國立政治大學財務管理學研究所

碩士學位論文

PTT 股票板討論對股市之影響

The Influence of Discussion in Stock Section of PTT on Stock Market

指導教授：周 冠 男 博士

研究生：林 彥 丞 撰

中 華 民 國 105 年 6 月

國立政治大學財務管理學研究所

論文口試委員會審定書

林　彥　丞　撰之碩士學位論文

PTT 股票板討論對股市之影響

The Influence of Discussion in Stock Section of PTT on Stock Market

業經本委員會審議通過

論文口試委員會委員

（簽名）

指導教授

系主任（所長）　（簽名）

中華民國 105 年　6　月　3　日

摘要

過去總認為市場是有效率的，投資人是理性的，但近年來，行為財務學的興起，開始證明投資人經常做出不理性的行為，導致投資市場只有少數人為贏家，能長期獲得超額報酬。散戶的投資行為也是學術上時常探討的，而在過去幾年，國外開始有學者利用社群網站 twitter，並用情緒辨識軟體將發文者的文章做情緒的歸類，來看能不能提升對大盤或個股的預測能力。結論是其中幾類情緒的確能提升預測大盤或個股的機率。而在台灣，比較有名的股票社群為 PTT 的股票板，該板通常會同時有 1000 人以上在線上。本文做了兩項研究，其一為藉由本週看多及看空文章的總數來看是否能預測未來幾週的大盤報酬，其二為藉由個股看多或看空文章的回文數，來看是否看多或看空文章的回文數愈多，是否能預測個股未來幾天的報酬，而以上兩項的結果，即使有些部分在統計上為顯著，但並沒有一致性，舉例來說，有些個股可能在第三天顯著，並與回文數呈負相關，但另一些個股可能在第 5 天顯著，並與回文數呈正相關，因此並無法得到一致性，很難有合理的解釋，因此可說明，PTT 裡的發文及回文對整個投資市場來說，並無資訊內涵，並無法預測股市的報酬。

關鍵詞:社群網路、PTT

Abstract

The efficient market was popular before and most of the investors are seen as rationality. However, behavioral finance emerged in recent years. This theory tells us investors usually invest irrationally that there are only a few people in the market are winner in the long run. Retail investors' behaviors are usually discussed in theses in recent years. In foreign country, there were some people studying the social network-twitter and use some software to judge the different emotions in an article in twitter. Finally, they found some types of emotions could predict the movement of stock prices or stock index like Dow Jones. In Taiwan, the stock section in PTT is famous. Many people express their opinions in stock section of PTT. In this paper, author use the bullish or bearish article to forecast the stock index a few weeks later and use the number of responses of an article for a specific stock to predict the return of the stock a few days later. The results are that some of the responses of stocks can predict stocks returns five days later and there are negative relation between the numbers of responses and the returns. However, other responses of stocks can predict stocks returns three days later and there are positive relation between the numbers of responses and the returns. The author can't find the consistent results. There are more than1000 people stay in the stock section of PTT at the same time, but these people are only a small group in the stock market. Consequently, we can use PTT to predict the stock market.

Index

1. Introduction 1

2. Literature Review 3

2.1 Efficient Market Hypothesis 3

2.2 Behavioral finance 4

2.3 Emotions and Tweeter 4

2.4 Summary of Literature Review 7

3. Data and Methodology 8

3.1 Data Collection 8

3.2 Research Methodology 10

3.3 Tests of the Hypotheses 11

4. Empirical Results 12

4.1 Use the Atmosphere of PTT Stock Section to Predict the Return of TAIEX 12

4.2 Use the number of responses of a specific stock to predict the return of stocks 14

5. Conclusion 34

6. References 35

List of Tables

Table 1: The result of model 1 12

Table 2: The result of model 2 13

Table 3: The result of model 3 (ticker:4938) 14

Table 4: The result of model 3 (ticker:3481) 15

Table 5: The result of model 3 (ticker:2330) 16

Table 6: The result of model 3 (ticker:3008) 17

Table 7: The result of model 3 (ticker:2303) 18

Table 8: The result of model 3 (ticker:2357) 19

Table 9: The result of model 3 (ticker:1303) 19

Table 10: The result of model 3 (ticker:2002) 20

Table 11: The result of model 3 (ticker:2311) 20

Table 12: The result of model 3 (ticker:2317) 21

Table 13: The result of model 3 (ticker:1216) 22

Table 14: The result of model 3 (ticker:2382) 22

Table 15: The result of model 3 (ticker:2891) 23

Table 16: The result of model 3 (ticker:2325) 23

Table 17: The result of model 3 (ticker:2412) 24

Table 18: The result of model 3 (ticker:2490) 25

Table 19: The result of model 3 (ticker:1476) 26

Table 20: The result of model 3 (ticker:9904) 26

Table 21: The result of model 3 (ticker:2354) 27

Table 22: The result of model 3 (ticker:2474) 28

Table 23: The result of model 3 (ticker:2354) 29

Table 24: The result of model 3 (ticker:2474) 30

Table 25: The result of model 3 (ticker:2474) 31

Table 26: The result of model 3 (portfolio) 32

List of Figures

Figure 1 Famous Bulletin Board System (BBS) PTT..2

Figure 2 Stock market section of PTT..2

Figure 3 More than 1000 people stay in the stock market section at the same time 2

Figure 4 The bullish article in PTT (1) 9

Figure 5 The bullish article in PTT (2) 9

Figure 6 Numbers of responses for articles in PTT 10

1. Introduction

Stock market prediction has attracted much attention from academia as well as business. Fundamental and technical analyses are two of the most popular methods that people use to forecast the future trend of the stock market. The Fundamental analysis is a method of assessing securities that attempts to measure its intrinsic value by examining economic, financial and other qualitative and quantitative factors. Technical analysis uses a completely different way. Technicians do not care about the value of a company. They are only interested in the price movements of the market.

In recent years, there are studies examining networks like twitter to predict the stock market. The number of monthly active users of twitter is about 320 million in 2015. There are some communities of stock market within twitter that many people like retail investors or professional traders give their opinion of the price trend of the stock market. People may consider that there may be tons of noisy messages within twitter but it could be interesting to collect some of the messages and explore some methods to realize whether the messages influence the direction of the price movement.

There is a famous Bulletin Board System (BBS) named PTT in Taiwan. There are more than 100 thousand (Figure 1) people staying in PTT at the same time and one of the most popular sections in PTT is stock market section (Figure 2). There are more than 1000 people that stay in the stock market section at the same time. Many people express their ideas and criticize other people’s opinion about stock market or the specific stock. Some people give the opinion of the price direction of some stocks. They may say stock A is bullish because of their outperformance earning per share (EPS) or stock B is bearish because the price drops down and is lower than the 60 days moving average line. Such as the examples above, studying the messages like those in the stock market section of PPT is the objective of this thesis.

