Journal of Financial Economics 145 (2022) 1–17

Contents lists available at ScienceDirect

Journal of Financial Economics

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

News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies☆

Yoontae Jeon[a], Thomas H. McCurdy[b,*], Xiaofei Zhao[c]

[a] Ted Rogers School of Management, Ryerson University, 350 Victoria St, Toronto, ON M5B 2K3, Canada
[b] Rotman School of Management, University of Toronto, 105 St George St, Toronto, ON M5S 3E6, Canada
[c] McDonough School of Business, Georgetown University, 3700 O St NW, Washington, DC 20057, USA

ARTICLE INFO

Article history:
Received 8 February 2021
Revised 22 May 2021
Accepted 18 June 2021
Available online 12 August 2021

JEL classification:
G10
G14
G19

Keywords:
Jumps
News frequency
Textual analysis
News content
Sentiment

ABSTRACT

Material news events can be potentially important sources of jumps in stock returns. We collect 21 million news articles associated with more than 9000 publicly-traded companies and use textual analyses to derive measures to summarize the news. We find that stock return jumps (including time-variation in jump-size distributions and jump intensity) are significantly related to news flow frequency and content and those effects increase substantially over the last few decades. The sensitivity of jump probability to news is stronger for firms with higher media visibility, analyst coverage, and institutional ownership. This sensitivity also varies across different news categories.

1. Introduction

There is a long history of asset pricing theory that links the quantity and quality of information flows to changes in asset prices. For example, information that results in a resolution of uncertainty about a firm's future prospects can result in a revision in the current price. According to this view, an important process affecting price movements is the news arrival process. Nevertheless, the fundamental question concerning news sources of large price movements in capital markets, typically labeled as "jumps," remains relatively under explored. Recently, however, Baker et al., 2021 explore the news source of jumps at the market level in 16 different countries. Their results show that the news source of stock market jumps, and clarity of that source, matters for future volatility and that the clarity has increased over time. Gűrkaynak et al. (2020) analyze a framework in which jumps in asset prices around

☆ Bill Schwert was the editor for this article. We are grateful for very helpful suggestions from an anonymous referee and the editor. We thank Martin Boyer, Stephen Figlewski (discussant), Luis Goncalves-Pinto (discussant), Ryan Israelson (discussant), Kristine Hankins, Tim Loughran, Sumudu Watugala, Dacheng Xiu and participants at the 2019 Risk Management and Financial Innovation Conference, 2019 European Financial Association Annual Meeting, 2019 Northern Finance Association Annual Meeting, and the 2019 CAAA Annual Meeting for helpful comments and suggestions. We also thank Tim Loughran and Bill McDonald for making their sentiment word list and master dictionary available; and acknowledge financial support from the Social Sciences and Humanities Research Council Insight Grant number 435-2020-1333.

* Corresponding author.
E-mail addresses: yoontae.jeon@ryerson.ca (Y. Jeon), tom.mccurdy@rotman.utoronto.ca (T.H. McCurdy), xiaofei.zhao@georgetown.edu (X. Zhao).

https://doi.org/10.1016/j.jfineco.2021.08.002

macroeconomic news announcements reflect both the response to headline news and to latent factors that reflect other non-headline news in the release. They find that accounting for the non-headline news is crucial for explaining yield curve variations. We contribute to this growing line of research by analyzing a comprehensive news dataset of 21 million news articles, and analyzing how the (contents of) news flows affect return jumps for over 9000 publicly-traded individual stocks.

Jumps in asset prices have been recognized as important for many financial and economic decisions, such as portfolio re-balancing, derivatives pricing, and risk measurement and management. The intuitive idea that large movements in stock prices might be related to important information flows (such as earnings surprises) in the market has inspired many studies related to modelling jumps in stock returns, mostly treating information flow as latent.

In this paper, we analyze the relationship between news processes and stock return jumps, rather than treating that relationship as latent. We focus on nonparametric analyses of firm-specific news. In particular, after collecting a comprehensive set of news articles from the Factiva database, we use textual analyses to derive measures summarizing those news, including news frequency, tone, and uncertainty.[1] We then test to what extent nonparametric measures of jumps are related to those news variables. We analyze important determinants of the cross-sectional news-jump relationships, especially as they relate to firm characteristics (e.g., size, sector, institutional ownership, analyst coverage), news categories, and news coverage, as well as how those relationships have changed over time. We also use news measures from RavenPack where feasible for robustness checks and to extend our sample period to the end of 2020 for some of the analyses.

Given our comprehensive Factiva data, one contribution of our paper is a wealth of descriptive statistics about both news frequency and content (tone and uncertainty). For example, the number of news articles across firms is heavily skewed towards large firms.[2] Our textual analysis reveals a slightly negative news tone and moderate news uncertainty on average, and there is considerable variation in these news measures across individual companies. Although data intensive, our paper features four encompassing highlights on how news flows affect stock return jumps.

Combining the news measures with nonparametric identification of jumps in individual stock returns reveals the first and central results of the paper. Specifically, we show that the news measures are significantly related to jump probability/intensity and jump return characteristics (jump size mean and volatility).[3] For example, there is a significantly higher likelihood of jumps in daily stock returns when the news frequency and the absolute value of the news tone are higher on a particular day, with news frequency playing a dominant role. Conditional on at least one jump in daily returns, we find that positive jump returns are positively related to news intensity, whereas negative jump returns become more negative on average as news intensity increases. The other characteristic of jump-size distributions, jump size volatility, is positively and significantly related to both news frequency and the absolute value of the news tone. In general, our results suggest that the measures of news flow explain an important fraction of the variations in jump size distributions, and news intensity plays an especially critical role in driving jump probability.

Notably, given our long sample period of news, a second highlight is that we are able to uncover a new time trend in the effect of news on stock return jumps. Our results suggest that the impact and explanatory power of news flow on the stock return jumps have increased over time. For example, the coefficient estimate for the effect of news count on the probability of jumps has increased five-fold from 1980 to 2012 (from 0.1 to over 0.5).[4] We relate the increased price informativeness of news in the tail of the return distributions to a broad trend of improved data provision and transparency, and changes in technology and accounting practices. For example, the creation of EDGAR in 1993 — an improvement in information dissemination technology (e.g., Gao and Huang, 2020; Goldstein et al., 2020) — as well as the wide adoption of the Internet after 2000 could be two contributing factors. Extending our sample period to the end of 2020 using RavenPack data reveals that this trend has continued with interesting exceptions during the crisis periods of 2008–2009 and 2020. This discovery of the improvement in (tail) price informativeness for individual stock returns, utilizing daily news flow information, complements the increased "clarity" over time associated with market-level jumps reported by Baker et al., 2021.

Thirdly, after establishing the news-jump relationship, we further investigate in the cross-section how the sensitivity of jump probability to news is determined by firm characteristics. In univariate regressions, we find that the jump sensitivity to news is higher for larger firms or firms in more transparent and visible information environments (such as more analyst coverage, more media visibility, and higher institutional ownership), highlighting the importance of those channels with respect to quick incorporation of news into returns. However, the firm characteristic variables are correlated and in multivariate regressions, the media visibility, analyst coverage, and institutional

[1] Our analysis builds on the word list provided by Loughran and McDonald (2011). See Loughran and McDonald (2016) for a recent survey on textual analysis in accounting and finance. There is a growing literature on using textual analysis to understand the content of news or regulatory filings. Also see, for example, Gentzkow et al. (2019) and Kelly et al. (2019).

[2] For example, when sorting all the firms into three firm size groups (large, medium, and small), we find that the group of large firms is covered by 86% (or 18.49 million) of all news articles, and a representative group of 20 large firms alone is covered by 3.16 million news articles.

[3] Under the most stringent criteria using the Lee and Mykland (2008) methodology, there is one jump every 57 days on average across all firms. As discussed below, this is equivalent to identifying a daily jump if the absolute value of the daily return exceeds 5.1 times the time-varying daily spot volatility. For robustness, we also provide results for alternative methods of identifying jump days.

[4] In untabulated analyses, we find a similar increasing trend for the R^2 of those regressions that R^2 has increased from 1% in 1980 to 9% in 2012.

ownership remain the three most important determinants of the jump sensitivity to news.

Another cross-sectional variation we explore is the heterogeneous impact of the news categories on jump probabilities. This leads to our fourth main finding that, among ten major categories of news classifications, news associated with earnings and revenues, analyst ratings, capital structure changes, mergers and acquisitions, and marketing and investor relations are the most influential factors associated with stock return jumps.[5] In addition, we find that within each of the ten news categories, the subcategories of news also affect jumps differently. For example, while marketing and investor relations is one of the most important news categories, only conference call and shareholder disclosure related news in that category are statistically significant at the sub-category level. Overall, the results highlight the value of investigating the impact by both news categories and sub-categories.

In addition, we conduct many robustness tests, including using alternative definitions to identify jumps and using news articles that only mention a single firm. The RavenPack data also allow us to conduct additional robustness analyses, including using novel news instead of the entire set of news. Our results are robust to these alternatives.

Our paper adds to the large literature concerned with the effect of news on stock returns.[6] Recent research links news flows to the distributional properties of individual stock returns. For example, Engle et al. (2021) show that public-news counts are related to the volatility of stock returns for 28 large firms. Boudoukh et al. (2019) use textual analysis to identify the effects of news that are associated with fundamentals as opposed to information revealed by trading; and test these alternative sources of return volatility. In a study of 23 firms, Lee (2012) finds that jumps in returns are more likely to occur during scheduled news announcements.

Our study differs from these related papers in significant ways. For example, our sample is more comprehensive than Lee (2012) and Engle et al. (2021); we have over 21 million news articles associated with more than 9000 firms. Therefore, our paper provides insight on the impact of news on jumps for large, medium, and small companies, the cross-sectional determinants, and the heterogeneous effects of news on jumps. Although both Lee (2012) and Engle et al. (2021) focus on news arrival or counts, we include not only the frequency of the news articles, but also the textual content measures of the news articles, and investigate the impact of news on the distribution of stock return jumps that include jump intensity, jump-size mean, and jump-size volatility.[7]

Our analyses of the relationships between news flow measures and jumps in stock returns for a large panel could enrich the economic content of a large class of models of stock return jumps that typically treat the sources of jumps as latent. Explicitly incorporating news processes in models of stock return jumps can potentially help identify jumps due to information arrival as opposed to jumps due to other reasons such as liquidity and strategic trading based on private information. This may have broad implications for applications where stock return jump models are frequently used.[8]

In Section 2, we discuss our data collection and the summary statistics of those data including news frequency measures, tone, and uncertainty measures of the news based on textual analyses, and nonparametric measures of realized jumps in stock returns. In Section 3, we provide a detailed analysis of relationships between the realized jumps and the various measures of news flow, both with respect to jump frequency and statistics of the realized jump-size distribution. In Section 4, we analyze the cross-sectional determinants of the jump probability sensitivity to news variables, as well as the heterogeneous impact of disaggregated news categories on jump probability. In Section 5, we report a number of robustness analyses. Concluding comments are in Section 6.

2. Data

Our primary news dataset is constructed by accessing news from Factiva. Our initial search for news articles started in July 2012, so we fixed this date as the ending date for our Factiva searches, resulting in a Factiva news sample period from January 1980 to July 2012. This search returned 9020 companies with at least one news article in Factiva. In order to conduct robustness analyses, and to extend our time period, we also accessed news from 2000 to 2020 from the RavenPack news database. The corresponding stock return data were retrieved from CRSP.

The Factiva news collection involves the following steps. First, we start with the public companies from the CRSP/Compustat merged database. For each of these companies, we search for its CRSP/Compustat company name in Factiva to obtain the corresponding Factiva company name and company code. We then match CRSP/Compustat IDs with Factiva IDs, which allow us to obtain news articles for individual companies.

[5] This analysis is based on our sample of RavenPack news (2000–2020) for which we have commercially generated news classifications.

[6] For example, research focuses on the effect of news on returns or volatility includes, among many others, Cutler et al. (1989), Berry and Howe (1994), Mitchell and Mulherin (1994), Chan (2003), Antweiler and Frank (2004), Veldkamp (2006), Tetlock (2007), Tetlock et al. (2008), Fang and Peress (2009), Tetlock (2010), Tetlock (2011), Kyle et al., 2011, Neuhierl et al. (2013), Baker et al. (2016), Manela and Moreira (2017), Zhao (2017), Bali et al. (2018), Bybee et al., 2019, Ke et al., 2019, Baker et al., 2019, and Jiang et al. (2021).

[7] Engle et al. (2021) focus on the impact of news counts on stock return volatilities. The firms in Lee (2012) have an average of 49 *scheduled* news events every year over the sample period of 1993–2008. In contrast, our top 20 firms in the Online Appendix Table A.1, which are comparable to the firms in Lee (2012), have an average of 5499 news articles every year (daily mean of 21.82 times 252 trading days) during our sample period from 1980 to 2012, covering a much broader sets of news and corporate events.

[8] These include studies on pricing jump risk for returns, such as Maheu et al. (2013), as well as studies that fit options and the underlying assets jointly with a focus on solving option pricing puzzles, as in Pan (2002), Christoffersen et al. (2012), and Bégin et al. (2019). Our analyses suggest that asset pricing models of jumps may be enriched by incorporating the observed news measures explicitly.

Table 1
Summary statistics.
Panel A reports summary statistics for news counts compiled from the Factiva database for all firms and firms in three size groups with an equal number of firms in each group, sorted by their market capitalization at the end of the sample period. Panel B reports summary statistics for daily news tones and the percentage of uncertain words (times 1,000). Both variables are constructed by analyzing the first paragraph of each news article. We calculate the percentage of positive, negative, and uncertain words using the positive, negative, and uncertain word list from Loughran and McDonald (2011), respectively. The news tone is calculated as the difference between the percentage of positive and negative words. Then, the tone and uncertainty of individual articles are aggregated at the daily level using the total number of words in each article as weights. Panel C reports summary statistics for daily realized jumps. The daily return jump indicator is identified using 4 alternative statistics. The $J99$ and $J95$ indicators use the statistic from Lemma 1 of Lee and Mykland (2008) at 99% and 95% significance, respectively. The $J_0 99$ and $J_0 95$ indicators use a less tight bound from the normal distribution as in Theorem 1 of Lee and Mykland (2008). We also use the correction term in Gilder et al. (2014). Each of four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a jump day if the absolute value of daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility. The sample period is from January 1980 to July 2012.

	Panel A: News counts summary statistics				
Company	Total	Mean	Median	Std. Dev.	% Post-2000
All Firms	21,510,023	1.06	0	10.45	79.48%
Large Firms	18,490,700	1.95	0	14.88	80.44%
Medium Firms	2,208,347	0.37	0	3.86	77.71%
Small Firms	810,976	0.17	0	1.36	62.29%
	Panel B: News tones summary statistics				
		Average Tone	Std. Dev. Tone	Average Uncertain Words	Std. Dev. Uncertain Words
All Firms		-0.315	17.122	2.185	14.661
Large Firms		-0.506	18.445	2.959	14.817
Medium Firms		0.042	15.074	1.519	13.318
Small Firms		-0.385	16.801	1.480	15.844
	Panel C: Realized jumps summary statistics				
	N	Jump Days $J99$	Jump Days $J95$	Jump Days $J_0 99$	Jump Days $J_0 95$
All Firms	20,079,694	351,374	480,489	1,096,893	2,055,081
Large Firms	9,426,014	117,683	172,332	445,746	894,227
Medium Firms	5,950,623	111,943	152,270	341,381	627,889
Small Firms	4,703,057	121,748	155,887	309,766	532,965

Second, for each of the companies with a Factiva company code, we search for the total number of news articles available in the Factiva database for our sample period. The total number of news articles is helpful in determining our search strategy for individual companies in the next step.

Third, we retrieve the news items for each Factiva firm. Due to the quota limitation of Factiva, we break down the searches for firms with a large number of news items into searches by firm-month. We then use a set of Python scripts to organize all the news items retrieved from Factiva. For each news item, we obtain the headline (title) and the first paragraph (but not the full text due to data downloading limitations), the date of the news article, the media outlet where the news article is published, and the total number of words in the article. We impose a Factiva filter to exclude the news articles discussing the market and stock price movements. This filter helps to alleviate concerns about reverse causality of stock return jumps triggering the news reports.

Panel A of Table 1 presents the summary statistics of news counts for our sample, including for all firms as a group and for firms in three size groups. There are 21.51 million news articles for these firms. The number of news articles across firms is heavily skewed towards large firms.[9] The Online Appendix Table A.1 reports the same information for the 20 individual firms selected from different industries with the most news coverage, which we label “Top 20 News Firms.”[10] These top 20 firms alone are discussed in 3.16 million news articles. Not surprisingly, news articles concentrate in the post 2000 period. For all firms, close to 80% of the news articles are in the post 2000 period, although this coverage is smaller for small firms (62%) and much higher for the top 20 firms (with a few exceptions). For the top 20 firms, on average 84% of the news articles are in the post 2000 period.

We analyze the textual content of the first paragraph of the news articles to obtain our key textual measures (such as news tone and percentage of uncertain words) using the word lists developed by Loughran and McDonald (2011, 2013).[11] Specifically, for the first paragraph of each news article, we calculate the news tone as the dif-

[9] Size groups are defined by market capitalization at the end of the sample period. There are an equal number of firms in the three different size groups.

[10] Note that these 20 firms do not necessarily correspond to 20 firms with the most number of news articles during the sample period as we selected firms with the most news coverage from various industries to avoid concentration of firms in a particular industry.

[11] Loughran and McDonald (2011) propose the word lists that are appropriate for business communication. Many of the recent papers on the

ference between the percentage of positive words and the percentage of negative words (from Loughran and McDonald, 2011); and the news uncertainty measure as the percentage of uncertain words, using the LM uncertain word list.

Since most of our analyses are at daily level, we consolidate the article-level measures to generate corresponding daily measures. We use the number of words in each article as the weight to calculate the word-weighted daily news tone and daily percentage of uncertain words. The summary statistics of these daily textual measures are presented in Panel B of Table 1. For the whole sample, the average daily news tone is −0.0315% and the average percentage of uncertain words is 0.2185%. Across the three groups, large firms show more negative news tone and a higher percentage of uncertain words. This becomes more clear when examining the top 20 firms (the Online Appendix Table A.2), whose news tone ranges from around −1% to −0.1% and the percentage of uncertain words ranges from around 0.2–0.8%.

Realized jumps in stock returns are identified using a non-parametric approach (Lee and Mykland, 2008).[12] Panel C of Table 1 presents summary statistics associated with daily realized jumps. Each indicator variable takes a value of 1 if there is a jump identified using the specific criteria for stock i on day t, and is 0 otherwise. To identify jumps, we use two distributions from Lee and Mykland (2008), and use J to denote using the distribution of maximums and use J_0 to denote using the normal distribution to determine jump thresholds. The number next to J or J_0 refers to the percentile of the corresponding jump identification criteria. For example, $J99_{i,t}$ corresponds to the 99th percentile of the distribution of maximums as in Lemma 1 in Lee and Mykland (2008). Effectively, each of the four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a daily return as a jump if the absolute value of the daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility, respectively. As a result, the thresholds to identify jumps according to these criteria are also time-varying.

The total number of days where $J99$ equals 1 for all firms is 351,374 (out of around 20 million days with non-missing returns), indicating that there is on average one jump every 57(=20/0.35) days. For the other measures ($J95$, $J_0 99$, $J_0 95$), we see more jumps on average as the criteria used to identify jumps become less stringent. The frequency of daily jumps varies significantly inversely across the size groups. On average, there is one jump every 80, 53, and 39 days for the large, medium, and small-size firm groups when $J99$ is used. The gap in jump frequency becomes smaller if we use less stringent criteria. For example, in the case of $J_0 95$, the jump frequency becomes one jump every 11, 9.5, and 9 days for the large, medium, and small size firm groups. The Online Appendix Table A.3 reports daily jump measures for the top 20 firms.

3. Realized jumps and news flows

3.1. Realized jump intensity and news

We link the realized jumps to news flows. We start by using logistic regressions to examine how the probability of jumps is related to the news flows measured by the news count, the (absolute value of) news tone, and the percentage of uncertain words.

$$\mathrm{logit}(p_{it}) = a + b_1 \times NewsCount_{it} + b_2 \times |NewsTone_{it}| + b_3 \times UncWords_{it} + b_4 \times |Ret_{it-1}| + \epsilon_{it}, \quad (1)$$

where the dependent variable is the daily jump indicator variable. We also include the (absolute value of) lagged stock returns in the regressions. All of the explanatory variables are standardized to have the same mean and standard deviation across firms for these logistic regressions. By doing so, we primarily rely on the time series variation in the explanatory variables to explain the probability of jumps. We expect the coefficient b_1 for *NewsCount* to be positive and statistically significant. In addition, we expect jumps to be positively related to the absolute value of the news tone.[13] Although a very high percentage of uncertain words could induce negative jumps, a low percentage of uncertain words might not be related to positive jumps. Therefore, the relation between the percentage of uncertain words and the jump intensity would likely not be as strong as that for the news count and absolute value of news tone.

