The Information Content of Chinese News Sentiment around Earnings Announcements [*]

Yu-Chen Wei [**]
Department of Money and Banking
National Kaohsiung First University of Science and Technology

Abstract

This paper examines the information content of Chinese news on the announcement drift and investigates its application to portfolio management. The analysis of linguistic text mining is applied to extract various dimensions of the information content of Chinese news. Proxy variables of public information are measured by the media coverage (*MEDIA*), public news surprises (*SUR*) and news sentiment (*SR*). The Fama and French (1992) 3 factors, momentum (Jegadeesh and Titman, 1993), turnover (*TURN*) and the probability of informed trading (*PIN*) proposed by Easley, Hvidkjaer and O'Hara (2002) are adopted as the factors to calculate the abnormal return around the earnings announcement date. The variable of standardized unexpected earnings (*SUE*) is regarded as the control variable. The empirical results show that there is positive (negative) relationship between *SR* and the cumulative abnormal return before (after) the earnings announcement, thus confirming that the market response covers the relevant information regarding the company. The application of public news sentiment to the portfolio management suggests that "long stocks with low news sentiment and high public news surprises" and "short stocks with high news sentiment and high public news surprises" earn positive excess returns. This study proposes that the relevant news sentiment of individual stocks could be applied to the prediction of abnormal returns and portfolio management.

Keywords: Public News, Sentiment, Linguistic Text Mining, Abnormal Return, Earnings Announcement

EFM Classification Codes: 370, 720

---

[*] This work is supported by the project of the Department of Industrial Technology in Taiwan (Grant number: 99-EC-17-A-34-S1-149). The authors also appreciate the grateful support by the Chinese knowledge and information processing group of the Institute of Information Science and the Institute of Linguistics of Academia Sinica.

[**] Yu-Chen Wei is an Assistant Professor in the Department of Money and Banking of National Kaohsiung First University of Science and Technology in Taiwan. Email: ycwei@nkfust.edu.tw. Address: 2 Jhuoyue Rd., Nanzih, Kaohsiung City, 811, Taiwan, R.O.C.

1. Introduction

In an efficient market, securities prices at any given time fully reflect all available information. Rather than adjusting immediately to news surprises, stock prices tend to drift over time in the same direction as the initial surprise. Post-earnings announcement drift (PEAD) or earnings momentum is one of the anomalies. One of the arguments which support the existence of anomalies is that investors may have different explanations of the 'information' which then results in under- or over-reactions.

Information can be classified into two groups, namely, public or private information (Vega, 2006). Some investors believe that public information is noise. However some investors propose that there is relevant information contained in the public reports. How to extract such relevant information from the large number of news stories is a huge and technical process. Linguistic text mining is one of the methods used to filter out the noise and construct proxies for the quantitative indicators from the qualitative news reports.[1]

Demers and Vega (2010) point out that there are different algorithms to measure the net optimism in managers' earnings announcements in English documents, such as General Inquirer (GI), version 6.0 of the Diction text-analysis program, and the Loughran-McDonald dictionaries (L&M). They also apply principal components analysis to construct the aggregate indicator of news net optimism.

Up to now, there has still been very little research conducted on information sentiment in the case of Chinese documents. The research problems in this present study focus on extending the traditional event study of the earnings announcement to probe the signals of information sentiment by digging out a large number of Chinese

[1] The interested reader should refer to Vega (2006), Tetlock (2007), Tetlock, Saar-Tsechansky and Macskassy (2008) and Demers and Vega (2010).

financial news items. We empirically measure the effect of private and public information on the post-announcement drift. The proxy variables of public information are measured by the media coverage (*MEDIA*), public news surprises (*SUR*) and news sentiment (*SR*). The Fama and French (1992) 3 factors, namely, momentum (Jegadeesh and Titman, 1993), turnover (*TURN*) and the probability of informed trading (*PIN*) proposed by Easley, Hvidkjaer and O'Hara (2002), are adopted as the factors to calculate the abnormal return around the earnings announcement date. Furthermore, we investigate the relevance of public information to portfolio management applications.

The empirical results show that there is a positive relationship between the news sentiment and the cumulative abnormal returns prior to an earnings announcement. The higher the degree of the news sentiment, the higher the cumulative abnormal return. We can also find that the leading effect of the news sentiment (*SR*) that is constructed based on the one to two weeks of public news prior to the earnings announcement continues until two weeks after the earnings announcement. The robustness check that considers the alternative asset pricing factors, such as the Fama and French 3 factors, momentum, the probability of informed trading (*PIN*) and turnover, proposes that news sentiment prior to the earnings announcement is associated with abnormal returns around the earnings announcement date.

The main contributions of this study are that we extract the information content related to the 'sentiment' from the Chinese news by using linguistic text mining. The quantitative indicators describing the news sentiment (*SR*), media coverage (*MEDIA*) and public news surprises (*SUR*) that are made available to investors prior to earnings announcements are constructed from Chinese financial news. Then the *SR* and *MEDIA* are combined with the financial quantitative dataset to examine the

relationship with abnormal returns and predict post-earnings announcement drift around the announcement period.

The remainder of the paper proceeds as follows. Section 2 provides a literature review. In Section 3, we describe the data, the different proxy variables of public information and the factors for calculating abnormal returns. In Section 4, we present the idea of the research design and regression analysis. Section 5 summarizes the empirical results including the summary statistics, cross-sectional regression analysis and portfolio management. Finally we conclude in Section 6.

2. Literature review

Recent references support the view that the information extracted from text sources might be a useful indicator for the prediction of abnormal returns and could be a key factor in the examination of market reactions.[2] Das and Chen (2007) extract sentiment from text by developing a methodology for extracting small investor sentiment from stock message boards. The algorithms for constructing the sentiment from reports may be used to assess the impact on investor opinion of management announcements, press releases, third-party news and regulatory changes. In the framework of proxy variables of news sentiment, Truong and Shane (2009) introduce a proxy for expected value-relevant earnings: the most optimistic (pessimistic) forecast of earnings that is higher (lower) than the median of all analysts' earnings forecasts over the 90 days prior to the earnings announcement when the median falls short of (exceeds) actual earnings. They find that the new measure of earnings surprise contains unique value-relevant information regarding future earnings, some

[2] For more details, please refer to Vega (2006), Das and Chen (2007), Truong and Shane (2009) and Demers and Vega (2010).

of which generates a statistically significant immediate contemporaneous market response and some of which generates a statistically and economically significant amount of post-earnings announcement drift.

Vega (2006) examines how private and public information received by agents prior to an earnings announcement affects the post-earnings announcement drift. Consistent with rational uncertainty theories, she finds that information acquisition variables that are correlated with the arrival rate of uninformed traders (media and analyst coverage) have a different effect on stock prices than information variables associated with the arrival rate of informed traders (the *PIN* measure and *SUR*). She shows that the more information that (private or public) investors have about the true value of an asset, the more they will agree and trade on this information which results in a smaller abnormal return drift.

Tetlock (2007) measures the interactions between the media and the stock market using daily content from a popular Wall Street Journal column. He finds that high media pessimism predicts downward pressure on market prices followed by a reversion to fundamentals, and unusually high or low pessimism gives rise to predictions of high market trading volume. These results are consistent with the DeLong et al. (1990) and Campbell et al. (1993) theoretical models of noise and liquidity traders. On the other hand, the results are inconsistent with theories of media content due to either a proxy for new information about fundamental asset values, a proxy for market volatility, or an irrelevant noisy variable. Tetlock et al. (2008) further investigate whether the negative words in the financial press lead to low firm earnings forecasts. Stock market prices incorporate the information embedded in negative words with a slight delay. They also demonstrate potential profits from using a simple trading strategy based on the words in a timely news source and find that the profits could easily vanish after accounting for reasonable levels of transaction costs.

Differing from the studies of Tetlock (2007), Tetlock et al. (2008), and Engelberg and Parsons (2011) in that these earlier works examine the news stories in national/local newspapers are associated with substantial price responses. Demers and Vega (2010) examine whether the soft information contained in the texts of management's quarterly earnings press releases is incrementally informative compared to company-issued hard information. The indicator of net optimism is constructed by using several textual-analysis programs over the period 1998 to 2006. These programs provide insight by examining the conditions under which management-expressed soft information is incorporated in prices, both in the short-window announcement period and in the intermediate term, post-announcement drift period. Recently, Engelberg and Parsons (2011) have compared the behavior of investors with access to different media coverage of the same information event. For all earnings announcements of S&P 500 Index firms, they find that local media coverage strongly predicts local trading, after controlling for earnings, investor, and newspaper characteristics.

As mentioned by Engelberg and Parsons (2011), disentangling the impact of media reporting from the impact of the events being reported is challenging. Whether news reports are relevant information or noise and under what conditions the information is associated with the market reaction is worth further analysis. To sum up, this study tries to structure the Chinese text by filtering out the noise words, constructing the representative optimism and pessimism terms and calculating the quantitative proxy variables of Chinese financial news reports. Furthermore, we examine the relationship between news indicators and the announcement drift and investigate its application to portfolio management.

