Bulletin of Applied Economics, 2023, 10(2), 127-*140*
https://doi.org/10.47260/bae/1027

Forecasting TAIEX and FITX with Affirmative and Doubtful Investor Sentiments

Tzu-Pu Chang[1]*, Yu-Wei Chan[1] and Ping-Huang Wang[1]

Abstract

The existing literature have used media messages as a proxy variable for investor sentiment, but they mainly classify sentiment into positive or negative categories based on the words used in news articles, without much attention to the degree of affirmative or doubtful conveyed by the words used. Thus, in addition to classifying news content into positive or negative sentiment, this study also measures the degree of affirmative or doubtful expressed in the news articles in order to achieve more accurate predictive results. The study converts qualitative text to two quantitative scores (sentiment ratio and affirmative ratio) and investigates the predictive ability of these two variables on stock index returns and volatility in Taiwan's case. The empirical findings indicate that only affirmative ratio exhibits a significant and negative impact on the one-day ahead returns of the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and the Taiwan Stock Exchange Index Futures (FITX). Additionally, the volatility of returns in both future and spot markets is significantly influenced by both sentiment ratio and affirmative ratio.

JEL Classifications: C22; G10; G40.
Keywords: Investor sentiment; Text-mining; TAIEX; FITX; Affirmative ratio.

[1] Department of Finance, National Yunlin University of Science and Technology, Taiwan.

Article Info: Received: March 15, 2023. Revised: August 14, 2023. Published online: August 22, 2023.

1 Introduction

With the arrival of the age of big data, investors no longer worry about a lack of information; investors must now learn how to identify investment targets and opportunities with high investment yields in a financial market oversaturated with information. Investors equipped with more information are generally expected to make more rational decisions. However, the literature reports that investors are commonly subject to investment behavior bias. For example, investors often exhibit confirmation bias by ignoring crucial information and making decisions based on information that is consistent with their original positions. Recency bias, however, causes investors to place greater importance on the most recent reports or data. Under the influence of herd mentality, investors tend to follow existing trends and strategies without considering all relevant factors and emphasize information that is consistent with their beliefs, the newest media reports, or the actions of others.

Information is the basis for investor decision-making. Investors often make decisions based on the sentiment of news content. Research on investor sentiment must reference publicly available quantitative information and qualitative information from news media. Most literature on market information and investor sentiment categorizes news media into positive and negative news sentiment and considers this categorization to be a direct sentiment index. However, news sentiment comprises positive and negative opinions and implications as to whether the news is authentic and affirmative or doubtful and speculative. Two examples of news sentiment are presented in the following headings:

1. "Fed Must Do 'a Little More' to Tame Inflation, Mester Says"
2. "Fed May Need to Hike to 6.5% to Cool Prices, Study Says"

The first heading presents news as authentic and affirmative, whereas the second heading presents news as doubtful and speculative. Although both headings present the news related to inflation, the headings have different effects on investor sentiment. This study quantifies positive and negative news content and then categorizes the news based on whether the news is affirmative or doubtful for further quantification. Text-mining technology is used to sample financial news from the internet. The samples are then screened and quantified to facilitate the analysis and prediction of future benchmarks.

This study employs financial news as a proxy variable for investor sentiment. Moreover, the study references the research methods of Chen et al. (2021), employs the positive–negative sentiment dictionary established in their study, and compares and quantifies news samples to classify investor sentiment into either positive or negative sentiment.[2] Additionality, this study further classifies news content into either affirmative or doubtful sentiment by using affirmative and doubtful vocabulary lists from the E-HowNet website of Academia Sinica as the foundation.[3] The vocabulary extracted from news samples is then used as a supplement to construct an affirmative–doubtful sentiment dictionary for the financial field. Subsequently, this study establishes and uses a sentiment ratio (positive–negative sentiment ratio) and affirmative ratio (affirmative–doubtful sentiment ratio) as investor sentiment variables to quantify unstructured news content. Finally, empirical analysis is performed, and other control variables are input into the ARMAX–GARCH model to forecast the return rate of the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and Taiwan Stock Exchange Index Futures (FITX) for the next day. More specifically, this paper aims to examine the following hypotheses: (1) whether the sentiment ratio has an impact on the one-day ahead returns of TAIEX and FITX, (2) whether the affirmative ratio influences the one-day ahead returns of TAIEX and FITX, (3) whether the sentiment ratio affects the volatilities of TAIEX and FITX, and (4) whether the affirmative ratio affects the volatilities of TAIEX and FITX.

[2] Unfortunately, there is currently no specialized dictionary for categorizing financial and economic words in Chinese. Therefore, this study used the dictionary provided in Chen et al (2021) for calculating the frequency of positive and negative words.
[3] For detail, please refer to https://ckip.iis.sinica.edu.tw/project/ehownet.

This study collects samples that date from April 2013 to April 2017, accounting for a total of 1,001 days. In particular, data on investor sentiment is obtained from financial news collected online from the *Commercial Times*. Econometric models are used to explore whether sentiment and affirmative ratios influence the return rate and volatility of the TAIEX and FITX. The empirical findings reveal the following: (1) News containing higher positive sentiment has a positive effect on the return rate and volatility of TAIEX for the next day, and the affirmative ratio has a negative effect on the stock price index and FITX for the next day. However, when both variables are used in the regression analysis, the effect of sentiment ratio is nonexistent. (2) Both sentiment ratio and affirmative ratio affect the return volatility of TAIEX and FITX; a higher positive sentiment reduces return volatility, whereas a higher affirmative sentiment increases return volatility. (3) Regardless of the positive and negative sentiments of news content published 1 day before the market opens, if the news contains high affirmative sentiment and high authenticity, the return rate of the TAIEX and FITX for the next day will be negative, and the return volatility of both indices will be high.

The subsequent sections of this study are detailed as follows: Section 2 compiles literature related to investor sentiment and news. Section 3 introduces the research data and model. Section 4 presents the empirical results and performs a robustness analysis. Finally, Section 5 presents the research conclusion and suggestions.