Figure 1 Famous Bulletin Board System (BBS) PTT

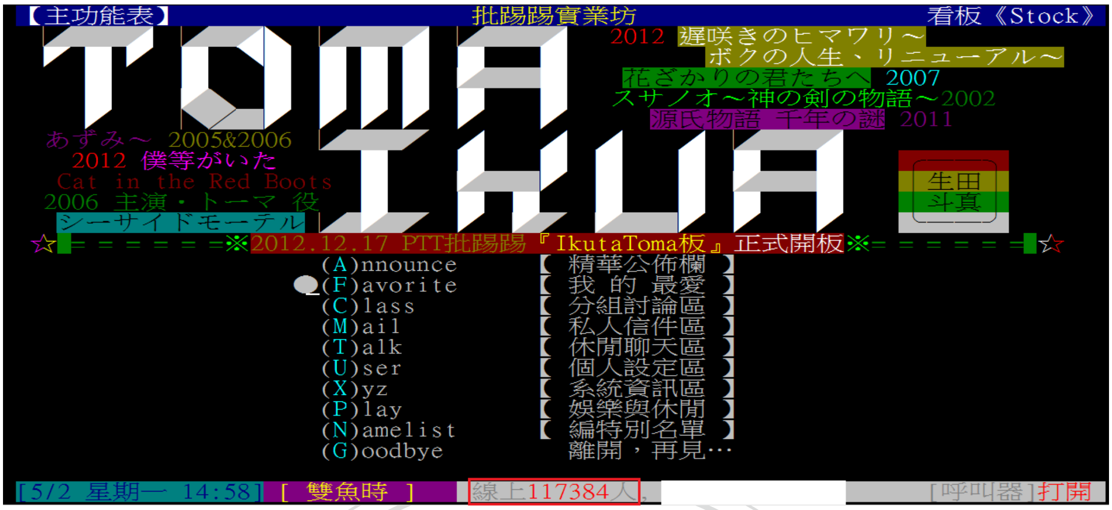


Figure 2 Stock market section of PTT

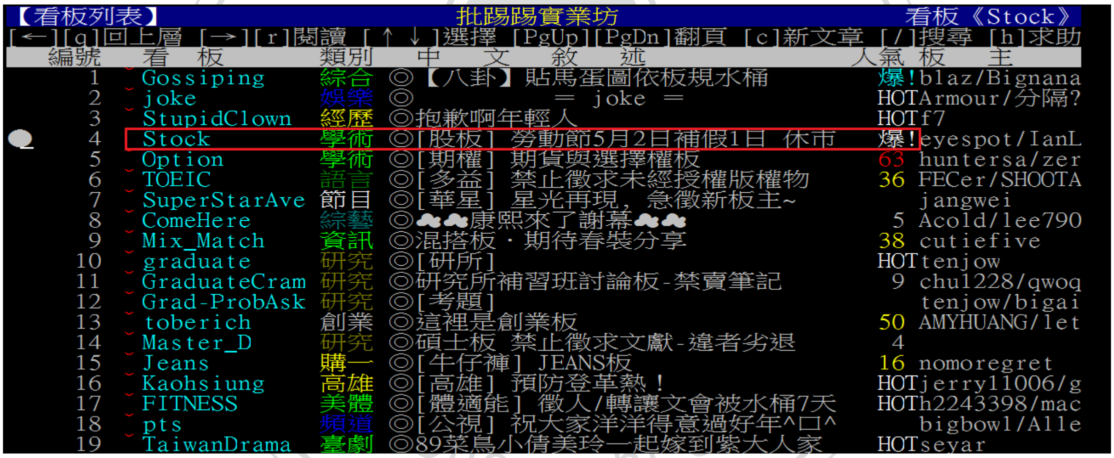


Figure 3 More than 1000 people stay in the stock market section at the same time

2. Literature Review

2.1 Efficient Market Hypothesis

Efficient market hypothesis is one of the most important theory in the finance field. The core of this theory is that asset price completely reflects all available information (Eugene Fama 1970). It means that no one can beat the market and it is really hard to earn the abnormal return in the long run. However, some critics say the market isn't efficient and rational. It's difficult to explain financial crises in recent years like dot com bubble in 2000, subprime mortgage crisis in 2007, European debt crisis in 2011. The efficient market hypothesis requires that agents have rational expectations that on average the population is correct which is rejected by the theory of behavioral finance.

2.2 Behavioral finance

Prospect theory is one of the most important theories in behavioral finance (Kahneman and Tversky 1979). There are four important points in this paper:(1) Reference dependence: When evaluating outcomes, the decision maker has a reference level. Outcomes are compared to the reference point and classified as "gains" if greater than the reference point and "losses" if less than the reference point. (2) Loss aversion: Losses bite more than equivalent gains. In their paper, they found the median coefficient of loss aversion to be about 2.25 times more than equivalent gains. (3) Non-linear probability weighting: Evidence indicates that decision makers overweight small probabilities and underweight large probabilities.(4) Diminishing sensitivity to gains and losses: As the size of the gains and losses relative to the reference point increase in absolute value, the marginal effect on the decision maker's utility falls.

According to the statements above, people aren't rational and the degree of gains or losses may easily influence one's mind and investment decisions. It means emotions is a significant factor for investors.

2.3 Information and Social Networks

The traditional financial theory mainly focuses on the efficient market hypothesis (EMH). One of the EMH assumptions tells us people are rational that people make their own investment strategies without any emotional problem. However, there is a completely different view in the behavioral finance. The Behavioral finance is a subject that tells us emotions can influence one's investment decision and there are lots of factors that may influence the investment decision meaning people are not rational. One of the factors is other people's opinions.

Not only retail investors but also institutional investors usually need to read tons of information from traditional social networks like TV programs of finance, news papers, financial magazines to realize the current condition of the industrial trend or pick up some stocks to put into their portfolios. For example, some professional people talk about their ideas and suggest some stocks to investors and investors may believe the people who are seen as professional people even though they don't know whether the people are really professional.

During 2000, Yahoo Finance and Raging Bull provided two of the largest and prominent sets of message boards. Werner Antweiler and Murray Z. Frank (2004) used these message boards to study how Internet stock message boards are related to stock markets. First, authors found that a positive shock to message board posting predicts negative returns on the next day. Second, a traditional hypothesis is that disagreement induces trading. Authors found significant evidence supporting this claim. Malcolm

Baker and Jeffrey Wurgler (2006) told us that when sentiment is estimated to be high, stocks that are attractive to optimists and speculators and at the same time unattractive to arbitrageurs. Wesley S. Chan (2002) told us stocks with public news in a given month experience momentum. Those that do not have public news show no momentum.

