3. THE CHINESE NEWS SENTIMENT AROUND EARNINGS ANNOUNCEMENTS[1]

Yang-Cheng LU[2]
Yu-Chen WEI[3]

Abstract

We examine the effect of Chinese news on announcement drift and investigate its application to portfolio management, applying a linguistic analysis to extract various dimensions of the information content. Our empirical results reveal a positive (negative) relationship between news sentiment and cumulative abnormal returns in the pre- (post-) earnings announcement period, thereby confirming that the market response takes into consideration all relevant information on the related firm. The application of public news sentiment to portfolio management indicates that long (short) stocks with low (high) news sentiment and high public news surprises will earn positive excess returns. We suggest that the relevant news of individual stocks could be applied to the prediction of abnormal returns and portfolio management.

Keywords: media coverage, news sentiment, abnormal returns, earnings announcements, linguistic analysis
JEL Classification: G12; G14; D82

I. Introduction

At any given time, securities prices in an efficient market should fully reflect all of the available information; however, as opposed to adjusting immediately to news surprises, stock prices will tend to drift, over time, in the direction of the initial surprise. Such 'post-earnings announcement drift' (PEAD) or 'earnings momentum' is an enduring phenomenon, with the arguments supporting the existence of this anomaly

[1] *The authors would like to show their appreciation for the support provided for this study by the Department of Industrial Technology at the Ministry of Economic Affairs in Taiwan (Grant number: 101-EC-17-A-34-S1-149), the Chinese knowledge and information processing group of the Institute of Information Science, Academia Sinica, and the Institute of Linguistics, Academia Sinica.*
This study was presented at 2013 Global Business, Economics, and Finance Conference, Wuhan University, China, 9-11 May 2013.
[2] *Department of Finance at Ming Chuan University, Taiwan.*
[3] *Department of Money and Banking at National Kaohsiung First University of Science and Technology, Taiwan. E-mail: claireycwei@gmail.com.*

being that the content of any news relating to the stock market may well be linked to investor psychology or sociology (Tetlock, 2007). Since investors may have different interpretations of the information that they acquire, this can subsequently lead to under- or over-reaction to such information.

There are generally two distinct classifications of information, namely, public and private information (Vega, 2006), and whilst some investors regard public information simply as 'noise', others contend that relevant information is contained within such public news reports. Clearly, a method of effectively extracting relevant information from the huge number of news stories that are available is an enormously complex and technical process. Hence, linguistic analysis has emerged as an appropriate method for filtering out the noise from such qualitative news reports and for constructing appropriate proxies for the quantitative indicators.[4]

As noted by Demers and Vega (2011), different algorithms are available in various English texts for measuring the net optimism in the earnings announcements made by managers, such as the General Inquirer (GI), the Diction textual-analysis program (Version 6.0) and the Loughran-McDonald (L&M) dictionaries. However, very little research has so far been carried out on information sentiment with regard to Chinese documents; thus, in contrast to the traditional event study method of analyzing earnings announcements to probe the signals of information sentiment, the research focus in the present study is on extending this field by exploring a substantial number of Chinese financial news items and empirically measuring the effects of public and private information on the post-earnings announcement drift.

The proxy variables of public information adopted for this study are 'media coverage' (*MEDIA*), 'public news surprises' (*SUR*) and 'news sentiment' (*SR*), with the three factors proposed by Fama and French (1992), comprising momentum (Jegadeesh and Titman, 1993), turnover (*TURN*) and the probability of informed trading (*PIN*, Easley, Hvidkjaer and O'Hara, 2002), being used to calculate abnormal returns around the earnings announcement dates. We also investigate the relevance of public information and its potential application to portfolio management.

Our empirical results indicate the existence of a positive relationship between *SR* and cumulative abnormal returns (*CARs*) prior to earnings announcements. Our results further reveal that the leading effect of *SR*, which is constructed on the basis of the public news available in the one- to two-week period prior to the earnings announcement, is found to continue for a further two weeks after the event. The main contribution of our study to the extant literature is our extraction of information content relating to Chinese news reports through the use of linguistic analysis and examining the explanations to the abnormal returns.

The remainder of this paper is organized as follows. A literature review is conducted in Section 2, followed in Section 3 by a description of the data. Our research design and regression analysis are described in Section 4, with a summary of the empirical results being provided in Section 5. Finally, the conclusions are presented in Section 6.

[4] Interested readers should refer to Vega (2006), Tetlock (2007), Tetlock, Saar-Tsechansky and Macskassy (2008) and Demers and Vega (2011).

