News Sentiment and Cross-Section of Stock Returns in the Taiwan Stock Market

Yu-Chen Wei*
Associate Professor, Department of Money and Banking, National Kaohsiung First University of Science and Technology

Yang-Cheng Lu
Professor, Department of Finance, Ming Chuan University

Yen-Ju Hsu
Ph.D. student, Department of Finance, National Taiwan University

Abstract

We measure the news sentiment, examining the ways in which such sentiment affects the cross-section of stock returns. We suggest that the news sentiment captures dimensions beyond the market, size premium factor, book-to-market premium factor, momentum and sentiment factors, and that it has effects on the characteristics of securities firms that are directly influenced by the prospections and trading behavior of investors. Our empirical results reveal that for portfolios formed on size, market-to-book value, age, revenue, EPS, volatility and high-low range, the subsequent returns will be high (low) when the news sentiment is relatively low (high), regardless of whether the portfolio is equally-weighted or value-weighted. We further confirm the supposition that when news sentiment is high, the returns over the subsequent month on low turnover, low volatility, low 'high-low range' and older firms will be relatively high.

JEL Classification: D82; G12; G14
Keywords: News Sentiment; Financial press; Linguistic analysis; Firm characteristics; Stock returns.

---

* This study was supported by a grant from the Ministry of Science and Technology in Taiwan (Grant number MOST 103-2410-H-327 -007). We would also like to express our appreciation to Professor Keh-Jiann Chen of the Institute of Information Science at Academia Sinica for the support of the Chinese Word Segmentation System.
** Yu-Chen Wei is an Associate Professor in the Department of Money and Banking at National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan. Email: claireycwei@gmail.com. Address: No. 1, University Road, Yanchao District, Kaohsiung City, 824 Taiwan (R.O.C). Yang-Cheng Lu is a Professor in the Department of Finance at Ming Chuan University, Taipei, Taiwan. Email: ralphyclu@gmail.com. Address: 250 Zhong Shan N. Rd., Sec. 5, Taipei 111, Taiwan. Yen-Ju Hsu is a Ph.D. student at the Department of Finance at National Taiwan University, Taipei, Taiwan. Email: yenj.hsu@gmail.com. Address: No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan (R.O.C). The corresponding author is Yu-Chen Wei.

1. Introduction

Numerous studies have confirmed that the trading behavior of investors and the changes in stock prices are both influenced by the information content of media reports,[1] with classical finance theory presenting evidence to show that various factors can have significant effects on the cross-section of stock prices.[2] In the present study, we propose that the degree of news sentiment that is inherent in financial news represents a public news channel easily accessible by investors prior to engaging in trading. Therefore any relevant information contained in such financial news may capture dimensions beyond the market, SMB, HML, momentum and sentiment factors, thereby significantly affecting the characteristics of securities firms that are directly influenced by the prospections and trading behavior of investors.

In their examinations of the role of the media in the stock markets, the prior studies have attempted to quantify the news relating to each stock; however, these studies have invariably concluded that the resultant quantitative index fits appropriately within the dimensions of investor sentiment. Thus, the major difference between the extant literature and the present study is that the degree of news sentiment is constructed in this study from firm level to market level and measured by soft information.

We argue that market-based soft information does not belong in the category of investor sentiment, but instead, that the positive or negative reports – which are invariably referred to as 'good' or 'bad' news – reveal the current level of market news sentiment to participants; thus, we propose that the market news sentiment index directly influences the trading behavior of investors. We suggest that our 'aggregate news sentiment index' (*ANSI*) measure, which refers to the degree of sentiment induced by such news, captures the

---

[1] Examples include Vega (2006), Tetlock (2007), Tetlock, Saar-Tsechansky and Macskassy (2008), Demers and Vega (2011) and Engleberg and Parsons (2011).

[2] See, for example, Fama and French (1993), Jegadeesh and Titman (1995), Easley, Hvidkjaer and O'Hara (2002), Baker and Wurgler (2006, 2007) and Chung, Hung and Yeh (2012).

optimism and pessimism indicators that are expressed in media reports, with such sentiment words potentially affecting the mood of investors, and ultimately influencing their trading decisions.

Despite several recent studies having clearly demonstrated that the sentiment extracted from financial news is of relevance to financial events,[3] to the best of our knowledge there have been very few studies which have set out with the purpose of constructing a representative proxy derived from financial news; there has also been a distinct absence of studies setting out to examine the relationship between such news and the cross-section of stock returns.

We aim to bridge this gap in the present study by analyzing a substantial volume of financial news reports relating to listed stocks in the Taiwan Stock Market. We create a new construct ‘aggregate news sentiment index’ (*ANSI*) and examine the differences between portfolio returns sorted under different news sentiment regimes. Our primary aim is to investigate the predictability of long and short portfolio returns, and to compare the results with other risk factors.

An important issue within the current finance literature is that of determining ways of analyzing and applying the content of public information.[4] Data mining and knowledge discovery in databases have attracted significant attention in research fields over recent years, with content analysis representing a valuable research technique designed to make replicable and valid inferences by interpreting and coding textual material. By systematically evaluating texts (such as documents, oral communication and graphics), qualitative text can be converted into quantitative data, with the transference of such quantitative data from text potentially being applied to further analyses and subsequent

---

[3] See, for example, Tetlock (2007), Tetlock, Saar-Tsechansky and Macskassy (2008), Demers and Vega (2011) and Engelberg and Parsons (2011).

[4] See for example, Vega (2006) Tetlock (2007), Tetlock et al. (2008), Engelberg and Parsons (2011), Demers and Vega (2011), Griffin, Hirschey and Kelly (2011) and Groß-Klußmann and Hautsch (2011).

decision making.

Linguistic analysis involves the examination of huge amounts of data and subsequently converting it into statistical data as a reference for future decision making. We compare the measures used in the prior analyses of the role of the media, and employ linguistic analysis to capture keywords in the news, quantifying the words contextually. We analyze the words commonly found in the financial press and construct the degree of news sentiment directly induced by financial news.

Our study contributes to the extant literature in this field by examining the cross-sectional predictability patterns of the market aggregate news sentiment index (*ANSI*) on stock returns. As compared to the prior studies, in addition to considering the traditional risk and behavioral factors of investor sentiment, we go on to verify the predictability of the degree of *ANSI* with regard to the returns of long and short portfolios. We also compare the average equally-weighted and value-weighted portfolio returns as a check for the robustness of our findings.

The remainder of this paper is organized as follows. A summary of the related literature is provided in Section 2, followed in Section 3 by a description of the method of measurement of news sentiment in financial news using linguistic analysis. A comparison of the portfolio returns, sorted by firm characteristics, is undertaken in Section 4, focusing on whether the aggregate news sentiment index is in the high or low regime. Section 5 introduces the models used in this study and goes on to present the predictive regressions for long and short portfolios, along with the results. Finally, the conclusions drawn from this study are presented in Section 6.

2. Literature review

Several of the prior studies have provided significant evidence of strong correlations between media reports and stock market reaction, with these studies often concluding with similar findings; that is, press releases prior to financial events (such as earnings announcements) have significant impacts on market reaction.[5] Vega (2006), for example, demonstrated that the more information that investors have available to them with regard to the value of an asset, then the more aligned their beliefs with regard to the 'true' value of that asset, as a result of which they will tend to trade on such information with smaller abnormal returns.

Demers and Vega (2011) extended the concept of the 'public information' variable, initially proposed in Vega (2006), employing textual-analysis programs to extract various dimensions and specific words relating to sentiment from more than 20,000 corporate announcements of abnormal earnings. Their empirical results suggested that the higher the level of net optimism (or certainty) with regard to the news content, the greater the cumulative returns. This clearly implies that the degree of net optimism towards news content can be an important factor in the prediction of cumulative abnormal returns following such announcements.

The geographical location of the associated media has been found to have different influences on the coverage of each enterprise; for example, from their exploration of the behavior of investors towards the same earnings announcement news reports released by different media outlets, Engelberg and Parsons (2011) found the existence of a significant relationship between stock returns and local media coverage.

This would seem to go some way towards explaining why the existing public information literature considers both the different types of public news and the diverse influences on different types of investors. The response by individual and institutional investors towards macroeconomic news and company-specific news reported by the likes of the Wall Street Journal is also likely to differ significantly, whilst these different types of

[5] See, for example,Vega (2006), Tetlock (2007), Tetlock et al. (2008), Demers and Vega (2011) and Engelberg and Parsons (2011).

investors are very likely to have distinct trading strategies in response to the release of ‘good’ or ‘bad’ news (Nofsinger, 2001).

Several recent empirical findings have revealed the existence of irrational behavior, pointing out that individual investors tend to actively engage in stock purchases on ‘high-attention’ days. Barber and Odean (2008), for example, examined the attention-driven buying of investors by sorting the stocks based upon specific events, abnormal trading volume, extreme one-day returns and whether or not a firm was currently involved in news coverage. They found that individual investors who exhibited attention-driven buying behavior tended to actively trade in stocks on those days that were characterized by both extremely negative and positive one-day returns with high trading volume, as well as in those stocks that were currently mentioned in news reports.

A very recent study revealed the existence of significant implied volatility spillovers both across European markets and between the US and European markets (Jiang et al., 2012). Although scheduled news announcement clearly help to resolve information uncertainty, unscheduled news events can often give rise to information uncertainty (that is, an increase in implied volatility).

Groß-Klußmann and Hautsch (2011) recently used automated text analysis tools to quantify high-frequency news-implied market reaction, employing a high-frequency VAR model to examine the reaction to intraday news amongst stocks traded on the London Stock Exchange. They found that a classification of news based on indicated relevance was crucial to filtering out noise (including distinct responses in returns, and the volatility, trading volume and bid-ask spreads attributable to news arrivals).

Fischer and Verrecchia (1999) demonstrated that steady-state heuristic trading was reduced by improved public disclosure, based upon a reduction in the rents available to informed traders, whilst the empirical results reported by Brown and Cliff (2005) provided strong support for the view that effectively predicting irrational investor sentiment can indeed have direct impacts on asset price levels; thus, even when controlling for factors commonly associated with rational pricing, some mispricing can be explained by investor sentiment.

The predictive ability between sentiment and stock market trends can be traced back to the irrational decisions made by investors. When they overreact to public information, there will be a discernible rise in the probability of irrational trading amongst such investors, thereby leading to abnormal returns and volatility. For example, based upon an investigation into the relationship between investor sentiment and future stock returns in 18 industrialized countries, Schmeling (2009) found that sentiment was generally found to be a significant predictor of expected returns across the countries examined. Furthermore, the predictive power of sentiment was found to be most pronounced for short- and medium-term horizons of one to six months, subsequently diminishing over longer horizons of 12 to 24 months.

The consumer confidence index also appears to be an effective indicator of the impacts on stock returns, as demonstrated by Jansen and Nahuis (2003); they found positive correlations in nine European countries, over the period from 1986 to 2001, between stock returns and changes in sentiment, with stock returns being found to Granger-cause consumer confidence over very short horizons (two to four weeks).

An examination of portfolio returns by Zhang (2006), based upon proxies for price momentum and information uncertainty, revealed that information uncertainty has a strongly positive correlation with the one-month-ahead returns of past winners. A stronger momentum effect was also found for firms with high levels of information uncertainty. The recent analysis by Griffin et al. (2011) of the differences in the information content of news announcement in 56 markets also found noticeably different responses to such news releases; indeed, the results indicated the presence of insider trading prior to takeovers in the emerging markets, as well as some of the smaller developed markets.

In summary, public news releases on specific firms may influence subsequent responses relating to such firms. By referring to the related literature, we aggregate such specific news reports in the present study and construct the deep-seated news sentiment induced by financial press. In the following sections, we go on to describe the method of measurement, portfolio returns and empirical results.

3. Measurement of news sentiment

The samples for examination in this study comprise of the daily transaction data on all firms listed on the Taiwan Stock Exchange (TWSE). The dataset was obtained from the Taiwan Economic Journal (TEJ), whilst the news-corpus information was collected from the InfoTimes database, a leading source of daily news reports in Taiwan. Our study period runs from 1 January 2003 to 31 December 2012 (providing a total of 2,489 trading days) a period during which 1,155,757 news report samples were collected from the Commercial Times and China Times. The summary statistics on the news report samples are presented in Table 1.

<Table 1 is inserted about here>

Table 2 provides comprehensive statistical data on all of the reported corporate samples, in addition to the proportion of reported corporate samples to the total number of listed firms (in percentage terms). The annual average number of reports for the corporate samples was 557, whilst the relative proportion of reported samples to all listed firms was 78.7 per cent. As shown in Tables 1 and 2, regardless of the number of news reports or the proportion of reported corporate samples to all listed corporate samples, no significant differences are discernible for each month or year.

<Table 2 is inserted about here>

3.1 Firm-level News Sentiment

It is generally found to be much more difficult to analyze the content of words in Chinese documents, as compared to analysis in English vocabulary, and indeed, there are very few models or tools available for effectively revealing or constructing information sentiment indicator based upon Chinese documents. Our news sentiment indicator relating to each stock from the news reports is therefore constructed by referring to the concept of Demers and Vega (2011) and Lu and Wei (2014).

The degree of firm-level news sentiment is measured by determining ‘net optimism’ and ‘net pessimism’. According to Diction,[6] optimism is defined as ‘language endorsing some person, group, concept or event, or highlighting their positive entailments’. The Diction formula for the degree of net optimism (the difference between ‘optimism’ and ‘pessimism’) is expressed as follows:

(Praise + Satisfaction + Inspiration) – (Blame + Hardship + Denial)

The News Sentiment Index is essentially an indicator as reflected by the individual stock; this indicator is measured as follows:

$$NSI_{i,d,m} = \frac{\sum_{p=1}^{P} tf_{i,d,p} - \sum_{n=1}^{N} tf_{i,d,n}}{TF_{i,d,m}} \times 100\% \qquad (1)$$

$$NSI_{i,d} = \frac{1}{M} \sum_{m=1}^{M} NSI_{i,d,m} \qquad (2)$$

where $NSI_{i,d,m}$ is the news sentiment of the m^{th} news for the i^{th} firm on day d; $tf_{i,d,p}$ is the term frequency of the p^{th} optimism characteristic term for the i^{th} firm on day d; $tf_{i,d,n}$ is the term frequency of the n^{th} pessimism characteristic term for the i^{th} firm on day d; $TF_{i,d,m}$ is the total term frequency of the m^{th} news for the i^{th} firm on day d; $NSI_{i,d}$ is the average score for the m^{th} news for the i^{th} firm on day d which could be a proxy of the news sentiment level based upon public information relating to the i^{th} firm.

Table 3 presents parts of the sample characteristic terms for both optimism (Panel A) and pessimism (Panel B). The process from quality to quantity, which begins with the financial news collection, then uses word processing technology to construct the news

[6] Diction is a computer-aided text analysis program for determining the tone of a verbal message. For details of Diction, readers can refer to the website introduction at: http://www.dictionsoftware.com/.

sentiment indicators. For example, if twelve characteristic terms for optimism and two characteristic terms for pessimism are found to appear in a financial news report, the news sentiment index for that news sample would be 71.4286 per cent, based upon a calculation process of (12-2) / (12+2) ∗ 100 per cent = 71.4286 per cent. This result indicates that the news sample is relatively optimistic. The above sample is presented in table 3.

<Table 3 is inserted about here>

3.2 Market-level News Sentiment

The measure of ‘Aggregate news sentiment index in financial news’ (*ANSI*) is calculated by considering both the level of news sentiment and the market value of each stock. Our measurement of monthly aggregate news sentiment index is as follows:

$$ANSI_d = \sum_{i=1}^{I}(NSI_{i,d} \times k_{i,d-1}),\; k_{i,d} = MV_{i,d} / \sum_{i=1}^{I} MV_{i,d} \qquad (3)$$

$$ANSI_{\mathrm{m}} = \frac{1}{N}\left(\sum_{n=1}^{N} ANSI_{\mathrm{m}-n}\right) \qquad (4)$$

where $ANSI_d$ is the market aggregate news sentiment in financial news on day d; $NSI_{i,d}$ is the news sentiment level for the i^{th} firm on day d; $k_{i,d}$ is the weight of the market value for the i^{th} firm on day d; $MV_{i,d}$ is the market value of the i^{th} firm on day d; $ANSI_m$ is the m^{th} aggregate market monthly sentiment in financial news, which is calculated by the average mean of the $ANSI_{m-n}$, n = 1…N, where N equals the calendar days in each month, ranging from 28 to 31.

