Journal of Business Research 69 (2016) 2087–2092

Contents lists available at ScienceDirect

Journal of Business Research

# Examining the influence of stock market variables on microblogging sentiment☆

Juan Ramón Piñeiro-Chousa *, M. Ángeles López-Cabarcos, Ada María Pérez-Pico

*Faculty of Business Administration and Management, University of Santiago de Compostela, s/n Alfonso X El Sabio Ave., E27002 Lugo, Spain*

ARTICLE INFO

*Article history:*
Received 1 August 2015
Received in revised form 1 October 2015
Accepted 1 November 2015
Available online 31 December 2015

*Keywords:*
Fuzzy-set QCA
Microblog sentiment
Tobin's *Q*
P/E ratio
Number of followers
Experience of users

ABSTRACT

This research investigates different combinations of causal conditions that may relate to microblogging sentiment. The study considers financial variables, such as VIX, Tobin's *Q*, capitalization, or P/E ratio, and variables related to social media activity, such as number of twits, number of followers, or the experience of the users. Using a fuzzy-set qualitative comparative analysis, this research analyzes the sentiment of stock-related microblogging messages of S&P 500 Index posted between October 2009 and October 2014. The model where the message sentiment increases finds three causal recipes and all of them contain user experience and number of followers: (1) the configuration within user experience, number of followers, and Tobin's *Q*; (2) the configuration within user experience, number of followers, and lower P/E ratio, lower capitalization, and lower Tobin's *Q*; (3) and the configuration within user experience, number of followers, P/E ratio, and Tobin's *Q*. The study discusses the results.


## 1. Introduction

In recent years, social networks and blogs have become increasingly important to the functioning of financial markets because they provide great amounts of information to users in a very short period of time. In this regard, some recent studies analyze the use of Twitter by investors to share information and opinions that in the end influence the markets although not in great depth (Sprenger, Tumasjan, Sandner, & Welpe, 2014; Zhang, Fuehres, & Gloor, 2011). The advance in social networks is unstoppable in this field, and recently, another social network specializing in the financial markets, namely StockTwits.com, has come online with the aim of analyzing the possible predictable relationship between the performance of shares and the activity on social networks (Oh & Sheng, 2011; Rao & Srivastava, 2012). The fact that these relationships exist does not mean that the data obtained from social network activity can predict market movements because researchers must consider many more variables, such as investor sentiment, market risk (Lee, Ribeiro, Olson, & Roig, 2007; Ribeiro-Soriano & Roig-Dobón, 2009; Ribeiro-Soriano & Urbano, 2010), the type of company, the sector as well as other information of a financial nature. However, information gathered from social media can help to obtain a more efficient prediction of market movements (Bissattini & Christodoulou, 2013; Oliveira, Cortez, & Areal, 2013). In this sense, external factors (Guzmán-Cuevas, Cáceres-Carrasco, & Soriano, 2009; Roig-Dobón & Ribeiro-Soriano, 2009) can affect the investor's behavior, which in turn affects the market, being one of this external factors the investor sentiment, that is, how investors form beliefs (Barberis, Shleifer, & Vishny, 1998). Recently, some studies used social network sentiment to try to predict market movements (Asur & Huberman, 2010; Makrehchi, Shah, & Liao, 2013; Oh & Sheng, 2011).

This study analyzes the relationship between the behavior of financial markets and social network sentiment, specifically between the S&P 500 Index and the sentiment of the messages on StockTwits.com using Stanford CoreNLP Natural Language Processing Toolkit. This analysis takes into account different financial variables, such as the measure of market volatility or VIX index, the measure of the valuation of the companies or Tobin's *Q*, the capitalization value of the company, and the share price of the company relative to its earnings per share or P/E ratio. The study also considers variables related to social media activity, such as the number of twits, the number of followers and the experience of the users. The time period of the sample is from October 2009 to October 2014, which is longer than other studies covering approximately six months (Sprenger et al., 2014; Zhang et al., 2011), four months (Oh & Sheng, 2011; Makrehchi et al., 2013), or three months (Asur & Huberman, 2010). In contrast with other studies that use daily data (Zhang et al., 2011; Sprenger et al., 2014) or intraday data (Antweiler & Frank, 2004), this study uses aggregate data per month to add a further dimension to the results.

