

This article was translated by the author and reprinted from the April 2014 issue of the Securities Analysts Journal® with the permission of the Securities Analysts Association of Japan (SAAJ).

Stock Market Predictability Using News Indexes

Tatsuyoshi Okimoto*

Crawford School of Public Policy

Australian National University

and

Eiji Hirasawa†, CMA

Financial Technology Research Institute

NIKKEI Group

Contents

1. Introduction
2. Tag Information and News Indexes
3. News Theories and Their Verification
4. Analysis of Individual Evaluation Items
5. Conclusion

Abstract

This study examines market predictability based on news indexes constructed by quantifying Nikkei news distributed at QUICK terminals. QUICK is a member of the Nikkei Group and one of the leading financial information vendors in Japan. Our analysis indicates that news indexes have significant power to explain stock returns and trading volumes on the next business day, and that the kinds of rebounds reported in previous research are not observed. This suggests that the news indexes used in this study contain essential information with respect to stock prices, a result of substantial interest.

*Associate Professor, Crawford School of Public Policy, Australian National University, 132 Lennox Crossing, ANU, Acton, ACT 2601, Australia. Tel: +61-2-6125-4763. e-mail: tatsuyoshi.okimoto@anu.edu.au.

†Senior Executive Managing Director, Financial Technology Research Institute, Inc. 19th Floor, Nihonbashi 1-chome Mitsui Bldg., 1-4-1 Nihonbashi, Chuo-ku, Tokyo. Tel: +81-3-3276-3440. e-mail: ehirasawa@ftri.co.jp

1. Introduction

When it comes to stock investments, decisions are often a synthesis not only of numerical values such as stock prices, financial indices, and so on, but also of “text information” such as news.¹ It makes sense, therefore, to use both numerical and text-based information in a study of decision making. However, such studies have not been conducted in sufficient numbers to date because using text-based information requires conversion of text into information that is quantifiable and subject to analysis, a task which presents significant technical and financial hurdles.

In this paper, we try to examine the relationship between text information and the stock market, in particular market predictability, by quantifying “news” text information using news tags formulated by the Financial Technology Research Institute, Inc. (FTRI). Although text may include information other than news, such as SNS information from the Internet, our analysis is focused on news that might be given priority use by investors.

Currently, a large amount of news about stock markets is distributed through newspapers and information terminals everywhere. Although news is expected to contain some fundamental information about stock prices, few studies have found a strong relationship between the two, as Roll [1988] found almost no relationship between news and stock returns. Tetlock [2007], for example, created news indexes from Wall Street Journal (WSJ) columns and found that they could be a market sentiment index, but they did not contain information essential to stock prices.

Not much research has been done concerning the effects of news on the Japanese stock market, although Griffin et al. [2011] showed that news may cause an increase in stock market volatility. In this study we quantify Nikkei news distributed at QUICK terminals to construct news indexes and then examine whether those indexes contain essential information in relation to stock prices. Specifically, following Tetlock [2007], we use a three-variable VAR model consisting of the news indexes, stock returns, and trading volumes, and investigate three theories, namely information theory, sentiment theory, and no-information theory. In addition, we classify news into individual evaluation items, and attempt to clarify what kinds of news have the biggest effect on stock returns.

The remainder of the paper is structured as follows. Section 2 provides a general overview of the news indexes and the news and tag information upon which they are based. Section 3 introduces three news theories based on Tetlock [2007] and

¹ In this paper, information that is expressed by way of text and difficult to process by machines such as personal computers is referred to as “text information.”

investigates them in relation to the news indexes based on a three-variable VAR model consisting of news indexes, stock returns, and trading volumes. Section 4 classifies news into individual evaluation items and examines what kinds of news have the greatest effect on stock returns. Finally, Section 5 presents conclusions.

