Investor Disagreement: The Secret Fuel Behind Stock Price Jumps *

Hasan Fallahgoul†
Monash University

Xin Lin‡
Monash University

This Version: February 28, 2023

Link to Most Recent Version

ABSTRACT

Our study offers a comprehensive analysis of the factors contributing to the occurrence of stock price jumps, with a focus on the role of investor disagreement and news tone. We find that higher levels of disagreement among investors are associated with a greater likelihood of jumps in daily stock returns, with negative jumps more likely to occur than positive jumps when investor disagreement is higher. Our results suggest that information differences are more important for jump occurrences than differences in interpretation, and that disagreement between sophisticated and unsophisticated investors can play a significant role in triggering stock jumps.

JEL classification: C58; G4; C32

Keywords: Investor Disagreement; News; Jumps; Hawkes Processes;.

*We are grateful to the StockTwits company for assistance and support. We thank Giovanni Barone-Adesi, Loriano Mancini, and seminar participants at the Swiss Finance Institute at University of Lugano for comments.

†Hasan Fallahgoul, Monash University, School of Mathematics and Centre for Quantitative Finance and Investment Strategies, 9 Rainforest Walk, 3800 Victoria, Australia. E-mail: hasan.fallahgoul@monash.edu.

‡Xin Lin, Monash University, School of Mathematics and Centre for Quantitative Finance and Investment Strategies, 9 Rainforest Walk, 3800 Victoria, Australia. E-mail: xin.lin@monash.edu

The efficient market hypothesis posits that the stock market fully and immediately incorporates all available information into stock prices, making it impossible to outperform the market through stock picking or market timing (Fama, 1970). However, empirical research has shown that certain events, such as earnings announcements, can lead to large and sudden movements in stock prices that are difficult to explain based on publicly available information alone. One possible explanation for these sudden price jumps is the existence of investor disagreement, where market participants have differing views on the value of a stock and its future prospects (Daniel, Hirshleifer, and Subrahmanyam, 1998; Hong and Stein, 1999, 2003, 2007; Diether, Malloy, and Scherbina, 2002). Recent studies have investigated the impact of investor disagreement on stock price movements, and have found that disagreement can lead to increased volatility and higher trading volume (Baker and Wurgler, 2006; Lamont and Thaler, 2003; Cookson and Niessner, 2020; Chang, Hsiao, Ljungqvist, and Tseng, 2022). These findings suggest that investor disagreement may play an important role in the dynamics of stock prices, and that it is an important factor to consider when analyzing the stock market.

The aim of this study is to investigate the relationship between investor disagreement and stock price jumps, and to explore the underlying mechanisms involved. Specifically, the research questions addressed in this study are: (i) whether investor disagreement leads to stock price jumps; (ii) if so, what is the role of different sources of disagreement, such as different information sets and different interpretations of information, in causing stock price jumps?; and (iii) what is the impact of information flow diffusion through disagreement on stock price jumps? To answer these questions, we analyze a large dataset of over 173 million tweets from StockTwits covering approximately 3,000 publicly-traded companies using an empirical approach. We employ a contagion model (Aït-Sahalia, Laeven, and Pelizzon, 2014; Aït-Sahalia, Cacho-Diaz, and Laeven, 2015) to capture the dynamics of stock returns and measure the intensity and distribution of stock price jumps in response to investor disagreement and news events.

We begin by employing the Hawkes jump-diffusion (HJD) model proposed by (Aït-Sahalia et al., 2014, 2015) to investigate the impact of investor disagreement on stock price jumps. This parametric model allows us to recover unobservable features of stock jumps and provides a better understanding of the determinants of various jumps, including memoryless jumps, self-exciting jumps, and market-exciting jumps. In particular, we use the bivariate HJD model between individual and market return processes to decompose jumps into different components. Memoryless jumps are exogenous, while self-exciting jumps depend on the past jumps of the stock, and market-exciting jumps depend on the jumps in the market. This decomposition enables a deeper understanding of the sources of jumps, such as instantaneous, self-reinforcing, and market-induced jumps, and has important implications for risk management and portfolio diversification (Aït-Sahalia et al., 2014)).[1]

We proceed by constructing a measure of investor disagreement based on sentiment extracted from a large corpus of real-time tweets posted on the StockTwits platform, totaling more than 173 million tweets. This measure provides a rich and multimodal characterization of investor disagreement, enabling us to assess granular variation within and across various groups of investors according to different criteria such as their number of followers, experience, investment strategy, approach, and horizon. The StockTwits platform has been the subject of growing interest in the literature, as it allows for the measurement of investor sentiment in real-time (see, for example, Deng, Huang, Sinha, and Zhao (2018); Cookson and Niessner (2020); Cookson, Engelberg, and Mullins (2022)). Our measure offers the advantage of being transparent, granular, and real-time compared to other available measures of investor sentiment. Market-based measures of investor sentiment, such as those proposed by Baker and Wurgler (2006), have the advantage of being readily available at a relatively high frequency. However, such measures also have the disadvantage of being the equilibrium outcome of many economic forces other than investor sentiment (see Baker and Wurgler (2007) for a comprehensive survey of the literature). Other studies have explored how investor sentiment can be directly measured through the internet search behavior of households, as shown by Da, Engelberg, and Gao (2015). We then investigate the extent to which the time-varying estimated jump intensity

[1] Instantaneous jump risk refers to a one-time risk that does not have an impact on future jump risks. In contrast, self-reinforcing jump risk occurs when a jump in one stock increases the likelihood of that stock jumping again in the future, while market-induced jump risk describes the risk that a jump in the market leads to multiple stocks jumping in the future.

processes and realized jump-size distributions are related to the disagreement and news variables.

Our study reveals a significant positive relationship between investor disagreement and the likelihood of stock price jumps, indicating that higher levels of disagreement among investors are associated with a greater likelihood of jumps in daily stock returns. This finding is consistent with previous studies in the literature (Hong and Stein, 2007). Importantly, our study shows that investor disagreement is positively related to the likelihood of self- and market-exciting jumps, implying that self-reinforcing jump risk and market-induced jump risk are greater when investor disagreement is higher. Furthermore, we find that negative jumps are more likely to occur than positive jumps when investor disagreement is higher, and this asymmetry is more pronounced for firms with larger market capitalizations. This result is consistent with previous researches that have also found asymmetry in stock prices (Chang et al., 2022) and asymmetry in market reactions to news (Engle, Ghysels, and Sohn, 2013).

In addition to investor disagreement, we also investigate the role of news tone in explaining the variations in estimated jump intensities. Our study finds that news tone is positively related to the estimated intensity process for all jumps, but its significance is weaker than investor disagreement. This result is consistent with previous research that has also found that news sentiment plays a role in stock market volatility (Baker and Wurgler, 2006; Tetlock, 2007). Moreover, we find that negative news tone is associated with a higher likelihood of negative jumps, while positive news tone is associated with a higher likelihood of positive jumps. This result is consistent with the notion that bad news has a stronger impact on stock market reactions than good news (Tetlock, 2007).[2]

Our study's findings suggest that investor disagreement and news tone play complementary roles in explaining the variations in the estimated jump intensities. Specifically, investor disagreement reflects differences in information and differences in interpretations among investors, while news tone captures the overall sentiment conveyed by the news media. Together, these factors contribute to the occurrences of stock price jumps.

[2]There is a positive relationship between economic activity and stock market volatility, but that this relationship is asymmetric, with negative economic shocks having a larger effect on volatility than positive economic shocks.

To further investigate the impact of disagreement on stock price jumps, we examine two distinct measures of disagreement. Our analysis reveals that both disagreement within-group (referring to differences in opinion among investors who follow the same investment strategy) and disagreement cross-group (referring to differences in opinion among investors who have different investment strategies) are positively related to the estimated intensity of all jumps in daily stock returns. However, disagreement within-group has a greater impact on jump occurrences than disagreement cross-group, indicating that differences in information among investors are more important for jumps than differences in interpretation. These results complement the findings of a prior study (Cookson and Niessner, 2020) that observed a similar pattern for trading volume.

Moreover, we find that the relationship between disagreement within-group and jump occurrences is stronger than the relationship between disagreement cross-group and jump occurrences for any type of jumps. Disagreement within-group is positively related to any estimated jump intensity, indicating that jumps of any type are more likely to occur when investors disagree more due to larger information differences. However, disagreement cross-group becomes significantly negatively related to the estimated intensity of market-exciting jumps (jumps that are excited by past jumps in the market) while still being positively related to other estimated jump intensities. These results suggest that both information differences and interpretation differences are economically and statistically important for any type of jumps, but that the impact of information differences is greater.

Based on our analysis, we also investigate the impact of disagreement between sophisticated and unsophisticated investors on stock price jumps. Our findings suggest that disagreement between sophisticated and unsophisticated investors with the same investment philosophy plays a significant role in triggering stock jumps, particularly negative jumps, and in determining the size of these jumps. This result is consistent with the gradual information diffusion hypothesis, which posits that sophisticated investors acquire information earlier than unsophisticated investors, leading to disagreement between them. The significant and positive relationship between the estimated intensity of all jumps and the disagreement between sophisticated and unsophisticated investors, as well as the relationship between jump size moments and this disagreement, provide support for the gradual information diffusion hypothesis.

This result is consistent with previous research on the role of information in financial markets. For example, Hong and Stein (2007) found that gradual diffusion of information can cause prices to gradually respond to that information. Similarly, Cookson and Niessner (2020) found that gradual information diffusion plays an important role in trading volume. Our study extends this literature by examining the effect of disagreement between sophisticated and unsophisticated investors with the same investment philosophy on stock jumps, and its implications for jump size.

Our study employs a bivariate HJD process calibrated to approximately 3,000 pairs of individual stocks and SPDR S&P 500 ETF Trust (SPY), yielding rich empirical insights into the nature of stock price jumps.[3] We utilize two algorithms from Lee and Mykland (2008) to detect jumps and our results are robust to different jump detection algorithms. Based on our analysis, the average estimated jump intensity (AEJI) of all jumps for all firms ranges from a minimum of 1.84 jumps to a maximum of 23.76 jumps per year. About 50% of firms have more than 8.65 jumps per year. Self-exciting jumps contribute the highest number of jumps, with the median AEJI for self-exciting jumps across firms being 5.94 jumps per year. This is followed by memoryless and market-exciting jumps, with median AEJIs of 1.08 and 0.98 jumps per year, respectively.

Our findings also indicate that for approximately 50% of firms, the percentage contribution of self-exciting jumps is more than 60%, followed by market-exciting and memoryless jumps, where their percentage contribution for 50% of firms is about 16% and 15%, respectively. When we split firms based on their size, we observe that the self-exciting jumps component is the highest contributor of jumps across firms with different sizes, followed by market-exciting (memoryless) jumps for large (small) firms. This suggests that large-size firms have a higher exposure to the market than small-size firms. Note that we use AEJI to measure the contribution of different jump components in overall jumps. Our results are obtained from a rigorous analysis of a large dataset of stock and ETF pairs, and our findings are robust to various jump detection algorithms.

[3] The SPY is an exchange-traded fund that tracks the performance of the S&P500 stock market index, and is used instead of the Standard and Poor's 500 because StockTwit's users use the cashtag $SPY.

Recent research has highlighted the importance of understanding the drivers of stock price jumps. Studies by Jeon, McCurdy, and Zhao (2021) and Baker, Bloom, Davis, and Sammon (2022) have identified news flow as a key factor in triggering jumps at the firm and market levels. Specifically, they find that the frequency, content, and source of news flow are significantly related to stock price jumps, and that the clarity of news sources is an important factor in predicting future volatility. These findings suggest that investors are becoming more adept at processing news information.

While previous research has shown that investor disagreement is positively related to stock price jumps (Hong and Stein, 2007; Vayanos and Wang, 2012; Chang et al., 2022), our study builds on this research by exploring the simultaneous role of news and disagreement in various types of stock jumps using a bivariate HJD model. This model allows us to decompose jumps into three distinct components and gain a more nuanced understanding of the mechanisms driving stock price jumps and how investor disagreement may affect each component differently.

Our study provides novel and important insights into the relationship between news flow, investor disagreement, and stock price jumps. In particular, we find that negative jumps are more likely to occur than positive jumps when investor disagreement is higher, consistent with previous research that has found asymmetry in stock market reactions to news (Engle et al., 2013). Furthermore, our investigation of the impact of disagreement between sophisticated and unsophisticated investors on stock price jumps provides further support for the gradual information diffusion hypothesis, which has been extensively studied in finance literature (Barberis, Shleifer, and Vishny, 1998; Dasgupta and Prat, 2008). By uncovering these complex relationships, our study sheds new light on the dynamics of stock price movements and provides important insights for investors and policymakers.

The rest of the paper is organized as follows. Section I introduces the contagion framework and the HJD model of Aït-Sahalia et al. (2015). Section II discusses our data collection. Section III present the empirical results. Section IV concludes.

I. The Model

In this section, we present a thorough analysis of extreme returns, commonly referred to as jumps. We begin by examining empirical regularities in jump behavior and then introduce a conceptual framework for modeling jumps. Specifically, we discuss the bivariate Hawkes-Jump-Diffusion (HJD) process, as described in Aït-Sahalia et al. (2015). We also detail our approach to estimating unobserved daily jump intensities from observed daily stock returns.

Subsequently, we provide a decomposition of the jump intensity process, allowing us to quantify the contribution of each component to the overall jump intensity process. *It is worth noting that while the HJD process is a reduced-form model and cannot explain the sources of contagion or jump clustering observed in the data, we can utilize implied jump intensity processes to test news and disagreement sources of jumps.*

A. *Jump clustering in asset returns*

Jump clustering in time and across assets is widely observed in empirical studies of asset returns, as documented in previous research (Aït-Sahalia et al., 2014, 2015; Aït-Sahalia and Hurd, 2015). As shown in Figure 1, this phenomenon is also evident in the returns of SPDR S&P 500 ETF Trust (SPY), Boeing (BA), and First Trust Small Cap Core AlphaDEX Fund (FYX) firms over a long period from June 01, 2010 until August 31, 2021.[4] Specifically, the ellipse in the figure marks a period of time when jumps occur in close succession, indicating jump clustering in time, while the rectangles indicate periods when jumps tend to occur across several assets, revealing jump clustering across assets. Notably, jump clustering in time and across assets is observed during both non-crisis and crisis periods.

Given the empirical regularity of jump clustering in asset returns, it is essential to work with a model of asset returns that incorporates the time-dependency of jumps across both time and stocks. By doing so, we can gain a more nuanced understanding of the underlying mechanisms driving stock price jumps and better inform investment decisions.

[4]SPY is an exchange-traded fund that closely tracks the S&P 500 index.

Insert Figure 1 about here

B. A conceptual framework

Figure 2 provides an animated example of the framework of interest. At time t_1, an exogenous shock, such as an unexpected public announcement, triggers an initial jump in SPY, which we refer to as a memoryless jump. However, the impact of this shock can be far-reaching and can lead to self-exciting jumps in SPY and market-exciting jumps in other assets, such as the jump in BA observed at time t_2 in the figure. These self- and market-exciting jumps can further excite jumps in themselves and other assets, leading to a contagion effect.

Insert Figure 2 about here

Conceptual example 2 is a possible illustration for a series of events. In practice, it is often difficult to identify the source of jumps, whether they are memoryless, self-exciting, or market-exciting. Instead, what we are interested in are the probabilities of observing these three types of jumps and the possibility of contagion.

C. The Hawkes jump-diffusion process

Multivariate Hawkes processes provide a way to model the dependency of jumps over time and the interaction of jumps across different assets, allowing a jump in one asset to increase the likelihood of future jumps both in itself and in other stocks. This is an important stylized fact of stock return jumps, which are known to cluster over time and across stocks. In order to capture this behavior, we employ the HJD process proposed by Aït-Sahalia et al. (2015) to model stock returns. Specifically, we use the bivariate case of the HJD process, which is given by

$$\begin{aligned} dX_{i,t} &= \mu_i dt + \sqrt{\theta_i} dW_{i,t} + Z_{i,t} dN_{i,t} \qquad (1) \\ dX_{j,t} &= \mu_j dt + \sqrt{\theta_j} dW_{j,t} + Z_{j,t} dN_{j,t} \\ d\lambda_{i,t} &= \alpha_i(\lambda_{i,\infty} - \lambda_{i,t})dt + \beta_{i,i} dN_{i,t} + \beta_{i,j} dN_{j,t} \end{aligned}$$

$$d\lambda_{j,t} = \alpha_j(\lambda_{j,\infty} - \lambda_{j,t})dt + \beta_{j,j}dN_{j,t} + \beta_{j,i}dN_{i,t} \quad j \neq i.$$

In this model, $X_{i,t}$ and $X_{j,t}$ represent the log-returns of two different assets i and j, respectively. The drift terms μ_i and μ_j capture the long-term mean returns of the assets, while the constant volatility terms θ_i and θ_j capture their respective volatilities. The Brownian motion processes $W_{i,t}$ and $W_{j,t}$ represent the continuous part of the price processes, while $Z_{i,t}$ and $Z_{j,t}$ represent the size of the jumps, which follow a pre-defined cumulative probability distribution.

The jump processes $N_{i,t}$ and $N_{j,t}$ are Hawkes processes, with $\lambda_{i,t}$ and $\lambda_{j,t}$ representing the corresponding jump intensities. The long-run means of the jump intensities are denoted by $\lambda_{i,\infty}$ and $\lambda_{j,\infty}$, while α_i and α_j are the mean-reversion rates. The self-exciting (SE) coefficients $\beta_{i,i}$ and $\beta_{j,j}$ capture the influence of past jumps on the probability of future jumps for the same asset, while the cross-exciting (XE) coefficients $\beta_{i,j}$ and $\beta_{j,i}$ capture the influence of past jumps in one asset on the probability of future jumps in the other asset.

When one of the assets is a market index (e.g., the S&P 500 index), we introduce the market-exciting (ME) coefficient $\beta_{i,SPY}$ to capture the influence of market-level jumps on the probability of future jumps in asset i.[5]

D. *Estimated jump intensity and its decomposition*

We can classify jumps into three groups: memoryless jumps, self-exciting jumps, and cross-exciting jumps. Memoryless jumps are independent of past jumps and have a constant intensity (e.g., Poissonian jumps). Self-exciting jumps are those that are excited by the stock's own past jumps, while cross-exciting jumps are those that are excited by the past jumps of another stock. If jumps in a stock are excited by past jumps in the market, we call these cross-exciting jumps "market-exciting" jumps. We consider the following jump intensity process

$$\lambda_{i,t+1} = \lambda_{i,t} + \alpha_i(\lambda_{i,\infty} - \lambda_{i,t})dt + \beta_{i,i}dN_{i,t} + \beta_{i,j}dN_{j,t} \tag{2}$$

[5]For further details about the Hawkes process and its properties, as well as a detailed discussion about this particular model, we refer interested readers to Aït-Sahalia et al. (2014, 2015); Aït-Sahalia and Hurd (2015).

$$\lambda^{+}_{i,t+1} = \lambda^{+}_{i,t} + \alpha_i(\lambda_{i,\infty} - \lambda^{+}_{i,t})dt + \beta_{i,i}dN^{+}_{i,t} + \beta_{i,j}dN^{+}_{j,t} \tag{3}$$

$$\lambda^{-}_{i,t+1} = \lambda^{-}_{i,t} + \alpha_i(\lambda_{i,\infty} - \lambda^{-}_{i,t})dt + \beta_{i,i}dN^{-}_{i,t} + \beta_{i,j}dN^{-}_{j,t} \tag{4}$$

$$\lambda^{ExcSE}_{i,t+1} = \lambda^{ExcSE}_{i,t} + \alpha_i(\lambda_{i,\infty} - \lambda^{ExcSE}_{i,t})dt + \beta_{i,j}dN_{j,t}$$

$$\lambda^{ExcXE}_{i,t+1} = \lambda^{ExcXE}_{i,t} + \alpha_i(\lambda_{i,\infty} - \lambda^{ExcXE}_{i,t})dt + \beta_{i,i}dN_{i,t}$$

$$\lambda^{SE}_{i,t+1} = \lambda_{i,t+1} - \lambda^{ExcSE}_{i,t+1} \tag{5}$$

$$\lambda^{XE}_{i,t+1} = \lambda_{i,t+1} - \lambda^{ExcXE}_{i,t+1} \tag{6}$$

$$\lambda^{EXO}_{i} = \lambda_{i,\infty}, \quad t = 0, 1, 2, ..., T \tag{7}$$

To estimate the time-varying intensity of these different types of jumps, we can use the parameters of the bivariate HJD model with the pair (i, j). Specifically, we can estimate the intensity of all jumps, positive jumps, and negative jumps at day $t+1$ using equations (2), (3), and (4), respectively. In these equations, $\lambda_{i,t+1}$, $\lambda^{+}_{i,t+1}$, and $\lambda^{-}_{i,t+1}$ are the estimated intensity of all, positive, and negative jumps at day $t+1$, respectively. We estimate these intensities by adding the drift term, $\alpha_i(\lambda_{i,\infty} - \lambda_{i,t})dt$, and the contribution of past jumps to the intensity at day $t+1$. The contribution of past jumps is given by the jump indicators $dN_{i,t}$ and $dN_{j,t}$, which are 1 if i's or j's daily return is identified as a jump at day t, respectively. If a jump occurs in stock i at day t, this increases the jump intensity of stock i at day $t+1$ by $\beta_{i,i}$. If a jump occurs in stock j at day t, this increases the jump intensity of stock i at day $t+1$ by $\beta_{i,j}$.

We can also estimate the intensity of self-exciting jumps and cross-exciting jumps using equations (5) and (6), respectively. Here, $\lambda^{ExcSE}_{i,t+1}$ and $\lambda^{ExcXE}_{i,t+1}$ are the estimated intensity of jumps when conditioning on the absence of self-excitation and cross-excitation from 0 to t, respectively. The intensity of self-exciting jumps at day $t+1$ is given by the difference between the estimated intensity of all jumps at day $t+1$ and the estimated intensity of jumps conditioned on the absence of self-excitation up to time t. The intensity of cross-exciting jumps at day $t+1$ is given by the difference between the estimated intensity of all jumps at day $t+1$ and the estimated intensity of jumps conditioned on the absence of cross-excitation up to time t. Finally, we can estimate the intensity of exogenous jumps as the constant intensity $\lambda_{i,\infty}$ given by equation (7). All of these intensity estimates are updated recursively from time t to time $t+1$ for $t = 0, 1, 2, \ldots, T$.

The intensity process of a jump-diffusion Hawkes process can be described by (1) or, equivalently, by the jump intensity dynamic given

$$\lambda_{i,t} = \lambda_{i,\infty} + \int_{-\infty}^{t} \beta_{i,i} e^{-\alpha_i (t-s)} dN_{i,s} + \int_{-\infty}^{t} \beta_{i,j} e^{-\alpha_i (t-s)} dN_{j,s}.$$

The coefficients $\lambda_{i,\infty}$, α_i, $\beta_{i,i}$, and $\beta_{i,j}$ are assumed to be identical for all jumps and signed jumps. The expected jump intensity, $\mathbb{E}[\lambda_{i,t}]$, can be computed as shown by

$$\mathbb{E}[\lambda_{i,t}] = \lambda_{i,\infty} + \frac{\beta_{i,i}}{\alpha_i} \times \mathbb{E}[\lambda_{i,t}] + \frac{\beta_{i,j}}{\alpha_i} \times \mathbb{E}[\lambda_{j,t}]$$

which involves the average number of self-exciting and cross-exciting jumps, denoted by $\frac{\beta_{i,i}}{\alpha_i}$ and $\frac{\beta_{i,j}}{\alpha_i}$, respectively. These coefficients quantify the expected intensity of self-exciting and cross-exciting jumps, respectively. The exogenous, self-exciting, and cross-exciting components of the jump intensity process are defined as $\lambda_{i,\infty}$, $\frac{\beta_{i,i}}{\alpha_i}\mathbb{E}[\lambda_{i,t}]$, and $\frac{\beta_{i,j}}{\alpha_i}\mathbb{E}[\lambda_{j,t}]$, respectively, with $\frac{\beta_{i,j}}{\alpha_i}\mathbb{E}[\lambda_{j,t}]$ representing the expected intensity of cross-exciting jumps from stock j to stock i. When stock j is the market index (e.g., the SPY), the cross-exciting component is referred to as the market-exciting component and is denoted by $ME_i\%$. To assess the contribution of the three components to the jump intensity process in stock i, we define the exogenous, self-exciting, and cross-exciting components as a percentage of the total jump intensity, as given by

$$\begin{aligned}
EXO_i\% &= \frac{\lambda_{i,\infty}}{\mathbb{E}[\lambda_{i,t}]} \times 100 \\
SE_i\% &= \frac{\beta_{i,i}}{\alpha_i} \times 100 \\
XE_i\% &= \frac{\beta_{i,j}}{\alpha_i} \times \frac{\mathbb{E}[\lambda_{j,t}]}{\mathbb{E}[\lambda_{i,t}]} \times 100, \quad i \neq j.
\end{aligned}$$

Specifically, $EXO_i\%$, $SE_i\%$, and $XE_i\%$ represent the percentage contribution of the exogenous, self-exciting, and cross-exciting components to the jump intensity process in stock i, respectively. These definitions allow for a more nuanced understanding of the dynamics of the jump intensity process and its underlying components.[6]

II. Data

A. StockTwits

Our measure of investor disagreement is constructed from data collected from a social media investing platform, StockTwits. StockTwits allows users to express their opinions about stocks and to provide their investment approach, holding period, and experience level on their profiles. Specifically, users can post a message, link it to one or several particular firms using "cashtags" with the symbol ($), and label that message with a sentiment, which can be bullish, bearish, or unspecified. We collected messages posted between June 1, 2010 and August 31, 2021, eliminating messages with multiple cashtags from our analysis. To classify unclassified messages, we used the maximum entropy method.

