# Text Based Categorization of Governance Styles of Firms and The Validity: Evidence from MD&A of Japanese Firms

Ying Chen [1, @], Kotaro Inoue [1, *, @], Kimura Yosuke [1, @]

**1 :** Tokyo Institute of Technology [Tokyo]

* : Corresponding author

Authors' Contact Information
Chen Ying: chen.y.bh@m.titech.ac.jp
Kotaro Inoue: inoue.k.aq@m.titech.ac.jp
Kimura Yosuke: kimura.y.bq@m.titech.ac.jp

# 1. Introduction

Stakeholder concerns vary across firms. There have been disputes about why companies exist and strive to operate for a long time. From the beginning, shareholder interest maximization has been the main goal of corporate management. However, the last two decades witness an evolution that the unilateral pursuit of financial returns and profits might be not conducive to realizing long-term value (Donaldson and Preston, 1995), especially when marginal cost uncertainty is greater than demand uncertainty (Allen et al., 2007). Instead, based on the concept of CSR (Corporate Social Responsibility) and sustainable development, balancing and meeting the interests of all stakeholders is regarded as the fundamental way for enterprises to develop in a long-term period (Jain et al., 2017; Martínez-Ferrero and García-Meca, 2020). In empirical experience, an ideal picture is expected that the economic and social benefits of enterprises are maximized simultaneously (Jensen, 2001). In this paper, we ask an empirical question: is there a relationship between corporate governance styles based on stakeholder-oriented and corporate performance? To answer this question, one of the difficulties is identifying companies with different corporate governance styles according to clear financial indicators because a company's corporate governance-related variables depend on its historical performance and can be manipulated by the management. More specifically, it may be difficult to identify whether the corporate governance characteristics of enterprises is more inclined to stakeholder-oriented value maximization by using numerical data from financial reports. This study attempts to categorize listed firms based on their corporate governance style using textual information from the MD&A (Management Discussion and Analysis) of Japanese firms and examines the validity of this classification.

Cespa et al., (2007) indicates that stakeholders other than shareholders can influence the CEO replacement, and the external stakeholders influence the company through by the threat of costly boycotts and media campaigns. In the institutionalization of stakeholder protection, CEO will use stakeholder protection as a consolidation strategy while CEO with strong take-over defenses and who can dominate the board of directors do not need to rely on stakeholders to strengthen their management. Therefore, the interest of shareholders and stakeholders are not always at loggerheads, and they share more common interests in an environment where management is deeply entrenched than people have expected. In Japan, several companies are often part of a network of affiliated companies called “keiretsu”, which is a system that allows firms to share their cross-shareholding and business relationships with their “Allies” (Yoshimori and Masaru, 1995). This system helps achieve a dominant position within their industry and market and creates a complex web of shareholdings that can make it difficult for outsiders to influence. Managers of these entrenched companies have a lower motivation to take a risk for future growth and avoid deciding to restructure or investment (Ikeda et al., 2018). In addition, the board of directors in Japanese companies is often made up of insiders, and with a higher level of CEO duality. It leads to higher information costs and a lack of supervision (Yang et al., 2014). For those companies with a high level of director shareholdings proportion, the executive

compensation structure is more inclined to stock option incentive than cash, which on one hand, encourages managers to make efforts to improve the stock price and take risks to invest (Mehran and Hamid, 1995). On the other hand, excessive director ownership will also lead to another result that those directors pay more attention to their interests than the interests of other stakeholders (Sakawa et al., 2018). Managerial Entrenchment leads to a result of favoring control contestability and managerial turnover. These entrenched companies often maintain a negative attitude toward external stakeholders' supervision of social responsibility, such as media and social activities, which affects their entrenched position (Surroca et al., 2008).

Different from these two types of companies, market-oriented companies focus on understanding and meeting the needs of customers and complying with the rules and preferences of the capital market. Compared with the interests of internal stakeholders, they may pay more attention to external stakeholders, such as the needs of customers, the provisions of the capital market, the preferences of external investors, and the social environment (Kirca et al., 2005). In our classification result, the firms in NN (Non-Entrenched and Non-Director-Dominated) group are further subdivided into high external supervision and low external supervision based on the independent director's ratio. We assume market-oriented companies have a lower proportion of cross-shareholdings and director shareholdings and more independent directors. These companies are more modern, and they emphasize a commitment to continuous innovation and adaptation to changing market conditions and have higher motivation to make effort on ESG activities (or just claim their attention). Japanese government and industry leaders follow *Davos Agenda of 2021* [1] and strongly recommend stakeholder capitalism instead of market-oriented capitalism. However, our results indicate that both ESG disclosure and shareholder-oriented capitalism are likely to be promoted by market-oriented firms.

Our main contributions are as follows:

Existing studies are underestimating the potential of text information to play an important role in corporate identification. The MD&A part of the company's financial report contains the manager's expectation of the company's future development direction, which to some extent reflects the concept of the company's operation. This paper uses Natural Language Processing (NLP) technology and Text Mining to cluster the MD&A texts of Japanese listed companies. Combined with the correlation analysis of numerical financing data, we proved the effectiveness of text information in solving the problem of company classification based on different stakeholder orientation. In addition, we also describe the characteristics of text disclosure in market-oriented group with high supervised level and low supervised level. We also establish ESG dictionary by conducting topic analysis in the MD&A documents from potential market-oriented companies. The regression result shows those non-entrenched companies who pay more attention to external stakeholders disclose more ESG information.

Many Japanese companies are considered deeply entrenched. These companies enjoy a quiet life through high director shareholding or cross-shareholding. Even

[1] https://www.weforum.org/events/the-davos-agenda-2021

though stakeholder interest maximization has been regulated in Japan, those entrenched companies insisted on paying more attention to the interests of internal stakeholders and may lack the motivation to focus on external stakeholders. After clustering companies based on text information and numerical data, this study uses OLS regression clustering by industry and year to estimate its impact on the financial performance of different companies. We find that market-oriented companies disclose more external stakeholder-related (Environment and Society) information than entrenched companies. The panel data regression results support that non-entrenched companies who focus on stakeholders discloses more ESG information and leads to better market evaluation than other firms. In addition, we find that strong external supervision can promote better financial performance for market-oriented companies that pay more attention to external stakeholders.

This study uses new data and new methods to supplement the gap between stakeholders and corporate governance theory. In the future, our methods still have great blank space for improvement, but our contributions are creative and new. It is expected to provide a new idea for the application of the text in the financial fundamental analysis. At the end of this paper, we clarify the insufficient parts that need to be improved in the future and explained the shortcomings of the classification method in this paper.

## 2. Methodology

### 2.1 Data and Samples

The textual data and corporate governance dataset are obtained from BDTI (Board Director Training Institute of Japan). Financial data were obtained from the Refinitiv Eikon database. The MD&A section was selected for this study, based on the classification of companies in the BDTI database, with PRIME/TSE1 companies as a sample, year from 2017-2021. The observation is 10,096. The similarity threshold is set at 50%. There are some boilerplate descriptions in the text information that appears in each text, and these boilerplate words obscure the results of text classification to a degree that cannot be ignored. In this paper, we define a word as a boilerplate word if it appears in more than 80% of the documents. The template words in the MD&A section include 課題、経営方針、対処、経営環境、当社、事業、経営、強化、向上、グループ、拡大、将来、判断、環境、現在、事項、企業、成長、目標. This paper also sets up auxiliary words that are commonly used in Japanese expressions but do not have a special meaning as a stop-words list.

