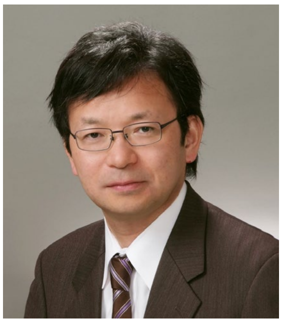
Analysis of the relationship between corporate organizational culture and financial performance using company employee reviews

Hironori Nishiie / Hiroshi Tsuda

Hironori Nishiie (CMA・CIIA)
Senior Consultant, Credit Pricing Corporation, concurrently undertaking PhD at Yokohama National University. Graduated from the Department of Science and Engineering, Tokyo University of Science, 2009; obtained Master's degree in Science, Tokyo University, 2011. Joined NTT Data Financial Solutions in 2011. In current position since 2014.



Hiroshi Tsuda (CMA)
Professor, Department of Mathematical Sciences, Doshisha University, PhD.
Delegate Past President, Japanese Association of Financial Econometrics and Engineering.
Visiting Professor, Institute of Statistical Mathematics.
Part-time lecturer, Graduate School of Medicine, Kyoto University.
PhD, Graduate University for Advanced Studies.
Master's degree in Engineering, Tokyo University.
Bachelor's degree in Engineering, Kyoto University.

Analysis of the relationship between corporate organizational culture and financial performance using company employee reviews

Hironori Nishiie / Hiroshi Tsuda

Abstract

This paper analyzes the relationship between "Corporate Organizational Culture Score", which is quantified by text mining and machine learning with respect to online employee reviews of Japanese listed companies, and the financial/equity performance of firms. We find (1) sales decrease with a low score, (2) the debt ratio increases with a worsening score, and (3) the long-short portfolio, constructed using the group showing improvement and that aggravation (as defined by score change), has statistically significant positive alpha(α), and, in particular, the aggravation group portfolio has negative alpha(α) measured by Fama-French three or five factor models.

Our research suggests that online company reviews contain useful information for the purpose of corporate valuation.

1. Introduction

The widespread availability of the Internet allows individuals to publish online reviews of various services. Prominent examples in Japan include tabelog[1] and kakaku.com[2], which review, respectively, restaurants and primarily household appliances. Visitors to these websites are able to obtain original information and opinions on restaurants and manufacturers, in addition to information from the companies themselves. Such immediate user feedback can be seen as a proxy for the market reputation of services and products .

Employees also review their employers using various evaluation criteria on company review websites, and visitors use these reviews when searching for jobs. In this study, we presume that employee reviews serve as a proxy for a company's corporate culture, and analyze their relationship to corporate finance and stock performance by using scores about public companies' corporate culture quantified reviews by text mining.

Some studies that examined the relationship between employer review sites and corporate performance used review data from US company Glassdoor[3]. Luo [9] argues that a positive relationship exists between Glassdoor's "overall rating" and corporate performance. Symitsi et al. [10], who used Glassdoor reviews in a panel regression that was adjusted for various attributes and self-correlation, demonstrated a positive relationship between return on assets (ROA) and employee satisfaction, which the authors also measured, as well as the possibility of significant excess returns in companies with high levels of satisfaction, based on an analysis using a factor model. Ji [6], who also used Glassdoor review data, showed that a higher risk of lawsuits relating to companies' legal disclosures correlated with low levels of employee satisfaction and poor corporate culture.

One major difference between the present study and prior research from the US is that the latter employed aggregated, five-stage Glassdoor ratings, whereas we reviewed raw text and employed an analysis based on text mining and machine learning.

This paper is organized as follows. First, we explain the research data and present the analysis and quantification methodology that was applied to the reviews. We then offer our conclusions regarding the relationships between corporate culture scores, a metric that quantifies reviews, and corporate financial performance.

2. Review Data

This study makes use of 69,707 reviews posted between July 2007 and November 2017 to the "organization and corporate culture" category of OpenWork, an employer review site for job seekers run by OpenWork. OpenWork, launched in July 2007, is Japan's largest employer review site, with 1.9

1 http://tabelog.com/
2 http://kakaku.com/
3 https://www.glassdoor.com/index.htm

million users and 4.9 million reviews and ratings as of November 2017. OpenWork users post ten quantitative ratings and ten corresponding reviews about their employer companies, including under the "organization and corporate culture" category[4]. OpenWork controls the quality of posted reviews manually and in a systematic manner. Statistics for "organization and corporate culture" reviews used in this study are shown in Table 1 and Appendix 9.