2. Literature Review

Recent studies support the view that the information extracted from text sources may provide valuable indicators for use in the prediction of abnormal returns, and indeed could be a key factor in the examination of market reaction. Vega (2006) examined the ways in which the public and private information acquired by agents prior to an earnings announcement could affect the PEAD. Consistent with the theories on rational uncertainty, Vega found that the more public or private information that investors acquire with regard to the true value of an asset, the more they will agree on and trade based on this information, which will result in a smaller drift in abnormal returns. In the investigation of the quarterly earnings press releases provided by managers, Demers and Vega (2011) further examined whether the 'soft' information contained within the text was incrementally informative, as compared to the 'hard' information issued by the firm.

Das and Chen (2007) demonstrated how sentiment could be extracted from various texts by developing a methodology for identifying small investor sentiment from stock message boards. Truong and Shane (2009) introduced a proxy for expected value-relevant earnings in their framework of proxy variables for news sentiment.

Based upon measurements obtained from the daily content of a popular Wall Street Journal column on the interactions between the media and the stock market, Tetlock (2007) found that high media pessimism predicted downward pressure on market prices, followed by a reversion to fundamentals, with unusually high or low pessimism giving rise to predictions of high market trading volume. These results are consistent with the theoretical models of 'noise' and 'liquidity trading' proposed by De Long, Shleifer, Summers and Waldmann (1990) and Campbell, Grossman and Wang (1993); however, the results are not consistent with the theories on media content due to a proxy for new information on fundamental asset values, a proxy for market volatility or an irrelevant noisy variable.

From the subsequent investigation into whether any negative words in the financial press could result in lowering firm earnings forecasts, Tetlock, Saar-Tsechansky and Macskassy (2008) found that stock market prices incorporated the information embedded in negative words with a slight delay. They also demonstrated potential profits from the use of a trading strategy based upon the words in a timely news source.

Engelberg and Parsons (2011) performed a comparison of the behavior of investors based upon their access to coverage of the same informative event through different media channels. They found that after controlling for earnings, investors and newspaper characteristics, local media coverage strongly predicted local trading. As noted by Engelberg and Parsons (2011), disentangling the impact of media reporting from the impact of the events being reported is challenging; thus further analysis is warranted to determine whether news reports represent relevant information or are simply noise, as well as the conditions under which the media are associated with the market reaction.

In the present study, we attempt to structure Chinese text by filtering out 'noise' words, to construct representative terms for 'optimism' and 'pessimism' and then calculate the quantitative proxy variables of Chinese financial news reports. We go on to further

examine the relationship that exists between news indicators and announcement drift, and investigate its potential application to portfolio management.

3. Data Sample and Description

Our analysis is carried out using corporate earnings announcements over the years from 2001 to 2005, with the study sample comprising electronic stocks traded on the Taiwan Stock Exchange (TWSE) and in the GreTai Securities Market (GTSM). The financial data are primarily obtained from the quarterly financial statements of listed companies in the Taiwan Economic Journal (TEJ) database, whilst the daily news reports on the firms are collected from the InfoTimes database.[5] The proxy variables for public information comprise 'media coverage' (*MEDIA*), 'public news surprises' (*SUR*) and 'news sentiment' (*SR*), with 'standardized unexpected earnings' (*SUE*) being considered as the control variable.

3.1. Abnormal and Cumulative Abnormal Returns

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

The *MTM* variable for each month is obtained in a similar way to the approach of Ku (2005), as the average returns from months $t-7$ to $t-12$. Since the turnover variable is considered in this study to be an explanatory variable of returns, we also include *TURN* in our model for the estimation and calculation of abnormal returns. All of the remaining variables are summarized in the following sub-sections.

3.1.1. The Fama and French three-factor model

We follow the approaches of Fama and French (1992), Easley et al. (2002) and Lu and Wong (2009) to calculate the betas in the present study. For each stock, we regress the monthly stock returns for a period of at least one year before the test month (two years where possible) on the contemporaneous and lagged value-weighted Taiwan stock index (TAIEX), from which we obtain the 'pre-ranking' portfolio betas as the sum of the two coefficients.

Ten portfolios are sorted for every month based upon the estimated betas, with the monthly portfolio returns being calculated by averaging the individual stock returns on an equal-weighted basis. The full sample period portfolio returns are then regressed on contemporaneous and lagged values of the TAIEX returns, from which the two

[5]Detailed introductions to the TEJ and InfoTimes database can be found at http://www.tej.com.tw/ and http://www.infotimes.com.tw/.

coefficients are summed to provide our 'post-ranking' portfolio betas (*PBeta*). Individual stock betas are therefore taken as the post-ranking betas of the portfolio to which they belong. Since the portfolio compositions change each month, the individually constructed stock betas will vary over time.

