SOCIAL MEDIA AND ASSET PRICES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Di Wu

August 2017

SOCIAL MEDIA AND ASSET PRICES

Di Wu, Ph.D.

Cornell University 2017

This thesis investigates the effect of social media on asset prices. The three chapters in the thesis each target one aspect of the social media effect.

Chapter 1 and 2 look at social media and post-earnings-announcement drift in response to companies' quarterly earnings announcements. In Chapter 1, I attempt to build a theoretical model to estimate the price response that is caused by investors' attention through utilizing Bayesian learning. Using data from quarterly earnings, Twitter and StockTwits data (17 quarters from the fourth quarter of 2010 to the fourth quarter of 2014), I utilize Twitter volume and a residual methodology to generate an attention proxy that is orthogonal to the growth of Twitter accounts. In Chapter 2, I demonstrate how the new attention brought about by social media after the earnings announcements, positively affects the cumulative abnormal returns, resulting in magnitudes that are larger than the earnings surprise effects. Finally, I find that even companies reporting bad news can still have positive immediate cumulative abnormal returns if they attract enough attention from investors after an earnings announcement.

Chapter 3 examines social media effects from a practitioner's point of view. I develop a method for measuring profits for a pairs trading strategy which has previously been used by institutional and hedge fund investors. Building on Engelberg et al. (2009), who provides possible explanations for pairs trading profits, I identify social media as another driver of pairs trading profits. The work builds on a large body of literature that investigates the economic drivers of stock return co-movement.

BIOGRAPHICAL SKETCH

Di Wu was born on September 17, 1987 in Xinjiang, China. He is the only son of the family. He joined the Charles H. Dyson School of Applied Economics and Management at Cornell University as a Ph.D. student in 2012. Prior to joining in Cornell, he received a bachelor's in Applied Economics from Purdue University with the highest distinction in 2010 and a master's in Applied Economics and Management from Cornell University in 2012.

Wu works extensively on analysis of investors' behavior in financial markets. His research primarily addresses issues of financial market anomalies, information processing, and the interaction with social media. The concentration of his research falls in the areas of behavioral finance, empirical asset pricing, and risk management. His recent publication "Earnings Announcements in the Hospitality Industry: Do You Hear What I Say?" has featured in CHR/SAS webinar in February 2013.

To Mom and Dad,

who always encourage me to go on every adventure,

especially this one

ACKNOWLEDGMENTS

My deepest sense of gratitude goes to the chair of special committee, my advisor, Professor Vicki L. Bogan, who expertly guided me through my graduate education and who shared the excitement of five years of discovery. Her unwavering enthusiasm for economics kept me constantly engaged with my research, and her personal generosity helped make my time at Cornell enjoyable.

My appreciation also extends to my dissertation committee members, Professor Pamela C. Moulton and Professor Byoung-Hyoun Hwang for their early insights which launched the greater part of this dissertation.

Above ground, I am indebted to my family, whose value to me only grows with age.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

THE MODEL

Modern financial theories assume that the stock market is efficient and the stock price can reflect all available information in a timely and effective manner. However, a large number of empirical studies on financial markets find that many modern financial theories cannot explain the anomalies. The reason for this is that the financial models used in these studies generally assume that investors have unlimited cognitive resources and that they are always able to give enough attention to the financial market. This includes two assumptions: first, complete symmetric information; second, the ability of the individual to quickly and accurately understand the above information (Grossman and Stiglitz, 1980; Kyle, 1985). Obviously, the actual financial market participants' knowledge, skills, and attention is limited and cannot satisfy the above two conditions. A large amount of psychological literature indicates that the human brain's central cognitive processing capacity is limited (Pashler and Johnston, 1998). In the stock market, investors' limited attention is due to their limited time and energy. As a result, investors cannot consider all of the investment options. Limited attention could also restrict the amount of information investors can analyze (Aboody et al., 2008).

In order to explain these anomalies in financial markets, some economists use cognitive psychology insights to analyze the behavior of investors. Specifically, investors' attention is an area of interest. Investor attention is a psychological phenomenon that could have a significant impact on financial markets. Specifically, investors influence stock price and trading volume through investment behavior changes. These changes will cause stock price fluctuations in the short term. For individual and institutional investors, understanding investors' attention can effectively help make profits (Barber and Odean, 2008). Therefore, the events that attract investors to pay attention to the financial markets could produce abnormal trading behavior which leads to the stock price reaction and affects stock returns.

Previous theoretical studies show that investor attention plays an important role in trading behavior and the equity premium (Hirshleifer and Teoh, 2003; Hirshleifer et al., 2004; Peng and Xiong, 2006; Abel et al., 2007; Huang and Liu, 2007). Many recent empirical studies show that since retail investors rarely short stocks, news that commands their attention will on average lead to retail purchases and put positive pressure on stock prices (Barber and Odean, 2008; Da et al., 2001). Nonetheless, few theoretical studies directly explore how prices respond to attention-grabbing trading. The purpose of this study is to develop a theoretical model of how prices react to attention that is brought on by public events.

Several existing empirical studies have heterogeneous public information as a major factor in asset pricing. For example, Collins and Kothari (1989) claim that the response of the stock prices to earnings surprise variation is both cross-sectional and temporal. Malatesta and Thompson (1985) investigate stock market reactions to partially expected acquisition announcements. Hand et al. (1992) study announcements of rating changes by Moody's and Standard and Poor's, and Nayak and Prabhala (2001) document the stock split announcement effects. Conrad et al. (2002) investigate the price response and the sign of earnings surprises. News, in general, carries information relevant for asset pricing. Of course, the valuation of the news depends on the prior expectations of the market participants and how the prior expectations are built. In this regard, recent literature focuses on the importance of information reaching the market. For example, Veronesi (1999) uses a dynamic equilibrium model of asset prices to show that in equilibrium, investors' willingness to hedge against changes in their level of uncertainty makes them overreact to bad news and underreact to good news, making the price of the asset more sensitive to bad news.

Conrad et al. (2002) provide empirical evidence of these effects. They document that stock price reactions to earnings surprises depend on the overall market level. However, little is known about how investors react to news and how they build their prior beliefs to incorporate the essence of the news. According to Bayes' theorem, investors' posterior beliefs should be proportional to the likelihood times their prior beliefs. There are only a few attempts to examine price response utilizing Bayesian learning. For example, Lang (1991) utilizes Bayesian learning to argue that the magnitude of the stock price response to earnings surprises decreases over time due to investor uncertainty decreasing as they observe more earnings signals. Meanwhile, Krueger and Fortson (2003) use a Bayesian model to show that there was an increase in the sensitivity of market interest rates to the unemployment rate after the Bureau of Labor Statistics increased the size of the sample. In addition, Chen et al. (2004) model investors as Bayesian to test investor learning about the predictive ability of security analysts.

To our knowledge, no study has yet established a theoretical model for asset pricing that considers public news events, attention-driven traders and Bayesian learning. The goal of this paper is to fill this gap and to build a theoretical model to estimate the price response that is caused by news events by adapting Bayesian learning. Our model is built from the public event model established by Kim and Verrecchia (1991).

We construct our theoretical model by focusing on examining two questions. The first question is how prices actually respond to attention-grabbing trading brought on by a public news event. In particular, we examine whether the Bayesian learning method is appropriate for modelling investors' perceptions of news event and market participants' valuations. In addition, we investigate whether the consideration of such effects significantly contributes to asset pricing. The second question is whether 'good' or 'bad' news affects the price. Our results show that the attention created by a public news event positively affects the stock return regardless of the nature (positive or negative) of the news.

1. The model

In this section we introduce the assumptions underlying our model and our problem set-up. We define *public news events as any events, decisions, or occurrences that change the true value of the firm. e.g., quarterly earnings announcement*. We assume pure exchange and there are two periods. Trades occur in period 1. The economy has two assets—a risky asset and a riskless asset. The riskless interest rate is assumed to be zero.

Suppose that non-attention-driven traders have homogeneous information about a company's earnings, y, before the company release its earnings announcement. Let $g(y)$ denote these prior beliefs about y and assume that the beliefs are normally distributed, i.e., $g(y) \sim N(\mu_f, \frac{1}{\rho_f})$, where μ_f is the mean of analyst forecast of the company's earnings and ρ_f is the precision of this forecast defined as the inverse of the variance. The return of the risky asset is a random variable that contains *y*, denoted by $R = y + \delta$, $R \sim N\left(\bar{r}, \frac{1}{h}\right)$ where y is observable, and δ is unobservable. It is assumed that R is normally distributed with mean $\bar{r}$ and precision h. Both y and δ are random variables. There are two agents, those who observe y (non-attention-driven traders), and those who observe only price (attention-driven traders). In our simple model, all individuals are, *ex ante*, identical. The only difference between the two agents is whether they observe y. Therefore, the demands of non-attention-driven traders will depend on y and the price of the risky asset P. Attention-driven traders' demands will depend only on P.

In period 1, trader i, $i \in [0,1]$, is endowed with C_i cash and X_i risky asset. It is convenient to measure a $[0,1]$ continuum of traders since the sum of the traders are averages (Kim and Verrecchia, 1991). The results of the paper are not affected by assuming a countable, infinite, number of traders, i.e., $i = 1, 2, \cdots$. The aggregate risky endowment, denoted by $X = \int_0^1 X_i d_i$, $X \sim N\left(0, \frac{1}{t}\right)$, is normally distributed with mean 0 and precision t. Each individual trader does not observe the aggregate risky endowment. Assuming a nonzero mean of X does not affect the results.

We assume pure exchange in our system. There are a total of two time periods in our model. Trading activities occur in time period 1 and consumption occurs in time period 2. The noise portion of the risky asset captures the randomness of the markets due to the fact that securities markets are often subject to random demand and supply shocks such as changing liquidity needs, weather, and political situations (Grossman and Stiglitz 1980; Diamond and Verrecchia 1981). Suppose that a public signal (e.g., companies' earnings announcements) provides traders with a normally distributed Y_1.

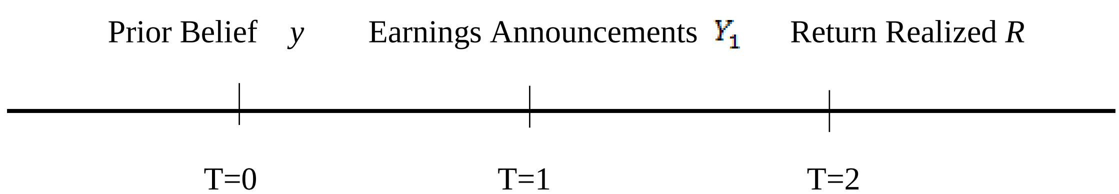


Figure 1. Model Timeline

Figure 1 presents the timeline of the model. Everyone observes Y_1 at time T=1, and only non-attention-driven traders observe *y*. We assume that the noisy estimate Y_1 is normally distributed with mean 0 and precision m:

$Y_1 = y + \tilde{\eta},\ where\ \tilde{\eta} \sim N\left(0, \frac{1}{m}\right)\ and\ E[y \cdot \tilde{\eta}] = 0$. Trader i also has his/her own attention-driven buying $\tilde{b}_i \sim N\left(Z, \frac{Z}{\varphi^2}\right)$ and selling $\tilde{s}_i \sim N\left(H, \frac{H}{\varphi^2}\right)$. We assume that the variances of attention-driven buying and selling are proportional to the means. In time period 2, attention-driven trader buying depends upon the attention generated by Y_1. In actual markets, we have non attention-driven trader activity that does not depend on attention. We set $E(\tilde{b}_i|Y_1) = z(M + Y_1^2)$, where $Y_1 = y + \tilde{\eta}$, $\tilde{\eta} \sim N\left(0, \frac{1}{m}\right)$ and M is a constant term; the variable z is the measure of the attention-driven trading intensity, where zY_1^2 is the expected attention-driven buying brought on by the earnings announcement. When $zM > 0$, trader i is a non-attention-driven trader and his/her attention is a constant term which does not depend on the public news event Y_1. We assume that attention-driven traders will be net buyers on good news (i.e., $Y_1 > 0$) or bad news (i.e., $Y_1 < 0$). Therefore, setting attention-driven buying in period 2 as proportional Y_1^2 captures our assumption and is consistent with the previous literature (i.e. Barber and Odean, 2008; Kyle, 1985). We set $E(\tilde{s}_i|Y_1) = z(M + kY_1^2 + (1-k)\vartheta^2)$, where $k, 0 \leq k < 1$, determines how much attention affects selling compared to buying. Finally, in time period $T = 1$ where trading activities happen, traders buy and sell securities at the competitive market prices. In this period, attention-grabbing traders trade based on the attention brought by the public signal Y_1. The asset return is realized in $T = 2$. We are particularly interested in the price P_2.

We assume that all random variables are mutually independent. Hence, the conditional probability density function of Y_1, given y, is $f(Y_1|y) \sim N(\mu_A, \rho_A)$. Recall that prior beliefs $g(y) \sim N(\mu_f, \frac{1}{\rho_f})$. It is assumed that all traders have of information of μ_f and ρ_f before the announcement. The actual announcement reveals both μ_A and ρ_A. Traders' posterior beliefs after observing the public announced signal Y_1. According to Bayes' rule:

$$g(y|Y_1) = \frac{f(Y_1|y)g(y)}{\int_{-\infty}^{+\infty} f(Y_1|y)g(y)dy'} \qquad (1)$$

Traders' posterior beliefs therefore are normally distributed with mean:

$\mu_P = E[y|Y_1] = \frac{\mu_f \rho_f}{\rho_f + \rho_A} + \frac{\mu_A \rho_A}{\rho_f + \rho_A}$ and precision $\rho_P = \frac{1}{Var[y|Y_1]} = \rho_f + \rho_A$

After the earnings announcement Y_1, the return of the risky asset is realized at time period 2. We assume the terminal value of the return is normally distributed $R = y + Y_1,\ R \sim N\left(\bar{r}, \frac{1}{h}\right)$. We assume that traders preferences can be represented by CARA[1] exponential utility functions $U_i(\widetilde{W_i}) = 1 - e^{-\frac{\widetilde{W_i}}{r_i}} = 1 - e^{-r_{Ai}\widetilde{W_i}}$, where for trader i, $\widetilde{W_i}$ is final wealth, r_i is risk tolerance and r_{Ai} is risk aversion. Trader i's final wealth $\widetilde{W_i}$ can be written as $\widetilde{W_i} = C_i + P_1 X_i + (P_2 - P_1)\widetilde{D_{1i}} + (R - P_2)\widetilde{D_{2i}}$, where P_1 and P_2 are the prices of the risky asset in time periods 1 and 2, respectively, and $\widetilde{D_{1i}}$ and $\widetilde{D_{2i}}$ are the quantity of the risky asset trader i holds for time periods 1 and 2, respectively.

Traders are heterogeneous in terms of risk aversion r_{Ai} and they differ because of their attention-driven intensity in period 1, $\tilde{Z}$. Thus, we model that some traders are attention-driven and others hold different expectations. Traders maximize their expected utilities conditional on the announcement, market prices and attention-driven intensity.

---

[1] To be consistent with Kim and Verrecchia (1991), we assume CARA utility.

Trader i's holding in period 2:

$$\max_{\widetilde{D_{2i}}} E\left[U_i\left(\widetilde{W_i}\right) \mid Y_1, y, \tilde{Z}, P_1, P_2\right]$$

$U_i\left(\widetilde{W_i}\right) = 1 - e^{-r_{Ai}\widetilde{W_i}}$, where $\widetilde{W_i} = C_i + P_1 X_i + (P_2 - P_1)\widetilde{D_{1i}} + (R - P_2)\widetilde{D_{2i}}$.

Hence, $U_i\left(\widetilde{W_i}\right) = 1 - e^{-r_{Ai}[C_i + P_1 X_i + (P_2 - P_1)\widetilde{D_{1i}} + (R - P_2)\widetilde{D_{2i}}]}$.

$$E[U_i\left(\widetilde{W_i}\right) \mid Y_1, y, \tilde{Z}_v, P_1, P_2] = E\left[1 - e^{-r_{Ai}[C_i + P_1 X_i + (P_2 - P_1)\widetilde{D_{1i}} + (R - P_2)\widetilde{D_{2i}}]} \mid Y_1, y, \tilde{Z}_v, P_1, P_2\right]$$

$$= 1 - E\left[e^{-r_{Ai}[C_i + P_1 X_i + (P_2 - P_1)\widetilde{D_{1i}} + (R - P_2)\widetilde{D_{2i}}]} \mid Y_1, y, \tilde{Z}_v, P_1, P_2\right],$$

where $E\left[e^{-r_{Ai}[C_i + P_1 X_i + (P_2 - P_1)\widetilde{D_{1i}} + (R - P_2)\widetilde{D_{2i}}]} \mid Y_1, y, \tilde{Z}, P_1, P_2\right]$ is the moment generating function for wealth. Suppose, for example, that the individual faces a continuous decision where wealth is distributed normally with mean μ_w and variance Var_w. Then, by looking at the moment generating function for a normal distribution, we find $E(e^{zt}) = e^{\mu t + \frac{t^2 Var}{2}}$. Thus, we have a closed form expression for the expected utility, $E[U_i(\widetilde{W_i})] = 1 - e^{-r_{Ai}[\mu_w - \frac{r_{Ai} Var_w}{2}]}$. One simple monotonic transformation of this expected utility function is $\mu_w - \frac{r_{Ai} Var_w}{2}$, which can be used as the objective function.

Thus, $\max_{\widetilde{D_{2i}}} E\left[U_i\left(\widetilde{W_i}\right) \mid Y_1, y, \tilde{Z}, P_1, P_2\right]$ is equivalent to $\max_{\widetilde{D_{2i}}} \mu_w - \frac{r_{Ai} Var_w}{2}$. We assume the return is normally distributed $R \sim N\left(\bar{r}, \frac{1}{h}\right)$. Let $\widetilde{D_{2i}}$ be trader i's risky asset demand in period 2. Then, we have $\widetilde{D_{2i}} = \frac{1}{r_{Ai}} h_i(\bar{r_i} - P_2)$. Trader i's holding is determined by the difference between his/her assessment of the return of the risky asset, $\bar{r}$, and the market price, P_2. The magnitude of his/her holding is determined by his/her risk aversion, r_{Ai}, and the precision of his/her information, h.

By equating the supply and the demand of the risky asset:

$$X = \widetilde{D_2} = \tilde{b}_i - \tilde{s}_i = \int \widetilde{D_{2\iota}}\, d_i = \int \frac{1}{r_{Ai}} h_i(\overline{r_\iota} - P_2) di$$

$$= \int \frac{1}{r_{Ai}} h_i \overline{r_\iota} d_i - \int \frac{1}{r_{Ai}} h_i P_2\, di$$

Rewriting the above:

$$P_2 = y - \frac{Z-H}{1+\frac{h}{L}\frac{(Z+H)}{\varphi^2}} + \frac{h^2\varphi^2(b-s)}{L\varphi^2+(Z+H)h} = \frac{A}{B} - \frac{1}{B}X \qquad (2)$$

where $A = \int \frac{1}{r_{Ai}} h_i \overline{r_\iota} d_i$, $B = \int \frac{1}{r_{Ai}} h_i d_i$, $L = \frac{1}{h\varphi}(z(2M + (1+k)Y_1^2 + (1-k)h)^{\frac{1}{2}}$.

