Physica A 451 (2016) 519–524

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Trading and non-trading period Internet information flow and intraday return volatility

Dehua Shen, Wei Zhang, Xiong Xiong *, Xiao Li, Yongjie Zhang

College of Management and Economics, Tianjin University, Tianjin, 300072, PR China
China Center for Social Computing and Analytics, Tianjin University, Tianjin, 300072, PR China

HIGHLIGHTS

- We employ Baidu News as a proxy for information flow.
- A positive impact of Internet information on volatility persistence is discovered.
- Trading period Internet information is a better proxy for information flow.
- The results provide evidences for the MDH.

ARTICLE INFO

Article history:
Received 7 June 2015
Received in revised form 21 November 2015
Available online 8 February 2016

Keywords:
Internet information
Mixture of Distribution Hypothesis
Trading and non-trading period information
Volatility persistence
GARCH

ABSTRACT

This paper employs the news appeared in Baidu News as the proxy for Internet information flow, separates them into trading period and non-trading period information and provides alternative evidence for the Mixture of Distribution Hypothesis (MDH). The empirical results show that the contemporary information can effectively reduce the volatility persistence; meanwhile, the lead information and the aggregate information also show some explanatory power. Some future directions are pointed out in the concluding remarks.

1. Introduction

The Mixture of Distribution Hypothesis (MDH) argues that the volatility of stock return is proportional to the rate of information flow at a given interval [1–4]. Although this is a reasonable hypothesis linking the information flow and the "stylized facts" of stock return. Empirical validation is difficult in quantifying the intangible form of information as well as capturing the volatility of return. Recent studies show evidence that the Autoregressive Conditional Heteroskedasticity (ARCH) model can capture well the variation in the volatility process [5,6]. In support of the argument of MDH, Lamoureux and Lastrapes [7] firstly introduce the daily trading volume as the proxy for information flow and show a significant explanatory power to the variance of daily returns. Wagner and Marsh [8] use the surprise volume (unexpected above-average trading activity) as the proxy for information flow and show that surprise volume can explain volatility persistence and excess kurtosis. Kalev et al. [9] employ the firm-specific announcements as the proxy for information flow and find that the inclusion of this variable in the conditional variance equation of the GARCH model reduces the volatility

* Corresponding author at: College of Management and Economics, Tianjin University, Tianjin, 300072, PR China. Tel.: +86 13920704381.
E-mail address: xxpeter@tju.edu.cn (X. Xiong).

http://dx.doi.org/10.1016/j.physa.2016.01.086
0378-4371/

persistence. Although the above-mentioned empirical studies strongly support the MDH, they are subject to criticism. Firstly, the trading volume-based proxies may suffer from the nonstationary properties and endogeneity. Secondly, firm-specific announcement is not sufficient, which only represents one aspect of the information flow.

In a seminal work by Zhang et al. [10], they employ the number of information appeared in Baidu News as the proxy for information flow, make a comprehensive comparison with the existing proxy and find that this Internet information-based proxy can reduce the volatility persistence of the SZSE SME PRICE INDEX. Our paper is a naturally extension of their work and contribute to the literature in two aspects. Firstly, we extract the information from Baidu News for individual stock, rather than focus on the stock index. Observing the explanatory for individual stock provides additional evidence for MDH. Secondly, given to the availability of the Internet information, we divide the information into trading period information and non-trading period information, include each of them into the conditional variance equation of the GARCH model and observe the decreases of the volatility persistence.

This paper is also in line to a broad literature on using Internet information for financial economics research [11–15]. Specifically, Da et al. [16] use the search frequency in Google Trends as the proxy for investor attention and find that this direct proxy can predict stock prices in the next 2 weeks and long-run underperformance of IPO. In a similar way, Zhang et al. [17] use the search frequency of stock name in Baidu Index as the proxy for investor attention and show that the quantified variable can predict the abnormal return and trading volume. Preis et al. [18] analyze changes in Google query volumes for search item related to finance and find patterns can be viewed as "early warning signs" of stock market movement. Moat et al. [19] also document changes in Wikipedia contain early signs of stock market movement.

The rest of the article is organized as follows. Section 2 describes the data. Section 3 presents the mode setup as well as the empirical results. Section 4 is the concluding remarks.

2. Data description

The dataset comprises daily returns for 10 actively traded stocks in CSI 300 Index,[1] which is the first index released by both Shenzhen Stock Exchange (SZSE) and Shanghai Stock Exchange (SSE).[2] The CSI 300 Index is revised on six months basis and thus generate the sample of stocks exist from the beginning date of CSI 300 Index to the end of July 2013. The rationale for using this sample is that the CSI 300 Index reflects the overall performance of Chinese stock market and the actively traded stocks are likely to have sufficient information flow to satisfy the precondition for MDH [1,2]. Daily stock returns are obtained from the WIND Financial Database. We use the opening price in the morning, closing price in the morning, opening price in the afternoon and closing price in the afternoon to calculate the stock return. The sample period is from 1 April 2010 to 1 July 2013.

The Internet information flow data is from the Baidu News, which is the production of Chinese largest search engine. Baidu News is the world's largest Chinese news platform. Since Baidu News does not provide us with the directly download service. We write a java script to automatically download the data. Given to the unique characteristic of Chinese stock market, we acquire the information in four different intervals for each trading days, i.e., D1 (from 15:00 p.m to 9:30 a.m.), D2 (from 9:30 a.m.to 11:30 a.m.), D3 (from 11:30 a.m. to 13: 00 p.m.) and D4 (from 13:00 p.m. to 15:00 p.m.). D1 represents the overnight Internet information flow. D2 represents the Internet information flow in the trading period in the morning. D3 represents the Internet information flow between the closing of the morning trading period and the beginning of afternoon trading period. D4 represents the Internet information flow in the trading period in the afternoon. The about-mentioned time is given in local time (GMT + 8). To control the time interval of D1 for each trading day, we only consider the week has five consecutive trading days, i.e. from Monday to Friday. For example, if Tuesday is the only non-trading day in a certain week, we exclude all the week from our sample. Because if Tuesday is missed, the information flow of D1 for Wednesday is not appropriately captured. For the information flow of D1 on each Monday, the search interval is also set from 15:00 p.m on Sunday to 9:30 a.m. on Monday. The period of Internet information flow is the same as the capital data.

Fig. 1 depicts the information appeared in Baidu News. We search the stock names and Baidu News gives the number of relevant news. For example, we search the stock name of "BaoTou Steel Union Company" in Chinese with Baidu News. It tells us that "Baidu News finds 143 relevant news for you" (the Chinese sentence in the blue ellipse). We use the number of relevant news as Internet information flow for individual stock. Baidu News provides us with "Advanced Setting" to get the number of news at a given interval. Therefore, we can get the intraday Internet information flow, i.e., the number of relevant news during D1, D2, D3 and D4 for each stock.

Table 1 shows the statistics properties of Baidu News in D1, D2, D3 and D4, respectively. The value in the table is the number of news provided by Baidu News when certain stock name is searched. The value in the Max (Min) column means the largest (smallest) number of news for the 10 stocks at a given interval. The value in the Std column is the average value of standard deviation of the 10 stocks at a given interval. We find that information in overnight period (D1) is

[1] These 10 stocks are: Ping An Bank Co. Ltd. (000001), Vanke Co. Ltd. (000002), Baoan Group Co. Ltd. (000009), China International Marine Cntnrs Gp Ltd. (000039), Shenzhen Agricultural Products Co. Ltd. (000061), Zoomlion Heavy Industry Sci. & Tch. Co. Ltd. (000157), Tangshan Jidong Cement Co. Ltd. (000401), XCMG Construction Machinery Co. Ltd. (000425), Guangxi Liugong Machinery Co. Ltd. (000528), Lu Zhou Lao Jiao Co. Ltd. (000568).

[2] We performed the Lagrange Multiplier (LM) test for their stock returns and they all showed ARCH effect.