Detailed information about the StockTwits data can be found in Cookson and Niessner (2020) and Fallahgoul (2021). After filtering messages with only one symbol per message, we were left with approximately 19,469 unique symbols and 173,898,340 messages posted by 1,104,849 unique users. We then took the intersection between those 19,469 symbols in StockTwits and all symbols in the Center for Research in Security Prices (CRSP), which gave us approximately 11,267 unique symbols and 154,860,211 messages posted by 995,046 unique users. We further filtered symbols with non-missing daily returns during the sample period. As a result, we were left with 2,679 stocks and 34,373,324 messages posted by 94,827 unique StockTwits users over 2,834 trading days.[7]

Table 1 provides the characteristics of messages and active users for remaining stocks after filtering. Panel A shows the characteristics for investor experience. Intermediate, novice, and professional investors represent 10.77%, 6.83%, and 4.89% of users, respectively, while professional investors post the largest proportion of messages, 21.78%, fol-

[6]Further details on the calculation of these components can be found in Filimonov and Sornette (2012) and Kim, Paini, and Jurdak (2019).

[7]Table IA.1 illustrates how we filtered stocks to obtain the final sample and how many observations were lost due to each filtering action.

lowed by intermediate and novice investors, 17.60% and 4.62%, respectively. Panel B reports the characteristics for investor philosophy. The most common investment approach is technical, representing 7.63% of users and about 18.91% of messages. Growth and momentum investors make up the second and third largest proportion of users (4.53% and 4.29%), followed by fundamental and value investors (2.84% and 2.31%). However, fundamental and values investors post the second and third largest proportion of messages (7.04% and 5.53%), followed by momentum and growth investors (6.01% and 4.50%). The least common investment approach is global macro, which represents only 0.63% of users and 0.72% of messages. Panel C shows the characteristics for holding period. In terms of investment horizon, the most common one is swing trader, representing the 8.97% of users and 16.94% of messages, followed by long term investor (4.84% of users and 9.77% of messages) and day trader (4.63% of users and 8.25% of messages). Position Trader only represents 4.16% of users and 8.39% of messages.

Insert Table 1 about here

To construct a sentiment and disagreement measure from bullish and bearish messages, we first labeled each bearish message as -1 and each bullish message as 1. We then took the arithmetic average of these classifications at the $group1 \times day \times group2$ level

$$\text{AvgSentiment}_{i,t,g} = \frac{N^{Bullish}_{i,t,g} - N^{bearish}_{i,t,g}}{N^{Bullish}_{i,t,g} + N^{bearish}_{i,t,g}}$$

where $N^{Bullish}itg$ and $N^{bearish}itg$ are the number of bullish and bearish messages per group1, day, and group2, respectively. Group1 can be either all firms, sectors, industries, or a specific firm, sector, or industry. Group2 can either be all investors or investors with a given investment philosophy, experience level, or holding period.

Next, we compute a measure of disagreement for a given firm × day × group as

$$\text{Disagreement}_{i,t,g} = \sqrt{1 - \text{AvgSentiment}^2_{i,t,g}} \tag{8}$$

where i is a particular firm and g can be a group of investors or all investors.[8]

B. Firm characteristics

To investigate potential drivers of jumps in financial markets, this study examines seven key firm characteristics (Jeon et al. (2021) and Chang et al. (2022)) from three comprehensive databases: StockTwits, RavenPack, and CRSP/Compustat. Specifically, we analyze the *Number of Messages* and *Disagreement* variables, which represent the number of investor-posted messages and overall disagreement among all investors or a given group of investors at the firm-day level in StockTwits, respectively. In addition, we investigate the *Number of News* and *News Tone* variables, which measure the number of news articles and overall news sentiment at the firm-day level, respectively, sourced from RavenPack. Furthermore, we examine *Size*, *12-month Momentum*, and *Idiosyncratic Volatility*, which represent market capitalization, past 12-months return, and idiosyncratic volatility, respectively, retrieved from CRSP/Compustat.

Note that *News Tone* is the variable *event_sentiment_score* in RavenPack, while *Size* is the variable *me* retrieved from CRSP. The *Idiosyncratic Volatility* variable is calculated as the variance of residuals from the Fama and French (1993) three-factor model, based on daily returns over the past three months. It is worth noting that *Size* and *12-month Momentum* are measured at the firm-day level, whereas *Idiosyncratic Volatility* is measured at the firm-month level.

We address missing values of firm characteristics in the following manner. Firstly, we group firms into three categories based on their daily market capitalization during the sample period: large, medium and small. This is achieved by sorting the average market capitalization and ensuring an equal number of firms are allocated to each group.[9] Subsequently, for each firm and on each day, any missing characteristic value is replaced with the cross-sectional median for that particular characteristic within the relevant size group to which the firm belongs. This methodology ensures that we retain the size effect

[8] For more details about this disagreement measure and its properties, please see Cookson and Niessner (2020).

[9] The allocation of firms to size groups is determined by sorting their average daily market capitalization during the sample period and allocating an equal number of firms to each size group.

on the value of other firm characteristics, which is of interest to us in examining whether the impact of the same characteristic varies across different size groups with respect to jumps.

Table 2 presents the summary statistics of the average of daily observations of firm characteristics for the 2,679 stocks, with *Number of Messages* and *Disagreement* calculated from all tweets available after taking successive filtering actions described in IA.1. Panel A and B illustrate the distribution of characteristics before and after filling missing characteristics. Notably, the range of *Disagreement* is between 0.00 and 0.85, with NFLX and PWZ having the maximum and minimum values, respectively. AAPL has the highest values for both *Number of Messages* (1,070) and *Number of News* (766), as well as the largest *Size* (8.04E+08). On the other hand, CHIE has the smallest *Size* (3.27E+03), *Number of Messages* (0), and *Number of News* (0).

Insert Table 2 about here

These findings are consistent with the well-established notion that companies with larger market capitalization generally receive more media and investor attention. This is further supported by Figure 3, which displays the correlation matrix between the average of daily observations of firm characteristics for the 2,679 stocks. The three most highly correlated pairs of characteristics are (*Size*, *Number of News*), (*Number of Messages*, *Number of News*), and (*Size*, *Number of Messages*), with correlation coefficients of 0.81, 0.67, and 0.58, respectively.

To visually inspect the behavior of the firm characteristics over time, we present Figure 4, which shows the time series of stock price, disagreement, news tone, and number of messages of two firms, Microsoft Corporation (MSFT) and BA, from October 1, 2019 to August 31, 2022. Panel (a) shows that the price of MSFT started to fall sharply at around late January 2021, with the volume of investor posted messages surging, overall disagreement among investors increasing and remaining high, and the news tone dropping greatly twice and rising greatly twice during this period. The price rebounded at around the end of March. In panel (b), we observe that the number of investor posted messages also surged during periods of sharp price declines and increases, while the link between the price and news tone is unclear. Interestingly, panel (b) shows that overall disagreement among investors increased (decreased) when the price sharply fell (rose).

This visual inspection of the price and these three firm characteristics provides insight into the potential role of disagreement, news tone, and the number of messages in driving jumps. Theoretical models, such as those proposed by Hong and Stein (1999, 2003); Hong, Kubik, and Stein (2004), may guide further exploration of the interactions among these firm characteristics and their contribution to jumps.

Insert Figure 4 about here

III. Empirical Results

This section presents the empirical study of jumps and their determinants. We begin by outlining our methodology for jump detection and providing statistics of realized jumps, as well as the estimated intensities of various types of jumps. We then delve into the role of disagreement and news in various types of jumps, exploring their relationship with the estimated jump intensity and the realized jump-size distribution.

To shed further light on the determinants of jumps, we examine the roles of two key sources of disagreement – information differences and differences in information interpretation – in various types of jumps. Additionally, we investigate the importance of gradual information diffusion from sophisticated investors to unsophisticated investors for different types of jumps.

To ensure the robustness of our findings, we provide several robustness analyses. Overall, our study offers novel insights into the nature and determinants of jumps in financial markets and highlights the importance of incorporating information heterogeneity in modeling jump dynamics.

A. *Jump detection*

To implement the generalized method of moments (GMM) proposed by Aït-Sahalia et al. (2015), it is necessary to first detect jumps. In this study, we use a well-established approach developed by Lee and Mykland (2008) and implemented in Jeon et al. (2021) to detect jumps in stock prices. Specifically, we consider a daily return to be a jump if its absolute value exceeds a certain threshold defined as the instantaneous volatility multiplied by a confidence level derived from a normal distribution with mean zero and standard deviation $\frac{1}{0.7979}$.

To ensure the robustness of our results, we employ two confidence levels, referred to as LM99 and LM95, which correspond to thresholds of 3.22 and 2.45 times the instantaneous volatility, respectively. The instantaneous volatility is estimated using the realized bipower variation.[10] By employing this rigorous approach, we aim to provide accurate and reliable statistics of realized jumps and the estimated intensities of various types of jumps.

Table IA.2 presents summary statistics for the number of detected jumps and signed jumps using LM99 and LM95. The average number of jumps across all firms is 123 in LM99, which implies an average jump intensity of 11.18 jumps per year (123 divided by 11 years). As expected, the jump intensity increases when a less strict criterion is used to detect jumps, with the average jump intensity in LM95 being 23.27 jumps per year.

We find that the jump intensity varies inversely with firm size, with small firms having a higher jump intensity compared to medium and large firms. Specifically, large, medium, and small firms have an average of 10.72 (22.81), 10.90 (23.00), and 11.90 (24.09) jumps per year in LM99 (LM95), respectively.

In terms of signed jumps, we observe that negative jumps are more prevalent than positive jumps overall. Across all firms, there are on average 5.91 (11.91) negative jumps per year and 5.27 (11.45) positive jumps per year in LM99 (LM95). Furthermore, the difference between the number of negative jumps and the number of positive jumps is larger for larger firms. For example, in LM99, large, medium, and small firms have on average 12, 10, and 3 more negative jumps than positive jumps, respectively. These findings are crucial to understanding the dynamics of jumps and signed jumps in different firms, which is an essential component of our subsequent empirical analyses using the GMM

[10] For further details on the jump detection methodology, refer to Definition 1 and Theorem 1 in Lee and Mykland (2008).

method proposed by Aït-Sahalia et al. (2015).

Table IA.3 and IA.4 present summary statistics of realized jump sizes for firms of different sizes, as measured by the mean and variance. Our analysis, using LM99 and LM95, reveals that the jump-size mean is negative for all, large, and medium firms, but positive for small firms on average. Additionally, both the jump-size mean and variance decrease as firm size increases. For instance, in LM99, the mean jump size is -5.95E-03, -4.86E-03, and 9.79E-04 for large, medium, and small firms, respectively. The corresponding variance is 3.30E-03, 5.00E-03, and 6.63E-03 for large, medium, and small firms, respectively. These findings suggest that smaller firms have a jump-size distribution with a more positive mean and higher variance. Furthermore, using a less stringent criteria to detect jumps (i.e., LM95) increases the jump-size mean and decreases the jump-size variance. Our results provide valuable insights into the behavior of jump sizes across firms of different sizes, which can have important implications for risk management and portfolio allocation strategies.

B. Estimation results

In this section, to provide some insights into the estimation results, we present the estimation details for three pairs and illustrate some estimation results that share similar patterns across all firms. We use the proposed GMM in Aït-Sahalia et al. (2015) to estimate parameters in equation (1). Specifically, the firms we consider are AEO (American Eagle Outfitters Inc), CRK (Comstock Resources Inc), and WSBF (Waterstone Financial Inc), which are classified as large, medium, and small firms, respectively, based on their average daily market capitalization during the sample period (June 01, 2010 - August 31, 2021).

Table 3 reports the GMM estimates for the parameters of the bivariate HJD model for the three pairs (SPY, AEO), (SPY, CRK), and (SPY, WSBF). The estimation results highlight several key aspects. Firstly, the estimated values for $\mathbb{E}[\lambda_{2,t}]$ suggest that the expected jump intensities are different across the three firms, with WSBF having the highest expected jump intensity, followed by CRK and AEO. Additionally, there are significant estimated values for $\lambda_{2,\infty}$, indicating the presence of memoryless jumps. The overall percentage of memoryless jumps across all jumps is highest for AEO, suggesting a more significant exogenous component in AEO than in the other two stocks.

Insert Table 3 about here

Secondly, all the estimated values of β_{11} and β_{22} are statistically significant, indicating the presence of self-excitation in SPY and the three stocks. Furthermore, estimates suggest a more pronounced self-excitation in CRK and WSBF than in AEO. The contribution of the self-exciting component to the jump intensity is highest for WSBF, followed by CRK and AEO.

Thirdly, all estimated values of β_{21} are statistically significant, indicating the presence of market-excitation in the three stocks. The contributions of the market-exciting component to the jump intensity are highest for AEO, followed by WSBF and CRK. Moreover, estimates of β_{21} provide strong statistical evidence for shock transmission from SPY to the three stocks. However, estimates of β_{12} are statistically insignificant, suggesting the absence of significant reverse transmission. This asymmetric shock transmission is not surprising because it is generally harder for a stock to have an impact on the movement of the market index (SPY tracks the S&P 500 index), especially if the stock does not have a large market capitalization.

Lastly, all the estimated values of α_1 and α_2 are statistically significant, ensuring that jump intensities decay to their steady-state levels after each increase. Furthermore, all the p-values of the J-test statistic suggest that the bivariate HJD model fits the data well for all three pairs.[11]

C. *Estimated jump intensities: An empirical inspection*

Figure 5 presents the estimated time series of jump intensity for the BA symbol using the LM99 method. The first row displays the estimated intensity of all jumps, $\lambda_{BA,t}$. Each spike in $\lambda_{BA,t}$ is caused by an earlier jump in BA and/or an earlier jump in the market. Notably, the most significant spikes tend to occur during periods of high investor disagreement. For example, the earliest and most noticeable spike occurs between August

[11] For further details, please refer to Technical Appendix A, C, and Aït-Sahalia et al. (2015).

4, 2011 and August 12, 2011, when the average investor disagreement was approximately 0.90. The most recent and noticeable spike occurs between March 11, 2020 and March 18, 2020, when the average investor disagreement was about 0.97. On August 11, 2011, $\lambda_{BA,t}$ peaked at around 137 jumps per year, potentially due to the United States Federal Reserve's announcement to keep interest rates near zero two days prior. Similarly, on March 17, 2020, $\lambda_{BA,t}$ peaked at around 105 jumps per year, possibly due to travel restrictions during the COVID-19 pandemic.

Insert Figure 5 about here

The second and third rows display the estimated intensity of self-exciting and market-exciting jumps, $\lambda_{BA,t}^{SE}$ and $\lambda_{BA,t}^{ME}$, respectively. Due to the jump intensity decomposition (i.e., equations (5) and (6)), $\lambda_{BA,t}^{SE}$ exhibits large spikes, while $\lambda_{BA,t}^{ME}$ has smaller spikes. This suggests that self-exciting jumps are the primary component of BA jumps. The fourth row illustrates an apparent excitation mechanism by visualizing different jump intensity dynamics in a shorter time window. In mid-June, both $\lambda_{BA,t}^{SE}$ and $\lambda_{BA,t}^{ME}$ increase in response to a BA jump and an SPY jump from the previous day, respectively, causing $\lambda_{BA,t}$ to peak. Subsequently, $\lambda_{BA,t}^{ME}$ gradually decays to 0 and then rises again in early July. In contrast, the strong self-excitation exhibited by BA causes $\lambda BA, t$ and $\lambda_{BA,t}^{SE}$ to spike four more times within two weeks before returning to their steady-state levels, i.e., λ_{BA}^{EXO} and 0, respectively.

The comparison of jump intensity between different firms is challenging due to the time-varying nature of the jump intensity process. To address this issue, we introduce a measure called the average estimated jump intensity (AEJI), which is computed as the mean of the estimated jump intensity across all trading days during the sample period. Specifically, the AEJI of all jumps for firm i is defined as $\frac{1}{T}\sum_{t=1}^{T}\lambda_{i,t}$, where T is the number of trading days in the sample period. For jumps of a specific type, we use superscripts to denote the average estimated intensity. For example, $AEJI_i^{SE}$ refers to the average estimated intensity of self-exciting jumps for firm i, which is computed as $\dfrac{1}{T}\sum_{t=1}^{T}\lambda_{i,t}^{SE}$. For brevity, we use the term "AEJI" to refer to the average estimated jump intensity throughout the rest of the paper.

Table 4 presents various statistics of the AEJI for different types of jumps across all firms and firms in three size groups, using LM99 and LM95. Panel A reports the statistics for all jumps. For LM99, the AEJI of all jumps ranges from 1.84 to 23.76 jumps per year across all firms, with around 50% of firms having more than 8.65 jumps per year. The mean AEJI for all jumps is 8.90 jumps per year for LM99, which more than doubles to 19.34 jumps per year for LM95. Notably, the AEJI of all jumps is inversely related to firm size, with averages of 8.11, 8.67, and 9.91 jumps per year for large, medium, and small firms in LM99, respectively.

Insert Table 4 about here

Columns 3-9 in Panel B and C report statistics for the AEJI of signed jumps using LM99. For all firms, the AEJI of negative jumps is larger than that of positive jumps, with averages of 5.65 and 5.07 jumps per year, respectively. This suggests that negative jumps are more frequent than positive jumps in stock returns. The AEJI of negative and positive jumps also varies inversely with firm size, with averages of 5.10 (4.30), 5.59 (4.91), and 6.25 (6.00) negative (positive) jumps per year for large, medium, and small firms, respectively. This indicates that smaller firms tend to have more negative and positive jumps, and therefore more jumps overall, compared to larger firms.

The results presented in Panel A, B, and C are consistent with those in Table IA.2, indicating that jump intensity tends to decrease with firm size, and negative jumps are more frequent than positive jumps.

Panel D, E, and F of Table 4 provide statistics on the AEJI of self-exciting jumps, memoryless jumps, and market-exciting jumps when using LM99. Specifically, for all firms, the AEJI of self-exciting jumps ranges from 0.27 to 14.47 jumps per year, with about 50% of firms having more than 5.94 self-exciting jumps per year. These findings suggest a prominent self-excitation phenomenon at the firm level, which complements prior research by Aït-Sahalia et al. (2015) that demonstrates the self-excitation phenomenon at the market level.

In comparison to self-exciting jumps, memoryless jumps have a wider range of the AEJI (from 0 to 21.53 jumps per year), while market-exciting jumps have a smaller range of the AEJI (from 0.78 to 8.60 jumps per year). The minimum AEJI of market-exciting jumps and memoryless jumps suggest the presence of market-exciting jumps in stock returns of all firms, but the absence of memoryless jumps in stock returns of some firms. These summary statistics also suggest that self-exciting jumps are more frequent than memoryless jumps and market-exciting jumps in stock returns. For instance, the mean AEJI of self-exciting jumps is 5.95, which is approximately three times that of memoryless jumps (1.82) and five times that of market-exciting jumps (1.13).

Examining the estimated jump intensities across the three size groups, we find that the AEJI of self-exciting jumps and memoryless jumps vary inversely across the size groups. Specifically, large, medium, and small firms have an average of 5.53 (1.29), 5.70 (1.83) and 6.60 (2.35) self-exciting (memoryless) jumps per year. In contrast, the AEJI of market-exciting jumps vary proportionally across the size groups. On average, there are 1.28, 1.15, and 0.96 market-exciting jumps per year for large, medium, and small firms. These results suggest that the stock prices of smaller firms are more vulnerable to exogenous shocks and their own performances than larger firms, while the stock prices of larger firms are more exposed to market performance than smaller firms.

Furthermore, Table IA.5 supplements the findings in Panel D, E, and F of Table 4 and suggests that, in terms of jump intensity: (i) the self-exciting component, exogenous component, and market-exciting component contribute the most, the moderate and the least; (ii) the exogenous component is absent for some firms since its minimum contribution is zero, and; (iii) the self-exciting (market-exciting) component is more important in smaller (larger) firms than larger (smaller) firms. These results provide important insights into the nature of jumps in stock prices and their implications for firms of different sizes.

Figure 6 presents the empirical distributions of the AEJI for different types of jumps across all firms. Upon visual inspection of panels d-f, several noteworthy observations can be made. First, the results support the dominance of self-exciting jumps in jumps. Specifically, the AEJI of self-exciting jumps is most concentrated between 2.5 to 8.5 jumps per year, while the AEJI of memoryless jumps and market-exciting jumps are most concentrated in the range of less than 2.5 jumps per year. This suggests that the self-exciting phenomenon plays a salient role in explaining jumps in stock prices, consistent with previous studies.

Insert Figure 6 about here

Second, the empirical distributions of the three types of jumps exhibit distinct characteristics. For instance, the empirical distributions of memoryless and market-exciting jumps are skewed to the right, while that of self-exciting jumps is symmetric. This implies that the nature of these jumps and their underlying mechanisms may differ from one another.

Overall, the findings suggest that self-exciting jumps are a dominant feature in the empirical distribution of jumps in stock prices, with distinct characteristics from memoryless and market-exciting jumps. These insights may have important implications for understanding the dynamics of jumps in financial markets.

D. The role of disagreement and news in jumps

In this section, we investigate the influence of disagreement and news on the likelihood of jumps in financial markets. To this end, we estimate the following regression model

$$\begin{aligned}\lambda_{i,t} = c_0 + \gamma_i + c_1 NumberOfMessages_{i,t} + c_2 Disagreement_{i,t} \\ + c_3 NumberOfNews_{i,t} + c_4 |NewsTone_{i,t}| + \eta Controls_{i,t} + \epsilon_{i,t}\end{aligned} \tag{9}$$

where $\lambda_{i,t}$ denotes the estimated intensity of all jumps for firm i on day t. The variables $NumberOfMessages_{i,t}$ and $NumberOfNews_{i,t}$ capture firm-date-specific investor attention and media attention, respectively. $Disagreement_{i,t}$ measures the overall disagreement among investors who specify their investment philosophy as one of Fundamental, Technical, Momentum, Growth, and Value for firm i on day t. We also include the absolute value of news tone, denoted by $|NewsTone_{i,t}|$, and firm fixed effects γ_i in the regression. Furthermore, we control for firm size, 12-month momentum, and idiosyncratic volatility through the vector of variables $Controls_{i,t}$.

To examine the role of different types of jumps, we estimate separate regression equations for self-exciting and market-exciting jumps, denoted by $\lambda_{i,t}^{SE}$ and $\lambda_{i,t}^{ME}$, respectively. Additionally, we estimate separate regression equations for negative and positive jumps, denoted by $\lambda_{i,t}^{-}$ and $\lambda_{i,t}^{+}$, respectively. To do so, we replace $\lambda_{i,t}$ in (9) with the corresponding intensity variable and use $NewsTone_{i,t}$ rather than $|NewsTone_{i,t}|$ as an independent variable.

D.1. *All, self- and market-exciting jumps*

Table 5 presents the regression results for the estimated intensity of different types of jumps for all firms using the LM99 and LM95 methods in Panel A and Panel B, respectively. The All column displays the regression result for the estimated intensity of all jumps.

Insert Table 5 about here

We find that the coefficient for overall disagreement (c_2) is significantly and positively related to the estimated intensity of all jumps, with a value of 0.648 in Panel A and 0.853 in Panel B. This suggests that a one-standard-deviation increase in overall disagreement is associated with an average increase of 64.8% and 85.3% in the estimated intensity of all jumps when detecting jumps using LM99 and LM95, respectively. The relationship between overall disagreement and the estimated intensity of all jumps is not only statistically significant, but also economically strong.

We address the concern that the link between overall disagreement and the estimated intensity of all jumps may be driven by media attention and investor attention by including the number of messages and the number of news in the regression to control for investor attention and media attention, respectively. We find that the relationship between overall disagreement and the estimated intensity of all jumps remains highly significant after including these control variables.[12]

Furthermore, we find that the absolute value of news tone is also significantly and positively related to the estimated intensity of all jumps. However, its economic significance is weaker than overall disagreement, with a one-standard-deviation increase in the

[12]To save space, we have not explored various sub-classes of model (9), i.e., considering a smaller number of covariates.

absolute value of news tone increasing the estimated intensity of all jumps by an average of 18.3% in Panel A and 23.5% in Panel B. This is less than one-third of the comparable increases in overall disagreement.

We also find a significant and positive relationship between the estimated intensity of all jumps and the number of messages, indicating that stock jumps are more likely when investors post more messages to express their bearish or bullish views about the stock. However, the estimated intensity of all jumps is not statistically significant related to the number of news. This could be due to the fact that including the number of messages in the regression reduces the significance of the number of news, since both variables are proxies for attention of firms.

Our results are consistent with Jeon et al. (2021), who find that the probability of realized jumps is significantly and positively associated with the absolute value of news tone. Overall, our findings suggest that both overall disagreement and news tone are important drivers of stock jumps, and that investor attention and media attention also play a role in this relationship.

In Column SE of Table 5, we present the regression results for the estimated intensity of self-exciting jumps. As self-exciting jumps are the most important component in stock jumps, the findings for the estimated intensity of self-exciting jumps share several similarities to those for all jumps.[13]

Firstly, the estimated intensity of self-exciting jumps is significantly and positively associated with both overall disagreement and the absolute value of news tone. Secondly, overall disagreement has a stronger influence on the estimated intensity of self-exciting jumps than the absolute value of news tone. The coefficient for overall disagreement, c_2, is approximately *3.5 times* the coefficient for the absolute value of news tone, c_4, indicating that the estimated intensity of self-exciting jumps is more sensitive to overall disagreement than to news tone. Furthermore, overall disagreement has the largest absolute value of t-statistic, suggesting that it is the most statistically significant characteristic for the estimated intensity of self-exciting jumps among all covariates. Thirdly, the estimated intensity of self-exciting jumps is significantly and positively related to the number

[13] (See Section III.C).

of messages, but not significantly related to the number of news. This suggests that self-exciting jumps are more likely to be triggered by their own past jumps when there is a higher level of disagreement among investors or when the news tone is more extreme. Overall, our findings suggest that stock jumps are more likely to occur and be self-exciting when there is greater disagreement among investors and when the news tone is more extreme.

Column ME presents the regression results for the estimated intensity of market-exciting jumps. The findings suggest that overall disagreement and the absolute value of news tone are significantly and positively associated with the estimated intensity of market-exciting jumps, although to a lesser extent than with self-exciting jumps. Specifically, in Panel A, a one-standard-deviation increase in overall disagreement is associated with a 4.7% average increase in the estimated intensity of market-exciting jumps. By comparison, the corresponding increase for self-exciting jumps was much larger, at 60.2%.

The study also finds that the number of messages and news articles are significantly associated with the estimated intensity of market-exciting jumps, but in opposite directions. Specifically, a one-standard-deviation increase in the number of messages is associated with a 1.6% average increase in the estimated intensity of market-exciting jumps, while a one-standard-deviation increase in the number of news articles is associated with a 1.8% average decrease in the estimated intensity of market-exciting jumps. These results suggest that stock jumps are more likely to be excited by past market jumps when there are more investor-posted messages, but less likely when there are more news articles available.

These findings may imply that investor-posted messages and news articles have different impacts on jump risk. Investor-posted messages tend to amplify a firm's exposure to systematic jump risk, while news articles tend to reduce that exposure. This is consistent with the model proposed by Bégin, Dorion, and Gauthier (2020), which incorporates a constant exposure coefficient and the product of the jump component in the market return process into the stock return process to capture a stock's exposure to systematic jump risk. The results suggest that when the exposure coefficient is large, the probability of market-excited jumps is higher, which may contribute to the observed relationship between the number of investor-posted messages and market-exciting jumps.

D.2. Negative and positive jumps

Regression analysis (9) considers the estimated intensity of negative and positive jumps as dependent variables, as reported in the Neg and Pos columns of Table 5, respectively. The results show that both jump intensities are significantly and positively related to the overall disagreement, but with an asymmetric relationship. Specifically, a one-standard-deviation increase in overall disagreement leads to an average increase of 48.2% in the estimated intensity of negative jumps, which is about three times the average increase, i.e., 17.0%, in the estimated intensity of positive jumps, as shown in Panel A. This asymmetry is consistent with the model proposed by Hong and Stein (2003), which suggests that increased differences in investor opinions can result in greater negative skewness in stock returns in both time series and cross-sections. The finding for negative jumps is also consistent with recent research by Chang et al. (2022), who demonstrate that reducing disagreement can lower the incidence of extreme negative returns.

Furthermore, the analysis indicates that news tone is significantly associated with the estimated intensity of negative and positive jumps, but with opposite effects. As reported in Panel A, a one-standard-deviation increase in news tone leads to an average decrease of 11.1% in the estimated intensity of negative jumps, while the estimated intensity of positive jumps rises by an average of 9.0%. This is because negative jumps are more likely when the news media becomes more pessimistic, whereas positive jumps are more likely when the news media becomes more optimistic. This relationship also explains the positive association between the absolute value of news tone and the likelihood of all jumps. As noted by Jeon et al. (2021), this relationship also holds true for the probability of realized jumps, regardless of the sign of the jumps.

D.3. Jump-size distribution

We analyze the distribution of realized jump sizes in financial markets, focusing on the role of disagreement and news tone. We use a regression framework to investigate the relation between jump sizes and various covariates. Specifically, we estimate the following equations

$$r_{i,t}|Jump = c_0 + \gamma_i + c_1 NumberOfMessages_{i,t} + c_2 Disagreement_{i,t} + c_3 NumberOfNews_{i,t}$$

$$+ c_4 NewsTone_{i,t} + \eta Controls_{i,t} + \epsilon_{i,t} \tag{10}$$

$$r^2_{i,t}|Jump = c_0 + \gamma_i + c_1 NumberOfMessages_{i,t} + c_2 Disagreement_{i,t} + c_3 NumberOfNews_{i,t}$$

$$+ c_4 |NewsTone_{i,t}| + \eta Controls_{i,t} + \epsilon_{i,t} \tag{11}$$

where $r_{i,t}|Jump$ and $r^2_{i,t}|Jump$ are the return and the squared return for firm i on jump day t. Equations (10) and (11) imply the following relation for the first two moments of jump size.

$$\begin{aligned} \mathbb{E}[r_{i,t}|Jump = 1] &= c_0 + \mathbb{E}[\gamma_i] + c_1 \mathbb{E}[NumberOfMessages_{i,t}] + c_2 \mathbb{E}[Disagreement_{i,t}] \\ &\quad + c_3 \mathbb{E}[NumberOfNews_{i,t}] + c_4 \mathbb{E}[NewsTone_{i,t}] + \eta \mathbb{E}[Controls_{i,t}] \\ \mathbb{E}[r^2_{i,t}|Jump = 1] &= c_0 + \mathbb{E}[\gamma_i] + c_1 \mathbb{E}[NumberOfMessages_{i,t}] + c_2 \mathbb{E}[Disagreement_{i,t}] \\ &\quad + c_3 \mathbb{E}[NumberOfNews_{i,t}] + c_4 \mathbb{E}[|NewsTone_{i,t}|] + \eta \mathbb{E}[Controls_{i,t}] \end{aligned}$$

The results are reported in Table IA.6. We find that the jump-size mean is significantly and positively related to news tone but negatively related to overall disagreement, indicating that returns become more negative on jump days when news tone is more bearish or overall disagreement is higher. This negative association between overall disagreement and the jump-size mean is consistent with previous research, such as Diether et al. (2002), who found that stocks with greater analyst disagreement about future earnings tend to have lower returns on average. Moreover, the jump-size mean is significantly and positively related to the number of messages but negatively related to the number of news, suggesting that returns become more positive on jump days when investors post more messages but more negative on jump days when more news articles are published.

Regarding the jump-size volatility, we find that it is significantly and positively related to overall disagreement, the absolute value of news tone, and the number of news, but is more sensitive to overall disagreement. For all firms in LM99, a one-standard-deviation increase in overall disagreement, the absolute value of news tone, and the number of news increases jump-size volatility by an average of 0.19

Taken together, our results suggest that both disagreement and news tone play a significant role in the distribution of realized jump sizes. Specifically, our findings highlight the importance of considering both the sentiment conveyed in news articles and the degree of disagreement among market participants when analyzing the occurrence and characteristics of market jumps.

E. Disagreement sources and jumps

In this section, we provide a detailed analysis of the relationship between stock jumps and two sources of disagreement among investors, namely information differences and differences in information interpretation (Hong and Stein, 2007). We first decompose the overall disagreement into two components: disagreement within-group (DWG), which arises from differences in information, and disagreement cross-group (DCG), which arises from differences in interpretation (Cookson and Niessner, 2020). These measures are given by

$$\mathrm{DWG}_{i,t} = \frac{\sum_{g \in A} n_{i,t,g} \mathrm{Disagreement}_{i,t,g}}{\sum_{g \in A} n_{i,t,g}} \tag{12}$$

$$\mathrm{DCG}_{i,t} = \sqrt{\frac{\sum_{g \in A} n_{i,t,g} (\mathrm{AvgSentiment}_{i,t,g} - \mathrm{AvgSentiment}_{i,t})^2}{\frac{N_A - 1}{N_A} \sum_{g \in A} n_{i,t,g}}} \tag{13}$$

where A denotes a set of investment philosophies including Fundamental, Technical, Momentum, Value, and Growth. N_A is the number of investment philosophies in A, and $n_{i,t,g}$ is the number of investors for firm i with investment philosophy g on day t. $\mathrm{AvgSentiment}_{i,t}$ represents the firm-date-specific average sentiment among investors who specify their investment philosophy as any one in A.

To further examine the role of DWG and DCG in the likelihood of stock jumps, we conduct a regression analysis. Specifically, we estimate the intensity of all jumps for firm

i on day t, denoted by $\lambda_{i,t}$, using the following regression equation

$$\lambda_{i,t} = c_0 + \gamma_i + c_1 NumberOfMessages_{i,t} + c_{2,w} DWG_{i,t} + c_{2,c} DCG_{i,t} \\ + c_3 NumberOfNews_{i,t} + c_4 |NewsTone_{i,t}| + \eta Controls_{i,t} + \epsilon_{i,t} \tag{14}$$

where $NumberOfMessages_{i,t}$ represents the number of posted messages for firm i on day t by investors who specify their investment philosophy as one of Fundamental, Technical, Momentum, Growth, and Value. $DWG_{i,t}$ and $DCG_{i,t}$ denote disagreement within-group and disagreement cross-group among those investors for firm i on day t, constructed using equations (12) and (13), respectively. The coefficients c_1 to c_4 capture the effects of other variables, including the number of news articles and the sentiment of those articles, as well as other controls represented by $Controls_{i,t}$, i.e., firm size, momentum, and volatility. We standardize all independent variables to compare their explanatory power. We extend our analysis to consider the estimated intensity of self-exciting (market-exciting) jumps and negative (positive) jumps, by replacing $\lambda_{i,t}$ with $\lambda^{SE}_{i,t}$ and $\lambda^{-}_{i,t}$ ($\lambda^{ME}_{i,t}$ and $\lambda^{+}_{i,t}$), respectively. In the case of negative (positive) jumps, we use $NewsTone_{i,t}$ as an independent variable instead of $|NewsTone_{i,t}|$. This allows us to investigate how disagreement sources impact the likelihood of specific types of jumps.

In particular, the estimated intensity of self-exciting (market-exciting) jumps is obtained by replacing $\lambda_{i,t}$ with $\lambda^{SE}_{i,t}$ ($\lambda^{ME}_{i,t}$). We use the same regression equation as in equation (14), but replace the dependent variable with $\lambda^{SE}_{i,t}$ ($\lambda^{ME}_{i,t}$), which represents the estimated intensity of self-exciting (market-exciting) jumps.

Similarly, we investigate the estimated intensity of negative (positive) jumps by replacing $\lambda_{i,t}$ with $\lambda^{-}_{i,t}$ ($\lambda^{+}_{i,t}$). We also use the same regression equation as in Equation (14), but replace the dependent variable with $\lambda^{-}_{i,t}$ ($\lambda^{+}_{i,t}$) to represent the estimated intensity of negative (positive) jumps. In addition, we use $NewsTone_{i,t}$ as an independent variable instead of $|NewsTone_{i,t}|$ to further explore the impact of disagreement sources on the likelihood of negative (positive) jumps.

Table 6 reports the estimated intensity of various types of jumps for all firms, obtained through regression (14). In Panel A and Panel B, we provide the results for LM99 and

LM95, respectively.

Insert Table 6 about here

Our results reveal that disagreement among investors is positively related to the intensity of all types of stock market jumps. Specifically, our findings suggest that both disagreement within-group (i.e., differences in information) and disagreement cross-group (i.e., differences in the interpretation of information) are significant determinants of stock market jumps. However, the impact of disagreement within-group is notably stronger than that of disagreement cross-group. For instance, a one-standard-deviation increase in within-group disagreement is associated with a significant average increase of 52.0% in the estimated intensity of all jumps, which is about *two* times greater than the comparable increase of 25.1% in cross-group disagreement.

Moreover, we find that information differences are economically more important than interpretation differences in the occurrence of stock market jumps. This result is consistent with prior research, which finds that both sources of disagreement have a significantly positive effect on trading volume, and that the within-group effect is several times greater than the cross-group effect (Cookson and Niessner, 2020).

Additionally, our analysis reveals that the relationship between interpretation differences and stock market jumps varies across different types of jumps. Specifically, we find that interpretation differences are negatively related to the intensity of market-exciting jumps, while positively related to other types of jumps. This suggests that market-exciting jumps are less likely to occur when investors disagree more due to differences in the interpretation of information, while other types of jumps are more likely to occur.

In addition to examining the likelihood of jumps, we also investigate the relationship between the distribution of realized jump sizes and disagreement within-group and disagreement cross-group. To this end, we estimate regression equations for the jump-size mean and jump-size volatility by replacing the jump intensity measure $\lambda_{i,t}$ and absolute NewsTone measure $|NewsTone|_{i,t}$ in Equation (14) with $r_{i,t}|Jump$ and $NewsTone_{i,t}$, respectively.

Table IA.7 presents the results for all firms as well as for firms divided into three size groups. Consistent with our findings regarding overall disagreement, we find that both disagreement within-group and disagreement cross-group are significantly positively related to jump-size mean and negatively related to jump-size volatility.

Our results suggest that, regardless of firm size, the link between disagreement within-group and the distribution of jump sizes is much stronger than the link between disagreement cross-group and the distribution of jump sizes. Thus, we conclude that differences in information are more crucial in explaining the occurrence of stock jumps, both in terms of their intensity and size, than differences in the interpretation of information.

F. *Information diffusion and jumps*

In this section, we present additional empirical analysis to investigate the relationship between gradual information diffusion and stock price jumps, building on the findings from Section III.E that disagreement within investor groups is a crucial determinant of stock jumps. Specifically, we distinguish between sophisticated (S) and unsophisticated (U) investors based on their self-reported experience levels as professional (S), novice (U), or intermediate (U). We measure overall disagreement among S-type and U-type investors, as well as disagreement between S-type and U-type investors with the same investment philosophy, using the following equations

$$\text{Disagreement (S)}_{i,t} = \frac{\sum_{g \in A} m_{i,t,g} \text{Disagreement (S)}_{i,t,g}}{\sum_{g \in A} m_{i,t,g}}$$

$$\text{Disagreement (U)}_{i,t} = \frac{\sum_{g \in A} m_{i,t,g} \text{Disagreement (U)}_{i,t,g}}{\sum_{g \in A} m_{i,t,g}}$$

$$|\text{Disagreement (S-U)}|_{i,t} = \frac{\sum_{g \in A} m_{i,t,g} |\text{AvgSentiment (S)}_{i,t,g} - \text{AvgSentiment (U)}_{i,t,g}|}{\sum_{g \in A} m_{i,t,g}}$$

where A={Fundamental, Technical, Momentum, Value, Growth}, is a set of investment philosophies. $Disagreement(S)_{i,t,g}$ and $AvgSentiment(S)_{i,t,g}$ are the disagreement and the average sentiment among S-type investors for firm i of investment philosophy g on day t, respectively. $Disagreement(U)_{i,t,g}$ and $AvgSentiment(U)_{i,t,g}$ are the disagreement and the average sentiment among U-type investors for firm i of investment philosophy g on day t, respectively. $|Disagreemen(S-U)|$ is the weighted average of the absolute value of the difference between S-type investors' average sentiment and U-type investors' average sentiment for each investment philosophy in A. $m_{i,t,g}$ is the number of messages for firm i of investment philosophy g on day t. We then run the following regression

$$\begin{aligned}\lambda_{i,t} = c_0 + \gamma_i + c_{1,s} NumberOfMessages(S)_{i,t} + c_{1,u} NumberOfMessages(U)_{i,t} \\ + c_{2,s} Disagreement(S)_{i,t} + c_{2,u} Disagreement(U)_{i,t} + c_{2,s-u} |Disagreement(S-U)|_{i,t} \\ + c_3 NumberOfNews_{i,t} + c_4 |NewsTone_{i,t}| + \eta Controls_{i,t} + \epsilon_{i,t} \end{aligned} \tag{15}$$

where $\lambda_{i,t}$ is the estimated intensity of all jumps for firm i on day t. $NumberOfMessages(S)_{i,t}$ and $NumberOfMessages(U)_{i,t}$ are the number of posted messages by S-type and U-type investors, respectively. The regression includes firm fixed effects (γ_i), controls for the number of news articles about the firm ($NumberOfNews_{i,t}$) and the sentiment of those news articles ($|NewsTone_{i,t}|$), and additional controls for firm-specific characteristics such as market capitalization (size), 12-month momentum, and idiosyncratic volatility ($Controls_{i,t}$).

Table 7 presents the results of regression (15) for all firms. Panel A, Column "All" shows that an increase in $|Disagreement(S-U)|_{i,t}$ of one standard deviation is associated with an average increase of 12.3% in the estimated intensity of all jumps. This finding suggests that disagreement between sophisticated and unsophisticated investors who share the same investment philosophy is significantly and positively associated with the estimated intensity of all jumps.

Insert Table 7 about here

The gradual information diffusion hypothesis posits that sophisticated investors have an informational advantage, which enables them to acquire information earlier than unsophisticated investors, and trade on this advantage (Hong and Stein, 1999, 2007). This leads to information differences, even though investors have the same investment philosophy, resulting in disagreement between sophisticated and unsophisticated investors. Therefore, the significant and positive relationship between the estimated intensity of all jumps and $|Disagreement(S-U)|_{i,t}$ implies that the gradual information diffusion hypothesis plays an essential role in triggering stock jumps.

Our findings are in line with previous research by Hong and Stein (2007), which suggests that gradual diffusion of information can cause prices to gradually respond to that information. Moreover, our results complement the findings of Cookson and Niessner (2020), who have highlighted the crucial role of gradual information diffusion in trading volume.

In Panel A, the coefficient on $|Disagreement(S-U)|_{i,t}$ in column SE is significantly positive, indicating that disagreement between sophisticated and unsophisticated investors with the same investment philosophy is positively related to the estimated intensity of self-exciting jumps. This finding suggests that gradual information diffusion may be partly responsible for the existence of the self-excitation phenomenon in stock jumps.

Furthermore, the coefficients in columns Neg and Pos reveal that an increase in |Disagreement$(S-U)|_{i,t}$ is associated with, on average, a 24.0% increase in the estimated intensity of negative jumps and a 21.1% decrease in the estimated intensity of positive jumps. These results suggest that disagreement between sophisticated and unsophisticated investors with the same investment philosophy plays *opposite roles* in triggering stock jumps with different signs. Specifically, as disagreement increases, we expect more negative jumps but fewer positive jumps in the future. This finding provides insight into the complex interplay between market participants and sheds light on the mechanisms underlying stock price movements.

We extended our analysis by regressing the distribution of the realized jump sizes on disagreement between sophisticated and unsophisticated investors with the same investment philosophy. To obtain the regression equation for jump-size mean and volatility, we replaced $\lambda_{i,t}$ in equation (15) with $r_{i,t}|Jump$ and $r^2_{i,t}|Jump$, respectively. When conducting the regression on jump-size volatility, we also replaced $|NewsTone|i,t$ in equation (15) with $NewsTonei,t$.

Our findings are reported in Table IA.8, which shows a significant relationship between $|Disagreement(S-U)|i,t$ and the first two jump-size moments. Specifically, $|Disagreement(S-U)|i,t$ is negatively related to jump-size mean and positively related to jump-size volatility. This indicates that when there is higher disagreement between sophisticated and unsophisticated investors with the same investment philosophy, jumps tend to be more negative and more extreme on average.

In conclusion, since gradual information diffusion can lead to disagreement between sophisticated and unsophisticated investors even when they share the same investment philosophy, our results imply that i) gradual information diffusion plays an important role in the self-excitation phenomenon in stock jumps and triggering negative stock jumps, and ii) gradual information diffusion has an adverse effect on jump size. Therefore, in markets where information diffuses slowly, jumps tend to be more frequent, negative, and extreme than in markets where information is fully transparent. These results contribute to the understanding of the role of gradual information diffusion in stock price dynamics and have important implications for market participants and policymakers. For example, our findings suggest that policymakers should focus on improving the transparency of information in financial markets to reduce the frequency and severity of negative stock price jumps.

G. *Robustness analysis*

G.1. *Disagreement from all tweets*

In this section, we investigate the impact of including all available messages, irrespective of users' specified investment philosophy, on our analysis of the association between investor disagreement and stock price jumps. We conduct our analysis on a vast dataset comprising 494,827 users and 34,373,324 messages, obtained via successive filtering actions. We construct novel variables for the number of messages and overall disagreement and employ regression analysis on these variables to estimate jump intensities using equation (9). The results of our analysis are presented in the Appendix, specifically in Table IA.10.

Overall, our findings are consistent with our previous analysis (see Table 5). We observe that the number of messages and overall disagreement are significantly and positively associated with different estimated jump intensities. In particular, the relationship between overall disagreement and the estimated jump intensity is even stronger for all jumps and self-exciting jumps when using all available messages obtained through filtering. Specifically, we find that the coefficient of overall disagreement increases by approximately 0.05 in column All and SE in Panel A when compared to the results in Table 5. These findings provide further evidence for the importance of investor disagreement in explaining stock price jumps and contribute to the growing literature on asset pricing and risk management.

G.2. Disagreement and firm size

This section aims to investigate the influence of investor disagreement on estimated jump intensities across firms of different sizes. Table IA.9 reports the results of the regression analysis that examines the impact of overall disagreement on estimated jump intensities for firms of different sizes. The firms are divided into three size groups with an equal number of firms based on the average daily market capitalization during the sample period. Specifically, there are 893 firms in each size group.

As shown in Panel A of Table IA.9, we find a significant and positive relationship between overall disagreement and the estimated intensity of all jumps for all three size groups, consistent with the results in column All of Table 5. Moreover, the results suggest that overall disagreement is more influential for the estimated intensity of all jumps in larger firms than in smaller ones. Specifically, in LM99 (LM95), a one-standard-deviation increase in overall disagreement leads to an average increase in the estimated intensity of all jumps of 72.0% (94.2%) for large firms and 52.6% (70.3%) for small firms. This indicates that the sensitivity of the estimated intensity of all jumps to overall disagreement increases with firm size.

The findings have important implications. During a crisis, when investor disagreement tends to be higher than usual, the stock prices of larger firms are more likely to experience significant jumps than those of smaller firms. Additionally, we find that the estimated intensity of all jumps is much more sensitive to the absolute value of news tone in small firms than in large ones. Specifically, in LM99, a one-standard-deviation increase in the absolute value of news tone leads to an average increase in the estimated intensity of all jumps of 25.7% in small firms, which is over six times the corresponding increase of 3.8% in large firms.