The market value of equity is the price multiplied by the number of outstanding shares 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. The TEJ uses the most recently available value of the book value of common equity (reported quarterly) to construct the book-to-market (*BM*) variable. As in Fama and French (1992) and Easley et al. (2002), those firms with negative book values are excluded from our sample.

3.1.2. The probability of informed trading (PIN)

The probability of informed trading (*PIN*) proposed by Easley et al. (2002) could be a proxy of information asymmetry and the measurement is as follow:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s}. \tag{1}$$

If a news event occurs with a probability of α, there will be a $(1 - \delta)$ chance that the news will be classified as good; thus, a good news event occurs with a probability of $\alpha(1 - \delta)$. Similarly, a bad news event occurs with a probability of $\alpha\delta$ Traders are assumed to arrive according to Poisson processes throughout the day, with orders from informed traders arriving at the rate of μ, and orders from uninformed buyers (sellers) arriving at the rate of ε_b (ε_s). In very simple terms, informed traders will buy if they perceive good news and sell if the news is perceived as being otherwise. The beliefs of market participants are updated by extracting information from the buy and sell trades, with new prices being formed, trades evolving and the price dynamics ultimately reflecting the changing beliefs of the market participants.

According to Easley et al. (2002), the likelihood function induced by this simple model of the trading process for a single trading day is as follows:[6]

$$\begin{aligned} L(\theta \mid B,S) = {} & (1-\alpha)\cdot e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}\cdot e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!} \\ & + \alpha\delta\cdot e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}\cdot e^{-(\mu+\varepsilon_s)}\frac{(\mu+\varepsilon_s)^S}{S!} \\ & + \alpha(1-\delta)\cdot e^{-(\mu+\varepsilon_b)}\frac{(\mu+\varepsilon_b)^B}{B!}\cdot e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!}. \end{aligned} \tag{2}$$

where the first line refers to the likelihood weighted by the probability of a day with no event $(1-\alpha)$; the second line refers to the probability of a 'bad news' day ($\alpha\delta$); and the third line refers to the probability of a 'good news' day. Let $\theta = (\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta)$ be the parameter vector and let (Bi, Si) denote the buy and sell trade data for period $i \in (1,\ldots,I)$. Assuming that sufficient independence conditions are held across *I* trading days, the parameters could be estimated by maximizing the likelihood function. Let *B*(*S*)

[6] *Refer also to Ellis, Michaely and O'Hara (2000) for further results on the calculation of buy and sell trades.*

denote the number of buy (sell) trades in a single trading day. 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.

3.1.3. Calculation of abnormal returns

The abnormal returns can be calculated as the actuarial return minus the expected return, which is expressed as:

$$AR_{it} = R_{it} - E(\hat{R}_{it}) \tag{3}$$

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

$$R_{id} = \gamma_{i0} + \gamma_{i1} PBeta_{pmd} + \gamma_{i2} BM_{imd} + \gamma_{i3} SIZE_{imd} + \eta_{id} \tag{4}$$

where $\gamma_{i0},\ldots,\gamma_{i3}$ 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* in which the trading day *d* belongs; BM_{imd} is the book-to-market value for month *m* in which the trading day *d* belongs; $SIZE_{imd}$ is the log of the market value for month *m* in which the trading day *d* belongs; and η_{id} is the mean-zero error term. The estimated excess returns can then be calculated as:

$$E(\hat{R}_{ie}) = \hat{\gamma}_{i0} + \hat{\gamma}_{i1} PBeta_{pme} + \hat{\gamma}_{i2} BM_{ime} + \hat{\gamma}_{i3} SIZE_{ime}\,. \tag{5}$$

where $E(\hat{R}_{ie})$ is the expected daily excess return of stock *i* on a specific date of *e* around the earnings announcement; $PBeta_{pme}$, BM_{ime} and $SIZE_{ime}$ are calculated accordingly. Once the coefficients of the factors have been estimated, the cumulative abnormal returns around the earnings announcement can be expressed as follows:

$$CAR_{i[-t1,t2]} = \sum_{e=-t1}^{t2} \left[R_{ie} - E(\hat{R}_{ie}) \right],\ t1 = 1,2,\ldots T1,\ t2 = 1,2,\ldots T2 \tag{6}$$

where *T*1 and *T*2 are the periods around the earnings announcement. The expected excess returns are also individually estimated with the incorporation of momentum (*MTM*), turnover (*TURN*) and the probability of informed trading (*PIN*) as a check for robustness. The cross-sectional regressions for the calculation of abnormal returns are expressed in Equations (7) to (9) as follows:

$$R_{id} = \gamma_{i0} + \gamma_{i1} PBeta_{pmd} + \gamma_{i2} BM_{imd} + \gamma_{i3} SIZE_{imd} + \gamma_{i4} MTM_{imd} + \eta_{id} \tag{7}$$

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

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

3.2. Public Information Proxies

Three proxy variables of public information are constructed based upon textual analysis, with all of the news variables being estimated *T* days prior to the earnings announcement in order to further identify the relevant information contained in the public news reports.