2 Literature review

Investor sentiment reflects investment willingness and market expectations; therefore, whether investor sentiment influences stock prices through trade volume is worthy of discussion. An increasing number of researchers have explored the effects of investor sentiment on the market. Fisher and Statman (2000) proposes that the sentiments of small (individual investors), medium (investment newsletters), and large investors (Wall Street strategists) differ under the same conditions and explores the correlation between changes in the sentiment of the three types of investors and future stock returns. Their results reveal that the changes in sentiment of the three groups of investors did not align; although a significant positive correlation is observed between changes in the sentiment of individual investors and the investment newsletters, no significant correlation is observed between the changes in the sentiment of Wall Street strategists and those of individual investors and investment newsletters.

Baker and Wurgler (2000) explore whether a correlation exists between new equity issues and equity returns in the United States stock market from 1928 to 1997. Their empirical results reveal that equity shares in new issues have a high predictive power for unusually high returns in the stock market. Additionally, equity shares in new issues exhibit high correlation with closed-end fund discounts and other investor sentiment variables. Baker and Stein (2004) report that in a market with short sales constraints, the turnover rate can be considered a proxy variable for measuring investor sentiment. High market liquidity prompts more active sentiment among irrational investors. However, excessively high investor sentiment results in future drops in stock prices. Brown and Cliff (2005) explore the relationship between investor sentiment and asset valuation. The study reveals that overly positive sentiment may result in the market price of stocks exceeding the intrinsic value. Additionally, high investor sentiment in the market is often accompanied by low returns, thereby reversing the market price of stocks and returning it to the actual value.

Studies have used principal component analysis to construct investor sentiment indices. Brown and Cliff (2004) compile numerous variables into four dimensions, namely market performance, transaction activity, derivative financial products, and other sentiment-related variables, and consider these variables to be indirect determinants of sentiment. Baker and Wurgler (2006) incorporate numerous variables, including closed-end fund discount, market turnover rate, number of IPOs, IPO first-day returns, equity shares in new issues, and the advance–decline line (A/D line), to propose an investor sentiment index and explore the effect of firm characteristics on stock returns when investor sentiment is high and low.

Of the studies exploring the effect of news information on stock prices, Liu et al. (1990) collect data on the "Heart on the Street" column between 1982–1985 and research the value of stocks appraised by

professional investors. The researchers reveal that the column affects stock prices on publication day and 2 days prior to publication. The abnormal returns on these days are associated with a higher trading volume. Gillam et al. (2002) compare news time series with financial time series to explore the underlying relationship between the two. The study employs financial, industrial, and political news announcements in the United Kingdom published between 9:00 a.m. and 5:30 p.m. as the sample. The content and tone of the news are compiled into "good word frequency" and "bad word frequency" and used for modeling market sentiment, which is then used to predict the daily closing value of the FTSE 100. The empirical findings reveal a correlation between the news and the stock market.

Tetlock (2007) is the first to construct sentiment indicators by using text-mining. The study performs vector autoregression (VAR) on data compiled from the "Abreast of the Market" column of the Wall Street Journal from 1984–1999 to explore the correlation between media pessimism and stock market performance. The results reveal the following: (1) Media sentiment influences stock market pricing. High levels of media pessimism induce downward pressure on market prices in the stock market, followed by a reversion to fundamentals. (2) Unusually high or low pessimism predicts high market trading volume. (3) Low market returns lead to high media pessimism. Overall, media sentiment influences the investment behavior of investors. Garcia (2013) employs data from 1905–1958 to explore the effect of sentiment on asset prices. The author categorizes positive and negative words into two columns of financial news from the New York Times to propose an investor sentiment indicator. Their empirical results reveal that after controlling for other time-series variables, predicting stock returns by using news content is effective during recessions.

In addition to positive and negative sentiments expressed by media optimism and pessimism, some scholars have proposed text-mining methods based on forecasting. In an empirical study on forward-looking statements (FLS), Li (2010) examines the FLS in the Management Discussion and Analysis sections of 10-K and 10-Q filings from 1994–2007 by using a Naïve Bayesian machine learning algorithm. The study divides FLS into categories based on tone and content and proposes the following empirical findings: (1) firms with better current performance, lower accruals, a smaller size, lower market-to-book ratio, less return volatility, lower Management Discussion and Analysis Fog index scores, and a longer history tend to have more positive FLSs; and (2) after controlling for other determinants of future performance, on average, the FLS is positively associated with future earnings. Loughran and McDonald (2011) propose that positive and negative word lists developed for other disciplines often misclassify common words in financial texts. In addition to creating an alternative negative word list, the study establishes five other word lists based on positive, uncertainty, litigious, strong modal, and weak modal word categories to expand the word classifications. Each of the word lists, namely negative, positive, uncertainty, litigious, strong modal, and weak modal word lists, are respectively used for researching event-period excess returns, event-period abnormal volume, postevent return volatility, fraud, material weakness, and unexpected earnings. The empirical findings reveal that uncertainty words have a negative effect on event-period excessive return.

3 Data and methodology

This study employs the TAIEX and FITX as research targets and collects samples from April 2013 to April 2017, accounting for a total of 1,001 days. The samples encompass data taken during the 2015 stock market crisis, followed by half a year of market shock, and the rise of the new bull market in 2016. The study collects data recorded daily to capture the effect of investor sentiment on the stock market.

The data samples comprise conventional numerical data (i.e., structured data) and textual data (i.e., unstructured data). The dependent variables are the daily returns of TAIEX and FITX. Samples of structured data are taken as independent variables based on market transaction and macroeconomic data, including the futures long–short ratios of qualified foreign institutional investors, investment trust, and security dealers (QFII_FB, IT_FB, and SD_FB), trading volume (Vol), turnover rate (Tor), short selling (Short), margin transaction (Margin), securities lending (SL), the ratio of number of upward to downward stocks (UpDown), buy/sell ratio of qualified foreign institutional investors and investment trusts (QFII_Buy,

IT_Buy), option put/call ratio (Option), and foreign exchange rates (Forex). The numerical data sources include the Taiwan Stock Exchange, Taiwan Futures Exchange, and the Central Bank of Taiwan.

