Quantifying the cross-sectional relationship between online sentiment and the skewness of stock returns

Dehua Shen a,b,*, Lanbiao Liu c, Yongjie Zhang a,d

a College of Management and Economics, Tianjin University, Tianjin, 300072, China
b China Center for Social Computing and Analytics, Tianjin University, Tianjin, 300072, China
c School of Finance, Nankai University, Tianjin, 300350, China
d Key Laboratory of Computation and Analytics of Complex Management Systems, Tianjin, 300072, China

HIGHLIGHTS

• Daily happiness sentiment (DHS) extracted from Twitter is investigated.
• The relationship between DHS and skewness of stock returns is investigated.
• Skewness of the Most-happiness subgroup is significantly larger than that of the Least-happiness subgroup.
• Skewness around the HHD is significantly larger than that of the LHD.

ABSTRACT

The constantly increasing utilization of social media as the alternative information channel, e.g., Twitter, provides us a unique opportunity to investigate the dynamics of the financial market. In this paper, we employ the daily happiness sentiment extracted from Twitter as the proxy for the online sentiment dynamics and investigate its association with the skewness of stock returns of 26 international stock market index returns. The empirical results show that: (1) by dividing the daily happiness sentiment into quintiles from the least to the most happiness days, the skewness of the Most-happiness subgroup is significantly larger than that of the Least-happiness subgroup. Besides, there exist significant differences in any pair of subgroups; (2) in an event study methodology, we further show that the skewness around the highest happiness days is significantly larger than the skewness around the lowest happiness days.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Over the past two decades, the landscape of information diffusion in stock market has been dramatically altered by the information and communication technology. The constantly increasing utilization of social media as the alternative information channel, e.g., Twitter, provides great opportunity for financial economists to broaden our horizons on the dynamics of the complex financial systems. For example, the discussion form of Twitter not only changes the “single-direction information diffusion mode” into “multi-sources information diffusion mode” in stock market, the user-generated content in Twitter from hundreds of millions of its users also illustrates the collective behavior in a previously inconceivable manner [1,2].

* Corresponding author at: College of Management and Economics, Tianjin University, Tianjin, 300072, China
E-mail address: dhs@tju.edu.cn (D. Shen).

http://dx.doi.org/10.1016/j.physa.2017.08.036
0378-4371/© 2017 Elsevier B.V. All rights reserved.
Recently, scholars begin to focus on the relationship between online activity and stock market performance and two distinct lines of literature stand out. The first line refers to the concern on the association between investors’ online activity, e.g., investors’ posting behavior in social message boards and searching behavior with the search engines, and the predictability of stock returns [3–15]. In particular, Bollen et al. [16] employ the OpinionFinder and Google-Profile of Mood States (GPOMS) to extract sentiment from Twitter feeds and find that the constructed sentiment can predict the daily directional changes in the closing prices of the DJIA with an accuracy of 86.7%. Da et al. [17] propose a proxy for investor attention using the search frequency in Google Trends and find that this novel proxy predict stock returns in the following 2 weeks. In a similar vein, Zhang et al. [18] also construct a proxy for investor attention with the search frequency of stock name in Baidu Index and show that this proxy can explain the abnormal stock returns. The second line of the literature mainly focuses on the impact of online activity on the dynamics properties of the stock markets [19–28]. Connecting with the information asymmetry theory, Blankespoor et al. [29] find that firm with the additional diffusion of firm-specific information via Twitter is related to greater abnormal depths and lower abnormal bid–ask spreads. Dimpfl and Jank [30] find a strong co-movement between Dow Jones’ realized volatility and the search frequency of its name in Google Trends. Shen et al. [31] retrieve the data from the Baidu News, exploit it as internet information flow and show that internet information flow can significantly reduce the volatility clustering.

