Measuring the real-time stock market impact of firm-generated content

DOI: 10.1177/00222429211042848

JOURNAL of
Marketing

Measuring the real-time stock market impact of firm-generated content

Journal:	*Journal of Marketing*
Manuscript ID	JM.20.0109.R4
Manuscript Type:	Revised Submission
Research Topics:	Digital Marketing - Firms, Social Media, Firm-level Impact of Marketing Actions
Methods:	Analytical Models, Data Mining, Stock Market Models, Time Series Models

Measuring the real-time stock market impact of firm-generated content

Abstract: Firms increasingly follow an 'always on' philosophy, producing multiple pieces of firm-generated content (FGC) throughout the day. Current methodologies used in marketing are unsuited to unbiasedly capturing the impact of FGC disseminated intermittently throughout the day in stock markets characterized by ultra-high frequency trading. They also neither distinguish between the permanent (i.e. long-term) and temporary (i.e. short-term) price impacts nor identify FGC attributes capable of generating these price impacts. In this study, the authors define price impact as the impact on the variance of stock price. Employing a market microstructure approach to exploit the variance of high frequency changes in stock price the authors estimate the permanent and temporary price impacts of the firm-generated Twitter content of S&P 500 IT firms. The authors find that firm-generated tweets induce both permanent and temporary price impacts, which are linked to tweet attributes; valence and subject matter. Tweets reflecting only valence or subject matter concerning consumer or competitor orientation result in temporary price impacts, while those embodying both attributes generate permanent price impact; negative valence tweets about competitors generate the largest permanent price impacts. Building on these findings, the authors offer suggestions to marketing managers on the design of intraday FGC.

Keywords: real-time marketing; microstructure; high frequency data; firm-generated content; Twitter

With an excess of $37 billion in investment by U.S. firms in 2020 (Statista, 2020), social media is one of the most pervasive communication channels used by marketers (Berger et al., 2020; Hewett et al., 2016). One by-product of this investment is the creation of corporate social media accounts supporting the dissemination of firm-generated content (FGC) – a firm's communication disseminated through its own online communication tools (Kumar et al., 2016). Many firms have adopted an 'always on' approach in their social media marketing, disseminating multiple pieces of FGC throughout the day. Figure 1 illustrates the high frequency approach to FGC dissemination using the example of IT firms' activity on Twitter. Each piece of FGC is characterized by its attributes (e.g. Figure 1 shows FGC's valence and subject matter as key attributes),[1] as well as a timestamp reflecting the FGC's dissemination time. FGC and its timestamp can be accurately recorded to the second and mapped against the corresponding timestamp of trading activity that takes place at ultra-high frequency, i.e. at sub-second intervals (Hasbrouck and Saar, 2013). As a result of these high frequency activities, large volumes of intraday data emerge. For example, an S&P 500 IT firm can issue in excess of 6,000 tweets in a given month and trading in a single firm's stock often yields well over 10 million trading-related messages (e.g. quotes, cancellations, and transactions) during the same interval (see Web Appendix A).

<Insert Figure 1 about here>

Currently used marketing methodologies are challenged when analyzing high-frequency data because the trading data, which is used in capturing the impact of FGC on firm value, is characterized by unequal time intervals. Low-frequency event studies using end-of-day price measures and time-series analysis, such as standard vector autoregressive (VAR)

[1] The two examples in Figure 1 illustrate FGC attributes. First, the valence of FGC varies from positive (e.g. 'We were so happy to be a part of it.' – CA Technologies) to negative (e.g. 'Have you lost trust in tech?' – CA Technologies) across the day. Second, the subject matter of FGC also varies: Adobe's subject matter ranges from focusing on consumer (e.g. 'We want to know what inspires you' – Adobe) to focusing more on competitive positioning (e.g. 'As of today, all fonts included with Creative Cloud can be used on iOS13.1' – Adobe).

models, are unable to effectively address these problems (see Web Appendix B). These methods rely on discrete and uniform time intervals that do not describe the time intervals associated with high frequency trading data, which encapsulate the intraday evolution of firm stock price. Often these methods aggregate trading data to regular intervals, which can lead to the elimination of upwards of 99% of intraday trading observations for studies employing end-of-day price data (see Web Appendix C). The frequency of FGC as an intraday variable and the effects of other non-FGC events potentially further bias low-frequency analysis employing standard low frequency analytical methods. Furthermore, the richness of such an assessment and marketing researchers' understanding of 'always on' strategies are compromised unless the research identifies both the short- and long-term impacts of this form of marketing (Gordon et al., 2021). Consequently, the findings deriving from current examinations lack detailed ex-post insights on the impact of FGC generated at intraday frequencies, which leaves marketing managers unable to effectively design future intraday marketing resource allocation strategies (Kanuri, Chen and Sridhar, 2018).

Employing the market microstructure approach, which relies on ultra-high frequency trading data analysis, we investigate the stock price impact of FGC where price impact is defined as the impact on the variance of stock price. This approach of estimating price impact of FGC as impact on the variance of stock price rather than on changes in the level of stock price is driven by both methodological and theoretical necessity. Although estimating level changes in stock price as a result of an event like FGC could be approached from the perspective of computing simple price impact measures, when working with high frequency data, the approach is inadequate for at least two reasons. Firstly, simple price impacts are misleading when trades in stock markets are serially correlated. Secondly, in the presence of transient impacts, as is the case in this study (e.g. we capture price impact at second-by-second intervals), simple price impacts rely on getting the timing just right, which is methodologically unfeasible when investigating large datasets as is again the case here. Our methodological approach is in line with market microstructure theory, addresses the above outlined issues, and is consistent with the microstructure literature (Brogaard, Hendershott and Riordan, 2014; Rzayev and Ibikunle, 2019). Based on this approach and assessing S&P 500 IT firms' Twitter activity, we contribute to the marketing literature at three levels.

Firstly, by using high frequency data, this is the first study to document the sub-minute impact of individual pieces of FGC disseminated during the day, thus the insights presented are unlikely to be affected by confounding effects that using end-of-day data is susceptible to. We estimate the price impact of FGC at the second and minute levels by computing the variance of fast-paced (e.g. sub-second-by-sub-second) intraday changes in stock price as they occur in financial markets (Brogaard, Hendershott and Riordan, 2014; Budish, Cramton and Shim, 2014; Kirilenko et al., 2017), thereby demonstrating the instantaneous impact of FGC. The variance of intraday changes in price is obtained through state space modeling with Kalman filtering. By doing so, we contribute to the emerging stream of marketing research studying the impact of FGC on firm financial outcomes (Borah et al., 2020; Colicev et al., 2018) and real-time marketing (Rust et al, 2021), and respond to research priorities established by the Marketing Science Institute (2018-2020), emphasizing the need to help marketers 'get marketing right' by providing insights into the instantaneous impact of FGC.

Secondly, with the microstructure perspective, we reveal both the permanent and temporary price impact of FGC as new forms of FGC impact on firm-level performance. In the process, the research addresses Gordon et al.'s (2021) call for research capable of identifying both the short- and long-term financial impact of marketing activity that thus far remain difficult to quantify. By being able to distinguish between temporary and permanent price impacts at the fine-grained level of analysis, marketing managers can demonstrate both the short- and long-term impacts of FGC on firm financial performance (Magill, Moorman, Avdiushko, 2019). This in turn, will allow them to overcome short-termism in marketing, improving long-term growth initiatives (Du et al, 2021; Moorman and Kirby, 2019).

Finally, to support intraday actionable FGC design, we examine the extent to which key attributes of FGC, including content valence and subject matter, influence the occurrence of FGC permanent and temporary price impacts. Although FGC valence and subject matter have been examined by previous research (e.g. Elliott, Grant and Hodge, 2018; Groß-Klußmann, König, and Ebner, 2019; Hewett et al., 2016), and are recognized as key components of marketing excellence (Homburg, Theek and Hohenberg, 2020), what constitutes the 'right content' is largely unknown according to research priorities recently published by the Marketing Science Institute (2020-2022). We show that FGC reflecting only one of the attributes: valence (positive or negative) or subject matter (consumer or competitor orientation), generates temporary price impact, while FGC that incorporates both valence and subject matter is associated with permanent price impacts in stock price and thus, they correlate with long-term firm performance.

Using a two-stage least squares (2SLS) estimation framework to examine S&P 500 IT firms' Twitter activity, we find that negative or positive valence tweets are consistently linked with a reduction in permanent price impact and an increase in temporary price impact. Similar findings are obtained when relating tweets that only reflect a consumer or competitor subject matter, although the reduction in permanent price impact and increase in temporary price impact they elicit are of smaller magnitudes. These findings indicate that tweets reflecting only valence (positive or negative) or subject matter (consumer or competitor orientation) result in temporary price impacts, which is commonly associated with the incorporation of noise into the price discovery process (O'Hara, 2003). This type of effect has not been studied previously in the marketing literature; however, given that it is a symptom of uncertainty in the value of firms that can increase a firm's cost of capital (Diamond and Verrecchia, 1991), it demands attention. Employing the market microstructure approach to exploit the variance of high frequency changes in stock price, this is the first study that reveals tweets' temporary price impacts and identifies tweet attributes that elicit such short-term price impacts.

We further find that tweets that embody both attributes; valence and subject matter, generate permanent price impacts; however, this impact varies based on the type of valence and subject matter. The results evidence the importance of interaction effects between tweet valence and subject matter in generating a higher permanent price impact. The average negative and positive valence tweet when viewed through the lens of consumer or competitor orientation generates a permanent price impact, while a competitor-oriented tweet with a negative valence is likely to have the highest permanent price impact. This is a crucial finding from the perspectives of marketing practice and intraday social media marketing strategy design because valence as a singular attribute is associated with decreasing permanent price impact. Our research shows that information-rich tweets that include both variance and subject matter can result in permanent price impacts, and underscores investors' ability to act on information contained in FGC at sub-second levels (Hendershott et al., 2011; Rzayev and Ibikunle, 2019).

To illustrate the relevance and magnitude of these findings, in Figure 2, tweets A and B are characterized by negative and positive valence, respectively; while tweets C and D reflect only consumer and competitor orientation, respectively. In line with our research findings, the permanent price impact associated with these tweets are more than three standard deviations lower than the average permanent price impact estimate for all the 153,041 tweets in our sample and are therefore below the 10th percentile of the estimates. By contrast, tweets E to H reflect varying combinations of both valence and subject matter (consumer or competitor orientation). Consistent with the research findings, these tweets are shown to generate permanent price impact estimates ranked above the 90th percentile in our sample of tweet trades' permanent price impact estimates. The temporary and permanent impact estimates for the average tweet are as large as 279 and 187 times, respectively, of what we document for the average regular intraday transaction in our sample.

<Insert Figure 2.>

Theoretical Background

Firm-generated content and high frequency trading data

Social media is increasingly used by firms because it provides greater reach and can be less costly than traditional channels for firm-generated content (FGC) dissemination (Kumar et al., 2016). FGC is often posted several times a day (Kanuri, Chen and Sridhar, 2018) and serves as a valuable source of high frequency marketing data that can offer insights into the growth potential of a firm (Du et al., 2021). With the advancement of data collection tools (Wedel and Kannan, 2016), marketing researchers can now record each piece of FGC and create large datasets depicting FGC attributes and their dissemination time. Recorded with accuracy to the second, FGC can then be mapped against the corresponding trading activity that takes place at sub-second intervals and used to study FGC's financial impact. However, measuring the impact of FGC sampled at intraday levels requires marketing researchers to be able to utilize high frequency trading data with observations occurring at unequal time intervals.

The market microstructure approach to estimating price impact offers marketing researchers tools to, piece-by-piece, algorithmically link FGC to time-specific trading activity at a fine-grained level of analysis (i.e., sub-seconds, seconds, minutes etc.). Unlike symmetrical asset pricing models, market microstructure recognizes that a firm's stock price is only informationally efficient to the extent that it reflects all available and relevant information (Fama, 1970). A firm's stock price, while reflecting information, is also distorted by noise generated by (temporary) non-information-based factors, such as trading frictions occurring due to low levels of liquidity defined as the ability to trade large quantities of a firm's stock quickly with little or no price impact (Amihud, 2002; Grossman and Miller, 1988) or the activity of traders lacking adequate information regarding the value of a stock, the so-called uninformed traders in the market microstructure literature (Glosten and Milgrom, 1985; Kyle, 1985). Estimating the proportion of stock price driven by information (relevant to the value of a firm), and the proportion driven by noise, on the other hand, is a critical aspect of the analyses many studies conduct in the market microstructure literature area (see Web Appendix D). However, this holistic view of both temporary and permanent price impacts is often missing from marketing research. The market microstructure approach allows marketing researchers to estimate the price impact of FGC at high frequencies, and to identify both types of price impacts. A crucial step in such analysis is knowing the so-called 'event time' (i.e. a timestamp), which refers to the time an event occurs, such as the FGC dissemination time. By deploying time series models to estimate changes in the components of price at high frequency intervals (e.g. seconds), and then linking the FGC timestamp to the components, the instantaneous impact of FGC on firm stock price can be estimated.

FGC's impact on firm outcomes

The existing research has primarily focused on linking FGC with consumer behavior (Colicev et al., 2018; Colicev, Kumar and O'Connor, 2019; Hewett et al., 2016; Kumar et al., 2016; Meire et al., 2019; Tellis et al., 2019), and firm performance (Borah et al., 2020; Colicev et al., 2018; Rust et al., 2021) (see Web Appendix E).

With the focus on firm performance, Colicev et al. (2018) and Borah et al. (2020) and most recently Rust et al. (2021) show that FGC effects firm value. Colicev et al. (2018) document an indirect effect of FGC volume on shareholder value measured based on abnormal returns and idiosyncratic risks. Borah et al. (2020) were the first to demonstrate the direct impact of humorous FGC on firm value as measured by abnormal stock market returns. To demonstrate these impacts, they employ an event study estimating abnormal returns and VAR modeling. These methods, as deployed, however, focus on daily activity, which can result in aggregation bias and misevaluation of FGC's impact on firm value (Rust et al, 2021). Moreover, the richness of such an assessment is compromised because short- and long-term impacts of FGC are not estimated (Gordon et al., 2021; Moorman and Kirby, 2019). Finally, current marketing methods do not examine FGC attributes at a fine level of granularity (i.e. intraday frequencies), preventing marketing managers from moving beyond a 'throw it on the wall and see what sticks' strategy (Hoffman and Fodor, 2010: p 47) in the design and dissemination of intraday FGC (Hewett et al., 2016).