3. Sample and Data Description

Our analysis is conducted on the corporate earnings announcements over the period 2001 to 2005. The study samples are the electronic stocks traded on the Taiwan Stock Exchange (TWSE) and the GreTai Securities Market (GTSM). The financial data are obtained mainly from the quarterly financial statements of listed companies in the Taiwan Economic Journal (TEJ) database and the daily news reports regarding the research company are collected from the InfoTimes database.[3] The proxy variables of public information are measured by the media coverage (*MEDIA*), public news surprises (*SUR*) and news sentiment (*SR*). The variable of standardized unexpected earnings (*SUE*) is considered as the control variable.

3.1 Abnormal returns and cumulative abnormal returns

The abnormal return used in this study is calculated by using the Fama and French (1992) three factors of portfolio beta (*PBeta*), firm size (*SIZE*) and book-to-market value (*BM*). Other asset pricing factors including momentum (*MTM*) (Jegadeesh and Titman, 1993), turnover (*TURN*) and the probability of informed trading (*PIN*) proposed by Easley, Hvidkjaer and O'Hara (2002) are also considered for the robustness check of the relevant information regarding the news sentiment around the earnings press releases. The event date is the quarterly earnings announcement date. The cumulative abnormal returns T days before and after the event date are defined as $CAR_{[-T,-1]}$ and $CAR_{[1,\,T]}$, respectively.

The momentum (*MTM*) variable in each month is obtained in a manner similar to Ku (2005) and is taken as the average returns in months t-7 to t-12. The turnover

[3] A detailed introduction to the TEJ and InfoTimes database can be found at http://www.tej.com.tw/ and http://www.infotimes.com.tw/, respectively.

(*TURN*) is considered to be an explanatory variable for returns. Therefore *TURN* is also included in the model to estimate and calculate the abnormal returns. The other variables for calculating the abnormal returns are summarized as shown below.

3.1.1 The Fama and French Three Factors

We calculate the betas by following the approach of Fama and French (1992), EHO (2002) and Lu and Wong (2009). For each stock, we regress before the test month at least one year, and when possible, two years of monthly stock returns on the contemporaneous and lagged value weighted Taiwan stock index (TAIEX). *Pre-ranking portfolio betas* are then obtained as the sum of the two coefficients. Ten portfolios are sorted every month on the basis of the estimated betas, and monthly portfolio returns are calculated by averaging individual stock returns on an equal-weighted basis. The full sample period portfolio returns are then regressed on contemporaneous and lagged values of TAIEX index returns, from which the two coefficients are summed to give rise to our *post-ranking portfolio betas*, *PBeta*. So, individual stock betas are taken as the post-ranking betas of the portfolio to which they belong. Since the portfolio compositions change each month, the constructed individual stock betas vary over time.

The market value of equity is the price multiplied by the number of outstanding shares as at the end of month *t*. The *SIZE* variable at month *t* is simply defined as the log of the market value of each firm.

TEJ uses the most recently available value of the book value of common equity (reported quarterly) to construct the book-to-market (*BM*) variable. Similar to Fama-French and EHO, firms with negative book values are excluded.

3.1.2 The probability of informed trading (*PIN*)

The distinguishing feature of the Taiwan equity market lies in there being no market maker at the Taiwan Stock Exchange and the market is purely order-driven. So the probability of the informed trading (*PIN*) model by EHO (2002) can be viewed as a microstructure model in which market participants observe trading activities and draw inferences about the underlying true value of an asset. Each trade conveys private information that updates the beliefs of market participants, who in turn set the trading prices based on the updated beliefs. Over time, the process of trading, learning and price setting results in prices converging to full information levels.

<Figure 1 is inserted about here>

Figure 1 depicts the simple sequential trade tree diagram considered by EHO. First, a news event occurs with probability α. There is a $(1 - \delta)$ chance that the news is classified as good, and so a good news event occurs with probability $\alpha(1 - \delta)$. Similarly, a bad news event occurs with probability $\alpha\delta$. Traders are assumed to arrive according to Poisson processes throughout the day. Orders from informed traders arrive at the rate μ whereas orders from uninformed buyers and sellers arrive at the rates ε_b and ε_s, respectively. Basically, informed traders buy if they perceive good news, and sell if the news is perceived to be otherwise. Market participants update their beliefs by extracting information from the buy and sell trades. New prices are formed, trades evolve, and the price dynamics reflects the market participants' changing beliefs. Crucial to the *PIN* model is its estimate of the probability of trade based on private information about the stock. Let *B* and *S* denote the number of buy and sell trades in a single trading day, respectively. We use the mid-quote and transaction price rules of Lee and Ready (1991) and Lu and Wei (2009) to calculate

the required *B* and *S* for each firm on each day.[4] According to EHO, the likelihood function induced by this simple model of the trade process for a single trading day is given below:

$$\begin{aligned} L(\theta \mid B,S) = (1-\alpha)\cdot e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}\cdot e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} \\ + \alpha\delta\cdot e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}\cdot e^{-(\mu+\varepsilon_s)}\frac{(\mu+\varepsilon_s)^S}{S!} \\ + \alpha(1-\delta)\cdot e^{-(\mu+\varepsilon_b)}\frac{(\mu+\varepsilon_b)^B}{B!}\cdot e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!}. \end{aligned} \tag{1}$$

In (1) above, the first line refers to the likelihood weighted by the probability of a day with no event (1 – α), the second line refers to that of a "bad news day" ($\alpha\delta$), and the last line refers to that of a "good news day". Let $\theta = (\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta)$ be the parameter vector and use (B_i, S_i) to denote the buy-sell trade data for period $i \in (1,\ldots,I)$. Assuming that sufficient independence conditions are held across the *I* trading days, the likelihood function for the period is

$$V = L(\theta \mid M) = \prod_{i=1}^{n} L(\theta \mid B_i, S_i), \tag{2}$$

where $M = \{(B_i, S_i)\}_{i=1}^{I}$ refers to the data set. Maximizing the likelihood above will give us the ML estimator for θ, from which we can estimate the probability that the trade is information-based as follows:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s}. \tag{3}$$

3.1.3 The Regression Model for the Calculation of Abnormal Returns

The abnormal returns could be calculated as the actuarial returns minus the expected return which could be expressed as

[4] See also Ellis et al. (2000) for further results on the calculation of buys and sells.

$$AR_{it} = R_{it} - E(\hat{R}_{it}) \quad (4)$$

where AR_{it} is the abnormal return of stock *i* on day *t*, R_{it} is the excess returns of stock *i* on day *i* and $E(\hat{R}_{it})$ is the expected excess returns of stock *i* on day *t*. The expected excess returns used in this study are estimated by the Fama and French three factor models which can be summarized as follows:

$$R_{id} = \gamma_{i0} + \gamma_{i1} PBeta_{pmd} + \gamma_{i2} BM_{imd} + \gamma_{i3} SIZE_{imd} + \eta_{id} \quad (5)$$

where $\gamma_{i0},...,\gamma_{i3}$ are the estimated coefficients of stock *i*, R_{id} is the daily excess return of stock *i* on day *d*, $PBeta_{pmd}$ is the portfolio beta estimated for month *m* to which the trading day *d* belongs, BM_{imd} is the book-to-market value for month *m* to which the trading day *d* belongs, $SIZE_{imd}$ is the log of market value for month *m* to which the trading day *d* belongs, and η_{id} is the mean-zero error term. Then the estimated excess returns can be calculated as

$$E(\hat{R}_{ie}) = \hat{\gamma}_{i0} + \hat{\gamma}_{i1} PBeta_{pme} + \hat{\gamma}_{i2} BM_{ime} + \hat{\gamma}_{i3} SIZE_{ime}. \quad (6)$$

where $E(\hat{R}_{ie})$ is the expected daily excess return of stock *i* on the specific date *e* around the earnings announcement, $PBeta_{pme}$ is the portfolio beta collected for month *m* to which the specific date *e* around the earnings announcement belongs, BM_{ime} is the book-to-market value for month *m* to which the specific date *e* around the earnings announcement belongs, and $SIZE_{ime}$ is the log of market value for month *m* to which the specific date *e* around the earnings announcement belongs.

After the coefficients of the factors are estimated, the cumulative abnormal returns around the earnings announcement could be expressed as shown below.

$$CAR_{i[-t,t]} = \sum_{e=-t}^{t} \left[R_{ie} - E(\hat{R}_{ie}) \right], \; t = 1, 2, ...T. \tag{7}$$

where *T* is the period around the earnings announcement.

We also estimate the expected excess returns with the incorporation of momentum (*MTM*), turnover (*TURN*) and the probability of informed trading (*PIN*) individually for the robustness check. The cross-sectional regression for the calculation of abnormal returns is expressed as in equations (8) to (10).

$$R_{id} = \gamma_{i0} + \gamma_{i1} PBeta_{pmd} + \gamma_{i2} BM_{imd} + \gamma_{i3} SIZE_{imd} + \gamma_{i4} MTM_{imd} + \eta_{id} \tag{8}$$

$$R_{id} = \gamma_{i0} + \gamma_{i1} PBeta_{pmd} + \gamma_{i2} BM_{imd} + \gamma_{i3} SIZE_{imd} + \gamma_{i4} TURN_{imd} + \eta_{id} \tag{9}$$

$$R_{id} = \gamma_{i0} + \gamma_{i1} PBeta_{pmd} + \gamma_{i2} BM_{imd} + \gamma_{i3} SIZE_{imd} + \gamma_{i4} PIN_{imd} + \eta_{id} \tag{10}$$

3.2 Proxy of Public Information

Three proxy variables of public information are constructed by using textual-analysis. All of the news variables are estimated prior to the earnings announcement and divided into different frequencies including 7 days, 14 days, 21 days and 28 days to further identify the relevant information in the public news.