In recent years, use of social networks like twitter and PTT has grown exponentially. A person talks about their ideas and suggest some stocks that can be bought or sold on the internet and other people will agree or disagree with that. The people who agree may really buy or sell the stocks suggested. It means that people's emotions may be triggered or changed by other people's opinions.

Huina Mao, Scott Counts and Johan Bollen (2011) survey a range of online data sets (Twitter feeds, news headlines, and volumes of Google search queries) and sentiment tracking methods (Twitter Investor Sentiment, Negative News Sentiment and Tweet & Google Search volumes of financial terms), and compare their value for financial prediction of market indices such as the Dow Jones Industrial Average, trading volumes, and market volatility (VIX), as well as gold prices. . The results show that traditional surveys of Investor Intelligence are lagging indicators of the financial markets. However, weekly Google Insight Search volumes on financial search queries do have predictive value. An indicator of Twitter Investor Sentiment and the frequency of occurrence of financial terms on Twitter in the previous 1-2 days are also found to be very statistically significant predictors of daily market log return. Survey sentiment indicators are however found not to be statistically significant predictors of financial market values, once we control for all other mood indicators as well as the VIX. Z. Da, J. Engelberand, and P. Gao (2010) told us that fear sentiment predicts (i) short-term return reversals, (ii) temporary increases in volatility, and (iii) mutual fund flows out of equity funds and into bond funds.

There are two mood tracking tools called OpinionFinder (OF) that measures positive and negative moods and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Johan Bollen, Huina Mao and Xiao-Jun Zeng (2010) used the two softwares above to test the sentence of twitter and investigated the hypothesis that public mood states are predictive of changes in DJIA closing values. Their results indicate that the accuracy of DJIA predictions can be significantly improved by the specific public mood but not others. They find an accuracy improvement that predicts the daily up and down changes in the closing values of the DJIA. OF is a publicly available software package for sentiment analysis that can be applied to determine sentence is positive or negative. However, there are only two kinds of sentences (positive and negative) if they use OF. Consequently, to obtain additional dimensions of public mood, they used another mood analysis tool GPOMS that can measure human mood including 6 different mood dimensions, namely Calm, Alert, Sure, Vital, Kind and Happy. Finally, they do not observe relation between DJIA and OpinionFinder's assessment of public mood states (positive and negative) but the GPOMS dimension labeled "Calm". The calmness of the public is predictive of the DJIA which can improve the predictive accuracy of closing values of DJIA.

Tushar Rao and Saket Srivastava (2012) conducted over a period of 14 months between (2010.6.2~2011.7.29) and analyzed 4 million tweets for DJIA, NASDAQ-100 and 13 other big cap technological stocks (Amazon, Apple, AT&T, Dell, EBay, Google, IBM, Intel, Microsoft, Oracle, Samsung Electronics, SAP, Yahoo). These companies are some of the highly traded and discussed technology stocks having very high tweet volumes. They used JSON API from Twittersentiment which is a service provided by Stanford NLP research group to divide all tweets into two parts, positive and negative tweets. The results show that negative and positive dimensions of public mood carry high correlation (up to 0.88 for returns) between stock prices and twitter sentiments.

2.4 Summary of Literature Review

In the foreign papers, they use twitter and some software to analyze sentiments. However, this paper may choose the local social network PTT as the target to study and to know if the result might be different or not. Before studying, we may expect two topics. The first is that the popular bullish articles in PTT may cause the poor performance of the stock index or some stocks mentioned in articles in the future and popular bearish articles in PTT may cause the great performance of the stock index or some stocks mentioned in articles in the future. The second is if there are more bullish articles than bearish articles now, the stock index will be with poor performance in the future. On the contrary, if there are more bearish articles than the bullish articles now, the stock index will be with great performance in the future.

3. Data and Methodology

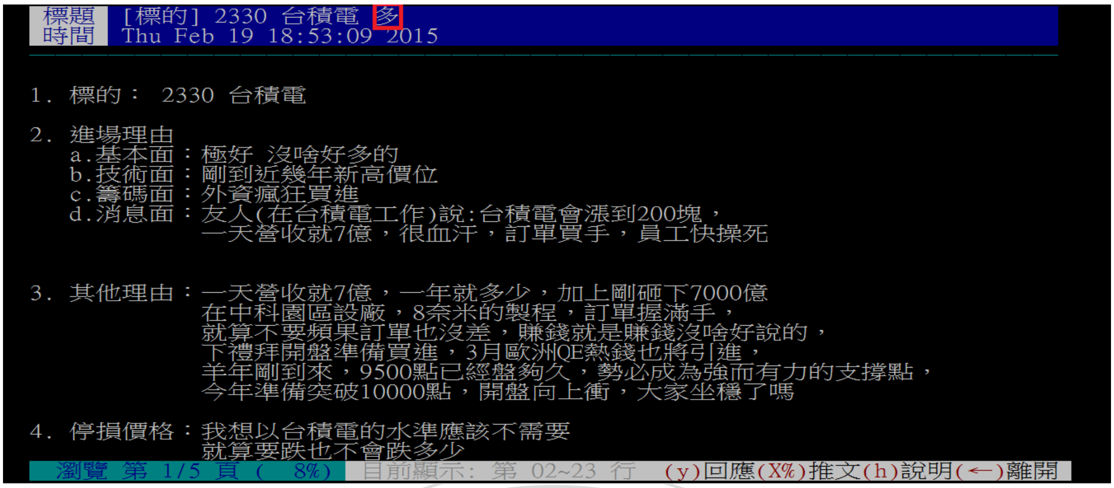
3.1 Data Collection

There are two topics discussing in this paper. The first topic is that there is a type of article called “target” in the stock section of PTT. This paper chooses only this type of articles and read these articles mentioning a specific stock. Then judge the stock of the article for the author is bullish or bearish. If the stock is bullish, the independent variable (X) is one (+1). However, the stock is bearish and the independent variable is minus one (-1). Sum up all the bullish and bearish articles this week to predict the returns of TAIEX (dependent variable: Y) next week to next seven week. For example, it means if there are 5 bullish articles and 3 bearish articles this week, the X is 2.The period of this topic is 2012/10~2016/2.

The method how to judge an article is bullish or bearish is that read an article and the author of the article may mention some statements that can easily know the author is bullish or bearish to a specific stock. For example, the statement is that he/she would stop loss if the price was under a number or stop loss if the stock price was under the 20-day moving average line and we can know that this is a bullish article.

According to the Figure 4, the author obviously told us that the stock 2330 will be bullish in the future. According to the Figure 5, the author didn’t obviously told us the stock is bullish or bearish for us, but we can see that the author said if the price broke through the bottom, he/she would stop loss. Consequently, we can know the stock 2330 is bullish for the author.