Textual data are collected from the finance news of the *Commercial Times*. Because this study classifies investor sentiment variables by using two different classification criteria (i.e., positive–negative and affirmative–doubtful), two textual classification dictionaries are established. After establishing the positive–negative sentiment dictionary, the comprehensiveness of the definitions provided in the dictionary and whether the definitions are feasible in the financial field must be considered to eliminate bias in the forecasting results. Therefore, this study references and integrates the sentiment dictionaries established by Chen et al. (2021) into the positive–negative sentiment dictionary proposed in this study. Subsequently, this study employs affirmative and doubtful vocabularies provided by the E-HowNet website of Academia Sinica as the foundation and extracts vocabulary from news samples as a supplement to establish the affirmative–doubtful sentiment dictionary with comprehensiveness and feasible definitions.

This study establishes two investor sentiment variables: (1) Sentiment ratio (Sentiment): A proxy variable representing investor sentiment. After tokenizing, comparing, and quantifying the news, the number of positive sentiment words is divided by the number of negative sentiment words to obtain the natural logarithm, which is subjected to first-order difference transformation to produce the rate of change. This variable reflects investors' expectations of the stock market. If the news is more positive, investors tend to have a bullish outlook on the stock market. Contrarily, if the news is more negative, investors tend to have a bearish outlook on the stock market. (2) Affirmative rate ratio (Sure): A variable calculated by dividing the affirmative sentiment score with the doubtful sentiment score to obtain the natural logarithm, which is subjected to first-order difference transformation to produce the rate of change. This variable reflects investors' expectations for the stock market. If the news is more affirmative, investors have a clearer outlook on the stock market. Alternatively, if the news is more doubtful, investors have a more uncertain outlook on the stock market.

A time-series autoregressive moving average–generalized autoregressive conditional heteroskedasticity (ARMAX–GARCH) model is used for the empirical analysis. The mean equation of the return rate considers the influence of autoregression, moving averages, and other independent variables. For the variance equation of the return rate, the GARCH model is used to examine the effect of variance heterogeneity. On the basis of Akaike's Information Criteria and Bayesian Information Criteria, this study recognizes ARMAX(0,0)–GARCH(1,1) as the optimal model and a lagged value of 1 as optimal for the independent variables. By using the TAIEX as an example, the ARMAX–GARCH model is presented as follows:

$$TAIEX_t = \beta_0 + \beta_1 QFII_{FB\,t-1} + \beta_2 IT_{FB\,t-1} + \beta_3 SD_{FB\,t-1} + \beta_4 Vol_{t-1} + \beta_5 Tor_{t-1}$$

$$+\beta_6 Short_{t-1} + \beta_7 Margin_{t-1} + \beta_8 SL_{t-1} + \beta_9 UpDown_{t-1} + \beta_{10} QFII_{Buy\,t-1} + \beta_{11} IT_{Buy\,t-1}$$

$$+\beta_{12} Option_{t-1} + \beta_{13} Forex_{t-1} + \beta_{14} Sentiment_{t-1} + \beta_{15} Sure_{t-1} + \varepsilon_t$$

$$\varepsilon_t = u_t \sqrt{h_t}$$

$$u_t \sim^{i.i.d} (0,1)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} + \alpha_3 Sentiment_{t-1} + \alpha_4 Sure_{t-1}$$

4 Empirical results

4.1 Descriptive statistics

First, descriptive statistics and correlation coefficients are used to describe the research data. Table 1 presents the descriptive statistics of each variable. According to the table, the means of all variables are approximately zero. The coefficients of skewness indicate that the data distributions of all variables are near symmetrical aside from Short, which is left skewed. Furthermore, Short has the largest coefficient of kurtosis of the variables. The two investor sentiment variables, Sentiment and Sure, have means close to zero and symmetrical, normal data distributions.

Table 1: Descriptive statistics

Variable	Samples	Mean	SD	Minimum	Median	Maximum	Skewness	Kurtosis
TAIEX	1001	0.0002	0.0081	-0.0496	0.0005	0.0352	-0.4264	2.8420
FITX	1001	0.0002	0.0088	-0.0529	0.0005	0.0444	-0.4517	3.4951
QFII_FB	1001	0.0009	0.1458	-0.7801	-0.0032	0.7746	0.2849	3.3223
IT_FB	1001	-0.0031	0.1451	-0.9554	-0.0033	0.8230	-0.0957	9.9104
SI_FB	1001	0.0003	0.2589	-1.3758	0.0195	1.0273	-0.5809	2.4595
VOL	1001	0.0006	0.1727	-0.7634	-0.0062	0.7439	0.2721	1.9772
TOR	1001	-0.0004	0.3128	-1.1835	-0.0039	1.0770	-0.0502	0.5470
Margin	1001	-0.0004	0.0348	-0.2404	0.0019	0.1526	-1.2651	7.9105
Short	1001	-0.0002	0.0057	-0.0936	0.0005	0.0185	-5.5796	77.7922
SL	1001	0.0005	0.0053	-0.0295	0.0019	0.0131	-1.2041	7.8694
Updown	1001	-0.0008	1.1560	-3.7992	0.0013	5.5476	-1.1771	3.7586
QFII_Buy	1001	-0.0002	0.2833	-1.6434	-0.0723	1.4476	0.4238	1.2034
IT_Buy	1001	0.0007	0.3644	-1.1438	-0.0087	1.3718	0.0077	2.4393
Option	1001	0.0002	0.0733	-0.4779	0.0009	0.2884	0.1393	0.2074
Forex	1001	0.00001	0.0026	-0.0118	-0.0039	0.0128	-0.0502	0.5470
Sentiment	1001	0.0001	0.1317	-0.5017	0.0023	0.6386	-0.7069	4.4590
Sure	1001	-0.0002	0.1475	-0.8528	0.0001	0.6093	0.0060	3.2052