Table 1 Review Data Statistics

		Attributes of Contributors				TSE 1st and 2nd Section (Threshold by Posts)			
Year[5] (See below)	Cumulative posts	Percentage employed	Percentage male	Percentage new graduates	Average years in the job	1 or more	5 or more	10 or more	15 or more
2007	191	52.9%	85.9%	71.7%	5.0	92	7	2	1
2008	761	48.8%	81.5%	70.0%	5.0	267	34	17	10
2009	2,406	45.1%	80.0%	67.7%	5.6	578	107	45	33
2010	5,203	45.9%	80.1%	65.8%	5.7	836	229	108	70
2011	7,978	46.5%	80.8%	66.6%	6.0	1,100	330	175	102
2012	11,145	48.2%	81.4%	67.1%	6.3	1,290	444	246	151
2013	16,542	48.6%	81.4%	66.7%	6.6	1,518	638	348	240
2014	24,387	48.7%	81.3%	66.1%	6.8	1,730	851	499	342
2015	34,168	48.8%	81.5%	65.6%	6.8	1,987	1,070	689	474
2016	49,069	49.8%	80.8%	65.4%	6.9	2,228	1,333	893	660
2017	69,707	50.8%	80.0%	65.1%	6.9	2,442	1,582	1,127	873

3. Generating the Corporate Culture Score

Data review information consists of text-based information from contributors, which we divided by punctuation mark and converted periodically into sentence-level information. Subsequently, we used a sentiment analysis model to assign a positivity probability to each sentence, combined these probabilities for each contributor, and then aggregated them by company to calculate a "corporate culture score". This is how we quantified the review information.

3.1 Sentiment Analysis Model

A sentiment analysis model is designed to distinguish between sentences that exhibit positive and negative sentiments. Although pre-trained models are generally available, we built a custom model that joined two types of learning data. The first was created by randomly extracting 20,000 sentences from the review data and manually flagging each as either positive, negative, or neutral, based on a

[4] Appendices 7 and 8 show OpenWork's ratings and review categories. Ratings are based on a five-stage scale, and reviews are free-form textual information.
[5] This data is as of 30 November 2017.

reading of the sentence by numerous individuals to eliminate individual bias. After removing "neutral" sentences, 13,509 valid entries remained. The second type of data treated reviews under the "company's strengths and weaknesses" category of OpenWork's review data as "positive" for sentences under "strengths" and "negative" for those under "weaknesses". We randomly selected the same number (13,509) of entries from this data. This yielded 27,018 learning data entries[6].

Features (explanatory variables) in the discriminant models of Japanese text information are generally taken from data whose input is divided into word-like segments. We used a model segmented into characters because Zhang [11] and Dos [5] have shown that such a model displays a high degree of accuracy.

We adopted a categorization model called fastText [8][4], developed by Facebook AI Research, as the sentiment analysis model's algorithm[7]. We tested the model by using training data and then evaluated its positive/negative categorization performance by using test data, which included 73,885 sentences under the "company's strengths and weaknesses" category. To evaluate accuracy, we used the AUC (area under the curve) approach commonly applied to reliability risk models.[8] The evaluation, through the use of our test data, yielded an AUC of 0.906, which we determined to be good.

We assigned a positivity probability in the range [0,1] to the 280,720 review sentences from "organization and corporate culture", not included in the test data[9].

3.2 Calculating the Corporate Culture Score

By using a positivity probability assigned to each sentence, we calculated the corporate culture score $VC_c(T)$ for company c at month-end T. A positivity probability for sentence s, from poster p, about company c, for posts made at time $t \leq T$, was denoted as $X_{c,p,s} \in [0,1]$, where $c = \{1,2,\cdots,n_c | t \leq T\}$, and $p = \{1,2,\cdots,n_{p|c} | t \leq T\}$, $s = \{1,2,\cdots,n_{s|p,c} | t \leq T\}$. We calculated the average value $\overline{X}_{c,p}(T)$ of a positivity probability from poster p, about company c, as follows:

$$\overline{X}_{c,p}(T) = \frac{1}{n_{s|p,c}} \sum_{s} X_{c,p,s} \tag{1}$$

[6] A single post can have more than one sentence. The number of sentences in the "organization and corporate culture" review data to which we applied the sentiment analysis model was 280,720.

[7] https://github.com/facebookresearch/fastText

[8] AUC takes a value between zero and one. A random categorization would yield an AUC of 0.5. The higher the AUC value, the greater the discrimination.

[9] Appendix 3 gives a word co-occurrence network, with words included in sentences from learning reviews that lean toward the positive. "Positive", in this model, corresponds to statements such as "can grow", "easy to work with", and "good company".

Subsequently, we estimated the population distribution of $\overline{X}_{c,p}(T)$ for all companies and all contributors, where $t \leq T$[10]. For population distribution, we assumed a three-valued, multivariate, normal distribution with adjusted initial values and constraints, so that the expected value for each distribution is either (1) positive (near 1), (2) neutral (near 0.5), or (3) negative (near 0). We then estimated value by using the EM algorithm. We denoted each normal distribution within the three-valued, multivariable distribution as $P_i(\overline{X}_{c,p})$ and the mix ratio as w_i, where $i \in \{1,2,3\}$. Finally, we calculated the corporate culture score $VC_c(T)$ by company. By using a Bayesian estimation approach, we revised the average $\mu_i(T)$ and the variance $\sigma_i^2(T)$ for each normal distribution $P_i\left(\overline{X}_{c,p}\right)$, based on the set of average positivity probabilities of posters for company c[11]. By using the average $\mu_{i,c}(T)$ of each company c's normal distribution, we solved for its corporate culture score $VC_c(T)$ as follows:

$$VC_c(T) = \sum_i w_i \, \mu_{i,c}(T) \tag{2}$$

We calculated each company's corporate culture score at month-end by sliding T every month.