2. Tag Information and News Indexes

The news data used in this study amounts to 809,667 Nikkei news articles distributed at QUICK terminals between 1 July 2007 and 31 December 2011, involving news reports on 3,535 listed companies. The news data was provided to us by Nikkei and Quick Co. Ltd. This news is read in real time by institutional investors and individual investors participating in the securities market, and has substantial power to communicate information to the market.

To enable quantified processing of this news data, we use tag information created by FTRI. The tag information chiefly used comprises company names (including codes) that are mentioned in the news as well as positive or negative information with reasons attached.² Such positive and negative information is defined in **Figure 1** and formulated from news headlines and bodies using the definitional criteria of FTRI. **Figure 2** shows the distribution of positive and negative information. Market commentary communicating the state of the market, and (positive) news conveying information about new products and services make up the major portion of distribution.

We will now discuss our data limitations. Time stamp information in units of hours, minutes, and seconds was not available for this news data. While one approach is to process news that comes after market closing (after 3:00 p.m.) as news of the following day, in this study, news occurring up to midnight is considered news of the same day and analyzed as such.³ To construct news indexes, we perform a daily count of the number of positive and negative news items appearing for each listed company, and, referring to Tetlock, et al. [2008], calculate the following two indexes.⁴

$$NI1 = \frac{P - N}{P + N + 1}, \quad NI2 = \frac{-N}{P + N + 1} \quad (1)$$

Here, P is the number of positive news items reporting positive information and N is the number of negative news items reporting negative information. In short, $NI1$ is

² According to FTRI, the accuracy of tag formulation exceeds 99% for both company names and positive/negative information, but the formulations are based on models and so may be less accurate when there is a relative increase in news of an unanticipated form or nature.

³ Separate from the data used in the paper, we chose (four) sample weeks from 2013, and in looking at the proportion of news from closing up until midnight found it amounted to about 30% of all news.

⁴ For this purpose, news not containing company names is excluded from analysis.

the difference in the number of positive and negative news items divided by the total number of news items, and $NI2$ is the calculated percentage of negative news in the total number of news items.⁵ It might be possible to use numerical values included in news to weight N and P , but since ideal weights may vary by brand attributes such as sector, we believe this requires further separate discussion. Note, therefore, that this study deals simply with the number of news items in order to investigate whether the number of news items is significant information for stock market predictions, without regard to any numerical information contained therein.

Other data used in the analysis, such as stock price indexes and volume of shares traded on the market, is obtained using NEEDS-Financial QUEST 2.0 from Nikkei Digital Media, Inc.

3. News Theories and Their Verification

Tetlock [2007] considered the following three theories about information contained in news. The first, information theory, holds that news contains essential information with respect to stock prices. Under this theory, news has a significant effect on future stock prices, and that effect continues permanently. The second theory is sentiment theory, according to which news does not possess information essential to stock prices but does influence market sentiment. Under this theory, news has a short-term significant effect on future stock prices, but this effect disappears over the long term. In the third theory, no-information theory, news has no effect whatsoever on either stock prices or market sentiment. In this case, news has neither short-term nor long-term effects on future stock prices.

Tetlock [2007] investigated these three theories in relation to WSJ columns. Specifically, Tetlock [2007] used a three-variable VAR model consisting of news indexes, stock returns, and trading volumes to determine whether the news indexes significantly influence future stock returns and whether such influence disappears with time. The results showed that the indexes constructed from WSJ columns had significant positive effects on stock prices in the short term, but that rebounds occurred and there was no significant long-term effect. In other words, these results are consistent with the sentiment theory, which suggests that WSJ columns do not contain essential information in relation to stock prices but may contain information that relates to

⁵ In Tetlock, et al. [2008], the explanatory power of positive news is weak, and only results based on indices similar to $NI2$ are reported. The index used in Tetlock, et al. [2008] and the news index in this study differ in that 1 was added to the denominator to avoid denominators of zero, and that $NI2$ was negated to keep signs identical for results using $NI1$ and $NI2$.

market sentiment.