High frequency approach to FGC analysis

The market microstructure approach responds to calls for more powerful methodological approaches that allow marketing researchers to harness the potential of rich data sources and develop insights capable of advancing theory and informing contemporary marketing practice (e.g. Du et al., 2021; Hewett et al., 2016; Lamberton and Stephen, 2016; Wedel and Kannan, 2016). The fine-grained level of analysis available with a market microstructure approach overcomes the limitations of low-frequency methodologies, such as VAR and daily event studies to study FGC and its impact on firm value (i.e. it estimates the variance in firm stock price following FGC dissemination). Utilizing high frequency intraday data, it adds richness to the assessment of FGC's financial impacts by distinguishing between permanent and temporary price impacts.

The temporary and permanent price impacts

Temporary price impacts are short-term impacts that result in momentary changes in price before returning to its pre-event (e.g. pre-FGC) value, and are the result of uninformed trader activity (see Web Appendix D). Uninformed trader activity could be driven by several factors; for example, it could be linked to investor uncertainty about the relevance of information (Hedge and McDermott, 2003; Holthausen, Leftwich and Mayers, 1990), or trading friction due to liquidity constraints (Amihud, 2002; Chordia, Roll and Subrahmanyam, 2008). Ignoring temporary price impacts can lead to misunderstanding the total impact of FGC, with prior research suggesting temporary price impacts result in larger transaction costs (Chan and Lakonishok, 1993) and firm cost of capital (Diamond and Verrecchia, 1991). In contrast, an event (e.g. FGC) can generate a permanent price impact and result in the price attaining an enduring new value after the event; this occurs when the event provides information that updates informed investor/trader expectations related to a firm's long-term performance (Madhavan, Richardson and Roomans, 1997). Importantly, the microstructure approach also supports intraday actionability by assessing how the attributes of information signaled by these events influence temporary and permanent price impacts.

A state-space decomposition of firm stock price

Consistent with the market microstructure literature, this study estimates the permanent and temporary price impact of FGC by first conducting a state-space decomposition of firm stock price into its efficient (permanent) and inefficient/noise (temporary) components and then linking the changes in these components to individual pieces of FGC. State-space modeling is a tool for modeling an observed variable as the sum of unobserved variables (Hendershott and Menkveld, 2014), and it is commonly used for the decomposition of price (Brogaard, Hendershott and Riordan, 2014; Hendershott and Menkveld, 2014; Menkveld, Koopman and Lucas, 2007; Rzayev and Ibikunle, 2019). Due to its efficiency when applied to ultra-high frequency data like stock price movements, the state-space modeling approach for decomposing price has significant economic and methodological advantages over other commonly used methods (Hasbrouck, 1991), such as VAR models.

An assumption underlying a standard VAR model is that data are sampled at regular frequencies since variables at time t are regressed on variables dated at t-1, t-2, etc. However, FGC and intraday trading data are often sampled at unequal time intervals, which suggests that there will be many instances of missing variables in a model calibrated on regular time intervals (Rao, 1986). The modeling of such data using VAR requires the alignment of variables misaligned in time either downward, by aggregating the data to a lower frequency, or upward, by interpolating the high frequency data with heuristic rules such as polynomial fillings. Downward alignment eliminates potentially valuable information in the high frequency data. Data aggregation is problematic (Silvestrini and Veredas, 2008): it can alter the lag order of autoregressive moving average (ARMA) models (Amemiya and Wu, 1972), reduce the efficiency of the parameter estimation and forecast (Tiao and Wei, 1976), affect Granger-causality and cointegration among component variables (Marcellino, 1999), and induce spurious instantaneous causality (Breitung and Swanson, 2002). Upward alignment is also deemed inefficient and dubious (Pavia-Miralles, 2010) because a VAR approach assumes that the model specifies the high frequency data-generating process. However, interpolation is not based on the multivariate model that generates the data, but on heuristic rules, which, at a minimum, inevitably incorporate noise into the data and distort it.

State-space modeling offers a solution to the irregular frequency challenge inherent in intraday transactions data (Qian, 2013). Specifically, the use of state-space modeling with a Kalman filter in maximum likelihood estimation of parameter estimates ensures maximum efficiency in dealing with unequal time intervals or irregular frequency in data. The use of a Kalman filter accounts for changes across periods of analysis with missing observations. This is a critical consideration in the use of state-space modeling for decomposing high frequency time series since standard approaches do not deal with the ‘missing observations’ caused by unequal data intervals. For example, estimating a standard autoregressive (AR) framework implies truncation of the lag structure and could potentially discount valuable information in high frequency data. Using the Kalman filter facilitates the decomposition of any realized change in the time series (e.g. variance in the stock prices), such that the estimated permanent or temporary component at any interval is estimated using all past, present, and future observations in the series. Thus, the purpose of filtering is to ensure that estimates are updated with the introduction of every additional observation (Durbin and Koopman, 2012).

Heterogeneously informed traders and FGC attributes

With the estimation of FGC's temporary and permanent price impacts, marketing researchers can explore how FGC attributes influence the occurrence of these two types of price impact, which are driven by the existence of heterogeneously informed trading agents in financial markets (Grossman and Stiglitz, 1980; O'Hara, 2003). Thus, the ways in which information events, such as FGC, are observed and deciphered vary significantly between the two main groups of agents in financial markets, i.e. the informed and uninformed traders/investors (see Web Appendix D for a discussion on how the activities of informed and uninformed traders drive the asymmetric effects of information events in financial markets). The valence and subject matter of FGC are attributes that should provide information signals to (informed) investors, and thus generate a permanent price impact. This is because FGC subject matter (consumer and competitor orientation) relate to a firm's competitive advantage (Kumar et al., 2011; Lam, Kraus and Ahearne, 2010), which is not often public and can be difficult to observe because it is embedded in a firm's culture (Gebhardt, Carpenter and Sherry, 2006). The role of valence has also been documented in the literature (Sul, Dennis and Yuan, 2017; van Heerde, Gijsbrechts and Pauwels, 2015), with the impact of negative valence appearing to be stronger than that of positive valence and thus more commonly associated with permanent price impacts (Tirunillai and Tellis, 2012). There is also reason to expect that FGC valence may interact with FGC subject matter and induce a permanent price impact. The basis for this expectation comes from a branch of signaling theory recognizing that signal recipients combine information signals to make more informed decisions (Bhagwat et al., 2020; Tellis and Wernefelt, 1987).

While there is ample evidence to suggest that FGC valence and subject matter may generate a permanent price impact, it is important to note that the price impact of FGC cannot occur without trading in financial markets. It is trading activity that incorporates the information and/or noise content of an event (e.g. FGC) into price. Therefore, since trading agents in financial markets are heterogeneously informed due to the way in which they observe and decipher the information content of events, their trading activities also generate varied price impacts. Specifically, a permanent price impact will arise as a result of trading activity by traders/investors who have been able to correctly decipher the information content of FGC; these are the informed traders. Conversely, the trading activity of those unable to decipher the information content of FGC (i.e. uninformed traders) will only induce temporary price impacts (Glosten and Milgrom, 1985) since their trading activity is uncorrelated with firm value (Barclay and Warner, 1993; Grossman and Stiglitz, 1980). Accordingly, FGC that incorporates valence and subject matter can be associated with both permanent and temporary price impact simply because of heterogeneously informed trading agents. The trading activity of informed traders thus contributes to the efficient component of price (i.e. permanent price impact), which is driven by information, while the activity of uninformed traders incorporates noise, which is uncorrelated with firm-relevant information (i.e. temporary price impact).

S&P 500 IT firms' use of Twitter

We examine the instantaneous stock market impact of FGC by studying S&P 500 IT firms' activity on Twitter. Twitter is a social media communication channel characterized by 'fast-paced and short-lived information flows' (Lambrecht, Tucker and Wiertz, 2018: 177), which is said to derive deep insights once appropriate methods are developed (Webel and Kannan, 2016). In addition, with 92% of firms tweeting multiple times a day (Brandwatch, 2020), Twitter FGC is an example of high frequency intraday marketing data. Finally, the Exchange Commission's Regulation Fair Disclosure recognizes Twitter FGC as potentially carrying 'useful' information for investors. For these reasons, Twitter provides a suitable context to study. We study the IT sector because IT firms are often considered to be early adopters of trends (Blankespoor, Miller and White, 2014). The IT sector provides a comprehensive sample of firms disseminating multiple pieces of FGC throughout the day (see Web Appendix F). It is a major driver of economic activity, with the leading five IT firms in the US accounting for more than 22% of the S&P 500 (Hargreaves Lansdown, 2020). Globally, the IT sector is valued at $11.5 trillion, representing over 15.5% of the global GDP (Brookings, 2019). Finally, the diverse consumer base of IT firms is useful for characterizing the relevance of FGC subject matter (consumer and competitor orientation), and its interaction effects with valence. A review of 10-K filings for each firm shows that 7% of the sample consists of firms marketing solely in B2C markets, 72% solely in B2B markets, and 21% selling in both B2C and B2B markets.

Dataset construction

We obtain a sample of tweets using an Application Programming Interface (API) to access data from Twitter. In line with previous research (Lambrecht, Tucker and Wiertz, 2018; Vermeer et al., 2019), and following Chan et al. (2016), we employ the API to access tweets from corporate accounts for S&P 500 IT firms. In total, we obtain 153,041 firm-generated tweets from 64 firms, which are then used in our analysis. On average, this is 2,391.2 tweets per firm over our sample period spanning January 2013 to August 2018. It should be noted that these are tweets that fall within the limits of Twitter API in terms of the maximum number of tweets that can be accessed over a given time period. Seven of the IT firms initially selected either do not have established corporate social media accounts on Twitter or Twitter API limited access to the data. Firms included in the sample engage in high frequency intraday marketing activity; on average they generate a minimum of 1.07 to a maximum of 37.03 tweets a day, with the total average equaling 4.42 tweets per firm per day (see Web Appendix F), which confirms the appropriateness of the selected sample. Table 1 shows the sample of 10 S&P 500 IT firms generating the highest number of tweets per day.

<Insert Table 1 about here>

Each tweet is recorded with a timestamp to the nearest millisecond.[2] These timestamps are used to obtain corresponding ultra-high frequency stock trading activity data from the Thomson Reuters Tick History (TRTH) v2 database in Datascope for each tweet in the sample (see Table 2); the stock trading data supplements the Twitter dataset. Our dataset includes data for the trading days between January 2013 and August 2018. After performing data cleaning using the criteria consistent with that of Chordia, Roll and Subrahmanyam (2001) and Ibikunle (2015), the stock trading data includes 8,177,183,865 instances of trading activity or messages (i.e. quotes, cancellations and transactions), which includes 520,356,393 transactions, and 7,656,827,472 orders.[3]

<Insert Table 2 about here>

After excluding days (and tweets generated on these days) with comparatively high levels of price volatility, the descriptive statistics show that the average time between trades is 7.159 seconds, and the mean number of tweets per firm over the sample period is 2,377.22. The mean number of tweets per day is 54.03 and the mean number of tweets per day per firm is 0.844 (see Table 3 for details).

<Insert Table 3 about here>

[2] 1,179 tweets, 0.77% of the total sample (see Web Appendix F), are excluded from the analysis because of excessive stock price return volatility on the days they occur. This is a standard approach commonly employed in market microstructure literature. We also define an exclusion criterion in order to exclude tweets that occur within 60 seconds of each other. However, none of the tweets in our data occur within 60 seconds of each another; therefore, no tweet is excluded on the basis of the exclusion criteria.

[3] There are three types of observations in our dataset. The first are the buy/bid and sell/offer quotes (or orders), while the second are the transactions or trades, which feature directly in the state space modeling and are generated as a result of the orders being executed. For the model, we only employ the prices of the 520,356,393 transactions in the dataset. The third type of observations are cancellations issued to cancel previously submitted orders. All the observations are captured using timestamps to the nearest millisecond.

Investigating the permanent and temporary price impact of tweets

To investigate the impact of tweets from a perspective of permanent and temporary price impact respectively, we first use state-space modeling to estimate the permanent and temporary components of price at a given time interval using trading observations within that time interval.[4] The primary interval of interest is one second; however, we estimate for one minute as well for robustness. Next, we link these estimates to firms' tweet activity using tweets' time stamps, which are labelled to the second. Thereafter, we estimate the impact of each tweet on the temporary and permanent components of price by estimating the corresponding change in the components following each tweet as the respective temporary and permanent price impacts. The methodological steps are outlined below:

Step one (model characterization): The first step involves modeling price as the sum of a non-stationary permanent (information-driven) component and a stationary temporary (noise) component.[5] In this step, the only relevant observations are the prices of the 520,356,393 transactions obtained from the TRTH v2 database; these prices are defined as the prices of stocks at intraday periods and intervals. In its simplest form, the structure of the state-space model for price, a multiple of S stock prices, T intraday periods, and N intervals, are expressed as:

$$v_{s,t,\tau} = m_{s,t,\tau} + i_{s,t,\tau} \tag{1}$$

and

$$m_{s,t,\tau} = m_{s,t,\tau-1} + u_{s,t,\tau}, \tag{2}$$

[4] While v (stock price in our model) is observable, its permanent and transitory components, which we aim to characterize, are unobservable, i.e. we cannot acquire them as we would observable variables, such as stock price or volume. Thus, we aim to observe the evolution of that one variable – v – that we could observe and use this evolution within time intervals (e.g. one second and one minute) to estimate its components.

[5] In addition to modeling the natural logarithm of price as an observable variable in the state-space representation, for robustness, we also employ percentage change in price, and first difference of price. Our inferences are unchanged irrespective of the approach we employ; indeed, all the estimates obtained are qualitatively similar.

where

$$v_{s,t,\tau} = \ln(p_{s,t,\tau}), \tag{3}$$

for s = 1,…,S, τ = 1,…,T, and t = 1,…,N; τ and t index event and clock times respectively (Menkveld, 2013); an event occurs when a transaction is recorded. Hence, T =520,356,393, while N equals the number of one second or one minute intervals during a stock trading day. $p_{s,t,\tau}$ is the price of stock s at interval t and period τ, $m_{s,t,\tau}$ is a non-stationary permanent component of the price of stock s at interval t and period τ, $i_{s,t,\tau}$ is a stationary transitory component of the price of stock s at interval t and period τ, and $u_{s,t,\tau}$ is an idiosyncratic disturbance error in the permanent price component of stock s at interval t and period τ. $i_{s,t,\tau}$ and $u_{s,t,\tau}$ are assumed to be mutually uncorrelated and normally distributed. [6]

The model captured in Equations (1) – (3) is a special case of the general state-space representation. The standard state-space model is formulated for a vector of time series $\mathbf{v_t}$ with a frequency/time interval t and this is given by (for simplicity, we temporarily ignore the stock notation s and period τ):

$$\mathbf{v_t} = \mathbf{W_t}\delta + \mathbf{Z_t m_t} + \mathbf{i_t}, \qquad \mathbf{m_{t+1}} = \mathbf{D_t m_t} + \mathbf{R_t u_t}, \quad t = 1,..,N, \tag{4}$$

where disturbances $\mathbf{i_t} \sim N(\mathbf{0}, \mathbf{I_t})$ and $\mathbf{u_t} \sim N(\mathbf{0}, \mathbf{U_t})$ are mutually and serially uncorrelated. The initial state vector $\mathbf{m_1} \sim N(\mathbf{a}, \mathbf{P})$ is also uncorrelated with the disturbances. The mean vector $\mathbf{a}$ and variance matrix $\mathbf{P}$ are usually implied by the dynamic process for $\mathbf{m_t}$ in Equation (4) (Menkveld, Koopman and Lucas, 2007). The remaining terms, $\mathbf{W_t}$, $\mathbf{Z_t}$, $\mathbf{D_t}$, $\mathbf{R_t}$, $\mathbf{I_t}$ and $\mathbf{U_t}$, are called system matrices and are generally assumed to be fixed for $t = 1,..,N$. The elements of these system matrices are usually known; however, some elements that are functions of the fixed parameter vector need to be estimated. Equations (1) and (2) can be represented as the

[6] According to Merton's (1986) model, when investors hold under-diversified portfolios, idiosyncratic risk should be priced. $u_{s,t,\tau}$ in Equation (2) captures the effect of idiosyncratic risk as a function of information, and it is different from non-information-based (temporary) evolution in stock price captured by $i_{s,t,\tau}$.

state-space Equation (4) by choosing $\mathbf{v_t}$ as a single time series (in this study this is the stock price series), where $\mathbf{W_t} = 0$, $\mathbf{Z_t} = \mathbf{D_t} = \mathbf{R_t} = 1$, $\mathbf{I_t} = \sigma^{2_i}$ and $\mathbf{U_t} = \sigma^{2_u}$. We note that σ^{2_i} and σ^{2_u} vary for each frequency t for $t = 1,..,N$. Unlike standard variable decomposition approaches, this model naturally deals with irregular frequency/missing observation issues since the Kalman filter is used for its estimation, which is critical in a high frequency analysis.[7]

Step two (model outputs): The structure of the model shows that only changes in $u_{s,t,\tau}$ (now reinstating the stock notation s and period τ) affect price permanently; $i_{s,t,\tau}$ is temporary because its effects are transient and hold no significance for long-term firm performance. This is because this model decomposes price into two parts. The first, $u_{s,t,\tau}$, captures smoothed (constant) changes in price, which is driven by informed trading activity, while the second captures irregular changes in price, which deviates from the smoothed evolution and is therefore driven by uninformed trading activity (noise or friction in the pricing process). By using maximum likelihood (constructed using the Kalman filter), we estimate $\sigma^{2_u}_{s,t}$ (i.e. permanent component) and $\sigma^{2_i}_{s,t}$, (i.e. temporary component) where t is equal to either one second or minute. Specifically, we first partition our sample into one second and one minute (clock) intervals, then estimate $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$ for these intervals by using the prices at different event periods (τ) during the intervals. This suggests that, as in Menkveld, Koopman and Lucas, (2007), our permanent and temporary components ($\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$), as estimated using the state-space model, are time variant (see Table 4 in Menkveld, Koopman and Lucas, (2007: 220)). We impose the time-variant structure to be consistent with the time intervals studied in

[7] Some adjustments are required. When there are instances of missing (or irregularly spaced) observations in $\mathbf{v_t}$, the Kalman filter is unable to use the measurement equation (Equation 1); however, the transition equation (Equation 2) can be used since it depends on the previously estimated state ($\mathbf{m_{t+1}}$depends on $\mathbf{m_t}$). Indeed, Kalman filtering suggests that with missing observations in $\mathbf{v_t}$, the best estimation for $\mathbf{m_t}$ is simply the evaluation of the transition equation The estimated state-space model's source code in SAS is presented in Selukar (2011).

subsequent multivariate regressions, which is in line with Brogaard, Hendershott and Riordan (2014) who also compute time-variant permanent and transitory components of price.

Step three (estimation with the Kalman filter): We use the Kalman filter to evaluate the conditional mean and variances of the state vector $\mathbf{m_t}$ (ignoring the stock notation s and period τ) given past observations $V_{t-1} = \{\mathbf{v_1},..,\mathbf{v_{t-1}}\}$: $\mathbf{a_{t|t-1}} = E(\mathbf{m_t}|V_{t-1})$, $\mathbf{P_{t|t-1}} = \mathrm{var}(\mathbf{m_t}|V_{t-1})$, $t = 1,..,N$. To initialize the Kalman filter, we also have $\mathbf{a_{1|0}} = \mathbf{a}$ and $\mathbf{P_{1|0}} = \mathbf{P}$, where $\mathbf{m_1} \sim N(\mathbf{a}, \mathbf{P})$. This initialization only works if $\mathbf{m_t}$ is a stationary process. However, as in our case, $\mathbf{m_t}$ is often not a stationary process due to its being obtained from stock price series, which is inherently non-stationary given the rational expectation of economic growth over time. Hence, 'diffuse initialization' (i.e. infinite variance distribution – see Koopman and Durbin, 2003) is used and estimated by numerically maximizing the log-likelihood. This is evaluated with the Kalman filter due to prediction error decomposition. According to the structure of the state-space model, our estimated outputs, $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$, are modeled as variances of permanent and temporary components of price respectively. $\sigma^{2_u}_{s,t}$ is a proxy for information reflected in the price, i.e. the permanent component of price, and $\sigma^{2_i}_{s,t}$ is a proxy for noise reflected in the price, i.e. the temporary component of price. Stock prices should only experience permanent movements due to the arrival of new information, thus we would expect $\sigma^{2_u}_{s,t}$ to be higher than $\sigma^{2_i}_{s,t}$. The two estimated coefficients are variances; hence, the coefficient encapsulating information$\left(\sigma^{2_u}_{s,t}\right)$, which is the primary driver of price from an efficient market perspective, should be larger; $\sigma^{2_i}_{s,t}$ captures frictions/noise and should therefore have a lower value.[8]

Step four (linking $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$ to tweets): Our empirical framework requires linking an individual intraday tweet to a corresponding trade/transaction with price p_t in our sample. We

[8] The code we estimate is made available via a public GitHub repository here: https://github.com/akataehonda/Twitter-Project.git

call each tweet-linked trade a 'tweet-trade'; t, in this case, corresponds to both trade and time. Accordingly, we designate a trade in the stock of a firm as a 'tweet-trade' if it is the first trade to occur immediately after a tweet in our sample, and if it occurs within 60 seconds of the tweet. For robustness, we vary this threshold but find the inferences to be consistent if the threshold is reduced to 30 and 45 seconds, suggesting that the occurrence of tweet-trade and trading in stock is not merely coincidental. The tweet-trade's time of occurrence allows us to link a tweet to a corresponding pair of $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$, which we estimate for the one-second and one-minute intervals covered by our sample period, including those with no tweet-trades. Thus, each second and minute in our sample has a corresponding set of $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$. We can therefore determine the information reflected in the price (i.e. price efficiency) and noise contained in price at every second or minute. This information allows us to estimate whether the change in both components is occasioned by the posting of a tweet. Table 4 presents the descriptive statistics for the one-minute intervals, including a tweet-trade. $\sigma^{2_u}_{s,t}$ is higher than $\sigma^{2_i}_{s,t}$, which is consistent with our expectation that most of the observed tweets at time t reflect fundamental information rather than frictions or transitory components of price. This is also in line with microstructure literature (Brogaard, Hendershott and Riordan, 2014; Hendershott and Menkveld, 2014; Menkveld, Koopman and Lucas, 2007; Rzayev and Ibikunle, 2019).

<Insert Table 4 about here>

Step five (estimating changes in $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$ following a tweet): The next step in our analysis is determining how a tweet/tweet-trade changes the composition of price with regards to $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$, which is required in further analysis when examining the impact of tweet valence and subject matter (i.e. consumer and competitor orientation). We link each tweet-trade to a pair of $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$ using the tweet-trade timestamps at the second level and then compute 30-second percentage absolute changes for both $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$; the changes in $\sigma^{2_u}_{s,t}$ and $\sigma^{2_i}_{s,t}$ following a tweet are designated as $\Delta\sigma^{2_u}_{s,t}$ (permanent price impact) and $\Delta\sigma^{2_i}_{s,t}$ (temporary price impact) respectively (45- and 60-second percentage changes are also computed for robustness):[9]

$$\Delta\sigma^{2_u}_{s,t} = \left| \frac{\sigma^{2_u}_{s,t+30s} - \sigma^{2_u}_{s,t-1}}{\sigma^{2_u}_{s,t-1}} \right| \tag{5}$$

$$\Delta\sigma^{2_i}_{s,t} = \left| \frac{\sigma^{2_i}_{s,t+30s} - \sigma^{2_i}_{s,t-1}}{\sigma^{2_i}_{s,t-1}} \right| \tag{6}$$

Thereafter, we also construct $\Delta\sigma^{2_u}_{s,t}$ and $\Delta\sigma^{2_i}_{s,t}$ for each non-tweet-trade in our sample. Using these measures, we construct daily ratios of the impact of non-tweet-trade relative to that of an average tweet-trade in stock s during day d. We then test the null that the mean daily ratios in stock s equal one on average across our sample period by using their standard errors for statistical inference. We expect to reject the null if the tweet-trades on average generate a larger or lower price impact than all the trades on an average day.[10] We present the results of the hypotheses testing in Table 5. The ratios employed in the analysis are winsorized at 0.5 and 99.5 percentiles within each stock. This statistical approach is consistent with prior marketing research (Boyd and Kannan, 2018), and it allows us to eliminate outliers or extreme values and improve the chance of obtaining statistically significant estimates. Winsorization is also necessary due to the inherent noisiness of high frequency trading data used in estimating t$\Delta\sigma^{2_u}_{s,t}$ and $\Delta\sigma^{2_i}_{s,t}$.

<Insert Table 5 about here>

[9] Note that the percentage change is from a trade at t-1 before the tweet-trade to 30 seconds after the tweet-trade; varying this measurement for up to five trades t-5 before the tweet-trade does not significantly impact the estimates obtained, neither does varying the time threshold to include 45-second and 60-second percentage changes. Estimating the effects of tweet-trades within sub-minute to minute windows avoids the methodological issues associated with the occurrence of confounding events. Given the fine-grained level of measurement, it is highly unlikely that any other relevant event could be driving the effects we capture.

[10] In order to ensure that the results are not driven by unusual trading days, we exclude days where stock return volatility is greater than the one standard deviation of the average stock return volatility over the surrounding (-30, +30) trading days. Daily volatility is measured as the standard deviation of intraday stock return.

The estimates in Table 5 show that, on average, tweet-trades generate larger permanent and temporary intraday price impacts than other non-tweet-trades. All the estimates are statistically significant at the 0.01 level, thus rejecting the null hypothesis that there is no difference between the impact generated by tweet-trades and other trades. The price impact of a tweet-trade is 150–300 times larger than that of the average trade. Using the 60-second threshold, $\Delta\sigma^{2_i}_{s,t}$, which corresponds to the temporary price impact generated by the average tweet-trade, it is 279.67 times that of the average trade, suggesting that tweets results in large but momentary movements of price. This finding suggests that FGC generates temporary effects that can induce increases in the cost of trading a firm's stock and the firm's cost of capital Chan and Lakonishok, 1993; Diamond and Verrecchia, 1991). $\Delta\sigma^{2_u}_{s,t}$, the permanent impact of the average tweet-trade, is about 146.83 times larger than the average trade's permanent impact, suggesting that FGC can cause investors to update their expectations about a firm's future performance and this leads to price movement. Overall, the analysis indicates that, on average, tweet-trades occurring in the wake of a potentially information-laden tweet substantially impact stock price in both permanent and temporary effects relative to non-tweet trading activity.

Estimating the effects of tweet-trades within sub-minute to minute windows addresses methodological issues associated with the occurrence of confounding events. Therefore, to a very high level of accuracy, we can attribute estimated temporary and permanent price impacts to the observed FGC. Given the fine-grained level of analysis, it is highly unlikely that any other relevant event could be driving the effects we capture. The sampling at high frequency intervals also raises the question of whether investors and other trading agents could digest and act on the contents of tweets within the price impact windows we examine. Addressing this question requires an understanding of the nature of trading in financial markets today, especially in the case of highly traded stocks, such as the S&P 500 stocks in our sample.

Today's markets are dominated by algorithmic traders/'algos' capable of digesting and acting on information in FGC (e.g. tweets) within the windows we examine in our analysis. The effects of this speed of activity are evidenced by the findings of Rzayev and Ibikunle (2019), who, by using S&P 500 stock data, show that information arriving in the US markets is exploited within seconds and that this activity is driven by algorithmic trading.

