Journal of Banking and Finance 156 (2023) 107019

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Why does option-implied volatility forecast realized volatility? Evidence from news events ☆

Sipeng Chen, Gang Li *

Faculty of Business, the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

JEL classification:
G12
G13

Keywords:
Implied volatility
Realized volatility
News intensity and magnitude
Scheduled and unscheduled news
Fundamental and non-fundamental news

ABSTRACT

This study examines the information content of stock option-implied volatility. We measure the arrival intensities and magnitudes of scheduled and unscheduled news as well as fundamental and non-fundamental news. Most of these news measures exhibit strong and positive associations with contemporaneous stock return volatility, and many of them can be predicted by implied volatility. Approximately one third of the predictive power of implied volatility on future realized volatility can be attributed to its ability to predict these news measures, with the majority of the predictive power arising from its capacity to predict the arrival intensities of both scheduled and unscheduled news. The predictive power is higher for fundamental news than for non-fundamental news.

1. Introduction

Numerous studies suggest that the volatility implied from option prices offers a more efficient forecast of future stock volatility compared with alternative approaches, such as historical volatility and model-based methods. In other words, option prices subsume the information contained in other forecasting variables. For instance, studies on the S&P 100 index options include works by Christensen and Prabhala (1998) and Fleming et al. (1995). In addition, studies on individual stock options include those conducted by Cao et al. (2010) and Taylor et al. (2010).[1] The superior predictive power of implied volatility is attributable to the fact that it reflects the market expectations of future volatility. However, studies have not examined the type and extent of volatility-related information contained within option prices. This study aims to fill this gap.

This study delves into the reasons for the ability of option-implied volatility to predict future volatility. We present empirical evidence that option-implied volatility contains information regarding future news events related to underlying stock returns' volatility. We utilize a comprehensive database of public news announcements to measure total news variation, a variable conceptually similar to realized return volatility but tailored to news. Using a sample of S&P 500 index constituents from January 2004 to June 2019, we investigate the joint dynamics of the total news variation, realized volatility, and implied volatility through structural vector autoregressions (SVARs). Our findings reveal a positive and significant association between total news variation and contemporaneous realized volatility for the majority of firms. For nearly 60% of the firms in our sample, implied volatility at the end of the previous month positively and significantly predicts total news variation. This finding suggests that the predictive power of

☆ We would like to thank the editor (Geert Bekaert), an anonymous referee, Guanglian Hu, Chu Zhang, and the participants at the Australasian Finance and Banking Conference and the World Finance Conference for their helpful comments on earlier versions. Li acknowledges the financial support by the Research Grants Council of the Hong Kong Special Administrative Region, China (grant no. PolyU 15506118).

* Corresponding author.
E-mail addresses: s1chen@polyu.edu.hk (S. Chen), garyli@polyu.edu.hk (G. Li).

[1] Latané and Rendleman (1976) and Chiras and Manaster (1978) are the first to report that option-implied volatility is a better predictor of future volatility than historical volatility is. The literature on this subject is extensive. Poon and Granger (2003) conduct a comprehensive review of early studies and determine that implied volatility outperforms both historical volatility and GARCH type of models when forecasting future volatility in the majority of studies. Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) derive the model-free implied volatility, which is an enhancement over the original Black-Scholes implied volatility because it does not require the assumptions found in the Black-Scholes model. Oikonomou et al. (2019) examine the more recently developed weekly options on the S&P 500 index and find that the implied variance of these weekly options is a strong predictor of the weekly realized variance.

https://doi.org/10.1016/j.jbankfin.2023.107019
Received 11 May 2022; Accepted 2 October 2023

Available online 6 October 2023

implied volatility concerning future realized volatility is partially explained by its ability to predict future news events.

Our primary analysis focuses on the decomposition of the total news variation into different components and the dynamic relationships of these components with realized and implied volatility. Some public announcements are scheduled in advance, including earnings announcements, whereas others are unscheduled, such as analyst recommendations. For scheduled news, the arrival timing is known prior to its announcement, but the magnitude of its impact remains unknown. For unscheduled news, both its arrival timing and impact of a news event are unknown ahead of time. Because the arrival time and impact of scheduled and unscheduled news have different implications for predictability, we examine variations in these two types of news separately. In addition, we decompose each of the news variation measures into two components, namely the arrival intensity of news and the magnitude of news, which is defined as the average expected impact per news event. By performing SVARs and incorporating more refined news measures, we find that the arrival intensities of scheduled and unscheduled news and the magnitude of unscheduled news exhibit a strong association with contemporaneous realized volatility. Furthermore, we classify news into fundamental and non-fundamental types and determine that variations in fundamental news have greater explanatory power for realized volatility than those in non-fundamental news have. Moreover, we find that implied volatility positively predicts many of these news measures. As expected, the predictability of arrival intensity of scheduled news is the strongest. In addition, we observe robust predictability for the arrival intensities of unscheduled fundamental and non-fundamental news. However, the prediction of the magnitude of news is generally more challenging.

We quantify the importance of the news channel through which implied volatility predicts future realized volatility by conducting a mediation analysis. In this analysis, we use the news measures as mediators. The results indicate that the news channel accounts for approximately one third of the overall predictive power of implied volatility on future realized volatility. The strength of the channel is greater for the arrival intensity of news than for the magnitude of news and is also greater for fundamental news than for non-fundamental news. Our time-series analysis reveals a negative relationship between the average strength of the news channel across firms and market volatility. Furthermore, the strength of the non-fundamental news channel relative to the fundamental news channel increases over time. These findings indicate that option prices incorporate both public and private information on stock return volatility, particularly information pertaining to the arrival of news events. This is a key reason why implied volatility is a more efficient predictor of future volatility than alternative volatility forecasting approaches, such as historical realized volatility, GARCH models, and stochastic volatility models. These alternative approaches do not rely on information concerning future events. Instead, they depend on statistical properties, such as volatility persistence and mean reversion.

Our paper contributes to the literature on the role of news in determining stock volatility.[2] Neuhierl et al. (2013) examine the impact of various types of corporate press releases on stock price reactions. Manela and Moreira (2017) perform a text analysis of news content to construct news-implied volatility of the US market and find that it effectively predicts the volatility index (VIX) of the S&P 500 and the realized volatility of market returns. Boudoukh et al. (2019) show that the arrival of firm news accounts for approximately 50% (12%) of idiosyncratic volatility during overnight (trading hours). Glasserman and Mamaysky (2019) find that atypical negative (positive) news predicts an increase (decrease) in return volatility at both stock and market levels. Calomiris and Mamaysky (2019) use context-specific measures of news flow to forecast equity market volatility around the world. Engle et al. (2021) propose an econometric specification for firm-specific return volatility, emphasizing that the arrival of news affects both contemporaneous and future volatility. Jeon et al. (2022) find that news flow is related to contemporaneous jump volatility in stock prices. Our results reveal that the arrival intensity of news has a greater impact on return volatility than the magnitude of news has. Furthermore, fundamental news has a greater impact than non-fundamental news. Our finding that the arrival intensity of news exhibits the strongest association with return volatility provides supporting evidence for the specifications of volatility models, such as those proposed by Boudoukh et al. (2019) and Engle et al. (2021).

Our study complements a stream of literature that indicates the predictive power of option trading for future volatility. Ni et al. (2008) find that the non-market maker net demand for volatility, as derived from the trading volume of individual equity options, is informative about the future realized volatility of underlying stocks and positively affects option prices. Using an alternative approach, Fahlenbrach and Sandås (2010) determine that the order flow in volatility-sensitive option strategies contains information on future realized volatility. Roll et al. (2010) document that the option-to-stock trading volume ratio is higher around earnings announcements, and post-announcement absolute stock returns are positively related to the pre-announcement volume ratio. Furthermore, Bali and Hovakimian (2009) find that both option volumes and option returns provide information on the future realized volatility of underlying stocks. These studies suggest that volatility trading is common in the options market and that such trading is informed. Going beyond these findings, our study investigates the specific nature of information embedded in this informed trading. Our results reveal that the information on future news events, especially the arrival intensity of news, is the source of informed trading.

This paper is also related to the literature examining the potential of implied volatility and option volume to predict stock returns ahead of corporate and macroeconomic announcements. Corporate events include analysts' recommendations (Hayunga and Lung, 2014), earnings announcements (Xing et al., 2010; Atilgan, 2014), mergers and acquisitions (Cao et al., 2005; Chan et al., 2015; Augustin et al., 2019), spin-offs (Augustin et al., 2020), and private equity buyouts (Acharya and Johnson, 2010). Han and Li (2021) find that the aggregate implied volatility spread of call and put equity options is significantly and positively related to future stock market returns. The return predictive power of the aggregate implied volatility spread is attributed to its ability to forecast macroeconomic news. These studies exclusively focus on directional information regarding underlying stock prices. However, studies focusing on volatility-related information are scant. The fact that option price and volume predict stock returns implies that they also contain information on future volatility because volatility can be regarded as unsigned stock returns. In this paper, we present direct evidence indicating that option prices contain information on the future arrival intensity of news, thereby predicting future realized volatility.

The remainder of the paper is organized as follows. Section 2 presents the data, describes the preliminary analysis of the impact of news events on stock price movements, and provides the summary statistics of the main variables used in the study. Section 3 delves into the joint dynamics of news variations, realized volatility, and implied volatility and further decomposes news variations into various components. Section 4 quantifies the strength of the news channel through which implied volatility predicts realized volatility and conducts a time-series analysis of the strength of the news channel. Section 5 concludes the paper.

[2] In an early study, without using data from actual news, French and Roll (1986) infer the impact of public information on volatility by examining differences in volatilities during trading and non-trading hours. They find little evidence regarding the role of public information. Kalev et al. (2004) investigate the impact of news on volatility modeling by using a sample of stocks and a stock index in Australia. In other markets, Ederington and Lee (1993) document the significant impact of scheduled macroeconomic news announcements on futures markets for interest rates and foreign exchange. Furthermore, Jones et al. (1998) find that the release of US macroeconomic news is associated with substantial bond market volatility.

2. Data description and preliminary analysis

In this section, we describe the data and examine the impact of news events on contemporaneous stock price movements. Subsequently, we define the variables used in our main empirical analysis, namely realized volatility, implied volatility, and new variation, and examine their properties.

2.1. News data

We utilize a comprehensive news database, RavenPack Analytic 1.0 (or RavenPack). This database records and transforms a vast number of unstructured news stories from various sources into structured and quantified data. We use the news stories of US companies in our empirical analysis. RavenPack sequentially introduced three mutually exclusive data packages, namely the Dow Jones Package, the Press Releases Package, and the Web Package. The Dow Jones Package encompasses all news from Dow Jones Newswires, Wall Street Journals, Barron's, and MarketWatch since 2000. The Press Releases Package tracks over 100,000 press releases and regulatory disclosures through various information distribution networks starting from 2004. The Web Package monitors over 19,000 sources of leading publishers and websites since 2007. We consider the following factors while selecting our sample period. Initially, 20% of firm-month-level observations lack any news data before 2004. However, this percentage declines to approximately 10% after 2004 and remains relatively stable thereafter. Because we cannot determine whether there is no news available or news is not covered in the database during the early period, to be on the conservative side, we use the database from 2004 and onward. Thus, our sample period spans from January 2004 to June 2019. This period completely covers sources tracked by the Dow Jones Package and Press Releases Package and partially covers sources tracked by the Web Package.

To identify news stories that exert the strongest impact on stock prices, we select those with maximum relevance and novelty. For news relevance, RavenPack generates a relevance score (RS) ranging from 0 to 100. This score reflects the extent to which a news story is pertinent to an individual equity. Scores over 90 indicate that the company is quoted in the main title or headline, whereas scores below 90 indicate that the company appears in the body text. A score of 100 suggests that this company plays a crucial role in the headline of the story. For example, in a news story of a rating company offering stock recommendations in the headline, the rating company would receive a score of 90 and the rated company would receive a score of 100. We only retain news with the maximum RS of 100. RavenPack provides event similarity days (ESD) as a measure of news novelty. The ESD of a news story is the number of days that have elapsed since the detection of a previous event with similar content, indicating the absence of similar events within the past ESD days. The maximum ESD, as defined by RavenPack, is 365, demonstrating that either the news is the most recent or that no similar news has been published over the past year. We select the news with maximum ESD, ensuring that we retain the initial news and discard all subsequent rebroadcasted news within the following year.