3.3 Bias in Review Data

Reviews gathered from the Internet include a number of biases. First and foremost, the number of contributors differ widely from company to company[12]. When considering how to aggregate positivity probabilities for quantification, the law of large numbers means that a simple, average calculation will result in different variances of aggregate values, depending on whether the company has a large or small number of contributors. The larger the company, the closer the aggregates are. Second, the attributes of contributors vary from company to company, and reviews are not uniform—they differ by industry, form of employment, years of experience, and gender. Taking these two issues into account, our aggregation methodology estimated a distribution of positivity probabilities for all reviews at a particular point in time, which was treated as prior information to correct for differences in the number of contributors per company and their attributes[13].

Third, since reviews are obviously limited to contributors who are motivated to post, they do not represent a comprehensive set of opinions about a company. This bias is an unavoidable issue for data collected passively over the Internet, and it is a topic for future research. Fourth, reviews can be forged or fraudulent. To ensure the quality of reviews, OpenWork removes all such posts by using both

[10] See Appendix 5 for the distribution of average positivity probabilities, by poster.
[11] The Bayes revision formula for the parameters of a normal distribution follows Tango [1].
[12] In general, larger companies have more reviewers and smaller ones fewer.
[13] Appendix 4 shows the relationship between number of contributors and corporate culture score.

system-based and human-eye checks. Nevertheless, although this study only makes use of reviews that passed these checks, the data may include some fraudulent or otherwise problematic posts.

4. Corporate Culture Score and Corporate Performance

We now consider the relationship between each company's corporate culture score, obtained from a sentiment analysis of the reviews of OpenWork's "organization and corporate culture" and corporate performance. We denote the corporate culture score of company i at the end of month t as $VC_i(t)$. Our analysis only included companies listed on the 1st or 2nd sections of the Tokyo Stock Exchange (TSE), excluding financial entities[14], with a minimum of 15 contributors at each month-end, from January 2010 through November 2017 (see Table 1 and Appendix 9). For financial, stock price, and valuation data, we employed "Securities Reports Data" and "Monthly Adjusted Stock Price and Valuation Data" in Japanese from Toyo Keizai, Inc.

4.1 Corporate Culture Score and Corporate Finance

We define a metric for annual change rates in corporate culture score to represent improvement or deterioration in corporate culture scores, as follows:

$$RVC_i(t) = \frac{VC_i(t) - VC_i(t-12)}{VC_i(t-12)} \tag{3}$$

To understand whether the level of the corporate culture score is good or poor, we considered the previous year's score, $LVC_i(t) = VC_i(t-12)$, that defined how the corporate culture was and how it had changed from. Since factors such as sector and size make it impossible to directly compare corporate finance, we followed Barber [3] and Yamada [2], who monitored each company against a reference portfolio (a group of similar companies)[15,16]. For companies whose fiscal years ended in March, June, September, and December (between 2010 and 2016), we constructed a 3 × 3, nine-quantile portfolio with $LVC_i(t)$ and $RVC_i(t)$. Each cell compared the following period's financial metric $Y_i(t+12)$ for company *i* and its expected financial metric $E[Y_i(t+12)] = Y_i(t) + [\widehat{Y}_i\,(t+12) - \widehat{Y}_i\,(t)]$ from the median financial metric for the reference portfolio $\widehat{Y}_i\,(t+12)$ and $\widehat{Y}_i\,(t)$, and performed a Wilcoxon signed-rank test as a non-parametric test, to see if a significant

[14] Banking, insurance, brokerages, and other financial industries.

[15] For the sake of brevity, we provide the specific methodology for constructing reference portfolios in the Appendix. Essentially, we used a group of companies in the same sector, with similar financial profiles, for the reference portfolio.

[16] There is no guarantee that this approach can completely eliminate the effect of other control variables, given the myriad factors affecting each company's performance.

difference against the expected financial metric existed[17]. Table 2 shows the results by evaluation metric. In each cell, the values in the Table are the medians of $Y_i(t+12) - E[Y_i(t+12)]$.

Table 2 Period-Ratio Medians for Similar Companies Based on a 3 × 3 Portfolio of Corporate Culture Scores

				Poor	Previous-Period Score	Good
Metric	Company	Change in Score		LV01	LV02	LV03
ROE	1,140	Worse	RV01	0.098%	0.424% .	0.132%
			RV02	0.847% *	0.664% **	0.378%
		Better	RV03	0.333%	0.303%	0.201%
Net profit on sales	1,140	Worse	RV01	0.243%	0.198%	0.239%
			RV02	0.163%	0.455% **	0.073%
		Better	RV03	0.052%	0.352% *	0.002%
Asset turnover	1,140	Worse	RV01	−0.769%	−0.552%	−0.166%
			RV02	−0.740%	−0.545%	−0.154%
		Better	RV03	−0.658%	0.615%	−0.304%
Financial leverage	1,140	Worse	RV01	−0.016	−0.022	0.013 *
			RV02	−0.012	−0.011	0.009 *
		Better	RV03	−0.004	0.002	−0.006
ROA	1,140	Worse	RV01	0.110%	0.335% .	0.090%
			RV02	0.091%	0.293% *	0.248%
		Better	RV03	0.161%	0.253% .	−0.021%
Change in sales	1,140	Worse	RV01	−1.571% **	−0.790%	−0.511%
			RV02	−1.743% *	−0.504%	0.402%
		Better	RV03	−0.944%	1.487% *	0.071%
Debt ratio	1,140	Worse	RV01	−1.285%	−1.815%	3.475% ***
			RV02	1.556%	−1.731%	1.391% **
		Better	RV03	−0.218%	2.274% .	−1.024%

Table 2 shows that the debt-to-equity ratio is significantly positive when previous years' scores are good. Although Yamada [2] found similar results, the debt-to-equity ratio here tends to grow in companies whose good scores deteriorated. It is likely that sharp drops in corporate culture had an impact on future financial risk. In addition, when a poor previous year score declined further, it was accompanied by a significant deterioration in sales. This suggests that a poor corporate culture can inhibit future growth.

