ARTICLE IN PRESS

Economic Modelling xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Economic Modelling

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

Daily happiness and stock returns: The case of Chinese company listed in the United States

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

[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] Key Laboratory of Computation and Analytics of Complex Management Systems, Tianjin 300072, China
[d] Carroll School of Management, Boston College, MA 02467, USA

ARTICLE INFO

Keywords:
Twitter
Daily happiness sentiment
Home bias
Cross-listed companies
Statistical analyses
Granger causality

ABSTRACT

Existing literature exclusively focuses on the association between local investor sentiment and local stock market performance. In this paper, we investigate the contemporaneous and the lead-lag relationship between local daily happiness sentiment extracted from Twitter and stock returns of cross-listed companies, i.e., the Chinese companies listed in the United States. The empirical results show that: 1) by respectively controlling for the firm capitalization, liquidity and volatility, there exists the largest skewness on the Most-happiness subgroup. (2) There exist bi-directional relationships between daily happiness sentiment and market variables, i.e., the stock return, range-based volatility and excess trading volume. (3) There are significantly positive stock returns, higher excess trading volume and higher range-based volatility around the daily happiness sentiment spike days. These findings not only suggest that there exists significant interdependence between online activities and stock market dynamics, but also provide evidence for the existence of "home bias".

1. Introduction

Over the past twenty years, the Internet has profoundly altered the landscape of information generation, dissemination and interaction. Various social media platforms provide us with unique datasets which are hard to obtain or unable to be recorded before the invention of the Internet (Da et al., 2011; Edelman, 2012; Jin et al., 2016; Zhang et al., 2016b). Among these, Twitter has more than 310 million monthly active users and stock market is among the hottest topics in Twitter.[1] There exist millions of tweets related to stock market, e.g., commentaries, recommendations and rumors, which are posted by individual investors, institutional investors, news agencies as well as regulators. Since the contents of the tweets are generated by hundreds of millions of its online users, it may reflect the collective opinion towards the stock market. Meanwhile, the "home bias" puzzle is a well-documented phenomenon in the US stock market (Coval and Moskowitz, 1999; Coval and Moskowitz, 2001; Ivković and Weisbenner, 2005; Ivković et al., 2008). For example, French and Poterba (1991) conclude that investors in the US are far more optimistic about domestic stocks than those foreign ones, thereby leading to the tendency to substantially overweight domestic stocks when constructing investment portfolios. Considering these two aspects, it is an empirical question to ask whether local investor sentiment extracted from Twitter will influence the prices of the non-local stocks, i.e., the Chinese companies listed in the United States.

In this paper, we investigate this issue by focusing on the Chinese companies listed in the United States and adopting an exogenous proxy for local investor sentiment extracted from 10% of all tweets using English language on a daily basis. As we all know that Twitter is forbidden in China, therefore the contents of the tweets cannot be generated by the potential Chinese investors. This unique mismatch provides us with a rare opportunity to investigate the impact of local investor sentiment on non-local companies, i.e., the Chinese companies listed in the United States. Therefore, we contribute to the existing literature in the following three aspects. To start with, unlike previous studies focusing on the relationship between local online sentiment and local market performance (Bollen et al., 2011; Siganos et al., 2014; Zhang et al., 2014; Zhang et al., 2016b), we are the first to investigate the impact of local investor sentiment on the cross-listed companies. In particular, the empirical results show that the local online sentiment has some impacts on the cross-listed companies, whose headquarters are not located in the United States. Therefore, we further provide a piece of evidence for the existence of "home bias" in the U.S stock market. Secondly, motivated by the empirical studies showing that

* Corresponding author at: College of Management and Economics, Tianjin University, Tianjin 300072, China.
[1] For the report, see: https://about.twitter.com/company.

http://dx.doi.org/10.1016/j.econmod.2017.03.002
Received 14 October 2016; Received in revised form 22 January 2017; Accepted 4 March 2017
0264-9993/

Please cite this article as: Li, X., Economic Modelling (2017), http://dx.doi.org/10.1016/j.econmod.2017.03.002

investor sentiment can affect the determinants of decision-making, e.g., risk perception and risk tolerance (Johnson and Tversky, 1983; Schwarz and Clore, 1983), we investigate the impact of the daily happiness sentiment on the distribution of stocks returns, i.e., the skewness. The reason is that skewness is often considered as the inverse measurement for market crash risk reflecting investors' risk perception (Hong and Stein, 2003; Bris et al., 2007; Kim et al., 2011). Thirdly, following the intuition that online sentiment spikes may contain important information affecting the stock market, we define the daily happiness sentiment spike (*dhs*-spike) days as days corresponding to the 5% (10%) of the highest daily happiness sentiment and observe the market performance around *dhs*-spike days. We find that there are significantly positive stock return, excess trading volume and higher range-based volatility around the *dhs*-spike days. These results have practical implications for constructing trading strategies based on the Twitter volume spike days as trading signals.

The remainder of this paper proceeds as follows. Section 2 illustrates the literature review on investor sentiment. Section 3 describes the daily happiness sentiment, capital data and variables definition. Section 4 presents the models and empirical results. Section 5 concludes.