We rely on the composite sentiment score (CSS) provided by RavenPack to quantify positive or negative news. The CSS ranges from -1 to 1, representing how positive or negative the projected price impact is for a news story. A score of 0 indicates that the news is neutral with no projected price impact, and a positive (negative) CSS indicates a positive (negative) price impact. RavenPack uses text analysis to create a universal scoring model and trains it using the intraday price responses of 100 large-cap stocks. This model comprises five domains of analytics: general wordings, earnings evaluation, commentary, M&A, and corporate action. The final CSS is the average of these five analytics.[3]

2.2. Sample selection

We empirically analyze the constituents of the S&P 500 index, focusing on ordinary common shares (share codes 10 and 11) that are traded on the NYSE, NASDAQ, and AMEX. Options on these stocks are actively traded. To ensure the accuracy of our SVARs involving many news components, we require a stock to have at least 10 years of available data. We obtain daily stock return data from the CRSP and data on individual stock options from OptionMetrics IvyDB US. Our final sample consists of 604 unique stocks, covering the period from January 2004 to June 2019.

2.3. Impacts of news on stock prices

Financial theory suggests that unanticipated public information affects stock prices. Recent studies have indicated that news and stock price movements are contemporaneously related. For example, Boudoukh et al. (2019) find that fundamental information in news accounts for a substantial proportion of stock volatility. Jeon et al. (2022) document that stock return jumps are significantly related to news flow frequency. To verify that news events in our data affect stock prices, we examine the relationship between stock price movements and the CSS at a daily frequency. We define the daily CSS for firm i on day d as the sum of news-level CSS for firm i on that day,

$$\mathrm{CSS}_{i,d} = \sum_{j=1}^{n_{i,d}} \mathrm{CSS}_{j,i,d} \tag{1}$$

where $n_{i,d}$ is the number of news stories for firm i on day d and $\mathrm{CSS}_{j,i,d}$ is the CSS for the news j of firm i on day d.[4] Furthermore, we define the daily sum of squared sentiment as follows:

$$\mathrm{SSS}_{i,d} = \sum_{j=1}^{n_{i,d}} \mathrm{CSS}^2_{j,i,d}. \tag{2}$$

Both $\mathrm{CSS}_{i,d}$ and $\mathrm{SSS}_{i,d}$ are defined as zero when there is no news event for firm i on day d.

Panel A of Table 1 shows the pooled distribution of CSS observations at the news story level. Our sample contains 3,103,245 news stories. The CSS ranges from -1 to 1, with an average CSS of approximately zero, indicating no price impact. The standard deviation is 0.105. Notably, 54% of these news stories have CSS of zero. The distribution of the CSS is left-skewed, with a skewness value of -2.513. This finding is consistent with that in the literature, indicating that negative news exerts a greater price impact. Panel B shows the cross-sectional averages of the statistics for the time-series distribution for each firm of each variable. The daily CSS, $\mathrm{CSS}_{i,d}$, is slightly positive on average and is skewed to the left, similar to the news-level CSS. $\mathrm{SSS}_{i,d}$ is skewed to the right. Both daily CSSs and SSSs exhibit substantial time-series variations, as indicated by their high standard deviations. The daily log return, $r_{i,d}$, is close to zero on average and is slightly negatively skewed. The squared log return, $r^2_{i,d}$, is right-skewed, similar to $\mathrm{SSS}_{i,d}$.

To examine the impact of news events on daily stock price movements, we run the following time-series regressions for each firm i,

$$r_{i,d} = b_{0,i} + \sum_{k=1}^{3} b_{k,i} \mathrm{CSS}_{i,d+1-k} + \epsilon_{i,d}, \tag{3}$$

[3] RavenPack offers an alternative approach, the event sentiment score (ESS), which is commonly used in the literature. The ESS is constructed by financial experts who apply different scoring systems to various event topics. Thus, for news on the same topic, the ESS provides a more accurate measure of price impact than the CSS. However, for news across different topics, the CSS outperforms the ESS because the same scoring model is implemented for all types of news and CSS is comparable for news across different topics. Thus, given that our study examines all news events, we choose CSS as our measure of price impact.

[4] If news stories are recorded after the market closes, we consider the next trading day as the date of the news events.

Table 1

Descriptive Statistics of News Measures and Stock Returns. Panel A presents the pooled distribution of the composite sentiment score for news j, denoted as CSS_j. Panel B illustrates the cross-sectional averages of the statistics for the time-series distributions for each firm of the following variables: $\text{CSS}_{i,d}$ is the sum of news level CSS for firm i on day d; $\text{SSS}_{i,d}$ is the sum of squared news level CSS for firm i on day d; $r_{i,d}$ is the log stock return for firm i on day d; and $r_{i,d}^2$ is the squared log return. The mean, standard deviation (std), skewness (skew), the 5th, 50th, and 95th percentiles (p5, p50, and p95) are reported. Variables in Panel B are multiplied by 100. The sample spans from January 2004 to June 2019.

A. News level						
	mean	std	skew	p5	p50	p95
CSS_j	0.005	0.105	-2.513	-0.160	0.000	0.120
B. Firm-day level						
	mean	std	skew	p5	p50	p95
$\text{CSS}_{i,d}$	0.704	26.852	-3.562	-4.853	0.000	12.427
$\text{SSS}_{i,d}$	1.521	12.434	17.866	0.000	0.006	3.646
$r_{i,d}$	0.032	2.223	-0.339	-3.154	0.048	3.141
$r_{i,d}^2$	0.055	0.244	16.450	0.000	0.011	0.202

$$r_{i,d}^2 = c_{0,i} + \sum_{k=1}^{3} c_{k,i}\text{SSS}_{i,d+1-k} + \varepsilon_{i,d}. \tag{4}$$

Equation (3) investigates the impact of news on directional changes in stock prices, and equation (4) examines the impact of news on unsigned stock price movements. We include the lagged terms of news measures for a possible delayed response of stock prices to news events.

The average coefficients across firms are shown in Table 2. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic $\geq$ 1.96) and negative and significant coefficients (t-statistic $\leq$ -1.96). The t-statistic is adjusted to account for heteroscedasticity and autocorrelation of 20 lags by using the Newey and West (1987) procedure. In simple regressions, we observe that $\text{CSS}_{i,d}$ is positively and significantly related to $r_{i,d}$ for 74.6% of the firms, and $\text{SSS}_{i,d}$ is positively and significantly related to $r_{i,d}^2$ for 73.6% of the firms. The addition of lagged terms to the model has a slight impact on the estimated coefficients of $\text{CSS}_{i,d}$ and $\text{SSS}_{i,d}$ as well as R^2. The average coefficients of the lagged terms are considerably smaller than those of the contemporaneous terms. The coefficients of the lagged terms are significant only for a small number of firms, and these significant coefficients can be either positive or negative. The results indicate that stock prices respond to contemporaneous news events with minimal delay, consistent with recent findings in the literature. The strong contemporaneous relationship between stock price changes and news measures at the daily frequency justifies our approach of examining the contemporaneous relationship between news measures and realized volatility at the monthly frequency in the subsequent analysis. The contemporaneous relation at the low frequency is not driven by the lead-lag relationship at the daily frequency.

2.4. Realized volatilities, implied volatilities, and news variations

Short-term options, such as those with maturities of 1 or 2 months, tend to be the most liquid ones. To examine the predictive power of implied volatility on realized volatility, we match the tenor of implied volatility with that of realized volatility. For our primary analysis, we use the monthly observations of realized and implied volatility with a 1-month maturity at the end of each month. The choice of monthly observations aligns with the existing literature, including studies conducted by Christensen and Prabhala (1998), Britten-Jones and Neuberger (2000), Chernov (2007), Taylor et al. (2010), and Bekaert and Hoerova (2014).

The annualized realized variance for firm i observed at the end of month t, $\text{RV}_{i,t}$, is computed as follows:

$$\text{RV}_{i,t} = \ln\left(\frac{252}{h_{i,t}}\sum_{d=1}^{h_{i,t}} r_{i,t,d}^2\right) \tag{5}$$

where $r_{i,t,d}$ is the daily log return for firm i on day d in month t and $h_{i,t}$ is the number of trading days for firm i in month t. We require a minimum of 10 trading days in a month to calculate the realized variance. Because we use the log version of variance, the terms variance and volatility (i.e., the square root of variance) are interchangeable in the paper unless an explicit differentiation is necessary. We examine the log version of variance because of its more favorable statistical properties, including a more symmetric distribution with fewer extreme values, compared with the variance itself.

At the end of each month t, we identify the most at-the-money call and put option pairs for each stock i that have the same strike price. This choice is made because at-the-money options tend to be the most liquid.[5] We calculate the average implied volatility across the pairs of call and put options. This averaging reduces errors in implied volatility that may arise from nonsynchronous trading between the underlying stock and options.[6] We select options expiring at the end of month $t+1$ to align the tenor of implied volatility (i.e., 1 month) with that of realized volatility, as calculated above. If options with such maturity dates are not available, we interpolate the implied volatilities of options with maturity dates immediately before and after the end of month $t+1$. Specifically, let $\sigma_{i,t,\tau}$ be the call and put average at-the-money implied volatility for stock i, observed at the end of month t, with a remaining life of τ, and the target time to maturity be τ^*. We select two options with the remaining lives of τ_1 and τ_2 such that $\tau_1 < \tau^* < \tau_2$, and we interpolate the implied volatility as follows:

$$\sigma_{i,t,\tau^*}^2 = \sigma_{i,t,\tau_1}^2 + (\sigma_{i,t,\tau_2}^2 - \sigma_{i,t,\tau_1}^2)\frac{\tau^* - \tau_1}{\tau_2 - \tau_1}. \tag{6}$$

The implied volatility used in the following empirical analysis is the log version, $\text{IV}_{i,t} = \ln(\sigma_{i,t,\tau^*}^2)$.

We measure monthly total news variation, a quantity conceptually similar to realized return variance. The total news variation for firm i in month t is defined as follows:

$$\text{NV}_{i,t} = \ln\left(\overline{\text{CSS}^2} + \sum_{j=1}^{n_{i,t}} \text{CSS}_{j,i,t}^2\right) \tag{7}$$

where $n_{i,t}$ is the number of news stories for firm i in month t, $\text{CSS}_{j,i,t}$ is the CSS for the news event j of firm i in month t, and $\overline{\text{CSS}^2}$ is the average of $\text{CSS}_{j,i,t}^2$ for the entire sample.[7]

Our sample contains 106,584 firm-month observations. Table 3 provides the summary statistics. The implied volatility, $\text{IV}_{i,t}$, has an average of -2.488 (or 28.8% in the annual percentage term), and this value is higher than the average of realized volatility, $\text{RV}_{i,t}$, of -2.707 (or 25.8% in the annual percentage term). This discrepancy is attributable to the volatility risk premium embedded in option prices. The time-series variation in $\text{IV}_{i,t}$ is lower than that in $\text{RV}_{i,t}$. The average total news variation, $\text{NV}_{i,t}$, is -2.321. Both return volatilities and news variation are positively skewed.

[5] We do not use model-free implied volatility because its construction requires out-of-the-money options that are either unavailable or not very liquid for stock options. Taylor et al. (2010) find that for stocks, at-the-money implied volatility generally outperforms model-free implied volatility in forecasting realized volatility.

[6] Implied volatility is calculated using the binomial tree approach in OptionMetrics. This ensures that the early exercise premiums for individual stock options are adjusted for.

[7] The addition of a small positive number, $\overline{\text{CSS}^2}$, makes $\text{NV}_{i,t}$ well-defined when no news is available for firm i in month t (i.e., $n_{i,t} = 0$). This definition of $\text{NV}_{i,t}$ yields the following decomposition that can be used in the subsequent analysis, $\text{NV}_{i,t} = \text{NI}_{i,t} + \text{NM}_{i,t}$, where $\text{NI}_{i,t} = \ln(1 + n_{i,t})$ is the news arrival intensity and $\text{NM}_{i,t} = \ln[(\overline{\text{CSS}^2} + \sum_{j=1}^{n_{i,t}} \text{CSS}_{j,i,t}^2)/(1 + n_{i,t})]$ is the average magnitude of news.

Table 2
Contemporaneous Relationship between Daily News Measures and Stock Price Movements. Panel A reports the results of the time-series regression $r_{i,d} = b_{0,i} + \sum_{k=1}^{3} b_{k,i}\text{CSS}_{i,d+1-k} + \epsilon_{i,d}$ and Panel B reports the results of the time-series regression $r_{i,d}^2 = c_{0,i} + \sum_{k=1}^{3} c_{k,i}\text{SSS}_{i,d+1-k} + \epsilon_{i,d}$ for each stock i, where $\text{CSS}_{i,d}$ is the sum of news level CSS for firm i on day d, $\text{SSS}_{i,d}$ is the sum of squared news level CSS for firm i on day d, $r_{i,d}$ is the log stock return for firm i on day d, and $r_{i,d}^2$ is the squared log return. The average coefficients across stocks multiplied by 100 are reported. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic ≥ 1.96) and negative and significant coefficients (t-statistic ≤ -1.96). The t-statistic is adjusted to account for heteroscedasticity and autocorrelation of 20 lags by using the Newey and West (1987) procedure. The last column shows the average R^2 across stocks. The sample spans from January 2004 to June 2019.