4.2 Corporate Culture Score and Stock Performance

[17] *** represents p-value of 0.001, ** 0.01, * 0.05, . 0.1.

4.2.1 Fama-MacBeth Cross-Sectional Regression

The Fama-MacBeth regression [7] measures whether annual change rates in a corporate culture score affect the following month's stock return. The following month's return is the dependent variable, and the annual change rate in corporate culture score is the explanatory variable. The control variables are market value (SIZ); the price-book value ratio (PBR); 3-, 6-, and 12-month momentum (MOM3, MOM6, and MOM12); and 12-month volatility (VOL12). We performed monthly cross-sectional regressions and evaluated the estimated cumulative sum and time-series average of the coefficients each month. Further, we used the latest available financial data recognized for the month and normalized the explanatory variables to have a mean of zero and a variance of one for the same month.

Figure 1 presents the cumulative sum of the monthly coefficients, estimated by cross-sectional regressions. The coefficients of annual change rate in corporate culture score, as represented by *RVC*, move steadily in a positive direction. Table 3 shows the results of the Fama-MacBeth regression analysis when control variables are changed. Model (5) derives from Figure 1. Regarding the significance of the coefficients, SIZ (SIZE of company), which represents market value, is significant, and RVC is highly significant. The information in RVC differs from that in the control variables.

Figure 1 Cumulative Sum of Monthly Coefficients

(Vertical axis label: Cumulative sum of coefficients. Horizontal axis label: Date)

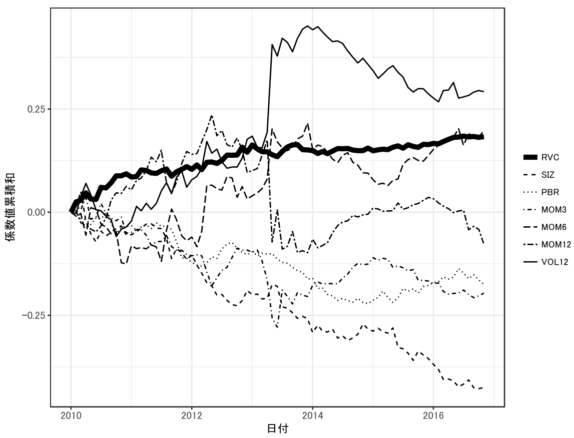


Table 3 Results of a Fama-MacBeth Regression Analysis

	Model									
	(1)		(2)		(3)		(4)		(5)	
Constant	0.014	*	0.014	*	0.014	*	0.014	*	0.014	*
t-value	2.351		2.349		2.360		2.368		2.369	
RVC	0.002	.	0.002	*	0.002	.	0.001		0.002	**
t-value	1.717		2.052		1.860		1.556		2.656	
SIZ	−0.006	**	−0.006	*	−0.006	**	−0.006	**	−0.005	**
t-value	−2.716		−2.624		−2.758		−2.938		−2.644	
PBR	−0.002	.	−0.003	.	−0.001		−0.002	.	−0.002	
t-value	−1.671		−1.957		−1.065		−1.729		−1.645	
MOM3			0.000		−0.002		−0.002		−0.002	
t-value			0.015		−0.508		−0.618		−0.962	
MOM6					0.002		0.003		0.002	
t-value					0.499		0.658		0.716	
MOM12							−0.000		−0.001	
t-value							−0.041		−0.254	
VOL12									0.004	
t-value									1.008	

4.2.2 Quantile Portfolios of Annual Change Rates and Evaluation Using a Fama-French Factor Model

Subsequently, we constructed a five-quantile portfolio index based on the annual change rate in corporate culture score $RVC_i(t)$ at the end of August every year and held for one year in equal amounts. We managed this portfolio monthly from August 2010 through November 2017. Figure 2 shows the five-quantile portfolio index and the difference between the portfolio with the highest (RV05) and lowest (RV01) $RVC_i(t)$. The difference between RV05 and RV01 widens over time. Tables 5 and 6 include the analysis results for the portfolio of each quantile for Fama-French three- and five-factor models[18]. It can be seen that the RV05 and RV01 long/short portfolios show a significant α, indicating positive excess return. In addition, the portfolios with poor annual change rates in corporate culture score, represented by RV01, exhibit large negative excess return.