### 2.2 Classification Based on Text Information

（1） Similarity calculation

Doc2Vec (Document Embedding with Paragraph Vectors) is a machine-learning technique used to generate vector representations for documents in a corpus. We train Doc2Vec to generate 300 Noun vectors (the vector size is 300) for each MD&A text, the epoch is set to 20. To calculate the similarity between two documents, we use the cosine similarity function to take the dot product of the two document vectors and divide it by the product of their length. The resulting value will be between -1 and 1,

where values closer to 1 indicate higher similarity.

（2） Topic Analysis

The Latent Dirichlet Allocation model, LDA model, is a probabilistic generative approach to identifying topics in a collection of documents. As shown in Picture 1, LDA assumes that each document is generated by firstly deciding a distribution over the topic from Dirichlet distribution with parameter, and associated with a latent topic z. These documents give a topic word distribution in the corpus. By using this model, we visualize the text characteristics in each group.

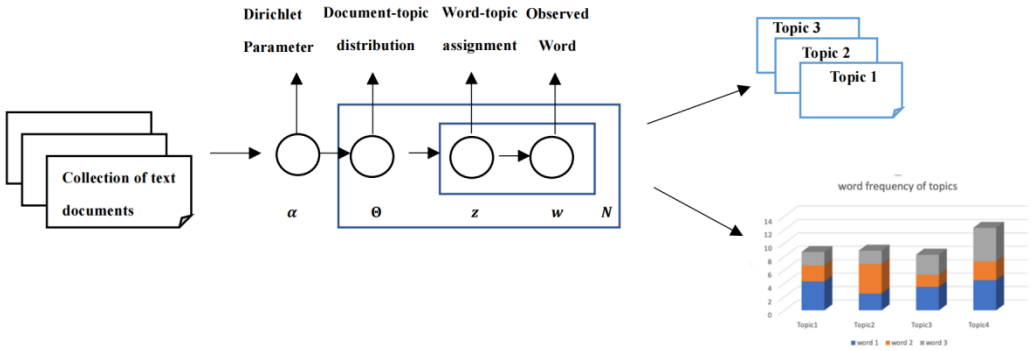



**Figure 1. LDA Model.** This figure plots the entire LDA space and its dataset. Alpha (α) controls per-document topic distribution, and the first box refers to all the documents in the corpus. The second box is the number of words in a document, given by N. Finally, we can get per topic word distribution and the frequency of each word in topic.

（3）Sample Filter Process:

The sample selection process is as follows：

- **Filter extreme value:** drop off firms with total assets larger than the value of (*mean+standard deviation*) or less than the value of (*mean-standard deviation*). The reason we do this is that companies with high cross shareholding ratios are generally larger in size, while companies with high director shareholding ratios are generally smaller in size. To avoid classification results based on company size, we minimize the gap among the size of samples.
- **Select three 'Entrenched (EN)' samples:** select the top 3 firms with highest the cross-shareholding ratio.
- **Filter family firms:** drop off firms with more than 34% directors' holdings. These firms are more likely to be family firms rather than director-dominated firms.
- **Select three 'Director-Dominated (DD)' samples:** After excluding family firms, three "shareholder oriented" samples were selected through the top 3 directors' shareholding ratio.
- **Sample screening:** Ensure that all EN samples are clear EN&Non-DD companies, and all DD samples are Non-EN& DD companies (More specifically, the similarity between an EN sample and all DD samples is less than 50%).

Finally, we select 3 'EN' samples and 3 'DD'samples from Materials (6/1104), Industrials (6/2990), Consumer Discretionary (6/2150), Consumer Staples (6/948), Health Care (6/509), Information Technology (6/1612), Communication Services

(6/541), Real Estate (6/257). We select 1 'EN' samples and 1 'DD'samples from Energy (2/105) and Utilities (2/133) due to the total number of firms in these two industries is relatively less than other industries.

(4) Group Classification

As shown in Picture 2, we calculate the similarity between each matching firm and sample firm by using Doc2Vec. If the similarity between the matching firm and any of the three EN samples is higher than 50%, the EN score is 1; if the similarity with all three samples is less than 50%, the EN score is 0. Likewise, if the similarity between the matching firm and any of the three DD samples is higher than 50%, the DD score is 1; if the similarity with all three samples is less than 50%, the DD score is 0. That is, each matching firm receives two scores, the EN score and the DD score.

We will obtain four groups:

If the EN score of firm A is 1, and DD score is 0, firm A will be classified to EN group. The feature of this group is that they have higher cross-shareholdings ratio.

If the EN score of firm B is 0, and DD score is 1, firm B will be classified to DD group. The feature of this group is that they have higher directors 'shareholdings ratio.

If the EN score of firm C is 1, and DD score is 1, firm C will be classified to ED group. The feature of this group is that they have both higher directors 'shareholdings ratio and higher cross-shareholdings.

If the EN score of firm D is 0, and DD score is 0, firm D will be classified to NN group. The feature of this group is that they have both lower directors 'shareholdings ratio and lower cross-shareholdings, we expect this group has higher level of independent directors.

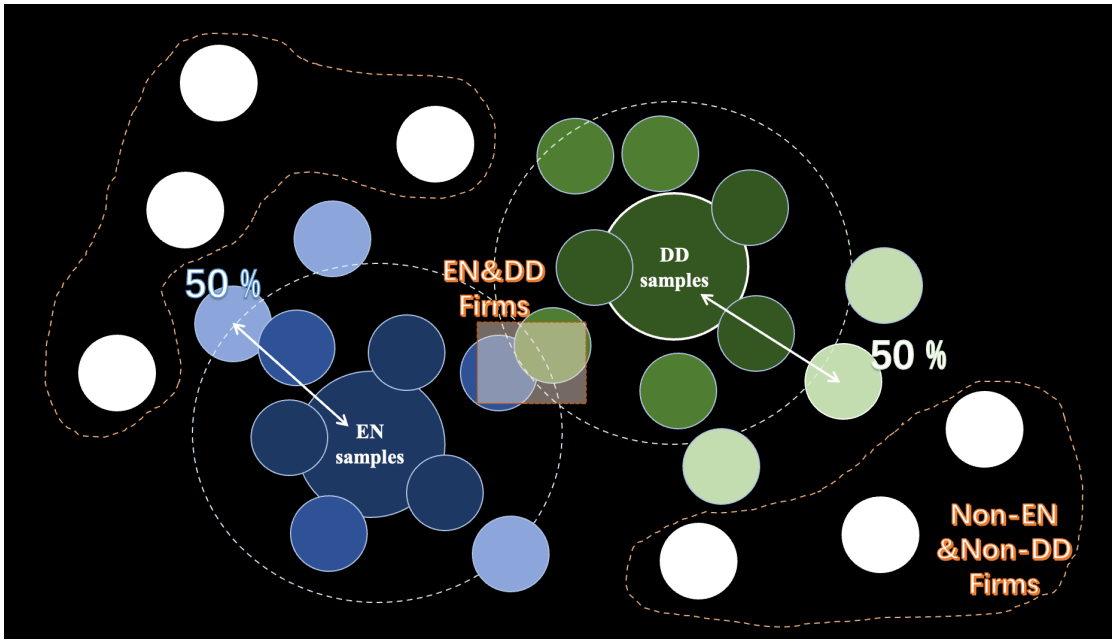



**Figure 2. Group Classification.** This figure plots the principle by which we perform the classification. With the EN sample and DD sample as the center, we calculate the cosine distance from the documents of other firms to the center. Companies within 50% similarity have darker colors, and the further the distance, the lighter the color. All blue companies belong to EN

companies and green companies belong to DD companies. Companies in the blue and green cross-over area we define as ED companies, and white companies are defined as NN companies.