A. Dependent variable: $r_{i,d}$				B. Dependent variable: $r_{i,d}^2$			
$\text{CSS}_{i,d}$	$\text{CSS}_{i,d-1}$	$\text{CSS}_{i,d-2}$	$\bar{R}^2(\%)$	$\text{SSS}_{i,d}$	$\text{SSS}_{i,d-1}$	$\text{SSS}_{i,d-2}$	$\bar{R}^2(\%)$
1.303			2.1	0.466			5.0
[74.6/0.0]				[73.6/0.0]			
1.305	-0.003	-0.022	2.2	0.466	0.023	0.010	5.1
[74.5/0.0]	[5.5/5.6]	[3.8/5.0]		[73.1/0.0]	[2.8/7.0]	[0.8/11.1]	

Table 3
Descriptive Statistics of Implied Volatility, Realized Volatility and News Variation. This table presents the cross-sectional averages of the statistics for the time-series distributions for each firm of the following variables: $\text{IV}_{i,t}$, which denotes the log at-the-money and 1-month maturity implied variance for firm i observed at the end of month t, and $\text{RV}_{i,t}$, which represents the log realized variance for firm i in month t. Both $\text{IV}_{i,t}$ and $\text{RV}_{i,t}$ are expressed in annual terms. $\text{NV}_{i,t}$ is the total variation of news for firm i in month t. The mean, standard deviation (std), skewness (skew), as well as the 5th, 50th, and 95th percentiles (p5, p50, and p95) are reported. The sample spans from January 2004 to June 2019.

	mean	std	skew	p5	p50	p95
$\text{IV}_{i,t}$	-2.488	0.661	0.764	-3.373	-2.576	-1.251
$\text{RV}_{i,t}$	-2.707	0.910	0.585	-4.011	-2.799	-1.049
$\text{NV}_{i,t}$	-2.321	1.416	0.364	-4.210	-2.461	0.109

3. Empirical analysis of the information content of implied volatility

In this section, we examine the dynamics of implied volatility, realized volatility, and news variations through SVARs. We demonstrate the contemporaneous relationship between realized volatility and news variations and examine the predictive power of implied volatility for news variations. Subsequently, we extend our analysis by separating unscheduled news from scheduled news and fundamental news from non-fundamental news. We examine the impacts of different types of news variations on realized volatility and evaluate how effectively implied volatility can predict each type of news variation.

3.1. Analysis based on total news variation

Let $Y_{i,t} = \begin{bmatrix}\text{NV}_{i,t} & \text{RV}_{i,t} & \text{IV}_{i,t}\end{bmatrix}^T$, where $\text{NV}_{i,t}$ is the total news variation in month t, $\text{RV}_{i,t}$ is the realized volatility in month t, and $\text{IV}_{i,t}$ is the implied volatility at the end of month t for firm i.[8] The SVARs are specified as follows:

$$Y_{i,t} = C_i + B_i Y_{i,t} + A_i Y_{i,t-1} + \epsilon_{i,t}, \tag{8}$$

where $C_i = \begin{bmatrix}c_{1,i} & c_{2,i} & c_{3,i}\end{bmatrix}^T$, a vector of intercepts,

$$B_i = \begin{bmatrix} 0 & 0 & 0 \\ b_{21,i} & 0 & 0 \\ b_{31,i} & b_{32,i} & 0 \end{bmatrix},$$

$$A_i = \begin{bmatrix} a_{11,i} & a_{12,i} & a_{13,i} \\ a_{21,i} & a_{22,i} & a_{23,i} \\ a_{31,i} & a_{32,i} & a_{33,i} \end{bmatrix},$$

$$\Sigma_i = \begin{bmatrix} \sigma_{1,i}^2 & 0 & 0 \\ 0 & \sigma_{2,i}^2 & 0 \\ 0 & 0 & \sigma_{3,i}^3 \end{bmatrix},$$

and Σ_i is the covariance matrix of $\epsilon_{i,t} = [\epsilon_{1,i,t}\ \epsilon_{2,i,t}\ \epsilon_{3,i,t}]^T$.

We now clarify some aspects of the model specification. First, B_i represents contemporaneous relationships between the elements of $Y_{i,t}$. To ensure identification, B_i is typically assumed to be a lower triangular matrix, with elements on and above the main diagonal restricted to be zero. When B_i is unrestricted, the system will contain too many parameters and cannot be identified. The restrictions on B_i are equivalent to specifying causal directions between the elements of $Y_{i,t}$. In our context, a nonzero value for $b_{21,i}$ implies that shocks in $\text{NV}_{i,t}$ can affect $\text{RV}_{i,t}$. This is consistent with the evidence that news events affect stock prices presented earlier in the paper. Both $b_{31,i}$ and $b_{32,i}$ are not restricted because $\text{IV}_{i,t}$ at the end of month t can be affected by both $\text{NV}_{i,t}$ and $\text{RV}_{i,t}$ during month t. In other words, investors may form expectations about future realized volatility $\text{RV}_{i,t+1}$ based on $\text{NV}_{i,t}$ and $\text{RV}_{i,t}$ and trade options so that information in $\text{NV}_{i,t}$ and $\text{RV}_{i,t}$ is incorporated in $\text{IV}_{i,t}$. The upper triangular part of B_i is restricted to be zero, indicating that neither $\text{RV}_{i,t}$ nor $\text{IV}_{i,t}$ affects $\text{NV}_{i,t}$ and that $\text{IV}_{i,t}$ does not affect $\text{RV}_{i,t}$. These assumptions are natural and not restrictive in our context. Second, A_i represents the predictive relationship between the elements of $Y_{i,t}$ and $Y_{i,t-1}$. For instance, the element $a_{13,i}$ shows the predictive relationship between implied volatility $\text{IV}_{i,t-1}$ at the end of month $t-1$ on total news variation $\text{NV}_{i,t}$ in month t, which is the coefficient of interest in this study. No restrictions are imposed on any elements of A_i. We consider SVARs with various orders from 1 to 3 and find that the optimal order is 1 according to the average Bayesian information criterion across all firms in our sample. Third, Σ_i is a diagonal matrix, indicating that structural shocks are uncorrelated across equations in (8). The error term $\epsilon_{i,t}$ is also assumed to be uncorrelated across time t.

The average coefficients across firms are shown in Table 4. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic ≥ 1.96) and negative and significant coefficients (t-statistic ≤ -1.96). In the first row, where $\text{NV}_{i,t}$ is the dependent variable, the average coefficient of $\text{IV}_{i,t-1}$ is 0.611. For approximately

[8] $\text{IV}_{i,t}$ is a biased estimator of $\text{RV}_{i,t+1}$ because of the variance risk premium embedded in $\text{IV}_{i,t}$. We do not include the variance risk premium in the model because of the following factors. First, we focus on the informational content of $\text{IV}_{i,t}$ instead of the unbiasedness of $\text{IV}_{i,t}$ as a predictor of $\text{RV}_{i,t+1}$, where the adjustment of the variance risk premium in $\text{IV}_{i,t}$ becomes crucial. Second, the estimates of the variance risk premium can be spanned by $\text{IV}_{i,t}$ and $\text{RV}_{i,t}$.

Table 4
Dynamics of Total New Variation, Realized Volatility, and Implied Volatility. This table reports the results of the SVARs, $Y_{i,t} = C_i + B_i Y_{i,t} + A_i Y_{i,t-1} + \varepsilon_{i,t}$, for each stock i, where $Y_{i,t} = \left[\mathrm{NV}_{i,t} \ \ \mathrm{RV}_{i,t} \ \ \mathrm{IV}_{i,t}\right]^T$, $\mathrm{IV}_{i,t}$ is the log at-the-money and 1-month maturity implied variance for firm i observed at the end of month t, $\mathrm{RV}_{i,t}$ is the log realized variance for firm i in month t, and both are in annual terms. $\mathrm{NV}_{i,t}$ is the total variation of news for firm i in month t. The dependent variable in each regression is indicated under Depvar. The average coefficient across stocks is reported. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic ≥ 1.96) and negative and significant coefficients (t-statistic ≤ -1.96). The sample spans from January 2004 to June 2019.

Depvar	$\mathrm{NV}_{i,t}$	$\mathrm{RV}_{i,t}$	$\mathrm{NV}_{i,t-1}$	$\mathrm{RV}_{i,t-1}$	$\mathrm{IV}_{i,t-1}$
$\mathrm{NV}_{i,t}$			-0.027	-0.367	0.611
			[9.3/15.7]	[0.2/48.5]	[57.6/0.2]
$\mathrm{RV}_{i,t}$	0.089		-0.043	0.156	0.781
	[66.1/0.0]		[0.5/22.8]	[43.0/0.3]	[98.0/0.0]
$\mathrm{IV}_{i,t}$	-0.068	0.292	-0.026	0.104	0.378
	[0.0/81.1]	[98.7/0.0]	[0.8/26.2]	[62.1/0.3]	[96.5/0.0]

58% of firms, this coefficient is positive and statistically significant, indicating that implied volatility has strong predictive power for future total news variation. The average coefficients of $\mathrm{NV}_{i,t-1}$ and $\mathrm{RV}_{i,t-1}$ are negative, which can be attributed to the mean reversion of $\mathrm{NV}_{i,t}$. In months with a high number of news releases, both realized volatility and news variations are elevated. Immediately following busy months with numerous news releases, firms tend to have less news to release, resulting in low news variation. This situation results in negative relationships between $\mathrm{RV}_{i,t-1}$ and $\mathrm{NV}_{i,t}$ and between $\mathrm{NV}_{i,t-1}$ and $\mathrm{NV}_{i,t}$.

In the second row, where realized volatility $\mathrm{RV}_{i,t}$ is the dependent variable, $\mathrm{NV}_{i,t}$ exhibits a positive and significant relationship with realized volatility $\mathrm{RV}_{i,t}$ for 66.1% of the firms. The results shown in Table 2 indicate that the majority of stock price variations can be explained by news variations on the same day, whereas news in the past few days explains a minor proportion of stock price movements. This finding suggests that the contemporaneous relationship observed between monthly realized volatility and news variations is minimally influenced by the lagged response of stock prices to news events. The average negative coefficient of $\mathrm{NV}_{i,t-1}$ is attributable to the mean reversion of $\mathrm{NV}_{i,t}$, as discussed earlier. For 98% (43%) of firms, the coefficient of $\mathrm{IV}_{i,t-1}$ ($\mathrm{RV}_{i,t-1}$) is positive and significant, with an average of 0.781 (0.156). This finding suggests that both $\mathrm{IV}_{i,t-1}$ and $\mathrm{RV}_{i,t-1}$ can predict $\mathrm{RV}_{i,t}$ and that $\mathrm{IV}_{i,t-1}$ is an informationally more efficient predictor than $\mathrm{RV}_{i,t-1}$ is. The predictive power of $\mathrm{IV}_{i,t-1}$ persists even after accounting for contemporaneous news variation $\mathrm{NV}_{i,t}$, suggesting that the ability of implied volatility to forecast realized volatility extends beyond its predictive power on news variation.

The third row shows how implied volatility at the end of month t, $\mathrm{IV}_{i,t}$, is influenced by news variations and realized volatility in the current and preceding months as well as its own lagged value. The average coefficients of $\mathrm{RV}_{i,t}$, $\mathrm{RV}_{i,t-1}$, and $\mathrm{IV}_{i,t-1}$ are all positive. For the majority of firms, these coefficients are positive and statistically significant. These positive coefficients indicate the presence of volatility clustering. The coefficients of $\mathrm{NV}_{i,t}$ and $\mathrm{NV}_{i,t-1}$ are negative, suggesting that $\mathrm{IV}_{i,t}$ incorporates information about the expected $\mathrm{RV}_{i,t+1}$, which is related to the mean reversion property of $\mathrm{NV}_{i,t}$.

The results of SVARs align with the concept of uncertainty resolution. Implied volatility tends to be high before the release of news due to escalated uncertainty. When news is released, it resolves this uncertainty regarding a firm's future prospects, leading to revisions in stock prices and thus generating realized volatility. Once the uncertainty is resolved, both implied and realized volatility decrease. Thus, implied volatility predicts future new variations, which are contemporaneously related to realized volatility.