The evolution of the monthly *ANSI* within the Taiwan Stock Exchange Capitalization Weighted Stock Index (*TAIEX*) is presented in Figure 1.

<Figure 1 is inserted about here>

4. Sentiment news and stock returns

4.1 Data and Sample

This section begins with a description of our measure of 'aggregate news sentiment index' (*ANSI*), which is constructed by including the news reports relating to all listed stocks in the Taiwan Stock Exchange. We go on to investigate and apply the measure by incorporating the following monthly firm characteristics:

Turnover; *Age*; market value (*SIZE*, which is the natural logarithm of market value); market value to book value (*MVBV*); revenue (*Rev*, which is the natural logarithm of monthly revenue); earnings per share (*EPS*); total risk (*Sigma*, which is the annual standard deviation of returns from t–12 to t–1); and high-low range (*HL*, which is calculated as the highest closing price minus the lowest closing price and adjusted to the monthly closing price).

The sample firms are selected at the end of each month and then individually sorted into groups of five and ten portfolios for each of the firm characteristics, with the equally-weighted and value-weighted portfolio returns subsequently being calculated. The summary statistics of firm characteristics according to their quintile rankings are presented in Table 4. Panel A of Table 4, which provides the summary statistics of the 'aggregate news sentiment index' (*ANSI*) measure under high and low regimes, shows that the standard deviation of *ANSI* in the low regime is higher than that in the high regime; thus, the sentiment induced by financial news is associated with higher volatility when the news sentiment level is relatively low.[7]

Panel B of Table 4 provides the summary statistics when matching the data for *ANSI* in the previous month under high and low regimes. As we can see from the differences in the level of *ANSI* between the *SIZE*, *MV*/*BV*, *Turnover* and *Sigma* variables, the average mean in the high *ANSI* level is greater than that in the low *ANSI* level. Conversely, the means of

[7] "High *ANSI*" means *ANSI* is higher than average level and "Low *ANSI*" means *ANSI* is lower than average level.

Age and *Rev* in the high *ANSI* level are smaller than the average values in the low *ANSI* level.

<Table 4 is inserted about here>

The correlation analysis between all of the firm characteristics, provided in Table 5, reveals that higher *Sigma* and *HL* are found to have significantly negative relationships with *Age*, *Rev* and *EPS*, thereby implying that younger, lower revenue and lower EPS firms will be found to have higher deviations in both returns and price changes.

<Table 5 is inserted about here>

Furthermore, higher *Turnover* is found in the present study to have significant associations with smaller *SIZE*, higher *MV*/*BV*, younger firms, lower revenue and higher EPS. Finally, the profitability characteristic, *Rev*, reveals significantly positive relationships with *SIZE*, *MV*/*BV* and *Age*.

Obvious correlations are revealed between most of the firm characteristics in the correlation analysis presented in Table 5. We therefore go on to further investigate whether the aggregate news sentiment is capable of predicting the variations between the future returns of long and short portfolios.

4.2 Future Returns

The future returns of the portfolios based upon the firm characteristics and the aggregate news sentiment index, *ANSI*, are presented in Table 6, with the returns being calculated for each of the firm characteristics according to the five portfolio quintiles at the beginning of the month. We present the results for both equally-weighted and value-weighted portfolio returns, which are respectively reported in Panels A and B.

<Table 6 is inserted about here>

We classify *ANSI* in high and low regimes and compare the differences between the five portfolios. As shown in Panel A of Table 6, with a low *ANSI*, the equally- weighted returns for firms in the bottom *SIZE* quintile are found to have an average of 2.1208 per cent per month, as compared to 1.0318 per cent for firms in the top *SIZE* quintile. A similar pattern is discernible for all other firm characteristics, with the average returns across high and low *ANSI* indicating that, across most of the cross-equity, subsequent returns tend to be higher when *ANSI* is low.

The results calculated on the basis of value-weighting are summarized in Panel B of Table 6. With a low *ANSI*, the value-weighted returns of firms in the bottom *SIZE* quintile are found to be 1.6534 per cent, as compared to the lower returns of 1.3074 per cent per month for firms in the top *SIZE* quintile. Similar findings are again discernible for all other firm characteristics when the samples are sorted into five portfolio groups.

All of the results reported above are found to be consistent with those reported in the prior related studies in which the market regime is also classified as higher or lower sentiment, such as Baker and Wurgler (2006). The robustness analysis is done for groups of ten equally-weighted and value-weighted portfolio returns with high or low *ANSI* in the previous month. The results support the findings in Table 6 and are eliminated for the space consideration. The details are available from the authors upon request.

The returns constructed during periods of low levels of news sentiment are found to be higher than those constructed during periods of high levels of news sentiment, a finding which is consistent for all groups classified by *SIZE*, *MV*/*BV*, *Age*, *Rev*, *EPS*, *Turnover*, *Sigma* and *HL*, with notable exceptions in the low *ANSI* portfolios for *Turnover* and *Sigma*.

These results indicate that firms can earn higher returns in those cases where the *ANSI* in the previous month is found to be relatively low; furthermore, the portfolios returns constructed by firm characteristics and *ANSI* are found to be similar, regardless of whether they are calculated by equally-weighted or value- weighted measures, and irrespective of whether the returns are classified under groups of five or ten portfolios.

Our portfolio analyses suggest that when engaging in their portfolio decision making and portfolio management, participants in the Taiwan Stock Market should construct their portfolios based upon either equal or value weighting. We suggest that such participants should focus on younger firms with smaller size, lower *MV/BV*, lower *Rev* and lower *Turnover* when market sentiment is relatively low; conversely, they should construct their portfolios based upon higher *EPS*, higher *Sigma* and higher *HL* when the aggregate news sentiment index is lower.

Regardless of the specific characteristics that are to be taken into consideration during portfolio selection, we suggest that investors should analyze the aggregate news sentiment index prior to engaging in their portfolio decision making.

5. Empirical analyses

5.1 Regression Models

An alternative way of carrying out an examination of the conditional characteristic effects is to use the degree of news sentiment reports as a means of forecasting portfolios that are long (short) on stocks in the highest (lowest) group. In the present study, we deal with the question of whether the aggregate news sentiment index can predict the various long or short portflios using the following regression model:

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + \alpha_1 ANSI_{t-1} + u_{it} \qquad (5)$$

where the dependent variable is the monthly return on long and short portfolios, such as *SIZE*, and the monthly returns are regressed on the ‘aggregate news sentiment index’, *ANSI*, which prevailed during the previous month.

We also distinguish between ‘novel predictability’ effects and other risk factors using the following multivariate regressions:

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + \alpha_1 ANSI_{t-1} + \alpha_2 SI_t + u_{it} \qquad (6)$$

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + \alpha_1 ANSI_{t-1} + \alpha_2 RMRF_t + \alpha_3 SMB_t + \alpha_4 BVMV_t + \alpha_5 SI_t + u_{it} \quad (7)$$

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + \alpha_1 ANSI_{t-1} + \alpha_2 RMRF_t + \alpha_3 SMB_t + \alpha_4 BVMV_t + \alpha_5 SI_t + \alpha_6 MTM_t + u_{it} \quad (8)$$

where *RMRF* is the excess return of the market portfolio return minus the riskless return; *SMB* is the size premium factor; *BVMV* is the book-to-market premium factor, which is constructed in order to isolate the difference between high and low book- value to market-value portfolios; *SI* is the sentiment index which is constructed by means of principal component analysis; and *MTM* is the momentum factor which is the moving average monthly return from month −12 to −2. The *SMB* and *BVMV* variables are excluded from the right side of the equations when *SIZE* and *MV/BV* are the dependent variables.

The details of the *RMRF, SMB, BVMV* and *MTM* factors used in the present study were obtained from the Taiwan Economic Journal (TEJ) database, whilst our sentiment index (*SI*) was constructed by referring to Baker and Wurgler (2006). We adopt five sentiment indicators comprising of the ‘market turnover ratio’ (*MTO*), the ratio of the number of ‘advancing issues to declining issues’ (*ADVDEC*), the ‘short sales ratio’ (*SSR*), the ‘variation in spread between futures and spot prices’ (*FS*) and the ‘volatility index’ (*VIX*). A principal component analysis was subsequently adopted in order to construct our sentiment index.