In Panel B and C, we observe a significant and positive relationship between overall disagreement and the estimated intensity of self-exciting and market-exciting jumps for all three size groups. We also find that the relationship between overall disagreement and the estimated intensity of self-exciting jumps is strongest for large firms. These results indicate that when disagreement among investors is higher, stock jumps are expected to exhibit a stronger self-excitation and a stronger market-excitation, and the self-excitation phenomenon is most noticeable for firms with large market capitalization.

In Panel D and E, we find that as firm size increases, overall disagreement has a greater impact on the estimated intensity of negative jumps but less on the estimated intensity of positive jumps. This suggests that negative jumps are more likely than positive jumps when there is a higher disagreement among investors, especially for firms with larger market capitalization. For instance, in column Large in LM99, the increase (57.6%) in the estimated intensity of negative jumps is approximately four times the increase (14.4%) in the estimated intensity of positive jumps for each unit increase in overall disagreement. The gap narrows to about 2.2 (=36.7%/16.5%) times for medium firms and to about 1.7 (=33.8%/19.9%) times for small firms.

Moreover, we find that the estimated intensity of negative jumps is more sensitive to news tone for larger firms. As news tone decreases by one standard deviation, the estimated intensity of negative jumps increases by an average of 12.5%, 9.9%, and 8.1% for large, medium and small firms, respectively. This suggests that larger firms are more likely to experience negative jumps than smaller firms when the news media becomes more pessimistic.

IV. Conclusion

In this study, we analyzed a large dataset of over 173 million tweets from StockTwits to investigate the relationship between investor disagreement and stock price jumps, and explore the underlying mechanisms involved. Our study aimed to answer three research questions: (i) does investor disagreement lead to stock price jumps?; (ii) if so, what is the role of different sources of disagreement in causing stock price jumps?; and (iii) what is the impact of information flow diffusion through disagreement on stock price jumps?

We found that higher levels of disagreement among investors are associated with a greater likelihood of jumps in daily stock returns, and this relationship was stronger than the impact of the number of posted messages on jump intensity. Negative jumps are more likely to occur than positive jumps when investor disagreement is higher, and this asymmetry is more pronounced for firms with larger market capitalizations. News tone is positively related to the estimated intensity process for all jumps, but its significance is weaker than investor disagreement. We also found that both within-group and cross-group disagreement are positively related to the estimated intensity of all jumps in daily stock returns, but within-group disagreement has a greater impact. Finally, disagreement between sophisticated and unsophisticated investors with the same investment philosophy plays a significant role in triggering stock jumps, particularly negative jumps, and in determining the size of these jumps.

In terms of future research, one potential direction could be to explore the impact of news sentiment on stock price jumps, particularly in the context of social media and other online forums where news spreads rapidly and investor sentiment can be quickly influenced. Another direction could be to investigate the role of specific types of investors in driving stock price jumps, such as institutional investors or retail investors, and how their differing investment strategies and philosophies may contribute to disagreement and subsequent stock price movements.

Overall, our study provides insight into the complex relationship between investor disagreement, news sentiment, and stock price movements, and opens up new avenues for further research in this area. These findings have significant implications for risk management, informational efficiency, derivative pricing, and the explanation of anomalies in financial markets.

Table 1: Frequencies of User Profile Characteristics

This table presents frequency distributions of users and messages by user profile characteristics, i.e., investment philosophy, holding period, and experience for 2,679 stocks during the sample period (June 01, 2010 - August 31, 2021).

Panel A: Experience	# Users	Users (%)	# Messages	Messages (%)
Intermediate	53,297	10.77	6,051,254	17.60
Novice	33,812	6.83	1,589,282	4.62
Professional	24,204	4.89	7,487,803	21.78
Missing	383,514	77.50	19,244,985	55.99
Total	494,827	100.00	34,373,324	100.00
Panel B: Investment Philosophy	**# Users**	**Users (%)**	**# Messages**	**Messages (%)**
Technical	37,770	7.63	6,500,765	18.91
Growth	22,438	4.53	1,547,805	4.50
Momentum	21,234	4.29	2,066,586	6.01
Fundamental	14,054	2.84	2,419,987	7.04
Value	11,437	2.31	1,899,162	5.53
Global Macro	3,113	0.63	246,424	0.72
Missing	384,781	77.76	19,692,595	57.29
Total	494,827	100.00	34,373,324	100.00
Panel C: Holding Period	**# Users**	**Users (%)**	**# Messages**	**Messages (%)**
Swing Trader	44,392	8.97	5,822,757	16.94
Long Term Investor	23,925	4.84	3,359,901	9.77
Day Trader	22,909	4.63	2,836,031	8.25
Position Trader	20,581	4.16	2,885,558	8.39
Missing	383,020	77.40	19,469,077	56.64
Total	494,827	100.00	34,373,324	100.00

Table 2: Summary Statistics of Firm Characteristics

This table presents summary statistics of the average of daily observations of firm characteristics for 2,679 stocks during the sample period (June 01, 2010 - August 31, 2021). Panel A and B show summary statistics before and after filling missing characteristics values, respectively. *Number of Messages* and *Disagreement* are the number of investor-posted messages and the overall disagreement among investors at the firm-day level in StockTwits. They are calculated from all tweets available after taking successive filtering actions described in IA.1. The construction of *Disagreement* is described in II.A *Number of News* and *News Tone* are the number of news articles and the overall news tone at the firm-day level from RavenPack. *News Tone* is the variable *event_sentiment_score* in RavenPack. *Size*, *12-month Momentum*, and *Idiosyncratic Volatility* are the market capitalization, the past 12-months return and the idiosyncratic volatility from CRSP/Compustat, respectively. Size and 12-month Momentum are at the firm-day level, while Idiosyncratic Volatility is at the firm-month level. *Size* is the variable *me* retrieved from CRSP. *Idiosyncratic Volatility* is computed as the variance of residual from Fama and French (1993) three-factor model based on daily return in the past 3 months. The missing values of characteristics are filled as the following. We assign firms to three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period. On each day, a firms missing characteristic is replaced by the cross-sectional median within the size group to which the firm belongs.

Panel A: Before filling missing values							
	mean	std	min	25%	50%	75%	max
Number of Messages	5	28	0	0	1	3	1,070
Disagreement	0.25	0.13	0.00	0.18	0.24	0.30	0.85
Number of News	8	23	0	1	4	7	766
News Tone	0.11	0.07	−0.46	0.07	0.10	0.14	0.54
Size	7.98E+06	3.14E+07	3.27E+03	2.74E+05	1.02E+06	4.37E+06	8.04E+08
12-month Momentum	0.01	0.00	−0.02	0.00	0.01	0.01	0.03
Idiosyncratic Volatility	4.37E-04	6.67E-04	3.22E-07	1.03E-04	2.19E-04	5.08E-04	8.88E-03
Panel B: After filling missing values							
	mean	std	min	25%	50%	75%	max
Number of Messages	5	28	0	0	1	3	1,070
Disagreement	0.14	0.14	0.00	0.04	0.11	0.20	0.85
Number of News	8	23	0	1	4	7	766
News Tone	0.10	0.04	−0.01	0.08	0.09	0.11	0.32
Size	7.98E+06	3.14E+07	3.27E+03	2.74E+05	1.02E+06	4.37E+06	8.04E+08
12-month Momentum	0.01	0.00	−0.02	0.00	0.01	0.01	0.03
Idiosyncratic Volatility	4.33E-04	6.58E-04	3.22E-07	1.05E-04	2.19E-04	5.03E-04	8.88E-03

Table 3: This table reports the GMM estimates for the parameters of the bivariate Hawkes jump-diffusion (HJD) model for three stocks in pairs with the SPY. Panel A and B report summary statistics when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) and 2.45 (in LM95) times the instantaneous volatility, respectively. AEO (American Eagle Outfitters Inc) is a large firm, CRK (Comstock Resources Inc) is a medium firm, and WSBF (Waterstone Financial Inc) is a small firm, according to their average of daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are reported in parentheses. The pval of the J-test is also reported.

Panel A: LM99				Panel B: LM95			
1 2	SPY AEO	SPY CRK	SPY WSBF	1 2	SPY AEO	SPY CRK	SPY WSBF
μ_1	0.233***	0.272***	0.236***	μ_1	0.242***	0.297***	0.258***
	(0.044)	(0.041)	(0.045)		(0.052)	(0.051)	(0.044)
μ_2	0.217**	-0.610***	-0.036	μ_2	0.278***	-0.801***	-0.091
	(0.087)	(0.209)	(0.089)		(0.089)	(0.232)	(0.098)
θ_1	0.016***	0.014***	0.016***	θ_1	0.015***	0.012***	0.015***
	(0.002)	(0.002)	(0.003)		(0.003)	(0.003)	(0.003)
θ_2	0.135***	0.470***	0.075***	θ_2	0.126***	0.298***	0.063***
	(0.006)	(0.035)	(0.007)		(0.006)	(0.071)	(0.007)
ρ	0.545***	0.481***	0.471***	ρ	0.600***	0.673***	0.545***
	(0.075)	(0.049)	(0.069)		(0.106)	(0.159)	(0.081)
α_1	235.36***	98.51***	108.96***	α_1	209.93**	121.27***	98.92**
	(88.91)	(28.02)	(42.10)		(81.52)	(43.14)	(45.09)
α_2	58.82**	40.16**	48.23***	α_2	70.06**	46.06***	43.57***
	(29.18)	(16.48)	(10.18)		(27.63)	(11.84)	(11.55)
β_{11}	205.22**	86.28***	91.19**	β_{11}	183.26**	105.71***	83.84**
	(81.80)	(25.28)	(36.81)		(74.42)	(40.25)	(40.57)
β_{12}	8.50	0.00	0.00	β_{12}	3.23	0.00	0.00
	(26.33)	(2.18)	(1.61)		(10.68)	(1.35)	(1.41)
β_{21}	2.88*	0.48*	2.50**	β_{21}	5.97**	1.43*	2.32*
	(1.74)	(0.29)	(1.22)		(2.68)	(0.87)	(1.30)
β_{22}	41.13**	33.95**	42.88***	β_{22}	52.10**	40.60***	38.70***
	(19.61)	(13.83)	(8.89)		(20.88)	(10.20)	(10.10)
$\lambda_{1,\infty}$	1.64*	2.98***	2.59**	$\lambda_{1,\infty}$	2.24*	3.88***	3.33**
	(0.95)	(0.72)	(1.15)		(1.27)	(1.36)	(1.47)
$\lambda_{2,\infty}$	1.07*	1.20**	1.25*	$\lambda_{2,\infty}$	1.63*	2.24*	2.01*
	(0.63)	(0.58)	(0.76)		(0.88)	(1.28)	(1.17)
$\mathbb{E}[\lambda_{1,t}]$	14.47**	24.04***	15.87**	$\mathbb{E}[\lambda_1, t]$	19.18**	30.24***	21.87***
	(6.16)	(5.21)	(8.05)		(9.19)	(9.89)	(7.48)
$\mathbb{E}[\lambda_{2,t}]$	5.93***	9.59***	18.65***	$\mathbb{E}[\lambda_2, t]$	12.74***	26.85***	28.36***
	(1.00)	(2.88)	(4.10)		(2.16)	(8.54)	(6.24)
pval	0.94	0.50	0.70	*pval*	0.70	0.30	0.85

Table 4: Summary Statistics of Average Estimated Jump Intensity (AEJI)

This table shows the summary statistics of the average of estimated daily jump intensity for all firms and for firms in three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). The calculation of AEJI for different jumps is described in III.C. Panel A, B, C, D, E, and F report summary statistics of the AEJI of all, negative, positive, self-exciting, memoryless, and market-exciting jumps, respectively. Column 2 is the number of firms in each size group. Column 3-9 and column 10-16 report summary statistics when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) and 2.45 (in LM95) times the instantaneous volatility, respectively. Row All, Large, Medium, and Small are summary statistics for all, large, medium and small firms, respectively.

Panel A: All jumps

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	8.90	2.56	1.84	7.20	8.65	10.24	23.76	19.34	4.37	1.41	16.25	19.24	22.00	39.35
Large	893	8.11	1.79	3.17	6.98	8.10	9.29	13.76	18.17	3.51	6.18	15.65	18.24	20.61	28.47
Medium	893	8.67	2.35	3.29	7.05	8.53	9.92	23.76	19.14	4.09	3.73	16.20	19.25	21.78	37.28
Small	893	9.91	3.05	1.84	7.80	9.60	11.72	22.98	20.72	5.00	1.41	17.37	20.64	23.92	39.35

Panel B: Negative jumps

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	5.65	2.05	1.27	4.44	5.26	6.29	22.80	11.86	3.52	0.88	9.69	11.36	13.10	35.26
Large	893	5.10	1.25	1.90	4.24	4.98	5.73	12.31	10.99	2.43	3.55	9.37	10.77	12.37	24.17
Medium	893	5.59	2.04	2.10	4.35	5.22	6.09	22.80	11.83	3.42	2.82	9.66	11.36	12.98	31.75
Small	893	6.25	2.50	1.27	4.77	5.76	6.90	22.11	12.77	4.24	0.88	10.27	11.95	14.10	35.26

Panel C: Positive jumps

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	5.07	2.28	0.98	3.63	4.52	5.85	22.49	11.53	3.75	0.78	9.17	10.77	12.96	35.35
Large	893	4.30	1.25	1.44	3.53	4.11	4.86	12.05	10.43	2.48	2.89	8.88	10.13	11.56	24.15
Medium	893	4.91	2.16	1.14	3.60	4.47	5.43	22.49	11.40	3.57	2.73	9.24	10.71	12.66	31.37
Small	893	6.00	2.81	0.98	4.01	5.63	7.24	21.81	12.75	4.55	0.78	9.65	12.08	14.68	35.35

Table 4 Continue.

		LM99							LM95						
Panel D: Self-exciting jumps															
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	5.95	2.40	0.27	4.06	5.94	7.73	14.47	12.16	4.68	0.32	8.41	12.09	16.01	24.68
Large	893	5.53	2.09	0.40	3.83	5.50	7.20	12.26	11.37	4.44	0.69	7.71	10.91	15.19	23.63
Medium	893	5.70	2.21	0.57	3.86	5.76	7.48	12.00	11.85	4.49	0.75	8.10	11.72	15.79	22.41
Small	893	6.60	2.70	0.27	4.54	6.65	8.52	14.47	13.28	4.90	0.32	9.51	13.40	17.07	24.68
Panel E: Memoryless jumps															
		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	1.82	2.27	0.00	0.48	1.08	2.23	21.53	4.05	4.22	0.00	1.24	2.67	5.34	32.48
Large	893	1.29	1.32	0.00	0.41	0.86	1.70	10.78	3.24	3.06	0.00	1.09	2.35	4.22	20.54
Medium	893	1.83	2.33	0.00	0.44	1.02	2.19	21.53	4.10	4.30	0.01	1.18	2.57	5.32	28.42
Small	893	2.35	2.78	0.00	0.62	1.38	3.00	20.94	4.80	4.93	0.03	1.50	3.10	6.24	32.48
Panel F: Market-exciting jumps															
		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	1.13	0.78	0.13	0.59	0.98	1.46	8.60	3.13	2.00	0.26	1.64	2.81	4.13	13.71
Large	893	1.28	0.73	0.15	0.76	1.17	1.62	7.15	3.56	1.89	0.39	2.15	3.32	4.72	12.01
Medium	893	1.15	0.78	0.16	0.65	1.01	1.45	8.60	3.19	1.88	0.26	1.86	2.91	4.18	13.51
Small	893	0.96	0.81	0.13	0.42	0.76	1.24	8.14	2.64	2.11	0.27	1.13	2.23	3.44	13.71

Table 5: Disagreement, News and estimated jump intensity for all firms

This table reports results of panel regressions of daily disagreement and news variables on the estimated intensity of various jumps for all firms during the sample period (June 01, 2010 - August 31, 2021). Column All, SE, ME, Neg, and Pos report the regressions results on the estimated intensity of all, self-exciting, market-exciting, negative, and positive jumps, respectively. Panel A and B reports the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. *NumberOfMessages* is the number of messages posted by investors who specify their investment philosophy as one of Fundamental, Technical, Momentum, Growth, and Value. *Disagreement* is the overall disagreement among those investors, constructed using equation (8). Other independent variables are described in II.B. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include the firm fixed effect and a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

	Panel A: LM99					Panel B: LM95				
	All	SE	ME	Neg	Pos	All	SE	ME	Neg	Pos
Number of Messages	0.305**	0.289**	0.016***	0.167**	0.139**	0.400**	0.369**	0.030**	0.235**	0.165**
	(0.128)	(0.123)	(0.006)	(0.068)	(0.062)	(0.160)	(0.148)	(0.012)	(0.100)	(0.064)
Disagreement	0.648***	0.602***	0.047***	0.482***	0.170***	0.853***	0.812***	0.041***	0.658***	0.198***
	(0.015)	(0.015)	(0.002)	(0.011)	(0.007)	(0.018)	(0.018)	(0.003)	(0.014)	(0.009)
Number of News	−0.006	0.012	−0.018***	0.027**	−0.034***	−0.013	0.039**	−0.052***	0.017	−0.031**
	(0.014)	(0.013)	(0.003)	(0.012)	(0.010)	(0.019)	(0.017)	(0.007)	(0.015)	(0.013)
News Tone				−0.111***	0.090***				−0.208***	0.163***
				(0.005)	(0.004)				(0.007)	(0.006)
|News Tone|	0.183***	0.176***	0.006***			0.235***	0.229***	0.006***		
	(0.007)	(0.007)	(0.001)			(0.010)	(0.009)	(0.002)		
Size	−0.016	−0.009	−0.007	0.045*	−0.060*	−0.097	−0.013	−0.084**	−0.028	−0.067*
	(0.043)	(0.041)	(0.004)	(0.026)	(0.034)	(0.086)	(0.060)	(0.038)	(0.049)	(0.039)
12-month Momentum	−0.945***	−0.666***	−0.279***	−2.476***	1.535***	−1.764***	−1.132***	−0.632***	−4.223***	2.464***
	(0.044)	(0.037)	(0.009)	(0.057)	(0.028)	(0.071)	(0.054)	(0.020)	(0.098)	(0.043)
Idiosyncratic Volatility	0.181***	0.168***	0.013***	0.241***	−0.060***	0.147***	0.099***	0.049***	0.348***	−0.200***
	(0.043)	(0.043)	(0.004)	(0.048)	(0.018)	(0.020)	(0.015)	(0.008)	(0.057)	(0.043)
N	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286
R^2(%)	0.75	0.54	0.89	6.45	3.18	0.98	0.61	1.32	8.70	3.50

Table 6: Disagreement sources, News and estimated jump intensity for all firms

This table reports results of panel regressions of daily disagreement sources and news variables on the estimated intensity of various jumps for all firms during the sample period (June 01, 2010 - August 31, 2021). Column All, SE, ME, Neg, and Pos report the regressions results on the estimated intensity of all, self-exciting, market-exciting, negative, and positive jumps, respectively. Panel A and B reports the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. *NumberOfMessages* is the number of messages posted by investors who specify their investment philosophy as one of Fundamental, Technical, Momentum, Growth, and Value. *Disagreement Within-Group* and *Disagreement Cross-Group* are disagreement within-group and cross-group among those investors, constructed using equations (12) and (13), respectively. Other independent variables are described in II.B. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include the firm fixed effect and a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

	Panel A: LM99					Panel B: LM95				
	All	SE	ME	Neg	Pos	All	SE	ME	Neg	Pos
Number of Messages	0.293**	0.279**	0.015***	0.159**	0.135**	0.384**	0.356**	0.028**	0.226**	0.160***
	(0.123)	(0.118)	(0.005)	(0.065)	(0.061)	(0.154)	(0.143)	(0.012)	(0.097)	(0.062)
Disagreement Within-Group	0.520***	0.471***	0.049***	0.386***	0.135***	0.682***	0.628***	0.054***	0.515***	0.169***
	(0.016)	(0.015)	(0.002)	(0.011)	(0.007)	(0.019)	(0.019)	(0.003)	(0.014)	(0.009)
Disagreement Cross-Group	0.251***	0.254***	−0.003**	0.160***	0.093***	0.318***	0.343***	−0.025***	0.213***	0.108***
	(0.008)	(0.007)	(0.001)	(0.005)	(0.005)	(0.010)	(0.009)	(0.003)	(0.007)	(0.006)
Number of News	−0.010	0.008	−0.018***	0.026**	−0.037***	−0.017	0.035**	−0.051***	0.018	−0.035***
	(0.014)	(0.013)	(0.003)	(0.012)	(0.010)	(0.018)	(0.017)	(0.007)	(0.015)	(0.014)
News Tone				−0.110***	0.090***				−0.207***	0.163***
				(0.005)	(0.004)				(0.007)	(0.006)
∣News Tone∣	0.184***	0.177***	0.007***			0.236***	0.230***	0.007***		
	(0.007)	(0.007)	(0.001)			(0.010)	(0.009)	(0.002)		
Size	−0.015	−0.008	−0.007	0.046*	−0.060*	−0.096	−0.011	−0.084**	−0.026	−0.068*
	(0.042)	(0.040)	(0.004)	(0.026)	(0.034)	(0.084)	(0.058)	(0.037)	(0.048)	(0.039)
12-month Momentum	−0.946***	−0.668***	−0.279***	−2.477***	1.535***	−1.766***	−1.134***	−0.632***	−4.224***	2.463***
	(0.044)	(0.037)	(0.009)	(0.057)	(0.028)	(0.071)	(0.054)	(0.020)	(0.098)	(0.043)
Idiosyncratic Volatility	0.182***	0.169***	0.013***	0.243***	−0.060***	0.149***	0.099***	0.050***	0.351***	−0.201***
	(0.043)	(0.043)	(0.004)	(0.048)	(0.018)	(0.020)	(0.015)	(0.008)	(0.058)	(0.043)
N	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286
R^2(%)	0.74	0.53	0.89	6.43	3.19	0.97	0.60	1.33	8.68	3.50

Table 7: Disagreement and sophistication, News and estimated jump intensity for all firms

This table reports results of panel regressions of daily sophistication-related disagreement and news variables on the estimated jump intensity of various jumps for all firms during the sample period (June 01, 2010 - August 31, 2021). Column All, SE, ME, Neg, and Pos report the regressions results on the estimated intensity of all, self-exciting, market-exciting, negative, and positive jumps, respectively. Panel A and B reports the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. *NumberOfMessages(S)*, *NumberOfMessages(U)*, *Disagreement(S)*, *Disagreement(U)*, and ∣*Disagreement(S-U)*∣ are described in III.F. Other independent variables are described in II.B. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include the firm fixed effect and a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and reported in parentheses.