3.2.1. Media coverage

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:

$$MEDIA_{i,t} = \sum_{k=1}^{T} NEWS_{i,t-k} \tag{10}$$

where $NEWS_{i,t-k}$ is a dummy variable which is equal to 1 if firm *i* is mentioned in the headline or leading paragraph of an article on day *t*. $MEDIA_{i,t}$ is measured using the InfoTimes database, which includes representative Chinese news.

3.2.2. Public news surprises

In order to consider the reaction of the stock market 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 TAIEX, *SUR* is adjusted to capture the abnormal return and abnormal market turnover as:

$$\begin{aligned} SUR_{i,t} = \sum_{k=1}^{30} \{ NEWS_{i,t-k} \left[I(AR_{i,t-k} \geq AR_{i,XT}) + I(AR_{i,t-k} \leq AR_{i,XB}) \right] \} \\ \times \left\{ I(DV_{i,t-k} \geq DV_{i,YT}) + I(DV_{i,t-k} \leq DV_{i,YB}) \right\}, \end{aligned} \tag{11}$$

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

3.2.3. News sentiment (SR)

Referring to Diction, we find that 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 provides the difference between 'optimism' and 'pessimism'. Since there is no specific optimism or pessimism in Chinese documents, we translate the sentiment classification and collect the related Chinese sentiment words.[7] The measure of news sentiment used in this study, representing the meaning of 'net optimism', is calculated as follows:

$$SR_{i,t} = \frac{\sum_{j=1}^{p} ptf_{ij,t} - \sum_{j=1}^{n} ntf_{ij,t}}{TF_{i,t}} \times 100\% \tag{12}$$

[7] The keywords, 'optimism' and 'pessimism', are translated using the Academia Sinica Bilingual Ontological Wordnet (http://bow.sinica.edu.tw/).

where $SR_{i,t}$ is the news sentiment of the i^{th} firm at time t; $ptf_{ij,t}$ is the term frequency of the j^{th} characteristic term for optimism for the i^{th} firm at time t; $ntf_{ij,t}$ is the term frequency of the j^{th} characteristic term for pessimism for the i^{th} firm at time t; and $TF_{i,t}$ is the total term frequency for the i^{th} firm at time t.

3.3 . Unexpected Earnings Surprises (SUR)

Following Demers and Vega (2011), we use a seasonal random walk model to capture *SUR*, defining 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 expectation of earnings which is measured by the earnings per share for the same quarter in the previous year in the TEJ sample. Unexpected earnings are standardized by dividing the surprise by the firm-specific standard deviation of the forecast error, with the standardized unexpected earnings associated with firm *i* for quarter *q* at time *t* then being defined as SUE_{iqt}.

4. Research Design

4.1 . Pre-Announcement Period Tests

We first of all investigate the *CAR* response to the news information surprises prior to the earnings announcement. The dependent variable is defined as the *CARs* calculated during the *T* calendar-day period prior to the earnings announcement date which corresponds to the period in which the public financial news was collected.

$$CAR_{iq[-T,-1]} = \beta_0 + \beta_1 SR_{iq[-T,-1]} + \beta_2 MEDIA_{iq[-T,-1]} + \beta_3 SUR_{iq[-T,-1]} + \beta_4 SUE_{iq} + \varepsilon_{iq} \quad (13)$$

where the *iq* subscript refers to firm *i* for event quarter *q*; [−*T*, −1] is the estimated *T* calendar-day period prior to the earnings announcement; $CAR_{iq[-T,-1]}$ is calculated based upon the expected abnormal return of actual trading days during the *T* calendar days prior to the earnings announcement; $SR_{iq[-T,-1]}$, $MEDIA_{iq[-T,-1]}$ and $SUR_{iq[-T,-1]}$ are the proxies for public information calculated *T* calendar days prior to the earnings announcement (since news reports might be published on non-trading days); and 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 unexpected earnings surprises by analyzing the relationship between the *CARs* after the earnings announcement and the news information collected prior to the earnings announcement day. The post-earnings announcement drift is calculated from 1 to 22 trading days after the announcement date. This leads to the following pooled regression model:

$$CAR_{iq[1,T']} = \beta_0 + \beta_1 SR_{iq[-T,-1]} + \beta_2 MEDIA_{iq[-T,-1]} + \beta_3 SUR_{iq[-T,-1]} + \beta_4 SUE_{iq} + \varepsilon_{iq} \quad (14)$$

where the *iq* subscript 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. $CAR_{iq[1,T']}$ is the cumulative abnormal return estimated during the post-earnings announcement period and T' is equal to 1 to 22 trading days.