3.2.1 Media Coverage

In referring to Vega (2006), we define media coverage (*MEDIA*) as the number of days a particular firm is mentioned in the news prior to its earnings announcement:

$$MEDIA_{i,t} = \sum_{k=1}^{30} NEWS_{i,t-k} \tag{11}$$

where $NEWS_{i,t-k}$ is a dummy variable equal to one if firm *i* is mentioned in the headline or lead paragraph of an article on day *t*. To measure $MEDIA_{i,t}$ we use the database of InfoTimes which includes the representative Chinese news.

3.2.2 Public News Surprises

In considering the stock market's reaction to headline news, we construct the measure of public news surprises (*SUR*) prior to a firm's earnings announcement by referring to Vega (2006). By incorporating the properties of the Taiwan stock market, *SUR* is adjusted to capture the abnormal return and abnormal market turnover as,

$$SUR_{i,t} = \sum_{k=1}^{30} \{NEWS_{i,t-k} \left[I(AR_{i,t-k} \geq AR_{i,XT}) + I(AR_{i,t-k} \leq AR_{i,XB}) \right]\} \times \left\{ I(DV_{i,t-k} \geq DV_{i,YT}) + I(DV_{i,t-k} \leq DV_{i,YB}) \right\}, \quad (12)$$

where $I(AR_{i,t-k} \geq AR_{i,XT})$ and $I(AR_{i,t-k} \leq AR_{i,XB})$ are indicator functions equal to one if the abnormal stock return for firm *i* on day *t-k* or day *t-k*+1 is above the top *XT*% and below the bottom *XB*%, respectively, of daily abnormal stock returns for that firm. $I(DV_{i,t-k} \geq DV_{i,YT})$ and $I(DV_{i,t-k} \leq DV_{i,YB})$ are indicator functions equal to one if the turnover for firm *i* on day *t-k* or day *t-k*+1 is above the top *YT*% and below the bottom *YB*%, respectively. The percentages of *XT*% and *YT*% are the top 20% and the percentages of *XB*% and *YB*% are the bottom 20%.

3.2.3 News Sentiment (*SR*)

There is no representative system revealing or constructing the information sentiment for Chinese documents. Therefore, we collect the Chinese financial news and conduct the process of information quantification presented in Figure 2.

<Figure 2 is inserted about here>

In referring to Diction, optimism is defined as 'language endorsing some person, group, concept or event or highlighting their positive entailments'. The Diction formula for net optimism is [praise + satisfaction + inspiration] - [blame + hardship + denial] which is the difference between 'optimism' and 'pessimism'. Considering

that there is no certain 'optimism' and 'pessimism' in Chinese documents, we translate the sentiment classification and collect the related Chinese sentiment words.[5] The measure of news sentiment used in this study is calculated as shown below and it represents the meaning of net optimism:

$$SR_{i,t} = \frac{\sum_{j=1}^{p} ptf_{ij,t} - \sum_{j=1}^{n} ntf_{ij,t}}{TF_{i,t}} \times 100\% \quad (13)$$

where $SR_{i,t}$ is the news sentiment of the ith firm at time t, $ptf_{ij,t}$ is the term frequency (tf) of the jth optimism characteristic term of the ith firm at time t, $ntf_{ij,t}$ is the term frequency (tf) of the jth pessimism characteristic term of the ith firm at time t, and $TF_{i,t}$ is the total term frequency of the ith firm at time t.

3.3 Unexpected Earnings Surprise

Following Demers and Vega (2010), we use a seasonal random walk model to capture the unexpected earnings surprise. We define the unexpected earnings as $UE_{iqt}=A_{iqt}-E_{iqt}$, where A_{iqt} is the earnings per share of firm i for fiscal quarter q announced on day t, and E_{iqt} is the proxy for the market's expectation of earnings which can be measured by the last year's same quarter earnings per share for the TEJ sample. We standardize the unexpected earnings by dividing the surprise by the firm-specific standard deviation of the forecast error, and then the standardized unexpected earnings associated with firm i for quarter q at time t is defined as SUE_{iqt}.

4. Research Design

4.1 Pre-Announcement Period Tests

We first investigate the cumulative abnormal returns response to the news

[5] We translate the key words of 'optimism' and 'pessimism' by using the Academia Sinica Bilingual Ontological Wordnet. (http://bow.sinica.edu.tw/)

information surprises contained before the earnings announcement. The dependent variable is defined as the cumulative abnormal returns calculated during the *T*-calendar-days prior to the earnings announcement date which corresponds to the period in which we collect the public financial news.

$$CAR_{iq[-T,-1]} = \beta_0 + \beta_1 SR_{iq[-T,-1]} + \beta_2 MEDIA_{iq[-T,-1]} + \beta_3 SUR_{iq[-T,-1]} + \beta_4 SUE_{iq} + \varepsilon_{iq} \quad (14)$$

where the subscript *iq* refers to firm *i* for event quarter *q*, [-*T*,-1] is the estimated period of the *T*-calendar-days prior to the earnings announcement, and *T* is 7-, 14-, 21-, and 28- calendar days. $CAR_{iq[-T,-1]}$ is calculated based on the expected abnormal return of actual trading days during the *T*-calendar-days prior to the earnings announcement. The $SR_{iq[-T,-1]}$, $MEDIA_{iq[-T,-1]}$ and $SUR_{iq[-T,-1]}$ are proxies of public information calculated *T*-calendar-days prior to the earnings announcement since news reports might be published on non-trading dates. SUE_{iq} is the standardized unexpected earnings associated with firm *i* for quarter *q*.

4.2 Post-Announcement Period Tests

We examine the leading effect of the news information surprises by analyzing the relationship between news information collected 7, 14, 21, and 28 calendar days prior to the earnings announcement day and the cumulative abnormal returns after the earnings announcement. The post-announcement drift is calculated from 1 to 22 trading days after the announcement date. This leads to the following pooled regression model:

$$CAR_{iq[1,T']} = \beta_0 + \beta_1 SR_{iq[-T,-1]} + \beta_2 MEDIA_{iq[-T,-1]} + \beta_3 SUR_{iq[-T,-1]} + \beta_4 SUE_{iq} + \varepsilon_{iq} \quad (15)$$

where the subscript *iq* refers to firm *i* for event quarter *q*, and [-*T*,-1] is the estimated period of the news variables *T*-calendar-days prior to the earnings announcement,

where T is 7, 14, 21, and 28 days. $CAR_{iq[1,T']}$ is the cumulative abnormal return estimated during the post-announcement period and T' equals 1 to 22 trading days.

4.3 Portfolio Management

If we can confirm whether the Chinese news reports contain relevant information content, we can then provide the average excess returns for the portfolios of stocks sorted independently by news sentiment (*SR*) and the public news surprises (*SUR*). For each event quarter, stocks are sorted into five *SR* groups, ranging from low to high, and five *SUR* groups, also ranging from low to high. The portfolio excess returns are averages of the relevant variables over the sample period.

5. Empirical Results

5.1 Summary Statistics

Before analyzing the relationship between the news information and the pre-/post-earnings announcement drift, we first show the summary statistics of the variables. To further classify the relevance of news information before the earnings announcement, we construct different frequencies of the proxies of news information including 7, 14, 21, and 28 calendar days prior to the earnings announcement. Table 1 presents the minimum, maximum, mean, median and standard deviation of the soft information and the control variable. Table 2 shows the correlation coefficient between the news information and the cumulative abnormal return during the pre-earnings announcement.

<Table 1 is inserted about here>

<Table 2 is inserted about here>

5.2 Pre-Announcement Period Tests

In this section we examine the market's response to the unexpected information in managerial earnings announcements including the news information and control variable. We first present the explanatory power of the news information collected prior to the earnings announcement in relation to the pre-earnings announcement drift by using equation (14). The empirical results of the pre-earnings announcement drift are shown in Table 3. The cumulative abnormal returns during the pre-announcement periods are calculated based on the Fama and French three-factor model expressed as equations (5) to (7).

<Table 3 is inserted about here>

Panel A in Table 3 presents the cross-sectional regression analysis 7 calendar days prior to the earnings announcement. It means that news information, including *MEDIA*, *SUR* and *SR*, are constructed by collecting public news for one week, or 7 calendar days, prior to the earnings announcement. The empirical results of model (1) in Panel A show that the standardized unexpected earnings associated with a firm simplified as *SUE* is negatively significant in regard to the pre-earnings announcement drift. In other words, there is a significant negative relationship between the *SUE* and the cumulative abnormal return regardless of whether the study period prior to the earnings announcement is one week, two weeks, three weeks, or one month.