In this paper, we perform an analysis similar to Tetlock [2007], and, using the news indexes outlined above, examine the three news theories. Specifically, using a three-variable VAR model consisting of news indexes (NI_t), stock returns (Tpx_t), and trading volumes (Vol_t), we investigate the possibility that news indexes contain essential information that relates to stock prices. For stock returns we use TOPIX's daily logarithmic rate of return (%), and for trading volumes we use a logarithm of the total volume of all stocks traded on the TSE 1st Section. Also, the news indexes are normalized so that they have an average of 0 and a standard deviation of 1. As in Tetlock [2007], the VAR lag is assumed to be 5, which is to say that the marginal effect of each variable disappeared within the period of a week. Under this assumption, the regression model for stock returns would look as follows⁶:

$$Tpx_t = \alpha_1 + \sum_{j=1}^5 \beta_{1j} Tpx_{t-j} + \sum_{j=1}^5 \gamma_{1j} NI_{t-j} + \sum_{j=1}^5 \delta_{1j} Vol_{t-j} + \varepsilon_{1t} \quad (2)$$

An important parameter for examination of the three news theories is γ_1 in (2). If the information theory is correct, γ_{11} should be positive and significant, and the remaining components of γ_1 should not be negative and significant. However, if the sentiment theory is correct, γ_{11} should be positive and significant, and γ_{1j} ($2 \leq j \leq 5$) should be negative and significant for at least one j . Finally, if the no-information theory is correct, no component of γ_1 should be significant.

Assuming (2), **Figure 3** summarizes estimated values for γ_1 together with Newey-West standard errors.⁷ As the figure shows, γ_{11} is significantly positive with a significance level of 5%, for every news index. This suggests that news indexes have a significant positive effect on stock returns the next business day, and that the no-information theory is not valid. In addition, other elements of γ_1 do not become significant for either news index. In contrast to Tetlock [2007], in which some of γ_1 was

⁶ Note that, due to data limitations stated above, this VAR model is not a structural VAR model treating hours after market closing (after 3:00 p.m.) as part of the following day and taking into account contemporaneous dependence. Also, the present study looks at the number of news articles in a day, and that number cannot be determined until the day is over, so in this sense the model in (2) is quite natural. A valuable future step would be to use time stamps and perform more precise analysis with observation periods set to shorter periods of time.

⁷ It must be noted that much of the news is stock market news and there may be multicollinearity due to a correlation between Tpx and $NI1$. Indeed, correlation between the two is high at around 0.6, but the centralized variance inflation factor as a result of estimating (2) is not greater than 4 for any variable. Since multicollinearity only becomes suspected at 10, it is not likely to be an issue.

negative and significant after two business days and rebounds were observed, these results show that news indexes have a permanent effect on stock prices. In other words, this indicates that the theory most likely to be valid is not the sentiment theory, but the information theory.

Interpreting the results of $NI1$ ($NI2$) more specifically, this demonstrates that if the news indexes of a given day are one standard deviation larger than average values, there is a tendency for stock returns on the next business day to be 0.167% (0.197%) larger in comparison to days when news indexes have average values. This value is also very significant economically, strongly suggesting the potential for news indexes to contain information essential to stock prices. In addition, though the results are not reported here to save space, the estimated value of β_1 does not become significant at any order, highlighting the significance of news indexes.

Next, we examine whether the news indexes in this study are not consistent with the sentiment theory from the perspective of the market's influence on news. To this end, we focus on the news index regression model in the three-variable VAR model, namely

$$NI_t = \alpha_2 + \sum_{j=1}^5 \beta_{2j} Tpx_{t-j} + \sum_{j=1}^5 \gamma_{2j} NI_{t-j} + \sum_{j=1}^5 \delta_{2j} Vol_{t-j} + \varepsilon_{2t}. \quad (3)$$

If the sentiment theory were correct, past stock market performance should have a positive effect on news indexes, which means β_2 should have a tendency to be positive and significant.

