THAMMASAT REVIEW
PRINT ISSN: 0859-5747 ONLINE ISSN: 2630-0303

Volume 24 No 1 (January-June) 2021 [Page 317-333]

# Can Sentiment from News Headlines Explain Stock Market Returns?: Evidence from Thailand

Sapphasak Chatchawan *

*Faculty of Commerce and Accountancy, Thammasat University, Thailand*

Received 18 November 2020; Received in revised form 3 May 2021
Accepted 5 May 2021; Available online 30 June 2021

## Abstract

The objective of this study was to investigate whether sentiment from financial market headline news explains equity returns. Techniques from computational linguistics were employed to extract the news-based sentiment from a corpus of financial market headlines collected from a newsfeed of a financial newswire.

In this study, news sentiment was classified as the overall sentiment and included both the positive and the negative sentiment. Using daily financial market data from 2017-2019, the overall sentiment and the positive sentiment were found to explain the equity returns of the Stock Exchange of Thailand (SET) while the negative sentiment did not explain the returns.

## Keywords

News-based sentiment index, Stock markets, Textual analysis

*** Corresponding author**: sapphasak@tbs.tu.ac.th **DOI**: 10.14456/tureview.2021.14

## Introduction

Leading brokerage firms enjoy the benefit of real-time financial market news from international newswires, such as Eikon and Bloomberg. These newswires electronically transmit financial news to their customers and help financial market participants monitor real-time market conditions. Newswires spread up-to-the-minute financial news and, hence, they have naturally created a voluminous amount of textual data.

This paper asks the following empirical question: to what extent does the sentiment of financial market headline news explain stock market returns in Thailand? A growing body of literature empirically explores the relationship between financial market news and stock market movements. The sentiment extracted from financial newspapers, such as *The Wall Street Journal*, *The New York Times* and *The Guardian*, has some predictive power over stock market movements and corporate earnings (Tetlock, 2007; Tetlock et al., 2008). The content of news influences investor decisions. During economic recessions, such as the Great Depression, the unfavorable tone of the financial news had more pronounced effects on stock market prices and stronger predictive power than the positive tone (Garcia, 2013). In addition, there exists a number of studies that have explored the effects of media newswires on financial markets (Engelberg & Parsons, 2011; Marty et al., 2020). The sentiment constructed from the media newswires produces more accurate forecasts of stock market movements than the use of macroeconomic variables (Uhl ,2014). The different frequency range of market sentiment indices also yields different results. That is, a daily market sentiment index gives an accurate prediction for a very short period of time. On the contrary, a weekly sentiment makes a better prediction of stock market returns in a longer period. Moreover, in terms of volatility forecasting, the use of the sentiment index provides a better forecast than the use of stock market prices (Allen et al., 2015; Atkins et al.,2018; Heston & Sinha, 2017). In addition to stock markets, news sentiment can be used to make a prediction and to explain the term structure of interest rates in bond markets (Gotthelf & Uhl, 2018). The sentiment from news also matters in the foreign exchange market (Uhl, 2017).

Instead of making use of financial market news stories like the pertinent literature, this study analyzes whether the financial market headline news in the newsfeed of Thomson Reuters Eikon can explain stock market returns in the Stock Exchange of Thailand (Hereafter, SET) from 2017-2019. By leveraging the techniques from Natural Language Processing (NLP), the news-based sentiment index is extracted from a collection of more than 1,000 headlines that are related to SET and, thus, the whole corpus consists of more than 20,000 words. The news sentiment from the newsfeed may provide timely information for financial market watchers and may influence their decisions. The study finds that the overall market sentiment and the positive market sentiment can explain stock market returns calculated from the SET, SET50 and SET100. However, there is no evidence that the negative sentiment can explain stock market returns between 2017-2019.

## Literature Review

This paper contributes to the existing literature which is pertinent to the study of market sentiment in two aspects. First, the study employs the user-defined dictionary method in Loughran and McDonald (2011) to characterize news sentiment into positive and negative sentiment. The technique is automated and replicable. Unlike Uhl (2014), Uhl (2017), and Gotthelf and Uhl (2018) that use Reuters sentiment index, the news sentiment of SET in this study is constructed from a collection of stock market headline news, which is linked to SET, in Thomson Reuters Eikon's newsfeed. Thus, the approach used in this paper is relatively flexible compared to Reuters sentiment index, which is currently no longer available. Second, the paper studies the effects of positive and negative news sentiment on stock market returns. The study also incorporates the results from the overall market sentiment on market returns.

