure could increase processing costs as it may increase the difficulty of separating relevant from irrelevant information. Existing empirical studies generally support this view and link discretionary redundancy to managerial obfuscation rather than facilitating communication (Li, 2008; Loughran and McDonald, 2014; Lang and Stice-Lawrence, 2015; Cazier and Pfeiffer, 2017; Blankespoor et al., 2020). On the other hand, information and communication theory predicts that adding redundancy should lower information processing costs because it can increase the message's salience and effectively reduce equivocation and communication error (Shannon and Weaver, 1949; Nakano, 1972; Hsia, 1977; Stephens, 2007; Stephens and Rains, 2011). This paper attempts to reconcile this discrepancy by decomposing redundancy into news redundancy and stale-news redundancy and further explores the asset pricing implications of news redundancy in the cross-section.

Measuring the redundancy of a document is essentially measuring the amount of information that is not redundant. This paper quantifies non-redundant information using Shannon's Entropy,[2] which can be proxied by the smallest possible compressed

[1]Examples of redundancy-related linguistic properties in financial reports include repetitiveness of phrases, textual similarity across different parts, and degrees of boilerplate language. These language features are partially due to regulatory requirements, but management discretion plays a key role.

[2]People characterize the amount of information in a non-random sequence using Kolmogorov complexity, which denotes the length of the shortest description for the sequences (Li et al., 2008). In this paper, we use Shannon's entropy to denote the amount of information for convenience. Using entropy in quantifying information has a long history in theoretical economics and finance literature (Sims, 2003; Peng and Xiong, 2006; Van Nieuwerburgh and Veldkamp, 2009; Mackowiak and Wiederholt, 2009; Caplin and Dean, 2015; Dewan and Neligh, 2020). This literature generates insightful results and inspires this project, but it is not easy to empirically calculate an as-rigorous entropy for numerical variables.

length in bits achieved by lossless compression algorithms (Shannon, 1948). Thus, I define a document's information redundancy ratio as the fraction of the raw file length that can be compressed without information loss. The complementary quantity of redundancy is information density, measured by the length of the optimally compressed file divided by the original file's length. Information redundancy and density add up to one. The more repeated phrases, the more redundant the file is. Highly repetitive disclosure gives the financial report a higher redundancy ratio.

For an investor without any prior knowledge of the company, most of the financial report information is news to her. Its redundancy (density) quantifies the sparseness (richness) of news in the file. For an average marginal investor who is reasonably sophisticated, however, the financial report contains both news and stale-news. The redundancies of news and stale-news may play different roles and have different implications on firm value because communicating news is the primary concern for managers. Decomposing redundancy into news redundancy and stale-news redundancy could provide more insights.

To decompose redundancy, we use the previous year's annual report as the benchmark to quantify news and stale-news in the current annual report. Given news and stale-news are likely to be mixed even in the same sentence, it is not practical to directly identify news and stale news. However, it is possible to indirectly measure the quantity of news using the incremental amount of information in the appended documents relative to the benchmark document. Thus, the current annual report's news redundancy ratio is measured as one minus the news density, which is the ratio of the quantity of news to the original file length. The higher the news redundancy of a filing, the less news per page it contains. Stale-news redundancy is constructed similarly, elaborated in section 2.3.

After some validation exercises, I disentangle the effects of news redundancy (NRR) and stale-news redundancy (SNR) on disclosure processing costs by examining their relationships with earnings, filing announcement returns, and analyst following. The first two tests are based on the assumption that managers tend to reduce (raise) processing costs when firm performance is good (bad) (Blankespoor et al., 2020). I regress current and future earnings on NRR and SNR, and I find positive and statistically significant coefficients on NRR, negatively significant coefficients on SNR (Table 2 and 3). These results suggest that adding redundancy to news may facilitate communication while adding redundancy to stale-news conveys a bad signal about firm performance. The total redundancy ratio behaves similarly to SNR, consistent with the evidence of managerial obfuscation in previous literature (Cazier and Pfeiffer, 2017).

The second test corroborates the above findings and helps confirm that high news (stale-news) redundancy is indeed associated with good (bad) news. I regress filing announcement returns on NRR and SNR, respectively. The coefficients on NRR are significantly positive, while SNR's coefficients are significantly negative (Table 4). These results are consistent with that managers generally[3] disclose good news with more redundancy than bad news and that high stale-news redundancy conveys bad signals.

The third test is based on the assumption that analysts respond to investors' demand for investment information for firms whose disclosures are more costly to process (Lehavy et al., 2011). If filings with higher news redundancy have higher (lower) processing costs, then news redundancy should positively (negatively) predict future analyst following. I find that NRR is negatively predicting analyst following in the next period, but the effect of SNR is insignificant (Table 5). These results suggest that filings with a higher news redundancy ratio have lower processing costs, and stale-news redundancy does not matter for analyst coverage.

Despite the value-relevant information in the news redundancy of filings, investors are not fully aware of this information for a long time. This claim is supported by a battery of tests on the return prediction effects. The equal-weighted quintile portfolio alphas are nearly monotonic in NRR (Table 6), and the long-short portfolio shows that stocks with high NRR outperform stocks with low NRR 0.23%-0.44% per month (2.76%-5.28% per year) depending on the factor model used. The t-statistics for equal-weighted long-short alphas are from 1.99 to 4.41. The economic magnitudes of excess return and alphas of the value-weighted long-short portfolio are larger than equal-weighted, ranging from 0.45% to 0.61% per month (5.4% to 7.32% per year). The

[3]Earnings management incentives would generate the opposite patterns of filing announcement return, which works against my results.

t-statistics are smaller but still significant, ranging from 1.97 for Stambaugh-Yuan 4-factor alpha to 2.73 for Fama-French 3-factor alpha. The strategy generates most of the profits in the long leg (Q5). The results from Fama-MacBeth regressions (Table 8) confirm the findings of portfolio sorting (Table 6). I find NRR remains a significant predictor of future returns even after including the prevailing known predictors in asset pricing literature. I also do a series of robustness checks on the NRR effect. The return predictability does not reverse (Table 10), cannot be fully explained by limits to arbitrage (Table 7,15) and remains robust for large stocks (Table 9), in crisis periods year 2007-2009 (Table 11), and after controlling for previously documented textual measures (Table 12).

Next, I explore the potential mechanisms behind the return prediction effect of NRR. First, I regress the changes in news redundancy on textual characteristics changes, including positive tone and negative tone. I find that the proportion of positive words has a significantly positive association with news redundancy. In contrast, the changes in the proportion of negative words have an insignificant or negatively significant association with NRR (Table 13). These findings are consistent with managers strategically disclose good (bad) news more (less) redundantly.[4] Then I test to what extent investors are surprised by subsequent earnings realizations (Table 14). The Fama-MacBeth regressions of future earnings announcement return on the current news redundancy ratio generate significant coefficients, indicating investors are surprised by the subsequent good (bad) news associated with high (low) NRR. These results suggest that the prediction effect of NRR is more likely due to mispricing rather than risk. A natural question followed is whether the predictability is due to limits to arbitrages. I do a series of double sorting tests dependent on measures of difficulty in valuing or trading the stocks (Table 15). The differences of Q5-Q1 alpha between high and low groups of limits-to-arbitrage proxies are not always consistent with predictions by limits to arbitrage. The results on an equal-weighted basis are not always consistent with those on a value-weighted basis. Arguments of limits to arbitrage cannot fully

[4]In other words, adding entropy (novel, unusual, and complicated language) when presenting bad news contributes to obfuscation because high-entropy information is much more attention-costly. These results also serve as evidence for entropy-based learning capacity constraints commonly used by rational inattention theories.

address the NRR effect.

In the following group of tests, I construct a measure of look-back attention, LBA, which is the moving average of $ln(1 + HESV)$ for the past six months. $HESV$ is the monthly aggregate value of historical Edgar search volume defined as the sum of human user requests for 10-Ks/Qs that are publicly available for greater than or equal to 360 days following Drake et al. (2016). Then I define a dummy variable indicating a stock belonging to the low LBA group, $LowLBA_{i,t}$, which equals one if $LBA_{i,t}$ is below the median LBA of the cross-section in month t. I run Fama-MacBeth regressions of future stock returns on the NRR, $LowLBA$, and an interaction term, $LowLBA \times NRR$. The results (Table 16) show that the NRR effect is stronger for stocks receiving lower look-back attention from investors. This result supports the limited attention channel from the supply side of investor attention. Further evidence shows that stocks with more lengthy filings exhibit a more substantial NRR effect (Table 17), which corroborates the limited attention channel from the demand side. The above return prediction results suggest that investors face attention constraints and acquire the information in news redundancy with a delay.

The remainder of the paper is organized as follows. Section 1.1 reviews the prior literature and discusses the contributions of this paper. Section 2 describes the data sources, construction of information density and redundancy, and the decomposition into news and stale-news components. Section 3 validates news redundancy and stale-news redundancy and compares their relationships with analyst following, earnings, and filing announcement returns. Section 4 examines the return prediction effect of the news redundancy ratio. Section 5 explores the potential mechanisms behind the return prediction effect of news redundancy. Section 6 concludes.

1.1 Related literature and contributions

This paper contributes to several growing strands of literature, including but not limited to: (a) topics of under-reaction in stock prices and the impact of investor limited attention, (b) management incentives and asymmetric disclosure processing costs regarding good news and bad news, and (c) textual measures to quantify information flow in the financial market.

The return prediction effect of news redundancy ratio contributes to the literature on stock price underreaction and investor inattention. Previous literature has documented that the extent of price underreaction is positively related to investor inattention, as is reviewed by Tetlock (2014). Ben-Rephael et al. (2017) use Bloomberg search activity to measure institutional attention and show that price drift is more pronounced for stocks with less institutional attention. Cohen et al. (2020) show a large magnitude of return predictability by textual changes in financial reports, and investors are systematically inattentive to changes in major firm disclosures. By contrast, I extract return prediction signals based on the subtle relationship between the richness of news and firm performance resulting from management incentives and investors' processing costs. My results propose a mechanism where managers' private information is revealed through their asymmetric disclosure behaviors.

The comparison between news and stale-news in their relationships to information processing cost and management incentives reconciles the discrepancy between the previous empirical evidence and the predictions of information and communication theory. It provides more granular evidence on management disclosure behaviors. Cazier and Pfeiffer (2017) find a negative relationship between earnings and disclosure redundancy and interpret this result as indicating higher redundancy leads to higher processing cost. I confirm their findings using my total redundancy measure and further decompose total redundancy into news and stale-news components. I find that news redundancy is positively related to earnings, while stale-news is negatively associated with earnings. Since reducing processing cost for the good news is compatible with management incentives (Blankespoor et al., 2020), my results indicate that adding redundancy actually lowers the processing cost, consistent with information theory (Shannon, 1948; Hsia, 1977).

Finally, the data compression method to quantify information content adds to the toolbox of empirically measuring news flow. The existing literature generally uses textual similarity to capture the news richness or novelty of media articles (Tetlock, 2011), modification score of MD&A disclosure in annual reports (Brown and Tucker, 2011), degrees of product similarity and differentiation (Hoberg and Phillips, 2016), changes in the language of 10-K and 10-Q filings (Cohen et al., 2020), news content in annual reports (Belo et al., 2018), among others. In an attempt to quantify news, Belo et al. (2018) explore the hump-shaped stock return volatility dynamics around 10-K fillings and find that larger 10-K reports lead to higher uncertainty in the short run due to its complexity (Loughran and McDonald, 2014) but eventually result in lower uncertainty, which is consistent with larger annual reports carrying more information content. They use cosine similarity between the current report and the previous year's report to help quantify news, while I use incremental compression ratio, which could be a useful complement to their method. To see the difference between my news redundancy measure and textual similarity, one could imagine an extreme case where the current annual report contains no news but only stops reporting a significant amount of boilerplate contents (common in each year but not informative) due to regulation changes. Then there will be a sharp drop in document similarity due to a shrink in overall filing content relative to the previous year, but this decrease in similarity does not imply more news. Although this extreme case is unlikely in real settings, the above comparison shows that low similarity does not necessarily imply low news redundancy, especially when the target document is smaller than the benchmark document in overall content.

A few recent studies also use cross-entropy calculated from N-grams to capture unusualness or informativeness in a bunch of news articles. Glasserman and Mamaysky (2019) measure the unusualness of Thomson Reuters news articles about the largest 50 financial companies and interact this measure with sentiment to forecast market-level and firm-specific volatility. Glasserman et al. (2019) measure the informativeness of Thomson Reuters (TR) articles that mention S&P 500 firms and examine how news informativeness relates to price underreaction to the news. In contrast, my paper focuses on financial reports rather than news articles. Moreover, the news redundancy ratio intends to capture the richness of news by the incremental value of entropy rather than the cross-entropy. The compression algorithm can be seen as a nonparametric way that does not depend on the choice of language model (3-grams or 10-grams).

2 Data and Redundancy Decomposition

2.1 Data sources

This paper uses all complete 10-K, 10-K405, 10-KSB, and 10-Q filings from the SECs Electronic Data Gathering, Analysis, and Retrieval (EDGAR) website from 1996 to 2018. All complete 10-K and 10-Q filings are in txt format and contain an aggregation of all information submitted with each firms file, such as exhibits, graphics, XBRL files, PDF files, and Excel files. Following Loughran and McDonald (2017), the analyses are concentrated on the textual content of the document[5]. I only extract the main 10-K and 10-Q texts in each document and remove all tables (if their numeric character content is greater than 10%), HTML tags, XBRL tables, ASCII-encoded PDFs, graphics, XLS, and other binary files.

I use stock returns and market-related information from the Center for Research in Security Prices (CRSP) and firm characteristics based on information from Compustat, and factors from the Fama-French data library[6]. I obtain analyst data from the Institutional Brokers Estimate System (IBES), institutional ownership data from Thomson Reuters 13f, sentiment category identifiers from Loughran and McDonald (2011) Master Dictionary[7], and the volume of requests on historical 10-K and 10-Q filings from SECs EDGAR log file data set [8]. Detailed variable descriptions can be found in later sections or Appendix A.1.

2.2 Measure total redundancy by compression

Measuring the redundancy of a document is essentially measuring the amount of information that is not redundant. In this paper, we quantify non-redundant information using Shannon's Entropy or Kolmogorov Complexity, which is proxied by the small-

[5]Bill McDonald provides a very detailed description of how to strip 10-K/Q down to text files: https://sraf.nd.edu/data/stage-one-10-x-parse-data/.

[6]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html.

[7]https://sraf.nd.edu/textual-analysis/resources/.

[8]https://www.sec.gov/dera/data/edgar-log-file-data-set.html. Detailed discussions on the usage of this data set can be found in Loughran and McDonald (2017), Ryans (2017), Dechow et al. (2016), Drake et al. (2017), and Chen et al. (2020)

est possible compressed length in bits achieved by lossless compression algorithms[9] (Shannon, 1948).

