HONG KONG MONETARY AUTHORITY

Research Memorandum 05/2020

29 October 2020

WHAT CAN WE LEARN FROM ANALYSING LISTED FIRMS' ESG REPORTS? – OBSERVATIONS FROM HONG KONG BASED ON TEXTUAL ANALYSIS

Key points:

- *ESG investing refers to the consideration of non-financial factors involving Environmental (E), Social (S) and Governance (G) issues alongside financial factors in the investment decision-making process. ESG funds are found to be relatively resilient to the market turbulence during the COVID-19 pandemic, suggesting that ESG factors significantly differentiate between stock valuations among firms.*

- *To broaden the understanding of potential channels through which ESG factors may affect firms' stock valuation, this study analyses the annual ESG-related textual disclosure of firms listed in Hong Kong since the introduction in 2016 of mandatory disclosure requirement by the Stock Exchange of Hong Kong.*

- *Using a range of computer-based textual analysis techniques, this study first examined the key attributes of ESG disclosures of these firms and found four key observations: (i) the length of ESG disclosures has broadly increased since 2016, with firms in sectors more exposed to environmental issues, such as utility and energy, tending to disclose more information; (ii) the importance of environmental issues has increased in ESG disclosures, while social and governance-related disclosures continue to dominate; (iii) forward-looking information accounted for an important part of ESG disclosures; and finally (iv) the comparability of ESG reports among firms has increased over time, while there was little improvements in reports' readability.*

- *This study further examined how the key attributes of ESG disclosure would affect the bid-ask spread and return volatility of stocks. Empirical findings suggest that firms can benefit from ESG disclosure, because it improves informational efficiency and reduces the uncertainty of stock valuation. Specifically, firms with more ESG disclosure and more forward-looking information are found to have a lower bid-ask spread of the*

stock price. In addition, firms that disclosed more environmental-related and forward-looking information will reduce return volatility.

- *These findings together show that firms' exposure to ESG risks are one important source of uncertainty in their stock valuations, and such uncertainty can be reduced effectively by their ESG disclosure. The fact that firms can benefit from being more transparent in ESG issues gives strong support to regulators' continuing efforts to improve firms' ESG disclosure.*

Prepared by: *Gabriel Wu*
Market Research Division, Research Department
Hong Kong Monetary Authority

* The author would like to thank Eric Wong and Tom Fong for their comments and suggestions.

1. INTRODUCTION

ESG investing refers to the consideration of non-financial factors related to Environmental (E), Social (S) and Governance (G) matter alongside financial factors in the investment decision-making process (Table 1).[1] ESG investing is found to be more resilient to the widespread market volatility caused by COVID-19, suggesting that ESG factors significantly differentiate between stock valuations among firms.[2] In particular, continuous inflows to ESG equity funds were observed during the extreme turbulence in March 2020 and continued to outpace conventional equity funds in the second quarter (Chart 1).[3] In addition, investment index tracking firms with good ESG performance outperformed the corresponding benchmark index during COVID-19 (Chart 2).

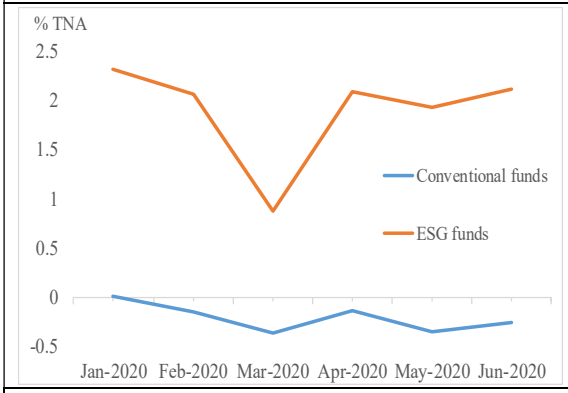
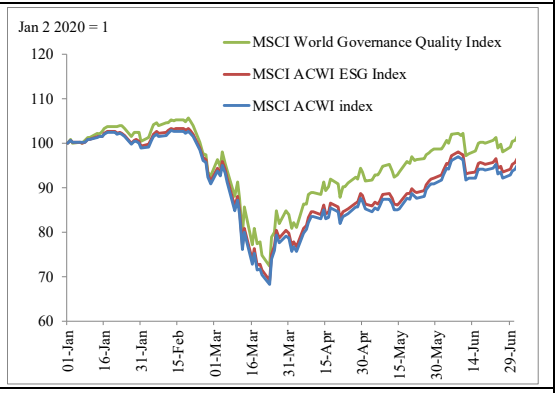
Table 1: Example of ESG issues