## Research Methodology

### Sources and Data Description

All financial data are from Eikon. The sample spans from 9/11/2017 to 8/02/2019, which is the longest trading period in which financial market headlines related to SET are available on the newswires. All headlines are compiled by Routers Instrument Code (RIC): "BKstmad.BKW". There are 308 observations of both financial and news-based sentiment variables.

### Dependent Variables

All dependent variables are the percentage change of stock market indices and can be computed by equation (1), which is the continuously compounded return formula.

$$\mathrm{RET}_t = \ln\left(\frac{X_t}{X_{t-1}}\right) \times 100 \quad (1)$$

$X_t$ represents SET, SET50 and SET100 accordingly. In Table 1, SET_RET, SET50_RET, and SET100_RET represent calculated returns from SET, SET50 and SET100 index. In Table 2, descriptive statistics for dependent variables are provided.

**Table 1** shows dependent variables.

| Variable | Description |
|---|---|
| SET_RET | SET returns |
| SET50_RET | SET50 returns |
| SET100_RET | SET100 returns |

**Table 2** shows descriptive statistics of dependent variables.

| | SET_RET | SET50_RET | SET100_RET |
|---|---|---|---|
| Mean | -0.01 | 0.00 | 0.00 |
| Median | 0.05 | 0.06 | 0.06 |
| Maximum | 2.27 | 2.81 | 2.71 |
| Minimum | -2.42 | -2.60 | -2.65 |
| Std. Dev. | 0.71 | 0.80 | 0.80 |
| Skewness | -0.32 | -0.17 | -0.22 |
| Kurtosis | 4.12 | 4.31 | 4.26 |
| Jarque-Bera | 21.51 | 23.53 | 23.07 |
| Probability | 0.00 | 0.00 | 0.00 |
| Sum | -3.74 | 1.14 | -1.21 |
| Observations | 308 | 308 | 308 |

**Independent Variables**

In this study, independent variables are news sentiment data from financial market news headlines. The news sentiment variables are the overall sentiment, the positive and the negative sentiment.

Table 3 provides the variable description and Table 4 exhibits descriptive statistics for each independent variable.

**Table 3** presents independent variables.

| Variable | Description |
|---|---|
| MKT_SEN | Overall market sentiment |
| POS | Positive market sentiment |
| NEG | Negative market sentiment |

**Table 4** reports descriptive statistics of news-based variables.

| | MKT_SEN | POS | NEG |
|---|---|---|---|
| Mean | -0.02 | 0.03 | 0.05 |
| Median | 0.00 | 0.00 | 0.04 |
| Maximum | 0.36 | 0.36 | 0.29 |
| Minimum | -0.29 | 0.00 | 0.00 |
| Std. Dev. | 0.08 | 0.05 | 0.06 |
| Skewness | 0.14 | 2.22 | 1.05 |
| Kurtosis | 4.16 | 11.03 | 3.66 |
| Jarque-Bera | 18.18 | 1080.08 | 62.06 |
| Probability | 0.00 | 0.00 | 0.00 |
| Sum | -5.79 | 9.84 | 15.63 |
| Observations | 308 | 308 | 308 |

## Textual Analysis

### Analytical Preprocessing

Financial news headlines are textual data, which are unstructured in nature. That is, they contain numbers, punctuations, alphanumeric and non-alphanumeric characters together with or without ordering. Unstructured data can be transformed to structured data, which are quantifiable, through the techniques in textual analysis. In Natural Language Processing, textual pre-processing is the procedure of cleaning text and transformation of textual characters into the bag-of-words model. In this paper, the corpus is a daily collection of financial headline news from Thomson Reuters Eikon newswires. One day is equivalent to one document in the corpus. For example, in Figure 1, there are two documents. One is 8-Feb-19 and another is 7-Feb-19. Text data pertaining to each trading day will be used in textual analysis. Figure 2 graphically illustrates the word count of each news headline.

| | | | |
|---|---|---|---|
| | | | 8-Feb-19 |
| 17:01:21 | RTRS | AXIA.KL .SETI | SE Asia Stocks-Most end lower as trade talks hit new bump on Trump's remarks |
| 11:00:05 | RTRS | DIS.N .SETI | SE Asia Stocks-Tumble on fresh trade war woes; Philippines worst hit |
| | | | 7-Feb-19 |
| 17:09:12 | RTRS | .SETI .VNI | SE Asia Stocks-Most rise on hopes of trade deal, Malaysia leads pack |
| 13:46:59 | RTRS | THB=TH PHP= | UPDATE 1-Fed, elections, oil align to revive demand for Philippine, Indonesian stocks |
| 13:14:46 | RTRS | THB=TH PHP= | Fed, elections, oil align to revive demand for Philippine, Indonesian stocks |
| 12:31:26 | RTRS | dMISX00000P .SET | BUZZ-Thailand's SET Index hits over 2-month high; technicals suggest further upside |
| 10:46:00 | RTRS | .SETI .VNI | SE Asia Stocks-Most rise on trade deal hopes, Singapore leads |