[18] We employed TOPIX returns for market factors. Other factors were obtained from Ken French's web site at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Figure 2　Five-Quantile Portfolios, Based on Annual Change Rates in Corporate Culture Scores

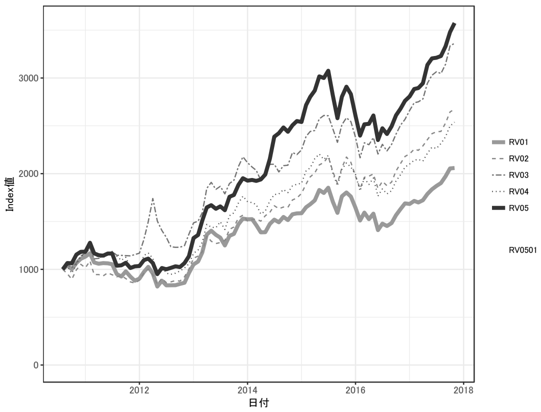


Table 4　Number of Issues in Each Quantile Portfolio

Month	RV01	RV02	RV03	RV04	RV05
2010/08	11	10	11	10	11
2011/08	15	14	14	14	14
2012/08	22	22	21	22	22
2013/08	35	34	35	34	35
2014/08	55	54	53	53	54
2015/08	72	71	71	71	72
2016/08	108	106	107	107	106
2017/08	145	144	145	144	145

Table 5　Fama-French Three-Factor Model Results

	RV01	RV02	RV03	RV04	RV05	RV05-RV01
α (YoY %)	−5.146 *	−0.491	2.474	−3.018	3.715	9.299 **
t-value	−2.035	−0.172	0.527	−1.221	1.296	2.660
MKT	1.083 ***	0.956 ***	0.897 ***	1.051 ***	0.996 ***	−0.087
t-value	23.750	19.000	10.981	23.827	20.074	−1.469
SMB	0.269 *	0.359 **	0.497 **	0.458 ***	0.216 .	−0.052
t-value	2.569	3.106	2.649	4.524	1.902	−0.385
HML	−0.212 *	−0.018	−0.222	−0.134	−0.318 **	−0.106
t-value	−2.233	−0.175	−1.305	−1.456	−3.078	−0.862
Adj. R2	0.873	0.815	0.579	0.871	0.830	0.007

Table 6　Fama-French Five-Factor Model Results

	RV01	RV02	RV03	RV04	RV05	RV05-RV01
α (YoY %)	−5.146 *	−0.491	2.474	−3.018	3.715	9.299 **
t-value	−2.035	−0.172	0.527	−1.221	1.296	2.660
MKT	1.083 ***	0.956 ***	0.897 ***	1.051 ***	0.996 ***	−0.087
t-value	23.750	19.000	10.981	23.827	20.074	−1.469
SMB	0.269 *	0.359 **	0.497 **	0.458 ***	0.216 .	−0.052
t-value	2.569	3.106	2.649	4.524	1.902	−0.385
HML	−0.212 *	−0.018	−0.222	−0.134	−0.318 **	−0.106
t-value	−2.233	−0.175	−1.305	−1.456	−3.078	−0.862
Adj. R2	0.873	0.815	0.579	0.871	0.830	0.007

5. Summary and Topics for Future Research

This study used text-based reviews of "organization and corporate culture", posted to the OpenWork company review site, as a proxy for a company's corporate culture. The investigation employed text mining and machine-learning technologies to calculate metrics for corporate culture scores and performed a comparative analysis with companies' corporate performance. To our knowledge, this is the first study to comprehensively analyze the relationship between text-based employee reviews and corporate performance. One distinguishing feature of our study is the use of time-series metrics for corporate culture scores, which enables a more detailed examination of either improvement or deterioration in a corporate culture. A deteriorating corporate culture can cause an increase in debt ratios, and, if further deterioration is likely, will reduce revenues. An analysis of stock performance, by using annual change rates in corporate culture scores, measured via Fama-MacBeth cross-sectional regressions, showed a significantly positive alpha from such changes. A statistically significant positive excess return on long/short portfolio of improvement/deterioration corporate culture occurred.This suggests that employee reviews contain important information not currently being captured. Although some issues remain, first it is necessary to establish better methods to compensate for biases, such as the number of reviewers and their attributes, motivation, and reliability. Second, since the market climbed steadily during our short analysis period of 2010-17, a more extended analysis of corporate performance is necessary.

References

[1] T. Tango, T. Becque, "Actual Bayesian Statistical Analysis (in Japanese)", Asakura Bookstore, 10 2011.

[2] T. Yamada, T. Usui, S. Goto, "Employee Satisfaction and Firm Performance (in Japanese)", Securities Analysts Journal, Vol.55, No.11, pp. 75-86, 2017.

[3] B. M. Barber, J. D. Lyon, "Detecting Abnormal Operating Performance: The empirical power and specification of test statistics", Journal of Financial Economics,Vol.41, No.3, pp. 359-399, 1996.

[4] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, "Enriching Word Vectors with Subword Information", arXiv preprint arXiv:1607.04606, 2016.

[5] C. N. dos Santos, M. Gatti, "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts", COLING pp. 69-78, 2014.

[6] Y. Ji, O. Rozenbaum, K. T. Welch, "Corporate Culture and Financial Reporting Risk: Looking through the glassdoor", Available at SSRN 2945745, 2017.

[7] E. F. Fama, J. D. MacBeth, "Risk, Return, and Equilibrium: Empirical tests", Journal of

Political Economy Vol.81, No.3, pp. 607-636, 1973.

[8] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, “Bag of Tricks for Efficient Text Classification”, arXiv preprint arXiv:1607.01759, 2016.