We can measure a document's information density by the length of the optimally-compressed file divided by the original length. Specifically, I define **Information Density Ratio (IDR)** for the the filing issued by stock i in period t ($Doc_{i,t}$) as

$$IDR_{i,t} \equiv \frac{CL(Doc_{i,t})}{OL(Doc_{i,t})} \tag{1}$$

where CL stands for compressed length, and OL stands for original length. If there are many repeated messages or duplicative sentences in the document, then we say the document has low information density. Information redundancy is just a complementary quantity of information density. A document's information redundancy could be calculated as the fraction of the raw file length that can be compressed without information loss. For the the filing issued by stock i in period t ($Doc_{i,t}$), I define its **Total Redundancy Ratio (TRR)** as

$$TRR_{i,t} \equiv 1 - IDR_{i,t} = \frac{OL(Doc_{i,t}) - CL(Doc_{i,t})}{OL(Doc_{i,t})} \tag{2}$$

Information redundancy and density add up to one. The more repeated phrases, the more redundant the file is. Highly repetitive disclosure gives the financial report a higher redundancy ratio. A hypothetical extreme text document composed of the phrase "Good News, Good News, Good News..." repeated 1,000 times has a TRR of nearly 100%, which is the natural upper bound. The lower bound of text redundancy, however, is not close to 0, but rather somewhere above 50% [10]. An example of an extremely low-redundancy text file could be the first 100,000 digits of Pi, a stream of unpredictable numbers "3.14159265358979...", with a TRR of 55.7%. The TRR of a

[9] I choose the 7-Zip software that is powerful enough to get close to the entropy limit, namely the highest possible compression ratio (original length divided by compressed length). I use Python scripts to shut down the time-saving mechanism in order to get the full power.

[10] The reason is that digital and alphabetical characters are stored as binary (combinations of machine code 0 and 1), unavoidably with common partial combinations occur in multiple characters' encodings. Common examples of character encoding systems include Morse code, the Baudot code, the American Standard Code for Information Interchange (ASCII), and Unicode. Under ASCII or Unicode, for example, the binary for "A" is "01000001", for "B" is "01000010", for "x" is "01111000", and for "y" is "01111001".

typical real-world document lies between the two extremes. For example, the classic paper Jensen and Meckling (1976) has a TRR of 71.2%, while Apple Inc.'s 10-K filing for the fiscal year 2018 has a TRR of 80.6%.

2.3 Decompose into news and stale-news redundancies

For an investor without any prior knowledge of the company, most of the financial report information is news. Its redundancy (density) quantifies the sparseness (richness) of news in the file. For an average marginal investor who is reasonably sophisticated, however, the financial report contains both news and stale-news. The redundancies of news and stale-news may play different roles and have different implications on firm value because communicating news is the primary concern for managers. Decomposing redundancy into news redundancy and stale-news redundancy could provide more insights.

In order to separately measure news and stale-news, we need a benchmark information set. In this paper, I choose the filing of year t-1 as the benchmark. Thus, we can quantify news in Doc_t relative to Doc_{t-1} by the Shannon's entropy of information in Doc_t that is not redundant to Doc_{t-1}. This quantity can be captured by the difference between Shannon's entropy of the appended text, $Doc_{t-1} + Doc_t$, and Shannon's entropy of the benchmark text, Doc_{t-1}.

Using compressed length of the text to approximate Shannon's entropy, I now define the **News Density Ratio (NDR)** for the 10-K(Q) filing by stock i in period t relative to $t-1$ as

$$NDR_{i,t} \equiv \frac{CL(Doc_{i,t-1} + Doc_{i,t}) - CL(Doc_{i,t-1})}{OL(Doc_{i,t})} \tag{3}$$

Since the total amount of information equals the sum of news quantity and stale-news quantity, we scale both sides by the document length and get the relationship $IDR \equiv NDR + SDR$. Then we get the **Stale-News Density Ratio (SDR)** as the total information density minus the news density

$$SDR_{i,t} \equiv IDR_{i,t} - NDR_{i,t} = \frac{CL(Doc_{i,t})}{OL(Doc_{i,t})} - NDR_{i,t} \tag{4}$$

Their compliments are **News Redundancy Ratio (NRR)**

$$NRR_{i,t} \equiv 1 - NDR_{i,t} \tag{5}$$

and the **Stale-News Redundancy Ratio (SNR)**

$$SNR_{i,t} \equiv 1 - SDR_{i,t} \tag{6}$$

for the 10-K(Q) filing by stock i in period t relative to $t-1$. The higher the NRR of a filing, the less news per page it contains. For example, if the Apple 2019 10-K is exactly the same as that in 2018, then the news redundancy for the 2019 filing is 100% [11].

The decomposition of redundancy into news redundancy and stale-news redundancy can be expressed as follows:

$$1 + TRR_{i,t} = NRR_{i,t} + SNR_{i,t} \tag{7}$$

In the appendix, I do validation tests to confirm that these measures indeed capture news and stale-news. I examine the different roles of news and stale-news in resolving information asymmetry and updating investors' beliefs. *Given the raw document size*, higher news redundancy (low news density) should go with higher levels of information asymmetry (Loughran and McDonald, 2014) and smaller price adjustments (Francis and Schipper, 1999; Roychowdhury and Sletten, 2012) upon announcement of the filing. Using idiosyncratic volatility and analyst forecast dispersion as proxies for information asymmetry, I find that news redundancy is positively associated with changes in information asymmetry. In contrast, stale-news redundancy does not have a significant effect (Table A.2.1). Using the absolute value of filing announcement return to capture the magnitude of price adjustments, I find that higher news redundancy is associated with smaller price impact and that stale-news redundancy (density) is pos-

[11] To differentiate my news redundancy measure from document similarity measures like cosine similarity used by Tetlock (2011) and Cohen et al. (2020), one may consider a hypothetical filing 2019 composed by duplicating some but not all paragraphs from filing 2018 by different multiples. Then the cosine similarity between filing 2019 and 2018 is not 100%, but the news redundancy ratio is still 100%.

itively (negatively) associated with the magnitude of price movements (Table A.2.2). These results validate that these two measures capture the variation in the amount of news and stale-news empirically.

3 News Redundancy vs. Stale-news Redundancy

This section studies the relationship between disclosure repetitiveness (redundancy) and processing costs using total redundancy ratio (TRR), news redundancy ratio (NRR) and stale-news redundancy ratio (SNR).

3.1 Relationships with firm performance

In this section, I attempt to pin down whether adding redundancy will increase or decrease processing costs at the news message level by examining the relationship between disclosure redundancy and firm performance. I assume that public firms' managers have an incentive to reduce processing costs for good news and raise processing costs for bad news. Under this assumption, if adding redundancy increases processing costs, we should find a significantly negative relationship between news redundancy and earnings. If adding redundancy lowers processing costs, we should find news redundancy positively related to earnings. As for stale-news, it should already be priced in given a long time interval between two filings. Repeating stale-news is unnecessary, thus a bad signal.

I run the following regressions in order to test in what way disclosure redundancy is related to current period earnings:

$$E_{i,t} = \alpha + \beta Red_{i,t} + X_{i,t-1} \cdot \gamma + \theta_i + \delta_t + \epsilon_{i,t} \tag{8}$$

and the next quarter earnings:

$$E_{i,t+1} = \alpha + \beta Red_{i,t} + X_{i,t} \cdot \gamma + \theta_i + \delta_t + \epsilon_{i,t+1} \tag{9}$$

where I use two measures for earnings ($E_{i,t}$): income before extraordinary items scaled by lagged total assets, $IB_t = IBQ_t/ATQ_{t-1}$, and cash-based operating profitability ($CBOP_t$), following Ball et al. (2016). The variable of interest, $Red_{i,t}$, represents news redundancy ratio (NRR_t), stale-news redundancy ratio (SNR_t), and total redundancy ratio (TRR_t). Control variables in the two equations are measured at quarter t-1 ($X_{i,t-1}$) and quarter t ($X_{i,t}$) respectively. The vector $X_{i,t}$ include earnings (E_t), a dummy for negative earnings ($E_t \leq 0$), the natural logarithm of total assets ($ln(Asset_t)$), a dummy for zero dividends ($Div_t = 0$), dividends scaled by total assets (Div_t), and accruals ($Accruals_t$), following Fama and French (2006) and Hou et al. (2012). All regressions include firm fixed effects (θ_i) and calendar quarter fixed effects (δ_t).

The results for current quarter earnings and future earnings are reported in Table 2 and Table 3 respectively. From column (1) and (4) of both tables, we can see that the coefficients on NRR_t are significantly positive. From column (2) and (5) of both tables, we see significantly negative coefficients on SNR_t. The result that high news redundancy conveys a good signal about firm fundamentals is consistent with the notion that adding redundancy to news disclosure should lower the processing cost. The results that both stale-news redundancy and total redundancy are negatively related to earnings are consistent with prior research. Stale-news redundancy is likely to drive the previous findings of information obfuscation.

3.2 Relationships with announcement returns

From section 3.1, we know how NRR, SNR, and TRR are associated with current and future operating performance. Then at least some of the investors should comprehend the underlying information content and trade accordingly.

I test how investors response to the contents of 10-K or 10-Q filings with different NRR, SNR, and TRR through Fama-MacBeth regressions:

$$BHAR^f(0,h)_{i,t} = \alpha + \beta Red_{i,t} + X_{i,t} \cdot \gamma + \epsilon_{i,t,h} \tag{10}$$

where $BHAR^f(0,h)_{i,t}$ is the buy-and-hold stock return minus the CRSP value-weighted buy-and-hold market index return over trading day window [0, h], with h=1, 10, relative to filing announcement day for stock i in quarter t. The variable of interest, $Red_{i,t}$, represents news redundancy ratio (NRR_t), stale-news redundancy ratio (SNR_t), and total redundancy ratio (TRR_t). Following Loughran and McDonald (2011), the control variables in vector $X_{i,t}$ also include natural logarithms of market capitalization ($ln(ME)$), book-to-market ratio of equity ($ln(BM)$), share turnover ratio averaged over the past 12 months ($ln(Turnover)$), the percent of institutional ownership reported in filing 13f ($IO\%$), the pre-filing Fama-French 3-factor alpha based on a regression using trading days [-250, -6], a dummy for Nasdaq-listed stocks. Fama and French (1997) industry dummies (based on 48 industries) and a constant are included in each first-stage cross-sectional regression for quarter t. The time-series averages of coefficients are reported with Newey and West (1987) adjusted (3 lags) t-statistics. I expect β on NRR to be positive if high news redundancy signals more good news. I expect β on SNR to be negative if high stale-news redundancy signals more bad news.

From Table 4, we can see that the filing announcement effect of NRR is significantly positive, while SNR's effect is significantly negative. When time-horizon expands, NRR dominates SNR in determining the sign of filing announcement returns. This result also suggests that earnings management (highlighting bad news) does not prevail in the cross-section.

3.3 Relationships with analyst following

Previous tests are based on management incentives to lower the processing cost for good news and raise the cost for bad news. In this section, I attempt to address this question at the firm-quarter level by examining the effect of disclosure redundancy on analyst following. Lehavy et al. (2011) finds that less readable disclosure is associated with more analysts following afterward, which indicates that high processing cost increases the demand for analyst coverage. Thus, I assume that analysts respond to investors' demand for investment information for firms whose disclosures are more costly to process. If filings with higher news redundancy have higher processing costs, then news redundancy should positively predict future analyst following. If filings with higher news redundancy have lower processing costs, we should find news redundancy negatively predicts future analyst following. If we assume analysts can tell news from stale-news, then the stale-news density or redundancy should not matter.

I run the regressions with the following specification:

$$Analyst_{i,t+1} = \alpha + \beta Red_{i,t} + X_{i,t} \cdot \gamma + \theta_i + \delta_t + \epsilon_{i,t+1} \tag{11}$$

where the dependent variable, $Analyst_{i,t+1}$, stands for the total number of analysts making quarterly earnings forecast before the next quarter filing announcement ($\#AF_{t+1}$) or the natural logarithms of one plus $\#AF_{t+1}$, ($\ln(1{+}\#AF_{t+1})$). The variable of interest, $Red_{i,t}$, represents news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and total redundancy ratio (TRR). The vector of control variables, $X_{i,t}$, contains the total number of analysts making quarterly earnings forecast before the current quarter filing announcement (#AF), the natural logarithms of one plus #AF (ln(1+#AF)), the natural logarithms of firm market capitalization (ln(ME)), the natural logarithms of book-to-market ratio of equity (ln(BM)), the growth rate on total asset following Cooper et al. (2008) (AG), the percent of institutional ownership reported in filing 13f ($IO\%$). All control variables are measured at period t. Firm fixed effects and calendar quarter fixed effects are included.

The results are reported in Table 5. We can see that news redundancy (NRR) is negatively predicting analyst following in the next period, but the effect of stale-news redundancy (SNR) is insignificant. Total redundancy (TRR) produces similar results to NRR. In column (1), the coefficient on NRR is -2.991 (t-statistic -12.83), which implies that one standard deviation increase in NRR is associated with a decrease of around 0.12 raw number of analysts following the firm in the next quarter. In column (4), the coefficient on NRR is -0.49 (t-statistic -12.10), which suggests that one standard deviation increase in NRR is associated with around 2% drop in analyst coverage in the next period. The coefficients on the control variables are consistent with prior research. The above results suggest that filings with higher news redundancy ratios have lower processing costs.

4 News Redundancy and Future Stock Returns

In this section, I conduct the asset pricing tests of the news redundancy ratio (NRR). First, I conduct standard calendar-time portfolio sorts, and then I control for additional determinants of returns by employing Fama-MacBeth monthly cross-sectional regressions.

4.1 Calendar-time portfolio sorts

I conduct the univariate portfolio sorting as follows: At the beginning of each month, from January 1997 to December 2018, all stocks are sorted into quintiles based on the latest available news redundancy ratio in the previous month. Stocks enter the portfolio in the month after the 10-K or 10-Q filings' public release and exit the portfolio after 3 months. I compute the time-series of each NRR-quintile portfolio's return over each month, both equal-weighted and value-weighted, and then examine the alphas, factor loadings, and firm characteristics across portfolios.

In Table 6, I report the average return of each quintile excess of the risk-free rate, the CAPM alpha, the Fama-French three-factor alpha (Fama and French, 1993), the Fama-French-Carhart four-factor alpha (Carhart, 1997), the Fama-French five-factor alpha (Fama and French, 2016), Hou-Xue-Zhang q-factor alpha (Hou et al., 2015), Hou-Mo-Xue-Zhang five-factor alpha (Hou et al., 2020), and Stambaugh-Yuan mispricing factor alpha (Stambaugh and Yuan, 2017) for each quintile portfolio and the Q5-Q1 portfolio which is a zero-investment portfolio that buys the stocks in the highest NRR quintile (Q5) and shorts the stocks in the lowest NRR quintile (Q1). Newey and West (1987) adjusted (3 lags) t-statistics are reported below the estimates.