**Figure 1** shows a sample of stock market news headlines from Reuters Eikon.

The goal of the analytical preprocessing is to reduce the dimensionality in the bag-of-words model. In this paper, the standard procedure of textual preprocessing includes word concatenation, removing white space, numbers, punctuation marks, non-alpha numeric (symbols), unicode characters and a list of the English stopwords, stemming words with the Porter stemmer and, finally, tokenization. The financial corpus contained more than 20,000 terms before analytical preprocessing was conducted.

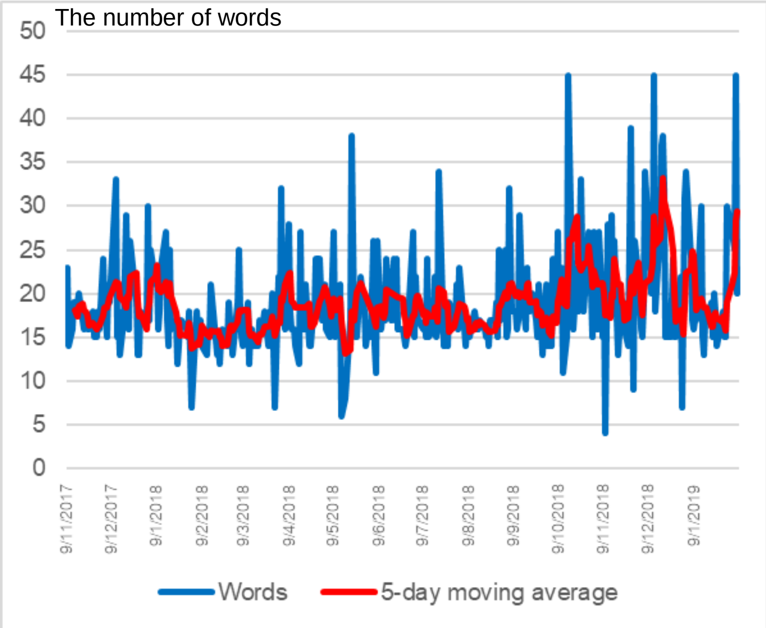


**Figure 2** shows the number of words in financial market headlines.

Some terms in financial headlines reoccurred with a high frequency but they did not contribute significant information to financial market watchers. According to Shannon's Entropy, high frequency terms carry little information with them. Hence, I employed the term-frequency inverse document frequency (TF-IDF) weighting scheme as a means to drop those terms off the document. Words that rarely occurred in the document gained a high weight score because they contained meaningful information to the reader. I eliminated the terms with a TF-IDF score equal to zero from the corpus and kept the terms with a TF-IDF greater than zero. Figure 3 illustrates a word cloud of high frequency terms and Figure 4 shows a word cloud of high frequency terms after trimming the corpus with the TF-IDF weighting scheme.


**Figure 3** depicts a word cloud of stock market headlines before the TF-IDF weighting scheme.

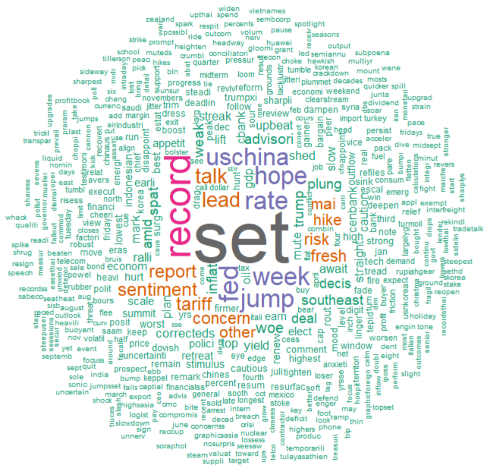


**Figure 4** shows a word cloud of stock market headlines after the TF-IDF weighting scheme.

### Dictionary Method and Market Sentiment

The dictionary technique ensures the replicability and the consistency of the results throughout the process. This method is also used in Tetlock (2007), Tetlock et al. (2008) and Loughran and McDonald (2011). Therefore, this study employed such techniques to extract the news sentiment from the corpus of financial market headlines.

In order to construct the news sentiment, a dictionary, which is a list of predetermined words in finance proposed by Loughran and McDonald (2011), is adopted as the user-defined dictionary. Then, word counting was executed and the overall sentiment of the headlines was calculated according to equation (2).

$$MKT\_SEN_t = \frac{n_{h,t}^{POS} - n_{h,t}^{NEG}}{n_{h,t}} \quad (2)$$

"$t$" is a trading day. $MKT\_SEN_t$ is the news-based market sentiment extracted from Reuters Eikon's newswires. $n_{h,t}^{POS}$ is the number of positive words occurring in the headline "$h$" on date "$t$". $n_{h,t}^{NEG}$ is the number of negative words in the headline "$h$" on date "$t$". $n_{h,t}$ is the total number of all words in headline $h$. The (net) sentiment is plotted in Figure 5. One may observe that the news-based market sentiment takes on the negative value 43.05% and the positive value 56.95% of all financial headlines.

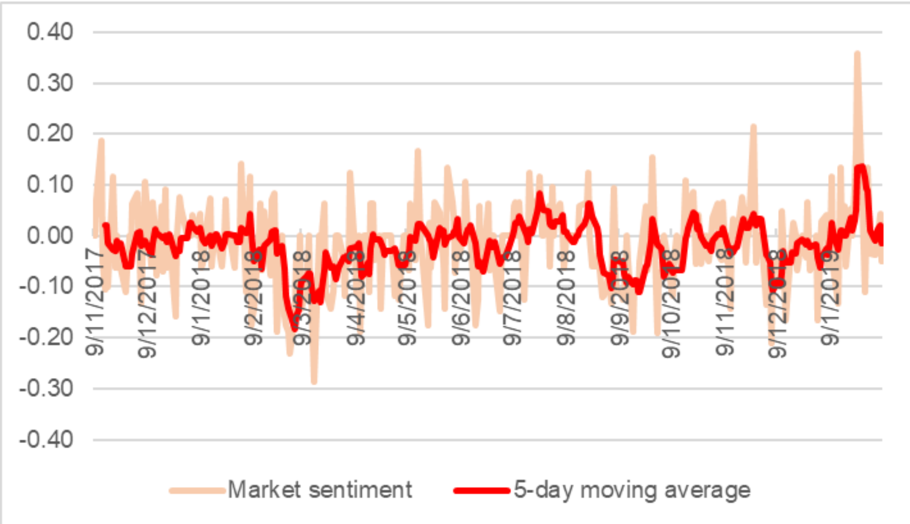


**Figure 5** presents the news-based sentiment index from stock market headline news.

In addition to the overall sentiment, the positive and the negative sentiment are also calculated by equation (3) and (4). The positive and the negative sentiment are the proportion of positive terms and of negative terms to the total number of all terms in each headline. Both positive and negative terms are from the Loughran-McDonald Sentiment Word Lists.

$$POS = \frac{n_{h,t}^{Pos}}{n_{d,t}} \tag{3}$$

$$NEG = \frac{n_{h,t}^{Neg}}{n_{d,t}} \tag{4}$$

Figure 6 and 7 depict the positive and the negative news sentiment from financial market headline news. The negative tone tends to be more pronounced than the positive tone from 2017 – 2019.