	Panel A: LM99					Panel B: LM95				
	All	SE	ME	Neg	Pos	All	SE	ME	Neg	Pos
Number of Messages (S)	0.132	0.150	−0.018***	0.078	0.057	0.163	0.240*	−0.077***	0.063	0.104
	(0.104)	(0.104)	(0.005)	(0.051)	(0.058)	(0.132)	(0.135)	(0.018)	(0.077)	(0.064)
Number of Messages (U)	0.117*	0.096*	0.022***	0.048	0.067**	0.166**	0.082	0.083***	0.118*	0.045
	(0.062)	(0.058)	(0.006)	(0.042)	(0.030)	(0.080)	(0.065)	(0.022)	(0.070)	(0.034)
Disagreement (S)	0.226***	0.234***	−0.008***	0.148***	0.079***	0.285***	0.318***	−0.033***	0.185***	0.101***
	(0.018)	(0.018)	(0.001)	(0.010)	(0.009)	(0.022)	(0.022)	(0.004)	(0.014)	(0.011)
Disagreement (U)	0.618***	0.527***	0.091***	0.461***	0.158***	0.801***	0.678***	0.123***	0.606***	0.197***
	(0.020)	(0.018)	(0.003)	(0.015)	(0.008)	(0.025)	(0.022)	(0.005)	(0.019)	(0.010)
∣Disagreement (S-U)∣	0.123***	0.125***	−0.001*	0.240***	−0.121***	0.322***	0.207***	0.115***	0.616***	−0.299***
	(0.007)	(0.007)	(0.001)	(0.006)	(0.005)	(0.011)	(0.010)	(0.003)	(0.011)	(0.008)
Number of News	−0.012	0.005	−0.016***	0.025**	−0.038***	−0.017	0.026	−0.043***	0.024**	−0.042**
	(0.014)	(0.013)	(0.002)	(0.010)	(0.012)	(0.020)	(0.018)	(0.005)	(0.012)	(0.017)
News Tone				−0.100***	0.087***				−0.184***	0.155***
				(0.005)	(0.004)				(0.007)	(0.006)
∣News Tone∣	0.190***	0.183***	0.007***			0.248***	0.239***	0.009***		
	(0.007)	(0.007)	(0.001)			(0.010)	(0.010)	(0.002)		
Size	−0.009	−0.005	−0.003	0.057*	−0.065*	−0.081	−0.009	−0.072**	−0.001	−0.080*
	(0.044)	(0.043)	(0.003)	(0.030)	(0.036)	(0.080)	(0.063)	(0.033)	(0.042)	(0.045)
12-month Momentum	−0.935***	−0.657***	−0.278***	−2.460***	1.528***	−1.742***	−1.118***	−0.624***	−4.183***	2.446***
	(0.044)	(0.037)	(0.009)	(0.057)	(0.028)	(0.071)	(0.054)	(0.019)	(0.097)	(0.042)
Idiosyncratic Volatility	0.180***	0.168***	0.012***	0.240***	−0.059***	0.146***	0.099***	0.047***	0.347***	−0.200***
	(0.042)	(0.043)	(0.004)	(0.048)	(0.018)	(0.019)	(0.015)	(0.007)	(0.057)	(0.042)
N	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286
R^2(%)	0.79	0.56	0.95	6.54	3.20	1.03	0.62	1.42	8.91	3.55

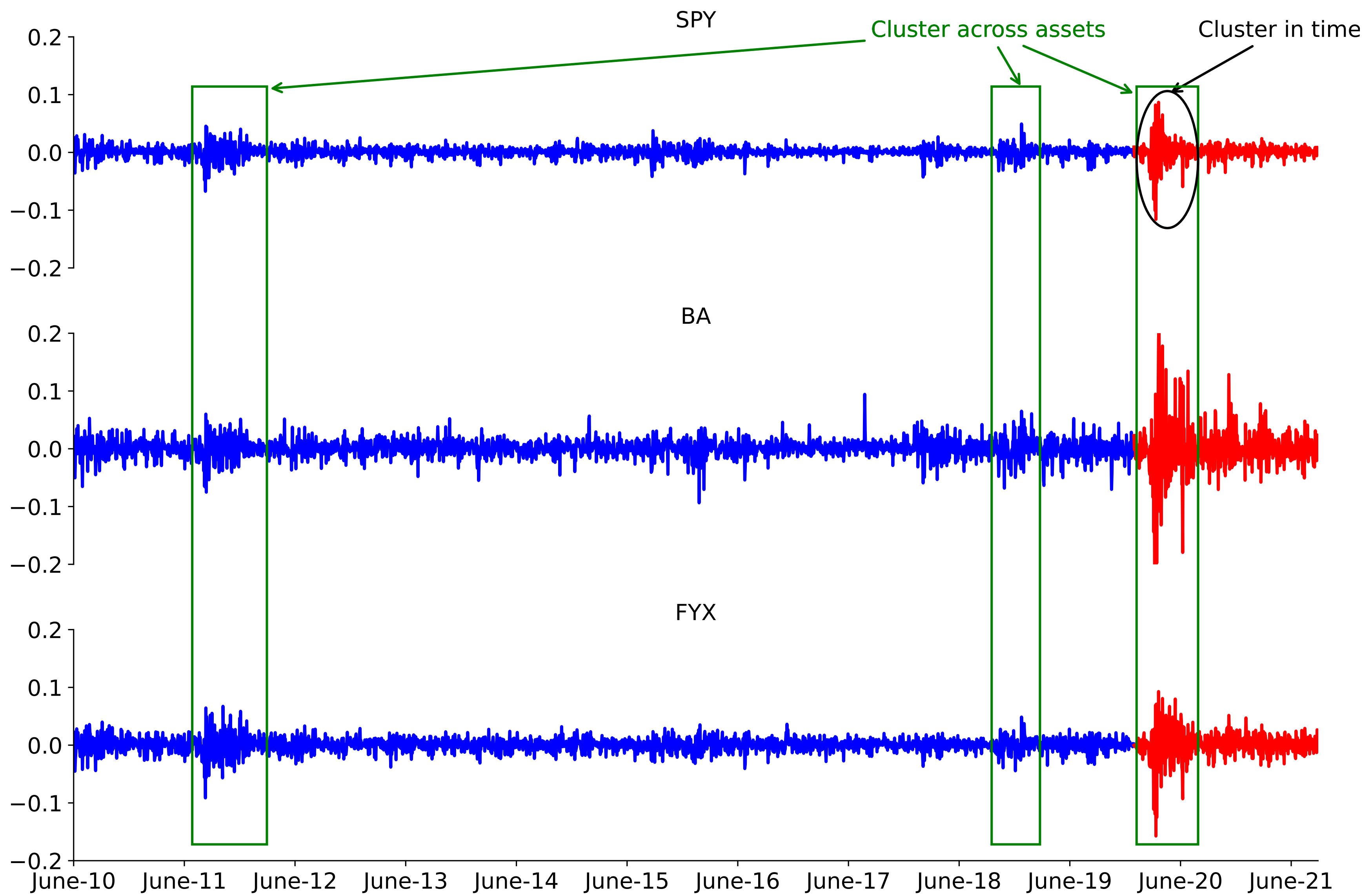


Figure 1: A stylised fact of asset returns: jump clustering in time and across assets. This figure depicts the stylised fact of asset returns, jump clustering in time and across assets, by plotting the daily returns of SPY (SPDR S&P 500 ETF Trust), BA (Boeing), and FYX (First Trust Small Cap Core AlphaDEX Fund) during the sample period (June 01, 2010 - August 31, 2021). The black ellipse marks a period when jumps tend to cluster in time. The green rectangles mark periods when jumps tend to cluster across assets. The blue part shows returns in pre-COVID period (June 01, 2010 - December 31, 2019), and the red part shows returns in COVID period (January 01, 2020-August 31, 2021).

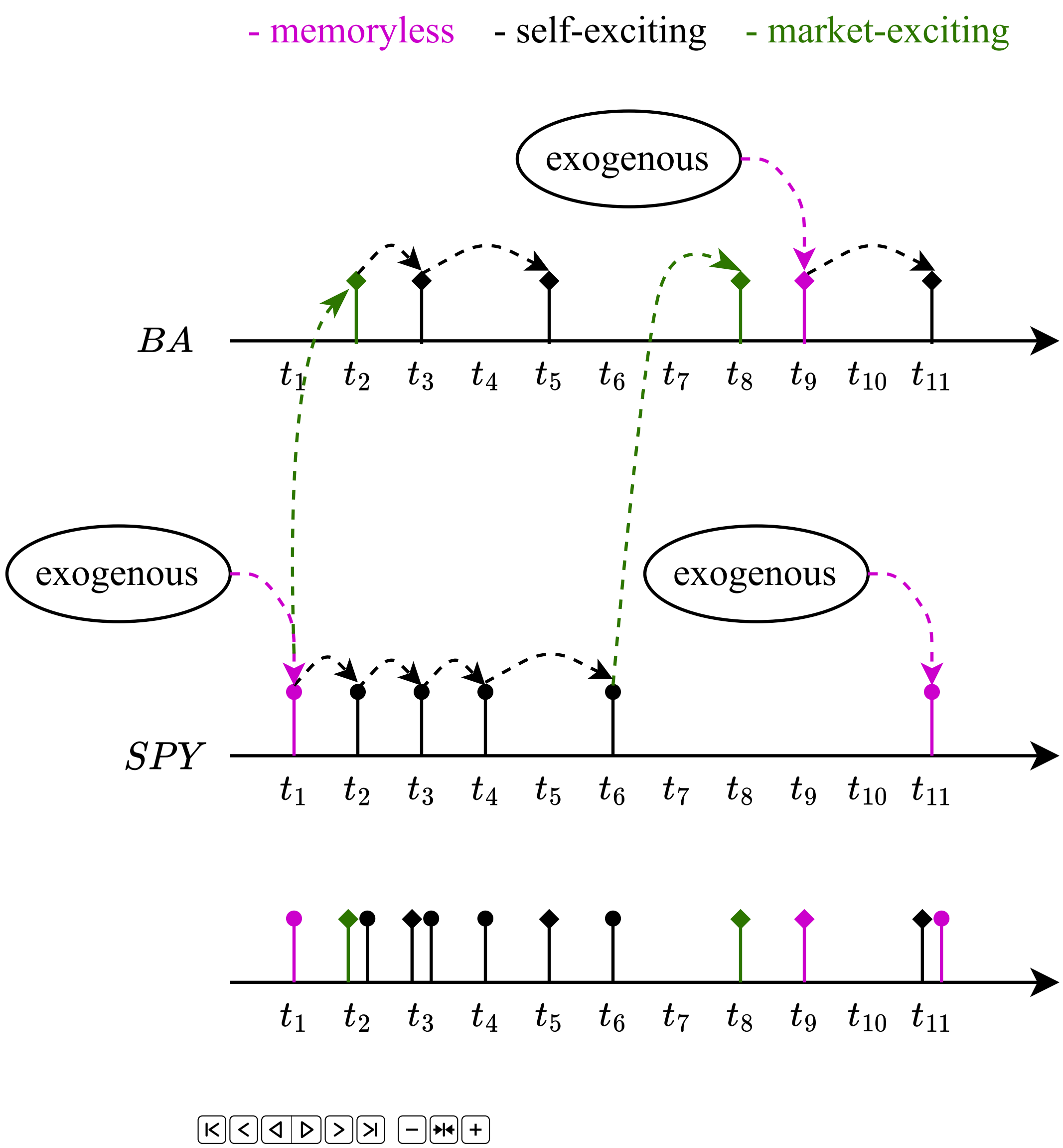


Figure 2: A conceptual framework of jump propagation over time and across assets. This figure illustrates a conceptual framework of jumps propagating over time and jumps propagating from SPY (SPDR S&P 500 ETF Trust) to BA (Boeing). The vertical line with a circle or rectangle on top marks the occurrence of a jump in SPY or in BA, respectively. The ellipse represents an exogenous shock, which can be an important announcement or a big event. The magenta vertical line refers to a memoryless jump, the black vertical line refers to a self-exciting jump, and the green vertical line refers to a market-exciting jump. The magenta dash line connects the exogenous shock with the memoryless jump that the shock triggers. The black dash line connects a jump in BA (SPY) to an earlier jump in BA (SPY) that excites that jump (jumps propagate over time). The green dash line connects a jump in BA to an earlier jump in SPY that excites the jump in BA (jumps propagate from SPY to BA).

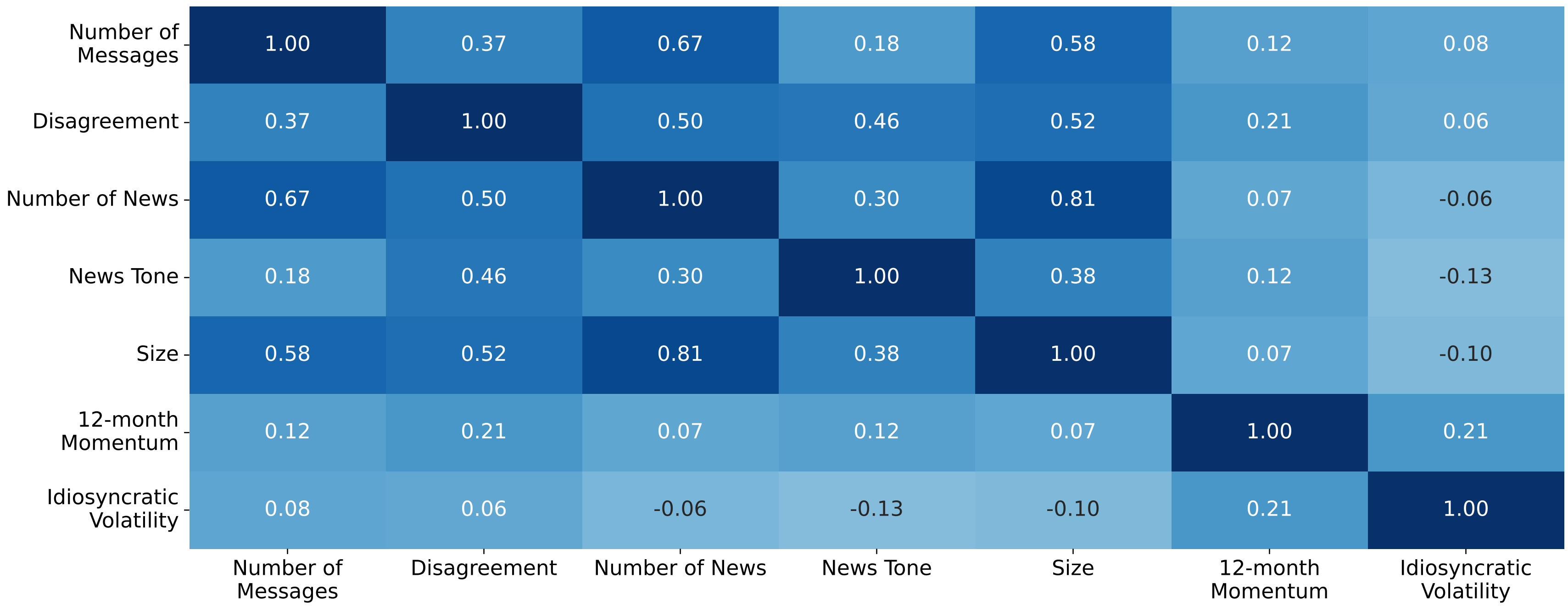
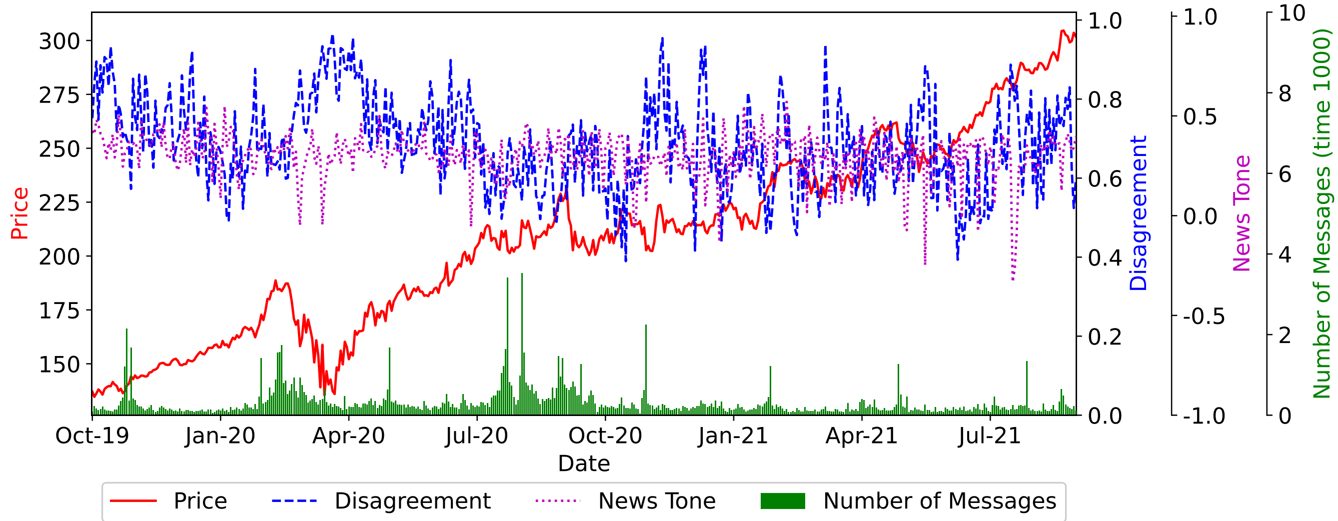
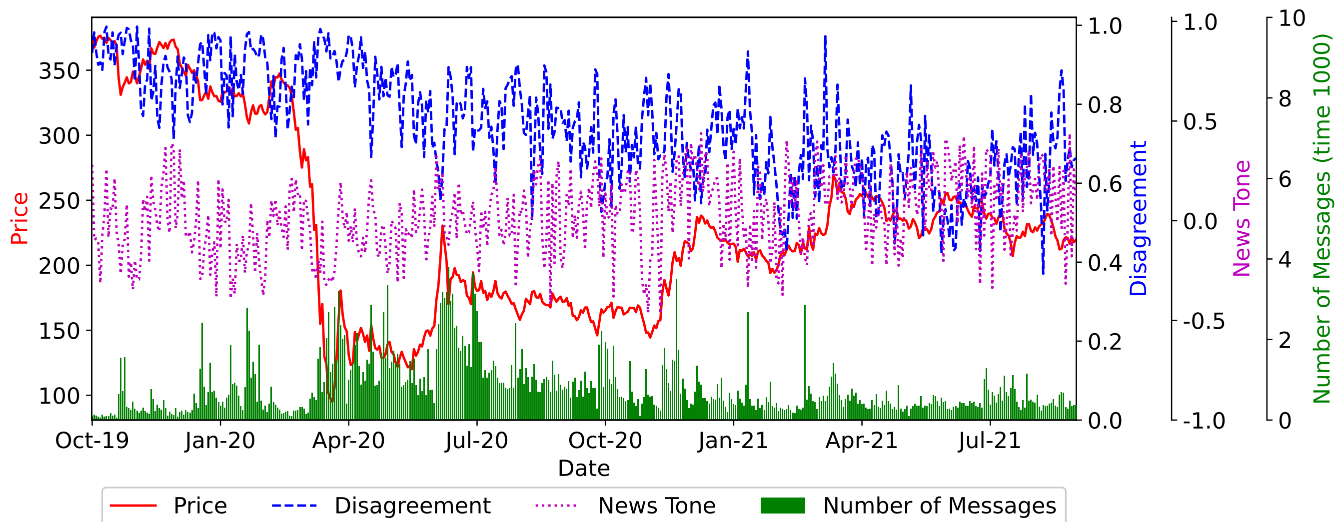


Figure 3: Correlation between firm characteristics. This figure shows the correlation between the average of daily observations of firm characteristics for 2,679 stocks during the sample period (June 01, 2010 - August 31, 2021). *Number of Messages* and *Disagreement* are the number of investor-posted messages and overall disagreement among investors at the firm-day level in StockTwits. The construction of disagreement is described in II.A *Number of News* and *News Tone* are the number of news articles and the overall news tone at the firm-day level from RavenPack. *News Tone* is the variable *event_sentiment_score* in RavenPack. *Size*, *12-month Momentum*, and *Idiosyncratic Volatility* are the market capitalization, the past 12-months return and the idiosyncratic volatility from CRSP/Compustat, respectively. Size and 12-month Momentum are at the firm-day level, while Idiosyncratic Volatility is at the firm-month level. *Size* is the variable *me* retrieved from CRSP. *Idiosyncratic Volatility* is computed as the variance of residual from Fama and French (1993) three-factor model based on daily return in the past 3 months. The missing values of characteristics are filled as the following. We assign firms to three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period. On each day, a firms missing characteristic is replaced by the cross-sectional median within the size group to which the firm belongs.



(a) MSFT



(b) BA

Figure 4: Stock price, disagreement, news tone, and the number of messages. This figure shows the daily stock price, disagreement, news tone, and the number of messages of MSFT (Microsoft) in (a) and BA (Boeing) in (b). The period is from October 01, 2019 to August 31, 2021. The red solid line is stock price. The blue dashed line is disagreement. The magenta dotted line is news tone. The green bar is the number of messages. The number of messages and disagreement are the number of investor-posted messages and the overall disagreement among investors at the firm-day level in StockTwits. The construction of disagreement is described in II.A.

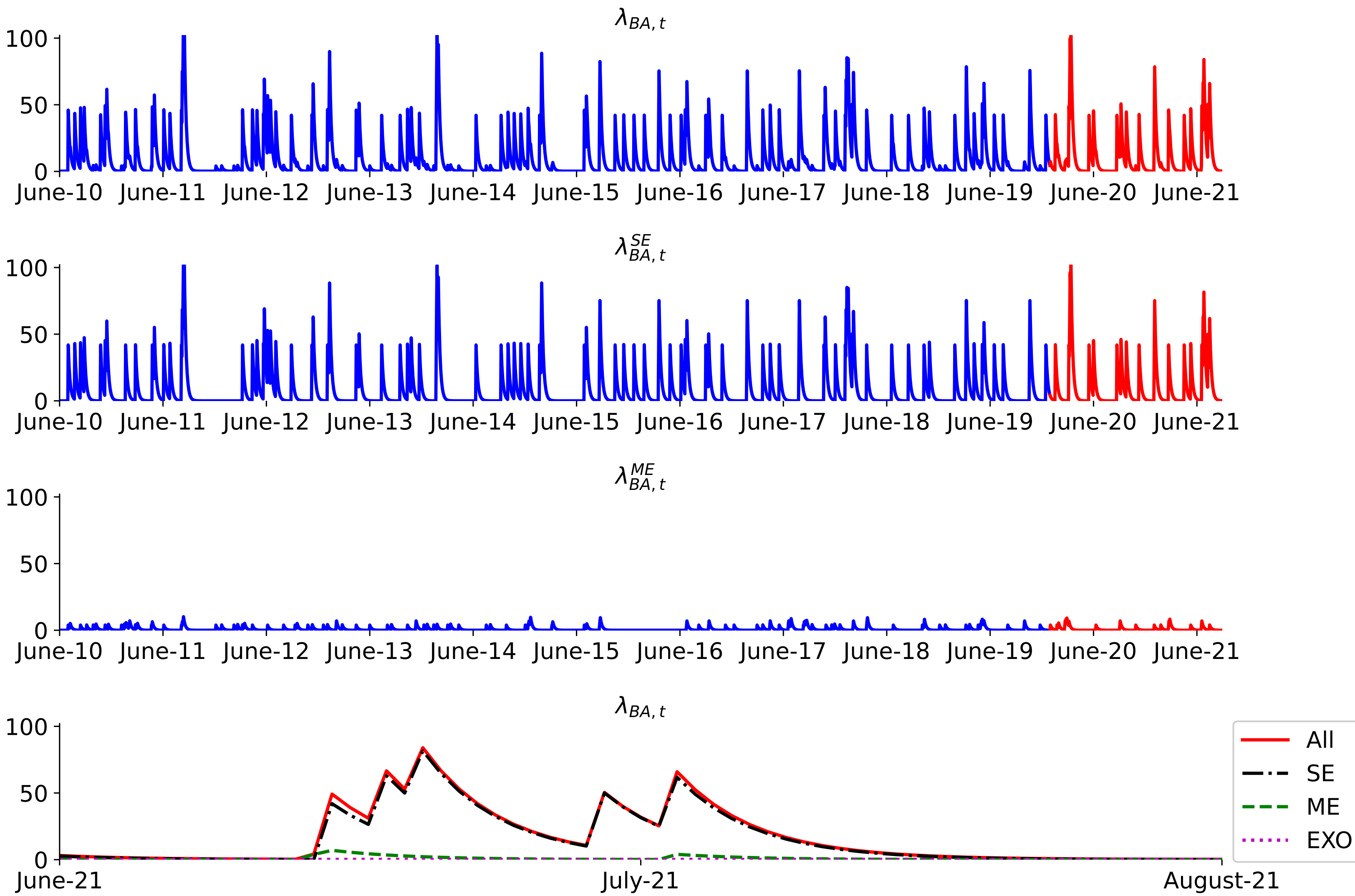
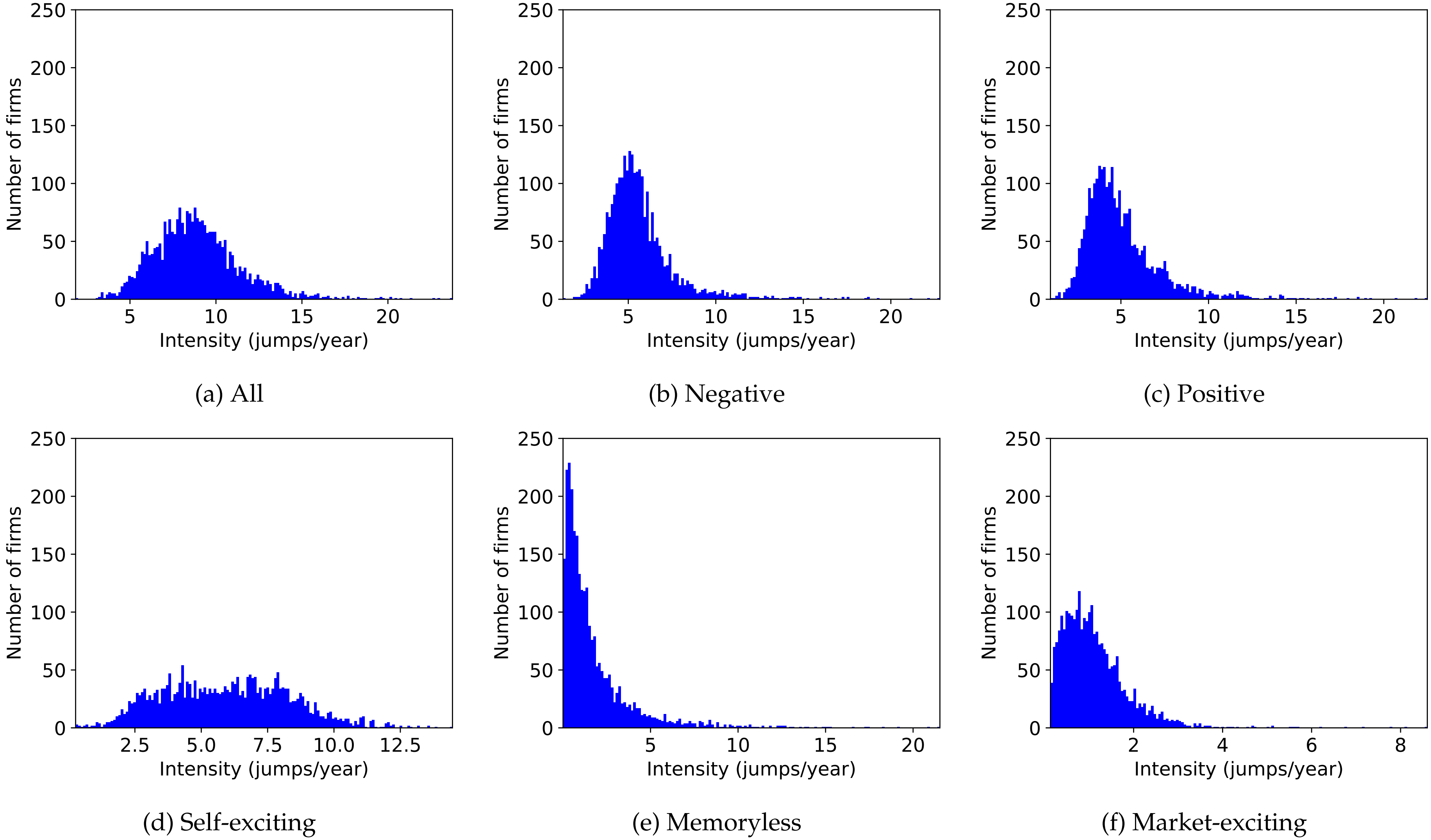


Figure 5: Estimated time series of jump intensity. This figure plots the estimated daily jump intensity for BA during the sample period (June 01, 2010-August 31, 2021), when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) times the instantaneous volatility. The procedures for estimating daily intensity of different jumps is described in III.C. The first, second and third rows show the estimated intensity of all (All), self-exciting (SE), and market-exciting (ME) jumps during the sample period, respectively. The forth row shows the estimated intensity of All, SE, ME and memoyless (EXO) jumps in a shorter period (June 01, 2021-August 31, 2021). The blue and red parts refer to the pre-COVID period (June 01, 2010 - December 31, 2019) and the COVID period (January 01, 2020 - August 31, 2021), respectively.