Trader i's holding in period 1:

$$\max_{\widetilde{D_{1\iota}}} E[U_i(\widetilde{W_\iota})|\, y, \tilde{Z}, P_1]$$

where $\widetilde{D_{2\iota}} = \frac{1}{r_{Ai}} h_i(\overline{r_\iota} - P_2)$.

$$\widetilde{W_\iota} = C_i + P_1X_i + (P_2 - P_1)\widetilde{D_{1\iota}} + (R - P_2)\frac{1}{r_{Ai}} h(\bar{r} - P_2)$$

$$E[U_i(\widetilde{W_\iota})\ |y, \widetilde{Z_\iota}, P_1] = 1 - E\left[e^{-r_{Ai}[C_i + P_1X_i + (P_2-P_1)\widetilde{D_{1\iota}} + (R-P_2)\widetilde{D_{2\iota}}]} \mid Y_1, \tilde{Z}\right]$$

$$= 1 - E\left[e^{-r_{Ai}[C_i + P_1X_i + (P_2-P_1)\widetilde{D_{1\iota}} + (R-P_2)\frac{1}{r_{Ai}}h_i(\overline{r_\iota} - P_2)]} \mid Y_1, \tilde{Z}\right]$$

$$= 1 - E\left[e^{-r_{Ai}[C_i + P_1X_i + (P_2-P_1)\widetilde{D_{1\iota}}] - (R-P_2)h_i(\overline{r_\iota} - P_2)} \mid Y_1, \tilde{Z}\right]$$

and thus, $E\left[e^{-(R-P_2)h(\overline{r_\iota} - P_2)}\right] = e^{-(\overline{r_\iota} - P_2)h_i(\overline{r_\iota} - P_2) + \frac{h}{2}(\overline{r_\iota} - P_2)} = e^{-\frac{h_i}{2}(r_i - P_2)^2}$

then, $E[U_i(\widetilde{W_\iota})|\, Y_1, \tilde{Z}, X] = 1 - E[e^{-r_{Ai}[C_i + P_1X_i + (P_2-P_1)\widetilde{D_{1\iota}}] - \frac{h_i}{2}(\overline{r_\iota} - P_2)^2}]$

$$= 1 - E[e^{-r_{Ai}[C_i + (\frac{A}{B} - \frac{1}{B}X)X_i + (\frac{A}{B} - \frac{1}{B}X - P_1)\widetilde{D_{1\iota}}] - \frac{h_i}{2}(\overline{r_\iota} - \frac{A}{B} - \frac{1}{B}X)^2}$$

The only random variable in this expression given $Y_1, \widetilde{Z}_v, X$ is P_1.

$$W_i = [C_i + (\frac{A}{B} - \frac{1}{B}X)X_i + \left(\frac{A}{B} - \frac{1}{B}X - P_1\right)\widetilde{D_{1\iota}}] - \frac{h_i}{2}\left(\overline{r_\iota} - \frac{A}{B} - \frac{1}{B}X\right)^2$$

$$\mu_w = [C_i + (\frac{A}{B} - \frac{1}{B}X)X_i + \left(\frac{A}{B} - \frac{1}{B}X - \mu_{P_1}\right)\widetilde{D_{1\iota}}] - \frac{h_i}{2}\left(\overline{r_\iota} - \frac{A}{B} - \frac{1}{B}X\right)^2$$

$$Var_w = \widetilde{D_{1\iota}}^2 Var_{P_1}$$

By Taylor Series:

$$E[U_i(\widetilde{W_i}) |\, Y_1, \tilde{Z}, X] \approx U_i(\mu_w) + \frac{U_i''(\mu_w)}{2} Var_w$$

$$= 1 - e^{-r_{Ai}\mu_w} - \frac{r_{Ai}^2 e^{-r_{Ai}\mu_w}}{2} Var_w$$

$$= 1 - \frac{2 - r_{Ai}^2 \widetilde{D_{1\iota}}^2 Var_{P_1}}{2} e^{-r_{Ai}\{[C_i + (\frac{A}{B} - \frac{1}{B}X)X_i + \left(\frac{A}{B} - \frac{1}{B}X - \mu_{P_1}\right)\widetilde{D_{1\iota}}] - \frac{h_i}{2}\left(\overline{r_\iota} - \frac{A}{B} - \frac{1}{B}X\right)^2\}}$$

So the problem:

$$\max_{\widetilde{D_{1\iota}}} E\left[U_i(\widetilde{W_i}) \middle|\, Y_1, \tilde{Z}, P_1\right]$$

The first order condition for this problem is given by

$$r_{Ai} e^{-r_{Ai}\mu_w} \left(\frac{A}{B} - \frac{1}{B}X - \mu_{P_1}\right) - r_{Ai}^2 Var_{P_1} \widetilde{D_{1\iota}} e^{-r_{Ai}\mu_w} + \frac{r_{Ai}^3 e^{-r_{Ai}\mu_w}}{2} \widetilde{D_{1\iota}}^2 Var_{P_1} \left(\frac{A}{B} - \frac{1}{B}X - \mu_{P_1}\right) = 0$$

Rewriting the above: $\frac{r_{Ai}^3 Var_{P_1} H}{2} \widetilde{D_{1\iota}}^2 - r_{Ai} Var_{P_1} \widetilde{D_{1\iota}} + H = 0$, where

$H = \frac{A}{B} - \frac{1}{B}X - \mu_{P_1}$.

$$\widetilde{D_{1\iota}} = \frac{1}{r_{Ai}^2 (\frac{A}{B} - \frac{1}{B}X - \mu_{P_1})} + \frac{\left(1 - \frac{2r_{Ai}}{Var_{P_1}}\right)^{1/2}}{r_{Ai}^2 (\frac{A}{B} - \frac{1}{B}X - \mu_{P_1})}$$

Since each trader observes the prices for risky securities in both trading periods, P_1 and P_2, traders have their own conjectures about the prices based on what information they have (Kim and Verrecchia, 1991). Thus, P_1 and P_2 can be written as:

$$P_1 = \alpha_1 R + \theta_1 Y_1 + \beta_1 y - \Upsilon_1 X \qquad (3)$$

Similarly,

$$P_2 = \alpha_2 R + \theta_{21} Y_1 + \beta_{21} \tilde{Z} + \beta_2 y - \Upsilon_2 X \qquad (4)$$

Since, $Y_1 = y + \tilde{\eta}$, $\tilde{Z} = \frac{h^2\varphi^2 + L\varphi^2(Z-H)}{L\varphi^2 + (Z+H)h}$, equation (3) can be written as :

$$P_1 = \alpha_1 R + \theta_1(y + \tilde{\eta}) + \beta_1 y - \Upsilon_1 X = (\alpha_1 R + \theta_1 \tilde{\eta} - \Upsilon_1 X) + (\theta_1 + \beta_1) y = v \cdot \mu_f$$

$$\mu_{P_1} = (\alpha_1 R - \Upsilon_1 X) + (\theta_1 + \beta_1) y$$

$$Var_{P_1} = \frac{\theta_1^2}{m}$$

equation (3) can be written as:

$$P_2 = \alpha_2 R + \theta_{21} Y_1 + \beta_2 y + \beta_{21} \tilde{Z} - \Upsilon_2 X = v \cdot \mu_P$$

$$\mu_{P_2} = (\alpha_2 R - \Upsilon_2 X) + (\theta_{21} + \theta_2 + \beta_2) y$$

$$Var_{P_2} = \frac{\theta_{21}^2}{m} + \frac{\theta_2^2}{n}$$

In equilibrium, equation (2) and (3) are identical:

$P_2 = \alpha_2 R + \theta_{21} Y_1 + \beta_2 y + \beta_{21} \tilde{Z} - \Upsilon_2 X = \frac{A}{B} - \frac{1}{B} X$ and thus:

$\alpha_2 R + \theta_{21} Y_1 + \beta_2 y + \beta_{21} \tilde{Z} = \frac{A}{B}$ and $\Upsilon_2 = \frac{1}{B}$

The prices P_1 and P_2 are linear functions of the mean of the public signals and the noise.

$$P = \begin{cases} v \cdot \mu_f, \ before\ the\ announcement \\ v \cdot \mu_p, \ after\ the\ announcement \end{cases}$$

Proposition 1.

The return after the public event is proportional to the average posterior beliefs of traders and unanticipated information in the event plus attention brought by the public event.

That is:

$R = f($Average posterior beliefs, Unanticipated information, Attention)

From equation (3) and (4) the price change due to the public information event Y_1 is:

$$\Delta P = P_2 - P_1 = (\alpha_2 - \alpha_1)R + (\theta_{21} + \theta_2 + \beta_2 - \theta_1 - \beta_1)y + \beta_{21}\tilde{Z} + (\theta_{21} - \theta_1)\tilde{\eta} - (\Upsilon_2 - \Upsilon_1)X$$

$$= v \ (\mu_p - \mu_f) \frac{\rho_A}{\rho_f + \rho_A} \qquad (5)$$

We rearrange equation (5):

$$R = v \ (\mu_p - \mu_f) \frac{\rho_A}{\rho_f + \rho_A} + \frac{\theta_{21} + \theta_2 + \beta_2 - \theta_1 - \beta_1}{\alpha_1 - \alpha_2} y + \frac{\theta_{21} - \theta_1}{\alpha_1 - \alpha_2} \tilde{\eta} - \frac{\Upsilon_2 - \Upsilon_1}{\alpha_1 - \alpha_2} X + \frac{\beta_{21}}{\alpha_1 - \alpha_2} \tilde{Z} \qquad (6)$$

Proposition 2.

The attention effect on the returns is larger than the news effect (e.g., earnings surprises).

Proof:

From equation (6),

$$R = v \ (\mu_p - \mu_f) \frac{\rho_A}{\rho_f + \rho_A} + \frac{\theta_{21} + \theta_2 + \beta_2 - \theta_1 - \beta_1}{\alpha_1 - \alpha_2} y + \frac{\theta_{21} - \theta_1}{\alpha_1 - \alpha_2} \tilde{\eta} - \frac{\Upsilon_2 - \Upsilon_1}{\alpha_1 - \alpha_2} X + \frac{\beta_{21}}{\alpha_1 - \alpha_2} \tilde{Z}$$

where $\tilde{Z} = \frac{h^2\varphi^2 + L\varphi^2(Z-H)}{L\varphi^2 + (Z+H)h}$; $L = \frac{1}{h\varphi}(z(2M + (1+k)Y_1^2 + (1-k)h)^{\frac{1}{2}}$; and $\mu_p = E[y|Y_1] = \frac{\mu_f \rho_f}{\rho_f + \rho_A} + \frac{\mu_A \rho_A}{\rho_f + \rho_A}$

What we want to show is $\frac{\beta_{21}}{\alpha_1 - \alpha_2}\tilde{Z} - \frac{v}{\alpha_2 - \alpha_1}(\mu_p - \mu_f)\frac{\rho_A}{\rho_f + \rho_A} > 0$

$$\frac{\beta_{21}}{\alpha_1 - \alpha_2}\tilde{Z} - \frac{v}{\alpha_2 - \alpha_1}(\mu_p - \mu_f)\frac{\rho_A}{\rho_f + \rho_A} = \frac{\beta_{21}}{\alpha_1 - \alpha_2}\frac{h^2\varphi^2 + L\varphi^2(Z-H)}{L\varphi^2 + (Z+H)h} - \frac{v}{\alpha_2 - \alpha_1}(\mu_p - \mu_f)$$

$$= \frac{\beta_{21}}{\alpha_1 - \alpha_2} * \frac{h^2\varphi^2 + \frac{\varphi^2(Z-H)}{h\varphi}(z(2M+(1+k)Y_1^2+(1-k)h)^{\frac{1}{2}}}{\frac{\varphi^2}{h\varphi}(z(2M+(1+k)Y_1^2+(1-k)h)^{\frac{1}{2}} + (Z+H)h} - v\ (\mu_p - \mu_f)\frac{\rho_A}{\rho_f + \rho_A}$$

$$= \frac{(\rho_f + \rho_A)\beta_{21}h^2\varphi^2 + (\rho_f + \rho_A)\beta_{21}\frac{\varphi^2(Z-H)}{h\varphi}(z(2M+(1+k)Y_1^2+(1-k)h)^{\frac{1}{2}} + v\rho_A(\mu_f - \mu_p)(\frac{\varphi^2}{h\varphi}(z(2M+(1+k)Y_1^2+(1\cdot}{(\rho_f + \rho_A)(\frac{\varphi^2}{h\varphi}(z(2M+(1+k)Y_1^2+(1-k)h)^{\frac{1}{2}} + (Z+H)h)}$$

> 0

The difference between the attention effect and the news effect is positive, which is what we wished to show.

2. Empirical Strategies

The detailed empirical study for the model above is presented in Chapter 2. This section introduces the empirical measures to test the model. From equation (6), we get the stock return $R = f($Average posterior beliefs, Unanticipated information, Attention). The public news event we choose to estimate using the theoretical model is the companies' quarterly earnings announcement. The impact of earnings announcements on stock prices has been examined by many studies (Ball and Brown, 1968; Beaver, 1968; Abarbenell and Bernard, 1992; Barberis et al., 1998; Daniel et al., 1998; Hong and Stein, 2000; Bernard and Thomas, 1990; Foster, 1981; and Han and Wild, 1990). Furthermore, a few recent studies investigate the relationship between investors' attention and earnings announcements (Hirshleifer et al., 2009; Chakrabarty et al., 2015; Chakrabarty and Moulton, 2012; Von Bommel, 2003; Wysocki, 1998; Antweiler and Frank, 2004; Dewally, 2003; and Sabherwal et al., 2008).

For our empirical analysis, we adapt the strategies used in most of the post-earnings-announcement drift literature and utilize cumulative abnormal returns. Abnormal returns during the earnings announcement period are computed as the difference between the return of the sample firm and the return on the index. Similar to Foster et al. (1984) and Bernard and Thomas (1989), the estimates of cumulative abnormal stock returns (CAR) for announcing and nonannoucing firms at earnings release dates are:

$$\mathrm{CAR}_{j,\{t_1,t_2\}} = \sum_{t_1}^{t_2} u_{j,t}$$

Where $u_{j,t} = R_{j,t} - (\partial_0 + \partial_1 * R_{M,t})$ is daily abnormal stock return for firm *j* on day *t*, and $\{t_1, t_2\}$ is the period from day t_1 to day t_2 inclusive. $R_{j,t}$ is the daily stock return for firm *j* on day *t*, and $R_{M,t}$ is the return on the CRSP value-weighted index for day *t*. The parameters, ∂_0 and ∂_1 are estimated using OLS.

Another component in stock return function from equation (6) is the unanticipated information after the public event. We utilize the unexpected earnings as the proxy for this component. Unexpected earnings are calculated by subtracting analyst forecast earnings per share from actual earnings per share deflated by the absolute value of forecast. Unexpected earning (UE) is defined as the difference between reported earnings and expected earnings:

$$UE_{j,t,q} = (AE_{j,t,q} - EE_{j,t,q})/|EE_{j,t,q}|$$

Where $AE_{j,t,q}$ is the actual earnings for firm *j* in quarter *q* of year *t*, and $EE_{j,t,q}$ is the analyst's forecast expected earnings for firm *j* in quarter *q* of year *t*.

Both direct and indirect investors' attention proxies have been used in the recent studies. Google's search Volume Index (SVI) is considered a direct measure of investors' attention (e.g., Da et al., 2011; DeHaan et al., 2015; Chakrabarty et al., 2015). Twitter volume has also been used as a direct attention proxy in a few recent studies (e.g., Bhagwat et al., 2014; Curtis et al., 2014). Several recent studies use different indirect attention proxies: for example, extreme returns (Barber and Odean, 2008); trading volume (Barber and Odean, 2008; Gervais et al., 2001; and Hou et al., 2008) and news and headlines (Barber and Odean, 2008; and Yuan, 2008). Moreover, advertising expense (Chemmanur and Yan, 2009; Grullon et al., 2004; and Lou, 2008) and price limits (Seasholes and Wu, 2007) are also used as indirect proxies for investors' attention. In our empirical analysis, we utilize Twitter volume as a direct measure of investors' attention. The attention component in the model equation (6) is the added attention from the public event. In order to estimate the new attention brought by the earnings announcements, we focus on the Twitter volume change one day after the earnings announcements.

3. Conclusion

This chapter theoretically analyzes the impact of attention on the stock prices. In theoretical models, trading by not fully rational attention-driven investors can put positive pressure on prices. Informed investors cannot eliminate mispricing due to limits of arbitrage. Attention-driven investors could affect asset prices by their trading behavior. When they are actively buying (selling) after a public event, it will put positive (negative) pressure on asset prices. Eventually, asset prices are likely to be brought towards their fundamental values. Most empirical studies of price response to earning news focus exclusively on market reaction instead of investor behavior. However, Bayesian learning claims that the price reaction is driven by the investors' beliefs before and after an event. Few studies theoretically explore the Bayesian learning models in the context of public news event and investor attention. The main purpose of the paper is to fill this gap in the literature. Focusing on the theoretical framework, we are able to model the change in price before and after an event using the average change in investors' beliefs and investors' attention. The results in this study are consistent with existing empirical findings. The main result of this paper is that, the attention-driven investor beliefs could positively affect stock prices. However, the sign of the public news event does not affect the price response to the attention-driven trading.

REFERENCES

Abarbanell, J., and V. Bernard. "Test of Analysts' Overreaction/Underreaction to Earnings Inforemation as an Explanation for Anomalous Stock Price Behavior."Journal of Finance 47(1992):1181-1207

Aboody, David, John Hughes and Jing Liu. 2005. Earnings quality, insider trading and cost of capital. Journal of Accounting Research, 43 (5): 651-673.

Antweiler, Werner, and Frank, Murray Z., 2004, Is all that talk just noise? The information content of internet stock message boards, Journal of Finance 59, 1259-1294.

Ball, R. and P. Brown. 1986. An Empirical Evaluation of Accounting Income Numbers. Journal of Accounting Research 6 (2):159-178.

Barber, Brad M., and Terrance Odean. 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies, 21(2), 785–818.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2006, Do noise traders move markets? working paper, University of California - Davis.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Beaver, William H. 1986. The Information Content of Annual Earnings Announcements. Journal of Accounting Research. 6, 67-92.

Bhagwat, Vineet, and Timothy R. Burch. 2014. Pump it Up? Tweeting to Manage Investor Attention to Earnings News. Working paper. University of Oregon and University of Miami

Bommel Van, Jos, 2003, Rumors, The Journal of Finance 58, 1499–1520.

Chemmanur, Thomas and An Yan. 2009. Product Market Advertising and New Equity Issues. Journal of Financial Economics. 92 (2) 40-65.

Chen, K., Chen, Z., and Wei, K., 2009. Legal protection of investors, corporate governance, and the cost of equity capital. Journal of Corporate Finance, 15: 273–289

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, The Quarterly Journal of Economics, 108, 905-939.

Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, Order imbalance and individual stock returns: theory and evidence, Journal of Financial Economics, 72, 485-518.

Collins, D.W., and S. P. Kothari. "An Analysis of the Intertemporal and Cross-Sectional Determinants of Earnings Response Coefficients." Journal of Accounting and Economics, 11 (1989), 143–181.

Cohen, Lauren, and Andrea Frazzini, 2006, Economic links and predictable returns, working paper, Yale University.

Conrad, J.; B. Cornell; and W. R. Landsman. "When Is Bad News Really Bad News?" Journal of Finance, 57 (2002), 2507–2533.

Curtis, A., V. J. Richardson, and R. Schmardebech. 2014. Investor Attention and the Pricing of Earnings News. Working paper, University of Washington, University of Arkansas, and University of Missouri.

Daniel, S., Levine, C., and Sivaramakrishnan, K., 1998. Earnings predictability and bias in analysts' earnings forecasts. The Accounting Review 73, 277–294.
DeHaan, Ed, Terry Shevlin, and Jacob Thornock, 2015, Market (in)attention and the strategic scheduling and timing of earnings announcements, Journal of Accounting and Economics 60: 36-55.
Dewally, Michael, 2003, Internet investment advice: Investing with a rock of salt, Financial Analysts Journal 59, 65-77.
DellaVigna, Stefano, and Joshua Pollet, 2006, Investor inattention, and Friday earnings announcements, working paper, University of California - Berkeley.
Diamond, Douglas and Robert E. Verrecchia, 1981. Information Aggregation in a Noisy Rational Expectations Economy. Journal of Financial Economics. 9(3): 221-235.
Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin, 2001, The high-volume return premium, Journal of Finance 56, 877-919. Grinblatt, Mark, and Matti Keloharju, 2001, What makes investors trade? Journal of Finance, 56, 589-616.
Grossman, Sanford J. and Joseph E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets. American Economics Reivew. 70(3): 393-408.
Grullon, Gustavo, George Kanatas, and James P. Weston. 2004. Advertising, Breadth of Ownership, and Liquidity. Review of Financial Studies. 17(2):439-461.
Hand, J.; R. Holthausen; and R. Leftwich. "The Effect of Bond Rating Agency Announcements on Bond and Stock Prices." Journal of Finance, 47 (1992), 733–752.
Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets? Journal of Accounting and Economics, 38, 297-331.
Hirshleifer, David, Seongyeon Lim, and Siew Hong Teoh, 2004, Disclosure to an audience with limited attention, working paper, Ohio State University.
Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, Journal of Accounting and Economics, 36, 337-386.
Hong, Harrison, and Valkanov Stein, 2000, Do industries lead the stock market? Journal of Financial Economics, 83, 367-396.
Hou, Kewei, Lin Peng, and Wei Xiong, 2008, A Tail of two anomalies: The implications of investor attention for price and earnings momentum, working paper, Ohio State University.
Kim, Oliver and Robert E. Verrecchia. 1991. Trading Volume and Price Reactions to Public Announcements. Journal of Accounting Research. 29(2): 302-321.
Krueger, Alan B. and Kenneth N. Fortson, 2003. Do Markets Respond More to More Reliable Labor Market Data? A Test of Market Rationality. Journal of the European Economic Association. 1(4), 931-957.
Kyle, Albert S. 1985. Continuous Auctions and Insider Trading. Econometrica. 54(6): 1315-1336.
Lang, Larry H.P. 1991. Insider Trading around Dividend Announcements: Theory and Evidence. Journal of Finance. 46(4): 1361-1389.
Lou, Xiaoxia. 2008. The Divergence of Liquidity Commonality in the Cross-Section

of Stocks. Journal of Financial Economics. 89(3): 444-466.
Lee, Charles M. C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, Journal of Financial Markets, 3, 83-111.
Lee, Charles M. C., and Mark J. Ready, 1991, Inferring investor behavior: evidence from TORQ data, Journal of Finance, 46, 733-746
Malatesta, Paul and Rex Thompson. 1985. Partially anticipated events: A model of stock price reactions with an application to corporate acquisitions. Journal of Financial Economics. 14(2): 237-250.
Nayak, S., and N. R. Prabhala. "FYI: The Interest Rate Sensitivity of Stock Prices." Review of Financial Studies, 14 (2001), 1083–1116.
Pashler, Horald, and James C. Johnson, 1998, Attentional limitations in dual-task performance, Psychology Press, Hove, UK. Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, Journal of Financial Economics, 80, 563-602.
Peng, Lin and Wei Xiong. 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics, 80, 563–602.
Sabherwal, Sanjiv, Sarkar, Salil K., and Zhang, Ying, 2008, Online talk: does it matter?, Managerial Finance 34, 423-436.
Seasholes, Mark, and Guojun Wu, 2007, Predictable behavior, profits, and attention, Journal of Empirical Finance, 14, 590-610.
Wysocki, Peter D. 1998. Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards. Working paper. University of Michigan.

CHAPTER 2

DOES SOCIAL MEDIA GET YOUR ATTENTION?

1. Introduction

With the growth of social media in recent years, companies are trying to capitalize on the financial value of their businesses by utilizing it (Divol et al. 2012). Social media captures the ‘wisdom of the crowd’. Intuitively, customers’ decisions drive the firm’s value.

Enabled by information technology (IT) advances, social media has rapidly integrated itself into customers’ decision-making since traditional media is not as effective in reaching their ideal customers (Brynjolfsson et al., 2002; Gao and Hitt, 2012). Furthermore, the content of social media is rapidly updated ahead of other sources of investor information (Aral and Walker, 2011). Thus, social media helps investors to absorb a large set of information in order to predict companies’ future business value. It may serve as an indicator of firm equity value (Divol et al. 2012).