## 2.2 Ordinary Least Squares Regression Analysis

(1) OLS regression

Firstly, by controlling the industry fixed effect and year fixed effect, we use simple OLS regression model to study the effect of different group types on financial performance.

$(Equation\ 1)$

$ROA_{i,t} = \alpha + y_1 Corporate\ Governance\ Styles + \beta_1 BSIZE + \beta_2 EXERTO + \beta_3 IDW\_RTO + \beta_4 IND + \beta_5 LEV + \beta_6 FSIZE + Industry\_FE + Year\_FE + \mathcal{E}_{it}.$

$(Equation\ 2)$

$ROA_{i,t} = \alpha + y_1 Corporate\ Governance\ Styles + y_2 Corporate\ Governance\ Styles \times IND + y_3 IND + \beta_1 BSIZE + \beta_2 EXERTO + \beta_3 IDW_{RTO} + \beta_4 LEV + \beta_5 FSIZE + Industry\_FE + Year\_FE + \mathcal{E}_{it}.$

Where corporate type respectively substituted into EN, DD, ED, and NN companies. When $Corporate\ Governance\ Styles$=1, it indicates that the company belongs to the corresponding group. In addition, i and t represent firm i at time t, respectively, α represents the constant, y（1-3）and β (1-6) is the slope of the Endogenous Variables and control variables which reflects a partial or prediction for the value of dependent variable, *Industry_FE* and *Year_FE* represents our regression result clustering by industry and year, and Ɛit is a random error term. $Corporate\ Governance\ Styles \times IND$ is the interaction of Corporate Governance Styles and independent directors' ratio.

(2) Variables

- Dependent Variable:

We measure ROA as operating income before interest and taxes to total assets divided by end of year total assets to represent financial performance.

- Independent Variables:

EN_similarity: A dummy variable which take a value of 1 if the similarity with EN samples is larger than 50%, and 0 otherwise. EN firms have higher cross-shareholdings. Entrenched managers of EN firms prefer to enjoy their quiet life (Ikeda, 2018) and have negative impact on firm value (Lucian,2005).

DD_similarity: A dummy variable which take a value of 1 if the similarity with DD samples is larger than 50%, and 0 otherwise. DD firms have higher director ownership. On one hand, we assume that managers of DD firms pay more attention to maximize the shareholder value and make effort on the revenue growth. On the other hand, excessive director ownership can lead to the same result as EN companies, that is, these managers only focus on the interests of internal stakeholders, the company's transparency is reduced.

- Corporate Governance

Good corporate governance might be correlated with unobservable risk management, higher transparency, and better internal control. The relationship between the governance structure and financial performance is proved to be endogenous. In this paper, some strong indicates of corporate governance are treated as control variables in our model.

(1) BSIZE: The number of board members. A larger board can allow for a wider range of expertise and experience and may have more accountability and responsibility for the company's financial performance.
(2) EXERTO: The ratio of directors who concurrently serve as executive officers. This indicate is higher, the independence of board decreases, the excessive power of directors will have negative impact on financial corporate.
(3) IDW_RTO: The ratio of the number of female directors to the total number of directors. A greater diversity of board may allow for making decisions from different perspectives and ultimately improved financial performance.

- Other Control Variables

In addition, we control two financial indicates:

(1) ln(asset): the natural logarithm of total assets.
(2) LEV: The ratio of total debt to total asset is represented as the financial leverage.

## 3 Results

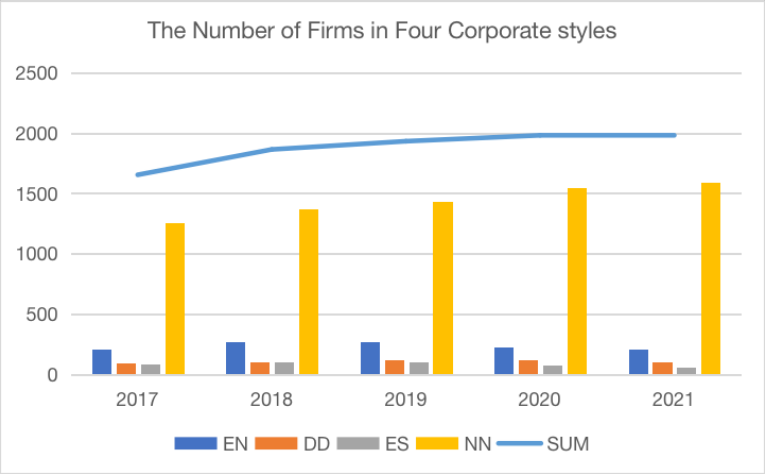
### 3.1 Classification Results

As shown in Table I, the number of EN firms decrease after the Tokyo Stock Exchange, Inc. (TSE) incorporates the fundamental principles for corporate governance established in "Japan's Corporate Governance Code" (the Code) into its listing rules in April 2019, which is to contribute to effective corporate governance in Japan.

**Table I.**

**Change in the number of firms in each group from 2017 to 2021.**

Panel A: The Number of Firm in Four Corporate Styles




Panel B: The Number and Percentage of Four Corporate Styles

| | 2017 | | 2018 | | 2019 | | 2020 | | 2021 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **num** | **percentage** | **num** | **percentage** | **num** | **percentage** | **num** | **percentage** | **num** | **percentage** |
| EN | 209 | 13% | 272 | 15% | 274 | 14% | 230 | 12% | 207 | 11% |
| DD | 99 | 6% | 109 | 6% | 119 | 6% | 120 | 6% | 105 | 5% |
| ED | 84 | 5% | 109 | 6% | 105 | 5% | 80 | 4% | 61 | 3% |
| NN | 1261 | 76% | 1373 | 74% | 1433 | 74% | 1550 | 78% | 1593 | 81% |
| SUM | 1653 | 100% | 1863 | 100% | 1931 | 100% | 1980 | 100% | 1966 | 100% |

In addition, NN firms (non-entrenched firms and non-director-oriented firms) are the group with the largest number, characterized by having lower cross-shareholdings, lower director ownerships, and in the regression analysis section we compare the corporate governance styles of each group after controlling for total assets, leverage, and firm value, NN firms have a higher proportion of independent directors. We define NN firms as market-oriented firms, that is, (1) these firms are more concerned with the interests of external stakeholders than internal stakeholders (e.g., the board of directors, directors, and major shareholders). (2) With less equity concentration and a higher percentage of independent directors, these companies are more attractive to outside investors due to their relative transparency of information. (3) These companies have less incentive for their managers to enjoy a quiet life than EN and DD companies and may be more concerned with environmental and social issues, as well as value growth. The number of NN firms is increasing after 2019.

We selected EN firms from those with highest cross-shareholdings and DD firms from those with highest director ownerships (<34%). Generally, the firm size of EN firms will be far larger than DD firms, therefore, we avoid choosing some companies with extreme values of total assets, even so, in the sample, we still find that the total assets of the EN samples are 8.93 times higher than those of the DD samples, the

director ownership is approximately 0.11 times higher, the cross-shareholding is 5.37 times higher, and there is almost no difference in the ratio of independent directors.