However, to the best of our knowledge, no empirical studies investigated the relationship between online sentiment and the skewness of stock returns. The skewness reflects the degree of asymmetry of stock returns, with right (left) skewness representing the possibility of large holding gains (losses) [32]. Besides, skewness is also regarded as the reverse measurement for stock returns crash risk reflecting investors’ risk perception [33,34]. Therefore, skewness is important for a variety of academic and practical implications, e.g., portfolios management and option trading. In this paper, we give the first empirical study on the cross-sectional relationship between online sentiment and the skewness of stock returns by employing a novel daily happiness sentiment extracted from Twitter as well as focusing on 26 international stock index.

This paper is organized as follows. Section 2 describes the data and the model setup. Section 3 presents the empirical results and Section 4 performs the robustness check. Section 5 concludes.

2. Data and models

2.1. Data description

There are mainly two sources of data. The first refers to the daily happiness data, which is directly downloaded from the website: http://hedonometer.org/index.html. The Hedonometer employs the Amazon’s Mechanical Turk Service to scale 10,000 words contained in the sample of roughly 50 million of all messages posted in Twitter, resulting in 100 GB of JSON each day. The daily happiness covers the time range from 0:00 a.m. to 24:00 p.m. while trading periods of the stock markets span more than one day in one time zone. For example, the trading period of US stock market ranges from 9:30 a.m. to 16:00 p.m. in Eastern Standard Time (EST) and the trading period of Australian stock market ranges from 10:00 a.m. to 16:00 p.m. in Sydney Times. The time difference is 14 h. To reduce the seasonality effect, we deseasonalized the raw daily happiness data with the weekday dummies and keep the residuals as the daily happiness sentiment employed in the paper. There are 2451 trading days and 1751 calendar days, respectively, covering the calendar date from September 10th, 2008 to May 27th, 2015. Fig. 1 illustrate the daily happiness sentiment with the max value of 0.1948, the min value of −0.1317 and the standard deviation value of 0.0338.

The second source of data is the capital data from the Choice database (http://choice.eastmoney.com/Product/index.html) including the returns of 26 international index: Amsterdam Exchange index (AEX), Australian Exchange index (AS51), Austrian Traded Index (ATX), Belgium 20 Index (BFX), Dow Jones Industrial Average (DJIA), CAC 40 Index (FCHI), Financial Times Stock Exchange 100 Index (FTSE), DAX Index (GDAXI), Hang Seng Index (HSI), Nasdaq Index (IXIC), Jakarta Composite Index (JKSE), Kuala Lumpur Stock Exchange (KLSE), KOSPI Index (KS11), Milano Italia Borsa (MIB), Mexico IPC Index (MXX), Nikkei 225 Index (N225), NASDAQ 100 (NDX), NZX 50 Index (NZ50), Oslo Stock Exchange All Share Index (OSEAX), Philippine Stock Exchange Index (PSI), Russian Trading System Index (RTS), Bombay Stock Exchange Index (SENSEX), S&P 500 (SPX), Swiss Market Index (SSMI), Singapore Straits Times Index (STI) and Taiwan Weighted Index (TWII). The rationale why we focus on these stock market indices is two-fold: firstly, English is the most widely used language in the world and the focus of this manuscript is on the impact of investor sentiment on the skewness of stock returns. To calculate the value of skewness, we need more than one observation, therefore, we have to include other stock market indices into the sample. In particular, we are very cautious to select the stock market indices and the selected 26 indices are from the regions have a large number of Twitter users. In particular, we do not include the Chinese stock index for reason that the Twitter is forbidden in China. Secondly, even if some countries do not speak English, they have lots of cross-listed companies, i.e., the headquarters are located in their countries, but listed and traded in the US and UK. Therefore, the sentiment extracted from the English content may also contains some country-specific sentiment. Table 1 reports the statistical properties of these 26 international stock market index return. We can find that the US and Russian has the negative returns in the sample period and the US stock market has the smallest negative skewness and kurtosis.

---

1 For the report, see https://www.forbes.com/sites/victorlipman/2014/05/24/top-twitter-trends-what-countries-are-most-active-whos-most-popular/#450044036652. Though China has the largest population in the world, the Twitter user is few. Therefore, we exclude the Chinese stock market index from our sample.
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات