Baidu news information flow and return volatility: Evidence for the Sequential Information Arrival Hypothesis

Dehua Shen[a,b], Xiao Li[c,*], Wei Zhang[a,d]

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

ARTICLE INFO

ABSTRACT

This paper employs Baidu News as the proxy for information flow and investigates competing hypotheses on the relationships between information flow and return volatility in Chinese stock market. The empirical results show that: (1) trading volume and return volatility are not driven by the same variable, i.e., the information flow, and thus contradicts the predication of the Mixture of Distribution Hypothesis (MDH); (2) there exist significant lead-lag relationships between information flow and return volatility, which is in accordance with the Sequential Information Arrival Hypothesis (SIAH); (3) these findings are robust to alternative measurement of return volatility and subsample analysis. Generally speaking, these findings contradict the prediction of MDH and support the SIAH.

1. Introduction

Return volatility can be the result of the rate of information flow, the reflection of the private information as well as the irrational behavior of the noise trader (Grossman and Stiglitz, 1980; Kyle, 1985; Glosten and Milgrom, 1985; French and Roll, 1986; Ross, 1989). Among the various explanations proposed, the rate of information flow has been widely considered as the major factor uncovering the underlying mechanism on the changes of asset prices (Andersen, 1996; Bergemann et al., 2015). Investigation of the relationships between information flow and return volatility is important. For example, discovering significant linear and nonlinear relationships between information flow and return volatility may be of interest to policy makers as they decide on the transparency and quality of information diffusion. Two competing hypotheses based on the rate of information flow explanation have been proposed: the Mixture of Distribution Hypothesis (MDH) and Sequential Information Arrival Hypothesis (SIAH).

However, the interactions and dynamics between the information flow and return volatility remain debatable for the reason that the intangible form of the information makes it hard to construct desired proxies for the information flow. Existing literature mainly relies on trading volume (Lamoureux and Lastrapes, 1990; Bohl and Henke, 2003; Le and Zurbruegg, 2010) and its adjusted forms (Wagner and

Marsh, 2005; Fleming et al., 2006; Park, 2010), and number of firm-specific announcements (Kalev et al., 2004) as the proxies for the information flow. They usually incorporate the constructed proxy into the conditional variance equation of the GARCH model (Bollerslev, 1986) and observe a significant reduction in the volatility persistence. In that sense, they provide evidence for the MDH, which claims that return volatility and trading volume are generated by a mixture of distribution, in which the rate of information flow is the stochastic mixing variable (Clark, 1973; Epps and Epps, 1976; Harris, 1987; Andersen, 1996). The alternative hypothesis for the information flow and return volatility relationships is the SIAH, which claims that the formation of new market equilibrium is not instantaneous and requires some time for investors to react to new information, producing the lead-lag relationships between information flow and return volatility (Copeland, 1976; Jennings et al., 1981; Smirlock and Starks, 1988; Darrat et al., 2003 and 2007).

Although both the MDH and SIAH hypotheses have experienced extensive investigation by examining the relationship between trading volume (is considered as the proxy for information flow) and return volatility, the overall empirical results are mixed (Darrat et al., 2003; Wagner and Marsh, 2005; Fleming et al., 2006; Park, 2010). This paper revisits this relationships between information flow and return volatility by employing a more appropriate proxy for information flow in Chinese stock market. This proxy is constructed

by acquiring information from the Internet. In particular, we use the search frequency of the stock names in Baidu News as the direct proxy for the information flow. Compared with the prevailing proxies for the information flow, e.g., trading volume, our proxy is more direct, which makes the rate of information flow tangible. With very few exceptions, Zhang et al. (2014) and Shen et al. (2016) also employ such proxy for information flow. But their studies focus exclusively on the explanatory power of the information flow on volatility clustering, while we aim to address the underlying hypotheses, i.e., MDH and SIAH, of the relationships between information flow and return volatility.

Our paper contributes to the existing literature in three aspects. Firstly, the information flow proxy is acquired from the Baidu News, which can naturally rule out other non-informational factors. Specifically, the prevailing proxy for information flow is trading volume. However, trading volume is not solely driven by information and it can also be driven by private information, irrational trading and liquidity shocks (French and Roll, 1986; Andersen, 1996; Shen et al., 2017). In that sense, trading volume is not a satisfied proxy for information flow. Besides, given to the high-frequency property of Baidu News, we can further divide the information flow into trading periods' information flow and non-trading periods' information flow and only the trading periods' information flow is considered as daily information flow employed in subsequent analysis. This further rules out the impact of second-hand information on the changes of asset prices (Davies and Canes, 1978). Secondly, this new proxy for information flow provides us with an alternative way to test the MDH by making comparisons between the contemporaneous correlation coefficient for information flow-return volatility (IF-RV) relationship and the contemporaneous correlation coefficient for information flow-trading volume (IF-TV) relationship. Existing literature mainly relies on the investigation of the contemporaneous correlation coefficients between trading volume (is considered as the proxy for information flow) and return volatility (Kalev et al., 2004; Wagner and Marsh, 2005; Fleming et al., 2006; Park, 2010). Employing the trading volume as the proxy for information flow seems to "test a theory that is about from inputs to outputs with an output measure" (Qiu and Welch, 2004). In fact, we find conflicting results when employing the Baidu News and trading volume as the proxy for information flow, respectively. Thirdly, the empirical investigation focusing on contemporaneous and lead-lag relationships between information flow and return volatility in Chinese stock market is seldom investigated and the only study is Lee and Rui (2000). Since 2000, both the investor structure and the trading rules have changed dramatically. For example, the Qualified Foreign Institutional Investors (QFII) are allowed to trade stocks in July 2003 and Share Price Index Futures is launched in April 2010. What's more, we also provide an alternative evidence on the relationships between information flow and return volatility in an emerging stock market, rather than focusing on the mature markets, e.g., the NYSE (Darrat et al., 2003, 2007), DAX, FTSE, CAC and TPX (Wagner and Marsh, 2005).

The remainder of this paper is organized as follows. Section 2 describes the Chinese stock market, the proxy for information flow and capital data. Section 3 gives the empirical methodology and Section 4 presents the main results based on correlation coefficients analysis and Granger causality test as well as the robustness. Section 5 concludes.

2. Data description

There are mainly two sources of data in this paper. The first refers to the capital data retrieved from RESSET Financial Research Database. The second is the data on the proxy for information flow, i.e., the Baidu News. A brief introduction of Chinese stock market is given to describe the trading periods.

2.1. Trading periods in Chinese stock market

The Chinese stock market composes of Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE), which were established on 19 December 1990 and 3 July 1991, respectively. Except for some national holidays (e.g., National Days, Mid-autumn Festival, Dragon Boat Festival and Lunar Chinese New Year), both the exchanges are open five days a week (from Monday to Friday) from 9:30 a.m. to 15:00 p.m. with a trading break from 11:30 a.m. to 13:00 p.m. in GMT +8 time zone. Therefore, the trading periods are from 9:30 a.m. to 11:30 and from 13:00 p.m. to 15:00 p.m. and there are 4 h for trading on each trading day. There are some evidence suggesting that Chinese stock market is significantly different from other developed stock markets, e.g., the New York stock exchange and the London stock exchange. Firstly, according to a report released by Shanghai stock exchanges, individual investors account for more than 90% of all the accounts in Chinese stock market (Zhang et al., 2016). Secondly, there exist price limits for the individual stocks where the prices cannot change more than 10% from the opening prices at each trading day. Thirdly, for the majority of the stocks, short sale is still constrained. On 31 March 2010, the China Security Regulatory Commission (CSRC) approved the margin trading and securities lending program to remove the restrictions on short selling for selected stocks. Up to the end of 2015, only 892 (less than one third) stocks are allowed to be sold short, but the naked short sale is still prohibited for all the stocks on both Shanghai and Shenzhen stock exchanges.

2.2. Proxy for information flow

The proxy for information flow is obtained from the Baidu News, which is a service provided by Baidu. According to a survey on searching behavior of Chinese netizen released by China Internet Network Information Center, Baidu dominates the Chinese searching market with a market share more than 80%. In that sense, Baidu News is an ideal source of the rate of information flow. Baidu News retrieves the news from more than 500 authoritative websites and provides 24/7 update service to its users. Besides, given the "Advanced Settings" provided by Baidu News, we can restrict the news to a certain interval. As a consequence, we can obtain the intraday information flow from Baidu News. For individual stock at a given trading day, we could confine the searching results to the trading periods (from 9:30 a.m. to 11:30 and from 13:00 p.m. to 15:00 p.m.). Table 1 reports the statistical properties of news in different intervals. In this paper, we search the stock names with Baidu News and employ the aggregated number of news appeared in trading periods as the daily information flow for individual stock. The rationale for employing this aggregated number of news is that if the stock market is efficient, the news in the non-trading periods should have been reflected into the prices. It is admitted that there may exists some overlapped news, i.e., the same news reported by different information sources, the proposed proxy has

Table 1
Statistical properties of news in T1, T2, T3 and T4.

This table reports the statistical properties of the news appeared in T1, T2, T3 and T4. "Std.", "Max" and "Min" denote the standard deviation, maximum value and minimum value, respectively. According to the trading rules in Chinese stock market, each trading days can be divided into four sub-periods, i.e., T1 (from 15:00 p.m. in previous trading day to 9:30 a.m.), T2 (from 9:30 a.m.to 11:30 a.m.), T3 (from 11:30 a.m. to 13:00 p.m.) and T4 (from 13:00 p.m. to 15:00 p.m.).

Variables	T1	T2	T3	T4
Mean	48.09	8.169	1.512	2.008
Median	8	2	0	0
Std.	122.9	22.28	2.704	3.649
Kurtosis	25.83	31.91	11.42	10.01
Max	819	162	15	19
Min	0	0	0	0