For the proxies of the news information, *MEDIA* and *SUR* belong to the degree of the disclosure of the specific firm on the public news. Models (3) and (4) in Panel A reveal that *MEDIA* and *SUR* exhibit no significant explanatory power in relation to the pre-earnings announcement drift while *MEDIA* and *SUR* are included in the model

individually. Model (2) in Panel A is used to derive the empirical results of the proxy of news sentiment, the net optimism of the specific firm revealed in the public news. The analysis shows that there is positive relationship between the net optimism and the cumulative abnormal returns during 7 to 14 days prior to the earnings announcement. The higher the degree of the net optimism, the higher the cumulative abnormal return. Model (5) in Panel A compares the explanatory power of the news information and *SUE* prior to the earnings announcement day. While *SR* and *SUE* are still significant in relation to the pre-earnings announcement drift, *MEDIA* and *SUR* are, however, not significant.

Panel B, Panel C, and Panel D show the cross-sectional regression analysis two weeks, three weeks, and one month prior to the earnings announcement day. 14 days of *SR* and *SUE* as presented in Panel B are still significant to the pre-earnings announcement drift. However if the information collection days are extended to 21 and 28 days, the explanatory power of *SR* is not consistent. On the other hand, *MEDIA* and *SUR* contain relevant information during the 21 to 28 days prior to the earnings announcement.

5.3 Post-Announcement Period Tests

We examine the relationship between the news information and the post-earnings announcement drift by using equation (15). To further investigate the signal effect of the news information on the post-earnings announcement drift, we use the 1-day to 22-day post-announcement period to calculate the cumulative abnormal return. The significance of the news information prior to the earnings announcement to the 1-day to 22-day post-announcement market response is presented in Figure 3.

<Figure 3 is inserted about here>

Panel A to Panel D in Figure 3 represent the relationship between the one week to one month of news information prior to the earnings announcement and the one day to 22 days of post-earnings announcement drift. We can find that the leading effect of the news sentiment (*SR*) constructed based on the one week public news prior to the earnings announcement continues for two weeks after the earnings announcement. The news sentiment collected two weeks, three weeks or one month prior to the earnings announcement presents a similar leading effect.

The *MEDIA* and *SUR* variables constructed one week, two weeks, three weeks or one month prior to the earnings announcement exhibit a significant leading effect. However the findings for *MEDIA* and *SUR* are the opposite. There is a positive (negative) relationship between *MEDIA* (*SUR*) and the post-earnings announcement drift and the relationship is significant one to two weeks after the earnings announcement.

5.4 Robustness Check

The regression analysis during the pre-announcement period is analyzed for the robustness check by using the Fama and French three-factors model with the incorporation of *MTM*, *TURN* or *PIN* individually. The news information used in the robustness check is calculated by using the 7 days of public information. The empirical results of news the information constructed based on the 14, 21, and 28 days prior to the earnings announcement lead to similar results and, for space considerations, we omit the details; however, the results of the regression analysis of the other frequencies of the news information are available from the authors upon request. Table 4 presents the results of the regression analysis of the news information and cumulative abnormal returns prior to the earnings announcement with the

incorporation of other asset pricing factors.

<Table 4 is inserted about here>

Panel A to Panel C in Table 4 present the explanatory power of news information and the standardized unexpected earnings in relation to the cumulative abnormal returns calculated based on the Fama and French three-factors model with the incorporation of *MTM*, *PIN* and *TURN* for the robustness check. The findings confirm the positive relationship between *SR* and the cumulative abnormal returns. The leading effect of the news information on the cumulative abnormal returns during the post-announcement period with the incorporation of *MTM*, *PIN* and *TURN* is analyzed and the results are presented in Figure 4.

<Figure 4 is inserted about here>

Panel A to Panel C in Figure 4 depict the relationship between the one week news information prior to the earnings announcement and the one day to one month post-earnings announcement drift which is calculated with the incorporation of *MTM*, *PIN* and *TURN* individually. We can find that the leading effect of the news sentiment (*SR*), media coverage (*MEDIA*) and public news surprises (*SUR*) constructed based on the one week public news prior to the earnings announcement continues to two weeks after the earnings announcement. The findings further confirm that the news reports contain relevant information for the prediction of cumulative abnormal returns during the earnings announcement.

5.5 Application of Portfolio Management

Table 5 shows the average excess returns for portfolios of stocks sorted independently by news sentiment (*SR*) and public news surprises (*SUR*) to confirm the application of the information sentiment.

<Table 5 is inserted about here>

Table 5 contains results for portfolios of stocks sorted independently on each earnings announcement quarter by information sentiment (*SR*) constructed based on the linguistic text mining and public news surprises (*SUR*). *SR* and *SUR* are calculated based on news reports 7 days prior to the earnings announcement. The holding periods of the portfolio are 5, 10, 15 and 22 trading days after the earnings announcement and Panel A to Panel D shown in Table 5 present the portfolio excess returns individually. The application of portfolio management supports the empirical results of Figure 3 that *SR* and *SUR* could be reference indicators of portfolio management and the holding periods are suggested to be one to two weeks after the earnings announcement. The application of public news sentiment to portfolio management suggests that “long stocks with low news sentiment and high public news surprises” and “short stocks with high news sentiment and high public news surprises” earn positive excess returns.

6. Conclusions

This paper examines the information content of Chinese news information on the announcement drift and investigates its application to portfolio management. The indicators of public information are constructed by using linguistic text mining in

referring to Vega (2006) and Demers and Vega (2010). This study confirms whether the public information is news or noise.

The research problems in this present study focus on extending the traditional event study of the earnings announcement to probe the signals of information sentiment by digging out a large number of financial news items. We empirically measure the effect of news information on the post-announcement drift. The proxy variables of news information are measured by the media coverage (*MEDIA)*, public news surprises (*SUR*) and news sentiment (*SR*). The Fama and French (1992) 3 factors, momentum (Jegadeesh and Titman, 1993), turnover (*TURN*) and the probability of informed trading (*PIN*) proposed by Easley, Hvidkjaer and O'Hara (2002) are adopted as the factors to calculate the abnormal return around the earnings announcement date.

The empirical results show that there is positive relationship between *SR* and the cumulative abnormal return before the earnings announcement which confirms that the market response covers the relevant information regarding the company. The application of news sentiment to the portfolio management suggests that "long stocks with low news sentiment and high public news surprises" and "short stocks with high news sentiment and high public news surprises" earn positive excess returns. This study proposes that the information sentiment of individual stocks could be applied to the warning model of abnormal returns and portfolio management.

References

Campbell, J. Y., S. J. Grossman, and J. Wang, 1993, Trading Volume and Serial Correlation in Stock Returns, *Quarterly Journal of Economics*, 108, 905–939.

Das, S. and M. Chen, 2007, Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, *Management Science*, 53, 1375-1388.

Demers, E. and C. Vega, 2010, Soft Information in Earnings Announcements: News or Noise? *Working paper*, International Finance Discussion Papers, Board of Governors of the Federal Reserve System, No. 951.

De Long, J.B., A. Shleifer, L.G. Summers, and R.J. Waldmann, 1990, Noise Trader Risk in Financial Markets, *Journal of Political Economy*, 98, 703–738.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is Information Risk a Determinant of Asset Returns? *Journal of Finance*, 57(5), 2185-2221.

Ellis, K., R. Michaely, and M. O'Hara, 2000, The Accuracy of Trade Classification Rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis*, 35, 529–551.

Engelberg, J. and C. Parsons, 2011, The Causal Impact of Media in Financial Markets, *Journal of Finance*, 66, 67-97.

Fama, E. F. and K. R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance*, 47, 427-286.

Jegadeesh, N. and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, 48, 65-91.

Ku, K. P., 2005, One-factor, Three-factor, or Four-factor Models? *Review of Securities and Futures Markets*, 17, 101-146.

Lee, C. M. C. and M. J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance*, 46, 733-746.

Lu, Y. C. and W. K. Wong, 2009, Probability of Information-Based Trading as a Pricing Factor in the Taiwan Stock Market, *International Research Journal of Finance and Economics*, Forthcoming.

Lu Y. C. and Y. C. Wei, 2009, Classification of Trade Direction for an Equity Market with Price Limit and Order Match: Evidence from the Taiwan Stock Market, *Investment Management and Financial Innovations*, 6 (2), 23-35.

Tetlock, P. C., 2007, Giving Content to Investors Sentiment: The Role of Media in the Stock Market, *Journal of Finance*, 62, 1139-1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than Words: Quantifying Language to Measure Firms' Fundamentals, *Journal of Finance*, 63(3), 1437-1467.

Truong, C. and P. B. Shane, 2009, Stock Market Efficiency with Respect to a New Measure of Earnings News.

Available at SSRN: http://ssrn.com/abstract=1472885.

Vega, C., 2006, Stock Price Reaction to Public and Private Information, *Journal of Financial Economics*, 82, 103-133.