Note: TAIEX and FITX denote the daily returns of the Taiwan Stock Exchange Capitalization Weighted Stock Index and the Taiwan Stock Exchange Index Futures, respectively. QFII_FB, IT_FB, and SI_FB denote the long–short ratios of futures of qualified foreign institutional investors, investment trusts, and security dealers, respectively. VOL and TOR denote the trading volume and turnover rate of the spot market, respectively. Margin, Short, and SL denote the margin transaction, short selling, and securities lending transaction of the spot market, respectively. Updown denotes the ratio of number of upward to downward stocks. QFII_Buy and IT_Buy denote the buy/sell ratio of qualified foreign institutional investors and investment trusts, respectively. Option denotes the put/call ratio of options market. Forex denotes the exchange rate of New Taiwan Dollar to US Dollar. Sentiment is the sentiment ratio that denotes the ratio of the number of positive sentiment words to the number of negative sentiment words. Sure, is the affirmative ratio that denotes the ratio of the number of affirmative sentiment words to the number of doubtful sentiment words.

Table 2 presents the correlation coefficient matrix. The table indicates that the correlation coefficients of Sentiment and Sure with both TAIEX and FITX are nonsignificant and significant, respectively. Higher return rates are observed when the news contains higher affirmative sentiment. Additionally, a negative correlation exists between Sentiment and Sure. Finally, an overview of the other control variables reveals that aside from SL and Vol, the correlation coefficients of all variables with TAIEX and FITX are significant. In particular, the correlation coefficients of Options with TAIEX and FITX exceed 0.6; therefore, Options has the highest level of correlation with the dependent variables.

Table 2: Correlation coefficient matrix

	TAIEX	FITX	QFII_FB	IT_FB	SI_FB	VOL	TOR	Short	Margin	SL	Updown	QFII_Buy	IT_Buy	Option	Forex	Sentiment	Sure
TAIEX	—																
FITX	0.939*	—															
QFII_FB	0.374*	0.405*	—														
IT_FB	0.401*	0.389*	0.206*	—													
SI_FB	0.152*	0.111*	-0.407*	0.040	—												
VOL	-0.004	-0.018	-0.033	0.041	-0.019	—											
TOR	0.389*	0.399*	0.288*	0.187*	0.039	-0.066*	—										
Short	0.298*	0.259*	0.157*	0.197*	0.035	-0.059	0.094*	—									
Margin	0.075*	0.068*	-0.136*	-0.024	0.107*	-0.171*	-0.076*	0.094*	—								
SL	-0.059	-0.062	-0.053	0.031	0.046	0.058	0.010	0.049	-0.068*	—							
Updown	0.606*	0.593*	0.337*	0.248*	0.097*	-0.098*	0.605*	0.210*	-0.044	0.020	—						
QFII_Buy	0.550*	0.561*	0.420*	0.208*	-0.065*	0.080*	0.332*	0.172*	-0.251*	-0.033	0.479*	—					
IT_Buy	0.316*	0.274*	0.125*	0.148*	0.081*	0.095*	0.211*	0.131*	-0.081*	0.017	0.327*	0.215*	—				
Option	0.616*	0.602*	0.243*	0.293*	0.168*	-0.025	0.263*	-0.190*	-0.001	0.012	0.387*	0.411*	0.197*	—			
Forex	-0.366*	-0.354*	-0.126*	-0.143*	-0.025	-0.071*	-0.110*	-0.098*	-0.017	0.046	-0.206*	-0.212*	-0.135*	-0.273*	—		
Sentiment	0.053	0.044	0.006	0.067*	0.047	-0.021	-0.084*	0.066*	0.051	0.029	-0.140*	-0.101*	0.054	0.091*	-0.058	—	
Sure	0.067*	0.074*	0.006	0.025	0.024	0.039	-0.059	-0.006	0.018	-0.029	0.049	0.006	-0.015	0.079*	-0.037	-0.168*	—

Note: * denotes the significant levels at 0.05. TAIEX and FITX denote the daily returns of the Taiwan Stock Exchange Capitalization Weighted Stock Index and the Taiwan Stock Exchange Index Futures, respectively. QFII_FB, IT_FB, and SI_FB denote the long–short ratios of futures of qualified foreign institutional investors, investment trusts, and security dealers, respectively. VOL and TOR denote the trading volume and turnover rate of the spot market, respectively. Margin, Short, and SL denote the margin transaction, short selling, and securities lending transaction of the spot market, respectively. Updown denotes the ratio of number of upward to downward stocks. QFII_Buy and IT_Buy denote the buy/sell ratio of qualified foreign institutional investors and investment trusts, respectively. Option denotes the put/call ratio of options market. Forex denotes the exchange rate of New Taiwan Dollar to US Dollar. Sentiment is the sentiment ratio that denotes the ratio of the number of positive sentiment words to the number of negative sentiment words. Sure is the affirmative ratio that denotes the ratio of the number of affirmative sentiment words to the number of doubtful sentiment words.

4.2 Regression analysis

Table 3 lists the regression results of the TAIEX obtained using the ARIMAX–GARCH model. Model (1) only contains the control variables and does not consider the two sentiment variables. The results reveal that the put/call option ratio of foreign investors in the futures market significantly influences the return rate of TAIEX for the next day. Additionally, a higher futures long–short ratio predicts a higher return rate of TAIEX for the next day. The turnover rate has a negative relationship with the return rate. The general assumption is that the market turnover rate reflects investor sentiment; a higher investor sentiment predicts a lower return rate. Finally, the Forex rate has a significant negative influence on the return rate for the next day, suggesting that an increase in Forex rate predicts a drop in stock prices.

