Received: 23 March 2020 | Accepted: 5 October 2022
DOI: 10.1111/jfir.12310

ORIGINAL ARTICLE

The Journal of Financial Research

Social media and price discovery: The case of cross-listed firms

Rui Fan[1] | Oleksandr Talavera[2] | Vu Tran[3]

[1]School of Management, Swansea University, Swansea, United Kingdom

[2]Birmingham Business School, University of Birmingham, Birmingham, United Kingdom

[3]ICMA Centre, Henley Business School, University of Reading, United Kingdom

Correspondence
Vu Tran, ICMA Centre, Henley Business School, University of Reading, Whiteknights, Reading, RG6 6BA, UK.
Email: v.tran@icmacentre.ac.uk

Abstract

In this article, we examine whether social media information affects the price-discovery process for cross-listed companies. Using over 29 million overnight tweets mentioning cross-listed companies, we examine the role of social media for a link between the last periods of trading in the US markets and the first periods in the UK market. Our estimates suggest that the size and content of information flows on social networks support the price-discovery process. The interactions between lagged US stock features and overnight tweets significantly affect stock returns and volatility of cross-listed stocks when the UK market opens. These effects weaken and disappear 1 to 3 hr after the opening of the UK market. We also develop a profitable trading strategy based on overnight social media, and the profits remain economically significant after considering transaction costs.

JEL CLASSIFICATION
G12, G14, L86

1 | INTRODUCTION

Information flows and investors' decisions to trade can affect the process by which new information is incorporated in stock prices. Previous research has examined whether information flows during the trading period can affect the price-discovery process (e.g., Frijns et al., 2015), trading patterns (e.g., Lou et al., 2019), and return patterns (e.g., Gao et al., 2018; Renault, 2017). However, news also arrives outside of regular trading hours (e.g., Barclay &

Hendershott, 2003). With the increasing popularity of social media, this high-speed stream of information becomes even more important than before, as it could have a potentially continuous effect on investors and analysts. We investigate whether social media flows during nontrading hours contribute to the price-discovery process.

We exploit British dual-listed stocks to address this research question for several reasons. First, United States has the largest number of Twitter users as of October 2021, followed by Japan, India, Brazil, and the United Kingdom.[1] Second, British firms cross-listed in the United States are likely to attract more attention from US investors and analysts (see, e.g., Frijns et al., 2015; Werner & Kleidon, 1996). Therefore, these dual-listed UK firms have the most relevant Twitter feeds in the United States. Third, lagged stock indicators in the United States should offer a good proxy of informative signals for opening prices on the London Stock Exchange (LSE). Interactions between overnight social media activity and lagged US stock indicators provide an intuitive direction to tackle the debate regarding the informativeness of social media information.

Using a unique data set of social media messages related to British companies that are dual-listed in the UK and US markets, we find that overnight information on social media affects opening stock indicators on the LSE. The volume, aggregate positiveness, and agreement of tweets generate the impact by overnight social media information. For example, the sentiment embedded in overnight tweets can be used to predict returns during the first 1 to 3 hr of the next trading day. These results suggest that overnight tweets contain informative signals and support the process through which information is incorporated into prices. In addition, the effects of overnight tweets are more pronounced if there were significant changes in the corresponding indicators of their US counterparts during the previous day's last trading periods. This may indicate that there were increasing levels of interest on overnight social media relating to firms that have recently experienced shocks. Furthermore, we develop a trading strategy based on overnight tweets, and our trading strategy outperforms two alternative benchmarks and remains profitable after deducting transaction costs.

Our article is related to several strands of the literature. First, many studies examine the effect of information on discovering the opening price in financial markets (see, e.g., Barclay & Hendershott, 2003).[2] Corporate event news, in particular, plays an important role in influencing the price-formation process in financial markets (see, e.g., Baruch et al., 2017). This type of information on company news and trading activities is frequently mentioned on social media. Therefore, we contribute to the literature by investigating whether the stock features of dual-listed UK stocks in the United States could assist the price discovery of cross-listed UK stocks in the United Kingdom through social media information flows during the overnight period.

Moreover, our article is linked to the literature on risk discovery during intraday trading. The first 30 min of trading receives the most attention because most macroeconomic news and company earnings are announced before stock markets open. Markets then start at different levels from the previous day's closing prices to reflect the new information. It typically takes around 30 min for stock prices to adjust. The plots of intraday trading volatility often exhibit U-shapes, and these patterns have been documented in the literature about intraday asset prices (see, e.g., Bogousslavsky, 2016; Gao et al., 2018; Heston et al., 2010; Murphy & Thirumalai, 2017). Furthermore, some studies emphasize the importance of last-hour trading (see, e.g., Cushing & Madhavan, 2000) as institutional investors calculate mutual fund values and portfolio returns using closing prices. Additionally, dealers typically sell in the last 30 min to eliminate the risks associated with holding stocks overnight. Motivated by these studies, we examine whether stock indicators during the last trading period in one market affect the opening stock features in another market in the case of dual-listed stocks.

Additionally, research focuses on overnight information and price discovery. For example, some studies find that corporate news announced overnight can affect price discovery in the opening period of the next

[1] https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

[2] Other papers have investigated this issue. For example, Biais et al. (1999) examine learning and price discovery on the Paris Bourse opening. Flood et al. (1999) study the influence of price disclosure on market performance in an experimental market. Madhavan and Panchapagesan (2000) explore the price-discovery process at the New York Stock Exchange (NYSE) opening.

day (e.g., Berkman & Truong, 2009; Doyle & Magilke, 2009; Moshirian et al., 2012). Other studies use overnight returns to proxy overnight information by examining price movement during the following trading day. For instance, Lou et al. (2019) focus on overnight returns and trading strategies, and show both continuation and reversal effects. Aboody et al. (2018) use overnight returns to proxy firm-level investor sentiment and confirm short-term persistence and long-term reversal effects. The effects are more pronounced for high-attention stocks and during high-investor-sentiment periods (see, e.g., Berkman et al., 2012). Our approach is different in that we use the information on social networks to measure investor sentiment directly.

Furthermore, multiple studies on multimarket trading investigate price discovery, and the empirical evidence is mixed. For example, some studies (see, e.g., Frijns et al., 2015; Korczak & Phylaktis, 2010) find that the US markets dominate in the stock price-discovery process. Other works on cross-listed stocks (see, e.g., Eun & Sabherwal, 2003) show that the home market is better at processing information than the US market in the case of dual-listed stocks. Conversely, Lockwood et al. (2018) document that price discovery is highly dependent on execution quality and is evenly split between the US and home markets. Our article is different as we focus on the information flows on social media when both markets are closed. In particular, we are interested in whether overnight social media information interacts with positive/negative information in the United States during the lagged trading period. The interaction (if any) could assist or hinder the price-discovery process. In other words, we examine whether overnight social media information could facilitate the price discovery of cross-listed UK stocks.