Assuming (3), **Figure 4** summarizes estimated values of β_2 together with Newey-West standard errors. As seen in the figure, no matter what the news index, there is no significance from β_{21} to β_{24} at the 5% significance level. Also, β_{25} is significant but it has a negative value. Thus, no statistical evidence was observed for past market performance having a positive effect on news indexes. This result is also in contrast to the Tetlock [2007] result with significantly positive β_{21} and is not consistent with the sentiment theory.

Finally, we explore the three theories from the effects of news indexes on trading volumes. On the basis of theories in DeLong et al. [1990] and Campbell et al. [1993], when there is a shock in market sentiment, noise traders conduct transactions with rational traders. As a result when sentiment is normalized to an average of zero, trading volumes rise as the absolute value of sentiment increases. In consideration of this observation, Tetlock [2007] considered the following type of regression model for trading volume, which includes an absolute value of news indexes:

$$Vol_t = \alpha_3 + \sum_{j=1}^5 \beta_{3j} Tpx_{t-j} + \sum_{j=1}^5 \gamma_{3j} NI_{t-j} + \sum_{j=1}^5 \psi_{3j} |NI_{t-j}| + \sum_{j=1}^5 \delta_{3j} Vol_{t-j} + \varepsilon_{3t} \quad (4)$$

If the sentiment theory is correct, ψ_{31} should be positive and significant. Assuming (4), **Figure 5** summarizes estimated values for γ_3 and ψ_3 together with Newey-West standard errors. As seen in the figure, ψ_{31} is not significant for either news index. In contrast, whatever the news index, γ_{31} is positive and significant. This tells us that positive news tends to increase trading volumes, and negative news tends to decrease trading volumes. More specifically, this suggests that if $NI1$ ($NI2$) goes up by one standard deviation, trading volumes the next business day increase around 1.5% (1.3%) on average. This result shows that news indexes have a powerful influence on the next business day's trading volumes, which is also consistent with the information theory.

To summarize the above results, the information theory is more consistent with the news indexes used in this study than either the sentiment theory or the no-information theory. This is in contrast to the results of Tetlock [2007], and suggests the possibility that the news indexes in this study contain information that is essential to stock prices. One reason why results of this kind were obtained may be that while the news indexes in Tetlock [2007] were created based on WSJ columns, the news indexes in this study were constructed from Nikkei news distributed at QUICK terminals. WSJ columns describe the entirety of market trends, and while the information there is powerful in its summarizing aspect, it can also be strongly nuanced, so though it may affect market sentiment temporarily, it is likely to contain little essential information in relation to stock prices. In contrast, the Nikkei news analyzed in this study has more detailed information, mainly positive or negative news about individual listed companies, and it has been confirmed to have high potential for including essential information that has an effect on stock prices.

4. Analysis of Individual Evaluation Items

Section 3 suggested the potential of the news indexes used in this study to contain essential information relating to stock prices. This section will categorize news into items to be evaluated and attempt to clarify what kinds of news have the greatest effect on stock returns.

To that end, for each evaluation item excepting corporate governance, we construct a news index 1 ($SNI1$) on the basis of (1) and estimate the following regression model.⁸

$$Tpx_t = c + \phi Tpx_{t-1} + \sum_j \gamma_j SNI1_{j,t-1} + \varepsilon_t \quad (5)$$

⁸ Since only negative news is reported on the subject of corporate governance, it was excluded from analysis in this section. Also, due to space limitations we are reporting only results relating to news index 1, but largely the same results were obtained when news index 2 was used.

Assuming (5) to be model 1 of all evaluated items, the second line in **Figure 6** summarizes estimated values. As the results indicate, corporate action and market commentary are positive and significant at a significance level of at least 10%.⁹ When there is a lot of positive news involving corporate action or market commentary, stock prices tend to rise the next business day. In other words, investors find useful information in news concerning corporate action or market commentary, and this apparently has great potential to influence stock prices.