Source: CFA Institute, Franklin Templeton and author's adoption

[1] MSCI (2019) broadly divides ESG investing into three main areas based on different investment objective, which include i) improving risk-return characteristics of a portfolio, ii) aligning his portfolio with investors' norms and beliefs and iii) triggering change for social or environmental purposes.

[2] Firms with a sound ESG policy can better withstand the immediate impact of the pandemic as well as its longer-term effects. On the environment front, the pandemic demonstrates the damages brought by a large-scale public crisis, which will fundamentally change the long-term risk perspective. From the social perspective, corporates showing responsibility and commitment to society will likely win public support and be rewarded with brand loyalty. Finally, firms with strong governance will have disaster recovery and business continuity plans in place, making them agile enough to adapt to unforeseen crises.

[3] That said, the size of ESG equity funds remain a very small portion of total assets of all equity investment funds (3.4% as at Jun 2020 according to EPFR data).

Chart 1: Inflows into ESG equity funds in 2020	Chart 2: Performance of "Good-ESG" firms during the COVID-19
	
Note: Conventional and ESG funds are based on definition by source. Source: EPFR.	Source: Bloomberg

The successful implementation of ESG investing requires comprehensive information on ESG-related risk and opportunities faced by firms. The CFA Institute (2015) supports the systematic consideration of ESG issues as it leads to more complete investment analyses and better-informed decisions. In the wake of growing demand for corporates' transparency on ESG issues, international organisations such as the Financial Stability Board (FSB) and market regulators in different jurisdictions have been compiling requirements and guidelines on ESG disclosure (detailed in the next section).

In this study, a wide range of computer-based textual analysis techniques are applied to analyse ESG reports issued by listed firms in Hong Kong (HK). Unlike financial factors that are well captured by financial figures, the textual content in an ESG report is a key platform for firms to communicate their performance on ESG and resulting impacts. Textual ESG disclosures can be analysed with either a manual or computer-based approach. The manual approach requires human effort in methodically examining the contents and extracting the key information for analysis. An annual review published by the HKEX is one example of manual analysis on ESG reports by HK-listed firms. While manual analysis could be more precise and detailed, the time and efforts are so great such that they are often confined to a small sample.[4] In contrast, the computer-based approach can process a

[4] For instance, the latest review by HKEX covered 400 reports issued for 2018 financial year, as compared

large amount of text within a short period of time, making big-sample analysis possible. In addition, computer-based techniques can also transform textual contents into different attributes and express them in numeric form (henceforth referred to as numerical attributes) for objective comparison and quantitative analysis (Li, 2010). Our study can complement existing manual analysis in these two dimensions.

We also study empirically how information in ESG disclosures could affect the stock valuation of a firm, focusing on the effect through informational efficiency and volatility. From the financial stability perspective, inadequate information about ESG-related risks can lead to a mispricing of assets and misallocation of capital, and can potentially give rise to concerns about financial stability as markets can be vulnerable to abrupt corrections (TCFD, 2016). Therefore, this study examines whether more transparency on ESG-related issues through firms' ESG disclosure could improve the functioning of the stock market (Dhailwal et al., 2012 and Siew et al., 2016). More specifically, we study the effect of ESG disclosure on price information efficiency and volatility. The former helps to inform whether a more transparent disclosure on ESG could reduce the information asymmetry between listed firms and investors, and promote price discovery in the market. Our study also examines whether ESG disclosures reduce the return volatility of firms' stocks, as arguably, ESG disclosures may reduce uncertainty arising from the exposure of firms to ESG risks.

The study is organised as follow. The next section provides a brief overview on the development of the ESG disclosure practice and requirements around the world. Section 3 lays out our textual analysis of ESG reports, and presents key observations. Section 4 discusses our empirical models and results of the effect of ESG disclosures on listed firms' stock valuation. The last section concludes.

to near 2,000 reports published for the same financial year (Chart 5).