Approximately 80% of institutional investors use social media as part of their regular information flow (Greenwich, 2015). The creation, sharing and exchange of information through social media has introduced new avenues for reaching wider investor audiences. As a result, participation in social media is expanding rapidly. Shiller and Pound (1989) suggest that most trading decisions involve interpersonal communication. Social media provides a platform for interpersonal communication and exchanging first-hand information.

Twitter's news feed is now viewed as a key source of information by investors. To our knowledge, only a few recent studies (e.g. Bhagwat et al., 2014; Curtis et al., 2014) use Twitter and StockTwits as a main source for creating a direct attention proxy and we believe Twitter and StockTwits volume can provide a measure of the attention of individuals which could potentially affect stock markets. This chapter will therefore attempt to use Twitter and StockTwits as a unique direct attention proxy to explain how social media affects asset prices and post-earnings-announcement drift (PEAD) in response to companies' quarterly earnings announcements.

Using quarterly earnings data and Twitter and StockTwits data (17 quarters from the fourth quarter of 2010 to the fourth quarter of 2014), we first estimate a model similar to Dellavigna and Pollet (2009), Hirshleifer et al. (2009) and Chakrabarty et al. (2015) to test how Twitter volume affects prices. The results support the prior literature's findings that limited attention earnings announcements have a lower immediate price response and higher PEAD. However, simply using Twitter and StockTwits volume (e.g. Bhagwat et al., 2014; Curtis et al., 2014) could be problematic due to the autocorrelation with Twitter and StockTwits account growth. Therefore, we utilize Twitter and StockTwits volume and a residual methodology to generate an attention proxy that is orthogonal to the growth of Twitter and StockTwits accounts. We find that new attention brought by social media after earnings announcements positively affects cumulative abnormal returns and the magnitudes are larger than the earnings surprise effects. Furthermore, the Twitter based attention effects differ across different industries. Specifically, we find that the new attention effect among unpopular industries is statistically significant and is muted among popular industries. Finally, we find that even companies reporting bad news can still have positive immediate abnormal returns if they attract enough attention from investors after an earnings announcement.

The rest of the chapter is organized as follows. Section two presents the literature review and our hypotheses. Section three discusses the data and sample selection. Section four presents our empirical methodology. Section five contains our main regression specifications for testing social media effects and section six presents our main results. Robustness checks are presented in section seven, while section eight concludes.

2. Background and Literature

The Efficient Market Hypothesis (EMH) states that market expectations are subject to change based only on new pieces of information and, as a result, stock prices fluctuate randomly (Fama, 1970; Samuelsonm, 1965). Early research on stock market prediction was based on the random walk theory and the EMH (Fama, 1970; Fama, 1991; Fama, 1965). EMH suggests that once new information has been revealed, stock prices should reflect all of the information. Since new information is unpredictable, stock prices should not be predicted with more than 50 percent accuracy (Qian et al., 2007).

There are two major branches of literature about the EMH. One branch focuses on arguing that the random walk assumption may not apply since there is evidence that stock prices can be predicted to some degree given a certain time period (Gallagher and Taylor, 2002;

Kavussanos and Dockery, 2001; Bulter and Malaikah, 1992). The other branch suggests that news may be unpredictable but news extracted from online platforms (e.g. discussion boards, blogs and Twitter feeds) can be used as an indicator to predict changes in financial markets and business valuations. For example, Mishne and Glance (2006) attempt to predict movie sales using online blog sentiments. There are a few recent studies that apply this idea to the stock market. For example, Schumaker and Chen (2009) investigate stock price movement using breaking financial news articles.

Post-Earnings-Announcement Drift

The first to document post-earnings-announcement drift (PEAD) are Jones and Litzenberger (1970). They suggest that after a company's earnings announcement, the market needs time to digest the earnings surprises and it may take months. This is because earnings surprises will be gradually spread through advisory services, such as stock brokers, to all other market participants.

Ball and Brown (1968) find empirical evidence that companies that release good earnings news, i.e. those with high standardized-unexpected earnings (SUE), outperform bad-news (low-SUE) stocks. This is due to investors underreacting to the earnings news, thereby generating a cumulative abnormal return. This is known as the PEAD anomaly. Some argue that the PEAD is due to the methodological shortcomings of financial analyst forecasts. Financial analysts are important players in the stock market. They provide guidance to investor decision making through earnings forecast and stock recommendations. However, the information analysts collect is from various sources. Some may argue that financial analysts tend to be too optimistic, over-react to some information and under-react to other information (Easterwood and Nutt, 1999; De Bondt and Thaler, 1990; Francis and Phiblbrick, 1993; Das et al., 1998; Lim, 2001). Others believe that the PEAD is due to investors underreacting to value-relevant earnings information (Abarbenell and Bernard, 1992; Barberis et al., 1998; Daniel et al., 1998; Hong and Stein, 2000). Bernard and Thomas (1990) show that if a company releases good news (positive earnings surprises) for their earnings announcement, their stock returns will increase from the earnings announcement date until the next quarter's earnings announcement date. Furthermore, Foster (1981), and Han and Wild (1990) present evidence that the PEAD exists within industries.

Fama (1970) assumes that all the information is timely, accurate and fully reflected in stock prices including the present and future value of the stocks, unless there is market manipulation. Otherwise investors cannot use the analysis of the previous prices to earn excess profits higher than the market average. Meanwhile, Bernard and Thomas (1990) present perhaps the most persuasive evidence for the PEAD anomaly, arguing that it is due to investors significantly underestimating (or being completely unaware of) the autocorrelation in seasonally-differenced earnings. However, others suggest that such an anomaly may arise from the misspecification of the stock-return-generating process. To examine this perspective, a number of studies (Foster et al., 1984; Bernard and Thomas, 1989; Han and Wild, 1990; Chan et al., 1996) use unexpected earnings, not announcing firms' unsystematic stock returns, to test for information transfers. Using earnings surprises as a proxy for the informative nature of earnings would reduce the likelihood of information transfers attributable to returns model misspecification and would support the hypothesis that earnings surprises, not the misspecification of the stock-return-generating process, drive information transfers on earnings release dates.

For the reasons given above, one might speculate that since investors have their own money on the line, they would follow the new information more closely than financial analysts whose compensation is not directly affected by reacting to the earnings news. Therefore, the market should adjust after the earnings announcement faster than financial analysts. However, the literature does not support this argument. For example, Abarbanell and Bernard (1992) find both financial analysts and market participants underreact to earnings announcements. However, their results show that the market participants underreact more than financial analysts because analysts' under-reaction cannot fully account for PEAD. Alford and Berger (1997), and Brous and Shane (2001) find consistent empirical evidence for this.

Social Media

According to Greenwich (2015), "Almost 80% of institutional investors use social media as part of their regular work flow and 48% of the investors said social media prompted them to do additional research on an industry issue or topic" [2]. Additionally, "approximately 40% of the institutions globally expect to increase their use of social media in the coming year". Greenwich (2015) further suggests that social media platforms have become the key source of information for institutional investors during their investment

[2] Greenwich Associates research is contained in a report entitled Institutional Investing in the Digital Age: How Social Media Informs and Shapes the Investing Process (Greenwich, 2015)

decision making process.

O'Connor et al. (2010) find that Twitter messages could be a leading indicator for predicting the Index of Consumer Sentiment. Zhang et al. (2010) find that using a random subsample of Twitter's public timeline messages can predict market indices such as the Dow Jones Industrial Average. Bollen et al. (2010) find similar empirical evidence with the S&P 500 index. Meanwhile, Chen et al. (2014) find that articles and commentaries that are published on social media platforms predict future stock returns and earnings surprises. There are many different types of investors in the market. The number of informed investors matters because they contribute to the limitation of arbitrage. Social media can play a crucial role in determining whether news reaches a large number of investors and how they perceive the information. Hence, it can have an impact on asset prices through this information role. The media can have an impact also because it creates common knowledge among investors. Morris and Shin (2002) show that increasing public information precision has an ambiguous effect on welfare because investors tend to over-react to public information, allowing the prices to deviate from fundamentals. A number of recent studies show evidence that the media could potentially shift public opinion, e.g. voting behaviour and political opinions (DellaVigna and Kaplan, 2007; Gerber, Karlan, and Bergan, 2009). These findings are surprising because the EMH assumes that the media should not be able to affect stock prices when there is no new information arrives to the market. Many studies present evidence that the media can in fact impact prices and investor behaviour (Huberman and Regev, 2001; Reuter and Zitzewitz, 2006).

Within the social media literature, most studies focus on a particular industry. For example, Liu (2006) and Chintagunta et al. (2010) investigate social media effects on movie box office revenue. Dellarocas et al. (2007) focus on the entertainment industry and find that social media is statistically significant in forecasting entertainment good sales, while Chevalier and Mayzlin (2006) and Dewan and Ramaprasad (2012) both focus on the publishing industry. Chevalier and Mayzlin (2006) find that social media has an important impact on book sales, and Dewan and Ramaprasad (2012) investigate the music publishing industry and show that social media affects consumption, while Godes and Mayzlin (2004) argue that in the motion picture industry, social media has explanatory power for TV ratings. Finally, Luo (2007, 2009) finds that social media impacts stock returns and cash flows in the airline industry.

Investigating PEAD within industries is important because earnings announcements provide information not only about the announcing firms but also peer firms in the same industry. There is a substantial body of literature focusing on the stock price reactions of announcing firms and peer firms within the same industry. For example, Foster (1981), and Han and Wild (1990) show that the PEAD exists in the same industry for the stock prices of competing firms. These studies provide evidence that earnings surprises affect both the stock price of the announcing firms and the non-announcing firms in the same industry. Freeman and Tse (1992) suggest that investors could use the information from earnings announcements in one industry to update their expectations of stock returns in the same industry, while Ramnath (2002) investigates whether investors and financial analysts can incorporate the earnings news in the industry.

Prior research documents the general movement for a stock's cumulative abnormal returns to drift in the direction of earnings surprises following an earnings announcement. Therefore, as the first step in this chapter, we verify whether the industries we utilize have such tendencies. Additionally, positive earnings surprises and negative earnings surprises may have different effects on an investor's portfolio. Thus, we examine the differences between the effects of positive and negative earnings surprises on the corresponding cumulative abnormal returns in each of the two industry categories that we use for our empirical analysis.

Hypotheses

Twitter was launched on March 21, 2006. Unlike other social networking websites such as Facebook, Twitter allows users to create, distribute and discover content without reciprocation[3]. Every second there are approximately 6,000 tweets, which corresponds to over 350,000 tweets sent per minute, 200 million tweets per day and around 200 billion tweets per year[4]. As of the third quarter of 2015, Twitter averaged 307 million monthly active users with 1 billion unique monthly visits to sites with embedded Tweets[5].

Social media is considered as a platform for firms to communicate and target their ideal customers (Gallaugher and Ransbotham, 2010). The rapid development of social media does not only gradually change people's daily lives, but also promotes a fundamental

---

[3] Source: rikorian, Raffi. (VP, Platform Engineering, Twitter Inc.). 'New Tweets per second record, and how!' Twitter Official Blog. August 16, 2013

[4] Source: Twitter Engineering. '200 million Tweets per day.' Twitter Official Blog. June 30, 2011

[5] Source: Twitter Company Facts: https://about.twitter.com/company

change in the dissemination of information. Online discussion websites such as consumer review boards and blogs are being used by investors to collect the newest information regarding companies' future predictions (Chen and Xie, 2008; Gu et al., 2012). Several recent studies document various ways in which social media can affect financial markets by using Twitter data. For example, Blankespoor et al. (2014) find that the tweets of companies that are related to press releases can affect stock liquidity and Chen et al. (2013) claim that corporate executives' personal tweets can help predict stock returns and liquidity. Meanwhile, Chawla et al. (2014) find evidence that Twitter messages are correlated with bid-ask spreads and stock prices on the news day. They conclude that social media does not introduce new information to the stock market, rather it spreads the news broadly. Our paper differs from these in its focus on using Twitter as an attention proxy regardless of the content of the tweets. We propose that Twitter volume can provide a measure of the attention of investors which could potentially affect stock markets.

Indirect proxies for investors' attention that have been used in the literature are extreme returns, news and headlines, and trading volume (Gervais et al., 2001; Barber and Odean, 2008; Hou et al., 2008; Yuan, 2008). Moreover, some unconventional indirect proxies such as advertising expense and price limits are also used in existing studies (Grullon et al. 2004; Seasholes and Wu, 2007; Lou, 2008; Chemmanur and Yan, 2009). All these indirect proxies have strong assumptions. For example, as long as a stock's return or turnover is extreme or the companies' names are mentioned in the news, investors will pay attention. However, this assumption can be easily disproved. There are many factors such as liquidity risk that could result in extreme stock return or turnover, which has nothing to do with investors’ attention. It is not appropriate to assume the consequences of investors paying attention can be equivalent to extreme stock return or turnover. Companies’ names being mentioned in news and headlines may not necessarily grab investors’ attention if the news cannot reach a broad audience.

Hence, we propose a direct measure of investors’ attention using Twitter volume. The mechanics of the attention measure creation will be discussed in section 5.

Several recent studies focus on investors’ attention and earning news. Chakrabarty and Moulton (2012) show that market makers’ limited attention could affect stock liquidity. Specifically, they find evidence that when an earnings announcement happens in the stocks handled by a market marker, the liquidity is lower for non-announcing firms because of his/her limited attention. In addition, Chakrabarty et al. (2015) investigate limited attention effects by comparing high-frequency traders with non-high-frequency traders, while Van Bommel (2003) finds that informed investors are more likely to spread ‘imprecise rumors’ for the stocks they trade and that it may positively impact the prices. In line with this argument, Wysocki (1998) finds that increasing message volume could yield positive immediate abnormal returns. Meanwhile, Antweiler and Frank (2004) find that the internet message boards’ volume has an impact on stock returns and market volatility. However, Dewally (2003) argues that the stock recommendations made on internet forums are not significantly correlated with cumulative abnormal returns. Sabherwal et al. (2008) find evidence that the message volume of online message boards such as Yahoo! Finance are correlated with immediate abnormal returns and positive stock returns on the next day.

Barber and Odean (2008) argue that retail investors rarely short stocks and that news that grabs their attention will on average lead to retail purchases and positive price pressure, while Da et al. (2011) use weekly frequency of Google searches as a direct measure of retail investors' attention to show that attention-driven investors put positive price pressure on stocks. Therefore, the first hypothesis is as follows:

Hypothesis 1: *Increasing Twitter and StockTwits volume due to an earnings announcement is associated with immediate positive stock returns*.

We select two groups of industries for analysis (popular industries and unpopular industries) based on their Twitter and StockTwits volume ranking. The popular industries have the highest ranking with regard to Twitter and StockTwits volume and the unpopular industries have the lowest Twitter and StockTwits volume. Hirshleifer et al. (2009) find that limited investor attention leads to market under-reaction, while DellaVigna and Pollet (2009) show that limited attention has a less immediate price response but a higher drift after the earnings announcements. As such, we expect:

Hypothesis 2: *The magnitude of the increasing Twitter volume effect on CAR will be different between high Twitter volume and low Twitter volume industries.*

3. Data and Sample Selection

For our analysis, we use both financial data and social media data.

3.1 Financial Data

(Figure 1 Here)

We rank S&P 500 companies by their Global Industry Classification Standard (GICS) sectors and examine their average tweet volume over the time period 2010-2014 (Figure 1). The information technology sector has the highest cross-sectional average Twitter volume among the 10 sectors in our sample. The sector that has the lowest cross-sectional average Twitter volume is the utilities sector.

(Figure 2 Here)

For each sector, we calculate tweet speed for each GICS sub-industry. Tweet speed is defined as the number of tweets generated per day. We use tweet volume over 2010-2014 to calculate the tweet speed for each sub-industry in each sector. In order to examine social media effects, we want to select the industries that have frequently been mentioned in social media and intensively engage social media participants (measured by retweets). We sort the data into ten deciles based on average tweet speed in each sub-industry (Tweet speed rank=1 : the lowest average tweet speed; Tweet speed rank=9 : the highest average tweet speed). Figure 2 shows the average daily Twitter volume decile ranking for sub-industries over the time period 2010-2014. The industries in the 10^{th} percentile have average daily Twitter volume at 66. For 90^{th} percentile, industries have average daily Twitter volume at 440. We select the industries with the highest tweet speed rank and name them *popular industries* (90^{th} percentile). For the industries with the lowest tweet speed rank, we name them *unpopular industries* (10^{th} percentile). We examine the social media effects on PEAD by comparing popular industries with unpopular industries.

Our sample consists of all firm-quarter earnings observations for the popular and unpopular industries (by GICS code) for which complete data are available. All the firms with non-calendar fiscal quarters and foreign firms are excluded from the sample. In addition, firms chosen must satisfy the following selection criteria:

(1) In order to match the quarterly earnings data and daily stock prices and returns, companies have to be listed on both the Standard & Poor's COMPUSTAT database and CRSP database;

(2) Companies' quarterly earnings announcement dates, and actual earnings are available from Standard & Poor's COMPUSTAT database; and

(3) Financial analysts' forecasts of quarterly earnings are available from the Institutional Brokers' Estimate System (IBES) database immediately prior to the earnings announcement dates and the reporting dates and corresponding actual earnings announcement data are available as well.

We examine all quarterly earnings announcements for the total of 17 quarters, from the fourth quarter of 2010 to the fourth quarter of 2014. We use data from the Center for Research in Security Prices (CRSP), COMPUSTAT, IBES and Bloomberg. The ownership data come from the Thomson Reuters CDA/Spectrum database on SEC 13F filings[6].

(Table 1 Here)

[6] Form 13F is filed on a quarterly basis by institutional investment managers who exercise investment discretion over accounts holding at least $100 million in eligible equity securities. These managers report the total long positions in each eligible security, aggregated across all accounts over which they exercise investment discretion.

Table 1 presents the frequency by GICS code of sub-industry. To examine the social media effect on PEAD, we select S&P 500 companies as our sample. There are 102 out of 500 companies in the unpopular industries and they represent 22 different GICS sub-industries. Popular industries are made up of 15 different GICS sub-industries with 77 companies out of 500 companies. The frequency gives the number of firms under each GICS code followed by the percentage of the total number of firms in the category. For example, the multi-utility industry has 14 out of 102 companies which is 12.75% of the companies within unpopular industries.

3.2 Social Media Data

For our set of firms identified above, we obtain social media data from Quandl Financial and Economic Data (Quandl). Quandl offers access to financial, economic and social datasets from multiple sources. We collect all the Tweets that are related to each company in your sample. We start by carefully creating filters to select the relevant tweets for a specific company using companies' stock ticker symbol. We do the search for each of the companies in our sample and set filters to obtain all the relevant tweets from a Quandl search engine. From Quandl, we obtain the daily total message volume from Twitter and StockTwits[7] from December 2010 to December 2014. Firms' Twitter and StockTwits data from Quandl allow us to gather all the daily tweets which including retweets, images, and URL links that related to the company. Therefore, we are able to collect all the tweets we need around companies' earnings announcement dates. The total daily Twitter volume is

[7] StockTwits is a social media platform designed for sharing ideas between investors, traders, and entrepreneurs. As of June 2013 StockTwits currently has 230,000 active members.

calculated as the summation of the total number of the tweets from both of the platforms. By convention, Twitter and StockTwits in discussions about a particular stock usually include the stock symbol prefixed by a dollar sign (e.g., $AAPL for Apple Inc).

In addition, in order to make sure that a prefixed dollar sign plus the stock ticker symbol would be a unique filter to obtain the relevant messages for each stock, we randomly selected 30 tweets for each stock from the time period 2010-2014. To be acceptable, at least 50% of tweets have to be related to the company, e.g., mentioning the company or their financial situation. We remove any company from our sample that does not meet this rule. As a result, 1.6% of stocks have been removed by this filter.

(Table 2 Here)

Table 2 reports the descriptive statistics for the sample, which includes stock's daily closing price, volume, actual earnings, daily return, unexpected earnings (UE), cumulative abnormal returns (CAR) and cumulative Twitter and StockTwits volume. The cross-sectional mean of stock price is $66.29 for popular industries, and $60.60 for unpopular industries; the full sample average stock price is $66.09. The average trading volume is 14.70 million dollars for popular industries and 5.58 million dollars for unpopular industries, while the full sample has average trading volume at 9.44 million dollars. The mean of unexpected earnings for popular industries is $1.80, however, the average unexpected earnings for unpopular industries is $1.07. For popular industries, average post one-day, 20-day, 40-day, cumulative abnormal returns are 0.29%, -0.04%, and -0.58%, respectively. For unpopular industries, the average post one-day, 20-day, 40-day, cumulative abnormal returns are -0.01%, -0.67%, and -1.08%, respectively. Institutional ownership (IO) is the percentage of shares owned by institutions at the end of the most recent calendar quarter constructed from the CDA/Spectrum 13F database[8]. Average IO is high in our full sample which is 73.72%. Average IO for popular industries is 76.38% which is higher than the unpopular industries average IO of 69.82%.

4. Empirical Analysis

In examining the cumulative abnormal returns reaction to earnings announcements, we assume that companies' earnings information after they reveal on the announcement date could be disseminated during the sixty-day announcement period (approximately 3 months) before the next earnings announcement date. That is, we expect companies' earnings information is incorporated into financial analyst forecasts throughout the quarter forecasts so that these forecasts can represent the investors prior beliefs for companies' earnings. Similar to many previous studies[9], we assume only that the previous quarter's earnings announcement has incremental information content. We also assume that during the earnings announcement period market participants receive the earnings news but the market may not react fully to adjust the prices. Evidence from the PEAD literature suggests that the market reacts to earnings news gradually rather immediately. Our empirical strategy incorporates this possibility.

[8] The institutional ownership of these stocks is also quite high for most of the sample and in some instances exceeds 100%. Because shares that are shorted are owned by more than one party (the original lender plus the purchaser on the other side of the short sale), institutional ownership can exceed 100%. If a share sold short is re-borrowed and sold again, short interest ratios can also exceed 100%.