From Panel A of Table 6, we can see that the equal-weighted portfolio alphas are nearly monotonic in NRR, and the long-short portfolio shows that stocks with high NRR outperform stocks with low NRR by 0.23% to 0.44% per month (2.76% to 5.28% per year). The t-statistics for equal-weighted long-short alphas are from 1.99 to 4.41. From Panel B, we can see that the economic magnitudes of excess return and alphas of the value-weighted long-short portfolio are larger than equal-weighted, ranging from 0.45% to 0.61% per month (5.4% to 7.32% per year). The strategy generates most of the profits in the long leg (Q5). The t-statistics are smaller but still significant, ranging from 1.97 for SY 4-factor alpha to 2.73 for FF 3-factor alpha.

Table 7 reports the time-series average of monthly cross-sectional means of firm characteristics in each quintile portfolio. The last column reports the time-series average of monthly t-statistic from pooled t-test with the null hypothesis that the mean firm characteristic in Q1 (short leg portfolio) minus the mean characteristic among the rest cross-section equals 0. The firm characteristics include market capitalization in millions by the end of month t (ME), the stock turnover ratio averaged from month t-11 to t (Turnover12), average daily bid-ask spread in month t (SpreadM), Idiosyncratic volatility by Fama-French 3-factor model on daily returns in month t (IVOL), the natural logarithm of 1 plus the number of analysts following a firm (Analyst), total institutional ownership as a percent of shares outstanding (IO%), and the fraction of shares held short (ShortInt). The average number of stocks in one quintile portfolio is 354 in the sample from 1997-01 to 2018-12.

From Table 7, we can see that the market capitalization, turnover rate, bid-ask spread, idiosyncratic volatility, analyst following, institutional ownership, and short interest are similar across the five NRR quintiles. From the last column, we can see that stocks' characteristics in the short leg portfolio do not differ significantly from the rest. The above results suggest that the NRR spread we observe is not likely to result from limits to arbitrage directly.

4.2 Fama-MacBeth regressions

I use the Fama and MacBeth (1973) regression test to examine the predictive power of the news redundancy ratio (NRR) while controlling for known return predictors. Each month t from January 1997 to December 2018, I run a cross-sectional regression of stock returns (in percentage) in month t+1 on the news redundancy ratio (NRR_t) with different specifications. Fama and French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all models. In Table 8, I report the time-series average of the periodic cross-sectional regression coefficients, average R-squared, and the number of observations. I report Newey and West (1987) adjusted (12 lags) t-statistics below the estimates. Panel

A of 8 presents the results for all stocks, and Panel B presents the results for all-but-microcaps. Microcaps are stocks with a market capitalization below the 20th percentile of the NYSE market capitalization distribution. If the NRR has incremental information on future stock returns beyond the common predictors, we should reject the null hypothesis that the average coefficient of NRR is equal to 0. If this effect is not limited to small stocks, we should observe significant results for the all-but-microcaps sample.

The results in Panel A of Table 8 confirm the findings of calendar-time portfolio analysis. In column (1), the coefficient on the NRR is 4.04, with a t-statistic of 5.24. Given the average NRR is 0.84 and 0.925 for the lowest and the highest quintile portfolios, respectively. The Fama-MacBeth regression coefficient implies a return spread of 4.04*(0.925-0.84)=0.341 percent per month. The magnitude of return spread is comparable to the equally-weighted alphas in the time-series portfolio analysis. The results are slightly weaker (coefficient 2.91) but remain significant (t-statistic 2.84) if I remove the microcaps, as is shown in Panel B.

The NRR variable remains a significant predictor of future returns (R_{t+1}) even after I include the major known predictors in asset pricing literature. In column (2), I include beta (Scholes and Williams, 1977), market capitalization (Size), and book-to-market ratio for equity (BM) following Fama and French (2008). In column (3), I add month t return as short term reversal (Rev), the cumulative return from t-11 to t-1 as momentum (MOM) following Jegadeesh and Titman (1993), and the cumulative return from month t-59 to t-12 as the long-term reversal (LTRev) following De Bondt and Thaler (1985). In column (4), I add Amihud (2002) illiquidity measure (ILLIQ) and idiosyncratic volatility by the Fama-French 3-factor model (IVOL) based on Ang et al. (2006). In column (5), I add asset growth (AG) by Cooper et al. (2008), cash-based operating profitability (CBOP) following Ball et al. (2016), and the Accruals (Sloan, 1996). In column (6), I add Cumpustat-based standardized unexpected earnings (SUE) following Livnat and Mendenhall (2006). From Panel A, we can see as I add more controls, the regression coefficients in column (2) to (6) decrease from 3.847 to 2.703, the corresponding t-statistics ranging from 4.59 to 5.15. For the multi-variate results in Panel B, the coefficients are smaller but still significant with t-statistics ranging from

2.3 to 2.56. The coefficients of the other variables are consistent with prior literature.

4.3 Robustness

4.3.1 Different future horizons

In order to check whether the return prediction by the NRR is transitory or permanent, I do Fama-MacBeth regressions of cumulative returns from the end of month t to the end of month t+m divided by m (CR(m)/m, m=2, 4, 6, 8, 10) on the NRR and prior predictors including industry dummies measured by the end of month t.

The results in Table 10 show statistically significant results for all horizons with t-statistics ranged from 4.54 to 6.76. The coefficient for the NRR in regression of CR(m)/m is larger than 3.112/m for m=2, 4, 6, 8, 10, 12. The sign and magnitude of coefficients imply a drift from month t to month t+12. The effect is not transitory as we do not observe systematic reversals. The results suggest that the predictive power of the NRR is likely to be based on news about firm fundamentals, which is consistent with the positive relationship between earnings and the NRR presented in Table 2.

4.3.2 Different sample periods

To see whether the NRR effect depends on some special period like financial crisis, I run the Fama-MacBeth univariate and multivariate regressions for different sub-periods, 1997-2006, 2007-2009, and 2010-2018. The results are shown in Table 11. We can see the NRR effect holds in all three sub-periods, weaker but still significant in the 2007-2009 financial crisis, and remains stronger in recent years.

4.3.3 Controlling for existing textual measures

In order to explore to what extent other document characteristics drive the NRR effect, I redo the main test controlling for change in the proportion of positive words (ΔPOS), change in the proportion of negative words (ΔNEG), original size of filing measured as natural logarithm (ln(FSize)), the change of filing size ($\Delta ln(FSize)$), and Jaccard similarity following Loughran and McDonald (2016) and Cohen et al. (2020). The results are shown in Table 12. The coefficients remain significant and comparable to those in the main results. The NRR effect is not driven out by changes in sentiments, variations in file size of filings, or document similarity.

5 Mechanism Discussions

In this section, I discuss the potential mechanisms behind the NRR effects on future stock returns.

5.1 News redundancy and textual characteristics

To better understand what document characteristics help explain changes in news redundancy across years for a given firm's financial reports, I regress quarter-on-quarter changes in NRR on changes in a host of textual characteristics.

Given news resolving uncertainty, I expect changes in NRR (ΔNRR) should be positively associated with changes in the proportion of uncertain words (ΔUNC) and weak modal words (ΔMDW). The results in Table 13 confirm my expectations. From Panel A of Table 13, we can see that the correlations between ΔNRR and the other document characteristics are generally weak, with the highest Spearman correlation with ΔMDW 0.082. In contrast, the Spearman correlation between ΔUNC and ΔMDW is 0.702, consistent with both measures are proxies for uncertainty reported by the 10-K or 10-Q filing. From Panel B of Table 13, we can see that the coefficients for ΔUNC and ΔMDW are positively significant with t-statistics 3.08 and 6.53 in column (1) and (2), respectively.

In addition to the validation tests, some interesting results emerged in Panel B. In column (3), (6), and (7), I find changes in litigious words (ΔLIT) is negatively associated with ΔNRR, which indicates that litigation events are important components of news captured by news redundancy ratio. In column (4) to (7), I find asymmetric results for changes in the proportion of positive words (ΔPOS) and changes in the proportion of negative words (ΔNEG). The coefficient for ΔPOS in column (4), 1.960, is the largest among the five textual measures, and so is the t-statistics, 10.94. In contrast, the coefficient for ΔNEG in column (5), 0.104, is the smallest and least significant in the first five columns. When the news redundancy gets higher, the proportion of positive words increases significantly while the proportion of negative words stays the same or decreases (see column (7)). This result indicates that news redundancy, as a quantitative measure for news, is also associated with news content.

5.2 News redundancy and future earnings announcement returns

To check whether investors fully comprehend the information in the NRR before the next quarterly earnings announcement, I test to what extent investors are surprised by subsequent earnings realizations. This method is widely used to provide evidence of biased expectations in previous studies including Sloan (1996), Shleifer and Vishny (1997), Engelberg et al. (2018). I conduct the Fama-MacBeth regressions with the following specification:

$$CAR^e(-2,2)_{i,t+1} = \alpha + \beta NRR_{i,t} + X_{i,t} \cdot \gamma + \epsilon_{i,t+1} \quad (12)$$

where $CAR^e(-2,2)_{i,t+1}$ refers to the market model cumulative abnormal return over trading days [-2, 2] relative to the next quarterly earnings announcement day ($EAD_{i,t+1}$) following the 10-K or 10-Q filing by firm i in quarter t. If investors fully understand the information in the NRR within one quarter after the filing announcement, or the return prediction effect is due to some kind of risk taken by the investors, the earnings announcement returns should be similar across stocks with different NRR. If the return prediction results from mispricing, I predict that the coefficient β should be positive as investors are surprised by the subsequent good (bad) news associated with high (low) NRR.

Table 14 reports the Fama-MacBeth regressions of future earnings announcement return, $CAR^e(-2,2)_{i,t+1}$, on the news redundancy ratio, $NRR_{i,t}$, and the control variables used in previous tables. In column (1), the coefficient on the NRR is 4.376 (t=3.73). The coefficient's magnitude does not change much after I control for all other predictors, including textual measures. The statistical significance slightly drops to t=2.15, as shown in column (6).

The results in Table 4 and Table 14 together with Table 2 provide support to the mispricing explanation that investors do not fully incorporate the NRR's implication for future profitability into their earnings forecast and are surprised by subsequent realizations of earnings.

5.3 Limits to arbitrage explanations

Table 7, we see no significant difference in the average values of several trading cost measures across the NRR-quintile portfolios. In this section, I do a more thorough analysis to examine the extent of the NRR effect explained by limits to arbitrage(Shleifer and Vishny, 1997). I do a series of double sorting tests dependent on measures of difficulty in valuing or trading the stocks. Following Baker and Wurgler (2006), I investigate the role of these limits-to-arbitrage measures: market capitalization (Size), idiosyncratic volatility (IVOL), illiquidity (ILLIQ), Bid-Ask Spread, number of analysts making quarterly earnings forecast (Analyst), institutional ownership in 13f (IO%), number of years recorded in CRSP (Age), Asset Growth, cash-based operating profitability (CBOP), financing expenses (XFIN), and percentage of tangible assets (PPE%). For each limits-to-arbitrage variable X, I first sort all the stocks into High/Low groups (relative to monthly cross-sectional median) based on X. Then, within each X group, I further sort stocks into NRR quintiles and calculate the Fama-French five-factor alpha (in percentage), on both an equal-weighted (EW) and value-weighted (VW) basis, for each quintile and the high-minus-low portfolio (Q5-Q1).

From Table 15, we can see the differences of Q5-Q1 alpha between the high and low groups of X are not always consistent with predictions by limits to arbitrage. The results on an equal-weighted (EW) basis are not always consistent with those on a value-weighted (VW) basis. For Asset Growth, both VW and EW results are consistent with limits to arbitrage explanations, with long-short alpha only significant in the high asset growth group. For IVOL, CBOP, and PPE%, the relative magnitude of EW and VW alphas are consistent with limits to arbitrage while there are sizable and significant alphas in both high and low X groups. For Size, ILLIQ, Bid-Ask Spread, Age, and XFIN, either EW or VW alpha is consistent with limits to arbitrage, but not both. For Analyst and IO%, both EW alpha and VW alphas are against the predictions of limits to arbitrage. To sum up, the mixed evidence suggests that the

NRR effect is not fully-explained by limits to arbitrage.

5.4 Investor limited attention

Given limits to arbitrage do not fully explain the NRR effect, investor inattention may play a key role since it leads to under-reaction to news, as documented in Cohen and Frazzini (2008), Dellavigna and Pollet (2009), Hirshleifer et al. (2009), deHaan et al. (2015), Ben-Rephael et al. (2017), and Cohen et al. (2020). Given the NRR is measuring news relative to historical filings in the same quarter last year, I expect the NRR effect is weaker for stocks with a larger investor base paying attention to historical financial reports. In order to test this hypothesis, I construct a measure of look-back attention, LBA, which is the moving average of $ln(1+HESV)$ for the past six months where $HESV$ is the monthly aggregate value of historical Edgar search volume, defined as the sum of human user requests for 10-Ks/Qs that are publicly available for greater than or equal to 360 days following Drake et al. (2016). Then I define a dummy variable indicating a stock belonging to the low LBA group, $LowLBA_{i,t}$, which equals 1 if $LBA_{i,t}$ is no bigger than the median LBA of the cross-section in month t. I run Fama-MacBeth regressions of future stock returns on the NRR, $LowLBA$, and an interaction term, $LowLBA \times NRR$. If the NRR effect is due to limited attention, then I expect the effect is stronger for stocks receiving lower look-back attention from investors.

From Panel A of Table 16, we can see that the interaction terms are consistently positive and are significant in column (4), (6), and (8). The magnitude of the coefficient for $LowLBA \times NRR$ is 50% larger than NRR in column (4), and more than 100% larger in column (6) and (8), implying the NRR effect is indeed stronger for stocks with lower look-back attention, the difference gets more dramatic when measured in longer horizons up to 12 months. Panel B of Table 16 confirms the above patterns using separate regressions with control variables in high and low LBA groups.

I also explore textual measures to see what kind of document characteristics could enhance or weaken the NRR effect. I double sort depending on a series of measures proposed by Loughran and McDonald (2011), Loughran and McDonald (2014), and Cohen et al. (2020). I expect the NRR effect to be stronger for those stocks with more attention-consuming information, like larger filing size, low similarity to previous year filings, more litigation descriptions, and more uncertainty to be resolved. The results in Table 17 are consistent with my expectations, which provide evidence for the limited attention channel from the demand side.

6 Conclusion

This paper proposes a novel method to quantify the information content in disclosure documents like annual and quarterly reports using information theory and data compression technology. Using the previous year's filing as a benchmark, one can isolate news from stale-news in the current filing and calculate their density and redundancy accordingly. I do a series of tests to examine the relationship between disclosure redundancy and investors' processing costs, analyze how they are related to firms' performance given managers' incentive to obfuscate bad news, and design a trading strategy to show the economic value of information conveyed by news redundancy.

Considering managers' incentive to make good news salience and bad news opaque, I predict that high NRR should imply good performance and more good news. I find that NRR is positively associated with earnings and filing announcement returns, while SNR shows the opposite patterns. I find NRR is negatively predicting analysts following, while SNR does not matter for analyst coverage. Given that the demand for analysts' service is lower for firms with low information acquisition costs, this evidence suggests filings with high NRR have lower processing costs. To sum up, the findings related to news redundancy are consistent with information and communication theory, while the findings on stale-news redundancy support the obfuscation hypothesis and are likely to drive the previous findings in the literature.