This empirical study contributes to exploring the combined antecedents of the social network sentiment, trying to achieve greater financial depth of knowledge not limited to aspects of computer software

☆ The authors thank Norat Roig Tierno, Politécnica de Valencia University and Domingo Ribeiro Soriano, Valencia University, for their careful reading and suggestions.
* Corresponding author. Tel.: +34 981 563100.
*E-mail addresses:* j.pineiro@usc.es (J.R. Piñeiro-Chousa), angeles.lopez.cabarcos@usc.es (M.Á. López-Cabarcos), adamaria.perez@rai.usc.es (A.M. Pérez-Pico).

http://dx.doi.org/10.1016/j.jbusres.2015.12.013
0148-2963/

development. In this regard, the method of analysis is different from those used in previous research in supposing that different combinations of causal conditions may be linked to social network sentiment. Therefore, instead of linear regression models, this study uses configurational comparative methods to uncover multiple paths to explain social network sentiment (Mas-Verdú, Ribeiro-Soriano, & Roig-Tierno, 2015; Ragin, 1987, 2000, 2008; Rihoux & Ragin, 2009). In contrast to previous studies, which analyze if the microblog sentiment can predict stock returns, this study analyzes the links between different combinations of causal conditions (market variables and variables related to social media activity) and the outcome (the microblog sentiment), analyzing aggregate data per month during five years, using a qualitative comparative analysis and fuzzy sets methodology.

The structure of the study is as follows: Section 2 presents the theoretical background. Section 3 contains the method. Section 4 exposes the results from the analysis. Finally, Section 5 presents the discussion and future research and limitations.

## 2. Theoretical background

### 2.1. Research of online stock blogs

The Internet contains a great variety of sources related to the financial market field, such as financial market activity forums, blogs, social networks, magazines, and newspapers. Obviously, studies which analyze the relationship between social networks and the financial market do not always use the same data sources. In this regard, the study by Wysocki (1998), the first to carry out this type of research, used Yahoo! Finance posts mentioning the fifty most cited companies in this forum. In contrast, Antweiler and Frank (2004) used posts on general-topic blogs, and Tumarkin and Whitelaw (2001) used data from the RagingBull.com blog. De Choudhury, Sundaram, John, and Seligmann (2008) made a model of the information roles and contextual properties of four technological companies, finding that communication activities in the blogosphere closely correlate with movements in the financial market. Bordino et al. (2012) gathered information from the Yahoo! Search engine, showing the existence of a positive correlation between web searches regarding shares and their transaction volume in subsequent days. Da, Engelberg, and Gao (2011) used the frequency of searches on Google (SVI) to measure the attention of investors, and they concluded that an increase in the SVI for the Russell 3000 share index precedes high share prices in the following two weeks. Besides blogs or search engines on the Internet, researchers also use news websites as a source for the analysis of microblogging and the financial market. Gilbert and Karahalios (2010) analyzed 20 million posts on LiveJournal to create an anxiety index in the United States and demonstrate that an estimation of sentiment on web blogs provides information on the future price of shares. Likewise, Schumaker and Chen (2009) analyzed 9211 financial news articles and 10,259,042 mentions of shares from the S&P 500 Index over five weeks; and Tetlock (2007), who used the daily content of the *Wall Street Journal*, demonstrated that high levels of pessimism in the media predict a downward pressure on market prices followed by a restoration of fundamentals.