While all the ratios are statistically significant and suggest that FGC influences the permanent and temporary components of price, the obvious question is how economically meaningful they are when compared to other events impacting stock price. To answer this question, we conduct further analysis to examine the corresponding ratios of other large impact non-tweet-trades in the same period by computing ratios similar to the ones presented in Table 5. This involves substituting a permanent price impact measure for each tweet-trade with that of other trades generating price impacts corresponding to one standard deviation or more above the daily mean in each stock. The obtained average ratios for the three thresholds are 7.9, 5.2, and 1.3 for the 30-, 45-, and 60-second windows respectively. The inference drawn from this analysis is that the information content of tweet-trades is several times higher than that of the average non-tweet high impact trade. In comparing the temporary price impacts associated with the same trades with those of the tweet-trades, we find that tweet-trade ratios are again several times higher. This suggests that tweet-trades tend to be noisier when compared to other trades associated with a more permanent price impact and this provides a basis for demonstrating to marketers the significance of the relatively high levels of both permanent and temporary price impacts that can be generated in financial markets with the use of tweets. A robustness comparative analysis based on Frino, Jarnecic and Lepone (2007) is consistent with the presented findings (please see Web Appendix G).

Investigating the temporary and permanent price impact of tweet valence and subject matter

To add intraday actionability, we use $\Delta\sigma^{2_u}_{s,t}$ and $\Delta\sigma^{2_i}_{s,t}$, encapsulating the permanent and temporary impacts of intraday tweets on firm value, as dependent variables to determine how tweet valence and subject matter (consumer and competitor orientation) influence the impact of tweets on stock price. To investigate whether tweet valence and subject matter drive the price impact of tweet-trades, we estimate Equation (7):

$$\text{Price Impact}_{s,t} = \alpha_s + \beta_t + \gamma_1 \text{consumer}_{s,t} + \gamma_2 \text{competitor}_{s,t} + \gamma_3 \text{consumer} * -\text{ve}_{s,t} + \gamma_4 \text{competitor} * -\text{ve}_{s,t} + \gamma_5 \text{consumer} * +\text{ve}_{s,t} + \gamma_6 \text{competitor} * +\text{ve}_{s,t} + \gamma_7 -\text{ve}_{s,t} + \gamma_8 +\text{ve}_{s,t} + \sum_{k=1}^{7} \varphi_k C_{k,s,t} + \epsilon_{s,t} \quad (7)$$

where $\text{Price Impact}_{s,t}$ corresponds to $\Delta\sigma^{2_u}_{s,t}$ or $\Delta\sigma^{2_i}_{s,t}$ respectively for a tweet-trade t in stock s. α_s and β_t are stock and time fixed effects. We use the VADER rule-based algorithm (Hutto and Gilbert, 2014) to determine the valence of the tweets. VADER out-performs other commonly used benchmark methodologies such as LIWC, ANEW, and the machine learning algorithm SVM in the literature as well as in our robustness tests. We also utilize Saboo and Grewal's (2013) library and follow their method in measuring the competitor ($\text{competitor}_{s,t}$) and consumer ($\text{consumer}_{s,t}$) orientation for each tweet, which is in line with Atuahene-Gima (2005) and Voss and Voss (2000). Consumer and competitor subject matter are dummy variables equaling one when a tweet-trade's content is about consumer and/or competitors. We also study the interaction effects of these attributes; $\text{competitor} * +\text{ve}_{s,t}$ and $\text{competitor} * -\text{ve}_{s,t}$ refer to positive valence tweets related to competitors and negative valence tweets related to competitors respectively for a tweet-trade t in stock s, and $\text{consumer} * +\text{ve}_{s,t}$ and $\text{consumer} * -\text{ve}_{s,t}$ refer to positive valence tweets related to consumers and negative valence tweets related to consumers respectively for a tweet-trade t in stock s.

To avoid omitted variable bias and to ensure completeness, the model also includes $C_{k,s,t}$, which reflects a vector of known determinants of price impact based on past research in the market microstructure literature, as well as the natural logarithm of the number of an account's followers at the time of a tweet-trade t's tweet ($\#followers_{s,t}$). $C_{k,s,t}$ includes the natural logarithm of trading volume ($lnvolume_{s,t}$), the natural logarithm of average trade size ($lntradesize_{s,t}$), volatility ($volatility_{s,t}$), effective spread ($Effectivespread_{s,t}$), the natural logarithm of a high frequency trading proxy ($HFT_{s,t}$), and order imbalance ($OIB_{s,t}$). We measure trading volume as the dollar volume of transactions executed in stock s prior to a corresponding tweet-trade t. Average trade size is computed as the trading volume prior to tweet-trade t divided by the number of transactions just prior to a corresponding tweet-trade t in stock s. $volatility_{s,t}$ is the standard deviation of mid-point dollar price returns from the start of the trading day up to the trade just before the corresponding tweet-trade t in stock s. $Effectivespread_{s,t}$ (in basis points) is computed as twice the absolute value of the last trade price less the prevailing price mid-point prior to the corresponding tweet-trade t in stock s divided by the prevailing price mid-point; price mid-point is the average of the prevailing best bid and ask prices. $HFT_{s,t}$ is the ratio of the number of messages (quotes, cancellations and transactions) to actual transactions from the start of the trading day until prior to a corresponding tweet-trade t in stock s. Finally, $OIB_{s,t}$ is the ratio of the difference between the number of sell and buy orders and the average of both from the start of the trading day until prior to a corresponding tweet-trade t in stock s. To eliminate outliers in the data due to the characteristic noisiness of high frequency trading data, all variables are winsorized at 0.5 and 99.5 percentiles within each stock.

We estimate Equation (7) using both panel least squares and two-stage least squares (2SLS) instrumental variable (IV) estimation approaches. PCSE errors are computed in order to obtain heteroscedasticity and autocorrelation robust standard errors. The IV estimation is undertaken in order to account for the likelihood of endogeneity due to selection bias caused by firm decision whether or not to use Twitter (Gong et al., 2017). The instrumental variable approach we employ is based on approaches adopted by an increasing number of studies in marketing literature (Whitler, Krause and Lehmann, 2018). For a particular firm in our sample of S&P 500 IT firms, it involves first identifying the firms in the same two-digit SIC that have sent a corresponding tweet on the prior or same day as the firm and then estimating the mean value of the potentially endogenous variables (consumer and competitor orientation) for these firms; the mean estimates are employed as an instrument for the particular firm. This variable meets the requirements for an instrument because price impact in the other firms' stocks is unlikely to be driven by tweeting in the particular firm and tweeting activity can be shown to be correlated for firms in similar industries. In each of the first stage regressions, we regress each of the consumer and competitor variables separately on the corresponding instrumental variables and the control variables defined above for each firm/stock and obtain the F-statistics as tests of the null of weak instruments. The fitted values for each of the measures from the first stage regressions are then employed as the variables in place of the consumer and competitor orientation variables in the second stage regressions.

The first-stage F-statistics, testing the null of weak instruments, show that our IV model does not suffer weak instrument issues. The test statistic is higher than the threshold of 10 needed for 2SLS inferences to be reliable when instrumenting for endogenous variables (Stock, Wright and Yogo, 2002). We also conduct further tests to examine the instruments' relevance and the validity of the over-identifying restrictions in the IV regressions. The Cragg-Donald and Kleibergen-Paap LM statistics we obtain reject the nulls of weak instruments and under-identification, based on the Hausman, Stock and Yogo (2005) critical values respectively. Essentially, these test the null hypothesis that the instruments we use have insufficient explanatory power to predict the endogenous variables in the model for identification of the parameters. All the p-values of the Sargan χ^2 test obtained also indicate that we cannot reject the null that the over-identifying restrictions are valid. All the 2SLS estimates for Equation (7) are presented in Table 6, while the results of the panel least squares estimations are presented in Web Appendix H.

<Insert Table 6 about here>

The results presented in Table 6 show the importance of tweet valence and subject matter in determining the permanent and temporary impact of tweets on firm value. The existence of permanent and temporary price impacts associated with tweet attributes support the signal theory perspective (Kirmani and Rao, 2000) and show that investors pay attention to the tweet attributes of valence and subject matter. The estimates of permanent price impact for γ_7 and γ_8 are negative and statistically significant (-.063 p<.05 and -.108 p<.0,1 respectively). This suggests that tweets displaying only one of positive or negative valence are linked to less permanent price impacts in stock price. The positive and statistically significant γ_7 and γ_8 estimates of the temporary price impact estimation also indicate that they are linked to increasing temporary price impact (.032 p<.05 and <.033 p<.05, respectively), and suggest that tweet valence generally contributes more noise to stock price than stock-relevant information. The findings reinforce the role of positive and negative valence FGCs and their impact on firm value (Tirunillai and Tellis, 2012; van Heerde, Gijsbrechts and Pauwels, 2015).

With respect to tweet subject matter, only tweets conveying information about competitors generate statistically significant permanent price impacts (-.034 p<.05). On average therefore, tweets about a firm's competitors generate lower permanent price impact relative to other tweets. Conversely, the positive and statistically significant estimates for γ_2 for temporary price impact (.026 p<.05) show that these types of tweets are more likely to contribute to the noise component of price, i.e. they generate a larger temporary price impact than other tweets on average. Thus, tweets conveying competitor orientation appear to result in a lower permanent price impact, suggesting that this form of subject matter is comparatively less impactful and relevant to investors' expectations about a firm's future performance (Lam, Kraus and Ahearne, 2010). It is of note that tweets about consumers do not yield any permanent price impact that is statistically different from that of other tweets and thus by themselves do not appear to offer a signal capable of causing investors to permanently update their firm performance expectations. Consumer-related tweets, like those about competitors, also generate more temporary price impact than other tweets on average, which suggests that their potential for inducing noise in stock price is higher than that of the average tweet in our sample. The γ_1 and γ_2 estimates of the temporary price impact are positive and statistically significant (.009 p<.05 and (.026 p<.05, respectively). This finding implies that, as is the case with valence, tweets reflecting only one of competitor or consumer orientation generate noise in the price discovery or trading processes and lower permanent price impact.

Inferring from information-based market microstructure models (Kyle, 1985; Glosten and Milgrom, 1985), the more information about a firm that investors observe, the more they become informed about the valuation of the firm. In line with this expectation, the interaction variables we include in Equation (7) should yield positive estimates for the $\Delta\sigma^{2_u}_{s,t}$ estimations. As expected, all the γ_3, γ_4, γ_5 and γ_6 estimates of permanent price impact are positive and statistically significant (.047 p<.05, .606 p<.01, .088 p<.05 and .220 p<.01, respectively), even though, as already stated above, all of γ_1, γ_2, γ_7 and γ_8 are negative and statistically significant (except for γ_1). Thus, increases in both negative and positive valence, when viewed through the lens of subject matter, are linked with increased permanent price impact. These estimates show that tweet valence, when contextualized by subject matter or vice versa, is seen by investors/traders as firm-relevant information. In the context of these findings, the incorporation of valence and subject matter into FGC can yield increases in permanent price impact.

Furthermore, the findings suggest that tweets about competitors with a negative valence are likely to have the highest permanent price impacts (.606 p<.001). The finding is crucial from the perspective of marketing practice and intraday social media marketing strategy design because valence and competitor subject matter as singular attributes of FGC are independently associated with decreasing permanent price impact. The findings, underscore the view that investors seek additional information while making trading decisions (Bhagwat et al., 2020; Tellis and Wernefelt, 1987), and they are in line with classical market microstructure models; for example, Kyle (1985) and Glosten and Milgrom (1985) emphasize the crucial importance of information to price discovery in financial markets. This also confirms Li, van Dalen and van Rees' (2018) findings that information from microblogging platforms, such as Twitter, impact investors' decisions.

To illustrate the relevance of these findings, Figure 3 presents tweets A and B as examples of FGC characterized by negative and positive valence, respectively, but not containing any subject matter related to competitor or consumer orientation. Consistent with the research findings, the permanent price impact estimates for the tweet trades corresponding to both tweets are more than three standard deviations lower than the average permanent price impact estimate and are thus below the 10th percentile of the estimates; the estimates for the negative and positive tweets' tweet trades are .0017% and .0035% respectively. In contrast to A and B, tweets C, D and E reflect varying combinations of both valence and subject matter. The findings suggest that these tweets should generate significant permanent price impact, and indeed the permanent price impact estimates for the tweet trades corresponding to tweets C, D and E are above the 90th percentile in our sample of tweet trades' permanent price impact estimates. The estimates are 3.74%, 2.84% and 1.32% for tweets C, D and E respectively.

<Insert Figure 3>

The effects of the tweet attributes we study on temporary price impact, $\Delta\sigma^{2_i}_{s,t}$, also deserve attention. The results suggest that the relationship between valence and temporary price impact is generally magnified when combined with subject matter. For example, the γ_7 and γ_8

estimates, which capture the relationship between $\Delta\sigma^{2_i}_{s,t}$ on the one hand and $-\mathrm{ve}_{s,t}$ and $+\mathrm{ve}_{s,t}$ on the other, are positive and statistically significant (.032 p<.05 and <.033 p<.0,5 respectively), while the estimates for γ_3 and γ_6, which capture the relationship between $\Delta\sigma^{2_i}_{s,t}$ on the one hand and $\mathrm{consumer}_{s,t} * -\mathrm{ve}_{s,t}$ and $\mathrm{competitor}_{s,t} * +\mathrm{ve}_{s,t}$ on the other, a are also positive and statistically significant (.072 p<.05 and <.077 p<.05, respectively). The latter set of estimates is at least two times larger than the former. The overall implication of these positive and statistically significant coefficient estimates related to temporary price impacts is that, although tweets reflecting both valence and subject matter are likely to generate permanent price impact, these attributes may also be associated with increased temporary price impact. Thus, on average, tweets would inject noise (uncertainty) into the prices of stocks traded in financial markets.

In conclusion, the estimates presented in Table 6 highlight the relevance of tweet attributes for the price discovery process in financial markets and reinforce the importance of studying the multifaceted nature of FGC (Kumar et al., 2016). We find that tweets, as with many events observed in relation to trading in financial markets, generate both permanent and temporary price impacts. However, while tweets containing singular attributes; either positive or negative valence, or consumer or competitor orientation, readily inject noise into the price discovery process and thus generate temporary price impact, those that include more than one attribute generate permanent price impact, and thus generally enhance the efficiency of the price discovery process.