Table 3: Regression analysis of TAIEX using ARIMAX–GARCH

Mean Equation	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	0.0004 (0.0003)	0.0004 (0.0003)	0.0004 (0.0003)	0.0004* (0.0002)
QFII_FB	0.0060*** (0.0022)	0.0062*** (0.0022)	0.0060*** (0.0022)	0.0063*** (0.0021)
IT_FB	-0.0008 (0.0015)	-0.0009 (0.0015)	-0.0008 (0.0015)	-0.0007 (0.0015)
SI_FB	-0.0017 (0.0011)	-0.0014 (0.0011)	-0.0017 (0.0011)	-0.0014 (0.0011)
VOL	-0.0005 (0.0015)	-0.0003 (0.0015)	-0.0002 (0.0015)	-0.0001 (0.0014)
TOR	-0.0021* (0.0011)	-0.0021** (0.0011)	-0.0024** (0.0010)	-0.0023** (0.001)
Short	-0.0066 (0.0081)	-0.0093 (0.0076)	-0.0066 (0.0079)	-0.0082 (0.0075)
Margin	-0.0662 (0.0628)	-0.0826 (0.0608)	-0.0633 (0.0625)	-0.0836 (0.0605)
SL	-0.0499 (0.0463)	-0.0648 (0.0431)	-0.0522 (0.0462)	-0.0669 (0.0435)
Updown	-0.0001 (0.0004)	0.0002 (0.0004)	0.0001 (0.0004)	0.0002 (0.0004)
QFII_Buy	0.0013 (0.0012)	0.0012 (0.0012)	0.0011 (0.0011)	0.0008 (0.0012)
IT_Buy	0.0011 (0.0007)	0.0010 (0.0007)	0.0011 (0.0007)	0.0010 (0.0007)
Option	0.0022 (0.0040)	0.0020 (0.0040)	0.0034 (0.0041)	0.0032 (0.0041)
Forex	-0.2145* (0.1111)	-0.1830* (0.1018)	-0.2075* (0.1101)	-0.1779* (0.0981)
Sentiment		0.0031 (0.0020)		0.0021 (0.0020)
Sure			-0.0049*** (0.0016)	-0.0046*** (0.0016)
Variance Equation Intercept	0.0001 (0.0001)	-12.6636*** (0.4754)	-12.9392*** (0.7341)	-12.807*** (0.4190)
ARCH	0.0639** (0.0291)	0.07036*** (0.0261)	0.0639*** (0.0286)	0.0643*** (0.0243)
GARCH	0.8553*** (0.0515)	0.8475*** (0.0476)	0.8968*** (0.0506)	0.8531*** (0.0398)
Sentiment		-7.1670*** (1.1585)		-7.1564*** (1.1669)
Sure			-0.2624 (1.7991)	2.4333* (1.4355)
Log likelihood	3455.428	3465.365	3460.193	3471.156
MAE	0.005829	0.005832	0.005822	0.005818
RMSE	0.00795	0.00794	0.00792	0.00790

Note: *, **, and *** denote the significant levels at 0.1, 0.05, and 0.01, respectively. Robust standard errors are in parentheses. TAIEX and FITX denote the daily returns of the Taiwan Stock Exchange Capitalization Weighted Stock Index and the Taiwan Stock Exchange Index Futures, respectively. QFII_FB, IT_FB, and SI_FB denote the long–short ratios of futures of qualified foreign institutional investors, investment trusts, and security dealers, respectively. VOL and TOR denote the trading volume and turnover rate of the spot market, respectively. Margin, Short, and SL denote the margin transaction, short selling, and securities lending transaction of the spot market, respectively. Updown denotes the ratio of number of upward to downward stocks. QFII_Buy and IT_Buy denote the buy/sell ratio of qualified foreign institutional investors and investment trusts, respectively. Option denotes the put/call ratio of options market. Forex denotes the exchange rate of New Taiwan Dollar to US Dollar. Sentiment is the sentiment ratio that denotes the ratio of the number of positive sentiment words to the number of negative sentiment words. Sure is the affirmative ratio that denotes the ratio of the number of affirmative sentiment words to the number of doubtful sentiment words.

Model (2) adds the sentiment variable into Model (1). The regression results reveal that although the effect of sentiment ratio on the return rate of TAIEX is not significant, the regression coefficients of sentiment ratio with all other variables are positive. Thus, news containing more positive and optimistic sentiments predicts a higher return rate for the next day, which is consistent with the literature and reveals that positive investor sentiment predicts short-term increases in stock prices. The effect of sentiment ratio on volatility is significant, which implies that optimistic sentiment predicts lower stock price volatility for the next day. The results align with the findings of Kumari and Mahakud (2015) and Audrino et al. (2020), who report that sentiment can be effectively used to predict future price volatility.

Model (3) adds the affirmative variable into Model (2). The regression results indicate that a higher affirmative ratio (i.e., news containing more affirmative and authentic sentiments) predicts a lower stock return rate, which is consistent with the findings of Loughran and McDonald (2011) and suggests that uncertainty increases investors' expectations, thereby indirectly influencing stock prices. In Model (4), the sentiment ratio and affirmative ratio are separately added to the mean equation and the variance equation of the return rate. The mean equation reveals that the effect of the sentiment ratio on the return rate is not significant, which implies that the affirmative ratio has greater forecasting power on the return rate of TAIEX. Table 3 provides two indices for assessing predictability, namely mean absolute error (MAE) and root mean square error (RMSE). The indices also suggest that adding the affirmative ratio into the ARMAX–GARCH model can increase the predictability accuracy for the return rates of TAIEX.

FITX serves as the dependent variable for the four sets of regression equations in Table 4. The results of Model (1), which does not consider the sentiment ratio and affirmative ratio, are similar to the results presented in Table 3, which implies that: (1) a higher put/call option ratio of foreign investors in the futures market significantly influences the return rate of FITX for the next day; and (2) a higher turnover rate results in a higher return rate of FITX for the next day. However, some differences from the results of Table 3 are also observed; for example, the put/call option ratio of investment trusts exhibits a significant positive effect on the return rate of TAIEX. Therefore, a high put/call option ratio of foreign investors in the spot market results in higher return rates of FITX for the next day. Overall, the leverage of foreign investors and investment trusts dictates the return rate.

