A social-media-based approach to predicting stock comovement

Ling Liu [a], Jing Wu [a], Ping Li [a], Qing Li [a,b,*]

[a] *School of Economics Information Engineering, Southwestern University of Finance and Economics, China*
[b] *Management Information Systems Department, University of Arizona, USA*

ARTICLE INFO

ABSTRACT

Stock return comovement analysis is important to financial analysts, decision makers, and academic researchers and has many financial implications, such as portfolio management, style investing, and market risk detecting. This paper proposes a novel model to both identify homogeneous stock groups and predict stock comovement with respect to firm-specific social media metrics. One of the innovations of the social media platform is that it breaks traditional media intermediation. A firm with an official Twitter account can publish information and interact with its users directly. Such direct information is largely reflected on firm-specific metrics, e.g., the firm's number of followers and number of tweets sent. To the best of our knowledge, this paper is the first to reveal the impact of social media metrics on stock return comovement studies. By analyzing samples from the NYSE and NASDAQ stock exchanges, we find that firms with official Twitter accounts have a much higher comovement than those without such accounts. Furthermore, we classify the former set of firms into homogeneous groups by their specific microblogging metrics. The results demonstrate that these metrics cannot only predict the comovement of stocks but also notably increase the accuracy of comovement predicting, compared with industry categories.

© 2015 Published by Elsevier Ltd.

1. Introduction

Stock return comovement analysis refers to identifying homogeneous groups of stocks that have similar movement of returns. Such an analysis is very important to financial analysts, decision makers, and academic researchers. On a practical level, the analysis of homogeneous stock groups can help investors construct investment portfolios. To minimize risks, many investors allocate their funds across different homogeneous assets groups (Barberis & Shleifer, 2003; Chan, Lakonishok, & Swaminathan, 2007; Swensen, 2009). Stock return comovement is also inversely associated with economic resources allocation (Chelley-Steeley, Lambertides, & Savva, 2013; Durnev, Morck, & Yeung, 2004; Wurgler, 2000).

Although many factors lead to similar stock returns, such as firm size, industry, investor behavior, and market characteristics (Aghabozorgi & Teh, 2014; Brockman, Liebenberg, & Schutte, 2010; Chelley-Steeley et al., 2013; Claessens & Yafeh, 2012), we believe that it is still necessary to propose a new model to identify homogeneous stock groups from a social media perspective. Social media has changed how society communicates and organizes itself

(Aral, Dellarocas, & Godes, 2013; Golbeck, Grimes, & Rogers, 2010; Yu & Kak, 2012). In particular, microblogging service platforms allow users to follow each other and, thus, to bypass the intermediate channel of information distribution between firms and the public. Communication among the public is also accelerated and enhanced by microblogging services. It is reasonable to infer that the stock market is not immune to the impact of such activities. In fact, a great deal of research has shown that social media users' activities can be used to predict firms' financial performances (Bollen, Mao, & Zeng, 2011; Yu, Duan, & Cao, 2013). We further suggest that such public behavior toward firms can be used as a proxy to predict the comovement between stocks.

By identifying official Twitter accounts for each publicly traded firm, our empirical results demonstrate that firms with official Twitter accounts have much higher comovement than firms without such accounts. This finding is similar to the results of Barberis, Shleifer, and Wurgler (2005), which showed that when added to a major index, firms experience higher comovement because investors have better information access to the firms after they were added to the index. This might indicate that Twitter can enhance the diffusion of firm information, resulting in the same effects those caused by major stock indexes.

Furthermore, this paper proposes a novel model to identify homogeneous stock groups and to predict stock comovement with respect to firm-specific microblogging metrics. Unlike most

* Corresponding author at: Management Information Systems Department, University of Arizona, USA.

Q4

E-mail address: liq_t@swufe.edu.cn (Q. Li).

previous studies (see Bollen et al., 2011; Luo, Zhang, & Duan, 2013; Yu et al., 2013), which focus on firms' overall social media metrics, we concentrate on firm-specific metrics. Overall metrics refers to a firm's broad metrics on the entire platform, e.g., how many tweets have mentioned the firm's name, or the public mood associated with all firm-related tweets. Our research identifies each listed firm's microblogging account and examines their specific metrics. This approach has several advantages. First, retrieving overall metrics requires creating filters that use stock symbol, company name, and other keywords to identify firm-related information (Ruiz, Hristidis, Castillo, Gionis, & Jaimes, 2012). Such identifiers and criteria can be ambiguous, and a great deal of noise information might be collected, which in turn should be carefully handled. Firm-specific metrics, however, are entirely relevant to each firm. Second, retrieving overall metrics requires managing a large volume of data. Thus, most studies can only choose companies in limited industries to study, which may impact the generalization of results (Luo et al., 2013). Our approach enables us to analyze a larger dataset of firms from all of the industries listed on NYSE and NASDAQ stock exchanges.

