

Quantifying Sentiment for the Japanese Economy and Stock Price Prediction

Hiroshi Ishijima¹, Takuro Kazumi², and Akira Maeda³

¹Graduate School of International Accounting, Chuo University, Shinjuku, Tokyo, Japan
and Center on Japanese Economy and Business, Columbia Business School, New York, NY, USA

²CyberAgent, Inc., Chiyoda, Tokyo, Japan

³Graduate School of Arts and Sciences, The University of Tokyo, Meguro, Tokyo, Japan

Abstract - *The purpose of this paper is to re-examine the analysis of Ishijima et al. [2], providing a more comprehensive analysis on the sentiment of the Japanese economy that appears on the Nikkei articles. We extend their analysis in two dimensions: One is to expand data coverage, and the other is to create variations of the original sentiment index. Granger causality test results indicate the usefulness of our sentiment indexes.*

Keywords: Big data, Text mining, Nikkei, Sentiment, Stock market

1 Introduction

Sentiment analysis is gaining increasing interest in both academia and businesses. As the sentiment invisibly reflects the atmosphere of economic activities and the psychology of economic agents, analyzing the sentiment helps us understand the economy and security markets in a more sophisticated way. With such a conception, Ishijima et al. [2] analyzed the sentiment towards the Japanese economy that might appear in daily news articles. In fact, they created an index that accounts for the frequency of occurrence of words that affirmatively or negatively describe the current economic situation. Articles were taken from the *Nikkei*, a popular business newspaper in Japan. They then performed statistical analysis to examine correlations between the sentiment index and Tokyo Stock Exchange prices. Interestingly, they concluded that the index significantly predicts stock prices of three days in advance.

The purpose of this paper is to re-examine the analysis of Ishijima et al. [2], providing a more comprehensive analysis on the sentiment of the Japanese economy that appears on the *Nikkei* articles. To this end, we extend their analysis in two dimensions: One is to expand data coverage, and the other is to create variations of the original sentiment index.

(1) Data that covers 29-year-horizon:

In their study, Ishijima et al. [2] only covered a period of five years that ranges from January 2007 to September 2012. In contrast, we work on a longer sample period that covers years from March 1984 to September 2012. On a yearly basis, we examine the predictability of stock prices by our sentiment indexes.

(2) Variations of sentiment indexes:

We reconsider the methodology of creating index and newly propose four indexes. The detail of our methodology is the following:

Scoring process

We quantify the sentiment along one-dimensional semantic axis; that is from negative to positive feelings. For every single word that appears in the *Nikkei*, we match it to the prescribed semantic dictionary developed by Takamura [3]. If it matches, we record the score paired with the word that represents how much it associates the negative feeling with the Japanese people. In this scoring process, there are two aspects that we can deal with. One is about how to score on each of matched words and the other is about how far we will cover the *Nikkei* pages – just headlines or entire set of articles. We will elaborate on each of these aspects.

Scoring method

The semantic dictionary (Takamura [3]) provides, to each of contained words, the score that ranges from -1 to 1. The score shows the closer the score becomes to -1, the more negative feeling people associates the word with, and vice versa. We then exploit the score in two ways: Using the raw score or rounding to the nearest integer score that is either +1 or -1. We call the former scoring “real-valued” and the latter “integer-valued.” In the latter integer-valued case, we round to -1 if the raw score ranges between -1 and 0, and otherwise +1.

Coverage of source in the Nikkei

Then we summed up these scores over the following two sources: One is limited to headlines, the other covers the entire set of articles. We call the former coverage “Headlines Only” and the latter “Entire Set of Articles.” These two ways of counting scores make us understand how important the sentiment exhibited in headlines is in predicting the stock prices, as comparing the stock price predictability of sentiment in the entire set of articles.

With the above-mentioned methodology, we can have two scoring methods and two extents of coverage in the *Nikkei*. It results in four ways to create the sentiment index. While, in Ishijima et al. [2], they only created and examined one of four sentiment indexes; that is the integer-valued article sentiment index in our category.