4.3 . Application to Portfolio Management

If we are indeed able to confirm whether relevant information is contained within Chinese news reports, we can then go on to provide the average excess returns for the portfolios of stocks sorted independently by news sentiment and public news surprises. For each event quarter, the stocks can be sorted into five news sentiment groups, ranging from low to high, and five public news surprise groups, again ranging from low to high. The portfolio excess returns obtained will be the averages of the relevant variables over the sample period.

5. Empirical Results

5.1. Summary Statistics

To further classify the relevance of news information prior to the earnings announcement, we construct the news information proxies comprising the 7-day period prior to the earnings announcement. The summary statistics are presented in Table 1, whilst Table 2 shows the correlation coefficient between the news information and the cumulative abnormal returns in the pre-earnings announcement period. Table 1 shows that the mean of *SR* is positive, which means that the media reports prior to the earnings announcement attempt to announce the positive reports. Table 2 indicates that *SUE* is only significantly correlated to *SR*. The higher the news sentiment, the higher the unexpected earnings surprise. Besides, there is a positive relationship between *SR* and *MEDIA*.

Table 1

Summary statistics

Variables	Min.	Max.	Mean	Median	S.D.	No. of Obs.
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	

Note: SUE refers to unexpected earnings surprises; *MEDIA* denotes media coverage; *SUR* is the measure of public news surprises which takes into consideration the reaction of the stock market to headline news; and *SR* is the news sentiment. $CAR_{[-7,1]}$ is the cumulative abnormal returns for the 7-day period prior to the earnings announcement.

Table 2

Correlation analysis

Variables	$SR_{[-7,-1]}$	$MEDIA_{[-7,-1]}$	$SUR_{[-7,-1]}$	$CAR_{[-7,-1]}$
SUE	0.1933 ***	0.0457	0.0442	−0.0762 **
$SR_{[-7,-1]}$		0.114 ***	−0.0141	0.1421 ***
$MEDIA_{[-7,-1]}$			0.6450 ***	0.0224
$SUR_{[-7,-1]}$				0.0175

Notes: SUE refers to unexpected earnings surprises; *SR* is the news sentiment; *MEDIA* denotes media coverage; *SUR* is the measure of public news surprises which takes into consideration the reaction of the stock market to headline news. ** denotes significance at the 5% level; and *** denotes significance at the 1% level.

5.2 . Pre-earnings Announcement Period Tests

We go on in this section to assess the market response to any unexpected information contained in the earnings announcements. We begin by presenting the explanatory power of the news information collected prior to the earnings announcement in relation to the pre-earnings announcement drift using Equation (13). The empirical results on the pre-earnings announcement drift are shown in Table 3, with the CARs during the pre-announcement periods being calculated based upon the Fama and French three-factor model, as expressed in Equations (4) to (6).

Model (1) in Table 3 indicates that the *SUE* associated with a firm has a significantly negative correlation with pre-earnings announcement drift. The news information proxies, namely, *MEDIA* and *SUR*, refer to the degree of disclosure in the public news related to a specific firm. Models (3) and (4) in Table 3 reveal that when these variables are individually included in the model, they do not exhibit significant explanatory power in regard to the pre-earnings announcement drift. Model (2) in Table 3 is used to derive the empirical results of the proxy of news sentiment, *SR*, which is the net optimism of the specific firm revealed in the public news. Our analysis reveals the existence of a positive relationship between the net optimism and the *CARs*; that is, the higher the net optimism, the higher the *CARs*.[8] Model (5) in Table 3 compares the explanatory power of the news information and *SUE* prior to the earnings announcement day. Only the *SR* and *SUE* variables are found to be significant with regard to pre-earnings announcement drift.

Table 3

Regression analysis of the pre-earnings announcement period

Variables	Models				
	(1)	(2)	(3)	(4)	(5)
Constant	–1.031 ***	–2.244 ***	–1.378 ***	–1.173 ***	–2.314 ***
	(–4.081)	(–6.381)	(–2.705)	(–3.471)	(–4.246)
SUE	–0.625 **	–0.883 ***	–0.634 **	–0.632 **	–0.893 ***
	(–2.304)	(–3.236)	(–2.337)	(–2.329)	(–3.268)
$SR_{[-7,-1]}$		0.113 ***			0.114 ***
		(4.896)			(4.892)
$MEDIA_{[-7,-1]}$			0.347		–0.169
			(0.783)		(–0.291)
$SUR_{[-7,-1]}$				0.072	0.113
				(0.632)	(0.761)

Notes: SUE refers to unexpected earnings surprises; *SR* is the news sentiment; *MEDIA* denotes media coverage; *SUR* is the measure of public news surprises which takes into consideration the reaction of the stock market to headline news. The values in the parentheses are *T*-statistics. ** denotes significance at the 5% level; and *** denotes significance at the 1% level.