Table 4: Regression analysis of FITX using ARIMAX–GARCH

Mean Equation	(1)	(2)	(3)	(4)
Intercept	0.0003 (0.0003)	0.0003 (0.0002)	0.0003 (0.0003)	0.0004 (0.0003)
QFII_FB	0.0058** (0.0025)	0.0056** (0.0025)	0.0059** (0.0025)	0.0057** (0.0024)
IT_FB	-0.0027 (0.0017)	-0.0026 (0.0017)	-0.0027 (0.0017)	-0.0021 (0.0016)
SI_FB	-0.0006 (0.0013)	-0.0004 (0.0012)	-0.0006 (0.0013)	-0.0005 (0.0012)
VOL	-0.0005 (0.0016)	-0.0005 (0.0016)	-0.0004 (0.0016)	-0.0003 (0.0015)
TOR	-0.0024* (0.0012)	-0.0025** (0.0011)	-0.0028** (0.0014)	-0.0027** (0.0011)
Short	-0.0051 (0.0080)	-0.0075 (0.0075)	-0.0059 (0.0080)	-0.0057 (0.0074)
Margin	-0.0650 (0.0675)	-0.0857 (0.0649)	-0.0649 (0.0675)	-0.0909 (0.0653)
SL	-0.0478 (0.0513)	-0.0635 (0.0471)	-0.0502 (0.0513)	-0.0638 (0.0469)
Updown	0.0000 (0.0004)	0.0002 (0.0004)	0.0001 (0.0004)	0.0002 (0.0004)
QFII_Buy	0.0009 (0.0012)	0.0007 (0.0011)	0.0006 (0.0012)	0.0003 (0.0011)
IT_Buy	0.0016** (0.0008)	0.0015** (0.0007)	0.0017** (0.0008)	0.0015** (0.0007)
Option	0.0052 (0.0054)	0.0045 (0.0054)	0.0060 (0.0054)	0.0049 (0.0054)
Forex	-0.1677 (0.1209)	-0.1460 (0.1107)	-0.1618 (0.1233)	-0.1427 (0.1045)
Sentiment		0.0025 (0.0021)		0.0018 (0.0021)
Sure			-0.0052*** (0.0018)	-0.0051*** (0.0019)
Variance Equation				
Intercept	0.0000 (0.0000)	-12.6202*** (0.5320)	-12.9669*** (1.5599)	-12.8628*** (0.4865)
ARCH	0.0635 (0.0396)	0.07187** (0.0286)	0.0584 (0.0488)	0.0541** (0.0267)
GARCH	0.8949*** (0.0741)	0.8565*** (0.0485)	0.8984*** (0.0998)	0.8632*** (0.0404)
Sentiment		-7.1223*** (1.3479)		-6.4964*** (1.3978)
Sure			5.0003 (3.2317)	5.1535** (2.4212)
Log likelihood	3375.445	3384.309	3380.231	3393.151
MAE	0.006156	0.006161	0.006130	0.006128
RMSE	0.00862	0.00861	0.00859	0.00858

Note: *, **, and *** denote the significant levels at 0.1, 0.05, and 0.01, respectively. Robust standard errors are in parentheses. TAIEX and FITX denote the daily returns of the Taiwan Stock Exchange Capitalization Weighted Stock Index and the Taiwan Stock Exchange Index Futures, respectively. QFII_FB, IT_FB, and SI_FB denote the long–short ratios of futures of qualified foreign institutional investors, investment trusts, and security dealers, respectively. VOL and TOR denote the trading volume and turnover rate of the spot market, respectively. Margin, Short, and SL denote the margin transaction, short selling, and securities lending transaction of the spot market, respectively. Updown denotes the ratio of number of upward to downward stocks. QFII_Buy and IT_Buy denote the buy/sell ratio of qualified foreign institutional investors and investment trusts, respectively. Option denotes the put/call ratio of options market. Forex denotes the exchange rate of New Taiwan Dollar to US Dollar. Sentiment is the sentiment ratio that denotes the ratio of the number of positive sentiment words to the number of negative sentiment words. Sure is the affirmative ratio that denotes the ratio of the number of affirmative sentiment words to the number of doubtful sentiment words.

Table 2 indicates that a positive but insignificant relationship exists between the sentiment ratio and the return rate for the next day. The regression coefficient of the sentiment ratio with volatility is negative, which indicates that news containing positive sentiment published 1 day before the market opens predicts lower volatility of FITX on the day the market opens. This finding is consistent with the results listed in Table 3. Model (3) calculates the effect of the affirmative ratio on the futures return rate for the next day. The results reveal that the affirmative ratio has a negative effect on the return rate of TAIEX, which is consistent with the findings of Loughran and McDonald (2011) that a low affirmative ratio denotes greater uncertainty, which in turn increases investors' expectations and further increases stock prices. Finally, Model (4) presents the regression coefficients of all variables.

The results are consistent with those listed in Table 3. In the mean equation, the forecasting power of the affirmative ratio on the return rate of TAIEX is greater than that of the sentiment ratio. In the variance equation, a higher sentiment ratio predicts a lower return volatility, and a higher affirmative ratio predicts a higher return volatility. Finally, adding the affirmative ratio to the ARMAX–GARCH model can increase the predictability accuracy for return rates, as measured using MAE and RMSE.