In recent years, research using Twitter as a source of information for the analysis of the influence of social networks on financial markets has proliferated. Twitter is a microblogging service created in 2006, which allows its users to make short posts on whatever they like (up to 140 characters) and to share them. Subscriptions to accounts as followers establish the relationships between users. The main characteristic of this service is the immediacy of its information, because once a user posts a tweet, all users following the sending account will receive the tweet instantly. This characteristic allows the broadcasting of information in real time. Another characteristic of Twitter is the use of keywords (hashtags) that allow the immediate location of conversations by topics. These characteristics make Twitter one of the best-suited social networks for analyzing how social activity affects the financial markets. For example, Asur and Huberman (2010) show how to use sentiments extracted from Twitter to predict the market. Sprenger et al. (2014) analyze about 250,000 tweets on financial markets concluding that the sentiments expressed in tweets have an association with the abnormal performance of shares, and posts concerning volume predict the volume of transactions during the following day. Zhang et al. (2011) analyze the mood on Twitter, comparing that mood with the Dow Jones, S&P 500, and NASDAQ to predict the indexes' performance through the tweets. Other authors analyze the mood or feeling of the masses on Twitter and their power of prediction regarding market behavior (Bollen, Mao, & Zeng, 2011; Chekanskiy, 2012). In recent years, researchers have been using StockTwits.com as a suitable source of data. StockTwits.com is a social network similar to Twitter, but with an exclusively financial scope. Just like Twitter, users write and share posts (also a maximum of 140 characters), but the posts are about stocks, indexes, and financial markets. Some studies use Twitter and StockTwits.com as a suitable data source for financial purposes, finding evidence of the relationship between activity on the two social networks and stock prices or trading volume (Irvine & Giannini, 2012; Ruiz, Hristidis, Castillo, Gionis, & Jaimes, 2012). Compared with Twitter, StockTwits offers a set of advantages to conquer the objective of this research: (a) all posts are of a financial nature, therefore separating financial from non-financial contents is unnecessary; (b) users indicate in their profile their experience as investors (novice, intermediate, professional), information which is highly relevant for the analysis; and (c) the user can catalog the post as bullish or bearish. Ultimately, scholars consider data from StockTwits.com to be better for predicting market performance than data from Google Trends (Loughlin & Harnisch, 2014; Oh & Sheng, 2011).

### 2.2. Investor sentiment

No consensus exists on a definition of investor sentiment because the context can modify its meaning. Broadly, investor sentiment is how investors form beliefs (Barberis et al., 1998), or as the propensity to speculate (Baker & Wurgler, 2006). Methods to measure this construct include surveys of consumer confidence or measures extracted from the closed-end fund discount (CEFD) (Qiu & Welch, 2004). Other measures are indices, such as the Equity Market Sentiment Index (EMSI) (Bandopadhyaya & Jones, 2006). Baker and Wurgler (2007) use an index based on the six proxies appearing in their earlier study (Baker & Wurgler, 2006), where the authors used NYSE turnover, the dividend premium, the CEFD, the number and amount of first-day returns on IPOs, and the equity share in new issues to measure volume.

In recent years, a new measure of sentiment has arisen, microblogging sentiment. Researchers extract this sentiment from microblog posts on social networks like Twitter or StockTwits.com; and studies confirm the existence of relationships between market activity and microblog activity (Antweiler & Frank, 2004; Tumarkin & Whitelaw, 2001). Some authors use sentiment analysis of stock microblogging posts to try to predict stock price directional movements. In this sense, Asur and Huberman (2010) show that sentiments extracted from Twitter can predict the market. The authors demonstrate that the use of sentiments obtained from Twitter can improve the forecasting power of social networks. Other authors analyze the role of stock microblog sentiment forecasting stock price directional movements with success, finding that sentiment has a significant predictive power for stock market movement (Bissattini & Christodoulou, 2013; Makrehchi et al., 2013; Oh & Sheng, 2011). In this sense, Bollen et al. (2011) show that some specific public mood dimensions derived from Twitter can improve the accuracy of DJIA predictions.

However, the predictive power of microblog sentiment depends on the sentiment index used, because, for example, Logunov and Panchenko (2011) use an index based on the emoticons of Twitter messages that fails to detect any predictability. This study analyzes the sentiment of the messages on StockTwits.com with Stanford CoreNLP Natural Language Processing Toolkit, developed by Manning et al.