In panel B (Table I), we find that the difference in asset size between DD and EN decreases from 9 times increases to 2 times. In addition, there is almost no difference in firm size between EN and NN. Among the four groups, EN has the lowest director ownership and the highest cross-shareholders. DD firms have the highest director ownership. NN firms have the highest independent director ratio.

## Table II

## Summary Statistics

In panel A, we report the mean value and median value of total asset, director ownership, cross-shareholdings, and the ratio of independent directors in EN and DD sample firms. Ratio to EN firms (TS) represents the ratio of the mean (median) of the total assets of each group to the mean (median) of the total assets of EN. In panel B, we report the mean value and median value of total asset, director ownership, cross-shareholdings, and the ratio of independent directors in four groups in matching results.

**Panel A: The summary statistics of four groups in sample firms**

| | EN | | DD | | EN/DD | |
|---|---|---|---|---|---|---|
| | MEAN | MEDIAN | MEAN | MEDIAN | MEAN | MEDIAN |
| **Total Asset** | 4,083,500,376 | 760,818,165 | 456,900,944 | 235,215,704 | 8.93 | 3.23 |
| **Director Ownership (%)** | 2.74 | 0.30 | 24.37 | 28.25 | 0.11 | 0.01 |
| **Cross-shareholdings (%)** | 18.42 | 18.10 | 3.42 | 0 | 5.37 | - |
| **independent directors (%)** | 38.56 | 37.50 | 39.55 | 40.00 | 0.97 | 0.93 |

**Panel B: The summary statistics of four groups in matching result**

| | EN | | DD | | ED | | NN | |
|---|---|---|---|---|---|---|---|---|
| | MEAN | MEDIAN | MEAN | MEDIAN | MEAN | MEDIAN | MEAN | MEDIAN |
| **Total Asset** | 4,328,816,581 | 850,583,588 | 1,974,416,290 | 464,236,676 | 2,315,768,807 | 683,635,997 | 4,949,067,178 | 757,487,879 |
| **Ratio to EN firms (TS)** | 1 | 1 | 0.46 | 0.55 | 0.53 | 0.80 | 1.14 | 0.89 |
| **Director Ownership (%)** | 6.35 | 0.92 | 12.36 | 4.37 | 6.31 | 1.86 | 5.24 | 0.67 |
| **Cross-shareholdings (%)** | 8.36 | 4.10 | 5.75 | 1.20 | 8.48 | 4.15 | 8.17 | 4.90 |
| **independent directors (%)** | 31.12 | 30.00 | 29.86 | 27.92 | 28.59 | 28.57 | 32.20 | 33.33 |

## Table III.
## The Firm Nature of Each Group

This table reports the estimated coefficients and t-statistics (in brackets) of the regression result firm nature of 4 groups. EN_similarity is the dummy varible indicates whether the document is more than 50% similar to EN samples, and the DD_similarity is the same. EN_S×DD_S is interaction term of EN_similarity and DD_similarity. If EN_similarity = 0 and DD_similarity = 0, the coefficient of EN_S×DD_S represent the nature of ED company. This regression is run at the industry and year level with control variables of total asset, leverage and TobinQ. This result indicates the nature of a company that belongs to EN, DD, ED, NN in the same industry in the same year, with the same company size, leverage, and company value. In columns (1) and (2), the dependent variable is cross-shareholdings. The coefficient of EN_similarity is 2.22, and it indicates if EN_similarity =1, DD_similarity =0 and EN_S×DD_S =0, the EN companies have 2.22 higher level of cross shareholdings than NN companies. The coefficient of NN is -0.96, that means NN companies has 0.96 lower level of cross shareholdings than other groups. In columns (3) and (4), The dependent variable is director shareholdings. A DD company has 4.93 higher level of director shareholdings than a NN company. In columns (5) and (6), NN company has 0.0209 higher level of independent directors than other companies. Asterisks denote significance levels (*** =1%, ** =5%, *=10%). R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by independent variables.

| Dependent Variable: | cross_share _ratio | P>t | Cross_share _ratio | P>t | Directors _share_ratio | P>t | Directors _share_ratio | P>t | Independent Directors | P>t | Independent Directors | P>t |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | | (2) | | (3) | | (4) | | (5) | | (6) | |
| EN_similarity | 2.22 | *** | | | -0.68 | ** | | | -0.0240 | *** | | |
| | (7.43) | | | | (-2.13) | | | | (-7.67) | | | |
| DD_similarity | -1.36 | *** | | | 4.93 | *** | | | 0.0022 | | | |
| | (-3.32) | | | | (11.28) | | | | (0.47) | | | |
| EN_S×DD_S | -0.37 | *** | | | -3.95 | *** | | | -0.0159 | ** | | |
| | (-0.57) | | | | (-5.71) | | | | (-2.28) | | | |
| NN | | | -0.96 | *** | | | -0.94 | *** | | | 0.0209 | *** |
| | | | (-4.02) | | | | (-3.69) | | | | (7.97) | |
| Ln (Total Asset) | 0.5 | *** | 0.56 | *** | -2.32 | *** | -2.40 | *** | 0.0171 | *** | 0.0167 | *** |
| | (8.38) | | (9.39) | | (-36.67) | | (-38.07) | | (29.31) | | (29.18) | |
| Leverage | -0.03 | *** | -0.04 | *** | 0.05 | *** | 0.06 | *** | -0.0003 | *** | -0.0003 | *** |
| | (-6.44) | | (-7.22) | | (9.54) | | (10.64) | | (-5.93) | | (-5.86) | |
| TobinQ | -0.94 | *** | -0.94 | *** | 1.57 | *** | 1.58 | *** | 0.0067 | *** | 0.0068 | *** |
| | (-18.11) | | (-18.04) | | (28.4) | | (28.27) | | (14.88) | | (14.93) | |
| R-squared | 0.1335 | | 0.1285 | | 0.3065 | | 0.2977 | | 0.1118 | | 0.1094 | |
| Num of Obs. | 10,096 | | | | 10,096 | | | | 10,096 | | | |
| Industry_Fixed | YES | | | | YES | | | | YES | | | |
| Year_Fixed | YES | | | | YES | | | | YES | | | |

Our approach is to cluster firms based on textual similarity, and to ensure that the clustering results are consistent with our sample, we control for total assets, leverage and firm value and investigate the correlation between firm type and corporate governance indicators using industry and year fixed effects models. This result strongly demonstrates the value of textual data in the company clustering problem. The independent variables in Table III (EN_similarity and DD_similarity) are different from the EN and DD in Table IV. In the point (4) of 2.2, we explained the classification method and the company will be classified to EN group only when EN_similarity =1 and DD_similarity=0 simultaneously. EN_similarity and DD_similarity are dummy variables that initially measures whether the similarity score between the text and the EN or SO samples exceed 50%.

In column (1), the coefficient of EN_ similarity is 2.22, which means that the cross shareholding of EN company in the industry is 2.22 higher than that of NN company; In the same way, DD company is 1.36 lower than NN company, and ED company is 0.49 (2.22-1.36-0.37) higher than NN company. In column (2), the coefficient of NN is -0.96, which means that the cross-shareholding level of NN company is 0.96 lower than that of other companies (EN&DD&NN). Therefore, we can conclude that after controlling the company size, leverage, and market value, the cross-shareholding level of EN company is higher than that of other types.