Discussion

In this research, we examine the real-time impact of FGC on the variance of firms' stock price. In the current fast-paced online communication landscape, marketers must understand the financial impact of firms' 'always on' marketing (Rust et al, 2021). The assessment of FGC's financial impacts, however, is in an early stage (Borah et al., 2020;

Colicev et al., 2018; Rust el al, 2021). This research contributes to this emerging stream of marketing research and addresses multiple calls for new methods able to develop real-time insights from online data (Berger et al., 2020; Lamberton and Stephen, 2016; Moorman et al., 2019; Wedel and Kannan, 2016). Employing the market microstructure approach to study S&P 500 IT firms' Twitter activity, this study contributes to the marketing literature and practice.

Research contribution

This study offers several implications to marketing research. Firstly, aligning with the work by Colicev et al. (2018), Borah et al. (2020) and Rust et al (2021), it advances our understanding of FGC's financial impact by providing an assessment of FGC's impact on the variance of stock price in real-time (i.e. seconds). By employing a market microstructure approach, we show how to algorithmically link piece-by-piece FGC to time-specific trading activity at a fine-grained level of analysis. In the process, we demonstrate the limitations of low-frequency methodologies, such as daily event studies that are subject to aggregation bias and which may yield bias estimates of the impact of FGC on firms' financial outcomes, while offering an alternative and more robust method of analysis for studying intraday marketing activity. In our examination of the impact of FGC on variance, we fully utilize high frequency transactions data characterized by unequal time intervals, and demonstrate how to retain data that otherwise would have been eliminated in studies employing end-of-day stock price. By doing so, we provide marketing researchers with a new approach that allows them to harness the potential of online data.

Secondly, we distinguish between FGC's temporary and permanent price impact. Specifically, we show that FGC impacts investor expectations related to a firm's future performance, thus generates permanent price impact, and also injects uncertainty about a firm's value into its stock price, hence induces temporary price impact. Our research therefore, adds a new perspective to the marketing literature stream on the financial impact of FGC. This assessment of FGC's temporary and permanent price impacts adds richness to the examination of marketing's financial impact and enables the quantifying of the long- and short-term financial impacts of marketing activity.

Finally, this research has implications for the design of intraday marketing strategies. By examining FGC valence and subject matter (consumer and competitor orientation), we advance a growing body of research documenting the complex nature through which marketing signals impact financial markets and firm financial outcomes. We show that, in isolation, FGC valence and subject matter are more prone to injecting uncertainty about a firm's stock price into the market, and thus they generate temporary price impacts, than permanently change investors and traders' belief about firm value. Holistically speaking, FGC valence and subject matter both hold statistically significant and economically meaningful relevance for price discovery in financial markets. In other words, they can influence investors' expectations related to firms' future performance and thus result in permanent price impacts. Recent research by Bhagwat et al. (2020) provides evidence of interactions between marketing signals, this research shows that the interaction between FGC valence and subject matter can also impact firm stock price.

Managerial implications

Thus far, firms have struggled to demonstrate financial accountability for the impact of FGC on firm value (Colicev et al., 2018; Kumar et al., 2016;), or evidence its immediate contribution to their financial outcomes (Magill, Moorman, Avdiushko 2019; Moorman and Kirby, 2019). We provide marketing managers with evidence of FGC's impact on the variance in firms' stock price. Specifically, we show that tweets can generate both permanent and temporary price impacts. By selecting tweet attributes, such as valence and subject matter, marketing managers can design Twitter content to generate varying degrees of permanent or temporary impact. From a market quality perspective, firm managers should prefer tweets that generate a permanent price impact, and our research provides some useful indications on how to achieve this outcome. We show that tweets expressing degrees of positive or negative valence about either consumers or competitors generate a permanent price impact. We, therefore, encourage marketing managers to design information-rich tweets concerning consumers or competitors as well as reflecting valence. The results suggest that firms should reflect valence and subject matter in their tweets if they would like their stock to be more informative with respect to their value. Our analysis suggests that tweets about competitors with a negative valence are likely to have the highest permanent price impacts. Thus, by using the permanent price impact as a metric to evaluate the longer-term impact of tweets, social media managers can design campaigns that have a sustainable impact on firm financial outcomes. The design recommendations from this study complement Kanuri, Chen and Sridhar's (2018) work on social media content scheduling and Rust et al.' (2021) work on real-time social media marketing, informing firms on which tweets to disseminate during a day for long-term effectiveness. We recognize that not all intraday tweets will, nor should they, have permanent impacts on firms' stockprice. Some tweets are aimed at the creation of social media 'buzz', which is similar to the temporary price impacts we examine in this study. Firms can achieve social media 'buzz' by disseminating tweets as the findings reveal that tweets, in aggregate, mostly generate temporary price impacts. We urge caution, however, because temporary price impact is linked with larger transaction costs (Chan and Lakonishok, 1993) and increases in firm cost of capital (Diamond and Verrecchia, 1991). This suggests that the benefits of designing tweets to generate 'buzz' and incorporate information into stock price must be carefully managed. To support marketing managers in their intraday social media strategy design, Table 7 is designed as a set of insights based on our findings.

<Insert Table 7 about here>

Limitations

We conclude by encouraging future research to address the limitations of our empirical study. One potential limitation of our analysis is its focus on firms belonging to the IT sector. We recognize that these findings may not apply to other sectors. Future research could extend our analysis to other sectors to confirm whether similar price impacts hold. Secondly, the impact of the tweets could depend on whether Twitter was the first source through which a firm released an important piece of news. Tweets could have been published in response to a competitor's tweet. In some cases, a firm's tweet could lead to a number of successive tweets, in which case the subsequent tweets might not be as impactful as the first. We do not discount the possibility that there could be some carry-over or dampening effect in such situations. We note that, if this is the case, it would be highly unlikely for the magnitude of the effects we observe to occur, especially given the granular level of analysis that our market microstructure approach entails. Thirdly, future work could explore high frequency data generated by firms' use of FGC other than Twitter, such as Facebook posts, where it has been reported firms post up to 80 times a day. It would be interesting to see if the effect of FGC across social media platforms is consistent or whether it varies. In addition to social media, it would be useful to examine firms' use of other online communication tools, such as webpages and blogging platforms, and to explore various types of FGC, including video content, and its characteristics including emotions (Tellis et al., 2019). As Hewett et al. (2016) show, there is an array of online marketing communication practices; future research could therefore study the 'echoverse' at a fine-grained level of analysis. Finally, we note that market microstructure can be applied to study user-generated content (UGC) in future research. We welcome future research that addresses the following questions: what is the real-time impact of UGC on firm value? What are the UGC attributes capable of generating permanent and temporary price impacts? Are they the same as FGC attributes, or do they differ? Our research highlights the importance of interaction effects when examining the impact of FGC attributes on firm value; therefore, investigating the optimal mix of UGC attributes capable of generating temporary and permanent price impacts should be an interesting endeavor.

References

Amemiya, Takeshi, and Roland Y.Wu, (1972),"The effect of aggregation on prediction in the autoregressive model,"*Journal of the American Statistical Association*, 67(339), 628-632.

Amihud, Yakov (2002),"Illiquidity and stock returns: Cross-section and time-series effects," *Journal of Financial Markets*, 5(1), 31-56.

Atuahene-Gima, Kwaku (2005),"Resolving the capability-rigidity paradox in new product innovation," *Journal of Marketing*, 69(4), 61-83.

Barclay, Michael J., and Jerold B.Warner, 1993), "Stealth trading and volatility: Which trades move prices? "*Journal of Financial Economics*, 34(3), 281-305.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020),"Uniting the tribes: Using text for marketing insights," *Journal of Marketing*, 84(1), 1-25.

Bhagwat, Yashoda, Nooshin L.Warren, Joshua T.Beck, and George F. Watson IV, G. F. (2020),"Corporate sociopolitical activism and firm value,"*Journal of Marketing*, 84 (5), 1-21.

Blankespoor, Elizabeth, Gregory S. Miller, and Hal D. White, (2014),"The role of dissemination in market liquidity; evidence from firms' use of Twitter,"*The Accounting Review*, 89(1), 79-112.

Borah, Abhishek, Sourindra Banerjee, Yu-Ting Lin, Apurv Jain, and Andreas B. Eisingerich (2020),"Improvised marketing interventions in social media,"*Journal of Marketing*, 84(2), 69-91.

Boyd, D. Eric, and P. K. Kannan, (2018),"(When)Does third-party recognition for design excellence affect financial performance in business-to-business markets?," *Journal of Marketing*, 82(3), 108-123.

Brandwatch (2020),"60 Incredible and Interesting Twitter Stats and Statistics," (accessed April 16, 2021) www.brandwatch.com/blog/twitter-stats-and-statistics/

Breitung, Jorg, and Norman S. Swanson, (2002),"Temporal aggregation and spurious instantaneous causality in multiple time series models," *Journal of Time Series*, 23(6), 651-665.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, (2014),"High-frequency trading and price discovery," *The Review of Financial Studies*, 27, 2267-2306.

Brookings (2019),"Trends in the Information Technology sector" (accessed April, 19 2021) www.brookings.edu/research/trends-in-the-information-technology-sector/

Budish, Eric, Peter Cramton, and John Shim, (2015),"The high-frequency trading arms race: Frequent batch auctions as a market design response," *Quarterly Journal of Economics*, 130(4), 1547-1621.

Chan, Louis. K. C., and Josef Lakonishok, (1993), "Institutional trades and intraday stock price behavior,"*Journal of Financial Economic*s, 33(2), 173-199.

Chan, Hing Kai, Xiaojun Wang, Ewelina Lacka, and Min Zhang, (2016),"A mixed-method approach to extracting the value of social media data,"*Production and Operations Management*, 25(3), 568-583.

Colicev, Anatoli, Ashwin Malshe, Koen Pauwels, and Peter O'Connor, (2018),"Improving consumer mindset metrics and shareholder value through social media: The different roles of owned and earned media,"*Journal of Marketing*, 82, 37-56.

Colicev, Anatoli., Ashish Kumar, and Peter O'Connor, (2019),"Modeling the relationship between firm and user generated content and the stages of marketing funnel," *International Journal of Research in Marketing*, 36(1), 100-116.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, (2001),"Market liquidity and trading activity," *The Journal of Finance*, 56(2), 501-530.

Chordia, Torun., Richard Roll, and Avanidhar Subrahmanyam, (2008), " Liquidity and market efficiency,"*Journal of Financial Economics*, 87(2), 249-268.

DeKimpe, Marnik G., and Dominique M. Hanssens, (1995),"The persistance of marketing effects on sales," *Marketing Science*, 14(1), 1-21.

Diamond, Douglas W., and Robert E. Verrecchia, (1991), "Disclosure, liquidity, and the cost of capital," *Journal of Finance*, 46(4), 1325-1359.

Du, Rex Yuxing, Oded Netzer, David A. Schweidel, and Debanjan Mitra, (2021), "Capturing marketing information to fuel growth," *Journal of Marketing*, 85(1) 163-183.

Durbin, J., Koopman, S. (2012). *Time series analysis by state space models*. Oxford University Press, Oxford, UK.

Elliott, W. Brooke, Stephanie M. Grant, and Frank D. Hodge, (2018),"Negative news and investor trust: The role of $Firm and #CEO Twitter use,"*Journal of Accounting Research*, 56(5), 1483-1519.

Fama, Eugene F. (1970),"Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, 25(2), 383-417.

Fama, Eugene. F. (1998),"Market efficiency, long-term returns, and behavioural finance," *Journal of Financial Economics*, 49(3), 283-306.

Frino, Alex, Elvis Jarnecic, and Andrew Lepone, (2007), "The determinants of the price impact of block trades: Further evidence," *Abacus*, 43(1), 94-106.

Gebhardt, Gary F., Gregory S. Carpenter, and John F. Sherry (2006),“Creating a market orientation: A longitudinal, multifirm, grounded analysis of cultural transformation,” *Journal of Marketing*, 70, 37-55.

Glosten, Lawrence. R., and Paul R. Milgrom, (1985),“Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,”*Journal of Financial Economics*, 14(1), 71-100.

Gong, Shiyang, Juanjuan Zhang, Ping Zhao, and Xuping Jiang, (2017),“Tweeting as a marketing tool: A field experiment in the TV industry,”*Journal of Marketing Research*, 54(6), 833-850.

Gordon, Brett R., Kinshuk Jerat, Zsolt Katona, Sridhar Narayanan, Juwoong Shin, and Kenneth Wilbur, (2021),“Inefficiencies in digital advertising markets,”*Journal of Marketing*, 85(1),7-25.

Groß-Klußmann, Alex, Stephan König, and Markus Ebner, (2019),“Buzzwords build momentum: Global financial Twitter sentiment and the aggregate stock market,”*Expert Systems with Applications*, 136, 171-186.

Grossman, Sanford J. and Joseph E. Stiglitz, (1980),“On the impossibility of informationally efficient markets,” *The American Economic Review*, 70(3), 393-408.

Grossman, Sanford J. and Merton H. Miller, (1988),“Liquidity and market structure,”*The Journal of Finance*, 43(3), 617-633.

Hargreaves Lansdown (2020)“How tech dominates the US stock market” (accessed on April, 16,2021) https://www.hl.co.uk/news/articles/how-tech-dominates-the-us-stock-market

Hausman, Jerry, James H.Stock, and Motohiro Yogo, (2005),“Asymptotic properties of the Hahn-Hausman test for weak-instruments,”*Economics Letters*, 89(3), 333-342.

Hasbrouck, Joel, (1991),“Measuring the information content of stock trades,” *The Journal of Finance*, 46(1), 179-207.

Hedge, Sjantaram P.,and John B. McDermott, (2003),"The liquidity effects of revisions to the S&P 500 index: An empirical analysis, "*Journal of Financial Markets*, 6(3), 413-459.

Hendershott, Terrence, and Albert J. Menkveld, (2014),"Price pressures,"*Journal of Financial Economics*, 114(3), 405-423.

Hewett, Kelly., William Rand, Roland T. Rust, and Harald J. van Heerde (2016),"Brand buzz in echoverse,"*Journal of Marketing*, 80, 1-24.