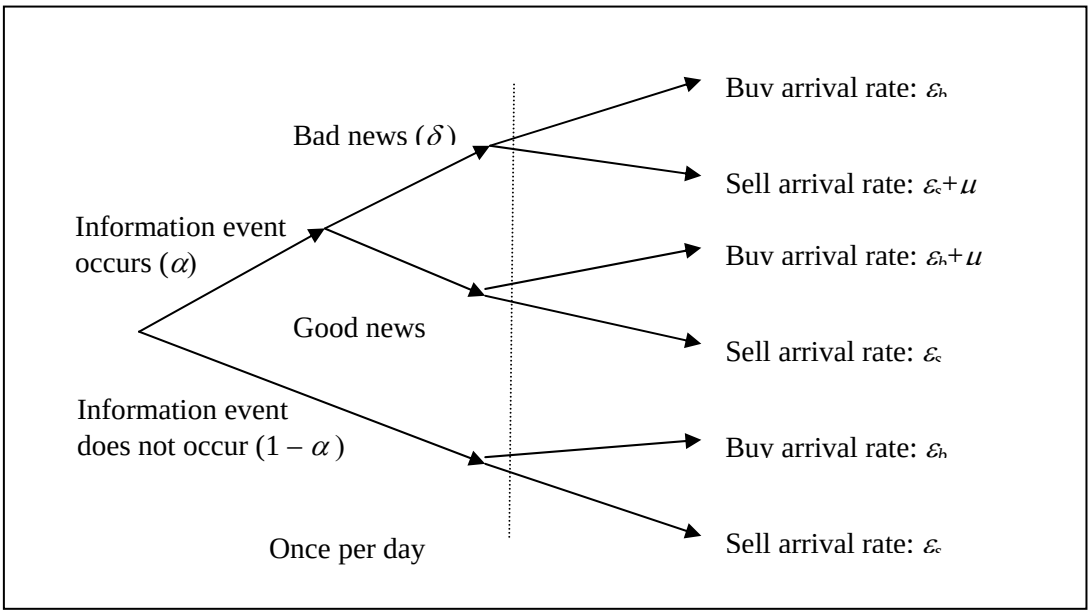


Figure 1 Microstructure tree diagram of the trading process

Notes: In the diagram above, α refers to the probability of an information event whereas δ refers to the probability of a bad news event. μ, ε_b and ε_s are the arrival rates of informed trade, uninformed buy and uninformed sell orders, respectively. Events to the left of the dotted line occur once per day.

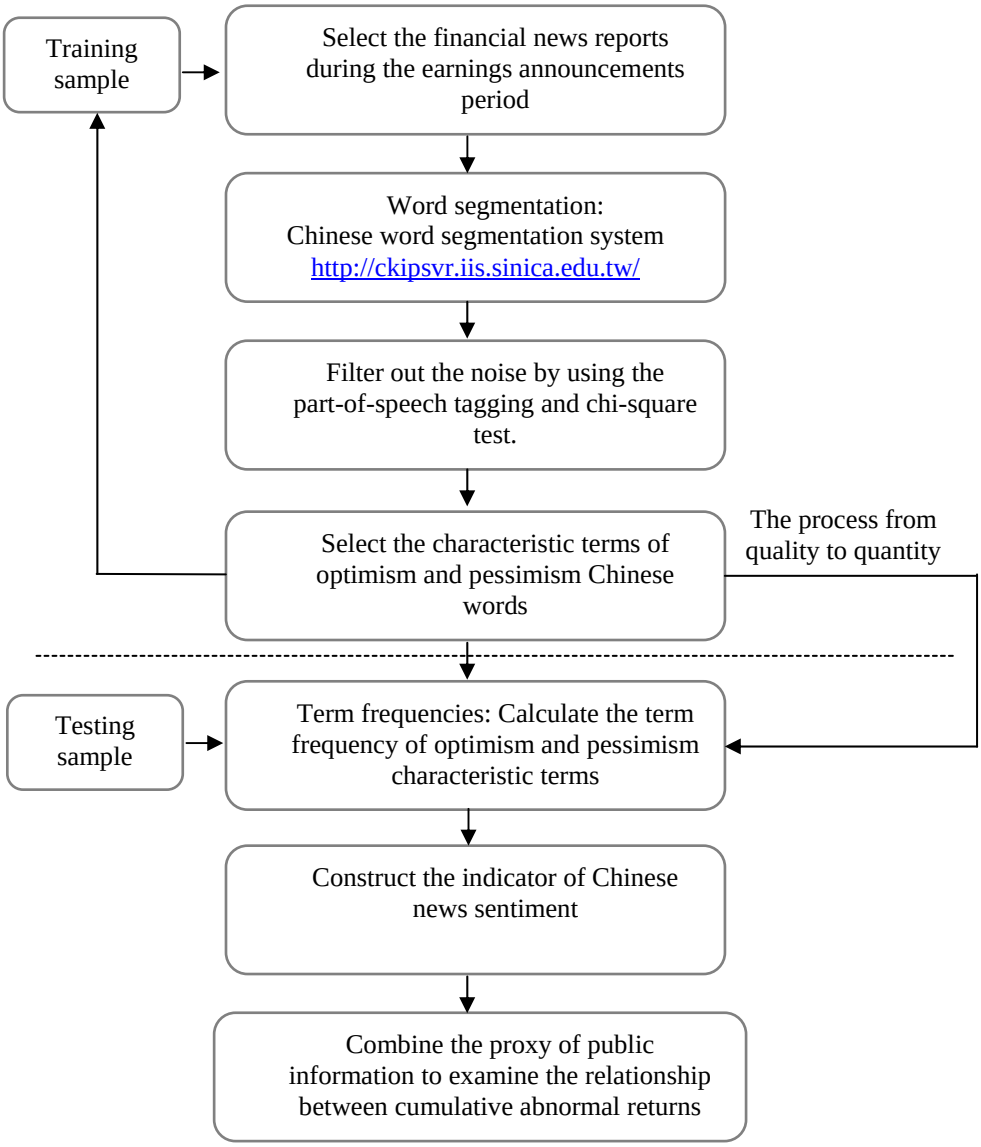
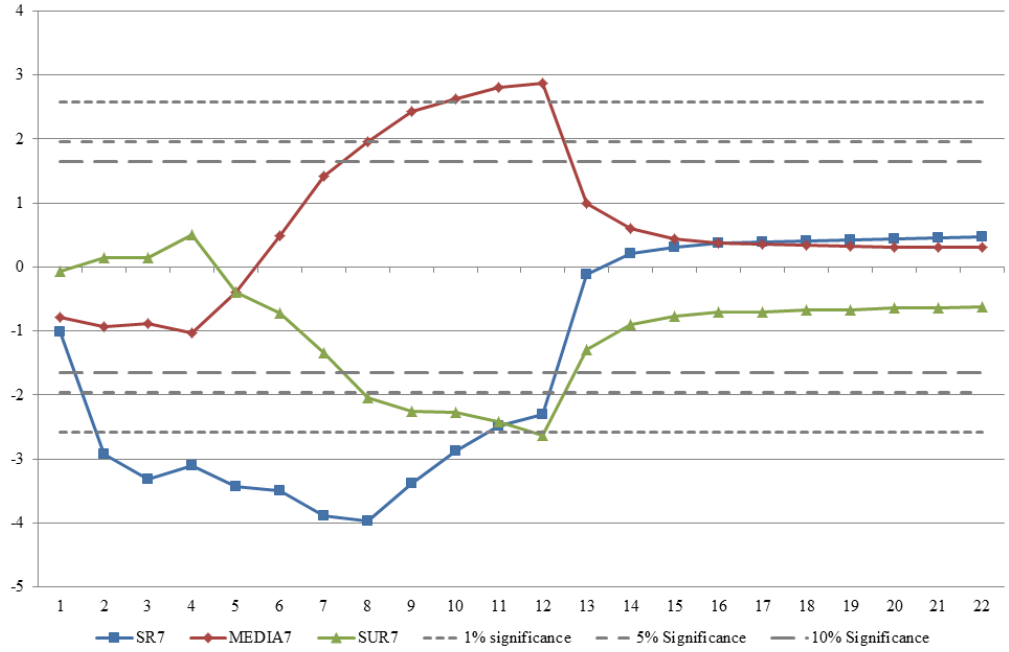


Figure 2 Diagram of the process of information quantification

Panel A The relationship between the news information 7 days prior to the earnings announcement and the post-earnings announcement drift



Panel B The relationship between the news information 14 days prior to the earnings announcement and the post-earnings announcement drift