In column (3), the coefficient of EN_ similarity is -0.68, which means that the director ownership of EN company in the industry is 0.68 lower than that of NN company; DD company is 4.93 lower than NN company, and ED company is 0.3 (-0.68+4.93-3.95) higher than NN company. In column (2), the coefficient of NN is -0.94, which means that the director ownership level of NN company is 0.94 lower than that of other companies (EN&DD&NN). Therefore, we can conclude that after controlling the company size, leverage, and market value, the director ownership of DD company is higher than that of other types.

In column (5), the coefficient of EN_ similarity is -0.024, which means that the independent director ratio of EN company is 0.024 lower than NN company. The coefficient of DD company is insignificant. The ED company is 0.0377 (-0.0240+0.0022-0.0159) lower than NN company. The column (6) shows the independent director ratio of NN company is 0.0209 higher than other groups. It is consistent with our hypothesis that the NN company is potentially regarded as market-oriented company.

Having distinguished the significant differences between EN and DD companies, we need to characterize NN companies, that is, the nature of NN firms. This is a group with a very high-volume share, and its importance is consistent with that of EN and DD firms. We hypothesize that a firm with lower cross-shareholding and executive ownership is more independent and may be more concerned with the interests of external stakeholders. In the part of 3.2, to verify whether NN firms are more concerned with the interests of external stakeholders (environmental, social and governance), we visualize the textual features of NN companies with high independent director ratio and low independent director ratio.

## 3.2 The differences of Text Characteristics between High and Low independent NN companies

We calculate the index of significance for each word by the ratio of word frequency on the number of documents.

$$Significance = \frac{the\ number\ of\ word\ appearance}{the\ number\ of\ total\ documents\ in\ the\ group}$$

As the Panel A (Table IV) shows, we compared the differences of topic's significance between NN group (non-entrenched company) and other groups (entrenched company). We obtained several conclusions as follows. (1) EN companies pay more attention to corporate governance and disclose more information about take-over defense than DD companies, which is to preserve the entrenchment characteristics of the company. (2) DD companies' topics revolve around customers and services, with little mention of topics related to external stakeholders. (3) NN companies do pay more attention to external stakeholders than other companies, and their disclosure texts mainly talk about environment, social and governance topics. They also show a strong focus on shareholders and firm value. This type of company fits well with the characteristics of a market-oriented company.

In addition, we divided NN companies into NN_High_Independence (higher than median value) and NN_Low_Independence (lower than median value) based on independent director ratio. As Panel B (Table IV) shows the differences between the NN companies with high independence and low dependence. The high dependent NN companies disclose more Society, and Environment related information than low dependent NN companies. The low dependent NN companies disclose pay more attention to Corporate Governance (especially shareholders) than high dependent NN company.

**Table IV.**

This table represent the result on text analysis. We first used LDA topic analysis model on the texts in the NN group and demonstrate the words related to ESG. In addition, the word frequency in each topic is achievable and we list the five most import words in each topic. In Panel A, we display the different significances of topics and corresponding words in NN, EN and DD group separately. To test our hypothesis, we divide NN firms into two groups according to different levels of monitoring. NN_High_Independence are those firms with independent director ratios greater than the median, and NN_Low_Independence are those firms with independent director ratios less than the median. We visualize the importance of the words contained in the topic by using a heat map, with green representing very important, yellow representing medium, and red representing low importance. The darker the green, the stronger the significance of the word; the darker the red, the stronger the insignificance of the word.

**Panel A: The Differences of ESG Disclosure among Four Corporate Styles**

| Topic | Significance = Word Frequency / The Number of Documents | | | |
|---|---|---|---|---|
| | Words | NN | EN | DD |
| Client Service | サービス(Service) | 3.3604 | 2.2149 | 3.0670 |
| | お客様(Customer) | 2.1323 | 1.8315 | 1.9986 |
| | 人材(Human Resources) | 1.5673 | 1.0383 | 1.6405 |
| | 開発(R&D) | 2.8573 | 2.0670 | 2.0937 |
| | 技術(Technology) | 2.5390 | 1.4485 | 0.7723 |
| Society Development | 社会(Society) | 2.3396 | 1.5631 | 1.3552 |
| | 推進(Advance) | 3.0101 | 2.1696 | 1.9301 |
| | 影響(Impact) | 1.5153 | 0.9609 | 1.0901 |
| | 実現(Implementation) | 2.4505 | 1.7684 | 1.5036 |
| | 提供(Provide) | 2.8020 | 1.9477 | 2.1967 |
| Enterprise Value | 企業価値(Firm Value) | 2.2185 | 1.7270 | 1.1484 |
| | 株主(Shareholders) | 3.1122 | 1.9806 | 1.0007 |
| | 利益(Profit) | 1.8497 | 1.2889 | 0.7968 |
| | 大規模(Large Scale) | 1.5498 | 0.9152 | 0.3285 |
| | 方針(Policy) | 1.2560 | 0.8218 | 0.6304 |
| Environment Risk | 社会(Society) | 2.3396 | 1.5631 | 1.3552 |
| | 投資(Investment) | 1.4017 | 0.8869 | 1.0007 |
| | リスク(Risk) | 0.5342 | 0.3071 | 0.3105 |
| | 気候変動(Climate Change) | 0.1734 | 0.0817 | 0.0382 |
| | サステナビリティ (Sustainability) | 0.2286 | 0.0852 | 0.0497 |
| Corporate Governance | 株主(Shareholders) | 3.1122 | 1.9806 | 1.0007 |
| | 買い付け(Take Over) | 2.6088 | 1.2820 | 0.3710 |
| | 取締役会(Board of Directors) | 1.7912 | 0.8656 | 0.2363 |
| | 株式(Stock) | 0.8677 | 0.4919 | 0.1988 |
| | コーポレート・ガバナンス (Corporate Governance) | 0.1722 | 0.3780 | 0.3581 |

**Panel B: The significance of topic words in different groups**

| | Topic1: Environment Risk | | | | | |
|---|---|---|---|---|---|---|
| word_list | 社会 (Society) | 持続 (Sustainable) | リスク (Risk) | 気候変動 (Climate Change) | サステナビリティ (Sustainability) | SUM |
| NN_High_Independence | 3.4121 | 1.3635 | 0.8074 | 0.4367 | 0.5274 | 6.54 |
| NN_Low_Independence | 2.2249 | 0.9183 | 0.3947 | 0.1017 | 0.1544 | 3.79 |
| | Topic2: Society Development | | | | | |
| word_list | 社会 (Society) | 推進 (Advance) | 実現 (Implementation) | 影響 (Impact) | 貢献 (Contribution) | SUM |
| NN_High_Independence | 3.4121 | 3.7164 | 3.3581 | 1.9439 | 2.0082 | 14.43 |
| NN_Low_Independence | 2.2249 | 3.0101 | 2.3714 | 1.2763 | 1.5496 | 10.43 |
| | Topic3: Corporate Governance | | | | | |
| word_list | 株主 (Shareholders) | 買い付け (Take-Over) | 取締役会 (Board of Directors) | 株式 (Stock) | コーポレート・ガバナンス (Corporate Governance) | SUM |
| NN_High_Independence | 3.0402 | 2.1201 | 1.6724 | 0.8435 | 0.5684 | 8.24 |
| NN_Low_Independence | 4.0531 | 3.612 | 2.3179 | 1.099 | 0.4728 | 11.55 |