The rest of this paper is organized as follows. Section 2 elaborates on how to create sentiment indexes. Section 3 builds

our models. Section 4 implements an empirical analyzes in the Japanese stock markets. Section 5 concludes the paper.

2 Creating Sentiment Indexes

Source: the coverage of pages to pick words

Every page in newspapers is comprised of pairs of headlines and articles. The place where the word appears either in headlines or in articles might affect the impact of how much the reader will invoke positive or negative sentiment on that word. In this aspect, we strictly distinguish the words in articles from the words in headlines. We clarify on this by introducing some notations.

In the newspapers delivered at day t , we have $n_t^{(H)}$ headlines and $n_t^{(A)}$ articles. Each headline and article are denoted by $H_{i,t}$ ($i = 1, \dots, n_t^{(H)}$) and $A_{l,t}$ ($l = 1, \dots, n_t^{(A)}$). Each headline and article have $n_{i,t}^{(H)}$ and $n_{l,t}^{(A)}$ words, respectively. The words in headline $H_{i,t}$ and article $A_{l,t}$ are denoted by $W_{ij,t}^{(H)}$ ($j = 1, \dots, n_{i,t}^{(H)}$) and $W_{lm,t}^{(A)}$ ($m = 1, \dots, n_{l,t}^{(A)}$), respectively.

At this point, we introduce the aggregate notation to represent whichever the word that comprises either headlines or articles. This enables us easier to articulate how to quantify the sentiment in the discussion that will follow. The coverage of pages to pick words is limited to either headlines or articles and is denoted by $\mathcal{G} := \{H, A\}$. We simply call \mathcal{G} as the “source.” We then let $G \in \mathcal{G}$ to show either headline H or article A . Then in the newspaper delivered at day t , $W_{ij,t}^{(G)}$ denotes the j -th word ($j = 1, \dots, n_{i,t}^{(G)}$) that comprises the i -th source $G_{i,t}$ ($i = 1, \dots, n_t^{(G)}$).

Semantic dictionary

The semantic dictionary (Takamura, [3]) is denoted by $\mathcal{D} := \{(D_k, S(D_k)) | k = 1 \dots K\}$. That is, the dictionary comprises pairs of word and its semantic score that ranges from -1 to +1. Regarding the semantic score, the closer to -1 (or +1) the score becomes, the more negative (or positive) feeling the word invokes to the Japanese people.

Semantic index: two methods to quantify the positive or negative feelings

We define the indicator function to count if the word that appears in the source matches one of listed words in the dictionary.

$$I_{ij,t}^{(G)}(k) := \begin{cases} 1 & \text{(if } W_{ij,t}^{(G)} \text{ matches } D_k) \\ 0 & \text{(otherwise)} \end{cases} \quad (1)$$

In the aspects of how to score the positive or negative feelings, we introduce two methods to create the sentiment indexes.

(1) Real-valued sentiment index

The first way is to exploit the semantic score $S_k = S(D_k)$ that is assigned to the listed word D_k . We define this first way as the “real-valued sentiment index:”

$$x_t^{(G,R)} := \sum_{i=1}^{n_t^{(G)}} \sum_{j=1}^{n_{i,t}^{(G)}} \sum_{k=1}^K I_{ij,t}^{(G)}(k) \cdot S_k \quad (2)$$

For the source $G = H$ that the coverage of pages to pick words is limited to headlines, the real-valued sentiment index is given by $x_t^{(G,R)} = x_t^{(H,R)}$. In the same way, for the source of $G = A$ (articles), the real-valued sentiment index is given by $x_t^{(G,R)} = x_t^{(A,R)}$.

(2) Integer-valued sentiment index

The second way is to round the semantic score S_k to the nearest binary integer that is either -1 or +1. Introducing the integer variable for each of semantics scores:

$$J_k := \begin{cases} +1 & \text{(if } 0 < S_k \leq 1) \\ -1 & \text{(if } -1 \leq S_k < 0) \end{cases} \quad (3)$$

We then define the second way as “integer-valued sentiment index”:

$$x_t^{(G,I)} := \sum_{i=1}^{n_t^{(G)}} \sum_{j=1}^{n_{i,t}^{(G)}} \sum_{k=1}^K I_{ij,t}^{(G)}(k) \cdot J_k \quad (4)$$

Reminding that each of two sentiment indexes has the option in picking the source, – either headlines ($G = H$) or entire set of articles ($G = A$) – we thus have four types of sentiment indexes in the analysis.