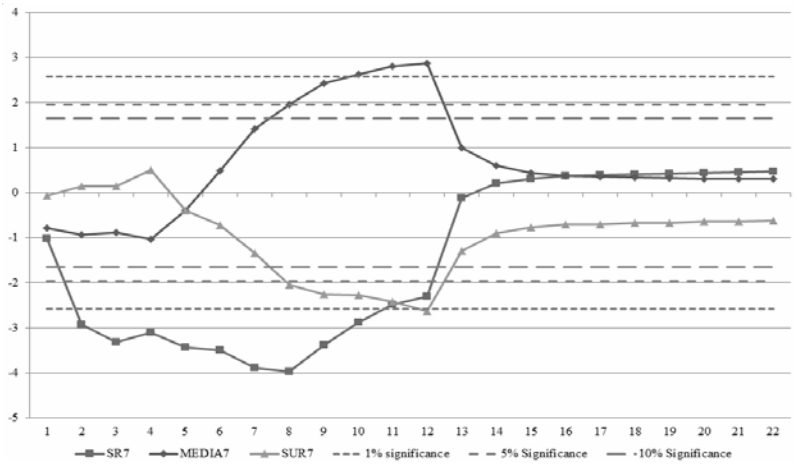
[8] The 14-day results are omitted for space considerations and the results are available from the authors upon request.

5.3. Post-earnings Announcement Period Tests

The relationship between the news information and post-earnings announcement drift is examined using Equation (14), and in order to further investigate the signaling effect of the former on the latter, we calculate the *CARs* using the 1- to 22-day post-earnings announcement period. Figure 1 illustrates the significance of the news information prior to the earnings announcement to the market response in the 1- to 22-day post-earnings announcement periods.

Figure 1

Relationship between public information and post-earnings announcement drift



The relationships that exist between the news information in the 7-day period prior to the earnings announcement and the 1 day to 22 days of post-earnings announcement drift are respectively illustrated in Figure 1. As we can see from the figure, the leading effect of *SR* constructed from the public news revealed one week prior to the earnings announcement is found to continue for two weeks after the earnings announcement, with the *SR* measures for the two-week, three-week and one-month periods prior to the earnings announcement also indicating similar leading effects. The results are omitted for space considerations, but are available from the authors upon request.

Although significant leading effects are also exhibited by the *MEDIA* and *SUR* variables constructed on the one-week, two-week, three-week and one-month periods prior to the earnings announcement, the findings for these two variables are directly opposite; that is, a positive (negative) relationship is found between post-earnings announcement drift and *MEDIA* (*SUR*), with this relationship again being found to be significant one to two weeks after the earnings announcement.

5.4. Robustness Check

As a check for robustness, we carry out a regression analysis on the pre-announcement period using the three-factor model of Fama and French (1992) with the separate incorporation of *MTM*, *TURN* and *PIN*, and the results are shown in Table 4.

Table 4

Robustness check of the cross-sectional regression analysis in the pre-earnings announcement period