Hoffman, Donna L., and Marek Fodor, (2010),"Can you measure the ROI of your social media marketing?"*MIT Sloan Management Review*, 52(1), 41-49.

Holthausen, Robert W., Richard W. Leftwich, and David Mayers,(1990),"Large-block transactions, the speed of response, and temporary and permanent stock-price effects"*Journal of Financial Economics*, 26(1), 71-95.

Homburg, Christian, Marcus Theel, and Sebastian Hohenberg, (2020),"Marketing excellence: Nature, measurement and investor valuations"*Journal of Marketing*, 84(4), 1-22.

Hutto, C. J., and Eric Gilbert, (2014),"VADER: A parsimonious rule-based model for sentiment analysis of social media text," Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI, June 2014.

Ibikunle, Gbenga (2015),"Opening and closing price efficiency: Do financial markets need the call auction?"*Journal of International Financial Markets, Institution and Money*, 24(0), 208-227.

Kanuri, Vamsi K.,Yixing Chen, and Shrihari Sridhar, (2018),"Sheduling content on social media: Theory, evidence, and application,"*Journal of Marketing*, 82(6), 89-108.

Kirmani, Amna, and Akshay R. Rao, 2020),"No pain, no gain: A critical review of the literature on signaling unobservable product quality,"*Journal of Marketing*, 64, 66-79.

Kirilenko, Andrei, Albert S.Kyle, Mehradad Samadi, and Tugkan Tuzun, (2017),"The flash crash: High-frequency trading in an electronic market,"*Journal of Finance*, 72 (3), 967-998.

Koopman, S. J., and J. Durbin, (2003),"Filtering and smoothing of state vector for diffuse state-space models,"*Journal of Time Series Analysis*, 24(1), 85-98.

Kumar V., Vikram Bhaskaran, Rohan Mirchandani and Malap Shah (2013),"Creating a measurable social media marketing strategy: Increasing the value and ROI of intangibles and tangibles for hokey pokey,"*Marketing Science*, 32(2), 191-363.

Kumar, V., Eli Jones, RajkumarVenkatesan, and Robert P. Leone, (2011),"Is market orientation a source of sustainable competitive advantage or simply the cost of competing?,"*Journal of Marketing*, 75 (1), 16-30.

Kumar, Ashish, Ram Bexawada, Rishika Rishika, Ramakumar Janakiraman, and P.K. Kannan, (2016),"From social to sale: The effects of firm-generated content in social media on consumer behaviour,"*Journal of Marketing*, 80, 7-25.

Kyle, Albert. S. (1985), "Continuous auctions and insider trading,"*Econometrica*, 53(6), 1315-1335.

Lam, Son K., Florian Kraus, and Michael Ahearne, (2010),"The diffusion of market orientation throughout the organization: A social learning theory perspective,"*Journal of Marketing,* 74(5), 61-79.

Lamberton, Cait, and Andrew T. Stephen, (2016),"A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry," *Journal of Marketing*, 80, 146-172.

Lambrecht, Anja, Catherine Tucker, and Caroline Wiertz, (2018),"Advertising to early trend propagators; evidence from Twitter,"*Marketing Science*, 37(2), 177-199.

Li, Ting, Jan van Dalen, and Pieter Jan van Rees, (2018),"More than just noise? Examining the information content of stock microblogs on financial markets," *Journal of Information Technology*, 33(1), 50-69.

Madhavan, Ananth, Matthew Richardson, and Mark Roomans, (1997),"Why do security prices change? A transaction-level analysis of NYSE stocks,"*The Review of Financial Studies*, 10(4), 1035-1064.

Magill, Paul, Christina Moorman, and Nikita Avdiushko, (2019),"8 ways marketers can show their work's financial results,"*Harvard Business Review*.

Marcellino, Marcellino (1999),"Some consequences of temporal aggregation in empirical analysis,"*Journal of Business and Economic Statistics*, 17(1), 129-136.

Meire, Matthijs, Kelly Hewett, Michel Ballings, V.Kumar, and Dirk Van den Poel, (2019),"The role of marketer-generated content in customer engagement marketing,"*Journal of Marketing*, 83(6), 21-42.

Menkveld, Albert J. (2013),"High frequency trading and the new market makers,"*Journal of Financial Markets*, 16, 712-741.

Menkveld, Albert J., Suem Jan Koopman, and Andre Lucas, (2007) ,"Modeling around-the-clock price discovery for cross-listed stocks using state space methods," *Journal of Business and Economic Statistics*, 25(2), 213-225.

Moorman, Christine, and Lauren Kirby,(2019),"How marketers can overcome short-termism,"*Harvard Business Review*.

Moorman, Christine, Harald J. van Heerde, C. Page Moreau, and Robert W. Palmatier, (2019),"Challanging the boundaries of marketing,"*Journal of Marketing*, 83(5), 1-4.

O'Hara, Maureen (2003),"Presidential address: Liquidity and price discovery,"*The Journal of Finance*, 58(4), 1335-1354.

Pavia-Miralles, Jose Manuel, (2010),"A survey of methods to interpolate, distribute and extrapolate time series,"*Journal of Service Science and Management*, 3, 449-463.

Qian, Hang (2013),"Vector autoregression with mixed frequency data,"Munich Personal RePEc Archive. Working Paper.

Rao, Ram C. (1986),“Estimating continuous time advertising-sales models,”*Marketing Science*, 5(2), 125-142.

Rust RolandT., William Rand, Ming-Hui Huang, Andrew T. Stephen, Gillian Brook, and Timur Chabuk (2021),“Real-Time Reputation Tracking Using Social Media,”J*ournal of Marketing*, 85(4), 21-43

Rzayev, Khaladdin, and Gbenga Ibikunle, (2019),“A state space modeling of the information content of trading volume,”*Journal of Financial Markets*, 46(100507),1-19.

Saboo, Alok R., and Rajdeep Grewal, (2013),“Stock market reactions to customer and competitor orientations: The case of initial public offerings,” *Marketing Science*, 32(1), 70-88.

Selukar, Rajesh (2011),“State space modeling using SAS,”*Journal of Statistical Software*, 41(12), 1-13.

Sharpe, William F. (1964),“Capital asset prices: A theory of market equilibrioum under conditions of risk,”*Journal of Finance*, 19(3), 442-256.

Silvestrini, Andrea, and David Veredas, (2008),“Temporal aggregation of univariate and multivariate time series models: A survey,”*Journal of Economic Surveys*, 22(3), 458-497.

Statista (2020),“Social media spend,” (accessed October 25, 2020) www.statista.com/statistics/736971/social-media-ad-spend-usa

Stock, James H., Jonathan H. Wright, and Motohiro Yogo, (2002),“A survey of weak instruments and weak identification in generalized method of moments,”*Journal of Business and Economic Statistics*, 20(4), 518-529.

Sridhar, Shrihari, Murali K. Mantrala, Prasad A.Naik and Esther Thorson (2011),“Dynamic marketing budgeting for platform firms: Theory, evidence, and application,”*Journal of Marketing Research*, 48, 929-943.

Sul, Hong Kee, Alan R. Dennis, and Lingyao Yuan, (2017),“Trading on Twitter: Using social media sentiment to predict stock returns,”*Decision Sciences*, 48(3).

Sun, Yuxin, and Gbenga Ibikunle (2017),“Informed trading and the price impact of block trades: A high frequency trading analysis,”*International Review of Financial Analysis*, 54(0) 114-129.

Tellis, Gerard J., and Birger Wernerfelt, (1987),“Competetive price and quality under asymmetric information,”*Marketing Science*, 6, 240-253.

Tellis, Gerard J., Deborah J. MacInnis, Seshadri Tirunillai, and Yanwei Zhang, (2019),“What drives virality (sharing) of online digital content? A critical role of informaition, emotion and brand prominence,”*Journal of Marketing*, 83(4), 1-20.

Tiao, G. C., and W. S.Wei, (1976),“Effect of temporal aggregation on the dynamic relationship of two time series variables,”*Biometrika*, 63(3), 513-523.

Tirunillai, Seshadri, and Gerard J. Tellis, (2012),“Does chatter really matter? Dynamics of user-generated content and stock performance,” *Marketing Science*, 31(2), 198-215.

van Heerde, Harald, Els Gijsbrechts, and Koen Pauwels, (2015),“Fanning the flames? How media coverage of a price war affects retailers, consumers and investors,”*Journal of Marketing Research*, LII, 674-693.

Vermeer, Susan A.M., Theo Araujo, Stefan F. Bernritter, and Guda van Noort, (2019),“Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media,”*International Journal of Research in Marketing*, 36(3), 492-508.

Voss, Glenn B., and Zannie G.Voss, (2000),“Strategic orientation and firm performance in an artistic Environment,” *Journal of Marketing*, 64(1), 67-83.

Webel, Michel, and P.K. Kannan, (2016),“Marketing analytics for data-rich environments,”*Journal of Marketing*, 80, 97-121.

Whitler, Kimberly A., Ryan Krause, and Donald R. Lehmann, (2018),"When and how board members with marketing experience facilitate firm growth,"*Journal of Marketing*, 82(5), 86-105.

Table 1. Twitter data sample

S&P 500 IT firm	No. of tweets	No. of tweet days	Min. no. of tweets per day	Max. no. of tweets per day	Average no. of tweets per day	Single tweet days (%)	No. of tweets excluded*	No. of days excluded*
Red Hat	3,102	204	2	155	15.13	0	20	25
DXC Technology	2,239	155	3	30	14.35	0	5	8
CA	3,052	232	1	69	13.09	0.26	0	27
Cognizant	3,094	294	1	32	10.48	0.65	34	36
Oracle	3,187	331	1	105	9.59	0.85	20	24
F5 Networks	3,041	355	2	32	8.54	0	15	30
Gartner	3,229	393	1	85	8.19	0.77	16	24
FLIR Systems	3,228	413	1	30	7.79	1.33	29	27
ANSYS	3,141	425	1	42	7.37	1.97	22	20
PAYCHEX	3,072	428	1	68	7.16	1.46	10	30

*Excluded because of excessive return volatility.

Table 2. Trading data descriptive statistics

Before cleaning		
Messages	Transactions	Orders
8,182,063,205	522,403,178	7,659,660,027
After cleaning**		
8,177,183,865	520,356,393	7,656,827,472
Percentage of trading data removed from the sample after data cleaning		
.06%	.39%	.04%

**Data cleaning is completed using the criteria outlined by Chordia, Roll and Subrahmanyam(2001) and Ibikunle (2015).

Table 3. Descriptive statistics of tweet activity

Tweets per day per firm		
Mean number of tweets per day per firm	Minimum number of tweets per day per firm	Maximum number of tweets per day per firm
.844	.00	6.79
Tweets per firm		
Mean number of tweets per firm	Minimum number of tweets per firm	Maximum number of tweets per firm
2,377.22	30.00	3,040
Tweets per day		
Mean number of tweets per day	Minimum number of tweets per day	Maximum number of tweets per day
54.03	0.00	353.00

Table 4. Permanent and temporary components of price: descriptive statistics

Price component	Mean	Median	St. Dev.	Minimum	Maximum
Temporary price component $\left(\sigma^{2_i}_{s,t}\right)$	.011	.008	.009	.000	.297
Permanent price component $\left(\sigma^{2_u}_{s,t}\right)$	.055	.010	.048	.000	.644

Table 5. Ratios of the price impact of tweet-trades to the price impact of other trades

Price Impact_t	60-second threshold	45-second threshold	30-second threshold
Temporary price impact $\left(\Delta\sigma^{2_i}_{s,t}\right)$	279.67*** (7.51)	230.12*** (9.58)	222.55*** (10.23)
Permanent price impact $\left(\Delta\sigma^{2_u}_{s,t}\right)$	146.83*** (4.33)	178.87*** (3.21)	189.04*** (5.17)

*** Statistical significance at the 0.001 level

Table 6. Permanent and temporary price impact and tweet valence and orientation

Variables	Permanent price impact $\left(\Delta\sigma^{2_u}_{s,t}\right)$	Temporary price impact $\left(\Delta\sigma^{2_i}_{s,t}\right)$	Key findings
$consumer_{s,t}$	-.007 (1.06)	.009** (2.39)	Consumer-related tweets are, on average, associated with a larger temporary price impact relative to other tweets.
$competitor_{s,t}$	-.034** (2.03)	.026** (2.46)	Competitor-related tweets are, on average, associated with a larger temporary price impact and lower permanent price impact relative to other tweets.
$-ve_{s,t}$	-.063** (2.37)	.032** (2.46)	Negative and positive valence only tweets are associated with increasing temporary price impact and decreasing permanent price impact.
$+ve_{s,t}$	-.108*** (3.11)	.033** (2.20)	
$consumer_{s,t} * -ve_{s,t}$	.047** (1.97)	.072** (2.50)	Tweets reflecting both valence and subject matter are associated with increases in both permanent and temporary price impact. The increase in permanent price impact contrasts the decrease in permanent price impact that tweets with only one of valence and subject matter are associated with. Except for tweets reflecting negative valence and consumer orientation $\left(consumer_{s,t} * -ve_{s,t}\right)$, permanent price impact is more pronounced than temporary price impact.
$competitor_{s,t} * -ve_{s,t}$	.606*** (3.69)	.011** (2.09)	
$consumer_{s,t} * +ve_{s,t}$	.088** (2.10)	.019** (2.21)	
$competitor_{s,t} * +ve_{s,t}$	.220*** (4.88)	.077** (2.43)	
$lnvolume_{s,t}$	-.039*** (-4.51)	-.034*** (-6.68)	Increases in firm stock trading activity are linked with reductions in both permanent and temporary price impacts.