(a) All (b) Negative (c) Positive

(d) Self-exciting (e) Memoryless (f) Market-exciting

Figure 6: Histograms of the average estimated jump intensity. This figure shows the histograms of the average estimated jump intensity (AEJI) for 2,679 stocks during the sample period (June 01, 2010 - August 31, 2021), when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) times the instantaneous volatility. The procedure for estimating daily intensity of different jumps is described in III.C and The AEJI of all jumps is the average of the estimated daily intensity of all jumps during the sample period. The AEJI of a given type of jumps is the average of the estimated daily intensity of that type of jumps during the sample period. (a), (b) and (c) are the histograms of the AEJI of all, negative, and positive jumps, respectively. (d), (e) and (f) are the histograms of the AEJI of self-exciting, memoryless, and market-exciting jumps, respectively.

Appendix A. GMM estimation

There are two main choices for conducting the estimation procedure of the Hawkes-jump-diffusion (HJD) processes depends on whether it has a diffusion component or not: the generalized method of moments (GMM) and Maximum Likelihood Estimation (MLE).[14] One can use the MLE approach for a pure-jump Hawkes process,[15] however, as the jump-diffusion Hawkes model does not have an explicit expression for its distribution function, the implementation of the MLE is not feasible. Instead of distribution function, the idea behind GMM estimation is to make the sample moments and model moments as close as possible, or equivalently find a parameter set that minimize the error between these moments. Aït-Sahalia et al. (2015) showed that it is possible to have an explicit expression for some important moment functions of the jump-diffusion Hawkes processes.[16] The closed-form moment functions will greatly simplify the GMM estimation, especially when our estimation involves a large number of parameters.

Given a time interval $[0, T]$, let ΔX_n be the stock returns between the time interval $[(n-1)\Delta,\ n\Delta]$. Then, we have a series of stock returns $\{\Delta X_n\}_{n>0}$ with $n = 1, 2, ..N$, $\Delta > 0$ and $N = \frac{T}{\Delta}$. Denote the vector of parameters by $\vartheta = [\vartheta_1, ..., \vartheta_p]'$, $\vec{0} = [0, ..., 0]'$ be a $m \times 1$ vector of zeros. Also, $h(\Delta X_n, \vartheta) = [h_1(\Delta X_n, \vartheta), ..., h_m(\Delta X_n, \vartheta)]'$, is a $m \times 1$ vector of moment conditions with $m \geq p$ and is differentiable with respect to ϑ. Each $h_i(\Delta X_n, \vartheta)$ represents the difference between the ith sample moment and the model moment of the stock returns. Denote the sample average of the moment conditions $h(\Delta x_n, \vartheta)$ by

$$g(\Delta x, \vartheta) = \frac{1}{N} \sum_{n=1}^{N} h(\Delta x_n, \vartheta)$$

[14] MLE requires the assumptions of the data generating process (DGP). In contrast, GMM requires prior knowledge about the moment functions but no assumptions of DGP, allowing a more flexible estimation, see Hansen (1982).

[15] Although, the MLE is theoretically feasible for the pure-jump Hawkes process, however, a large parameter space and low curvatures for the objective function are two main obstacles in the implementation of it, see Zhou, Zha, and Song (2013) and Xu, Morse, and González (2020). Bacry, Bompaire, Deegan, Gaïffas, and Poulsen (2017) develop an excellent *Python* library for the estimation of the pure-jump Hawkes process via parametric/non-parametric approaches.

[16] See Section 3 in Aït-Sahalia et al. (2015).

Then, the GMM estimation of ϑ is given by

$$\hat{\vartheta} = argmin_{\vartheta \in \Theta} g(\Delta x, \vartheta)' W g(\Delta x, \vartheta)$$

where W is a $m \times m$ weighting matrix. Hansen and Singleton (1982) shows the optimal weighting matrix is $W^{opt} = S^{-1}$, where S is the asymptotic covariance matrix of the sample average $g(\Delta x, \vartheta_*)$ where ϑ_* represents the true parameter vector

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}[h(\Delta x_n, \vartheta_*) h(\Delta x_{n-j}, \vartheta_*)']$$

A positive semi-definite (PSD) estimator of S is given in Newey and West (1986)

$$\hat{S} = \Gamma_0 + \sum_{j=1}^{m} w_j (\Gamma_j + \Gamma_j')$$

with $w_j = 1 - \frac{j}{m+1}$ and

$$\Gamma_j = \frac{1}{N} \sum_{n=j+1}^{N} h(\Delta x_n, \hat{\vartheta}) h(\Delta x_{n-j}, \hat{\vartheta})'$$

where $\hat{\vartheta}$ is the consistent estimate of ϑ_*.

Hansen, (Hansen, 1982, Theorem 3.1), shows that under the standard regularity conditions the following asymptotic normality holds

$$\sqrt{N}(\hat{\vartheta} - \vartheta_*) \sim N(0, \Sigma(\hat{\vartheta}))$$

where

$$\Sigma(\hat{\vartheta}) = (D'WD(D'WSWD)^{-1} D'WD)^{-1}$$

is the asymptotic variance of $\hat{\vartheta}$ with

$$D := \frac{\partial \mathbb{E}[h(\Delta X_n, \vartheta)]}{\partial \vartheta'} \Big|_{\vartheta = \vartheta_*} = \frac{\partial g(\Delta X_n, \vartheta)}{\partial \vartheta'} \Big|_{\vartheta = \vartheta_*}.$$

Hansen, (Hansen, 1982, Theorem 3.2) also shows that with the optimal weighting matrix, i.e. $W = W^{opt}$, the asymptotic variance becomes

$$\Sigma(\hat{\vartheta}) = (D'(S)^{-1}D)^{-1}. \tag{A1}$$

Appendix B. Hypothesis Testing

We assess the overall fitness of the model using the misspecification test, also known as the *J-test*. When the number of moment conditions is larger the number parameters – $m > p$, the J-test is used to testing the over-identifying restrictions. In this case, the null hypothesis is

$$\mathcal{H}_0^J : \text{the model is correctly specified}$$

where its asymptotic distribution is a chi-square with $m - p$ degree of freedom – the J-statistic satisfies

$$J \sim \chi^2(m - p)$$

where

$$J := Ng(\Delta X_n, \hat{\vartheta})'\hat{S}^{-1}g(\Delta X_n, \hat{\vartheta}).$$

A mis-specified model or a model with invalid moment conditions will generate a large J-statistic.

Appendix C. The GMM implementation: some details

We adopt iterative GMM estimation procedure from ACL, which consists of two basic steps and an iterative step:

- First step: we extract the continuous part from the daily stock returns by filtering a return if the absolute value of that return is smaller than 3.22 (for LM99) and 2.45 (for LM95) times its instantaneous volatility, which is estimated from the realized bipower variation. [17] We then fit the truncated data into the moment conditions that are derived from the continuous part of the model to estimate continuous-part parameters, i.e. $\{\mu_i, \mu_j, \theta_i, \theta_j, \rho\}$ in bivariate case.

- Second step: we fix the estimated continuous-part parameters and estimate the jump parameters, i.e. $\{\alpha_i, \alpha_j, \beta_{i,i}, \beta_{i,j}, \beta_{j,i}, \beta_{j,j}, \lambda_{i,\infty}, \lambda_{j,\infty}\}$ in bivariate case, on the original stock returns using the full moment conditions. We then obtain the initial estimates of all parameters.
- Iterative step: using the obtained initial estimates in the second step, we estimate the whole set of parameters on the original stock returns using the full moment conditions. We repeat this procedure until the GMM estimates do not change significantly from one iteration to the next iteration.

How do we tackle the optimization. We utilize the existing function `minimize` in the package `scipy.optimize`. We use the `Sequential Least Squares Programming` or `SLSQP` in bivariate case. SLSQP allows us to impose bounds and constraints on parameters. In implementation, the bounds of μ_i and μ_j are $(-\infty, +\infty)$. The bounds of θ_i and θ_j are $[0, +\infty)$. The bound of ρ is $[0, 1)$. The bounds of α_i, α_j, $\beta_{i,i}$, $\beta_{i,j}$, $\beta_{j,i}$, $\beta_{j,j}$, $\lambda_{i,\infty}$, and $\lambda_{j,\infty}$ are $[0, +\infty)$. Especially, we impose constraint, $(\alpha_i - \beta_{i,i})(\alpha_j - \beta_{j,j}) > \beta_{i,j}\beta_{j,i}$, to maintain the positive mean of jump intensity, i.e. $E[\lambda_{i,t}]$ and $E[\lambda_{j,t}]$.

Numerical derivatives. The derivatives of moment conditions are required to compute the Ω in (A1). We approximated the derivatives using forward finite difference method with the optimal step size $h = \sqrt{\epsilon}|x|$, where x is a vector of values at which function derivative is evaluated.

[17] The truncation filters out the jumps from stock returns. Therefore, the truncated data can be approximated as the continuous part.

REFERENCES

Aït-Sahalia, Yacine, Julio Cacho-Diaz, and Roger JA Laeven, 2015, Modeling financial contagion using mutually exciting jump processes, *Journal of Financial Economics* 117, 585–606.

Aït-Sahalia, Yacine, and Thomas Robert Hurd, 2015, Portfolio choice in markets with contagion, *Journal of Financial Econometrics* 14, 1–28.

Aït-Sahalia, Yacine, Roger JA Laeven, and Loriana Pelizzon, 2014, Mutual excitation in eurozone sovereign cds, *Journal of Econometrics* 183, 151–167.

Bacry, Emmanuel, Martin Bompaire, Philip Deegan, Stéphane Gaïffas, and Søren V Poulsen, 2017, Tick: a python library for statistical learning, with an emphasis on hawkes processes and time-dependent models, *Journal of Machine Learning Research* 18, 7937–7941.

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

Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129–152.

Baker, Scott R, Nicholas Bloom, Steven J Davis, and Marco C Sammon, 2022, What triggers stock market jumps?, Technical report, Tech.Rep, National Bureau of Economic Research.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Bégin, Jean-François, Christian Dorion, and Geneviève Gauthier, 2020, Idiosyncratic jump risk matters: Evidence from equity returns and options, *Review of Financial Studies* 33, 155–211.

Chang, Yen-Cheng, Pei-Jie Hsiao, Alexander Ljungqvist, and Kevin Tseng, 2022, Testing disagreement models, *The Journal of Finance* 77, 2239–2285.

Cookson, J Anthony, Joseph Engelberg, and William Mullins, 2022, Echo chambers, *Available at SSRN 3603107* .

Cookson, J Anthony, and Marina Niessner, 2020, Why don't we agree? evidence from a social network of investors, *Journal of Finance* 75, 173–228.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The sum of all fears investor sentiment and asset prices, *Review of Financial Studies* 28, 1–32.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under-and overreactions, *Journal of Finance* 53, 1839–1885.

Dasgupta, Amil, and Andrea Prat, 2008, Information aggregation in financial markets with career concerns, *Journal of Economic Theory* 143, 83–113.

Deng, Shuyuan, Zhijian James Huang, Atish P Sinha, and Huimin Zhao, 2018, The interaction between microblog sentiment and stock return: An empirical examination, *MIS quarterly* 42, 895–918.

Diether, Karl B, Christopher J Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *The Journal of Finance* 57, 2113–2141.

Engle, Robert F, Eric Ghysels, and Bumjean Sohn, 2013, Stock market volatility and macroeconomic fundamentals, *Review of Economics and Statistics* 95, 776–797.

Fallahgoul, Hasan, 2021, Inside the mind of investors during the covid-19 pandemic: Evidence from the stocktwits data, *Journal of Financial Data Science* 3, 134–148.

Fama, Eugene F, 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383–417.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Filimonov, Vladimir, and Didier Sornette, 2012, Quantifying reflexivity in financial markets: Toward a prediction of flash crashes, *Physical Review E* 85, 056108.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 1029–1054.

Hansen, Lars Peter, and Kenneth J Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 1269–1286.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, 2004, Social interaction and stock-market participation, *Journal of Finance* 59, 137–163.

Hong, Harrison, and Jeremy C Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of finance* 54, 2143–2184.

Hong, Harrison, and Jeremy C Stein, 2003, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies* 16, 487–525.

Hong, Harrison, and Jeremy C Stein, 2007, Disagreement and the stock market, *Journal of Economic Perspectives* 21, 109–128.

Jeon, Yoontae, Thomas H McCurdy, and Xiaofei Zhao, 2021, News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies, *Journal of Financial Economics* .

Kim, Minkyoung, Dean Paini, and Raja Jurdak, 2019, Modeling stochastic processes in disease spread across a heterogeneous social system, *Proceedings of the National Academy of Sciences* 116, 401–406.

Lamont, Owen A, and Richard H Thaler, 2003, Can the market add and subtract? mispricing in tech stock carve-outs, *Journal of Political Economy* 111, 227–268.

Lee, Suzanne S, and Per A Mykland, 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial Studies* 21, 2535–2563.

Newey, Whitney K, and Kenneth D West, 1986, A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix, Technical report, National Bureau of Economic Research.

Tetlock, Paul C, 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of finance* 62, 1139–1168.

Vayanos, Dimitri, and Jiang Wang, 2012, Liquidity and asset returns under asymmetric information and imperfect competition, *Review of Financial Studies* 25, 1339–1365.

Xu, Sharon, Steven Morse, and Marta C González, 2020, Modeling human dynamics and lifestyle using digital traces, *arXiv preprint arXiv:2005.06542* .

Zhou, Ke, Hongyuan Zha, and Le Song, 2013, Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes, in *Artificial Intelligence and Statistics*, 641–649.

Internet Appendix to:

Investor Disagreement: The Secret Fuel Behind Stock Price Jumps

Hasan Fallahgoul*

Monash University

Xin Lin†

Monash University

This Version: February 28, 2023

Link to Most Recent Version

*Hasan Fallahgoul, Monash University, School of Mathematics and Centre for Quantitative Finance and Investment Strategies, 9 Rainforest Walk, 3800 Victoria, Australia. E-mail: hasan.fallahgoul@monash.edu.

†Xin Lin, Monash University, School of Mathematics and Centre for Quantitative Finance and Investment Strategies, 9 Rainforest Walk, 3800 Victoria, Australia. E-mail: xin.lin@monash.edu

Table IA.1: Filtering Restrictions and the Size of the Analysis Sample

This table presents the number of messages, users, stocks and days during the sample period (June 01, 2010 - August 31, 2021) in StockTwits, when taking successive filtering actions. Step 1 is to keep user-posted messages that mention only 1 symbol per message in StockTwits. Step 2 is to merge the universe of StockTwits symbols left from Step 1 with the universe of CRSP symbols. Step 3 is to keep stocks with non-missing daily returns during the sample period.

Step	Messages	Users	Stocks	Days	Trading Days	Filtering Actions
1	173,898,340	1,104,849	19,469	4,124	2,834	keep messages with 1 symbol per message
2	154,860,211	995,046	11,267	4,112	2,834	merge with CRSP symbols
3	34,373,324	494,827	2,679	4,112	2,834	keep stocks with non-missing daily returns

Table IA.2: Summary Statistics of Number of Realized Jumps

This table shows summary statistics of the number of realized jump size, for all firms and for firms in three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A and B report summary statistics when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) and 2.45 (in LM95) times the instantaneous volatility, respectively. Row All, Large, Medium, and Small are summary statistics for all, large, medium and small firms, respectively.

Panel A: All jumps

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	123	17	64	111	122	134	205	256	18	196	244	255	267	338
Large	893	118	12	70	110	118	126	166	251	14	196	242	251	260	297
Medium	893	120	14	69	110	120	130	177	253	15	199	243	253	263	304
Small	893	131	21	64	117	132	147	205	265	21	199	251	265	280	338

Panel B: Negative jumps

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	65	9	36	60	66	71	114	131	11	97	123	130	138	170
Large	893	65	8	37	59	65	69	88	130	10	100	123	129	136	161
Medium	893	65	9	36	59	65	71	101	130	11	97	122	129	137	166
Small	893	67	9	36	61	67	73	114	132	11	104	124	132	140	170

Panel C: Positive jumps

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	58	15	20	47	56	67	107	126	16	79	115	125	137	182
Large	893	53	10	27	47	54	60	83	121	12	88	113	122	130	158
Medium	893	55	12	20	47	55	64	99	123	14	79	115	124	132	165
Small	893	64	18	25	50	67	79	107	133	19	82	119	136	147	182

Table IA.3: Summary Statistics of Mean of Realized Jump Size

This table shows summary statistics of the mean of realized jump size, for all firms and for firms in three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A and B report summary statistics when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) and 2.45 (in LM95) times the instantaneous volatility, respectively. Row All, Large, Medium, and Small are summary statistics for all, large, medium and small firms, respectively.

Panel A: LM99								
	count	mean	std	min	25%	50%	75%	max
All	2,679	−3.27E-03	9.77E-03	−7.23E-02	−9.20E-03	−4.40E-03	1.16E-03	6.02E-02
Large	893	−5.95E-03	5.19E-03	−2.44E-02	−9.34E-03	−5.98E-03	−2.67E-03	1.09E-02
Medium	893	−4.86E-03	8.06E-03	−5.33E-02	−1.00E-02	−5.04E-03	−3.06E-04	3.05E-02
Small	893	9.79E-04	1.29E-02	−7.23E-02	−7.73E-03	−1.37E-04	7.31E-03	6.02E-02
Panel B: LM95								
	count	mean	std	min	25%	50%	75%	max
All	2,679	−1.12E-03	5.28E-03	−3.52E-02	−4.38E-03	−1.78E-03	1.23E-03	3.20E-02
Large	893	−2.60E-03	2.84E-03	−1.29E-02	−4.59E-03	−2.66E-03	−5.84E-04	5.86E-03
Medium	893	−1.90E-03	4.16E-03	−2.81E-02	−4.63E-03	−2.04E-03	5.76E-04	1.75E-02
Small	893	1.15E-03	7.09E-03	−3.52E-02	−3.57E-03	3.25E-04	5.14E-03	3.20E-02

Table IA.4: Summary Statistics of Variance of Realized Jump Size

This table shows summary statistics of the variance of realized jump size, for all firms and for firms in three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A and B report summary statistics when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) and 2.45 (in LM95) times the instantaneous volatility, respectively. Row All, Large, Medium, and Small are summary statistics for all, large, medium and small firms, respectively.

Panel A: LM99								
	count	mean	std	min	25%	50%	75%	max
All	2,679	4.98E-03	5.77E-03	2.48E-06	1.57E-03	3.06E-03	6.23E-03	5.87E-02
Large	893	3.30E-03	3.19E-03	1.95E-04	1.61E-03	2.47E-03	3.97E-03	5.09E-02
Medium	893	5.00E-03	4.74E-03	2.48E-06	1.84E-03	3.57E-03	6.65E-03	3.37E-02
Small	893	6.63E-03	7.86E-03	2.68E-05	1.25E-03	3.94E-03	8.84E-03	5.87E-02
Panel B: LM95								
	count	mean	std	min	25%	50%	75%	max
All	2,679	3.41E-03	3.78E-03	2.00E-06	1.13E-03	2.17E-03	4.24E-03	3.92E-02
Large	893	2.26E-03	1.89E-03	1.33E-04	1.16E-03	1.72E-03	2.71E-03	2.56E-02
Medium	893	3.36E-03	2.99E-03	2.00E-06	1.36E-03	2.45E-03	4.57E-03	1.92E-02
Small	893	4.60E-03	5.26E-03	1.74E-05	9.30E-04	2.87E-03	6.10E-03	3.92E-02

Table IA.5: Summary Statistics of Jump Intensity Component Contributions

This table reports summary statistics of the percentage contribution of three components to the jump intensity, for all firms and for firms in three size groups with the same number of firms in each group, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). The calculation of the percentage contribution of each component to the jump intensity is described in **??**. Panel A, B, and C report summary statistics for the self-exciting, exogenous, and market-exciting component, respectively. Column 2 is the number of firms in each size group. Column 3-9 and column 10-16 report summary statistics when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 (in LM99) and 2.45 (in LM95) times the instantaneous volatility, respectively. Row All, Large, Medium, and Small are summary statistics for all, large, medium and small firms, respectively.