Finally, several studies investigate the association between information on social networks and stock features. For instance, Sprenger et al. (2014) show that various company-related news on Twitter could influence S&P 500 company stock prices differently. Other studies explore whether information from tweet messages can help predict stock market indexes. For example, Zhang et al. (2011) find a significant relation between the level of tweet emotionality and three US stock market indexes: Dow Jones, Nasdaq, and S&P 500.[3] However, as far as we know, no prior studies investigate how social media information could affect the price-discovery process through the use of cross-listed stocks. Therefore, we address this research question. Most prior studies focus on tweets from StockTwits, where the information is about stocks and trading, whereas we collect primary data from Twitter directly. This covers a wider scope of information compared to StockTwits.

2 | DATA

2.1 | Data description

We use Application Programming Interface (API) to collect data from Twitter; it can be viewed as an interface between users and the Twitter system. This interface forwards queries from users to the system and then sends responses back to users. We send requests to collect tweets with keywords (i.e., the name of a UK–US dual-listed company) and obtain a sample of Twitter messages containing the keywords. The Twitter messages contain information about the content of the tweets, username and ID, date, and follower counts. We harvest around 29.43 million tweets containing the name of a UK–US dual-listed firm over a 4-year period from 2015 to 2018. We obtain intraday stock data for UK–US dual-listed firms from Tickdatamarket. Because people normally do not like to tweet long company names, some firms are excluded from our sample due to insufficient data (e.g., Aberdeen Asset Management). Consequently, we have 20 dual-listed firms in the UK and US stock markets that have, on average,

[3]Recent studies about social media and stock markets include Renault (2017) and Behrendt and Schmidt (2018).

more than 100 daily Twitter messages.[4] Our Twitter messages are collected during overnight periods when both the UK and US markets are closed. Specifically, tweets from 21:00 UK time on day $t - 1$ to 08:00 UK time on day t are attributed to observations on day t.

Following Fan et al. (2020), we further clean the Twitter messages in three steps. First, we delete all special characters in the tweets, for instance, link tokens (starting with http, https, www), hashtag tokens (starting with #), and user identifier tokens (starting with @). Second, we exclude all Twitter messages with only links or URLs. Finally, all non-English tweets are eliminated.

2.2 | Tweet sentiment and volumes

Because of the importance of the sentiment of the tweets, we separate positive Twitter messages from negative Twitter messages. We produce a polarity score for each of the sampled tweets using a text-processing tool in Python called TextBlob.[5] This library is widely used for tasks such as noun–phrase extraction and sentiment analysis (e.g., Gorodnichenko et al., 2021; Li et al., 2019). We also experiment with three other dictionaries: B. Liu (2015), Loughran and McDonald (2011), and the social media lexicon of Renault (2017). When analyzing sentiment, TextBlob gives polarity scores between −1 and 1. We classify the tweets with negative (positive) scores as negative (positive) sentiment Twitter messages, and tweets with 0 scores as neutral sentiment. Both PatternAnalyzer and NaiveBayesAnalyzer from TextBlob are used to conduct sentiment analysis, and we receive the same sentiment score for every Twitter message.

The aggregated measures of Twitter volume, agreement, and sentiment (positiveness) are constructed using all tweets in the sample. First, Twitter message volume is the natural logarithm of the number of tweets with a UK–US dual-listed firm name on day t. Second, following Antweiler and Frank (2004), we define the positiveness sentiment measure as

$$Positiveness_t = \ln\left(\frac{1 + Message_t^{positive}}{1 + Message_t^{negative}}\right), \tag{1}$$

in which $Message_t^{positive}$ and $Message_t^{negative}$ give the number of positive and negative tweets on each day t, respectively. Finally, we denote the agreement measure as

$$Agreement_t = 1 - \sqrt{1 - \left(\frac{Message_t^{positive} - Message_t^{negative}}{Message_t^{positive} + Message_t^{negative}}\right)^2}. \tag{2}$$

This agreement measure equals one if all tweets are either positive or negative.

2.3 | Summary statistics

In total, we have 29.43 million Twitter messages with the name of a UK–US dual-listed firm. Table 1 reports summary statistics for the market and tweet features. The average number of daily tweets is around 1792, and the standard deviation is around 5225 Twitter messages per day. The large number of tweets relating to firms in our sample suggests that our data contains a good information flow.

[4]The 20 UK–US dual-listed firms in our sample include: AstraZeneca, Barclays, BHP Billiton, British American Tobacco, BT Group, BP, Carnival, Diageo, GlaxoSmithKline, HSBC, Intercontinental Hotels Group, Lloyds Banking Group, National Grid, Pearson, Prudential, Rio Tinto, Royal Bank of Scotland Group, Royal Dutch Shell, Unilever, and Vodafone Group.

[5]See Loria (2018) for more information about TextBlob.

TABLE 1 Descriptive statistics of overnight tweets features and F/X rate returns

Variable	Mean	SD	1st quartile	Median	3rd quartile
No. of tweets	1792	5225	42	203	1384
Positiveness	0.8698	0.9710	0.2412	0.8973	1.4543
Agreement	0.3017	0.3413	0.0572	0.1602	0.4000
F/X return	−0.0000	0.0026	−0.0011	0.0001	0.0011

Note: This table reports summary statistics of overnight tweets and foreign exchange (F/X) rate returns. The aggregate characteristics, that is, positiveness, agreement measures are calculated based on overnight tweets from midnight until 08:00 UK time. The collected tweets contain the sampled companies' names. F/X rate returns are calculated based on F/X rate until 08:00 UK time.

There are statistically significant relations between the sentiment embedded in overnight tweets and both return and volatility measures. Yet, these significant correlations are relatively low, indicating that the linear relation is weak.[6] Figure 1 plots the average 30-min returns and volatility of the sampled firms. In line with the literature, the volatility plots form a U-shape. This confirms that investors are active in the first 30 min due to news announcements and the last 30 min as they want to unload their inventory and reduce the exposure to the overnight risk of holding stocks.

Table 2 presents the descriptive statistics of intraday UK stocks' and corresponding US American Depository Receipt (ADR) returns and volatility during 30-min, 1-hr, and 3-hr intervals. The mean return decreases from a 30-min interval to a 3-hr interval for the UK first and last trading intervals, whereas it increases from a 30-min interval to a 3-hr interval for the US last trading intervals. The largest mean return is 0.0178 for the UK first 30-min interval and the smallest mean return is −0.0067 for the US last 30-min interval. All volatility and standard deviation measures increase as we include more observations from a 30-min interval to a 3-hr interval. The largest mean volatility is 0.8564 for the UK first 3-hr interval and the smallest mean volatility is 0.1831 for the US last 30-min interval.

3 | EMPIRICAL STRATEGY

We employ log return to get abnormal return, which is given as:

$$AR_{l,i,t} = R_{l,i,t} - E(R_{l,i,t}),$$

where $R_{l,i,t}$ is the log return for stock i on day t over window l, l = [30-min, 1-hr, 3-hr], and $E(R_{l,i,t})$ denotes the expected return for stock i on day t over window l. This expected intraday return equals the mean value of $R_{l,i,t}$ during a 100-day window from day $t-110$ (i.e., $[t-110, t-10]$).