3.2. Scheduled and unscheduled news

Some news events are scheduled in advance, including earnings and dividends announcements, whereas others are unscheduled, for which the exact timing of occurrence is unknown ahead of time, such as analyst ratings and insider trading. Scheduled news is expected to be easier to predict than unscheduled news. We separate unscheduled news from scheduled news and examine whether implied volatility can predict both types of news.

We depend on the classification of scheduled and unscheduled news provided by RavenPack. RavenPack determines the category of a news event based on its content and the predictability of its timing.[9] The news categories encompass various types, including earnings-above-expectations, dividend-below-expectations, analyst-ratings-change-negative, and credit-rating-downgrade. For the first two categories, since news events are associated with earnings announcements, they are classified as scheduled. For the last two categories, the timing is not known before actual events. Thus, news events in these categories are classified as unscheduled.

We define the total variation in scheduled news for firm i in month t, $\mathrm{NV}^s_{i,t}$, as

$$\mathrm{NV}^s_{i,t} = \ln\left(\overline{\mathrm{CSS}^{s2}} + \sum_{j=1}^{n^s_{i,t}} \mathrm{CSS}^{s2}_{j,i,t}\right) \tag{9}$$

where $n^s_{i,t}$ is the number of scheduled news stories for firm i in month t, $\mathrm{CSS}^s_{j,i,t}$ is the CSS for the scheduled news j for firm i in month t, and $\overline{\mathrm{CSS}^{s2}}$ is the average of $\mathrm{CSS}^{s2}_{j,i,t}$ for all scheduled news in the sample. The total variation in unscheduled news for firm i in month t, $\mathrm{NV}^u_{i,t}$, is defined in a similar manner.

For scheduled news, the arrival timing is known prior to its announcement, but the magnitude of its impact remains unknown.[10] Conversely, for unscheduled news, both the arrival timing and the impact of a news event are unknown ahead of time. Given that the arrival time and impact of scheduled and unscheduled news have different implications for predictability, we decompose the total news variation into the news arrival intensity and average news impact, which represents the magnitude of the news. This decomposition enables us to better understand the sources of the predictive power of implied volatility.

The arrival intensity of scheduled news for firm i in month t, $\mathrm{NI}^s_{i,t}$, is defined as follows:

$$\mathrm{NI}^s_{i,t} = \ln(1 + n^s_{i,t}), \tag{10}$$

where $n^s_{i,t}$ is the number of scheduled news stories for firm i in month t. The magnitude of scheduled news for firm i in month t, $\mathrm{NM}^s_{i,t}$, is defined as follows:

$$\mathrm{NM}^s_{i,t} = \ln\left(\frac{\overline{\mathrm{CSS}^{s2}} + \sum_{j=1}^{n^s_{i,t}} \mathrm{CSS}^{s2}_{j,i,t}}{1 + n^s_{i,t}}\right), \tag{11}$$

where $\mathrm{CSS}^s_{j,i,t}$ is the CSS for the scheduled news j of firm i in month t, and $\overline{\mathrm{CSS}^{s2}}$ is the average of $\mathrm{CSS}^{s2}_{j,i,t}$ for all scheduled news in the sample. The total variation in scheduled news, $\mathrm{NV}^s_{i,t}$, is then decomposed into two components:

[9] A news story occasionally contains multiple events. In such cases, the same news story is duplicated and recorded in various categories. In our sample, 14.1% of news stories have duplicated entries. To prevent double counting, we retain only one news event with the highest event relevance score, which measures the relevance of a news event to the category.

[10] In situations where scheduled news is set to be announced close to the end of a month, the arrival timing may not be known at the end of the previous month when implied volatility is observed.

Table 5
Descriptive Statistics of Scheduled and Unscheduled News. Panel A reports the pooled distribution of the composite sentiment score for scheduled news and unscheduled news j, denoted as CSS^s_j and CSS^u_j, respectively. Panel B presents the cross-sectional averages of the statistics for the time-series distributions for each firm of the following variables: $\mathrm{NV}^s_{i,t}$ ($\mathrm{NV}^u_{i,t}$) is the total variation of scheduled (unscheduled) news, $\mathrm{NI}^s_{i,t}$ ($\mathrm{NI}^u_{i,t}$) is the arrival intensity of scheduled (unscheduled) news, and $\mathrm{NM}^s_{i,t}$ ($\mathrm{NM}^u_{i,t}$) is the magnitude of scheduled (unscheduled) news, for firm i in month t. The mean, standard deviation (std), skewness (skew), as well as the 5th, 50th, and 95th percentiles (p5, p50, and p95) are reported. The sample spans from January 2004 to June 2019.

A. News level						
	mean	std	skew	p5	p50	p95
CSS^s_j	0.009	0.099	-3.106	-0.080	0.000	0.120
CSS^u_j	-0.002	0.116	-1.701	-0.220	0.000	0.120
B. Firm-month level						
	mean	std	skew	p5	p50	p95
$\mathrm{NV}^s_{i,t}$	-3.022	1.410	0.912	-4.333	-3.620	-0.281
$\mathrm{NV}^u_{i,t}$	-2.977	1.178	0.818	-4.321	-3.248	-0.732
$\mathrm{NI}^s_{i,t}$	1.963	1.493	0.184	0.070	1.666	4.201
$\mathrm{NI}^u_{i,t}$	1.877	0.817	-0.024	0.483	1.887	3.204
$\mathrm{NM}^s_{i,t}$	-4.986	0.864	-0.099	-6.466	-4.911	-3.729
$\mathrm{NM}^u_{i,t}$	-4.854	0.852	0.745	-6.012	-4.994	-3.234

$$\mathrm{NV}^s_{i,t} = \mathrm{NI}^s_{i,t} + \mathrm{NM}^s_{i,t}. \tag{12}$$

Similarly, we define the arrival intensity of unscheduled news, $\mathrm{NI}^u_{i,t}$, and the magnitude of unscheduled news, $\mathrm{NM}^u_{i,t}$.

In our sample, the number of scheduled news stories is 2,056,599, which is approximately double the number of unscheduled news stories, 1,046,646. This difference arises because over 58% of news stories in RavenPack pertain to earnings or revenue-related topics, which are predominantly scheduled news events. Panel A of Table 5 presents the pooled distributions of CSSs for both scheduled and unscheduled news at the news story level. On average, the CSS of scheduled (unscheduled) news is slightly positive (negative) at 0.009 (-0.002), suggesting a positive (negative) average price impact of scheduled (unscheduled) news. For both types of news, the CSS is skewed to the left, with scheduled news showing a more negative skewness. Panel B displays the cross-sectional averages of the statistics for the time-series distribution for each firm. On average, the variation in unscheduled news, $\mathrm{NV}^u_{i,t}$, is higher than that in scheduled news, $\mathrm{NV}^s_{i,t}$, but $\mathrm{NV}^s_{i,t}$ is more volatile than $\mathrm{NV}^u_{i,t}$. On average, the arrival intensity of scheduled news, $\mathrm{NI}^s_{i,t}$, exceeds that of unscheduled news, $\mathrm{NI}^u_{i,t}$, indicating a higher frequency of scheduled news events than unscheduled ones. The magnitudes of scheduled and unscheduled news, $\mathrm{NM}^s_{i,t}$ and $\mathrm{NM}^u_{i,t}$, respectively, are similar on average.

We examine the dynamics of the vector $Y_{i,t} = [\mathrm{NV}^s_{i,t}\ \mathrm{NV}^u_{i,t}\ \mathrm{RV}_{i,t}\ \mathrm{IV}_{i,t}]^T$ for each firm i through SVARs as expressed in equation (8).

In this case, C_i is a 4 by 1 vector of intercepts, and B_i is a 4 by 4 contemporaneous coefficient matrix, structured as follows:

$$B_i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ b_{21,i} & 0 & 0 & 0 \\ b_{31,i} & b_{32,i} & 0 & 0 \\ b_{41,i} & b_{42,i} & b_{43,i} & 0 \end{bmatrix},$$

A_i is an unrestricted 4 by 4 coefficient matrix representing 1-month predictive relationships, and $\varepsilon_{i,t}$ is a 4 by 1 vector of structural shocks. Similar to equation (8), B_i is assumed to be a lower triangular matrix for identification. The order of elements in $Y_{i,t}$ indicates causal directions, demonstrating that news measures affect $\mathrm{RV}_{i,t}$ and that both news measures and $\mathrm{RV}_{i,t}$ during month t affect $\mathrm{IV}_{i,t}$ at the end of month t. In addition, we assume no causal relationships between contemporaneous news measures, i.e., $b_{21,i} = 0$.

The first two rows of Table 6 reveal that implied volatility from the previous month, $\mathrm{IV}_{i,t-1}$, predicts the variation in both scheduled and unscheduled news, $\mathrm{NV}^s_{i,t}$ and $\mathrm{NV}^u_{i,t}$, respectively, for a significant portion of firms. In particular, $\mathrm{IV}_{i,t-1}$ predicts $\mathrm{NV}^s_{i,t}$ and $\mathrm{NV}^u_{i,t}$ with positive and significant coefficients for over 60% and 30% of the firms, respectively. $\mathrm{IV}_{i,t-1}$ has a stronger predictive power on $\mathrm{NV}^s_{i,t}$ than on $\mathrm{NV}^u_{i,t}$, likely because of the pre-known timing of scheduled news. Furthermore, a higher $\mathrm{NV}^u_{i,t-1}$ ($\mathrm{NV}^s_{i,t-1}$) tends to predict a higher (lower) $\mathrm{NV}^s_{i,t}$ and $\mathrm{NV}^u_{i,t}$. The third row of the table shows that $\mathrm{NV}^s_{i,t}$ and $\mathrm{NV}^u_{i,t}$ are positive and statistically significant for 25.9% and 56.1% of the firms, respectively, implying that both scheduled and unscheduled news hold strong explanatory power for contemporaneous realized volatility. The explanatory power of a regressor can be measured using the standardized coefficient, which is the product of the coefficient estimate and the standard deviation of the regressor, as shown in Table 5. The explanatory power of $\mathrm{NV}^u_{i,t}$ considerably surpasses that of $\mathrm{NV}^s_{i,t}$, as indicated by the standardized coefficients $0.087 \times 1.178 = 0.1025$ for $\mathrm{NV}^u_{i,t}$ and $0.045 \times 1.410 = 0.0635$ for $\mathrm{NV}^s_{i,t}$. The last row reveals that for the majority of firms, $\mathrm{IV}_{i,t}$ is positively and significantly related to $\mathrm{RV}_{i,t}$, $\mathrm{RV}_{i,t-1}$, and $\mathrm{IV}_{i,t-1}$, demonstrating the phenomenon of volatility clustering. In addition, the result also suggests that negative relationships between $\mathrm{IV}_{i,t}$ and $\mathrm{NV}_{i,t}$ and between $\mathrm{IV}_{i,t}$ and $\mathrm{NV}_{i,t-1}$ shown in Table 4 are mainly driven by the negative relationship between $\mathrm{IV}_{i,t}$ and variations in scheduled news (i.e., $\mathrm{NV}^s_{i,t}$ and $\mathrm{NV}^s_{i,t-1}$).

Table 7 presents the results of SVARs for $Y_{i,t} = [\mathrm{NI}^s_{i,t}\ \mathrm{NM}^s_{i,t}\ \mathrm{NI}^u_{i,t}\ \mathrm{NM}^u_{i,t}\ \mathrm{RV}_{i,t}\ \mathrm{IV}_{i,t}]^T$. We apply the same assumptions to B_i regarding the causality relationships between the elements of $Y_{i,t}$, as in the previous analysis. The first four rows display results related to the predictive power of implied volatility on the arrival intensities and magnitudes of scheduled and unscheduled news. $\mathrm{IV}_{i,t-1}$ is positively and significantly related to $\mathrm{NI}^s_{i,t}$ for 57.4% of the firms, which is expected given the known arrival timing of scheduled news. $\mathrm{IV}_{i,t-1}$ also demonstrates predictive power for $\mathrm{NI}^u_{i,t}$ in 38% of the firms, suggesting that option prices contain information about the arrival of news events that have not yet been publicly announced. However, $\mathrm{NM}^s_{i,t}$ and $\mathrm{NM}^u_{i,t}$ exhibit lower predictability by $\mathrm{IV}_{i,t-1}$. $\mathrm{NI}^s_{i,t}$ is negatively related to its lagged value because of the periodic release of scheduled news, such as earnings announcements. Both $\mathrm{NI}^s_{i,t}$ and $\mathrm{NI}^u_{i,t}$ are positively related to $\mathrm{NI}^u_{i,t-1}$, indicating that unscheduled news can trigger more news releases in the future. $\mathrm{NM}^s_{i,t}$ and $\mathrm{NM}^u_{i,t}$ exhibit little autocorrelation. The fifth row shows that all $\mathrm{NI}^s_{i,t}$, $\mathrm{NM}^s_{i,t}$, $\mathrm{NI}^u_{i,t}$, and $\mathrm{NM}^u_{i,t}$ are positively associated with $\mathrm{RV}^s_{i,t}$ on average, suggesting that all of these variables contain independent information related to realized volatility. News arrival intensities have a stronger association with realized volatility than news magnitudes do, as indicated by higher standardized coefficients, regardless of whether the news is scheduled or unscheduled. The average standardized coefficients of $\mathrm{NI}^s_{i,t}$, $\mathrm{NM}^s_{i,t}$, $\mathrm{NI}^u_{i,t}$, and $\mathrm{NM}^u_{i,t}$ are 0.0612, 0.0196, 0.1227, and 0.0503, respectively. The results in the last row reveal that the negative relationship between $\mathrm{IV}_{i,t}$ and variations in scheduled news shown in Table 6 are mainly driven by the negative relationship between $\mathrm{IV}_{i,t}$ and the arrival intensity of scheduled news.