Panel A: Self-exciting component

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	59.45	20.08	2.00	43.48	62.60	77.28	95.50	58.78	19.74	1.40	43.30	61.50	76.40	95.60
Large	893	58.30	20.32	4.00	41.70	60.40	77.00	92.90	56.79	20.54	2.50	39.20	58.20	75.50	94.60
Medium	893	59.03	19.96	6.80	42.35	62.90	76.95	93.00	58.25	19.73	3.70	42.40	60.65	76.90	90.40
Small	893	61.03	19.88	2.00	46.90	64.30	77.80	95.50	61.30	18.65	1.40	48.25	63.20	77.10	95.60

Panel B: Exogenous component

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	19.76	15.64	0.00	7.40	15.23	28.21	89.59	23.27	16.39	0.07	10.03	19.52	33.09	81.94
Large	893	17.00	13.50	0.00	6.37	12.95	24.01	71.92	21.94	15.89	0.13	8.97	17.82	31.62	80.50
Medium	893	20.41	16.58	0.02	7.34	14.97	29.93	78.54	24.04	17.43	0.07	10.42	19.36	33.95	81.94
Small	893	21.88	16.27	0.01	8.81	18.11	31.36	89.59	23.84	15.73	1.12	10.99	20.57	33.98	77.95

Panel C: Market-exciting component

		LM99							LM95						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
All	2,679	18.62	11.05	0.50	10.82	16.34	23.98	78.97	15.54	9.62	0.92	8.90	13.66	19.77	80.55
Large	893	21.91	11.47	2.52	13.97	19.35	27.17	76.97	18.12	9.85	2.11	11.32	15.96	22.57	66.29
Medium	893	18.64	10.28	2.08	11.96	16.56	23.29	78.97	15.49	8.67	1.44	9.77	13.79	19.17	57.66
Small	893	15.30	10.35	0.50	7.85	13.01	20.31	67.50	13.01	9.61	0.92	6.28	10.89	17.20	80.55

Table IA.6: Disagreement, News and daily jump-size distribution

This table reports results of regressions of daily disagreement and news variables on the jump-size distribution, for all firms and for firms in three size groups, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A and B report the result for jump size mean and jump size volatility, respectively. Column All, Large, Medium and Small report the results for all 2,679 firms, 893 large firms, 893 medium firms, and 893 small firms, respectively. Column 2-5 and column 6-9 report the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

Panel A: Mean	LM99				LM95			
	All	Large	Medium	Small	All	Large	Medium	Small
Number of Messages	1.30E-03***	2.00E-04	3.00E-04	3.70E-03***	1.30E-03***	3.00E-04	5.00E-04	3.30E-03***
	(5.00E-04)	(2.00E-04)	(5.00E-04)	(8.00E-04)	(5.00E-04)	(2.00E-04)	(5.00E-04)	(1.00E-03)
Disagreement	−1.50E-03***	−3.00E-03***	−1.50E-03***	7.00E-04**	−1.10E-03***	−2.60E-03***	−1.30E-03***	8.00E-04***
	(2.00E-04)	(2.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(3.00E-04)
Number of News	−6.00E-04***	0.00E+00	−6.00E-04***	−1.20E-03***	−6.00E-04***	−2.00E-04	−5.00E-04***	−8.00E-04***
	(2.00E-04)	(1.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(2.00E-04)
News Tone	1.54E-02***	1.51E-02***	1.71E-02***	1.24E-02***	1.02E-02***	9.70E-03***	1.15E-02***	8.60E-03***
	(3.00E-04)	(4.00E-04)	(5.00E-04)	(6.00E-04)	(2.00E-04)	(3.00E-04)	(3.00E-04)	(4.00E-04)
Size	−2.00E-04	−7.00E-04***	2.00E-04	3.00E-03***	1.00E-04	0.00E+00	5.00E-04**	2.10E-03***
	(3.00E-04)	(3.00E-04)	(4.00E-04)	(7.00E-04)	(2.00E-04)	(1.00E-04)	(2.00E-04)	(5.00E-04)
12-month Momentum	3.90E-03***	7.30E-03***	7.70E-03***	−1.70E-03*	1.10E-03***	3.20E-03***	3.20E-03***	−2.70E-03***
	(5.00E-04)	(7.00E-04)	(7.00E-04)	(9.00E-04)	(3.00E-04)	(4.00E-04)	(3.00E-04)	(6.00E-04)
Idiosyncratic Volatility	6.90E-03***	1.80E-03	1.30E-03	1.05E-02***	4.50E-03***	1.50E-03**	1.60E-03***	6.50E-03***
	(9.00E-04)	(1.20E-03)	(1.20E-03)	(1.90E-03)	(1.00E-03)	(6.00E-04)	(6.00E-04)	(1.80E-03)
N	329,983	105,322	107,437	117,224	687,051	224,038	226,178	236,835
R^2(%)	6.52	8.66	7.89	6.84	3.95	4.99	4.87	4.44

Table IA.6 Continue.

Panel B: Volatility	LM99				LM95			
	All	Large	Medium	Small	All	Large	Medium	Small
Number of Messages	1.20E-03***	4.00E-04*	7.00E-04	2.70E-03***	1.10E-03**	3.00E-04*	7.00E-04*	2.50E-03***
	(4.00E-04)	(2.00E-04)	(4.00E-04)	(7.00E-04)	(4.00E-04)	(2.00E-04)	(4.00E-04)	(8.00E-04)
Disagreement	1.70E-03***	8.00E-04***	1.00E-03***	1.50E-03***	1.30E-03***	7.00E-04***	8.00E-04***	1.30E-03***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)
Number of News	0.00E+00	4.00E-04***	9.00E-04***	8.00E-04***	0.00E+00	3.00E-04***	8.00E-04***	7.00E-04***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(4.70E-05)	(1.00E-04)	(1.00E-04)
∣News Tone∣	8.00E-04***	1.00E-04**	5.00E-04***	1.40E-03***	6.00E-04***	0.00E+00**	4.00E-04***	1.00E-03***
	(1.00E-04)	(2.70E-05)	(1.00E-04)	(1.00E-04)	(3.30E-05)	(1.40E-05)	(3.00E-05)	(1.00E-04)
Size	−1.00E-04	−2.00E-04	−7.00E-04***	−2.00E-03***	−1.00E-04	−1.00E-04	−5.00E-04***	−1.50E-03***
	(3.00E-04)	(2.00E-04)	(1.00E-04)	(6.00E-04)	(2.00E-04)	(1.00E-04)	(1.00E-04)	(4.00E-04)
12-month Momentum	−1.70E-03***	−2.00E-03***	−2.70E-03***	−1.60E-03***	−1.10E-03***	−1.40E-03***	−1.80E-03***	−9.00E-04***
	(2.00E-04)	(2.00E-04)	(2.00E-04)	(4.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)
Idiosyncratic Volatility	9.60E-03***	5.30E-03***	8.50E-03***	1.22E-02***	6.00E-03***	3.80E-03***	5.60E-03***	7.40E-03***
	(1.20E-03)	(1.00E-03)	(4.00E-04)	(1.70E-03)	(1.30E-03)	(6.00E-04)	(4.00E-04)	(1.80E-03)
N	329,983	105,322	107,437	117,224	687,051	224,038	226,178	236,835
$R^2(\%)$	27.75	31.77	31.44	34.65	19.39	25.71	25.05	25.98

Table IA.7: Disagreement sources, News and daily jump-size distribution

This table reports results of regressions of daily disagreement and news variables on the jump-size distribution, for all firms and for firms in three size groups, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A and B report the result for jump size mean and jump size volatility, respectively. Column All, Large, Medium and Small report the results for all 2,679 firms, 893 large firms, 893 medium firms, and 893 small firms, respectively. Column 2-5 and column 6-9 report the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

Panel A: Mean	Panel A: LM99				LM95			
	All	Large	Medium	Small	All	Large	Medium	Small
Number of Messages	1.30E-03***	2.00E-04	3.00E-04	3.60E-03***	1.30E-03***	3.00E-04	5.00E-04	3.30E-03***
	(5.00E-04)	(2.00E-04)	(5.00E-04)	(9.00E-04)	(5.00E-04)	(2.00E-04)	(5.00E-04)	(1.00E-03)
Disagreement Within-Group	−8.00E-04***	−2.00E-03***	−8.00E-04***	5.00E-04	−6.00E-04***	−1.70E-03***	−7.00E-04***	7.00E-04**
	(2.00E-04)	(2.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(3.00E-04)
Disagreement Cross-Group	−4.00E-04***	−7.00E-04***	−5.00E-04**	3.00E-04	−3.00E-04***	−7.00E-04***	−4.00E-04**	3.00E-04
	(1.00E-04)	(2.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(2.00E-04)
Number of News	−7.00E-04***	−1.00E-04	−6.00E-04***	−1.20E-03***	−7.00E-04***	−2.00E-04	−6.00E-04***	−8.00E-04***
	(2.00E-04)	(1.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(2.00E-04)
News Tone	1.54E-02***	1.51E-02***	1.70E-02***	1.24E-02***	1.02E-02***	9.70E-03***	1.15E-02***	8.60E-03***
	(3.00E-04)	(4.00E-04)	(5.00E-04)	(6.00E-04)	(2.00E-04)	(3.00E-04)	(3.00E-04)	(4.00E-04)
Size	−2.00E-04	−7.00E-04***	2.00E-04	3.00E-03***	0.00E+00	−1.00E-04	4.00E-04**	2.10E-03***
	(3.00E-04)	(3.00E-04)	(4.00E-04)	(7.00E-04)	(2.00E-04)	(1.00E-04)	(2.00E-04)	(5.00E-04)
12-month Momentum	4.00E-03***	7.40E-03***	7.70E-03***	−1.70E-03*	1.20E-03***	3.30E-03***	3.20E-03***	−2.70E-03***
	(5.00E-04)	(7.00E-04)	(7.00E-04)	(9.00E-04)	(3.00E-04)	(4.00E-04)	(3.00E-04)	(6.00E-04)
Idiosyncratic Volatility	6.90E-03***	1.80E-03	1.30E-03	1.05E-02***	4.50E-03***	1.50E-03**	1.60E-03***	6.50E-03***
	(9.00E-04)	(1.20E-03)	(1.20E-03)	(1.90E-03)	(1.00E-03)	(7.00E-04)	(6.00E-04)	(1.80E-03)
N	329,983	105,322	107,437	117,224	687,051	224,038	226,178	236,835
R^2(%)	6.49	8.54	7.85	6.84	3.93	4.88	4.84	4.43

Table IA.7 Continue.

Panel B: Volatility	LM99				LM95			
	All	Large	Medium	Small	All	Large	Medium	Small
Number of Messages	1.20E-03***	3.00E-04*	6.00E-04	2.60E-03***	1.10E-03**	3.00E-04*	7.00E-04*	2.50E-03***
	(4.00E-04)	(2.00E-04)	(4.00E-04)	(7.00E-04)	(4.00E-04)	(2.00E-04)	(4.00E-04)	(8.00E-04)
Disagreement Within-Group	1.40E-03***	8.00E-04***	8.00E-04***	1.30E-03***	1.10E-03***	6.00E-04***	7.00E-04***	1.10E-03***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)
Disagreement Cross-Group	6.00E-04***	3.00E-04***	3.00E-04***	7.00E-04***	5.00E-04***	2.00E-04***	3.00E-04***	6.00E-04***
	(4.30E-05)	(3.60E-05)	(1.00E-04)	(1.00E-04)	(3.50E-05)	(2.30E-05)	(4.20E-05)	(1.00E-04)
Number of News	0.00E+00	3.00E-04***	9.00E-04***	8.00E-04***	0.00E+00	3.00E-04***	8.00E-04***	7.00E-04***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(4.70E-05)	(1.00E-04)	(1.00E-04)
| News Tone |	9.00E-04***	1.00E-04**	5.00E-04***	1.40E-03***	6.00E-04***	0.00E+00**	4.00E-04***	1.00E-03***
	(1.00E-04)	(2.70E-05)	(1.00E-04)	(1.00E-04)	(3.30E-05)	(1.40E-05)	(3.00E-05)	(1.00E-04)
Size	−1.00E-04	−2.00E-04	−7.00E-04***	−2.00E-03***	−1.00E-04	−1.00E-04	−5.00E-04***	−1.50E-03***
	(3.00E-04)	(2.00E-04)	(1.00E-04)	(6.00E-04)	(2.00E-04)	(1.00E-04)	(1.00E-04)	(4.00E-04)
12-month Momentum	−1.70E-03***	−2.00E-03***	−2.70E-03***	−1.60E-03***	−1.10E-03***	−1.40E-03***	−1.80E-03***	−9.00E-04***
	(2.00E-04)	(2.00E-04)	(2.00E-04)	(4.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)
Idiosyncratic Volatility	9.60E-03***	5.20E-03***	8.50E-03***	1.22E-02***	6.00E-03***	3.80E-03***	5.60E-03***	7.40E-03***
	(1.20E-03)	(1.00E-03)	(4.00E-04)	(1.70E-03)	(1.30E-03)	(6.00E-04)	(4.00E-04)	(1.80E-03)
N	329,983	105,322	107,437	117,224	687,051	224,038	226,178	236,835
R^2(%)	27.85	31.84	31.49	34.77	19.48	25.75	25.09	26.11

Table IA.8: Disagreement and sophistication, News and daily jump-size distribution

This table reports results of regressions of daily disagreement and news variables on the jump-size distribution, for all firms and for firms in three size groups, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A and B report the result for jump size mean and jump size volatility, respectively. Column All, Large, Medium and Small report the results for all 2,679 firms, 893 large firms, 893 medium firms, and 893 small firms, respectively. Column 2-5 and column 6-9 report the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

Panel A: Mean	LM99				LM95			
	All	Large	Medium	Small	All	Large	Medium	Small
Number of Messages (S)	8.00E-04	−1.70E-03***	1.00E-04	5.20E-03***	6.00E-04	−1.40E-03***	2.00E-04	4.80E-03***
	(5.00E-04)	(5.00E-04)	(5.00E-04)	(8.00E-04)	(5.00E-04)	(4.00E-04)	(4.00E-04)	(7.00E-04)
Number of Messages (U)	7.00E-04*	1.50E-03***	3.00E-04	−1.50E-03*	7.00E-04*	1.30E-03***	4.00E-04	−1.30E-03
	(4.00E-04)	(5.00E-04)	(4.00E-04)	(8.00E-04)	(4.00E-04)	(5.00E-04)	(4.00E-04)	(8.00E-04)
Disagreement (S)	3.00E-04	−4.00E-04*	−2.00E-04	9.00E-04***	2.00E-04*	−3.00E-04**	−1.00E-04	8.00E-04***
	(2.00E-04)	(2.00E-04)	(3.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(3.00E-04)
Disagreement (U)	−2.00E-04	−8.00E-04***	−2.00E-04	5.00E-04**	−2.00E-04**	−9.00E-04***	−2.00E-04	5.00E-04**
	(1.00E-04)	(2.00E-04)	(2.00E-04)	(2.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(2.00E-04)
| Disagreement (S-U) |	−5.70E-03***	−5.50E-03***	−6.10E-03***	−4.80E-03***	−5.00E-03***	−5.10E-03***	−5.80E-03***	−3.80E-03***
	(1.00E-04)	(2.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)
Number of News	−9.00E-04***	1.00E-04	−7.00E-04***	−1.80E-03***	−8.00E-04***	0.00E+00	−7.00E-04***	−1.30E-03***
	(2.00E-04)	(1.00E-04)	(3.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(2.00E-04)
News Tone	1.51E-02***	1.48E-02***	1.67E-02***	1.21E-02***	1.00E-02***	9.50E-03***	1.11E-02***	8.40E-03***
	(3.00E-04)	(4.00E-04)	(5.00E-04)	(5.00E-04)	(2.00E-04)	(2.00E-04)	(3.00E-04)	(4.00E-04)
Size	−5.00E-04	−1.00E-03***	−9.00E-04**	2.60E-03***	−2.00E-04	−3.00E-04**	−7.00E-04***	1.70E-03***
	(3.00E-04)	(3.00E-04)	(4.00E-04)	(6.00E-04)	(2.00E-04)	(1.00E-04)	(2.00E-04)	(4.00E-04)
12-month Momentum	3.40E-03***	6.70E-03***	6.90E-03***	−2.30E-03**	7.00E-04**	2.70E-03***	2.60E-03***	−3.10E-03***
	(5.00E-04)	(7.00E-04)	(7.00E-04)	(1.00E-03)	(3.00E-04)	(4.00E-04)	(3.00E-04)	(6.00E-04)
Idiosyncratic Volatility	6.90E-03***	1.80E-03	1.10E-03	9.70E-03***	4.50E-03***	1.50E-03**	1.40E-03**	6.00E-03***
	(9.00E-04)	(1.20E-03)	(1.20E-03)	(1.80E-03)	(1.00E-03)	(7.00E-04)	(6.00E-04)	(1.80E-03)
N	329,983	105,322	107,437	117,224	687,051	224,038	226,178	236,835
R^2(%)	7.14	9.61	8.74	7.74	4.68	6.14	6.01	5.22

Table IA.8 Continue.

Panel B: Volatility	LM99				LM95			
	All	Large	Medium	Small	All	Large	Medium	Small
Number of Messages (S)	1.30E-03***	5.00E-04**	1.00E-03***	3.00E-03***	1.10E-03**	5.00E-04**	1.10E-03***	3.00E-03***
	(5.00E-04)	(2.00E-04)	(2.00E-04)	(6.00E-04)	(5.00E-04)	(2.00E-04)	(1.00E-04)	(5.00E-04)
Number of Messages (U)	0.00E+00	−1.00E-04	−2.00E-04	−3.00E-04	1.00E-04	−1.00E-04	−2.00E-04	−4.00E-04
	(2.00E-04)	(1.00E-04)	(3.00E-04)	(7.00E-04)	(2.00E-04)	(1.00E-04)	(3.00E-04)	(7.00E-04)
Disagreement (S)	1.20E-03***	6.00E-04***	6.00E-04***	9.00E-04***	9.00E-04***	5.00E-04***	4.00E-04***	7.00E-04***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)
Disagreement (U)	9.00E-04***	5.00E-04***	6.00E-04***	9.00E-04***	8.00E-04***	4.00E-04***	6.00E-04***	9.00E-04***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)
∣Disagreement (S-U)∣	4.00E-04***	0.00E+00	1.00E-04***	7.00E-04***	3.00E-04***	1.00E-04***	2.00E-04***	5.00E-04***
	(4.10E-05)	(3.90E-05)	(4.80E-05)	(1.00E-04)	(2.40E-05)	(2.20E-05)	(2.70E-05)	(4.30E-05)
Number of News	−2.00E-04**	2.00E-04***	7.00E-04***	6.00E-04***	−2.00E-04*	2.00E-04***	6.00E-04***	5.00E-04***
	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(4.20E-05)	(1.00E-04)	(1.00E-04)
∣News Tone∣	9.00E-04***	1.00E-04**	5.00E-04***	1.30E-03***	6.00E-04***	0.00E+00**	4.00E-04***	9.00E-04***
	(1.00E-04)	(2.70E-05)	(5.00E-05)	(1.00E-04)	(3.30E-05)	(1.40E-05)	(2.80E-05)	(1.00E-04)
Size	−1.00E-04	−2.00E-04	−8.00E-04***	−1.80E-03***	−1.00E-04	−1.00E-04	−6.00E-04***	−1.40E-03***
	(3.00E-04)	(2.00E-04)	(1.00E-04)	(4.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(3.00E-04)
12-month Momentum	−1.70E-03***	−2.00E-03***	−2.70E-03***	−1.70E-03***	−1.10E-03***	−1.40E-03***	−1.80E-03***	−1.00E-03***
	(2.00E-04)	(2.00E-04)	(2.00E-04)	(3.00E-04)	(1.00E-04)	(1.00E-04)	(1.00E-04)	(2.00E-04)
Idiosyncratic Volatility	9.50E-03***	5.20E-03***	8.40E-03***	1.17E-02***	6.00E-03***	3.80E-03***	5.50E-03***	7.00E-03***
	(1.20E-03)	(9.00E-04)	(4.00E-04)	(1.70E-03)	(1.30E-03)	(6.00E-04)	(4.00E-04)	(1.80E-03)
N	329,983	105,322	107,437	117,224	687,051	224,038	226,178	236,835
R^2(%)	28.32	32.12	32.34	36.41	19.92	26.02	26.04	28.10

Table IA.9: Disagreement, News and estimated jump intensity for three size groups

This table reports results of panel regressions of daily disagreement and news variables on the estimated intensity of various jumps for firms in three size groups, sorted by the average of their daily market capitalization during the sample period (June 01, 2010 - August 31, 2021). Panel A, B, C, D, and E report the results on the estimated intensity of all, self-exciting, market-exciting, negative, and positive jumps, respectively. Column Large, Medium, and Small report the results for 893 large firms, 893 medium firms, and 893 small firms, respectively. Column 2-5 and column 6-9 report the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include the firm fixed effect and a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

Panel A: All	LM99			LM95		
	Large	Medium	Small	Large	Medium	Small
Number of Messages	0.279**	0.383***	0.347***	0.361**	0.511***	0.478***
	(0.140)	(0.046)	(0.081)	(0.167)	(0.068)	(0.109)
Disagreement	0.720***	0.525***	0.526***	0.942***	0.709***	0.703***
	(0.019)	(0.020)	(0.022)	(0.023)	(0.025)	(0.029)
Number of News	0.049***	0.002	0.009	0.058***	0.004	0.031*
	(0.015)	(0.013)	(0.012)	(0.022)	(0.017)	(0.018)
$\lvert$News Tone$\rvert$	0.038***	0.255***	0.257***	0.061***	0.318***	0.346***
	(0.011)	(0.011)	(0.013)	(0.015)	(0.016)	(0.018)
Size	0.014	−0.017	−0.209***	−0.093	−0.273***	−0.414***
	(0.045)	(0.026)	(0.034)	(0.100)	(0.032)	(0.047)
12-month Momentum	−1.704***	−1.185***	−0.257***	−2.958***	−2.130***	−0.681***
	(0.059)	(0.056)	(0.059)	(0.100)	(0.089)	(0.093)
Idiosyncratic Volatility	0.176***	0.173***	0.077**	0.122***	−0.038	−0.019
	(0.042)	(0.052)	(0.034)	(0.025)	(0.044)	(0.034)
N	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762
R^2(%)	1.95	1.19	0.28	2.48	1.49	0.32

Table IA.9 Continue.

Panel B: SE	LM99			LM95		
	Large	Medium	Small	Large	Medium	Small
Number of Messages	0.266**	0.365***	0.333***	0.345**	0.474***	0.434***
	(0.135)	(0.044)	(0.077)	(0.158)	(0.066)	(0.102)
Disagreement	0.674***	0.480***	0.481***	0.915***	0.658***	0.653***
	(0.018)	(0.019)	(0.021)	(0.023)	(0.025)	(0.027)
Number of News	0.070***	0.021*	0.017	0.113***	0.066***	0.063***
	(0.015)	(0.012)	(0.011)	(0.020)	(0.016)	(0.017)
∣News Tone∣	0.042***	0.247***	0.240***	0.054***	0.313***	0.323***
	(0.011)	(0.011)	(0.013)	(0.014)	(0.015)	(0.017)
Size	0.021	−0.013	−0.220***	0.032	−0.095***	−0.323***
	(0.043)	(0.025)	(0.034)	(0.067)	(0.029)	(0.045)
12-month Momentum	−1.292***	−0.887***	−0.080	−2.047***	−1.453***	−0.273***
	(0.048)	(0.048)	(0.051)	(0.075)	(0.068)	(0.073)
Idiosyncratic Volatility	0.183***	0.189***	0.071**	0.115***	−0.016	−0.055
	(0.048)	(0.054)	(0.035)	(0.024)	(0.042)	(0.036)
N	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762
$R^2(\%)$	1.45	0.90	0.25	1.68	0.97	0.23

Panel C: ME	LM99			LM95		
	Large	Medium	Small	Large	Medium	Small
Number of Messages	0.013**	0.018***	0.015***	0.016	0.036***	0.044***
	(0.005)	(0.003)	(0.005)	(0.011)	(0.007)	(0.010)
Disagreement	0.045***	0.045***	0.045***	0.027***	0.051***	0.050***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)
Number of News	−0.021***	−0.019***	−0.008***	−0.055***	−0.061***	−0.032***
	(0.004)	(0.003)	(0.002)	(0.010)	(0.007)	(0.004)
∣News Tone∣	−0.004	0.008***	0.018***	0.007	0.005	0.023***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.003)
Size	−0.006	−0.004	0.011***	−0.126**	−0.178***	−0.091***
	(0.005)	(0.003)	(0.004)	(0.055)	(0.012)	(0.010)
12-month Momentum	−0.411***	−0.298***	−0.177***	−0.912***	−0.677***	−0.408***
	(0.015)	(0.012)	(0.012)	(0.032)	(0.025)	(0.026)
Idiosyncratic Volatility	−0.007	−0.016***	0.006	0.008	−0.023***	0.036***
	(0.008)	(0.004)	(0.006)	(0.007)	(0.008)	(0.007)
N	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762
$R^2(\%)$	1.54	1.15	0.44	2.18	1.75	0.70

Table IA.9 Continue.