As a summary, we use the following notation to represent these four sentiment indexes (s.i.).

$$x^{(G,\#)} := \begin{cases} x^{(H,I)} & \text{(integer – valued headline s. i.)} \\ x^{(H,R)} & \text{(real – valued headline s. i.)} \\ x^{(A,I)} & \text{(integer – valued article s. i.)} \\ x^{(A,R)} & \text{(real – valued article s. i.)} \end{cases} \quad (5)$$

where G denotes one of the sources H or A and $\#$ denotes the scoring method that is either integer-valued “I” scoring or real-valued “R” scoring.

Along the procedures that we described above, we created 29-year daily time-series of four sentiment indexes based on headlines and articles from the Nikkei. We remark that these sentiment indexes are normalized so that they have zero means and unit standard deviations. Due to space limitation, we omitted reporting summary statistics and time-series charts of our sentiment indexes.

3 Model

To explore whether or not our sentiment indexes are able to predict stock prices, we will exploit the vector auto-regression (VAR) model that is conventional one in the econometrics literature. For each of two cases in terms of the scoring method, we estimated three VAR(p) models which comprise either (1) Model H: headline sentiment index, (2) Model A: entire article sentiment index, (3) Model H&A: both headline and article sentiment indexes, as well as stock log-returns. Namely, each of three models is specified as follows:

$$\text{Model H} \quad y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \beta_i x_{t-i}^{(H,R \text{ or } I)} + \varepsilon_t \quad (6)$$

$$\text{Model A} \quad y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \gamma_i x_{t-i}^{(A,R \text{ or } I)} + \varepsilon_t \quad (7)$$

$$\text{Model H\&A} \quad y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \beta_i x_{t-i}^{(H,R \text{ or } I)} + \gamma_i x_{t-i}^{(A,R \text{ or } I)} + \varepsilon_t \quad (8)$$

Within these VAR(p) model specifications, the *Granger causality* can be stated if the sentiment indexes ($x^{(G,\#)}$) *Granger-cause* (G -cause) stock log-returns (y), the past sentiment indexes should help predict the stock log-returns, beyond prediction by past stock log-returns alone. Using these three VAR(p) models, we will implement the three Granger causality tests (G-tests).

(1) “*G-test for Model H*” tests whether or not the headline sentiment index G -causes stock log-returns by exploiting Equation (6). The null hypothesis is $\beta_i = 0$ ($i = 1 \dots p$).

(2) “*G-test for Model A*” tests whether or not the entire article sentiment index G -causes stock log-returns by exploiting Equation (7). The null hypothesis is $\gamma_i = 0$ ($i = 1 \dots p$).

(3-1) “*G-test 1 for Model H&A*” tests whether or not the headline sentiment index G -causes stock log-returns by exploiting Equation (8). The null hypothesis is $\beta_i = 0$ ($i = 1 \dots p$).

(3-2) “*G-test 2 for Model H&A*” tests whether or not the entire article sentiment index G -causes stock log-returns by exploiting Equation (8). The null hypothesis is $\gamma_i = 0$ ($i = 1 \dots p$).

4 Empirical Analysis in the Japanese

Stock Market

While the Nikkei is daily published and delivered with a few no-issue days, the Japanese stock market is closed every weekend. To handle such daily data set that is partially missing and hence inconsistent in frequency, we follow the approach of Bollen et al. [1]. That is, we eliminated every Saturday and Sunday from complete data set before implementing analysis.

Also, before estimating VAR models Eqs. (6)–(8), we implemented augmented Dickey–Fuller tests. For each year, we verified that all the time series of stock log-returns, headline and article sentiment indexes do not have unit root with 1% significance.