3.3. Fundamental and non-fundamental news

Stock return volatility can rise due to increases in the volatilities of earnings and sales. In this scenario, changes in volatility are attributable to the firm's fundamentals. In addition, changes in volatility may result from non-fundamental factors, such as institutional and insider trading. We investigate the roles of fundamental and non-fundamental news in determining realized volatility and assess how effectively implied volatility can predict each type of news.

We build upon the RavenPack taxonomy to classify fundamental and non-fundamental news. RavenPack broadly categorizes news events into 48 groups based on the content of the news. Examples of these groups are provided in Table 8. We classify these groups into either fundamental or non-fundamental groups. For the majority of groups, the

Table 6
Dynamics of Total Variation of Scheduled and Unscheduled News, Realized Volatility, and Implied Volatility. This table reports the results of the SVARs, $Y_{i,t} = C_i + B_i Y_{i,t} + A_i Y_{i,t-1} + \varepsilon_{i,t}$, for each stock i, where $Y_{i,t} = \left[\mathrm{NV}^s_{i,t}\ \ \mathrm{NV}^u_{i,t}\ \ \mathrm{RV}_{i,t}\ \ \mathrm{IV}_{i,t}\right]^T$. $\mathrm{NV}^s_{i,t}$ ($\mathrm{NV}^u_{i,t}$) is the total variation of scheduled (unscheduled) news for firm i in month t. The dependent variable in each regression is indicated under Depvar. The average coefficient across stocks is reported. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic ≥ 1.96) and negative and significant coefficients (t-statistic ≤ -1.96). The sample spans from January 2004 to June 2019.

Depvar	$\mathrm{NV}^s_{i,t}$	$\mathrm{NV}^u_{i,t}$	$\mathrm{RV}_{i,t}$	$\mathrm{NV}^s_{i,t-1}$	$\mathrm{NV}^u_{i,t-1}$	$\mathrm{RV}_{i,t-1}$	$\mathrm{IV}_{i,t-1}$
$\mathrm{NV}^s_{i,t}$				-0.195	0.122	-0.39	0.671
				[1.3/68.5]	[30.5/1.2]	[0.2/53.9]	[60.9/0.2]
$\mathrm{NV}^u_{i,t}$				-0.057	0.163	-0.189	0.284
				[1.3/18.1]	[48.6/0.3]	[0.3/24.7]	[31.7/1.2]
$\mathrm{RV}_{i,t}$	0.045	0.087		-0.025	-0.039	0.162	0.774
	[25.9/0.7]	[56.1/0.0]		[0.5/12.3]	[0.2/15.9]	[46.1/0.3]	[97.8/0.0]
$\mathrm{IV}_{i,t}$	-0.078	-0.021	0.293	-0.03	-0.006	0.097	0.389
	[0.0/81.1]	[2.2/18.9]	[98.7/0.0]	[0.7/33.3]	[3.5/7.1]	[56.1/0.3]	[97.5/0.0]

Table 7
Dynamics of Arrival Intensities and Magnitudes of Scheduled and Unscheduled News, Realized Volatility, and Implied Volatility. This table reports the results of the SVARs, $Y_{i,t} = C_i + B_i Y_{i,t} + A_i Y_{i,t-1} + \varepsilon_{i,t}$, for each stock i, where $Y_{i,t} = \left[\mathrm{NI}^s_{i,t}\ \ \mathrm{NM}^s_{i,t}\ \ \mathrm{NI}^u_{i,t}\ \ \mathrm{NM}^u_{i,t}\ \ \mathrm{RV}_{i,t}\ \ \mathrm{IV}_{i,t}\right]^T$. $\mathrm{NI}^s_{i,t}$ ($\mathrm{NI}^u_{i,t}$) is the arrival intensity of scheduled (unscheduled) news, and $\mathrm{NM}^s_{i,t}$ ($\mathrm{NM}^u_{i,t}$) is the magnitude of scheduled (unscheduled) news, for firm i in month t. The dependent variable in each regression is indicated under Depvar. The average coefficient across stocks is reported. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic ≥ 1.96) and negative and significant coefficients (t-statistic ≤ -1.96). The sample spans from January 2004 to June 2019.

Depvar	$\mathrm{NI}^s_{i,t}$	$\mathrm{NM}^s_{i,t}$	$\mathrm{NI}^u_{i,t}$	$\mathrm{NM}^u_{i,t}$	$\mathrm{RV}_{i,t}$	$\mathrm{NI}^s_{i,t-1}$	$\mathrm{NM}^s_{i,t-1}$	$\mathrm{NI}^u_{i,t-1}$	$\mathrm{NM}^u_{i,t-1}$	$\mathrm{RV}_{i,t-1}$	$\mathrm{IV}_{i,t-1}$
$\mathrm{NI}^s_{i,t}$						-0.296	-0.094	0.314	0.081	-0.453	0.663
						[0.8/83.3]	[4.8/21.4]	[54.7/2.2]	[11.6/1.2]	[0.2/59.2]	[57.4/0.8]
$\mathrm{NM}^s_{i,t}$						0.065	0.009	-0.077	-0.015	0.087	-0.018
						[26.4/3.0]	[4.5/4.8]	[2/21.2]	[1.8/5.6]	[16.3/1.8]	[7.3/10.8]
$\mathrm{NI}^u_{i,t}$						-0.028	-0.034	0.314	0.023	-0.182	0.223
						[4.8/17.6]	[3.5/13.3]	[79.3/0.3]	[7.6/3.2]	[0.2/41.6]	[38.0/1.2]
$\mathrm{NM}^u_{i,t}$						-0.062	0.000	0.015	0.024	-0.004	0.078
						[0.8/23.4]	[3.5/3.5]	[6.1/4.1]	[7.0/1.7]	[3.6/4.3]	[11.4/2.3]
$\mathrm{RV}_{i,t}$	0.041	0.024	0.142	0.059		-0.02	0.008	-0.087	-0.018	0.182	0.744
	[28.6/2.5]	[13.2/4.0]	[54.8/0.8]	[21.4/0.3]		[1.3/10.1]	[4.0/2.5]	[0.3/25.9]	[1.0/4.5]	[49.1/0.5]	[97.3/0.0]
$\mathrm{IV}_{i,t}$	-0.077	-0.041	0.006	-0.015	0.316	-0.034	0.003	0.002	-0.016	0.092	0.372
	[7.8/80.1]	[8.6/24.7]	[15.1/19.5]	[5.0/9.9]	[92.7/3.0]	[3.8/46.1]	[4.1/3.5]	[17.4/8.6]	[2.2/6.1]	[50.1/2.2]	[88.7/3.5]

determination of whether a group is fundamental or non-fundamental is straightforward. For instance, the "earnings," "revenues," and "analyst ratings" groups are considered fundamental, whereas "stock prices," "insider trading," and "technical analysis" groups are deemed non-fundamental. However, for a few groups, the classification is less clear. In such cases, if the proportion of fundamental news within a group exceeds half, we classify the group as a fundamental group; otherwise, we consider it as a non-fundamental group. For example, within the "equity actions" group, the largest proportion of news (49.8%) pertains to changes in ownership, which we classify as non-fundamental. The second-largest proportion of news (17.3%) is related to capital expenditure, which we consider fundamental. The third-largest proportion of news (6.28%) is related to changes in trading status, such as halting or resuming trading, which we regard as non-fundamental. Thus, we classify the "equity actions" group as a non-fundamental news group.

As demonstrated in Panel A of Table 8, "earnings" and "revenues" are the top two fundamental groups, accounting for more than half of the total news events. Among the fundamental groups, news on "earnings," "revenues," and "dividends" is predominantly scheduled, whereas news on "analyst ratings" and "acquisitions mergers" is mostly unscheduled. "Equity actions" and "stock prices" are the top two non-fundamental groups, accounting for more than 10% of the total news events, as shown in Panel B. Among the non-fundamental groups, news on "equity actions," "stock prices," and "insider trading" is mainly unscheduled, whereas news on "investor relations" is primarily scheduled. In general, fundamental news tends to be scheduled, whereas non-fundamental news tends to be unscheduled.

We define news measures for fundamental and non-fundamental news separately. $\mathrm{NI}^{sf}_{i,t}$ represents the arrival intensity of scheduled fundamental news for firm i in month t, defined as $\mathrm{NI}^{sf}_{i,t} = \ln\left(1 + n^{sf}_{i,t}\right)$, where $n^{sf}_{i,t}$ is the number of scheduled fundamental news for firm i in month t. $\mathrm{NM}^{sf}_{i,t}$ is the magnitude of scheduled fundamental news, calculated as $\mathrm{NM}^{sf}_{i,t} = \ln\left(\frac{\overline{\mathrm{CSS}^{sf2}} + \sum_{j=1}^{n^{sf}_{i,t}} \mathrm{CSS}^{sf2}_{j,i,t}}{1+n^{sf}_{i,t}}\right)$, where $\overline{\mathrm{CSS}^{sf2}}$ is the average of CSS^{sf2}_j, the squared CSS for the scheduled fundamental news event j, for the entire sample. $\mathrm{NI}^{sn}_{i,t}$, $\mathrm{NI}^{uf}_{i,t}$, and $\mathrm{NI}^{un}_{i,t}$ are the arrival intensities of scheduled non-fundamental news, unscheduled fundamental news, and unscheduled non-fundamental news, respectively. $\mathrm{NM}^{sn}_{i,t}$, $\mathrm{NM}^{uf}_{i,t}$, and $\mathrm{NM}^{un}_{i,t}$ are the magnitudes of scheduled non-fundamental news, unscheduled fundamental news, and unscheduled non-fundamental news, respectively, and all are defined in a similar manner.

Table 9 presents the summary statistics of fundamental and non-fundamental news. The average CSS of fundamental news is lower and exhibits a greater variation than non-fundamental news, regardless of whether it is scheduled or unscheduled. Furthermore, the arrival intensity of fundamental news is higher than that of non-fundamental news, aligning with the results shown in Table 8. Moreover, the magnitude of fundamental news is higher than that of non-fundamental news, regardless of whether it is scheduled or unscheduled.

Panel C of the table presents correlations between the news measures. The total news variation, $\mathrm{NV}_{i,t}$, is included to examine its relationship with various components of the news measure. The corre-

Table 8
Fundamental and Non-fundamental News. This table presents fundamental news groups in Panel A and non-fundamental news groups in Panel B. A brief description of each news group is provided in the second column. The last three columns represent the proportions of the numbers of news events, scheduled news events and unscheduled news events within each respective group, relative to the total number of news events.

