Microblogging Metrics and Stock Return Comovement

Ling Liu, Jing Wu, Qing Li
Southwestern University of Finance and Economics

Abstract

Identifying homogeneous groups of stocks, where these stocks have similar movement of returns is called stock return comovement analysis. Stock return comovement analysis is important to financial analysts, decision makers, and academic researchers, in many financial implications. This paper examines firms’ social media, in particular, microblogging metrics’ role on analyzing stock return comovement. The results show microblogging metrics can effectively identify homogenous stock groups.

Keywords
Social Media, Stock Return Comovement, Microblog, Homogeneous Stock Groups

1 Introduction

Identifying homogeneous groups of stocks, where these stocks have similar movement of returns is called stock return comovement analysis. Stock return comovement analysis is important to financial researchers, analysts, investors, and decision makers, in many financial implications, such as, portfolio management, style investing, and market risk detecting (Farrell 1975, Barberis & Shleifer 2003, King 1966, Chelley-Steeley et al. 2013). There are two sources of comovement, rational and behavioral. The former was based on Efficient Market Theory, which supposes all market participants are rational and the comovement between stocks are totally based on their similar fundamental values. However, behavioral finance suggests that investors are not rational and therefore, their behavior will influence the price of the stocks. For example, Rashes (2001) found that two assets have nothing in common but similar names have higher abnormal comovement. The comovement was intrigued by the confusion of their names from the investors.

Social media has changed the way society communicates and organizes itself (Aral et al. 2013). In particular, microblogging service platforms have provided us an automatic, fast, free, and large- scale addition to discover firms’ latent attributes (Bollen et al. 2011, Jansen et al. 2009). More and more researchers have devoted to find connections between social media information and firm financial outcomes. For example, using sentiment analysis, Bollen et al. (2011) find that the public’s mood on Twitter can predict the Dow Jones Industrial Average. Yu et al. (2013) suggest that overall social media metrics have a stronger relationship with firm
stock performance than conventional media, and Luo et al. (2013) find that social media metrics are significant indicators of firm equity value.

Follow previous research, this paper aims to whether the public behavior on social media can be used to identify homogeneous stock groups. We first identify the official microblogging accounts of listed companies and then use related metrics to cluster firms into different groups. The results show that controlling firms’ market capitalization, stocks divided by social media metrics have higher within-group correlations than the best industry classification scheme.

2 Data and Research Method

Previous work often focuses on firms’ overall social media metrics, which were estimated by collecting the tweets that including firms’ names. The approach may cause many noises and inaccuracy of the data. This paper uses firm-specified metrics instead. Firm-specified metrics are the metrics that related to firms’ official microblogging accounts, for example, the number of followers a firm’s account has, and the number of tweets the account has sent.

This paper takes the firms that are publicly traded in China’s Shanghai and Shenzhen stock exchanges as samples. Following criteria were set to identify a company’s official microblogging account. 1) It must be a verified account. Sina Weibo provides verification of accounts to establish the authenticity of identities. Thus, if a company’s account is not verified, it is not treated as its official microblogging account. 2) The product’s or department’s account is not treated as the firm’s account. A web crawler program was written to collect the data from Sina Weibo, which collected a set of cross-sectional data including the values of the number of followers. The program started from 30 Jun 2013 and ran continuously for two days (crawling was finished on 2 July 2013). Quarterly market capitalization and monthly stock returns of the firms between 01 Jan 2008 and 31 Dec 2012 were collected from the RESSET database and The China Stock Market & Accounting Research (CSMAR) database. Then we collected the number of followers of each account.

One of the most significant factor or microblogging is that users can follow each other. The following action indicates a favor of the user to the ones they followed. The number of followers is a well-known reflector of public opinions upon the firms. A larger number of followers indicates the firm is more popular in common users. According to previous research, public’s opinion, such as sentiment can influence firms’ stock price, we hence suggest that firms with similar number of followers may have higher comovement than the ones have less similar number of followers. Furthermore, from the information diffusion perspective, firms have more followers will have better information diffusion than the ones have less followers. In other words, the number of
followers can be seen as an indicator which can estimate their information diffusion effectiveness. Therefore, we use the number of followers as an indicator to identify homogenous stock groups.

Moreover, it is well accepted that firm size has a great influence on stock return comovement, i.e., firms have similar sizes have higher comovement. Thus, firms’ followers (FO) and market capitalization (MC) were used as K-means clustering observations to identify homogeneous stock groups. Then, we use 2008 to 2012 time-series stock monthly return data to test the effectiveness of the model. We followed Chan et al. (2007)’s method to compare the within-group pairwise stock return correlations to the GICS schemes’ within-industry pairwise stock return correlations. A better correlation indicates a better classification.

The within-group stock return correlation is calculated as follows. Suppose there are S groups in one classification scheme and each group has M subgroups, each of which has N stocks. The pairwise correlation between stock i and the other stocks in the same subgroup is:

\[ R_i = \frac{\sum_{j \in M, j \neq i} R_{ij}}{N - 1} \]

\( R_{ij} \) denotes the pairwise correlation between stock i and j. Then, the average pairwise stock return correlation of the classification scheme is calculated as:

\[ R_S = \frac{\sum_{k=1}^{S} \sum_{i=1}^{M} R_i}{M} \]

3 Results and Discussion

Table 1 Panel A presents the mean correlations of individual stocks’ returns within different GICS classification systems for all stocks and the ones that have microblogging accounts. The results of all the stocks show a higher correlation when moving from two-digit GICS codes to six-digit GICS codes. This is consistent with the findings of Chan et al. (2007). However, the firms that use microblogging platforms reflect a different outcome — the GICS industry system (four-digit code) has the best performance. One possible reason of this difference is that only a small portion of firms are using microblogging services. It can be found that Sina Weibo firms have a lower return comovement in comparison to all the Chinese stocks.