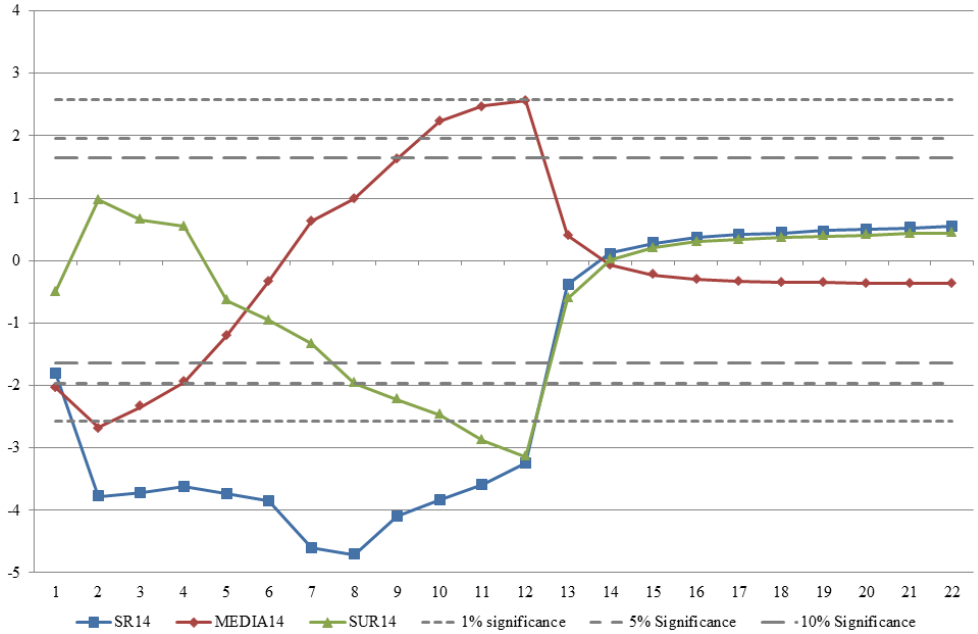
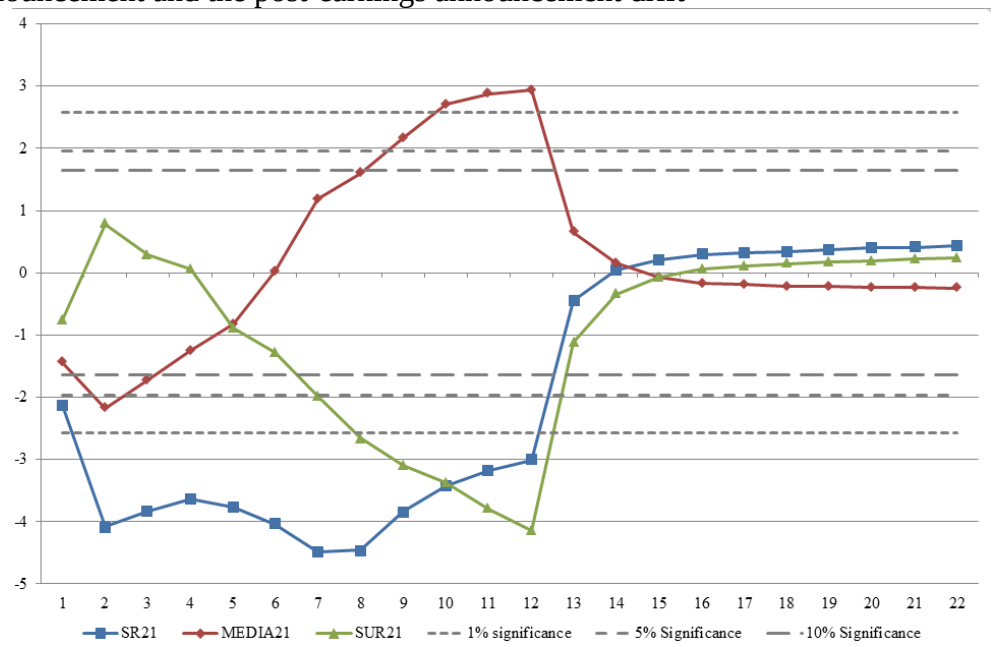


Figure 3 The relationship between the private and public information and the post-earnings announcement drift

Panel C The relationship between the news information 21 days prior to the earnings announcement and the post-earnings announcement drift



Panel D The relationship between the news information 28 days prior to the earnings announcement and the post-earnings announcement drift

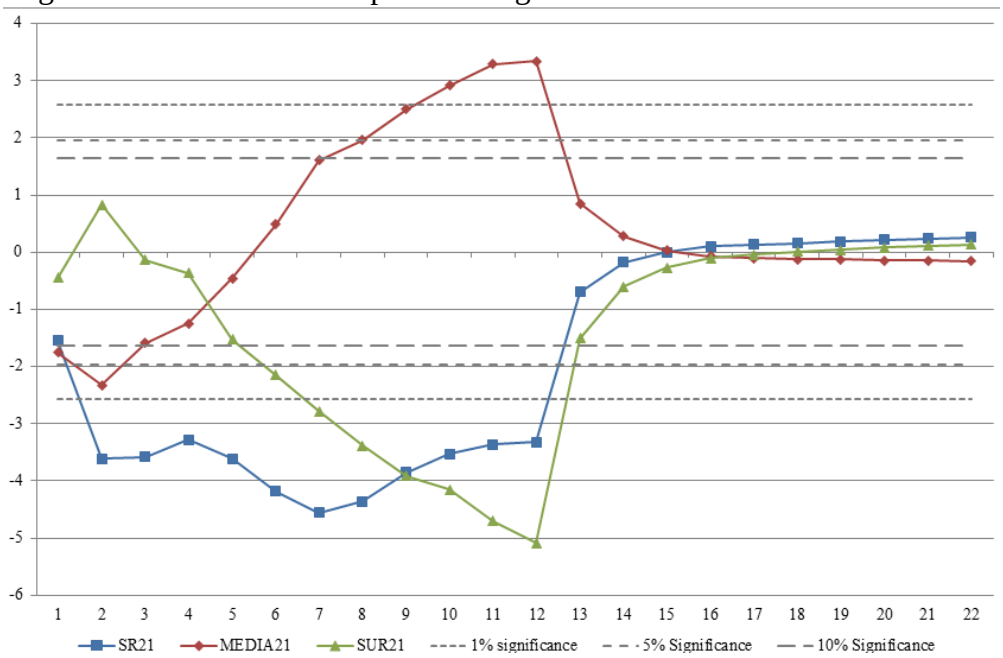
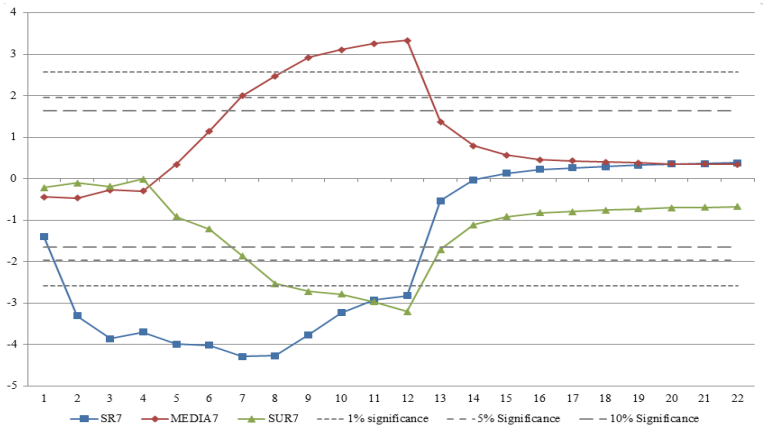
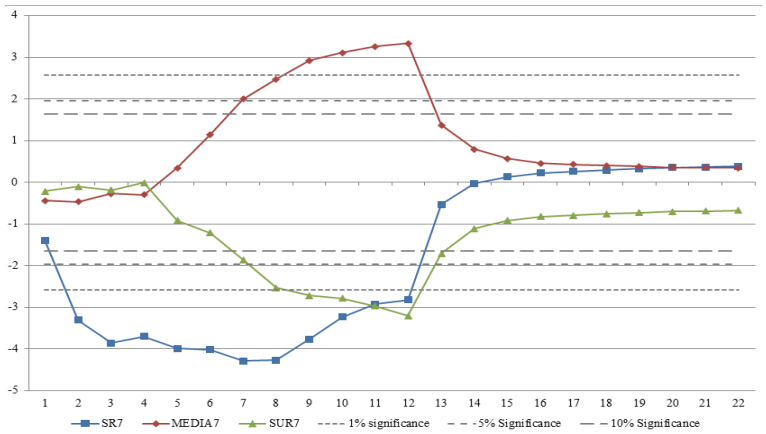


Figure 3 The relationship between the private and public information and the post-earnings announcement drift (Cont'd.)

Panel A The relationship between the news information 7 days prior to the earnings announcement and the post-earnings announcement drift with the incorporation of the Fama and French three factors and momentum



Panel B The relationship between the news information 7 days prior to the earnings announcement and the post-earnings announcement drift with the incorporation of the Fama and French three factors and PIN



Panel C The relationship between the news information 7 days prior to the earnings announcement and the post-earnings announcement drift with the incorporation of the Fama and French three factors and turnover

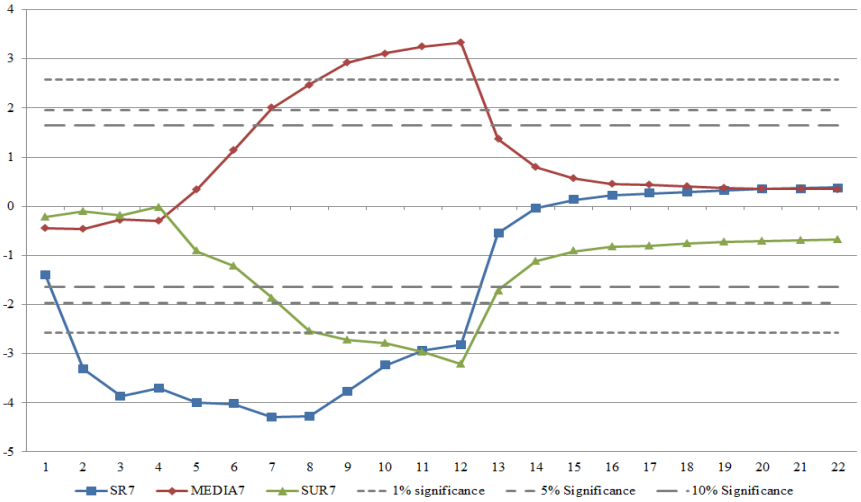


Figure 4 The relationship between the news information and the post-earnings announcement drift for the robustness check