[9] For example, Ball and Brown, 1968; Bernard and Thomas, 1990; De Bondt and Thaler, 1990; Francis and Phiblbrick, 1993; Das, Levine and Sivaramakrishnan, 1998; and Lim, 2001; Abarbenell and Bernard, 1992; Barberis et al., 1998; Daniel et al., 1998; and Hong and Stein, 2000.

4.1 Empirical Methodology

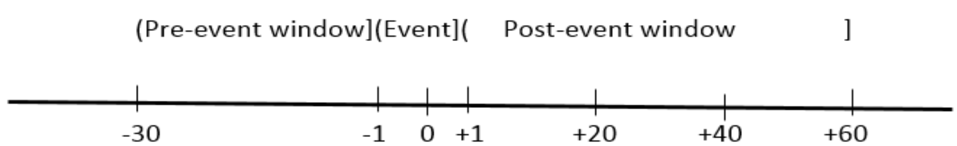
Unexpected Earnings (UE). The unexpected earnings are calculated by subtracting analyst forecast earnings per share from actual earnings per share and then dividing by the absolute value of the forecast. Unexpected earning (UE) is defined as the difference between reported earnings and expected earnings:

$$UE_{j,t,q} = (AE_{j,t,q} - EE_{j,t,q})/|EE_{j,t,q}| \qquad (1)$$

Where $AE_{j,t,q}$ is the actual earnings for firm *j* in quarter *q* of year *t*, $EE_{j,t,q}$ is the analyst's forecast of expected earnings for firm *j* in quarter *q* of year *t*. There is some debate as to which statistic and at which time point should be used as the most updated analysts' expectations for analyst forecasts. A number of studies argue that using the averages of analyst forecasts are more accurate than using a particular individual forecast. This is consistent with the idea of using averages could minimize the idiosyncratic errors.

Cumulative Abnormal Returns (CAR).

Time line:



Similar to most of the PEAD literature, we choose plus/minus one day around the earnings announcement as our estimation window. For the pre-event window we select 30 days before earnings announcement as the estimation window. For the post-event window, we use 20-day, 40-day and 60-day post earnings announcement (e.g., Campbell et al., 1997; Bernard and Thomas, 1990; Foster, 1981; Han and Wild, 1990).

We generate the cumulative abnormal returns for post earnings announcement periods of one-day, 20-day, 40-day and 60-day. The daily abnormal return is the actual firm return minus the return on the CRSP value-weighted index. We use a market model to construct the abnormal returns during the earnings announcement period. Estimates of cumulative abnormal stock returns (CAR) for announcing and non-announcing firms at earnings release dates are:

$$\mathrm{CAR}_{j,\{t_1,t_2\}} = \sum_{t_1}^{t_2} u_{j,t} \qquad (2)$$

Where $u_{j,t} = R_{j,t} - (\partial_0 + \partial_1 * R_{M,t})$ is daily abnormal stock return[10] for firm *j* on day *t*, and $\{t_1, t_2\}$ is the period from day t_1 to day t_2 inclusive. $R_{j,t}$ is the daily stock return for firm *j* on day *t*, and $R_{M,t}$ is the return on the CRSP value-weighted index for day *t*. The parameters, ∂_0 and ∂_1 are estimated using OLS. In terms of time interval $\{t_1, t_2\}$, we have one day, 20 days, 40 days and 60 days subsequent to the earnings announcement. We preserve comparability with Foster, Olsen and Shevlin (1984) and sum abnormal returns over time to obtain cumulative abnormal returns (CARs)[11]. Missing values of the returns on earnings announcement day from CRSP are dropped from our sample. If CRSP returns

[10] We use a market model as the base to calculate the abnormal stock return. In the robustness checks section, we also use Constant Mean-Return Model and market-adjusted-return model to calculate the abnormal stock return.

[11] Blume and Stambaugh (1983) argues that summing abnormal returns over time implicitly assumes daily rebalancing and leads to an upward bias in the returns cumulated over time periods. However, Bernard and Thomas (1989) conduct analyses that indicate that the difference between abnormal returns on extreme good news and bad news firms is similar. Since the bias affects both the primary and the companion portfolios, there is no bias in our estimated abnormal returns.

series does not encompass the 120 trading days surrounding the earnings announcement, the observations are also dropped.

Attention Proxy

Prior studies have used the Search Volume Index (SVI) in Google as a direct investors' attention proxy (e.g., Da et al., 2011; DeHaan et al., 2015; Chakrabarty et al., 2015). Twitter volume has been used as an attention proxy in a few recent studies (e.g., Bhagwat et al., 2014; Curtis et al., 2014).

(Figure 3 here)

We initially use Twitter volume as an attention proxy. Figure 3 shows the growth in Twitter accounts over our sample period. Since increasing Twitter volume is correlated with the number of newly opened Twitter accounts, we want to create another Twitter volume based attention proxy that is orthogonal to the growth of Twitter accounts.

We first identify the new attention gained one day after the earnings announcement date:

$$\mathrm{TweetVolPost1}_{j,t} = \alpha + \beta * \mathrm{TweetVolPre20}_{j,t} + \varepsilon, \qquad (3)$$

where $\mathrm{TweetVolPost1}_{j,t}$ is Twitter volume one day after the earnings announcement for company j at quarter t, and $\mathrm{TweetVolPre20}_{j,t}$ represents cumulative Twitter volume for 20 days before the earnings announcement for company j at quarter t. Then we create the residuals from the regression, which we call $\mathrm{NewAttPost1}_{j,t}$. Therefore,

$$\mathrm{NewAttPost1}_{j,t} = \mathrm{TweetVolPost1}_{j,t} - \mathrm{Tweet}\widehat{\mathrm{VolP}}\mathrm{ost1}_{j,t} \qquad (4)$$

where $\widehat{\mathrm{TweetVolPost1}}_{j,t}$ is the predicted value from equation (3), $\widehat{\mathrm{TweetVolPost1}}_{j,t} = \alpha + \hat{\beta} * \mathrm{TweetVolPre20}_{j,t}$, and $\hat{\beta}$ is the estimator from equation (3).

4.2 Verifying Post-Earnings-Announcement Drift in Each Industry

We examine the stock cumulative abnormal return reaction to the same stock's earnings surprise. The stock cumulative abnormal return reaction to quarterly earnings announcements is measured over different time intervals centered on the announcement date.

The linear relationship between the unexpected earnings and stock price movements is modeled as follows (Campbell et al., 1997; Bernard and Thomas, 1990; Foster, 1981; Han and Wild, 1990):

$$\mathrm{CAR}_{j,q} = \alpha + \beta_1 * \mathrm{UE}_{j,q} + +\sum \beta_k * X_{j,q,k} + \varepsilon_{j,q} \qquad (5)$$

where $\mathrm{CAR}_{j,q}$ is the firm *j*'s cumulative abnormal stock return at quarter *q* (the day after quarter *q* earnings announcement and continuing through the day of the firm's quarter *q+1* earnings announcement); $\mathrm{UE}_{j,q}$ is the firm *j*'s unexpected earnings at quarter *q*; and $X_{j,q,k}$ is a vector of the control variables including industry dummies and seasonal dummies.

4.3 Positive and Negative Earnings Surprise Effects

Previous studies of PEAD show that investors react to good news and bad news differently and there is an asymmetrically large negative price response to negative earnings surprises (Basu, 1977; Dreman and Berry, 1995; and Skinner and Sloan, 2002). Our analysis of splitting positive and negative earnings surprises is based on the assumption that investors have overly optimistic expectations for stocks in response to earnings surprises. Thus, this assumption could lead to an asymmetrically large negative stock return following negative earnings news. This is because the negative earnings surprise causes investors to revise downward their previously overly optimistic expectations (Skinner, 2002). Further, another reason for splitting positive and negative earnings surprises is investors react differently. They are likely to keep a long position if it is good news so that the stock returns will be realized during the positive earnings surprises period but not the negative earnings surprises period. To split the market earnings surprises into positive and negative, we use unexpected earrings interacted with positive and negative dummy variables. Then we have the regression model as follows:

$$\mathrm{CAR}_{j,q} = \alpha + \beta * \mathrm{UE} * \mathrm{POSITIVE}_{j,q} + \mu * \mathrm{UE} * \mathrm{NEGATIVE}_{j,q} + \varepsilon_{j,q} \qquad (6)$$

where $\mathrm{CAR}_{j,q}$ is the firm *j*'s cumulative abnormal stock return at quarter *q* (the day after quarter *q* earnings announcement and continuing through the day of the firm's quarter *q+1* earnings announcement); $\mathrm{UE} * \mathrm{POSITIVE}_{j,q}$ is the positive unexpected earnings of the firm *j* at quarter *q*; and $\mathrm{UE} * \mathrm{NEGATIVE}_{j,q}$ is the negative unexpected earnings of the firm *j* at quarter *q*.

5. Social Media Effects

5.1 Twitter Volume

Previous literature finds that for limited attention earnings announcements, the immediate price response is lower but PEAD is higher (e.g., DellaVigna and Pollet, 2009; Hirshleifer et al., 2009). To understand the possible effects of attention due to social media, we first use Twitter volume as an attention proxy in a specification similar to Hirshleifer et al.,

(2009) and Chakrabarty, Moulton and Wang (2015):

$$\mathrm{CAR}_{j,q} = \alpha + \beta_1 * \mathrm{UE}_{j,q} + \beta_2 * \mathrm{Log(TwTVol)}_{j,q} + \beta_3 * \mathrm{Log(TwTVol)}_{j,q} * \mathrm{IO}_{j,q} + \sum \beta_k * X_{j,q,k} + \varepsilon_{j,q}, \qquad (8)$$

where $\mathrm{CAR}_{j,q}$ is the firm *j*'s cumulative abnormal stock return at quarter *q* (the day after quarter *q* earnings announcement and continuing through the day of the firm's quarter *q+1* earnings announcement); $\mathrm{UE}_{j,q}$ is unexpected earnings of firm *j* at quarter *q* (as in Hirshleifer et al., 2009 and Chakrabarty, Moulton and Wang, 2015); $\mathrm{Log(TwTVol)}_{j,q}$ is the logarithm of the Twitter volume of firm *j* one day after the earnings announcement at quarter *q*; $\mathrm{IO}_{j,q}$ is institutional ownership for firm *j* at quarter *q* and $X_{j,q,k}$ is a vector of the control variables including industry dummies, seasonal dummies, and institutional ownership.

5.2 New Attention Residual

Due to the autocorrelation between Twitter account growth and twitter volume, we utilize the residual methods described previously to create another attention proxy using Twitter volume. From the above analysis the following empirical model specification to estimate the social media effects is:

$$\mathrm{CAR}_{j,q} = \alpha + \beta_1 * \mathrm{UE}_{j,q} + \beta_2 * \mathrm{NewAttPost}_{j,t} + \beta_3 * \mathrm{NewAttPost}_{j,t} * \mathrm{IO}_{j,q} + \sum \beta_k * X_{j,q,k} + \varepsilon_{j,q} \qquad (9)$$

where $\mathrm{CAR}_{j,q}$ is the firm *j*'s cumulative abnormal stock return at quarter *q* (the day after quarter *q* earnings announcement and continuing through the day of the firm's quarter *q+1* earnings announcement). $\mathrm{UE}_{j,q}$ is the firm *j*'s unexpected earnings at quarter *q*. $\mathrm{NewAttPost}_{j,t}$ is the attention proxy for firm *j* at quarter *q*. $\mathrm{IO}_{j,q}$ is institutional ownership for firm *j* at quarter *q*. $X_{j,q,k}$ is a vector of the control variables including industry dummies, seasonal dummies and institutional ownership. Similarly, we also split unexpected earrings in to positive and negative to test social media effects by using new attention proxy:

$$CAR_{j,q} = \alpha + \beta * UE * POSITIVE_{j,q} + \mu * UE * NEGATIVE_{j,q} + \gamma * NewAttPost_{j,t}$$
$$+\vartheta * NewAttPost_{j,t} * IO_{j,q} + +\sum \beta_k * X_{j,q,k} + \varepsilon_{j,q} \qquad (10)$$

6. Results

(Table 3 Here)

The empirical results are shown in Table 3. The dependent variable for all of the regressions is cumulative abnormal returns. We control for industry dummy variables and year-month dummy variables. Panel A presents results which replicate those in the existing literature. Panel B presents the results with the residual attention proxy. Columns 1-3 of Panel A are the results replicating the existing studies for PEAD which only has unexpected earnings as independent variable. In Columns 7-9 of Panel A, the independent variables include unexpected earnings, logarithm of the Twitter volume as an attention proxy, institutional ownership, and the interaction between institutional ownership and logarithm of the Twitter volume. The results are consistent with the literature (e.g., DellaVigna and Pollet, 2009; Hirshleifer et al., 2009; Chakrabarty, Moulton and Wang, 2015) that for post one day Twitter volume positively affects post one day abnormal returns and negatively impacts cumulative abnormal returns 20 days and 40 days post earnings announcement. For post 20 days and 40 days of the announcement, limited attention earnings announcements have higher PEAD. Columns 1-3 of Panel B are the results from running the regression in equation (10) which uses a residual attention proxy. As we can see, the results are consistent in terms of the sign and the significance between Panel A column 1-3 and Panel B column 1-3. Specifically, the attention effect by social media is positive for 1 day, negative for 20 days and 40 days post earnings announcement (0.33%, -2.35%, -2.88% respectively) with a 1% significant level. The results indicate that the more new attention being brought by social media after the earnings announcement, the higher the cumulative abnormal returns, are consistent with the findings in the existing literature that investors' attention puts upward pressure on the stock price.[12] Since retail investors rarely short stocks, news that commands their attention will on average lead to retail purchases and positive price pressure, as argued by Barber and Odean (2008). To test whether institutional ownership affects the attention effect, we include an interaction term (New Attention * Institutional Ownership). There is weak evidence that the attention effect is stronger among firms with low institutional ownership. For example, New Attention * Institutional Ownership is positive and significant at 1% level for the CAR 20 days after the earnings announcement, implying that the attention effect on post-announcement returns (negative New Attention) is muted among high institutional ownership firms. Furthermore, since both unexpected earnings and the attention are positively correlated with cumulative abnormal returns, the attention effect could potentially offset the bad news effect on cumulative abnormal returns. This indicates that even if the company has a negative earnings surprise, as long as the stock has been discussed broadly on social media up to one day after the announcement, the cumulative abnormal return will go up after the

[12] Many studies have similar findings that investors' attention can move up the prices. See Kumar and Lee (2006), Barber and Odean (2008), Barber, Odean, and Zhu (2009), Burch, Emery, and Fuerst (2014), and Hvidkjaer (2008).

earnings announcement. In this case, the sign of the unexpected earnings becomes less important.

What is important is the new stock attention being added by social media. In line with this argument, we also test to compare the differences among positive unexpected earnings, negative unexpected earnings and new attention brought by social media. In Table 3, Panel A, columns 4-5 are the results replicating the previous literature that separates positive and negative unexpected earnings. From Panel B columns 4-6 we find that all the signs for unexpected earnings are positive; however, the magnitudes are significantly different between positive and negative unexpected earnings. Negative unexpected earnings have larger marginal effects than positive unexpected earnings. This is consistent with the findings in the previous literature that there are larger price responses to negative earnings surprises than positive earnings surprises[13]. For one day post earnings announcement, the marginal effect for new attention is 0.29% where the negative unexpected earnings effect is 0.34%. While for post 20 days and 40 days, the signs of the marginal effects for new attention are negative.

(Table 4 Here)

Table 4 presents the results by industry, popular industries and unpopular industries. We find that for popular industries the new attention effects are not statistically significant for one day, 20 days and 40 days after the earnings announcement. However, for unpopular

[13] See Basu (1977), Dreman and Berry (1995) and Skinner and Sloan (2002).

industries, results are consistent with the full sample results in Table 3. For example, one day after the earnings announcement, increasing new attention by one unit decreases the cumulative abnormal return by 0.94% and it is statistically significant. Further, new attention effects among unpopular industries are larger than the popular industries. This implies investors are less sensitive to the popular industries which have significant higher average Twitter volume. On the other hand, the cumulative abnormal returns would be significantly affected by the new attention brought by the earnings announcements among the unpopular industries.

7. Robustness Checks

In our main model specification in section 4, we use a market model to calculate the abnormal stock return. In this section, we use two other commonly used models (constant mean-return model and market-adjusted-return model) to calculate the abnormal stock return.

7.1 Constant Mean-return model

$$R_{j,t} = \mu_j + \varepsilon_{j,t}$$

Where $E[\varepsilon_{j,t}]=0$, $var[\varepsilon_{j,t}]=\sigma_\varepsilon^2$. For daily data, $R_{j,t}$ is usually measured by nominal return.

7.2 Market-adjusted-return Model

$$u_{j,t} = R_{j,t} - R_{M,t}$$

$R_{j,t}$ is the daily stock return for firm *j* on day *t*, and $R_{M,t}$ is the return on the CRSP value-weighted index for day *t.*

Using these two models to calculate the cumulative abnormal returns, our results are qualitatively similar to the market model full sample results in Table 3 (results are presented in Appendix). In addition, checking the standardized residuals reveals, no significant outliers in our full sample.

8. Conclusion

This study focuses on whether social media affects PEAD and, if so, what determines the effects. We set up two different hypotheses based on the literature. First, we use Twitter and StockTwits volume as attention proxies to test the social media effects using a similar estimation model from the literature (e.g. DellaVigna and Pollet, 2009; Hirshleifer et al., 2009) and we get consistent results that show that limited attention announcements lower immediate price response but have higher PEAD. However, using volume as an attention proxy may have an autocorrelation problem due to the correlation between Twitter and StockTwits volume and the number of Twitter and StockTwits new accounts. Therefore, we create a new attention proxy using a residual methodology. By creating this orthogonalized direct measure of attention proxy, the major finding of our paper is that the new attention brought by social media after an earnings announcement positively affects the immediate cumulative abnormal returns and the magnitudes are larger than the earnings surprise effects. This finding is consistent with Barber and Odean (2008) who conclude that news that grabs investors' attention will lead to positive price pressure.

We rank industries into deciles based on their average Twitter volume in the time period 2010-2014. By comparing popular industries (the highest ranking industries) with unpopular industries (the highest ranking industries), we find that the new attention effects are different in both magnitude and statistical significance. In addition, we show that new attention effects are significant among unpopular industries rather than popular industries. Our findings show that even companies announcing ‘bad earning news’ can still have a positive immediate stock price response if the companies have provoked enough attention from investors after the earnings announcements. A topic for future research would be to examine the social media effects on PEAD within industries, using a similar time period to the prior research.

REFERENCES

Abarbanell, J., and V. Bernard. "Test of Analysts’ Overreaction/Underreaction to Earnings Inforemation as an Explanation for Anomalous Stock Price Behavior."Journal of Finance 47(1992):1181-1207

Akerlof, G. A. and R. J. Shiller. 2009. Animal spirits: How human psychology drives the economy, and why it matters for global capitalism, Princeton University Press.

Antweiler, Werner, and Frank, Murray Z., 2004, Is all that talk just noise? The information content of internet stock message boards, Journal of Finance 59, 1259-1294.

Aral S, Walker D, 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. Management Sci. 57(9):1623–1639.

Ball, R. and P. Brown. 1986. An Empirical Evaluation of Accounting Income Numbers. Journal of Accounting Research 6 (2):159-178

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343

Barber, Brad, and Terrence Odean, 2008, All that glitters: The effect of attention and news on thebuying behavior of individual and institutional investors, Review of Financial Studies 21:785-818.

Basu, S. 1977. "Investment Performance of Common Stocks in Relation to Their Price Earnings Ratios: A Test of the Efficient Market Hypothesis." Journal of Finance 32, 663–682.

Bernard, V. and J. Thomas. "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings." Journal of Accounting and Economics 13 (1990): 305-341.

Bernard, V. and J. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium?" Journal of Accounting Research Supplement 96(5): 1-36.

Beaver, William H. 1986. The Information Content of Annual Earnings Announcements. Journal of Accounting Research. 6, 67-92.

Bhagwat, Vineet, and Timothy R. Burch. 2014. Pump it Up? Tweeting to Manage Investor Attention to Earnings News. Working paper. University of Oregon and University of Miami

Bommel Van, Jos, 2003, Rumors, The Journal of Finance 58, 1499–1520.

Brynjolfsson E, Hitt L, Yang S. 2002. Intangible assets: Computers and organizational capital. Brookings Papers on Economic Activity: Macroeconomics 2002(1):137–181.

Butler, K. C., & Malaikah, S. J. 1992. Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia. Journal of Banking and Finance, 197-210.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, The Quarterly Journal of Economics, 108, 905-939.

Chakrabarty, B., and Moulton, P. 2012. Earnings announcements and attention constraints: The role of market design, Journal of Accounting and Economics 53: 612-634.

Chakrabarty, B., and Moulton, P and Wang, X., 2015. Attention Effects in a High-Frequency World, Working Paper.