Note that when constructing the measure of news count we do not drop the news articles that might be related to other news articles on the same day. By keeping all the news articles in the news count measure, we aim to capture the importance of the news flows underlying the news articles. Undoubtedly this introduces some noise in the news count measure. In Section 5, as a robustness check, we reconstruct the news count measure using novel news only. The impact of related news articles on news tone and uncertainty is smaller as these are percentage measures.

Table 2 reports the results. Panel A reports the coefficient estimates for the whole sample and Panel B reports the standardized odds ratio associated with the corresponding variable. The probability of a jump is statistically significantly related to the news count. The standardized odds ratio associated with the news count is 1.22 (for $J99$), suggesting that one standard deviation increase in news count increases the odds of a jump by 22%. Realized jumps are also positively and significantly related to the absolute value of news tone with a standardized odds ratio of 1.02 (for $J99$). The relationship between realized jumps and the percentage of uncertain words is also positive but the economic significance is weaker, with a stan-

textual content of news articles have adopted LM word lists when measuring sentiment or tone in the news articles. We follow this approach to avoid the subjectivity of creating our own word lists. See Loughran and McDonald (2016) for a survey on related studies.

[12] Alternatively, Bollerslev and Todorov (2011) estimate the market jump tail under the physical measure using high-frequency intra-day data and estimate the risk-neutral counterpart from index options.

[13] This is because a positive news tone is likely to be related to positive jumps, whereas a negative news tone is likely to be related to negative jumps.

Table 2
Effects of news measures on probability of daily jumps.
This table reports results from pooled logistic regressions of daily jump indicators, defined using Lee and Mykland (2008), on daily news measures for all firms in the sample. The explanatory variables, which are standardized to have the same mean and standard deviation across firms, are the total number of news articles reported on the Factiva database each day, the absolute value of news tone, the percentage of uncertain words, and the absolute value of the previous day's return. The news tone measure is constructed from the percentage of positive and negative words using the list in Loughran and McDonald (2011). The sample period is from January 1980 to July 2012. The t-statistics, computed using standard errors clustered at individual firm levels, are reported in parentheses. Panel B reports the odds ratios associated with each variable in brackets. Each of four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a jump day if the absolute value of daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility. All regression specifications include a constant term that is not reported for brevity.

	(1) $J99$	(2) $J95$	(3) $J_0 99$	(4) $J_0 95$
		Panel A: Coefficient estimates		
$NewsCount_t$	0.2003	0.1915	0.1597	0.1318
	(105.66)	(110.30)	(114.07)	(113.19)
$\|NewsTone_t\|$	0.0186	0.0156	0.0113	0.0080
	(11.07)	(10.33)	(9.91)	(8.70)
$UncWords_t$	0.0094	0.0093	0.0049	0.0037
	(5.68)	(6.23)	(4.41)	(4.17)
$\|Ret_{t-1}\|$	0.0809	0.0897	0.1016	0.1061
	(45.05)	(59.75)	(96.91)	(115.55)
N	20,079,694	20,079,694	20,079,694	20,079,694
$R^2_{McFadden}$	1.40%	1.24%	0.83%	0.61%
		Panel B: Odds ratios		
$NewsCount_t$	[1.222]	[1.211]	[1.173]	[1.141]
$\|NewsTone_t\|$	[1.019]	[1.016]	[1.011]	[1.008]
$UncWrods_t$	[1.009]	[1.009]	[1.005]	[1.004]
$\|Ret_{t-1}\|$	[1.084]	[1.094]	[1.107]	[1.112]

dardized odds ratio of 1.01. The R^2 of the regression in column 1 is 1.4%. The patterns are, in general, similar when we use other measures of realized jumps in columns 2–4.[14]

In our Online Appendix Tables A.4–A.6, we report results for the large, medium, and small firm-size groups. For all three groups, news count is the most significant variable that relates to realized jumps. However, there are a couple of differences. In the large firms, the absolute value of news tone is more important than the percentage of uncertain words in explaining jumps, whereas in medium and small firms, these two variables are of equal importance. Moreover, the news flows are much more important in explaining realized jumps for large firms. This is evident from the R^2 of 2.18% for large firms, followed by 1.36% and 0.79% for medium and small firms respectively (when $J99$ is used).[15]

To better understand the cross-sectional variations in the impact of news flows on realized jumps, we also run the logistic regressions for each of the top 20 firms using each firm's time series data. The estimation results are reported in the Online Appendix Table A.8. For brevity, we only tabulate the results for $J95$. In general, the positive relationship between news count and the probability of jumps holds for all firms and is statistically significant for 16 of the top 20 firms. Both the magnitude of the coefficient estimates and the R-squared have noticeable cross-sectional variations. For example, the R^2 is highest for Amazon (16.5%) and lowest for Bank of America (0.28%). The coefficient for news count is highest for Amazon (0.048) and lowest for Cisco (0.001).

Table 2 reports results investigating the relationship between daily news content and the probability of jumps, including all news sources and all firms. To explore whether or not mixing news sources affects our results, we reestimate that model using more homogeneous media sources. The results for the $J95$ jump indicator are reported in the Online Appendix Table A.9. Column 1 reports results for which the daily news measures were compiled using articles from the Dow Jones News Services and *The Wall*

[14] The analyses reported in Table 2 use contemporaneous news measures based on the expectation that public news will be reflected in prices quickly (Engle et al., 2021). As a robustness check, we construct five day moving average variables as our news measures. The results have the same signs as in Table 2 but are uniformly weaker. Given that the information value of news can decay quickly, using moving-average measures that incorporate stale news could weaken the estimate of the effect of news on jump probabilities. Nevertheless, even with those smoothed news measures, news frequency and the absolute value of the news tone are both statistically significant and positively related to the probability of jumps.

[15] Stock returns frequently exhibit jumps on quarterly earnings announcement dates and a significant portion of firm-specific news articles might be produced around these dates. To address the concern that our results might be mainly driven by the earnings announcements, we repeat the main analyses of Table 2 by excluding the [-3,+3] trading-day window around the quarterly earnings announcement dates, following Huang et al. (2019), who also analyze the Factiva news database. For example, in the Online Appendix Table A.7, we report the results corresponding to the $J95$ jump test statistic (using the other jump statistics does not change the conclusion qualitatively). As anticipated, removing earnings announcements slightly reduces the R^2 and the statistical significance associated with the news count variable. For example, the odds ratio associated with the news count variable decreases from 1.211 to 1.163 and the R^2 drops from 1.24% to 0.83%. Nevertheless, the statistical and economic significance remain robust after removing the earnings announcements.

Street Journal only, whereas the results in column 2 include additionally daily news content from articles in *The New York Times* and the results in column 3 include additionally *The Financial Times*. The significance of our news measures are similar across the different media sources and also with Table 2.

3.2. The impact of news flows on jumps over time

We utilize the measures of news in our sample period to understand how the association between the news flows and the stock return jumps evolves over time. We estimate the model in Eq. (1) using daily data within each year from 1980 to 2012. The top panel of Fig. 1 presents the plots for the time-series of the coefficient estimates on *NewsCount* from our Factiva samples (1980–2012). A clear trend emerges from the plots: the association between the news flow and the return jumps increased over time. For example, the coefficient on *NewsCount* ranges from 0.1 to 0.2 in the 1980s, increases to a 0.2–0.3 range in the 1990s, and further increases to be over 0.5 by 2012.

In order to control for Factiva's inclusion of additional media sources over time, we redo the analyses by estimating the time trend separately for the fixed media sources.[16] The results are reported in the bottom panel of Fig. 1 which reveals a similar pattern over time for the fixed media sources.

Fig. 2 reports the same analyses using our RavenPack sample from 2000 to 2020. The association between news frequency and return jumps continued to increase over this extended sample. Note that this relationship decreases during the 2008–2009 financial crisis and the start of the COVID-19 crisis in 2020. During crisis periods, many factors (e.g., liquidity) other than firm-specific news come into play, which can decrease the relationship between firm-specific news and jumps in stock returns.

The association between news flows and return jumps can be interpreted as one measure of (tail) price informativeness provided by news. Our findings provide new evidence that price informativeness increased over time. This identification of the improvement in (tail) price informativeness for individual stock returns, utilizing daily news flow information, complements the increased clarity over time associated with market-level jumps reported by Baker et al., 2021.

There are many potential reasons for this increase in tail price informativeness over time. It is consistent with a broad trend of improved data provision and transparency, especially the improved information dissemination technology (e.g., Gao and Huang, 2020; Goldstein et al., 2020).[17] Our findings suggest that the major price informativeness improvement occurs after 2000 with the wide adoption of the Internet.

[16] We thank the referee for suggesting this robustness check.

[17] Both Gao and Huang (2020) and Goldstein et al. (2020) use the staggered implementation of the EDGAR system from 1993 to 1996 as an identification of information dissemination technologies.

3.3. Realized jump-size distribution and news

Next, we analyze how the jump-size distribution is affected by news flows. We focus on the jump-size mean and jump-size volatility and expect the news content (e.g., news tone) to affect both moments significantly.