Table 1 and Table 2 show Granger test results with real- and integer-valued sentiment indexes, respectively. For each year, three models of Model H, A and H&A or relevant Eqs. (6), (7) and (8) are estimated and tested. On each model estimation, we searched the lag p from 1 to 7 to identify the best (i.e. lowest) p in terms of AIC. For each model estimation with the best p , the relevant test statistics (“Granger”-labeled columns) and AIC values are reported in Tables.

Table 1: Predictability of “real-valued sentiment index” in terms of Granger tests and AICs. Test statistics for Granger causality are given in the columns titled “Granger.” *, ** and *** mark the test statistics that are 10%, 5% and 1% significant, respectively.

Year	Headline Eq. (6)			Article Eq. (7)			Headline & Article Eq. (8)			
	Lag (p)	Granger	AIC	Lag (p)	Granger	AIC	Lag (p)	Granger H	Granger A	AIC
1984	3	1.12	-10.17	6	1.51	-10.10	6	4.99***	5.20***	-11.19
1985	7	1.54	-10.46	6	1.78	-10.65	6	3.60***	3.67***	-11.45
1986	2	1.69	-10.18	6	1.00	-10.44	6	1.91**	1.17	-11.60
1987	2	0.02	-9.21	6	1.78	-9.77	6	0.71	1.64*	-11.17
1988	2	1.12	-11.07	6	0.66	-11.38	6	1.88**	1.28	-12.88
1989	2	2.89*	-11.53	5	1.39	-12.01	5	1.98**	1.86**	-13.70
1990	2	0.14	-8.73	7	1.96*	-9.63	5	3.30***	4.75***	-11.21
1991	2	0.40	-9.37	5	0.52	-10.35	5	3.48***	4.93***	-11.72
1992	5	1.29	-8.69	5	0.83	-9.48	5	2.06**	2.69***	-10.79
1993	1	1.67	-9.18	5	0.21	-10.03	5	3.17***	3.69***	-11.29
1994	1	0.41	-9.46	5	0.37	-10.27	5	1.77*	5.12***	-11.52
1995	5	0.94	-9.19	5	0.37	-9.98	5	3.91***	4.71***	-11.48
1996	5	1.08	-10.05	5	0.53	-11.08	5	1.70*	2.24**	-12.61
1997	2	0.07	-8.78	5	1.18	-9.72	5	2.31**	2.49***	-11.21
1998	5	3.01**	-8.80	5	2.22*	-9.94	5	1.43	3.23***	-11.28
1999	1	0.64	-9.29	5	0.29	-9.99	5	1.60	2.20**	-11.22
2000	1	4.35**	-8.99	5	1.37	-10.04	6	1.77**	3.74***	-11.36
2001	2	0.32	-8.59	1	0.36	-9.72	1	2.19	2.31*	-11.17
2002	2	2.03	-8.90	5	0.94	-10.10	5	0.77	1.42	-11.74
2003	1	0.12	-8.89	5	0.25	-10.30	1	0.54	2.59*	-11.99
2004	1	0.16	-9.50	5	1.30	-10.70	1	0.19	2.74*	-12.26
2005	5	0.18	-10.10	5	0.30	-11.24	1	4.29***	8.65***	-12.81
2006	5	0.70	-9.37	5	1.74	-10.53	5	1.07	3.57***	-12.27
2007	6	0.42	-9.59	6	0.15	-10.70	5	2.27**	4.40***	-12.32
2008	6	0.37	-7.68	5	0.58	-8.59	5	1.92**	4.48***	-10.15
2009	6	0.95	-8.87	6	0.83	-9.88	6	1.51	1.62*	-11.36
2010	5	2.35**	-9.24	5	2.26**	-10.36	5	1.51	2.72***	-12.02
2011	1	0.19	-9.18	5	1.13	-10.00	5	1.07	1.45	-11.44
2012	2	3.88*	-9.71	5	1.13	-10.91	5	1.20	1.04	-12.32

Table 2: Predictability of “integer-valued sentiment index” in terms of Granger tests and AICs. Test statistics for Granger causality are given in the columns titled “Granger.” *, ** and *** mark the test statistics that are 10%, 5% and 1% significant, respectively.