Chan, L., N. Jegadeesh and J. Lakonishok. "Momentum Strategies." Journal of Finance 51(1996): 1681-1713

Chen H., De P. Hu Y, Hwang B, 2014. Wisdom of Crowds: The Value of Stock Opinions

Transmitted Through Social Media. Review of Financial Studies 27, 1367-1403.
Chemmanur, Thomas and An Yan. 2009. Product Market Advertising and New Equity Issues. Journal of Financial Economics. 92 (2) 40-65.
Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman 2010, ‘‘The Effects of Online User Reviews on Movie BoxOffice Performance: Accounting for Sequential Rollout and Aggregation across Local Markets,’’ Marketing Science 944-957.
Conrad, J.; B. Cornell; and W. R. Landsman. “When Is Bad News Really Bad News?” Journal of Finance, 57 (2002), 2507–2533.
Daniel, K., Hirshleifer, D., Subramanyam, A., 1998. Investor psychology and security market under- and overreactions.Journal of Finance 53, 1839–1885
De Bondt, W., Thaler, R., 1990. Do security analysts overreact? The American Economic Review 80, 52-57.
Deans PC. 2011. The impact of social media on C-level roles. MIS Quart. Executive 10(4):187–200.
DellaVigna S, Pollet JM, 2009. Investor inattention and Friday earnings announcements. Journal of Finance, 64(2):709–49
DeHaan, Ed, Terry Shevlin, and Jacob Thornock, 2015, Market (in)attention and the strategic scheduling and timing of earnings announcements, Journal of Accounting and Economics 60: 36-55.
Dewally, Michael, 2003, Internet investment advice: Investing with a rock of salt, Financial Analysts Journal 59, 65-77.
Dewan S, Ramaprasad J. 2012. Music blogging, online sampling, and the long tail. Inform. Systems Res. 23(3, Part 2):1056–1067.
Divol R, Edelman D, Sarrazin H. 2012. Demystifying social media. McKinsey Quart. (April). http://www.mckinseyquarterly.com/ Demystifying_social_media_2958.
Dreman, D. and M. Berry. 1995. Overreaction, Underreaction, and the Low-P/E Effect. Financial Analysts’ Journal 51, 21–30.
Fama, E. and J. MacBeth. 1973. Risk, return and equilibrium: empirical tests. Journal of Political Economy 38: 607-636.
Fama, E. F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance. 25: 383-417.
Fama, E., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.
Fama, E., 1991. Efficient capital markets: II. Journal of Finance 46, 1575–1617.
Foster,G. 1981. Intra-Industry information Transfers Associated with Earnings Releases." Journal of Accounting and Economics 3 : 201-32.
Foster,G. 1977. Quarterly Accounting Data: Thime-Series Properties and Predictive Ability Results." The Accounting Review 3 : 1-21.
Foster,G., C. Olsen and T., Shevlin. 1984. Earnings releases, anomalies, and the behavior of security returns." The Accounting Review 58 : 574-603.
Francis, J., Philbrick, D., 1993. Analysts' decisions as products of a multi-task environment. Journal of Accounting Research 31, 216-230.
Freeman, R.N., and S.Y. Tse. 1992. An Earnings Prediction Approach to Examining Intercompany Information Transfers.” Journal of Accounting and Economics 15: 509-23
Gallagher, L.A., Taylor, M.P., 2002. The stock return-inflation puzzle revisited.

Economics Letters, 75(2), 147-156.
Gallaugher, J., and Ransbotham, S. 2010. Social media and customer dialog management at Starbucks," MIS Quarterly Executive 9(4).
Gao G, Hitt LM. 2012. Information technology and trademarks: Implications for product variety. Management Sci. 58(6):1211–1226.
Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin, 2001, The high-volume return premium, Journal of Finance 56, 877-919.
Grinblatt, Mark, and Matti Keloharju, 2001, What makes investors trade? Journal of Finance, 56, 589-616.
Grossman, Sanford J. and Joseph E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets. American Economics Reivew. 70(3): 393-408.
Grullon, Gustavo, George Kanatas, and James P. Weston. 2004. Advertising, Breadth of Ownership, and Liquidity. Review of Financial Studies. 17(2):439-461.
Gerber A, Karlan D, Bergan, D, 2009. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. American Economic Journal: Applied Economics, 1(2):35–52
Han, J. Y., and J. Wild. 1990. Unexpected Earnings and Intra-Industry Information Transfers: Further Evidence." Journal of Accounting Research 28 : 211-19
Hand, J.; R. Holthausen; and R. Leftwich. "The Effect of Bond Rating Agency Announcements on Bond and Stock Prices." Journal of Finance, 47 (1992), 733–752.
Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets? Journal of Accounting and Economics, 38, 297-331.
Hirshleifer, David, Seongyeon Lim, and Siew Hong Teoh, 2004, Disclosure to an audience with limited attention, working paper, Ohio State University.
Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, Journal of Accounting and Economics, 36, 337-386.
Hong, H., T. Lim and J. Stein, 2000. Bad News Travels Slowly: Size, Analysts Coverage and the Profitability of Momentum Strategies. Journal of Finance 55, 265–296.
Hou, Kewei, Lin Peng, and Wei Xiong, 2008, A Tail of two anomalies: The implications of investor attention for price and earnings momentum, working paper, Ohio State University.
Huberman G, Regev T, 2001. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. Journal of Finance, 56(1):387–96
Jones, P. C. and R. H. Litzenberger. 1970. Quarterly Earnings Reports and Intermediate Stock Price Trends." Journal of Finance 25 :143-148
Kavussanos, M & Dockery, E. 2001. A multivariate test for stock market efficiency: the case of ASE," Applied Financial Economics. Applied Financial Economics 11, 573–79.
Lim, T., 2001. Rationality and Analysts' Forecast Bias. Journal of Finance 56, 369–385.
Liu Y. 2006. Word-of-mouth for movies: Its dynamics and impact on box office revenue. J. Marketing 70(3):74–89.
Lou, Xiaoxia. 2008. The Divergence of Liquidity Commonality in the Cross-Section of Stocks. Journal of Financial Economics. 89(3): 444-466.
Mendenhall, R. 1991. Evidence of possible under weighting of earnings-related information." Journal of Accounting Research 29(2) : 397-417.

Mishne, G & Glance, N. 2006. Predicting Movie Sales from Blogger Sentiment. AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs

Morris S., and Shin H.S., 2002. Social Value of Public Information. The American Economics Review, 92(5): 1521-1534

O'Brien, P., 1988. Analysts' forecasts as earnings expectations." Journal of Accounting and Economics 10 : 53-83.

O'Connor, B, Ramnath B, Routledge B. R, Smith N.A. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. in Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010).

Peng, Lin and Wei Xiong. 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics, 80, 563–602.

Qian, Bo, Rasheed, & Khaled. 2007. Stock market prediction with multiple classifiers. Applied Intelligence 26, 25–33.

Ramnath, S. 2002. Investor and Analyst Reactions to Earnings Announcements of Related Firms: An Empirical Analysis. Journal of Accounting Research 40: 1351-1376

Reuter J, Zitzewitz E, 2006. Do ads influence editors? Advertising and bias in the financial media. The Quarterly Journal of Economics, 121(1):197–227

Sabherwal, Sanjiv, Sarkar, Salil K., and Zhang, Ying, 2008, Online talk: does it matter?, Managerial Finance 34, 423-436.

Seasholes, Mark, and Guojun Wu, 2007, Predictable behavior, profits, and attention, Journal of Empirical Finance, 14, 590-610.

Schumaker, R. P & Chen, H. 2009. A Quantitative. Stock Prediction System based on Financial News. Information Processing and Management, 45(5): 571-583.

Shane, P. and P. Brous. 2001. Investor and (Value Line) Analyst Underreaction to Information about Future Earnings: The Corrective Role of Non-Earnings-Surprise Information." Journal of Accounting Research 39 : 387-404

Shiller, R. J. 2000. Irrational exuberance, Princeton University Press.

Skinner, D. J., and R. G. Sloan. 2002. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Review of Accounting Studies 7:289-312

Tirunillai S, Tellis G. 2012. Does chatter matter? The impact of online consumer generated content on a firm's financial performance. Marketing Sci. 31(2):198–215.

Weinbaum, D. 2009. Assessing the Historical Performance of Hospitality Stocks: The Investor's Perspective. Cornell Hospitality Quarterly 50:113-127

Wysocki, Peter, 1998, Cheap talk on the web: The determinants of postings on stock message boards, Working paper. University of Michigan.

Zhang, X. Fuehres, H. Gloor, P. 2011. Predicting Asset Value Through Twitter Buzz. Springer Advances in Intelligent and Soft Computing, 112.

Figure 1 Average Twitter Volume by Sectors (2010-2014)

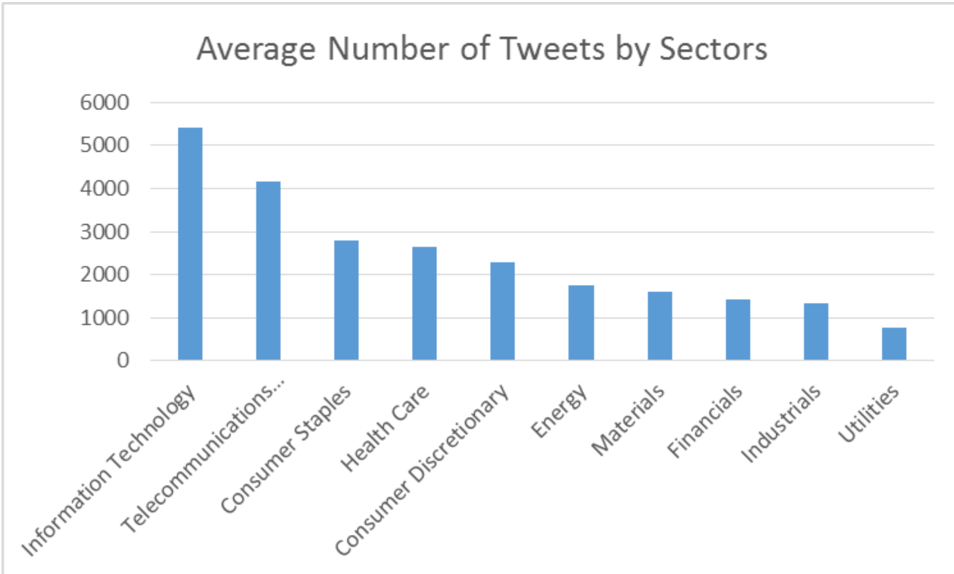


Figure 2 Average Twitter Volume Decile Ranking in Industries (2010-2014)

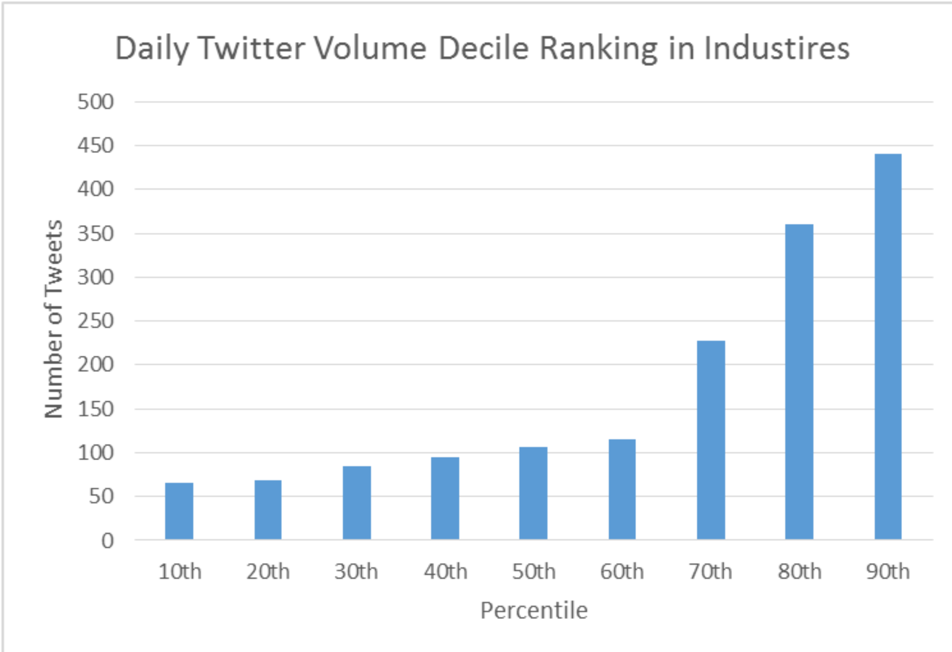


Figure 3 Number of Twitter Accounts from 1st Quarter 2010 to 4th Quarter 2014(in millions)

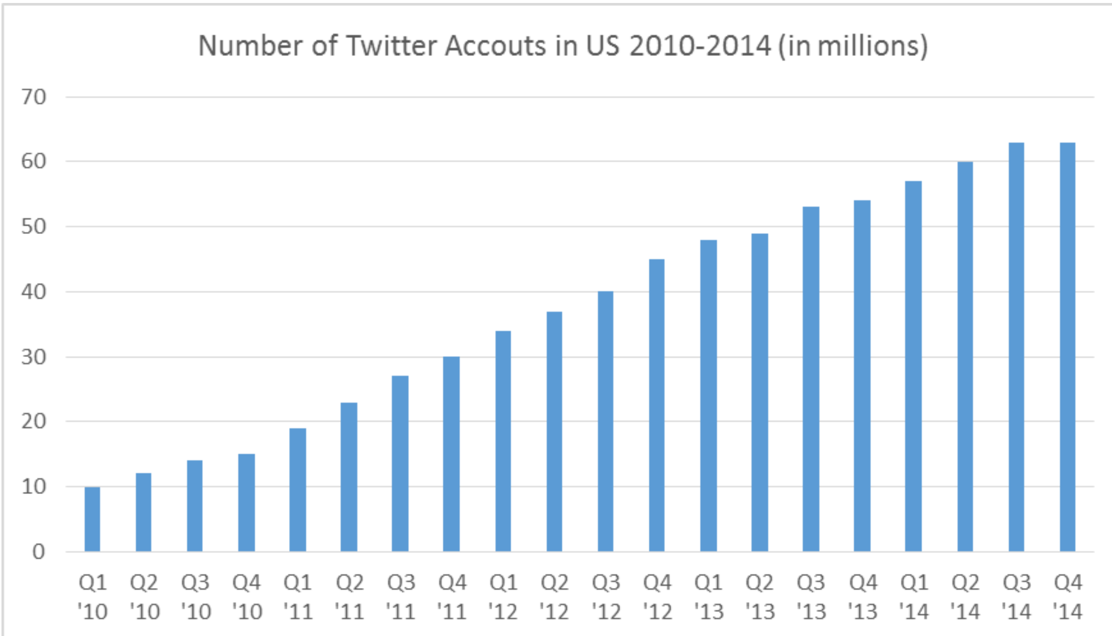


Table 1 Sample Sorted by GSIC Industries
This table present the GSIC code for unpopular industries and popular industries.

Panel A Unpopular Industries

GICS Code	Industry	Freq	Percent
20101010	Aerospace & Defence	11	10.78
20102010	Building Products	2	1.96
20103010	Construction & Engineering	3	2.94
20104010	Electrical Components & Equipment	3	2.94
20105010	Industrial Conglomerates	4	3.92
20106010	Construction & Farm Machinery & Heavy Trucks	4	3.92
20106020	Industrial Machinery	15	14.71
20107010	Trading Companies & Distributors	3	2.94
20201050	Environmental & Facilities Services	4	3.92
20201060	Office Services & Supplies	1	0.98
20201070	Diversified Support Services	1	0.98
20201080	Security & Alarm Services	3	2.94
20202010	Human Resource & Employment Services	1	0.98
20202020	Research & Consulting Services	4	3.92
20301010	Air Freight & Logistics	4	3.92
20302010	Airlines	3	2.94
20304010	Railroads	4	3.92
20304020	Trucking	2	1.96
55101010	Electric Utilities	13	12.75
55102010	Gas Utilities	2	1.96
55103010	Multi-Utilities	13	12.75
55105010	Independent Power Producers & Energy Traders	2	1.96
Total		102	100

Panel B Popular Industries

GICS Code	Industry	Freq	Percent
45101010	Internet Software & Services	7	9.09
45102010	IT Consulting & Other Services	5	6.49
45102020	Data Processing & Outsourced Services	11	14.29
45103010	Application Software	5	6.49
45103020	Systems Software	5	6.49
45103030	Home Entertainment Software	1	1.3
45201020	Communications Equipment	6	7.79
45202010	Computer Hardware	9	11.69
45203010	Electronic Equipment & Instruments	1	1.3
45203015	Electronic Components	2	2.6
45203020	Electronic Manufacturing Services	3	3.9
45301010	Semiconductor Equipment	3	3.9
45301020	Semiconductors	13	16.88
50101010	Alternative Carriers	1	1.3
50101020	Integrated Telecommunication Services	5	6.49
Total		77	100

Table 2 Sample Descriptive Statistics (Dec. 2010-Dec. 2014)
This table presents summary statistics for the full sample and two different industries over the time period Dec. 2010 to Dec. 2014. *Price* is the average daily closing price of the stock; *Volume* is the average daily trading volume; *Actual earnings* is the actual EPS for the stock; *Unexpected earnings* is the average earnings surprised for the stock; *Daily Return* is the average daily return for the stock; *CAR* is the average cumulative abnormal return for the stock; *Tweet Vol* is the average number of tweets for the stock; *Institutional Ownership* is the percentage of shares owned by institutions at the end of the most recent calendar quarter.

	Full Sample		Popular Industries		Unpopular Industries	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
Price ($)	66.09	71.75	66.29	105.31	60.60	37.29
Volume ($Millions)	9.44	29.50	14.70	20.50	5.58	11.60
Actual earnings	0.87	1.05	0.79	1.50	0.98	0.88
Unexpected earnings	1.51	17.65	1.80	4.89	1.07	2.86
Daily Return	0.13%	3.49%	0.09%	3.26%	0.11%	3.24%
CAR post 1 days	0.07%	5.02%	0.29%	6.50%	-0.01%	3.99%
CAR post 20 days	-0.41%	7.34%	-0.04%	8.76%	-0.67%	6.41%
CAR post 40 days	-1.00%	9.27%	-0.58%	10.46%	-1.08%	8.07%
Tweet Vol post 1 day	247.79	1012.15	717.55	2403.00	88.88	158.37
Tweet Vol post 20 days	686.44	2916.51	1938.43	7146.52	264.21	384.84
Tweet Vol post 40 days	1155.88	5094.33	3290.88	12525.33	444.16	627.30
Tweet Vol pre 20 days	636.75	2763.90	1734.45	6819.28	267.28	428.81
Institutional Ownership	73.72%	16.50%	76.38%	24.13%	69.82%	14.10%
Observation	7071		1042		1410	

Table 3 Regressions of Cumulative Abnormal Return on Unexpected Earnings and New Attention: Full Sample (S&P 500 Companies Quarterly Earnings 2010-2014)

The sample used in this table is S&P 500 companies' quarterly earnings data from 2010 to 2014. The dependent variables are the cumulative abnormal return 40 days, 20 days and 1 day after earnings announcement, respectively. *UE* is the unexpected earnings calculated by subtracting the mean analyst forecast from the actual earnings, and then dividing by the absolute value of the mean analyst forecast release. *UE*POSITIVE* is taking unexpected earnings times the positive dummy yields the positive unexpected earnings from full sample. *UE*NEGATIVE* is taking unexpected earnings times the negative dummy yields the negative unexpected earnings from full sample. *Log(Twitter Volume)* is the logarithm of the Twitter Volume one day after the earnings announcement. *New Attention* is residual attention proxy, calculated by subtracting the predicated Twitter volume one day after the earnings' announcement by using 20 days before earnings announcement Twitter Volume from the actual Twitter volume one day after the earnings' announcement. *Institutional Ownership* is the percentage of shares owned by institutions at the end of the most recent calendar quarter. Robust standard errors are reported in parentheses. *, **, *** significant at the 10%, 5%, 1% level, respectively

Regressions of Cumulative Abnormal Returns on Unexpected Earnings

Panel A: Replicating Literature

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}
	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]
Unexpected Earnings	0.0002***	0.0003***	0.0002***	-	-	-	-	-	-
	(0.0000)	(0.0001)	(0.0001)	-	-	-	-	-	-
Unexpected Earnings*Positive	-	-	-	0.0002***	0.0002***	0.0002**	-	-	-
	-	-	-	(0.0000)	(0.0001)	(0.0001)	-	-	-
Unexpected Earnings*Negative	-	-	-	0.0017***	0.0017***	0.0005*	-	-	-
	-	-	-	(0.0002)	(0.0003)	(0.0004)	-	-	-
Unexpected Earnings	-	-	-	-	-	-	0.0001***	0.0001**	0.0001*
	-	-	-	-	-	-	(0.0000)	(0.0000)	(0.0001)
Log(Twitter Volume)	-	-	-	-	-	-	0.0019***	-0.0007*	-0.0004
	-	-	-	-	-	-	(0.0005)	(0.0007)	(0.0009)
Institutional Ownership							-0.0110	-0.0550**	-0.1231***
							(0.0160)	(0.0234)	(0.0294)
Institutional Ownership*Log(Twitter Volume)							-0.0001	0.0001**	0.0001*
							(0.0000)	(0.0000)	(0.0000)
Industry Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	7071	7071	7071	7071	7071	7071	7071	7071	7071
R2	0.0164	0.0166	0.0152	0.0231	0.0184	0.0152	0.0057	0.0070	0.0077

Panel B: Social Media Effects

	{1}	{2}	{3}	{4}	{5}	{6}
	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]
Unexpected Earnings	0.0001***	0.0001**	0.0001*	-	-	-
	(0.0000)	(0.0000)	(0.0001)	-	-	-
Unexpected Earnings*Positive	-	-	-	0.0001**	0.0001	0.0000
	-	-	-	(0.0000)	(0.0000)	(0.0001)
Unexpected Earnings*Negative	-	-	-	0.0031***	0.0034***	0.0038***
	-	-	-	(0.0003)	(0.0005)	(0.0006)
New Attention	0.0033**	-0.0235***	-0.0288***	0.0029*	-0.0235***	-0.0288***
	(0.0017)	(0.0034)	(0.0043)	(0.0017)	(0.0034)	(0.0043)
Institutional Ownership	-0.0032	-0.0198	-0.0853***	-0.0051	-0.0147	-0.0796***
	(0.0450)	(0.0246)	(0.0309)	(0.0447)	(0.0245)	(0.0308)
Institutional Ownership*New Attention	-0.0178	0.0219***	0.0257***	-0.0155	0.0215***	0.0252***
	(0.0297)	(0.0041)	(0.0051)	(0.0296)	(0.0041)	(0.0051)
Industry Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observation	7071	7071	7071	7071	7071	7071
R2	0.0251	0.0457	0.0489	0.0368	0.0521	0.0540