A. Fundamental news				
Group	Description	Total(%)	Sch(%)	Unsch(%)
earnings	earnings announcement or news related to earnings expectation	45.11	44.07	1.04
revenues	revenues announcement or news related to revenues expectation	12.87	11.31	1.57
dividends	dividend announcement or dividend guidance announcement	5.72	5.63	0.09
analyst ratings	analyst recommendation or change of recommendation	5.69	0.00	5.69
acquisitions mergers	merger and acquisition events or rumors	3.87	0.43	3.44
products services	sign a new business contract or launch a new product or service, etc.	3.00	0.59	2.41
labor issues	executive appointment, resignation, retirement, salary, board of director appointment, etc.	1.94	0.00	1.94
marketing	participate in or organize a conference or business event, launch a new advertisement campaign	1.30	1.27	0.02
others	including price target, credit rating, asset, partnership, legal, and bankruptcy, etc.	4.68	0.13	4.55
B. Non-fundamental news				
Group	Description	Total(%)	Sch(%)	Unsch(%)
equity actions	actions related to change of ownership, trading status, and etc.	5.57	1.10	4.48
stock prices	stock price rise or drop makes headline	5.30	0.00	5.30
investor relations	schedule a conference call, shareholder meeting, board of director meeting, disclose largest shareholder, etc.	2.74	1.76	0.98
insider trading	insider buy, sell, register to sell, surrender, gift, lawsuit, etc.	1.70	0.00	1.70
technical analysis	Relative Strength Index (RSI), technical price level, bullish or bearish technical view	0.24	0.00	0.24
stock picks	stock recommendation in the headlines	0.13	0.00	0.13
order imbalances	an excess of Market On Close (or Open) buy (or sell) orders cannot be matched with the order of the opposite side	0.11	0.00	0.11
indexes	stock added or removed from the index's components	0.02	0.00	0.02

lations between the components and total news variation are mostly positive, except for the magnitude of scheduled and non-fundamental news, $\mathrm{NM}^{sn}_{i,t}$. The correlations with $\mathrm{NV}_{i,t}$ are generally higher for the arrival intensity of news than for the magnitude of news. Furthermore, the correlations between the components of the total news variation differ based on the nature of the components. The arrival intensities are positively correlated within themselves, and the correlations between news magnitudes are mostly positive. However, the correlations across the two groups of news measures are mostly negative. Overall, the correlations between the components are not strong, indicating that they measure different dimensions of news variations.

We examine the joint dynamics of all the news measures, realized volatility, and implied volatility through SVARs, where $Y_{i,t} = \left[\mathrm{NI}^{sf}_{i,t}\ \ \mathrm{NM}^{sf}_{i,t}\ \ \mathrm{NI}^{sn}_{i,t}\ \ \mathrm{NM}^{sn}_{i,t}\ \ \mathrm{NI}^{uf}_{i,t}\ \ \mathrm{NM}^{uf}_{i,t}\ \ \mathrm{NI}^{un}_{i,t}\ \ \mathrm{NM}^{un}_{i,t}\ \ \mathrm{RV}_{i,t}\ \ \mathrm{IV}_{i,t}\right]^T$. We apply the same identification assumptions regarding causal relationships between the elements of $Y_{i,t}$ as in the previous analysis. We focus on the predictive power of implied volatility on news measures and the explanatory power of news measures on contemporaneous realized volatility. To save space, we report only the coefficients of $\mathrm{IV}_{i,t-1}$ in the regressions of news measures and the coefficients of contemporaneous news measures in the regression of $\mathrm{RV}_{i,t}$ in Table 10. Panel A reports the average coefficients of $\mathrm{IV}_{i,t-1}$. For a large portion of firms, $\mathrm{IV}_{i,t-1}$ positively and significantly predicts all the news intensity variables, regardless of whether the news is fundamental or non-fundamental and scheduled or unscheduled. The predictive relationship is stronger for fundamental news intensity than for non-fundamental news intensity. $\mathrm{IV}_{i,t-1}$ exhibits poor performance in predicting the magnitudes of news. For approximately 29% of firms, $\mathrm{IV}_{i,t-1}$ is negatively and significantly associated with $\mathrm{NM}^{sn}_{i,t}$. Panel B of this table shows that $\mathrm{NM}^{sn}_{i,t}$ has a negative relationship with $\mathrm{RV}_{i,t}$. $\mathrm{NM}^{sn}_{i,t}$ is negatively correlated with the total news variation $\mathrm{NV}_{i,t}$, as shown in Table 9. Thus, the presence of average negative coefficients for $\mathrm{IV}_{i,t-1}$ does not contradict the fact that $\mathrm{IV}_{i,t-1}$ predicts news measures and thus predicts $\mathrm{RV}_{i,t}$.

Panel B shows that for the regression of $\mathrm{RV}_{i,t}$, the average coefficients of news measures are generally positive, excluding $\mathrm{NM}^{sn}_{i,t}$. This finding suggests a predominantly positive association between realized volatility and these components of news variations. $\mathrm{NM}^{sn}_{i,t}$ is the only variable that exhibits a negative correlation with the total news variation $\mathrm{NV}_{i,t}$, as shown in Table 9. The explanatory power of each news measure can be determined by calculating the standardized coefficient (i.e., the product of the coefficient estimate of the news measure and the standard deviation of that news measure) as reported in Table 9. Among the variables related to fundamental news, $\mathrm{NI}^{uf}_{i,t}$ has the highest standardized coefficient, calculated as $0.129 \times 0.726 = 0.1001$. This is followed by $\mathrm{NI}^{sf}_{i,t}$, with a standardized coefficient of $0.047 \times 1.506 = 0.0708$. The explanatory power is non-trivial because the average standard deviation of $\mathrm{RV}_{i,t}$ is 0.910, as reported in Table 3. Among the variables related to non-fundamental news, $\mathrm{NI}^{un}_{i,t}$ has the largest standardized coefficient, calculated as $0.101 \times 0.835 = 0.0843$. This is followed by $\mathrm{NM}^{un}_{i,t}$, with a standardized coefficient of $0.057 \times 0.682 = 0.0389$.

4. Strength of channels through which implied volatility forecasts realized volatility

The results indicated in the previous section suggest that implied volatility forecasts future realized volatility by predicting the arrival intensity and magnitude of future news. We refer to this as the "news channel" through which implied volatility predicts realized volatility. In this section, we quantify the strength of the overall news channel and its individual components by using a multiple mediator model, as proposed by Preacher and Hayes (2008). We allow the strength of this relationship to vary over time for a given firm. For this, we run the following time-series regressions with rolling windows for each firm i,

$$N^{(p)}_{i,t-k} = \phi^{(p)}_{0,i,t} + \phi^{(p)}_{1,i,t}\mathrm{IV}_{i,t-1-k} + \phi^{(p)}_{2,i,t}\mathrm{RV}_{i,t-1-k} + \varepsilon^{(p)}_{i,t-k}, \tag{13}$$

$$\mathrm{RV}_{i,t-k} = \gamma_{0,i,t} + \gamma_{1,i,t}\mathrm{IV}_{i,t-1-k} + \sum_{p=1}^{8} \gamma_{p+1,i,t} N^{(p)}_{i,t-k} + \gamma_{10,i,t}\mathrm{RV}_{i,t-1-k} + \epsilon_{i,t-k}, \tag{14}$$

where $k = 0, 1, \cdots, 35$ (i.e., a window of 3 years) and $N^{(p)}_{i,t}$ for $p = 1, \cdots, 8$ are $\mathrm{NI}^{sf}_{i,t}$, $\mathrm{NM}^{sf}_{i,t}$, $\mathrm{NI}^{uf}_{i,t}$, $\mathrm{NM}^{uf}_{i,t}$, $\mathrm{NI}^{sn}_{i,t}$, $\mathrm{NM}^{sn}_{i,t}$, $\mathrm{NI}^{un}_{i,t}$, and $\mathrm{NM}^{un}_{i,t}$, respectively. These are the mediators in mediation analysis.[11] The rolling regressions are estimated on a quarterly basis.

[11] The mediation analysis differs from a two-stage regression analysis where equation (13) represents the first-stage regression and equation (14) represents the second-stage regression. In the two-stage regression, $N^{(p)}_{i,t}$ for $p = 1, \cdots, 8$ is substituted with the fitted values estimated from equation (13). The two-stage regression analysis does not estimate each individual $\gamma_{p,i,t}$ for $p = 1, \cdots, 9$, which is essential for calculating the mediation effect.

Table 9

Descriptive Statistics of Fundamental and Non-fundamental News. Panel A reports the pooled distribution of the composite sentiment score for news j. Panel B presents the cross-sectional averages of the statistics for the time-series distributions for each firm of the news arrival intensity, $\mathrm{NI}_{i,t}$, and of the news magnitude, $\mathrm{NM}_{i,t}$, for firm i in month t. The superscripts, s, u, f, and n, denote scheduled, unscheduled, fundamental, and non-fundamental, respectively. The mean, standard deviation (std), skewness (skew), as well as the 5th, 50th, and 95th percentiles (p5, p50, and p95) are reported. Panel C presents the cross-sectional average of the time- series correlations between the news measures for each firm. The sample spans from January 2004 to June 2019.

A. News level

	mean	std	skew	p5	p50	p95
CSS_j^{sf}	0.009	0.100	-3.097	-0.080	0.000	0.120
CSS_j^{sn}	0.016	0.068	-2.223	-0.060	0.000	0.100
CSS_j^{uf}	-0.003	0.119	-1.794	-0.220	0.000	0.120
CSS_j^{un}	-0.001	0.109	-1.487	-0.160	0.000	0.120

B. Firm-month level

	mean	std	skew	p5	p50	p95
$\mathrm{NI}_{i,t}^{sf}$	1.902	1.506	0.189	0.060	1.577	4.162
$\mathrm{NI}_{i,t}^{sn}$	0.372	0.536	1.317	0.001	0.069	1.424
$\mathrm{NI}_{i,t}^{uf}$	1.493	0.776	-0.032	0.240	1.522	2.718
$\mathrm{NI}_{i,t}^{un}$	1.018	0.835	0.574	0.009	0.895	2.537
$\mathrm{NM}_{i,t}^{sf}$	-4.930	0.869	-0.166	-6.445	-4.817	-3.706
$\mathrm{NM}_{i,t}^{sn}$	-5.907	0.455	0.456	-6.703	-5.756	-5.540
$\mathrm{NM}_{i,t}^{uf}$	-4.660	0.801	0.658	-5.789	-4.739	-3.154
$\mathrm{NM}_{i,t}^{un}$	-4.998	0.682	0.833	-5.997	-5.034	-4.008

C. Correlation

	$\mathrm{NV}_{i,t}$	$\mathrm{NI}_{i,t}^{sf}$	$\mathrm{NM}_{i,t}^{sf}$	$\mathrm{NI}_{i,t}^{uf}$	$\mathrm{NM}_{i,t}^{uf}$	$\mathrm{NI}_{i,t}^{sn}$	$\mathrm{NM}_{i,t}^{sn}$	$\mathrm{NI}_{i,t}^{un}$
$\mathrm{NI}_{i,t}^{sf}$	0.718							
$\mathrm{NM}_{i,t}^{sf}$	0.075	-0.393						
$\mathrm{NI}_{i,t}^{uf}$	0.603	0.441	-0.148					
$\mathrm{NM}_{i,t}^{uf}$	0.313	-0.022	0.085	-0.095				
$\mathrm{NI}_{i,t}^{sn}$	0.465	0.590	-0.210	0.319	0.013			
$\mathrm{NM}_{i,t}^{sn}$	-0.250	-0.382	0.200	-0.196	0.019	-0.517		
$\mathrm{NI}_{i,t}^{un}$	0.434	0.398	-0.143	0.390	-0.038	0.242	-0.146	
$\mathrm{NM}_{i,t}^{un}$	0.142	0.012	0.088	-0.005	0.077	0.046	-0.002	-0.302

The strength of each mediator through which implied volatility forecasts realized volatility is calculated as the product of the coefficient of $\mathrm{IV}_{i,t-1-k}$ in equation (13), where $N_{i,t-k}^{(p)}$ is the dependent variable, and the coefficient of $N_{i,t-k}^{(p)}$ in equation (14). For example, the strength through the channel $\mathrm{NI}_{i,t-k}^{sf}$ for firm i and month t is calculated as $\phi_{1,i,t}^{(p)}\gamma_{p+1,i,t}$ for $p=1$. The estimates in multiple mediation analysis can be explained as follows: the variation in $\mathrm{RV}_{i,t-k}$ due to a change in $N_{i,t-k}^{(p)}$ is measured by $\gamma_{p+1,i,t}N_{i,t-k}^{(p)}$, as shown in equation (14), and a fraction of the change in $N_{i,t-k}^{(p)}$ is due to the change in $\mathrm{IV}_{i,t-1-k}$, which is measured by $\phi_{1,i,t}^{(p)}\mathrm{IV}_{i,t-1-k}$, as shown in equation (13). Thus, the variation in $\mathrm{RV}_{i,t-k}$ due to the change in $\mathrm{IV}_{i,t-1-k}$ via $N_{i,t-k}^{(p)}$ is the product $\gamma_{p+1,i,t}\phi_{1,i}^{(p)}\mathrm{IV}_{i,t-1-k}$. The variation in $\mathrm{RV}_{i,t-k}$ due to the change in $\mathrm{IV}_{i,t-1-k}$ not via any of $N_{i,t-k}^{(p)}$ for $p=1,\cdots,8$ is $\gamma_{1,i,t}\mathrm{IV}_{i,t-1-k}$, as shown in equation (14). The strength of the channel can be defined through the coefficients of $\mathrm{IV}_{i,t-1-k}$. $\sum_{p=1}^{8}\phi_{1,i,t}^{(p)}\gamma_{p+1,i,t}$ represents the strength of the overall news channel for firm i and month t, and $\gamma_{1,i,t}$ refers to the strength of a non-news-related channel. The overall strength of both news- and non-news-related channels for firm i and month t is measured by $\gamma_{1,i,t}+\sum_{p=1}^{8}\phi_{1,i,t}^{(p)}\gamma_{p+1,i,t}$. The relative strength of the channel through the pth news variation variable is measured by

$$\mathrm{NC}_{i,t}^{(p)}=\frac{\phi_{1,i,t}^{(p)}\gamma_{p+1,i,t}}{\gamma_{1,i,t}+\sum_{p=1}^{8}\phi_{1,i,t}^{(p)}\gamma_{p+1,i,t}}, \tag{15}$$

for $p=1,\cdots,8$.