2. REVIEW ON THE DEVELOPMENT OF ESG DISCLOSURE PRACTICE AND REQUIREMENT

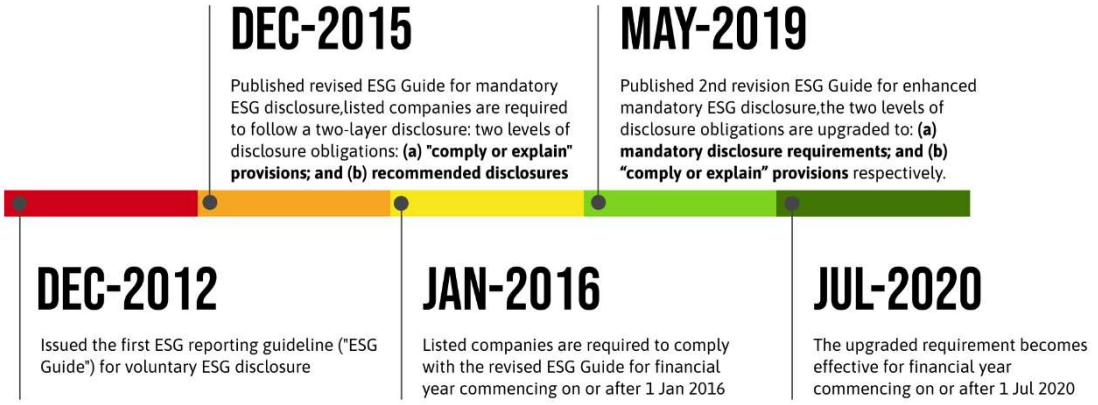
ESG was first raised in 2004 in a joint initiative by 23 major financial institutions, invited by then United Nations Secretary-General, Kofi Annan, to develop guidelines and recommendations on how to better integrate ESG issues in asset management, securities brokerage services and associated research functions. One of the key recommendations made by the initiative was that firms should provide information and reports on related performance in a more consistent and standardised format. To achieve this, international organisations and local market regulators have been developing guidelines or standards on firms' ESG disclosure. On the international front, the Global Reporting Initiative (GRI) launched the first global standards for ESG reporting in 2016 and remained the most widely adopted standard by stock market regulators.[5] In 2017, the Task Force on Climate-related Financial Disclosures (TCFD), created by the FSB,[6] issued a final recommendation on firms' voluntary financial disclosure on climate-related risks.

Stock market regulators have adopted a different pace on regulating firms' ESG disclosures. According to the Sustainable Stock Exchange Initiative (SSE), a United Nation partnership programme on enhancing ESG disclosure, 55 of the 103 stock exchanges (accounting for 64% of total market capitalisation) tracked by the SSE have so far published written ESG guidance for their listed firms. Among these stock exchanges, 24 (including HK, Chart 3) have a listing rule requirement that listed firms must publish ESG reports (i.e. compulsory ESG disclosure).

[5] Established in 1997, GRI is an independent international organization that pioneers ESG reporting. It published the first ESG reporting guideline in 2000 before announcing the first global standard for ESG reporting in 2016. According to SSE, 91% of the stock exchange who published written ESG reporting guidance have referenced GRI.

[6] TCFD was created in 2015 by the FSB to develop consistent climate-related financial risk disclosures for use by firms, banks, and investors in providing information to stakeholders.

Chart 3: Timeline on ESG disclosure requirements by the HKEX



Notes: (1) HKEX's ESG Guide focused two ESG subject areas, specifically Environmental and Social, where they are further divided into several provisions where each provision is subject to either one layer of disclosure obligations. Governance is addressed separately in the Corporate Governance Code. (2) For provisions falling under "comply or explain" requirement, issuer of the ESG report must provide reasons in its ESG report if the issuer does not report on a particular provision.