Table 1 Summary Statistics

	Min	Max	Mean	Median	Std Dev	Sample Size
Panel A: Sample Data 7 days prior to the earnings announcement						
SUE	-3.09	2.86	-0.16	-0.08	0.92	911
$SR_{[-7,-1]}$	-44.94	47.14	10.41	9.41	10.91	
$MEDIA_{[-7,-1]}$	0.00	4.33	0.99	1.00	0.56	
$SUR_{[-7,-1]}$	0.00	23.00	1.95	1.00	2.19	
$CAR_{[-7,-1]}$	-28.65	25.53	-0.93	-1.16	7.53	
Panel B: Sample Data 14 days prior to the earnings announcement						
SUE	-3.13	2.86	-0.16	-0.08	0.93	1235
$SR_{[-14,-1]}$	-44.94	50.53	11.14	10.63	10.15	
$MEDIA_{[-14,-1]}$	0.00	3.50	1.00	1.00	0.48	
$SUR_{[-14,-1]}$	0.00	35.00	2.64	2.00	2.96	
$CAR_{[-14,-1]}$	-60.39	46.61	0.45	-0.50	11.98	
Panel C: Sample Data 21 days prior to the earnings announcement						
SUE	-3.13	2.86	-0.17	-0.09	0.92	1461
$SR_{[-21,-1]}$	-35.00	47.14	11.58	11.11	9.37	
$MEDIA_{[-21,-1]}$	0.00	4.00	0.98	1.00	0.44	
$SUR_{[-21,-1]}$	0.00	62.00	3.38	2.00	4.22	
$CAR_{[-21,-1]}$	-76.46	115.94	1.13	-0.66	16.00	
Panel D: Sample Data 28 days prior to the earnings announcement						
SUE	-3.13	2.86	-0.18	-0.10	0.91	1608
$SR_{[-28,-1]}$	-35.00	46.67	11.67	11.51	9.11	
$MEDIA_{[-28,-1]}$	0.00	4.27	0.97	1.00	0.42	
$SUR_{[-28,-1]}$	0.00	89.00	4.05	2.00	5.35	
$CAR_{[-28,-1]}$	-102.25	148.83	1.54	0.08	17.41	

Notes: *SUE* is the unexpected earnings surprises. *MEDIA* is the media coverage and *SUR* is the measure of public news surprises considering the stock market's reaction to headline news. *SR* is the news sentiment. $CAR_{[-7,1]}$, $CAR_{[-14,1]}$, $CAR_{[-21,1]}$, $CAR_{[-28,-1]}$ are the cumulative abnormal returns 7, 14, 21, and 28 days prior to the earnings announcement, respectively.

Table 2 Correlation Analysis

Panel A: 7 days prior to the earnings announcement				
	SRC7	$MEDIA_{[-7,-1]}$	$SUR_{[-7,-1]}$	$CAR_{[-7,-1]}$
SUE	0.1933***	0.0457	0.0442	-0.0762**
$SR_{[-7,-1]}$		0.114***	-0.0141	0.1421***
$MEDIA_{[-7,-1]}$			0.645***	0.0224
$SUR_{[-7,-1]}$				0.0175
Panel A: 14 days prior to the earnings announcement				
	$SRC_{[-14,-1]}$	$MEDIA_{[-14,-1]}$	$SUR_{[-14,-1]}$	$CAR_{[-14,-1]}$
SUE	0.1892***	0.0484*	0.0421	-0.1339***
$SR_{[-14,-1]}$		0.1073***	-0.0584**	0.038
$MEDIA_{[-14,-1]}$			0.484***	0.0253
$SUR_{[-14,-1]}$				-0.0206
Panel A: 21 days prior to the earnings announcement				
	$SRC_{[-21,-1]}$	$MEDIA_{[-21,-1]}$	$SUR_{[-21,-1]}$	$CAR_{[-21,-1]}$
SUE	0.2049***	0.0644**	0.0567**	-0.1496***
$SR_{[-21,-1]}$		0.1333***	-0.0631**	-0.0403
$MEDIA_{[-21,-1]}$			0.4559***	0.0254
$SUR_{[-21,-1]}$				-0.0501*
Panel A: 28 days prior to the earnings announcement				
	$SRC_{[-28,-1]}$	$MEDIA_{[-28,-1]}$	$SUR_{[-28,-1]}$	$CAR_{[-28,-1]}$
SUE	0.2155***	0.0606**	0.0391	-0.1735***
$SR_{[-28,-1]}$		0.1369***	-0.0509**	-0.0841***
$MEDIA_{[-28,-1]}$			0.4435***	0.0215
$SUR_{[-28,-1]}$				-0.017

Notes: *SUE* is the unexpected earnings surprises. *MEDIA* is the media coverage and *SUR* is the measure of public news surprises considering the stock market's reaction to headline news. *SR* is the news sentiment. The *, **, and *** symbols denote significance at the 10%, 5% and 1% levels, respectively.

Table 3 Cross-Sectional Regression Analysis Prior to the Earnings Announcement

	(1)	(2)	(3)	(4)	(5)
Panel A: Cross-Sectional Regression Analysis 7 days prior to the earnings announcement					
C	-1.0314***	-2.2443***	-1.3775***	-1.1731***	-2.314***
	(-4.0805)	(-6.3812)	(-2.7048)	(-3.4707)	(-4.2456)
SUE	-0.6246**	-0.8828***	-0.6343**	-0.6322**	-0.8934***
	(-2.3044)	(-3.2358)	(-2.3373)	(-2.3293)	(-3.2682)
$SR_{[-7,-1]}$		0.1126***			0.114***
		(4.8956)			(4.8915)
$MEDIA_{[-7,-1]}$			0.3471		-0.169
			(0.7829)		(-0.2911)
$SUR_{[-7,-1]}$				0.0719	0.1127
				(0.6319)	(0.7609)
Panel B: Cross-Sectional Regression Analysis 14 days prior to the earnings announcement					
C	0.1693	-0.7206	-0.6371	0.3308	-1.3381
	(0.4935)	(-1.3912)	(-0.8039)	(0.721)	(-1.5686)
SUE	-1.7275***	-1.8879***	-1.7474***	-1.7194***	-1.8814***
	(-4.7452)	(-5.1008)	(-4.7948)	(-4.7173)	(-5.0751)
$SR_{[-14,-1]}$		0.0776**			0.0701**
		(2.29)			(2.0418)
$MEDIA_{[-14,-1]}$			0.8028		1.0299
			(1.1288)		(1.2552)
$SUR_{[-14,-1]}$				-0.0607	-0.1247
				(-0.5302)	(-0.9473)

Notes: *SUE* is the unexpected earnings surprises. *MEDIA* is the media coverage and *SUR* is the measure of public news surprises considering the stock market's reaction to headline news. *SR* is the news sentiment. Values in the table and the parentheses are regression betas and *T*-statistics, respectively. The *, **, and *** symbols denote significance at the 10%, 5% and 1% levels, respectively.

Table 3 Cross-Sectional Regression Analysis prior to the earnings announcement (Cont'd.)

	(1)	(2)	(3)	(4)	(5)
Panel C: Cross-Sectional Regression Analysis 21 days prior to the earnings announcement					
C	0.6787	0.8842	-0.5814	1.2202**	-0.4269
	(1.6109)	(1.2916)	(-0.5704)	(2.2638)	(-0.3867)
SUE	-2.6113***	-2.5752***	-2.6508***	-2.57***	-2.5273***
	(-5.7807)	(-5.5783)	(-5.8578)	(-5.6834)	(-5.4732)
$SR_{[-21,-1]}$		-0.0172			-0.0431
		(-0.3809)			(-0.9379)
$MEDIA_{[-21,-1]}$			1.2744		2.6464**
			(1.3575)		(2.4764)
$SUR_{[-21,-1]}$				-0.1581	-0.2911***
				(-1.6094)	(-2.6137)
Panel D: Cross-Sectional Regression Analysis 28 days prior to the earnings announcement					
C	0.9395**	2.0684***	-0.3454	1.0758**	0.5458
	(2.1541)	(2.8513)	(-0.3206)	(1.9721)	(0.4661)
SUE	-3.3153***	-3.1137***	-3.3524***	-3.3076***	-3.1101***
	(-7.0601)	(-6.4808)	(-7.1277)	(-7.0366)	(-6.4693)
$SR_{[-28,-1]}$		-0.0936*			-0.1122**
		(-1.9466)			(-2.297)
$MEDIA_{[-28,-1]}$			1.3168		2.3189**
			(1.3043)		(2.0323)
$SUR_{[-28,-1]}$				-0.0333	-0.1259
				(-0.4162)	(-1.4015)

Notes: *SUE* is the unexpected earnings surprises. *MEDIA* is the media coverage and *SUR* is the measure of public news surprises considering the stock market's reaction to headline news. *SR* is the news sentiment. Values in the table and the parentheses are regression betas and *T*-statistics, respectively. The *, **, and *** symbols denote significance at the 10%, 5% and 1% levels, respectively.