### 3.3 Corporate Governance and Financial Performance

By using the year and industry fixed effect, we firstly use simple OLS regression to test the effect of company types on financial performance in Panel A (Table V), the ROA increases by 1.59 if the corporate governance style stated in the MD&A section is more relevant to DD companies. This result shows that DD companies have better financial performance compared to other companies in the same industry. The columns (1),(3),(4) show that the relationship between the types of EN, ED and NN and financial performance is insignificant. In Panel B (Table V), the TobinQ increases by 0.1137 if the corporate governance style stated in the MD&A section is more relevant non-entrenched companies. This result is also strong evidence that NN Company is a potential market -oriented company

According to Ran Duchin and John G. Matsusaka (2010), companies are required to increase the number of outside directors by regulation. When the information cost is low, external supervision will improve performance, while when the information cost is high, the impact of external directors will make situation worse. In panel C, after introducing the interaction term of independent directors, the impact of EN and NN companies becomes significant. EN is a typical company with asymmetric information, while NN is the company with the best information transparency. The coefficient of EN is 5.07 and the coefficient of interaction is - 22.25, which means that for every 1% increase in the independent director's ratio, EN's financial performance decreases by 15.18. The coefficient of NN is -3.34 and the coefficient of interaction is 12.38, which means that for every 1% increase in the independent executive ratio, the financial performance of NN company increases by 9.04. In other words, for market oriented NN companies, the higher the degree of external supervision, the better the financial performance of the enterprise. However, a higher level of independence did not prove to have a significant impact on the performance of NN in terms of market value.

## Table V
## The Impact of Corporate Governance Style on Financial Performance

This table reports the estimated coefficients and t-statistics (in brackets) of the regression that describes the effect of corporate governance style that the managers disclosed in MD&A text on financial performance. The EN, DO, ED, NN is dummy variable indicates whether the company belongs to the corresponding company type (a company can only correspond to one type). The dependent variable in Panel A is ROA and the dependent variable in Panel B is TobinQ, and this regression run at industry and year level. We also control variables of corporate governance, leverage, and the logarithm of total asset. R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by independent variables. In Panel C and D, we introduce the interaction of the corporate type clustered by the concern of stakeholders and independent director ratio. The dependent variable in Panel C is ROA and the dependent variable in Panel D is TobinQ. The independent variable is EN in column (1), DD in column (2), ED in column (3), and NN in column (4). We run this regression model at industry and year level, which includes Board Size, EXERTO, Female Directors Ratio, ln (Total Asset) and Leverage. Asterisks denote significance levels (*** =1%, ** =5%, *=10%).

Panel A: The Impact of corporate Governance Style on Financial Performance

| Dependent Variable | ROA (1) | P>t | ROA (2) | P>t | ROA (3) | P>t | ROA (4) | P>t |
|---|---|---|---|---|---|---|---|---|
| EN | 0.5496 (0.25) | | | | | | | |
| DD | | | 1.7836 (5.97) | *** | | | | |
| ED | | | | | -0.8503 (-2.68) | *** | | |
| NN | | | | | | | -0.3941 (-2.24) | ** |
| Leverage | -0.1167 (-29.98) | *** | -0.1184 (-30.57) | *** | -0.1167 (-30.16) | *** | -0.1164 (-30.07) | *** |
| Board Size | 1.7031 (4.96) | *** | 1.7178 (5.01) | *** | 1.6922 (4.93) | *** | 1.7082 (4.97) | *** |
| EXERTO | 0.0063 (3.23) | *** | 0.0064 (3.28) | *** | 0.0064 (3.27) | *** | 0.0063 (3.20) | *** |
| Female directors | 0.0896 (8.66) | *** | 0.0898 (8.71) | *** | 0.0896 (8.67) | *** | 0.0906 (8.76) | *** |
| ln (Total Assets) | -0.3476 (-7.93) | *** | -0.3145 (-7.15) | *** | -0.3521 (-8.05) | *** | -0.3429 (-7.84) | *** |
| Independent directors | 3.2764 (4.96) | *** | 3.3793 (5.12) | *** | 3.2157 (4.87 ) | *** | 3.3375 (5.05) | *** |
| R-squared | 0.1565 | | 0.1594 | | 0.1571 | | 0.1569 | |
| Observation | 10,251 | | | | | | | |
| Industry_Fixed | YES | | | | | | | |
| Year_Fixed | YES | | | | | | | |

| Panel B: The Impact of Corporate Governance Style on Market Value | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dependent Variable | TobinQ (1) | P>t | TobinQ (2) | P>t | TobinQ (3) | P>t | TobinQ (4) | P>t |
| EN | -0.0271 | | | | | | | |
| | (-0.51) | | | | | | | |
| DD | | | -0.0238 | | | | | |
| | | | (-0.32) | | | | | |
| ED | | | | | -0.2257 | *** | | |
| | | | | | (-2.88) | | | |
| NN | | | | | | | 0.0961 | ** |
| | | | | | | | (2.20) | |
| Leverage | -0.0086 | *** | -0.0086 | *** | -0.0086 | *** | -0.0087 | *** |
| | (-9.02) | | (-8.97) | | (-9.00) | | (-9.10) | |
| Board Size | 0.1950 | ** | 0.1945 | ** | 0.1914 | ** | 0.1937 | ** |
| | (2.30) | | (2.29) | | (2.25) | | (2.28) | |
| EXERTO | 0.0018 | *** | 0.0018 | *** | 0.0018 | *** | 0.0018 | *** |
| | (3.79) | | (3.79) | | (3.83) | | (3.82) | |
| Female directors | 0.0221 | *** | 0.0221 | *** | 0.0221 | *** | 0.0218 | *** |
| | (8.63) | | (8.67) | | (8.69) | | (8.55) | |
| ln (Total Assets) | -0.2275 | *** | -0.2283 | *** | -0.2292 | *** | -0.2288 | *** |
| | (-20.92) | | (-20.89) | | (-21.12) | | (-21.09) | |
| Independent directors | 1.5444 | *** | 1.5442 | *** | 1.5288 | *** | 1.5294 | *** |
| | (9.42) | | (9.42) | | (9.32) | | (9.32) | |
| R-squared | 0.1374 | | 0.1374 | | 0.1381 | | 0.1378 | |
| Observation | 10,251 | | | | | | | |
| Industry_Fixed | YES | | | | | | | |
| Year_Fixed | YES | | | | | | | |

### 3.4 Corporate Governance Style and ESG disclosure

From column (1) (Table VI), we can see that EN company is EN_ similarity=1, DD_ Similarity=0, with a coefficient of -6.6143, indicating that the level of disclosure of social and environmental related information by EN company is 6.61 lower than that of NN company. Similarly, the DD company is EN_ similarity=0, DD_ Similarity=1, with a coefficient of -7.0584, indicating that the level of disclosure of social and environmental related information by EN company is 7.0584 lower than that of NN company. ES company is EN_ similarity=1, DD_Similarity=1, with a coefficient of -9.915 (-6.6143-7.0584+3.7577), indicating that the level of disclosure of social and environmental related information by ED companies is 9.915 lower than that of NN companies. At the same time, it can be seen from column (2) that NN company is 7.4409 better than Non-NN company.