4.3 Robustness

To reinforce the empirical results, this section explores whether the affirmative ratio has an asymmetric effect on the return rates of TAIEX and FITX. Moreover, this section introduces two variables, namely the optimistic affirmative ratio and pessimistic affirmative ratio, which are established using two dummy variables, D1 and D2. D1 is used to distinguish the daily optimistic sentiment. If the sentiment ratio is greater than 1, then D1 equals 1; otherwise, D1 equals 0. D2 is used to distinguish the daily pessimistic sentiment. If the sentiment ratio is less than 1, D2 equals 1; otherwise, D2 equals 0. D1 and D2 are multiplied by the affirmative ratio to obtain the optimistic affirmative ratio and pessimistic affirmative ratio, respectively. The main purpose of introducing these two variables is to determine (1) whether the affirmative ratio exhibits an asymmetric effect and (2) the effects of the optimistic affirmative ratio and the pessimistic affirmative ratio on the return rate and volatility of the dependent variables. In this section, the optimistic affirmative ratio and the pessimistic affirmative ratio are separately added to the ARIMAX–GARCH model to predict the return rates of TAIEX and FITX. When selecting independent variables for the model, collinearity problems may occur if the affirmative ratio is added to the model alongside the optimistic affirmative ratio and pessimistic affirmative ratio. Therefore, the discussion on investor sentiment variables in this section does not consider the affirmative ratio.

TAIEX and FITX serve as the dependent variables for the two sets of regression equations in Table 5. The effects of the sentiment ratio and optimistic affirmative ratio on both dependent variables are not significant. However, the pessimistic affirmative ratio has a significant negative effect on both dependent variables, and its effect on FITX is stronger than that on TAIEX. Therefore, the affirmative ratio exhibits an asymmetrical effect, which implies that if news containing pessimistic affirmative sentiment is published 1 day before the market opens, the drop in the return volatility of FITX may surpass that of the market. Pessimistic affirmative news passes through numerous personnel before publication, causes early stock price responses, and induces panic among investors after its publication, which produces immense pressure to sell. Consequently, the return rates of TAIEX and FITX 1 day after the news is published is predicted to be negative.

This result is consistent with the findings of Loughran and McDonald (2011), which reveals that uncertainty words have a significant negative correlation with unusually high returns. The findings also align with the results of Tetlock (2007), who reports that media sentiment influences stock market trends, pessimistic news induces downward pressure on market prices in the stock market, and that futures products have greater leverage than spot products. Thus, the influence of news on the market is relatively greater than spot products.

Table 5: Asymmetrical effect of affirmative ratio

	TAIEX	FITX
Mean Equation		
Intercept	0.0004 (0.0002)	0.0004 (0.0003)
Sentiment ratio	0.0021 (0.0021)	0.0018 (0.0024)
Optimistic affirmative ratio	-0.0026 (0.0024)	-0.0028 (0.0026)
Pessimistic affirmative ratio	-0.0067** (0.0025)	-0.0076** (0.0027)
Control Variables	Yes	Yes
Variance Equation		
Intercept	-12.7812*** (0.3164)	-12.8522*** (0.2733)
Sentiment ratio	-7.0932*** (1.0333)	-6.5874*** (0.8642)
Optimistic affirmative ratio	1.0771 (5.9183)	5.6332** (1.8904)
Pessimistic affirmative ratio	2.3343 (1.4632)	5.0510*** (1.1023)
ARCH	0.0633*** (0.0132)	0.0545*** (0.0121)
GARCH	0.8532*** (0.0257)	0.8611*** (0.0229)
Log likelihood	3472.033	3393.01
MAE	0.005812	0.006123
RMSE	0.00789	0.00856

Note: *, **, and *** denote the significant levels at 0.1, 0.05, and 0.01, respectively. Robust standard errors are in parentheses. TAIEX and FITX denote the daily returns of the Taiwan Stock Exchange Capitalization Weighted Stock Index and the Taiwan Stock Exchange Index Futures, respectively. Sentiment is the sentiment ratio that denotes the ratio of the number of positive sentiment words to the number of negative sentiment words. Optimistic affirmative ratio denotes the affirmative ratio when sentiment ratio is larger than unity; Pessimistic affirmative ratio denote the affirmative ratio when sentiment ratio is less than unity.

In addition to the ARCH and GARCH variables, this study added the sentiment ratio, optimistic affirmative ratio, and pessimistic affirmative ratio variables to the ARIMAX–GARCH model to determine whether the optimistic affirmative ratio and pessimistic affirmative ratio variables directly influence the return rate and volatility of the TAIEX and FITX. The results in Table 5 indicate that the effects of the optimistic affirmative ratio and pessimistic affirmative ratio on the return volatility of TAIEX are not significant; however, the effects of all other variables on the return volatility of TAIEX and FITX are significant. The regression coefficients of the sentiment ratio with TAIEX and FITX are negative, which is consistent with the previous empirical findings of this study. Publishing news containing positive sentiment 1 day before the market opens predicts low return volatility of FITX, which indirectly implies that negative news can more easily induce volatility in investor sentiment than positive news, which in turn drives price volatility. The regression coefficients of the optimistic affirmative ratio and pessimistic affirmative ratio with FITX are both positive, but the optimistic affirmative ratio is higher. This finding indicates that the effect of the affirmative ratio is asymmetric. Thus, publishing news containing affirmative sentiment 1 day before the market opens, regardless of whether the news is optimistic or pessimistic, predicts a high return volatility of FITX for the day the market opens. Optimistic affirmative news induces a higher return volatility than pessimistic affirmative news; optimistic affirmative news generates greater volatility in investor sentiment than pessimistic affirmative news, which in turn drives price volatility. A comparison between the correlation coefficients of GARCH for the two regression equations reveals that the sentiment ratio has a greater effect on the return volatility of TAIEX than on the return volatility of FITX. Additionally, only the effects of the optimistic affirmative ratio and pessimistic affirmative ratio on the return volatility of FITX are significant. In other words, the responses of FITX investors are more influenced by the affirmative sentiment in the news than by those of TAIEX investors.

