Implied Volatility and Its Information Content of Future News

CHEN Sipeng

PhD

Hong Kong Polytechnic University

2019

The Hong Kong Polytechnic University
School of Accounting and Finance

Implied Volatility and Its Information Content of Future News

CHEN Sipeng

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy

March 2019

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

____________________ (Signed)

CHEN Sipeng (Name of student)

Abstract

My study examines the information content of implied volatility (IV) and finds out that implied volatility(IV) can predict future realized volatility (RV) through the channel of predicting future news.

There are two pieces of supporting evidence in my results. The first evidence is the "RV-relevance" of news, which refers to the fact that RV can be explained by the concurrent news in a batch of time-series regressions for each individual firm. The cross-sectional average of the regression coefficients is significantly positive with a t-statistic over 35, and more than 90% of the stocks have positive coefficients. The second evidence is the "IV-predictability" of news, which refers to the fact that the future news can be forecasted by the IV in a batch of time series regressions for each individual firm. The cross-sectional average of the regression coefficients is significantly positive with a t-statistic about 20, and around 65% of the stocks have positive coefficients. Both pieces of evidence hold up robustly in different measures of news and in most kinds of news.

More specifically, I devise two measures to quantify the news. One is the news intensity (N) by counting the news occurrence(s) within a month. The other one is the news volatility (NV) which measures the total news impact magnitude, as defined by the sum of squares of the Composite Sentiment Score(s) (CSS) within a month. IV has a greater prediction power for NV than for N. Consistently, NV is greater in RV-relevance than N.

Furthermore, I apply different news classification methods to find: which kind of news is more RV-relevant; which kind of news is more IV-predictable; and which kind of news drives the RV_t–IV_{t-1} forecasting relation the most.

In terms of timing predictability, the unscheduled news is more RV-relevant while the scheduled news is more IV-predictable. The scheduled news drives the forecasting relationship more.

In terms of news formats as proxy for different information roles of media, both news-flash (proxy for information dissemination role) and full-article (proxy for information creation role) are similarly strong in RV-relevance. But news-flash is more IV-predictable, and thus drives the forecasting relation more.

In terms of different news content groups, the overall results suggest that the strength of the IV-predictability is monotonically increasing with the strength of the RV-relevance. Thus, I can select all the news content groups that have high RV-relevance to form new measures of news. These new measures of news drive the RV_t–IV_{t-1} forecasting relation significantly more than the original news measures of all the news.

Eventually, through a two-stage mediation analysis, I am able to quantify the strength of the predictability through different kinds of news channels over the total predictability of IV_{t-1} on RV_t. The best mediation model captures that about one-third of the total predictability comes from the news channel.

JEL Classification: G12, G14, G17

Keywords: Volatility Forecasting, Implied Volatility, News, Uncertainty

Acknowledgements

I firstly and primarily would like to thank my chief supervisor Dr LI Gang for inspiring me with the research ideas, and his patience in training me so intensively with critical thinking to be a serious scholar, and spending an extensive amount of time and energy in helping me. I thank my co-supervisor Prof Agnes Cheng for supporting me in many oversea training opportunities. I thank my first boss Prof Chu Zhang for lighting up my research path in the very beginning. I thank Dr Steven Wei for offering me with career opportunities. I thank my friends Fu Jiajia, David Pecha, Mariem, Emna and Ye Jiayao for their company over the years.

Finally, I would like to thank my parents for their support and encouragement. I thank my wife for her deep understanding of me, cleaving to me even though life is tough, and forcing me to be upbeat at all times. I thank my church-mates for praying for my family and me as always.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Equity volatility can proxy for risk, and its forecasting would be an interesting topic: to investors, because they can minimize risks while maximizing returns; to regulators, because they can have an ex-ante measure of the economic uncertainty; and to managers, because they can plan their investment ahead with better risk management.

In the past, when forecasting equity volatility, we can utilize its characteristic of clustering(Mandelbrot, 1963), meaning that we can simply use the past volatility to extrapolate future volatility. Later researchers (Latane & Rendleman Jr, 1976; Day & Lewis, 1992; Lamoureux & Lastrapes, 1993; Canina & Figlewski, 1993) set eyes on the option implied volatility to see whether it could be a better predictor than the past volatility. The comparison results were mixed and undetermined until when Christensen and Prabhala (1998) fixed the mismatch problem between the option maturity and its forecasting period in the empirical settings. Thanks to these later findings (Christensen & Prabhala, 1998; Jiang & Tian, 2005), academics can confirm that the option implied volatility is a better predictor and subsumes the forecasting power of past volatility.

However, academics have not yet taken any effort to explain why it is so, but simply assume that implied volatility is, by definition, the market's forecast on future volatility of the underlying asset. My study is going to validate this assumption, extend the previous results, and further nail down its cause that the options market can predict future volatility through the channel of predicting future news. Plus, this future news is highly relevant to the concurrent realized volatility.

In the empirical setup, I retrieve the stock data from CRSP and filter them with share code 10 & 11 and the three primary US stock exchanges(NYSE, AMEX, NASDAQ). The option data comes from Option Metrics IvyDB, and the news data is obtained from RavenPack Analytics 1.0. The sample period spans from 2004 to 2017.

Specifically, the realized volatility(**RV**) is the standard deviation of the daily returns within a month and then annualized. The implied volatility(**IV**) is an interpolated IV of the closest being at-the-money options by matching the maturity date to the forecasting period of the next month. I further adjust for stale price bias by averaging the IV(s) in each put-call pair. For the news data, I devise two measures to quantify the news. One is the news intensity(**N**) by counting the news occurrence(s) within a month. The other one is the news volatility(**NV**) which measures the total news impact magnitude, as defined by the sum of squares of the Composite Sentiment Score(s)(**CSS**) within a month. Additionally, CSS is an ex-ante sentiment score provided by RavenPack, which measures how positive or negative the news story is in terms of its projected short-term impact on equity prices.

My major empirical analysis can be structured into three parts: first, prove that the RV is highly relevant to the concurrent news, and refer this part as the analysis of "RV Relevance" of news; second, show that IV can predict future news, and refer this part as the analysis of the "IV Predictability" of news. Taken together, IV can forecast future RV through forecasting the future RV-relevant news. In other words, the two key features of news – RV relevant and IV predictable – bridge the connection between the current IV and the future RV. Third, I implement a two-stage mediation analysis suggested by Preacher and Hayes (2008) to quantify the strength of the news channel predictability over the total predictability of IV_{t-1} on RV_t.

In the analysis of RV-Relevance of news, I perform a batch of time-series regressions for each individual firm by using their news to explain their concurrent RV. In the results, the regression coefficients are positive for over 90% of the firms and the overall cross-sectional average is significantly positive with a t-statistic over 35. The coefficients become even more positive when controlling for the past RV. These results suggest that the RV is highly relevant to its concurrent news, and rule out the alternative explanation that our sampled news is a mere reflection of the past RV.

In the analysis of IV-Predictability of news, I conduct a batch of time-series regressions for each individual firm by using their IV to predict the future news. Eventually, around 65% of the stocks have positive coefficients and the cross-sectional coefficient average is significantly positive with a t-statistic about 20. Even though these results become weaker when controlling for the past news, the new average can still maintain its statistical significance with a t-statistic over 14. These results indicate that IV can predict the novel part of news whose information is not repeated from the past news.

In the two-stage mediation analysis, the mediators are my interested news variables at the current month(i.e., N_t or NV_t). In the first stage, I use the past implied volatility (IV_{t-1}) to predict each mediator one by one. In the second stage, I conduct the regression test of IV_{t-1} predicting RV_t and controlling for all the mediators. The non-news channel predictability is represented by the coefficient of IV_{t-1} in stage two, while the news channel predictability of each mediator is measured by the product of the coefficient of IV_{t-1} in stage one and the coefficient of the corresponding mediator in stage two. Total equality exists between the sum of decomposed predictability components and the total predictability, which can be acquired from the coefficient of regressing RV_t on IV_{t-1} without controlling the mediators. Eventually, we can quantify the strength of the news channel predictability as a proportion over the total predictability. Our best mediation model suggests that about a quarter of the total predictability comes from the news channel.

Both pieces of evidence of the RV-relevance and IV-predictability of news hold up robustly in different news measures and in most kinds of news. In our mediation analysis, some kinds of news apparently stand out from the others in terms of their proportion over the news predicting channel.

Regarding to different measures of news, IV has greater prediction power for NV than for N, and consistently, RV is more relevant to NV than to N. Contrasting to N, the analyses of IV-predictability and RV-relevance of NV present the following key results: more positive t-statistic for the regression coefficient average, more proportion of the firms with positive relation, and a higher R^2.

Regarding to different kinds of news, I can categorize the news data based on a variety of news characteristics: (1) the timing predictability – scheduled news and unscheduled news; (2) the news format – news-flash, full-article, press-release, tabular-material and SEC filings; and (3) the news content groups[1] – "earnings", "analyst-ratings", "insider-trading", "equity-actions", and etc. I then apply these classification methods one by one to find: 1 which kind of news is more RV-relevant; 2 which kind of news is more IV-predictable; and 3 which kind of news drives the RV_t–IV_{t-1} forecasting relation the most.

In terms of timing predictability, the option traders can strikingly predict the unscheduled news as well as the scheduled news. Controlling the previous news can weaken the predictability of scheduled news, but strengthen the predictability of unscheduled news. Even after adding the controls, the results suggest that the scheduled news is more IV-predictable, while the unscheduled news is more RV-relevant. These results of the unscheduled news can be attributed to its charac-

[1]The groups are at the same level as the RavenPack's content classification. Figure A1 depicts the whole taxonomy.

teristic of unpredictable timing, which is hard to predict its occurrence and adds an additional factor to cause market shocks. In the mediation analysis, the strength of scheduled news intensity is slightly higher than the unscheduled news intensity. However, the strength of unscheduled news volatility is over 20% high than the strength of scheduled news volatility. This suggests that even though the investors are poor in predicting the occurrence of unscheduled news on average, they are able to predict the unscheduled news that is important, as measured by its news volatility. As a result, the high RV-relevance of this unscheduled news was not totally offset its not-so-poor IV-predictability.

In terms of news format, news-flash can purely proxy for the information dissemination role of media because it contains no body text, while full-article can roughly proxy for the information creation role because it has a propensity to have additional editorial or analytical content in their body text. Both news-flash (proxy for information dissemination role) and full-article (proxy for information creation role) are similarly strong in RV-relevance. But news-flash is more IV-predictable and thus influence the forecasting relation between RV_t and IV_{t-1} the most. On the other hand, the relatively weaker IV-predictability of news article implies that the in-depth analytical skills of the reporters are different from the expertise of the option traders, which adds value to the market. Unlike the previous two news formats, press-releases are firm-initiated, and its disclosure can be voluntary and discretionary. Surprisingly, its RV-relevance and IV-predictability is also significant, but not as strong as the previous two. The tabular material and SEC filings are less important due to their small sample size. On the strength of the news channel, the mediation analysis generates a consistent ranking of the results: news-flash > full-article > press-releases, and the remaining two formats lag behind with a large gap.

In terms of different news content groups, the overall results across groups suggest that the strength of the IV-predictability is monotonically increasing with the strength of the RV-relevance. The two most RV-relevance groups are "earnings" and "revenue". In particular, these two groups consist of 68% of the total news stories, and thus I further decompose these two groups into two independent sets of subgroups: 12 different accounting items and nine different information formats. The accounting items are the variants of "earnings"(e.g. "ebita" or "ebit") and "revenue" (e.g. "operating margin" or "same-store-sales"). Information formats are the ways that earnings news (e.g. "guidance") and revenue news (e.g. "volume") are presented. Eventually, the results across different subgroups also depict a similar pattern as across groups: the higher RV-relevance, the higher IV-predictability. Therefore, I can select all the news content groups or the earnings & revenue subgroups, which have high RV-relevance with t-statistics over 5 to form a new sample. This new sample of news drives the RV_t–IV_{t-1} forecasting relation significantly more than the original full sample of news. Additionally, in the mediation analysis, I find a consistent top three most important news in terms of its strength over the overall news channel: earnings news, analyst-ratings news and revenue news.

In the end, using a two-stage mediation analysis, I am able to quantify the strength of the news channel through which the implied volatility can forecast the future realized volatility. The strength of the channel is represented by its proportion over the total predictability of implied volatility on the future realized volatility. I filter out the unimportant news based on the knowledge of my previous discovered results. I combine the three classification methods to form a new grouping. Based on the new grouping, 32.8% of the total predictability comes from the overall news channel.

My study differs from most lead-lag relation papers which discuss option price leading stock price around some specific significant information events (Jin, Livnat, & Zhang, 2012; Hayunga & Lung, 2014; Qing, 2016; Gharghori, Maberly, & Nguyen, 2017; Zhang, 2018), and instead, I focus only on the volatility without considering the return direction. My goal is modest by trying not to claim that some informed option traders have special market insights of forecasting good news or bad news, but trying to say that the majority of option traders are sophisticated in planning whether there are significant information events ahead regardless of the news sentiment. This paper contributes to implied volatility forecasting literatures by trying to understand how it can forecast future volatility and through what kinds of news.

The remainder of this dissertation is organized as follows. Chapter 2 goes through the past literature with motivations. Chapter 3 describes the data and variables. Chapter 4 presents the baseline empirical evidences and Chapter 5 breaks them down in detail through different news classification methods. Chapter 6 concludes.

Chapter 2

Literature Review

2.1 Relation between Implied and Realized Volatility

The theoretical hypothesis to draw the relation between implied and realized volatility is that: in an efficient market, implied volatility, like option price, should contain all the forward-looking information. Therefore, it should subsume all the information that predicts the future realized volatility. This line of literature tries to document this phenomenon empirically.

Latane and Rendleman Jr (1976) is the first paper to claim that implied volatility is generally a better predictor than historical volatility in forecasting future volatility, but they construct their work in a static cross-sectional setting. In other words, stocks with higher implied volatility tend to have higher future realized volatility. The later researchers switch their focus to a dynamic setting of using time-series regression because they are interested in the information content of implied volatility, which is measured by how good the implied volatilities can forecast the future volatility of the underlying asset over the remaining period of the option.

In the early days, the time-series regression research used to generate mixed results. Both Day and Lewis (1992) and Lamoureux and Lastrapes (1993) find that implied volatility can forecast better as a biased and inefficient predictor. "Biased" means the IV coefficient of its univariate regression is not equal to one, and the intercept is not equal to zero. "Inefficient" means the residual is not white noise and there are missing variables beyond IV to predict future RV. Fleming, Ostdiek, and Whaley (1995) find that implied volatility is an efficient but biased predictor. In contrast, Canina and Figlewski (1993) refute all these results with a completely opposite one that implied volatility could not forecast future volatility. However, all of these early time-series papers fall into a pitfall of 'maturity mismatch' problem in their empirical settings. They use implied volatility with much longer maturity to predict the next week's or the next day's realized volatility.

Christensen and Prabhala (1998) are the first to implement a careful empirical setting to clearly identify that implied volatility can forecast better than historical volatility. Implied volatility is even an efficient and unbiased predictor in some of their specification in which they apply an instrumental variable framework to address the error-in-variable issue. Furthermore, Jiang and Tian (2005) calculate the model-free implied volatilities (Britten-Jones & Neuberger, 2000) and show a stronger result as an unbiased and efficient predictor. Poon and Granger (2005) compares 93 studies on volatility forecasting tests. In terms of forecasting accuracy, their survey ranks different volatility methods in the following order: implied volatility, historical volatility, GARCH, and stochastic volatility.

Busch, Christensen, and Nielsen (2011) expand the test and confirm that implied volatility can forecast stock index as an efficient and unbiased predictor in different market and a different frequency. For individual equity options, Taylor, Yadav, and Zhang (2010) study 149 US firms where 85% of these firms' implied volatility forecasts are more informative than their historical volatilities. Furthermore, the more actively the options traded, the more informative these options are. In terms of forecast power, they further point out that the at-the-money options implied volatility generally outperforms the model-free option implied volatility.

Yet, there is no paper discussing in which channel the implied volatilities can forecast the future realized volatilities. My paper is the first to bridge the gap and tackle this problem by using news information.

2.2 Equity Index and Individual Equities Options

Although implied volatility preponderates in forecasting power of the future volatilities in both equity index (Day & Lewis, 1992; Fleming et al., 1995; Christensen & Prabhala, 1998; Jiang & Tian, 2005) and individual equity (Latane & Rendleman Jr, 1976; Lamoureux & Lastrapes, 1993), the options of these two kinds of underlying assets can have subtle differences in term of trading, hedging and risk.

Equity index options are traded more frequent than the underlying basket of equities (Fleming et al., 1995), whilst the individual equity options are traded less frequent than its underlying stock. Due to the potential infrequent trading problems, the less frequent trading asset may lag from its counterpart to reflect the true values. As a result, in a rising market, the implied volatility of the call(put) option tend to be upward(downward) bias for the index option. The biases reverse for the individual equity options. Therefore, in the empirical settings, the researchers need to adjust for this kind of "stale price bias" by averaging the implied volatility in each put-call pair.

Both the equity index and individual equities are affected by macro-economic events. However, the implied volatility of individual equity option is more importantly driven by the event risk at the firm level, such as earnings announcement(Dubinsky, Johannes, Kaeck, & Seeger, 2018), while the individual effects of these firm-level events may be mixed or canceled out with each other at the index level. Therefore, the relationship between news and volatility can be identified more clearly at the firm level than at the index level.

Even though there is more noise at the firm level than at the index level, my paper focuses on analyzing the individual equity options because their relationship with the news can be identified more clearly. Meanwhile, I will adjust the infrequent trading problems by averaging the call and put options, whose bias can cancel out with each other.

2.3 The Creation and Resolution of Market Uncertainty

Ederington and Lee (1996) are the first to discuss how the scheduled and unscheduled news influence the implied volatility of options. They consider the relationship between the implied volatility of T-Bond, Eurodollar, and Deutsch-mark options with the macroeconomics news such as employment report and PPI. They find that before the scheduled macroeconomic announcement, the pre-release IV will increase by incorporating the expected price-volatility of the scheduled event; after the announcement, the IV will drop for the uncertainty has resolved. As for the unscheduled news, the market will adjust the IV upward for the remaining life of the option. Beber and Brandt (2006) test on the U.S. Treasury option with the regularly scheduled macroeconomic announcement, and find a similar result that IV declines after the announcement regardless of the sentiment of the news.

Besides macroeconomic news, at the firm level, Patell and Wolfson (1979, 1981) discover a similar result that IV increases before the earnings announcement events, and decline discontinuously after the announcement. Donders and Vorst (1996) and Donders, Kouwenberg, and Vorst (2000) find a similar result in the European Dutch market. Isakov and Perignon (2001) replicate the results in the Swiss option market, and further claim that the post-announcement IV path may depend on the sentiment of the news (i.e., good news or bad news). Dubinsky et al. (2018) extend the Patell and Wolfson (1979, 1981)'s sample period, and associate the increase/decrease of IV to the direct estimate of the earning announcement risk.

There are three important features making my study different from the previous papers: first, I extend the sample and focus on the individual stock options which is more clearly driven by the event risks; second, instead of focusing a specific type of news, I generalize the IV–news results to all kinds of firm news and identify the most volatility relevant news among them; third, I can draw a conclusion on the relationship between different strength of IV predictability and RV relevance by comparing different kinds of news.

2.4 Lead-lag Relation between Option and Stock

Informed traders choose to trade in the options market prior to the underlying stock market mainly because options can offer the trading benefits of greater financial leverage. Such benefits are even more lucrative when the options market has better liquidity than the underlying stock market (Easley, O'hara, & Srinivas, 1998). Chan, Chung, and Fong (2002) add another reason that the informed traders who only have private insights on the volatility can only bet in the options market.

Prior literatures indicates that both scheduled and unscheduled news may hit options market before the stock market: earnings announcement (Jin et al., 2012), merger/takeover announcement (Cao, Chen, & Griffin, 2005; Chan, Ge, & Lin, 2015), dividend change announcement(Zhang, 2018), stock split announcement (Gharghori et al., 2017), repurchase announcement (Qing, 2016), and analyst forecast revisions (Hayunga & Lung, 2014).

This strand of lead-lag relation between options and stock literature focuses on using options information to predict the stock return direction conditional on the informational significant events. They conjecture that informed options traders have the information advantage to forecast good news or bad news. But the goal of my paper modestly differs. I shift the focus to using the IV to predict the advent of volatility relevant event and thus forecast the future realized volatility. This study does not require the assumption that some options traders have special market insights of forecasting good news or bad news, but trying to show that the majority of options traders are sophisticated in predicting the existence of information-significant events ahead regardless of the news sentiment. This paper contributes to implied volatility forecasting literatures by trying to understand what information and how strong the IV can forecast.

2.5 The Information Roles of Media

Drake, Guest, and Twedt (2014) distinguish the two roles of media: the information dissemination role and the information creation role. Both roles supplement annual accounting information to impact the capital market. Their evidence suggests that the role of information dissemination has a higher impact than the role of information creation in mitigating the cash flow mis-pricing. Specifically, Twedt (2015) use different news formats to proxy the two roles. He regards the news-flash news as a fast rebroadcast of the headline information, and thus as a clear proxy for the role of information dissemination, whilst full-article may add editorial or analytical contents to the news story and can roughly proxy for the role of information creation. Instead of studying the supplement effect of accounting information, it would be interesting to see in an option implied volatility setting whether the options traders can predict the advent of the full article which proxy for information creation role as well as the news flash which proxy information dissemination role. In other words, my paper will test whether the options trader would have duplicate expertise as the reporters in analyzing the events in-depth.

Bushman, Williams, and Wittenberg Moerman (2016) also discuss the information role of media by utilizing the RavenPack data. They distinguish the press-releases news as another type of information source different from the media because press-releases are firm-initiated. Its advent can be voluntary and even discretionary, creating more challenges for the options trader to predict. In my dissertation, it would be interesting to test whether the options traders can handle this kind of challenge and successfully predict its advent.

Chapter 3

Data and Variables

3.1 Stock Data

I download the stock data from CRSP and filter them with ordinary common shares (share code 10 and 11) and within the three primary exchange (NYSE, AMEX, and NASDAQ). Using the stock data, we are able to construct the variable of Realized volatility (RV) as the standard deviation of daily return (r_t) within a month (m):

$$\text{Realized Volatility: RV} = \frac{\sum_{t \in m}(r_t - \bar{r})^2}{n}$$

where n is the number of trading days in the month m, $\bar{r}$ is the average return of the month. We require there are at least ten trading days of a month to calculate a valid RV data.

3.2 Option Data

Option data comes from the Option Metrics IvyDB US, which provides the calculated implied volatility (IV). Since all the individual equity options are American options, the database applies the Cox-Ross-Rubinstein binomial tree model (Cox, Ross, & Rubinstein, 1979) to numerically obtain the value of implied volatility, which can make the model price to converge to the market price. Some individual stock options may have a thin trading problem, which makes option price lags behind from its stock price. In a rising market, the implied volatility of the call(put) option tends to be downward(upward) bias. Therefore, I will calculate the average of the put and call of the nearest at-the-money options for a certain maturity. Since by convention options expire on the third Friday of each month, I need to align the maturity date to the end date of my forecasting month. Let the end of the current month to be date t, and the objective is to construct a hypothetical option that expires at the end date of next month T. First, select two options whose expiration dates are T_1 and T_2 and they are closest to T with $t < T_1 \leq T < T_2$. Let IV be the implied volatility. We can then use the following formula to interpolate the implied volatility that expires at date T:

$$\text{Implied Volatility: } \mathrm{IV}_{t,T} = \mathrm{IV}_{t,T_1} + (\mathrm{IV}_{t,T_2} - \mathrm{IV}_{t,T_1})\frac{T - T_1}{T_2 - T_1}$$

where $\mathrm{IV}_{t,T}$ is implied volatility at date t that expires at date T, and $T_1(T_2)$ is the expiration date of the two existing options whose expiration date are the closest to T and $t < T_1 \leq T < T_2$. The last available option record is December 2017, which is also the last month of my sample.

3.3 News Data

3.3.1 About RavenPack Database

The news data comes from RavenPack Analytic 1.0[1] ("RavenPack" for short), which covers a wide range of entities: companies, organizations, commodities, places, products, nationalities, and currency. My study only focuses on US company news.

[1]RavenPack Analytic 1.0 is the newest version with extensive expansion in firm coverage as well as news story coverage. Most of the prior literature used an old version called RavenPack News Analytics 4.0 where some variables, such as ENS(event novelty score), are now deprecated and replaced with an improved one.

Shevlin and Thornock (2015) mentions that there are four advantage of using RavenPack: (1) RavenPack can identify a specific content group of the news; (2) RavenPack gather news from multiple sources comparing to prior literature often with only single source; (3) easy for massive download comparing to restricted download in using Factiva; (4) RavenPack provides relevance and novelty measures to avoid duplicates or news with less impact.

The database converts the vast amount of unstructured company news stories into structured quantitative and categorical information(RavenPack, 2018). The conversion process can be boiled down into the following steps: 1) identify which company is mentioned in the news story; 2) summarize what theme is the story about; 3) detect what role the company plays in the story[2]; 4) determine the news novelty to avoid duplicate records of stories of the same theme.

Specifically, RavenPack designs a taxonomy to categorize the contents' themes into four major levels of classification: Topic, Group, Type, Sub-type. Figure A1 summarizes the taxonomy structure. At the highest level, there are only five Topics – "business", "environment", "society", "politics" and "economy". Under Topics, related events are grouped together into 56 Groups. For examples, under the business topic, we have "earnings", "legal", "analyst ratings", "credit ratings", "acquisitions-mergers", and etc. Under Groups, events with similar characteristics are classified as one Type. The data universe contains 495 Types. For instance, under the "earnings" Group, there are Types "earnings", "earnings-expectation", "earnings-guidance", "earnings-estimates", and "earnings revision". Furthermore, the Types are subdivided into 146 Sub-types which may sometimes indicate whether the news is positive, negative or neutral. Figure A1 further depicts the type "earnings-guidance" as an example in detail. The type "earnings-guidance" contains four sub-types "up", "down", "suspended", and "unchanged".

[2]If there are multiple companies participating in the same news story, only one company is identified as the principal

—— Insert Figure A1 ——

Depending on the context, in each subtype, RavenPack further defines two attributes to refine the structured information on its properties and fact-level. Properties may indicate that the role of the company plays in the news story. For example, in the same piece of earning-estimate news, company JP Morgan plays a role of "rater"(property = "rater") to give the stock recommendation on company Apple(empty property). For the other attribute, fact-level determines whether the news is a fact, an opinion or a forecast.

In the end, RavenPack uses a variable called Category to summarize all the classification information together with the refined attributes of properties and fact-level. At this level, Ravepack further determines whether the timing of the news is predictable and assigns an indicator of "scheduled news" or "unscheduled news". For example, in Figure A1, all the "earnings guidance" news are scheduled news except for the "earning guidance suspended" news, which is unscheduled news. Furthermore, Table A2 (RavenPack, 2018) explains all the categories under the news type "earnings guidance" in details.

3.3.2 Filter the News Data

RavenPack sequentially introduced three mutually exclusive packages in three different year: Dow Jones Package since 2000; Press Releases Package since 2004, and Web Source Package since 2007[3]. As depicted in Table1, the number of firm-months with missing news could account for over 20% before 2004, but this figure declines to around 10% after 2004. Since I cannot determine whether the missing values of a month is due to no news or no data records, I try to avoid the missing record problem by choosing the starting year from 2004. Thus, I choose the sample period

[3]the Dow Jones Packages covers all the news information from Dow Jones Newswires, Wall Street Journals, Barron's and MarketWatch; the Press Releases Package keeps tracks of over 100,000 press releases and regulatory disclosures on a daily basis through different press releases distribution networks; the Web Package monitors more than 19,000 sources of leading publishers and websites.

spanning from 2004 to 2017, which will have full coverage on Dow Jones Package and Press Releases Package, and partial coverage on Web Package.

RavenPack stores all the global equity news data in the UTC time zone. I select the news data of US equity, and convert the UTC time to the New York time. The conversion adds 5 hours to take care of the daylight saving adjustment from the second Sunday of March to the first Sunday of November, and only add 4 hours for the remaining dates without the adjustment. The US exchange trading hours are from 9:30 to 16:00. If the news stories occur after trading hours, I adjust the date to the next trading date.

In order to identify a clear news-return response, I need to further filter the news data with maximum relevance and maximum novelty. Relevance score, ranging from 0 to 100, represents how relevant the individual equity is mentioned in the news story context. Score over 90 indicates the company is in the main title or headline, whereas lower scores mean it is in the body of the story. Score 100 means this company plays a key role in the headline of the story; otherwise not playing a key role, such as rater. For example, a rating company that gives stock recommendations in the headline: the rating company receives score 90, and the rated company receives score 100. To exclude raters, I choose the maximum relevance of score 100.

On the other hand, to measure novelty, RavenPack analytics 1.0 deprecated the old variable ENS or G_ENS, which were often used in prior literature, but introduce a more flexible variable called Event_Similarity_Days. This new variable measures the number of days elapses after having detected a similar event over the last 365 days. Therefore, I pick the news with maximum novelty by choosing Event_Similarity_Days = 365 (365 is the maximum number in the data). In other words, I am choosing the very first news covered and discard all the following repeated news over the year. If the following news updates its sentiment, say from positive to negative, regarding the same event, this news will be counted as novel news again and Event_Similarity_Days will be reset to its default value of 365.