Table 4 Cross-Sectional Regression Analysis prior to the earnings announcement for the robustness check

	(1)	(2)	(3)	(4)	(5)
Panel A：Fama and French 3 factors with *MTM*					
C	-1.0331***	-2.0439***	-1.5803***	-1.1904***	-2.3707***
	(-3.9596)	(-5.611)	(-3.0125)	(-3.4218)	(-4.1948)
SUE	-0.6792**	-0.898***	-0.6962**	-0.6873**	-0.9088***
	(-2.4381)	(-3.1861)	(-2.4964)	(-2.4643)	(-3.2185)
$SR_{[-7,-1]}$		0.0943***			0.0935***
		(3.9433)			(3.8534)
$MEDIA_{[-7,-1]}$			0.547		0.2227
			(1.2023)		(0.3705)
$SUR_{[-7,-1]}$				0.079	0.0567
				(0.684)	(0.3734)
Panel B：Fama and French 3 factors with PIN					
C	-1.0331***	-2.0439***	-1.5803***	-1.1904***	-2.3707***
	(-3.9596)	(-5.611)	(-3.0125)	(-3.4218)	(-4.1948)
SUE	-0.6792**	-0.898***	-0.6962**	-0.6873**	-0.9088***
	(-2.4381)	(-3.1861)	(-2.4964)	(-2.4643)	(-3.2185)
$SR_{[-7,-1]}$		0.0943***			0.0935***
		(3.9433)			(3.8534)
$MEDIA_{[-7,-1]}$			0.547		0.2227
			(1.2023)		(0.3705)
$SUR_{[-7,-1]}$				0.079	0.0567
				(0.684)	(0.3734)
Panel C：Fama and French 3 factors with *TURN*					
C	-1.0331***	-2.0439***	-1.5803***	-1.1904***	-2.3707***
	(-3.9596)	(-5.611)	(-3.0125)	(-3.4218)	(-4.1948)
SUE	-0.6792**	-0.898***	-0.6962**	-0.6873**	-0.9088***
	(-2.4381)	(-3.1861)	(-2.4964)	(-2.4643)	(-3.2185)
$SR_{[-7,-1]}$		0.0943***			0.0935***
		(3.9433)			(3.8534)
$MEDIA_{[-7,-1]}$			0.547		0.2227
			(1.2023)		(0.3705)
$SUR_{[-7,-1]}$				0.079	0.0567
				(0.684)	(0.3734)

Notes: *SUE* is the unexpected earnings surprises. *MEDIA* is the media coverage and *SUR* is the measure of public news surprises considering the stock market's reaction to headline news. *SR* is the news sentiment. Values in the table and the parentheses are regression betas and *T*-statistics, respectively. The *, **, and *** symbols denote significance at the 10%, 5% and 1% levels, respectively.

Table 5 Portfolio Excess Returns Using Information Sentiment and Public News Surprises

Panel A：Hold the portfolio 5 trading days after the earnings announcement

SR/*SUR*	Low	2	3	4	High	0	Low	2	3	4	High
Panel A: Excess Returns						Panel B: Number of Stocks					
Low	-0.19	0.29	-3.42	-3.71	0.43	Low	4.17	2.67	3.67	2.67	3.00
2	-3.67	-0.84	-0.57	-2.72	-1.10	2	3.50	2.67	2.33	3.33	4.33
3	-4.32	0.07	-2.25	-0.46	-1.30	3	3.00	3.50	2.67	2.83	4.17
4	-4.34	-1.99	-4.75	-1.09	-1.37	4	2.83	3.83	2.67	3.67	3.17
High	-1.15	-0.06	-2.36	-0.40	-6.03	High	2.67	3.50	4.83	3.67	3.83
Panel C: *SR*						Panel D: *SUR*					
Low	-1.19	-3.09	-2.56	-1.74	-2.32	Low	0.41	0.94	1.19	2.28	4.36
2	6.12	7.00	6.25	5.89	5.35	2	0.38	0.88	1.06	2.07	6.90
3	11.41	10.79	10.98	9.83	10.75	3	0.58	1.00	1.28	1.97	5.18
4	15.91	15.91	15.65	15.21	15.79	4	0.58	0.92	1.11	2.09	4.26
High	25.08	25.22	25.93	23.25	22.32	High	0.50	0.92	1.19	1.98	3.74

Panel B：Hold the portfolio 10 trading days after the earnings announcement

SR/*SUR*	Low	2	3	4	High		Low	2	3	4	High
Panel A: Excess Returns						Panel B: Number of Stocks					
Low	-2.52	1.79	-5.03	-4.05	2.44	Low	4.17	2.67	3.67	2.67	3.00
2	-6.23	-3.42	-3.11	-2.58	-3.61	2	3.50	2.67	2.33	3.33	4.33
3	-4.87	0.42	-4.50	0.51	-3.10	3	3.00	3.50	2.67	2.83	4.17
4	-4.02	-1.48	-5.65	-2.12	-1.58	4	2.83	3.83	2.67	3.67	3.17
High	-2.48	-1.93	-3.95	-7.52	-5.88	High	2.67	3.50	4.83	3.67	3.83
Panel C: *SR*						Panel D: *SUR*					
Low	-1.19	-3.09	-2.56	-1.74	-2.32	Low	0.41	0.94	1.19	2.28	4.36
2	6.12	7.00	6.25	5.89	5.35	2	0.38	0.88	1.06	2.07	6.90
3	11.41	10.79	10.98	9.83	10.75	3	0.58	1.00	1.28	1.97	5.18
4	15.91	15.91	15.65	15.21	15.79	4	0.58	0.92	1.11	2.09	4.26
High	25.08	25.22	25.93	23.25	22.32	High	0.50	0.92	1.19	1.98	3.74

Notes: This table contains results for portfolios of stocks sorted independently on each earnings announcement quarter by information sentiment (*SR*) constructed based on the linguistic text mining and *SUR*. The holding periods of the excess returns are one to four weeks after the earnings announcement. The reported results are averages of the relevant variables over the sample period 2001 to 2005 for the Taiwan stock market. Panels A, B, C and D respectively report in each portfolio the average excess returns, the average number of stocks, and the average *SR* and *SUR* of the stocks.

Table 5 Portfolio Excess Returns Using Information Sentiment and Public News Surprises (Cont'd.)

Panel C: Hold the portfolio 15 trading days after the earnings announcement											
SR/*SUR*	Low	2	3	4	High		Low	2	3	4	High
Panel A: Excess Returns							Panel B: Number of Stocks				
Low	-2.15	-1.61	-5.91	-9.87	-0.09	Low	4.17	2.67	3.67	2.67	3.00
2	-6.42	-4.77	-9.92	-8.15	-4.67	2	3.50	2.67	2.33	3.33	4.33
3	-6.12	-2.68	-7.18	-1.94	-7.20	3	3.00	3.50	2.67	2.83	4.17
4	-5.90	-6.66	-4.45	-2.23	-4.88	4	2.83	3.83	2.67	3.67	3.17
High	-7.83	-6.13	190.35	-15.16	-10.66	High	2.67	3.50	4.83	3.67	3.83
Panel C: *SR*							Panel D: *SUR*				
Low	-1.19	-3.09	-2.56	-1.74	-2.32	Low	0.41	0.94	1.19	2.28	4.36
2	6.12	7.00	6.25	5.89	5.35	2	0.38	0.88	1.06	2.07	6.90
3	11.41	10.79	10.98	9.83	10.75	3	0.58	1.00	1.28	1.97	5.18
4	15.91	15.91	15.65	15.21	15.79	4	0.58	0.92	1.11	2.09	4.26
High	25.08	25.22	25.93	23.25	22.32	High	0.50	0.92	1.19	1.98	3.74
Panel D: Hold the portfolio 22 trading days after the earnings announcement											
SR/*SUR*	Low	2	3	4	High		Low	2	3	4	High
Panel A: Excess Returns							Panel B: Number of Stocks				
Low	0.91	2.02	-2.22	-9.45	0.37	Low	4.17	2.67	3.67	2.67	3.00
2	-4.19	-2.87	-13.23	-7.20	-3.25	2	3.50	2.67	2.33	3.33	4.33
3	-3.29	-5.70	-5.61	0.80	-7.60	3	3.00	3.50	2.67	2.83	4.17
4	-6.65	-3.00	-4.81	-2.38	-5.79	4	2.83	3.83	2.67	3.67	3.17
High	-4.48	-6.37	646.72	-24.48	-15.48	High	2.67	3.50	4.83	3.67	3.83
Panel C: *SR*							Panel D: *SUR*				
Low	-1.19	-3.09	-2.56	-1.74	-2.32	Low	0.41	0.94	1.19	2.28	4.36
2	6.12	7.00	6.25	5.89	5.35	2	0.38	0.88	1.06	2.07	6.90
3	11.41	10.79	10.98	9.83	10.75	3	0.58	1.00	1.28	1.97	5.18
4	15.91	15.91	15.65	15.21	15.79	4	0.58	0.92	1.11	2.09	4.26
High	25.08	25.22	25.93	23.25	22.32	High	0.50	0.92	1.19	1.98	3.74

Notes: This table contains results for portfolios of stocks sorted independently on each earnings announcement quarter by information sentiment (*SR*) constructed based on the linguistic text mining and *SUR*. The holding periods of the excess returns are one to four weeks after the earnings announcement. The reported results are averages of the relevant variables over the sample period 2001 to 2005 for the Taiwan stock market. Panels A, B, C and D respectively report in each portfolio the average excess returns, the average number of stocks, and the average *SR* and *SUR* of the stocks.