Figure 4 The bullish article in PTT (1)


標題 [標的] 2330 台積電 多
時間 Thu Feb 19 18:53:09 2015

1. 標的： 2330 台積電

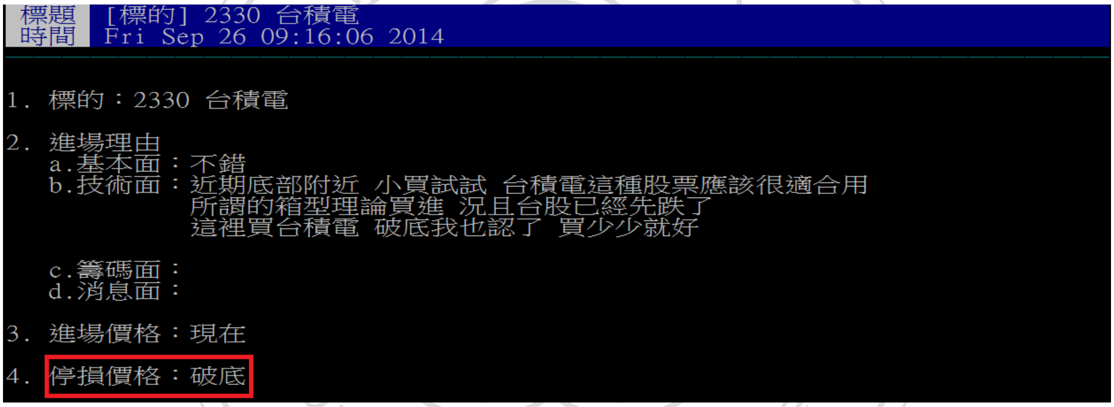
2. 進場理由
a.基本面：極好 沒啥好多的
b.技術面：剛到近幾年新高價位
c.籌碼面：外資瘋狂買進
d.消息面：友人(在台積電工作)說:台積電會漲到200塊，
一天營收就7億，很血汗，訂單買手，員工快操死

3. 其他理由：一天營收就7億，一年就多少，加上剛砸下7000億
在中科園區設廠，8奈米的製程，訂單握滿手，
就算不要蘋果訂單也沒差，賺錢就是賺錢沒啥好說的，
下禮拜開盤準備買進，3月歐洲QE熱錢也將引進，
半年剛到來，9500點已經盤夠久，勢必成為強而有力的支撐點，
今年準備突破10000點，開盤向上衝，大家坐穩了嗎

4. 停損價格：我想以台積電的水準應該不需要
就算要跌也不會跌多少

瀏覽 第 1/5 頁 (8%) 目前顯示: 第 02~23 行 (y)回應(X%)推文(h)說明(←)離開

Figure 5 The bullish article in PTT (2)


標題 [標的] 2330 台積電
時間 Fri Sep 26 09:16:06 2014

1. 標的：2330 台積電

2. 進場理由
a.基本面：不錯
b.技術面：近期底部附近 小買試試 台積電這種股票應該很適合用
所謂的箱型理論買進 況且台股已經先跌了
這裡買台積電 破底我也認了 買少少就好

c.籌碼面：
d.消息面：

3. 進場價格：現在

4. 停損價格：破底

The second topic is that we want to know whether the number of responses of a bullish or bearish article for a specific stock today can influence the return of the specific stock next day or few days from now. According to the Figure 6, look at the left hand side of the figure, we can see the numbers of responses.

We pick up component stocks of Taiwan 50 (0050) whose samples in the stock section of PTT are at least 10 and we also only choose the type of target article in the stock section. Finally, there are 23 component stocks in Taiwan 50 matching the criteria and the tickers of these stocks are 4938, 3481, 2330, 3008, 2303, 2357, 1303, 2002, 2311, 2317, 1216, 2382, 2891, 2325, 2412,2490, 1476, 9904, 2354, 2474, 2454, 2887 and 3474. The period of this topic is 2012/10~2016/3.

The independent variable (R) is the number of response for a bullish or bearish article of one of the 23 stocks that last paragraph mentioned and the dependent variable (S) is the returns of one of the 23 stocks next day or few days later.

Figure 6 Numbers of responses for articles in PTT

3.2 Research Methodology

In this paper, we put the independent variables (X and R) of the two topics in natural logarithm. The regressions of the first topic are below:

$$\text{Model1}: Y(t)=b+B1*\ln X(t-1)+B2*\ln X(t-2)+B3*\ln X(t-3)+B4*\ln X(t-4)+B5*\ln X(t-5)+ B6*\ln X(t-6)+ B7*\ln X(t-7) + \varepsilon \ldots.(0)$$

$$\text{Model 2}: Y(t) = a1+ A1*\ln X(t-1)+ \varepsilon \ldots.(1)$$

$$Y(t) = a2+ A2*\ln X(t-2)+ \varepsilon \ldots.(2)$$

$$Y(t) = a3+ A3*\ln X(t-3)+ \varepsilon \ldots.(3)$$

$$Y(t) = a4+ A4*\ln X(t-4)+ \varepsilon \ldots.(4)$$

$$Y(t) = a5+ A5*\ln X(t-5)+ \varepsilon \ldots.(5)$$

$$Y(t) = a6+ A6*\ln X(t-6)+ \varepsilon \ldots.(6)$$

$$Y(t) = a7+ A7*\ln X(t-7)+ \varepsilon \ldots.(7)$$

Where X= sum up all the bullish and bearish articles of a specific stock this week (bullish+1 and bearish-1), Y= return of TAIEX, Ai (i=1, 2,…,7) and Bi (i=1, 2,…,7) = regression coefficients , b and ai (i=1, 2,…,7) = intercept, ε= error term.

The regressions of the second topic are as below:

Model 3: $S(t) = c1+ C1*\ln R(t-1)+M(t)+M(t-1)+M(t-2)+M(t-3)+ \varepsilon$….(8)

$S(t) = c2+ C2*\ln R(t-2)+M(t)+M(t-1)+M(t-2)+M(t-3)+ \varepsilon$….(9)

$S(t) = c3+ C3*\ln R(t-3)+M(t)+M(t-1)+M(t-2)+M(t-3)+ \varepsilon$….(10)

Where R= the number of responses, M= market return, S= return of a stock, Ci(i=1, 2, 3)= regression coefficients, ci (i=1, 2, 3) = intercept, ε= error term.

According to the regression above, the number of lags is three. However, not all the stocks have enough samples and the least required number of samples is ten so that the numbers of lags are not all three. Besides, we used the portfolio comprising the 23 stocks to run the regression. The regression is as below:

Model 4: $PS(t) = c1+ C1*\ln AR(t-1)+M(t)+M(t-1)+M(t-2)+M(t-3)+ \varepsilon$….(11)

$PS(t) = c1+ C1*\ln AR(t-2)+M(t)+M(t-1)+M(t-2)+M(t-3)+ \varepsilon$….(12)

$PS(t) = c1+ C1*\ln AR(t-3)+M(t)+M(t-1)+M(t-2)+M(t-3)+ \varepsilon$….(13)

Where AR= the average number of responses, M= market return, PS= return of the portfolio, Ci (i=1, 2, 3)= regression coefficients, ci (i=1, 2, 3) = intercept, ε= error term.