[9] N. Luo, Y. Zhou, J. Shon, “Employee Satisfaction and Corporate Performance: Mining employee reviews on glassdoor.com”,AIS eLibrary icis2016, 2016.

[10] E. Symitsi, P. Stamolampros, G. Daskalakis, “Employees’ Online Reviews and Equity Prices”, Economics Letters Vol.162, pp. 53-55, 2018.

[11] X. Zhang, J. Zhao, Y. LeCun, “Character-level Convolutional Networks for Text Classification”, Advances in neural information processing systems pp. 649-657, 2015.

Appendix 1

Ten quantitative ratings and ten textual categories on OpenWork

Table 7　Quantitative Ratings

No	Rating category	Answer
1	Satisfaction of treatment	5 grades rating
2	Goodness of ventilation	5 grades rating
3	20's growth environment	5 grades rating
4	Legal compliance awareness	5 grades rating
5	Employee morale	5 grades rating
6	Mutual respect for employees	5 grades rating
7	Long-term human resource development	5 grades rating
8	Appropriateness of personnel evaluation	5 grades rating
9	Average overtime hours	overtime hours
10	Average paid digestion rate	paid digestion rate

Table 8　Textual Categories

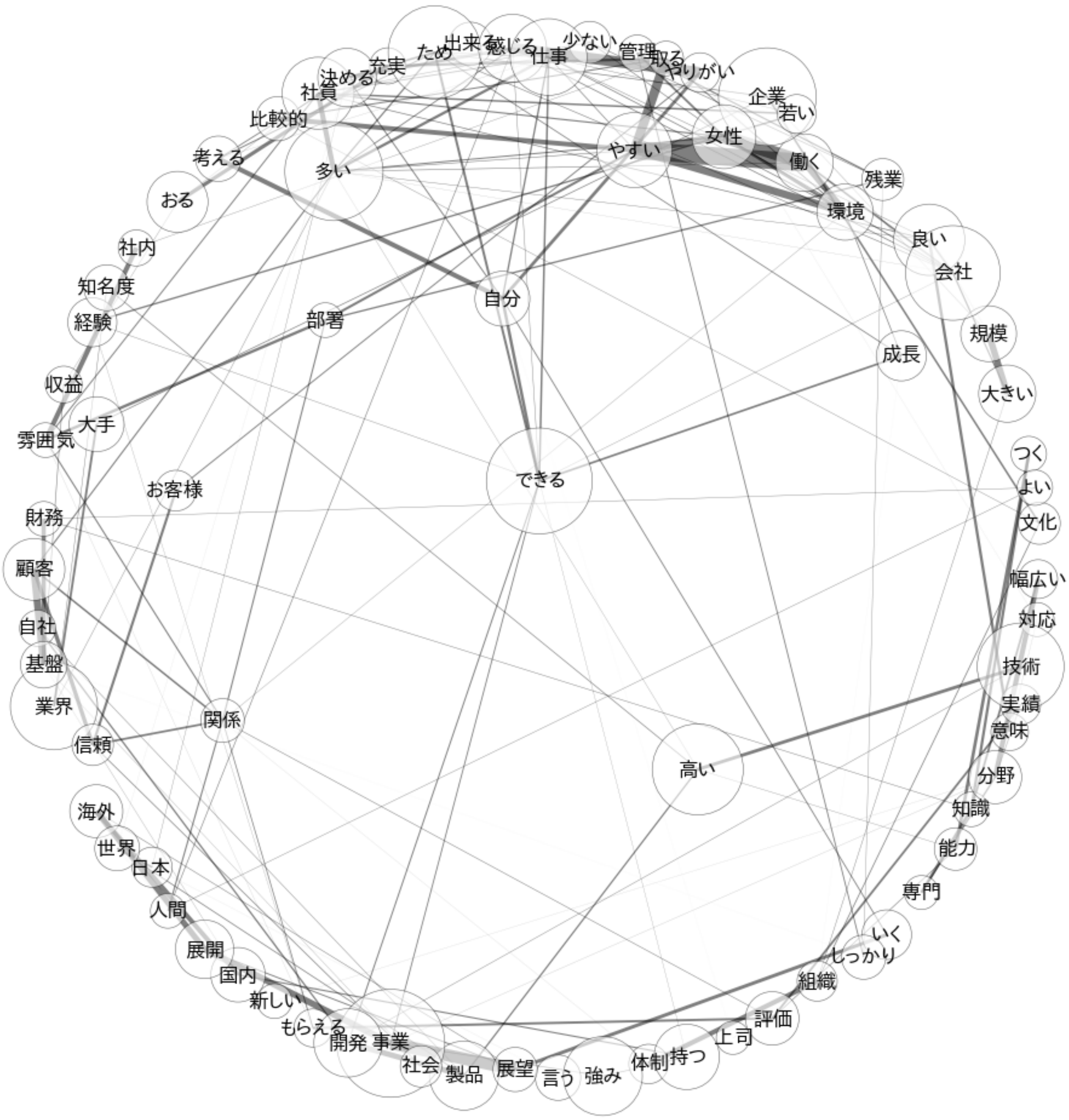
No	Textual category
1	Organizational and corporate culture
2	Annual salary and system
3	Reasons for joining and gaps
4	Motivation and evaluation system
5	Ease of working for women
6	Work life balance
7	Growth and career development
8	Reason for retirement
9	Strengths, weaknesses, and prospects
10	Recommendations for management

Appendix 2

Co-occurrence network of words in learning data

Below we give a co-occurrence network of words in positive learning data. This figure indicates that the link between two words is thick when they appear easily in the same sentence, and the node is large when the word appears frequently.