3. TEXTUAL ANALYSIS ON ESG REPORTS

3.1 What can we obtain with textual analysis?

There is a variety of "outputs" we could extract from textual analysis for characterising ESG disclosures. As shown in Chart 4, we adopt a two-layered approach where we first convert the textual contents into a number of numerical attributes for characterising and comparing different reports. Next, we take a step further and analyse the contents in more detail. Table 2 describes each analysis performed, including the measurement used, interpretation and relevance in understanding the ESG reports. Full technical details are given in Appendix A

Chart 4: Schematic view of the two-layers textual analysis

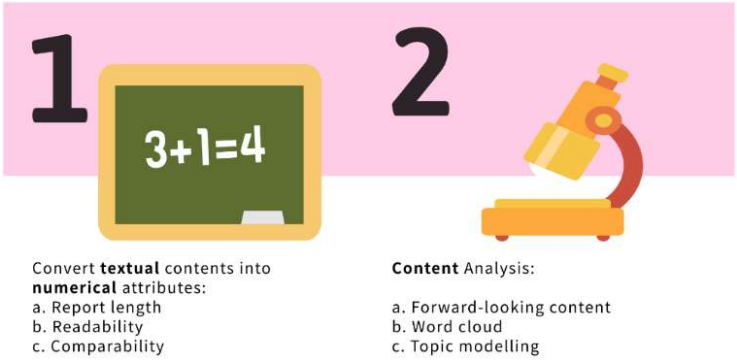


Table 2: Summary on textual analysis performed

Category	Analysis	Description
Converting textual contents into numerical attributes	Report length	Measured by the number of non-stop[7] words in the report. This provides a direct measure on how informative an ESG report is. All else equal, we expect a longer ESG report to be more informative.
	Readability	Measured by the Flesh Reading Ease (FRE) that quantifies the complexity of text based on (i) average sentence length and (ii) average number of syllables per word. The core idea is that longer sentences and longer words are harder to read, and therefore a lower FRE. Readability of text affects how effective the ESG-information could be communicated with external stakeholders.
	Comparability	Measured by the similarity of two ESG reports which compares the relative word frequencies across them. The comparability of an ESG report is then defined as median value of the similarity against its peer ESG reports. Comparability between ESG reports affect investors' ability to make fair comparison on the ESG-risk borne by firms.
Content analysis	Forward-looking content	Measured by the share of forward-looking sentences in the report, it gauges how much future ESG opportunities and risks are discussed in the report.
	Word cloud	A popular visualisation tool for textual analysis, word cloud allows us to easily identify the words/topics that are most commonly discussed in a given set of ESG reports.
	Topic modelling	This statistical method condenses the textual contents into a number of topics, where each topic is presented as a weighted combination of key words. Each sentence is then assigned with the topic that has the highest probability of matching the words contained in that sentence, where we could further identify the dominant topic of each ESG report as the one that represents most of the sentences in that report.

[7] Stop words are the commonly used words that do not add much meaning to a sentence, such as "a", "an", "we" and "the". The word count with stop words excluded could better reflect the "informative" contents in the report. The full stop word list is obtained using the "NLTK" package in Python.

3.2 Sample of ESG reports

We gather HK-listed firms' annual ESG reports from HKEX's HKEXnews (披露易) website,[8] a centralised platform that provides access to all disclosures and announcements by firms listed in the HKEX. In particular, the platform allows us to filter the ESG report released by each listed company. A total of 6385 English ESG reports published as at 29 May 2020 were downloaded from the platform.[9] The ESG reports issued by HK-listed firms come into two main forms, either a standalone ESG report or an "ESG" section in a firm's annual report. The ESG reports obtained are fed into Python programmes for textual contents extraction and analysis.

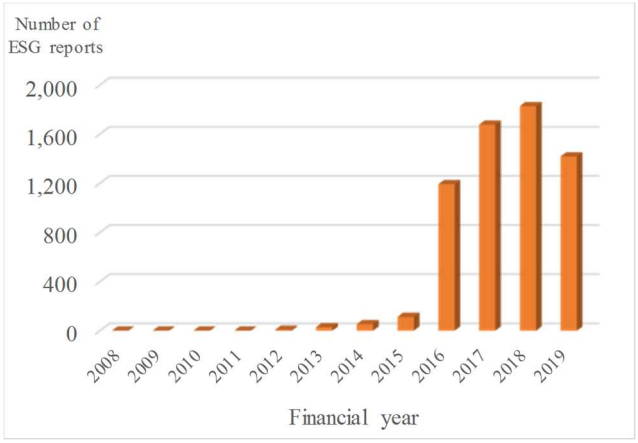
Chart 5 depicts the number of ESG reports issued by each financial year. A sharp jump in the number of ESG reports issued was observed in 2016, the year when HKEX's mandatory ESG disclosure requirement came into force. Since then, with the number of reports has kept increasing.[10] Given the few reports issued prior to 2016 financial year, this study focuses on ESG reports issued since then.