NRR also has strong implications for firms' future returns. A portfolio that buys stocks with high NRR and shorts stocks with low NRR generates value-weighted alphas of 5%-11% per annum. This return predictability does not reverse, cannot be fully explained by limits to arbitrage, and remains robust for large stocks, in crisis periods, and after controlling for known predictors and previous textual measures. These results are consistent with the notion that managers present good (bad) news with higher (lower) redundancy, and investors face attention constraints and acquire information with a delay.

These measures are also useful practically for attention allocation by anyone facing an information overload. The length of an average 10-K (60,000+ words, 100+ pages) has grown more than six times as long as that in the year 1995 (Loughran and McDonald, 2014; Cohen et al., 2020). Seeking value-relevant information in a thick financial report can be an exercise with time wasted wading through repetitive, verbose text with little actual news acquired ex-post. If an investor or analyst has many firms to follow within one or two weeks, it is better to ex-ante allocate more time and energy to the filings with a more actual amount of news indicated by low news redundancy.

References

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

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

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The journal of Finance* 61, 1645–1680.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Belo, Frederico, Jun Li, Xiaoji Lin, and Xiaofei Zhao, 2018, The opposing effects of information complexity and information content on return volatility, *Available at SSRN 2535469* .

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It Depends on Where You Search: Institutional Investor Attention and Underreaction to News, *The Review of Financial Studies* 30, 3009–3047.

Blankespoor, Elizabeth, Ed deHaan, and Ivan Marinovic, 2020, Disclosure processing costs, investors information choice, and equity market outcomes: A review, *Journal of Accounting and Economics* 101344.

Brown, Stephen V, and Jennifer Wu Tucker, 2011, Large-sample evidence on firms year-over-year md&a modifications, *Journal of Accounting Research* 49, 309–346.

Caplin, Andrew, and Mark Dean, 2015, Revealed preference, rational inattention, and costly information acquisition, *American Economic Review* 105, 2183–2203.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of finance* 52, 57–82.

Cazier, Richard A., and Ray J. Pfeiffer, 2017, 10-k disclosure repetition and managerial reporting incentives, *Journal of Financial Reporting* 2, 107–131.

Chen, Huaizhi, Lauren Cohen, Umit Gurun, Dong Lou, and Christopher Malloy, 2020, Iq from ip: Simplifying search in portfolio choice, *Journal of Financial Economics* .

Cohen, Lauren, and Andrea Frazzini, 2008, Economic Links and Predictable Returns, *Journal of Finance* 63, 1977–2011.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2020, Lazy prices, *The Journal of Finance* 75, 1371–1415.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset Growth and the CrossSection of Stock Returns, *Journal of Finance* 63, 1609–1651.

De Bondt, Werner F M, and Richard Thaler, 1985, Does the stock market overreact?, *Journal of Finance* 40, 793–805.

Dechow, Patricia M, Alastair Lawrence, and James P Ryans, 2016, Sec comment letters and insider sales, *The Accounting Review* 91, 401–439.

deHaan, Ed, Terry Shevlin, and Jacob Thornock, 2015, Market (in)attention and the strategic scheduling and timing of earnings announcements, *Journal of Accounting and Economics* 60, 36–55.

Dellavigna, Stefano, and Joshua M. Pollet, 2009, Investor Inattention and Friday Earnings Announcements, *Journal of Finance* 64, 709–749.

Dewan, Ambuj, and Nathaniel Neligh, 2020, Estimating information cost functions in models of rational inattention, *Journal of Economic Theory* 187, 105011.

Drake, Michael S., Jared Jennings, Darren T. Roulstone, and Jacob R. Thornock, 2017, The comovement of investor attention, *Management Science* 63, 2847–2867.

Drake, Michael S, Darren T Roulstone, and Jacob R Thornock, 2016, The usefulness of historical accounting reports, *Journal of Accounting and Economics* 61, 448–464.

Engelberg, Joseph, R David McLean, and Jeffrey Pontiff, 2018, Anomalies and news, *The Journal of Finance* 73, 1971–2001.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of* .

Fama, Eugene F, and Kenneth R French, 1997, Industry costs of equity, *Journal of financial economics* 43, 153–193.

Fama, Eugene F, and Kenneth R French, 2006, Profitability, investment and average returns, *Journal of financial economics* 82, 491–518.

Fama, Eugene F., and Kenneth R. French, 2008, Dissecting Anomalies, *Journal of Finance* 63, 1653–1678.

Fama, Eugene F, and Kenneth R French, 2016, Dissecting anomalies with a five-factor model, *The Review of Financial Studies* 29, 69–103.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Francis, J, and K Schipper, 1999, Have financial statements lost their relevance?, *Journal of Accounting Research* 37, 319–352.

Glasserman, Paul, Fulin Li, and Harry Mamaysky, 2019, Time variation in the news-returns relationship, *Available at SSRN 3420981* .

Glasserman, Paul, and Harry Mamaysky, 2019, Does unusual news forecast market stress?, *Journal of Financial and Quantitative Analysis* 54, 1937–1974.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American economic review* 70, 393–408.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh, 2009, Driven to Distraction: Extraneous Events and Underreaction to Earnings News, *Journal of Finance* 64, 2289–2325.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2020, An augmented q-factor model with expected growth, *Review of Finance* .

Hou, Kewei, Mathijs A Van Dijk, and Yinglei Zhang, 2012, The implied cost of capital: A new approach, *Journal of Accounting and Economics* 53, 504–526.

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

Hsia, H. J., 1977, Redundancy: Is it the lost key to better communication?, *AV Communication Review* 25, 63–85.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jensen, Michael C., and William H. Meckling, 1976, Theory of the firm: Managerial behavior, agency costs and ownership structure, *Journal of Financial Economics* 3, 305–360.

Lang, Mark, and Lorien Stice-Lawrence, 2015, Textual analysis and international financial reporting: Large sample evidence, *Journal of Accounting and Economics* 60, 110 – 135.

Lehavy, Reuven, Feng Li, and Kenneth Merkley, 2011, The effect of annual report readability on analyst following and the properties of their earnings forecasts, *The Accounting Review* 86, 1087–1115.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

Li, Ming, Paul Vitányi, et al., 2008, *An introduction to Kolmogorov complexity and its applications*, volume 3 (Springer).

Livnat, Joshua, and Richard R. Mendenhall, 2006, Comparing the PostEarnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts, *Journal of Accounting Research* 44, 177–205.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* 66, 35–65.

Loughran, Tim, and Bill McDonald, 2014, Measuring readability in financial disclosures, *The Journal of Finance* 69, 1643–1671.

Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.

Loughran, Tim, and Bill McDonald, 2017, The use of edgar filings by investors, *Journal of Behavioral Finance* 18, 231–248.

Mackowiak, Bartosz, and Mirko Wiederholt, 2009, Optimal sticky prices under rational inattention, *American Economic Review* 99, 769–803.

Nakano, Isao, 1972, Noise and redundancy in accounting communications, *The Accounting Review* 47, 693–708.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563 – 602.

Roychowdhury, Sugata, and Ewa Sletten, 2012, Voluntary Disclosure Incentives and Earnings Informativeness, *The Accounting Review* 87, 1679–1708.

Ryans, James, 2017, Using the edgar log file data set, *Available at SSRN 2913612* .

Scholes, Myron, and Joseph Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* 5, 309–327.

Shannon, Claude E, 1948, A mathematical theory of communication, *The Bell system technical journal* 27, 379–423.

Shannon, Claude E., and Warren Weaver, 1949, *The Mathematical Theory of Communication* (University of Illinois Press).

Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *The Journal of finance* 52, 35–55.

Sims, Christopher A, 2003, Implications of rational inattention, *Journal of monetary Economics* 50, 665–690.

Sloan, Richard G, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting review* 289–315.

Stambaugh, Robert F, and Yu Yuan, 2017, Mispricing factors, *The Review of Financial Studies* 30, 1270–1315.

Stephens, Keri, 2007, The successive use of information and communication technologies at work, *Communication Theory* 17, 486–507.

Stephens, Keri, and Stephen Rains, 2011, Information and communication technology sequences and message repetition in interpersonal interaction, *Communication Research - COMMUN RES* 38, 101–122.

Tetlock, Paul C., 2011, All the News That's Fit to Reprint: Do Investors React to Stale Information?, *The Review of Financial Studies* 24, 1481–1512.

Tetlock, Paul C., 2014, Information transmission in finance, *Annual Review of Financial Economics* 6, 365–384.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2009, Information immobility and the home bias puzzle, *The Journal of Finance* 64, 1187–1215.

Table 1: Data Summary

This table presents summary statistics for variables used in the main analysis. For each cross-section for month t from January 1998 through December 2018, the pairwise Pearson correlations with NRR (ρ_{NRR}) , mean (Mean), standard deviation (SD), skewness (Skew), 1th, 5th, 25th,50th, 75th, 95th, and 99th percentiles of each variable is calculated. The table presents the time-series means for each cross-sectional value. The column labeled N indicates the average number of stocks for which the given variable is available.

	ρ_{NRR}	Mean	SD	Skew	P5	P25	P50	P75	P95	N
NRR	1.00	0.88	0.03	-0.42	0.83	0.86	0.89	0.91	0.93	1936.8
Beta	-0.02	1.00	1.72	0.22	-1.34	0.16	0.92	1.78	3.62	1936.6
Size	0.00	6.23	1.93	0.27	3.18	4.85	6.17	7.48	9.60	1936.8
BM	0.02	-0.84	0.93	-0.79	-2.45	-1.35	-0.75	-0.23	0.50	1874.6
Rev	0.01	0.02	0.15	2.80	-0.17	-0.06	0.01	0.07	0.24	1936.8
MOM	0.00	0.18	0.64	4.45	-0.46	-0.14	0.08	0.34	1.13	1893.2
LTRev	-0.01	0.72	1.90	6.80	-0.72	-0.20	0.31	1.03	3.38	1878.7
ILLIQ	0.01	0.82	3.13	5.98	0.00	0.00	0.02	0.21	3.82	1936.7
IVOL	-0.04	0.03	0.02	5.17	0.01	0.01	0.02	0.03	0.06	1936.6
AG	-0.04	0.14	0.38	3.01	-0.22	-0.03	0.06	0.18	0.81	1928.5
CBOP	0.05	0.03	0.12	-2.06	-0.06	0.01	0.03	0.06	0.12	1919.4
Accruals	0.00	0.00	0.15	2.27	-0.18	-0.04	-0.01	0.03	0.16	1935.9
SUE	0.04	0.16	1.76	-0.37	-2.55	-0.44	0.13	0.87	2.90	1855.5
ΔNEG	-0.03	0.01	0.40	0.21	-0.61	-0.19	0.00	0.20	0.68	1679.0
ΔPOS	0.03	0.00	0.16	0.04	-0.26	-0.08	0.00	0.08	0.26	1679.0
ln(FSize)	-0.07	13.60	0.81	0.21	12.36	13.03	13.55	14.13	14.97	1889.6
$\Delta ln(FSize)$	-0.03	-0.01	0.82	-0.03	-1.40	-0.40	-0.05	0.44	1.34	1664.7
Sim_Jaccard	0.38	0.72	0.17	-1.02	0.37	0.62	0.75	0.85	0.92	1889.6

Table 2: News (stale-news) redundancy and current performance

This table reports the regressions of current-quarter earnings (E_t) on news redundancy ratio (NRR_t), stale-news redundancy ratio (SNR_t), and total redundancy ratio (TRR_t). Earnings (E_t) are measured as income before extraordinary items divided by lagged total asset, $IB_t = IBQ_t/ATQ_{t-1}$, for (1) to (3); and cash-based operating profitability ($CBOP_t$), following Ball, Gerakos, Linnainmaa, and Nikolaev (2016), for (4) to (6). Control variables include earnings, a dummy for negative earnings, the natural logarithm of total assets, a dummy for zero dividends, dividends scaled by total assets, and accruals. All control variables measured at quarter $t-1$. All the variables are winsorized quarter by quarter at the 1% level for both tails. All regressions include firm and calendar quarter fixed effects. t-statistics are in parentheses with standard errors clustered by firm and calendar quarter. Statistical significance at 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) IB_t	(2) IB_t	(3) IB_t	(4) $CBOP_t$	(5) $CBOP_t$	(6) $CBOP_t$
NRR_t	0.014***			0.053***		
	(2.92)			(8.42)		
SNR_t		-0.041***			-0.032***	
		(-10.57)			(-7.15)	
TRR_t			-0.070***			0.004
			(-10.12)			(0.43)
E_{t-1}	0.323***	0.322***	0.323***	0.008	0.007	0.008
	(15.20)	(15.16)	(15.18)	(0.53)	(0.51)	(0.55)
$E_{t-1} \leq 0$	-0.005***	-0.005***	-0.005***	0.004***	0.004***	0.004***
	(-6.11)	(-6.02)	(-6.07)	(4.55)	(4.50)	(4.44)
$ln(Asset_{t-1})$	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(4.04)	(4.25)	(4.33)	(3.31)	(3.40)	(3.24)
$Div_{t-1} = 0$	-0.001**	-0.001**	-0.001**	-0.002***	-0.002***	-0.002***
	(-2.33)	(-2.27)	(-2.34)	(-3.12)	(-3.15)	(-3.19)
Div_{t-1}	0.022**	0.022**	0.022**	0.050***	0.049***	0.049***
	(2.15)	(2.15)	(2.11)	(3.44)	(3.42)	(3.42)
$Accruals_{t-1}$	0.020***	0.020***	0.020***	0.010***	0.010***	0.010***
	(10.44)	(10.48)	(10.36)	(4.61)	(4.57)	(4.54)
Firm YQ FE	Yes	Yes	Yes	Yes	Yes	Yes
YrQt FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.596	0.596	0.596	0.273	0.273	0.273
Observations	167,192	167,192	167,192	165,530	165,530	165,530