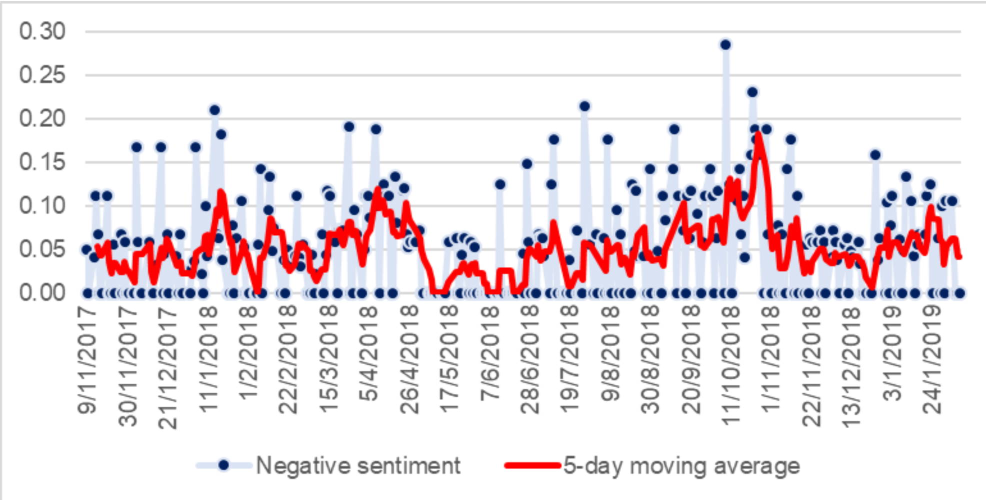


**Figure 6** depicts the negative sentiment index from stock market headline news.

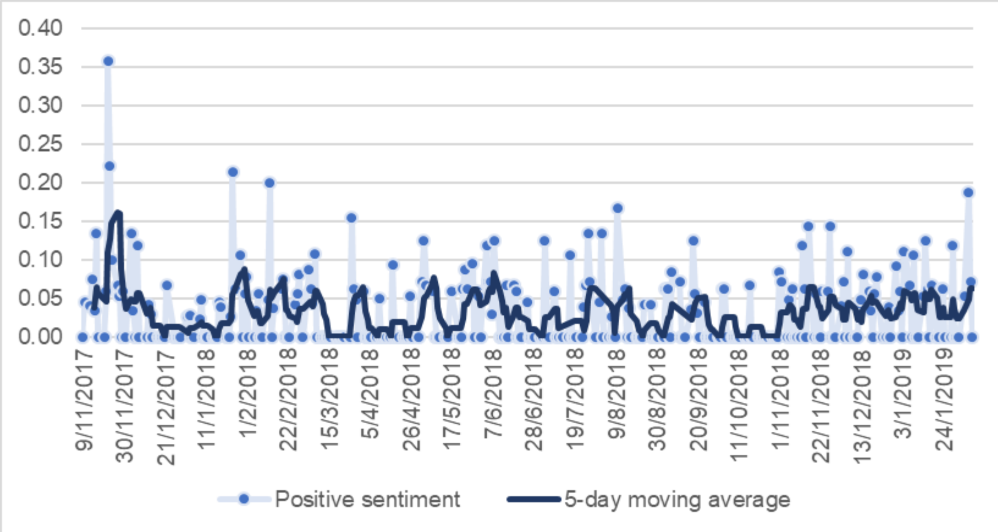


**Figure 7** presents the positive sentiment index from stock market headline news.

Due to the availability of the headline news that I extracted by the key word "BKstmad.BK", which was recommended by Eikon, the length of the headlines related to the Stock Exchange of Thailand covers from 2017-2019 during the time of the study. Nevertheless, I constructed the monthly news-based market sentiment index and compared the index with the Investor Sentiment Index (ISI) published by the Capital Market Research Institute (CMRI), which is shown in Figure 8.

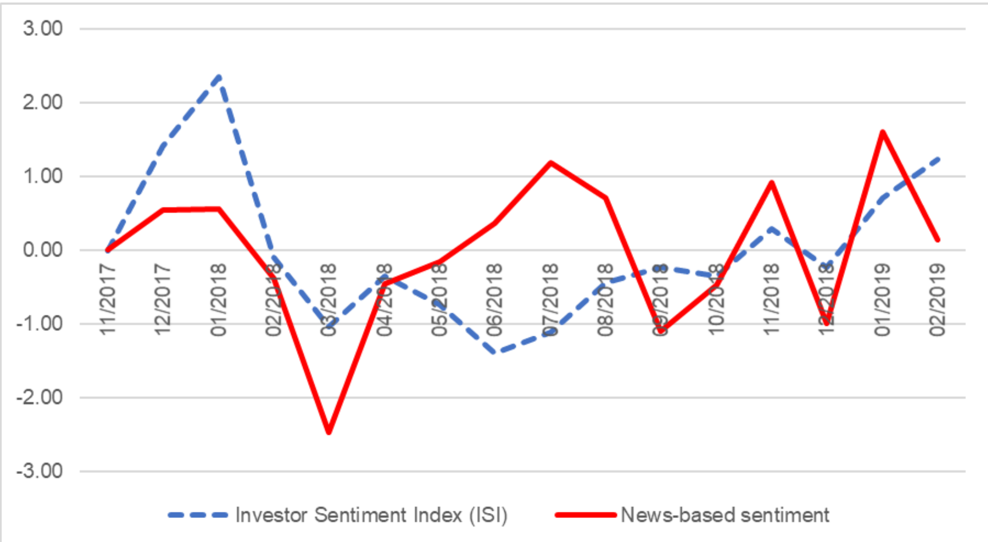


**Figure 8** compares the news-based sentiment index with
the Investor Sentiment Index (ISI) from 2017-2019.

The investor sentiment index of the Stock Exchange of Thailand was constructed by the method presented in Baker and Wurgler (2006). Both indices exhibited an upward trend even though the news sentiment fluctuated. The linear correlation between the two indices was 0.3343. One advantage of the news sentiment index is that the data frequency is daily, which can capture the immediate changes of the market sentiment. Nevertheless, this comes at the expense of the length of the index.