We focus on the observations of realized jumps to understand the impact of news on the first two moments of jump-size distributions. For the jump-size mean, we run the following regressions:

$$\begin{aligned} r_{it}|\text{Jump} = b_0 + b_1 \times NewsCount_{it} + b_2 \times NewsTone_{it} \\ + b_3 \times UncWords_{it} + b_4 \times r_{it-1} + \epsilon_{it}, \end{aligned} \tag{2}$$

where $r_{it}|\text{Jump}$ measures the realized returns on jump days and r_{it-1} is the realized returns on the previous day. Note that each of the news measures is standardized to have the same mean and standard deviation across firms to be consistent with our logistic regressions. The above regression implies the following relation between the jump-size mean and news content measures:

$$\begin{aligned} &\mathbb{E}[r_{it}|\text{Jump=1}] \\ &= b_0 + b_1 \times \mathbb{E}[NewsCount_{it}] + b_2 \times \mathbb{E}[NewsTone_{it}] \\ &\quad + b_3 \times \mathbb{E}[UncWords_{it}] + b_4 \times \mathbb{E}[r_{it-1}]. \end{aligned} \tag{3}$$

Table 3 reports the results for all firms. Panel A reports the results for all jumps (regardless of the sign of jumps). The jump-size mean is statistically significantly related to the news content: positively related to news counts, the news tone and the percentage of uncertain words. We explore this further in Panel B and Panel C in which we split the analyses for positive versus negative jumps. As shown in Panel B, positive jump returns are significantly and positively related to the news counts: more good news is associated with higher positive returns on jump days. In Panel C, negative jump returns are significantly and negatively related to news counts: more bad news is associated with more negative returns on jump days. When we put all the jump returns in the same regressions, the effects offset, resulting in the smaller but still positive coefficient on news counts in Panel A. The R^2 of news flows explaining negative jump returns, identified by the $J99$ jump indicator, is 7.90%, whereas it is 2.74% for positive jump returns.[18]

We also conduct the same analysis for the top 20 firms and present the results in the Online Appendix Table A.13. Jump returns are positively related to the news tone but are mainly driven by negative jump returns. The percentage of uncertain words is no longer statistically significant in explaining the jump returns, at least for jumps identified by the $J99$ and $J95$ jump indicators. The result on news count is similar to the results using all firms: when news is negative, more news is related to more negative jump returns; and when news is positive, more news is related to more positive jump returns. But the association is more pronounced for negative jumps. If we do not separate out the positive jump returns from negative jump returns, we

[18] The results for large, medium, and small firms are similar, as reported in the Online Appendix Tables A.10– A.12.

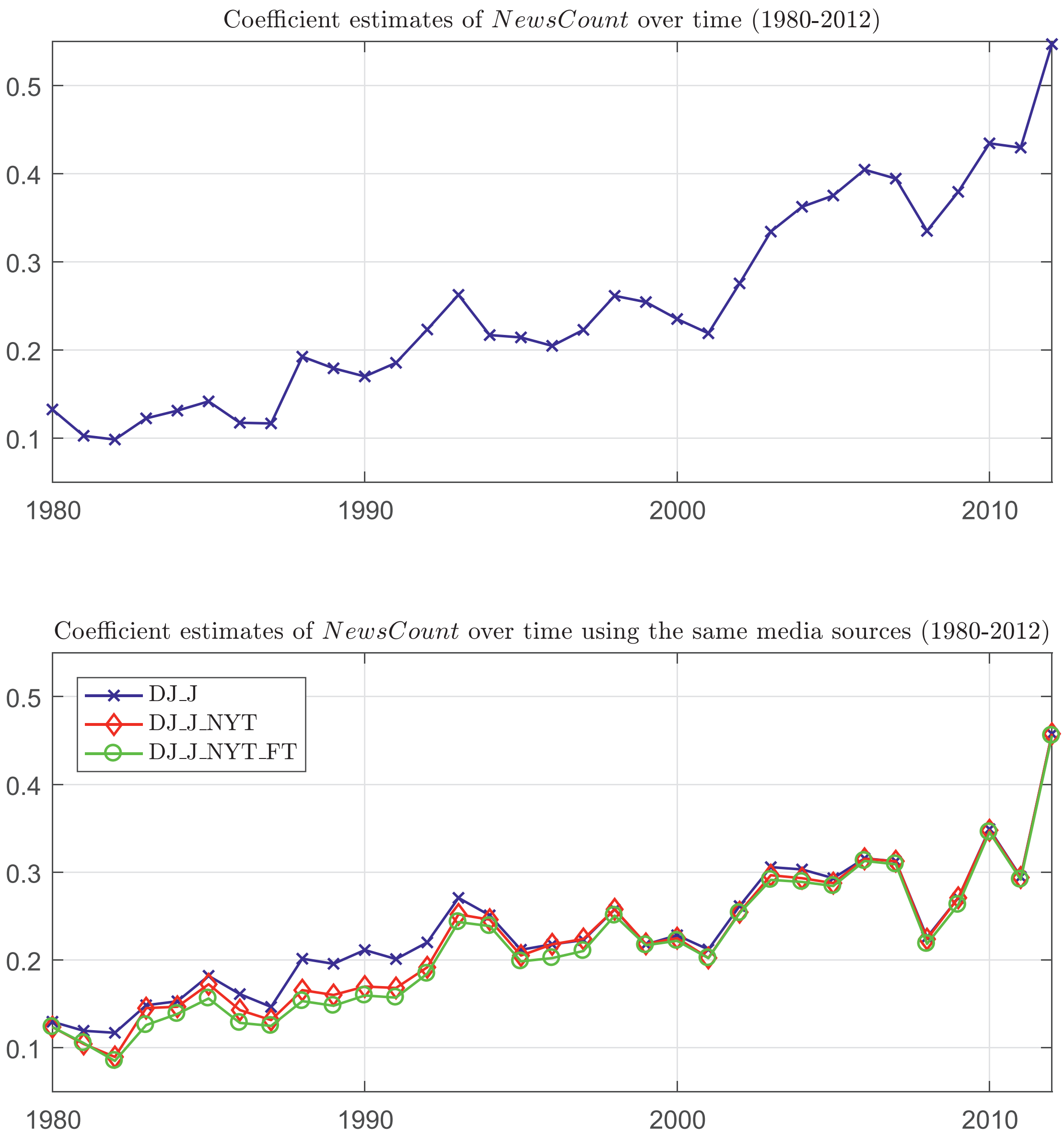

Fig. 1. Effects of news on the probability of daily jumps over time.
In this figure we plot the effect of news on the probability of daily jumps over time. We estimate the same model as in Table 2 within each year from 1980 to 2012 using Factiva data. In the top panel, we plot the time-series of coefficient estimates for the *NewsCount* variable. In the bottom panel, we plot the results obtained using news items coming from alternative media sources. DJ_J refers to the sample using the Dow Jones and Wall Street Journal news articles. DJ_J_NYT refers to the DJ_J sample augmented using the New York Times news articles, and DJ_J_NYT_FT refers to the DJ_J_NYT sample augmented using the Financial Times news articles. All models are estimated using *J*95 as the jump indicator.

see an unconditionally negative association between news counts and jump returns. That is, as news counts increase, the jump returns, on average, become more negative. The R^2 of news flows explaining negative jump returns is over 30%, compared to an R^2 of around 11% for positive jump returns (for *J*99).

For the jump-size volatility, we use the same set of explanatory variables as for the jump-size mean, and run the following regressions:

$$\log(r_{it}^2|\text{Jump}) = c_0 + c_1 \times NewsCount_{it} + c_2 \times |NewsTone_{it}| + c_3 \times UncWords_{it} + c_4 \times |r_{it-1}| + \epsilon_{it}, \quad (4)$$

where $r_{it}^2|\text{Jump}$ measures the realized jump return variance. The above regression implies the following relation between the jump-size second moment and news:

$$\begin{aligned}&\mathbb{E}[\log(r_{it}^2|\text{Jump=1})]\\&\quad = c_0 + c_1 \times \mathbb{E}[NewsCount_{it}] + c_2 \times \mathbb{E}[|NewsTone_{it}|]\\&\qquad + c_3 \times \mathbb{E}[UncWords_{it}] + c_4 \times \mathbb{E}[|r_{it-1}|]. \quad (5)\end{aligned}$$

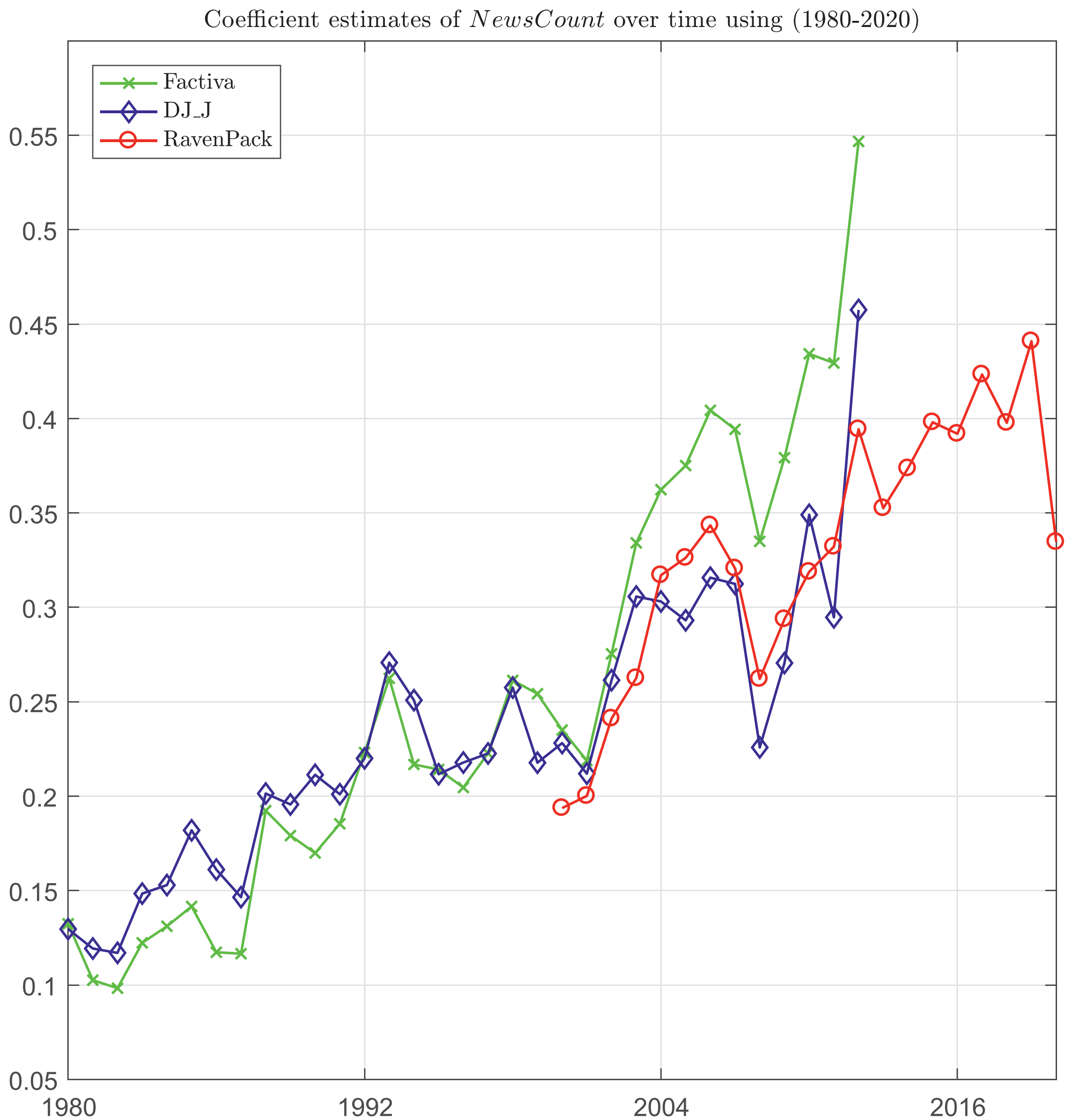

Fig. 2. Effects of news on the probability of daily jumps over time (1980–2020).
This figure plots the effect of news on the probability of daily jumps over time. We estimate the model similar to Table 2 within each year from 1980 to 2020, then plot the time-series of coefficient estimates for the *NewsCount* variable. We provide three different time-series corresponding to all Factiva sample (1980–2012), labelled as Factiva; Dow Jones and Wall Street Journal news articles from Factiva (1980–2012), labelled as DJ_J; and RavenPack DJ Edition sample (2000–2020), labelled as RavenPack. All models are estimated using *J*95 as the jump indicator.