Table 4 Regressions of Cumulative Abnormal Return on Unexpected Earnings and New Attention in Industries

This table presents the regression results for two different industries. The dependent variables are the cumulative abnormal return 40 days, 20 days and 1 day after earnings announcement, respectively. *UE* is the unexpected earnings calculated by subtracting the mean analyst forecast from the actual earnings, and then dividing by the absolute value of the mean analyst forecast release.. *UE*POSITIVE* is taking unexpected earnings times the positive dummy yields the positive unexpected earnings from full sample. *UE*NEGATIVE* is taking unexpected earnings times the negative dummy yields the negative unexpected earnings from full sample. *New Attention* is residual attention proxy, calculated by subtracting the predicated Twitter volume one day after the earnings' announcement by using 20 days before earnings announcement Twitter Volume from the actual Twitter volume one day after the earnings' announcement. Robust standard errors are reported in parentheses. *, **, *** significant at the 10%, 5%, 1% level, respectively

Panel A

	Unpopular Industries			Popular Industries		
	{1}	{2}	{3}	{4}	{5}	{6}
	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]
Unexpected Earnings	0.0034***	0.0040***	0.0044***	0.0019***	0.0016***	0.0019***
	(0.0004)	(0.0006)	(0.0008)	(0.0004)	(0.0005)	(0.0006)
New Attention	0.0094***	-0.0192**	-0.0278**	-0.0108	-0.0117	-0.0146
	(0.0031)	(0.0090)	(0.0114)	(0.0126)	(0.0082)	(0.0098)
Institutional Ownership	-0.0033***	-0.0406	-0.0574	-0.0062	-0.0314***	-0.0542***
	(0.0010)	(0.0397)	(0.0500)	(0.0064)	(0.0182)	(0.0163)
Institutional Ownership*New Attention	-0.0023***	0.0139	0.0245*	-0.0040	0.0088	0.0089
	(0.0006)	(0.0119)	(0.0150)	(0.0035)	(0.0075)	(0.0089)
Industry Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1410	1410	1410	1042	1042	1042
R2	0.1058	0.0834	0.0805	0.0431	0.0437	0.0458

Panel B

	Unpopular Industries			Popular Industries		
	{1}	{2}	{3}	{4}	{5}	{6}
	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]
Unexpected Earnings*Positive	0.0038***	0.0045***	0.0039***	0.0031***	0.0019***	0.0020**
	(0.0005)	(0.0008)	(0.0010)	(0.0006)	(0.0008)	(0.0009)
Unexpected Earnings*Negative	0.0026***	0.0030**	0.0053***	0.0006	0.0013	0.0018*
	(0.0008)	(0.0012)	(0.0016)	(0.0006)	(0.0008)	(0.0010)
New Attention	0.0095***	-0.0189**	-0.0280**	-0.0118	-0.0115	-0.0146
	(0.0031)	(0.0090)	(0.0114)	(0.0126)	(0.0082)	(0.0098)
Institutional Ownership	-0.0033***	-0.0423	-0.0558	-0.0037	-0.0371**	-0.0642***
	(0.0010)	(0.0397)	(0.0501)	(0.0063)	(0.0185)	(0.0223)
Institutional Ownership*New Attention	-0.0024***	0.0135	0.0249*	-0.0023	0.0086	0.0088
	(0.0006)	(0.0119)	(0.0150)	(0.0035)	(0.0075)	(0.0089)
Industry Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1410	1410	1410	1042	1042	1042
R2	0.1068	0.0839	0.0808	0.0510	0.0439	0.0458

Appendix

Panel A: Market-adjusted-return Model						
	{1}	{2}	{3}	{4}	{5}	{6}
	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]
Unexpected Earnings	0.0001***	0.0001**	0.0001*	-	-	-
	(0.0000)	(0.0000)	(0.0001)	-	-	-
Unexpected Earnings*Positive	-	-	-	0.0001**	0.0001	0.0001
	-	-	-	(0.0000)	(0.0000)	(0.0001)
Unexpected Earnings*Negative	-	-	-	0.0032***	0.0034***	0.0038***
	-	-	-	(0.0003)	(0.0005)	(0.0006)
New Attention	0.0032**	-0.0214***	-0.0265***	0.0029*	-0.0214***	-0.0266***
	(0.0017)	(0.0035)	(0.0043)	(0.0017)	(0.0035)	(0.0043)
Institutaional Ownership	-0.0009	-0.0205	-0.0867***	-0.0029	-0.0151	-0.0807***
	(0.0450)	(0.0246)	(0.0310)	(0.0449)	(0.0246)	(0.0310)
Institutaional Ownership*New Attention	-0.0163	0.0205***	0.0242***	-0.0139	0.0201***	0.0239***
	(0.0298)	(0.0041)	(0.0052)	(0.0296)	(0.0041)	(0.0052)
Industry Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7071	7071	7071	7071	7071	7071
R2	0.0250	0.0450	0.0485	0.0366	0.0516	0.0535

Panel B: Constant Mean-return Model						
	{1}	{2}	{3}	{4}	{5}	{6}
	CAR[0,1]	CAR[2,20]	CAR[21,40]	CAR[0,1]	CAR[2,20]	CAR[21,40]
Unexpected Earnings	0.0001***	0.0002***	0.0002**	-	-	-
	(0.0000)	(0.0001)	(0.0001)	-	-	-
Unexpected Earnings*Positive	-	-	-	0.0001**	0.0001**	0.0001*
	-	-	-	(0.0000)	(0.0001)	(0.0001)
Unexpected Earnings*Negative	-	-	-	0.0032***	0.0032***	0.0037***
	-	-	-	(0.0003)	(0.0006)	(0.0007)
New Attention	0.0026*	-0.0017*	-0.0039*	0.0022*	-0.0016*	0.0036*
	(0.0018)	(0.0014)	(0.0018)	(0.0018)	(0.0004)	(0.0019)
Institutaional Ownership	-0.0188	-0.0319	-0.1085***	-0.0171	-0.0273	-0.1032***
	(0.0470)	(0.0283)	(0.0346)	(0.0469)	(0.0283)	(0.0346)
Institutaional Ownership*New Attention	-0.0037	0.0016	0.0011	-0.0013	0.0012	-0.0014
	(0.0311)	(0.0047)	(0.0058)	(0.0310)	(0.0047)	(0.0058)
Industry Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Dummy Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7071	7071	7071	7071	7071	7071
R2	0.0275	0.0717	0.0602	0.0384	0.0758	0.0639

CHAPTER 3

DOES YOUR ATTENTION DRIVE YOUR PROFITS?

1. Introduction

Along with the development of mathematical finance and computer technology, pairs trading originated from the practice of securities traders. The embryonic form of pairs trading can be traced back to the twentieth century. This strategy is simple, but it is still widely used today. The idea of pairs trading is to look for a pair of stocks with similar historical trends. When the price of one stock in the pair starts to rise from the historical trends and/or the price of the other stock falls, traders take a long position in the undervalued stock and short the overvalued stock. When the divergence of the spread between the two stocks returns to the historical level, the profits will be realized by selling the long position stock and buying the short position stock (Gatev et al., 2006). The pairs trading strategy is considered a market neutral strategy with stable income. The key to the success of pairing trading is that the stocks with similar historical trends will always return to equilibrium after a divergence.

In stock investment, investors need to consider a certain number of stocks in a limited time. Why should they consider certain stocks instead of other stocks? Recent studies in investors' attention can explain the specific financial market anomalies, and the research results have been widely applied to corporate finance, asset pricing and other areas (Mankiw and Reis, 2002; Hirshleifer and Teoh, 2003; Graham and Kumar, 2004). Attention is a rare cognitive resource (Kahneman, 1973). When there is a large amount of available information as part of investment decision-making, there are inevitable limits on investors' cognitive resources. Investors could overreact to specific events causing major fluctuations in stock markets (Engelberg et al., 2009). The stock market reactions such as abnormal trading volume and excess returns can be explained by investors' attention towards specific news events, which involve certain industries and companies (Barber and Odean, 2008). For example, on January 28th, 2016, the World Health Organization (WHO) announced that the Zika virus was 'spreading explosively' in the Americas [14]. Consequently, the healthcare sector index went up by 4.34% in a month after WHO's announcement[15]. Investors expected that the demand for healthcare industry products would increase, making the relevant stocks (pharmaceuticals, biotechnology, health care providers and equipment) go up without fundamental changes in firms. This is the stock market overreaction to the investors' attention as a result of investors' limited ability to process and absorb all available information (Engelberg et al., 2009). For individual and institutional investors, understanding investors' attention can effectively avoid the loss (Barber and Odean, 2008). Therefore, it is necessary to know and study investors' attention. Twitter volume has been used as an attention proxy in a few recent studies (e.g. Bhagwat and Burch, 2014; Curtis et al., 2014). Given that the empirical evidence in this area is still limited, extending the analysis to pairs trading strategies should help bring some valuable insights. Therefore, the purpose of this study is to investigate the economic drivers of pairs trading profits. Specifically, we examine whether the pairs trading profits are driven by social media after accounting for determinants of the return comovement, given that prior studies have identified a large set of variables that are correlated with stock return

[14] Source: http://www.who.int/en/
[15] Source: http://www.spindices.com/indices/equity/sp-500-health-care-sector

comovement such as earnings comovement, industry membership, size, exchange membership and trading volume (Fama and French, 1992, 1993; Sloan, 1996; Engelberg, et al., 2009; Shiller, 1989; Campbell and Mei, 1993; Barberis et al., 2005; Kumar and Lee, 2006; Greenwood, 2008; Pirinsky and Wang, 2006; Israelsen, 2009; Green and Hwang, 2009; Boyer, 2010; Gao, 2010; Hameed et al., 2010).

Our main contribution is the examination of social media as one of the economic drivers of the pairs trading strategy. Building on Engelberg et al. (2009), who provide possible explanations for pairs trading profits, we identify social media as another driver of pairs trading profits. Our work builds on a large body of literature that investigates the economic drivers of stock return comovement. One of the advantages for utilizing firm-level pairwise return is that we are able to control for a rich set of economic factors that have been identified from previous studies while examining social media effects on return comovement.

Accurately measuring trading profits provides us with higher statistical power when analyzing the sources of the profit. Several recent studies explore certain aspects of pairs trading such as the effectiveness of the basic algorithm and methodology of pair selection (Gatev et al., 2006; Do and Faff, 2009; Lin et al., 2006). Little research has been done measuring pairs trading profits. When calculating pairs trading profits, there are a few steps involved: (1) trading pairs selection, (2) the length of training[16] and trading[17] period and (3) portfolio return realization. This work contributes to the literature by extending the

[16] Training period refers to the time period used in pairs trading to identify trading pairs.
[17] Trading period refers to the time period that traders act on opening and closing trades.

cointegration approach of Lin et al. (2006), Vidyamurthy (2004), Gillespie and Ulph (2001), and Hong and Susmel (2003) by using the technique developed by Johansen (1988) to access the profitability of a pairs trading strategy.

The chapter proceeds as follow. In Section 2, we review and compare different methods for pair selection and introduce our method and profit measure. In Section 3, we review the economic drivers of the pairs trading returns and introduce social media as a new practical driver. Section 4 summarizes the data and sample selection, while Section 5 presents the empirical results and section 6 concludes.

2. Trading Pairs Selection and Measuring Profits

When it comes to pairs trading, the key to its success is to identify trading pairs. There are several statistical methods have been used in the literature to find the trading pairs and the relationships between the pairs. The main methods include Distance Trading, Stochastic Spread, and Cointegration. In this section, we briefly review the first three methods and then provide a detailed discussion of the cointegration method, which is used in this study.

2.1 Distance Method and Stochastic Spread Method

The distance method refers to a nonparametric method of minimizing the distance between the two stocks in the pair. Gatev et al. (2006) describes the distance methods as follows: first, choose an appropriate training period and normalize the stock prices; second, calculate the square distance between normalized prices. Trade criteria is based on the distance between the two stocks. When a pair is found in which the distance exceeds the preset threshold, the trade is executed. The main disadvantage for the distance method is that it purely establishes the statistical relationship between the two stocks rather than the economic relationship.

Nath (2003) applies the pairs trading strategy for U.S. Treasury securities to show that traders in the secondary market could profit by using the simple pairs trading strategy compared to equities. Gatev et al. (2006) employ the distance method to test the pairs trading strategy over the time period 1962-2002 and find that it generates annualized excess portfolio returns of up to 11% on average. Furthermore, Do and Faff (2009) test the distance method used in Gatev et al. (2006) and they find the pairs trading profits have been declining by extending the sample to 2008. Mori and Ziobrowski (2011) suggest that applying the pairs trading strategy in the U.S. REITs market could yield a higher profit than common stocks. Huck (2013) shows that, within the distance method, the length of the training period in pair selection is highly sensitive to the portfolio excess returns.

Meanwhile, Elliott et al. (2005) develop a mean-reverting stochastic model for the pairs trading strategy. One of the major criticisms of the stochastic spread is that this method is not very practical in reality. The number of pairs found after running this method is minimal (Huurman, 2012). Do et al. (2006) claim that the stochastic spread method captures mean reversion and it helps forecasting with convergence time, while Mudchanatongsuk et al. (2008) propose a stochastic model using the Ornstein-Uhlenbeck process and provide a maximum-likelihood model for empirical estimation. Kanamura et al. (2010) propose a price spread model using stochastic method. Applying this model into the energy futures market, they show that the profitability of the natural gas futures is higher than WTI crude oil and heating oil. In addition to the distance method and the stochastic method, Huck (2009) propose a combined forecast model that ranks the stocks. This model helps identify underperforming and outperforming stocks proved by an empirical analysis of S&P 100 index stocks.

2.2 Cointegration Method

The cointegration method is based on cointegration theory in time series and is currently the most widely used method by practitioners. The classical regression model is built on the basis of the stationary variables. Many financial time series are non-stationary, which cannot be used in the classical regression models since it potentially leads to spurious regressions. However, cointegration describes the long-term equilibrium between non-stationary time series. The equilibrium refers to the stationarity of the linear combination of these non-stationary time series. That is, the mean and variance of the linear combination is constant and the covariance is only related to the time interval. This assumes that some financial time series show convergence due to the fact that they are affected by common economic factors (Vidyamurthy, 2004).

Engle and Granger (1987) establish a method of cointegration, which provides an effective way for modelling non-stationary time series. Although some financial time series are non-stationary time series, the linear combination of them may be stationary. That is to say, despite the various financial time series having long-term fluctuations respectively, each series' moments, such as the mean, variance and covariance, will change over time, some linear combinations of them are stable. Thus, there exists a long-term equilibrium relationship between these non-stationary time series (Engle and Granger, 1987).

Compared with the distance method and the stochastic spread method, a large number of empirical studies have shown that the cointegration method is better in terms of the stability of trading pair selection (Vidyamurthy, 2004; Lin et al., 2006; Galenko et al., 2007; Schmidt, 2008; Puspaningrum et al, 2009; Chiu and Wong, 2012). Vidyamurthy (2004) points out that the stock price is often assumed to be a random walk, or a non-stationary time series. Hence, the cointegration method is appropriate to analyse two stocks' long-term equilibrium by taking the logarithm of the prices. Lin et al. (2006) apply the cointegration method to the pairs trading strategy with a minimum profit condition and show that pairs trading cannot be protected by this condition. Meanwhile, Galenko et al. (2007) explore the new properties of cointegration and form the pairs trading strategy. They show the evidence of the empirical results from 2001 to 2006 and claim that the profits from the portfolio are positive. Schmidt (2008) tests daily stock prices from 2002 to 2007 for pairs trading using the cointegration method and shows that the pairs found in the sample can be profitable. For the trading pair selection, Schmidt (2008) utilizes the Johansen test for the cointegration relationship between the two stocks in the pair. Puspaningrum et al. (2009) investigate the trading trigger criteria, the length of training and trading period, and the optimal preset threshold while maximizing the profits, while Chiu and Wong (2012) apply a dynamic cointegration model for pairs trading to show that there is a relationship between cointegration and statistical arbitrage.

It will be recalled that one of the disadvantages of the distance method is that the measure for the distance between the two stocks is static, which is obviously inconsistent with the time-varying stock market. That is why the distance method works well with short training and trading periods. However, the cointegration method looks for a long-term relationship between cointegrated time series which increases the stability. Alexander and Dimitriu (2005) argue that the cointegration method increases the stability of the pairs trading strategy to overcome issues like volatility clustering.

Certain economic variables have a long-run equilibrium relationship. This equilibrium relationship reflects the fact that there is no destruction of the intrinsic mechanism of equilibrium in the economic system. If the variables are disturbed in a certain time period causing the deviation from its long-term equilibrium, then the system mechanism will adjust to bring it back to the equilibrium state in the next period (Vidyamurthy, 2004). We call a stationary time series an *I(0)* process. For a non-stationary time series, if its first order difference becomes stationary, then we call the original time series an *I(1)* process. The linear combination of *I(1)* processes are also *I(1)* (Granger, 1986). Consider a set of time series $\{X_t\}$, $\{Y_t\}$; they are cointergreated if the following conditions are met:

1. $\{X_t\}$ and $\{Y_t\}$ are both I(1) processes which indicate they are non-stationary but their first order differences are stationary.
2. There is a nonzero constant a , so that we can find the relationship: $Y_t - aX_t = \epsilon_t \sim I(0)$, where ϵ_t is stationary.

Suppose the long-term equilibrium is described by the following:

$$Y_t = \mu + aX_t + \epsilon_t \qquad (1)$$

where a is a nonzero constant, μ is the long-term equilibrium and ϵ_t is a noise term also known as the cointegration error.

The cointegration error can be represented by:

$\epsilon_t = Y_t - \mu - aX_t$ (2)

The cointegration error ϵ_t represents the deviation from the long-term equilibrium.

At time $T = t - 1$, there are three situations: (1) Y equals the equilibrium value, $Y_{t-1} = \mu + aX_{t-1}$; (2) Y is smaller than the equilibrium value, $Y_{t-1} < \mu + aX_{t-1}$; (3)Y is greater than the equilibrium value, $Y_{t-1} > \mu + aX_{t-1}$.

At time $T = t$, we have $\Delta Y_t = a\Delta X_t + \vartheta_t$,

where $\Delta Y_t = Y_t - Y_{t-1}$, $\Delta X_t = X_t - X_{t-1}$ and $\vartheta_t = \epsilon_t - \epsilon_{t-1}$

If at $T = t - 1$, Y is smaller than the equilibrium value, $Y_{t-1} < \mu + aX_{t-1}$, then ΔY_t is larger compared to the third situation mentioned above. That is, if $Y_{t-1} > \mu + aX_{t-1}$ then ΔY_t is smaller.

If the equation (1) correctly indicates the long-term equilibrium relationship between X and Y, then the deviation of Y from its equilibrium is essentially temporary. Thus, an important assumption is that the cointegration error ϵ_t is an *I(0)* process. Obviously, if ϵ_t is a random term, then any deviation from the equilibrium will be accumulated and cannot be eliminated. As mentioned above, in equation (1), both X_t and Y_t are *I(1)* processes and the linear combination of the two, cointegration error ϵ_t, is an *I(0)* process. Then, we say that X_t and Y_t are cointegrated.

Consider that $X_{1i}, X_{2i}, \dots, X_{ki}$ are *I(1)* processes.

There is a vector $a = (a_1, a_2, \dots, a_k)$,

$Z_t = aX_t{}'$ ~ *I(d-b),*

where $b > 0, X_t = (X_{1i}, X_{2i}, \ldots, X_{ki})'$,

Then, we say that $X_{1i}, X_{2i}, \ldots, X_{ki}$ are cointegrated of order (d, b), denoted by $X_t \sim CI(d, b)$. Vector a is called a cointegrated vector. If both of the two time series are integrated, they are cointegrated only when they have the same order of integration. If they don't have the same order of integration, it is impossible to cointegrate. If there are three or more variables with a different order of integration, the linear combination of them may have a lower order of integration.

To test for cointegration of two time series, there are two major methodologies: the Engle-Granger test (1987) and Johansen test (1988). Chiarella et al. (2008) point out the potential problems caused by using the Engle-Granger test for cointegration. They argue that the Engle-Granger test may result in incorrectly rejecting cointegration between the two time series and incorrectly accepting when the two have no cointegration relationship. However, Johansen (1988) proposes an approach using the maximum likelihood strategy to estimate cointegrating vectors. In our empirical analysis, we use the Johansen test.

Consider the following case:

$y_t = a_1 y_{t-1} + \varepsilon_t$

First difference:

$\Delta y_t = (a_1 - 1) y_{t-1} + \varepsilon_t$

If $(a_1 - 1)$ is zero we then conclude that $\{y_t\}$ has a unit root which indicates a non-stationary process. However, if $a_1 - 1 \neq 0$ then $\{y_t\}$ is a stationary process.

If we now generalize to the two variable case:

Consider two stock prices $X_{s1,t}$ and $X_{s2,t}$ as following a simple vector auto-regression (VAR) model:

$$X_{s1,t} = a_{11}X_{s1,t-1} + a_{12}X_{s2,t-1} + \varepsilon_{s1,t}$$

$$X_{s2,t} = a_{21}X_{s2,t-1} + a_{22}X_{s1,t-1} + \varepsilon_{s2,t}$$

Both $X_{s1,t}$ and $X_{s2,t}$ are non-stationary.