3.3 Tests of the Hypotheses

We will test our hypotheses using ordinary least squares (OLS) and use a significance level of 1%, 5% and 10% (two-tailed). Before running the data, we delete the outliers of the data (1% of right tail and 1% of left tail) to avoid these data influencing the results.

4. Result

4.1 Use the Atmosphere of PTT Stock Section to Predict the Return of TAIEX

In 3.2 (Research Methodology), we use the independent variable X representing the total number of bullish articles of specific stocks minus the bearish number and Y is the return of TAIEX. We assume that X represents the atmosphere of the stock market now and we want to know if this factor can predict the future return of TAIEX. The model 1 and 2 are used to exam the results. Results of model 1 and model 2 in chapter 3 are as below:

Table 1: The result of model 1

Variable	(0)
Constant	0.00212
	(0.00348)
lag1	0.00037
	(0.00100)
lag2	-0.00063
	(0.00113)
lag3	-0.00144
	(0.00111)
lag4	-0.00102
	(0.00112)
lag5	0.00120
	(0.00112)
lag6	0.00036
	(0.00112)
lag7	0.00012
	(0.00101)
Observation	165
R square	0.0356

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The number of samples for model 1 is 165 and the result of model 1 show that the atmosphere of PTT Stock Section can't predict the return of TAIEX. Multicollinearity might exist among X(t-1), X(t-2),……, X(t-7) and then we use the model 2 in chapter 3. The result is as below:

Table 2: The result of model 2

Variable	(1) lag1	(2) lag2	(3) lag3	(4) lag4	(5) lag5	(6)lag6	(7)lag7
Constant	**0.00004742**	**0.0019**	**0.00342**	**0.00203**	**-0.00183**	**-0.00147**	**-0.00090736**
	(0.00239)	**(0.00250)**	**(0.00255)**	**(0.00265)**	**0.00258**	**0.00246**	**0.00238**
lnX	**-0.0003884**	**-0.00099**	**-0.00148**	**-0.00103**	**0.000223**	**0.000106**	**-0.00007719**
	(0.00076)	**(0.00080)**	**(0.00081)**	**(0.00084)**	**0.00082**	**0.00079**	**0.00076**
Observation	**165**	**165**	**165**	**165**	**165**	**165**	**165**
R square	**0.00160**	**0.00930**	**0.01990**	**0.00900**	**0.00040**	**0.00010**	**0.00010**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The number of samples for model 2 is also 165 and the model 2 can avoid the multicollinearity. The result of model 2 shows there is significantly negative relation between lnX (t-3) and Y (t) (Table 2). It means the bullish atmosphere in stock section of PTT now will cause the negative return of TAIEX three weeks later and bearish atmosphere in stock section of PTT now will cause the positive return of TAIEX three weeks later. It means that the atmosphere in PTT delayed for three weeks.

4.2 Use the number of responses of a specific stock to predict the return of stocks

We pick up 23 stocks from the component stocks of Taiwan 50 whose samples are at least 10. In ch3, we use the independent variable (R) to represent the number of responses in an article mentioning one of the 23 stocks – the author of article told us the stock is bullish or bearish. The dependent variable (S) is the returns of one of the 23 stocks next day or more days later. The model 3 in ch3 is used to exam whether the R can predict the S or not and the results are as below:

Table 3: The result of model 3 (ticker:4938)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	-0.00544	0.00216	-0.03840**
	(0.01661)	(0.01785)	(0.01674)
lnR	0.00068	-0.00276	0.01545**
	(0.00738)	(0.00786)	(0.00732)
M(t)	0.00675	0.00757	0.00342
	(0.00611)	(0.00652)	(0.00569)
M(t-1)	-0.00922	-0.00925	-0.01320*
	(0.00715)	(0.00692)	(0.00650)
M(t-2)	-0.00238	-0.00202	-0.00010
	(0.00728)	(0.00732)	(0.00661)
M(t-3)	-0.00569	-0.00550	-0.00910
	(0.00693)	(0.00693)	(0.00641)
Observation	24	24	24
R square	0.20490	0.20996	0.36214

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 3 above shows that there is positive relationship between the number of responses and the returns 3 days from now of 4398.

Table 4: The result of model 3 (ticker:3481)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**0.013089**	**0.039979**	**-0.03759**
	(0.02480)	**(0.02416)**	**(0.02383)**
lnR	**-0.00337**	**-0.01336**	**0.015953***
	(0.00927)	**(0.00881)**	**(0.00884)**
M(t)	**-0.0094**	**-0.01132***	**-0.0121***
	(0.00708)	**(0.00641)**	**(0.00631)**
M(t-1)	**0.004623**	**0.007167**	**0.005**
	(0.00593)	**(0.00575)**	**(0.00544)**
M(t-2)	**0.005226**	**0.004176**	**0.002356**
	(0.00601)	**(0.00573)**	**(0.00578)**
M(t-3)	**0.001646**	**0.002387**	**0.003797**
	(0.00593)	**(0.00563)**	**(0.00560)**
Observation	**25**	**25**	**25**
R square	**0.1613**	**0.2466**	**0.2791**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 4 above shows that there is positive relationship between the number of responses and the returns 3 days from now of 3481.

Table 5: The result of model 3 (ticker:2330)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	-0.0011	0.011279	0.004166
	(0.00704)	(0.00697)	(0.00747)
lnR	0.000717	-0.00444	-0.00149
	(0.00276)	(0.00274)	(0.00298)
M(t)	0.001833	0.00061	0.001459
	(0.00297)	(0.00293)	(0.00309)
M(t-1)	0.00348	0.00434	0.00300
	(0.00313)	(0.00301)	(0.00312)
M(t-2)	-0.00384	-0.00507	-0.00362
	(0.00311)	(0.00305)	(0.00314)
M(t-3)	0.004675	0.004297	0.004293
	(0.00310)	(0.00297)	(0.00320)
Observation	35	35	35
R square	0.1498	0.2183	0.1551

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 5 above shows that there is no significant relationship between the number of responses and returns of 2330.