We adopt principal components analysis to form an aggregate investor sentiment index from January 2003 to December 2012 by referring to Baker and Wurgler (2006). First, we exclude the influence of macroeconomic cycle variables including Industrial Production index (*IPI*), the Unemployment rate (*UNE*), a Monitoring indicator (*MI*), the Consumer Price index (*CPI*), Foreign Exchange in US dollars (*USE*), and the Term Structure of Interest Rates (*TSIR*). Each sentiment proxy is regressed by the macroeconomic variables and we get the standardized macro-orthogonalized sentiment proxies. Second, we run the first principal components analysis with the consideration of the current and lag term of each sentiment proxy and define the larger weight in prin1 of each sentiment proxy to choose the current or lag term. Third, we proceed to the second principal component and we retain an eigenvalue of above 0.8 from the results of the second principal components analysis, for the eigenvalues of the two principal components analyses are 1.54, 1.22 and 0.82 and can explain the 71.64% sample variance of the orthogonalized sentiment proxies. We can obtain the weight and formulate the sentiment index equation as follows:

$$\begin{aligned} SI_t = {} & (-0.2099)MTO_{t\text{-}1} + 0.2788ADVDEC_{t-1} + 0.2646SSR_t \\ & +0.1923FS_t + (-0.3532)\Delta VIX_t \end{aligned} \tag{9}$$

Table 7 presents the correlation analysis between the risk factors which reveals that *ANSI* is found to be positively correlated to *SI*, *RMRF* and *BVMV*, whereas a negative correlation is found between *ANSI, SMB* and *MTM*, although the results are not found to be significant.

<Table 7 is inserted about here>

5.2 Empirical Results

In this study, we consider both long and short portfolios, ‘$R_{x=Decile10}$-$R_{x=Decile1}$’ and ‘$R_{x=Quintile5}$-$R_{x=Quintile1}$’; that is, those that are long in the highest characteristic group and short in the lowest characteristic group. The predictive regressions, show the coefficient estimates on the news sentiment. As a check for robustness, we also go on to examine the long and short portfolio returns using equal weighting and value weighting, with the respective results being reported in Table 8 and Table 9, both of which provide strong support for the preliminary findings on the various factors.

The results on trading behavior reported in Panel A of Table 8 show that when the aggregate news sentiment index is high, returns on low turnover, low volatility and low ‘high-low range’ firms are relatively high over the subsequent month. As shown in Model

(4), the coefficient on *ANSI* remains consistent once controls are put in place for *SI RMRF*, *SMB*, *BVMV*, and *MTM*.

The results on firm age, reported in Panel D of Table 8, show that when the aggregate news sentiment index is high, the returns of older firms over the subsequent month are relatively high. In terms of magnitude, the coefficient predicting *Turnover*, for example, indicates that a one-unit increase in *ANSI* is associated with a −0.4584 per cent lower monthly return on the high minus low turnover portfolio.

<Table 8 is inserted about here>

The results reported in Table 8 also reveal very similar coefficients on both '$R_{x=Decile10}$-$R_{x=Decile1}$' and '$R_{x=Quintile5}$-$R_{x=Quintile1}$'. The characteristics of firm size (Panel B) and profitability (Panel C) indicate that the aggregate news sentiment index has no significant predictability when controlling for all other factors.

The results summarized in Table 9 are found to be very similar to those reported in Table 8, with the main difference between the two tables being that long and short portfolio returns are calculated individually with equal and value weightings. Although the results reported by Baker and Wurgler (2006) used only equally- weighted calculations, the calculation of the results in the present study makes use of value weighting as a check for the robustness of our earlier findings.

<Table 9 is inserted about here>

The results in Table 9 are found to provide further support for the view that when the aggregate news sentiment index is high, returns on low turnover, low volatility, low 'high-low range' and older firms will be relatively high over the subsequent month, as shown in Model (4), which includes controls for the *SI*, *RMRF*, *SMB*, *BVMV* and *MTM* factors in the long and short portfolios, '$R_{x=Decile10}$-$R_{x=Decile1}$' and '$R_{x=Quintile5}$-$R_{x=Quintile1}$'.

These results provide general support for our supposition that the aggregate news sentiment index has stronger effects on stocks where firm characteristics are influenced by the prospections and trading behavior of investors, with the results effectively complementing the findings of Baker and Wurgler (2006, 2007).

6. Conclusions

We set out in the present study with the primary aim of measuring the aggregate news sentiment index (*ANSI*); we aim to achieve this by incorporating public news relating to each individual stock listed on the Taiwan Stock Exchange and examining the effects on the cross-section of stock returns whilst controlling for market, *SMB*, *BVMV*, momentum and sentiment factors.

The major differences between the present study and the extant related literature are that *ANSI* is derived in this study using linguistic analysis to extract all representative and related information from news reports. To the best of our knowledge, our study is the first to attempt to create a market index of the news sentiment relating to each of the traded stocks. Our empirical results, which report the average returns across high and low *ANSI* samples, indicate that across most of the cross-equity, the subsequent returns will tend to be higher when *ANSI* is low.

These results are consistent with those reported in the prior studies where the market regime is classified as either high or low sentiment. Furthermore, our results are found to remain robust, regardless of whether the portfolios are classified into groups of five or ten, or whether the portfolio returns are calculated using equal weighting or value weighting.

The predictive regressions carried out on the long and short portfolios reveal that, when controlling for market, *SMB*, *BVMV*, momentum and sentiment factors and when the aggregate news sentiment index is high, returns on low turnover, low volatility, low ‘high-low range’ and older firms are found to be relatively high over the subsequent month. These results provide strong support for the argument that the aggregate news sentiment index will have stronger effects on those stocks with firm characteristics that are influenced by the prospections and trading behavior of investors, thereby extending the findings reported by Baker and Wurgler (2006, 2007).

We suggest that our study complements the existing line of research into the trading behavior of investors and the cross-section of stock returns and further recommend that the application of linguistic analysis should be considered as a means of extracting the degree of sentiment from public news, provided that appropriate filtering is applied. Investors could use this process to further analyze the aggregate news sentiment index prior to making any decisions on their proposed portfolios. Finally, this study also provides a further contribution to the prior related studies by demonstrating that the aggregate news sentiment index could also be considered to be an important behavioral factor in the examination of the cross-section of stock returns.

References

Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 2, 785-818.

Baker, M., Stein, J., 2004. Market liquidity as a sentiment indicator. Journal of Financial Market 7, 271-299.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 4, 1645-1680.

Baker, M., Wurgler J., 2007. Investor sentiment in the stock market. Journal of Economic Perspectives 21, 129-151.

Baker, M., Wurgler, J., Yuan, Y., 2012. Global, local and contagious investor sentiment. Journal of Financial Economics 104, 272-287.

Brown, G.W., Cliff, M.T., 2004. Investor sentiment and the near-term stock market. Journal of Empirical Finance 11, 1-27.

Brown, G.W., Cliff, M.T., 2005, Investor sentiment and asset valuation. Journal of Business, 78, 405-440.

Chung, S.L., Hung, C.H., Yeh, C.Y., 2012. When does investor sentiment predict stock returns? Journal of Empirical Finance 19, 217-240.

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

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns?. Journal of Finance 157, 2185-2221.

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

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

Fischer, P.E., Verrecchia, R.E., 1999. Public information and heuristic trade. Journal of Accounting and Economics 27, 89-124.

Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J., 1992. Knowledge discovery in databases: An Overview. AI Magazine 13(3), 57-70.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3), 424-438.

Griffin, J.M., Hirschey, N.H., Kelly, P.J., 2011. How important is the financial media in global markets? Review of Financial Studies 24(12), 3941-3992.

Groß-Klußmann, A., Hautsch, N., 2011. When machines read the news: using automated text analytics to quantify high-frequency news-implied market reactions. Journal of Empirical Finance 18, 321-340.

Jansen, W.J., Nahuis, N.J., 2003. The stock market and consumer confidence: European evidence. Economics Letters 79, 89-98.

Jegadeesh, N., Titman, S., 1995. Overreaction, delayed reaction and contrarian profits.

Review of Financial Studies 8, 973-993.

Jiang, G.J., Konstantinidi, E., Skiadopoulos, G., 2012. Volatility spillovers and the effect of news announcements. Journal of Banking and Finance 36, 2260-2273.

Kumar, A., Lee, C.M.C., 2006. Retail investor sentiment and return co-movements. Journal of Finance 61(6), 2451-2486.