Applying lag operator L and rearranging we get:

$$(1 - a_{11}L)X_{s1,t} - a_{21}LX_{s2,t} = \varepsilon_{s1,t}$$

$$-a_{21}LX_{s1,t} + (1 - a_{22}L)X_{s2,t} = \varepsilon_{s2,t}$$

We then rewrite them in matrix form:

$$\begin{bmatrix} 1 - a_{11}L & -a_{21}L \\ -a_{21}L & 1 - a_{22}L \end{bmatrix} \begin{pmatrix} X_{s1,t} \\ X_{s2,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{s1,t} \\ \varepsilon_{s2,t} \end{pmatrix}$$

By applying Cramer's rule we get the following:

$$X_{s1,t} = \frac{(1 - a_{11}L)\varepsilon_{s1,t} + a_{12}L\varepsilon_{s2,t}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$

$$X_{s2,t} = \frac{a_{21}L\varepsilon_{s1,t} + (1 - a_{11}L)\varepsilon_{s2,t}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$

By taking the differences we get:

$$\Delta X_{s1,t} = X_{s1,t} - X_{s1,t-1} = (a_{11} - 1)X_{s1,t-1} + a_{12}X_{s2,t-1} + \varepsilon_{s1,t}$$

$$\Delta X_{s2,t} = X_{s2,t} - X_{s2,t-1} = a_{21}X_{s1,t-1} + (a_{22} - 1)X_{s2,t-1} + \varepsilon_{s2,t}$$

We then rewrite them in matrix form:

$$\begin{pmatrix} \Delta X_{s1,t} \\ \Delta X_{s2,t} \end{pmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12}L \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{pmatrix} X_{s1,t-1} \\ X_{s2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{s1,t} \\ \varepsilon_{s2,t} \end{pmatrix}$$

$$\tilde{\Pi} = \begin{bmatrix} a_{11} - 1 & a_{12}L \\ a_{21} & a_{22} - 1 \end{bmatrix}$$

If $\tilde{\Pi} = 0$, then $\{X_{s1,t}\}$ and $\{X_{s2,t}\}$ are not cointegrated. Therefore, we test for the rank of the matrix $\tilde{\Pi}$. If $rank(\tilde{\Pi}) \neq 0$, then $\{X_{s1,t}\}$ and $\{X_{s2,t}\}$ are cointegrated. The rank of the matrix $\tilde{\Pi}$ is the number of cointegrating vectors.

If $a_{12} \neq 0$ and $a_{21} \neq 0$, then by normalizing the cointegrating vector with respect to $X_{s1,t-1}$:

$$\Delta X_{s1,t} = -\frac{a_{12}a_{21}}{1-a_{22}}(X_{s1,t-1} - \frac{1-a_{22}}{a_{21}}X_{s2,t-1}) + \varepsilon_{s1,t}$$

$$\Delta X_{s2,t} = a_{21}(X_{s1,t-1} - \frac{1-a_{22}}{a_{21}}X_{s2,t-1}) + \varepsilon_{s2,t}$$

The long-term equilibrium is $X_{s1,t-1} - \frac{1-a_{22}}{a_{21}}X_{s2,t-1}$.

Note that both $\{X_{s1,t}\}$ and $\{X_{s2,t}\}$ are non-stationary; the linear combination $X_{s1,t-1} - \frac{1-a_{22}}{a_{21}}X_{s2,t-1}$ is stationary, with the normalized CV by $[1, -\frac{1-a_{22}}{a_{21}}]$.

(Figure 1 here)

Figure 1 shows an example of two cointegrated shares - Boston Properties (BXP) and AvalonBay Communities (AVB) - and their cointegration error. We then compute the mean of the cointegration error. The preset upper bound is two standard deviations above the cointegration error mean, while the lower bound is two standard deviations below the mean. A trade is opened whenever the cointegration error crosses the preset threshold. Normally, when the cointegration error bounces back to the mean we close the trade. However, when a trading period ends before the cointegration error crosses the mean we force the trade to close at the end of the trading period. For example, on April 5, 2012, the cointegration error of the two stocks (BXP and AVB) is higher than the upper bound, so the trade is opened by selling BXP and buying AVB. At t=September 6, 2012, the cointegration error crosses the mean, however, trading period 3 ends at May 30, 2012 so the trade is closed at the end of trading period 3 by taking the opposite position. Another incomplete trade opens during trading period 7 and it ends at the end of the trading period (May 30, 2014).

3. Pairs Trading Profits

3.1 Measuring Profits and Returns

Suppose two stock prices are non-stationary and their first differences are stationary, denoted by $P_{S1} \sim I(1)$ and $P_{S2} \sim I(1)$. If the two prices are cointegrated, then there is a vector $(1, \beta)$ such that a cointegration relationship can be constructed as follows

$$P_{S1,t} - \beta P_{S2,t} = \mu + \epsilon_t \ ,$$

where ϵ_t represents cointegration errors and μ is the long-term equilibrium of the cointegration.

Denote N_{S1} and N_{S2} as the quantity of the two stocks respectively. When the cointegration error is higher than the preset threshold, a trade opens which we short N_{S1} of $S1$ stocks and long N_{S2} of $S2$ stocks. The trade closes when the cointegration error bounces back to zero. Then, we buy N_{S1} of $S1$ stocks and sell N_{S2} of $S2$ stocks to close the trade. Therefore, the profits, π , in this trade are:

$$\pi = N_{S2}\left(P_{S2,t_c} - P_{S2,t_o}\right) + N_{S1}\left(P_{S1,t_o} - P_{S1,t_c}\right)$$

$$= \beta\left(P_{S2,t_c} - P_{S2,t_o}\right) + \left(P_{S1,t_o} - P_{S1,t_c}\right)$$

$$= \beta\left(P_{S2,t_c} - P_{S2,t_o}\right) + [\left(\mu + \epsilon_{t_o}\right) + \beta P_{S2,t_o}\right) - \left(\mu + \epsilon_{t_c}\right) + \beta P_{S2,t_c}\right)$$

$$= (\epsilon_{t_o} - \epsilon_{t_c}),$$

where t_o and t_c are times for opening and closing the pair trade.

3.2 Determinants of Pairs Trading Profits

Prior studies find that many economic variables are related to return comovement. Engelberg et al. (2009) argue that their pairs trading strategy may be driven by both the delay in information diffusion and the short-term liquidity provision. Meanwhile Barberis et al. (2005) suggest that investors' trading behaviour contributes to stock comovement.

Thus, the variables we include in our empirical analysis are as follows:

Social media volume. Whether by accident or by design, a firm's tweets may affect the level of attention investors are paying to firm news, even if such news is communicated elsewhere. For example, if a firm's tweets keep the firm at the forefront of the investor's mind, even unrelated tweets by the firm may increase the likelihood an investor is 'tuned in' and notices (or even seeks out) earnings news from a variety of sources and considers trading. A firm's tweets during the days following earnings news may also increase trading by prompting investors who found it inconvenient to trade when they first encountered the earnings news. Note that the general attention channel could be operative whether or not the firm is aware of its effect. Wysocki (1998) illustrates that message volume forecasts next-day trading volume and abnormal returns, while Blankespoor et al. (2014) show that companies' tweets that link to press releases will increase stock liquidity. We include the total Twitter volume for two firms in a pair during the trading period.

Earnings correlation. Engelberg et al. (2009) suggest that the correlation between the earnings of the two stocks is linked to the stock return comovement. Hence, we include earnings correlation as one of the independent variables in our analysis. To construct the earnings correlation variables, we first calculate the return on equity ratio, ROE, as companies' earnings per share divided by the book value of equity per share. Second, we compute the correlation of the quarterly ROE ratio between the two stocks in a pair for the corresponding trading period.

Earning surprise correlation. To capture the correlations of two stocks' cash flow news, we calculate the correlation coefficient of their earnings surprises. We measure the quarterly earnings surprises for a stock as the IBES actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the absolute value of the analyst forecast of earnings. Earning surprise correlation is then computed as the correlation of the quarterly earnings surprises between a pair of stocks.

Industry. Kumar and Lee (2006) show that firms in the same industry are more likely to have the return comovement since within the same industry, firms have similar business conditions, cash-flow, and discount-rate shocks. Therefore, we include dummy variables for a stock pair that equals 1 if they are from the same industry and 0 otherwise. Bhojraj et al. (2003) claim that using a Global Industry Classifications Standard (GICS) code is better than a Standard Industrial Classification (SIC) code or the North American Industry

Classification System (NAICS) code for explaining stock return comovements. Therefore, we use GICS to construct this variable. We create an industry dummy variable that equals one if the two stocks in the pair have the same eight-digit GICS code and zero otherwise.

Size. Fama and French (1993) suggest the expected stock returns are related to firm level characteristics such as firm size. If two firms are in the same market capitalization category, they are exposed to similar risk factors. In addition, investors' trading behaviour is also linked to firm size (Barberis and Shleifer, 2003). Regardless of the interpretation of size as a risk factor or a mispricing factor, prior literature suggests that firms with a similar size tend to comove in stock returns. We group all the firms in our sample into two categories, large cap and small cap. We define large cap stocks as those stocks with above median market capitalization, and the rest of the stocks are small cap stocks. We then construct a dummy variable equal to one if two firms in the pair are in the same market capitalization category, zero otherwise.

Geographic location. Ivkovic and Weisbenner (2005) show that investors prefer local stocks due to their familiarity. Pirinsky and Wang (2006) show the evidence that firms in the same geographic location are subject to similar returns. Therefore, if two stocks in the pair are located in the same geographic location, they are exposed to common shocks. We utilize Compustat's state codes to create a location dummy variable, which equals one if two firms are located in the same state and zero otherwise.

Exchange listing. Huddart et al. (1999) show that public disclosure requirements by exchanges affect firms making decisions for listing. In other words, if two firms are listed in the same exchange, they will deal with similar exchange requirement issues. Hence, exchange listing plays a role in explaining stock return comovement. We construct a dummy variable that equals to one if two stocks are listed in the same stock exchange, and zero otherwise.

Trading volume. Amihud (2002) develops a liquidity measure using trading volume and shows that the measure positively impacts expected stock return, while Campbell and Mei (1993) provide empirical evidence that liquidity highly corresponds to trading volume-related returns. Consequently, two stocks' returns could comove with each other because of similar liquidity situations. Combining the ideas of using trading volume as a proxy for liquidity, we calculate the correlation between two firms' trading volumes, and include it as a determinant of pairwise return correlation.

4. Data and Sample Selection

For our analysis, we use both financial data and social media data.

4.1 Financial Data

Our sample consists of 60 months of S&P500 stocks, from December 2009 to December 2014. We constructed the training period to be 12 months (December 2009 to December 2010) to identify cointegration relationship between two stocks. We start the first trading period from December 2010. The duration for each trading period is six months. Therefore, from December 2010 to December 2014, there are a total of 8 trading periods in our sample.

We use data from the Center for Research in Security Prices (CRSP), COMPUSTAT, and Institutional Brokers' Estimate System (IBES).

In addition, firms chosen must satisfy the following selection criteria:

(1) Firms are present in both the Standard & Poor's COMPUSTAT File and the CRSP Daily Returns File;

(2) Quarterly and annual earnings release dates, and reported values, are available from Standard & Poor's COMPUSTAT File; and

(3) Analysts' forecasts of quarterly and annual earnings are available from the IBES database immediately prior to the earnings release and corresponding actual earnings data are available as well.

4.2 Social Media Data

We obtain social media data for our sample stocks from Quandl Financial and Economic Data (Quandl). Quandl offers access to financial, economic, and social datasets from multiple sources. First, we carefully filter all the relevant tweets for a specific company using companies' ticker symbols. We include the original tweets that contains companies' ticker symbols, retweets of the company, and URL links of the company. We set filters to obtain all the relevant tweets from a Quandl search engine. From Quandl, we obtain the daily total message volume from Twitter and StockTwits from December 2010 to December 2014[18]. Firms' Twitter and StockTwits data from Quandl allow us to track the

[18] The Twitter data in our sample matches the entire trading periods from December 2010 to December 2014. During the training period December 2009 to December 2010, no Twitter data is needed for constructing pairs portfolios.

tweets 20 days before and 20 days after the earning announcement date. The total message volume is the summation of the total number of the messages from both of the platforms. By convention, Twitter and StockTwits usually include the stock symbol prefixed by a dollar sign (e.g., $AAPL for Apple Inc) in discussions about a stock.

In addition, to ensure that a prefixed dollar sign plus the stock ticker symbol would be a unique filter to obtain the relevant messages for each stock, we randomly selected 30 tweets from the time period 2010-2014. To be acceptable, at least 50% of tweets have to be related to the company, e.g., mentioning the company or its financial situation. We remove any company from our sample that does not pass this test. As a result, 1.6% of stocks have been removed by this filter.

(Table 1 Here)

Table 1 provides summary statistics for the sample, including variables such as portfolio return, logarithm of the Twitter volume, trading volume correlation, ROE correlation, earning surprise correlation, industry dummy variable, company size dummy variable, company location dummy variable, and exchange listing dummy variable. Our sample consists of 41 pairs[19] of stocks over the sample period December 2010 to December 2014. On average, the pair portfolio semi-annual[20] return is 3.16% and the standard deviation is 4.96%. The table presents the summary statistics of the potential determinants of the pairs

[19] The rest of the stocks do no pass the Johansen test and AR(1) fitting.
[20] The average is calculated based on trading period. We have 8 six-month trading periods in total from December 2010 to December 2014.

trading strategy returns. The mean value of trading volume correlation is 0.286, the highest negative correlation is -0.070, and the highest positive correlation is 0.840. The average ROE correlation between two stocks in a pair is 0.158 and the average earnings surprise correlation is 0.056. 43.90% of our sample pairs are from the same GICS code industry. The mean of the size dummy variable is 56.10%, which suggests that about 56.10% of the firm-pairs are from the same size category (large cap or small cap) in terms of their market capitalization. The mean value of location dummy variable is 7.32%, which suggests that 7.32% of our sample pairs are from the same state. About 76% of the stock pairs belong to the same exchange.

5. Empirical Results

5.1 Trading Pairs Selection

The key for a pairs trading strategy is trading pair selection. We utilize the cointegration method to identify trading pairs. To determine whether a cointegration relationship exists between two stocks in the pair, we use the technique developed by Johansen (1988). The following steps are executed for our pair selection using our sample:

Step 1: Run cointegration Johansen tests for each stock in the S&P500 to find the pairs that are significantly cointegrated during the training period. Remove the pairs if they are not significantly cointegrated using the Johansen test.

Step 2: Obtain residuals (cointegration errors) ϵ_t of the pairs from the equation:

$P_{S1,t} - \beta P_{S2,t} = \mu + \epsilon_t$,

where $P_{S1,t}$ and $P_{S2,t}$ are the prices of stock S_1 and S_2 respectively and μ is a constant.

Step 3: For the cointegration errors we obtain from Step 2, analyze whether an AR(1) model is appropriate. Remove the pairs which cannot be fitted with an AR(1) model.

To combine the Johansen test and cointegration errors, first, the Johansen test is used to make sure that the pairs of stocks have a long-run equilibrium. Second, obtain the cointegration errors ϵ_t and test if ϵ_t is an AR(1) process in order to confirm the cointegration. By doing the steps above, the type 2 spurious regression problem in the Engle-Granger method will be avoided and we can confirm that the pairs are cointegrated after step 3.

(Table 2 here)

Using 12-month data from December 2009 to December 2010, we identify 41 pairs from S&P 500 companies as shown in Table 2 by completing all three steps mentioned above. Column 1 in Table 2 reports the p-value from Johansen test for cointegration relationship between the two stocks. The pairs selected in our sample are at a 5% significant level. The null hypothesis of the Johansen test is "no cointegrating relationships between the two". Column 2 from Table 2 shows the p-value from the Dickey-Fully test of the cointegration errors. We obtain cointegration errors (residuals series) from each pair and do the Dickey-Fuller test for unit root. We reject the null hypothesis at the 5% level and conclude that the cointegration errors fit with an AR(1) model. Table 2 Column 3 is the correlation coefficient between the two stocks in each pair. The average correlation coefficient across pairs is 0.793. We note that not all pairs in our sample are in the same sector. A pairs trading strategy is built upon a high degree of similarity of the two stocks. This is in line with the Law of One Price (LOP) in economics. Gatev et al (2006) attempt to use the LOP theory to explain the profitability of pairs trading. The LOP states: "Assuming other conditions remain unchanged, the price of any homogeneous item should be equal in the case of an efficient market." Gatev et al (2006) believe that constructing pairs in the same industry is necessary to for pairs trading strategy to be profitable. Therefore, if two stocks are in the same industry, they will be subject to the LOP. However, the industry constraint is not necessary for identifying cointegration relationships. Von Hagen (1989) finds cointegration relationships among different commodities using data from 1900-1986. Similarly, Pindyck and Rotemberg (1990) confirm that unrelated commodity prices have a tendency to be cointegrated. Despite the LOP theory, these empirical studies suggest that cointegration relationships could exist without the industry constraint. Thus, in our pairs selection process, the restriction of pairs within sectors will not be applied.

5.2 Determining Training Period and Trading Period

The durations of the training and trading period are very important with the cointegration method. For the training period, it needs to cover a sufficient amount of time to let the two stocks demonstrate the cointegration relationship. However, we need enough range of time to establish trading periods. As for the trading period, we need the length of the trading period to be appropriate to open and close trades. We choose a 12-month training period from December 2009 to December 2010 and 8 six-month trading periods from December 2010 to December 2014.

5.3 Computing Profits and Returns

In deriving the profits and returns, we assume that all trades are completed before the end of period T and that there are no transaction costs. If a trade has not been closed by the end of the trading period, we force it to be closed at the end of the period. In our empirical analysis, we utilize a realistic measure of return used in Hong and Susmel (2003). We define:

$$Total\ Profit\ (\pi_{i,t}) = N_{long}\left(P_{long,i,t_c} - P_{long,i,t_o}\right) + N_{short}\left(P_{short,i,t_o} - P_{short,i,t_c}\right)$$

$$Average\ Return\ (R) = \frac{\sum_{t=1}^{T} r_{i,t}}{T},$$

where $r_{i,t} = \frac{\pi_{i,t}}{P_{long,i,t}+P_{short,i,t}}$ is return for portfolio *i* at time *t*. $P_{long,i}$ is the stock value in the long position, $P_{short,i}$ is the stock value in the short position. t_o and t_c are times for opening and closing the pair trade. *T* is total number of the trades.

5.4 Empirical Analysis for Social Media Effect

To explore the economic drivers of the pairs trading profits, we now examine the social media effect on the trading profits based on the pairs portfolio constructed in section 5.2. Social media and microblogs are fundamentally changing interactions between investors and firms (Gallaugher and Ransbotham, 2010). Investors' attention puts upward pressure on the stock price (Kumar and Lee, 2006; Barber and Odean, 2008; Barber et al., 2009; Burch et al., 2014; and Hvidkjaer, 2008). If investors implement a pairs trading strategy to form portfolios, we would expect increasing investors' attention to positively affect the returns of the portfolios. To test this hypothesis, we estimate the following regression model:

$$R_{i,t} = \alpha + \beta * \log(Twtvol)_{i,t} + \rho * \mathrm{TwtVolCorr}_{i,t} + \mu * UEcorr_{i,t} + \gamma * ROEcorr_{i,t}$$

$$+\theta * VOLcorr_{i,t} + \sum \beta_k * X_{i,t,k} + \varepsilon_{j,q}$$

where $R_{i,t}$ is the portfolio *i*'s return at trading period *t*. $log(Twtvol)_{i,t}$ is the logarithm of the total Twitter volume of the two stocks in portfolio *i* at trading period *t*. $\mathrm{TwtVolCorr}_{i,t}$ is the Twitter volume correlation between the two stocks in portfolio *i* at trading period *t*. $UEcorr_{i,t}$ is the unexpected earnings correlation between the two stocks in portfolio *i* at trading period *t*. $ROEcorr_{i,t}$ is the return on earnings correlation between the two stocks in portfolio *i* at trading period *t*. $VOLcorr_{i,t}$ is the trading volume correlation between the two stocks in portfolio *i* at trading period *t*. $X_{i,t,k}$ is a vector of the control variables including sector dummies, location dummies, companies' size dummies, exchange listing dummies, and year-month dummies.

(Table 3 here)

We estimate the regression of pairs portfolio returns on the variables we hypothesize as predicting pairwise correlations. In Table 3, we present the results for the OLS pooled regression of portfolio returns on the determinants using data from December 2010 to December 2014. Our first trading period following the 12-month training period starts in December 2010. To account for autocorrelation, we adjust the standard errors by three-way clustering by the id number of the first stock, the id number of the second stock, and year. The three-way clustering method is based on Cameron et al. (2010).

Most of the variables explain the pairs portfolio returns in the ways we expect, and they have statistically significant coefficients. In Table 3, Column 1, we include the determinants that have been tested in the literature: ROE correlation, earnings surprise correlation, trading volume correlation and control variables. All the three variables that capture the correlation in earnings and trading volume load up statistically significantly. ROE correlation coefficient is 0.0485 (t-statistic 6.94) and the correlation of earnings surprise correlation coefficient is 0.0684 (t-statistic 7.04). Trading volume correlation coefficient is 0.0058 and it is statistically significant at 1% level (t-statistic 3.81). The adjusted R-squared of this regression is 2.69%. This indicates that pairwise earnings correlations and trading volume correlation explain less than 3% of the total variation in the pairs trading portfolio returns.

In Table 3, Column 2, we include the social media variable, the logarithm of the total Twitter volume of the two stocks in the pair, in the regression. The logarithm of the total Twitter volume shows up as positive and statistically significant. The coefficient on log(Twitter Vol) is 0.0654 (t = 4.12), which suggests that if the logarithm of the two stocks' Twitter volume increases by one unit and their portfolio return is going up by 6.54%. For other three variables both the magnitudes and the significant levels are the same with Column 1. The adjusted R-squared also increases significantly when we include the logarithm of the Twitter volume (the adjusted R-squared in Column 2 is 6.29% compare to 2.69% from Column 1).

In Column 3 of Table 3, we use a different social media variable, Twitter volume correlation, in our regression analysis. It shows up as statistically insignificant. The other three determinates ROE correlation, earnings surprise correlation and trading volume correlation are still consistent with Column 1 and Column 2 in both magnitude and significant level.