Market commentary is information indicating movements in stock prices such as a sustained rise or a rally, but note that it is not regarded as simply the effect of momentum. Indeed, equation (5) controls the momentum effect by including past TOPIX returns as a regressor and, as seen in **Figure 6**, the effect of momentum is not significant. In addition, many studies have pointed out, and Fama and French [2012] also recently confirmed, that there is no momentum effect in the Japanese stock market. Therefore, in market commentary the Nikkei (reporters) selects company names from all and attaches reasoning for movements in their stock prices, which presumably, like corporate actions, becomes useful information for investors.

As for other evaluation items, there are none that are positive and significant. Recommendations from brokers' analysts were negative and significant at a significance level of 5%, meaning that given positive news involving a recommendation from a broker analyst, stock prices tend to decline the next business day. One may posit that often a broker analyst's recommendation is reflected more than is necessary in the stock price that same day, and on the next business day a rebound is observed. Often, a consensus is formed regarding profits and sales in advance of the flow of news. Therefore, even if the content of news is simply evaluated, it will not be evaluated against the consensus. This could be a reason why news about profits and sales has no significant effect on stock prices. A credit rating is thought to be information that by its nature impacts stock prices, but in this study it has no impact on stock prices the next business day. One reason for this could be that information concerning (potential) changes in rating has previously been fully reflected in stock prices, but we hope to uncover a more definite reason through further study. Finally, the results suggest that news about products and services and macro indexes does not necessarily include new information in relation to stock prices. For example, when we narrow our focus down to news about new products related to health or GDP-related news, there was some

⁹ Market commentary comprises news items with headlines such as “◇<Tokyo Stock Exchange>Taisei Co. stock reaches seven-year high, stocks of companies related with Tokyo Olympics 2020 continue to post gains,” in which the trading status of stocks is circulated with reasons for companies selected by journalists.

potential, based on news categorization, to extract information useful for stock prices, but this exceeded the objective of this study and is left for future research.

In order to confirm the robustness of the above results, we estimate a multi-regression model (model 2) which extracted from explanatory variables only those that were significant, and simple regression models (model 3 through model 5) which use only significant explanatory variables, and the results are summarized in the third through sixth columns in **Figure 6**.¹⁰ As is shown, the results of model 2 through model 5 are completely consistent with the results of model 1, confirming once again that news about corporate actions and market commentary have strong potential to impact stock prices the next business day.

5. Conclusion

In this paper we investigated whether news indexes constructed on the basis of Nikkei news distributed at QUICK terminals had an effect on the future stock market. Specifically, having used tag information to enable quantified processing of news data and the creation of news indexes, we analyzed the predictability of stock returns and trade volumes with the use of those news indexes.

The results of the analysis confirmed that the news indexes had a significant impact on the stock returns of the next business day and that the rebounds reported in previous research were not observed. In addition, this study showed that the news indexes had a significant positive effect on trading volumes the next business day. This was in contrast to the results of Tetlock [2007], in which news indexes were created based on WSJ columns, a similar analysis was performed, and the news indexes contained information about market sentiment, but were not found to have information essential to stock prices. In short, the news indexes used in this study did not express market sentiment, but did suggest the potential of having essential information in relation to stock prices, a finding of great interest.

Furthermore, when we categorized and analyzed individual evaluation items, we observed that news concerning corporate actions or market commentary did contain essential information related to stock prices. In other words, our study demonstrated that investors found useful information in news concerning corporate actions or market commentary, and such news had high potential to impact stock price formation.

¹⁰ The results are not included to save space, but in simple regression models using variables that were not significant, the variables did not become significant in any of the models.

The above findings suggest that the news indexes used in this study potentially include information that can predict stock prices and trading volumes, and it can reasonably be said that the potential of the use of news indexes has been demonstrated. However, the analyses conducted in this study are all in-sample analyses, and it is not certain that these results can be applied to actual investment strategies. In the future we hope to see additional research, such as constructing investment strategies that use news indexes, so that the use of news indexes can continue in more practical applications.