Lu, Y.C., Wei, Y.C., 2014. Media impacts around earnings announcement dates with consideration of investor types and market scenarios. Journal of Financial Studies 22(3), 73-104.

Nofsinger, J.R., 2001. The impact of public information on investors. Journal of Banking and Finance 25, 1339-1366.

Schmeling, M., 2009. Investor sentiment and stock returns: some international evidence. Journal of Empirical Finance 16, 394-408.

Shu, H.C., 2010. Investor mood and financial markets. Journal of Economic Behavior and Organization 76, 267-282.

Sims, C.A., 1972. Money, income and causality. American economic review 62(4), 540-552.

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

Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. Journal of Finance 63(3), 1437-1467.

Vega, C., 2006. Stock price reaction to public and private information. Journal of Financial Economics, 82, 103-133.

Zhang, X.F., 2006. Information uncertainty and stock returns. Journal of Finance 61(1), 105-137.

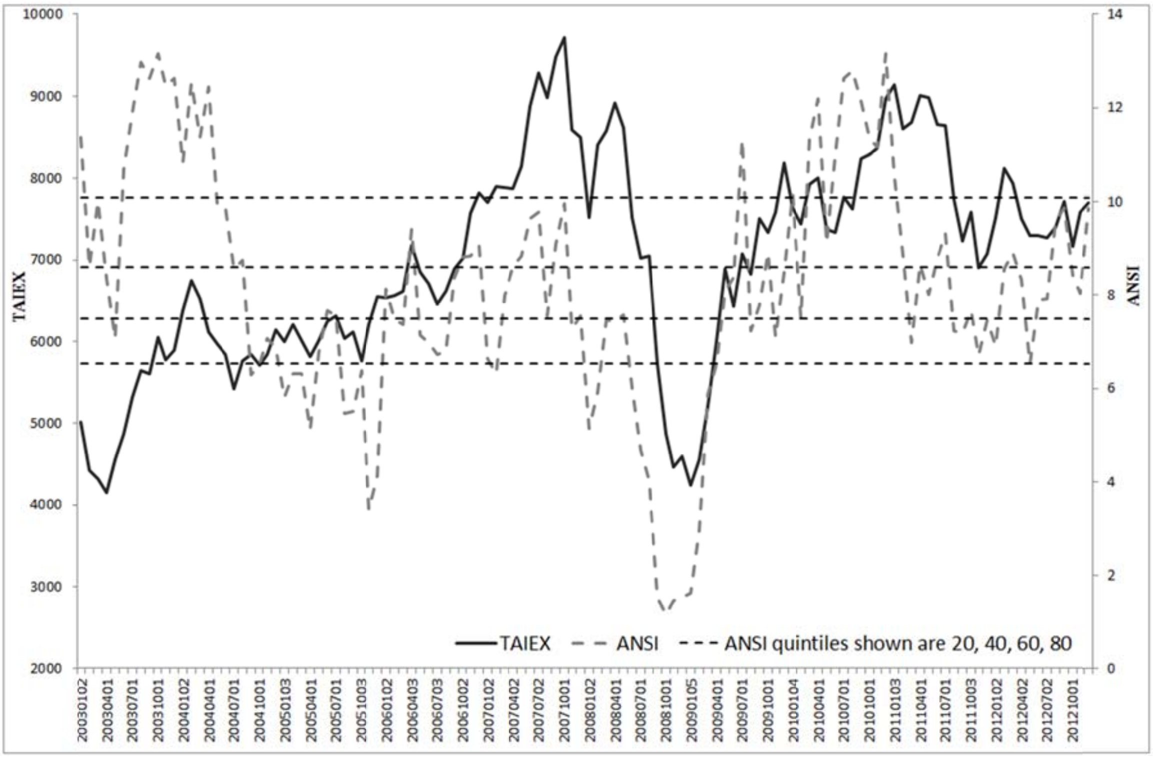

Figure 1 Monthly evolution of degree of aggregate news sentiment index and the TAIEX

Note: The figure illustrates the historical pattern on the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), the TAIEX returns and the aggregate news sentiment index (*ANSI*) for the period from January 2003 to December 2012; the *ANSI* quintiles shown are 20, 40, 60 and 80.

Table 1 Summary statistics of the news report samples, January 2003-December 2012

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2003	9,016 (17)	7,146 (14)	10,705 (20)	9,677 (18)	9,574 (18)	12,890 (23)	10,787 (20)	10,677 (20)	13,296 (25)	10,639 (20)	9,476 (17)	12,877 (24)	126,760 (237)
2004	6,995 (14)	9,017 (17)	12,456 (23)	9,508 (17)	9,204 (17)	12,376 (22)	10,626 (19)	9,794 (17)	7,743 (14)	5,808 (11)	5,527 (11)	8,329 (16)	107,383 (201)
2005	5,892 (12)	4,190 (9)	8,529 (16)	5,529 (11)	6,810 (13)	9,760 (17)	6,595 (12)	6,401 (12)	7,716 (15)	7,489 (15)	3,939 (10)	5,833 (14)	78,683 (157)
2006	6,568 (13)	6,194 (12)	9,257 (17)	7,843 (15)	8,421 (15)	11,052 (20)	8,231 (15)	8,978 (16)	10,958 (19)	8,553 (15)	8,762 (16)	11,183 (20)	106,000 (195)
2007	7,746 (14)	5,290 (11)	9,516 (17)	7,482 (13)	8,379 (15)	10,436 (19)	8,351 (15)	8,402 (15)	10,308 (18)	8,023 (14)	8,343 (15)	11,578 (21)	103,854 (186)
2008	7,050 (13)	4,942 (10)	6,837 (12)	6,408 (12)	6,779 (12)	6,239 (12)	8,351 (14)	6,959 (13)	7,327 (13)	6,781 (12)	7,017 (14)	7,486 (14)	82,176 (151)
2009	5,249 (10)	6,081 (12)	7,229 (13)	6,648 (12)	7,146 (13)	7,227 (13)	7,691 (13)	7,219 (12)	7,063 (12)	6,824 (12)	6,093 (11)	6,323 (11)	80,793 (145)
2010	9,685 (17)	7,618 (15)	11,866 (20)	10,210 (17)	9,950 (17)	11,503 (20)	10,131 (17)	13,211 (22)	12,289 (20)	10,421 (18)	10,639 (18)	13,423 (23)	130,946 (224)
2011	11,185 (19)	8,650 (16)	13,206 (22)	9,813 (16)	11,255 (18)	12,806 (21)	10,860 (18)	11,662 (19)	11,746 (20)	16,781 (26)	16,992 (27)	17,981 (28)	152,937 (252)
2012	12,486 (26)	13,510 (27)	14,658 (30)	15,226 (23)	15,769 (24)	15,903 (24)	16,120 (25)	16,250 (25)	15,097 (24)	16,824 (25)	16,843 (26)	17,539 (26)	186,225 (304)
Totals	81,872 (155)	72,638 (142)	104,259 (191)	88,344 (157)	93,287 (164)	110,192 (192)	97,743 (168)	99,553 (173)	103,543 (182)	98,143 (172)	93,631 (171)	112,552 (202)	1,155,757

Note: The table presents the summary statistics of the news reported samples, with the figures in parentheses reporting the average news reported samples for each of the reported corporate samples. The news reported samples are collected from the China Times and the Commercial Times data contained in the InfoTimes database.

