

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Textual Analysis of Stock Market Prediction Using Financial News Articles and Google Search Queries

Tseng-Chung Tang⁺, Hsu-Tong Deng, Li-Chiu Chi

Dept. of Business Administration, National Formosa University, Yunlin, Taiwan

Dept. of Business Administration, Jinwen University of Science and Technology, New Taipei City, Taiwan

Dept. of Applied Foreign Languages, National Formosa University, Yunlin, Taiwan

ABSTRACT: To do a Google search query is the main feature of symbolic behavior for an investor on his early stage of choosing stocks. Also this mirrors the extent of an investor's concern and preference about specific investment decisions. As such, keyword search volumes can be used as early warning signs of stock market movements. The main purpose and contribution of this study is to employ the Google search volume index as a proxy for investor attention, thereby enabling the observation of price trends as well as the translation into profitable stock investment strategies. This study further investigates the characteristic of time distribution that investors spend on Google search queries under the effects of limited attention to predict market trends.

KEYWORDS: Google search queries; News; Text mining; Stock selection; Google search

I. INTRODUCTION

The aim of this study is to examine any possibility of the profitable stock investment strategies from the financial news article for firms publicly traded on the Taiwan Stock Exchange (TWSE) and the Taipei Exchange (TPEX). A large body of existing literature suggests that the financial news articles contain timely, accurate and profitable information (Nofsinger, 2001, Tetlock, 2007; Tetlock, Maytal, and Macskassy, 2008) and therefore generate and immediate excess returns (see, for example, Lee, 1986; Mitchell and Mulherin, 1994; Womack, 1996, 2012; Chan, 2003; Schumaker and Chen, 2009; Bollen, Mao, and Zeng, 2011). However, in contrast to the previous examinations, empirical results of Fromlet (2001) and Kutan and Yuan (2002) show that the financial news articles are not important information intermediaries. While the extant empirical evidence is conflicting and non-conclusive, this study sets out to reinvestigate the information content of the financial news coverage.

The remainder of this study is divided into the following four sections. In the Section 2, we lay out the sampling procedure and data. Section 3 describes research methods. Section 4 then reports the main findings. Finally, a brief conclusion follows.

II. SAMPLE SELECTION AND DATA COLLECTION

This study consists of the financial news articles of IPOs (initial public offerings) in Taiwan crawled from the Google News during the period between 2014 and 2015. Many financial news articles report not only one but multiple firms' information, even describing overall market movement or market sectors of stocks. Such coverage is excluded from the news articles of this study. In addition, this study further eliminates the sample firm which had inadequate data or faced regulatory enquires.

The historical daily high, low, opening, and closing prices, trade volume, and accounting variables for sample firms listed on the Taiwan Stock Exchange (TWSE) and the Taipei Exchange (TPEX), formerly known as Gre Tai Securities Market (GTSM) were derived from the Taiwan Economic Journal (TEJ) databank. The Formosa Stock Index is

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

selected as a market index for the present study. The Formosa Stock Index is created by converting the existing TAIEX and TPEX into one new benchmark index. The Formosa Stock Index can measure the aggregate performance of two local stock markets, namely the TWSE and the TPEX (Taipei Exchange, 2015).

III. EMPIRICAL METHODOLOGY

3.1 Methodology for measuring stock return behavior

Following the methodology employed by Tang and Chi (2014), we make use of the abnormal rate of return (AR, also known as excess return or alpha) as a proxy for the computation of stock price change in response to the financial news on coverage release date (t_0). Since the market-adjusted return model is applied to estimate the daily returns of the Formosa Stock Index, sample abnormal rate of return AR_{t_0} are analyzed for each sample by subtracting the predicted returns of the market-adjusted return model from its observed returns (see detailed in Chi, et al., 2012; Tang and Chi, 2014):

$$AR_{t_0}(\%) = \frac{P_{t_0} - P_{t-1}}{P_{t-1}} - \frac{FS_{t_0} - FS_{t-1}}{FS_{t-1}} \quad (1)$$

where $AR_{t_0}(\%)$, $\frac{P_{t_0} - P_{t-1}}{P_{t-1}}$, and $\frac{FS_{t_0} - FS_{t-1}}{FS_{t-1}}$ are the abnormal rate of return, daily returns, and daily returns of the Formosa Stock Index on day zero ($t=0$), respectively.