Table 3: News (stale-news) redundancy and future performance

This table reports the regressions of next-quarter earnings (E_{t+1}) on news redundancy ratio (NRR_t), stale-news redundancy ratio (SNR_t), and total redundancy ratio (TRR_t). Earnings (E_{t+1}) are measured as income before extraordinary items, $IB_{t+1} = IBQ_{t+1}/ATQ_t$, for (1), (2) and (3); and cash-based operating profitability ($CBOP_{t+1}$), following Ball, Gerakos, Linnainmaa, and Nikolaev (2016), for (4), (5), and (6). Control variables include earnings, a dummy for negative earnings, the natural logarithm of total assets, a dummy for zero dividends, dividends scaled by total assets, and accruals. All control variables measured at quarter t. All the variables are winsorized quarter by quarter at the 1% level for both tails. All regressions include firm and calendar quarter fixed effects. t-statistics are in parentheses with standard errors clustered by firm and calendar quarter. Statistical significance at 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) IB_{t+1}	(2) IB_{t+1}	(3) IB_{t+1}	(4) $CBOP_{t+1}$	(5) $CBOP_{t+1}$	(6) $CBOP_{t+1}$
NRR_t	0.024***			0.019***		
	(4.78)			(3.06)		
SNR_t		-0.024***			-0.037***	
		(-6.43)			(-7.55)	
TRR_t			-0.019***			-0.054***
			(-3.72)			(-6.24)
E_t	0.321***	0.320***	0.321***	0.007	0.007	0.008
	(15.40)	(15.37)	(15.40)	(0.51)	(0.51)	(0.55)
$E_t \leq 0$	-0.005***	-0.005***	-0.005***	0.004***	0.004***	0.004***
	(-6.36)	(-6.27)	(-6.30)	(4.44)	(4.51)	(4.50)
$ln(Asset_t)$	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(4.02)	(4.13)	(4.05)	(3.22)	(3.40)	(3.39)
$Div_t = 0$	-0.001**	-0.001**	-0.001**	-0.002***	-0.002***	-0.002***
	(-2.36)	(-2.36)	(-2.41)	(-3.18)	(-3.15)	(-3.21)
Div_t	0.023**	0.023**	0.023**	0.049***	0.049***	0.049***
	(2.27)	(2.26)	(2.25)	(3.45)	(3.44)	(3.42)
$Accruals_t$	0.020***	0.020***	0.020***	0.010***	0.010***	0.010***
	(10.43)	(10.46)	(10.40)	(4.57)	(4.63)	(4.57)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
YrQt FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.596	0.596	0.596	0.272	0.273	0.273
Observations	167,192	167,192	167,192	165,530	165,530	165,530

Table 4: News (Stale-news) redundancy and filing announcement return

This table reports the Fama-MacBeth regressions of filing period abnormal return on news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and total redundancy ratio (TRR). The dependent variables are buy-and-hold stock return minus the CRSP value-weighted buy-and-hold market index return (BHAR) over trading day window [0,1] for column (1) to (3) and window [0,10] for column (4) to (6). The returns are in percentages. See the Appendix for the other variable definitions. All the variables are winsorized quarter by quarter at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Newey-West (1987) adjusted (3 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) BHAR(0,1)	(2) BHAR(0,1)	(3) BHAR(0,1)	(4) BHAR(0,10)	(5) BHAR(0,10)	(6) BHAR(0,10)
NRR	1.306***			3.812***		
	(2.79)			(3.58)		
SNR		-0.832**			-1.672*	
		(-2.00)			(-1.86)	
TRR			0.448			2.057*
			(0.58)			(1.69)
ln(ME)	0.055	0.058*	0.054	0.020	0.028	0.014
	(1.62)	(1.70)	(1.56)	(0.44)	(0.60)	(0.29)
ln(BM)	0.069*	0.071*	0.069*	0.172***	0.181***	0.168***
	(1.69)	(1.73)	(1.67)	(2.83)	(2.92)	(2.74)
ln(Turnover)	-0.028	-0.026	-0.030	-0.077	-0.073	-0.091
	(-0.22)	(-0.20)	(-0.24)	(-0.63)	(-0.60)	(-0.75)
IO%	0.461***	0.490***	0.497***	0.548**	0.639***	0.555**
	(4.42)	(4.35)	(4.24)	(2.59)	(2.69)	(2.57)
pre_Alpha	-0.061	-0.061	-0.059	-0.274*	-0.278*	-0.272*
	(-1.17)	(-1.18)	(-1.15)	(-1.77)	(-1.80)	(-1.76)
pre_RMSE	-0.085***	-0.085***	-0.087***	-0.032	-0.031	-0.037
	(-3.90)	(-3.84)	(-3.95)	(-0.55)	(-0.54)	(-0.64)
Nasdaq	0.057	0.057	0.062	0.108	0.107	0.124
	(1.39)	(1.39)	(1.55)	(1.19)	(1.20)	(1.39)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.058	0.059	0.068	0.068	0.079	0.079
Observations	153,429	153,429	153,297	153,297	153,160	153,160

Table 5: News (stale-news) redundancy and future analyst following

This table reports the regressions of next-quarter analyst following on news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and total redundancy ratio (TRR). The dependent variables are the total number of analysts making quarterly earnings forecast before the next quarter filing announcement ($\#AF_{t+1}$) and the natural logarithms of one plus $\#AF_{t+1}$, ($\ln(1+\#AF_{t+1})$). Control variables are total number of analysts making quarterly earnings forecast before the current quarter filing announcement (#AF), the natural logarithms of one plus #AF (ln(1+#AF)), the natural logarithms of firm market capitalization (ln(ME)), the natural logarithms of book-to-market ratio of equity (ln(BM)), the growth rate on total asset following Cooper et al. (2008) (AG), the percent of institutional ownership reported in filing 13f ($IO\%$). All control variables are measured at period t. All the variables are winsorized quarter by quarter at the 1% level for both tails. All regressions include firm and calendar quarter fixed effects. t-statistics are in parentheses with standard errors clustered by firm and calendar quarter. Statistical significance at 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) $\#AF_{t+1}$	(2) $\#AF_{t+1}$	(3) $\#AF_{t+1}$	(4) $\ln(1+\#AF_{t+1})$	(5) $\ln(1+\#AF_{t+1})$	(6) $\ln(1+\#AF_{t+1})$
NRR	-2.991***			-0.491***		
	(-12.83)			(-12.10)		
SNR		-0.039			-0.016	
		(-0.26)			(-0.59)	
TRR			-4.173***			-0.711***
			(-11.15)			(-10.11)
#AF	0.868***	0.867***	0.868***			
	(50.49)	(50.27)	(50.42)			
ln(1+#AF)				0.820***	0.819***	0.821***
				(75.06)	(74.67)	(75.22)
ln(ME)	0.263***	0.263***	0.265***	0.056***	0.056***	0.056***
	(13.11)	(12.98)	(13.18)	(11.70)	(11.67)	(11.71)
ln(BM)	-0.036***	-0.038***	-0.033***	-0.008***	-0.008***	-0.008***
	(-3.34)	(-3.56)	(-3.10)	(-3.06)	(-3.21)	(-2.88)
AG	0.055**	0.059***	0.061***	0.015***	0.016***	0.016***
	(2.62)	(2.79)	(2.89)	(4.19)	(4.36)	(4.43)
IO%	0.405***	0.402***	0.394***	0.086***	0.086***	0.084***
	(3.01)	(2.97)	(2.91)	(3.49)	(3.46)	(3.39)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
YrQt FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.967	0.967	0.967	0.962	0.962	0.962
Observations	160,418	160,418	160,418	160,418	160,418	160,418

Table 6: Calendar time portfolio returns

This table reports average monthly excess returns and alphas (in percentage) on both equal-weighted and value-weighted stock portfolios sorted by news redundancy ratio (NRR). Each month from 1997-01 to 2018-12, all stocks are sorted into quintiles based on the latest available NRR in previous month. Stocks are held in the portfolio for 3 months. For each of the decile portfolios, Q1 (low) through Q5 (high), we report the average excess return, Fama-French three-factor alpha, Fama-French-Carhart four-factor alpha, Fama-French five-factor alpha, Hou-Xue-Zhang q-factor alpha, Hou-Mo-Xue-Zhang five-factor alpha, and Stambaugh-Yuan mispricing-factor alpha. The right-most column reports the excess returns and alphas of the High-minus-Low portfolios. Panel A reports equalweight portfolio returns, and Panel B reports valueweight portfolio returns. Newey-West (1987) adjusted (3 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Panel A: Equally Weighted						
Excess Return	1.17***	1.36***	1.43***	1.42***	1.54***	0.36***
	(2.91)	(3.35)	(3.76)	(3.71)	(3.98)	(3.73)
CAPM Alpha	0.31	0.52**	0.64***	0.60***	0.75***	0.44***
	(1.50)	(2.34)	(2.98)	(3.04)	(3.56)	(4.21)
FF 3-Factor Alpha	0.23**	0.45***	0.52***	0.52***	0.67***	0.44***
	(2.01)	(3.31)	(4.72)	(4.59)	(5.16)	(4.41)
FFC 4-Factor Alpha	0.33***	0.52***	0.62***	0.61***	0.75***	0.42***
	(2.95)	(3.91)	(5.44)	(5.64)	(5.71)	(4.07)
FF 5-Factor Alpha	0.26**	0.47***	0.46***	0.48***	0.63***	0.37***
	(1.98)	(3.21)	(4.17)	(4.25)	(4.78)	(3.79)
HXZ 4-Factor Alpha	0.35**	0.57***	0.58***	0.60***	0.70***	0.36***
	(2.34)	(3.46)	(4.10)	(4.41)	(4.45)	(3.58)
HMXZ 5-Factor Alpha	0.42***	0.57***	0.56***	0.60***	0.65***	0.23**
	(2.72)	(3.40)	(4.00)	(4.79)	(4.36)	(2.18)
SY 4-Factor Alpha	0.52***	0.67***	0.69***	0.71***	0.76***	0.25**
	(3.29)	(3.63)	(4.44)	(4.60)	(4.34)	(1.99)
Panel B: Value Weighted						
Excess Return	0.72**	0.65**	0.66**	0.80***	1.32***	0.60***
	(2.53)	(2.23)	(2.16)	(2.68)	(4.49)	(2.72)
CAPM Alpha	0.00	-0.07	-0.03	0.09	0.61***	0.60***
	(0.03)	(-0.64)	(-0.13)	(0.74)	(3.61)	(2.64)
FF 3-Factor Alpha	-0.01	-0.08	-0.04	0.09	0.60***	0.61***
	(-0.07)	(-0.72)	(-0.25)	(0.76)	(3.81)	(2.73)
FFC 4-Factor Alpha	0.03	-0.10	-0.09	0.11	0.55***	0.53**
	(0.26)	(-0.86)	(-0.52)	(0.88)	(3.44)	(2.37)
FF 5-Factor Alpha	-0.16	-0.24**	-0.15	-0.02	0.42***	0.58***
	(-1.42)	(-2.15)	(-0.96)	(-0.15)	(3.10)	(2.67)
HXZ 4-Factor Alpha	-0.04	-0.17	-0.15	0.04	0.49***	0.53**
	(-0.35)	(-1.48)	(-0.95)	(0.33)	(3.43)	(2.45)
HMXZ 5-Factor Alpha	-0.12	-0.25**	-0.29	-0.06	0.33**	0.45**
	(-0.94)	(-2.06)	(-1.46)	(-0.43)	(2.32)	(1.99)
SY 4-Factor Alpha	-0.10	-0.25*	-0.31	-0.02	0.39***	0.49**
	(-0.71)	(-1.93)	(-1.54)	(-0.11)	(2.64)	(1.97)

Table 7: Portfolio characteristics

This table reports the time-series average of monthly cross-sectional means of firm characteristics in each quintile portfolio. The last column reports the time-series average of monthly t-statistic from pooled t-test with the null hypothesis that the mean firm characteristic in Q1 (short leg portfolio) minus the mean characteristic among the rest of cross-section equals 0. The firm characteristics include market capitalization in millions by the end of month t (ME), Stock turnover ratio averaged from month t-11 to t (Turnover12), average daily bid-ask spread in month t (SpreadM), Idiosyncratic volatility by Fama-French 3-factor model on daily returns in month t (IVOL), the natural logarithm of 1 plus the number of analysts following a firm (Analyst), total institutional ownership as percent of shares outstanding (IO%), and fraction of shares held short (ShortInt). The average number of stocks in one quintile portfolio is 354 in the sample from 1997-01 to 2018-12.

	Q1	Q2	Q3	Q4	Q5	**t-stat**
ME	4461.6	5106.8	5036.0	4742.9	4575.9	**-0.19**
Turnover12	0.187	0.184	0.174	0.168	0.161	**1.23**
SpreadM	0.011	0.011	0.011	0.011	0.012	**0.34**
IVOL	0.028	0.027	0.026	0.026	0.026	**1.50**
Analyst	1.568	1.619	1.627	1.625	1.584	**-0.86**
IO%	0.562	0.581	0.588	0.587	0.579	**-1.24**
ShortInt	0.045	0.044	0.044	0.043	0.042	**0.49**

Table 8: Main results: Fama-MacBeth regressions

This table reports the results of Fama-MacBeth regressions on New Redundancy Ratio (NRR) and a host of known return predictors (Definitions see Appendix 1). The dependent variable is future one month return (in percentage). All the accounting variables are winsorized month by month at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Panel A presents results for full sample, and Panel B presents results for all-but-microcaps. Microcaps are stocks with market capitalization below the 20th percentile of the NYSE market capitalization distribution. Newey-West (1987) adjusted (12 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

Panel A: full sample						
	(1)	(2)	(3)	(4)	(5)	(6)
NRR	4.040***	3.847***	3.323***	3.238***	2.973***	2.703***
	(5.24)	(4.66)	(5.15)	(5.14)	(4.79)	(4.59)
Beta		0.060	-0.022	0.018	0.027	-0.043
		(1.01)	(-0.45)	(0.33)	(0.49)	(-0.84)
Size		-0.195**	-0.097*	-0.094*	-0.099**	-0.110**
		(-2.51)	(-1.74)	(-1.89)	(-2.03)	(-2.39)
BM		0.274**	0.302***	0.264***	0.271***	0.337***
		(2.19)	(3.07)	(2.98)	(2.97)	(2.88)
Rev			-2.610***	-2.887***	-2.923***	-2.933***
			(-5.60)	(-5.80)	(-5.84)	(-5.76)
MOM			-0.038	-0.003	-0.022	-0.018
			(-0.15)	(-0.01)	(-0.09)	(-0.08)
LTRev			-0.135**	-0.163*	-0.161*	-0.126**
			(-2.46)	(-1.96)	(-1.93)	(-2.40)
ILLIQ				0.235	0.236	0.112**
				(1.54)	(1.53)	(2.24)
IVOL				-1.085	-0.199	0.234
				(-0.24)	(-0.05)	(0.05)
AG					-0.214**	-0.236**
					(-2.17)	(-2.36)
CBOP					2.737***	2.747***
					(4.00)	(4.57)
Accruals					-0.050	-0.021
					(-0.22)	(-0.10)
SUE						0.067***
						(2.71)
Intercept	-2.299***	-0.777	-1.025	-0.929	-0.633	-0.272
	(-2.91)	(-0.90)	(-1.06)	(-1.03)	(-0.71)	(-0.31)
R-squared	0.071	0.091	0.104	0.113	0.117	0.120
Observations	490,222	474,774	463,639	463,634	460,751	450,760

Table 9: Main results: Fama-MacBeth regressions (continued)

This table reports the results of Fama-MacBeth regressions on New Redundancy Ratio (NRR) and a host of known return predictors (Definitions see Appendix 1). The dependent variable is future one month return (in percentage). All the accounting variables are winsorized month by month at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Panel A presents results for full sample, and Panel B presents results for all-but-microcaps. Microcaps are stocks with market capitalization below the 20th percentile of the NYSE market capitalization distribution. Newey-West (1987) adjusted (12 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