To ensure the meaningfulness of the news channel strength measures, we apply the following criteria. First, if the coefficient of $\mathrm{IV}_{i,t-1-k}$ in equation (14) is non-positive, we exclude the variable $\mathrm{IV}_{i,t-1-k}$ from the regression (i.e., we set the coefficient of $\mathrm{IV}_{i,t-1-k}$ to be zero). This step ensures that the strength of the news channels does not exceed the overall strength, which includes both news and non-news channels. Second, if the coefficient of $\mathrm{IV}_{i,t-1-k}$ in equation (13) for a given dependent variable $N_{i,t-k}^{(p)}$ is non-positive, $\mathrm{IV}_{i,t-1-k}$ fails to predict $N_{i,t-k}^{(p)}$. Thus, the strength of the news channel through $N_{i,t-k}^{(p)}$ is zero. Therefore, we exclude $N_{i,t-k}^{(p)}$ from equation (14). Third, if a non-positive coefficient of $N_{i,t-k}^{(p)}$ appears in equation (14), it suggests that $N_{i,t-k}^{(p)}$ is not a suitable mediator through which implied volatility positively predicts realized volatility. In this case, we also exclude $N_{i,t-k}^{(p)}$ from equation (14), setting the strength of the news channel through $N_{i,t-k}^{(p)}$ to zero.[12] When multiple combinations of mediators meet the aforementioned criteria, we choose the one with the highest R^2 value in equation (14).

[12] The negative coefficients are mostly statistically nonsignificant. The inclusion of these negative coefficients can result in the strength of a news channel being less than zero or greater than one, which is not interpretable. To avoid such an undesirable property of the strength, we exclude relevant variables from the regressions. Approximately 1% of the coefficients are negative and significant. The inclusion of these coefficients reduces the median strength of news channels by only 1%, and our results remain essentially the same.

Table 10
Dynamics of Arrival Intensities and Magnitudes of Fundamental and Non-fundamental News, Realized Volatility, and Implied Volatility. This table reports the results of the SVARs, $Y_{i,t} = C_i + B_i Y_{i,t} + A_i Y_{i,t-1} + \varepsilon_{i,t}$, for each stock i, where $Y_{i,t} = \left[\mathrm{NI}^{sf}_{i,t} \;\; \mathrm{NM}^{sf}_{i,t} \;\; \mathrm{NI}^{sn}_{i,t} \;\; \mathrm{NM}^{sn}_{i,t} \;\; \mathrm{NI}^{uf}_{i,t} \;\; \mathrm{NM}^{uf}_{i,t} \;\; \mathrm{NI}^{un}_{i,t} \;\; \mathrm{NM}^{un}_{i,t} \;\; \mathrm{RV}_{i,t} \;\; \mathrm{IV}_{i,t}\right]^T$. $\mathrm{NI}_{i,t}$ and $\mathrm{NM}_{i,t}$ denote the news arrival intensity and the news magnitude, respectively, for firm i in month t. The superscripts, s, u, f, and n, denote scheduled, unscheduled, fundamental, and non-fundamental, respectively. The average coefficient across stocks is reported. The numbers in square brackets denote the percentages of positive and significant coefficients (t-statistic $\geq$ 1.96) and negative and significant coefficients (t-statistic $\leq$ -1.96). Panel A shows the coefficients of $\mathrm{IV}_{i,t-1}$ in the regressions of news measures, and Panel B shows the coefficients of contemporaneous news measures in the regression of $\mathrm{RV}_{i,t}$. The sample spans from January 2004 to June 2019.

A. News Measures as Dependent Variable and $\mathrm{IV}_{i,t-1}$ as Independent Variable							
$\mathrm{NI}^{sf}_{i,t}$	$\mathrm{NM}^{sf}_{i,t}$	$\mathrm{NI}^{sn}_{i,t}$	$\mathrm{NM}^{sn}_{i,t}$	$\mathrm{NI}^{uf}_{i,t}$	$\mathrm{NM}^{uf}_{i,t}$	$\mathrm{NI}^{un}_{i,t}$	$\mathrm{NM}^{un}_{i,t}$
0.635	-0.03	0.202	-0.093	0.193	0.029	0.192	0.034
[56.6/0.7]	[6.6/10.8]	[45.3/0.3]	[1.5/28.9]	[31.3/0.8]	[7.3/3.5]	[30.8/1.0]	[8.1/4.3]
B. $\mathrm{RV}_{i,t}$ as Dependent Variable and News Measures as Independent Variables							
$\mathrm{NI}^{sf}_{i,t}$	$\mathrm{NM}^{sf}_{i,t}$	$\mathrm{NI}^{sn}_{i,t}$	$\mathrm{NM}^{sn}_{i,t}$	$\mathrm{NI}^{uf}_{i,t}$	$\mathrm{NM}^{uf}_{i,t}$	$\mathrm{NI}^{un}_{i,t}$	$\mathrm{NM}^{un}_{i,t}$
0.047	0.017	0.057	-0.011	0.129	0.045	0.101	0.057
[38.1/7.1]	[10.6/5.6]	[23.4/11.9]	[10.0/13.4]	[50.1/0.8]	[12.4/0.3]	[40.3/2.3]	[19.7/2.0]

We calculate $\mathrm{NC}^{(p)}_{i,t}$ averaged across month t for each firm i and present the cross-sectional distribution of $\mathrm{NC}^{(p)}_{i}$ in Table 11. The strength of overall news channels, represented by $\sum_{p=1}^{8} \mathrm{NC}^{(p)}_{i}$, accounts for an average of 32.2% of the total strength. In the breakdown, the channel of fundamental news accounts for 18.2% and the channel of non-fundamental news accounts for 14.0%. The most important channel for both fundamental and non-fundamental news is the arrival intensity of scheduled news, accounting for 8.4% for fundamental news and 6.0% for non-fundamental news. This result is intuitive because the arrival of scheduled news is already known prior to the actual announcement and option prices contain the available information in the market. The second most important channel is the arrival intensity of unscheduled news, accounting for 6.3% for fundamental news and 4.7% for non-fundamental news. This result suggests that option prices play an important role in information discovery by incorporating information related to return volatility beyond public announcements. The strengths of the channels associated with the magnitudes of news are relatively low. The channel associated with the magnitudes of unscheduled news is stronger than that associated with scheduled news.

The time-series variation in the strength of all news channels as a proportion of the total strength, represented by $\sum_{p=1}^{8} \mathrm{NC}^{(p)}_{i,t}$, is depicted in the upper panel of Fig. 1. The solid line represents the cross-sectional average, and the dashed lines correspond to the first and third quartiles. The average strength of all news channels fluctuates between 20% and 50%. The strength is higher at the beginning of the sample period, declines to a low level from 2009 to 2012, and then increases again, remaining at a high level since then. The first and third quartiles move in tandem with the average, and the cross-sectional variation is positively related to the average. The lower panel of Fig. 1 shows the 3-year moving average of VIX, the implied volatility of the S&P 500 index. The strength of all news channels tends to move in the opposite direction with VIX. During the periods of high volatility, especially the 2008-2009 financial crisis, numerous factors unrelated to firm-specific news, for example, market-wide liquidity, become more important in determining realized volatility. Thus, the strength of the non-news channel is higher during high volatility periods. Our results are consistent with those reported by Jeon et al. (2022), who document a weaker impact of news on jumps in stock returns during the 2008-2009 financial crisis.

The upper panel of Fig. 2 displays the time-series variation in the strength of the fundamental news channels as a proportion of the total strength, represented by $\sum_{p=1}^{4} \mathrm{NC}^{(p)}_{i,t}$. The middle panel shows the time-series variation in the strength of the non-fundamental news channels, represented by $\sum_{p=5}^{8} \mathrm{NC}^{(p)}_{i,t}$. The solid line represents the cross-sectional average, and the dashed lines correspond to the first and third quartiles. The pattern of the time-series variation in the average strength of fundamental news channels resembles that in the average strength of all news channels, as shown in Fig. 1. However, the overall trend is slightly downward for fundamental news channels, in contrast to the upward trend observed for all news channels. Although the strength of non-fundamental news channels is lower than that of fundamental news channels, it tends to increase over time. The lower panel of Fig. 2 shows that the proportion of non-fundamental news events relative to the total number of news events increases. We identify factors that contribute to the increase in the strength of non-fundamental news channels over time. The increase in the strength can be due to the increase in $\phi^{(p)}_{1,i,t}$ for $p = 1, \cdots, 8$, which indicates the predictive power of implied volatility on the news measures, or the increase in $\gamma_{p+1,i,t}$ for $p = 1, \cdots, 8$, which represents the extent to which the news measures are associated with contemporaneous realized volatility, or both. We find that both $\phi^{(p)}_{1,i,t}$ and $\gamma_{p+1,i,t}$ related to non-fundamental news measures increase over time, with a more pronounced increase observed in $\phi^{(p)}_{1,i,t}$. These results are consistent with the explanation that non-fundamental news is more important in determining realized volatility and that option prices incorporate more information about future non-fundamental news.

5. Conclusion

This study uses a comprehensive database of news events to examine volatility-related information embedded in the prices of stock options. We contribute to the literature by constructing news measures from various dimensions, namely the arrival intensity of news versus the magnitude of news, scheduled news versus unscheduled news, and fundamental news versus non-fundamental news. We examine the joint dynamics of these news measures, realized volatility, and implied volatility through SVARs. We find that most of these news measures exhibit a positive and significant association with contemporaneous realized volatility. Among these news measures, the explanatory power of the arrival intensities of scheduled and unscheduled fundamental news as well as the arrival intensity of unscheduled non-fundamental news is the highest. We determine that implied volatility positively predicts many of these news measures, with predictive power being particularly

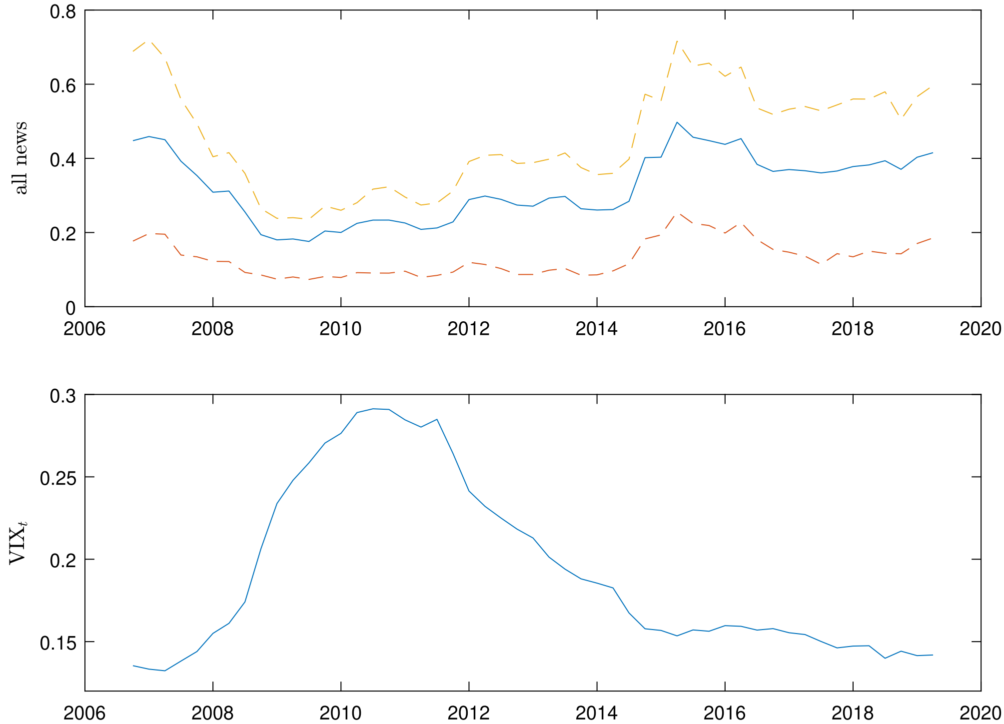

Fig. 1. Strength of All News Channel. The upper panel displays the quarterly strength of all news channels as a proportion of the total strength in a 3-year rolling window. The solid line represents the cross-sectional average, and the dashed lines correspond to the first and third quartiles. The lower panel illustrates the 3-year moving average of VIX. The sample period is from January 2004 to June 2019.

strong for the news arrival intensities. The finding of implied volatility predicting the arrival intensities of scheduled news is expected because the occurrence of these news events is mostly known before the actual announcements. In addition, we find strong predictability for the arrival intensity of unscheduled news. This result suggests that option prices incorporate both public and private information related to the return volatility of underlying stocks. However, predicting the magnitudes of news is generally challenging.