Following Parkinson (1980), we use intraday high and low stock prices $S_{t,high}$ and $S_{t,low}$ to measure daily volatility and define $Vol^{Park} = \ln(S_{t,high}/S_{t,low})/2\sqrt{\ln 2}$. Log range is argued as an efficient and robust range-based volatility measure (see, e.g., Alizadeh et al., 2002; Hendershott et al., 2011).

LSE opens from 08:00 to 16:30 UK time, while the opening hours of NYSE and Nasdaq are between 09:30 and 16:00 Eastern time; thus, there are 11 hours after the US markets close and before the UK market opens the next day. We examine how the message volume, sentiment, and agreement of tweets posted after the US markets close can affect UK dual-listed stocks after the UK market opens. A simple timeline for illustration purposes is shown in Figure 2.

[6]Refer to Online Appendix Table A1 for detailed results.

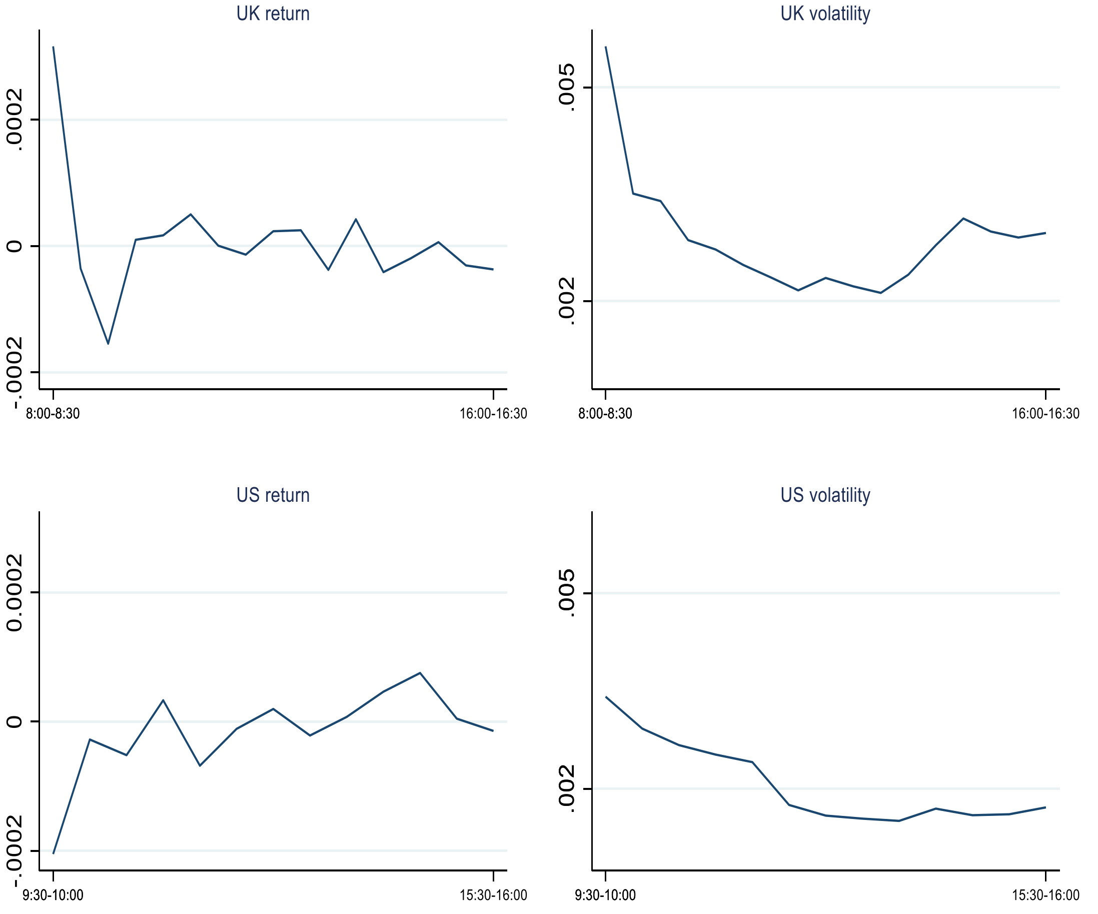

FIGURE 1 Intraday returns and volatility. This figure depicts 30-min return and volatility across a trading day. The *y*-axis shows the value of UK (US) stock return and volatility, and the *x*-axis shows the time, that is, aggregated half-hour intervals when the UK (US) stock market opens [Color figure can be viewed at wileyonlinelibrary.com]

We investigate how the stock features of the ADRs in the US market can affect the corresponding UK companies dual-listed in the UK market through overnight messages on social networks. The baseline regression is given as follows:

$$y^{UK}_{F,i,t} = \alpha + y^{US}_{L,i,t-1} + \beta_1 Positiveness_{i,t} + \beta_2 Message_{i,t} + \beta_3 Agreement_{i,t} + \beta_4 Positiveness_{i,t} \times y^{US}_{L,i,t-1} + \beta_5 Message_{i,t} \times y^{US}_{L,i,t-1} + \beta_6 Agreement_{i,t} \times y^{US}_{L,i,t-1} + \delta C_t + u_i + \varepsilon_{i,t}, \quad (3)$$

where i denotes firm and t stands for time, F and L stand for the first and last trading intervals in a day. $\{F, L\}$ = {30-min, 1-hr, 3-hr}, that is, the first or last 30 min, 1 hr, and 3 hr during the trading period. y is a vector of stock indicators including return and volatility. We employ Parkinson's (1980) intraday high–low range as a measure of volatility. $Positiveness_{i,t}$ captures aggregate positiveness embedded in overnight tweets, and $Agreement_{i,t}$ measures the degree to which tweets agree with one another. In terms of aggregation methods, we follow the literature to calculate $Positiveness_{i,t}$ and $Agreement_{i,t}$ using Equations (1) and (2), respectively. $Message_t$ is calculated as the natural logarithm of the number of tweets. $Positiveness_{i,t} \times y^{US}_{L,i,t-1}$, $Message_{i,t} \times y^{US}_{L,i,t-1}$, and $Agreement_{i,t} \times y^{US}_{L,i,t-1}$ are the interaction terms between positiveness, message volume, agreement, and the stock features of the ADRs in US markets on the previous day. We use the FTSE 100 index return as the market return. C_t represents a vector of control variables such as market return,

TABLE 2 Descriptive statistics of intraday stock indicators

	UK, *F*		UK, *L*		US, *L*	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
30-min interval: {*F*, *L*} = 0.5 hr						
Return (%)	0.0178	1.0529	0.0008	0.3257	−0.0067	0.2223
Volatility (%)	0.4586	0.4501	0.3312	1.1532	0.1831	0.1283
1-hr interval: {*F*, *L*} = 1 hr						
Return (%)	0.0172	1.0783	−0.0031	0.4501	−0.0045	0.3054
Volatility (%)	0.5361	0.5889	0.4565	1.2180	0.2461	0.1768
3-hr interval: {*F*, *L*} = 3 hr						
Return (%)	0.0150	1.3287	−0.0039	0.7606	−0.0012	0.5036
Volatility (%)	0.8564	3.0578	0.7918	4.0961	0.4073	0.2958

Note: This table reports summary statistics of UK stock and corresponding US American Depository Receipt returns volatility during 30-min, 1-hr, and 3-hr intervals. *F* and *L* denote the first and last trading intervals, {*F*, *L*} = {0.5, 1, 3} in hours.