Panel D: Neg	LM99			LM95		
	Large	Medium	Small	Large	Medium	Small
Number of Messages	0.137**	0.176***	0.227***	0.159**	0.265***	0.393***
	(0.061)	(0.017)	(0.061)	(0.077)	(0.030)	(0.089)
Disagreement	0.576***	0.367***	0.338***	0.795***	0.509***	0.439***
	(0.015)	(0.013)	(0.016)	(0.018)	(0.017)	(0.021)
Number of News	0.068***	0.049***	0.031***	0.070***	0.069***	0.041***
	(0.016)	(0.009)	(0.008)	(0.016)	(0.011)	(0.012)
News Tone	−0.121***	−0.095***	−0.079***	−0.200***	−0.189***	−0.165***
	(0.008)	(0.008)	(0.009)	(0.011)	(0.011)	(0.013)
Size	0.107***	0.170***	−0.023	0.011	0.001	−0.158***
	(0.041)	(0.019)	(0.028)	(0.050)	(0.021)	(0.035)
12-month Momentum	−2.937***	−2.611***	−2.126***	−5.044***	−4.478***	−3.566***
	(0.069)	(0.064)	(0.092)	(0.119)	(0.107)	(0.154)
Idiosyncratic Volatility	0.189***	0.183***	0.223***	0.213***	0.124***	0.320***
	(0.045)	(0.038)	(0.053)	(0.032)	(0.032)	(0.060)
N	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762
$R^2(\%)$	8.85	7.73	4.45	12.10	10.36	5.94
Panel E: Pos	**LM99**			**LM95**		
	Large	Medium	Small	Large	Medium	Small
Number of Messages	0.142*	0.209***	0.125***	0.202**	0.248***	0.091
	(0.080)	(0.036)	(0.047)	(0.091)	(0.051)	(0.055)
Disagreement	0.144***	0.165***	0.199***	0.146***	0.209***	0.280***
	(0.010)	(0.011)	(0.012)	(0.014)	(0.015)	(0.018)
Number of News	−0.019	−0.042***	−0.007	−0.013	−0.058***	0.010
	(0.016)	(0.007)	(0.007)	(0.021)	(0.013)	(0.015)
News Tone	0.096***	0.088***	0.099***	0.155***	0.167***	0.184***
	(0.007)	(0.007)	(0.008)	(0.011)	(0.010)	(0.011)
Size	−0.092**	−0.170***	−0.172***	−0.102*	−0.252***	−0.235***
	(0.046)	(0.017)	(0.022)	(0.056)	(0.024)	(0.029)
12-month Momentum	1.236***	1.430***	1.871***	2.088***	2.354***	2.887***
	(0.024)	(0.031)	(0.067)	(0.037)	(0.045)	(0.099)
Idiosyncratic Volatility	−0.014	−0.009	−0.143***	−0.092***	−0.162***	−0.335***
	(0.015)	(0.026)	(0.033)	(0.031)	(0.037)	(0.078)
N	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762	2,530,762
$R^2(\%)$	2.17	3.23	4.08	2.57	3.54	4.38

Table IA.10: Disagreement from all tweets, News and estimated jump intensity for all firms

This table reports results of panel regressions of daily disagreement and news variables on the estimated intensity of various jumps for all firms during the sample period (June 01, 2010 - August 31, 2021). Column All, SE, ME, Neg, and Pos report the regressions results on the estimated intensity of all, self-exciting, market-exciting, negative, and positive jumps, respectively. Panel A and B reports the results when detecting a daily return as a jump if the absolute value of that return is larger than 3.22 and 2.45 times the instantaneous volatility, respectively. All independent variables are standardized to mean 0 and standard deviation 1. All regressions include the firm fixed effect and a constant term that is not reported for saving space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm and are reported in parentheses.

	Panel A: LM99					Panel B: LM95				
	All	SE	ME	Neg	Pos	All	SE	ME	Neg	Pos
Number of Messages (all)	0.327***	0.308***	0.019***	0.191***	0.135***	0.420***	0.383***	0.037***	0.285***	0.135***
	(0.068)	(0.065)	(0.004)	(0.045)	(0.029)	(0.084)	(0.077)	(0.009)	(0.065)	(0.032)
Disagreement (all)	0.699***	0.655***	0.044***	0.528***	0.175***	0.879***	0.877***	0.002	0.690***	0.195***
	(0.011)	(0.011)	(0.002)	(0.008)	(0.006)	(0.014)	(0.014)	(0.003)	(0.011)	(0.009)
Number of News	0.011	0.028	−0.017***	0.034*	−0.024***	0.015	0.063**	−0.048***	0.025	−0.011
	(0.023)	(0.023)	(0.002)	(0.019)	(0.008)	(0.028)	(0.029)	(0.005)	(0.025)	(0.011)
News Tone				−0.115***	0.089***				−0.213***	0.161***
				(0.005)	(0.004)				(0.007)	(0.006)
| News Tone |	0.179***	0.172***	0.006***			0.231***	0.224***	0.007***		
	(0.007)	(0.007)	(0.001)			(0.010)	(0.009)	(0.002)		
Size	−0.092***	−0.081***	−0.011***	−0.001	−0.090***	−0.194***	−0.103***	−0.091**	−0.095**	−0.097***
	(0.029)	(0.025)	(0.004)	(0.018)	(0.030)	(0.069)	(0.038)	(0.036)	(0.039)	(0.033)
12-month Momentum	−0.944***	−0.665***	−0.279***	−2.475***	1.535***	−1.764***	−1.131***	−0.633***	−4.222***	2.465***
	(0.044)	(0.037)	(0.009)	(0.057)	(0.028)	(0.071)	(0.054)	(0.020)	(0.098)	(0.043)
Idiosyncratic Volatility	0.171***	0.158***	0.012***	0.234***	−0.062***	0.136***	0.086***	0.050***	0.338***	−0.201***
	(0.040)	(0.041)	(0.004)	(0.046)	(0.018)	(0.019)	(0.014)	(0.008)	(0.055)	(0.043)
N	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286	7,592,286
R^2(%)	0.79	0.58	0.89	6.50	3.18	0.99	0.64	1.32	8.73	3.50

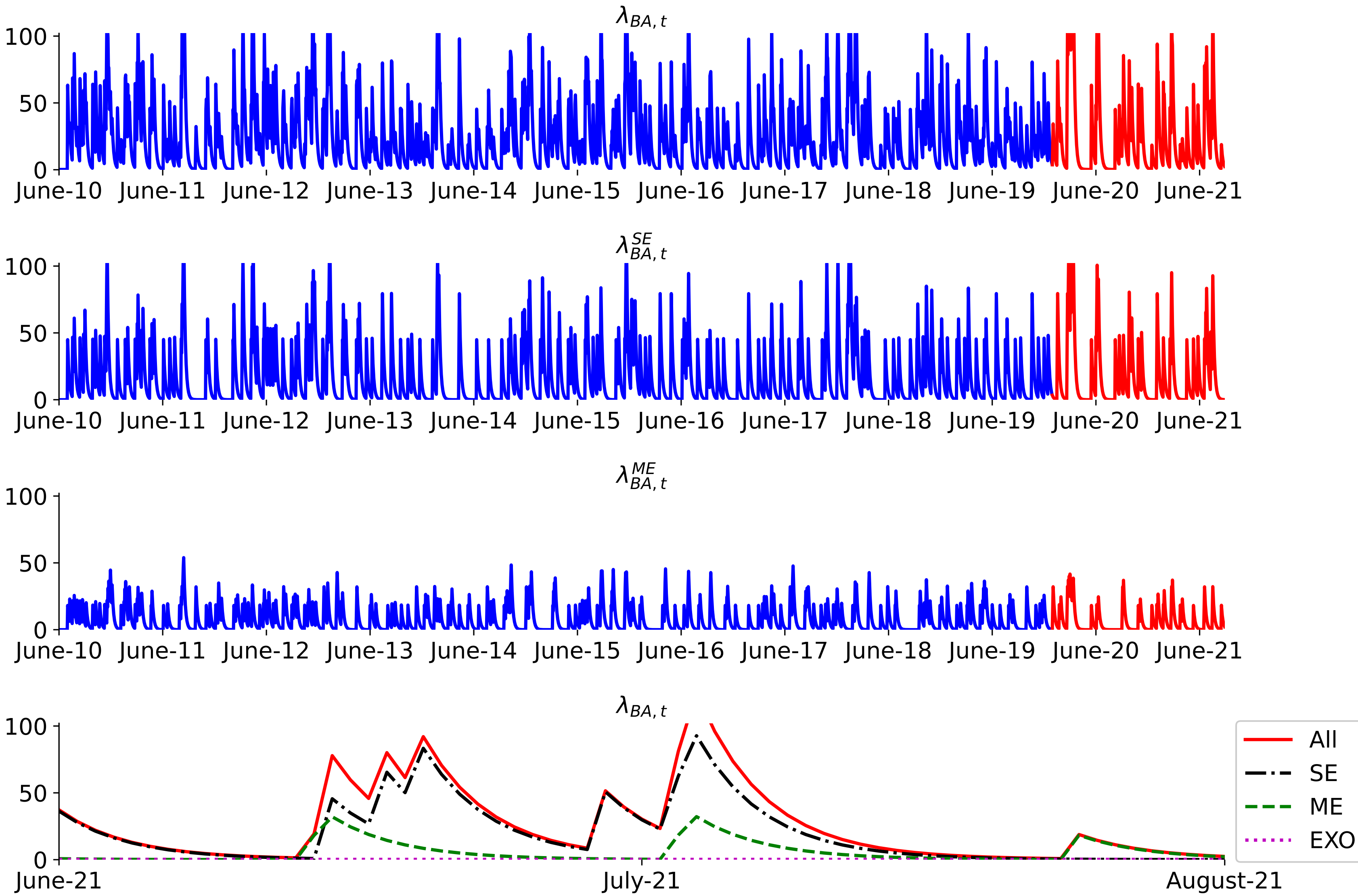


Figure IA.1: This figure shows estimated jump intensity of BA in LM95

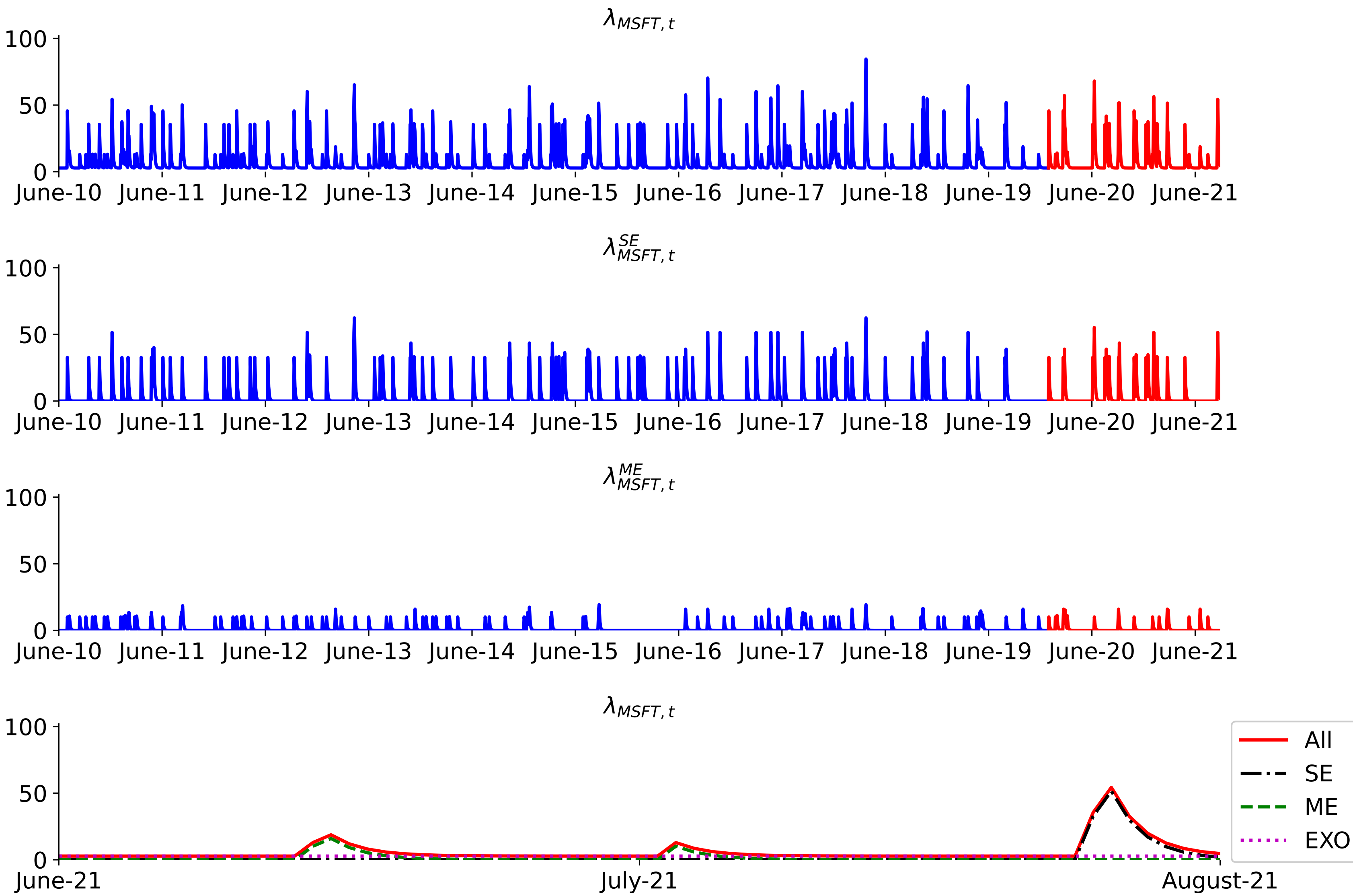


Figure IA.2: This figure shows estimated jump intensity of MSFT in LM99

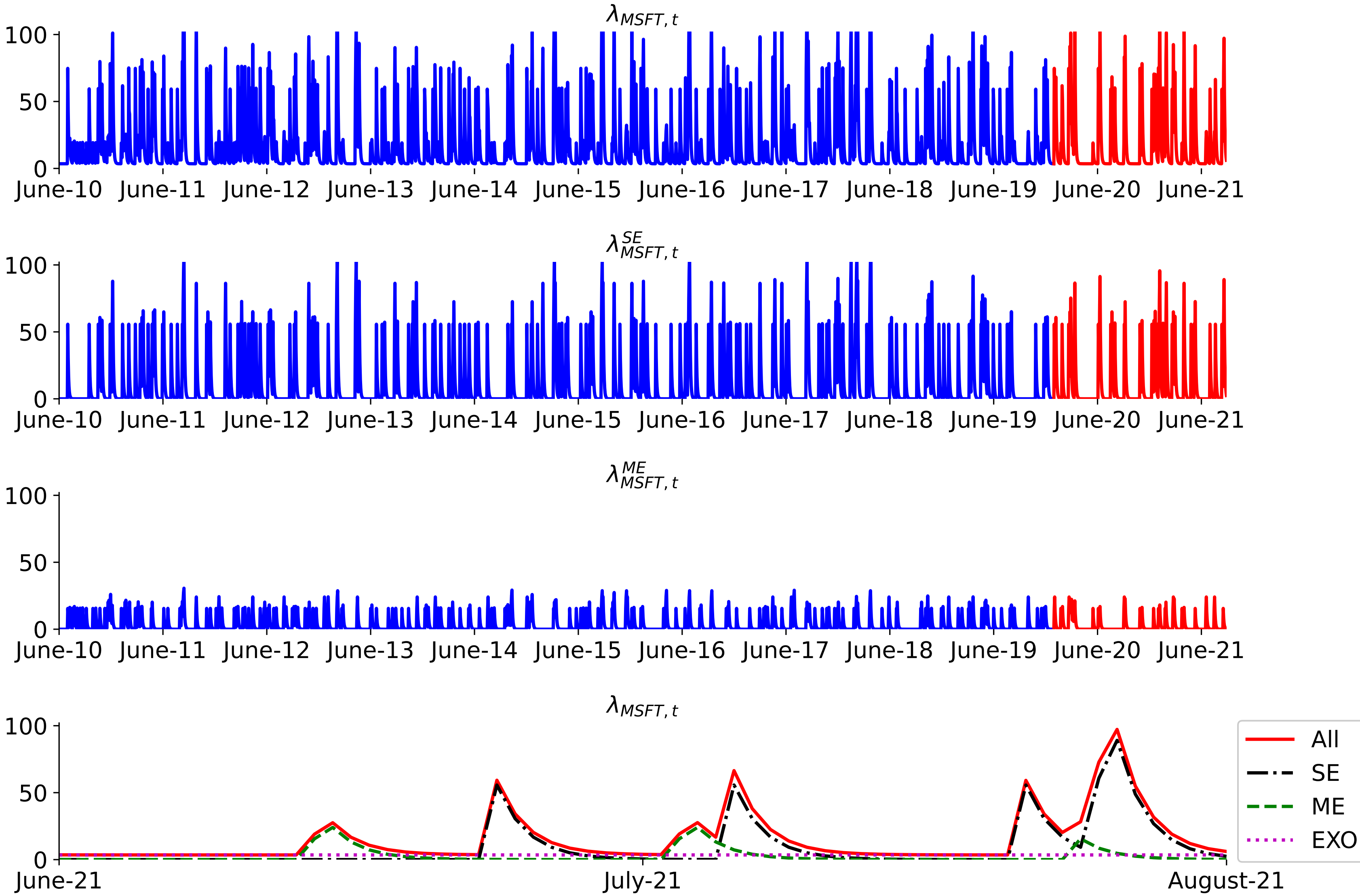


Figure IA.3: This figure shows estimated jump intensity of MSFT in LM95

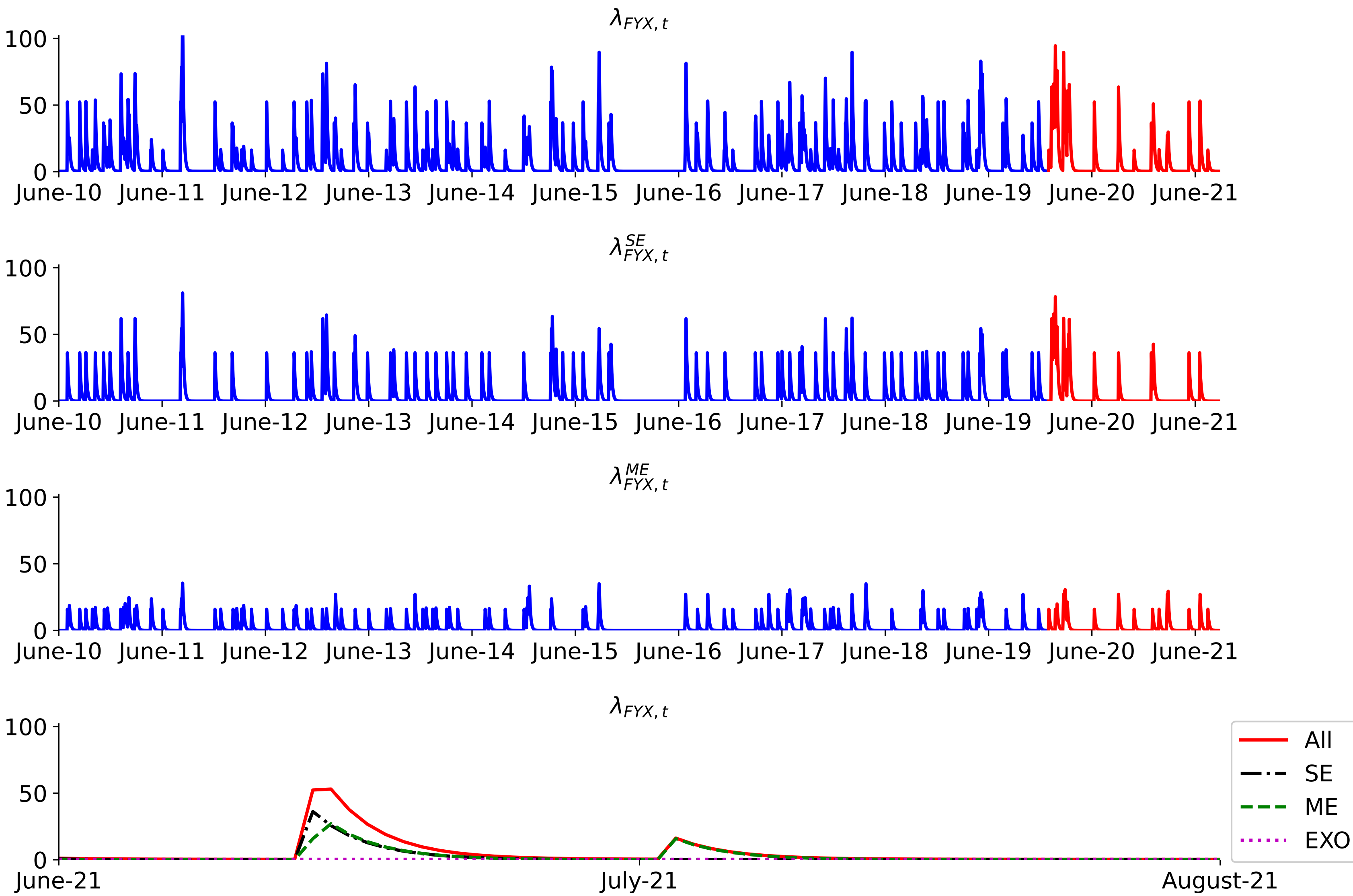


Figure IA.4: This figure shows estimated jump intensity of FYX in LM99

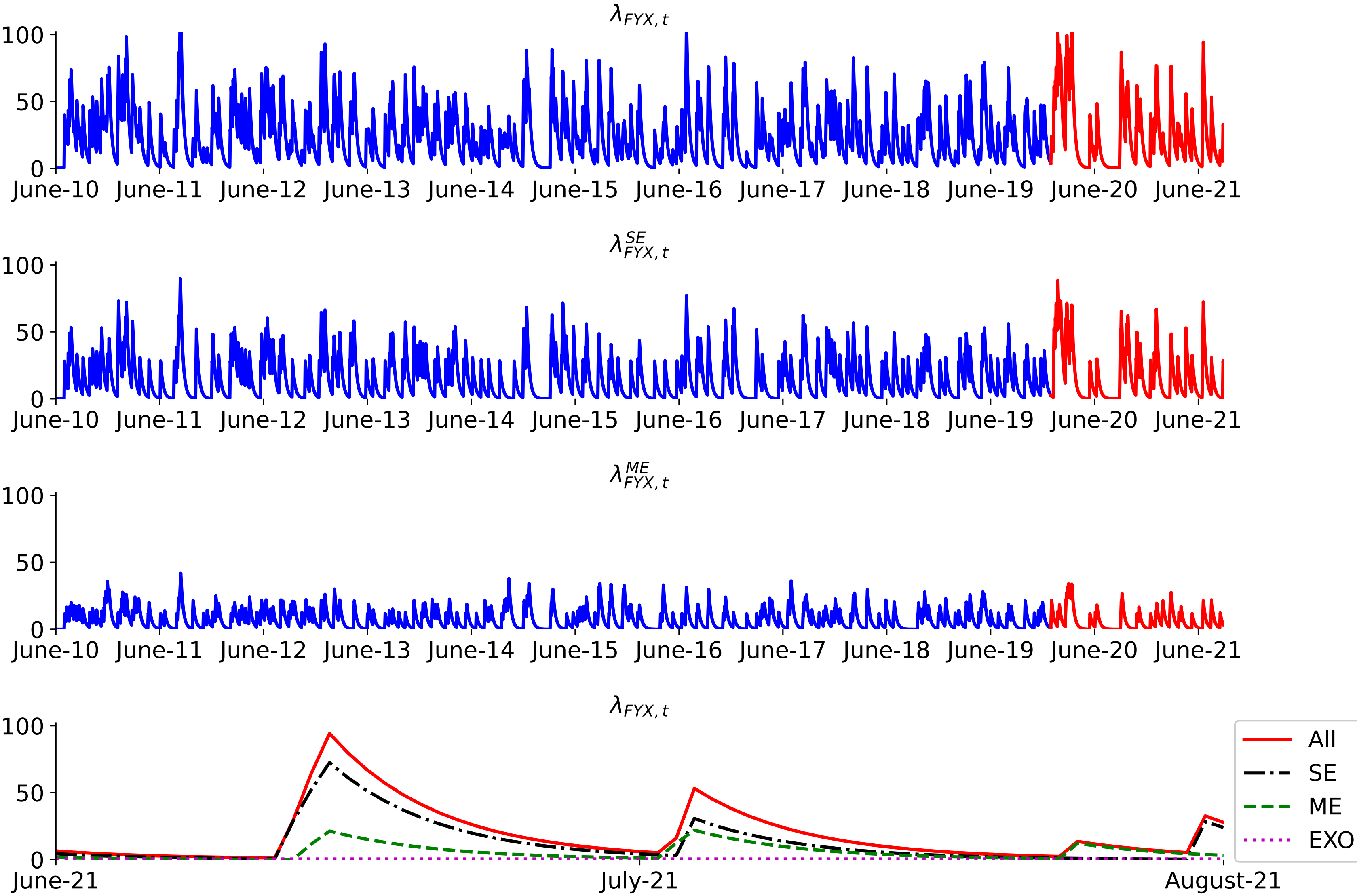


Figure IA.5: This figure shows estimated jump intensity of FYX in LM95