[8] https://www.hkexnews.hk/index.htm

[9] Like a full annual report, ESG reports issued by HK-listed firms are available in both English and Chinese. We adopt English reports in this study for two reasons. First, most computer programmes on textual analysis are written for analysing English text, as the range of analysis would be limited if we adopted Chinese reports instead. Second, those interested in the reports include not only local but also international investors, and the findings inferred from the English reports are more informative.

[10] The count for 2019 financial year is lower than previous year is due the fact that we only covered reports issued on or before 29 May 2020, a number of reports will only be issued in July or later based on published schedule in previous years.

Chart 5: Number of ESG reports issued by financial year

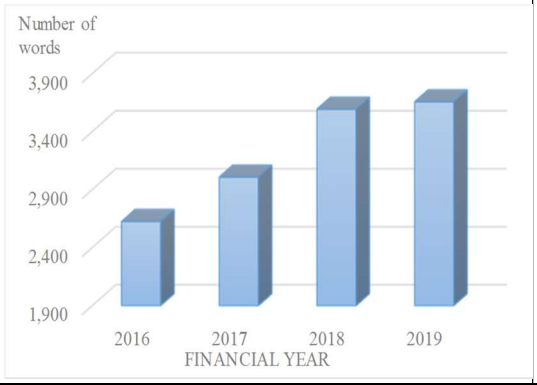
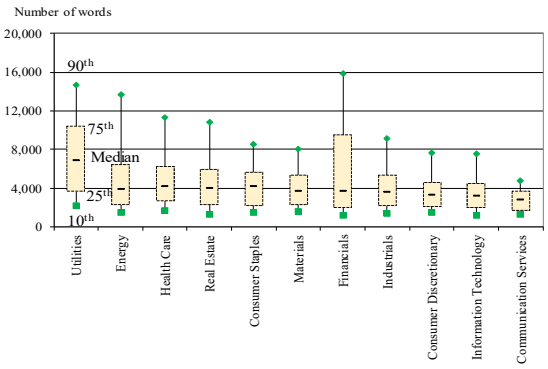


Note: Count for 2019 financial year is based on reports issued on or before 29 May 2020.

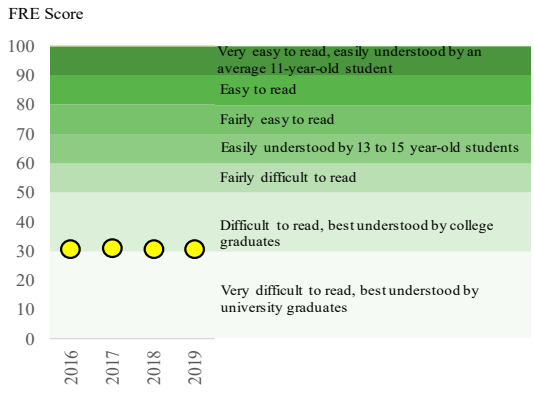
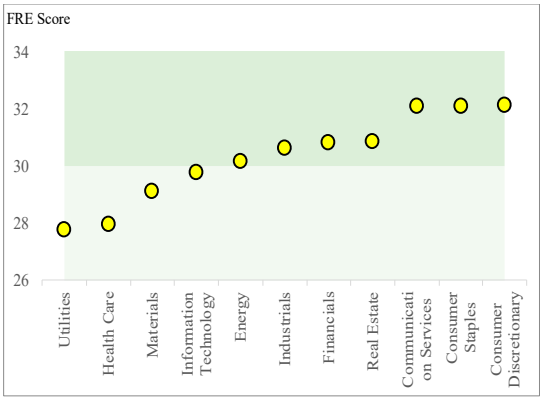
3.3 What can we tell from the numerical attributes on ESG reports?