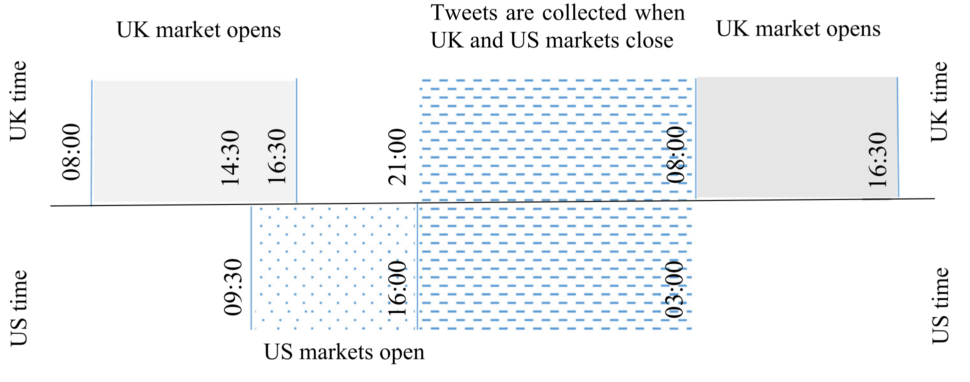

FIGURE 2 Timeline of events in UK and US stock markets. This figure shows the timeline of UK market opens, US market opens, both UK and US markets open, and both UK and US markets close (tweets are collected) [Color figure can be viewed at wileyonlinelibrary.com]

GBP/USD (British pound/US dollar) exchange rate return, and lagged return (volatility). Moreover, Lockwood et al. (2018) show that execution quality plays an important role in the price-discovery process. Therefore, we include relative volume as a control variable to account for stock-specific order-execution differences.[7] Relative volume is denoted as the money value of stocks traded on US exchanges over the money value of shares traded on both the UK and US stock markets. Money value is calculated as the product of the number of shares traded and the close price of the trading interval. Specifically, relative volume for each cross-listed stock i at time t is

$$Relative\ Volume_{i,t} = \frac{Money\ Volume_{i,US,t}}{Money\ Volume_{i,US,t} + Money\ Volume_{i,UK,t}}. \tag{4}$$

We use firm fixed effects u_i and $\varepsilon_{i,t}$ is the error term.

[7]Relative cost and price impact advantage are included as control variables in Section 4.3.

In line with the literature (see, e.g., Sprenger et al., 2014), we expect a positive (negative) coefficient of positiveness to explain returns (volatility). Antweiler and Frank (2004) find that the message volume on online stock message boards can predict stock volatility. De Long et al. (1990) argue that trading by noise traders can result in an increase in volatility, as the uncertainty of noise traders' beliefs could create a risk that prevents arbitrageurs from getting the true fundamental value. Hence, we expect a positive coefficient of *Message* in explaining volatility. $Agreement_t$ captures the degree to which tweets agree with each other, that is, similar or different numbers of positive versus negative tweets. If the agreement measure among investors is low, there should be more uncertainty among investors in the market and hence a higher level of volatility. Jones et al. (1994) empirically show that volatility can reflect disagreement among market participants. Therefore, we expect negative coefficients of *Agreement* in explaining volatility.

4 | RESULTS DISCUSSION

4.1 | Timeliness of information propagation on social media

We use regressions to investigate whether aggregate Twitter features (positiveness, message volume, and agreement) can help explaining stock trading indicators (stock returns and volatility). Lagged returns, GBP/USD exchange rate returns, and relative volume are employed as control variables. Consistent with the literature (see, e.g., Chen et al., 2014), there are positive and significant relations between the aggregate positiveness measure of tweets and stock returns as shown in Table 3. The magnitude of positiveness in the first 30-min interval (0.0361) is economically meaningful, as the average of the first 30-min return is 0.0178 (see Table 2). This suggests that more optimistic information on social networks is associated with positive abnormal returns.

In line with prior papers (e.g., Sprenger et al., 2014), we find a negative coefficient of the agreement measure. The negative and statistically significant relation between returns and the agreement measure of tweets indicates that there are positive abnormal returns when there is an increased level of disagreement among investors in the market. However, contrary to Wysocki (1998), we do not find a similar relation between the volume of Twitter messages and stock returns. This is consistent with the literature in the sense that information on social networks potentially contains a certain degree of noise compared to information from professional investor discussion boards.

The interaction terms between returns and the tweet features are included as extra independent variables at different intervals (i.e., first 30 min, first 1 hr, and first 3 hr of trading). The coefficients of the interaction terms are statistically significant and economically meaningful. The signs of these coefficients are consistent with those of the tweet features, suggesting that lagged returns of ADRs in US stock markets interact and strengthen the impact of social media. Overnight information on social networks could reinforce the stock returns of ADRs in US stock markets in explaining the changes in stock returns of cross-listed UK firms in the UK stock market.

We find that statistically significant coefficients of Twitter features explain the changes in volatility in Table 4. A 1% increase in Twitter message volume is linked to 0.10% increase in volatility during the first 30 min of trading. Consistent with Antweiler and Frank (2004), our results imply that more messages are posted on social media when there is more uncertainty and risk in the market. The interaction terms between volatility and Twitter message volume are also included in the regressions. We find evidence that the volume of tweet information on social media could strengthen the volatility of ADRs in US stock markets in explaining the changes in the volatility of dual-listed UK companies in the UK stock market. There are also strong associations between the positiveness sentiment measure of tweets and volatility. These results are in line with our expectations and previous findings (see, e.g., Sprenger et al., 2014).