$\ln tradesize_{s,t}$	.089*** (6.64)	.061*** (7.25)	Larger firm stock trade sizes induce larger permanent and temporary price impacts.
$volatility_{s,t}$	-.123*** (-3.62)	.015** (2.13)	Firm stock volatility is linked with increasing temporary price impact and is decreasing permanent price impact
$Effectivespread_{s,t}$	.241** (2.66)	.014** (2.06)	Deterioration in firm stock liquidity is associated with increasing permanent and temporary price impact.
$\ln HFT_{s,t}$	-.000 (-.26)	-.021*** (-3.83)	Algorithmic and high frequency trading is linked with decreases in temporary price impact; its effect on permanent price impact is benign.
$OIB_{s,t}$	-.371*** (-6.89)	.046** (2.39)	Order imbalance is linked with reductions in permanent price impact and increases in temporary price impact.
$\ln\#followers_{s,t}$	-.082** (-2.43)	.037** (2.43)	The number of followers of firm's twitter accounts amplifies the propensity for tweets to generate larger temporary price impact and reduce permanent price impact.
$\overline{R^2}$	.35	.49	
Observations	139,997	139,997	
Kleibergen-Paap LM	31.32***	110.24***	*Tests the null that the employed instruments have insufficient explanatory power to predict the endogenous variables in the model for identification of the parameter*
Cragg-Donal	79.08***	88.66***	*Tests the same null hypothesis as the Kleibergen-Paap LM test*
Sargan's χ^2 *p*-value	.37	.46	*Tests the null that the over-identifying restrictions are valid*

Table 7. Suggested insights for marketing managers

		Permanent Price Impact	Temporary Price Impact	Research Findings	Recommendation	Expected outcome	Interaction effects	Permanent Price Impact	Temporary Price Impact	Research Findings
Valence	Positive Valence			Positive and negative valence-only FGC contributes to more noise in a firm's stock prices	Add subject matter (e.g. competitor orientation such as 'competition', 'peer')	Increased permanent price impact	Positive valence & competitor orientation			Interaction of valence and subject matter increases/generates permanent price impact and amplifies temporary price impact. Permanent price impact is more pronounced than temporary price impact. The financial implication of these outcomes is a reduction in the transaction and firm capital costs.
	Negative Valence				Add subject matter (e.g. consumer orientation such as 'customer', 'consumer', 'buyer')	Increased permanent price impact	Negative valence & consumer orientation			
Subject Matter	Consumer Orientation			Subject matter-only FGC contributes to the noise component in a firm's stock prices	Add valence (e.g. positive valence such as 'help', 'solution', 'best')	Increased permanent price impact elicited	Positive valence & consumer orientation			
	Competitor Orientation				Add valence (e.g. negative valence such as 'attack', 'stop', ' threat')	Increased permanent price impact	Negative valence & competitor orientation			

No price impact; Negative impact on stock price component; Positive impact on stock price component; Increased positive impact on stock price component

Figure 1. A high frequency approach to FGC dissemination on the example of IT firms' activity on Twitter

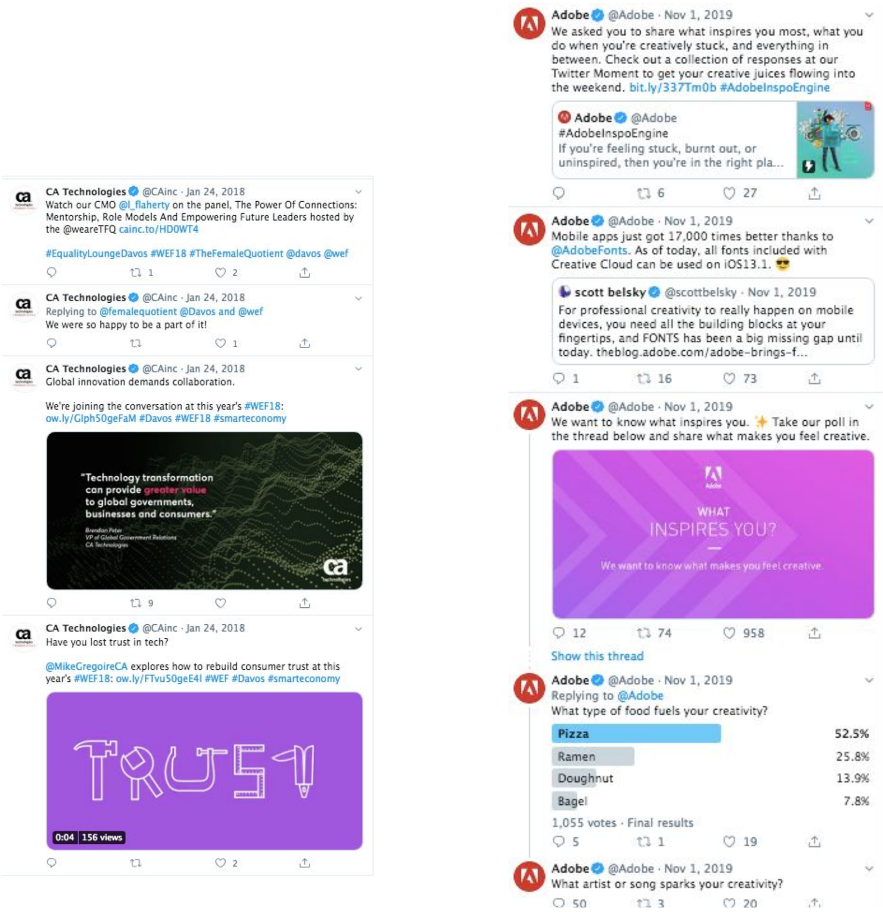
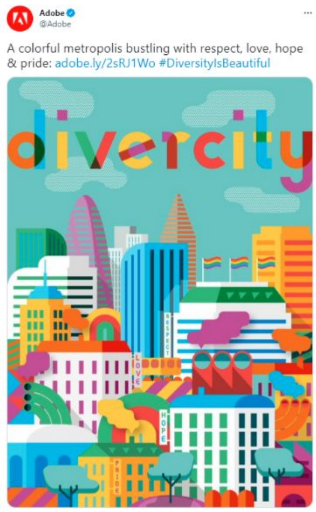


Figure2. Examples of tweets characterized by valence and subject matter (consumer and competitor orientation)

A.Negative valence tweet

B. Positive valence tweet



C.Consumer orientation

D. Competitor orientation

E.Negative valence and consumer orientation

F.Negative valence and competitor orientation

G Positive valence and consumer orientation

H.Positive valence and competitor orientation

Figure 3. Examples of tweets generating temporary and permanent price impacts

A

B

C

D

E

Measuring the real-time stock market impact of firm-generated content

Ewelina Lacka
Assistant Professor in Digital Marketing & Analytics
University of Edinburgh
ewelina.lacka@ed.ac.uk

D. Eric Boyd
Associate Professor of Marketing, University of Central Florida
Professorial Fellow, University of Edinburgh
eric.boyd@ucf.edu

Gbenga Ibikunle
Professor of Finance, University of Edinburgh
Fellow, RoZetta Institute Sydney
gbenga.ibikunle@ed.ac.uk

P.K. Kannan
Professor of Marketing Science
Robert H. Smith School of Business
University of Maryland
pkannan@rhsmith.umd.edu

Web Appendix

These materials have been supplied by the authors to aid in the understanding of their paper. The AMA is sharing these materials at the request of the authors

Web Appendix A. Sample of high frequency Twitter data generated during June 2018 by S&P 500 IT firms

S&P 500 firm	No. of tweets	S&P 500 firm	No. of tweets	S&P 500 firm	No. of tweets
Accenture	55	F5 Networks	253	NVIDIA	56
Activision Blizzard	20	Facebook	75	Oracle	156
Adobe	127	Fidelity National Information Services	76	Paychex	155
Akamai	58	Fiserv	51	PayPal	46
Alliance Data Systems	8	FLIR Systems	188	Qorvo	32
Automatic Data Processing	53	Gartner	226	QUALCOMM	68
Alphabet A (ex. Google)	N/A	Global Payments	13	Red Hat	394
Alphabet C (ex. Google)	N/A	Harris	71	Salesforce	131
AMD	55	HP	2	Seagate	35
AMPHENOL	N/A	Hewlett Packard Enterprise	61	Skyworks Solutions	9
Analog Devices	76	IBM	N/A	Symantec	163
ANSYS	151	Intel	35	Synopsys	73
Apple	N/A	Intuit	26	TE Connectivity	35
Applied Materials	15	Juniper Networks	49	Texas Instruments	73
Autodesk	62	KLA-Tencor	17	The Western Union Company	74
Broadcom	18	Lam Research	80	Time Warner	N/A
CA	490	Mastercard	83	Total System Services	45
Cadence Design Systems	27	Microchip Technology	128	Twitter	6
Cisco	117	Micron Technology	N/A	VERISIGN	16
Citrix Systems	118	Microsoft	N/A	Visa	6
Cognizant	315	Motorola Solutions	96	Western Digital	53
Corning	43	NetApp	247	Xerox	80
DXC Technology	537	Netflix	136	Xilinx	76
Electronic Arts	151	Newell Brands	31		

N/A Data not available for June 2018

Web Appendix B. Low-frequency methodologies: event studies and standard vector autoregressive (VAR) models

	Theory	Firm Value Metric(s)	Estimation Method	Time Treatment	Counterfactual
Low-frequency event study	Efficient market hypothesis	Level stock return change within a specified event window (usually days)	Stock return measure based on subtracting post-event stock from pre-event expected stock return based on estimates of the expected returns as a function of risk factors that reflect the general stock market, size, the relative importance of intangibles (book-to market ratio), and momentum	Discrete with time interval length treated as equal	Quasi-experimental using pre-event price over a pre-determined period as counterfactual
VAR		Smoothed level change based on time series of first differences of the logarithm of stock prices	Stock return measure based on subtracting post-event stock from pre-event expected stock return based on estimates of the expected returns as a function of risk factors that reflect the general stock market, size, the relative importance of intangibles (book-to market ratio), and momentum	Discrete with time interval length treated as equal	Quasi-experimental using pre- and post-event return as counterfactual in testing whether the return goes back to the mean, and how long it takes for the model to go back to the mean 'dust settling period'

Web Appendix C. Daily event study methodology

The daily event study methodology assumes that any change in stock price occurs due to the arrival of new information (e.g. FGC/ tweet) (Sharpe, 1964; Fama, 1998). We use Web Appendix F describing the number of tweets sent by S&P 500 IT firms in our sample during the period under investigation. This reveals that the average number of tweets per day for all firms is greater than one. As shown in Web Appendix F, with the exception of KLA-Tencor (56.67%), all firms in the sample disseminate multiple tweets in a day. If the problem of linking FGC with firm value was to be examined by employing an event study methodology, marketing researchers would be bound to compare the end-of-day stock price on the day of a single tweet with the expected stock price estimated over a period of time in the past to calculate an abnormal stock return. The abnormal stock return would then be tested for significance in determining whether FGC (i.e. a tweet) impacts firm value.

An important requirement for ensuring internal validity of the event study analysis is the removal of information events, such as tweets, other than the focal one on the event day or a window of time surrounding the event day. Considering our sample, this would severely reduce the sample size. With the exception of one firm, firms from our sample would have approximately 50% of their tweets removed from the examination. In fact, only 11.47% of the actual tweeting activity would be analyzed. In comparison, the sample size reduction with the microstructure approach involves tweets excluded due to excessive return volatility. As shown in Web Appendix F on average, 18.42 tweets per firm are excluded from the sample, which represents.77% of the total sample.

The microstructure approach enables marketing researchers to access high frequency data with little to no sample size reduction. This is because the microstructure approach investigates the impact of an event at a fine-grained level of analysis. This then allows us to differentiate tweets that generate temporary and permanent price impacts. This is critical because intraday tweets, like most data generated at high frequency, are largely noisy. To illustrate this point, we use state-space modeling, as previously outlined. We obtain $\sigma_{s,t}^{2u}$ from the decomposition of the price of stock s at t = 60 seconds and employ it as an inverse measure of noise in the price discovery process, i.e. a direct measure of pricing efficiency, in the following regression to check whether tweets inject noise into the price discovery process or generally aid it:

$$\sigma_{s,t}^{2u}=\alpha_s+\beta_t+\gamma \text{tweet}_{s,t-1}+\sum_{k=1}^{5}\delta_k C_{k,S,t}+\epsilon_{s,t} \quad \text{(W1)}$$

where $\text{tweet}_{s,t-1}$is a dummy variable equalling 1 if a tweet occurs during time t-1, and α_s and β_t are stock and time fixed effects. All other variables are as previously defined. Standard errors are robust to heteroscedasticity and autocorrelation. There are two key differences between the earlier estimated Equation (9) and Equation (W1). Firstly, Equation (W1) is a predictive regression estimating the predictive power of tweets for $\sigma_{s,t}^{2u}$, and, secondly, γ in Equation (W1) captures the full effects of all tweets in a single coefficient – unlike the series of coefficients included in Equation (9), which captures the effects of tweets with various attributes on price impact estimates, $\Delta\sigma_{s,t}^{2u}$ and $\Delta\sigma_{s,t}^{2i}$.

Table W1 presents Equation (W1)'s estimated coefficients. The results show that there is a statistically significant negative relationship between the efficiency of the price discovery process and the average tweet. Specifically, the coefficient for $\text{tweet}_{s,t-1}$ (γ) is negative and highly statistically significant (-8.00x10^{-7}, p<.01). This suggests that an average tweet injects noise into the price discovery process. Most of the tweets are therefore 'noisy'. While exploiting high frequency intraday data, the microstructure approach allows for the relative impact of each tweet and its permanent or temporary price impact to be captured. This in turn allows us to capture the information signal and differentiate it from noise in the average tweet without compromising sample size.

Table W1. The effect of tweeting on the efficiency of the price discovery process

Variables	Coefficient estimates
$\mathrm{tweet}_{s,t-1}$	-.008*** (-5.78)
$\mathrm{lnvolume}_{s,t}$	-.006*** (-5.16)
$\mathrm{lntradesize}_{s,t}$	-.010*** (-5.71)
$\mathrm{volatility}_{s,t}$	.237*** (6.73)
$\mathrm{Effectivespread}_{s,t}$	.589*** (5.90)

*** Corresponds to statistical significance at the 0.01 level.

Another drawback to the event study methodology is the level of analysis. The event day focuses on the end-of-day price, but firms produce FGC throughout the day and there are respective prices throughout the day that align with each intraday FGC. The end-of-day price may not necessarily be reflective of the price reaction to the intraday event in question.