Figure 3 Co-occurrence Network in Positive Learning Data



Appendix 3

Organizational culture score and distribution of average positive sentiment probability

We show a scatter plot between log number of reporting and organizational culture score for each company, and distribution of average positive sentiment probability for each reviewee.

Figure 4 Log Numbers of Reporting and Organizational Culture Score

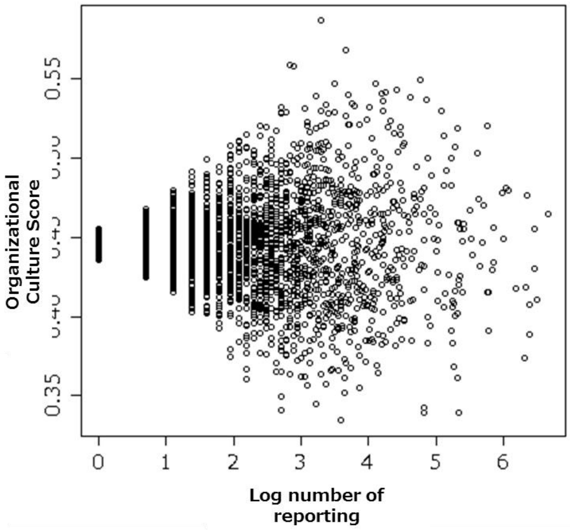
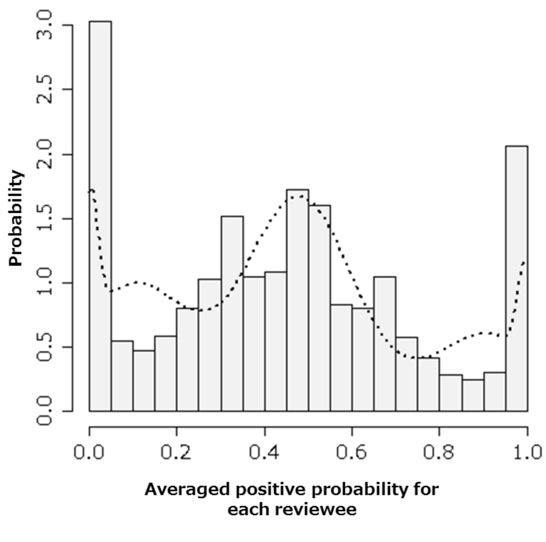


Figure 5 Distribution of Average Positive Sentiment Probability



Appendix 4

Market capitalization coverage of review data

Table 9　Review Data Statistics

	TSE 1st and 2nd Section Coverage				TSE 1st and 2nd Section Market Cap. Coverage			
Year	1 or more	5 or more	10 or more	15 or more	1 or more	5 or more	10 or more	15 or more
2007	3.2%	0.2%	0.1%	0.0%	32.0%	5.8%	2.5%	2.3%
2008	9.3%	1.2%	0.6%	0.3%	46.1%	12.1%	6.7%	5.6%
2009	20.7%	3.8%	1.6%	1.2%	60.9%	35.5%	20.2%	12.9%
2010	32.2%	8.8%	4.2%	2.7%	71.9%	48.5%	34.2%	29.1%
2011	42.6%	12.8%	6.8%	3.9%	78.6%	55.0%	42.3%	33.2%
2012	49.9%	17.2%	9.5%	5.8%	83.2%	64.9%	52.8%	42.3%
2013	58.0%	24.4%	13.3%	9.2%	85.9%	72.6%	61.4%	54.7%
2014	65.8%	32.4%	19.0%	13.0%	88.2%	77.6%	68.5%	62.1%
2015	71.6%	38.5%	24.8%	17.1%	90.2%	80.3%	73.5%	66.0%
2016	78.4%	46.9%	31.4%	23.2%	92.2%	83.6%	77.2%	71.4%
2017	84.7%	54.9%	39.1%	30.3%	94.5%	87.2%	82.6%	78.3%