Table 2 Summary statistics of the corporate samples, January 2003-December 2012

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean
2003	522 (84.2)	515 (83.1)	528 (84.8)	526 (83.9)	533 (85.0)	555 (88.7)	537 (85.5)	545 (86.4)	542 (85.4)	545 (85.4)	544 (84.7)	528 (81.6)	535 (84.9)
2004	503 (77.6)	525 (80.9)	534 (81.7)	552 (84.1)	545 (83.1)	560 (85.4)	567 (86.3)	560 (85.0)	535 (81.2)	511 (77.5)	499 (75.7)	512 (77.5)	534 (81.3)
2005	505 (76.1)	473 (71.3)	527 (79.2)	506 (75.7)	514 (76.9)	565 (84.3)	551 (82.1)	536 (79.9)	517 (76.8)	510 (75.9)	388 (57.5)	415 (61.2)	501 (74.7)
2006	516 (76.1)	501 (73.8)	535 (78.7)	536 (78.6)	553 (81.0)	545 (79.8)	555 (81.1)	547 (79.9)	562 (81.9)	562 (81.7)	554 (80.3)	563 (81.6)	544 (79.5)
2007	553 (79.9)	498 (72.0)	559 (80.8)	565 (81.6)	561 (81.1)	563 (81.4)	570 (82.3)	563 (81.1)	568 (81.7)	563 (80.8)	566 (80.4)	555 (78.4)	557 (80.1)
2008	530 (74.5)	497 (69.8)	548 (76.8)	542 (75.8)	556 (77.8)	531 (74.3)	582 (81.4)	553 (77.3)	566 (79.1)	550 (76.8)	517 (72.2)	540 (75.1)	543 (75.9)
2009	519 (72.0)	512 (70.9)	544 (74.6)	558 (76.4)	560 (76.6)	564 (77.0)	589 (80.5)	592 (80.7)	589 (80.2)	567 (76.9)	549 (74.4)	556 (74.7)	558 (76.3)
2010	586 (78.7)	524 (70.3)	580 (77.7)	608 (81.4)	579 (77.5)	577 (77.0)	597 (79.5)	606 (80.6)	600 (79.7)	590 (78.4)	578 (76.5)	582 (76.6)	584 (77.8)
2011	583 (76.6)	554 (72.8)	609 (79.8)	597 (78.1)	617 (80.7)	611 (79.8)	617 (80.5)	599 (78.0)	577 (74.9)	646 (83.4)	630 (81.0)	648 (83.2)	607 (79.1)
2012	473 (60.7)	504 (64.7)	493 (63.1)	650 (83.1)	654 (83.4)	658 (83.9)	656 (83.6)	652 (83.1)	640 (81.5)	661 (83.9)	639 (80.9)	662 (83.3)	612 (78.0)
Mean	529 (75.4)	510 (72.7)	546 (77.4)	564 (79.9)	567 (80.3)	573 (81.0)	582 (82.2)	575 (81.1)	570 (80.2)	571 (80.1)	546 (76.4)	556 (77.4)	557 (78.7)

Note: The table presents the summary statistics of the reported corporate samples, with the figures in parentheses reporting the media coverage (in percentage terms). The news reported samples are collected from the Commercial Times and China Times data contained in the InfoTimes database.

Table 3 Characteristic terms for optimism and pessimism[a]

Panel A: Characteristic Terms for Optimism[b]					
Abundant (豐富)	Active (積極)	Advantage (優勢)	Amazing (驚人)	Upgrade (調升)	Upturn (好轉)
Benefit (受惠)	Best (最佳)	Breakthrough (突破)	Conducive (有助於)	Success (成功)	Transcend (優於)
Exceed (超越)	Excellent (出色)	Explode (激增)	Favor (看好)	Steady (穩定)	Stimulate (激勵)
Fluency (暢旺)	Growth (成長)	Increase (提升)	Influx (湧進)	Rebound (反彈)	Rise (上漲)
Lead (領先)	New Height (新高)	Optimistic (樂觀)	Overbought (買超)	Profit (獲利)	Prospect (展望)
Turn Loss into Gain (轉虧為盈)	Upper Price Limit (漲停)	Recovery (復甦)	Victory (捷報)	Jumped (躍升)	Powerful (強勁)

Panel B: Characteristic Terms for Pessimism[b]					
Bankruptcy (破產)	Capital-reducing (減資)	Cheapen (跌價)	Crisis (危機)	Serious (嚴重)	Weak (疲弱)
Decline (衰退)	Decrease (減少)	Deficit (赤字)	Depreciate (貶值)	Plunge (重挫)	Reverse (失利)
Depression (不景氣)	Deteriorate (惡化)	Dispirit (不振)	Down (下滑)	Pessimistic (悲觀)	Over-Fall (跌破)
Downgrade (調降)	Lower Price Limit (跌停)	Encumber (拖累)	Fail (告吹)	One Disaster After Another (雪上加霜)	Not Good Enough (不佳)
Fall (下跌)	Impact (衝擊)	Involve (波及)	Loss (虧損)	New Low (新低)	Mournful 悽慘
Negative (負面)	Oversold (賣超)	Embroil (拖累)	Frustrated (受挫)	Thump (重擊)	Downturn (低迷)

Notes:

[a] The 'characteristic terms' refer to those terms which can adequately describe the characteristics of the optimism and pessimism groups.

[b] The characters in parentheses are the Chinese equivalents for each of the special terms. A complete list of special terms for each group is available from the authors upon request.

Table 4 Summary statistics of ANSI and firm characteristic variables[a]

Variables[b]	Full Sample				Quintile Means				
	Mean	S.D.	Min	Max	1	2	3	4	5
Panel A: *ANSI*									
High *ANSI*	10.18	1.61	8.16	13.18	–	–	–	–	–
Low *ANSI*	6.20	1.83	1.16	8.06	–	–	–	–	–
Difference	3.98	–0.22	7.00	5.12	–	–	–	–	–
Panel B: Firm Characteristic Variables									
SIZE									
High *ANSI*	8.70	1.47	3.47	14.66	6.89	7.87	8.53	9.26	10.89
Low *ANSI*	8.57	1.51	3.58	14.75	6.70	7.72	8.40	9.16	10.83
Difference	0.14	–0.04	–0.12	–0.09	0.19	0.16	0.13	0.11	0.06
MV/BV									
High *ANSI*	1.73	2.01	0.09	169.61	0.69	1.02	1.36	1.89	3.62
Low *ANSI*	1.50	1.41	0.00	94.10	0.58	0.87	1.18	1.65	3.20
Difference	0.23	0.61	0.09	75.51	0.11	0.15	0.19	0.24	0.42
Age									
High *ANSI*	12.52	9.37	0.99	50.72	4.04	7.21	9.74	13.75	26.88
Low *ANSI*	12.59	9.27	0.99	50.88	4.22	7.30	9.81	13.85	26.96
Difference	–0.06	0.10	0.00	–0.17	–0.18	–0.09	–0.07	–0.10	–0.08
Rev									
High *ANSI*	15.07	1.64	4.13	21.96	13.07	14.27	14.95	15.74	17.31
Low *ANSI*	15.16	1.66	4.13	22.09	13.13	14.33	15.03	15.84	17.42
Difference	–0.09	–0.02	0.00	–0.12	–0.06	–0.06	–0.08	–0.10	–0.11
EPS									
High *ANSI*	1.20	2.37	–13.06	73.32	–0.85	0.31	0.85	1.63	4.01
Low *ANSI*	1.17	2.56	–52.32	73.32	–1.05	0.26	0.83	1.64	4.15
Difference	0.03	–0.20	39.26	0.00	0.20	0.05	0.01	0.00	–0.14
Turnover									
High *ANSI*	19.30	22.93	0.00	291.06	2.54	6.99	12.87	22.57	52.54
Low *ANSI*	14.63	19.49	0.00	257.56	1.70	4.62	8.68	16.17	43.01
Difference	4.67	3.44	0.00	33.50	0.83	2.37	4.18	6.40	9.52
Sigma									
High *ANSI*	43.67	23.45	2.50	296.40	21.51	31.76	39.74	49.83	77.07
Low *ANSI*	42.06	21.42	4.29	411.03	21.13	31.18	38.91	48.20	71.32
Difference	1.61	2.04	–1.79	–114.63	0.38	0.57	0.82	1.63	5.75
HL									
High *ANSI*	17.10	9.81	0.00	149.93	7.92	12.01	15.49	19.95	30.77
Low *ANSI*	18.94	11.99	0.00	298.22	8.91	13.55	17.31	22.11	33.46
Difference	–1.84	–2.18	0.00	–148.30	–0.99	–1.54	–1.82	–2.16	–2.69

Notes:

[a] The table reports the summary statistics of the aggregate news sentiment index (*ANSI*) and firm characteristic variables using monthly data running from January 2003 to December 2012.

[b] *ANSI* is the monthly 'aggregate news sentiment index'; *SIZE* is the natural logarithm of market value; *MV/BV* is the market-to-book value; *Rev* is the natural logarithm of monthly revenue; *EPS* is earnings per share; *Sigma* is the annual standard deviation of returns from *t*-12 to *t*-1; and *HL* is the measure of the high-low range.