Based on the dataset, we take key firms' microblogging metrics, e.g., the number of followers, as observations; and we then cluster those firms into homogeneous groups. The idea of clustering stocks into groups is already well accepted. For example, Farrell (1975) heuristically partitions stocks into four homogeneous groups and discusses their implications on portfolio construction, performance measurement, and group rotation. Aghabozorgi and Teh (2014) propose a three-phase clustering method to automatically categorize stock markets into different groups. Liao and Chou (2013) adopt a data mining approach to investigating the comovement between Taiwan and Hong Kong stock markets. By applying the model, our results show that firms in the same clustered groups have higher comovements than those in the same industry categories. Our research proves that with simple metrics, social media can produce better results compared to other traditional industry classifications. Another advantage of our model is that our proposed metrics are easy to be understood and verify, and they might attract more practical usage than complex models.

The remainder of the paper is structured as follows. Section 2 introduces the background and theory of the paper, and is followed by Section 3, which describes the model. Section 4 describes the data collection and research framework. Section 5 discusses the results. Finally, Section 6 summarizes the paper.

2. Theory and related work

2.1. Stock comovement

Comovement, as a component of market beta in the form of covariance, exerts an important influence over portfolio performance, the market risk premium, and the cost of capital (Chelley-Steeley et al., 2013; Roll, 1988). Thus, it has attracted a great deal of attention in many areas and is surrounded by a long-standing debate on its causes. Efficient Market Hypothesis (EMH) theory suggests that a firm's stock price reflects all the information and is thus only influenced by the firm's fundamental values (Fama, 1970). It is grounded on two basic assumptions: (1) all investors are rationals, and (2) stock markets are informationally efficient. According to this theory, the prices of different stocks move together only in response to common movements in their fundamental values.

However, these two assumptions have been consistently challenged by behavioral finance research. For instance, Rashes (2001) has discovered that even two assets with nothing in common but their similar names have abnormal comovement, caused by investors who have misidentified the assets' names. This research indicates that investors are not always rational and that their behavior may cause stock comovement. Pindyck and Rotemberg (1993) have found that company size and degree of institutional ownership also influence stock comovements. The work of Dutt and Mihov (2013) has shown that markets with similar industries demonstrate higher stock comovements. Such effects occur because many investors prefer to group stocks into categories based on their basic characteristics, such as firm size and industry categories, and then allocate funds at the level of those categories (Barberis et al., 2005). Interactions among investors are also found to have an impact on stock comovement. Based on Internet message discussion boards, Antweiler and Frank (2004) have demonstrated that investors' activities influence trading volume and return volatility. Chue, Gul, and Mian (2014) have find that aggregated stock market comovement becomes higher when investor sentiment is extreme.

Uneven information diffusion is another major factor that is considered to have an impact on stock return comovement. The media, the major information diffusion intermediate, plays an important role in stock price fluctuation (Li et al., 2014a). Quantified media content is proven to be able to predict stock prices (Chan, 2003; Tetlock, 2007). Researchers also have found that uneven diffusion of information can inspire stock comovement. When added to a major index, stocks experience significant increase in their betas (Vijh, 1994). The reason for such a change is that investors may have better access to information related to stocks in those major indexes (Claessens & Yafeh, 2012).

A great deal of research is based on media analysis which was once the only channel for information distribution and acquisition. Furthermore, prior to the social media era, it was difficult to collect investors' firm-related activities toward, which makes media a good anchor. The emergence of social media changed this situation.

2.2. Social media and firm financial performance

Social media has dramatically changed how people obtain information. In particular, microblogging services, which allow users to follow each other, to make comments and to forward tweets, provide a more flattened and efficient way for people to acquire and distribute information. Moreover, microblogging platforms also enable us to directly collect the public's opinions and behavior toward firms. Although tweeting may not provide new firm information to the market, firm tweets are found to increase information dissemination (Alexander & Gentry, 2014). Recent research has demonstrated that users' social-media activities can be used to predict firms' financial outcomes (Goh, Heng, & Lin, 2013; Tirunillai & Tellis, 2012; Zheludev, Smith, & Aste, 2014). Based on joint mentions in Twitter's public tweets, Sprenger and Welpe (2011) suggest that news-based relatedness can help delineate meaningful industry groups.

The stream of research analyzing the relationship between public activities on social media platforms and firms' financial performance is very popular. Researchers from marketing, information systems, economics, and many other domains have explored the effectiveness of social media information. A great deal of research has found that consumers' social media activities and interactions can affect their consumption intentions, which in turn will influence firms' sales (Dellarocas, Awad, & Zhang, 2004; Lee, Shi, Cheung, Lim, & Sia, 2011; Phang, Zhang, & Sutanto, 2013; Zhu & Zhang, 2010). Bollen et al. (2011) find that the public's mood on Twitter can predict the Dow Jones Industrial Average. Luo et al. (2013) discovered that social media metrics are significant indicators of firm equity value. Sul, Dennies, and Yuan (2014) collect data from public Twitter posts about S&P 500 firms and assert that the