In Column 4, to further address the social media effect, we include both logarithm of the Twitter volume and Twitter volume correlation into our regression model. The logarithm of Twitter volume is consistent with Column 2, with a coefficient of 0.0619, significant at 1% (t=3.97). Twitter volume correlation is statistically insignificant. All three other determinants, ROE correlation, earnings surprise correlation and trading volume correlation, are statistically significant and the magnitudes are consistent with Columns 1, 2, and 3. Overall, when we include all the determinants in Column 4, there is substantial variation in the pairs portfolio returns that cannot be explained; when we include all the explanatory variables, the adjusted R-squared is 7.21%.

Since each pairs portfolio contains a long position stock and a short position stock, the logarithm of the total Twitter volume of each portfolio also carries both stocks' number of Tweets. We decompose this independent variable into the logarithm of the total Twitter volume of the long stock at time *t*, $log(Totvol\_Long)_{i,t}$, and the logarithm of the total Twitter volume of the short stock at time *t*, $log(Totvol\_Short)_{i,t}$.

We then estimate the following regression model:

$$R_{i,t} = \alpha + \beta * log(Totvol\_Long)_{i,t} + \delta * log(Totvol\_Short)_{i,t} + \rho * \mathrm{TwtVolCorr}_{i,t} + \mu * UEcorr_{i,t} + \gamma * ROEcorr_{i,t} + \theta * VOLcorr_{i,t} + \sum \beta_k * X_{i,t,k} + \varepsilon_{j,q}$$

The results are presented in Table 3 Column 5. For the long position stocks, we find a coefficient of 0.0724 for the logarithm of the total Twitter volume and it is statistically significant at 1% (t-statistic 4.07). This is consistent with Barber and Odean (2008) which suggests investors' attention may put a positive pressure on stock price. However, the coefficient of the logarithm of the total Twitter volume of the short stock is not statistically significant. This indicates the attention of short position stocks does not contribute to the pairs trading profits.

6. Conclusion

The purpose of this study was to develop a method for selecting trading pairs in a pairs trading strategy and measure the portfolio returns in order to estimate the social media impact on the returns. We employ the Johansen test (1988) to identify cointegration relationship between stock pairs. Trading pair selection is limited to S&P 500 companies. Gatev et al, 2006 suggest that according to the LOP, imposing industry constraint when selecting pairs is necessary for pairs trading to be profitable. However, we do not constraint our search for pairs in the same industry. Applying the cointegration method after a Johansen test, we first construct the cointegration errors from the long-term equilibrium between each of the pairs. By running augmented Dickey-Fuller tests, we are able to discover whether the cointegration errors are stationary. To be selected as trading pairs, their cointegration errors have to be stationary so that we can conclude they are cointegrated. The study finally identifies 41 trading pairs which are cointegrated using both daily prices from the training period December 2009 to December 2010. After forming the pairs, we measure the portfolio returns by the method of Hong and Susmel (2003).

We explore the economic drivers of the pairs portfolio returns. We provide several pieces of evidence that are consistent with the literature for example, ROE correlation between the two stocks in a pair positively affects the portfolio returns. Earnings surprise correlation and trading volume correlation between the two stocks in a pair also are economic drivers of the portfolio returns. Finally, we find evidence that suggests social media has a positive impact on the portfolio returns, using the logarithm of the total Twitter volume of the two stocks in a pair as a proxy for social media attention. This study introduces social media as one of the economics driver of the portfolio returns. Future research could look at the implementation of social media in trading strategies.

REFERENCES

Alexander Carol, Dimitriu Anca. 2005. Indexing and statistical arbitrage: tracking error or cointegration? Journal of Portfolio Management, 10, 1-19

Amihud, Yakov. 2002. Illiquidity and stock returns: cross-section and time-series effects, Journal of Financial Markets, 5, 31-56.

Barber, Brad M., and Terrance Odean. 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies, 21(2), 785–818.

Barber, Brad M., Terrance Odean, and Ning Zhu. Forthcoming. Do Retail Trades Move Markets?. Review of Financial Studies, 22(1), 151–186.

Barberis, Nicholas, and Andrei Shleifer. 2003. Style investing, Journal of Financial Economics. 68, 161-199.

Barberis, Nicholas, Andrei Shleifer, and Jeffery Wurgler, 2005, Comovement, Journal of Financial Economics. 75, 283-317.

Bhagwat, Vineet, and Timothy R. Burch. 2014. Pump it Up? Tweeting to Manage Investor Attention to Earnings News. Working paper. University of Oregon and University of Miami

Bhojraj, Sanjeev, Charles M. Lee, and Derek K. Oler, 2003, What's my line? A comparison of industry classification schemes for capital market research, Journal of Accounting Research. 41, 745-774.

Blankespoor, Elizabeth, Gregory Miller, and Hal White. 2014. The Role of Dissemination in Market Liquidity: Evidence From Firms' Use of Twitter. The Accounting Review. 89, 79-112

Boyer, H. Brian, 2010, Style-Related Comovement: fundamentals or labels?. Journal of Finance. 66 (1), 307-332

Burch, Timothy, Douglas Emery, and Michael Fuerst. 2014. Who Moves Markets in a Sudden Market-Wide Crisis? Evidence from Nine-Eleven. Journal of Financial and Quantitative Analysis, forthcoming.

Cameron, A. Colin, Jonah Gelbach and Douglas Miller, 2010, Robust inference with multi-way clustering, Journal of Business and Economic Statistics.

Campbell, John Y., and Jianping Mei, 1993, Where do betas come from? Asset price dynamics and the sources of systematic risk, Review of Financial Studies 6(3), 567-592.

Chiarella, C., S. Gao and M. Stevenson, 2008, Resolving the Price-Earnings Puzzle and Related Econometric Issues. Working paper. University of Sydney.

Chiu, Mei Choi and Hoi Ying Wong. 2012. Mean-Variance Asset-Liability Management: Cointegrated Assets and Insurance Liability. European Journal of Operational Research. 223(3), 785-793

Curtis, A., V. J. Richardson, and R. Schmardebech. 2014. Investor Attention and the Pricing of Earnings News. Working paper, University of Washington, University of Arkansas, and University of Missouri.

Do, Binh. and Robert W. Faff. 2009. Does simple pairs trading still work?. Financial Analysts Journal. 66, 83–95

Do, Binh, Robert W. Faff. and Kais Hamza. 2006. A new approach to modeling and estimation for pairs trading. In Proceedings of the Proceedings of 2006 Financial Management Association European Conference.

Elliott, Robert J., John Van Der Hoek, and William P. Malcolm. 2005. Pairs trading. Quantitative Finance. 5(3), 271–276.

Engelberg, Joseph, Pengjie Gao, Ravi Jagannathan, 2009, An anatomy of pairs trading: The role of idiosyncratic news, common information and liquidity. In Proceedings of the Third Singapore International Conference on Finance.

Engle, Robert F. and Clive W. Granger. 1987. Co-integration and Error Correction: Representation, Estimation, and Testing. Econometrica. 55 (2), 251-276.

Fama, Eugene F., and Kenneth R. French, 1992, The Cross Section of Expected Stock Returns, Journal of Finance 47, 427-465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3-56.

Galenko, Alexander., Elmira Popova, and Ivilina Popova. 2007. Trading in the Presence of Cointegration. The Journal of Alternative Investments. 15, 85–97

Gallaugher, John M. and Sam Ransbotham. 2010. Social media and customer dialog management at Starbucks. MIS Quarterly Executive. 9(4): 197-212.

Gao, Pengqin, 2010, Characteristic-based covariances and cross-sectional expected returns, working paper.

Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. 2006. Pairs trading: Performance of a relative value arbitrage rule. Review of Financial Studies 19, 797-827.

Gillespie, Timothy. and Chris Ulph. 2001. Pair trades methodology: a question of mean reversion, Proceedings of International Conference on Statistics, Combinatorics and Related Areas and the 8th International Conference of Forum for Interdisciplinary Mathematics, NSW.

Graham, John R., and Alok Kumar, 2006, Do dividend clienteles exist? Evidence on dividend preferences of retail investors, Journal of Finance. 61, 1305–1336

Granger, Clive J. 1986. Developments in the Study of Cointegrated Economic Variables. Oxford Bulletin of Economics and Statistics, 48, 213-228.

Green, T. Clifton, and Byoung-Hyoun Hwang, 2009, Price-based return comovement, Journal of Financial Economics 93, 37-50.

Greenwood, Robin, 2008, Excess comovement: Evidence from cross-sectional variation in Nikkei 225 weights, Review of Financial Studies 21, 1153-1186.

Hameed, Allaudeen, Randall Morck, Jianfeng Shen, and Bernard Yeung, 2010, Information, analysts, and stock return comovement, Working Paper, National University of Singapore.

Hirshleifer, David, and Siew Hong Teoh. 2003. Limited attention, information disclosure, and financial reporting. Journal of Accounting and Economics 36, 337–386.

Hong, Gwangheon and Raul Susmel. 2003. Pairs-Trading in the Asian ADR Market. Working paper. Saginaw Valley State University and University of Houston.

Huck, Nicolas. 2009. Pairs selection and outranking: An application to the S&P 100 index. European Journal of Operational Research.196(2), 819-825

Huck, Nicolas. 2013. The high sensitivity of pairs trading returns. Applied Economics. 20(14), 1301-1304.

Huddart, Steven, John S. Hughes, and Markus K. Brunnermeier. 1999. Disclosure

Requirements and Stock Exchange Listing Choice in an International Context. Journal of Accounting and Economics. 26, 237-269

Huurman, Nick. 2012. Pairs Trading: An Implementation of the Stochastic Spread and Cointegration Approach. Working paper. University of Amsterdam.

Hvidkjaer, Soeren.. 2008. Small trades and the cross-section of stock returns. Review of Financial Studies. 21, 1123–1151.

Israelsen, Ryan, 2009, Does correlated analyst coverage explain excess comovement? Working Paper, Indiana University.

Ivković Zoran and Scott Weisbenner. 2005. Local Does as Local is: Information Content of the Geography of Individual Investors' Common Stock Investments. Journal of Finance. 60, 267-306.

Johansen, Soren. 1988. Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control. 12(2), 231-254.

Kahneman, Daniel. 1973. Attention and Effort. Prentice-Hall, Englewood Cliffs, NJ.

Kanamura, Takashi, Rachev, Svetlozar T., and Fabozzi, Frank J. 2010. A profit model for spread trading with an application to energy futures. Working Paper Series. Karlsruhe Institute of Technology (KIT), Department of Economics and Business Engineering.

Kumar, Alok, and Charles M.C. Lee, 2006, Retail investor sentiment and return comovements, Journal of Finance 61, 2451-2486.

Lin, Yan-Xia, Michael McRae, and Chandra Gulati. 2006. Loss Protection in Pairs Trading through Minimum Profit Bounds: A Cointegration Approach. Journal of Applied Mathematics and Decision Sciences. 6,1-14.

Mankiw, N. Gregory, and Ricardo Reis. 2002. Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve. Quarterly Journal of Economics 117(4): 1295-1328.

Mori, Masaki and Alan J. Ziobrowski. 2011. Performance of Pairs Trading Strategy in the U.S. REIT Market. Real Estate Economics. 39(3): 409-428.

Mudchanatongsuk, Supakorn, James A. Primbs, and Wilfred Wong. 2008. Optimal pairs trading: A stochastic control approach. American Control Conference, 1035-1039.

Pindyck, Robert S., and Julio J. Rotemberg. 1990. The Excess Co-Movement of Commodity Prices. The Economic Journal. 100(402): 1173-1189.

Pirinsky, Christo A. and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, Journal of Finance 61, 1991-2015.

Puspaningrum, Heni, Yan-Xia Lin, and Chandra M. Gulati. 2009. Finding the Optimal Pre-Set Boundaries for Pairs Trading Strategy Based on Cointegration Technique. Journal of Statistical Theory and Practice. 4(3): 391-419.

Schmidt, A. D. 2008. Pairs Trading: A Cointegration Approach, Finance Honours Thesis University of Sydney.

Shiller, Robert, 1989, Comovements in stock prices and comovements in dividends, Journal of Finance 44, 719-729.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings, The Accounting Review 71, 289-315.

Vidyamurthy, Ganapathy. 2004. Pairs Trading, Quantitative Methods and Analysis. New York: John Wiley.

Von Hagen, Juergen. 1989. Relative Commodity Prices and Cointegration. Journal of Business and Economic Statistics. 7(4): 497-503.

Wysocki, Peter D. 1998. Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards. Working paper. University of Michigan.

Table 1 Summary Statistics

Table 1 is the summary statistics from our sample from December 2009 to December 2010 with 41 selected pairs portfolio. All the stocks been used in each portfolio are S&P 500 companies. *Portfolio Return* is the semi-annual return for each pairs portfolio. *Log(Twitter Vol)* is the logarithm of the total Twitter volume for the two stocks in each portfolio. *Trading Vol Corr* is the trading volume correlation between the two stocks in each portfolio. *ROE Corr* is the return on earnings correlation between the two stocks in each portfolio. *Earnings Surprise Corr* is the earnings surprise correlation between the two stocks in each portfolio. *Industry Dummy* equals to one if two stocks are in the same industry, zero otherwise. *Size Dummy* equals to one if two stocks are in the same category in terms of company's size, zero otherwise. *Location Dummy* equals to one if two stocks are located in the same state, zero otherwise. *Exchange listing Dummy* equals to one if two stocks are listed in the same exchange, zero otherwise.

Variable	Mean	Stdev	Min	Max
Portfolio Return	3.16%	4.96%	-13.38%	18.55%
Log(Twitter Vol)	7.632	1.219	5.094	11.263
Trading Vol Corr	0.286	0.220	-0.070	0.840
ROE Corr	0.158	0.350	-0.766	0.990
Earnings Surprise Corr	0.056	0.226	-0.601	0.994
Industry Dummy	43.90%	50.24%	0.000	1.000
Size Dummy	56.10%	50.24%	0.000	1.000
Location Dummy	7.32%	26.04%	0.000	1.000
Exchange Listing Dummy	75.61%	42.94%	0.000	1.000
Observation			328	

Table 2 Pairs Selection Summary

Table 2 presents the pairs been selected after cointegration method. The training period used to select the pair is from December 2009 to December 2010. Pair number represents each portfolio. Each portfolio contains two stocks. Column 1 shows the p-value for Johansen test for cointegration. Column 2 presents the Dickey-Fuller test results for the residuals from the pair. Column 3 is the correlation between the two stocks.

			{1}	{2}	{3}
Pair #	Companies' Name (Ticker)		P-value for Johansen Test	P-value for Dickey-Fuller test	Correlation
1	Accenture PLC (ACN)	Smucker J M Co (SJM)	0.012	0.011	0.812
2	Alliance Data Systems Corp (ADS)	Disney Walt Co (DIS)	0.007	0.020	0.819
3	American Tower Corp (AMT)	Prologis (PLD)	0.019	0.043	0.910
4	AMP Inc (AMP)	3M Co (MMM)	0.044	0.035	0.859
5	Amazon Com Inc (AMZN)	Fossil Group Inc (FOSL)	0.022	0.036	0.579
6	Avantgo Inc (AVGO)	AmerisourceBergen (ABC)	0.015	0.032	-0.483
7	Autozone Inc (AZO)	Grainger W W Inc (GWW)	0.017	0.033	0.748
8	Boston Properties Inc (BXP)	Avalonbay Communities Inc (AVB)	0.001	0.025	0.985
9	Crestar Financial Corp (CF)	Praxair Inc (PX)	0.004	0.027	0.921
10	Chipotle Mexican Grill Inc (CMG)	Tiffany & Co New (TIF)	0.013	0.032	0.679
11	Columbia Hca Healthcare Corp (COL)	Edison International (EIX)	0.003	0.023	0.818
12	Delta Air Lines Inc (DAL)	Southwest Airlines Co (LUV)	0.012	0.004	0.506
13	E.I. Du Pont De Nemours And Co (DD)	Linear Technology Corp (LLTC)	0.008	0.005	0.873
14	Department 56 Inc (DFS)	Nisource Inc (NI)	0.048	0.014	0.874
15	Dollar General Corp (DG)	L Brands Inc (LB)	0.034	0.013	-0.648
16	Discovery Communications Inc (DISCK	Union Pacific Corp (UNP)	0.026	0.006	0.563
17	Dr Pepper Snapple Group Inc (DPS)	Sandisk Corp (SNDK)	0.013	0.009	0.802
18	Energysolutions Inc (ES)	Diamond Offshore Drilling Inc (DO)	0.018	0.002	0.894
19	Essex Property Trust Inc (ESS)	Simon Property Group Inc New (SPG)	0.021	0.048	0.944
20	Genworth Financial Inc (GNW)	General Growth Pptys Inc New (GGP)	0.009	0.011	0.986
21	Genuine Parts Co (GPC)	Stanley Black & Decker Inc (SWK)	0.001	0.016	0.800
22	Homebase Inc (HBI)	Johnson & Johnson (JNJ)	0.022	0.017	0.765
23	Hershey Co (HSY)	Hasbro Inc (HAS)	0.003	0.016	0.563
24	Interpublic Group Cos Inc (IPG)	Ford Motor Co Del (F)	0.035	0.021	0.597
25	Mead Johnson Nutrition Co (MJN)	Apartment Investment & Mgmt Co (AIV)	0.042	0.031	0.935
26	Priceline Group Inc (PCLN)	Perrigo Co Plc (PRGO)	0.038	0.037	0.531
27	Parker Hannifin Corp (PH)	Cummins Inc (CUM)	0.048	0.050	0.900
28	Polymedica Corp (PM)	IBM (IBM)	0.038	0.004	0.824
29	Public Storage (PSA)	Norfolk Southern Corp (NSC)	0.041	0.049	0.876
30	S L Green Realty Corp (SLG)	Vornado Realty Trust (VNO)	0.050	0.007	0.960
31	Scripps Networks Interactive Inc (SNI)	Ingersoll-Rand Plc (IR)	0.047	0.025	0.976
32	Teradata Corp De (TDC)	Whole Foods Market Inc (WFMI)	0.043	0.023	0.757
33	T C C Industries Inc (TELC)	Omnicom Group Inc (OMC)	0.036	0.007	0.975
34	Time Warner Cable Inc (TWC)	CenturyLink Inc (CTL)	0.011	0.018	0.826
35	United Technologies Corp (UTX)	Waste Management Inc (WMI)	0.043	0.012	0.773
36	Visa Inc (V)	Tenet Healthcare Corp (THC)	0.042	0.007	0.796
37	Viacom Inc (VIA)	Dover Corp (DOV)	0.039	0.045	0.792
38	Verisign Inc (VRSN)	EMC Corp MA (EMC)	0.023	0.042	0.531
39	Ventas Inc (VTR)	Health Care Reit Inc (HCN)	0.031	0.051	0.924
40	Wyndham Intl Inc (WYN)	Starbucks Corp (SBUX)	0.039	0.006	0.840
41	Zions Bancorp (ZION)	Suntrust Banks Inc (STI)	0.033	0.020	0.975
	Average		0.026	0.023	0.740

Table 3 Regression Results

Table 3 presents the results from the following regression model:

$$R_{i,t} = \alpha + \beta * log(Twtvol)_{i,t} + \rho * TwtVolcorr_{i,t} + \gamma * ROEcorr_{i,t} + \mu * UEcorr_{i,t} + \theta * VOLcorr_{i,t} + \sum \beta_k * X_{i,t,k} + \varepsilon_{j,q}$$

where $R_{i,t}$ is portfolio *i*'s return at trading period *t*. $log(Twtvol)_{i,t}$ is logarithm of the total Twitter volume of portfolio *i* at trading period *t*. $UEcorr_{i,t}$ is the unexpected earnings correlation between the two stocks in portfolio *i* at trading period *t*. $ROEcorr_{i,t}$ is the return on earnings correlation between the two stocks in portfolio *i* at trading period *t*. $VOLcorr_{i,t}$ is the trading volume correlation between the two stocks in portfolio *i* at trading period *t*. *TwtVolcorr* is the Twitter volume correlation between the two stocks in portfolio. $log(Twtvol\_Long)_{i,t}$ is logarithm of the total Twitter volume of the long position stock in portfolio *i* at trading period *t*. $log(Twtvol\_Short)_{i,t}$ is logarithm of the total Twitter volume of the short position stock in portfolio *i* at trading period *t*. $X_{i,t,k}$ is a vector of the control variables including sector dummies, location dummies, companies' size dummies, exchange listing dummies, and year-month dummies. The sample period is from December 2010 to December 2014 which is consist of 8 trading periods with 41 pairs portfolios selected from S&P 500 companies.

	{1}	{2}	{3}	{4}	{5}
Log(Twitter Vol)	-	0.0654	-	0.0619	-
	-	(4.12)***	-	(3.97)***	-
Log(Twitter Vol_Long)	-	-	-	-	0.0724
	-	-	-	-	(4.07)***
Log(Twitter Vol_Short)	-	-	-	-	0.0112
	-	-	-	-	(0.87)
Twitter Vol Corr	-	-	0.0577	0.0426	0.0400
	-	-	(0.77)	(0.58)	(0.23)
ROE Corr	0.0485	0.0430	0.0401	0.0274	0.0388
	(6.94)***	(5.83)***	(5.01)***	(4.91)***	(3.27)***
Earnings Surprise Corr	0.0684	0.0623	0.0615	0.0610	0.0639
	(7.04)***	(6.78)***	(6.41)***	(4.95)***	(5.81)***
Trading Vol Corr	0.0058	0.0052	0.0052	0.0049	0.0033
	(3.81)***	(3.69)***	(3.35)***	(3.17)***	(3.06)***
Industry Dummy	YES	YES	YES	YES	YES
Year-Month Dummy	YES	YES	YES	YES	YES
Size Dummy	YES	YES	YES	YES	YES
Location Dummy	YES	YES	YES	YES	YES
Exchange Listing Dummy	YES	YES	YES	YES	YES
Observation	328	328	328	328	328
Adjusted R^2	2.69%	6.29%	4.48%	7.21%	6.93%

Figure 1 Boston Properties (BXP) and AvalonBay Communities (AVB) Trading Pair