The results are reported in Table 4. In general, the jump-size volatility is positively and significantly related to both the news count and the absolute value of the news tone. The R^2s of the jump-size volatility regressions are also much higher than those for the jump-size mean regressions. For example, in the case of *J*99, the R^2 is 14.1% for all firms and 25% for the top 20 firms (in untabulated results). The results are similar for firms in the three size groups, as shown in the Online Appendix Table A.14. Overall, our results indicate that news flows are very important in explaining variations in the second moment of jump-size distributions.

4. Cross-sectional determinants and news categories

4.1. Determinants of the sensitivity of jump probability to news

The results so far suggest that jumps in individual stock returns are significantly related to the news process, particularly to news frequency. We now explore further the potential determinants of this sensitivity of jump probability to news counts in the cross-section of firms. Specifically, we first run logistic regressions, similar to Eq. (1), at the firm-by-firm level to obtain an estimate of the sensi-

Table 3
Effects of news measures on daily jump size.
This table reports results from regressions of daily jump sizes, conditional on the jump indicator being 1, on daily news measures for all firms in the sample. The explanatory variables are the total number of news articles reported on the Factiva database each day, news tone, the percentage of uncertain words, and the previous day's return. Each of the news variables is standardized to have the same mean and standard deviation across firms. The news tone measure is constructed from the percentage of positive and negative words using the list in Loughran and McDonald (2011). The sample period is from January 1980 to July 2012. The t-statistics, computed using standard errors clustered at individual firm levels, are reported in parentheses. Panels B and C report the results for positive jump sizes and negative jump sizes, respectively. Each of four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a jump day if the absolute value of daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility. All regression specifications include a constant term that is not reported for brevity.

	(1) $J99$	(2) $J95$	(3) $J_0 99$	(4) $J_0 95$
	Panel A: All jumps			
$NewsCount_t$	2.96E-03	2.64E-03	2.39E-03	2.26E-03
	(19.42)	(20.53)	(29.22)	(39.54)
$NewsTone_t$	0.0126	0.0110	0.0071	0.0050
	(40.03)	(43.38)	(50.52)	(55.75)
$UncWords_t$	0.0045	0.0037	0.0020	0.0012
	(13.50)	(14.06)	(13.65)	(13.38)
r_{t-1}	-0.2094	-0.2206	-0.2588	-0.2850
	(-48.10)	(-62.99)	(-130.33)	(-221.23)
N	351,374	480,489	1,096,893	2,055,081
R^2	1.24%	1.31%	1.84%	2.54%
	Panel B: Positive jumps			
$NewsCount_t$	1.04E-02	9.92E-03	8.69E-03	7.72E-03
	(56.70)	(64.59)	(90.11)	(114.45)
$NewsTone_t$	0.0014	0.0010	0.0006	0.0003
	(3.56)	(3.34)	(3.44)	(2.32)
$UncWords_t$	0.0031	0.0024	0.0015	0.0009
	(7.62)	(7.65)	(8.72)	(8.25)
r_{t-1}	-0.2575	-0.2515	-0.2345	-0.2222
	(-49.65)	(-60.57)	(-99.58)	(-142.82)
N	206,070	282,470	629,554	1,140,445
R^2	2.74%	2.75%	2.81%	2.86%
	Panel C: Negative jumps			
$NewsCount_t$	-1.07E-02	-9.93E-03	-8.17E-03	-6.75E-03
	(-96.96)	(-104.91)	(-130.51)	(-150.35)
$NewsTone_t$	0.0046	0.0043	0.0032	0.0025
	(21.04)	(23.83)	(30.56)	(37.37)
$UncWords_t$	2.75E-05	1.44E-04	-1.40E-04	-2.94E-05
	(0.12)	(0.77)	(-1.29)	(-0.42)
r_{t-1}	-0.0445	-0.0617	-0.0985	-0.1268
	(-13.96)	(-23.87)	(-65.66)	(-129.07)
N	145,304	198,019	467,339	914,636
R^2	7.90%	6.96%	5.24%	4.64%

tivity, $b_{i,1}$, of jump probability to news count for each firm i. Then, we run a cross-sectional regression to understand which firm characteristics are linked to this coefficient.

We consider firm characteristics similar to those in Fang and Peress (2009), computed as an annual average of end-of-year observations during the same sample period. The size variable is measured by market capitalization reported by CRSP, book value from Compustat, analyst coverage and dispersion data are collected from I/B/E/S summary files, and the fraction of institutional ownership is from aggregate Thompson Reuters 13F filings. We use the

Table 4
Effects of news measures on daily jump volatility.
This table reports results from regressions of daily jump volatilities, defined as the log of the squared jump size conditional on the jump indicator being 1, on daily news measures for all firms in the sample. The explanatory variables are the total number of news reported on the Factiva database each day, the absolute value of news tone, the percentage of uncertain words, and the absolute value of the previous day's return. Each of the news variables is standardized to have the same mean and standard deviation across firms. The news tone measure is constructed from the percentage of positive and negative words using the list in Loughran and McDonald (2011). The sample period is from January 1980 to July 2012. The t-statistics, computed using standard errors clustered at individual firm levels, are reported in parentheses. Each of four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a jump day if the absolute value of daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility. All regression specifications include a constant term that is not reported for brevity.

	(1) $J99$	(2) $J95$	(3) $J_0 99$	(4) $J_0 95$
$NewsCount_t$	0.1318	0.1261	0.1164	0.1072
	(102.74)	(110.74)	(134.37)	(151.34)
$\lvert NewsTone_t \rvert$	0.0398	0.0343	0.0236	0.0175
	(14.17)	(14.57)	(15.12)	(15.17)
$UncWords_t$	0.0044	0.0045	0.0070	0.0062
	(1.56)	(1.95)	(4.54)	(5.47)
$\lvert r_{t-1} \rvert$	8.4333	8.7466	9.8895	10.9520
	(215.75)	(262.77)	(430.42)	(619.27)
N	351,374	480,489	1,096,893	2,055,081
R^2	14.71%	15.10%	16.10%	16.86%

natural logarithm of one plus the number of analysts covering the company for the analyst coverage variable, and one minus the fraction of institutional ownership to represent individual ownership. We also construct a media visibility measure that captures the number of unique news sources covering an individual firm (News Sources).

We first run univariate regressions with each of the firm characteristic variables and report the results in the Online Appendix Table A.15. We find that, in general, the jump sensitivity to news is higher for larger firms or firms in a more transparent and visible information environment, such as more analyst coverage, more media visibility, and higher institutional ownership, highlighting the importance of those channels with respect to quick incorporation of news into returns. However, these firm characteristic variables are likely to be correlated. Therefore, we evaluate the marginal impact of these variables in multivariate regressions. We include 11 sector indicator variables in the regressions to understand the differential effect across the sectors.

Table 5 reports multivariate cross-sectional regression results in which the dependent variable $b_{i,1}$ is the coefficient from firm-by-firm logistic regressions estimating the sensitivity of jump probability to news counts for each firm, as explained above. The results indicate that, from our set of firm characteristics, media visibility proxied by the number of news sources covering a firm is the most significant determinant for the sensitivity of jump probability to news. The positive relationship suggests that the media visibility could help disseminate the information broadly and the information is more likely to lead to stock return jumps.

Table 5
Cross-sectional determinants of jump probability sensitivity to news.
This table reports cross-sectional regression results concerning potential firm characteristic determinants of the sensitivity of jump probabilities to news counts. The dependent variable in the cross-sectional regressions is $b_{i,1}$ estimated from firm-by-firm logistic regressions of Eq. 1, that is, the sensitivity of jump probability to news counts at the individual firm level. The seven firm characteristic regressors in the cross-sectional regressions are computed as annual averages of end-of-year observations during the sample period from January 1980 to July 2012. Eleven sector indicator variables are included in the regressions. The *t*-statistics are reported in parentheses.

	(1) $J99$	(2) $J95$	(3) $J_0 99$	(4) $J_0 95$
Intercept	0.0902	0.0192	-0.0616	-0.0805
	(0.30)	(0.07)	(-0.32)	(-0.44)
Size	-0.1214	-0.0961	-0.0176	0.0162
	(-4.22)	(-3.51)	(-0.94)	(0.92)
Book-to-Market Ratio	-0.1715	-0.1155	0.0565	-0.0273
	(-1.24)	(-0.87)	(0.62)	(-0.32)
Analyst Coverage	0.1601	0.0994	0.0303	0.0219
	(2.64)	(1.72)	(0.76)	(0.59)
Analyst Dispersion	0.0175	0.0203	0.0119	0.0065
	(0.55)	(0.67)	(0.57)	(0.33)
Individual Ownership	-0.0951	-0.0999	-0.0465	-0.0081
	(-1.86)	(-2.05)	(-1.39)	(-0.26)
Return Volatility	-0.1362	-0.0986	-0.0224	0.0125
	(-2.04)	(-1.55)	(-0.51)	(0.30)
News Sources	0.1901	0.1694	0.0650	0.0099
	(7.74)	(7.24)	(4.06)	(0.66)
Consumer Nondurables	0.0523	-0.0068	-0.1051	-0.0562
	(0.35)	(-0.05)	(-1.07)	(-0.61)
Consumer Durables	0.1405	0.0926	-0.0286	-0.0676
	(0.74)	(0.51)	(-0.23)	(-0.58)
Manufacturing	0.1352	0.0238	-0.0423	-0.0746
	(1.16)	(0.21)	(-0.55)	(-1.04)
Energy	0.1663	0.0846	-0.0485	-0.1007
	(1.06)	(0.57)	(-0.47)	(-1.05)
Chemicals	-0.0828	-0.1228	-0.0703	-0.0706
	(-0.37)	(-0.58)	(-0.49)	(-0.52)
Business Equipment	0.0028	-0.0390	-0.1034	-0.0701
	(0.03)	(-0.45)	(-1.74)	(-1.25)
Telecommunications	0.0847	0.0507	-0.1337	-0.1639
	(0.55)	(0.35)	(-1.33)	(-1.74)
Utilities	-0.0561	-0.0614	-0.1489	-0.1711
	(-0.28)	(-0.32)	(-1.13)	(-1.38)
Wholesale & Retails	-0.0233	-0.0487	-0.0167	-0.0536
	(-0.21)	(-0.47)	(-0.23)	(-0.80)
Healthcare & Medical	0.1726	0.1067	-0.0141	-0.0457
	(1.59)	(1.03)	(-0.20)	(-0.69)
Financial	-0.0204	-0.0530	-0.1519	-0.1844
	(-0.22)	(-0.60)	(-2.51)	(-3.24)
N	5026	5026	5026	5026
R^2	2.38%	1.90%	1.09%	0.56%