For the news content, I further filter out the news content group of "technical-analysis", "stock-prices", and "order-imbalances" as these news does not trigger shocks to the market but merely reflects the market shocks that have already happened.

3.3.3 The Key Variable: Sentiment Score

In the past literature, the two mostly used sentiment score measures provided by RavenPacks are: Event Sentiment Score (ESS) (Dang, Moshirian, & Zhang, 2015), and Composite Sentiment Score (CSS) (Bushman et al., 2016; Bonsall IV, Green, & Muller III, 2018).

RavenPack provides the Composite Sentiment Score (CSS) ranging from -1 to 1 to indicate how positive (negative) the news story is. The strength of the score is modeled by the intra-day stock price responses trained by using the tick data of 100 large cap stocks.

$$\text{Composite Sentiment Score (CSS)} = \begin{cases} (0, 1], & \text{Positive News.} \\ 0, & \text{Neutral News.} \\ [-1, 0), & \text{Negative News} \end{cases}$$

The composite sentiment model comprises of five different analytics: PEQ analytics identify general positive and negative words and phrases in articles about the equities; BEE analytics focus on news about earnings evaluation; BMQ analytics which specializes in analyzing short commentary and editorial; BAM analytics focus on news stories of mergers, acquisitions and takeovers; BCA analytics expertize in reports about corporate action announcement. RavenPack will further make sure that there is no sentiment disagreement that exists amongst these five analytics. The CSS would be the average of these five analytics scores.

Alternatively, RavenPack offers a similar measure called Event Sentiment Scores

(ESS), which has the same range as CSS. ESS is constructed by financial experts using a different scoring system for over 6700 different categories of contents, whereas CSS is a universal composite based on the five separate textual analysis algorithms mentioned previously. There are pros and cons in choosing the best sentiment score between ESS and CSS:

1. for news with a similar theme, ESS may perform more accurate than CSS because it relies on financial experts and customizes different scoring systems for different topics. For example, in earnings news, experts compare actual figures to estimated figures; in analyst rating news, experts compare the rating changes; in earthquake news, they analyze the Richter scale; and in terrorism attack news, they focus on the number of casualties.

2. for news across different themes, CSS may perform better because it implements a universal scoring system to all kinds of news, and this scoring system has the main focus on five factors that significantly impact stock prices. On the hand, the scoring algorithm for ESS on different topics may not be comparable.

Since my study utilizes all topics of news, I choose CSS as my major measure of news sentiment score but also keep ESS for the robustness analysis.

In the later chapters, I will devise two measures to quantify the news. One is the news intensity(**N**) by counting the news occurrence(s) within a month. The other one is the news volatility(**NV**) which measures the total news impact magnitude, as defined by the sum of squares of the Composite Sentiment Score(s)(**CSS**) within a month.

3.4 Data Merging

OptionMetrics provides external firm identifiers of CUSIPs and Tickers, and RavenPack offers a similar list of firm-identifiers of NCUSIP, TICKERS and International

Securities Identification Number (ISIN).

I merge CRSP and OptionMetrics IvyDB US in the following priority order: CUSIPs and TICKERs. Since the two identifiers are not permanent and may change over time, it is important to keep track of their beginning dates and end dates, and merge the data by both firm and time identifiers together.

I merge RavenPack and CRSP with the priority order of NCUSIPs, TICKERs and ISINs. The first two steps are the same as merging CRSP and OptionMetrics. Regards to the final step using ISINs, I rely on Capital IQ Identifiers database to bridge the firm-identifier gaps: (1) Capital IQ offers a link table between Gvkey and ISIN, (2) CRSP/Compustat Merged database provides a link table of Gvkey and PERMNO, and (3) PERMNO is the primary firm identifier of CRSP.

— Insert Table 1 —

Table 1 describes the data attrition during the process of data merging. First, we have the CRSP data filtered with share code 10 & 11 and three primary exchanges(NYSE, AMEX, and NASDAQ). Over the years, the number of stocks is shrinking from its peak of 7,241 firms in the year 2000 to its bottom of 3,862 in the year 2017. "Life in the public" (2017) explains this phenomenon with two reasons: IPOs are dropping and acquisitions are surging. After merging CRSP and OptionMetrics, the number of firms drops to an average of 2450 per year, which number is quite stable over the years. Due to the constantly declining number of listed firms, the coverage of option data exceeds 60% since 2011 from its original 33% in 2000. When adding the RavenPack, the data attrition is little and the numbers decline to an average of 2406 per year. I further calculate the percentage of firm-months with no news over the total firm-months. The ratio is constantly declining and is halved in 2004 from 20% to around 10% due to the introduction of the Press Releases Data Package by RavenPack. The last column of Table 1 shows the number of firms in my final dataset when I require a firm to have at least two years of data to be included in the data. Eventually, the annual average number of firms is around 2351 in my final dataset which has a period of 14-year coverage from 2004 to 2017.

3.5 Summary Statistics

Table 2 describes the distribution of all the data variables. My data sample contains 7,418,062 news stories and 363,456 firm-months covering the period from 2004 to 2017.

— Insert Table 2 —

Panel A shows the news data summary at the news story level. There are two sentiment scores provided by RavenPack: CSS and ESS. Both scores range from -1 to 1. The CSS is on average neutral at 0.003, while ESS has a significantly positive average of 0.13. At the intra-day news level, the number 0.13 is unreasonably too high when comparing to a much lower positive drift of CSS or the annualized 7% for the US stock market. ESS deploys different scoring rules on different news content categories, while CSS implements a universal consistent scoring rule. Thus, the disparity in mean value implies that the ESS scoring cannot be comparable across different content categories. In addition, ESS records a much larger standard deviation of 0.459 than CSS's 0.105. For the skewness, CSS is greater in left-skewed than ESS (-3.345 vs. -0.379). Since CSS is trained on the stock return, this is consistent with past literature that negative news is more return-relevant with higher price impacts. Therefore, I choose CSS to be my major variable to measure news sentiment for its superior performance across different content of news and in mimicking the return.

Panel B shows the summary statistics of the other variables at the firm-month level. Implied volatility (IV) has an average of 0.466 which is higher than realized volatility (RV)'s average of 0.394. RV and IV have a roughly similar standard deviation of around 0.29. Both volatility measures are skewed to the right because they only have positive values with the upper bound unlimited. Specifically, the RV has a much larger maximum value of 22.108 than IV's 2.944. The average volatility risk premium (VP) as devised by the difference of IV_{t-1} and RV_t is 0.072.

For news intensity(N), there is an average of 20.410 news stories per firm-month. Less than 10% of the firm-month records zero news, while the most intensive news occurrence for a firm-month is 1065. For news volatility(NV), the average is 1.055 which skews to the right with its maximum at 24.652. The news volatility is constructed as the sum of squares of the CSS which summarizes both the intensity and the magnitude information of the news impacts. Therefore, the distribution of NV is similar to N but on a smaller scale because the absolute sentiment scores are by definition, less or equal to one.

When breaking down the news category based on the timing, we have more scheduled news than unscheduled news on average for each firm-month (i.e., 15.475 vs. 4.934). Unscheduled news contributes much less volatilities than scheduled news on average (0.505 vs. 0.768). This can be attributed to the fact that scheduled news contains a large quantity of high impact news like earnings and revenues news.

In terms of categorizing by news format, most news is disseminated as news-flash (9.697 per firm-month) because this is the fastest rebroadcast channel with only headline information and without adding other editorial content. The second highest frequency is through the full-article (6.406 per firm-month) which adds additional editorial content. Press-releases is a voluntary firm-initiated channel and ranks the third (3.836 per firm-month). The least frequent ways of the broadcast are through tabular materials and SEC filings (only 0.053 and 0.458 per firm-month, respectively). About the news volatility, the full article has the highest impact of 0.640 with news flash closely following behind at 0.527. This may be because important news, proxy by high news volatility, is often covered by full-articles. Press-releases impact ranks third at 0.253. The impact of Tabular Material and SEC filings are insignificant at 0.009 and 0.024, respectively, because most of these types of news are neutral from the data.

— Insert Table 3 —

Table 3 runs correlation tests among all our interested variables including implied volatilities with one-month lag, denoted as IV_{t-1}. The lower(upper) triangle shows the average Pearson(Spearman) correlation coefficient constructed as follows: first, I run a correlation test for each firm independently; then, show the average of the correlation coefficients across all firms. The main regression tests in the later sections will also show the average of coefficients in the same way.

Since the results of these two kinds of correlations are similar, I only focus on Pearson's correlations. The first column describes the relationship between realized volatility with different variables. RV_t is highly correlated with IV_{t-1} at around 47%(see $\boxed{1}$3), which is higher than with RV_{t-1} at around 37%(see $\boxed{1}$2). But when it comes to news volatility, it drops to 22%(see $\boxed{1}$14). This implies that the market may contain some liquidity traders making non-information driven trading. Additionally, when it comes to news intensity, the correlation drops to 18%(see $\boxed{1}$6), which is lower than 22%. This implies that news intensity is weaker in RV-relevance than news volatility. Furthermore, RV_t is significantly negative correlated with VP_{t-1} at -0.13(see $\boxed{1}$5), which indicates that VP_{t-1} also has forecasting power on RV_t.

When comparing news volatilities with concurrent implied volatilities, there is almost no correlation (-2%)(see $\boxed{3}$14) between IV_t and NV_t. In the opposite, there is a relatively strong correlation (10%)(see $\boxed{4}$14) between IV_{t-1} and NV_t. It implies that implied volatilities are forward-looking so as to predict the future news. There is a similar pattern for news intensity. I further look into the correlation between news intensity and news volatility which has a moderate correlation of 70%(see $\boxed{6}$14), which indicates the two measures are similar in theme but measures news activities in different prospects.

In the end, we look at the break down of different kinds of news: scheduled news is the most highly correlated with news-flash which is the most realized volatility relevant, and unscheduled news is the most highly correlated with full articles. The news volatility of scheduled news has an average of only 30%(see 15 16) correlation with the one of unscheduled news. For the two most realized volatility relevant news-type, the average correlation of news volatility between news-flash and full-article is 41%(see 17 19).

Chapter 4

News Channel: IV Forecasting Future RV

The objective of this chapter is to prove that the implied volatility(IV) can predict future realized volatility(RV) through the channel of predicting future news. We decompose the logical process of our analysis into three parts:

(1) **RV Relevance of News**: Realized volatility (RV) is highly relevant to the concurrent news information

(2) **IV Predictability of News**: Implied volatility (IV) can predict future news information

(3) **The News Channel**: Thus, implied volatility can predict future volatility through predicting future news.

To measure the news impacts, I devise two variables with different emphasis: the first measure is news intensity (denoted by N) – how many news stories occur within a month – to measure the intensity solely regardless of the magnitude of the news impacts; the second measure is news volatilities (denoted by NV) – the sum of square of CSS within a month, annualized by multiplying 12, and finally take the square root. The NV measure captures the total magnitude of the news impacts.

The sections of this chapter are structured as follows: first, I validate the existence of prior results of IV predicting future RV; second, I look into whether realized volatility is relevant to news information; then, I examines whether implied volatility can forecast future news information; in the end, I combine the previous two parts to examine the prediction through news channel.

4.1 Results in Prior Literature

I first replicate Christensen and Prabhala (1998)'s main result in individual stock level with a new time span from 2004 to 2017, where RavenPack has a reliable coverage. They examine the information content of implied volatilities in predicting future volatility of the stock index as follows:

$$\mathrm{RV}_t = \alpha + \beta_1 \mathrm{RV}_{t-1} + \beta_2 \mathrm{IV}_{t-1} + \beta_3 \mathrm{VP}_{t-1} + \varepsilon_t \tag{4.1}$$

I perform a similar test on all the individual stocks and present the cross-sectional distribution with at 1% significance level of the test results as in Table 4. Panel A shows the whole sample of 365,456 firm-months where both implied volatilities and past volatilities can independently predict future volatility in a single variable regression. But the implied volatility has a higher coefficient (0.661 vs. 0.371), higher R^2 (28.5% vs. 20.7%), a larger number of stocks with significantly positive predictor (68% vs. 56%) than past volatility. This test, as well as the remaining parts, use the significance level of 1% with the 12-lag Newey-West adjusted standard errors. When putting the predictors RV_{t-1} and IV_{t-1} together, the coefficient of past volatility drops from 0.371 to 0.130. For 35% of the individual stocks, the predictability of past volatilities is subsumed by the information content of implied volatilities. If conditional on the original 56% of stocks having the past volatility's coefficient significant positive, then 62.5% of these stocks lose significance in using RV_{t-1} as a predictor. On the other hand, the majority of stocks (61%) have their implied volatilities IV_{t-1} remaining positively significant. Including past volatility risk premium (VP_{t-1}) can further subsume the forecasting power of the RV_{t-1}.

Since the past papers tend to focus on stock index(Christensen & Prabhala, 1998) or highly traded individual stocks(Taylor et al., 2010), I further filter out the stocks with its average options trading volume (over the whole life of the stock) below median in Panel B. The sample shrinks from 363,456 firm-month to 207,169 firm-month, more than half of the original sample size because stocks with highly traded options are big companies and lasts for longer years. The overall results consistently become stronger. When using a single predictor, past volatility is positively significant for 61% of the individual stocks, whereas implied volatility is positively significant for 84%. But when putting the implied volatility as the additional predictor, only 12% of these stocks have the past volatility remaining positively significant. The coefficient of past volatility drops from 0.408 to 0.075, but 78% of the stocks' implied volatilities remain positive significant at $p \leq 0.01$ level. However, the cross-sectional coefficient average of past volatility remains positively significant with t-statistics equal to 17.48. This figure is much lower than the implied volatility's t-statistics of 89.95. If we include the predictor of past volatility risk premium (VP_{t-1}), the forecasting power of the past volatility is completely subsumed.

In Panel C, I further restrict the data to keep the top 25% stocks with the most traded options. The number of firm-months shrinks to 112,261. In univariate regressions, both past volatility and implied volatility can predict future realized volatility for 67% and 90% of the stocks, respectively. But when putting together, the t-statistic of the average coefficient of past volatility drops to 9.00, while implied volatility still has a t-statistic of 75.85. Almost 85% of the stocks have their implied volatility to predict the future realized volatility with positive significant forecast power. Putting VP_{t-1} reverses the forecasting direction of RV_{t-1} into a negative.

— Insert Table 4 —

To sum up, implied volatility shows superior forecast power over the past volatility in predicting future volatility. For around 80% of the individual stocks, the forecast power of past volatility is subsumed by implied volatility. The results become stronger when I implement a more restrict requirement on the activeness of option trading or involving the VP_{t-1} as an additional predictor. The overall results validate the finding of prior literature and I extend the results to all the individual stocks' level.

4.2 RV Relevance of News

As long as news contains new information that surprises the stock market, it will generate volatility. In this section, I examine whether realized volatility is relevant to news information and further assess the strength of such relevance.

Table 5 examines how strong is the news relevance of the concurrent realized volatility. I measure the news impact in two ways: news intensity (N) which measures the intensity of the news occurrence and news volatility (NV) which measures the total magnitude of the news impacts. First, I regress realized volatility on these two news measures respectively for each firm. Then, I show the average and the distribution of the coefficients. In addition, I control the previous month's realized volatility to see whether the news information is a mere reflection of past volatility or contain new information for the concurrent volatility. The regression equations are as follows:

$$\mathrm{RV}_t = a + \beta_1 X_t + \epsilon_t \tag{4.2}$$

$$\mathrm{RV}_t = a + \beta_1 X_t + \beta_2 \mathrm{RV}_{t-1} + \epsilon_t \tag{4.3}$$

where X_t is substituted to be N_t (news intensity) or NV_t (news volatility).

For news intensity, 88% of the stocks have a positive relation with concurrent volatility. The coefficient average has a t-value of 35.58. When controlling the past volatility, this relation becomes more crystal: 94% of the stocks have a positive relation with the average's t-statistic at 42.13.

The results for news volatility are in general stronger than the one for news intensity in terms of higher R^2 (7.3% vs. 6.0%) and higher t-statistics for the coefficient average (56.95 vs. 35.58). The percentage of stocks with positive relation is about the same as the percentage in news intensity around 90%. Adding control of the past volatility will strengthen the result a little bit: another 9% of the total stocks become positively significant in news-volatility relevant. Intuitively, the stronger results of NV is understandable because NV is more accurate by adding consideration of the magnitude of the news impact.

— Insert Table 5 —

In summary, the news of about 90% of the stocks is positively relevant to concurrent realized volatility. Over one third of them depict significant positive relevance at 1% significant level. The results become even stronger if one adds the control of past realized volatility.

4.3 IV Predictability of News

Implied volatility is by construction the market forecast of the future volatility for the remaining life of the option. In this section, I will examine whether implied volatility can forecast future news.

Table 6 uses future realized news to test the implied volatility's forecasting power. I regress the future news intensity/volatility (N & NV) on implied volatility with adjustment to align the options' remaining life to be the same as the forecasting period. Furthermore, I control for the past news intensity/volatility (N & NV) because the market may get the inference from the past news.

$$X_t = a + \beta_1 IV_{t-1} + \epsilon_t \tag{4.4}$$

$$X_t = a + \beta_2 X_{t-1} + \epsilon_t \tag{4.5}$$

$$X_t = a + \beta_1 IV_{t-1} + \beta_2 X_{t-1} + \epsilon_t \tag{4.6}$$

$$X_t = a + \beta_1 IV_{t-1} + \beta_2 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \epsilon_t \tag{4.7}$$

where X can be substituted to be N_t (news intensity) or NV_t (news volatility).

— Insert Table 6 —

For news intensity N_t, the forecast power is positively significant because the t-stat of the coefficient average in the univariate regression is 25.46. In terms of distribution, 69% of the stocks have positive forecast power. When controlling the past news intensity, the result remains the same with t-statistics of 18.96, and 61% of the stocks have positive forecast power. The control variable N_{t-1} is significantly negative because there is more scheduled news than unscheduled news. For individual firms, their scheduled news concentrate around the quarterly earnings announcement period because a large quantity of scheduled news is earnings-related. When controlling for three lags of the news intensity, it is clear to see N_{t-3} flips the sign to be significantly positive and the result of implied volatility forecast power remains there with t-statistics equal to 14.21.

On the other hand, the results in news volatility are stronger than in news intensity. The coefficient of β_1 is 0.921 on average with a T statistics of 33.21, much higher than the one for the news intensity. 74% of the stocks' implied volatility can positively predict the news volatility. Controlling the past news volatility NV_{t-1} will not alter the results and t-statistics remains 31.60 and 72% of the stocks' implied volatility has positive forecast power. When controlling three lags, the t-statistic of β_1 drops to 26.92, and the ratio of the stocks keeping their positive forecast power of implied volatility declines to 70%, but r-square increases to 21.3%.

In summary, implied volatility can predict the intensity of the future occurrence of news as well as the magnitude of these future news' impact. But in general, news volatility has greater IV-predictability than news intensity. The forecasting power remains even after controlling the past news. This indicates that the option market can incorporate future news into the price which does not rely on past news information.

4.4 The News Effect

In the last section of this chapter, I put news intensity/volatility into Christensen and Prabhala (1998)'s equation (4.1) to examine the effect of news channel:

$$\mathrm{RV}_t = \alpha + \beta_1 \mathrm{RV}_{t-1} + \beta_2 \mathrm{IV}_{t-1} + \beta_3 \mathrm{VP}_{t-1} + \gamma X_t + \varepsilon_t \tag{4.8}$$

where X_t can be news intensity N_t or news volatility NV_t.

Table 7 presents the results which can be parallel compared with Table 4. Recall that in the original equation (4.1) of Table 7, the average of the coefficient β_2 of IV_{t-1} is 0.613 with t-stat of 81.88.

When I include the concurrent news intensity(N_t) into the regression, the forecast power of IV_{t-1} becomes weaker. N_t attenuates the β_2 from 0.613 to 0.531 with T-stat from 81.88 to 71.82. On the other hand, the forecast power of past volatility RV_{t-1} is strengthened from 0.077 to 0.164 in coefficient, from 13.47 to 28.96 in t-statistics. This indicates that the market tends to set aside some persistent uncertainty that is unrelated to news information.

— Insert Table 7 —

Next, I consider the magnitude of the news impact by using news volatility(NV_t). It generates similar results. Even though NV_t apparently shows stronger relevance to RV_t, the attenuation effect of β_2 is slightly weaker than using N_t by 0.011 difference in coefficient averages, and weaker by 1.49 difference in t-statistics.

Panel B and C of Table 7 restrict the sample with more active traded options. Panel B filters out stocks with option trading volume below the median. When adding N_t or NV_t to equation (4.1), the two results are similar: the coefficient average of implied volatility attenuates from 0.817 to about 0.725, and t-statistics attenuate from 84.31 to around 72.21. But when it comes to Panel C where I only keep the stocks with top 25% of the highest average option trading volume, the results using N and NV are similar: the coefficient average of implied volatility attenuates from 0.911 to about 0.823, and t-statistics attenuate from 71.57 to around 58.5.

Together with the evidence in the previous sections, all these results suggest that implied volatility can predict future realized volatility through the channels of predicting both news intensity and news volatility. When comparing these two channels, news intensity is consistently as strong as news volatility, even though it is relatively weaker both in volatility relevance and implied volatility predictability. On one side, news intensity may be easier to forecast regardless of the impact magnitude; on the flip side, even though impact magnitude is difficult to forecast, the news impact magnitude would be more relevant to realized volatility.

Furthermore, I implement a two-stage mediation analysis devised by Preacher and Hayes (2008) to quantify the strength of both news intensity and news volatility channel in Table 8. For each company i, the mediation model can involve a set of following three regression equations, but the first one is optional for checking purpose:

$$\begin{cases} \text{Total Effect:} & \text{RV}_{i,t} = \text{Int}_i + \beta_i \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1:} & \text{X}_{i,t} = \text{Int}_i + \theta_i \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 2:} & \text{RV}_{i,t} = \text{Int}_i + \beta_i' \text{IV}_{i,t-1} + \phi_i \text{X}_{i,t} + \text{Controls}_{i,t} + \varepsilon_{i,t} \end{cases}$$

where $X_{i,t}$ can be substituted as $N_{i,t}$ or $NV_{i,t}$.

— Insert Table 8 —

In this two-stage mediation analysis, the mediator $X_{i,t}$ is my interested news variables in the current month. For firm i, the total predictability of $IV_{i,t-1}$ forecasting $RV_{i,t}$ is represented by β_i in the first equation. In the first stage, I use the past implied volatility ($IV_{i,t-1}$) to predict the mediator. In the second stage, I conduct the regression test of IV_{t-1} predicting RV_t and controlling for the mediator. The non-news channel predictability for the firm i is represented by β_i' in stage two, while the news channel predictability is measured through the mediator as the product of θ_i in stage one and the corresponding mediator coefficient ϕ_i in stage two. The first equation is optional because there exists total equality that the total predictability is the sum of the non-news channel predictability and the news channel predictability: $\beta_i = \beta_i' + \theta_i\phi_i$. Eventually, we can quantify the strength of the news channel predictability as a proportion over the total predictability.

To further ensure that the decomposed components from the total predictability are non-negative, I first filter out stocks whose total predictability is negative, which is about 6% of the whole sample (see Table 4). The options implied volatility of these dropped stocks does not have any forecasting power on realized volatility, which is out of the scope of my study. Second, for each stock, I run an iteration of the two-stage mediation tests to determine whether to include the mediator. If $\theta_i \leq 0$ or $\phi_i \leq 0$, then I set the news channel predictability to be zero. In other words, the non-news predictability strength is 100%.

Table 8 illustrates cross-sectional summary statistics of the strength of the news channel predictability across firms. The average strength is about 12% of the total predictability. In specific, the average strength of the news intensity channel is slightly higher than the one of the news volatility channel, which is consistent with the results in Table 7. The distribution is right-skewed. It means that some firms' options are more heavily loaded with forward-looking information, where the maximum suggests about half of total predictability could come from the news channel.

Nevertheless, on the one hand, even though the forecast power of implied volatility attenuates in equation (4.8), such predictability of IV on RV remains strong; consistently, on the other hand, the news channel strength on average only explains about 12% of the total predictability. I posit that this may be because the whole news sample contains the noise of volatility irrelevant news. That is why in the next chapter, I categorize news based on different classification, devise multiple mediators for different types of news, and let the model choose the important news. In the end, I will also try to pick the most volatility relevant news to form a new sample of news to study.

Chapter 5

Distinguish Different Kinds of News

In order to find out the most volatility relevant news to explain the news channels, I further categorize the news based on three different classification method:

(1) **Classification on Timing Predictability**: scheduled news (Sch) and unscheduled news (UnSch)

(2) **Classification on News Format**: news flash (NF), full article (FA), press releases (PR), tabular materials (TM), and SEC filings (SEC)

(3) **Classification on News Content**: earnings news, M&A news, and insider trading news etc.

In the first three sections of this chapter, I will study these three classification one by one, look into the strength of both RV-Relevance of news and IV-Predictability of news, and lastly quantify the strength of the news channel of different types of news. In the last section, I pick out the best RV-relevant news and form a new sample of news to study.

5.1 Classification on Timing Predictability

Based on the timing predictability, one can classify news into two groups. One is scheduled(Sch) news whose timing is known beforehand. For example, earnings announcement news and dividends announcement news. The other is unscheduled news whose exact time of occurrence is unknown. For example, analyst rating news and insider trading news. RavenPack use a taxonomy like Figure A1 to categorize the news content. In the category level of the taxonomy, RavenPack would assign Scheduled or Unscheduled to the specific category of news. In the example of Figure A1, all the earnings-guidance news are scheduled except that earning-guidance-suspended news is unscheduled.

5.1.1 Summary Statistics in This Classification

In Table 9, we can have an overview of news frequency and news sentiment grouped by timing predictability classification. First, the number of scheduled news stories (5,624,626) is roughly three times larger than the number of unscheduled news (1,793,436). This is because over 68% of news content in RavenPack is earnings news and revenue news which are by and large scheduled news. However, over the total number of 363,456 firm months, the firm-month with unscheduled news (84.29%) is more than the firm-month with scheduled news (69.25%). The reason is apparent that most scheduled news is earnings-related and concentrate in earnings announcement month.

——Insert Table 9——

Regards to news sentiment, even though CSS and ESS give us different results, it is clear that scheduled news is, in general, better than unscheduled news. Since CSS is a better measurement when comparing across different news content, we can see that on average, unscheduled news tends to be bad news while scheduled news tends to be good news. Lastly, if we look at the News Impact Projection (NIP) score provided by RavenPack, scheduled news tends to generate higher news impact than unscheduled news in terms of projection of relative volatility increase in the next two hours.

5.1.2 RV Relevance of News Test

Table 10 first compares the volatility relevance between scheduled and unscheduled news. In general, the results show that unscheduled news (UnSch) has greater volatility relevance than scheduled news (Sch).

— Insert Table 10 —

Specifically, in univariate regressions of using news intensity(N), scheduled news (N.Sch_t) is weaker in volatility relevance than unscheduled news (N.UnSch_t) in terms of R^2 (5.0% vs. 8.2%), coefficient average (0.002 vs. 0.015) and t-statistics (33.14 vs. 40.28). When putting these two variables together to explain RV_t, R^2 increases to 10.9% and scheduled news (N.Sch_t) still remains weaker than unscheduled news (N.UnSch_t) in coefficient average (0.001 vs. 0.012), t-statistics (15.88 vs. 32.41), and ratio of stocks with positive significant coefficient (14% vs. 15%). These results indicate that the occurrence of unscheduled news can create more shocks to the market because the timing is unpredictable.

Adding consideration of the news impact magnitude to test volatility relevance, the result of scheduled news volatility (NV.Sch_t) is similar to the one of unscheduled news volatility (NV.UnSch_t). In the simple regressions, even though NV.Sch_t is lower than NV.UnSch_t in terms of the coefficient average (0.038 vs. 0.098) and R^2 (5.6% vs. 7.5%), yet the t-statistics is higher (46.02 vs. 39.53) and slightly more stocks with positive significant coefficient (28% vs. 25%). When regressing on both NV.Sch_t and NV.UnSch_t together, the R^2 rises to 11.1%. The cross-sectional distribution of these two coefficients for NV.Sch_t and NV.UnSch_t are similar: 18% vs. 16% is positive significance and 61% vs. 64% is positive insignificant. Although

NV.UnSch$_t$ is slightly higher in coefficient (0.082 vs. 0.026), their t-statistics is roughly the same (32.49 vs. 32.42). Comparing to results of news intensity, adding the consideration of magnitude increases the relative importance of scheduled news to be volatility relevant. Intuitively, scheduled news like earnings information is important in value relevance. This further implies that the magnitude of scheduled news may be less predictable, and it creates shocks to the market even though its timing is predictable.

In summary, the overall empirical evidence suggests that both scheduled news and unscheduled news are realized volatility relevant. But in terms of news intensity, the relation is stronger for unscheduled news than scheduled news. When considering magnitude by using news volatility, the scheduled news is roughly the same as unscheduled news in linking to realized volatility.