Table 10 Review Data Statistics

	1 or more			5 or more			15 or more		
業種	2010	2014	2017	2010	2014	2017	2010	2014	2017
ガラス・土石製品	40.7%	92.7%	97.2%	5.8%	80.1%	88.3%	0.0%	25.2%	58.9%
ゴム製品	76.4%	96.6%	99.0%	65.7%	94.0%	95.3%	63.3%	75.7%	90.6%
サービス業	61.2%	88.4%	97.1%	48.7%	70.8%	89.7%	22.1%	62.4%	63.9%
その他金融業	78.9%	88.6%	98.1%	60.1%	69.9%	70.1%	32.1%	55.2%	67.7%
その他製品	41.8%	92.3%	95.8%	23.6%	81.7%	91.9%	18.9%	41.2%	84.1%
パルプ・紙	21.8%	47.1%	95.4%	6.4%	37.5%	85.8%	0.0%	11.4%	36.9%
医薬品	80.2%	89.1%	87.4%	23.5%	80.5%	86.4%	0.0%	72.2%	82.1%
卸売業	85.0%	94.7%	97.4%	72.2%	85.6%	91.6%	47.9%	69.2%	81.8%
化学	47.9%	81.9%	93.8%	14.1%	53.3%	84.6%	0.0%	33.9%	69.1%
海運業	38.3%	86.2%	93.1%	0.0%	66.0%	89.6%	0.0%	0.0%	52.9%
機械	66.1%	92.2%	97.6%	33.9%	80.2%	87.7%	9.5%	55.1%	77.1%
金属製品	13.7%	35.9%	85.9%	0.0%	23.8%	33.4%	0.0%	0.0%	18.8%
銀行業	82.4%	93.9%	96.3%	68.2%	87.4%	94.3%	62.1%	74.7%	89.6%
空運業	2.3%	55.2%	48.9%	1.9%	54.8%	48.9%	0.0%	54.8%	48.0%
建設業	73.6%	92.5%	98.0%	44.1%	76.0%	89.6%	8.6%	59.5%	79.7%
鉱業	84.3%	93.8%	96.1%	0.0%	92.3%	93.1%	0.0%	0.0%	91.2%
小売業	60.4%	83.1%	96.9%	48.5%	73.0%	83.9%	14.3%	53.7%	72.2%
証券、商品先物取引業	18.3%	49.6%	97.5%	3.4%	44.1%	91.4%	0.0%	9.2%	36.3%
情報・通信業	89.0%	94.2%	95.1%	86.7%	90.7%	93.0%	50.7%	87.5%	89.7%
食料品	51.7%	74.9%	92.6%	1.2%	53.7%	61.5%	0.0%	42.5%	53.3%
水産・農林業	0.0%	97.9%	100.0%	0.0%	0.0%	63.1%	0.0%	0.0%	53.2%
精密機器	89.5%	95.0%	96.6%	75.9%	83.1%	93.0%	0.0%	67.2%	86.5%
石油・石炭製品	79.2%	78.0%	98.9%	58.4%	73.1%	95.1%	0.0%	68.0%	93.9%
繊維製品	58.2%	91.0%	100.0%	37.7%	71.2%	93.4%	0.0%	57.1%	69.0%
倉庫・運輸関連業	59.0%	85.6%	96.1%	0.0%	69.9%	88.8%	0.0%	31.9%	68.8%
鉄鋼	42.7%	65.1%	74.7%	8.5%	56.0%	58.4%	0.0%	55.6%	54.6%
電気・ガス業	85.2%	98.5%	99.6%	24.7%	79.1%	96.8%	0.0%	43.7%	90.2%
電気機器	87.1%	98.4%	99.4%	66.0%	91.0%	97.3%	54.1%	74.7%	94.1%
非鉄金属	58.2%	92.6%	94.5%	24.5%	73.6%	90.4%	0.0%	43.4%	76.8%
不動産業	71.5%	80.9%	90.7%	13.0%	72.3%	74.6%	0.4%	14.8%	63.4%
保険業	18.7%	21.7%	36.5%	18.7%	21.7%	30.0%	0.0%	21.7%	29.9%
輸送用機器	91.0%	98.0%	99.8%	80.0%	93.0%	99.0%	60.5%	90.2%	94.2%
陸運業	61.4%	80.9%	95.0%	46.2%	65.1%	80.0%	24.6%	48.3%	67.5%

Appendix 5

Method for building reference portfolio

This study examines the method for building a reference portfolio for Company *i* at a certain point in time for the settlement of accounts *d*. In Barber [3], when analyzing the impact of an event during the period from a certain point in time for the settlement of accounts *d* until the following settlement of accounts *d + 1*, companies within the same industry and those that are within *90%* to *110%* of the financial indices to be evaluated for Company *i* at the time of settlement of accounts are adopted for the reference portfolio for Company *i*. Further, this study also refers to Yamada [2], and the reference portfolio is built as follows.

First, the financial index to be evaluated *Y* is determined. Later, companies that have undergone settlements of accounts within one year prior to the settlement period *d* for Company *i* are selected from the group of companies listed on the 1st and 2nd sections of the Tokyo Stock Exchange.

1. Of the companies that are classified as being under the small industry group on the Tokyo Stock Exchange, select those whose financial indices are within *90%* to *110%* of the financial index *Yi(d)* to be evaluated for Company *i*.
2. If less than three companies are selected under the first condition, we select companies that are classified as falling under the mid-sized industry group on the Tokyo Stock Exchange in accordance with the criterion mentioned in the first condition.
3. If less than three companies are selected under the second condition, we select companies that are classified as being under the large industry group on the Tokyo Stock Exchange in accordance with the criterion mentioned in the first condition.
4. If fewer than three companies are selected under the third condition, we exclude the industry type condition and select three companies in order from those with the least errors in their absolute values for the financial index *Yi(d)* to be evaluated for Company *i*.

As a result of this selection using the above mentioned method, the reference portfolio was built with a composition ratio of *78%* for the first condition, *5%* for the second condition, and *9%* for the third condition.

Barber [3] asserts that rather than searching for similar companies based on company size, it is better to search for companies with similar past performance. Further, this study also adopts this method. However, as factors affecting performance vary, it is important to note that the impact of other control variables cannot be completely eliminated by using this method.