Consistent with prior findings, the interactions between the positiveness sentiment measure of overnight tweets and lagged US stock returns have a positive effect on UK stock returns the following day, as positive expectations about the market are related to positive returns. Contrarily, the interactions between the agreement measure (message volume) and lagged US stock returns are negatively related to UK stock returns the following

TABLE 3 Regressions of returns

	{L, F} = 0.5 hr (1)	{L, F} = 1 hr (2)	{L, F} = 3 hr (3)
$Positiveness_t$	0.0361***	0.0275***	0.0242**
	(3.71)	(2.80)	(2.55)
$Message_t$	−0.0321*	−0.0294	−0.0328**
	(−1.75)	(−1.59)	(−1.97)
$Agreement_t$	−0.0359***	−0.0322***	−0.0275**
	(−3.17)	(−2.82)	(−2.44)
$Positiveness \times Ret^{US}{}_{L,t-1}$	0.1304***	0.0738***	0.0573***
	(10.84)	(6.02)	(4.93)
$Message \times Ret^{US}{}_{L,t-1}$	−0.0571***	−0.0193	−0.0345***
	(−4.83)	(−1.63)	(−2.97)
$Agreement \times Ret^{US}{}_{L,t-1}$	−0.0730***	−0.0399***	−0.0551***
	(−5.60)	(−3.02)	(−4.21)
$Ret^{US}{}_{L,t-1}$	0.0512***	0.1155***	0.2279***
	(4.03)	(9.08)	(18.67)
$Ret^{UK}{}_{L,t-1}$	−0.0077	−0.0165**	−0.0037
	(−0.96)	(−2.02)	(−0.47)
$Relative\ Volume_{L,t-1}$	0.0062	0.0108	0.0184
	(0.65)	(1.12)	(1.36)
$FX\ ret_t$	0.2136***	0.1637***	0.1563***
	(26.39)	(20.09)	(19.84)
N	14,486	14,495	14,839
R^2	0.070	0.053	0.088

Note: This table reports fixed-effects regressions of UK stock returns during first 30-min to 3-hr trading by overnight tweets. Dependent variables are $Ret^{UK}{}_{F,t}$, which is (log) return during the first 30-min, 1-hr, and 3-hr trading in the United Kingdom. The main independent variables are aggregate tweet measures based on overnight tweets (*Positiveness*, *Message*, and *Agreement*). The interaction terms are between these tweet measures and prior-day returns of the corresponding US American Depository Receipts (ADRs) during the last intervals equivalent to those of the dependent variables. *Relative Volume* is the trading value of an ADR on US exchanges over the total trading value of the corresponding stock in both UK and US markets. *FX ret* is the log return of GBP/USD. Subscript *t* denotes day *t*, and *F* and *L* denote the first and last trading intervals, {*F*, *L*} = {0.5, 1, 3} in hours.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

day. Hence, there are positive abnormal returns when there is more disagreement among investors. Moreover, the interactions between Twitter message volume and lagged US stock volatility is positively associated with UK stock volatility the following day. Therefore, more messages are posted during more volatile periods. Overall, the interactions between overnight tweets features and lagged stock indicators provide an insightful way to examine the informativeness of social media information.

TABLE 4 Regressions of volatility

	{L, F} = 0.5 hr (1)	{L, F} = 1 hr (2)	{L, F} = 3 hr (3)
$Positiveness_t$	0.0582***	0.0584***	0.0151
	(4.12)	(4.06)	(0.91)
$Message_t$	0.1027***	0.0953***	0.0170
	(5.27)	(4.77)	(0.75)
$Agreement_t$	−0.0245	−0.0221	−0.0050
	(−1.52)	(−1.35)	(−0.26)
$Positiveness \times Volatility^{US}_{L,t-1}$	−0.1129***	−0.1145***	−0.0234
	(−7.35)	(−7.31)	(−1.30)
$Message \times Volatility^{US}_{L,t-1}$	0.0640***	0.0662***	0.0185
	(4.19)	(4.27)	(0.98)
$Agreement \times Volatility^{US}_{L,t-1}$	0.0410**	0.0319*	0.0044
	(2.50)	(1.90)	(0.22)
$Volatility^{US}_{L,t-1}$	0.2254***	0.2226***	0.0644***
	(20.71)	(19.73)	(4.91)
$Volatility^{UK}_{L,t-1}$	0.0215***	0.0420***	0.0050
	(3.18)	(5.95)	(0.61)
$Relative\ Volume_{L,t-1}$	0.5323***	0.4652***	0.0482***
	(66.57)	(55.86)	(3.42)
$FX\ ret_t$	−0.1886***	−0.2098***	−0.0538***
	(−27.86)	(−29.82)	(−6.55)
N	14,486	14,495	14,839
R^2	0.345	0.291	0.012

Note: This table reports fixed-effects regressions of UK stock volatility during first 30-min to 3-hr trading by overnight tweets. Dependent variables are $Volatility^{UK}_{F,t}$, which is volatility during the first 30-min, 1-hr, and 3-hr trading in the United Kingdom. The main independent variables are aggregate tweet measures based on overnight tweets (*Positiveness*, *Message*, and *Agreement*). The interaction terms are between these tweet measures and prior-day volatility of the corresponding US American Depository Receipts (ADRs) during the last intervals equivalent to those of the dependent variables. *Relative Volume* is the trading value of an ADR on US exchanges over the total trading value of the corresponding stock in both UK and US markets. *FX ret* is the log return of GBP/USD. Subscript *t* denotes day *t*, and *F* and *L* denote the first and last trading intervals, {F, L} = {0.5, 1, 3} in hours.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Figure 3 plots estimated coefficients of the interaction terms between tweet and stock features. Overall, the coefficients decrease (increase) for the positiveness (message volume, agreement) measure in the first 30-min of trading and then become persistent around 3 hr after the market opens. This is in line with the literature, demonstrating that investors trade actively during the first 30 min. Altogether, these findings are consistent with our prior results and confirm our assertion about the information dissemination role of social networks in supporting price discovery among different stock markets.