Table 6: The result of model 3 (ticker:3008)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**0.015061**	**0.007836**	**0.015109**
	(0.00920)	**(0.00941)**	**(0.00879)**
lnR	**-0.00556**	**-0.00242**	**-0.0057**
	(0.00368)	**(0.00377)**	**(0.00357)**
M(t)	**0.000239**	**-0.0007**	**-0.00105**
	(0.00361)	**(0.0037)**	**(0.00358)**
M(t-1)	**0.00043**	**0.00003**	**-0.00062**
	(0.00369)	**(0.00381)**	**(0.00365)**
M(t-2)	**0.00020**	**0.00140**	**0.00191**
	(0.00376)	**(0.00386)**	**(0.00374)**
M(t-3)	**-0.00053**	**-0.00163**	**-0.00063**
	(0.00365)	**(0.00376)**	**(0.00363)**
Observation	**36**	**36**	**36**
R square	**0.0763**	**0.0196**	**0.0841**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 6 above shows that there is no significant relationship between the number of responses and returns of 3008.

Table 7: The result of model 3 (ticker:2303)

Variable	(8) lag1	(9) lag2
Constant	-0.01268	0.00817
	(0.01373)	(0.00899)
lnR	-0.0028	-0.0158**
	(0.01014)	(0.00538)
M(t)	0.02786	0.02656*
	(0.02123)	(0.01328)
M(t-1)	-0.00755	-0.00852
	(0.02096)	(0.00896)
Observation	10	10
R square	0.2694	0.6963

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 7 above shows that there is negative relationship between the number of responses and the returns 2 days from now of 2303.

Table 8: The result of model 3 (ticker:2357)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**-0.01248**	**-0.00272**	**0.004495**
	(0.01166)	**(0.01229)**	**(0.01118)**
lnR	**0.009039**	**0.003297**	**-0.00142**
	(0.00645)	**(0.00735)**	**(0.00619)**
M(t)	**-0.00066**	**0.00212**	**0.00231**
	(0.00674)	**(0.00675)**	**(0.00678)**
M(t-1)	**-0.00863**	**-0.01161**	**-0.01054**
	(0.00641)	**(0.00712)**	**(0.00666)**
M(t-2)	**0.00158**	**0.00284**	**0.00200**
	(0.00654)	**(0.00734)**	**(0.00704)**
M(t-3)	**-0.00386**	**-0.00184**	**-0.00258**
	(0.00671)	**(0.00696)**	**(0.00731)**
Observation	**22**	**22**	**22**
R square	**0.2506**	**0.1692**	**0.1615**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 8 above shows that there is no significant relationship between the number of responses and returns of 2357.

Table 9: The result of model 3 (ticker:1303)

Variable	(8) lag1
Constant	**-0.00599**
	(0.01566)
lnR	**0.01038**
	(0.01456)
M(t)	**0.01162**
	(0.00832)
Observation	**10**
R square	**0.3180**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 9 above shows that there is no significant relationship between the number of responses and returns of 1303.

Table 10: The result of model 3 (ticker:2002)

Variable	(8) lag1
Constant	-0.00976
	(0.02202)
lag1	0.00013258
	(0.00788)
M	-0.00872
	(0.00486)
Observation	10
R square	0.34830

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 10 above shows that there is no significant relationship between the number of responses and returns of 2002.

Table 11: The result of model 3 (ticker:2311)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	-0.015978	0.026541	0.05443**
	(0.01994)	(0.01956)	(0.02097)
lnR	0.0088156	-0.015986	-0.02852**
	(0.01117)	(0.00954)	(0.00978)
M(t)	0.00176	0.00244	0.01124*
	(0.00620)	(0.00552)	(0.00545)
M(t-1)	-0.00181	-0.00095	-0.000001
	(0.00644)	(0.00593)	(0.00501)
M(t-2)	0.00171	-0.00127	0.00075
	(0.00685)	(0.00592)	(0.00497)
M(t-3)	0.0063981	0.0023567	0.00198
	(0.00676)	(0.00557)	(0.00467)
Observation	17	17	17
R square	0.1253	0.2638	0.4787

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 11 above shows that there is negative relationship between the number of responses and the returns 3 days from now of 2311.

Table 12: The result of model 3 (ticker:2317)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	0.008615*	-0.01164**	-0.00199
	(0.00510)	(0.00483)	(0.00526)
lnR	-0.003	0.005511***	0.00145
	(0.00202)	(0.00191)	(0.00206)
M(t)	0.00634**	0.00840***	0.00669**
	(0.00250)	(0.00241)	(0.00254)
M(t-1)	0.00503*	0.00670***	0.00621**
	(0.00262)	(0.00243)	(0.00269)
M(t-2)	0.00283	0.00400	0.00296
	(0.00263)	(0.00250)	(0.00271)
M(t-3)	0.004479*	0.00469*	0.005571**
	(0.00261)	(0.00241)	(0.00267)
Observation	48	48	48
R square	0.2567	0.3476	0.2270

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 12 above shows that there is positive relationship between the number of responses and the returns 2 days from now of 2317.

Table 13: The result of model 3 (ticker:1216)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	0.05526***	-0.02947	-0.00119
	(0.01490)	(0.02606)	(0.02514)
lnR	0.02471***	0.012567	-0.00046
	(0.00675)	(0.01180)	(0.01110)
M(t)	0.01839***	0.02562***	0.02079
	(0.00458)	(0.00783)	(0.00690)
M(t-1)	0.00482	-0.00172	0.00014
	(0.00485)	(0.00692)	(0.00799)
M(t-2)	-0.00792	0.00176	-0.00228
	(0.00479)	(0.00745)	(0.00681)
M(t-3)	0.003671	0.000223	0.00384
	(0.00451)	(0.00683)	(0.00706)
Observation	17	17	17
R square	0.7773	0.5524	0.5063

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 13 above shows that there is significantly positive relationship between the number of responses and the returns 1 day from now of 1216.

Table 14: The result of model 3 (ticker:2382)

Variable	(8) lag1
Constant	-0.00229
	(0.02304)
lnR	0.00547
	(0.01443)
M(t-1)	-0.01279
	(0.01574)
Observation	10
R square	0.08840

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 14 above shows that there is no significant relationship between the number of responses and returns in 2382.

Table 15: The result of model 3 (ticker:2891)

Variable	(8) lag1	(9) lag2
Constant	0.01226*	0.00962
	(0.00596)	(0.00614)
lnR	-0.00577*	-0.00474
	(0.00244)	(0.00256)
M(t)	0.00996**	0.00925**
	(0.00278)	(0.00304)
M(t-1)	0.00209	0.00039
	(0.00275)	(0.00286)
Observation	10	10
R square	0.7251	0.6622

(p-value<0.01: ***, p-value < 0.05: **, p-value < 0.1: *)

The table 13 above shows that there is significantly negative relationship between the number of responses and the returns 1 day from now of 2891.