Table 5 Correlation matrix of firm characteristic variables[a]

Variables[b,c]	*SIZE*	*MV/BV*	*Age*	*Rev*	*EPS*	*Turnover*	*Sigma*
MV/BV	0.1343***						
Age	0.0693***	–0.1322***					
Rev	0.5672***	0.0496***	0.0578***				
EPS	0.2062***	0.3455***	–0.1317***	0.1588***			
Turnover	–0.0635***	0.1503***	–0.1179***	–0.0591***	0.0667***		
Sigma	–0.1143***	0.0583***	–0.0327***	–0.0831***	–0.1338***	0.3010***	
HL	–0.0864***	0.0002	–0.0422***	–0.0548***	–0.1172***	0.3345***	0.3902***

Notes:

[a] The table reports the correlations between the firm characteristic variables using data running from January 2003 to December 2012.

[b] *SIZE* is the natural logarithm of market value; *MV/BV* is the market-to-book value; *Rev* is the natural logarithm of monthly revenue; *EPS* is earnings per share; *Sigma* is the annual standard deviation of returns from *t*-12 to *t*-1; and *HL* is the measure of the high-low range.

[c] *** indicates significance at the 1 per cent level.

Table 6 Future returns of five portfolios, by ANSI and firm characteristic variables

Variables	Quintiles					Comparison
	1	2	3	4	5	5-1
Panel A: Equally-weighted Returns (*Sentiment*$_{t-1}$)						
SIZE						
High ANSI	1.4241	0.7653	0.9355	0.6042	0.6763	–0.7479
Low ANSI	2.1208	1.8977	1.5156	1.6484	1.0318	–1.0889
Difference	–0.6966	–1.1324	–0.5801	–1.0441	–0.3556	0.3411
MV/BV						
High ANSI	2.0530	0.9585	0.6550	0.4286	0.3372	–1.7158
Low ANSI	2.2653	1.7346	1.5137	1.3498	1.3807	–0.8846
Difference	–0.2123	–0.7761	–0.8587	–0.9212	–1.0435	–0.8312
Age						
High ANSI	0.3037	0.6245	0.8670	1.1351	1.4733	1.1696
Low ANSI	1.8331	1.8871	1.4975	1.3963	1.6054	–0.2277
Difference	–1.5295	–1.2626	–0.6305	–0.2611	–0.1322	1.3973
Rev						
High ANSI	1.0835	0.8075	0.8734	0.7613	0.8798	–0.2037
Low ANSI	1.9009	1.9885	1.5669	1.5003	1.2610	–0.6399
Difference	–0.8175	–1.1811	–0.6936	–0.7390	–0.3812	0.4362
EPS						
High ANSI	0.0233	0.7193	1.0050	1.2933	1.3476	1.3243
Low ANSI	1.2088	1.3922	1.7367	1.9006	1.9561	0.7473
Difference	–1.1856	–0.6729	–0.7317	–0.6073	–0.6086	0.5770
Turnover						
High ANSI	1.6829	1.2993	1.1332	0.6044	–0.3100	–1.9929
Low ANSI	1.5430	1.5659	1.7256	1.9370	1.4470	–0.0960
Difference	0.1399	–0.2666	–0.5924	–1.3326	–1.7570	–1.8969
Sigma						
High ANSI	1.3366	1.2129	0.8994	0.6492	0.2983	–1.0383
Low ANSI	1.1176	1.6598	1.6905	1.7080	2.0470	0.9295
Difference	0.2190	–0.4468	–0.7911	–1.0588	–1.7487	–1.9678
HL						
High ANSI	1.0376	0.9707	0.7717	0.8590	0.7624	–0.2753
Low ANSI	1.2697	1.5061	1.5727	1.5598	2.3090	1.0393
Difference	–0.2321	–0.5354	–0.8010	–0.7008	–1.5467	–1.3145

Table 6 (*Contd.*)

Variables	Quintiles					Comparison
	1	2	3	4	5	5-1
Panel B: Value-weighted Returns (*ANSI*$_{t-1}$)						
SIZE						
High ANSI	1.0378	0.6723	0.9107	0.6503	0.8623	–0.1755
Low ANSI	1.6534	1.7171	1.4973	1.7650	1.3074	–0.3460
Difference	–0.6157	–1.0449	–0.5866	–1.1148	–0.4452	0.1705
MV/BV						
High ANSI	1.6790	0.9198	0.6383	0.5131	0.3831	–1.2959
Low ANSI	1.9184	1.6383	1.4809	1.3908	1.5120	–0.4064
Difference	–0.2394	–0.7185	–0.8427	–0.8777	–1.1289	–0.8895
Age						
High ANSI	0.2944	0.5917	0.7727	1.0440	1.4305	1.1361
Low ANSI	1.7845	1.7938	1.4266	1.3420	1.5935	–0.1910
Difference	–1.4900	–1.2022	–0.6540	–0.2979	–0.1630	1.3270
Rev						
High ANSI	0.8147	0.6975	0.8021	0.7543	1.0647	0.2500
Low ANSI	1.5849	1.7795	1.5490	1.5351	1.4919	–0.0931
Difference	–0.7702	–1.0819	–0.7470	–0.7808	–0.4271	0.3431
EPS						
High ANSI	–0.1245	0.5893	0.9112	1.2895	1.4678	1.5923
Low ANSI	1.1132	1.1884	1.6738	1.8885	2.0765	0.9633
Difference	–1.2377	–0.5990	–0.7625	–0.5990	–0.6088	0.6290
Turnover						
High ANSI	1.4064	1.2546	1.1450	0.6018	–0.2745	–1.6809
Low ANSI	1.3415	1.4509	1.6738	1.9267	1.5476	0.2062
Difference	0.0650	–0.1963	–0.5287	–1.3249	–1.8221	–1.8871
Sigma						
High ANSI	1.3527	1.1875	0.7989	0.6004	0.1938	–1.1589
Low ANSI	1.0804	1.5854	1.7006	1.6726	1.9013	0.8209
Difference	0.2723	–0.3980	–0.9017	–1.0722	–1.7075	–1.9798
HL						
High ANSI	1.0272	0.9162	0.7168	0.8267	0.6465	–0.3807
Low ANSI	1.1876	1.4494	1.5024	1.5560	2.2449	1.0573
Difference	–0.1604	–0.5333	–0.7856	–0.7293	–1.5984	–1.4380

Notes:

[a] The table reports the monthly equally-weighted and value-weighted future returns on five portfolios for the 'aggregate news sentiment index' (*ANSI*) and firm characteristic variables from January 2003 to December 2012.

[b] *SIZE* is the natural logarithm of market value; *MV/BV* is the market-to-book value; *Rev* is the natural logarithm of monthly revenue; *EPS* is earnings per share; *Sigma* is the annual standard deviation of returns from *t*-12 to *t*-1; and *HL* is the measure of the high-low range.

Table 7 Correlations of risk factors

	ANSI	*SI*	*RMRF*	*SMB*	*BVMV*
SI	0.1887**				
RMRF	0.2894***	0.3597***			
SMB	–0.0687	0.1397	0.1543*		
BVMV	0.2228**	0.1644*	0.3917***	0.0623	
MTM	–0.0198	0.0430	–0.2037**	–0.1011	–0.0224

Notes:

[a] The table reports the correlations of the risk factors from January 2003 to December 2012.

[b] *ANSI* is the monthly 'aggregate news sentiment index'; *SI* is the sentiment index, which is constructed using principal component analysis; *RMRF* is the excess return of the market portfolio return minus the riskless return; *SMB* is the size premium factor; *BVMV* is the book-to-market premium factor which is constructed in order to isolate the difference between high and low book-value to market-value portfolios; and *MTM* is the momentum factor which is the moving average monthly return from month -12 to -2.

[c] * indicates significance at the 10 per cent level; ** indicates significance at the 5 per cent level; and *** indicates significance at the 1 per cent level.