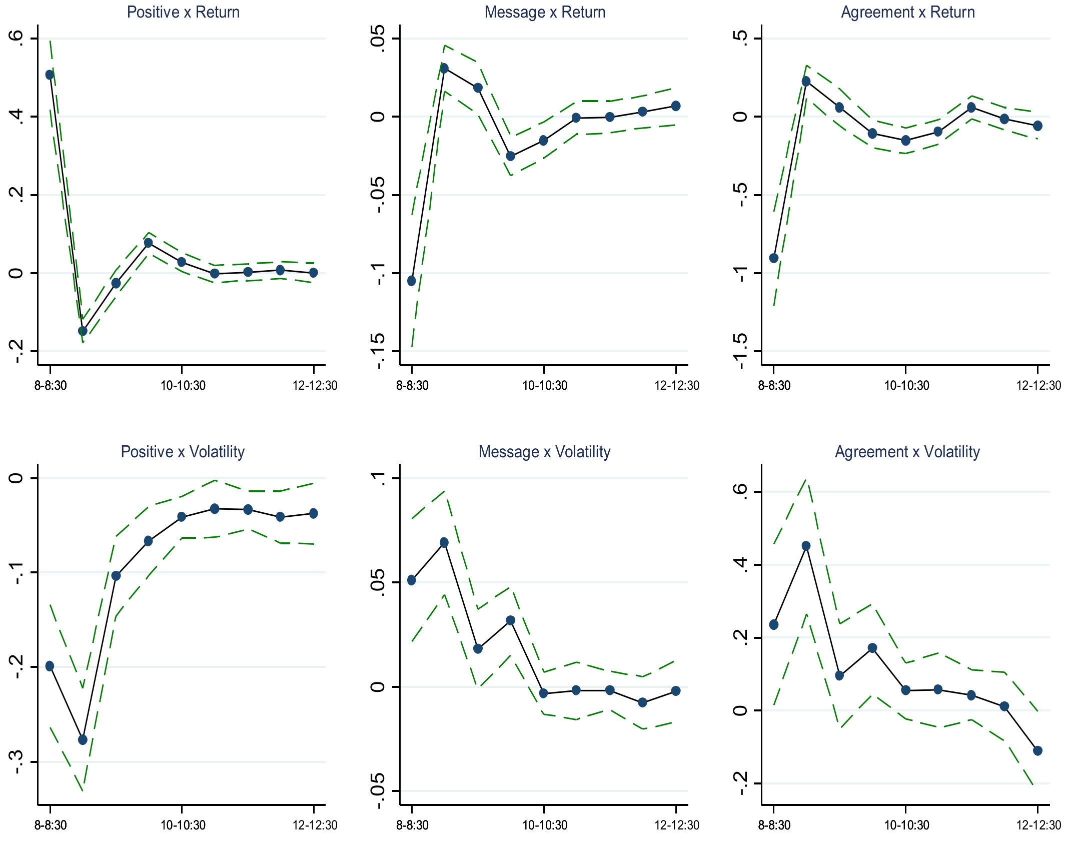

FIGURE 3 Interaction between tweet and stock features. This figure shows estimated coefficients of the interaction terms between tweet measures, that is, positiveness, message, agreement, yesterday return, and volatility of the corresponding American Depository Receipts (ADRs) during last 30 min of trading. The estimated coefficients are from 18 regressions: 9 regressions of returns (volatility) during 1st to 9th 30-min intervals, respectively. NYSE/Nasdaq opens at 09:30 Eastern Time, which is 14:30 UK time [Color figure can be viewed at wileyonlinelibrary.com]

4.2 | Trading strategies and transaction costs

We further develop a trading strategy to assess the value of the return predictive regression. A predictive regression model for each stock is estimated recursively using Equation (3). The predicted return during the first 30-min interval on each day is then used to form trading rules. A long (short) position is opened at 08:00 and closed at 08:30 UK time if the predicted return meets the following two criteria: (1) the predicted return is larger (smaller) than the top (bottom) 5%/10%/25% of daily distribution of predicted returns of all 20 firms (i.e., 20 observations each day are used to form the threshold values), and (2) the predicted return is larger (smaller) than 0.5% (−0.5%). Otherwise, we do not trade. The summary statistics for returns achieved from this strategy are reported in Panels A–C of Table 5. This trading strategy generates the highest returns when we use the bottom 25th and top 75th percentile thresholds. The average annual returns range from 19.49% to 53.52%.[8] These returns do not seem very large at first glance. However, they are actually enormous for 30-min trading per day when no overnight position and risk are involved.

[8]Returns are annualized by multiplying the number of trading days in a year (252), as we trade once a day, but only for 30 min.

TABLE 5 Predictive regression models for each stock are estimated recursively using Equation (3)

	Return (%)	*SD* (%)	Skewness	Kurtosis	Success (%)
Panel A: Results for bottom 5th and top 95th percentile thresholds					
No transaction costs					
Long	19.49	13.46	0.18	6.24	52.02
Short	21.64	12.81	−0.21	4.81	59.45
Long − short	24.83	13.70	0.00	5.71	57.69
20 bps transaction costs					
Long	10.55	12.05	0.26	7.79	41.61
Short	12.14	11.42	−0.16	6.01	46.25
Long − short	13.48	12.13	0.07	7.27	44.26
Panel B: Results for bottom 10th and top 90th percentile thresholds					
No transaction costs					
Long	27.18	20.14	0.58	8.98	52.19
Short	38.46	19.16	0.76	8.09	56.92
Long − short	39.67	21.03	0.34	8.85	55.19
20 bps transaction costs					
Long	13.69	18.06	0.69	11.17	41.03
Short	22.75	17.26	0.91	10.00	44.55
Long − short	21.72	18.77	0.44	11.14	42.21
Panel C: Results for bottom 25th and top 75th percentile thresholds					
No transaction costs					
Long	33.04	29.27	0.29	13.51	52.59
Short	53.52	42.31	−2.26	45.11	57.69
Long − short	52.38	40.82	−2.78	52.25	55.90
20 bps transaction costs					
Long	13.91	26.24	0.32	16.81	41.67
Short	29.69	38.09	−2.51	55.60	45.17
Long − short	26.03	36.42	−3.10	65.64	42.96
Panel D: Benchmark (without transaction costs)					
Average weighted portfolio	2.77	14.80	−0.12	5.74	
FTSE 100	0.59	14.10	−0.18	5.57	

Note: The predicted first 30-min return for each day is employed to form trading rules. We sell (buy) at the opening and buy (sell) after 30 min if the predicted return satisfies the following two criteria: (1) predicted return is smaller (larger) than the first (third) quartile of daily distribution of predicted returns of all 20 companies (i.e., 20 observations on each day are used to form quartile values), and (2) predicted return is smaller (larger) than 0.5%. Returns are annualized.

For comparison, Panel D of Table 5 reports the performance of two benchmark strategies. The first benchmark is a buy-and-hold strategy, in which we hold a long position in the weighted average portfolio of all 20 stocks from the beginning until the end of the sample period. The average return of this strategy is 2.77% annually; hence, our trading strategy significantly outperforms this passive strategy. The second strategy is always long of the FTSE 100 index, where we take a long position in the market index at the opening and close our position when the market closes to avoid overnight risk. The annualized average return of this strategy is 0.59%, which is lower than buy-and-hold and the average return given by our proposed trading strategy.

We also need to consider risk. The standard deviations (from 12.81% to 13.70%) of our trading strategy using the bottom 5th and top 95th percentile thresholds are comparable to that of an always-long strategy (14.10%). However, the standard deviations of our trading strategy using the bottom 10th (25th) and top 90th (75th) percentile thresholds become higher (19.16% to 21.03% and 29.27% to 42.31%, respectively). This indicates that although our proposed trading strategy can deliver higher returns than benchmark strategies, our trading strategy can be considered riskier. Additionally, our proposed trading strategy shows positive skewness (0.34 to 0.76) when employing the bottom 10th and top 90th percentile thresholds (compared to −0.12 and −0.18 of buy-and-hold and always-long strategies) and positive kurtosis between 4.81 and 52.25, indicating that our strategy usually generates high positive returns.

Additionally, we discuss success rates, which are the proportion of times that our predictions are correct in applying trading rules as compared to actual returns. The success rates of our trading strategy vary from 52.02% to 59.45% if we choose the bottom 5th and top 95th percentile thresholds. As we choose stricter critical values, the success rates of our trading strategy decrease and range from 52.19% to 56.92% (between 52.59% and 57.69%) if we choose the bottom 10th (25th) and top 90th (75th) percentile thresholds. Although the success rates decrease, implying a lower chance of making correct decisions, the average returns of our trading strategy using the bottom 10th (25th) and top 90th (75th) percentile thresholds vary from 27.18% to 39.67% (between 33.04% and 53.52%) and are higher than those obtained using the bottom 5th and top 95th percentile thresholds (19.49% to 24.83%). This is because our strategies deliver higher returns when we employ stricter thresholds; however, we have a lower number of trading days.