## Empirical Strategy

### Auto Regressive Model

In order to examine whether the sentiment contains significant information that can be used to explain stock market returns, the techniques in McLaren and Shanbhogue (2011) were employed. I began with the estimation of the following baseline ARMA(p,q) model.

**The Benchmark Model**

$$X_t=\beta_0+\beta_1X_{t-1}+\beta_2X_{t-2}+\ldots+\beta_{p-1}X_{t-p+1}+\beta_pX_{t-p}+\varepsilon_t+\gamma_1\varepsilon_{t-1}+\gamma_2\varepsilon_{t-2}+\ldots+\gamma_q\varepsilon_{t-q} \quad (5)$$

**The Overall Sentiment Equation**

$$X_t=\beta_0+\beta_1X_{t-1}+\beta_2X_{t-2}+\ldots+\beta_{p-1}X_{t-p+1}+\beta_pX_{t-p}+\varepsilon_t+\gamma_1\varepsilon_{t-1}+\gamma_2\varepsilon_{t-2}+\ldots+\gamma_q\varepsilon_{t-q}+\alpha MKT\_SEN_t \quad (6)$$

**The Positive Sentiment Equation**

$$X_t=\beta_0+\beta_1X_{t-1}+\beta_2X_{t-2}+\ldots+\beta_{p-1}X_{t-p+1}+\beta_pX_{t-p}+\varepsilon_t+\gamma_1\varepsilon_{t-1}+\gamma_2\varepsilon_{t-2}+\ldots+\gamma_q\varepsilon_{t-q}+\alpha POS_t \quad (7)$$

**The Negative Sentiment Equation**

$$X_t=\beta_0+\beta_1X_{t-1}+\beta_2X_{t-2}+\ldots+\beta_{p-1}X_{t-p+1}+\beta_pX_{t-p}+\varepsilon_t+\gamma_1\varepsilon_{t-1}+\gamma_2\varepsilon_{t-2}+\ldots+\gamma_q\varepsilon_{t-q}+\alpha NEG_t \quad (8)$$

Equation (5) is the benchmark equation. $X_t$ is a dependent variable at time t. $X_t \in \{SET\_RET_t, SET50\_RET_t, SET100\_RET_t\}$ and $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ are autoregressive terms. "p" is the order of lagged variable and is optimally chosen based on the Akaike Information Criteria value (Hereafter, AIC). $\varepsilon_t$ follows an independent, identical and normal distribution process. Equation (6), (7) and (8) extend equation (5) by incorporating the overall sentiment, the positive and the negative sentiment variable into the benchmark equation.

In order to analyze whether sentiment variables explain stock market returns, I fit equation (6) to (8) and compared the AIC values with the AIC from the benchmark equations. If the news sentiment from the financial market headline news contained the meaningful information that explains stock market variables, the estimated coefficient of $\alpha$ would be statistically significant and the AIC in equation (6) to (8) would need to be lower than the AIC in equation (5).

All financial market and news sentiment data passed the unit root test at the significance level of 1%. That is, empirical results being drawn from these data rule out the possibility of spurious relationships among variables. The unit root tests are in Table 5.

**Table 5** reports unit root tests.

| Variables | t-statistic | P-value |
|---|---|---|
| MKT_SEN | -14.27 | 0.00 |
| POS | -14.54 | 0.00 |
| NEG | -14.53 | 0.00 |
| SET_RET | -16.90 | 0.00 |
| SET50_RET | -17.32 | 0.00 |
| SET100_RET | -17.06 | 0.00 |

## Findings

ARMA (2,2) equations are estimated and the results of the benchmark equations and the overall news sentiment equations are reported in Table 6 and the positive and the negative sentiment equations are reported in Table 7.