Table 8 Predictive regressions on equally-weighted long and short portfolio returns

Variables[a]		Model (1)[b,c]		Model (2)[b,c]		Model (3)[b,c]		Model (4)[b,c]	
		Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.	Coeff.	*t*-stat.
Panel A: Trading Behavior									
$R_{x=Quintile5} - R_{x=Quintile1}$	*Turnover*	–0.4584***	–2.6447	–0.4612***	–2.5910	–0.4105**	–2.2131	–0.4127**	–2.2371
	Sigma	–0.3835*	–1.9328	–0.4135**	–2.0338	–0.4830**	–2.2523	–0.4841**	–2.2512
	HL	–0.3018*	–1.7831	–0.3435**	–1.9880	–0.3571**	–1.9632	–0.3585**	–1.9700
$R_{x=Decile10} - R_{x=Decile1}$	*Turnover*	–0.5017**	–2.2720	–0.5043**	–2.2240	–0.4461*	–1.8821	–0.4492*	–1.9107
	Sigma	–0.4199*	–1.6902	–0.4479*	–1.7577	–0.5444**	–2.0271	–0.5458**	–2.0269
	HL	–0.3583	–1.6161	–0.4099*	–1.8095	–0.4295*	–1.8079	–0.4308*	–1.8090
Panel B: Firm Size									
$R_{x=Quintile5} - R_{x=Quintile1}$	*SIZE*	0.0933	0.5707	0.1339	0.8023	0.2092	1.2407	0.2171	1.2749
	MV/BV	–0.1463	–0.7864	–0.1284	–0.6727	–0.0061	–0.0320	–0.0158	–0.0818
$R_{x=Decile10} - R_{x=Decile1}$	*SIZE*	0.0167	0.0777	0.0932	0.4265	0.2285	1.0544	0.2431	1.1117
	MV/BV	–0.1142	–0.4797	–0.0704	–0.2888	0.0929	0.3840	0.0886	0.3622
Panel C: Profitability									
$R_{x=Quintile5} - R_{x=Quintile1}$	*Rev*	0.1317	1.0349	0.1603	1.2326	0.1880	1.3658	0.1876	1.3574
	EPS	0.0962	0.4766	0.1654	0.8070	0.2566	1.1942	0.2564	1.1881
$R_{x=Decile10} - R_{x=Decile1}$	*Rev*	0.1869	1.0897	0.2290	1.3072	0.3011	1.6351	0.3004	1.6257
	EPS	0.1347	0.5716	0.2066	0.8614	0.3255	1.2997	0.3251	1.2925
Panel D: Other Factors									
$R_{x=Quintile5} - R_{x=Quintile1}$	*Age*	0.3133*	1.7843	0.3692**	2.0675	0.3202*	1.7116	0.3218*	1.7224
$R_{x=Decile10} - R_{x=Decile1}$	*Age*	0.3459	1.6086	0.4035*	1.8400	0.3524	1.5339	0.3540	1.5394

Notes:

[a] The table reports the results of the predictive regressions on equally-weighted long and short portfolio returns. The risk factors are: *SI*, the sentiment index which is constructed using principal component analysis; *RMRF*, the market portfolio excess return minus the riskless return; *SMB*, the size premium factor; *BVMV*, the book-to-market premium factor which is constructed to isolate the difference between high and low book-value to market-value portfolios; and MTM, the momentum factor which is the moving average monthly return from month -12 to -2. The firm characteristics are *Age*; *Turnover*; *SIZE*, the natural logarithm of market value; *MV/BV*, the market-to-book value; *Rev*, the natural logarithm of monthly revenue; *EPS*, earnings per share; Sigma, the annual standard deviation in returns from *t*-12 to *t*-1; and *HL*, the high-low range measure. *SMB* and *BVMV* are not included as control variables when *SIZE* and *MV/BV* are the dependent variables.

[b] *ANSI* is the monthly 'aggregate news sentiment index'; Model (1) refers to $ANSI_{t-1}$ with no controls; Model (2) refers to $ANSI_{t-1}$ with a control for *SI*; Model (3) refers to $ANSI_{t-1}$ with controls for *SI*, *RMRF*, *SMB* and *BVMV*; and Model (4) refers to $ANSI_{t-1}$ with controls for *SI*, *RMRF*, *SMB*, *BVMV* and *MTM.*

[c] * indicates significance at the 10 per cent level; ** indicates significance at the 5 per cent level; and *** indicates significance at the 1 per cent level.

Table 9 Predictive regressions on value-weighted long and short portfolio returns

Variables[a]		Model (1)[b,c] Coeff.	Model (1)[b,c] *t*-stat.	Model (2)[b,c] Coeff.	Model (2)[b,c] *t*-stat.	Model (3)[b,c] Coeff.	Model (3)[b,c] *t*-stat.	Model (4)[b,c] Coeff.	Model (4)[b,c] *t*-stat.
Panel A: Trading Behavior									
$R_{x=Quintile5}$ - $R_{x=Quintile1}$	*Turnover*	–0.5017**	–2.2720	–0.5043**	–2.2240	–0.4461*	–1.8821	–0.4492*	–1.9107
	Sigma	–0.4199*	–1.6902	–0.4479*	–1.7577	–0.5444**	–2.0271	–0.5458**	–2.0269
	HL	–0.3583	–1.6161	–0.4099*	–1.8095	–0.4295*	–1.8079	–0.4308*	–1.8090
$R_{x=Decile10}$ - $R_{x=Decile1}$	*Turnover*	–0.4850**	–1.9753	–0.5070**	–2.0124	–0.4809*	–1.8057	–0.4842*	–1.8302
	Sigma	–0.4293*	–1.8935	–0.4527*	–1.9463	–0.5218**	–2.1252	–0.5232**	–2.1274
	HL	–0.3794*	–1.7425	–0.4269*	–1.9179	–0.4313*	–1.8459	–0.4328*	–1.8504
Panel B: Firm Size									
$R_{x=Quintile5}$ - $R_{x=Quintile1}$	*SIZE*	0.0167	0.0777	0.0932	0.4265	0.2285	1.0544	0.2431	1.1117
	MV/BV	–0.1142	–0.4797	–0.0704	–0.2888	0.0929	0.3840	0.0886	0.3622
$R_{x=Decile10}$ - $R_{x=Decile1}$	*SIZE*	0.0555	0.2718	0.0914	0.4372	0.1468	0.6924	0.1524	0.7111
	MV/BV	–0.1273	–0.6207	–0.1097	–0.5209	0.0137	0.0650	0.0022	0.0103
Panel C: Profitability									
$R_{x=Quintile5}$ - $R_{x=Quintile1}$	*Rev*	0.1869	1.0897	0.2290	1.3072	0.3011	1.6351	0.3004	1.6257
	EPS	0.1347	0.5716	0.2066	0.8614	0.3255	1.2997	0.3251	1.2925
$R_{x=Decile10}$ - $R_{x=Decile1}$	*Rev*	0.2082	1.1601	0.2249	1.2211	0.2540	1.3023	0.2527	1.2941
	EPS	0.1401	0.6659	0.1880	0.8746	0.2684	1.1930	0.2677	1.1854
Panel D: Other Factors									
$R_{x=Quintile5}$ - $R_{x=Quintile1}$	*Age*	0.3459	1.6086	0.4035*	1.8400	0.3524	1.5339	0.3540	1.5394
$R_{x=Decile10}$ - $R_{x=Decile1}$	*Age*	0.3235	1.4961	0.3850*	1.7475	0.3374	1.4619	0.3387	1.4649

Notes:

[a] The table reports the results of the predictive regressions on value-weighted long and short portfolio returns. The risk factors are: *SI*, the sentiment index which is constructed using principal component analysis; *RMRF*, the market portfolio excess return minus the riskless return; *SMB*, the size premium factor; *BVMV*, the book-to-market premium factor which is constructed to isolate the difference between high and low book-value to market-value portfolios; and MTM, the momentum factor which is the moving average monthly return from month -12 to -2. The firm characteristics are *Age*; *Turnover*; *SIZE*, the natural logarithm of market value; *MV/BV*, the market-to-book value; *Rev*, the natural logarithm of monthly revenue; *EPS*, earnings per share; Sigma, the annual standard deviation in returns from *t*-12 to *t*-1; and *HL*, the high-low range measure. *SMB* and *BVMV* are not included as control variables when *SIZE* and *MV/BV* are the dependent variables.

[b] *ANSI* is the monthly ‘aggregate news sentiment index’; Model (1) refers to $ANSI_{t-1}$ with no controls; Model (2) refers to $ANSI_{t-1}$ with a control for *SI*; Model (3) refers to $ANSI_{t-1}$ with controls for *SI*, *RMRF*, *SMB* and *BVMV*; and Model (4) refers to $ANSI_{t-1}$ with controls for *SI*, *RMRF*, *SMB*, *BVMV* and *MTM.*

[c] * indicates significance at the 10 per cent level; and ** indicates significance at the 5 per cent level.