5.1.3 IV Predictability of News Test

Next, I compare the implied volatilities' forecast power on these two kinds of news in Table 11. As we can see in the table, the first salient difference between scheduled and unscheduled news is the sign of their autocorrelation on news intensity: scheduled news' autoregressive model(1)'s coefficient – denoted by AR(1) – is strongly negative, but unscheduled news' AR(1) is positive. For each individual firm, monthly aggregated scheduled news tends to intersperse unevenly within a year. For example, some scheduled earnings-related news concentrates in the cycle of the quarterly announcement periods. Intuitively, unscheduled news, in contrast, is prone to be evenly distributed within a year, for its occurrence is supposed to be unpredictable. However, when I increase the lags from AR(1) to AR(3), the coefficients of the lag 3 in both scheduled news and unscheduled news become positive. This implies the earnings announcement cycle could also influence the density of unscheduled news, but not as substantial as on scheduled news.

— Insert Table 11 —

Panel A tests the implied volatility predictability of future news intensity. Firstly, it is straightforward to presume that implied volatility should predict the scheduled news intensity (N.Sch). My results support this presumption: the average coefficient is 16.920 with t-statistics of 27.29. When I control the scheduled news intensity of the previous month (N.Sch_{t-1}), the IV_{t-1} coefficient drops to 11.337 with t-statistics of 20.43. If increasing the controls to three lags (N.Sch_{t-1}, N.Sch_{t-2} and N.Sch_{t-3}), the IV_{t-1} coefficient attenuates to 5.192 but remains positively significant with t-statistics of 12.64. Based on the change of t-statistics across model (1) to (3), we can see that most of the omitted variable bias comes from N.Sch_{t-3} because the quarterly cycle of earnings announcement creates a potent positive link between N.Sch_t and N.Sch_{t-3}.

Secondly, it is striking to see that implied volatility can predict the future intensity of unscheduled news (N.UnSch) in the empirical results: in the single variable regression, the coefficient average of IV_{t-1} is 1.005 with t-statistics of 6.58 indicating a significantly positive difference from zero. This result is strengthened if I control the lags of unscheduled news intensity. Controlling AR(1) will increase the coefficient average from 1.005 to 1.352, and t-statistics from 6.58 to 9.99. Controlling AR(3) will not alter the results a lot, but to the coefficient average of 1.608 with t-stat of 12.05. Comparing model (4) to (6), the most crucial control variable is N.UnSch_{t-1} to avoid omitted variable bias. The reason is as follows: the current unscheduled news releases will decrease the future uncertainty and also reduce the current implied volatility which is also forward-looking.

Panel B shows the test on implied volatility forecasting future news volatility. The results of IV_{t-1}'s forecasting power for news volatility is stronger than the previous one for news intensity.

For scheduled news volatility (NV.Sch_t), its AR(3) control variables show similar patterns as in their news intensity counterparts: NV.Sch_{t-1} and NV.Sch_{t-2} are significantly negative while NV.Sch_{t-3} is significantly positive. Most of the forecast power comes from controlling these three lags of previously scheduled news volatility which increase R^2 from 3.1% in the univariate regression to 26.3%. After controlling the AR(3), the IV_{t-1} remains positively significant to predict the future scheduled news volatility $NV.Sch_t$ with the coefficient average of 0.596 and the t-statistics of 25.19.

For unscheduled news volatility ($NV.UnSch_t$), its AR(1) or AR(3) control variables have little influences on the results of the univariate regression: the coefficient average of IV_{t-1} is stable at the level of around 0.228 with t-statistics of around 16.5. We can refer back to Table 3, $NV.UnSch_t$ has a very weak negative correlation with IV_t which is not as strong as the negative correlation between $N.UnSch_t$ and IV_t. This implies that the aggregated impact magnitude of unscheduled shocks can be independent with future uncertainty resolution. In other words, the realization of an aggregated heavy unscheduled shock would not lower much the probability of another aggregated heavy unscheduled shock in the next month; thus, there is little uncertainty resolution and no clear sign implication for the current implied volatility here. In contrast, recall the previous result in this subsection that the realization of the number of unscheduled news can lower future uncertainty.

In summary, the overall results suggest that the option investors have the ability to predict both scheduled news and unscheduled news in terms of its occurrences and impact magnitudes. Apparently, the forecast power on the scheduled news is much stronger than on the unscheduled news. Prediction on news volatility is in general more effective than a mere prediction on the news intensity.

5.1.4 The News Effect

In the last of this section, Table 12 examines the impact of putting the measures of scheduled news or unscheduled news into original RV-IV forecasting equation (4.1). The format is similar to Table 7.

— Insert Table 12 —

Both scheduled and unscheduled news variables can mitigate the forecast power of IV_{t-1} on future RV_t, but the scheduled news exert a stronger influence: N.Sch_t attenuates the average coefficient of IV_{t-1} from 0.613 to 0.543 with t-statistics decreasing from 81.88 to 72.69, while N.UnSch_t attenuates IV_{t-1}'s average coefficient to 0.561 and t-statistics to 77.73. The results for the counterparts of news volatility is similar. This weaker result of unscheduled news can be attributed to the following reasons:

(1) scheduled news is volatility relevant at a similar level as the unscheduled news;

(2) future scheduled news is much easier to predict by the implied volatility.

On the other hand, the coefficient averages of the past realized volatility are strengthened in every model of Table 12. There is more increase in using scheduled news than unscheduled news. The reason behind is that besides the strong volatility relevance of news, the relation between current news appearance and the past volatility tends to be negative because the news are dispersed around the year. The scheduled news is more dispersed than unscheduled news. Table 3 shows the correlation evidences: RV_{t-1} has an average correlation of -0.11 with N.Sch_t (see $\boxed{2}$ 7), -0.06 with N.UnSch_t (see $\boxed{2}$ 8), -0.07 with NV.Sch_t (see $\boxed{2}$ 15), and -0.02 with NV.UnSch_t (see $\boxed{2}$ 16).

Lastly, I implement a similar two-stage mediation analysis as in the section 4.4 to quantify the strength of both scheduled news and unscheduled news channel in Table 13. For each company i, the mediation model can involve multiple mediators with the following regression equations:

$$\begin{cases} \text{Total Effect:} & \text{RV}_{i,t} = \text{Int}_i + \beta_i \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.Sch):} & \text{X.Sch}_{i,t} = \text{Int}_i + \theta_{1,i} \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.UnSch):} & \text{X.UnSch}_{i,t} = \text{Int}_i + \theta_{2,i} \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 2:} & \text{RV}_{i,t} = \text{Int}_i + \beta_i' \text{IV}_{i,t-1} + \phi_{1,i} \text{X.Sch}_{i,t} + \phi_{2,i} \text{X.UnSch}_{i,t} + \text{Controls}_{i,t} + \varepsilon_{i,t} \end{cases}$$

where X can be substituted as N or NV.

— Insert Table 13 —

The model contains two mediator: scheduled news ($\text{X.Sch}_{i,t}$) and unscheduled news ($\text{X.UnSch}_{i,t}$) at the current month. For firm i, the total predictability of $\text{IV}_{i,t-1}$ forecasting $\text{RV}_{i,t}$ is represented by β_i in the first equation. In the first stage, I use the past implied volatility ($\text{IV}_{i,t-1}$) to predict each mediator separately. In the second stage, I conduct the regression test of IV_{t-1} predicting RV_t and controlling for all the mediators. Therefore, the non-news channel predictability for the firm i is represented by β_i' in stage two. The news channel predictability of each mediator is calculated by the product of the coefficients relevant to the corresponding mediator in stage one and two. In specific, we can acquire the scheduled-news-channel predictability as $\theta_{1,i}\phi_{1,i}$, and unscheduled-news-channel predictability as $\theta_{2,i}\phi_{2,i}$. The first equation is optional because there exists total equality that the total predictability is the sum of the non-news channel predictability and the news channel predictability of different types of news: $\beta_i = \beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i}$. Eventually, we can quantify the strength of the news channel predictability as a proportion over the total predictability.

To ensure the decomposed components of the total predictability to be non-negative, I implement a similar iteration process as in section 4.4: first filter out stocks whose total predictability is negative. Second, for each stock, I iterate the same two-stage mediation tests throughout all the possible combination of the choices whether to include a specific mediator and the $\text{IV}_{i,t-1}$ in stage 2. The combinations with non-positive θ or ϕ are dropped. Furthermore, I also drop the combinations with negative β_i' in stage 2, because it will generate negative non-news channel predictability. In the end, within all the valid combinations, I choose the best one with the highest R^2 in stage 2. If $\text{IV}_{i,t-1}$ in stage 2 is dropped, then β_i' is winsorized to be zero, which means that the non-news-channel predictability is zero. If any mediators are dropped, the predictability through this type of news channel is winsorized to be zero. Therefore, the strength of the scheduled-news and unscheduled-news channel can be quantified as $\theta_{1,i}\phi_{1,i}/(\beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i})$ and

$\theta_{2,i}\phi_{2,i}/(\beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i})$ respectively. β_i', $\theta_{1,i}\phi_{1,i}$, or $\theta_{2,i}\phi_{2,i}$ would get to zero at the time, because I request β_i to positive in the first place.

Table 13 illustrates the cross-sectional summary statistics of the strength of the scheduled and unscheduled news channel predictability across firms. The average strength of the overall news channel increase to about 18% (from 12% in section 4.4). The substantial improvement suggests that the classification of news can help assign different level of RV-relevance and IV-predictability to different types of news. The average strength of the news intensity channel is still slightly higher than the one of the news volatility channel. In the breakdown of scheduled and unscheduled news, the strength of scheduled news intensity is slightly higher than the unscheduled news intensity (10.1% vs. 9.2%), while the strength of scheduled news volatility is lower than the strength of unscheduled news volatility (7.5% vs. 9.7%). This result suggests that even though the investors are poor in predicting the occurrence of unscheduled news on average, they are able to predict the unscheduled news that is important, as measured by its news volatility. As a result, the high RV-relevance of this unscheduled news was not totally offset its not-so-poor IV-predictability.

5.2 Classification on News Formats

Based on news format, RavenPack classifies the news into five types:

(1) News Flash (NF): a headline without body text.

(2) Full Article (FA): a headline with one or more paragraph in the body text

(3) Press Releases (PR): firm initiates a corporate announcement

(4) Tabular Materials (TM): tabular data

(5) SEC Filings (SEC): SEC filings including 10K, 10Q, 13D, 13F, 144 and 8K.

The first three news formats set up important empirical distinctions in two ways. Firstly, past literature (Drake et al., 2014; Twedt, 2015) relies on this classification to distinguish the two information role of media: information creation and information dissemination. News Flash (NF) plays a pure role of information dissemination because such dissemination is most timely with no time to add the body text, while Full Article (FA) adds additional editorial contents and involves in the role of information creation. Secondly, it provides a way to distinguish the media initiated news and firm-initiated news. Press Releases (PR) is apparently firm-initiated, while the other news formats are media-initiated. The firm initiated news may involve a manager's voluntary disclosure and strategic plan of information dissemination, while there are little incentives for media to do the same.

The remaining two news formats are less important because they have relatively less impact or noisy impact on the stock price volatility.

5.2.1 Summary Statistics in This Classification

Table 14 presents the summary statistics based on the news format classification. Among the different news format, news flash(NF) is the most popular with 3,524,328 news stories and accounts for almost half of the sample, for this is the quickest and handiest way to disseminate information without extra efforts to write the body text. Full article (FA) ranks the second with 2,328,327(31.39%), followed by press-release (PR) with 1,394,111 (18.79%). The remaining two formats, tabular material(TM) and SEC filings(SEC), only comprise of about 2% of the total.

——Insert Table 14——

In terms of firm-month coverage, full-article has the highest coverage: 73.41% of the firm-months has full-article news, which is 6% higher than press-release and 8% higher than news-flash. But SEC filings also covers 34.14% of the firm-month because most of the filings are mandatory by regulation. If we focus on the composite sentiment score (CSS), which is a clearer measure across different topics, interestingly, news-flash on average tends to be bad news at the score -0.10 with t-statistics of -193.76, whilst press-releases on average is prone to be good news at score 0.038 with t-statistics of 750.94. Lastly, regards to news impact projection (NIP) provided by RavenPack, only news flash has a positive projection with score 0.012 and positive significant from 0 with t-statistics of 161.04. For the other formats, their projections are undetermined.

To sum up, the news flash is the most popular way to disperse information. It also tends to have the highest impact projection on price volatility in the next two hours. News-flash is on average bad news, while press-release is on average good news.

5.2.2 RV Relevance and IV Predictability of News Tests

Table 15 compares both the realized volatility relevance and the implied volatility predictability on news based on different news formats.

For realized volatility, all news formats are relevant except for the SEC filings whose coefficient average is negative (-0.008 with t = -6.46). Ranking the relevance by different criteria, we can get

- ranked by R^2: N.FA (5.8%) > N.NF (5.7%) > N.PR (4.9%) > N.TM (1.4%), and NV.NF (6.5%) > NV.FA (6.3%) > NV.PR (4.1%) > NV.TM (1.1%)

- ranked by t-statistics: N.NF (32.81) > N.PR (29.64) > N.FA(28.83) > N.TM (15.96), and NV.FA(34.18) > NV.NF(31.43) > NV.PR (31.15) > NV.TM (12.29)

- ranked by the percentage of stocks with positive coefficent: N.NF(91%) > N.PR(83%) > N.FA(72%) > N.TM(62%), and NV.NF(90%) > NV.FA (86%) > NV.FA (80%) > NV.TM (59%).

where I exclude SEC because its relation is negative. To summarize about volatility relevance, the ranking of news intensity is roughly the same as news volatility.

Specifically, news flash (NF) ranks the highest, closely followed by full-article (FA), then by press-releases (PR), and the last with a huge distance is the tabular material.

— Insert Table 15 —

For implied volatility, all news formats are predictable except for the news volatility of the SEC filings whose coefficient average is insignificant. Ranking the predictability by different criteria, we can get:

- ranked by R^2: N.FA (3.3%) > N.FA (3.2%) > N.PR (2.7%) > N.TM (1.9%), and NV.NF (3.3%) > NV.FA (2.9%) > NV.PR (2.7%) > NV.TM (1.7%)

- ranked by t-statistics: N.NF (34.19) > N.PR (24.49) > N.TM(9.18) > N.FA (2.18), and NV.NF(34.21) > NV.FA(21.84) > NV.PR (20.03) > NV.TM (7.56)

- ranked by the percentage of stocks with positive coefficent: N.NF(78%) > N.PR(66%) > N.TM(56%) > N.FA(50%), and NV.NF(86%) > NV.FA (63%) > NV.PR (62%) > NV.TM (53%).

where I exclude the SEC filings news due to its relatively low in significance. In summary, most of the predictability ranking agrees with previous volatility relevance ranking except for the full article. There is a huge gap between NF and FA in terms of IV predictability.

Taken together, the results suggest that it is relatively difficult to use implied volatility to predict the occurrence or impact magnitudes of the future full articles even though full-article is high in RV-relevance. Recall that full article can proxy the information creation role of media. This implies that the in-depth analytical skills of the reporters are different from the expertise of the option traders. The reporters add value to the market by creating new in-depth analytical information.

To conclude, across different news formats, implied volatility performs the best in forecasting news flash which is also the most realized volatility relevant format. This is because the news-flash is the most timely way to exclusively disseminate information without creating new information, such as information interpretation. It is relatively difficult for implied volatility to predict the future full article, a potential act of information creation. Even though press-release is firm-initiated, they are important in terms of RV-relevant and predictable in terms of IV-predictability.

5.2.3 The News Effect

In the last of this section, Table 16 examines the impact of putting the measures of different news formats into the original RV-IV forecasting equation (4.1), whose format is similar to Table 7. First, we should exclude the results of the SEC filings format for its relation with realized volatility is negative.

——Insert Table 16 ——

For the other news formats, all of them can attenuate the forecast power of IV_{t-1} on future RV_t. Among them, news flash (NF) has the most influence. The intensity of news flash decreases the IV_{t-1}'s coefficient average from 0.613 to 0.534, with t-statistics from 81.88 to 72.01. The results are similar for the volatility of news flash.

——Insert Table 17 ——

Similar to section 5.1.4, I implement the two-stage mediation analysis to quantify the strength of the news channel of different news formats. Table 17 summarizes the cross-section distribution of the strength across firms. The overall average further increases to about 22%. News intensity and news volatility have the same strength ranking of different news formats: news-flash (8.5%) > full-article (6.5%) > press-releases (5.5%), and the remaining two formats (about 1%) lag behind with a large gap. For each news format, the strength of NI is higher than NV except for the full-article format. This suggests that full-article is better in delivering news with high impact, as measured by the news volatility, with more in-depth analysis.

5.3 Classification on News Content Groups

This type of classification is based on news content groups. In the empirical setup, I utilize RavenPack Event Taxonomy to group the related events together at the level of "Group". As shown in Figure A1, "Group" is the second layer in the RavenPack's integrated content classification map.

Table 18 presents all the groups which have the number of news more than 300. In terms of frequency, Earnings news and Revenues news are the highest in total. The news content can partially determine the timing predictability, for the proportions of schedule news of most groups are either as high as over 80% or as low as below 20%. The last column NIP is a projected news impact factors provided by RavenPack: scores of 0 are the center point representing no impacts, higher than which means there is an impact on the following two trading hours after the news, scores lower than which means there is no impact or unknown impacts. We can see that only the NIP of earnings, revenues and dividends news are obvious and significant. The impacts for the other kinds of news are unknown or undetermined in general.

— Insert Table 18 —

Table 19 and 20 shows RV relevance and IV predictability for news intensity (N) and news volatility (NV) respectively. The groups are sorted by their frequency. From these two tables, we learn an interesting empirical fact that the higher the realized volatility relevance, the higher the implied volatility predictability, in terms of both t-statistics of average coefficient and R^2. Positively volatility relevant news, such Earnings News (E) and Labor-Issues News (LI), will also be positively predicted by implied volatility. On the opposite, negatively volatility relevant news tends to be negatively predicted by implied volatility, such as Corporate-responsibility (CR). These are most likely unscheduled news whose releases can be strategically controlled by the manager in order to smooth the price changes and lower the stock volatility. As for the insignificant volatility relevance news, implied volatility shows a propensity of failing to predict. Figure 1 graphically presents the same results by giving scatter plots of the T-statistics and R^2, respectively.

— Insert Table 19 —

— Insert Table 20 —

— Insert Figure 1 —

In the end, the two-stage mediation analysis helps to rank the strongest type of news content channel over the total predictability. The results on news intensity and news volatility are shown in Table 21 and Table 22 respectively. The details of the implementation using multiple mediators can be referred to as in section 5.1.4. Due to the computation limitation on the iterations, I first cut 30 news content groups into three test groups according to the descending frequency order, as in Table 18. Second, I pick those news content groups with the average strength not less than 1.1% and 0.9% for N and NV, respectively. As a result of both N and NV, there are 14 news content groups selected. Lastly, I re-run the mediation model again with these 14 news content groups and present the results in Panel B sorting by the average strength. The top three most important news in terms of its strength of the overall news channel are consistent between N and NV: earnings news, analyst-ratings news, revenue news with their strength over 4%.

— Insert Table 21 —

— Insert Table 22 —

5.3.1 Subgroups under the Earnings and Revenue Group

As in Table 18, the earnings group and the revenue group take up more than 63% of the total number of news stories. To filter out RV-irrelevant news in these two big groups, I further try to derive the subgroups under RavenPacks' group. The official lower level of content classification under group in RavenPacks are types. But there are more than 40 types under these two big groups. Some types may lack the number of news to qualify in running the regression analysis. Therefore, I devise two mapping system as in Figure A2 and Figure A3 to aggregate the types under earnings group and revenue group respectively into two meaningful classifications of subgroups: one includes 12 accounting items, and the other includes nine information formats. We can see that the types of RavenPacks Classification are the combinations of the 12 accounting items and the nine information formats.

—- Insert Table 23 —-

Subgroups based on Accounting Items

The accounting items represent the variants or alias of "earnings" or "revenue" mentioned in the news. For example, the variants of earnings news can be "pretax earnings", "earnings-per-share", or "ebita", etc. The variants of revenues include "revenue", "same-store-sales", and "operating-margin". Table 23 summarizes the statistics related to the accounting items. As we can see, the original name of "earnings" and "revenue" is the most popular in usage. Particularly, some accounting items like "pretax-earnings", "ebit", and "ebitda" have very low standard deviation of CSS scores. This causes contradictory results for low RV-relevance and low IV-predictability when using news volatility, and high RV-relevance and high IV-predictability when using the news intensity. That is why in the later section we consider to filter out these subgroups of low volatility news.

—- Insert Table 24 —-

—- Insert Table 25 —-

From the RV-relevance and IV-predictability results in Table 24 and Table 25, we can decide to keep "earnings", "earnings-per-share", "operating-earnings" and "revenues" as the highly RV-relevant accounting items. In order to see which accounting-item-related news drives the RV_t - IV_{t-1} forecasting relationship the most, I further running a similar two-stage mediation analysis in Table 26 and in Table 27 for N and NV, respectively. The results are aligned consistently that the high volatility relevance news is associated with its relatively high importance in terms of its channel strength through which IV_{t-1} can predict RV_t.

—- Insert Table 26 —-

—- Insert Table 27 —-

Subgroups based on Information Formats

On the other hand, the information formats represent the way that earnings or revenue news can be presented. As depicted in Table 28, under earnings news, there are "(general) earnings" news, "earnings-guidance" news, "earnings-expectation" news, "earnings-guidance-expectation" news, and "earnings-estimate" news. We can see that the general earnings news and the general revenues news, which contain no specific information formats, make up the largest proportion.

—- Insert Table 28 —-

From the RV-relevance and IV-predictability results in Table 29 and Table 30, all the information formats are highly RV-relevant except that "earnings-estimates" and "revenue-volume" are lowest in their groups. Consistently, all the information formats are IV-predictable except for "earnings-estimates" and "revenue-volume".

—- Insert Table 29 —-

—- Insert Table 30 —-

Together with consistent mediation results as in Table 31 and Table 32, we can choose to use (general) earnings news, earnings expectation news, earnings guidance news, (general) revenue news and revenue guidance news.

—- Insert Table 31 —-

—- Insert Table 32 —-

Figure 2 and 3 graphically presents the same results as in the previous RV-relevance and IV-predictability tables. We can see a similar pattern as in Figure 1 that the strength of the IV-predictability is monotonically increasing with the strength of the RV-relevance.

—- Insert Figure 2 —-

—- Insert Figure 3 —-

To sum up, all these consistently interrelated results point to the same take-away: the level of IV-predictability is monotonically increasing with the level of the RV relevance. Therefore, the high RV-relevance news is also important news through whose channel IV_{t-1} predicts RV_t.

5.4 The Effect of Important News with a New Grouping

Since the previous results suggest that the IV predictability of news increases along with RV relevances of news, we can form the best sample of news that drives RV_t–IV_{t-1} forecasting relation the most by simply selecting the highly RV-relevant news.

I set up criteria of selecting highly RV-relevant news in terms of having a more-than-five t-statistics of coefficient average in regressing the realized volatility (RV_t) on the concurrent news measures (N_t or NV_t) in the group.

As a result, a sample of highly RV-relevant news are selected as follows:

- For the subgroups under the news content groups of "Earnings" and "Revenue" : for the different accounting items used as variants, "earnings", "earnings-per-share", "operating-earnings" and "revenues" are selected. For the different information formats, "general", "expectations" and "guidance" are selected.

- For the other news content groups: "analyst ratings", "“investor-relations", "partnership", "price-targets", "equity action", "acquisitions-mergers", "products-services", "labor-issues", "legal", "assets", "insider-trading", "credit-ratings" are selected.

- exclude news format as tabular materials or SEC filings.

Recall from the previous mediation results that these selected news are also the important news, through which channel IV_{t-1} is strong in relationship to predict RV_t.

— Insert Table 33 —

Table 33 presents results by adding the control of the concurrent new measures of news, which can generate lower the forecasting power of IV_{t-1} than adding the control of the original measures of all the news as in Table 7. The attenuation effects are, in general, positively significant for the different samples of firms by using the concurrent news intensity. But there is little effect when using the concurrent news volatility. Even though I have only selected all the highly relevant news, the naive aggregation of the news volatility may cause potential inaccuracy. I conjecture that this is probably because option investors may assign different importance to different kinds of news for different companies. This motivates me to utilize the two-step mediation analysis, which gives the freedom of assigning different channel strength to different kinds of news, instead of simple aggregation.

— Insert Table 34 —

After filtering out the unimportant news, I combine the previous three classification methods together to form new groups of news. Based on the news content, I separate the earnings, revenue and analyst ratings due to their frequency at the top three. The remaining news contents are categorized as "Others" news. Then I add a layer of classification of the timing predictability. As depicted in Table 18, since most of the earnings and revenue news are scheduled news and all the analyst rating news are unscheduled, so we only need to classify the other news into scheduled or unscheduled news. Lastly, we add the third layer of the news format (i.e., News-Flash, Full-Articles and Press-Release). Table 34 presents the summary statistics based on this new grouping. The most frequent news type is the earnings news flash with a total number around 1.5 million, while the least frequent type is the press-release of analyst ratings news with a small amount of only 1,470, which I drop out from further mediation analysis.

— Insert Table 35 —

— Insert Table 36 —

Table 35 and Table 36 generate the two-step mediation analysis results for N and NV, respectively. In the breakdown of the new grouping, the most important news types are the news flash of the analyst rating news, the news flash of earnings or revenue news. We can see that option investors pay the most attention to the earnings-related news and value the timeliness of news by focusing on the news flash. Other Unscheduled news consistently has a higher ranking than other scheduled news, which is even so for the unscheduled news flash.

There are some differences between news intensity and news volatility results. The full article of earnings news volatility has a higher ranking than its counterpart in the news intensity. We can see that option investors may rely on the information creation role of the media's in-depth analysis to predict the impact of the news.

Eventually, I am able to pin down 32.8% of the overall predictability of IV_{t-1} on RV_t comes through the news channel, which is also the best model throughout the whole study.

Chapter 6

Conclusion

This dissertation looks into the implied volatility and studies its information contents of news at the individual equity level. With the help of RavenPack news data, I am able to empirically pin down the reason why the implied volatility(IV) can forecast future realized volatility(RV). The RV_t–IV_{t-1} forecasting relation can be attributed to the two key features of news – RV-relevant and IV-predictable. Firstly, RV is highly relevant to the contemporaneous news. Secondly, the IV can predict future news. Taken together, IV can predict future RV by predicting future RV-relevant news. These two pieces of evidence hold up robustly in different news measures and for most kinds of news.

In detail, the option traders can forecast the future news' occurrence(s) (measured by the news intensity) as well as its impact magnitude (measured by the news volatility). In the general results, IV has greater forecast power on the future news volatility, for news volatility is consistently more RV-relevant than the news intensity.

In terms of timing predictability of news, even though the timing of unscheduled news is unknown beforehand, I am able to show that the option market has the ability to predict the occurrence of unscheduled news. As a result, IV can predict both scheduled news and unscheduled news. But in comparison, the unscheduled news is more RV-relevant, whereas the scheduled news is more IV-predictable. This implies that the unpredictable timing relatively hinders the option traders' ability to predict and creates more shocks to the market when it happens.

In terms of different news formats, implied volatility can predict both media-initiated news flash and the discretionary firm-initiated press releases very well. Both news-flash (proxy for the information dissemination role) and full-article (proxy for the information creation role) are similarly the strongest in RV-relevance. But full-article is comparatively weak in IV-predictability. This implies that the in-depth analytical skills of the reporters are different from the expertise of the option traders, which adds value to the market.

In terms of different news content groups, I am able to discover a significantly positive relationship between realized volatility relevance and implied volatility predictability across the content groups. Therefore, one can pick the most influential news on the RV_t–IV_{t-1} forecasting relation by selecting the highly RV-relevant news.

Lastly, through a two-stage mediation analysis, I am able to quantify the predictability strength of the news channel, which is represented by its proportion over the total predictability of implied volatility on realized volatility. The best mediation model suggests that about one-third of the total predictability is attributed to the news channel.