We use K-means clustering models to classify firms with respect to their microblogging metric, the number of followers. K-means clustering is selected because it is simple to use yet still efficient. Furthermore, the clusters are non-hierarchical, and they do not overlap, which is the same as the GICS industry classification. The clusters are set between 60 – 70 groups to coordinate with the six-digit GICS sub-industries, which is the best industry classification scheme (Chan et al. 2007,
Furthermore, the play scheme of controlling the within-industry firms into narrower groups. Each four-digit industry was set to have three clusters to coordinate with the number of GICS 6-digits groups. In addition, we also exclude four-digit GICS groups that have less than three stocks. The small number of stocks in these groups may cause extreme correlation bias. Table 1 Panel B reports the within-group pairwise stock return of the K-means clustering models.

The first part of Panel B shows that the direct clustering groups, which dividing stocks without any industry, are less homogeneous than the GICS scheme. However, utilizing the microblogging metrics with the K-means clustering has improved the accuracy of the GICS scheme. The second part of Panel B shows that all the mean and minimum correlation are higher than the highest GICS scheme result (0.554).

Controlling firm’s market capitalization have improved microblogging metrics effect. It also should be noted that groups classified solely with the market capitalization have lower correlations than the ones classified involved microblogging metrics. This indicates that microblogging metrics play a more important role on identifying homogeneous stock groups than market capitalization. Furthermore, the results also show that the microblogging firms have a slightly lower pairwise correlation in the GICS 6-digits industries than in the GICS 4-digits industries. Our K-means clustering model results show a better than GICS 4-digits results.

Bollen et al. 2011). K-means clustering may produce different results each time it runs. Thus, to produce more reliable data, we run each model 10 times and use the mean of these tests as the final result.

Similar to Fama-French classification (Fama & French 1997), we also classify firms based on the four-digit GICS industries. In other words, we use the four-digit GICS industries classification scheme as the base and then further classify the within-industry firms into narrower groups. Each four-digit industry was set to have three clusters to coordinate with the number of GICS 6-digits groups. In addition, we also exclude four-digit GICS groups that have less than three stocks. The small number of stocks in these groups may cause extreme correlation bias. Table 1 Panel B reports the within-group pairwise stock return of the K-means clustering models.

The first part of Panel B shows that the direct clustering groups, which dividing stocks without any industry, are less homogeneous than the GICS scheme. However, utilizing the microblogging metrics with the K-means clustering has improved the accuracy of the GICS scheme. The second part of Panel B shows that all the mean and minimum correlation are higher than the highest GICS scheme result (0.554).

Controlling firm’s market capitalization have improved microblogging metrics effect. It also should be noted that groups classified solely with the market capitalization have lower correlations than the ones classified involved microblogging metrics. This indicates that microblogging metrics play a more important role on identifying homogeneous stock groups than market capitalization. Furthermore, the results also show that the microblogging firms have a slightly lower pairwise correlation in the GICS 6-digits industries than in the GICS 4-digits industries. Our K-means clustering model results show a better than GICS 4-digits results.

---

**Table 1: Average Pairwise Correlations between Individual Stocks’ Returns**

<table>
<thead>
<tr>
<th>Panel A: GICS Code Results</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-digit GICS Industry</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>Four-digit GICS</td>
<td>0.56</td>
<td>0.55</td>
</tr>
<tr>
<td>Six-digit GICS Sub-industry</td>
<td>0.57</td>
<td>0.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Direct clustering Based on GICS 4-digits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Fo</td>
</tr>
<tr>
<td>MC</td>
</tr>
<tr>
<td>Fo &amp; MC</td>
</tr>
</tbody>
</table>
4 Conclusion

This paper explores the question of whether firm-specific microblogging metrics can be used to identify the comovement of stocks. This paper makes several theoretical contributions. We found that social media metrics can help improve the performance of industry classification. The results of the model using the number of followers with market capitalization to classify the four-digit GICS industry groups into narrower groups demonstrate that the average pairwise within-group stock return correlation are higher than the six-digit GICS industries. This finding suggests that firm-specific metrics, such as the number of firms’ followers can be an anchor for identifying stock return comovement. In building the theory, this paper taps into the financial literature of stock return comovement. This extends the information system research and introduces a new approach for financial researchers. This study also has implications for financial analysts, investors, firm managers, and policy makers in the financial market. SEC and other regulators should welcome social media as a market risk indicator. Prior research has found stock return comovement is an identifier of market risk premium (Chelley-Steeley et al. 2013). Given the ease of monitoring and less latency period, regulators can view social media metrics as a new comovement tester and thus, as a market risk indicator.

Acknowledgement

This work was supported by the Fundamental Research Funds for the Central Universities, No. JBK120505; Ling Liu also acknowledges the support from the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References