To sum up, this paper finds that affirmative ratio is a better predictor than sentiment ratio in terms of the returns of TAIEX and FITX. This finding is in accordance with the work of Loughran and McDonald (2011), which concludes that investors are more sensitive to uncertainty. Moreover, this paper shows that both sentiment ratio and affirmative ratio can predict the volatilities of TAIEX and FITX well, that is

consistent with Kumari and Mahakud (2015) and Audrino et al. (2020). Finally, this study presents the asymmetric effects of affirmative ratio on the returns of TAIEX and FITX, implying that investors tend to sell their holdings when bad news are confirmed. This asymmetric effect is less studied in literature and we address this issue for future researchers.

5 Conclusions

This study employs financial news as a proxy variable for investor sentiment, references the research methods of Chen et al. (2021), uses positive–negative sentiment dictionaries established in the study, and compares and quantifies news samples to classify investor sentiment into positive and negative sentiment. Data samples from April 2013 to April 2017 are collected, accounting for a total of 1,001 days. The independent variables consist of four structured variables, namely the TAIEX, FITX, Taiwan Weighted Stock Index, and the overall economic performance, and an unstructured variable, namely investor sentiment, obtained from financial news collected online from the *Commercial Times*. Therefore, four multiple regression equations are established, and numerous time-series models are employed for forecasting the two dependent variables, namely TAIEX and FITX.

The empirical results reveal that adding the two investor sentiment variables to the forecasting model significantly improves the ability of the model to forecast the return rate and volatility of TAIEX and FITX for the next day. Additionally, the sentiment ratio only exhibits a significant positive effect on the return rate of TAIEX for the next day. However, if the affirmative ratio is also present in the forecasting model, the effect of the sentiment ratio on the return rates of TAIEX and FITX becomes significantly limited and nonsignificant. The results of the GARCH model reveal that the sentiment ratio exhibits negative effects on the return volatility of both TAIEX and FITX for the next day, and the effect on the return volatility of TAIEX exceeds that of the effect on the return volatility of FITX. In other words, publishing news containing positive sentiment 1 day before the market opens predicts a high return rate of FITX for the next day and relatively low volatility in market and futures prices. This phenomenon reflects how positive news causes investors to refrain from selling stocks and how negative news induces fear among investors, causing them to sell stocks.

This study proposes the following findings on the affirmative ratio. First, the affirmative ratio negatively affects the return rates of TAIEX and FITX for the next day, which is consistent with that of Loughran and McDonald (2011). In particular, the effect of the affirmative ratio on the return rate of FITX is greater than that on the return rate of TAIEX; the weights of the affirmative ratio on TAIEX and FITX are moderate and large, respectively. Subsequently, the results of the GARCH model reveal that the affirmative ratio positively affects the return volatility of both TAIEX and FITX for the next day, and the effect on the return volatility of FITX exceeds the effect on the return volatility of TAIEX. In other words, publishing news containing affirmative sentiment (i.e., news with higher authenticity), regardless of positive or negative sentiment, predicts negative return rates and a high return volatility of TAIEX and FITX for the next day. This finding may be attributed to the premature spread of affirmative news among insiders prior to its publication, which allows insiders to leverage accordingly before or after the market opens. Consequently, insiders can respond to news before general investors, which results in negative return rates and high volatility in market prices. Conversely, when encountering relatively doubtful news (i.e., news with low authenticity), most investors choose to refrain from selling stocks, resulting in low return and high volatility for Taiwan's spot and futures markets.

According to the conclusion, investors in both Taiwan's spot and futures markets are highly sensitive to affirmative and doubtful news. Hence, we suggest that policymakers and competent authority should put more efforts into confirming whether reports' contents are affirmative or not. Additionally, policymakers and competent authority can immediately request related companies to clarify any doubtful news in order to protect investors from news manipulation.

References

Audrino, F., Sigrist, F. and Ballinari, D. (2020). The Impact of Sentiment and Attention Measures on Stock Market Volatility. *International Journal of Forecasting*, 36, 334-357.

Baker, M. and Stein, J.C. (2004). Market Liquidity as a Sentiment Indicator. *Journal of Financial Markets*, 7, 271-299.

Baker, M. and Wurgler, J. (2000). The Equity Share in New Issues and Aggregate Stock Returns. *Journal of Finance*, 55, 2219-2257.

Baker, M. and Wurgler, J. (2006). Investor Sentiment and the Cross-section of Stock Returns. *Journal of Finance*, 61, 1645-1680.

Brown, G.W. and Cliff, M.T. (2004). Investor Sentiment and the Near-term Stock Market. *Journal of Empirical Finance*, 11, 1-27.

Brown, G.W. and Cliff, M.T. (2005). Investor Sentiment and Asset Valuation. *Journal of Business*, 78, 405-440.

Chen, K.C., Lin, C.I. and Chen, H.M. (2021). Relationship between News Sentiment Indicator and the Taiwan Weighted Stock Index. *Journal of Social Sciences and Philosophy*, 33, 383-423. (in Chinese)

Fisher, K.L. and Statman, M. (2000). Investor Sentiment and Stock Returns. *Financial Analysts Journal*, 56, 16-23.

Garcia, D. (2013). Sentiment during Recessions. *Journal of Finance*, 68, 1267-1300.

Gillam, L., Ahmad, K., Ahmad, S., Casey, M., Cheng, D., Taskaya, T., de Oliveira, P.C.F. and Manomaisupat, P. (2002). Economic News and Stock Market Correlation: A Study of the UK Market. Workshop at the International Conference on Terminology and Knowledge Engineering.

Kumari, J. and Mahakud, J. (2015). Does Investor Sentiment Predict the Asset Vvolatility? Evidence from Emerging Stock Market India. *Journal of Behavioral and Experimental Finance*, 8, 25-39.

Li, F. (2010). The Information Content of Forward-looking Statements in Corporate Filings - A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research*, 48, 1049-1102.

Liu, P., Smith, S.D. and Syed, A.A. (1990). Stock Price Reactions to the Wall Street Journal's Securities Recommendations. *Journal of Financial and Quantitative Analysis*, 25, 399-410.

Loughran, T. and McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66, 35-65.

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