References

Beber, A., & Brandt, M. W. (2006). The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, *53*(8), 1997-2039. 19

Bonsall IV, S. B., Green, J. R., & Muller III, K. A. (2018). Are credit ratings more rigorous for widely covered firms? *The Accounting Review*, *93*(6), 61-94. 28

Britten-Jones, M., & Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, *55*(2), 839-866. 17

Busch, T., Christensen, B. J., & Nielsen, M. r. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, *160*(1), 48-57. 17

Bushman, R. M., Williams, C. D., & Wittenberg Moerman, R. (2016). The informational role of the media in private lending. *Journal of Accounting Research*. 21, 28

Canina, L., & Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies*, *6*(3), 659-681. 9, 17

Cao, C., Chen, Z., & Griffin, J. M. (2005). Informational content of option volume prior to takeovers. *The Journal of Business*, *78*(3), 1073–1109. 20

Chan, K., Chung, Y. P., & Fong, W.-M. (2002). The informational role of stock and option volume. *The Review of Financial Studies*, *15*(4), 1049–1075. 20

Chan, K., Ge, L., & Lin, T.-C. (2015). Informational content of options trading on acquirer announcement return. *Journal of Financial and Quantitative*

Analysis, *50*(5), 1057–1082. 20

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility1. *Journal of financial economics*, *50*(2), 125–150. 9, 17, 18, 36, 37, 41, 75

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of financial Economics*, *7*(3), 229-263. 23

Dang, T. L., Moshirian, F., & Zhang, B. (2015). Commonality in news around the world. *Journal of Financial Economics*, *116*(1), 82-110. 28

Day, T. E., & Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, *52*(1-2), 267-287. 9, 16, 18

Donders, M. W., Kouwenberg, R., & Vorst, T. C. (2000). Options and earnings announcements: an empirical study of volatility, trading volume, open interest and liquidity. *European Financial Management*, *6*(2), 149-171. 19

Donders, M. W., & Vorst, T. C. (1996). The impact of firm specific news on implied volatilities. *Journal of Banking & Finance*, *20*(9), 1447-1461. 19

Drake, M. S., Guest, N. M., & Twedt, B. J. (2014). The media and mispricing: The role of the business press in the pricing of accounting information. *The Accounting Review*, *89*(5), 1673-1701. 21, 53

Dubinsky, A., Johannes, M., Kaeck, A., & Seeger, N. J. (2018). Option pricing of earnings announcement risks. *The Review of Financial Studies*, hhy060. 18, 19

Easley, D., O'hara, M., & Srinivas, P. S. (1998). Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, *53*(2), 431–465. 20

Ederington, L. H., & Lee, J. H. (1996). The creation and resolution of market uncertainty: the impact of information releases on implied volatility. *Journal of Financial and Quantitative Analysis*, *31*(4), 513–539. 19

Fleming, J., Ostdiek, B., & Whaley, R. E. (1995). Predicting stock market volatility: A new measure. *Journal of Futures Markets*, *15*(3), 265-302. 17, 18

Gharghori, P., Maberly, E. D., & Nguyen, A. (2017). Informed trading around stock split announcements: Evidence from the option market. *Journal of Financial and Quantitative Analysis*, *52*(2), 705–735. 14, 20

Hayunga, D. K., & Lung, P. P. (2014). Trading in the options market around financial analysts' consensus revisions. *Journal of Financial and Quantitative Analysis*, *49*(3), 725–747. 14, 20

Isakov, D., & Perignon, C. (2001). Evolution of market uncertainty around earnings announcements. *Journal of banking & finance*, *25*(9), 1769-1788. 19

Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, *18*(4), 1305–1342. 9, 17, 18

Jin, W., Livnat, J., & Zhang, Y. (2012). Option prices leading equity prices: do option traders have an information advantage? *Journal of Accounting Research*, *50*(2), 401–432. 14, 20

Lamoureux, C. G., & Lastrapes, W. D. (1993). Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *The Review of Financial Studies*, *6*(2), 293-326. 9, 16, 18

Latane, H. A., & Rendleman Jr, R. J. (1976). Standard deviations of stock price ratios implied in option prices. *The Journal of Finance*, *31*(2), 369-381. 9, 16, 18

"Life in the public". (2017). Life in the public eye (Schumpeter). *The Economist*, *423*(9037), 62. 30

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394-419. 9

Patell, J. M., & Wolfson, M. A. (1979). Anticipated information releases reflected in call option prices. *Journal of accounting and economics*, *1*(2), 117-140. 19

Patell, J. M., & Wolfson, M. A. (1981). The ex ante and ex post price effects of quarterly earnings announcements reflected in option and stock prices. *Journal of Accounting Research*, 434-458. 19

Poon, S.-H., & Granger, C. (2005). Practical issues in forecasting volatility. *Financial analysts journal*, 45-56. 17

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, *40*(3), 879–891. 11, 42

Qing, H. (2016). Is there information leakage prior to share repurchase announcements? evidence from daily options trading. *Journal of Financial Markets*, *27*(C), 79–101. 14, 20

RavenPack. (2018). Ravenpack analytics user guide and service overview. *RavenPack User Guide*, *1*, 1–38. 25, 26

Shevlin, T., & Thornock, J. (2015). Market (in) attention and the strategic scheduling and timing of earnings announcements. *Journal of Accounting and Economics*, *60*(1), 36-55. 25

Taylor, S. J., Yadav, P. K., & Zhang, Y. (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance*, *34*(4), 871-881. 17, 37

Twedt, B. (2015). Spreading the word: Price discovery and newswire dissemination of management earnings guidance. *The Accounting Review*, *91*(1), 317-346. 21, 53

Zhang, J. (2018). Informed options trading prior to dividend change announcements. *Financial Management*, *47*(1), 81–103. 14, 20

Table 1: Data Attrition

This table describes the data attrition for merging CRSP, OptionMetrics and RavenPack. The first column lists out all the years from 2000 to 2017. The second column shows the number of firms from CRSP with share code 10 or 11 and from the three primary stock exchanges: NYSE, AMEX, and NASDAQ. The third column shows the number of firms when merging CRSP and OptionMetrics data. The fourth column shows the number of firms when merging CRSP, OptionMetrics and RavenPack together. The fifth depicts the ratio of firm-months without news over the total number of firm-months. The last column shows the number of firms in my final dataset where I further request a firm to have at least two years of data to be included. The years below the horizontal dashed line are my selected period where RavenPack starts to include the package of Press Releases from 2004.

(1)	(2)	(3)	(4)	(5)	(6)
Year	CRSP	+ OptionMetrics	+ RavenPack	No News Month (%)	Final Dataset
2000	7241	2431	2249	21.2%	
2001	6549	2190	2130	21.9%	
2002	5834	2096	2058	21.1%	
2003	5414	1952	1924	18.8%	
2004	5211	2110	2077	9.2%	1781
2005	5153	2208	2172	10.7%	1957
2006	5069	2345	2306	10.6%	2179
2007	5051	2444	2404	9.7%	2266
2008	4744	2437	2406	8.7%	2285
2009	4472	2341	2307	10.0%	2187
2010	4274	2403	2357	9.5%	2272
2011	4101	2596	2544	8.7%	2442
2012	3954	2628	2587	9.7%	2497
2013	3905	2710	2678	9.8%	2610
2014	3993	2785	2758	11.5%	2686
2015	4024	2940	2914	10.0%	2739
2016	3932	2857	2836	4.7%	2615
2017	3862	2632	2615	2.7%	2395

Table 2: Descriptive Statistics

This table summarizes the distribution of all my used variables. Panel A shows the pool statistics at the news-story level. CSS is the Composite Sentiment Score which uses a universal integrated textual analysis to evaluate the news sentiment of a given story. ESS is the Event Sentiment Score, which is constructed by financial experts to assess news sentiment based on different content categories. The range of both sentiment scores is [-1, 1]. Panel B shows the pool statistics at the firm-month level. IV (RV) represents implied (realized) volatilities. VP is the volatility risk premium, calculated as the IV of the previous month subtracting the current month's RV. N stands for the number of news. NV stands for news volatility, which is constructed as follows: sum up the square of news-story level CSS within a month; if there is no news within a month, replace it with zero; in the end, take the square root and annualize it by multiplying $\sqrt{12}$. For both N and NV, we further categorize the news into – (1) based on the predictability of timing: scheduled news (Sch) and unscheduled news (UnSch); (2) based on the format of news: News-Flash (NF), Press-Release (PR), Full-article (FA), Tabular-Material(TM), and SEC filings (SEC). The period spans from 2004 to 2017.

Panel A: News Level									
Variable	FREQ	MEAN	STD	SKEW	MIN	P25	P50	P75	MAX
CSS	7,418,062	0.003	0.105	-3.345	-0.92	0	0	0.04	1
ESS	7,418,062	0.13	0.459	-0.379	-1	-0.2	0	0.48	1
Panel B: Firm-Month Level									
Variable	FREQ	MEAN	STD	SKEW	MIN	P25	P50	P75	MAX
IV	363,456	0.466	0.299	2.518	0.015	0.278	0.386	0.553	2.944
RV	363,456	0.394	0.281	5.263	0.005	0.223	0.323	0.479	22.108
VP	363,456	0.072	0.256	-1.620	-19.810	-0.016	0.055	0.138	2.590
N	363,456	20.410	27.506	3.570	0	2	8	32	1065
N.Sch	363,456	15.475	23.173	2.992	0	0	3	26	905
N.Unsch	363,456	4.934	7.969	15.056	0	1	3	6	673
N.NF	363,456	9.697	14.515	1.663	0	0	2	17	129
N.PR	363,456	3.836	5.396	3.163	0	0	2	6	169
N.FA	363,456	6.406	14.056	13.253	0	0	2	7	987
N.TM	363,456	0.053	0.463	35.987	0	0	0	0	70
N.SEC	363,456	0.458	0.723	3.040	0	0	0	1	26
NV	363,456	1.055	1.263	2.056	0	0.139	0.576	1.512	24.652
NV.Sch	363,456	0.768	1.180	2.390	0	0	0.208	1.078	24.484
NV.unSch	363,456	0.505	0.685	2.387	0	0	0.250	0.679	12.912
NV.NF	363,456	0.527	0.900	2.217	0	0	0.139	0.604	7.600
NV.PR	363,456	0.253	0.405	3.251	0	0	0	0.346	10.392
NV.FA	363,456	0.640	0.976	3.175	0	0	0.277	0.857	24.222
NV.TM	363,456	0.009	0.094	21.205	0	0	0	0	6.962
NV.SEC	363,456	0.024	0.148	16.156	0	0	0	0	6.772

Table 3: Correlation between Different Volatility and News Measures

This table summarizes the correlation among different volatility and news measures at the firm-month level. First, run the time-series correlation test for each firm. Then, show the average correlation coefficient across all firms. t the is the timestamp of a month. IV (RV) represents implied (realized) volatilities. VP is the volatility risk premium, calculated as the IV of the previous month subtracting the current month's RV. N stands for the number of news. NV stands for news volatility, which is constructed as follows: take the sum of squared of all CSS at a news-story level over a month; if there is no news within a month, replace it with zero; finally, take the square root and annualize it by multiplying $\sqrt{12}$. For both N and NV, we further categorize the news into – (1) based on the predictability of timing: scheduled news (Sch) and unscheduled news (UnSch); (2) based on the format of news: News-Flash (NF), Press-Release (PR), Full-article (FA), Tabular-Material(TM), and SEC filings(SEC). The lower triangle shows the Pearson correlations, and the upper triangle shows the Spearman correlations. The sample period spans from 2004 to 2017.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	RV_t	1	0.350	0.462	0.465	-0.082	0.198	0.174	0.161	0.241	0.181	0.135	0.0630	-0.012	0.214	0.182	0.165	0.229	0.143	0.171	0.0574	-0.025
2	RV_{t-1}	0.365	1	0.413	0.460	-0.567	-0.117	-0.133	-0.065	-0.092	-0.103	-0.116	0.0086	0.0153	-0.082	-0.110	-0.042	-0.061	-0.071	-0.084	0.0083	0.0028
3	IV_t	0.469	0.417	1	0.612	0.0085	-0.096	-0.086	-0.085	-0.063	-0.040	-0.123	0.0137	-0.016	-0.058	-0.070	-0.051	-0.021	-0.037	-0.078	0.0123	-0.038
4	IV_{t-1}	0.471	0.470	0.625	1	0.0321	0.0798	0.0832	0.0274	0.128	0.0944	0.0183	0.0531	0.0228	0.0843	0.0859	0.0283	0.113	0.0673	0.0491	0.0455	-0.005
5	VP_{t-1}	-0.133	-0.659	-0.053	-0.031	1	0.0345	0.0451	0.0011	0.0440	0.0399	0.0251	0.0025	-0.033	0.0278	0.0419	-0.003	0.0378	0.0204	0.0220	0.0009	-0.027
6	N_t	0.182	-0.111	-0.089	0.0787	0.0511	1	0.907	0.697	0.857	0.773	0.813	0.169	0.0944	0.819	0.808	0.529	0.660	0.576	0.714	0.148	0.0064
7	$N.Sch_t$	0.159	-0.112	-0.082	0.0832	0.0559	0.979	1	0.443	0.800	0.793	0.702	0.169	0.0255	0.730	0.872	0.345	0.588	0.585	0.625	0.148	-0.016
8	$N.UnSch_t$	0.196	-0.060	-0.082	0.0218	0.0069	0.570	0.413	1	0.551	0.431	0.692	0.0964	0.204	0.623	0.422	0.734	0.498	0.334	0.593	0.0822	0.0522
9	$N.NF_t$	0.183	-0.095	-0.068	0.102	0.0522	0.938	0.950	0.427	1	0.667	0.591	0.149	0.0214	0.737	0.722	0.452	0.778	0.505	0.571	0.131	-0.010
10	$N.PR_t$	0.152	-0.096	-0.055	0.0647	0.0443	0.779	0.778	0.400	0.677	1	0.509	0.158	0.0074	0.624	0.704	0.320	0.516	0.731	0.459	0.137	-0.006
11	$N.FA_t$	0.132	-0.101	-0.109	0.0116	0.0332	0.738	0.662	0.684	0.544	0.466	1	0.134	0.0205	0.714	0.665	0.539	0.476	0.382	0.826	0.119	-0.061
12	$N.TM_t$	0.0605	0.0093	0.0177	0.0472	0.0032	0.182	0.180	0.0974	0.155	0.164	0.127	1	0.0035	0.136	0.146	0.0709	0.112	0.112	0.115	0.859	-0.015
13	$N.SEC_t$	-0.010	0.0039	-0.021	0.0072	-0.021	0.0715	0.0428	0.174	0.0414	0.0188	0.0200	0.0003	1	0.0227	0.0202	0.0308	-0.007	0.0012	0.0145	0.0021	0.465
14	NV_t	0.215	-0.053	-0.020	0.0962	0.0328	0.704	0.672	0.497	0.641	0.559	0.583	0.130	0.0277	1	0.825	0.710	0.753	0.565	0.810	0.135	0.0395
15	$NV.Sch_t$	0.175	-0.067	-0.025	0.0981	0.0456	0.723	0.729	0.351	0.674	0.602	0.544	0.135	0.0319	0.904	1	0.357	0.630	0.627	0.681	0.146	0.0039
16	$NV.UnSch_t$	0.202	-0.019	-0.018	0.0404	-0.005	0.385	0.291	0.622	0.312	0.259	0.445	0.0654	0.0138	0.620	0.302	1	0.539	0.322	0.631	0.0696	0.0869
17	$NV.NF_t$	0.199	-0.030	0.0092	0.106	0.0329	0.546	0.527	0.366	0.548	0.432	0.364	0.0952	0.0023	0.797	0.722	0.461	1	0.403	0.491	0.102	-0.010
18	$NV.PR_t$	0.128	-0.060	-0.037	0.0544	0.0236	0.544	0.540	0.296	0.484	0.659	0.327	0.105	0.0141	0.508	0.531	0.272	0.324	1	0.365	0.133	0.0018
19	$NV.FA_t$	0.179	-0.051	-0.033	0.0589	0.0217	0.576	0.532	0.486	0.484	0.389	0.641	0.0997	0.0099	0.789	0.680	0.602	0.410	0.300	1	0.111	-0.046
20	$NV.TM_t$	0.0527	0.0073	0.0126	0.0380	-19E-5	0.156	0.155	0.0827	0.135	0.140	0.109	0.791	0.0027	0.138	0.145	0.0660	0.0892	0.133	0.102	1	-0.013
21	$NV.SEC_t$	-0.014	0.0003	-0.026	-0.004	-0.018	0.0044	-0.004	0.0378	-0.006	0.0027	-0.037	-0.010	0.470	0.0405	0.0342	0.0364	-0.005	0.0102	-0.022	-0.008	1

Table 4: Time-series Regression to Predict Realized Volatilities

This table extends Christensen and Prabhala (1998)'s result and examines the information content of implied volatilities for all the individual stocks. For each stock, the model runs the times series regression with the realized volatilities (RV_t) on the past implied volatilities (IV_{t-1}), the past realized volatility (RV_{t-1}) and the past volatility risk premium (VP_{t-1}). Panel A shows the whole sample. Panel B filters out the stocks with their average options trading volume below the median. Panel C only keeps the top 25% stocks with the highest options trading volume. The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative significant (significant if p-value ≤ 0.01, and the standard errors are newey-west adjusted with 12 lags). The last column shows the average R^2. The sample period is from 2004 to 2017.

Model: RV_t = Intercept $+\beta_1$ RV_{t-1} + β_2 IV_{t-1}+ β_3 VP_{t-1} + ε_t

Panel A: Whole Sample (363,456 firm-month)

	Intercept	RV_{t-1}	IV_{t-1}	VP_{t-1}	R^2
(1)	0.268	0.371			20.7%
	(100.07)	(83.16)			
		[++56%,+34%,-9%, - -1%]			
(2)	0.121		0.661		28.5%
	(39.02)		(101.21)		
			[++68%,+28%,-4%, - -0%]		
(3)	0.105	0.130	0.577		32.8%
	(37.72)	(36.25)	(87.15)		
		[++21%,+50%,-27%, - -3%]	[++61%,+34%,-5%, - -0%]		
(4)	0.111	0.077	0.613	-0.053	34.3%
	(36.39)	(13.47)	(81.88)	(-10.84)	
		[++14%,+44%,-39%, - -3%]	[++56%,+38%,-6%, - -0%]	[++2%,+37%,-54%, - -6%]	

Panel B: Sample with option trading volume above median (207,169 firm-month)

	Intercept	RV_{t-1}	IV_{t-1}	VP_{t-1}	R^2
(5)	0.264	0.408			23.1%
	(67.60)	(66.10)			
		[++61%,+32%,-7%, - -1%]			
(6)	0.067		0.823		36.5%
	(17.16)		(105.14)		
			[++84%,+15%,-1%, - -0%]		
(7)	0.063	0.075	0.758		38.6%
	(17.34)	(17.48)	(89.95)		
		[++12%,+52%,-33%, - -3%]	[++78%,+20%,-2%, - -0%]		
(8)	0.071	-0.004	0.817	-0.089	40.1%
	(18.13)	(-0.58)	(84.31)	(-12.16)	
		[++6%,+40%,-50%, - -4%]	[++73%,+24%,-3%, - -0%]	[++2%,+31%,-60%, - -7%]	

Panel C: Sample with option trading volume at top 25% (112,261 firm-month)

	Intercept	RV_{t-1}	IV_{t-1}	VP_{t-1}	R^2
(9)	0.234	0.454			26.8%
	(44.50)	(52.39)			
		[++67%,+27%,-5%, - -1%]			
(10)	0.034		0.905		42.2%
	(7.38)		(92.25)		
			[++90%,+9%,-0%, - -0%]		
(11)	0.035	0.053	0.849		43.7%
	(7.86)	(9.00)	(75.85)		
		[++9%,+52%,-35%, - -4%]	[++85%,+14%,-1%, - -0%]		
(12)	0.042	-0.025	0.911	-0.088	45.1%
	(8.52)	(-2.33)	(71.57)	(-7.90)	
		[++4%,+38%,-54%, - -4%]	[++82%,+17%,-1%, - -0%]	[++2%,+29%,-62%, - -7%]	

Table 5: Realized Volatility-News Relevance Test

This table examines the relationship between realized volatilities and news intensities(N) or news volatilities(NV). For each stock, the model runs the times series regression with the realized volatilities(RV_t) on the concurrent news measures(N_t or NV_t) and control for the past realized volatilities(RV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value ≤ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Depvar	Panel A: News Intensity (N) Model: RV_t = Intercept + β_1 N_t + β_2 RV_{t-1} + ε_t			
RV_t	**Intercept**	**N_t**	**RV_{t-1}**	**R^2**
(1)	0.383	0.002		6.0%
	(165.20)	(35.58)		
	[++100%,+0%,-0%, - -0%]	[++30%,+58%,-12%, - -0%]		
(2)	0.212	0.003	0.405	28.1%
	(86.52)	(42.13)	(95.75)	
	[++80%,+20%,-1%, - -0%]	[++44%,+50%,-6%, - -0%]	[++62%,+31%,-6%, - -0%]	

Depvar	Panel B: News Volatility (NV) Model: RV_t = Intercept + β_1 NV_t + β_2 RV_{t-1} + ε_t			
RV_t	**Intercept**	**NV_t**	**RV_{t-1}**	**R^2**
(3)	0.373	0.045		7.3%
	(161.50)	(56.95)		
	[++99%,+1%,-0%, - -0%]	[++34%,+57%,-9%, - -0%]		
(4)	0.215	0.046	0.390	28.2%
	(87.60)	(59.90)	(92.04)	
	[++80%,+19%,-1%, - -0%]	[++43%,+52%,-6%, - -0%]	[++61%,+32%,-7%, - -0%]	

Table 6: Implied Volatilities-News Predictability Test

This table examines the information content of implied volatilities which is represented by news predictability. For each stock, the model runs the times series regressions by using the past implied volatility(IV_{t-1}) to forecast news intensity(N_t) or news volatilities(NV_t), plus controlling for the past news. The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value $\leq$ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Panel A: News Intensity (N)

Model: N_t = Intercept + β IV_{t-1} + γ_1 N_{t-1} + γ_2 N_{t-2} + γ_3 N_{t-3} + ε_t

Depvar N_t	**Intercept**	**IV_{t-1}**	**N_{t-1}**	**N_{t-2}**	**N_{t-3}**	**R^2**
(1)	11.595	17.925				2.9%
	(42.66)	(25.46)				
	[++53%,+34%,-11%, - -2%]	[++12%,+57%,-28%, - -3%]				
(2)	24.782		-0.324			11.8%
	(131.84)		(-171.84)			
	[++100%,+0%,-0%, - -0%]		[++0%,+1%,-10%, - -88%]			
(3)	19.647	12.137	-0.323			14.4%
	(74.30)	(18.96)	(-174.25)			
	[++66%,+28%,-5%, - -1%]	[++10%,+51%,-34%, - -4%]	[++0%,+1%,-10%, - -88%]			
(4)	19.136	6.915	-0.337	-0.197	0.292	38.1%
	(80.99)	(14.21)	(-111.98)	(-72.92)	(62.90)	
	[++54%,+41%,-4%, - -0%]	[++7%,+50%,-39%, - -4%]	[++0%,+2%,-24%, - -74%]	[++1%,+9%,-50%, - -40%]	[++47%,+36%,-16%, - -1%]	

Table 6: Implied Volatilities-News Predictability Test (Continue)

Depvar NV_t	Panel B: News Volatility (NV) Model: NV_t = Intercept + β IV_{t-1} + γ_1 NV_{t-1} + γ_2 NV_{t-2} + γ_3 NV_{t-3} + ε_t					R^2
	Intercept	IV_{t-1}	NV_{t-1}	NV_{t-2}	NV_{t-3}	
(5)	0.615 (55.16) [++48%,+41%,-10%, - -1%]	0.921 (33.21) [++18%,+56%,-24%, - -2%]				3.2%
(6)	1.222 (143.30) [++99%,+1%,-0%, - -0%]		-0.195 (-86.68) [++0%,+8%,-35%, - -57%]			5.7%
(7)	0.847 (72.89) [++54%,+38%,-7%, - -1%]	0.876 (31.60) [++17%,+55%,-26%, - -2%]	-0.204 (-92.83) [++0%,+6%,-35%, - -58%]			9.1%
(8)	0.848 (63.88) [++45%,+47%,-8%, - -0%]	0.696 (26.92) [++13%,+57%,-28%, - -2%]	-0.213 (-77.27) [++0%,+7%,-45%, - -48%]	-0.122 (-48.95) [++1%,+17%,-58%, - -25%]	0.209 (65.69) [++38%,+48%,-13%, - -1%]	21.3%

Table 7: The News Channel

This table examines the news channel through which the implied volatility can predict future volatility. For each stock, the model runs the times series regression with the realized volatilities (RV_t) on the past volatilities (RV_{t-1}), the past implied volatility(IV_{t-1}) and the past volatility risk premium (VP_{t-1}), plus controlling for the concurrent news intensity(N_t) or news volatility(NV_t). Panel A shows the whole sample. Panel B filters out the stocks with their average options trading volume below the median. Panel C only keeps the top 25% stocks with the highest options trading volume. The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value $\leq$ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Model: RV_t = Intercept + β_1 RV_{t-1} + β_2 IV_{t-1} + β_3 VP_{t-1} + θ (N_t or NV_t) + ε_t

Panel A: Whole Sample (363,456 firm-month)

Depvar RVt	**Intercept**	**RV$_{t-1}$**	**IV$_{t-1}$**	**VP$_{t-1}$**		**R^2**
					N$_t$	
(1)	0.079	0.164	0.531	-0.005	0.002	39.1%
	(27.34)	(28.96)	(71.82)	(-1.04)	(37.14)	
		[++20%,+51%,-28%, - -1%]	[++48%,+44%,-8%, - -0%]	[++4%,+45%,-47%, - -4%]	[++28%,+62%,-10%, - -0%]	
					NV$_t$	
(2)	0.080	0.143	0.542	-0.016	0.038	39.5%
	(27.42)	(25.41)	(73.31)	(-3.32)	(47.96)	
		[++18%,+50%,-30%, - -2%]	[++50%,+41%,-8%, - -0%]	[++3%,+44%,-48%, - -5%]	[++29%,+61%,-9%, - -0%]	

Panel B: Sample with option trading volume above median (207,169 firm-month)

Depvar RVt	**Intercept**	**RV$_{t-1}$**	**IV$_{t-1}$**	**VP$_{t-1}$**		**R^2**
					N$_t$	
(3)	0.044	0.084	0.723	-0.033	0.002	43.7%
	(11.62)	(10.73)	(71.92)	(-4.35)	(22.82)	
		[++11%,+50%,-38%, - -2%]	[++65%,+31%,-4%, - -0%]	[++3%,+40%,-52%, - -4%]	[++22%,+65%,-12%, - -0%]	
					NV$_t$	
(4)	0.043	0.068	0.730	-0.042	0.035	44.3%
	(11.23)	(8.82)	(72.24)	(-5.70)	(32.48)	
		[++10%,+50%,-39%, - -2%]	[++67%,+29%,-4%, - -0%]	[++2%,+39%,-53%, - -5%]	[++25%,+65%,-10%, - -0%]	

Panel C: Sample with option trading volume at top 25% (112,261 firm-month)

Depvar RVt	**Intercept**	**RV$_{t-1}$**	**IV$_{t-1}$**	**VP$_{t-1}$**		**R^2**
					N$_t$	
(5)	0.017	0.059	0.819	-0.033	0.002	48.3%
	(3.68)	(5.26)	(59.27)	(-2.86)	(15.80)	
		[++8%,+49%,-42%, - -1%]	[++75%,+22%,-3%, - -0%]	[++3%,+39%,-53%, - -4%]	[++19%,+68%,-13%, - -0%]	
					NV$_t$	
(6)	0.014	0.047	0.825	-0.041	0.029	48.8%
	(3.06)	(4.22)	(58.34)	(-3.60)	(22.41)	
		[++7%,+48%,-44%, - -1%]	[++76%,+21%,-2%, - -0%]	[++2%,+37%,-56%, - -5%]	[++23%,+67%,-11%, - -0%]	

Table 8: The Strength of the News Channel

This table quantifies the strength of the news channel through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. Panel A summarizes the channel of News Intensity (N), and Panel B shows the distribution of the News Volatility (NV). The detail calculation involves a two-stage mediation analysis for each firm i as follows:

$$\begin{cases} \text{Total Effect:} & \mathrm{RV}_{i,t} = \mathrm{Int}_i + \beta_i \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1:} & \mathrm{X}_{i,t} = \mathrm{Int}_i + \theta_i \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 2:} & \mathrm{RV}_{i,t} = \mathrm{Int}_i + \beta_i' \mathrm{IV}_{i,t-1} + \phi_i \mathrm{X}_{i,t} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \end{cases}$$

where X can be substituted as N or NV, and the controls are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. β_i represents the total predictability, while β_i' represents the non-news-channel predictability. We can implement the two-stage regressions to acquire the mediator effect (i.e., the news-channel predictability) as $\theta_i\phi_i$. In these mediation models, we have total equality: $\beta_i = \beta_i' + \theta_i\phi_i$. Therefore, the strength of the news channel can be quantified as $\theta_i\phi_i/\beta_i$. In specific, I first filter out stocks with negative total predictability. Then, for each stock, through iteration, I determine whether to include the mediator to ensure the total predictability can be decomposed into non-negative components (i.e., news and non-news channel predictability) and that the variable choice is the best combination with the highest R^2 for stage 2. If the mediator $\mathrm{X}_{i,t}$ is not selected, the news channel predictability is set to be zero. The sample period is from 2004 to 2017.

	Panel A: News Intensity (N)							
	mean	std	skew	p5	p25	p50	p75	p95
N	12.5%	18.8%	2.703	0.0%	0.5%	5.9%	15.6%	51.0%
	Panel B: News Volatility (NV)							
	mean	std	skew	p5	p25	p50	p75	p95
NV	11.2%	17.3%	3.048	0.0%	0.2%	5.4%	14.1%	43.7%

Table 9: News Summary Based on Timing Predictability Classification

This table summarizes the news frequency and news sentiment grouped by the timing predictability: scheduled news (Sch) and unscheduled news (UnSch). The ① row counts the number of the type of news stories and the square bracket below shows the percentage of the type of news stories over the total 7,418,062 news stories in the sample. The ② row counts the number of firm-month that contains the type of news and the square bracket below shows the percentage of the type of firm-month over the total 363,456 firm-months in the sample. ③ and ④ shows the average Composite Sentiment Score (CSS) and Event Sentiment Score (ESS) respectively. ⑤ measures News Impact Projection. The parenthesis shows the T-statistics of the average's difference from zero. The sample period is from 2004 to 2017.