Firms have increased their ESG disclosure, as judged by the length of text. Chart 6 depicts the median length of ESG reports by financial year. As can be seen, there was a rapid growth in the length of ESG reports between 2016 and 2018, before slowing down in 2019.[11] The industry breakdown in the Chart 7 box plot shows that firms in the utilities, energy or health care sectors tend to disclose more ESG information, which may be explained by the fact these firms generally serve the public interest or are more exposed to environmental issues. As such, investors could be more concerned about their ESG performance. It is also worth noting that ESG reports published by the financial sector display a much larger deviation in report length within the sector when compared to other sectors.

[11] This may be explained by the COVID-19 that could possibly interrupt the incorporations of new elements into the ESG reports.

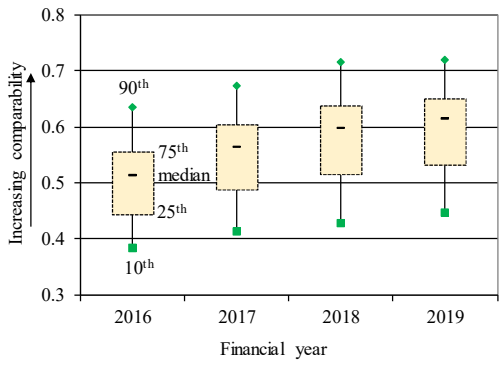
Chart 6: ESG Report length by year	Chart 7: ESG report length by sector
	
	Note: The box plots dipicts the 10th, 25th, 50th (median), 75th and 90th percentile of the ESG report length in each industry.

There is no substantial improvement in the readability of ESG reports. Chart 8 depicts the median FRE score of ESG reports by financial year, with a higher FRE score indicating better readability. However, there is no notable increase in the readability of ESG reports as the median FRE score stays at around 30 (i.e. the yellow spots in Chart 8). Indeed, based on the mapping of FRE scores to levels of readability (i.e. the shaded regions in Chart 8), it indicates that it generally requires at least college graduates to understand the contents of ESG reports. Next, Chart 9 suggests the readability of ESG reports is notably lower for firms in the health care and utility sector. These reports not only tended to have longer sentences, they were also inclined to use more complex words in their ESG disclosures (Appendix C).

Chart 8: Readability by year	Chart 9: Readability by industry
	
Note: Values refer to the sample median by financial year	Note: Values refer to the sample industry median.

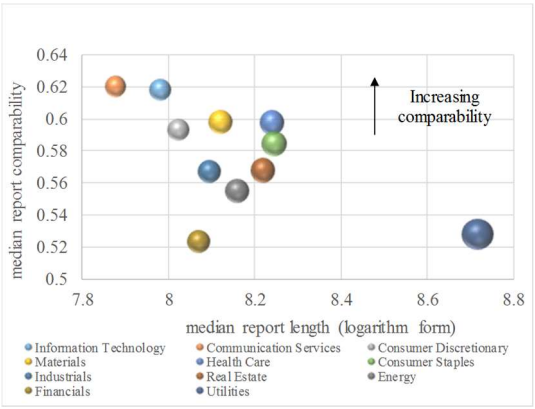
The comparability of ESG reports among firms has increased over time. Chart 10 depicts the box plots on the report comparability of ESG reports by financial year. In particular, the report comparability measure in Chart 10 refers to the median value of the cosine similarity against each of the other ESG reports issued in the same financial year. A broad-based improvement in comparability is observed, as the whole box plot shifts up across years. Despite the overall improvement in comparability over time, the scatter plot in Chart 11 shows that the comparability of ESG reports among firms within the same industry tends to be lower for those industries with a higher median length of ESG reports.

Chart 10: Report similarity by financial year



Notes: (1) The box plots dipicts the 10th, 25th, 50th (median), 75th and 90th percentile of report similarity of ESG reports issued in each fianncial year. (2) For each ESG report, the report similarity is defined as the median value of the consine similiarty against each of the other ESG reports issued in the same financial year.