Variables	Models				
	(1)	(2)	(3)	(4)	(5)
Panel A: Fama-French Three Factors with *MTM*					
Constant	–1.033 ***	–2.044 ***	–1.580 ***	–1.190 ***	–2.372 ***
	(–3.960)	(–5.611)	(–3.013)	(–3.422)	(–4.195)
SUE	–0.679 **	–0.898 ***	–0.696 **	–0.687 **	–0.909 ***
	(–2.438)	(–3.187)	(–2.496)	(–2.464)	(–3.219)
$SR_{[-7,-1]}$		0.094 ***			0.094 ***
		(3.943)			(3.853)
$MEDIA_{[-7,-1]}$			0.55		0.223
			(1.202)		(0.371)
$SUR_{[-7,-1]}$				0.080	0.057
				(0.684)	(0.373)
Panel B: Fama-French Three Factors with *PIN*					
Constant	–1.033 ***	–2.044 ***	–1.580 ***	–1.190 ***	–2.371 ***
	(–3.960)	(–5.611)	(–3.013)	(–3.422)	(–4.195)
SUE	–0.679 **	–0.898 ***	–0.696 **	–0.687 **	–0.909 ***
	(–2.438)	(–3.186)	(–2.496)	(–2.464)	(–3.219)
$SR_{[-7,-1]}$		0.094 ***			0.093 ***
		(3.943)			(3.853)
$MEDIA_{[-7,-1]}$			0.547		0.223
			(1.202)		(0.371)
$SUR_{[-7,-1]}$				0.079	0.057
				(0.684)	(0.373)
Panel C: Fama-French Three Factors with *TURN*					
Constant	–1.033 ***	–2.044 ***	–1.580 ***	–1.190 ***	–2.371 ***
	(–3.960)	(–5.611)	(–3.013)	(–3.422)	(–4.195)
SUE	–0.679 **	–0.898 ***	–0.696 **	–0.687 **	–0.909 ***
	(–2.438)	(–3.186)	(–2.496)	(–2.464)	(–3.219)
$SR_{[-7,-1]}$		0.094 ***			0.094 ***
		(3.943)			(3.853)
$MEDIA_{[-7,-1]}$			0.547		0.223
			(1.202)		(0.371)
$SUR_{[-7,-1]}$				0.079	0.057
				(0.684)	(0.373)

Notes: SUE refers to unexpected earnings surprises; *SR* is the news sentiment; *MEDIA* denotes media coverage; *SUR* is the measure of public news surprises which takes into consideration the reaction of the stock market to headline news. $CAR_{[-T,-1]}$ is calculated 7 days prior to the earnings announcement. The values in the parentheses are *T*-statistics. ** denotes significance at the 5% level; and *** denotes significance at the 1% level.

Although the calculation of the news information for our robustness check is based on seven days of public information, the empirical analyses related to the 14-, 21- and 28-day periods prior to the earnings announcement all provide similar results.[9] The results of the regression analysis on the news information and *CARs* prior to the earnings announcement are investigated by using equations (7) to (9); all of the other asset pricing factors are incorporated in this analysis as a check for robustness. Panels A to C show the effects of the results on the explanatory power of news information and of *SUE* on the *CARs*, with the calculation of these results being based upon the three-factor model of Fama and French (1992) with the incorporation of *MTM*, *PIN* and *TURN*. The findings confirm the positive relationship between *SR* and the *CARs*.

The leading effect of news information on the *CARs* during the post-announcement period, with the incorporation of *MTM*, *PIN* and *TURN*, are further analyzed and the results are similar to the *CARs* calculated by the three-factor model in Figure 1.[10] We find that the leading effect of *SR*, *MEDIA and SUR*, constructed on the basis of the public news revealed in the one-week period prior to the earnings announcement, continues for a further two weeks after the earnings announcement. This finding further confirms that such news reports contain relevant information capable of predicting *CARs* around earnings announcement periods.

5.5. Application to Portfolio Management

The average excess returns for the portfolios of stocks are reported in Table 5, with the returns being independently sorted for each earnings announcement quarter by public news surprises (*SUR*) and news sentiment (*SR*, the construction of which is based upon linguistic text mining) in order to confirm the application of the information sentiment. Our calculations of both *SUR* and *SR* are based upon news reports in the seven-day period prior to the earnings announcement. The portfolio excess returns are the average one-week performance after the earnings announcement.

Table 5 presents the individual portfolio excess returns, with the holding periods of the portfolios being set at 5-trading days after the earnings announcement. The application of the empirical results of Figures 1 to portfolio management provides support for our contention that *SUR* and *SR* could be effectively used as reference indicators in portfolio management, with suggested holding periods of one to two weeks after the earnings announcement. The results for two weeks to one month are omitted for space considerations. The application of public news sentiment to portfolio management suggests that long (short) stocks with low (high) news sentiment and high public news surprises will earn positive excess returns.

[9] Due to space considerations, the full details are omitted here; however, the results of the regression analysis including all frequencies of the news information are available from the authors upon request.

[10] The figures presenting the leading effect of *SR*, *MEDIA* and *SUR* on the *CARs* after the earnings announcement, calculated by using the Fama and French three factors and *MTM, PIN and TURN* individually, are omitted for space considerations.