IV. EMPIRICAL RESULTS

In this section of the paper, we delineate the empirical evidence of the stock performance differences between the two different datasets this study illustrated in the previous subsection. This study observes that there is an increase of 55.91% in the trading volume on day t_0 , with a t -value of 2.969 and a p -value of 0.031. However, the result reveals that the volume traded dramatically decreases 31.31% on day $t+1$, with a t -value of 2.829 and a p -value of 0.025.

The evidence of this study substantiates that the mean and median abnormal rates of returns of our sample data are relatively different. The mean return peaks on day t_0 with an AR of 7.46% while the median return peaks on day $t-1$ with an AR of 0.54%. The average AR for the stock with higher search queries from trading day $t-1$ to trading day $t+1$ is 1.91%, 10.58%, and 0.95% respectively. The average AR for low-search-query stock from day $t-1$ to day $t+1$ is 1.19%, 2.41%, and 1.19% respectively. Collectively, it can be seen that the AR of the former is higher than that of the latter. This is in accordance with the assertion in Bank, Larch and Peter's (2011) research that Google search queries are related to a rise in trading activity and stock performance. Synthesizing our findings, we accordingly suggest that Google search queries are temporarily associated with the value effect.

V. CONCLUSIONS

The purpose and contribution of this study is to employ the Google search volume index as a proxy for investor attention, thereby enabling the observation of price trends as well as the translation into profitable stock investment strategies. The co-movement of the changes in trading volume and volatility and volume of search queries is found. Therefore, the results of this study are consistent with our previous postulate that the Google search query mirrors the extent of an investor's concern and preference about specific investment decisions. The findings are useful in adding to the body of literature describing the relationship between investor attention and trading activity.

REFERENCES

1. Bank, M., Larch, M., and Peter, G. (2011), Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management*, 25(3), 239-264.
2. Bollen, J., Mao, H., and Zeng, X. (2011), Twitter mood predicts the stock market, *Journal of Computational Science*, 2(1), 1-8.
3. Chan, W.S. (2003), Stock price reaction to news and non-news: drift and reversal after headlines, *Journal of Financial Economics*, 70(2), 223-260.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

4. Chi, L.C., Tang, T.C., and Chin, M. (2012), Corporate transparency as a defense against a stock price plunge: Evidence from a market-crash context. *Chiao Da Management Review*, 32, 137-162.
5. Fromlet, H. (2001), Behavioral finance-Theory and practical application, *Business Economics*, 36(3), 63-69.
6. Kutan, A.M. and Yuan, S. (2002), Does public information arrival matter in emerging markets?: Evidence from stock exchanges in China, Working paper, Department of Economics and Finance, Southern Illinois University.
7. Lee, C.W. (1986), Information content of financial column, *Journal of Economics and Business*, 38, 27-39.
8. Mitchell, M.L. and Mulheri, J.H. (1994), The impact of public information on the stock market, *Journal of Finance*, 49(3), 923-950.
9. Nofsinger, J.R. (2001), The impact of public information in investors, *Journal of Banking & Finance*, 25(7), 1339-1366.
10. Tetlock, P.C. (2007), Giving content to investor sentiment: the role of media in the stock market, *Journal of Finance*, 62(3), 1139-1168.
11. Tang, T.C. and Chi, L.C. (2014), See what you are searching for?. The proceeding of IAOS 2014 Conference on Official Statistics Meeting the demands of a changing world.
12. Tetlock, P.C., Maytal, S., and Macskassy, S. (2008), More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance*, 63, 1437-1467.
13. Schumaker, R.P. and Chen, H. (2009), Textual analysis of stock market prediction using breaking financial news: The AZFin text system, *ACM Transactions on Information Systems*, 27(2), 1-19.
14. Womack, K.L. (1996), Do brokerage analysts' recommendations have investment value? *Journal of Finance*, 51, 137-167.
15. Womack, K.L. (2012), Do Brokerage Analysts' Recommendations Have Investment Value? *The Journal of Finance*, 51(1), 137-167.