		Sch	UnSch
①	# News Stories	5,624,626	1,793,436
		[75.82%]	[24.18%]
②	# Firm-Month	251,700	306,352
		[69.25%]	[84.29%]
③	CSS	0.006	-0.004
		(133.69)	(-52.59)
④	ESS	0.130	0.129
		(688.27)	(351.95)
⑤	NIP	0.023	-0.132
		(297.39)	(-830.21)

Table 10: Volatilities Relevance of Scheduled/Unscheduled News

This table examines the relation between concurrent realized volatilities and the intensities/volatilities of scheduled/unscheduled news. For each individual stock, the model runs the times series regression with the realized volatilities (RV_t) on the concurrent news measures which include: scheduled(unscheduled) news intensity denoted by $N.Sch_t$($N.UnSch_t$) and scheduled(unscheduled) news volatility denoted by $NV.Sch_t$($NV.UnSch_t$). The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value $\leq$ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Panel A: News Intensity (N)

Model: RV_t = Intercept + β_1N.Sch_t + β_2N.$UnSch_t$ + ϵ_t

Depvar RV_t	**Intercept**	**N.Sch_t**	**N.$UnSch_t$**	**R^2**
(1)	0.392	0.002		5.0%
	(164.35)	(33.14)		
	[++100%,+0%,-0%, - -0%]	[++26%,+60%,-14%, - -0%]		
(2)	0.368		0.015	8.2%
	(168.71)		(40.28)	
	[++99%,+1%,-0%, - -0%]		[++27%,+56%,-17%, - -0%]	
(3)	0.364	0.001	0.012	10.9%
	(167.26)	(15.88)	(32.41)	
	[++99%,+1%,-0%, - -0%]	[++14%,+59%,-26%, - -1%]	[++15%,+58%,-26%, - -1%]	

Panel B: News Volatility (NV)

Model: RV_t = Intercept + β_1NV.Sch_t + β_2NV.$UnSch_t$ + ϵ_t

Depvar RV_t	**Intercept**	**NV.Sch_t**	**NV.$UnSch_t$**	**R^2**
(4)	0.390	0.038		5.6%
	(163.31)	(46.02)		
	[++100%,+0%,-0%, - -0%]	[++28%,+58%,-13%, - -0%]		
(5)	0.378		0.098	7.5%
	(168.33)		(39.53)	
	[++99%,+1%,-0%, - -0%]		[++25%,+62%,-13%, - -1%]	
(6)	0.367	0.026	0.082	11.1%
	(164.30)	(32.42)	(32.49)	
	[++99%,+1%,-0%, - -0%]	[++18%,+61%,-20%, - -1%]	[++16%,+64%,-20%, - -1%]	

Table 11: Implied Volatilities Forecasting Scheduled/Unscheduled News

This table examines the information content of implied volatilities which is represented by its forecast power on scheduled/unscheduled news. For each individual stock, the model runs the times series regressions with the news measures on the past implied volatility (plus controlling for past news). The news measures include: scheduled(unscheduled) news intensity denoted by N.Sch$_t$(N.UnSch$_t$) and scheduled(unscheduled) news volatility denoted by NV.Sch$_t$(NV.UnSch$_t$). The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value ≤ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Panel A: News Intensity (N)

Model: $\mathbf{N.Sch_t = Intercept + \beta\, IV_{t-1} + \gamma_1\, N.Sch_{t-1} + \gamma_2\, N.Sch_{t-2} + \gamma_3\, N.Sch_{t-3} + \varepsilon_t}$

Depvar N.Sch$_t$	Intercept	IV$_{t-1}$	N.Sch$_{t-1}$	N.Sch$_{t-2}$	N.Sch$_{t-3}$	R^2
(1)	7.575	16.920				2.9%
	(33.73)	(27.29)				
	[++46%,+37%,-14%, - -2%]	[++12%,+59%,-26%, - -3%]				
(2)	14.811	11.337	-0.358			16.1%
	(67.31)	(20.43)	(-271.40)			
	[++61%,+32%,-7%, - -1%]	[++10%,+54%,-32%, - -4%]	[++0%,+0%,-4%, - -95%]			
(3)	16.770	5.192	-0.389	-0.245	0.257	40.5%
	(86.23)	(12.64)	(-140.26)	(-95.05)	(54.68)	
	[++56%,+40%,-4%, - -0%]	[++6%,+50%,-40%, - -4%]	[++0%,+1%,-15%, - -85%]	[++0%,+4%,-43%, - -53%]	[++41%,+37%,-20%, - -2%]	

Model: $\mathbf{N.UnSch_t = Intercept + \beta\, IV_{t-1} + \gamma_1\, N.UnSch_{t-1} + \gamma_2\, N.UnSch_{t-2} + \gamma_3\, N.UnSch_{t-3} + \varepsilon_t}$

Depvar N.UnSch$_t$	Intercept	IV$_{t-1}$	N.UnSch$_{t-1}$	N.UnSch$_{t-2}$	N.UnSch$_{t-3}$	R^2
(4)	4.020	1.005				2.9%
	(49.36)	(6.58)				
	[++58%,+35%,-6%, - -0%]	[++7%,+45%,-43%, - -4%]				
(5)	3.653	1.352	0.022			6.4%
	(59.21)	(9.99)	(7.10)			
	[++56%,+37%,-7%, - -0%]	[++7%,+46%,-43%, - -4%]	[++10%,+42%,-37%, - -11%]			
(6)	2.841	1.608	0.005	0.000	0.168	14.2%
	(47.04)	(12.05)	(1.52)	(0.01)	(38.97)	
	[++42%,+49%,-9%, - -0%]	[++6%,+48%,-42%, - -3%]	[++8%,+43%,-39%, - -10%]	[++6%,+46%,-42%, - -6%]	[++30%,+52%,-17%, - -1%]	

Table 11: Implied Volatilities Forecasting Scheduled/Unscheduled News (Continue)

	Panel B: News Volatility (NV)					
Depvar	**Model: $\mathbf{NV.Sch}_t = \mathbf{Intercept} + \beta\ \mathbf{IV}_{t-1} + \gamma_1\ \mathbf{NV.Sch}_{t-1} + \gamma_2\ \mathbf{NV.Sch}_{t-2} + \gamma_3\ \mathbf{NV.Sch}_{t-3} + \varepsilon_t$**					
NV.Sch$_t$	Intercept	IV$_{t-1}$	NV.Sch$_{t-1}$	NV.Sch$_{t-2}$	NV.Sch$_{t-3}$	R^2
(1)	0.361	0.901				3.1%
	(33.74)	(33.33)				
	[++36%,+46%,-16%, - -2%]	[++18%,+57%,-23%, - -2%]				
(2)	0.589	0.832	-0.257			11.0%
	(54.40)	(31.77)	(-135.69)			
	[++43%,+44%,-11%, - -1%]	[++17%,+56%,-25%, - -2%]	[++0%,+2%,-18%, - -80%]			
(3)	0.665	0.596	-0.270	-0.164	0.227	26.3%
	(54.19)	(25.19)	(-101.24)	(-67.79)	(66.51)	
	[++40%,+49%,-10%, - -1%]	[++12%,+56%,-29%, - -3%]	[++0%,+3%,-28%, - -69%]	[++0%,+10%,-52%, - -37%]	[++39%,+47%,-13%, - -1%]	
Depvar	**Model: $\mathbf{NV.UnSch}_t = \mathbf{Intercept} + \beta\ \mathbf{IV}_{t-1} + \gamma_1\ \mathbf{NV.UnSch}_{t-1} + \gamma_2\ \mathbf{NV.UnSch}_{t-2} + \gamma_3\ \mathbf{NV.UnSch}_{t-3} + \varepsilon_t$**					
NV.UnSch$_t$	Intercept	IV$_{t-1}$	NV.UnSch$_{t-1}$	NV.UnSch$_{t-2}$	NV.UnSch$_{t-3}$	R^2
(4)	0.374	0.219				2.7%
	(61.89)	(16.12)				
	[++43%,+48%,-9%, - -0%]	[++9%,+51%,-37%, - -4%]				
(5)	0.371	0.228	-0.003			4.6%
	(61.27)	(16.46)	(-1.32)			
	[++41%,+49%,-9%, - -0%]	[++8%,+51%,-37%, - -4%]	[++4%,+43%,-42%, - -10%]			
(6)	0.340	0.238	-0.010	-0.014	0.074	9.2%
	(53.82)	(16.84)	(-3.93)	(-5.64)	(27.59)	
	[++33%,+56%,-11%, - -1%]	[++8%,+52%,-38%, - -3%]	[++3%,+44%,-45%, - -8%]	[++3%,+42%,-48%, - -7%]	[++11%,+57%,-29%, - -3%]	

Table 12: The Channel of Scheduled/Unscheduled News

This table examines the scheduled or unscheduled news channel through which the implied volatility can predict future volatility. For each individual stock, the model runs the times series regression with the realized volatilities (RV_t) on the past volatilities(RV_{t-1}), the past implied volatility(IV_{t-1}) and the past volatility risk premium (VP_{t-1}), plus controlling for the concurrent news measures which include: scheduled(unscheduled) news intensity denoted by N.Sch_t(N.UnSch_t) and scheduled(unscheduled) news volatility denoted by NV.Sch_t(NV.UnSch_t). Panel A shows the whole sample. Panel B filters out the stocks with their average options trading volume below the median. Panel C only keeps the top 25% stocks with the highest options trading volume. The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value $\leq$ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Model: RV_t = Intercept + β_1 RV_{t-1} + β_2 IV_{t-1} +β_3 VP_{t-1} + θ (N(.Sch/.UnSch)_t or NV(.Sch/.UnSch)_t) + ε_t

Depvar RV*t*	**Intercept**	**RV_{t-1}**	**IV_{t-1}**	**VP_{t-1}**		**R^2**
					N.Sch_t	
(1)	0.089 (30.33)	0.151 (26.39) [++19%,+50%,-29%, - -2%]	0.543 (72.69) [++48%,+43%,-8%, - -0%]	-0.015 (-2.96)	0.002 (31.25) [++23%,+62%,-15%, - -0%]	38.3%
					N.UnSch_t	
(2)	0.063 (22.03)	0.130 (23.88) [++17%,+49%,-32%, - -2%]	0.561 (77.73) [++53%,+39%,-8%, - -0%]	-0.016 (-3.42)	0.015 (42.23) [++30%,+60%,-10%, - -0%]	41.1%
					NV.Sch_t	
(3)	0.094 (31.40)	0.133 (23.37) [++18%,+49%,-31%, - -2%]	0.554 (74.32) [++51%,+41%,-8%, - -0%]	-0.026 (-5.17)	0.031 (35.97) [++21%,+63%,-15%, - -0%]	38.2%
					NV.UnSch_t	
(4)	0.081 (27.52)	0.110 (19.81) [++16%,+48%,-34%, - -2%]	0.577 (78.65) [++54%,+38%,-7%, - -0%]	-0.029 (-6.17)	0.090 (36.66) [++26%,+64%,-10%, - -0%]	0.402

Table 13: The Strength of the Scheduled/Unscheduled News Channel

This table quantifies the strength of the scheduled(unscheduled) news channel through which the implied volatility at the end of the month $t-1$, $\text{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\text{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. Panel A summarizes the channel of News Intensity (N), and Panel B shows the distribution of the News Volatility (NV). The detail calculation involves a two-stage mediation analysis for each firm i as follows:

$$\begin{cases} \text{Total Effect:} & \text{RV}_{i,t} = \text{Int}_i + \beta_i \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.Sch):} & \text{X.Sch}_{i,t} = \text{Int}_i + \theta_{1,i} \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.UnSch):} & \text{X.UnSch}_{i,t} = \text{Int}_i + \theta_{2,i} \text{IV}_{i,t-1} + \text{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 2:} & \text{RV}_{i,t} = \text{Int}_i + \beta_i' \text{IV}_{i,t-1} + \phi_{1,i}\text{X.Sch}_{i,t} + \phi_{2,i}\text{X.UnSch}_{i,t} + \text{Controls}_{i,t} + \varepsilon_{i,t} \end{cases}$$

where X can be substitude as N or NV, and the controls are $\text{RV}_{i,t-1}$ and $\text{VP}_{i,t-1}$. β_i represents the total predictability, while β_i' represents the non-news-channel predictability. We can implement the two-stage regressions to acquire the mediator effect of the scheduled-news-channel predictability as $\theta_{1,i}\phi_{1,i}$ and the mediator effect of the unscheduled-news-channel predictability as $\theta_{2,i}\phi_{2,i}$. In these mediation models, we have total equality: $\beta_i = \beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i}$. Therefore, the strength of the scheduled-news and unscheduled-news channel can be quantified as $\theta_{1,i}\phi_{1,i}/(\beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i})$ and $\theta_{2,i}\phi_{2,i}/(\beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i})$ respectively. In specific, I first filter out stocks with negative total predictability. Then, for each stock, through iteration, I select the set of mediators to ensure the total predictability can be decomposed into non-negative components (i.e., different news and non-news channel predictability) and that it is the best combination with the highest R^2 for stage 2. The unselected mediator effects are set to zero. The sample period is from 2004 to 2017.

	Panel A: News Intensity (N)							
	mean	std	skew	p5	p25	p50	p75	p95
N.Sch	10.1%	17.2%	2.769	0.0%	0.0%	2.5%	13.2%	44.6%
N.UnSch	9.2%	17.4%	3.152	0.0%	0.0%	1.7%	10.4%	44.0%
sum	19.3%	23.6%	1.877	0.0%	2.7%	11.1%	25.9%	75.1%
	Panel B: News Volatility (NV)							
	mean	std	skew	p5	p25	p50	p75	p95
NV.Sch	7.5%	15.4%	3.769	0.0%	0.0%	1.4%	7.9%	35.9%
NV.UnSch	9.7%	16.2%	3.019	0.0%	0.0%	3.4%	12.4%	40.2%
sum	17.1%	22.1%	2.203	0.0%	2.5%	9.7%	21.5%	67.2%

Table 14: News Summary Based on News Format Classification

This table summarizes the news frequency and news sentiment grouped by news formats: news format (NF), full-article (FA), press-releases (PR), tabular-material (TM) and SEC filings (SEC). The ① row counts the number of the type of news stories and the square bracket below shows the percentage of the type of news stories over the total 7,418,062 news stories in the sample. The ② row counts the number of firm-month that contain the type of news and the square bracket below shows the percentage of the type of firm-month over the total 363,456 firm-months in the sample. ③ and ④ show the average Composite Sentiment Score (CSS) and Event Sentiment Score (ESS) respectively. ⑤ measures News Impact Projection. The round bracket shows the T-statistics of the average's difference from zero. The sample period is from 2004 to 2017.

		NF	FA	PR	TM	SEC
①	# News Stories	3,524,328	2,328,327	1,394,111	19,324	151,972
		[47.51%]	[31.39%]	[18.79%]	[0.26%]	[2.05%]
②	# Firm-Month	232,507	266,814	240,952	10,894	124,079
		[63.97%]	[73.41%]	[66.29%]	[3.00%]	[34.14%]
③	CSS	-0.010	0.003	0.038	0.022	-0.007
		(-193.76)	(38.99)	(750.94)	(25.85)	(-44.91)
④	ESS	0.141	0.128	0.128	0.113	-0.095
		(520.71)	(430.82)	(468.78)	(43.67)	(-139.10)
⑤	NIP	0.012	-0.029	-0.032	-0.000	-0.253
		(161.04)	(-173.02)	(-191.41)	(-0.18)	(-532.40)

Table 15: RV Relevance and IV Predictability in Different News Formats

This table studies both realized volatility relevance and implied volatility predictability in the classification of different news format: News Flash (NF), Full Article (FA), Press Release (PR), Tabular Materials (TM), and SEC filings (SEC). For each individual stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current news intensity/volatility(N_t/ NV_t), and then the news intensity/volatility(N_t/NV_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value $\leq$ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: News Intensity (N)						
	Model 1: $RV_t = \alpha_1 + \beta X_t + \varepsilon_t$			**Model 2: $X_t = \alpha_2 + \gamma IV_{t-1} + \varepsilon_t$**		
The X	α_1	X_t	R^2	α_2	IV_{t-1}	R^2
N.NF	0.390	0.004	5.7%	3.503	14.064	3.2%
	(164.71)	(32.81)		(23.96)	(34.18)	
		[++33%,+58%,-8%, - -0%]			[++14%,+64%,-20%, - -2%]	
N.PR	0.392	0.008	4.9%	2.499	3.011	2.7%
	(163.41)	(29.64)		(50.32)	(24.49)	
		[++24%,+59%,-16%, - -1%]			[++11%,+55%,-31%, - -3%]	
N.FA	0.390	0.010	5.8%	5.200	0.686	3.3%
	(171.15)	(28.83)		(34.15)	(2.18)	
		[++19%,+53%,-27%, - -2%]			[++8%,+42%,-46%, - -4%]	
N.TM	0.415	0.042	1.4%	0.009	0.098	1.9%
	(169.75)	(15.96)		(2.34)	(9.18)	
		[++25%,+37%,-22%, - -16%]			[++3%,+53%,-43%, - -1%]	
N.SEC	0.404	-0.007	2.1%	0.431	0.059	1.9%
	(130.45)	(-5.65)		(55.42)	(2.78)	
		[++3%,+39%,-53%, - -5%]			[++4%,+46%,-45%, - -5%]	
Panel B: News Volatility (NV)						
	Model 1: $RV_t = \alpha_1 + \beta X_t + \varepsilon_t$			**Model 2: $X_t = \alpha_2 + \gamma IV_{t-1} + \varepsilon_t$**		
The X	α_1	X_t	R^2	α_2	IV_{t-1}	R^2
NV.NF	0.388	0.060	6.5%	0.216	0.694	3.3%
	(164.79)	(31.43)		(25.07)	(34.21)	
		[++29%,+61%,-10%, - -0%]			[++18%,+58%,-22%, - -1%]	
NV.PR	0.398	0.087	4.1%	0.164	0.207	2.7%
	(165.08)	(31.15)		(41.78)	(20.03)	
		[++18%,+62%,-19%, - -1%]			[++10%,+52%,-34%, - -4%]	
NV.FA	0.387	0.061	6.3%	0.395	0.454	2.9%
	(167.43)	(34.18)		(49.05)	(21.84)	
		[++25%,+61%,-14%, - -1%]			[++12%,+51%,-33%, - -4%]	
NV.TM	0.415	0.186	1.1%	0.002	0.014	1.7%
	(170.04)	(12.44)		(2.33)	(7.56)	
		[++31%,+28%,-21%, - -20%]			[++2%,+51%,-46%, - -1%]	
NV.SEC	0.402	-0.043	1.5%	0.022	0.008	2.1%
	(132.84)	(-3.40)		(12.63)	(2.12)	
		[++12%,+24%,-44%, - -20%]			[++2%,+43%,-53%, - -2%]	

Table 16: The Channel of Different News Formats

This table examines whether the implied volatility can predict future volatility through different news formats: News Flash (NF), Full Article (FA), Press Release (PR), Tabular Materials (TM), and SEC filings (SEC). For each individual stock, the model runs the times series regression with the realized volatilities (RV_t) on the past volatilities (RV_{t-1}), the past implied volatility (IV_{t-1}) and the past volatiltiy risk premium (VP_{t-1}), plus controlling for the concurrent news intensity(N) or news volatility(NV) restricted in different news formats. The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentages of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative insignificant (significant if p-value ≤ 0.01, and the standard errors are newey-west adjusted with 12 lags). The sample period is from 2004 to 2017.

Panel A: News Intensity (N)

Model: RV_t = Intercept + β_1 RV_{t-1} + β_2 IV_{t-1} + β_3 VP_{t-1} +θ N.News_Type$_t$ + ε_t

Depvar RVt	**Intercept**	**RV$_{t-1}$**	**IV$_{t-1}$**	**VP$_{t-1}$**		**R^2**
					N.NF$_t$	
(1)	0.091	0.154	0.534	-0.012	0.003	0.386
	(30.90)	(27.06)	(72.01)	(-2.50)	(30.20)	
		[++19%,+50%,-29%, - -1%]	[++48%,+44%,-8%, - -0%]		[++26%,+62%,-11%, - -0%]	
					N.PR$_t$	
(2)	0.086	0.139	0.557	-0.020	0.008	0.381
	(29.01)	(24.52)	(75.24)	(-3.96)	(28.61)	
		[++18%,+50%,-30%, - -2%]	[++51%,+41%,-7%, - -0%]		[++22%,+63%,-15%, - -1%]	
					N.FA$_t$	
(3)	0.082	0.126	0.569	-0.022	0.010	0.389
	(28.16)	(22.63)	(77.07)	(-4.55)	(30.18)	
		[++17%,+49%,-33%, - -2%]	[++53%,+39%,-8%, - -0%]		[++19%,+64%,-16%, - -1%]	
					N.TM$_t$	
(4)	0.110	0.080	0.611	-0.051	0.022	0.352
	(36.12)	(13.86)	(81.28)	(-10.35)	(9.57)	
		[++23%,+36%,-31%, - -10%]	[++64%,+30%,-5%, - -1%]		[++23%,+36%,-23%, - -18%]	
					N.SEC$_t$	
(5)	0.104	0.073	0.641	-0.053	-0.008	0.371
	(26.10)	(9.26)	(64.48)	(-7.68)	(-6.44)	
		[++14%,+43%,-39%, - -4%]	[++55%,+37%,-7%, - -1%]		[++2%,+37%,-56%, - -5%]	

Panel B: News Volatility (NV)

Model: RV_t = Intercept + β_1 RV_{t-1} + β_2 IV_{t-1} + β_3 VP_{t-1} +θ NV.News_Type$_t$ + ε_t

Depvar RVt	**Intercept**	**RV$_{t-1}$**	**IV$_{t-1}$**	**VP$_{t-1}$**		**R^2**
					NV.NF$_t$	
(6)	0.096	0.123	0.558	-0.029	0.046	0.387
	(32.25)	(22.01)	(75.28)	(-5.94)	(29.27)	
		[++17%,+49%,-33%, - -2%]	[++52%,+40%,-8%, - -0%]		[++22%,+65%,-12%, - -1%]	
					NV.PR$_t$	
(7)	0.095	0.112	0.580	-0.034	0.077	0.373
	(31.66)	(19.67)	(77.71)	(-6.96)	(28.92)	
		[++17%,+47%,-34%, - -2%]	[++53%,+39%,-8%, - -0%]		[++15%,+66%,-19%, - -1%]	
					NV.FA$_t$	
(8)	0.087	0.122	0.568	-0.026	0.056	0.39
	(29.31)	(21.92)	(77.31)	(-5.31)	(31.81)	
		[++17%,+48%,-32%, - -2%]	[++52%,+40%,-7%, - -0%]		[++23%,+64%,-12%, - -1%]	
					NV.TM$_t$	
(9)	0.111	0.079	0.612	-0.052	0.098	0.35
	(36.07)	(13.60)	(81.22)	(-10.48)	(8.19)	
		[++24%,+35%,-31%, - -10%]	[++63%,+30%,-6%, - -1%]		[++29%,+28%,-22%, - -22%]	
					NV.SEC$_t$	
(10)	0.101	0.071	0.645	-0.053	-0.020	0.364
	(25.90)	(9.28)	(65.26)	(-7.93)	(-1.76)	
		[++19%,+38%,-36%, - -7%]	[++59%,+34%,-6%, - -1%]		[++11%,+25%,-45%, - -20%]	

Table 17: The Strength of the Channel of Different News Format

This table quantifies the strength of the channel of different news format through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. Panel A summarizes the channel of News Intensity (N), and Panel B shows the distribution of the News Volatility (NV). The detail calculation involves a two-stage mediation analysis for each firm i as follows:

$$\begin{cases} \text{Total Effect:} & \mathrm{RV}_{i,t} = \mathrm{Int}_i + \beta_i \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.NF):} & \mathrm{X.NF}_{i,t} = \mathrm{Int}_i + \theta_{1,i} \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.FA):} & \mathrm{X.FA}_{i,t} = \mathrm{Int}_i + \theta_{2,i} \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.PR):} & \mathrm{X.PR}_{i,t} = \mathrm{Int}_i + \theta_{3,i} \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.SEC):} & \mathrm{X.SEC}_{i,t} = \mathrm{Int}_i + \theta_{4,i} \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 1 (X.TM):} & \mathrm{X.TM}_{i,t} = \mathrm{Int}_i + \theta_{5,i} \mathrm{IV}_{i,t-1} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \\ \text{Stage 2:} & \mathrm{RV}_{i,t} = \mathrm{Int}_i + \beta_i' \mathrm{IV}_{i,t-1} + \phi_{1,i}\mathrm{X.NF}_{i,t} + \phi_{2,i}\mathrm{X.FA}_{i,t} \\ & +\phi_{3,i}\mathrm{X.PR}_{i,t} + \phi_{4,i}\mathrm{X.SEC}_{i,t} + \phi_{5,i}\mathrm{X.TM}_{i,t} + \mathrm{Controls}_{i,t} + \varepsilon_{i,t} \end{cases}$$

where X can be substitude as N or NV, and the controls are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. β_i represents the total predictability, while β_i' represents the non-news-channel predictability. We can implement the two-stage regressions to acquire the predictability of the news mediators: $\theta_{1,i}\phi_{1,i}$ for news-flash (NF), $\theta_{2,i}\phi_{2,i}$ for full-article (FA), $\theta_{3,i}\phi_{3,i}$ for press-release (PR), $\theta_{4,i}\phi_{4,i}$ for SEC filings (SEC) and $\theta_{5,i}\phi_{5,i}$ for tabular-materials (TM) respectively. In these mediation models, we have total equality: $\beta_i = \beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i} + \theta_{3,i}\phi_{3,i} + \theta_{4,i}\phi_{4,i} + \theta_{5,i}\phi_{5,i}$. Therefore, the strength of the channel of different news format can be quantified as the predictability of a news format over the total predictability (i.e., $\beta_i' + \theta_{1,i}\phi_{1,i} + \theta_{2,i}\phi_{2,i} + \theta_{3,i}\phi_{3,i} + \theta_{4,i}\phi_{4,i} + \theta_{5,i}\phi_{5,i}$) respectively. In specific, I first filter out stocks with negative total predictability. Then, for each stock, through iteration, I select the set of mediators to ensure the total predictability can be decomposed into non-negative components (i.e., different news and non-news channel predictability) and that it is the best combination with the highest R^2 for stage 2. The unselected mediator effects are set to zero. The period spans from 2004 to 2017.

Panel A: News Intensity (N)

	mean	std	skew	p5	p25	p50	p75	p95
N.NF	8.9%	16.9%	2.949	0.0%	0.0%	0.8%	10.5%	44.2%
N.FA	6.1%	15.0%	3.880	0.0%	0.0%	0.0%	4.7%	34.8%
N.PR	5.8%	13.2%	3.933	0.0%	0.0%	0.0%	5.9%	29.9%
N.SEC	1.6%	7.9%	8.784	0.0%	0.0%	0.0%	0.0%	7.4%
N.TM	1.0%	5.1%	10.953	0.0%	0.0%	0.0%	0.0%	5.0%
sum	23.6%	27.0%	1.601	0.0%	4.3%	13.7%	31.6%	100.0%

Panel B: News Volatility (NV)

	mean	std	skew	p5	p25	p50	p75	p95
NV.NF	8.0%	15.7%	3.485	0.0%	0.0%	1.7%	8.8%	37.0%
NV.FA	6.8%	13.9%	3.826	0.0%	0.0%	1.0%	7.8%	30.4%
NV.PR	5.4%	13.3%	4.527	0.0%	0.0%	0.0%	4.6%	26.4%
NV.SEC	0.9%	5.6%	10.628	0.0%	0.0%	0.0%	0.0%	2.7%
NV.TM	0.9%	5.3%	12.145	0.0%	0.0%	0.0%	0.0%	3.4%
sum	21.9%	25.7%	1.808	0.0%	4.4%	13.2%	27.5%	100.0%

Table 18: Summary Statistics of Different News Content Groups

This table summarizes the statistics of different news content group. The group is defined by RavenPack Taxonomy as in column (2). Column (1) denotes its short-form as G-ID for easy identification. Column (3) is the total number of news stories in a group. Column (4) is the percentage of scheduled news within a group. Column (5)-(7) shows the distribution of Composite Sentiment Scores(CSS). The last column depicts the projected news impact(NIP) within the next two hours. The sample period spans from 2004 to 2017.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	G-ID	GROUP	FREQ	Sch %	CSS					NIP
					avg	std	skew	min	max	
1	E	earnings	3,907,364	98.0%	0.00	0.11	-4.11	-0.92	1.00	0.03
2	R	revenues	1,147,321	86.0%	0.02	0.10	-1.56	-0.92	1.00	0.05
3	AR	analyst-ratings	392,618	0.0%	-0.03	0.17	-1.18	-0.92	1.00	-0.08
4	D	dividends	353,108	98.3%	0.04	0.06	-2.11	-0.92	1.00	0.04
5	EA	equity-actions	317,156	34.1%	0.01	0.07	-1.91	-0.92	1.00	-0.10
6	IR	investor-relations	292,947	66.6%	0.02	0.05	-2.27	-0.92	1.00	-0.18
7	AM	acquisitions-mergers	197,873	14.1%	0.02	0.07	-2.34	-0.92	1.00	-0.16
8	PS	products-services	166,306	18.0%	0.02	0.08	-0.69	-0.92	1.00	-0.20
9	IT	insider-trading	154,113	0.0%	-0.01	0.03	-0.62	-0.62	1.00	-0.19
10	LI	labor-issues	128,002	0.0%	0.01	0.07	-2.27	-0.92	1.00	-0.20
11	M	marketing	95,824	98.4%	0.01	0.04	2.11	-0.78	1.00	-0.19
12	PT	price-targets	59,120	0.0%	0.01	0.13	-0.74	-0.92	1.00	-0.06
13	Ct	credit-ratings	52,022	0.0%	-0.05	0.16	-1.83	-0.92	1.00	-0.23
14	A	assets	47,951	5.9%	0.01	0.08	-1.22	-0.92	1.00	-0.14
15	Pn	partnerships	34,142	0.0%	0.03	0.05	0.48	-0.92	1.00	-0.28
16	L	legal	31,926	10.6%	-0.01	0.09	-4.53	-0.92	1.00	-0.21
17	C	credit	21,364	0.1%	0.01	0.07	-2.39	-0.92	1.00	-0.21
18	Sk	stock-picks	5,525	0.0%	0.06	0.08	0.80	-0.78	1.00	-0.16
19	Rg	regulatory	5,148	2.7%	-0.01	0.07	-3.26	-0.92	0.42	-0.17
20	I	indexes	1,310	0.0%	0.02	0.06	-0.27	-0.62	0.30	-0.20
21	CR	corporate-responsibility	1,064	0.0%	0.01	0.05	-1.51	-0.38	0.30	-0.23
22	IA	industrial-accidents	939	0.0%	0.00	0.07	-3.34	-0.66	0.30	-0.21
23	WC	war-conflict	740	0.0%	0.01	0.06	-1.77	-0.38	0.14	-0.21
24	Ex	exploration	718	0.0%	0.03	0.08	-3.59	-0.78	0.30	-0.15
25	S	security	702	0.0%	0.00	0.07	-5.01	-0.92	0.14	-0.19
26	Cm	crime	550	0.0%	0.01	0.06	1.05	-0.24	0.56	-0.20
27	B	bankruptcy	470	0.0%	0.01	0.08	-0.76	-0.62	0.56	-0.16
28	Bp	balance-of-payments	457	70.0%	0.04	0.11	-2.06	-0.66	0.30	-0.15
29	T	transportation	456	0.0%	-0.01	0.06	-0.54	-0.24	0.30	-0.15
30	CU	civil-unrest	390	0.0%	0.01	0.07	-0.93	-0.66	0.56	-0.16

Table 19: News Intensity Regressions in Different News Content Groups

This table studies both realized volatility relevance and implied volatility predictability in the classification of different news content group using the measures of News Intensity. For each individual stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current news intensity(N_t), and then the news intensity(N_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis shows the T-statistics of the average's difference from zero. The sample period spans from 2004 to 2017.

	G id	GROUP	$RV_t = a + \beta N_t + \varepsilon_t$			$N_t = a + \beta IV_{t-1} + \varepsilon_t$			N(firm)	N(month)
			a	N_t	R^2	a	IV_{t-1}	R^2		
1	E	earnings	0.394	0.002	5.03%	4.596	13.697	2.92%	3855	94.3
			(164.79)	(29.13)		(28.64)	(29.30)			
2	R	revenues	0.393	0.009	5.34%	1.498	3.528	2.69%	3849	94.4
			(163.78)	(27.10)		(19.24)	(20.59)			
3	AR	analyst-ratings	0.391	0.033	6.61%	0.846	0.361	2.35%	3822	94.8
			(170.02)	(29.93)		(42.08)	(7.90)			
4	D	dividends	0.359	0.025	2.42%	1.306	0.102	1.91%	2306	108.4
			(134.18)	(7.13)		(32.67)	(0.92)			
5	EA	equity-actions	0.403	0.031	4.15%	0.843	-0.228	2.89%	3850	94.4
			(173.56)	(26.07)		(29.90)	(-3.59)			
6	IR	investor-relations	0.406	0.019	3.10%	0.623	0.384	2.08%	3835	94.5
			(166.42)	(19.31)		(39.33)	(11.24)			
7	AM	acquisitions-mergers	0.408	0.026	4.60%	0.605	-0.303	2.38%	3796	95.1
			(169.76)	(14.60)		(26.95)	(-6.12)			
8	PS	products-services	0.411	0.034	2.63%	0.370	0.076	1.76%	3789	95.2
			(170.62)	(20.21)		(15.92)	(2.00)			
9	IT	insider-trading	0.415	0.004	1.92%	0.427	-0.044	1.74%	3840	94.4
			(171.67)	(3.78)		(55.77)	(-2.27)			
10	LI	labor-issues	0.413	0.011	1.95%	0.270	0.128	1.78%	3793	95.2
			(168.95)	(9.59)		(33.46)	(7.00)			
11	M	marketing	0.423	0.004	1.62%	0.274	0.022	1.66%	3415	98.3
			(161.95)	(2.00)		(14.74)	(1.14)			
12	PT	price-targets	0.405	0.057	3.57%	0.134	0.056	1.95%	3369	101.0
			(165.62)	(19.26)		(21.40)	(3.79)			
13	Ct	credit-ratings	0.367	0.034	3.08%	0.130	0.236	2.09%	2143	110.1
			(127.36)	(10.86)		(21.61)	(11.33)			
14	A	assets	0.410	0.033	2.07%	0.114	0.026	1.73%	3549	98.3
			(164.99)	(12.39)		(20.65)	(2.28)			
15	Pn	partnerships	0.413	0.026	1.84%	0.102	-0.013	1.42%	3141	101.6
			(157.55)	(8.48)		(17.23)	(-1.37)			

Table 19: News Intensity Regressions in Different News Content Groups (Continue)

	G id	GROUP	$RV_t = a + \beta N_t + \varepsilon_t$ a	N_t	R^2	$N_t = a + \beta IV_{t-1} + \varepsilon_t$ a	IV_{t-1}	R^2	N(firm)	N(month)
16	L	legal	0.409	0.093	4.06%	0.087	0.016	1.79%	2921	102.5
			(152.89)	(18.10)		(10.66)	(1.17)			
17	C	credit	0.412	0.040	1.89%	0.051	0.051	1.59%	3155	101.4
			(150.21)	(12.82)		(14.24)	(5.83)			
18	Sk	stock-picks	0.393	0.025	1.54%	0.028	-0.011	1.29%	2149	115.3
			(130.75)	(4.52)		(15.99)	(-2.56)			
19	Rg	regulatory	0.408	0.070	2.30%	0.027	0.018	1.80%	1236	117.0
			(95.50)	(7.50)		(8.48)	(1.56)			
20	I	indexes	0.409	0.031	1.34%	0.014	0.009	1.34%	1042	114.1
			(90.66)	(3.62)		(7.18)	(1.66)			
21	CR	corporate-responsibility	0.341	-0.014	0.87%	0.017	-0.014	0.90%	677	133.2
			(77.09)	(-1.75)		(10.42)	(-2.93)			
22	IA	industrial-accidents	0.354	0.012	1.30%	0.023	-0.009	0.74%	361	134.3
			(50.36)	(1.05)		(6.77)	(-0.76)			
23	WC	war-conflict	0.331	0.017	1.09%	0.021	0.079	0.83%	190	141.5
			(36.70)	(1.03)		(1.23)	(0.78)			
24	Ex	exploration	0.469	0.048	1.67%	0.025	0.005	1.65%	216	116.3
			(41.04)	(2.68)		(4.65)	(0.45)			
25	S	security	0.329	0.011	0.98%	0.017	0.018	0.80%	281	138.6
			(47.60)	(0.96)		(1.51)	(0.39)			
26	Cm	crime	0.332	-0.030	0.66%	0.027	-0.021	0.71%	203	136.5
			(42.31)	(-3.27)		(5.30)	(-1.27)			
27	B	bankruptcy	0.402	0.143	2.91%	0.007	0.003	1.61%	361	122.7
			(49.29)	(5.28)		(2.99)	(0.50)			
28	Bp	balance-of-payments	0.363	0.001	0.62%	0.028	0.006	0.85%	144	137.0
			(30.70)	(0.05)		(3.20)	(0.34)			
29	T	transportation	0.388	0.003	0.88%	0.041	0.105	0.92%	32	134.1
			(15.34)	(0.10)		(0.84)	(0.88)			
30	CU	civil-unrest	0.358	0.021	1.07%	0.017	-0.002	1.15%	177	136.2
			(32.43)	(1.09)		(4.44)	(-0.18)			

Table 20: News Volatility Regressions in Different News Content Groups

This table studies both realized volatility relevance and implied volatility predictability in the classification of different news content group using the measures of News Volatility. For each stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current month's news volatility(NV_t), and then with the news volatility(NV_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis shows the T-statistics of the average's difference from zero. The period is from 2004 to 2017.

	G id	GROUP	$RV_t = a + \beta\ NV_t + \varepsilon_t$			$NV_t = a + \beta\ IV_{t-1} + \varepsilon_t$			N(firm)	N(month)
			a	NV_t	R^2	a	IV_{t-1}	R^2		
1	E	earnings	0.393	0.040	5.57%	0.238	0.865	3.20%	3851	94.3
			(164.94)	(38.18)		(24.39)	(34.52)			
2	R	revenues	0.399	0.073	4.62%	0.140	0.267	2.38%	3793	95.1
			(166.05)	(16.27)		(26.98)	(20.55)			
3	AR	analyst-ratings	0.394	0.106	6.56%	0.208	0.094	2.20%	3813	94.9
			(170.93)	(23.59)		(43.91)	(8.79)			
4	D	dividends	0.350	0.077	2.17%	0.108	0.018	1.81%	2064	110.8
			(127.90)	(4.32)		(32.04)	(1.90)			
5	EA	equity-actions	0.411	0.110	2.65%	0.074	-0.019	2.47%	3820	94.8
			(171.66)	(14.82)		(36.32)	(-4.00)			
6	IR	investor-relations	0.410	0.151	3.20%	0.040	0.045	2.36%	3717	95.9
			(167.20)	(13.27)		(27.86)	(12.77)			
7	AM	acquisitions-mergers	0.407	0.203	3.80%	0.051	-0.018	2.13%	3734	96.0
			(168.78)	(11.42)		(31.76)	(-4.44)			
8	PS	products-services	0.414	0.171	2.57%	0.035	0.021	1.72%	3683	96.6
			(169.31)	(15.29)		(23.64)	(7.22)			
9	IT	insider-trading	0.414	0.096	1.88%	0.021	0.005	1.68%	3802	95.0
			(170.90)	(8.45)		(37.48)	(3.29)			
10	LI	labor-issues	0.413	0.094	2.05%	0.024	0.025	1.73%	3659	96.9
			(167.46)	(10.04)		(24.01)	(9.65)			
11	M	marketing	0.424	0.037	1.57%	0.013	0.005	1.51%	3005	101.7
			(154.63)	(1.86)		(18.24)	(3.63)			
12	PT	price-targets	0.405	0.151	3.56%	0.035	0.030	1.85%	3273	102.2
			(162.64)	(16.34)		(21.02)	(7.47)			
13	Ct	credit-ratings	0.363	0.108	3.43%	0.017	0.112	2.71%	1995	112.4
			(124.74)	(7.85)		(6.91)	(13.50)			
14	A	assets	0.407	0.134	2.01%	0.015	0.008	1.60%	3215	101.9
			(158.74)	(8.53)		(15.64)	(3.40)			
15	Pn	partnerships	0.412	0.121	1.65%	0.013	-0.000	1.38%	2713	105.3
			(147.52)	(5.37)		(17.00)	(-0.11)			

Table 20: News Volatility Regressions in Different News Content Groups (Continue)

	G id	GROUP	$RV_t = a + \beta NV_t + \varepsilon_t$			$NV_t = a + \beta IV_{t-1} + \varepsilon_t$				
			a	NV_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
16	L	legal	0.406	0.487	3.06%	0.010	0.007	1.64%	2298	107.8
			(133.50)	(10.82)		(9.88)	(3.08)			
17	C	credit	0.397	0.168	1.67%	0.006	0.010	1.50%	2387	109.8
			(132.88)	(7.65)		(6.99)	(3.73)			
18	Sk	stock-picks	0.391	0.081	1.47%	0.007	-0.004	1.19%	1952	116.2
			(123.77)	(2.44)		(13.61)	(-3.08)			
19	Rg	regulatory	0.396	0.274	2.33%	0.004	0.006	1.69%	789	121.5
			(73.89)	(4.24)		(4.42)	(2.30)			
20	I	indexes	0.425	0.292	1.54%	0.001	0.005	1.44%	555	114.2
			(62.65)	(2.71)		(1.49)	(2.23)			
21	CR	corporate-responsibility	0.337	-0.104	0.74%	0.003	-0.002	0.82%	374	135.5
			(55.22)	(-1.91)		(4.96)	(-1.03)			
22	IA	industrial-accidents	0.341	0.017	1.06%	0.006	-0.006	0.82%	223	136.2
			(38.43)	(0.21)		(4.65)	(-2.24)			
23	WC	war-conflict	0.319	-0.039	0.94%	0.006	0.001	0.48%	128	145.2
			(30.95)	(-0.41)		(3.96)	(0.23)			
24	Ex	exploration	0.479	0.058	1.87%	0.006	-0.001	1.27%	162	115.1
			(35.31)	(0.68)		(3.24)	(-0.23)			
25	S	security	0.322	0.116	1.27%	0.004	-0.002	0.73%	168	141.2
			(38.93)	(1.15)		(2.92)	(-0.38)			
26	Cm	crime	0.329	-0.161	0.90%	0.002	0.005	0.81%	123	134.3
			(30.72)	(-2.08)		(1.30)	(0.73)			
27	B	bankruptcy	0.408	0.656	3.37%	0.001	0.003	1.03%	206	121.7
			(36.27)	(3.17)		(1.39)	(1.30)			
28	Bp	balance-of-payments	0.361	-0.029	0.58%	0.009	0.003	0.88%	114	138.8
			(26.03)	(-0.39)		(2.42)	(0.57)			
29	T	transportation	0.410	0.052	0.79%	0.001	0.013	0.96%	24	133.9
			(13.59)	(0.40)		(0.20)	(1.04)			
30	CU	civil-unrest	0.352	-0.001	0.64%	0.005	-0.004	0.87%	107	136.5
			(24.94)	(-0.01)		(3.66)	(-0.99)			

Table 21: The Strength of News Intensity Channel of Different News Content

This table quantifies the strength of the news intensity channel of different news contents through which the implied volatility at the end of the month $t-1$, $\text{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\text{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\text{RV}_{i,t-1}$ and $\text{VP}_{i,t-1}$. Panel A separates the 30 news content groups into three mediation analysis test groups. Panel B runs a final test with a set of selected news content groups whose average strength is not less than 0.09% in Panel A. The sample period is from 2004 to 2017.

Panel A: Test Groups

	mean	std	skew	p5	p25	p50	p75	p95
				Test Group 1				
N.R	6.6%	13.3%	3.248	0.0%	0.0%	0.0%	7.2%	33.7%
N.E	4.6%	11.3%	3.818	0.0%	0.0%	0.0%	2.8%	27.4%
N.AR	4.5%	10.1%	4.171	0.0%	0.0%	0.2%	4.5%	23.0%
N.IR	3.0%	8.8%	5.760	0.0%	0.0%	0.0%	1.7%	16.2%
N.EA	2.6%	8.4%	6.173	0.0%	0.0%	0.0%	0.7%	14.3%
N.AM	1.9%	7.4%	7.429	0.0%	0.0%	0.0%	0.1%	10.4%
N.PS	1.8%	7.3%	8.363	0.0%	0.0%	0.0%	0.4%	8.5%
N.LI	1.4%	5.9%	8.179	0.0%	0.0%	0.0%	0.1%	7.2%
N.IT	1.2%	6.4%	9.260	0.0%	0.0%	0.0%	0.0%	5.0%
N.D	0.8%	4.9%	11.208	0.0%	0.0%	0.0%	0.0%	3.4%
				Test Group 2				
N.PT	3.1%	9.1%	5.527	0.0%	0.0%	0.0%	1.8%	16.6%
N.L	1.8%	8.0%	7.851	0.0%	0.0%	0.0%	0.0%	8.5%
N.A	1.7%	7.5%	8.250	0.0%	0.0%	0.0%	0.2%	8.3%
N.Pn	1.2%	6.5%	10.339	0.0%	0.0%	0.0%	0.0%	4.5%
N.Ct	1.1%	5.8%	9.475	0.0%	0.0%	0.0%	0.0%	4.3%
N.C	1.1%	5.4%	10.931	0.0%	0.0%	0.0%	0.0%	5.4%
N.M	1.0%	5.7%	10.540	0.0%	0.0%	0.0%	0.0%	4.4%
N.Sk	0.7%	4.4%	10.619	0.0%	0.0%	0.0%	0.0%	2.1%
N.Rg	0.5%	3.9%	15.190	0.0%	0.0%	0.0%	0.0%	0.5%
N.I	0.3%	3.2%	20.380	0.0%	0.0%	0.0%	0.0%	0.1%
				Test Group 3				
N.B	0.2%	3.3%	19.943	0.0%	0.0%	0.0%	0.0%	0.0%
N.Ex	0.1%	1.8%	23.623	0.0%	0.0%	0.0%	0.0%	0.0%
N.CR	0.1%	1.7%	53.309	0.0%	0.0%	0.0%	0.0%	0.0%
N.IA	0.1%	1.2%	28.282	0.0%	0.0%	0.0%	0.0%	0.0%
N.S	0.0%	0.9%	36.872	0.0%	0.0%	0.0%	0.0%	0.0%
N.CU	0.0%	0.5%	34.566	0.0%	0.0%	0.0%	0.0%	0.0%
N.WC	0.0%	0.4%	36.737	0.0%	0.0%	0.0%	0.0%	0.0%
N.Bp	0.0%	0.4%	26.165	0.0%	0.0%	0.0%	0.0%	0.0%
N.Cm	0.0%	0.2%	26.072	0.0%	0.0%	0.0%	0.0%	0.0%
N.T	0.0%	0.2%	54.662	0.0%	0.0%	0.0%	0.0%	0.0%

Table 21: The Strength of News Intensity Channel of Different News Content (Continue)

	Panel B: Top 14 News Content Groups							
	mean	std	skew	p5	p25	p50	p75	p95
N.R	6.2%	12.4%	3.173	0.0%	0.0%	0.0%	6.8%	32.0%
N.E	4.4%	10.9%	3.891	0.0%	0.0%	0.0%	2.6%	26.0%
N.AR	4.1%	9.3%	4.293	0.0%	0.0%	0.0%	3.8%	21.2%
N.IR	2.8%	8.2%	5.545	0.0%	0.0%	0.0%	1.6%	15.6%
N.EA	2.3%	7.7%	6.388	0.0%	0.0%	0.0%	0.4%	12.6%
N.PT	1.9%	6.6%	7.000	0.0%	0.0%	0.0%	0.3%	10.5%
N.AM	1.6%	6.6%	7.689	0.0%	0.0%	0.0%	0.0%	9.3%
N.PS	1.5%	6.2%	8.430	0.0%	0.0%	0.0%	0.3%	7.9%
N.LI	1.4%	5.7%	8.320	0.0%	0.0%	0.0%	0.1%	7.0%
N.L	1.3%	6.1%	8.282	0.0%	0.0%	0.0%	0.0%	6.4%
N.A	1.2%	5.8%	9.919	0.0%	0.0%	0.0%	0.0%	6.4%
N.IT	1.1%	6.0%	9.424	0.0%	0.0%	0.0%	0.0%	4.7%
N.Pn	0.9%	4.7%	11.426	0.0%	0.0%	0.0%	0.0%	3.7%
N.Ct	0.8%	4.3%	10.787	0.0%	0.0%	0.0%	0.0%	3.9%
sum	31.4%	28.2%	1.201	2.1%	10.5%	22.6%	42.7%	100.0%

Table 22: The Strength of News Volatility Channel of Different News Content

This table quantifies the strength of the news volatility channel of different news contents through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. Panel A separates the 30 news content groups into three mediation analysis test groups. Panel B runs a final test with a set of selected news content groups whose average strength is not less than 0.09% in Panel A. The sample period is from 2004 to 2017.

Panel A: Test Groups

	mean	std	skew	p5	p25	p50	p75	p95
			Test Group 1					
NV.E	6.2%	12.1%	3.395	0.0%	0.0%	0.5%	7.5%	29.5%
NV.R	4.7%	10.8%	4.282	0.0%	0.0%	0.0%	4.7%	23.3%
NV.AR	4.4%	10.9%	4.754	0.0%	0.0%	0.1%	3.9%	22.5%
NV.IR	2.9%	8.9%	5.540	0.0%	0.0%	0.0%	1.3%	15.8%
NV.PS	1.8%	7.5%	8.079	0.0%	0.0%	0.0%	0.3%	8.8%
NV.AM	1.7%	6.7%	7.480	0.0%	0.0%	0.0%	0.1%	8.9%
NV.EA	1.4%	5.6%	7.296	0.0%	0.0%	0.0%	0.0%	6.9%
NV.IT	1.3%	6.5%	9.119	0.0%	0.0%	0.0%	0.0%	7.0%
NV.LI	1.3%	5.7%	8.694	0.0%	0.0%	0.0%	0.0%	7.2%
NV.D	0.8%	4.9%	10.541	0.0%	0.0%	0.0%	0.0%	3.4%
			Test Group 2					
NV.PT	3.1%	9.5%	5.929	0.0%	0.0%	0.0%	1.7%	15.8%
NV.A	1.4%	6.7%	9.145	0.0%	0.0%	0.0%	0.0%	6.3%
NV.L	1.2%	6.9%	9.946	0.0%	0.0%	0.0%	0.0%	4.2%
NV.Ct	1.1%	5.8%	10.688	0.0%	0.0%	0.0%	0.0%	5.2%
NV.M	0.9%	4.9%	10.104	0.0%	0.0%	0.0%	0.0%	3.3%
NV.Pn	0.8%	4.8%	11.935	0.0%	0.0%	0.0%	0.0%	3.3%
NV.C	0.7%	3.9%	12.716	0.0%	0.0%	0.0%	0.0%	2.8%
NV.Sk	0.6%	4.9%	13.181	0.0%	0.0%	0.0%	0.0%	1.4%
NV.Rg	0.3%	3.4%	19.678	0.0%	0.0%	0.0%	0.0%	0.0%
NV.I	0.1%	2.2%	33.642	0.0%	0.0%	0.0%	0.0%	0.0%
			Test Group 3					
NV.B	0.1%	2.4%	23.659	0.0%	0.0%	0.0%	0.0%	0.0%
NV.Ex	0.1%	2.1%	34.416	0.0%	0.0%	0.0%	0.0%	0.0%
NV.IA	0.0%	0.9%	32.602	0.0%	0.0%	0.0%	0.0%	0.0%
NV.S	0.0%	0.9%	38.238	0.0%	0.0%	0.0%	0.0%	0.0%
NV.CR	0.0%	0.3%	23.311	0.0%	0.0%	0.0%	0.0%	0.0%
NV.CU	0.0%	0.3%	36.475	0.0%	0.0%	0.0%	0.0%	0.0%
NV.Cm	0.0%	0.1%	24.453	0.0%	0.0%	0.0%	0.0%	0.0%
NV.WC	0.0%	0.1%	30.162	0.0%	0.0%	0.0%	0.0%	0.0%
NV.Bp	0.0%	0.1%	30.036	0.0%	0.0%	0.0%	0.0%	0.0%
NV.T	0.0%	0.1%	37.351	0.0%	0.0%	0.0%	0.0%	0.0%

Table 22: The Strength of News Volatility Channel of Different News Content (Continue)

Panel B: Top 14 News Content Groups								
	mean	std	skew	p5	p25	p50	p75	p95
NV.E	5.9%	11.7%	3.432	0.0%	0.0%	0.3%	7.0%	28.6%
NV.R	4.5%	10.4%	4.396	0.0%	0.0%	0.0%	4.4%	22.8%
NV.AR	4.0%	10.1%	4.991	0.0%	0.0%	0.0%	3.4%	19.8%
NV.IR	2.8%	8.6%	5.581	0.0%	0.0%	0.0%	1.3%	15.2%
NV.PT	2.1%	7.3%	6.417	0.0%	0.0%	0.0%	0.6%	11.4%
NV.PS	1.6%	6.8%	8.359	0.0%	0.0%	0.0%	0.2%	8.2%
NV.AM	1.6%	6.5%	7.602	0.0%	0.0%	0.0%	0.1%	8.2%
NV.IT	1.3%	6.3%	9.137	0.0%	0.0%	0.0%	0.0%	6.9%
NV.EA	1.3%	5.4%	7.538	0.0%	0.0%	0.0%	0.0%	6.2%
NV.LI	1.2%	5.2%	8.631	0.0%	0.0%	0.0%	0.0%	6.3%
NV.A	1.0%	4.8%	8.808	0.0%	0.0%	0.0%	0.0%	5.0%
NV.L	0.9%	5.0%	9.649	0.0%	0.0%	0.0%	0.0%	3.1%
NV.Ct	0.8%	4.2%	11.354	0.0%	0.0%	0.0%	0.0%	4.5%
NV.M	0.8%	4.6%	10.744	0.0%	0.0%	0.0%	0.0%	2.8%
sum	29.9%	27.4%	1.317	1.9%	10.1%	20.8%	39.9%	100.0%

Table 23: Summary Statistics of Different Accounting Items

This table presents the summary statistics of different accounting items related to earnings and revenue news. The Accounting Items are aggregated and mapped from RavenPack's "Type" in the content classification as in Figure A2 and Figure A3. Column (1) denotes a short-form as AC-ID for easy identification. Column (3) is the total number of news stories using the accounting item. Column (4) is the percentage of scheduled news within a group. Column (5)-(7) shows the distribution of Composite Sentiment Scores(CSS). The last column depicts the projected news impact(NIP) within the next two hours. The sample period spans from 2004 to 2017.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Panel A: "Earnings" Group							
	AC-ID	Accounting Item	FREQ	Sch %	CSS					NIP
					avg	std	skew	min	max	
1	e	earnings	1,569,026	99.9%	0.00	0.14	-3.15	-0.92	1.00	0.06
2	eps	earnings-per-share	1,423,678	94.9%	-0.01	0.11	-4.50	-0.92	1.00	0.02
3	pe	pretax-earnings	311,037	100.0%	0.00	0.02	-28.92	-0.92	1.00	0.00
4	et	ebit	291,708	100.0%	0.00	0.01	-27.01	-0.66	0.56	0.00
5	ed	ebitda	265,182	100.0%	0.00	0.03	-8.57	-0.92	1.00	0.00
6	oe	operating-earnings	34,189	100.0%	0.01	0.14	-3.16	-0.92	1.00	0.06
7	er	earnings-revision	4,821	0.0%	-0.03	0.14	-3.22	-0.78	0.66	-0.05
8	ii	interest-income	4,465	100.0%	0.02	0.07	-2.53	-0.78	0.66	0.24
9	ea	ebita	3,258	100.0%	0.00	0.02	-12.91	-0.66	0.30	0.00
			Panel B: "Revenue" Group							
	AC-ID	Accounting Item	FREQ	Sch %	CSS					NIP
					avg	std	skew	min	max	
10	r	revenue	1,062,681	85.8%	0.02	0.10	-1.55	-0.92	1.00	0.05
11	sss	same-store-sales	77,413	89.0%	0.02	0.13	-1.50	-0.92	1.00	0.07
12	om	operating-margin	7,227	82.8%	0.03	0.08	-1.50	-0.92	0.66	0.08

Table 24: News Intensity Regressions in Different Accounting Items

This table studies both realized volatility relevance and implied volatility predictability in the classification of different accounting items related to earnings and revenue news using the measures of News Intensity. For each individual stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current news intensity(N_t), and then the news intensity(N_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis shows the T-statistics of the average's difference from zero. The sample period spans from 2004 to 2017.

Panel A: "Earnings" Group

		$RV_t = a + \beta N_t + \varepsilon_t$			$N_t = a + \beta IV_{t-1} + \varepsilon_t$				
	AC-ID	a	N_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
1	e	0.393	0.007	5.21%	2.053	4.955	2.84%	3855	94.3
		(164.52)	(37.28)		(33.09)	(26.06)			
2	eps	0.396	0.006	4.68%	1.892	4.385	2.70%	3855	94.3
		(164.94)	(23.21)		(32.86)	(27.37)			
3	pe	0.399	0.019	4.19%	0.245	1.490	3.08%	3746	95.6
		(165.51)	(29.56)		(13.51)	(29.90)			
4	et	0.400	0.020	4.22%	0.219	1.442	3.06%	3698	96.0
		(165.24)	(25.22)		(11.80)	(27.45)			
5	ed	0.403	0.019	4.03%	0.164	1.542	3.15%	3425	97.8
		(164.12)	(18.60)		(7.79)	(24.96)			
6	oe	0.394	0.047	2.42%	0.050	0.171	2.01%	2902	105.7
		(152.14)	(16.12)		(10.51)	(12.17)			
7	er	0.377	0.017	1.58%	0.018	0.024	1.31%	1552	119.8
		(117.18)	(3.19)		(5.70)	(2.59)			
8	ii	0.399	0.032	1.47%	0.033	0.024	1.40%	1227	106.5
		(98.17)	(5.64)		(9.32)	(2.61)			
9	ea	0.338	0.018	1.77%	0.045	0.161	1.86%	245	132.9
		(42.60)	(2.21)		(2.14)	(2.49)			

Panel B: "Revenue" Group

		$RV_t = a + \beta N_t + \varepsilon_t$			$N_t = a + \beta IV_{t-1} + \varepsilon_t$				
	AC-ID	a	N_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
10	r	0.393	0.009	5.32%	1.385	3.323	2.67%	3849	94.4
		(164.03)	(27.70)		(19.20)	(21.69)			
11	sss	0.370	0.022	3.97%	0.844	1.516	2.84%	472	112.7
		(70.67)	(5.16)		(6.45)	(4.41)			
12	om	0.365	0.016	1.24%	0.027	0.055	1.00%	1307	125.3
		(113.90)	(3.53)		(7.50)	(4.82)			

Table 25: News Volatility Regressions in Different Accounting Items

This table studies both realized volatility relevance and implied volatility predictability in the classification of different accounting items related to earnings and revenue news using the measures of News Volatility. For each individual stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current news volatility(NV_t), and then the news volatility(NV_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis shows the T-statistics of the average's difference from zero. The sample period spans from 2004 to 2017.