A third restriction of the event day for studying intraday marketing multi-activity is the abnormal return measure used to describe the marketing impact. The event study makes no distinction in describing the effect of marketing activity, other than whether the effect is statistically significant and its directional impact. The microstructure approach distinguishes between information capable of generating permanent price impacts and temporary price impacts. By doing so, the microstructure adds new financial measures to the marketing-finance interface discussion.

Web Appendix D. How heterogeneous trading agents in financial markets enable the permanent and temporary price impacts of FGC

The ability of trading agents (traders, investors etc.) in financial markets to observe and effectively decipher information events, such as FGC (including their attributes such as valence and subject matter), in a timely manner varies significantly. This variation in ability to observe the information content of events is underscored based on the existence of heterogeneously informed agents in financial markets (Glosten and Milgrom, 1985); in the classical market microstructure literature, these are classified into two broad groups: informed and uninformed traders (see, as an example, Grossman and Stiglitz, 1980). Informed traders' trading activity conveys information to the market and therefore is linked to permanent price impact, while uninformed trading activity is linked with temporary price impact.

Informed traders are typically modeled as investing in the acquisition of information, which they then trade with for profit by adversely selecting uninformed traders (O'Hara, 2003). This is not necessarily negative because both informed and uninformed traders are crucial to the price discovery process in financial markets. Specifically, informed trading activity is needed for price discovery, while the presence of uninformed traders in financial markets, who can be taken advantage of, incentivizes informed traders to acquire the information with which they subsequently trade. Once the information held by informed traders is incorporated into price, a revealing equilibrium ensues, leading to a more efficient price (O'Hara, 2003). This process occurs at high speeds and a vast number of times throughout the average trading day in the global financial market.

The heterogeneous nature of market participants suggests that only informed traders will invest the necessary resources required to decipher their relevance for their linked securities. In a market driven by algorithmic trading, this suggests investment in the technological apparatus necessary to analyze the information and its content at high speed and subsequently to incorporate the information they convey into trading decisions. Thus, an informed trader is likely to be an institutional trader, such as an investment bank's algo trading desk or asset manager, while an uninformed trader could be the average retail trader trading through a broker app, such as *Robinhood*. This is the reality of trading in the US, European and several Asian markets today. Hence, if an event (e.g. FGC) contains information relevant to the value of a firm, only a rather small section of the market will observe that information and exploit it in a timely manner, while the rest of the market only becomes wiser with the attainment of the inevitable revealing equilibrium, i.e. permanent price impact. Prior to the revealing equilibrium, a large section of the market will largely conduct trading uncorrelated with the firm's value and thus generate a temporary price impact. This explains why both temporary and permanent price impacts could be observed around the release of an information-conveying event like a FGC. It is important to note that the price impact of FGC, or any event, cannot be estimated without trading; it is trading activity that incorporates the information or noise content of an event into price. Hence, approaches for documenting the effect of the stock price impact of FGC will inevitably capture both the temporary price impact, due to uninformed trading activity, and permanent price impact, due to informed trading activity.

Web Appendix E. FGC and its impacts

Reference	FGC Attributes	Performance Metric	Estimation Approach	Level of Analysis	Findings
Kumar et al. (2016) ***	Valence – positive or negative sentiment Receptivity – consumers' response to social media messages Consumer susceptibility – consumers' predisposition towards using social media	Consumer metrics: consumer spending (transaction value of consumer to the firm), cross-buying (number of different product categories consumer purchased) and consumer profitability (in-store transaction)	Propensity score matching (PSM); difference-in-differences (DID) analysis	N/A	FGC has a positive and significant effect on consumer behavior, including spending and cross-buying behavior The effects of FGC valence, receptivity and consumer susceptibility are significant, although the effect of receptivity is the greatest.
Colicev et al. (2018) *, **	Volume – number of FGC pieces	Consumer mindset metrics; brand awareness, purchase intent, consumer satisfaction Shareholder value: abnormal returns, idiosyncratic risk	VAR	Daily	FGC increases brand awareness and consumer satisfaction, but not purchase intent Consumer satisfaction (along with purchase intent) positively affects shareholder value
Colicev, Kumar and O'Connor (2019) *	Valence – positive or negative sentiment Vividness – content richness	Marketing funnel stages (awareness, consideration, purchase intent and satisfaction)	VAR	Daily	FGC has a strong relationship with consideration and purchase intent

	ranging from text to video				
Hewett et al. (2016) *, ***	Volume – number of FGC pieces Valence – positive or negative sentiment	Volume of word-of-mouth, consumer sentiment, advertising spend	VAR	Weekly	High-volume, consistent, moderately toned FGC helps manage word-of-mouth, and lift consumer sentiment and firm outcomes
Tellis et al. (2019)	Emotional versus informational content 60 ad characteristics	Virality – a number of views in a short time period due to sharing	Mixed-effects regression model	N/A	Information-focused content has a significantly negative effect on sharing, except in risky contexts Emotional ads are shared more on general platforms (Facebook, Google, Twitter) than on LinkedIn, and the reverse holds for informational ads
Meire et al. (2019) *	Emotional versus informational message	Sentiment of consumer digital engagement – positive or negative valence	Generalized linear mixed-effects model	Daily	Emotional FGC has a positive and significant influence on consumers' sentiment regardless of the event outcome Informational FGC, more so than emotional content, improves the sentiment of consumers' digital engagement following the negative event
Borah et al. (2020)	Improvised marketing intervention (IMI) – humorous message	Virality – the number of shares of a marketing message Abnormal stock market returns	DID, panel regression, event study	Daily	Humorous FGC has positive effects on virality and firm value

* In addition to FGC, UGC was examined.
** In addition to FGC, earned media was studied.
*** Synergic effects of FGC with offline communication, e.g. TV advertising, and email communication, e.g. press releases, were also studied.

Web Appendix F. Twitter data sample

The table reports the frequency statistics for a sample of tweets generated between January 8th 2013 and August 17th 2018 for 64 S&P 500 IT firms with stocks included in the S&P 500 index. The tweets and associated data are obtained using an Application Programming Interface (API) to scrape them from Twitter.

S&P 500 firm	No. of tweets	No. of tweet days	Min. no. of tweets per day	Max. no. of tweets per day	Average no. of tweets per day	Single tweet days (%)	No. of tweets excluded*	No. of days excluded*
Accenture	3,046	603	1	39	5.05	3.58	28	27
Activision Blizzard	317	171	1	11	1.84	30.6	6	34
Adobe	2,168	438	2	23	4.93	0	58	34
Akamai	3,103	847	1	55	3.65	3.54	25	28
Alliance Data Systems	3,189	790	1	96	4.03	9.09	32	29
Automatic Data Processing	1,245	554	1	33	2.24	19.92	0	25
Alphabet A (ex. Google)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Alphabet C (ex. Google)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AMD	2,244	1,033	1	24	2.17	21.57	1	10
AMPHENOL	33	21	1	4	1.5	42.42	1	19
Analog Devices	3,220	607	1	40	5.29	2.64	0	26
ANSYS	3,141	425	1	42	7.37	1.97	22	20
Apple	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Applied Materials	3,201	1,080	1	48	2.12	27.73	25	33
Autodesk	2,943	859	1	44	3.42	1.29	36	28
Broadcom	2,871	1,136	1	29	2.52	19.23	38	34
CA	3,052	232	1	69	13.09	0.26	0	27
Cadence Design Systems	3,241	1,052	1	28	3.07	9.66	28	29
Cisco	2,568	719	1	33	3.56	3.35	22	33
Citrix Systems	2,603	439	1	75	5.91	0.92	32	33
Cognizant	3,094	294	1	32	10.48	0.65	34	36
Corning	3,030	875	1	32	3.45	7.66	32	32
DXC Technology	2,239	155	3	30	14.35	0	5	8
Electronic Arts	2,109	541	1	43	3.89	2.7	0	28
F5 Networks	3,041	355	2	32	8.54	0	15	30
Facebook	240	84	1	19	2.82	14.17	0	29

Fidelity National Information Services	3,067	698	1	47	4.38	5.09	14	24
Fiserv	3,151	768	1	26	4.09	5.59	19	31
FLIR Systems	3,228	413	1	30	7.79	1.33	29	27
Gartner	3,229	393	1	85	8.19	0.77	16	24
Global Payments	1,920	874	1	25	2.19	23.28	26	24
Harris	3,201	843	1	20	3.79	5.19	21	22
HP	1,052	644	1	6	1.63	37.07	21	32
Hewlett Packard Enterprise	1,434	541	1	16	2.64	8.3	15	13
IBM	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Intel	989	393	1	40	2.51	20.02	22	39
Intuit	2,213	943	1	37	2.34	19.84	17	27
Juniper Networks	3,030	949	1	21	3.18	5.21	26	34
KLA-Tencor	2,106	1,597	1	5	1.31	56.17	35	28
Lam Research	3,228	942	1	25	3.42	4.89	27	31
Mastercard	3,157	545	1	38	5.78	2.25	26	25
Microchip Technology	2,710	451	1	21	5.99	0.15	28	24
Micron Technology	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Microsoft	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Motorola Solutions	2,470	531	1	63	4.64	3.12	18	26
NetApp	2,911	416	1	37	6.98	0.55	21	32
Netflix	1,599	269	1	50	5.92	1.19	25	36
Newell Brands	3,206	1,257	1	40	2.54	20.02	14	27
NVIDIA	2,893	778	1	92	3.71	7.43	19	35
Oracle	3,187	331	1	105	9.59	0.85	20	24
Paychex	3,072	428	1	68	7.16	1.46	10	30
PayPal	2,191	806	1	37	2.71	13.56	15	17
Qorvo	1,682	1,082	1	7	1.55	39.42	17	19
QUALCOMM	2,162	716	1	25	3.01	10.36	25	34
Red Hat	3,102	204	2	155	15.13	0	20	25
Salesforce	1,516	297	1	22	5.08	1.39	23	28
Seagate	1,140	441	1	15	2.57	11.05	0	30
Skyworks Solutions	1,296	707	1	8	1.83	31.48	22	26
Symantec	3,065	452	1	36	6.76	2.35	19	40

Synopsys	3,228	1,402	1	56	2.30	20.79	19	30
TE Connectivity	3,020	1,196	1	27	2.52	12.42	21	25
Texas Instruments	2,987	621	1	25	4.80	2.98	0	28
The Western Union Company	212	124	1	20	1.69	35.38	2	31
Time Warner	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Total System Services	3,224	972	1	28	3.31	6.27	20	28
Twitter	589	316	1	18	1.85	29.71	2	27
VERISIGN	3,160	731	1	20	4.31	4.27	16	19
Visa	241	146	1	9	1.63	41.91	7	29
Western Digital	1,241	411	1	48	3.01	13.54	0	25
Xerox	2,901	894	1	34	3.24	1.96	24	21
Xilinx	3,093	672	1	2	1.00	2.55	18	29

*Excluded because of excessive return volatility.

N/A: data not available due to API restrictions or companies not having established Twitter accounts.

Web Appendix G. Robustness analysis: the impact of FGC on stock returns

As an extension to the results of the comparative analysis of the impact of FGC, for robustness, following Frino, Jarnecic and Lepone (2007) we estimate the percentage return/price impact for each tweet-trade. We find that, on average, a tweet-trade has a mean permanent price return of 0.74%, with a median return of 0.05%. A few tweet-trades yield much higher returns, which explains the average being higher than the median. As a comparison, the price impact of block trades (which are typically highly informative – indeed these large trades generate the most intraday impact) estimated using the same measures as in Frino, Jarnecic and Lepone (2007) range from.14% to.40% and in Sun and Ibikunle (2017) they range from.011% to.020%. Comparing these return estimates with similar estimates from the investigations conducted by other studies further underscores the extent of the impact tweets can have on firm value, especially considering that our estimates are based on 60-second or shorter event windows. Nevertheless, such a consideration of the economic impact of tweets provides an incomplete picture of FGC's impact on price because it ignores the significant temporary price impact, which can increase the cost of capital and costs associated with trading a firm's stock (Diamond and Verrecchia, 1991; Chan and Lakonishok, 1993). A holistic understanding of the permanent and temporary price impacts of FGC alongside its economic impact provides a more robust and complex understanding of marketing's financial impact.

Web Appendix H. Permanent and temporary price impact and tweet valence and orientation; panel least squares

Variables	Permanent price impact $(\Delta\sigma^{2u}_{s,t})$	Temporary price impact $(\Delta\sigma^{2i}_{s,t})$
$consumer_{s,t}$	-.004 (1.13)	.005** (2.36)
$competitor_{s,t}$	-.063** (2.24)	.014** (2.39)
$-ve_{s,t}$	-.086** (2.42)	.029** (2.45)
$+ve_{s,t}$	-.098*** (3.09)	.030** (2.18)
$consumer_{s,t} * -ve_{s,t}$	.029* (1.83)	.065** (2.41)
$competitor_{s,t} * -ve_{s,t}$	.515*** (3.62)	.009** (2.01)
$consumer_{s,t} * +ve_{s,t}$	.082** (2.04)	.023** (2.39)
$competitor_{s,t} * +ve_{s,t}$	.161*** (4.79)	.068** (2.40)
$lnvolume_{s,t}$	-.039*** (-4.50)	-.018*** (-5.97)
$lntradesize_{s,t}$	.086*** (6.33)	.030*** (6.34)
$volatility_{s,t}$	-.114*** (-3.55)	.014** (2.13)
$Effectivespread_{s,t}$	.019** (2.58)	.005** (1.98)
$lnHFT_{s,t}$	-.000 (-0.25)	-.013*** (-3.80)
$OIB_{s,t}$	-.370*** (-6.85)	.045** (2.40)

$\ln\#followers_{s,t}$	-.074** (-2.35)	.030** (2.37)
$\overline{R^2}$	.34	.46
Observations	139,997	139,997
Kleibergen-Paap LM (tests the null that the employed instruments have insufficient explanatory power to predict the endogenous variables in the model for identification of the parameters)		
Cragg-Donald (tests the same null hypothesis as the Kleibergen-Paap LM test)		
Sargan's χ^2 p-value (tests the null that the over-identifying restrictions are valid)		