**Table 6** reports the estimated results of ARMA (2,2)

| **Models** | **Benchmark Equations** | | | **Overall Market Sentiment** | | |
|---|---|---|---|---|---|---|
| **Variables** | $SET\_RET_t$ | $SET50\_RET_t$ | $SET100\_RET_t$ | $SET\_RET_t$ | $SET50\_RET_t$ | $SET100\_RET_t$ |
| $MKT\_SEN_t$ | - | - | - | 0.97* | 1.06* | 1.07* |
| | | | | (0.50) | (0.56) | (0.56) |
| $POS_t$ | - | - | - | - | - | - |
| $NEG_t$ | - | - | - | - | - | - |
| AR(1) | -0.44** | -0.48** | -0.46** | -0.45** | -0.49** | -0.47** |
| | (0.17) | (0.17) | (0.18) | (0.18) | (0.18) | (0.19) |
| AR(2) | -0.80*** | -0.81*** | -0.80*** | -0.80*** | -0.80*** | -0.79*** |
| | (0.19) | (0.21) | (0.21) | (0.20) | (0.23) | (0.22) |
| MA(1) | 0.47*** | 0.51*** | 0.49** | 0.48** | 0.52*** | 0.50** |
| | (0.16) | (0.17) | (0.18) | (0.17) | (0.18) | (0.19) |
| MA(2) | 0.82*** | 0.82*** | 0.81*** | 0.82*** | 0.81*** | 0.81*** |
| | (0.18) | (0.21) | (0.21) | (0.20) | (0.23) | (0.22) |
| Durbin-Watson | 1.98 | 2.03 | 2.00 | 1.99 | 2.04 | 2.01 |
| AIC | 2.1747 | 2.4042 | 2.3984 | 2.1685 | 2.3987 | 2.3928 |

*, ** and *** denote significance at the 10, 5 and 1 percent level.

**Table 7** reports the estimated results of ARMA (2,2)

| **Model** | **Negative Sentiment** | | | **Positive Sentiment** | | |
|---|---|---|---|---|---|---|
| **Variables** | $SET\_RET_t$ | $SET50\_RET_t$ | $SET100\_RET_t$ | $SET\_RET_t$ | $SET50\_RET_t$ | $SET100\_RET_t$ |
| $MKT\_SEN_t$ | - | - | - | - | - | - |
| $POS_t$ | - | - | - | 2.26** | 2.22** | 2.25** |
| | | | | (0.87) | (0.97) | (0.97) |
| $NEG_t$ | -0.65 | -0.70 | -0.69 | - | - | - |
| | (0.72) | (0.81) | (0.81) | | | |
| AR(1) | -0.44** | -0.47** | -0.45** | -0.92*** | -0.54*** | -0.51*** |
| | (0.17) | (0.18) | (0.19) | (0.12) | (0.15) | (0.17) |
| AR(2) | -0.80*** | -0.80*** | -0.79*** | -0.60*** | -0.84*** | -0.82*** |
| | (0.19) | (0.22) | (0.21) | (0.13) | (0.19) | (0.20) |
| MA(1) | 0.47** | 0.50** | 0.48** | 0.97*** | 0.56*** | 0.54*** |
| | (0.17) | (0.18) | (0.18) | (0.11) | (0.14) | (0.16) |
| MA(2) | 0.82*** | 0.81*** | 0.80*** | 0.69*** | 0.86*** | 0.84*** |
| | (0.19) | (0.22) | (0.21) | (0.11) | (0.19) | (0.19) |
| Durbin-Watson | 1.99 | 2.04 | 2.00 | 2.02 | 2.03 | 2.00 |
| AIC | 2.1785 | 2.4082 | 2.4026 | 2.1631 | 2.3935 | 2.3870 |

*, ** and *** denote significance at the 10, 5 and 1 percent level.

In Table 6, under the 'Benchmark Equations' column, all AR and MA terms are statistically significant and the AIC values are equal to 2.1747, 2.4042, and 2.3984. Plus, Durbin-Watson statistics indicate that there is no autocorrelation left.

Under the column 'Overall Market sentiment' in Table 6, the estimated coefficients of the overall market sentiment variables are statistically significant at 10% level for all three equations and AIC values of three overall market sentiment equations decrease compared to those in the benchmark equations. Thus, the overall market sentiment from financial market headlines can explain the stock market returns of the SET, SET50 and SET100 indices.

In Table 7, the effects of the negative sentiment are depicted under the column 'Negative Sentiment'. The coefficients of the negative sentiment are not statistically significant in all ARMA (2,2) and their AIC values are greater than the AIC values of the benchmark models. Adding the negative market sentiment into ARMA (2,2) does not significantly explain the stock market returns from the SET, SET50 and SET100 indices. On the contrary, under the column 'Positive Sentiment', the coefficients of the positive market sentiment are statistically significant at 5%. Their AIC values are lower than those AIC values in the benchmark. The positive market sentiment therefore contains the information that can be used to explain the returns of the SET, SET50 and SET100. By and large, the overall and the positive market sentiment can explain returns of the SET, SET50 and SET100 during the period of the study.

## Discussion and Conclusion

This study constructed the sentiment index from the corpus of financial market headlines that are related to the Stock Exchange of Thailand from Eikon's newsfeed from 2017- 2019. Thus, this allowed the researcher to answer whether the sentiment from financial headline news can explain the stock market returns from 2017-2019 in Thailand.