Finally, we examine the effect of transaction costs on our proposed trading strategy. Transaction costs have decreased in recent years with the development of technology, and particularly after decimalization. Following Heston et al. (2010), we employ 20 basis points (bps) as transaction costs because all cross-listed stocks in the sample are large stocks.[9] The results reported in Panel A to C of Table 5 show that after deducting transaction costs, both returns and standard deviations decrease compared to before, but our trading strategy still remains profitable. Similar to before, we obtain the highest returns (from 13.91% to 29.69%) when choosing the bottom 25th and top 75th percentile thresholds. Contrarily, when we employ the bottom 5th (10th) and top 95th (90th) percentile thresholds, the returns are lower, between 10.55% and 13.48% (from 13.69% to 22.75%). In summary, our proposed trading strategy still generates economically meaningful profits after considering transaction costs.

4.3 | Robustness checks

Lockwood et al. (2018) document that price discovery is highly dependent on execution quality (e.g., relative cost, relative volume, and price impact advantage). As a result, we control for other stock-specific order execution differences (relative cost and price impact advantage) apart from relative volume. First, we denote relative cost as

[9]This is conservative, given that our sampled stocks are large and liquid stocks. Average bid–ask spreads for blue chips on LSE are usually significantly lower than 10 bps.

the fraction of the percentage bid–ask spread of the home market to the percentage bid–ask spread of the US markets. The percentage bid–ask spread is given as

$$BA_{i,j,t} = \frac{Ask_{i,j,t} - Bid_{i,j,t}}{0.5 \times (Ask_{i,j,t} + Bid_{i,j,t})}, \tag{5}$$

where $Bid_{i,j,t}$ and $Ask_{i,j,t}$ are the bid and ask prices of stock market j for cross-listed stock i at time t. Relative cost is then calculated as the percentage bid–ask spread of the UK stock market over the percentage bid–ask spread of US stock markets, that is,

$$Relative\ Cost_{i,t} = \frac{BA_{i,UK,t}}{BA_{i,US,t}}. \tag{6}$$

Second, a price impact difference measure is used. Price impact describes the trading volume required to change the stock price by a certain amount, and it captures the ability of the exchange to absorb trades without changing the price. We choose Kyle's lambda (Kyle, 1985) to measure price impact, which is the slope coefficient in

$$\Delta P_{i,j,t} = \lambda_{i,j} Volume_{i,j,t} + \varepsilon_{i,j,t}, \tag{7}$$

where $\Delta P_{i,j,t}$ and $Volume_{i,j,t}$ are the percentage price change and trading volume of each cross-listed stock i in stock market j at time t, and $\varepsilon_{i,j,t}$ is the error term. Price impact advantage is then calculated as the difference in price impact between the UK stock market and US stock markets, that is,

$$Price\ Impact\ Advantage_{i,t} = \hat{\lambda}_{i,UK} - \hat{\lambda}_{i,US}. \tag{8}$$

We obtain consistent results after controlling for relative cost and price impact advantage, as reported in Online Appendix Tables A3–A6.

Additionally, to verify our choice of using TextBlob to conduct sentiment analysis, we use three additional methods as robustness checks. First, we use the dictionaries in B. Liu (2015), which relate to sentiment analysis. Second, we employ the Loughran and McDonald (2011) dictionary, which is developed using 10-Ks to better reflect tone within a financial context. However, the language in formal documents differs from the content posted by individuals on social media. Therefore, third, we use the social media lexicon of Renault (2017), which is constructed using messages published on the investor Twitter platform StockTwits. All results are quantitatively similar, and Online Appendix Tables A7 and A8 report results for the 30-min interval.

It is probable that the profits of trading strategies are the result of aggregation and are dominated by only a small number of cross-listings. Therefore, in Online Appendix Table A9, we show the trading frequency and profit/loss for each of the 20 companies in the sample, to reassure that the trading strategy is robust across the sampled companies. Specifically, all firms are traded at least six times. National Grid has the lowest trading frequency, and we take a long position on this stock four times and go short twice. Other companies' shares are traded more frequently, and Barclays and Royal Bank of Scotland have the highest trading frequency. In terms of trading profit/loss, 7 companies suffer losses and 13 firms enjoy profits. The average magnitude of profits is larger than that of the losses, and Royal Bank of Scotland and BHP Billiton have the largest profits. Our strategy is also robust after excluding Royal Bank of Scotland and BHP Billiton from the sample.

Further alternative measures of stock features are used to check the robustness of our results. First, we use ordinary least squares (OLS) regressions to estimate the market model and expected return. We employ a 250-day estimation period from 260 days before the relevant date, following the literature (e.g., Fan et al., 2020). Second, as an alternative measure of volatility we subtract average volatility over the past 100 trading days from today's volatility (i.e. [−110, −10]). This captures the abnormal 30-min volatility compared to its typical value in past 100 days. All results in this section are in line with our main findings and are reported in Online Appendix Tables A10 and A11.

One possible explanation for our results could relate to the investor attention. Following prior papers (see, e.g., Da et al., 2011), we conduct investigations using the Google Search Volume Index (SVI) as a proxy for investor attention. The results remain strongly significant after controlling for SVI. This indicates that overnight social media messages contain certain informational values beyond behavioral explanations.

Moreover, we investigate whether social media messages convey new information or simply repeat what has already been circulated by traditional news outlets. We collect all news related to the sampled companies from the *Financial Times*. The *Financial Times* is a leading financial news outlet, at least within the UK market. We argue that because of competition between news outlets, any fundamental news related to sampled firms should be covered in the Financial Times. The investigations controlling for this proxy of traditional news yield qualitatively similar results. Further to this, we identify corporate announcements in the *Financial Times* and exclude all tweets on the days of these announcements and the days before and after the announcements. We repeat the estimation of our main specification based on the subsample of tweets. Additionally, we exclude tweets posted 3 days on either side of the earnings announcement dates from Institutional Broker's Estimate System (IBES). We obtain consistent results as reported in Online Appendix Tables A12–A15.

Finally, we conduct several additional robustness checks. First, we employ alternative 100-day and 2-year estimation periods when using OLS regressions to estimate the market model and expected return. Second, we employ 5-min realized variance (see, e.g., Andersen et al., 2007) during the first 30-min, 1-hr, and 3-hr of trading as the dependent variables instead of Parkinson (1980) volatility.[10] We get quantitatively similar results, as illustrated in Online Appendix Table A16. Third, we perform Fama–MacBeth regressions in addition to our baseline regressions. Lastly, we follow Gao et al. (2018) to use bid–ask spread as a proxy for transaction costs. All results from the robustness checks are quantitatively similar.