Panel A: "Earnings" Group

		$RV_t = a + \beta\ NV_t + \epsilon_t$			$NV_t = a + \beta\ IV_{t-1} + \epsilon_t$				
	AC-ID	a	NV_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
1	e	0.396	0.051	5.18%	0.170	0.639	3.01%	3849	94.4
		(165.27)	(43.01)		(23.33)	(32.41)			
2	eps	0.400	0.050	4.60%	0.118	0.445	3.03%	3841	94.4
		(168.16)	(22.95)		(17.69)	(27.40)			
3	pe	0.383	0.176	1.41%	0.006	0.014	1.29%	465	121.8
		(61.90)	(4.09)		(2.96)	(2.65)			
4	et	0.347	0.075	1.32%	0.003	0.021	0.71%	210	132.2
		(38.54)	(0.93)		(0.79)	(1.71)			
5	ed	0.404	0.135	2.20%	0.007	0.037	1.78%	1542	102.6
		(124.53)	(6.75)		(3.46)	(6.54)			
6	oe	0.394	0.164	2.18%	0.007	0.049	1.73%	2583	109.0
		(145.02)	(12.05)		(4.89)	(13.01)			
7	er	0.366	0.035	1.48%	0.005	0.005	1.19%	1233	123.9
		(104.23)	(1.31)		(4.40)	(1.53)			
8	ii	0.396	0.079	1.33%	0.006	0.003	1.13%	827	107.8
		(79.86)	(2.77)		(6.92)	(1.18)			
9	ea	0.401	0.085	1.01%	0.003	0.011	1.92%	14	109.4
		(15.48)	(0.53)		(0.27)	(0.62)			

Panel B: "Revenue" Group

		$RV_t = a + \beta\ NV_t + \epsilon_t$			$NV_t = a + \beta\ IV_{t-1} + \epsilon_t$				
	AC-ID	a	NV_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
10	r	0.400	0.074	4.51%	0.133	0.247	2.32%	3793	95.1
		(166.50)	(16.36)		(27.85)	(20.23)			
11	sss	0.368	0.111	4.44%	0.108	0.302	2.97%	440	110.9
		(68.04)	(6.17)		(6.35)	(7.40)			
12	om	0.354	0.091	1.31%	0.005	0.016	0.96%	938	130.2
		(96.40)	(3.60)		(4.60)	(4.25)			

Table 26: The Strength of News Intensity Channel of Different Accounting Items

This table quantifies the strength of the news intensity channel related to different accounting items through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. Panel A separates the earnings news into 9 different accounting items. Panel B distinguishes the revenue news into 4 accounting items. The sample period is from 2004 to 2017.

				Panel A: "Earnings" Group					
AC-ID	Accounting Items	mean	std	skew	p5	p25	p50	p75	p95
e	earnings	6.2%	13.9%	3.706	0.0%	0.0%	0.0%	5.8%	32.9%
eps	earnings-per-share	4.6%	12.3%	4.426	0.0%	0.0%	0.0%	2.6%	26.6%
pe	pretax-earnings	2.3%	8.8%	6.281	0.0%	0.0%	0.0%	0.0%	13.7%
ed	ebitda	2.2%	8.7%	6.387	0.0%	0.0%	0.0%	0.0%	13.7%
et	ebit	2.0%	8.3%	6.910	0.0%	0.0%	0.0%	0.0%	10.7%
oe	opearting-earnings	1.5%	6.1%	8.294	0.0%	0.0%	0.0%	0.0%	8.1%
er	earnings-revision	0.4%	3.5%	15.834	0.0%	0.0%	0.0%	0.0%	0.8%
ii	interest-income	0.4%	3.7%	19.032	0.0%	0.0%	0.0%	0.0%	0.5%
ea	ebita	0.1%	0.9%	22.513	0.0%	0.0%	0.0%	0.0%	0.0%
sum		19.6%	24.2%	1.837	0.0%	2.6%	10.7%	26.5%	76.7%
				Panel B: "Revenue" Group					
AC-ID	Accounting Items	mean	std	skew	p5	p25	p50	p75	p95
r	revenue	14.1%	20.4%	2.312	0.0%	0.0%	6.3%	19.2%	56.7%
sss	same-store-sales	0.5%	3.6%	12.315	0.0%	0.0%	0.0%	0.0%	0.2%
om	operating-margin	0.3%	1.8%	10.164	0.0%	0.0%	0.0%	0.0%	1.0%
sum		14.8%	20.8%	2.252	0.0%	0.1%	7.0%	20.3%	59.0%

Table 27: The Strength of News Volatility Channel of Different Accounting Items

This table quantifies the strength of the news volatility channel related to different accounting items through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. Panel A separates the earnings news into 9 different accounting items. Panel B distinguishes the revenue news into 4 accounting items. The sample period is from 2004 to 2017.

		Panel A: "Earnings" Group							
AC-ID	Accounting Items	mean	std	skew	p5	p25	p50	p75	p95
e	earnings	7.3%	13.6%	3.505	0.0%	0.0%	1.4%	9.5%	31.7%
eps	earnings-per-share	5.6%	14.4%	4.333	0.0%	0.0%	0.0%	4.3%	29.0%
oe	opearting-earnings	1.3%	6.1%	9.284	0.0%	0.0%	0.0%	0.0%	6.2%
ed	ebitda	0.8%	4.4%	11.627	0.0%	0.0%	0.0%	0.0%	3.4%
er	earnings-revision	0.3%	3.0%	17.886	0.0%	0.0%	0.0%	0.0%	0.6%
ii	interest-income	0.3%	3.5%	23.430	0.0%	0.0%	0.0%	0.0%	0.0%
pe	pretax-earnings	0.1%	1.3%	25.243	0.0%	0.0%	0.0%	0.0%	0.0%
et	ebit	0.0%	0.6%	27.891	0.0%	0.0%	0.0%	0.0%	0.0%
ea	ebita	0.0%	0.1%	42.808	0.0%	0.0%	0.0%	0.0%	0.0%
sum		15.7%	21.5%	2.294	0.0%	1.5%	8.2%	19.8%	63.2%
		Panel B: "Revenue" Group							
AC-ID	Accounting Items	mean	std	skew	p5	p25	p50	p75	p95
r	revenue	8.8%	16.3%	3.373	0.0%	0.0%	2.3%	10.6%	38.2%
sss	same-store-sales	0.5%	3.6%	12.678	0.0%	0.0%	0.0%	0.0%	0.5%
om	operating-margin	0.2%	1.9%	14.028	0.0%	0.0%	0.0%	0.0%	0.6%
sum		9.6%	16.7%	3.204	0.0%	0.0%	3.1%	12.1%	40.2%

Table 28: Summary Statistics of Different Information Formats

This table presents the summary statistics of different information formats related to earnings and revenue news. The Information Formats are aggregated and mapped from RavenPack's "Type" in Content Classification as in Figure A2 and Figure A3. Column (1) denotes a short-form as Info-ID for easy identification. Column (3) is the total number of news stories presented in the information formats. Column (4) is the percentage of scheduled news within a group. Column (5)-(7) shows the distribution of Composite Sentiment Scores(CSS). The last column depicts the projected news impact(NIP) within the next two hours. The sample period spans from 2004 to 2017.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Panel A: "Earnings" Group							
	Info-ID	Informaition Format	FREQ	Sch %	CSS					NIP
					avg	std	skew	min	max	
1	E.g	(general)	2,318,221	100.0%	-0.01	0.13	-3.63	-0.92	1.00	0.04
2	E.ex	expectations	1,163,427	100.0%	0.00	0.05	-6.87	-0.92	1.00	0.02
3	E.gd	guidance	233,208	68.2%	-0.01	0.13	-2.49	-0.92	1.00	0.04
4	E.ge	guidance-expectations	102,553	100.0%	-0.01	0.08	-3.88	-0.92	1.00	0.01
5	E.es	estimate	80,669	100.0%	0.01	0.08	-2.58	-0.92	1.00	-0.14
			Panel B: "Revenue" Group							
	Info-ID	Informaition Format	FREQ	Sch %	CSS					NIP
					avg	std	skew	min	max	
6	R.g	(general)	931,346	100.0%	0.02	0.09	-1.45	-0.92	1.00	0.04
7	R.gd	guidance	160,273	0.0%	0.00	0.11	-1.89	-0.92	1.00	0.08
8	R.v	volume	52,218	100.0%	0.03	0.12	-1.61	-0.92	1.00	0.01
9	R.es	estimate	3,484	100.0%	0.00	0.13	-2.02	-0.92	1.00	0.03

Table 29: News Intensity Regressions in Different Information Formats

This table studies both realized volatility relevance and implied volatility predictability in the classification of different information formats related to earnings and revenue news using the measures of News Intensity. For each individual stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current news intensity(N_t), and then the news intensity(N_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis shows the T-statistics of the average's difference from zero. The sample period spans from 2004 to 2017.

Panel A: "Earnings" Group

	Info-ID	$RV_t = a + \beta N_t + \epsilon_t$			$N_t = a + \beta IV_{t-1} + \epsilon_t$				
		a	N_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
1	E.g	0.394	0.004	4.94%	2.976	7.582	2.80%	3854	94.3
		(164.84)	(35.28)		(33.81)	(29.40)			
2	E.ex	0.399	0.006	4.47%	1.072	4.880	3.11%	3829	94.6
		(166.07)	(17.22)		(17.75)	(28.06)			
3	E.gd	0.400	0.029	4.48%	0.272	0.965	2.46%	3454	99.1
		(162.88)	(15.57)		(17.09)	(20.87)			
4	E.ge	0.389	0.032	3.49%	0.117	0.599	2.47%	2678	104.7
		(157.23)	(12.79)		(7.40)	(13.38)			
5	E.es	0.402	0.023	2.58%	0.250	-0.034	2.81%	3299	101.7
		(160.04)	(9.42)		(23.33)	(-1.37)			

Panel B: "Revenue" Group

	Info-ID	$RV_t = a + \beta N_t + \epsilon_t$			$N_t = a + \beta IV_{t-1} + \epsilon_t$				
		a	N_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
6	R.g	0.395	0.010	4.92%	1.287	2.702	2.57%	3848	94.4
		(164.34)	(29.95)		(21.58)	(18.75)			
7	R.gd	0.397	0.026	4.67%	0.130	0.912	2.69%	3234	100.2
		(164.26)	(14.11)		(8.34)	(21.46)			
8	R.v	0.388	0.013	1.34%	0.033	-0.030	1.32%	1741	116.9
		(128.21)	(2.73)		(12.55)	(-5.65)			
9	R.es	0.396	0.054	2.06%	0.133	0.112	1.71%	2524	109.4
		(147.56)	(13.79)		(4.72)	(5.49)			

Table 30: News Volatility Regressions in Different Information Formats

This table studies both realized volatility relevance and implied volatility predictability in the classification of different information formats related to earnings and revenue news using the measures of News Volatility. For each individual stock, the model runs the times series regressions, first with the realized volatilities(RV_t) on the current news volatility(NV_t), and then the news volatility(NV_t) on the last month's implied volatility(IV_{t-1}). The first row of each model shows the average coefficient across all firms. The parenthesis shows the T-statistics of the average's difference from zero. The sample period spans from 2004 to 2017.

Panel A: "Earnings" Group

		$RV_t = a + \beta\ NV_t + \varepsilon_t$			$NV_t = a + \beta\ IV_{t-1} + \varepsilon_t$				
	Info-ID	a	NV_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
1	E.g	0.396	0.042	4.97%	0.190	0.775	3.20%	3850	94.4
		(166.19)	(38.92)		(21.10)	(34.21)			
2	E.ex	0.407	0.078	3.36%	0.033	0.132	2.47%	3766	95.4
		(168.63)	(11.15)		(12.38)	(17.56)			
3	E.gd	0.402	0.107	4.19%	0.034	0.177	2.16%	3330	100.1
		(162.39)	(12.89)		(10.60)	(20.26)			
4	E.ge	0.382	0.114	2.29%	0.012	0.039	1.40%	2088	112.7
		(141.69)	(6.25)		(6.11)	(7.29)			
5	E.es	0.400	0.174	2.25%	0.028	0.001	2.36%	3050	105.2
		(153.44)	(10.52)		(19.81)	(0.38)			

Panel B: "Revenue" Group

		$RV_t = a + \beta\ NV_t + \varepsilon_t$			$NV_t = a + \beta\ IV_{t-1} + \varepsilon_t$				
	Info-ID	a	NV_t	R^2	a	IV_{t-1}	R^2	N(firm)	N(month)
6	R.g	0.402	0.076	4.10%	0.128	0.202	2.31%	3787	95.2
		(166.75)	(15.16)		(28.29)	(17.50)			
7	R.gd	0.401	0.110	3.69%	0.019	0.119	2.03%	3108	101.1
		(164.69)	(11.26)		(7.99)	(18.15)			
8	R.v	0.387	0.028	1.36%	0.002	0.008	1.35%	1322	120.9
		(114.57)	(0.89)		(2.65)	(3.91)			
9	R.es	0.392	0.195	1.88%	0.014	0.041	1.61%	2070	114.0
		(132.84)	(9.73)		(4.84)	(9.65)			

Table 31: The Strength of News Intensity Channel of Different Information Formats

This table quantifies the strength of the news intensity channel of different information formats through which the implied volatility at the end of the month $t-1$, $\text{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\text{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\text{RV}_{i,t-1}$ and $\text{VP}_{i,t-1}$. Panel A separates the earnings news into 9 information formats. Panel B distinguishes the revenue news into 4 information formats. The sample period is from 2004 to 2017.

Panel A: "Earnings" Group

Info-ID	Information Format	mean	std	skew	p5	p25	p50	p75	p95
E.g	(general)	6.6%	14.1%	3.637	0.0%	0.0%	0.0%	7.0%	32.4%
E.ex	expectation	5.0%	14.1%	4.203	0.0%	0.0%	0.0%	2.0%	30.3%
E.gd	guidance	4.6%	11.5%	4.593	0.0%	0.0%	0.0%	3.7%	24.2%
E.ge	guidance-expectation	1.9%	7.5%	6.601	0.0%	0.0%	0.0%	0.0%	11.0%
E.es	estimate	1.3%	5.6%	9.217	0.0%	0.0%	0.0%	0.0%	6.9%
sum		19.4%	24.0%	1.884	0.0%	2.4%	11.0%	26.0%	79.4%

Panel B: "Revenue" Group

Info-ID	Information Format	mean	std	skew	p5	p25	p50	p75	p95
R.g	(general)	10.2%	17.4%	2.893	0.0%	0.0%	3.1%	12.8%	46.5%
R.gd	guidance	5.2%	13.1%	4.279	0.0%	0.0%	0.0%	3.7%	28.7%
R.v	volume	1.1%	5.6%	11.325	0.0%	0.0%	0.0%	0.0%	5.1%
R.es	estimate	0.3%	2.8%	18.640	0.0%	0.0%	0.0%	0.0%	0.7%
sum		16.8%	22.5%	2.095	0.0%	1.1%	8.9%	22.4%	68.0%

Table 32: The Strength of News Volatility Channel of Different Information Formats

This table quantifies the strength of the news volatility channel of different information formats through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. Panel A separates the earnings news into 5 information formats. Panel B distinguishes the revenue news into 4 information formats. The sample period is from 2004 to 2017.

Panel A: "Earnings" Group

Info-ID	Information Format	mean	std	skew	p5	p25	p50	p75	p95
E.g	(general)	7.1%	14.2%	3.677	0.0%	0.0%	1.4%	8.1%	32.6%
E.ex	expectation	4.4%	11.4%	4.505	0.0%	0.0%	0.0%	3.3%	24.1%
E.gd	guidance	3.5%	9.3%	5.293	0.0%	0.0%	0.0%	2.7%	17.2%
E.es	estimate	1.1%	4.9%	9.933	0.0%	0.0%	0.0%	0.0%	5.8%
E.ge	guidance-expectation	1.0%	4.3%	7.744	0.0%	0.0%	0.0%	0.0%	5.1%
sum		17.1%	22.2%	2.141	0.0%	2.1%	9.6%	21.7%	67.4%

Panel B: "Revenue" Group

Info-ID	Information Format	mean	std	skew	p5	p25	p50	p75	p95
R.g	(general)	7.1%	14.4%	3.934	0.0%	0.0%	1.2%	8.2%	30.8%
R.gd	guidance	3.5%	10.4%	5.645	0.0%	0.0%	0.0%	1.9%	17.6%
R.v	volume	0.8%	4.2%	11.126	0.0%	0.0%	0.0%	0.0%	4.1%
R.es	estimate	0.3%	3.0%	21.596	0.0%	0.0%	0.0%	0.0%	0.5%
sum		11.6%	18.9%	2.878	0.0%	0.0%	4.6%	14.5%	48.0%

Table 33: The Attenuation of IV Forecasting Power by Using the Important News

This table presents the attenuation of IV forecasting power by selecting a sample of important news, where I filter out unimportant news as in section 5.4. The first row of each model shows the average coefficient across all firms. The parenthesis in the second line shows the T-statistics of the average's difference from zero. The four numbers in the square brackets count the percentage of firms which have coefficient: positive significant, positive insignificant, negative insignificant, negative significant (significant if p-value ≤ 0.01, and the standard errors are newey-west adjusted with 12 lags). The last column shows the attenuation of β_2 by comparing it to its counterpart in Table 7 in terms of mean difference and t-statistics. The sample period spans from 2004 to 2017.

Model: RV_t = Intercept + β_1 RV_{t-1} + β_2 IV_{t-1} + β_3 VP_{t-1} + θ (N_t or NV_t) + ϵ_t

Panel A: Whole Sample (363,456 firm-month)							
Depvar RVt	**Intercept**	**RV_{t-1}**	**IV_{t-1}**	**VP_{t-1}**		**R^2**	**Attenuate β_2**
					N_t		
(1)	0.076	0.169	0.526	-0.002	0.003	39.5%	0.005***
	(26.58)	(30.00)	(71.30)	(-0.46)	(39.52)		(11.21)
		[++20%,+51%,-27%, - -1%]	[++48%,+44%,-8%, - -0%]	[++4%,+45%,-47%, - -4%]	[++30%,+60%,-9%, - -0%]		
					NV_t		
(2)	0.081	0.143	0.542	-0.016	0.039	39.5%	-0.000
	(27.85)	(25.44)	(73.37)	(-3.29)	(48.63)		(-0.13)
		[++19%,+50%,-30%, - -1%]	[++50%,+42%,-8%, - -0%]	[++3%,+44%,-49%, - -5%]	[++29%,+61%,-10%, - -0%]		
Panel B: Sample with option trading volume above median (207,169 firm-month)							
Depvar RVt	**Intercept**	**RV_{t-1}**	**IV_{t-1}**	**VP_{t-1}**		**R^2**	**Attenuate β_2**
					N_t		
(3)	0.041	0.091	0.716	-0.029	0.003	44.1%	0.008***
	(10.99)	(11.65)	(70.89)	(-3.78)	(24.14)		(11.14)
		[++11%,+51%,-36%, - -2%]	[++64%,+32%,-4%, - -0%]	[++3%,+41%,-52%, - -4%]	[++25%,+64%,-11%, - -0%]		
					NV_t		
(4)	0.044	0.069	0.730	-0.042	0.036	44.3%	0.000
	(11.52)	(8.90)	(72.43)	(-5.65)	(32.31)		(0.35)
		[++10%,+50%,-39%, - -2%]	[++67%,+29%,-4%, - -0%]	[++3%,+39%,-54%, - -5%]	[++25%,+66%,-9%, - -0%]		
Panel C: Sample with option trading volume at top 25% (112,261 firm-month)							
Depvar RVt	**Intercept**	**RV_{t-1}**	**IV_{t-1}**	**VP_{t-1}**		**R^2**	**Attenuate β_2**
					N_t		
(5)	0.015	0.066	0.811	-0.028	0.002	48.7%	0.009***
	(3.21)	(5.94)	(57.99)	(-2.43)	(16.27)		(8.17)
		[++9%,+50%,-40%, - -1%]	[++75%,+22%,-3%, - -0%]	[++3%,+39%,-54%, - -4%]	[++21%,+68%,-11%, - -0%]		
					NV_t		
(6)	0.015	0.048	0.824	-0.040	0.030	48.8%	0.001
	(3.32)	(4.30)	(58.49)	(-3.54)	(22.27)		(1.07)
		[++7%,+48%,-44%, - -1%]	[++77%,+21%,-2%, - -0%]	[++3%,+37%,-56%, - -5%]	[++21%,+69%,-10%, - -0%]		

Table 34: Summary Statistics Based on the New Grouping

This table combines all the classification methods and presents the summary statistics on the important news, where I filter out unimportant news as in section 5.4. Column (1) denotes the new grouping: "Earnings" stands for the filtered earnings news, "Revenue" stands for the filtered revenue news, "AnalystRating" stands for Analyst Rating news, "Others" includes news content group id of IR, PT, EA, AM, PS, LI, L, A, IT and Ct. "NF", "FA" and "PR" represent the news format of News-Flash, Full-Article and Press-Releases, respectively. "Sch" ("Unsch") stands for scheduled (unscheduled) news. Column (2) is the total number of news stories. Column (3) is the percentage of scheduled news within a group. Column (4)-(6) shows the distribution of Composite Sentiment Scores(CSS). The last column depicts the average projected news impact(NIP) within the next two hours. The sample period spans from 2004 to 2017.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
New Grouping	FREQ	Sch %			CSS			NIP
			avg	std	skew	min	max	
Earnings.NF	1,534,060	96.2%	-0.02	0.12	-4.89	-0.92	1.00	0.04
Earnings.FA	792,609	98.5%	0.00	0.16	-2.16	-0.92	1.00	0.09
Earnings.PR	507,170	99.3%	0.04	0.06	0.23	-0.92	1.00	0.01
Revenue.NF	491,619	80.1%	0.01	0.06	0.44	-0.92	1.00	0.05
Revenue.FA	254,630	84.3%	0.01	0.15	-1.38	-0.92	1.00	0.04
Revenue.PR	253,512	95.3%	0.05	0.06	0.19	-0.92	1.00	0.05
AnalystRating.NF	182,626	0.0%	-0.04	0.16	-1.10	-0.92	0.56	-0.05
AnalystRating.FA	208,105	0.0%	-0.03	0.17	-1.24	-0.92	1.00	-0.11
AnalystRating.PR	1,470	0.0%	0.01	0.14	-2.14	-0.92	0.42	-0.16
Others.Sch.NF	87,778	100.0%	0.00	0.04	-6.75	-0.92	0.30	-0.04
Others.Sch.FA	114,695	100.0%	0.02	0.09	-2.84	-0.92	1.00	-0.12
Others.Sch.PR	161,236	100.0%	0.04	0.05	0.76	-0.92	1.00	-0.10
Others.Unsch.NF	155,931	0.0%	-0.01	0.09	-3.31	-0.92	1.00	-0.18
Others.Unsch.FA	592,949	0.0%	0.00	0.09	-1.78	-0.92	1.00	-0.17
Others.Unsch.PR	185,353	0.0%	0.02	0.07	-1.64	-0.92	1.00	-0.16

Table 35: The Strength of News Intensity Channel of Important News with New Grouping

This table quantifies the strength of the news intensity channel of different important news through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. We filtered out those unimportant news[1], and combined all the previous mentioned classification methods to form the New Grouping: "Earnings" stands for the filtered earnings news, "Revenue" stands for the filtered revenue news, "AnalystRating" stands for Analyst Rating news, "Others" includes news content group id of IR, PT, EA, AM, PS, LI, L, A, IT and Ct. "NF", "FA" and "PR" represent the news format of News-Flash, Full-Article and Press Releases, respectively. "Sch" ("Unsch") stands for scheduled (unscheduled) news. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. Panel A separates the 30 news content groups into three mediation analysis test groups. Panel B runs a final test with a set of selected news content groups whose average strength is not less than 0.09% in Panel A. The sample period is from 2004 to 2017.

New Grouping	mean	std	skew	p5	p25	p50	p75	p95
AnalystRating.NF	3.8%	8.6%	4.330	0.0%	0.0%	0.1%	3.7%	19.5%
Revenue.NF	3.6%	9.0%	3.917	0.0%	0.0%	0.0%	2.1%	21.2%
Others.Unsch.NF	2.8%	8.2%	5.840	0.0%	0.0%	0.0%	1.6%	15.2%
Revenue.PR	2.7%	7.7%	4.965	0.0%	0.0%	0.0%	0.9%	15.4%
Earnings.PR	2.6%	7.3%	4.954	0.0%	0.0%	0.0%	1.3%	15.3%
AnalystRating.FA	2.4%	7.9%	6.026	0.0%	0.0%	0.0%	0.9%	13.6%
Earnings.NF	2.3%	8.5%	5.723	0.0%	0.0%	0.0%	0.0%	14.1%
Earnings.FA	2.1%	7.2%	6.010	0.0%	0.0%	0.0%	0.0%	12.5%
Others.Unsch.PR	2.0%	6.7%	6.505	0.0%	0.0%	0.0%	0.9%	10.7%
Others.Sch.NF	1.9%	7.1%	6.587	0.0%	0.0%	0.0%	0.0%	11.2%
Others.Unsch.FA	1.9%	7.4%	7.021	0.0%	0.0%	0.0%	0.0%	10.6%
Others.Sch.PR	1.7%	6.2%	7.436	0.0%	0.0%	0.0%	0.5%	9.4%
Revenue.FA	1.5%	6.0%	8.145	0.0%	0.0%	0.0%	0.0%	7.8%
Others.Sch.FA	1.5%	6.0%	7.543	0.0%	0.0%	0.0%	0.0%	7.8%
sum	32.8%	28.6%	1.116	2.3%	11.3%	23.5%	45.5%	100.0%

[1] refer to Section 5.4

Table 36: The Strength of News Volatility Channel of Important News with New Grouping

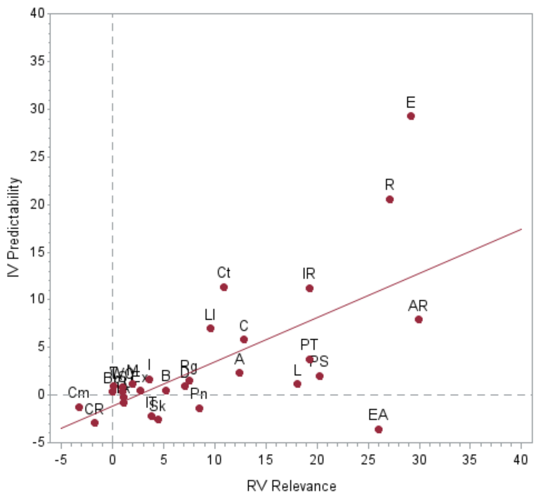
This table quantifies the strength of the news volatility channel of different important news through which the implied volatility at the end of the month $t-1$, $\mathrm{IV}_{i,t-1}$, forecasts the realized volatility at month t, $\mathrm{RV}_{i,t}$. The strength of the channel is represented by its proportion over the total predictability of implied volatility on realized volatility. We filtered out those unimportant news[2], and combined all the previous mentioned classification methods to form the New Grouping: "Earnings" stands for the filtered earnings news, "Revenue" stands for the filtered revenue news, "AnalystRating" stands for Analyst Rating news, "Others" includes news content group id of IR, PT, EA, AM, PS, LI, L, A, IT and Ct. "NF", "FA" and "PR" represent the news format of News-Flash, Full-Article and Press Releases, respectively. "Sch" ("Unsch") stands for scheduled (unscheduled) news. The mean, standard deviation(std), skewness (skew), the 5th, 25th, 50th, 75th, and 95th percentiles of the cross-sectional distribution of the strength are shown. The detail calculation involves a two-stage mediation analysis mentioned in the previous sections. The control variables are $\mathrm{RV}_{i,t-1}$ and $\mathrm{VP}_{i,t-1}$. Panel A separates the 30 news content groups into three mediation analysis test groups. Panel B runs a final test with a set of selected news content groups whose average strength is not less than 0.09% in Panel A. The sample period is from 2004 to 2017.