By performing a mediation analysis using these news measures as mediators, we quantify the strength of the news channel through which implied volatility predicts realized volatility. We find that approximately one third of the overall predictive power of implied volatility on future realized volatility is through the channel in which implied volatility predicts the news measures. The strength of this channel is greater for the arrival intensity of news than for the magnitude of news and is greater for fundamental news than for non-fundamental news. The results of the time-series analysis suggest that the average strength of the news channel is negatively related to market volatility. Moreover, the strength of the non-fundamental news channel relative to that of the fundamental news channel increases over time.

This study offers insights into why option-implied volatility outperforms historical volatility in forecasting future volatility, which is a well-documented but not well-explained empirical result in the literature. Historical volatility and volatility models, such as GARCH and stochastic volatility models, can predict future volatility because of the inherent persistence and mean reversion of volatility. Information contained in these forecasts is based on historical stock returns. However, implied volatility incorporates not only historical information but also public and private information about future events. This additional information explains why implied volatility is a more efficient predictor of future volatility compared with alternative volatility forecasting approaches.

Table 11
Strength of News Channels through which Implied Volatility Predicts Realized Volatility. This table reports the strength of various news channels through which the implied volatility predicts the realized volatility. The strength of each channel is measured as its proportion out of all channels, which include both news and non-news channels. The mean, standard deviation (std), skewness (skew), the 5th, 50th, and 95th percentiles (p5, p50, and p95) of the cross-sectional distribution across stocks are reported. NI_i represents the news arrival intensity, and NM_i denotes the news magnitude. The superscripts, s, u, f, and n, denote scheduled, unscheduled, fundamental, and non-fundamental, respectively. Fundamental, Non-fundamental and Overall indicate the strength through fundamental news channels, through non-fundamental news channels, and through all news channels, respectively. The sample spans from January 2004 to June 2019.

	mean	std	skew	p5	p50	p95
NI_i^{sf}	8.4%	6.7%	1.581	0.8%	7.0%	21.4%
NM_i^{sf}	1.1%	1.7%	3.197	0.0%	0.5%	4.2%
NI_i^{uf}	6.3%	5.2%	1.539	0.7%	5.1%	16.4%
NM_i^{uf}	2.5%	2.6%	1.922	0.0%	1.6%	7.8%
Fundamental	18.2%	9.0%	1.021	6.1%	16.7%	34.6%
NI_i^{sn}	6.0%	5.6%	1.427	0.3%	4.4%	17.3%
NM_i^{sn}	0.6%	1.3%	5.222	0.0%	0.1%	2.2%
NI_i^{un}	4.7%	4.7%	1.958	0.2%	3.2%	14.1%
NM_i^{un}	2.7%	2.7%	1.690	0.1%	1.9%	8.2%
Non-fundamental	14.0%	7.8%	0.903	4.1%	12.4%	28.6%
Overall	32.2%	12.2%	0.525	14.5%	31.9%	54.7%

CRediT authorship contribution statement

Sipeng Chen: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Gang Li:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

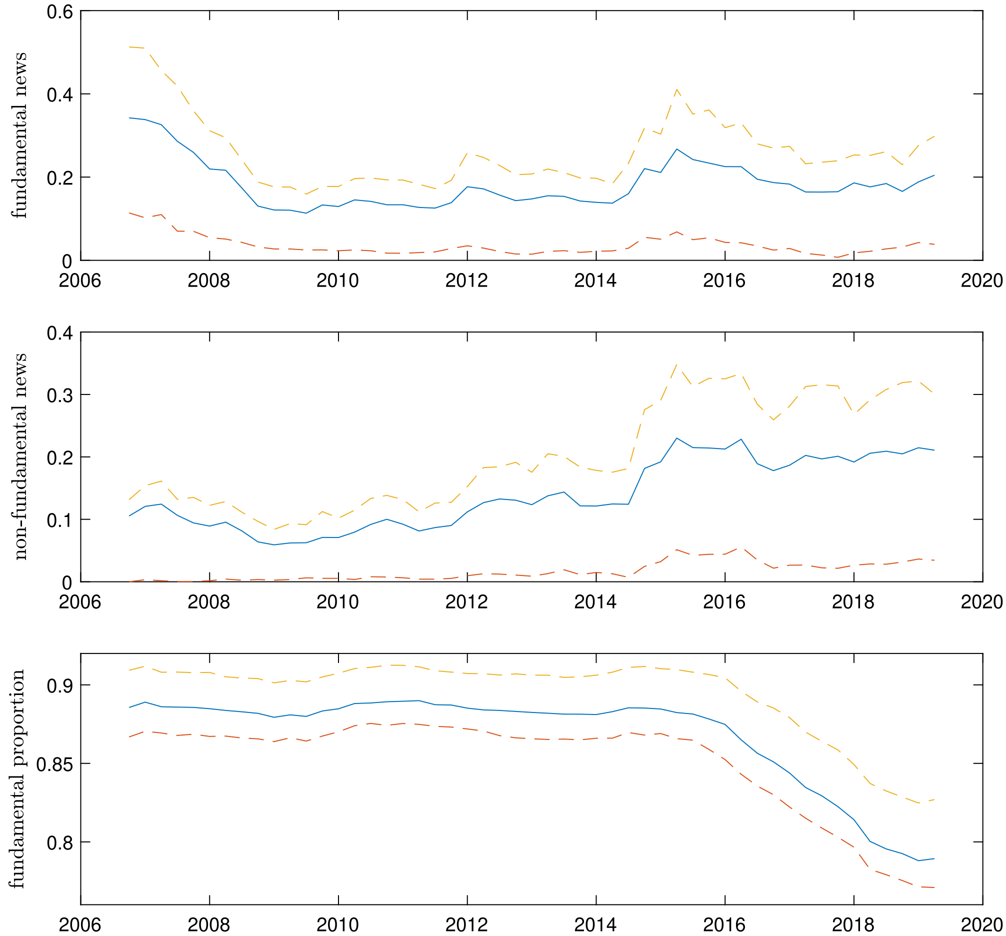

Fig. 2. Strength of Fundamental and Non-fundamental News Channels. The upper (middle) panel displays the quarterly strength of fundamental (non-fundamental) news channels as a proportion of the total strength, and the lower panel shows the proportion of the number of fundamental news events, relative to the total number of news events, in a 3-year rolling window. The solid line represents the cross-sectional average, and the dashed lines correspond to the first and third quartiles. The sample period is from January 2004 to June 2019.

References

Acharya, V.V., Johnson, T.C., 2010. More insiders, more insider trading: evidence from private-equity buyouts. J. Financ. Econ. 98, 415–650.

Augustin, P., Brenner, M., Subrahmanyam, M.G., 2019. Informed options trading prior to takeover announcements: insider trading? Manag. Sci. 65, 5697–5720.

Augustin, P., Brenner, M., Hu, J., Subrahmanyam, M.G., 2020. Are corporate spin-offs prone to insider trading? Crit. Finance Rev. 9, 115–155.

Atilgan, Y., 2014. Volatility spreads and earnings announcement returns. J. Bank. Finance 38, 205–215.

Bali, T.G., Hovakimian, A., 2009. Volatility spreads and expected stock returns. Manag. Sci. 55, 1797–1812.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. J. Econom. 183, 181–192.

Britten-Jones, M., Neuberger, A., 2000. Option prices, implied price processes, and stochastic volatility. J. Finance 55, 839–866.

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., 2019. Information, trading, and volatility: evidence from firm-specific news. Rev. Financ. Stud. 32, 993–1033.

Calomiris, C.W., Mamaysky, H., 2019. How news and its context drive risk and returns around the world. J. Financ. Econ. 133, 299–336.

Cao, C., Chen, Z., Griffin, J., 2005. Informational content of option volume prior to takeovers. J. Bus. 78, 1073–1109.

Cao, C., Yu, F., Zhong, Z., 2010. The information content of option-implied volatility for credit default swap valuation. J. Financ. Mark. 13, 321–343.

Chan, K., Ge, L., Lin, T.-C., 2015. Informational content of options trading on acquirer announcement return. J. Financ. Quant. Anal. 50, 1057–1082.

Chernov, M., 2007. On the role of risk premia in volatility forecasting. J. Bus. Econ. Stat. 25, 411–426.

Chiras, D.P., Manaster, S., 1978. The information content of option prices and a test of market efficiency. J. Financ. Econ. 6, 213–234.

Christensen, B.J., Prabhala, N.R., 1998. The relation between implied and realized volatility. J. Financ. Econ. 50, 125–150.

Ederington, L.H., Lee, J.H., 1993. How markets process information: news releases and volatility. J. Finance 48, 1161–1191.

Engle, R.F., Hansen, M.K., Karagozoglu, A.K., Lunde, A., 2021. News and idiosyncratic volatility: the public information processing hypothesis. J. Financ. Econom. 19, 1–38.

Fahlenbrach, R., Sandås, P., 2010. Does information drive trading in option strategies? J. Bank. Finance 34, 2370–2385.

Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: a new measure. J. Futures Mark. 15, 265–302.

French, K., Roll, R., 1986. Stock return variances: the arrival of information and the reaction of traders. J. Financ. Econ. 17, 5–26.

Glasserman, P., Mamaysky, H., 2019. Does unusual news forecast market stress? J. Financ. Quant. Anal. 54, 1937–1974.

Han, B., Li, G., 2021. Information content of aggregate implied volatility spread. Manag. Sci. 67, 1249–1269.

Hayunga, D.K., Lung, P.P., 2014. Trading in the options market around financial analysts' consensus revisions. J. Financ. Quant. Anal. 49, 725–747.

Jeon, Y., McCurdy, T.H., Zhao, X., 2022. News as sources of jumps in stock returns: evidence from 21 million news articles for 9000 companies. J. Financ. Econ. 145, 1–17.

Jiang, G.J., Tian, Y.S., 2005. The model-free implied volatility and its information content. Rev. Financ. Stud. 18, 1305–1342.

Jones, C.M., Lamont, O., Lumsdaine, R.L., 1998. Macroeconomic news and bond market volatility. J. Financ. Econ. 47, 315–337.

Kalev, P.S., Liu, W.-M., Pham, P.K., Jarnecic, E., 2004. Public information arrival and volatility of intraday stock returns. J. Bank. Finance 28, 1441–1467.

Latané, H.A., Rendleman, R.J., 1976. Standard deviations of stock price ratios implied in option prices. J. Finance 31, 369–381.

Manela, A., Moreira, A., 2017. News implied volatility and disaster concerns. J. Financ. Econ. 123, 137–162.

Neuhierl, A., Scherbina, A., Schlusche, B., 2013. Market reaction to corporate press releases. J. Financ. Quant. Anal. 48, 1207–1240.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Ni, S., Pan, J., Poteshman, A., 2008. Volatility information trading in the option market. J. Finance 63, 1059–1091.

Oikonomou, I., Stancu, A., Symeonidis, L., Wese Simen, C., 2019. The information content of short-term options. J. Financ. Mark. 46, 100504.

Poon, S.-H., Granger, C., 2003. Forecasting volatility in financial markets: a review. J. Econ. Lit. 41, 478–539.

Preacher, K.J., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav. Res. Methods 40, 879–891.

Roll, R., Schwartz, E., Subrahmanyam, A., 2010. O/S: the relative trading activity in options and stock. J. Financ. Econ. 96, 1–17.

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. J. Bank. Finance 34, 871–881.

Xing, Y., Zhang, X., Zhao, R., 2010. What does individual option volatility smirk tell us about future equity returns? J. Financ. Quant. Anal. 45, 641–662.