### Does the News Sentiment from the Financial Market Headlines Explain Stock Market Returns?

The empirical findings, ARMA (2,2), are used to answer the research question above. Adding the overall news sentiment index into ARMA (2,2) reduces AIC values across all market returns calculated from three indices. This means that incorporating the sentiment of headlines improves the model in terms of the information criterion.

Additionally, estimated coefficients of the overall sentiment are statistically significant. The news-based sentiment contains the information for market returns to some extent. As for the relationship between the sentiment index and returns, each market-sentiment coefficient has a positive sign. The more positive the market sentiment is, the higher the stock market returns across the SET, SET50 and SET100.

When categorizing the overall sentiment into the positive and the negative sentiment variables, this gives more insightful results. Only estimated coefficients of the positive sentiment index are statistically significant; whereas, coefficients of the negative sentiment index are not statistically significant. The signs of estimated coefficients of positive sentiment are also positive. These are consistent with the results from the overall market sentiment index. Unlike the positive sentiment, the estimated coefficients of the negative sentiment are negative, and they are insignificant. This may be because, during the period of the study, bad news was prevalent while good news was not. Thus, financial market participants may have responded to the good news when it was presented in the stock market headline news.

Above all, the overall market sentiment and the positive sentiment from the financial market headline news can explain stock market returns of the SET, SET50 and SET100 from 2017-2019.

## Conclusion

This study asked whether the news sentiment of stock market news headlines can explain stock market returns in the Stock Exchange of Thailand. Using the daily data from 2017-2019, the news-based sentiment was extracted from the corpus of stock market headlines from Eikon. The findings indicate that the overall market sentiment and the positive sentiment can significantly explain stock market returns, but the negative sentiment can not. Thus, the sentiment from financial market news headlines contains some crucial information for stock market returns during the period of the study.

## Acknowledgements

The previous version of this paper was awarded the research grant from the Stock Exchange of Thailand 's Capital Market Research Innovation Contest. The author would like to thank Mrs. Rojana Vanich, Mr. Manatchai and Mrs. Nuansri Chatchawan for their support throughout the project. The views expressed herein do not necessarily reflect the views of the Faculty of Commerce and Accountancy, Thammasat University.

## References

Allen, D. E., McAleer, M. J., & Singh, A. K. (2015). Chapter 19 - Machine News and Volatility: The Dow Jones Industrial Average and the TRNA Real-Time High-Frequency Sentiment Series. In G. N. Gregoriou (Ed.), *The Handbook of High Frequency Trading* (pp. 327-344). San Diego: Academic Press. doi.org/10.1016/B978-0-12-802205-4.00019-1

Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science, 4*(2), 120-137. doi.org/10.1016/j.jfds.2018.02.002

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance, 61*(4), 1645-1680. doi.org/10.1111/j.1540-6261.2006.00885.x

Engelberg, J. E., & Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. *The Journal of Finance, 66*(1), 67-97. doi.org/10.1111/j.1540-6261.2010.01626.x

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*. *68*(3), 1267-1300. doi.org/10.1111/jofi.12027

Gotthelf, N., & Uhl, M.W. (2018). News sentiment - a new yield curve factor. *Journal of Behavioral Finance*, *19*(3),31-41. doi.org/10.1080/15427560.2018.1432620

Heston, S. L., & Sinha, N.R. (2017). News versus sentiment: predicting stock returns from news stories, *Financial Analysts Journal, 73*(3)*,*67-83. doi.org/10.2469/faj.v73.n3.3

Loughran, T., & McDonald B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-Ks.*The Journal of Finance*, *66*(1), 35–65. doi.org/10.1111/j.1540-6261.2010.01625.x

Marty, T., Vanstone, B., & Hahn, T. (2020). News media analytics in finance: a survey. *Accounting & Finance, 60*(2), 1385-1434. doi:https://doi.org/10.1111/acfi.12466

McLaren, N. & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, 2011Q2, 134-140, Bank of England.

Tetlock, P. (2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*, *62*(3), 1139–1168. doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P., Saar-Tsechansky M., & Macskassy S. (2008). More than words: quantifying language to measure firms' fundamentals. *Journal of Finance*, *63*(3), 1437–1467.doi.org/10.1111/j.1540-6261.2008.01362.x

Uhl, M.W.(2014). Reuters sentiment and stock returns, *Journal of Behavioral Finance*, *15*(4), 287-298. doi.org/10.1080/15427560.2014.967852

Uhl, M.W. (2017). Emotions matter: sentiment and momentum in foreign exchange, *Journal of Behavioral Finance, 18*(3),249-257. doi.org/10.1080/15427560.2017.1332061