5 | CONCLUSIONS

Social media has become an increasingly popular channel for information dissemination, particularly for investors (see, e.g., Behrendt & Schmidt, 2018; Renault, 2017). The literature documents stock price discovery in multimarket trading (see, e.g., Eun & Sabherwal, 2003; Frijns et al., 2015). We examine whether social media serves as a vehicle for information flows when markets are closed. In other words, does social media facilitate the price-discovery process? Specifically, we use British stocks cross-listed in the UK and US markets to address this question. This unique laboratory enables us to uncover the economic forces behind the influence of Twitter information and the associated information transmission mechanisms.

Our results show that the volume, sentiment, and agreement of messages on social networks support price discovery by influencing the stock returns and volatility of cross-listed stocks. These effects weaken and disappear around 1 to 3 hr after the markets open. Additionally, dual-listing provides a natural environment to examine the information transmission role of social media between two stock markets, which may be neglected by investors and traditional news media. We focus on the substantial amount of information available on social media after one market closes and before the other market opens, that is, when people finish work and there is little information to be obtained from traditional news outlets. Our findings on the price-discovery enhancement role of social networks indicate that certain investors can benefit by carefully embedding social media information in their investment strategies. The findings confirm that social media is an effective vehicle for information transmission and supports the price-discovery process. We also propose a trading strategy based on the return relation, and our trading strategy outperforms two alternative benchmarks and is still profitable after considering transaction costs.

Altogether, our results suggest that investors do trade on the information spread on social media. However, investors are more cautious when assessing the information content of tweets about firms. Certain investors are able to distill such

[10]Liu, Patton, et al. (2015) study over 400 different estimators, using 11 years worth of data on 31 financial assets from five asset classes. They find that other measures seldom significantly outperform 5-min realized variance.

information and make profits. There are several areas of potential further research related to this study. First, although we provide empirical evidence on the information dissemination role of social media in stock markets, additional theoretical work could be carried out to better understand this transmission mechanism. Second, more research relating to social network trading strategies and associated portfolio management is important and promising. Finally, there could be increased benefits if social media data were to be made available for researchers to examine.

ACKNOWLEDGMENTS

We thank Dennis Philip, Patricia Chelley-Steeley, and two anonymous reviewers for their helpful comments. Standard disclaimer applies.

ORCID

Rui Fan http://orcid.org/0000-0002-8147-5758
Vu Tran http://orcid.org/0000-0001-9561-8118

REFERENCES

Aboody, D., Even-Tov, O., Lehavy, R., & Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, *53*(2), 485–505.
Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091.
Andersen, T. G., Bollerslev, T., & Dobrev, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics*, *138*(1), 125–180.
Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294.
Barclay, M. J., & Hendershott, T. (2003). Price discovery and trading after hours. *Review of Financial Studies*, *16*(4), 1041–1073.
Baruch, S., Panayides, M., & Venkataraman, K. (2017). Informed trading and price discovery before corporate events. *Journal of Financial Economics*, *125*(3), 561–588.
Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, *96*, 355–367.
Berkman, H., Koch, P. D., Tuttle, L., & Zhang, Y. J. (2012). Paying attention: Overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, *47*(4), 715–741.
Berkman, H., & Truong, C. (2009). Event day 0? After-hours earnings announcements. *Journal of Accounting Research*, *47*(1), 71–103.
Biais, B., Hillion, P., & Spatt, C. (1999). Price discovery and learning during the preopening period in the Paris Bourse. *Journal of Political Economy*, *107*(6), 1218–1248.
Bogousslavsky, V. (2016). Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance*, *71*(6), 2967–3006.
Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403.
Cushing, D., & Madhavan, A. (2000). Stock returns and trading at the close. *Journal of Financial Markets*, *3*(1), 45–67.
Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499.
Doyle, J. T., & Magilke, M. J. (2009). The timing of earnings announcements: An examination of the strategic disclosure hypothesis. *The Accounting Review*, *84*(1), 157–182.
Eun, C. S., & Sabherwal, S. (2003). Cross-border listings and price discovery: Evidence from U.S.-listed Canadian stocks. *The Journal of Finance*, *58*(2), 549–575.
Fan, R., Talavera, O., & Tran, V. (2020). Social media bots and stock markets. *European Financial Management*, *26*(3), 753–777.
Flood, M. D., Huisman, R., Koedijk, K. G., & Mahieu, R. J. (1999). Quote disclosure and price discovery in multiple-dealer financial markets. *Review of Financial Studies*, *12*(1), 37–59.
Frijns, B., Gilbert, A., & Tourani-Rad, A. (2015). The determinants of price discovery: Evidence from US-Canadian cross-listed shares. *Journal of Banking & Finance*, *59*, 457–468.
Gao, L., Han, Y., Zhengzi Li, S., & Zhou, G. (2018). Market intraday momentum. *Journal of Financial Economics*, *129*(2), 394–414.
Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *European Economic Review*, *136*, 103772.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, *66*(1), 1–33.
Heston, S. L., Korajczyk, R. A., & Sadka, R. (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance*, *65*(4), 1369–1407.
Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Transactions, volume, and volatility. *Review of Financial Studies*, *7*(4), 631–651.
Korczak, P., & Phylaktis, K. (2010). Related securities and price discovery: Evidence from NYSE-listed non-U.S. stocks. *Journal of Empirical Finance*, *17*(4), 566–584.
Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.
Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. *International Journal of Forecasting*, *35*(4), 1548–1560.
Liu, B. (2015). *Sentiment analysis: Mining sentiments, opinions, and emotions*. Cambridge University Press.
Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293–311.
Lockwood, J., Lockwood, L., & Lau, S. T. (2018). Price discovery of internationally cross-listed stocks during the 2008 financial crisis. *Journal of Financial Research*, *41*(3), 351–381.
De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738.
Loria, S. (2018). TextBlob documentation: Release 0.16.0. https://media.readthedocs.org/pdf/textblob/dev/textblob.pdf
Lou, D., Polk, C., & Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics*, *134*(1), 192–213.
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.
Madhavan, A., & Panchapagesan, V. (2000). Price discovery in auction markets: A look inside the black box. *Review of Financial Studies*, *13*(3), 627–658.
Moshirian, F., Nguyen, H. G., & Pham, P. K. (2012). Overnight public information, order placement, and price discovery during the pre-opening period. *Journal of Banking & Finance*, *36*(10), 2837–2851.
Murphy, D. P., & Thirumalai, R. S. (2017). Short-term return predictability and repetitive institutional net order activity. *Journal of Financial Research*, *40*(4), 455–477.
Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, *53*(1), 61–65.
Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, *84*, 25–40.
Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957.
Werner, I. M., & Kleidon, A. W. (1996). U.K. and U.S. trading of British cross-listed stocks: An intraday analysis of market integration. *Review of Financial Studies*, *9*(2), 619–664.
Wysocki, P. D. (1998). *Cheap talk on the web: The determinants of postings on stock message boards* (Working Paper 98025), University of Michigan Business School.
Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter "I hope it is not as bad as I fear." *Procedia - Social and Behavioral Sciences*, *26*, 55–62.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fan, R., Talavera, O., & Tran, V. (2022). Social media and price discovery: The case of cross-listed firms. *Journal of Financial Research*, 1–17. https://doi.org/10.1111/jfir.12310