From column (3), we can see that EN company is EN_ similarity=1, DD_ Similarity=0, with a coefficient of -9.1453, indicating that the level of disclosure of governance related information by EN company is about 9 words less than that of NN company. Similarly, the SO company is EN_ similarity=0, DD_ Similarity=1, with a coefficient of -7.0584, indicating that the level of disclosure of governance related information by EN company is about 7 words less than that of NN company. ED company is EN_ similarity=1, DD_Similarity=1, with a coefficient of -9.915 (-6.6143-7.0584+3.7577), indicating that the level of disclosure of governance related information by ED companies is 9 words less than that of NN companies. In column (4) that NN company is 13 words more than non-NN company.

## Table VI: ESG Information Disclosure

This table reports the estimated coefficients and t-statistics (in brackets) of the regression that describes the impact of firm nature on ESG disclosure clustering by 4 groups. Column (1) and (2) presents the result of the impact of corporate governance style on the disclosure number of Environment and Society related words. Column (3) and (4) presents the result of the impact of corporate governance styles on the disclosure number of Governance related words. In column (1), EN_similarity is the dummy varible indicates whether the document is more than 50% similar to EN samples, and the DD_similarity is the same. EN_S×DD_S is interaction term of EN_similarity and DD_similarity. In column (2), the independent variable is NN, which is a dummy variable whether the company belongs to NN groups. For example, if EN_similarity = 1 and DD_similarity = 1, the coefficient of EN_S×DD_S represent the nature of ED company. This regression is run at the industry and year level with control variables of total asset, leverage and Tobin Q. The regression run at industry and year fixed effect. Asterisks denote significance levels (*** =1%, ** =5%, *=10%).

| Dependent Variable | Num_ Environment_ Society (1) | P>t | Num_ Environment_ Societ (2) | P>t | Num_ Governance (3) | P>t | Num_ Governance (4) | P>t |
|---|---|---|---|---|---|---|---|---|
| EN_similarity | -6.6143 | *** | | | -9.1453 | *** | | |
| | (-13.90) | | | | (-6.52) | | | |
| DD_similarity | -7.0584 | *** | | | -15.8844 | *** | | |
| | (-10.76) | | | | (-8.22) | | | |
| EN_similarity× DD_similarity | 3.7577 | *** | | | 5.131 | * | | |
| | (3.72) | | | | (1.72) | | | |
| NN | | | 7.4409 | *** | | | 13.1439 | *** |
| | | | (19.6) | | | | (11.75) | |
| ln(asset) | 3.8042 | *** | 3.8284 | *** | 2.051 | *** | 2.2095 | *** |
| | (39.4) | | (39.99) | | (7.21) | | (7.83) | |
| Leverage | -0.0008 | | -0.0022 | | -0.0741 | *** | -0.0847 | *** |
| | (-0.10) | | (-0.26) | | (-2.95) | | (-3.39) | |
| Tobinq | 0.5359 | *** | 0.5434 | *** | -0.764 | *** | -0.7427 | *** |
| | (6.16) | | (6.24) | | (-2.98) | | (-2.89) | |
| R-squared | 0.2453 | | 0.2439 | | 0.0756 | | 0.0734 | |
| Observation | 10,489 | | | | 10,489 | | | |
| Industry_Fixed | YES | | | | YES | | | |
| Year_Fixed | YES | | | | YES | | | |

# 4 Conclusion and Discussion

## 4.1 Conclusion

The debate on whether maximizing the interests of shareholders or stakeholders should be followed has never ceased to be discussed. It is difficult to categorize companies by corporate governance characteristics based on governance-related variables alone because a company's corporate governance-related variables depend on its historical performance and can be manipulated by the management. In the annual reports of companies, managers give a very clear description of the challenges and strategies for the future development of the company in MD&A disclosure, and from this part of the textual information we can get a clearer picture of which stakeholders the company pays attention to, such as customers, shareholders or society and the environment. To solve this problem, this paper uses a text-based clustering method, which converts MD&A text into vectors using Doc2Vec after selecting the sample lists, and finally uses cosine similarity to get the distance between each matching firm and the sample firms. From the results, we can obtain the following conclusions:

(1) First, we performed a statistical test on the firm classification results. After controlling for total assets, leverage, and firm value, the text-based company clustering results are consistent with their financial data characteristics. In other words, firms with high text similarity to the EN sample also have higher cross-shareholdings and lower directors' shareholdings. Companies with high text similarity to the DD sample also have higher director ownership and lower cross-shareholdings, which are two very different types of companies. At the same time, we define the NN type of companies that are too large in number to be ignored as a market-oriented companies. These companies exhibit significantly lower director ownership and cross-shareholdings and higher levels of independent directors than the other groups. Our results demonstrate the potential value of text data in identifying company characteristics, and that it is possible and reliable to find companies of the same type utilizing text clustering.

(2) Secondly, in order to clarify the nature of NN companies. We divided NN companies into high independent group and low independent group based on the ratio of independent directors. we visualize the textual characteristics of two groups. According to the result of topic analysis, the high independent NN company pay more attention to external stakeholders, such as society and environment. The low independent NN company care about shareholders and stock prices at a very high level, and these companies are potential shareholder interest-oriented companies. The ESG dictionary based on the MD&A text from NN companies is established to calculate the word frequency related to Environment, Society and Governance in each document. The regression result shows NN company (potential market-oriented company) disclose more ESG related information than other groups.

(3) Finally, we analyze the effect of different firm types on financial performance using fixed effects OLS regression models. In the OLS regression model, we fixed industry and year, and the coefficient of DD firms is significantly positive to ROA, while all other firms are insignificant. This indicates that within the same year and industry, DD firms outperform other firms at a 99% confidence level. For EN

companies with the highest degree of information asymmetry, the introduction of external regulators has a negative impact on their financial performance. For market-oriented and highly transparent NN companies, the introduction of external regulators improves the company's performance. In other words, those modern enterprises with higher independence are more concerned about the interests of external stakeholders (such as society and environment) and have achieved better financial performance.

### 4.2 Discussion

The purpose of this study is to use textual information to distinguish between different types of companies that focus on the interests of shareholders and stakeholders. We used text clustering, an idea that is new and has much space for development in the future.

(1) In selecting DD samples, we expect that managers in this type of company are more interested in shareholders' interests and will make efforts to increase the stock price. Although we excluded family firms with more than 34% director ownership from the sample, this standard line may still be designed too high. Although director shareholding increases executives' incentives to raise share prices, the relationship between executive shareholding and firm performance may be inverted U-shaped, meaning that after a certain percentage, these executives may sacrifice the firm's interests for their own. In the next stage of the study, we will control the shareholding of the DD sample between 10% and 15% and introduce the executive compensation of the sample firms.

(2) The companies in the NN group in this paper are roughly defined as non-entrenched companies and potential market-oriented companies. Although we conducted a preliminary examination of its nature using the independent directors' ratio, we found that market-oriented companies with high independent directors' ratios may be more inclined to maximize the interests of external stakeholders (environmental and social), while market-oriented companies with low independent directors' ratios pay attention to shareholders at highest level might be the shareholder-oriented companies we are looking for. We will conduct further examination in the future work.

(3) We have obtained groupings of different company types, and we plan to design portfolios of the best-performing DD companies at the industry level and model and analyze the stock returns of these types of companies using time series models. This may enhance the practical value of our proposed new approach.