Table 5

Portfolio excess returns using news sentiment and news surprises

Variables		SUR					SUR				
		1 (Low)	2	3	4	5 (High)	1 (Low)	2	3	4	5 (High)
		a. Excess Returns					b. No. of Stocks				
SR	1 (Low)	–0.19	0.29	–3.42	–3.71	0.43	4.17	2.67	3.67	2.67	3
	2	–3.67	–0.84	–0.57	–2.72	–1.10	3.5	2.67	2.33	3.33	4.33
	3	–4.32	0.07	–2.25	–0.46	–1.30	3	3.5	2.67	2.83	4.17
	4	–4.34	–1.99	–4.75	–1.09	–1.37	2.83	3.83	2.67	3.67	3.17
	5 (High)	–1.15	–0.06	–2.36	–0.40	–6.03	2.67	3.5	4.83	3.67	3.83
		c. SR					d. *SUR*				
SR	1 (Low)	–1.19	–3.09	–2.56	–1.74	–2.32	0.41	0.94	1.19	2.28	4.36
	2	6.12	7	6.25	5.89	5.35	0.38	0.88	1.06	2.07	6.9
	3	11.41	10.79	10.98	9.83	10.75	0.58	1	1.28	1.97	5.18
	4	15.91	15.91	15.65	15.21	15.79	0.58	0.92	1.11	2.09	4.26
	5 (High)	25.08	25.22	25.93	23.25	22.32	0.5	0.92	1.19	1.98	3.74

Note: The results reported are for portfolios of stocks sorted independently on each earnings announcement quarter by information sentiment (*SR*, which is constructed based upon linguistic text mining) and news surprises (*SUR*). The holding periods of the excess returns are one week after the earnings announcement and the reported results are averages of the relevant variables on the Taiwan Stock Market over the 2001-2005 sample period. The results for two weeks to one month are omitted for space considerations.

6. Conclusions

We set out in this study to investigate the information content of Chinese media press on post-earnings announcement drift, and to examine its potential application to portfolio management. We follow Vega (2006) and Demers and Vega (2011) to construct the indicators of news sentiment using linguistic analysis. Our aim in this study is to verify whether public information is valuable news or simply noise, with our research focusing on an extension of the traditional event study approach to earnings announcements.

We empirically measure the effect of information obtained from news reports on post-earnings announcement drift, with the proxy variables of news information comprising of media coverage (*MEDIA)*, public news surprises (*SUR*) and news sentiment (*SR*). We adopt the three factors proposed by Fama and French (1992), 'momentum' (Jegadeesh and Titman, 1993), 'turnover' (*TURN*) and the 'probability of informed trading' (*PIN*, Easley et al., 2002) for the calculation of abnormal returns around the earnings announcement date.

Our empirical results reveal a positive relationship between *SR* and *CARs* prior to the earnings announcement. The application of news sentiment to portfolio management suggests that long (short) stocks with low (high) news sentiment and high public news surprises will earn positive excess returns. We therefore conclude by proposing that the news sentiment of individual stocks may well have some psychological influence on the trading behavior of investors, and that this could lead to the development of a warning model of abnormal returns and portfolio management.

References

Campbell, J.Y. Grossman, S.J. and Wang, J. 1993. Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, 108, pp. 905-939.

Das, S. and Chen, M. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53, pp. 1375-1388.

De Long, J.B. Shleifer, A. Summers, L.G. and Waldmann, R.J. 1990. Noise trader risk in financial markets. *Journal of Political Economy*, 98, pp. 703-738.

Demers, E.A. and Vega, C. 2011. Linguistic tone in earnings announcements: News or noise? FRB International Finance Discussion Paper No. 951.

Easley, D. Hvidkjaer, S. and O'Hara, M. 2002. Is information risk a determinant of asset returns? *Journal of Finance*, 57, pp.2185-2221.

Ellis, K. Michaely, R. and O'Hara, M. 2000. The accuracy of trade classification rules: Evidence from the Nasdaq. *Journal of Financial and Quantitative Analysis*, 35, pp. 529-551.

Engelberg, J. and Parsons, C. 2011. The causal impact of media in financial markets, *Journal of Finance*, 66, pp. 67-97.

Fama, E.F. and French, K.R. 1992. The cross-section of expected stock returns. *Journal of Finance*, 47, pp. 427-465.

Jegadeesh, N. and Titman, S. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, pp. 65-91.

Ku, K.P. 2005. One-factor, three-factor or four-factor models? *Review of Securities and Futures Markets*, 17, pp. 101-146.

Lee, C.M.C. and Ready, M.J. 1991. Inferring trade direction from intraday data. *Journal of Finance*, 46, pp. 733-746.

Lu, Y.C. and Wong, W.K. 2009. Probability of information-based trading as a pricing factor in the Taiwan stock market. *International Research Journal of Finance and Economics*, 33, pp. 31-49.

Lu Y.C. and Wei, Y.C. 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, pp. 23-35.

Tetlock, P.C. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62, pp. 1139-1168.

Tetlock, P.C. Saar-Tsechansky, M. and Macskassy, S. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63, pp. 1437-1467.

Truong, C. and Shane, P.B. 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, pp. 103-133.