Table 16: The result of model 3 (ticker:2325)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	0.09513***	0.00753	-0.02186
	(0.01797)	(0.0590)	(0.03947)
lnR	0.04260***	-0.00682	0.00714
	(0.00832)	(0.0251)	(0.01865)
M(t)	-0.0258***	-0.0225*	-0.02321**
	(0.00274)	(0.0083)	(0.00734)
M(t-1)	0.00497	0.00423	0.00299
	(0.00323)	(0.0082)	(0.00759)
M(t-2)	-0.00453	-0.0026	-0.00479
	(0.00284)	(0.0106)	(0.00768)
M(t-3)	0.00629*	0.00147	-0.00018
	(0.00283)	(0.0073)	(0.00799)
Observation	10	10	10
R square	0.9699	0.7767	0.7806

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 16 above shows that there is significantly positive relationship between the number of responses and the returns 1 day from now of 2325.

Table 17: The result of model 3 (ticker:2412)

Variable	(8) lag1	(9) lag2
Constant	-0.00841	-0.00205
	(0.00718)	(0.00531)
lnR	0.00291	0.00071
	(0.00252)	(0.00204)
M(t)	0.00558***	0.00549***
	(0.00071)	(0.00079)
M(t-1)	0.00054	0.00090
	(0.00084)	(0.00091)
M(t-2)	0.00029	0.00039
	(0.00074)	(0.00082)
Observation	10	10
R square	0.9325	0.9164

(p-value<0.01: ***, p-value < 0.05:**, p-value < 0.1: *)

The table 17 above shows that there is no significant relationship between the number of responses and returns in 2412.

Table 18: The result of model 3 (ticker:2490)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**0.00981**	**-0.01012**	**0.02060**
	(0.0241)	**(0.0266)**	**(0.0209)**
lnR	**-0.00535**	**0.00296**	**-0.01056**
	(0.0096)	**(0.0107)**	**(0.0088)**
M(t)	**0.00405**	**0.00381**	**0.00450**
	(0.0051)	**(0.0058)**	**(0.0048)**
M(t-1)	**-0.00360**	**-0.00257**	**-0.00057**
	(0.0050)	**(0.0052)**	**(0.0051)**
M(t-2)	**-0.00133**	**-0.00182**	**-0.00428**
	(0.0053)	**(0.0053)**	**(0.0051)**
M(t-3)	**-0.00491**	**-0.00342**	**-0.00318**
	(0.0088)	**(0.0084)**	**(0.00797)**
Observation	**20**	**20**	**20**
R square	**0.1161**	**0.1015**	**0.1800**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 18 above shows that there is no significant relationship between the number of responses and returns in 2490.

Table 19: The result of model 3 (ticker:1476)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	-0.00219	-0.02264*	0.008389
	(0.01295)	(0.01305)	(0.01101)
lnR	-0.00303	0.005707	-0.00757*
	(0.00502)	(0.00491)	(0.00409)
M(t)	0.02006***	0.02229***	0.01737**
	(0.00667)	(0.00656)	(0.00642)
M(t-1)	0.01202*	0.01401**	0.01170*
	(0.00666)	(0.00652)	(0.00617)
M(t-2)	0.00067	0.00467	0.00022
	(0.00727)	(0.00648)	(0.00614)
M(t-3)	0.001149	0.005787	-0.00004
	(0.00609)	(0.00691)	(0.00572)
Observation	27	27	27
R square	0.3590	0.3873	0.4392

(p-value<0.01: ***, p-value < 0.05: **, p-value < 0.1: *)

The table 19 above shows that there is significantly negative relationship between the number of responses and the returns 3 days from now of 1476.

Table 20: The result of model 3 (ticker:9904)

Variable	(8) lag1
Constant	-0.03027
	(0.02219)
lnR	0.01352
	(0.01086)
M	0.00274
	(0.01416)
Observation	10
R square	0.1874

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 20 above shows that there is no significant relationship between the number of responses and returns of 9904.

Table 21: The result of model 3 (ticker:2354)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	0.011796	0.03279	0.010732
	(0.02871)	(0.02670)	(0.03119)
lnR	-0.00871	-0.01894	-0.00761
	(0.01462)	(0.01265)	(0.01513)
M(t)	0.00018	-0.001	0.003516
	(0.01012)	(0.00891)	(0.01137)
M(t-1)	0.005509	-0.00047	0.00188
	(0.01076)	(0.00966)	(0.01135)
M(t-2)	0.009579	0.015187	0.010182
	(0.01103)	(0.00944)	(0.01095)
M(t-3)	0.005619	-0.00245	0.002736
	(0.01203)	(0.00947)	(0.01049)
Observation	12	12	12
R square	0.3181	0.4740	0.3069

(p-value<0.01: ***, p-value < 0.05: **, p-value < 0.1: *)

The table 21 above shows that there is no significant relationship between the number of responses and returns of 2354.

Table 22: The result of model 3 (ticker:2474)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	-0.01403	0.015708	0.010182
	(0.01408)	(0.01497)	(0.01534)
lnR	0.005135	-0.00986	-0.00704
	(0.00645)	(0.00711)	(0.00729)
M(t)	0.01809***	0.015336***	0.016977***
	(0.00524)	(0.00532)	(0.00519)
M(t-1)	0.002762	0.000415	0.000228
	(0.00537)	(0.00524)	(0.00548)
M(t-2)	-0.00061	-0.0004	-0.00118
	(0.00593)	(0.00582)	(0.00592)
M(t-3)	-0.0144**	-0.01282**	-0.01339**
	(0.00626)	(0.00612)	(0.00619)
Observation	37	37	37
R square	0.3846	0.4087	0.3904

(p-value<0.01: ***, p-value < 0.05: **, p-value < 0.1: *)

The table 22 above shows that there is no significant relationship between the number of responses and returns of 2474.

Table 23: The result of model 3 (ticker:2454)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**0.01273**	**-0.01463**	**0.01702**
	(0.0124)	**(0.0115)**	**(0.0128)**
lnR	**-0.00244**	**0.01013****	**-0.00457**
	(0.0052)	**(0.00491)**	**(0.0056)**
M(t)	**-0.00225**	**-0.00485**	**-0.00110**
	(0.0062)	**(0.0059)**	**(0.0062)**
M(t-1)	**0.00115**	**0.00180**	**0.00214**
	(0.0059)	**(0.00546)**	**(0.0059)**
M(t-2)	**-0.00406**	**-0.00444**	**-0.00417**
	(0.0058)	**(0.00535)**	**(0.0057)**
M(t-3)	**-0.00485**	**-0.00517**	**-0.00410**
	(0.0061)	**(0.0056)**	**(0.00598)**
Observation	**31**	**31**	**31**
R square	**0.0557**	**0.1862**	**0.1862**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 23 above shows that there is significantly positive relationship between the number of responses and the returns 2 days from now of 2454.