## Reference


Allen, Franklin, Elena Carletti, and Robert Marquez, 2007, "Stakeholder capitalism, corporate governance and firm value." Corporate Governance and Firm Value (September 16, 2009). EFA: 09-28.

Bárcena-Ruiz, JC, Sagasta, A, 202, Cross-ownership and corporate social responsibility. The Manchester School, 89: 367– 384.

Cespa, Giovanni, and Giacinta Cestone. "Corporate social responsibility and managerial entrenchment." Journal of Economics & Management Strategy 16.3 (2007): 741-771.

Donaldson, T., & Preston, L. E, 1995, The stakeholder theory of the corporation: Concepts, evidence, and implications. Academy of management review, 20(1), 65-91.

Freeman, R. Edward, 2010, Strategic management: A stakeholder approach. Cambridge university press.

Freeman, R. Edward, and David L. Reed, 1983, "Stockholders and stakeholders: A new perspective on corporate governance." California management review 25.3 88-106.

James A. Brickley, Jeffrey L. Coles, Gregg Jarrell, 1997, Leadership structure: Separating the CEO and Chairman of the Board, Journal of Corporate Finance, Volume 3, Issue 3, Pages 189-220

Jensen, M. C, 2001, Value maximization, stakeholder theory, and the corporate objective function. Journal of Applied Corporate Finance, 14(3), 8-21.

Jain, T., Aguilera, R.V. & Jamali, D, 2017, Corporate Stakeholder Orientation in an Emerging Country Context: A Longitudinal Cross Industry Analysis. J Bus Ethics 143, 701–719.

Kirca, Ahmet H., Satish Jayachandran, and William O. Bearden, 2005, "Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance." Journal of marketing 69.2: 24-41.

Lucian A. Bebchuk, Alma Cohen, 2005, The costs of entrenched boards, Journal of Financial Economics, Volume 78, Issue 2, Pages 409-433.

Martínez-Ferrero, J, García-Meca, E, 2020, Internal corporate governance strength as a mechanism for achieving sustainable development goals. Sustainable Development, 28: 1189– 1198.

Mehran, Hamid, 1995, "Executive compensation structure, ownership, and firm performance." Journal of financial economics 38.2: 163-184.

Naoshi Ikeda, Kotaro Inoue, Sho Watanabe, 2018, Enjoying the quiet life: Corporate decision-making by entrenched managers, Journal of the Japanese and International Economies, Volume 47, Pages 55-69.

Ran Duchin, John G. Matsusaka, Oguzhan Ozbas, 2010, When are outside directors effective?, Journal of Financial Economics,Volume 96, Issue 2, Pages 195-214,

Sanjai Bhagat, Brian Bolton, 2008, Corporate governance and firm performance, Journal of Corporate Finance, Volume 14, Issue 3, Pages 257-273

Wintoki, M. Babajide, James S. Linck, and Jeffry M. Netter, 2012, "Endogeneity and the dynamics of internal corporate governance." Journal of financial economics 105.3: 581-606.

Sakawa, Hideaki, and Naoki Watanabel, 2018, "Family control and ownership monitoring in Stakeholder-oriented corporate governance." Management Decision 57.7: 1712-1728.

Surroca, Jordi, and Josep A. Tribó, 2008, "Managerial entrenchment and corporate social performance." Journal of Business Finance & Accounting 35.5-6: 748-789.

Yoshimori, Masaru, 1995, "Whose company is it? The concept of the corporation in Japan and the West." Long range planning 28.4: 2-44.

Yang, Tina, and Shan Zhao, 2014, "CEO duality and firm performance: Evidence from an exogenous shock to the competitive environment." Journal of banking & finance 49: 534-552.

## Appendix I. Variables

| | Variables | Description | Definition |
|---|---|---|---|
| **Independent Variables** | EN | EN firms | A dummy variable which takes a value of one if the company is 'Entrenched-firm', and zero otherwise. (A firm will be clustered as EN firm only when EN_similarity = 1 and DD_similarity=0) |
| | DD | DO firms | A dummy variable which takes a value of one if the company is 'Directors-Dominated' firm, and zero otherwise. (A firm will be clustered as DD firm only when EN_similarity = 0 and DD_similarity=1) |
| | ED | ED firms | A dummy variable which takes a value of one if the company with both the nature of 'Entrenched-firm' and 'Directors-Dominated' firm, and zero otherwise. (A firm will be clustered as ED firm only when EN_similarity = 1 and DD_similarity=1) |
| | NN | NN firms | A dummy variable which takes a value of one if the company with neither the nature of 'Entrenched-firm' nor the characteristic of 'Directors-Dominated' firm, and zero otherwise. (A firm will be clustered as NN firm only when EN_similarity = 0 and DD_similarity=0) |
| | EN_similarity | Weather the similarity of text between the firm and the EN sample larger than 50%. | A dummy variable which takes a value of '1' if the similarity of text between the firm and the EN sample larger than 50%. EN = 1 when EN_similarity=1 and DD_similarity=0. |
| | DD_similarity | Weather the similarity of text between the firm and the DD sample larger than 50%. | A dummy variable which takes a value of '1' if the similarity of text between the firm and the DD sample larger than 50%. DD = 1 when EN_similarity=0 and DD_similarity=1. |
| **Corporate Governance** | BSIZE | Board size | The number of board members/ln(Total Asset) |
| | EXERTO | Ratio of directors who concurrently serve as executive officers | Number of directors who concurrently serve as executive officers / Number of directors x 100% Excluding Audit Committee (directors) |
| | IDW_RTO | female directors' ratio | The ratio of the number of female directors to the total number of directors. |
| | DIR | directors' shareholdings | directors' shareholding ratio |
| | IND | independent directors | the number of independent directors |
| | CROSS | cross-shareholding ratio | Total percentage of shares held by publicly traded companies with which mutual shareholdings are possible |
| **Control Variables** | LEV | Financial leverage | The ratio of total debt to total assets. |
| | FSIZE | Ln (Total Asset) | The natural logarithm of total assets. |
| **Dependent Variables** | Tobin Q | Market Value | (Market capitalization + total liabilities)/total assets |
| | ROA | ROA | The ratio of earnings before interest and taxes to total assets. |

## Appendix 2. The Dictionary for ESG Topic

We extract ESG related word list from the MD&A part of NN groups by using LDA topic model. Stop words which appeared in 80% documents are filtered out before processing of natural language. In the Table V, we calculated the number of words in the Dictionary of Environment and Society and Dictionary of Governance in each text as the explained variable.

| **ES_Dictionary** | **G_Dictionary** |
|---|---|
| 社会,実現,貢献,持続,変革,創出,創造,サステナビリティ,長期,人財,人材,削減,ソリューション,気候変動,多様,再生可能エネルギー,新型コロナウイルス感染症,グローバル,エネルギー,中長期,ESG,ステークホルダー,未来,理念,排出,SDGs,従業員,Sustainable,持続的,炭素 | コーポレート・ガバナンス,体制,取締役会,取締役,社外取締役,監査,中期経営計画,委員会,役員,ステークホルダー,監督,買収,再発防止,報酬,透明性,内部統制,海外投資家,買い付け,役職,開示,監査役,投資家,内部監査,措置,法令,社員,執行役員,持株会社,新株予約権,対抗措置,買収,ガバナンス,株主総会,定時株主総会,買付け,特別委員会,情報開示,議決権,選任,社外 |