New Grouping	mean	std	skew	p5	p25	p50	p75	p95
Earnings.NF	3.5%	9.3%	4.577	0.0%	0.0%	0.0%	2.3%	19.6%
AnalystRating.NF	3.3%	8.8%	5.225	0.0%	0.0%	0.0%	2.6%	16.5%
Earnings.FA	3.2%	8.1%	4.991	0.0%	0.0%	0.0%	2.9%	16.9%
Revenue.NF	2.9%	8.3%	5.334	0.0%	0.0%	0.0%	1.6%	15.7%
AnalystRating.FA	2.3%	7.8%	6.024	0.0%	0.0%	0.0%	0.7%	13.0%
Others.Unsch.FA	2.2%	7.7%	6.682	0.0%	0.0%	0.0%	0.7%	11.8%
Others.Unsch.NF	2.1%	7.2%	6.037	0.0%	0.0%	0.0%	0.6%	12.5%
Earnings.PR	2.1%	6.4%	5.789	0.0%	0.0%	0.0%	0.8%	11.9%
Revenue.PR	1.8%	6.3%	6.824	0.0%	0.0%	0.0%	0.2%	10.1%
Revenue.FA	1.8%	6.0%	6.548	0.0%	0.0%	0.0%	0.3%	10.5%
Others.Unsch.PR	1.7%	5.9%	7.129	0.0%	0.0%	0.0%	0.5%	9.5%
Others.Sch.FA	1.6%	6.4%	7.808	0.0%	0.0%	0.0%	0.2%	8.1%
Others.Sch.PR	1.6%	5.9%	7.315	0.0%	0.0%	0.0%	0.3%	7.9%
Others.Sch.NF	1.4%	5.5%	7.535	0.0%	0.0%	0.0%	0.0%	7.5%
sum	31.7%	27.8%	1.221	2.5%	11.3%	22.4%	43.3%	100.0%

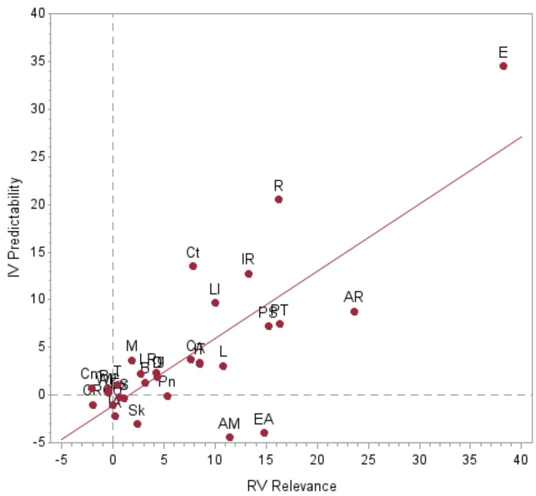
[2] refer to Section 5.4

Figure 1: RV-IV Relation across Different Content Groups

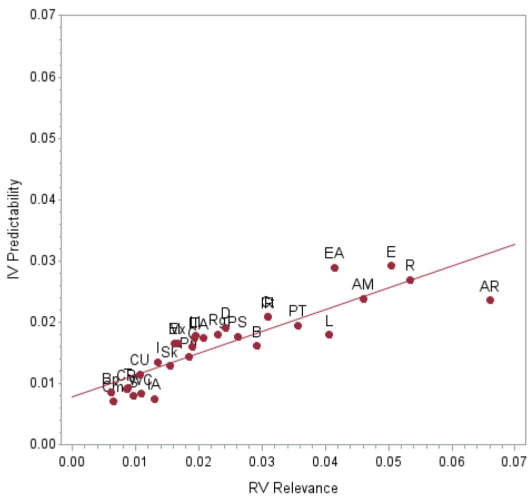
This figure draws four scatter plots to present the positive relation between realized volatility relevance and implied volatility predictability across different news content groups in four different settings: A. T-stat of average coefficients for using News Intensity(N) measure; B. T-stat of average coefficients for using News Volatility(NV) measure; C. R^2 for using News Intensity(N) measure; D. R^2 for using News Volatility(NV) measure. The x-axis represents the realized volatility relevance by regressing realized volatility (RV_t) on the concurrent news measures(N_t or NV_t), while Y-axis measures the implied volatility predictability by regressing the future news measures (N_{t+1} or NV_{t+1}) on the current implied volatility(IV_t). The capital letters are the G-ID denoting the news content group as in Table 18. The lines are regression lines. The sample period spans from 2004 to 2017.



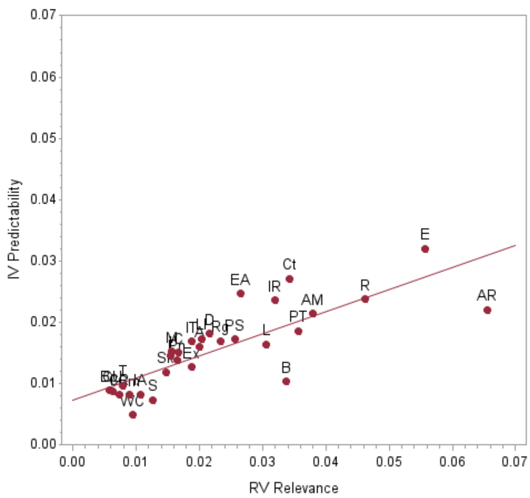
A. News Intensity (N) T-stat



B. News Volatility (NV) T-stat



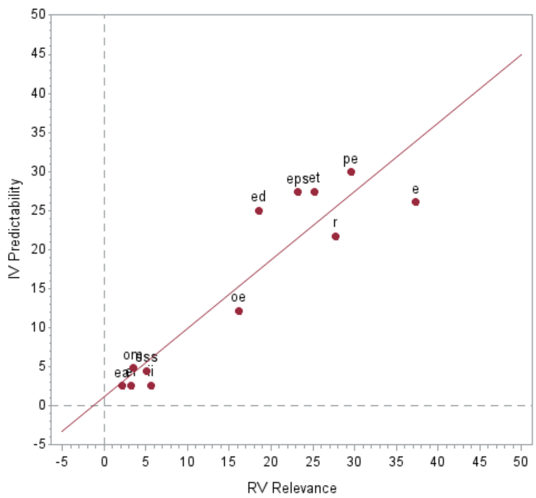
C. News Intensity (N) R^2



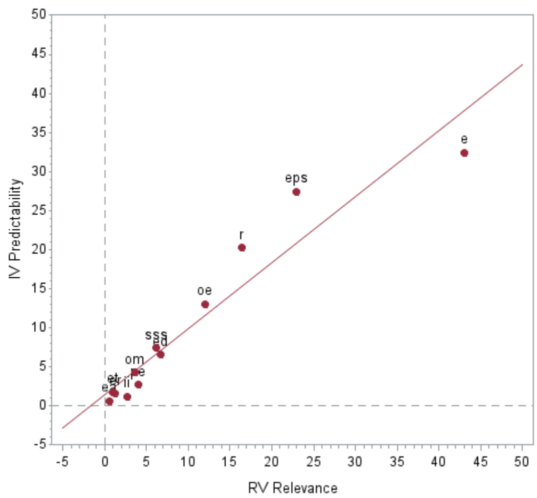
D. News Volatility (NV) R^2

Figure 2: RV-IV Relation across Different Accounting Items

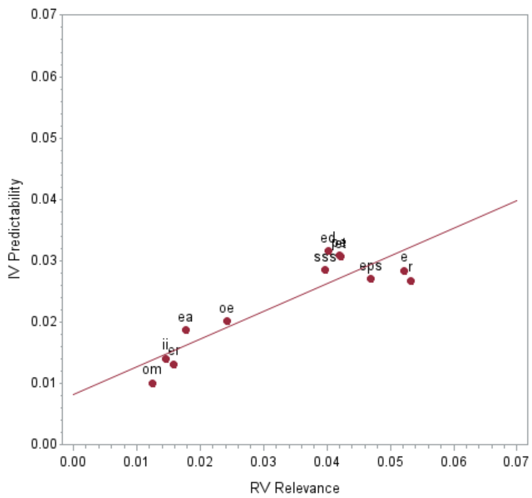
This figure draws four scatter plots to present the positive relation between realized volatility relevance and implied volatility predictability across different accounting items related to earnings or revenue news in four different settings: A. T-stat of average coefficients for using News Intensity(N) measure; B. T-stat of average coefficients for using News Volatility(NV) measure; C. R^2 for using News Intensity(N) measure; D. R^2 for using News Volatility(NV) measure. The x-axis represents the realized volatility relevance by regressing realized volatility (RV_t) on the concurrent news measures(N_t or NV_t), while Y-axis measures the implied volatility predictability by regressing the future news measures (N_{t+1} or NV_{t+1}) on the current implied volatility(IV_t). The lowercase letters are the AC-ID denoting the accounting items related to earnings or revenue news as defined in Table 23, Figure A2 and Figure A3. The lines are regression lines. The sample period spans from 2004 to 2017.



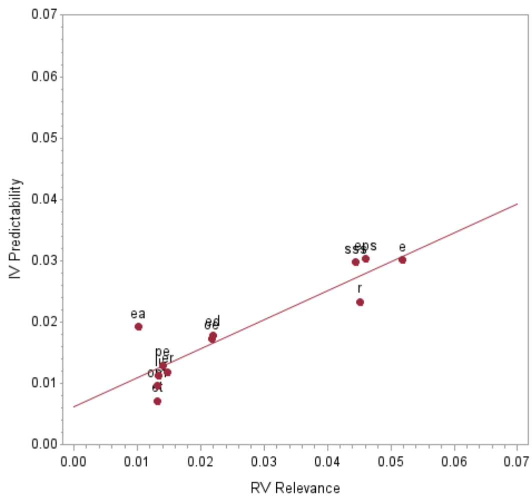
A. News Intensity (N) T-stat



B. News Volatility (NV) T-stat



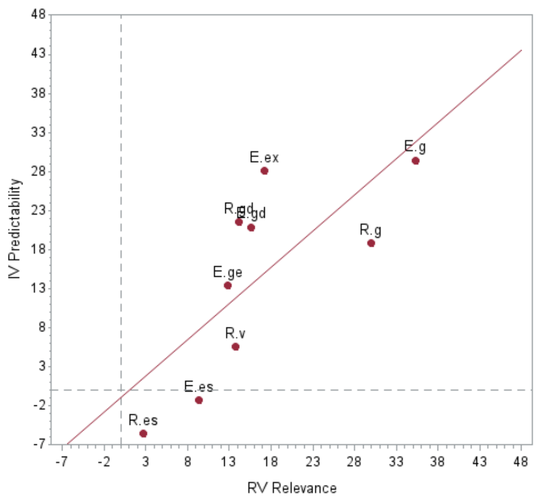
C. News Intensity (N) R^2



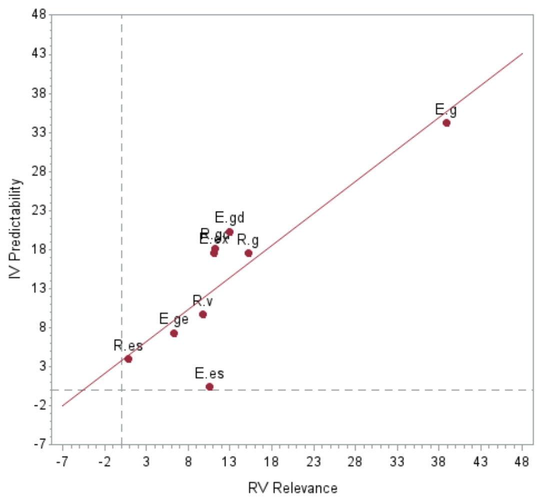
D. News Volatility (NV) R^2

Figure 3: RV-IV Relation across Different Information Formats

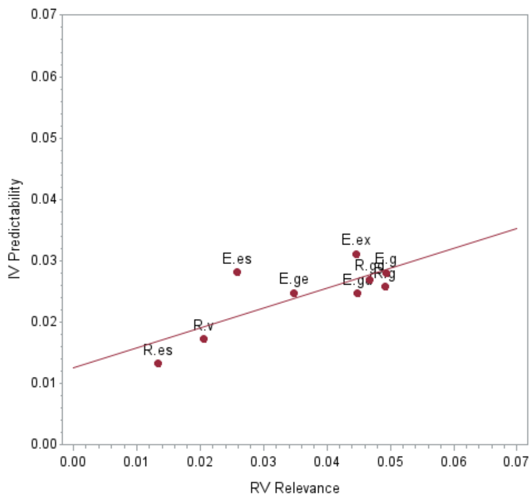
This figure draws four scatter plots to present the positive relation between realized volatility relevance and implied volatility predictability across different information formats related to earnings or revenue news in four different settings: A. T-stat of average coefficients for using News Intensity(N) measure; B. T-stat of average coefficients for using News Volatility(NV) measure; C. R^2 for using News Intensity(N) measure; D. R^2 for using News Volatility(NV) measure. The x-axis represents the realized volatility relevance by regressing realized volatility (RV_t) on the concurrent news measures(N_t or NV_t), while Y-axis measures the implied volatility predictability by regressing the future news measures (N_{t+1} or NV_{t+1}) on the current implied volatility(IV_t). The lowercase letters are the Info-ID denoting the information formats related to earnings or revenue news as defined in Table 28, Figure A2 and Figure A3. The lines are regression lines. The sample period spans from 2004 to 2017.



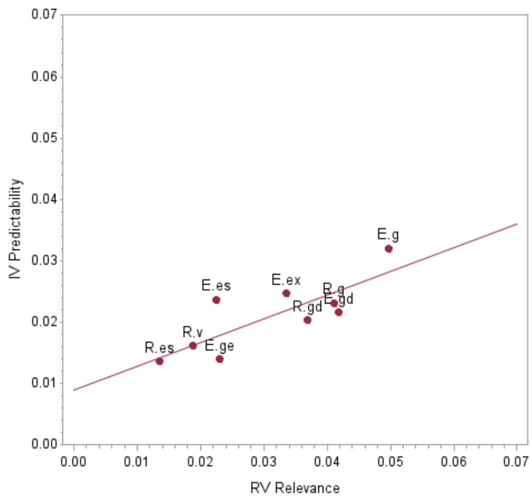
A. News Intensity (N) T-stat



B. News Volatility (NV) T-stat



C. News Intensity (N) R^2



D. News Volatility (NV) R^2

Appendix

Table A1: Variables List

This table explains all the variables I used in this paper. News data is from RavenPack Analytic 1.0, option data is retrieved from Option Metrics IvyDB US, and stock data is from CRSP. The sample period spans from 2004 to 2017.

Variables	Construction
RV	Realized volatility. Download the daily stock return data from CRSP, and filter with Share Code 10 & 11 and the three primary stock exchanges of NYSE, AMEX & NASDAQ. The realized volatility is the standard deviation of daily return(r_t) within a month (m) and then annualize it by multiplying $\sqrt{252}$: $$\mathrm{RV}_m = \sqrt{\frac{\sum_{t \in m}(r_t - \bar{r})^2}{n} \times 252}$$ where n is the number of trading days in the month m, $\bar{r}$ is the average return of the month. I require there are at least ten trading days of a month in order to calculate a valid RV data.
IV	Implied volatility. First, obtain the implied volatility of the closest to being at-the-money call and put options from Option Metrics IvyDB. Then, take the average of the put-call pair to adjust the stale price bias. Eventually, align their maturity date to the end of each month, as follows: let the end of the current month to be date t, the end of next month to be T. We select two options whose expiration dates are T_1 and T_2, which are closest to T with $t < T_1 \leq T < T_2$. One can interpolate the implied volatility at t with an expiration date at T with the following formula: $$\mathrm{IV}_{t,T} = \mathrm{IV}_{t,T_1} + (\mathrm{IV}_{t,T_2} - \mathrm{IV}_{t,T_1}) \frac{T - T_1}{T_2 - T_1}$$ where $t < T_1 \leq T < T_2$.

Table A1: Variables List (Continue)(2)

Variables	Construction
CSS	Composite Sentiment Score. The score ranges from -1 to 1. Score 0 represents the neutral news, and the amount above(below) 0 indicates how positive(negative) the news story is. The strength of the score is determined by a composite sentiment model on the intra-day stock price responses, which is trained by using the tick data of 100 large-cap stocks. The model comprises of five different analytics: • PEQ analytics identify general positive and negative words and phrases in articles about the equities; • BEE analytics focus on news about earnings evaluation; • BMQ analytics specialize in short commentary and editorial; • BAM analytics focus on news stories of mergers, acquisitions and takeovers; • BCA analytics expertize in reports about corporate action announcement; where CSS is the average of these five analytics, and RavenPack ensures that there is no sentiment disagreement amongst them. This news data is extracted from RavenPack Analytics 1.0 with the following filters: Maximum Novelty (Event Similarity Day = 365), Maximum Relevance (Relevance = 100), and drop the post-reflection news (i.e., the news content groups[3] of "order-imbalances", "stock-prices", and "technical-analysis").
ESS	Event Sentiment Score. Similar to CSS, the score ranges from -1 to 1. Score 0 represents the neutral news, and the amount above(below) 0 indicates how positive(negative) the news story is. The score is constructed by financial experts using the different scoring system on (over 6700) different categories of contents. For examples: in earnings news, experts compare actual figures to estimated figures; in analyst rating news, experts compare the rating changes; in earthquake news, they analyze the Richter scale; and in terrorism attack news, they focus on the number of casualties. This data variable is retrieved from RavenPack Analytics 1.0 with the same filtering method as for CSS.

[3] "Group" is a content classification level in RavenPack. The whole taxonomy is illustrated in Figure A1.

Table A1: Variables List (Continue)(3)

Variables	Construction
NIP	News Impact Projection. The score ranges from -1 to 1. It measures the news' projected impact in terms of relative volatility increase over the next two hours. The model analyzes the text used by journalist, especially for corporate action and analyst revision. The score is only meaningful if the score is above zero. The higher value indicates more impact on the volatility with higher confidence in the projection. Score zero or below indicates a low impact or the projected effect is unknown.
N	Number of News within a month
N.Sch	Number of Scheduled News within a month
N.unSch	Number of Unscheduled News within a month
N.NF	Number of News-Flash and Hot-News-Flash within a month
N.PR	Number of Press-Release within a month
N.FA	Number of Full-Article within a month
N.TM	Number of Tabular-Materials within a month
N.SEC	Number of SEC filings (10K, 10Q, 13D, 13F, 144 and 8K) within a month
NV	News Volatility. Sum up the square of news-story level CSS within a month, and then annualize it by multiplying 12. Finally, take the square root to get the NV. If there is no news within a month, the value is zero. $$NV = \sqrt{\sum_{i=1}^{\mathrm{N}} (CSS_i)^2 \cdot 12}$$ where N is the number of news within a month.
NV.Sch	News Volatility of Scheduled News.
NV.unSch	News Volatility of Unscheduled News.
NV.NF	News Volatility of News Flash
NV.PR	News Volatility of Press Release
NV.FA	News Volatility of Full Articles
NV.TM	News Volatility of Tabular Materials
NV.TM	News Volatility of SEC filings (10K, 10Q, 13D, 13F, 144 and 8K)

Table A2: Category Description of Type "earnings-guidance"

This table lists out all the categories in the Type "earnings-guidance" with detail descriptions provided by RavenPack. The order follows the category order as in Figure A1.

CATEGORY	DESCRIPTION
earnings-guidance-rater	The Entity announces or expresses an opinion about the company's earnings guidance
earnings-guidance-opinion	A view or opinion is expressed about the Company's earnings guidance
earnings-guidance	The Company announces its earnings guidance figures or projections
earnings-guidance-up-rater	The Entity announces or expresses an opinion about the increase in the company's earnings guidance
earnings-guidance-up-opinion	A view or opinion is expressed about the increase in the Company's earnings guidance
earnings-guidance-up	The Company announces an increase in its earnings guidance figures or projections
earnings-guidance-down-rater	The Entity announces or expresses an opinion about the decrease in the company's earnings guidance
earnings-guidance-down-opinion	A view or opinion is expressed about the decrease in the Company's earnings guidance
earnings-guidance-down	The Company announces a decrease in its earnings guidance figures or projections (e.g. Profit Warning)
earnings-guidance-suspended-rater	The Entity announces or expresses an opinion about the company suspending its earnings guidance
earnings-guidance-suspended-opinion	A view or opinion is expressed towards the Company suspending its earnings guidance
earnings-guidance-suspended	The Company suspends issuance or will not provide financial guidance for an upcoming earnings period
earnings-guidance-unchanged-rater	The Entity announces or expresses an opinion about the company's earnings guidance remaining unchanged
earnings-guidance-unchanged-opinion	A view or opinion is expressed about the Company's earnings guidance remaining unchanged
earnings-guidance-unchanged	The Company announces unchanged earnings guidance figures or projections

Figure A1: Content Classification with RavenPack Taxonomy

This figure plots the taxonomy structure of RavenPack's content classifcation. The three horizontal dashed lines delineate the four major layers of content classification: Topic, Group, Type, and Sub-type. The area below the last horizontal dashed lines depicts the type "earnings-guidance" as an example in details. The first three vertical dashed lines demarcate the two naming attributes: properties(i.e., extracted roles, numbers, and strings) and fact-level(i.e., fact, opinion or forecast). The third and fourth vertical dashed lines delimit the finest content classification: category. The last column indicates whether the category is scheduled or not.

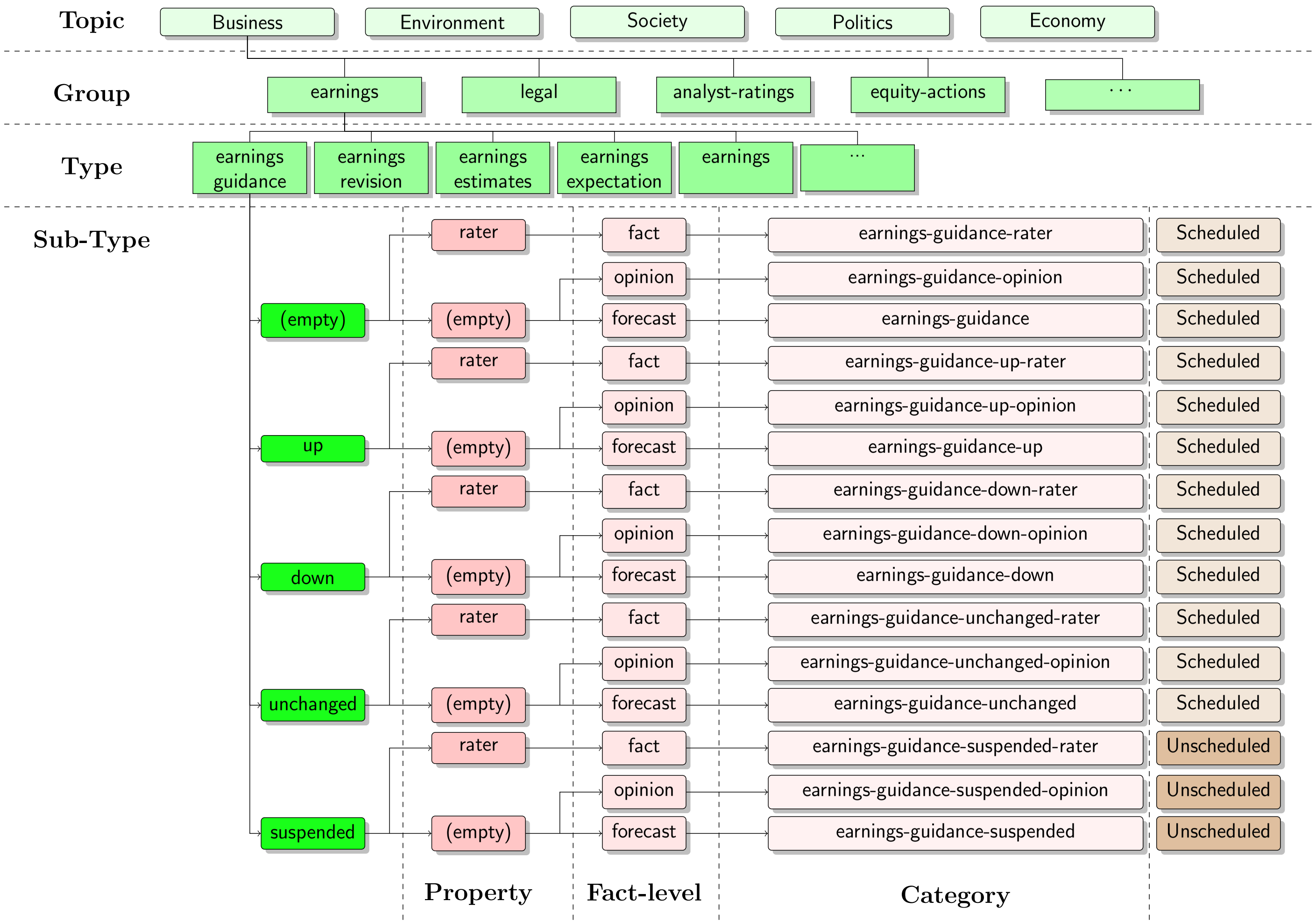


Figure A2: The Subgroups under the Earnings Group

This figure maps all the RavenPack's types under the Earnings Group as a combination of accounting items (Column ①) and information formats(Column ②). The accounting items are the variants of "earnings" mentioned in the news. Information formats are the ways that earnings news is presented. Column ③ lists all the Ravenpack's types under the Earnings Group. The parenthesis denotes the short-form of a specific accounting item or information format.

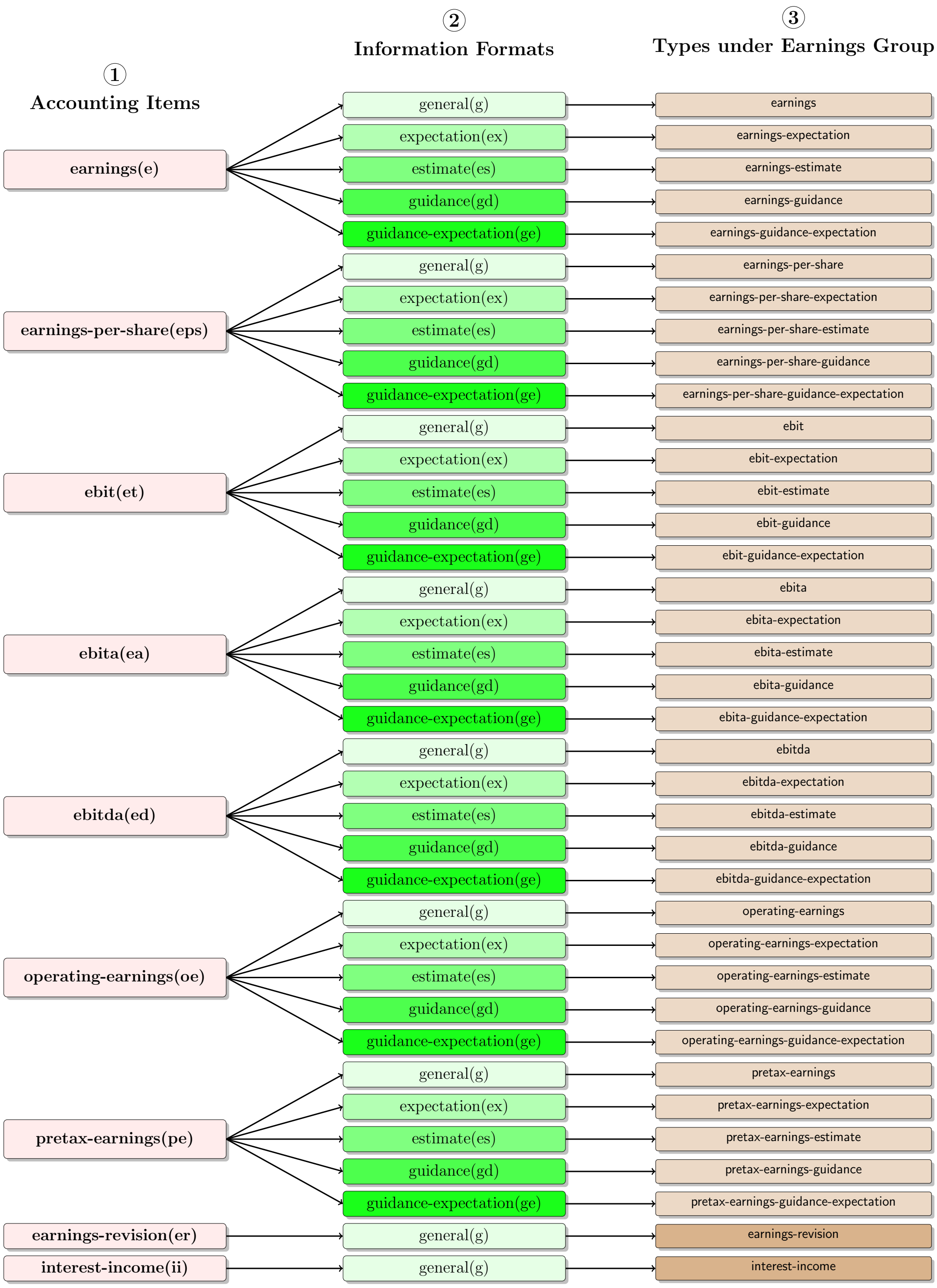


Figure A3: The Subgroups under the Revenue Group

This figure maps all the RavenPack's types under the Revenue Group as a combination of accounting items (Column 1) and information formats(Column 2). The accounting items are the variants of "revenue" mentioned in the news. Information formats are the ways that revenue news is presented. Column 3 lists all the Ravenpack's types under the Revenue Group. The parenthesis denotes the short-form of a specific accounting item or information format.