Table 24: The result of model 3 (ticker:2887)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**-0.01603**	**0.011135**	**0.000237**
	(0.01135)	**(0.00803)**	**(0.00859)**
lnR	**0.00598**	**-0.00395**	**-0.00002**
	(0.003875)	**(0.002458)**	**(0.00283)**
M(t)	**0.01483*****	**0.015001*****	**0.01466*****
	(0.00162)	**(0.00161)**	**(0.00196)**
M(t-1)	**-0.00239**	**-0.00072**	**-0.00081**
	(0.00207)	**(0.001778)**	**(0.0022)**
M(t-2)	**-0.00170**	**-0.00036**	**-0.00192**
	(0.00197)	**(0.00217)**	**(0.00239)**
M(t-3)	**-0.00115**	**0.00673**	**0.00252**
	(0.00733)	**(0.00733)**	**(0.00852)**
Observation	**11**	**11**	**11**
R square	**0.9577**	**0.9587**	**0.9375**

(p-value<0.01: ***, p-value < 0.05:, p-value < 0.1: *)**

The table 24 above shows that there is no significant relationship between the number of responses and returns of 2887.

Table 25: The result of model 3 (ticker:3474)

Variable	(8) lag1	(9) lag2	(10) lag3
Constant	**0.013083**	**0.002841**	**0.027658***
	(0.01557)	**(0.01552)**	**(0.01480)**
lnR	**-0.00557**	**-0.00144**	**-0.0114***
	(0.006019)	**(0.005981)**	**(0.00568)**
M(t)	**0.01059****	**0.01034****	**0.01062*****
	(0.00401)	**(0.004121)**	**(0.00389)**
M(t-1)	**-0.00336**	**-0.00206**	**-0.00348**
	(0.00429)	**(0.00412)**	**(0.00401)**
M(t-2)	**0.007***	**0.00692**	**0.007935***
	(0.00408)	**(0.004295)**	**(0.00396)**
M(t-3)	**-0.00254**	**-0.00163**	**-0.00403**
	(0.00414)	**(0.004058)**	**(0.004068)**
Observation	**55**	**55**	**55**
R square	**0.1674**	**0.1538**	**0.2172**

(p-value<0.01: ***, p-value < 0.05: **, p-value < 0.1: *)

The table 25 above shows that there is significantly negative relationship between the number of responses and the returns 3 days from now of 3474.

Table 26: The result of model 4 (portfolio)

Variable	(11) lag1	(12) lag2	(13) lag3
Constant	-0.01797*	0.01598	0.01994*
	0.01017	0.01029	0.01013
lnR	0.00858*	-0.00687	-0.00864*
	0.00454	0.00461	0.00455
M(t)	0.00779	0.00751	0.00854
	0.00517	0.00535	0.00533
M(t-1)	0.01407***	0.01074**	0.01047**
	0.00513	0.00497	0.00492
M(t-2)	0.00198	-0.00156	0.00263
	0.00476	0.00525	0.00494
M(t-3)	0.00611	0.00301	0.00035
	0.00473	0.00527	0.00555
Observation	55	55	55
R square	0.1867	0.1527	0.1789

(p-value<0.01: ***, p-value < 0.05: **, p-value < 0.1: *)

The table 26 above shows the portfolio that there is significantly positive relationship between the average number of responses and the returns 1 day from now and negative relationship 3 days from now.

According to the Table 3 to Table 25, we can find the independent variables of stocks 4938, 3481, 2303, 2311, 2317, 1216, 2891, 2325, 1476, 2454 and 3474 can predict the return of these stocks and the numbers of lags significant for all stocks above are not all the same. The relationship between independent and dependent variables can be positive or negative. Even if some of the independent variables can predict the dependent variables, the results are inconsistent. Besides, the table 26 shows the portfolio that there is significantly positive relationship between the average number of responses and the returns 1 day from now and negative relationship 3 days from now. We also can not explain why the significant numbers of lags are 1 and 3 but 2.

Consequently, the information can not give us consistent results. We can't explain rationally about these results and we can say that the people in PPT are just a specific group. Their comments or articles in PTT aren't able to influence the stock market. The information is insufficient.

5. Conclusion

This paper use the social network PTT to study if the articles of stock section of PTT can predict the return of TAIEX and specific stocks. According to the model 2, the result is if the atmosphere in PTT is bullish/bearish this week, the return of TAIEX will be negative/positive three weeks from now. However, it is hard to explain why the result doesn't show that other independent variables are significant but the X (t-3). Consequently, the information isn't enough so that we can't explain the result. According to the model 3, the components of Taiwan 50 like 4938, 3481, 2303, 2311, 2317, 1216, 2891, 2325, 1476, 2454 and 3474 can use the numbers of responses of article to predict their returns few weeks from now. However, not only some of the relations between independent and dependent are positive and other are negative but also the numbers of lags significant for every stock are different. These information can't give us consistent results. The model 4 shows the portfolio that there is significantly positive relationship between the average number of responses and the returns 1 day from now and negative relationship 3 days from now. We also can not explain why the significant numbers of lags are 1 and 3 but 2. We can't explain rationally about these results and we can say that the people in PPT are just a specific group. Their comments or articles in PTT are not able to influence the stock market. The information is insufficient.

Reference

1. Eugene Fama, 1970, Efficient capital markets: A review of theory and empirical work, The Journal of Finance 25, 383-417.
2. Kahneman and Tversky, 1979, Prospect Theory-Division of the Humanities and Social Science, Econometrica 47, 263-292.
3. Wesley S. Chan, 2002, Stock Price Reaction to News and No-News: Drift and Reversal After Headlines, Journal of Financial Economics 70, 223–236.
4. Werner Antweiler and Murray Z. Frank, 2004, Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, The Journal of Finance 3, 1259-1294.
5. Malcolm Baker and Jeffrey Wurgler, 2006, Investor Sentiment and the Cross-Section of Stock Returns, The Journal of Finance 4, 1645-1680.
6. Z. Da, J. Engelberand, and P. Gao, 2010, The sum of all fears: investor sentiment and asset prices, http://ssrn.com/abstract=1509162
7. Johan Bollen, Huina Mao and Xiao-Jun Zeng, 2010, Twitter mood predicts the stock market, Indiana University, Computer 1010, 1–8.
8. Huina Mao, Scott Counts and Johan Bollen, 2011, Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data, arXiv.org, Quantitative Finance Papers, 1051-1112
9. Tushar Rao and Saket Srivastava, 2012, Analyzing Stock Market Movements Using Twitter Sentiment Analysis, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
10. Eric D. Brown, 2012, Will Twitter Make You a Better Investor? A Look at sentiment, User Reputation and Their Effect on the Stock Market, Dakota State University

國立政治大學

財務管理研究所碩士論文

PTT股票板討論對股市之影響

林彥丞 撰

一零五年六月