References

Campbell, J. Y., S. J. Grossman, and J. Wang [1993] “Trading Volume and Serial Correlation in Stock Returns,” *Quarterly Journal of Economics* 108, pp. 905-939.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann [1990] “Noise Trader Risk in Financial Markets,” *Journal of Political Economy* 98, pp. 703-738.

Fama, E. F., K. R. French [2012] “Size, Value, and Momentum in International Stock Returns,” *Journal of Financial Economics* 105, pp. 457-472.

Griffin, J. M., N. H. Hirschey, and P. J. Kelly [2011] “How Important is the Financial Media in Global Markets?” *Review of Financial Studies* 24, pp. 3941-3992.

Roll, R. W. [1988] “R-Squared,” *Journal of Finance* 43, pp. 541–566.

Tetlock, P. C. [2007] “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” *Journal of Finance* 62, pp. 1139-1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy [2008] “More than Words: Quantifying Language to Measure Firms Fundamentals,” *Journal of Finance* 63, pp. 1437-1467.

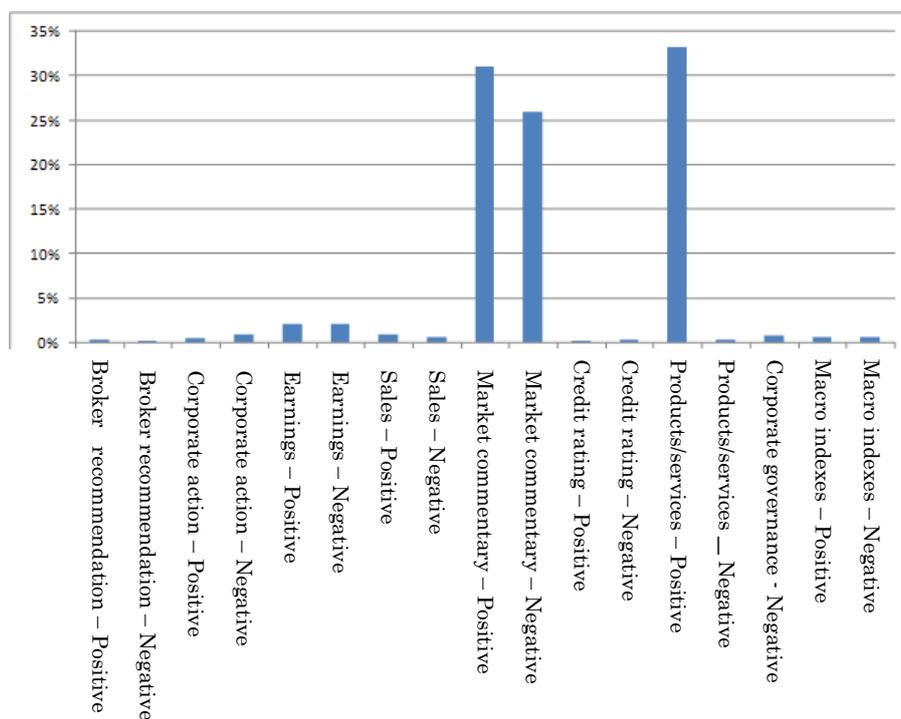
Figure1: Positive and Negative Information, Reasons Identified (Example)

Evaluation Item	Reason	Positive/Negative
Recommendation of broker analyst	Upward revision	Positive
	Downward revision	Negative
Corporate action	Share buyback, div. increase	Positive
	No dividends, div. decrease	Negative
Earnings	Upward revision, increase	Positive
	Downward revision, decrease	Negative
Sales	Increased sales, bullish	Positive
	Decreased sales, slump	Negative
Market commentary	Sustained rise, rebound	Positive
	Continued fall, drop	Negative
Credit rating	Upgrade	Positive
	Downgrade	Negative
Products/services	New services	Positive
	Recall	Negative
Corporate governance	Scandal, fraud	Negative
Macro indexes	TANKAN (Bank of Japan) positives	Positive
	Decreased orders for machinery	Negative

Note: Positive and negative evaluations are tied to the major companies mentioned in each news article. If several reasons are applicable to one news item, they are processed as multiple positive and negative evaluations generated for the major company.