Panel B: all but microcaps (NYSE 20 pctl)						
	(1)	(2)	(3)	(4)	(5)	(6)
NRR	2.906***	1.925**	2.255**	2.134**	2.031**	2.013**
	(2.84)	(2.30)	(2.56)	(2.39)	(2.36)	(2.34)
Beta		0.111	0.049	0.042	0.046	-0.018
		(1.30)	(0.65)	(0.62)	(0.68)	(-0.27)
Size		-0.092	-0.032	-0.040	-0.047	-0.062
		(-1.58)	(-0.60)	(-0.78)	(-0.91)	(-1.20)
BM		0.148	0.102	0.104	0.113	0.190*
		(1.20)	(1.28)	(1.33)	(1.49)	(1.73)
Rev			-1.981***	-2.043***	-2.056***	-2.106***
			(-3.03)	(-2.98)	(-3.01)	(-3.06)
MOM			-0.041	-0.007	-0.029	-0.056
			(-0.13)	(-0.02)	(-0.09)	(-0.18)
LTRev			-0.146	-0.142	-0.134	-0.099
			(-1.54)	(-1.50)	(-1.41)	(-1.58)
ILLIQ				1.452	1.929	5.303
				(0.21)	(0.26)	(0.66)
IVOL				-1.942	-0.860	-1.574
				(-0.28)	(-0.13)	(-0.24)
AG					-0.174	-0.166
					(-1.62)	(-1.49)
CBOP					1.542**	1.593**
					(2.27)	(2.40)
Accruals					0.222	0.278
					(0.85)	(1.00)
SUE						0.048*
						(1.70)
Intercept	-1.647*	-0.331	-0.827	-0.729	-0.778	-0.735
	(-1.70)	(-0.34)	(-0.83)	(-0.68)	(-0.77)	(-0.72)
R-squared	0.142	0.171	0.192	0.200	0.206	0.210
Observations	255,693	248,728	244,210	244,210	242,932	238,391

Table 10: Predicting long-horizon returns

This table reports the results from the Fama and MacBeth regressions of cumulative returns from month t to month m divided by m ($CR(m)/m$) on news redundancy ratio (NRR) and controls measured at the end of month t. All the accounting variables are winsorized month by month at the 1% level for both tails.Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Newey-West (1987) adjusted (12 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) CR(2)/2	(2) CR(4)/4	(3) CR(6)/6	(4) CR(8)/8	(5) CR(10)/10	(6) CR(12)/12
NRR	2.236***	1.904***	1.600***	1.642***	1.666***	1.746***
	(4.54)	(4.55)	(4.72)	(5.65)	(6.24)	(6.76)
Beta	-0.026	-0.021	-0.016	-0.022	-0.027	-0.025
	(-0.93)	(-0.90)	(-0.75)	(-1.07)	(-1.41)	(-1.44)
Size	-0.063	-0.013	0.007	0.023	0.037	0.047
	(-1.56)	(-0.34)	(0.20)	(0.68)	(1.15)	(1.51)
BM	0.254***	0.260***	0.244***	0.247***	0.246***	0.255***
	(3.19)	(3.63)	(3.51)	(3.60)	(3.65)	(3.83)
Rev	-1.090***	-0.434*	-0.023	0.159	0.217	0.197
	(-2.69)	(-1.75)	(-0.11)	(0.83)	(1.19)	(1.19)
MOM	-0.033	-0.072	-0.070	-0.090	-0.119	-0.146
	(-0.15)	(-0.36)	(-0.39)	(-0.57)	(-0.85)	(-1.21)
LTRev	-0.075***	-0.058**	-0.043**	-0.038**	-0.039**	-0.041**
	(-3.16)	(-2.54)	(-2.29)	(-2.11)	(-2.21)	(-2.19)
ILLIQ	0.120***	0.124***	0.107**	0.104**	0.109***	0.115***
	(2.84)	(2.60)	(2.15)	(2.33)	(2.62)	(2.88)
IVOL	-13.960***	-15.610***	-15.752***	-15.951***	-14.710***	-14.328***
	(-3.31)	(-4.28)	(-4.59)	(-4.86)	(-4.69)	(-4.76)
AG	-0.218**	-0.265***	-0.282***	-0.263***	-0.250***	-0.248***
	(-2.28)	(-2.94)	(-3.12)	(-3.28)	(-3.41)	(-3.74)
CBOP	3.304***	3.604***	3.522***	3.573***	3.558***	3.584***
	(6.78)	(8.73)	(8.22)	(8.42)	(8.16)	(8.38)
Accruals	-0.077	-0.233	-0.235	-0.258*	-0.178*	-0.102
	(-0.38)	(-1.16)	(-1.45)	(-1.89)	(-1.90)	(-1.21)
SUE	0.043***	0.034**	0.026*	0.018	0.023*	0.030**
	(2.62)	(2.30)	(1.81)	(1.27)	(1.75)	(2.29)
Intercept	-0.365	-0.611	-0.540	-0.707	-0.822	-0.946*
	(-0.46)	(-0.87)	(-0.85)	(-1.23)	(-1.43)	(-1.68)
Avg. R-squared	0.128	0.139	0.147	0.154	0.159	0.164
Observations	448,706	444,669	440,681	436,662	432,650	428,643

Table 11: Return prediction in different sample periods

This table reports Fama-MacBeth regressions for different sample periods on News Redundancy Ratio (NRR) and other known predictors (Definitions see Appendix 1). The dependent variable is future one month return (in percentage). All the accounting variables are winsorized month by month at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Newey-West (1987) adjusted (12 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	1997-2006		2007-2009		2010-2018	
	(1)	(2)	(3)	(4)	(5)	(6)
NRR	5.017***	2.368***	3.379**	3.401**	3.174***	2.842***
	(3.62)	(2.80)	(2.71)	(2.32)	(3.35)	(2.91)
Beta		-0.029		-0.129		-0.030
		(-0.30)		(-0.97)		(-0.74)
Size		-0.208**		-0.013		-0.033
		(-2.57)		(-0.31)		(-0.83)
BM		0.573***		0.284		0.093
		(2.85)		(0.72)		(1.09)
Rev		-4.344***		-1.854		-1.726***
		(-6.11)		(-1.35)		(-3.15)
MOM		0.396**		-1.505		0.019
		(2.07)		(-1.37)		(0.11)
LTRev		-0.237**		-0.017		-0.038
		(-2.31)		(-0.79)		(-0.97)
ILLIQ		0.126		0.138		0.088
		(1.49)		(1.03)		(1.36)
IVOL		2.084		-3.001		-0.743
		(0.32)		(-0.19)		(-0.12)
AG		-0.174		-0.081		-0.356***
		(-1.05)		(-0.31)		(-3.00)
CBOP		3.614***		1.054		2.348***
		(3.86)		(0.57)		(3.27)
Accruals		-0.100		0.198		-0.006
		(-0.26)		(0.39)		(-0.03)
SUE		0.138***		0.043		-0.004
		(3.16)		(1.48)		(-0.22)
Intercept	-2.543*	1.651	-3.174*	-3.549*	-1.736*	-1.316
	(-1.85)	(1.32)	(-2.00)	(-1.97)	(-1.69)	(-1.14)
Avg. R-squared	0.080	0.136	0.059	0.101	0.065	0.108
Observations	195,363	177,999	81,687	74,484	213,172	198,277

Table 12: Return prediction controlling for textual measures

This table reports Fama-MacBeth regressions on News Redundancy Ratio (NRR) and textual measures of tones, document size and Jacard similarity. All the textual and accounting variables are winsorized month by month at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Newey-West (1987) adjusted (12 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1)	(2)	(3)	(4)	(5)
NRR	3.319***	3.681***	3.013**	3.143***	2.785***
	(3.59)	(2.73)	(2.28)	(4.00)	(2.99)
ΔNEG	0.985	0.765	0.401	1.202	-0.018
	(0.95)	(0.80)	(0.93)	(1.04)	(-0.09)
ΔPOS		1.140	1.136	-0.059	0.022
		(1.55)	(1.29)	(-0.22)	(0.07)
ln(FSize)			0.001	0.305***	0.300***
			(0.01)	(2.62)	(2.83)
$\Delta ln(FSize)$				-0.090	-0.055
				(-0.50)	(-0.29)
Sim_Jaccard					0.058
					(0.16)
Beta	-0.424	-0.105	-0.203	-0.078	0.026
	(-0.85)	(-0.60)	(-0.82)	(-0.46)	(0.29)
Size	-0.069	-0.099	0.030	-0.179**	-0.140
	(-0.61)	(-1.20)	(0.15)	(-2.21)	(-1.57)
BM	0.739	0.412**	0.420**	0.293	0.079
	(0.92)	(2.27)	(2.16)	(1.62)	(0.44)
Rev	-4.931***	-2.834***	-2.864***	-1.679	-1.804
	(-2.72)	(-2.98)	(-3.54)	(-1.22)	(-1.29)
MOM	1.403	0.678	0.571	0.904	-0.301
	(0.69)	(0.75)	(0.76)	(0.84)	(-0.82)
LTRev	-0.341*	-0.234	-0.218	-0.375	-0.206**
	(-1.73)	(-1.40)	(-1.51)	(-1.62)	(-2.47)
Constant	0.488	-0.675	-1.942	-2.331	-2.529
	(0.38)	(-0.41)	(-1.31)	(-1.43)	(-1.38)
R-squared	0.138	0.139	0.141	0.141	0.142
Observations	341,490	341,490	338,723	338,453	338,453

Table 13: News redundancy and textual contents

This table reports how change in news redundancy ratio (NRR) is associated with changes in a host of textual characteristics. Panel A reports the Pearson product-moment (lower-diagonal) and Spearman rank (upper-diagonal) correlations between pairs of NRR, change in NRR (ΔNRR), and changes in the proportion of words denoting uncertainty (ΔUNC), weak modal words (ΔMDW), litigious words (ΔLIT), positive words (ΔPOS), and negative words (ΔNEG). Panel B reports the regressions of ΔNRR on textual contents measured as changes. All the variables are winsorized quarter by quarter at the 1% level for both tails. All regressions include firm and calendar quarter fixed effects. t-statistics are in parentheses with standard errors clustered by firm and calendar quarter. Statistical significance at 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

Panel A: correlation matrix

	NRR	ΔNRR	ΔUNC	ΔMDW	ΔLIT	ΔPOS	ΔNEG
NRR		0.434	0.054	0.082	0.028	0.059	0.012
ΔNRR	0.467		0.077	0.168	0.020	0.155	0.050
ΔUNC	0.044	0.047		0.702	0.077	0.147	0.391
ΔMDW	0.086	0.136	0.771		0.224	0.260	0.507
ΔLIT	0.002	-0.017	-0.073	0.161		-0.009	0.424
ΔPOS	0.060	0.152	0.135	0.260	-0.073		0.142
ΔNEG	0.006	0.032	0.406	0.521	0.402	0.146	

Panel B: regressions of ΔNRR_t

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔUNC_t	0.477***					0.320***	
	(3.08)					(2.84)	
ΔMDW_t		1.193***					1.262***
		(6.53)					(6.81)
ΔLIT_t			-0.293***			-0.227*	-0.272**
			(-3.90)			(-1.96)	(-2.33)
ΔPOS_t				1.960***		1.810***	1.497***
				(10.94)		(9.38)	(7.76)
ΔNEG_t					0.104	0.014	-0.254**
					(0.92)	(0.11)	(-2.06)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YrQt FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.072	0.078	0.072	0.078	0.070	0.080	0.085
Observations	131,085	131,085	131,085	131,085	131,085	131,085	131,085

Table 14: News redundancy and future earnings announcement return

This table reports the Fama-MacBeth regressions of future earnings announcement return on news redundancy ratio (NRR). The dependent variable (in percentage) is defined as the market model cumulative abnormal return over trading days [-2,2] relative to next earnings announcement day following the current 10-K or 10-Q filing. See the Appendix for the other variable definitions. All the variables are winsorized quarter by quarter at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Newey-West (1987) adjusted (3 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1)	(2)	(3)	(4)	(5)
NRR	3.035***	2.668***	2.601***	2.268***	2.025**
	(3.29)	(3.16)	(3.17)	(2.78)	(2.51)
Beta		0.059	0.056	0.068	0.063
		(1.33)	(1.25)	(1.57)	(1.42)
Size		-0.037	-0.037	-0.010	-0.014
		(-0.95)	(-0.94)	(-0.28)	(-0.38)
BM		0.296***	0.266***	0.238***	0.233***
		(4.75)	(4.24)	(3.81)	(3.89)
Rev			-0.169	0.041	0.048
			(-0.49)	(0.11)	(0.14)
MOM			-0.163	-0.132	-0.105
			(-1.57)	(-1.35)	(-1.05)
LTRev			-0.006	-0.012	-0.006
			(-0.24)	(-0.47)	(-0.24)
ILLIQ				0.059	0.054
				(0.97)	(0.88)
IVOL				-11.117***	-9.761***
				(-3.12)	(-2.68)
AG					-0.149
					(-1.52)
CBOP					1.084*
					(1.74)
Accruals					-0.150
					(-0.61)
SUE					-0.017
					(-1.08)
Constant	-1.950**	-1.239	-1.146	-0.921	-0.750
	(-2.38)	(-1.57)	(-1.43)	(-1.13)	(-0.91)
R-squared	0.044	0.051	0.054	0.058	0.062
Observations	163,246	156,576	152,465	152,464	150,180

Table 15: Limits to arbitrage

This table presents the results on limits to arbitrage. For each limits-to-arbitrage variable X, we first sort all the stocks into High/Low groups (relative to monthly cross-sectional median) based on X. Then within each X group, we further sort stocks into NRR quintiles and calculate the Fama-French five-factor alpha (in percentage), on both an equal-weighted (EW) and value-weighted (VW) basis, for each quintile and the high-minus-low portfolio (Q5-Q1). Newey-West (1987) adjusted (3 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