2. Literature review

In behavior economics, sentiment refers to the extent of investors' expectations diverge from the norm, either manifested as excessive optimism or pessimism. Recent studies tell us that human's emotions, in addition to information, have profound impacts on individuals' behavior and expectations through distancing them from being rational. As a pioneer in this field, De Long et al. (1990) relate sentiment to noise traders' unpredictable expectations and ascribe such unpredictability to limited arbitrage. Also, the costs of arbitrage reduce the effectiveness of arbitrageurs' role in eliminating mispricing as argued by standard asset pricing model (Shleifer and Vishny, 1997). Barberis et al. (1998) explain the over-reaction and under-reaction through constructing a model to capture the sentiment relating to representativeness as well as conservatism bias. Similarly, Daniel et al. (1998) propose a model for sentiment resulting from overconfidence and self-attribution. Recently, Falato (2009) develops a dynamic model to test how happiness maintenance derived from consumption and financial wealth utility affects risk preferences. These theoretical models present a general framework for the interpretation of the empirical results.

The empirical side of studies on this topic mainly focuses on constructing different proxies for sentiment and three strands of literature stand out. The first strand uses market-based proxies, including the closed-end fund discounts (Lee et al., 1991; Swaminathan, 1996; Neal and Wheatley, 1998), NYSE share turnover rate (Baker and Stein, 2004), variables related to IPO (Ritter, 1991), share of equity (Baker and Wurgler, 2000) and dividend premium (Baker and Wurgler, 2007). Notably, Baker and Wurgler (2006) construct a composite index for investor sentiment based on the measurements employed in previous literature and explore how sentiment affect the cross-sectional stock returns. Specifically, they find a higher sensitivity to sentiment for small stocks, young stocks, stocks with high volatility, stocks with financial distress or less profitability, stocks with no dividend payments and stocks that have extreme growth rates. However, these market-based proxies suffer from the endogenous problem and may lead to relatively biased results. The second strand of studies uses exogenous survey-based data to represent investor sentiment (Solt and Statman 1988; Fisher and Statman, 2000; Lee et al., 2002; Brown and Cliff, 2004; Lemmon and Portniaguina, 2006; Schmeling, 2009; Lux, 2011). The critics on survey-based data refer to the low participating rates and biased incentives of respondents that may result in less representativeness of the survey (Zhang et al., 2016a). The third strand derives investor sentiment from the news contents, social media as well as exogenous events that influence individuals' mood (Tumarkin and Whitelaw 2001; Bollen et al., 2011; Siganos et al., 2014; Da et al., 2015). In particular, Tetlock (2007) relates sentiment extracted from Wall Street Journal to index returns and trading volumes. Interestingly, Edmans et al. (2007) observe a significant decline for stock returns after the loss of sports games. Similarly, Hirshleifer and Shumway (2003) regard morning sunshine as a signal for positive sentiment and relate cloudiness to lower daily stock returns. Lepori (2015) and Zhang et al. (2017) regard the TV program as the proxy for sentiment and show that sentiment is closely related to stock return and can also explain some "anomalies" in stock market. Taken together, existing literature exclusively focuses on the impact of local investor sentiment on local market performance, while the impact of local investor sentiment on the cross-listed companies is less investigated.

3. Data and variables

3.1. Data description

The happiness data is derived from the website (http://hedonometer.org/index.html), which is based on the results of the natural language processing technique on around 10% of all tweets using English language. In particular, in order to quantify the magnitude of happiness, the Amazon's Mechanical Turk service is employed to score the magnitude of happiness of selected words appeared in Twitter, e.g., laughter, love, successful, winning, healthy and celebration. For each individual word, the scale is set from 1 to 9, with 1 representing extremely negative, 5 neutral and 9 extremely positive. The sample period spans from September 10, 2008 to May 27, 2015 with 2451 (1751) calendar (trading) days.[2] Since the happiness data shows a significant weekday effect (the happiness increases during the end of the week) and holiday effect (the happiness spikes during national holidays, e.g., "Thanksgiving", "FIFA World Cup" and "Christmas Eve"),[3] in line with Da et al. (2015), we also deseasonalize the raw happiness data by applying the OLS model to regress the raw happiness data on the weekday dummies and holiday dummies and keep the residuals as the daily happiness sentiment used in the following sections (see Panel A of Fig. 1 for the illustration). This deseasonalized time series allows us to analyze the trends of daily happiness sentiment independently of the predictable seasonal patterns and reveals the intrinsic relationship. This sentiment proxy has the advantages of deriving from hundreds of millions potential investors as well as avoiding the endogenous problem of the market-based variables. Besides, we match the daily happiness sentiment with the Financial and Economic Attitudes Revealed by Search (FEARS) advocated by Da et al. (2015) and find a negative correlation coefficient of −0.0576 with p-value=0.0963. Therefore, our proxy is correlated to the FEARS.[4]

As our focus is the Chinese companies listed in the United States, the initial sample consists of 99 stocks in NASDAQ, 58 stocks in NYSE and 5 stocks in AMEX. The capital data corresponding to the daily happiness sentiment is retrieved from Wind Economic Database, including the opening price, highest price, lowest price, closing price, firm capitalization, turnover and trading volume for each stock. All of the capital variables are on daily basis. In particular, we employ two procedures to select the appropriate stocks to meet the aim of this study. We first exclude stocks that have less than 1000 trading days

[2] The following das are missed: "2015-02-28", "2015-02-27", "2015-02-26", "2015-02-12", "2015-02-07", "2015-02-06", "2015-02-05", "2012-04-22", "2009-12-18", "2009-12-19", "2009-12-20", "2009-08-03", "2009-08-04", "2009-08-05", "2009-05-19", "2009-05-18", "2009-05-17", "2009-05-16", "2009-05-15" and "2009-05-14".

[3] For brevity, we do not illustrate the raw happiness data. For the illustration, see: Zhang et al. (2016).

[4] Recently, Chan et al. (in press) raise a question on the validity of investor sentiment proxies.