Analyst coverage is another significant variable in determining the sensitivity of jump probability to news. This relationship is positive, suggesting that firms covered by more analysts are more likely to jump when public news arrives. Fang and Peress (2009) show that analyst and media coverage are substitutes rather than complements. With this in mind, our findings suggest that when public news arrives for firms with higher analyst coverage, the information contained in the news may be more informative, possibly not previously covered by the analysts, thus creating a higher chance of inducing jumps in stock returns.

The fraction of individual ownership shows a significantly negative relationship with respect to the sensitivity of the jump probability to news. In other words, firms with higher institutional ownership tend to jump more frequently when public news arrives. This finding suggests that it is more likely that institutional traders are the ones who trade when public news arrives (e.g. Huang et al., 2019), at least on the same day. Individual investors may suffer from limited attention and may not be able to react on the same day for all public news that arrives.

Overall, the results in Table 5 suggest that the existence of more information intermediaries, such as media, analysts, or a higher fraction of institutional investors, helps to incorporate the firm-specific public news into stock return jumps.

4.2. Additional analysis using different news categories

So far, we have not differentiated between the type of news. In reality, specific type of news, such as earnings related news, may be more important in explaining stock return jumps than other types of news. To investigate the role of different types of news articles, we employ the RavenPack database. We rely on the news groups that RavenPack provides and regroup them into ten major news categories. We then allocate all news into those ten categories.

The results are presented in Table 6. Due to the unavailability of the RavenPack data for the first 20 years of our sample, the sample is from January 2000 to the end of 2020. The most important news category for both news counts and the absolute value of news tone (according to both *t*-stats and odd ratios) is Earnings and Revenues, followed by: Analyst Ratings; Capital Structure; M&A; Marketing and Investor Relations; Labor Issues and Executive Turnovers; and Products and Services information.

In addition to the news groups information, RavenPack also provides further classification of news items within each group, which we call news sub-categories. While the group information is about the type of news article, the sub-category item corresponds to more detailed information on the actual content. For example, news articles classified as having a group of Analyst Ratings can be further classified into specific sub-categories depending on the content, such as Change to Negative, Change to Positive, and so on. To investigate how specific sub-categories of news articles matter for jumps in stock returns, we further divide each of the ten categories we used in Table 6 into five different sub-categories. Within each category, we pick the four most important types of new sub-categories based on the frequency and place the rest into others.[19]

Table 7 reports the results. For brevity, we only report the coefficients and *t*-stats associated with the total number of news articles for each sub-category. Mostly, the sub-categories are statistically significant when the parent category is significant. However, there are a few interesting exceptions where some of the sub-categories no longer

[19] The precise mapping from the RavenPack group and category variables to our grouping is reported in the Online Appendix.

Table 6
Effects of different news categories on the probability of daily jumps.
This table reports results from pooled logistic regressions of the *J*95 daily jump indicator, defined using Lee and Mykland (2008), on daily news counts and absolute news tone by category using RavenPack data. The explanatory variables, which are standardized to have the same mean and standard deviation across firms, are the total number of news articles and the absolute value of news tone reported on the RavenPack database each day. The sample period is from January 2000 to December 2020. The *t*-statistics, computed using standard errors clustered at individual firm levels, are reported in parentheses. The odds ratios are reported in brackets. The regression specification include a constant term and lagged absolute return that are not reported for brevity.

	$NewsCount_t$	$\|NewsTone_t\|$
	All news	
M&A	0.055	0.022
	(55.31)	(21.67)
	[1.056]	[1.022]
Analyst Ratings	0.094	0.038
	(62.56)	(23.46)
	[1.098]	[1.039]
Assets	-0.007	0.021
	(-3.87)	(11.73)
	[0.993]	[1.021]
Capital Structure	0.075	0.024
	(75.29)	(24.31)
	[1.078]	[1.024]
Credit Ratings	0.018	0.009
	(7.74)	(3.58)
	[1.018]	[1.009]
Earnings and Revenues	0.174	0.097
	(110.88)	(75.66)
	[1.190]	[1.102]
Marketing and Investor Relations	0.073	-0.017
	(47.76)	(-6.20)
	[1.076]	[0.984]
Labor Issues including Executive Turnovers	0.030	0.002
	(19.10)	(1.09)
	[1.030]	[1.002]
Products and Services	0.033	0.023
	(15.88)	(10.32)
	[1.034]	[1.023]
Insider Trading	-0.015	0.009
	(-5.81)	(3.68)
	[0.985]	[1.009]
$R^2_{McFadden}$	4.67%	

show statistical significance even though their parent category is significant. For instance, while Capital Structure is the third most important group in Table 6, Buybacks are not statistically significant and Note Sale carries marginal significance at the sub-category level. As a further example, Marketing and Investor Relations is a significant category in Table 6, but only Shareholder Disclosure and Conference Call related news articles are significant as sub-categories in that group. Overall, these results highlight the value of disaggregating and investigating the impact by news sub-categories.

5. Robustness and additional analyses

In this section, we summarize some robustness analyses of our results. We also present results from additional non-parametric analyses related to our main results concerning the relationship between measures of news flow and stock return jumps.

5.1. Forecasting future volatility using news measures

Our analyses have focused on how daily news measures are related to individual stock return jumps identified probabilistically at various levels of confidence. Motivated by the results in Baker et al., 2021 for market-level jumps, we now check whether our frequency and content news measures can forecast future volatility. We use a test equation analogous to that in section 3.5 of Baker et al., 2021. In our case:

$$\begin{aligned}\sum_{n=1}^{5}\frac{r^2_{i,t+n}}{5} = a &+ b_1 NewsCount_{i,t} \\ &+ b_2|NewsTone_{i,t}| + b_3 UncWords_{i,t} \\ &+ c_1(r_{i,t}\times I(r_{i,t}>0)) \\ &+ c_2(|r_{i,t}|\times I(r_{i,t}\le 0)) + c_3 r^2_{i,t-1} \\ &+ c_4\sum_{n=1}^{5}\frac{r^2_{i,t-n}}{5} + c_5\sum_{n=1}^{22}\frac{r^2_{i,t-n}}{22} + \epsilon_t \qquad (6)\end{aligned}$$

Table 7
Effects of news sub-categories on the probability of daily jumps.
This table reports results from pooled logistic regressions of the *J*95 daily jump indicator, defined using Lee and Mykland (2008), on daily news count and absolute news tone by sub-categories using RavenPack data. The explanatory variables, which are standardized to have the same mean and standard deviation across firms, are the total number of news articles and the absolute value of news tone reported on the RavenPack database each day. We report the result for the total number of news only for brevity. The sample period is from January 2000 to December 2020. The *t*-statistics computed using standard errors clustered at individual firm levels are reported in parentheses. The regression specification include a constant term and lagged absolute return that are not reported for brevity.

	Acquirer	Acquiree	Merger	Interest	Others
M&A	0.023*** (10.43)	0.050*** (20.31)	0.033*** (7.94)	0.026*** (8.06)	0.007*** (3.61)
	Change to Negative	Change to Positive	Set Negative	Set Positive	Others
Analyst Ratings	0.065*** (35.54)	0.023*** (7.45)	0.055*** (2.69)	0.017*** (2.74)	0.016*** (7.99)
	Sale	Facility Open	Facility Close	Patent	Others
Assets	0.012** (2.22)	-0.001 (-0.18)	-0.004 (-0.54)	0.003 (0.18)	-0.004* (-1.89)
	Note Sale	Dividend	Buybacks	Public Offering	Others
Capital Structure	0.010** (2.04)	0.016*** (10.57)	0.004 (1.42)	0.025*** (5.79)	0.064*** (72.24)
	Set	Affirmation	Downgrade	Upgrade	Others
Credit Ratings	0.002 (0.54)	0.009** (2.20)	0.020*** (4.76)	-0.010 (-1.49)	0.019*** (8.38)
	Earnings Up	Earnings Down	Revenue Up	Revenue Down	Others
Earnings and Revenues	0.050*** (35.75)	0.028*** (19.88)	0.023*** (16.88)	0.012*** (7.21)	0.103*** (75.77)
	Shareholder Disclosure	Conference Call	Board Meeting	Conferences	Others
Marketing and Investor Relations	-0.018*** (-8.92)	0.090*** (65.66)	-0.002 (-0.14)	0.002 (1.11)	-0.002 (-0.15)
	Executive Appointment	Executive Resignation	Layoffs	Executive Salary	Others
Labor Issues including Executive Turnovers	0.005** (2.08)	0.017*** (6.03)	0.013*** (3.84)	-0.001 (-0.21)	0.003 (0.83)
	Business Contract	Product Release	Regulatory Approval	Award	Others
Products and Services	0.020*** (7.51)	0.017*** (3.41)	0.021*** (2.90)	0.036** (2.21)	0.018*** (7.29)
	Sell	Sell Registration	Buy	Surrender	Others
Insider Trading	-0.004 (-0.86)	-0.026*** (-11.01)	0.013*** (4.58)	0.009** (2.42)	-0.004 (-0.33)
$R^2_{McFadden}$			4.91%		

The explanatory variables include the current standardized news measures, as before, as well as indicators for positive and negative absolute values of returns plus lagged daily, weekly, and monthly proxies for realized variance. The dependent variable is average squared daily returns over the following five days.

The results are in Table 8. Higher news counts reduce future realized volatility, consistent with an interpretation that more news reduces uncertainty. This result is similar to that in Baker et al., 2021, who identify different clarity with respect to different sources of news.

5.2. Robustness: using single firm articles only

The Online Appendix Table A.19 reports the results of logistic regressions of daily jump indicators on daily news measures using the news articles that are unique to each firm. That is, only news articles that mention a single firm are included. For robustness, we compare these results with the results in Table 2, which include all news for each firm. The results are very similar. Small differences suggest that news tone is slightly less significant but that the percentage of uncertain words is more significant for the case of single firm news. The percentage of uncertain words may be a more precise measure when restricted to single firm news articles.

5.3. Robustness: alternative identification methods for jumps

As described in Section 2, we identify jump days using a nonparametric approach that accommodates time-varying spot volatility. The jump identification criteria cor-

Table 8
Effects of news flows on future realized volatility.
This table reports results from regressions of future realized volatility, defined as the following week's realized variance conditional on the jump indicator being 1, on daily news measures for all firms. The explanatory news variables include the total number of news reported on the Factiva database each day, the absolute value of news tone, and the percentage of uncertain words. Each of these news variables is standardized to have the same mean and standard deviation across firms. The news tone measure is constructed from the percentage of positive and negative words using the list in Loughran and McDonald (2011). Other regressors discussed in the text (Eq. 6) but not reported in this table for brevity, are indicators for positive and negative absolute value of returns plus lagged daily, weekly, and monthly realized variance. The sample period is from January 1980 to July 2012. The t-statistics, computed using standard errors clustered at individual firm levels, are reported in parentheses. Each of four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a jump day if the absolute value of daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility. All regression specifications include a constant term that is not reported for brevity.

	(1) $J99$	(2) $J95$	(3) $J_0 99$	(4) $J_0 95$
	Panel A: All jumps			
$NewsCount_t$	-6.65E-04	-6.10E-04	-4.96E-04	-3.96E-04
	(-8.01)	(-9.25)	(-12.83)	(-15.77)
$\|NewsTone_t\|$	-2.24E-04	-1.77E-04	-6.78E-05	-4.01E-05
	(-1.24)	(-1.31)	(-0.98)	(-0.99)
$UncWords_t$	2.66E-05	1.92E-05	-2.17E-05	-7.52E-06
	(0.15)	(0.14)	(-0.32)	(-0.19)
N	348,549	476,985	1,090,421	2,044,388
R^2	1.91%	2.05%	2.56%	3.20%
	Panel B: Positive jumps			
$NewsCount_t$	-4.17E-04	-3.98E-04	-2.74E-04	-3.02E-04
	(-5.07)	(-6.03)	(-9.02)	(-10.74)
$\|NewsTone_t\|$	-4.63E-05	-4.61E-05	1.88E-05	2.35E-05
	(-0.25)	(-0.33)	(0.25)	(0.50)
$UncWords_t$	1.66E-04	1.29E-04	6.55E-05	3.63E-05
	(0.90)	(0.93)	(0.87)	(0.79)
N	204,496	280,490	625,893	1,134,549
R^2	3.40%	3.62%	3.68%	4.36%
	Panel C: Negative jumps			
$NewsCount_t$	-1.27E-03	-1.13E-03	-7.29E-04	-5.58E-04
	(-7.25)	(-8.28)	(-9.86)	(-12.31)
$\|NewsTone_t\|$	-2.42E-04	-2.00E-04	-1.08E-04	-8.65E-05
	(-0.68)	(-0.76)	(-0.85)	(-1.23)
$UncWords_t$	-1.62E-04	-1.28E-04	-1.25E-04	-5.06E-05
	(-0.46)	(-0.49)	(-1.00)	(-0.73)
N	144,053	196,495	464,528	909,839
R^2	1.17%	1.23%	2.02%	2.59%

Table 9
Using alternative definitions of jumps.
This table reports results from pooled logistic regressions of the daily jump indicator, defined using alternative absolute thresholds of 8.92% and 10%, on daily news measures for all firms in the sample. The explanatory variables, which are standardized to have the same mean and standard deviation across firms, are the total number of news reported on the Factiva database each day, the absolute value of news tone, the percentage of uncertain words, and the absolute value of the previous day's return. The news tone measure is constructed from the percentage of positive and negative words using the list in Loughran and McDonald (2011). The sample period is from January 1980 to July 2012. The t-statistics, computed using standard errors clustered at individual firm levels, are reported in parentheses. Panel B reports the odds ratios associated with each variable in brackets. All regression specifications include a constant term that is not reported for brevity.

	(1) $J(8.92\%)$	(2) $J(10\%)$
	Panel A: Coefficient estimates	
$NewsCount_t$	0.1320	0.1366
	(86.72)	(84.91)
$\|NewsTone_t\|$	0.0152	0.0172
	(12.82)	(13.34)
$UncWords_t$	0.0101	0.0097
	(9.06)	(7.97)
$\|Ret_{t-1}\|$	0.4118	0.4108
	(220.33)	(217.20)
N	20,079,694	20,079,694
$R^2_{McFadden}$	4.40%	4.58%
	Panel B: Odds ratios	
$NewsCount_t$	[1.141]	[1.146]
$\|NewsTone_t\|$	[1.015]	[1.017]
$UncWords_t$	[1.010]	[1.010]
$\|Ret_{t-1}\|$	[1.510]	[1.508]

respond to different percentiles of the maximum return distribution from Lee and Mykland (2008). For example, the $J95$ indicator (corresponding to the 95th percentile) is 4.4881, which would identify a jump day if the absolute value of the daily return is greater than 4.4881 times that day's spot volatility. Total daily jump counts, associated with several alternative jump indicator statistics, are reported in the bottom panel of Table 1 for all firms and for our three firm-size groups. For example, the number of jump days for all firm days using the $J95$ indicator is 480,489.

One could, instead, use an unconditional threshold to identify jump days, for example, by assuming the spot volatility is constant over time. The median spot volatility for all firm-day observations in our sample is 1.9867%. Scaling our $J95$ jump indicator (4.4881) by the median value for spot volatility, we identify a jump if the absolute daily return is above 8.92%.[20] The number of jump days identified using the constant 8.92% threshold is 1,060,281, more than twice as many as identified using the $J95$ indicator with time-varying spot volatility. This is because large increases in spot volatility may be identified as jump days using the former method but not the latter.

Table 9 reports results for effects of our daily news measures on daily jump probabilities for all firms, using the $J(8.92\%)$ threshold and an alternative $J(10\%)$ threshold for identifying jump days. All three news measures continue to be significant. Comparing these results with the $J95$ column of Table 2, we see that the coefficients and statistical significance of the three news measures are similar, although the news frequency effect is stronger using the jump identification method that controls for time-varying spot volatility. For a related reason, the R^2 reported in Table 9 is higher, reflecting the significance of the abso-

[20] Baker et al., 2021 use 2.5% as a threshold for index returns which, due to diversification, would have a more concentrated distribution than our individual stock returns.

Table 10
Effects of news measures on probability of daily jumps: novel news.
This table reports results from logistic regressions of the daily jump indicator, defined using Lee and Mykland (2008), on daily news count and absolute news tone for all firms using novel news from the RavenPack database. The explanatory variables, which are standardized to have the same mean and standard deviation across firms, are the total number of news articles and the absolute value of news tone reported on the RavenPack database each day, and the absolute value of the previous day's return. The sample period is from January 2000 to December 2020. The t-statistics computed using standard errors clustered at individual firm levels are reported in parentheses. Panel B reports the odds ratios associated with each variable in brackets. Each of four statistics $\{J99, J95, J_0 99, J_0 95\}$ identifies a jump day if the absolute value of daily return is above $\{5.1024, 4.4881, 3.2283, 2.4565\}$ times the time-varying daily spot volatility. All regression specifications include a constant term that is not reported for brevity.

	(1) $J99$	(2) $J95$	(3) $J_0 99$	(4) $J_0 95$
	Panel A: Coefficient estimates			
$NewsCount_t$	0.3045	0.2862	0.2331	0.1907
	(140.44)	(140.67)	(137.28)	(132.79)
$\lvert NewsTone_t \rvert$	0.0776	0.0756	0.0639	0.0513
	(26.80)	(31.27)	(41.22)	(46.19)
$\lvert Ret_{t-1} \rvert$	0.1044	0.1057	0.1019	0.0941
	(70.30)	(82.15)	(102.94)	(107.11)
N	23,808,488	23,808,488	23,808,488	23,808,488
$R^2_{McFadden}$	4.96%	4.04%	2.22%	1.32%
	Panel B: Odds ratios			
$NewsCount_t$	[1.356]	[1.331]	[1.263]	[1.210]
$\lvert NewsTone_t \rvert$	[1.081]	[1.079]	[1.066]	[1.053]
$\lvert Ret_{t-1} \rvert$	[1.110]	[1.111]	[1.107]	[1.099]

lute value of the lagged return that is capturing the spot volatility clustering which is, to some extent, excluded using the jump identification method in Table 2.

5.4. Robustness analyses using novel news

In our benchmark analyses, we include all news articles. The implicit assumption is that the total news count (including the repeated ones) would capture the importance of news. Nonetheless, it is important to investigate how novel (innovative or surprising) news is related to stock market jumps. That is, does using only novel news as the measure of information flow result in a different relationship with stock market jumps?

To implement this analysis, we rely on the RavenPack news dataset. This dataset provides a variable that measures how "novel" a news article is by comparing the content of the news article with previous news articles about the same company. The highest novelty score is 100. We keep only the news articles with a novelty score of 100 for this analysis to focus on news that is most likely to be a surprise. In addition to the number of novel news, we also measure the tone of these news articles using the proprietary sentiment measure that RavenPack provides.[21]

Using these textual measures related to novel news flow, we repeat our baseline regressions as for Table 2 and present the results in Table 10. In general, our results in Table 10 are quite consistent with those in Table 2: more news is related to higher probabilities of jumps in stock returns.[22]

In summary, we can conclude that our results with respect to the effects of the news flow on stock market jumps is robust to whether we measure the news flow using all news articles or just novel news articles. The 20 extra years of data provided by the Factiva data collected for this paper allows us to evaluate a longer sample period and uncover the time trend of how news flows affect stock return jumps.

[21] The sentiment measure from RavenPack ranges from 0 to 100; we subtract 50 so that it ranges from −50 to 50, with a negative value of the re-centred measure representing negative sentiment and a positive value representing positive sentiment. Note that RavenPack does not provide the news text, so we are not able to generate the uncertain words measure which would be missing from the analyses with RavenPack data.

[22] To address the sample difference between the RavenPack and Factiva data. we repeat the same regressions using all news articles from the RavenPack dataset (results reported in the Online Appendix Table A.20). The results with all news articles in RavenPack are very close to the results with novel news articles only. The differences in results between all news articles by RavenPack in the Online Appendix Table A.20 and all news articles from Factiva (Table 2), can be explained by differences in news coverage. RavenPack's coverage is from 2000 to 2012 (we use the Dow Jones version of RavenPack), while Factiva's coverage is from 1980 to 2012 and Factiva covers more sources of news articles (according to Factiva, Dow Jones is part of the Factiva news sources and there are 32,000 sources in total for Factiva). To further address the sample-period difference, we present the results using Factiva articles between 2000 and 2012 (reported in the Online Appendix Table A.21). The results in the Online Appendix Table A.21 are closer to the Online Appendix Table A.20 than Table 2.

5.5. Robustness to the Factiva timestamp

A potential issue with using the Factiva database is that it does not provide accurate time stamps on when the news is released for all the news articles (only available for a subset of news wire articles). This could confound our results as the jumps are identified using close-to-close returns, while news-related variables are aggregated at the calendar day level. Therefore, our news-related variables would capture news released after the market close, although we expect the majority of news would be released during the day. We address this issue as follows (results are in the Online Appendix Table A.23). We use the RavenPack data, which includes precise time stamps for when each news article is released, to compare the differences associated with two methods. For method 1, we use the same timing criteria as with the Factiva data and aggregate news counts at the calendar day level, while we use a close-to-close definition to aggregate news counts for method 2. Note that we use RavenPack data for both methods and the only difference is the timing of the measurements for news. We find that the result using method 2 is more significant overall than the result using method 1. This indicates that our results using Factiva could be slightly underestimating (more importantly not overestimating) the true effect due to the unavailability of precise timestamps in Factiva.

6. Conclusion

Stock prices exhibit large, discrete changes, typically labelled as "jumps". A potential important source of jumps in stock returns can be material news events. In this paper, we test the relationship between jumps in stock returns and news flows. Explicitly analyzing the impact of news on return jumps requires firm-level news data. To do so, we collect 21 million news articles associated with more than 9000 publicly-traded companies from the Factiva database and use textual analyses to derive measures summarizing those news, including news frequency, tone, and uncertainty.

Our results show that these measures of news flow content are significantly related to nonparametric measures of jump intensity and jump-size distributions and explain an important fraction of variations in the jumps across individual companies. We analyze important determinants of the cross-sectional news-jump relationships, especially as they relate to firm characteristics (e.g., size, sector, institutional ownership, analyst coverage), news categories, and news coverage, as well as how those relationships have changed over time.

Our long sample period of news reveals that the impact and explanatory power of news flow on the stock return jumps have increased over time following a broad trend of improved data provision and transparency, particularly following the improvements in information dissemination technology associated with the creation of EDGAR in 1993 and the wide adoption of the Internet after 2000. This improvement in (tail) price informativeness associated with daily news for individual stock returns, complements the increased "clarity" over time associated with market-level jumps reported by Baker et al., 2021.

Equipped with our large cross-section of firms, we are also able to investigate determinants of the differential impact of news on stock return jumps across firms. We find important determinants with respect to firm characteristics, as well as news categories and sub-categories. For example, higher media visibility, higher analyst coverage, and a higher fraction of institutional ownership all increase the sensitivity of jump probabilities to news arrival, highlighting the importance of those channels with respect to quick incorporation of news into returns.

We provide results from a comprehensive range of robustness analyses which reveal that the effect of the daily news flow on daily stock return jumps and the differential effects over time and across companies is remarkably robust.

Our analyses of the relationships between news flow measures and stock return jumps for a large panel of companies could enrich the economic content of models for stock return dynamics with jumps which typically treat the sources of jumps as latent. Explicitly incorporating news processes in models of stock return jumps can potentially help identify jumps due to information arrival as opposed to jumps due to other reasons, such as liquidity or strategic trading based on private information. This may have broad implications for applications such as option pricing and risk management where stock return jump models are frequently used.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jfineco.2021.08.002.

References

Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. J. Financ. 59 (3), 1259–1294. doi:10.1111/j.1540-6261.2004.00662.x. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2004.00662.x

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Q. J. Econ. 131 (4), 1593–1636.

Kyle, A.S., Obizhaeva, A.A., Sinha, N.R., Tuzun, T., 2011. News arrival and trading game invariance. Unpublished working Paper.

Baker, S.R., Bloom, N., Davis, S., Sammon, M., 2021. What triggers stock market jumps? Unpublished working paper. Becker Friedman Institute for Economics, University of Chicago.

Bali, T.G., Bodnaruk, A., Scherbina, A., Tang, Y., 2018. Unusual news flow and the cross section of stock returns. Manag. Sci. 64 (9), 4137–4155.

Bégin, J.-F., Dorion, C., Gauthier, G., 2019. Idiosyncratic jump risk matters: evidence from equity returns and options. Rev. Financ. Stud. 33 (1), 155–211.

Berry, T.D., Howe, K.M., 1994. Public information arrival. J. Financ. 49 (4), 1331–1346.

Bollerslev, T., Todorov, V., 2011. Tails, fears and risk premia. J. Financ. 66 (6), 2165–2211.

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., 2019. Information, trading, and volatility: evidence from firm-specific news. Rev. Financ. Stud. 32 (3), 992–1033. doi:10.1093/rfs/hhy083. http://oup.prod.sis.lan/rfs/article-pdf/32/3/992/27711227/hhy083.pdf

Baker, S.R., Bloom, N., Davis, S.J., Kost, K., 2019. Public news and stock market volatility. Unpublished working paper.

Chan, W.S., 2003. Stock price reaction to news and no-news: drift and reversal after headlines. J. Financ. Econ. 70 (2), 223–260.

Christoffersen, P., Jacobs, K., Ornthanalai, C., 2012. Dynamic jump intensities and risk premiums: evidence from S&P500 returns and options. J. Financ. Econ. 106 (3), 447–472. doi:10.1016/j.jfineco.2012.05.017.

Cutler, D.M., Poterba, J.M., Summers, L.H., 1989. What moves stock prices? J. Portf. Manag. 15 (3), 4–12. doi:10.3905/jpm.1989.409212. http://jpm.iijournals.com/content/15/3/4.full.pdf
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), 1–38.
Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. J. Financ. 64 (5), 2023–2052.
Gao, M., Huang, J., 2020. Informing the market: the effect of modern information technologies on information production. Rev. Financ. Stud. 33 (4), 1367–1411.
Gentzkow, M., Kelly, B.T., Taddy, M., 2019. Text as data. J. Econ. Lit. 57 (3), 535–574.
Gilder, D., Shackleton, M.B., Taylor, S.J., 2014. Cojumps in stock prices: empirical evidence. J. Bank. Financ. 40, 443–459. doi:10.1016/j.jbankfin.2013.04.025.
Bybee, L., Kelly, B.T., Manela, A., Xiu, D., 2019. The structure of economic news. Unpublished working paper.
Goldstein, I., Yang, S., Zuo, L., 2020. The real effects of modern information technologies. Unpublished working paper.
Gűrkaynak, R.S., Kisacikoğlu, B., Wright, J.H., 2020. Missing events in event studies: identifying the effects of partially measured news surprises. Am. Econ. Rev. 110 (12), 3871–3912. doi:10.1257/aer.20181470.
Huang, A.G., Tan, H., Wermers, R., 2019. Institutional trading around corporate news: evidence from textual analysis. Rev. Financ. Stud. 33 (10), 4627–4675.
Jiang, H., Li, S.Z., Wang, H., 2021. Pervasive underreaction: evidence from high-frequency data. J. Financ. Econ. 141 (2), 573–599.
Kelly, B.T., Manela, A., Moreira, A., 2019. Text selection. J. Bus. Econ. Stat. doi:10.1080/07350015.2021.1947843. forthcoming
Ke, Z.T., Kelly, B.T., Xiu, D., 2019. Predicting returns with text data. Unpublished working paper.
Lee, S.S., 2012. Jumps and information flow in financial markets. Rev. Financ. Stud. 25 (2), 440–479.
Lee, S.S., Mykland, P.A., 2008. Jumps in financial markets: a new nonparametric test and jump dynamics. Rev. Financ. Stud. 21 (6), 2535–2563. doi:10.1093/rfs/hhm056. http://rfs.oxfordjournals.org/content/21/6/2535.full.pdf+html
Loughran, T., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. J. Financ. 66 (1), 35–65. doi:10.1111/j.1540-6261.2010.01625.x.
Loughran, T., McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. J. Financ. Econ. 109 (2), 307–326. doi:10.1016/j.jfineco.2013.02.017.
Loughran, T., McDonald, B., 2016. Textual analysis in accounting and finance: a survey. J. Account. Res. 54 (4), 1187–1230. doi:10.1111/1475-679X.12123. https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-679X.12123
Maheu, J.M., McCurdy, T.H., Zhao, X., 2013. Do jumps contribute to the dynamics of the equity premium? J. Financ. Econ. 110 (2), 457–477. doi:10.1016/j.jfineco.2013.07.006.
Manela, A., Moreira, A., 2017. News implied volatility and disaster concerns. J. Financ. Econ. 123 (1), 137–162. doi:10.1016/j.jfineco.2016.01.032.
Mitchell, M.L., Mulherin, J.H., 1994. The impact of public information on the stock market. J. Financ. 49 (3), 923–950.
Neuhierl, A., Scherbina, A., Schiusene, B., 2013. Market reaction to corporate press releases. J. Financ. Quant. Anal. 48 (4), 1207–1240.
Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. J. Financ. Econ. 63 (1), 3–50.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. J. Financ. 62 (3), 1139–1168. doi:10.1111/j.1540-6261.2007.01232.x.
Tetlock, P.C., 2010. Does public financial news resolve asymmetric information? Rev. Financ. Stud. 23 (9), 1437–1467.
Tetlock, P.C., 2011. All the news that's fit to reprint: do investors react to stale information? Rev. Financ. Stud. 24 (5), 1481–1512.
Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. J. Financ. 63 (3), 1437–1467. doi:10.1111/j.1540-6261.2008.01362.x. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2008.01362.x
Veldkamp, L.L., 2006. Media frenzies in markets for financial information. Am. Econ. Rev. 96 (3), 577–601.
Zhao, X., 2017. Does information intensity matter for stock returns? evidence from form 8-k filings. Manag. Sci. 63 (5), 1382–1404.