		EW						VW					
		Q1	Q2	Q3	Q4	Q5	Q5-Q1	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Size	**Low**	0.48**	1.08***	0.78***	0.89***	0.94***	0.45***	0.33*	0.83***	0.44***	0.52***	0.57***	0.24
		(2.38)	(4.32)	(4.38)	(5.11)	(4.81)	(2.84)	(1.75)	(3.46)	(2.94)	(3.46)	(3.69)	(1.27)
	High	-0.07	0.05	0.15*	0.13	0.19*	0.26**	-0.14	-0.33**	-0.15	-0.01	0.42***	0.56**
		(-0.68)	(0.41)	(1.68)	(1.25)	(1.73)	(2.51)	(-1.21)	(-2.46)	(-0.96)	(-0.06)	(3.03)	(2.49)
IVOL	Low	0.03	0.13	0.31***	0.23**	0.36***	0.34***	-0.16	-0.18	-0.16	-0.05	0.31**	0.47**
		(0.24)	(1.21)	(3.36)	(2.25)	(4.00)	(3.71)	(-1.30)	(-1.43)	(-0.94)	(-0.35)	(2.11)	(2.22)
	High	0.50**	0.84***	0.76***	0.68***	0.91***	0.42***	-0.26	0.14	0.27	0.07	0.45	0.73**
		(2.40)	(3.42)	(4.12)	(4.02)	(4.27)	(2.71)	(-1.18)	(0.45)	(1.24)	(0.34)	(1.51)	(2.06)
ILLIQ	Low	0.16	0.02	0.31**	0.18*	0.29**	0.11	-0.11	-0.36**	-0.15	-0.03	0.42**	0.52*
		(0.98)	(0.13)	(2.47)	(1.81)	(2.20)	(0.87)	(-0.77)	(-2.23)	(-1.11)	(-0.27)	(2.59)	(1.80)
	High	0.50**	0.79***	0.69***	0.79***	0.79***	0.29*	0.12	0.24	0.19	0.27*	0.28**	0.15
		(2.30)	(3.31)	(4.02)	(3.93)	(4.46)	(1.78)	(0.64)	(1.47)	(1.42)	(1.83)	(2.40)	(0.78)
Bid-Ask Spread	Low	0.14	0.13	0.29**	0.18*	0.32***	0.17	-0.06	-0.23*	-0.17	0.08	0.42***	0.48**
		(1.10)	(1.20)	(2.49)	(1.77)	(2.90)	(1.25)	(-0.42)	(-1.69)	(-1.06)	(0.59)	(2.78)	(1.98)
	High	0.37*	0.83***	0.72***	0.80***	0.90***	0.52***	-0.05	0.11	-0.02	0.11	0.30	0.35
		(1.78)	(3.81)	(4.24)	(4.63)	(4.57)	(3.29)	(-0.20)	(0.38)	(-0.11)	(0.53)	(1.38)	(1.51)
Analyst	**Low**	0.62***	0.80***	0.66***	0.72***	0.80***	0.18	0.17	0.19	0.34**	0.00	0.25*	0.07
		(3.24)	(3.94)	(4.91)	(4.66)	(5.00)	(1.31)	(0.91)	(0.98)	(2.15)	(0.02)	(1.78)	(0.38)
	High	0.03	0.15	0.09	0.12	0.42***	0.39***	-0.12	-0.16	-0.26	-0.11	0.44***	0.57**
		(0.19)	(1.20)	(0.60)	(0.72)	(2.86)	(3.29)	(-0.79)	(-1.31)	(-1.54)	(-0.69)	(2.67)	(2.00)
IO%	**Low**	0.63***	0.88***	0.85***	0.69***	0.88***	0.23*	-0.05	0.32	0.35	0.59*	0.40	0.44
		(3.23)	(3.81)	(5.41)	(4.77)	(5.45)	(1.74)	(-0.23)	(1.62)	(1.54)	(1.97)	(1.54)	(1.39)
	High	-0.12	-0.00	0.15	0.13	0.35**	0.46***	-0.22	-0.43***	-0.15	-0.02	0.33**	0.55***
		(-0.96)	(-0.03)	(1.15)	(1.12)	(2.58)	(4.13)	(-1.59)	(-3.06)	(-0.80)	(-0.11)	(2.12)	(2.63)
Age	**Low**	0.43**	0.68***	0.62***	0.53***	0.72***	0.27*	-0.05	0.02	0.15	0.34	0.82***	0.83***
		(2.20)	(3.58)	(4.29)	(3.16)	(4.27)	(1.94)	(-0.24)	(0.07)	(0.69)	(1.34)	(4.04)	(2.84)
	High	0.07	0.27**	0.25**	0.39***	0.52***	0.46***	-0.18	-0.25*	-0.21	-0.14	0.18	0.37
		(0.55)	(2.17)	(2.21)	(4.15)	(4.80)	(4.40)	(-1.23)	(-1.96)	(-1.05)	(-0.78)	(1.39)	(1.60)
Asset Growth	Low	0.59***	0.67***	0.58***	0.57***	0.69***	0.10	-0.14	0.03	-0.13	-0.18	0.21	0.35
		(3.33)	(4.18)	(4.73)	(4.37)	(4.61)	(0.60)	(-0.74)	(0.15)	(-0.78)	(-1.00)	(1.31)	(1.21)
	High	-0.01	0.25	0.38***	0.30**	0.47***	0.48***	-0.17	-0.27*	0.04	0.09	0.54**	0.71***
		(-0.06)	(1.56)	(2.67)	(2.32)	(3.28)	(4.07)	(-1.42)	(-1.67)	(0.26)	(0.49)	(2.50)	(2.67)
CBOP	**Low**	0.06	0.46**	0.43***	0.49***	0.52***	0.45***	-0.38**	-0.32	0.03	0.18	0.15	0.52**
		(0.36)	(2.58)	(3.09)	(3.42)	(3.12)	(3.22)	(-2.00)	(-1.50)	(0.14)	(1.17)	(0.98)	(2.10)
	High	0.41***	0.42***	0.51***	0.55***	0.60***	0.19	-0.07	-0.12	0.12	0.03	0.45**	0.52*
		(3.22)	(3.12)	(4.16)	(3.90)	(4.30)	(1.36)	(-0.54)	(-0.83)	(0.94)	(0.15)	(2.21)	(1.87)
XFIN	Low	0.25	0.43***	0.38***	0.45***	0.49***	0.23	-0.25	-0.14	-0.25	-0.09	0.44***	0.68***
		(1.45)	(4.21)	(3.46)	(3.69)	(3.76)	(1.45)	(-1.41)	(-0.86)	(-1.28)	(-0.53)	(2.61)	(2.62)
	High	0.12	0.38**	0.46***	0.39***	0.79***	0.65***	-0.00	-0.11	0.05	0.05	0.44**	0.43
		(0.77)	(2.18)	(2.92)	(2.69)	(4.60)	(4.71)	(-0.01)	(-0.55)	(0.26)	(0.23)	(1.99)	(1.52)
PPE%	**Low**	0.25	0.73***	0.60***	0.60***	0.69***	0.43***	-0.17	-0.14	0.17	0.15	0.39	0.54*
		(1.51)	(4.29)	(4.22)	(4.44)	(4.35)	(2.91)	(-1.02)	(-0.66)	(0.80)	(0.69)	(1.61)	(1.80)
	High	0.19	0.33**	0.25*	0.43***	0.51***	0.32***	-0.08	-0.20	-0.28	-0.02	0.33**	0.41*
		(1.26)	(1.97)	(1.91)	(2.87)	(3.78)	(2.79)	(-0.54)	(-1.33)	(-1.44)	(-0.10)	(2.36)	(1.80)

Table 16: Limited attention effects

This table reports the Fama-MacBeth regression results on how the NRR effect varies with investors' look-back attention (LBA). The dependent variables are the average monthly returns from the end of month *t* to the end of month t+1 for column (1) and (2), t+3 for (3) and (4), t+6 for (5) and (6), and t+12 for (7) and (8). Investors' look-back attention, LBA, is measured as the monthly sum of human user requests for 10-Ks/Qs from EDGAR that are publicly available for greater than or equal to 360 days following Drake, Roulstone, and Thornock (2016) averaged over month [t-6, t-1]. Low attention dummy (LowLBA) equals 1 if the look-back attention is lower than the median value of the monthly cross-section. The High/Low attention group is defined the same way. Panel A reports regressions with and without the dummy and interaction terms. Panel B reports subsample regressions in each group with control variables. See the Appendix for the other variable definitions. All the variables are winsorized quarter by quarter at the 1% level for both tails. Fama-French (1997) industry dummies (based on 48 industries) and a constant are also included in each first-stage regression for all columns. Newey-West (1987) adjusted (3 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) 1m	(2) 1m	(3) 3m	(4) 3m	(5) 6m	(6) 6m	(7) 12m	(8) 12m
Panel A: interacting with low attention dummy								
NRR	2.827***	2.053*	3.496***	1.634*	2.868***	1.181	3.250***	1.429*
	(4.19)	(1.92)	(5.64)	(1.80)	(4.71)	(1.32)	(5.33)	(1.95)
LowLBA		-1.431		-3.441***		-3.324***		-3.663***
		(-1.27)		(-4.29)		(-4.23)		(-5.16)
LowLBAxNRR		1.613		3.670***		3.473***		3.784***
		(1.30)		(4.00)		(3.96)		(5.00)
Constant	-1.178	-0.490	-2.187**	-0.461	-1.777**	-0.175	-2.289***	-0.541
	(-1.42)	(-0.46)	(-2.59)	(-0.43)	(-2.15)	(-0.17)	(-2.96)	(-0.66)
R-squared	0.061	0.063	0.070	0.072	0.075	0.079	0.082	0.087
Observations	362,691	362,691	359,203	359,203	354,062	354,062	343,952	343,952
Panel B: separate regressions in High/Low attention groups								
	Low	High	Low	High	Low	High	Low	High
NRR	3.443***	2.628***	4.230***	2.094**	3.495***	1.256	4.047***	1.207*
	(3.56)	(2.82)	(6.71)	(2.34)	(6.89)	(1.48)	(7.16)	(1.94)
Beta	-0.006	0.003	-0.050	-0.018	-0.047**	-0.029	-0.048***	-0.043*
	(-0.13)	(0.04)	(-1.55)	(-0.37)	(-2.04)	(-0.73)	(-2.99)	(-1.78)
Size	-0.180**	-0.176**	-0.013	-0.013	0.039	0.039	0.072	0.073*
	(-1.99)	(-2.51)	(-0.19)	(-0.23)	(0.62)	(0.80)	(1.36)	(1.96)
BM	0.365***	0.102	0.377***	0.077	0.355***	0.066	0.324***	0.044
	(3.58)	(1.29)	(4.62)	(1.06)	(5.25)	(1.00)	(5.15)	(0.75)
Rev	-2.163***	-1.577***	-0.973***	-0.428	-0.577**	-0.083	-0.290*	-0.072
	(-4.29)	(-3.34)	(-3.24)	(-1.10)	(-2.26)	(-0.24)	(-1.69)	(-0.29)
MOM	-0.138	-0.509	-0.155	-0.438	-0.166	-0.322	-0.141	-0.261
	(-0.59)	(-1.13)	(-0.75)	(-1.15)	(-0.81)	(-0.99)	(-0.90)	(-1.17)
LTRev	-0.093*	-0.044	-0.033	-0.031	-0.022	-0.024	-0.020	-0.030
	(-1.69)	(-1.18)	(-0.81)	(-0.94)	(-0.64)	(-0.85)	(-0.70)	(-1.50)
Constant	-1.006	0.328	-3.068***	-0.710	-2.895***	-0.438	-3.782***	-0.717
	(-0.73)	(0.25)	(-2.98)	(-0.53)	(-2.76)	(-0.36)	(-3.54)	(-0.80)
R-squared	0.095	0.146	0.107	0.158	0.113	0.169	0.123	0.173
Observations	167,307	177,522	166,026	175,852	163,907	173,590	159,453	169,292

Table 17: Bivariate dependent-sort portfolio on textual characteristics

This table reports calendartime portfolio Fama-French 5factor alphas (in percentage) for samples of high and low levels (relative to median) of tones, modal, uncertainty, litigiousness, file size, and document similarity. Newey-West (1987) adjusted (3 lags) tstatistics are shown below the estimates, and statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

		EW						VW					
		Q1	Q2	Q3	Q4	Q5	Q5-Q1	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Negative	Low	0.38**	0.54***	0.48***	0.39***	0.62***	0.20	-0.22	-0.07	-0.20	0.03	0.16	0.32
		(2.01)	(3.02)	(3.60)	(2.74)	(3.22)	(0.84)	(-0.92)	(-0.32)	(-1.29)	(0.14)	(0.73)	(1.00)
	High	0.39**	0.44**	0.47***	0.58***	0.62***	0.22	-0.26	-0.19	-0.20	-0.19	0.50**	0.76**
		(2.10)	(1.97)	(2.65)	(3.43)	(2.96)	(1.20)	(-1.32)	(-0.89)	(-0.99)	(-0.95)	(2.13)	(2.15)
Positive	**Low**	0.25	0.59**	0.65***	0.60***	0.64***	0.30*	-0.31	-0.26**	-0.02	-0.10	0.65**	0.85**
		(1.28)	(2.49)	(3.89)	(3.78)	(4.64)	(1.70)	(-1.12)	(-2.05)	(-0.10)	(-0.40)	(2.46)	(2.01)
	High	0.51***	0.59**	0.42**	0.52***	0.75***	0.24	-0.15	-0.10	-0.15	-0.04	0.64***	0.79**
		(2.72)	(2.46)	(2.29)	(3.58)	(3.58)	(1.47)	(-0.80)	(-0.52)	(-0.82)	(-0.17)	(2.68)	(2.01)
Modal Strong	Low	0.42**	0.31*	0.34**	0.49***	0.56***	0.14	-0.13	-0.45*	-0.08	0.16	0.14	0.26
		(2.37)	(1.86)	(2.24)	(3.33)	(3.45)	(0.92)	(-0.81)	(-1.82)	(-0.40)	(0.59)	(0.75)	(1.08)
	High	0.48**	0.80***	0.65***	0.64***	0.62***	0.15	-0.08	-0.03	-0.10	0.29	0.62**	0.70*
		(2.09)	(2.90)	(4.06)	(3.77)	(3.13)	(0.64)	(-0.36)	(-0.10)	(-0.53)	(1.24)	(2.58)	(1.85)
Modal Weak	Low	0.31*	0.41*	0.49***	0.69***	0.70***	0.40*	-0.31	-0.18	-0.26	0.17	0.25	0.52
		(1.81)	(1.78)	(3.45)	(4.22)	(3.77)	(1.84)	(-1.33)	(-0.93)	(-1.47)	(0.85)	(1.13)	(1.64)
	High	0.47**	0.66***	0.52***	0.57***	0.76***	0.35*	0.15	-0.31	0.16	0.02	0.72***	0.62*
		(2.27)	(2.84)	(2.60)	(3.59)	(3.66)	(1.71)	(0.89)	(-1.42)	(0.70)	(0.08)	(2.90)	(1.90)
Uncertain	Low	0.33**	0.67**	0.56***	0.71***	0.58***	0.25	-0.15	-0.19	-0.17	0.23	0.22	0.37
		(1.98)	(2.40)	(3.52)	(4.48)	(3.74)	(1.50)	(-0.77)	(-0.85)	(-0.91)	(1.16)	(1.23)	(1.28)
	High	0.21	0.44**	0.51***	0.48***	0.67***	0.35	-0.30	-0.42*	-0.04	-0.25	0.74***	0.92**
		(0.73)	(2.20)	(3.31)	(3.11)	(3.11)	(1.43)	(-1.23)	(-1.79)	(-0.20)	(-1.21)	(2.94)	(2.59)
Litigious	Low	0.23	0.57***	0.45***	0.54***	0.45**	0.22	-0.47**	0.22	-0.30	-0.17	-0.36	0.11
		(1.12)	(3.17)	(3.08)	(2.97)	(2.23)	(1.14)	(-1.99)	(0.91)	(-1.47)	(-0.84)	(-1.35)	(0.30)
	High	0.37**	0.70**	0.63***	0.54***	0.65***	0.28*	0.06	-0.03	0.00	0.31	0.88***	0.83**
		(2.23)	(2.24)	(3.68)	(3.48)	(4.17)	(1.73)	(0.29)	(-0.10)	(0.02)	(1.49)	(4.17)	(2.59)
File Size	Low	0.55**	0.39***	0.47***	0.53***	0.57***	0.17	-0.17	-0.20	-0.33*	0.12	0.10	0.44*
		(2.47)	(2.73)	(3.75)	(3.85)	(3.66)	(1.12)	(-0.64)	(-0.92)	(-1.85)	(0.73)	(0.58)	(1.79)
	High	0.24*	0.52**	0.47***	0.47***	0.72***	0.48***	-0.14	-0.24	-0.25	0.13	0.45**	0.59**
		(1.75)	(2.52)	(3.37)	(3.20)	(4.77)	(3.99)	(-1.16)	(-1.29)	(-1.12)	(0.76)	(2.49)	(2.42)
Cosine Similarity	**Low**	0.33**	0.43**	0.36***	0.61***	0.61***	0.30**	-0.20	-0.25	-0.29*	0.13	0.47**	0.69***
		(2.09)	(2.03)	(2.75)	(4.05)	(3.73)	(2.35)	(-0.93)	(-1.49)	(-1.71)	(0.85)	(2.48)	(2.96)
	High	0.50**	0.59***	0.59***	0.42***	0.64***	0.14	0.05	-0.06	0.07	0.02	0.29	0.24
		(2.58)	(3.98)	(4.38)	(3.01)	(4.32)	(0.83)	(0.25)	(-0.28)	(0.35)	(0.15)	(1.29)	(0.74)