Source: Created by authors. Same for below.

Figure 2: Distribution of Positive and Negative Information



Note: **Figure 2** displays the distribution of positive and negative news, with reasons. Much of it is market commentary communicating fluctuations in stock prices, reasons, etc. for listed companies, and new products and services news introducing products and services. There are also some items that had low frequency but a large impact on stock prices.

Figure 3: Estimated Values for γ_1

	NI1	NI2
γ_{11}	0.167** (0.078)	0.194** (0.084)
γ_{12}	-0.021 (0.068)	-0.016 (0.071)
γ_{13}	0.087 (0.073)	0.098 (0.076)
γ_{14}	-0.032 (0.070)	-0.043 (0.071)
γ_{15}	0.043 0.061	0.044 (0.065)
Adj R ²	0.011	0.014

Note: Newey-West standard errors appear in parentheses. *, **, *** represent significance at the 10%, 5%, and 1% levels respectively.

Figure 4: Estimated Values for β_2

	NI1		NI2	
β_{21}	-0.021	(0.020)	-0.029	(0.021)
β_{22}	-0.028	(0.022)	-0.035	(0.022)
β_{23}	-0.026	(0.018)	-0.031	(0.018)
β_{24}	0.024	(0.018)	0.025	(0.018)
β_{25}	-0.038**	(0.018)	-0.042**	(0.019)
Adj R ²	0.095		0.105	

Note: Newey-West standard errors appear in parentheses. *, **, *** represent significance at the 10%, 5%, and 1% levels respectively.

Figure 5: Estimated Values for γ_3 and ψ_3

	NI1		NI2	
γ_{31}	0.015***	(0.006)	0.013**	(0.006)
γ_{32}	-0.005	(0.006)	-0.003	(0.006)
γ_{33}	0.001	(0.006)	0.004	(0.006)
γ_{34}	-0.002	(0.005)	-0.002	(0.005)
γ_{35}	-0.006	(0.006)	-0.007	(0.006)
ψ_{31}	-0.007	(0.008)	-0.007	(0.008)
ψ_{32}	-0.010	(0.007)	-0.008	(0.007)
ψ_{33}	-0.002	(0.007)	-0.002	(0.008)
ψ_{34}	0.013*	(0.007)	0.013*	(0.007)
ψ_{35}	-0.012	(0.007)	-0.012*	(0.007)
Adj R ²	0.547		0.547	

Note: Newey-West standard errors appear in parentheses. *, **, *** represent significance at the 10%, 5%, and 1% levels respectively.

Figure 6: Analysis of Individual Evaluation Items

	Model 1	Model 2	Model 3	Model 4	Model 5
C	-0.083	-0.085*	-0.080	-0.081*	-0.086*
ϕ	-0.063	-0.061	0.006	0.001	-0.064
BR	-0.095**	-0.096**	-0.084*		
CA	0.093*	0.101*		0.109**	
ER	0.011				
SL	0.040				
MC	0.170*	0.170*			0.170*
CR	0.003				
PS	-0.054				
MI	-0.040				
Adj R ²	0.013	0.012	0.002	0.004	0.005

Note: *, **, *** represent significance at the 10%, 5%, and 1% levels respectively.

Key to Abbreviations:

BR : Broker analyst company recommendation

CA : Corporate action

ER : Earnings (net income)

SL : Sales

MC : Market commentary

CR : Credit rating

PS : Products/services

MI : Macro indicator