Year	Headline Eq. (6)			Article Eq. (7)			Headline & Article Eq. (8)			
	Lag (p)	Granger	AIC	Lag (p)	Granger	AIC	Lag (p)	Granger H	Granger A	AIC
1984	1	0.35	-9.76	1	1.06	-10.04	1	2.77*	2.72*	-10.19
1985	2	3.03**	-10.39	3	3.39**	-10.72	1	1.16	0.69	-10.83
1986	1	2.58	-9.62	1	2.21	-9.49	1	3.56**	0.49	-9.79
1987	1	0.20	-8.40	6	1.29	-8.70	1	0.28	2.11	-9.08
1988	2	0.60	-10.37	5	1.87*	-10.77	1	0.18	0.98	-11.15
1989	6	1.92*	-10.42	1	2.49	-10.86	1	0.12	1.50	-11.11
1990	2	0.04	-7.65	2	1.01	-8.16	2	1.16	0.89	-8.28
1991	1	0.37	-8.69	1	1.36	-9.03	2	2.86**	1.76	-9.39
1992	1	2.07	-8.28	5	1.04	-8.29	1	0.95	0.23	-8.80
1993	4	3.51***	-8.81	1	0.03	-9.06	1	0.87	0.03	-9.34
1994	5	1.32	-9.55	1	0.01	-9.59	1	0.14	0.42	-10.31
1995	1	0.21	-8.86	1	0.11	-9.28	1	0.19	0.52	-9.74
1996	1	2.12	-9.33	1	0.12	-9.73	1	1.70	0.44	-9.91
1997	2	1.04	-8.67	2	0.42	-8.82	2	0.53	0.57	-9.48
1998	2	0.78	-8.49	5	2.05*	-8.78	1	1.43	2.12	-9.40
1999	1	0.10	-9.35	1	0.05	-9.57	1	0.11	7.06***	-10.39
2000	1	0.31	-8.41	1	0.65	-9.11	1	0.86	0.67	-9.23
2001	1	1.43	-8.28	1	0.02	-8.48	1	3.02**	0.59	-9.13
2002	1	0.10	-8.82	1	0.91	-9.01	1	0.02	1.70	-9.78
2003	5	2.88*	-8.70	4	1.71	-8.63	1	2.73*	2.29	-9.10
2004	1	0.01	-9.38	1	0.00	-9.69	1	2.96*	0.23	-10.38
2005	1	0.48	-9.74	5	0.41	-9.67	5	1.13	1.45	-10.13
2006	1	0.17	-9.24	2	0.13	-9.70	2	0.86	1.02	-10.30
2007	1	1.28	-9.06	6	2.21**	-9.27	1	2.41*	3.40***	-9.66
2008	1	0.32	-7.62	3	0.41	-7.68	1	1.04	2.16	-8.31
2009	3	1.73	-8.38	6	1.57	-8.02	1	6.26***	0.91	-8.60
2010	3	1.55	-8.89	5	2.18*	-9.22	3	2.64**	1.35**	-9.56
2011	1	0.10	-8.61	4	2.43**	-8.95	1	1.34	3.48**	-9.21
2012	1	0.87	-9.56	3	0.95	-9.86	1	1.95	0.20	-10.36

Goodness of fit:

For both real- and integer-valued cases, Model H&A fits best in terms of AICs throughout 29 years. When comparing the real- and integer-valued cases, the former performs better in the aspects of AICs. Hence, in the following discussion, we will mainly focus on the results for estimating Model H&A on the basis of real-valued sentiment index.

Predictability of stock prices:

From the right most panel in Table 1, Model H&A of Eq. (8) persistently shows the predictability of stock prices on the basis of real-valued sentiment index. More specifically, the article index persistently and significantly Granger-causes the stock log-returns in conjunction with the headline index (“Granger A” column in that right most panel). Also, the headline index Granger-causes the stock log-returns in conjunction with the article index (“Granger H” column in the same panel). On the flipside, Models H and A (Eqs. (6) and (7)) do not seem to provide the persistent Granger causalities. These results imply that it is important for our VAR modeling to incorporate both headline and article sentiment indexes in order to predict stock prices; and that it is insufficient to separately introduce either headline or article sentiment indexes¹.