Appendix 1: Variable Definitions

Table A.1: Asset pricing variables in the main tests

Variable	Definition
NRR	News redundancy ratio
Beta	CAPM beta using daily returns in month t following Scholes and Williams (1977)
Size	The natural logarithm of market capitalization at the end of month t
BM	The natural logarithm of book-to-market of equity by the end of month t following Fama and French (2008)
Rev	Short-term reversal, buy-and-hold return in month t
MOM	Momentum, cumulative return from month t-11 to t-1 following Jegadeesh and Titman (1993)
LTRev	Long-term reversal, cumulative return from month t-59 to t-12 following De Bondt and Thaler (1985)
IVOL	Idoisyncratic volatility by Fama-French 3-factor model on daily returns in month t following Ang et al. (2006)
ILLIQ	Illiquidity using daily returns in month t following Amihud (2002)
AG	Growth rate on total asset following Cooper et al. (2008)
CBOP	Cash-based operating profitability (Ball et al., 2016)
Accruals	Accruals divided by total asset following Sloan (1996)
SUE	Standardized unexpected earnings following Livnat and Mendenhall (2006)
ΔNEG	Quarter-on-quarter change of the proportion of negative words defined by Loughran and McDonald (2011)
ΔPOS	Quarter-on-quarter change of the proportion of positive words defined by Loughran and McDonald (2011)
ln(FSize)	The natural logarithm of original file size (in bytes) following Loughran and McDonald (2014)
$\Delta ln(FSize)$	Quarter-on-quarter change of the natural logarithm of original file size (in bytes) following Loughran and McDonald (2014)
Sim_Jaccard	Jaccard similarity between the current filing and the previous year's filing following Cohen et al. (2020)

Appendix 2: Validation Tests

It is important to check whether these measures indeed capture news and stale-news. I examine the different roles of news and stale-news in resolving information asymmetry and updating investors' beliefs. *Given the raw document size*, higher news redundancy (low news density) should go with higher levels of information asymmetry (Loughran and McDonald, 2014) and smaller price adjustments (Francis and Schipper, 1999; Roychowdhury and Sletten, 2012) upon announcement of the filing. The amount of news, rather than the processing cost, is the first-order determinant of uncertainty resolution by the disclosure.

A.2.1: Higher NRR, less news, more uncertainty

More informative disclosure in corporate filings should resolve more uncertainty and lower the information asymmetry to a greater extent (Belo et al., 2018; Loughran and McDonald, 2014). A financial report with high news redundancy (low news density) should be less informative *given the document's length and other factors controlled.* Then we should observe a positive relationship between news redundancy and changes in information asymmetry (uncertainty resolution), which is equivalently a negative relationship between news density and resolution of uncertainty. On the other hand, stale-news density or redundancy should not affect the degree of information asymmetry or uncertainty resolution. Then we should observe no significant relationship between stale-news redundancy and changes in information asymmetry.

Using idiosyncratic volatility and analyst forecast dispersion as proxies for the uncertainty of firm value in the information environment following Loughran and McDonald (2014), I measure the degree of uncertainty resolution using the proportional change ($\Delta\%$) of these proxies around the filing announcement day.

I conduct the test using the following regression specification:

$$\Delta\%Uncertainty_{i,t} = \alpha + \beta Red_{i,t} + X_{i,t} \cdot \gamma + \theta_i + \delta_t + \epsilon_{i,t} \tag{13}$$

where $\Delta\%Uncertainty_{i,t}$ is measured in two ways: the proportional change of Fama-French 3-factor model root mean squared error (RMSE) estimated from trading days

[-48, -6] to [6,28][12] relative to filing announcement day; and the proportional change of analyst forecast dispersion from the current quarter to next. The variable of interest, $Red_{i,t}$, stands for different redundancy ratios of the filing: news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and the total redundancy ratio (TRR). Similar to Loughran and McDonald (2014), I include the natural logarithms of file size (ln(FSize)), firm market capitalization (ln(ME)), book-to-market ratio of equity (ln(BM)), one plus the total number of analysts making quarterly earnings forecast before the filing announcement (Analyst), market model alpha and RMSE estimated using trading days [-48, -6] relative to filing announcement date (pre_Alpha and pre_RMSE), and absolute value of market model cumulative abnormal return for trading days [0, 2] ($|CAR(0,2)|$). I also control firm fixed effects (θ_i) and quarter fixed effects (δ_t) for each regression. I report t-statistics in parentheses with the standard errors clustered by firm and calendar quarter.

Since news rather than stale-news resolves uncertainty, I expect the coefficient β be significantly positive for NRR and TRR, insignificant for SNR.

The results are consistent with my expectations. Table A.2.1 shows that the coefficients on NRR and TRR are positive and significant, while the coefficients on SNR are insignificant. In column (1) to (3), I report the regressions of proportional changes in RMSE on the redundancy measures. The coefficients on NRR and TRR are 0.134 (t-statistic 2.35) and 0.168 (t-statistic 2.46) respectively. While the coefficient on SNR is -0.031 (t-statistic -0.80). In column (4) to (6), I report the regressions of proportional changes in analyst forecast errors on the redundancy measures. The coefficients on NRR and TRR are 1.136 (t-statistic 4.02) and 1.902 (t-statistic 4.68). In comparison, the coefficient for SNR is -0.067 (t-statistic -0.30). These results verify that the news redundancy ratio can capture the richness of news in financial reports, while the stale-news redundancy ratio does not affect resolving information asymmetry.

[12] The results are robust using different windows to estimate RMSE

Table A.2.1: News (stale-news) redundancy and uncertainty resolution
This table reports regressions of uncertainty resolution measures on news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and total redundancy ratio (TRR). The dependent variables are proportional change of (Δ%) Fama-French 3-factor model RMSE estimated from trading days [-48, -6] to [6,28] ($\Delta\%RMSE_3$) for (1), (2), and (3), and Δ% analyst forecast dispersion from the current quarter to next ($\Delta\%Fdisp$) for (4), (5), and (6). Analyst forecast dispersion is measured as standard deviation of analysts' forecasts divided by the mean forecast. Control variables are the natural logarithms of file size (ln(FSize)), market capitalization (ln(ME)), book-to-market ratio of equity (ln(BM)), one plus the total number of analysts making quarterly earnings forecast before the filing announcement (Analyst), market model root mean squared error and alpha estimated using trading days [-48, -6] relative to filing announcement date (pre_RMSE and pre_Alpha) and absolute value of market model cumulative abnormal return for trading days [0, 2] ($|CAR(0,2)|$). All the variables are winsorized quarter by quarter at the 1% level for both tails. All regressions include firm and calendar quarter fixed effects. t-statistics are in parentheses with standard errors clustered by firm and calendar quarter. Statistical significance at 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) $\Delta\%RMSE_3$	(2) $\Delta\%RMSE_3$	(3) $\Delta\%RMSE_3$	(4) $\Delta\%Fdisp$	(5) $\Delta\%Fdisp$	(6) $\Delta\%Fdisp$
NRR	0.134**			1.136***		
	(2.35)			(4.02)		
SNR		-0.031			-0.067	
		(-0.80)			(-0.30)	
TRR			0.168**			1.902***
			(2.46)			(4.68)
ln(FSize)	0.013***	0.014***	0.010***	0.118***	0.118***	0.086***
	(3.53)	(3.32)	(2.72)	(7.40)	(6.69)	(5.47)
pre_Alpha	-0.042***	-0.041***	-0.041***	-0.868***	-0.864***	-0.865***
	(-3.51)	(-3.47)	(-3.45)	(-12.04)	(-12.00)	(-12.05)
pre_RMSE	-0.064***	-0.064***	-0.064***	0.049***	0.047***	0.045**
	(-13.28)	(-13.30)	(-13.35)	(2.77)	(2.67)	(2.58)
$\|CAR(0,2)\|$	0.739***	0.739***	0.737***	0.563**	0.546**	0.544**
	(15.77)	(15.75)	(15.72)	(2.46)	(2.38)	(2.36)
ln(ME)	-0.044***	-0.045***	-0.045***	0.076***	0.076***	0.075***
	(-10.73)	(-10.77)	(-10.76)	(3.86)	(3.83)	(3.84)
ln(BM)	-0.005	-0.005	-0.005	0.075***	0.076***	0.074***
	(-1.33)	(-1.29)	(-1.30)	(3.03)	(3.05)	(2.99)
Analyst	-0.019***	-0.019***	-0.019***	-0.029	-0.023	-0.025
	(-4.71)	(-4.63)	(-4.64)	(-1.07)	(-0.87)	(-0.92)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
YrQt FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.083	0.083	0.083	0.065	0.065	0.066
Observations	154,187	154,187	154,187	71,344	71,344	71,344

A.2.2: Higher NRR, less news, smaller magnitude of price adjustment

Stock prices should adjust in response to news arrivals. The magnitude of price adjustment around filing day should positively relate to the value relevance of the financial report (Francis and Schipper, 1999; Roychowdhury and Sletten, 2012). A filing with high news *density* should be more value-relevant on average. Thus, I expect that news *redundancy* is negatively associated with the magnitude of price adjustments[13]. Stale-news should already be incorporated into the stock prices; thus, high stale-news *density* should come with a smaller price response. So I expect stale-news *redundancy* is positively associated with the magnitude of price adjustment. The effect of total redundancy is unclear ex-ante.

I conduct the test as follows:

$$|CAR^f(0,h)_{i,t}| = \alpha + \beta Red_{i,t} + X_{i,t} \cdot \gamma + \theta_i + \delta_t + \epsilon_{i,t,h} \tag{14}$$

where the dependent variable is the absolute value of current filing period market model cumulative abnormal return (CAR^f) from filing date (day 0) to day h inclusive, with $h = 1, 5$, for stock i in quarter t. The variable of interest, $Red_{i,t}$, represents different redundancy ratios of the filing: news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and the total redundancy ratio (TRR). The control variables, $X_{i,t}$, are similar to those used by Loughran and McDonald (2014). I expect the β to be significant and negative for NRR but positively significant for SNR.

The results (Table A.2.2) and consistent with the above expectations. The coefficient on NRR is -1.977 (t-statistic -4.69) in column (1) and -2.326 (t-statistic -3.65) in column (4). The coefficient on SNR is 1.987 (t-statistic 6.53) in column (2) and 2.834 (t-statistic 5.94) in column (5). These results suggest that, given the original document's length, filings with higher NRR contain less news and lead to smaller magnitudes of price adjustments. Stale-news generates the opposite pattern.

[13] Good news and bad news should cancel out each other when determining the direction and magnitude of price adjustment following the financial report, but this will work against my results.

Table A.2.2: News (stale-news) redundancy and price adjustments
This table reports regressions of price adjustments on news redundancy ratio (NRR), stale-news redundancy ratio (SNR), and total redundancy ratio (TRR). The dependent variables are absolute values of market model cumulative abnormal returns estimated using trading day window [0, 1] relative to filing date ($|CAR(0,1)|$) for column (1), (2), and (3), and $|CAR(0,5)|$ estimated using window [0, 5] for column (4), (5), and (6). The dependent variables are in percentages. Control variables are the natural logarithms of file size (ln(FSize)), firm market capitalization (ln(ME)), book-to-market ratio of equity (ln(BM)), average daily turnover ratio (ln(Turnover), one plus the total number of analysts making quarterly earnings forecast before the filing announcement (Analyst), market model root mean squared error and alpha estimated using trading days [-48, -6] relative to filing announcement date (pre_RMSE and pre_Alpha). All the right-hand-side variables are winsorized quarter by quarter at the 1% level for both tails. All regressions include firm and calendar quarter fixed effects. t-statistics are in parentheses with standard errors clustered by firm and calendar quarter. Statistical significance at 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	(1) $\|CAR(0,1)\|$	(2) $\|CAR(0,1)\|$	(3) $\|CAR(0,1)\|$	(4) $\|CAR(0,5)\|$	(5) $\|CAR(0,5)\|$	(6) $\|CAR(0,5)\|$
NRR	-1.977***			-2.326***		
	(-4.69)			(-3.65)		
SNR		1.987***			2.834***	
		(6.53)			(5.94)	
TRR			1.394***			2.933***
			(2.65)			(3.81)
ln(FSize)	-0.105***	-0.140***	-0.126***	-0.039	-0.090**	-0.086***
	(-4.24)	(-5.37)	(-5.26)	(-1.21)	(-2.64)	(-2.74)
ln(ME)	-0.460***	-0.458***	-0.458***	-0.736***	-0.734***	-0.734***
	(-10.69)	(-10.67)	(-10.66)	(-13.00)	(-12.98)	(-12.93)
ln(BM)	-0.018	-0.018	-0.019	-0.047	-0.047	-0.049
	(-0.48)	(-0.49)	(-0.53)	(-0.86)	(-0.86)	(-0.90)
ln(Turnover)	0.326***	0.326***	0.331***	0.464***	0.463***	0.470***
	(9.75)	(9.76)	(9.88)	(9.61)	(9.61)	(9.74)
Analyst	0.087**	0.086*	0.080*	0.049	0.049	0.040
	(2.02)	(1.99)	(1.86)	(0.88)	(0.87)	(0.72)
pre_RMSE	0.355***	0.354***	0.355***	0.576***	0.575***	0.577***
	(17.39)	(17.33)	(17.44)	(20.39)	(20.32)	(20.45)
pre_Alpha	-0.339***	-0.339***	-0.340***	-0.592***	-0.590***	-0.592***
	(-5.69)	(-5.67)	(-5.69)	(-5.35)	(-5.34)	(-5.36)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
YrQt FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.209	0.209	0.209	0.209	0.209	0.209
Observations	151,571	151,571	151,571	151,433	151,433	151,433