It should be noted that the Granger causality seems to be weakened during some periods. We will elaborate on each of these. During the period from 1986 to 1988 that was right after the Plaza Accord on 1985, the Japanese market has been a bull market and on the way to its peak. In this period, the headline sentiment index has stronger Granger causality than the article’s except 1987 that has brought Black Monday.

In contrast, during the period from 2001 to 2004 that was right after the burst of the Internet bubble, in the year of 2009 that was right after the 2008 financial crisis, and during two years from 2011 to 2012 in which we experienced the Japan quake and had been trying to recover, the Japanese economy suffered from these unusual events. In those periods, the article index has stronger Granger causality than the headline index.

Significant lagged variables:

Table 3 exhibits the year-on-year estimates of Model H&A of Eq. (8) on the basis of real-valued sentiment index. Interestingly enough, we found the seven cyclical patterns in our estimation results.

Cycle 1 was significant: For the first 10 years from 1984 to 1993 except 1986 – that correspond to the 11th business cycle defined by Cabinet Office in Japan –, two sentiment indexes with shorter lags (1 to 2) and longer lags (5 to 6) served as significant variables.

Cycle 2 was not significant: In the next 5 years from 1994 to 1998 that correspond to the 12th business cycle in Japan, we couldn’t find any significant lagged variables on neither headline nor article indexes.

Cycle 3 was significant: In the next 2 years from 1999 to 2000 that correspond to the Internet bubble, two sentiment indexes with lag-of-one and middle lags (3 to 4) were significant.

Cycle 4 was not significant: During the period from 2001 to 2005 that was after the burst of the Internet bubble and before 2008 financial crisis, we couldn’t find any significant lagged variables in sentiment indexes, again.

Cycle 5 was significant: In the 3 years from 2006 to 2008 that brought the financial crisis, two sentiment indexes with lag-of-one and middle lags (3 to 4) were significant.

Cycle 6 was not significant: During the period from 2009 to 2010 that was right after the crisis and corresponded to the very first two years of the 15th business cycle in Japan, we couldn’t find any significant lagged variables in sentiment indexes.

Cycle 7 was significant: In recent two years after the Japan quake, two sentiment indexes with lag-of-two were significant.

Comparison with the relevant work:

Ishijima et al. [2] reported that during the period after the 2008 financial crisis, the integer-valued article sentiment index alone significantly predicts stock prices three-days-ahead. This can be found in the middle panel titled “Article Eq. (7)” on Table 2. Indeed, we can see the significant Granger causalities around 2008. Unfortunately, this finding does not seem to be persistent when we review this from 29-year-horizontal results that we have shown in this paper.

5 Conclusions

We created the 29-year daily time-series of four sentiment indexes that reflect the positive or negative feelings represented in the Nikkei newspaper. The analysis is based on Ishijima et al. [2], but is a sophisticated version of their analysis. We showed the persistent predictability of Japanese stock prices on the basis of two sentiment indexes that quantified the sentiment over headlines and entire articles, respectively.

6 References

- [1] J. Bollen, H. Mao and X. Zeng, “Twitter Mood Predicts the Stock market,” *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [2] H. Ishijima, T. Kazumi and A. Maeda, “Sentiment Analysis for the Japanese Stock Market,” *forthcoming in Global Business and Economics Review*, 2014.
- [3] H. Takamura. (2007). Tango Kanjyo Kyokusei Taiou Hyou (Semantic Orientations Dictionary). [Online]. Available URL: http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html (accessed 2014-05-31)

¹ Results shown here are those of Granger causality tests on the basis of OLS estimators. We remark that the data we handled may contain sample biases. Hence, to consider the heteroscedasticity due to such possible sample biases, we double-checked to conduct tests with robust covariance-matrix

estimators as well. The results were mostly the same in case of real-valued sentiment index, but there were found small differences in case of integer-valued sentiment index. The detail is omitted here due to space limitation.