Chart 11: Scatter plot on industry-wide report similarity against length

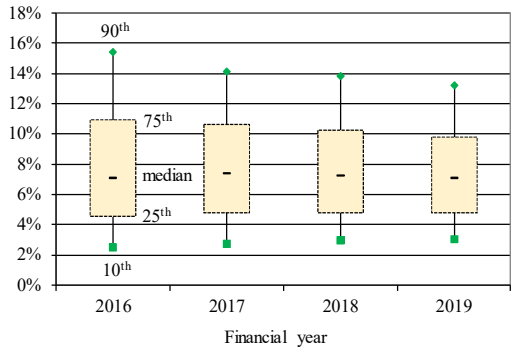
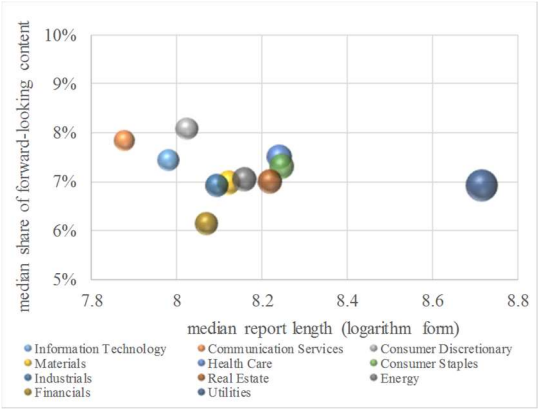


Note: Size of bubble denotes report median ESG report length in each industry. (2) For each ESG report, the report similarity is defined as the median value of the consine similiarty against each of the other ESG reports issued by firms in the same industry.

3.4 What are discussed in the ESG reports?

Forward-looking information accounted for a significant part of ESG reports. Chart 12 displays the box plots on the percentage of forward-looking sentences in ESG reports by financial year. As can be seen, the median

value remained steady at 7%. It is worth noting that the box plot narrows over time, suggesting there are less dispersions among firms on allocating the report to forward-looking information. The scatter plot in Chart 13 shows there is no notable difference in the share of forward-looking statement across industries (with the ratio spanning between 6% and 8%), and the share appears to have no strong association with the report length.

Chart 12: Share of forward-looking content by year	**Chart 13: Share of forward-looking content by sector**
	
Note: The box plots dipicts the 10th, 25th, 50th (median), 75th and 90th percentile of share of forward-looking contents of ESG reports issued in each fianncial year.	Note: Size of bubble denotes median report length in each industry.

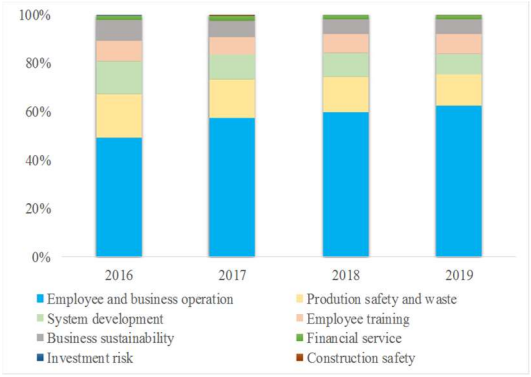
Environmental issues have gained importance in ESG disclosures, although social and governance-related disclosures continue to dominate. Chart 14 presents the results using word cloud analysis, with the more notable words and topics being more commonly discussed in ESG reports. Social or governance-related matters consistently dominated ESG reports over the years, with the majority of phrases including "law", "regulation", "health" and "safety" (see Table 1 on the issues classified under these two aspects). The dominance of social and governance issues is further confirmed by the results of topic-modelling analysis. Chart 15 shows that more than half of the ESG reports are dominated by the topic "Employee and business operation" (i.e. light blue portion), which contains not only employee (social-related), but also policy and law (governance-related) in its top 5 keywords (Appendix D). Nevertheless,

environmental issues have gained the importance in ESG reports in recent financial years. In particular, Chart 14 indicates that more environmental-related subjects such as “energy consumption” and “hazardous waste”, were commonly discussed in ESG reports in the two most recent financial years, while “greenhouse gas” and “gas emission” remained common environment-related topics covered by ESG reports.

Chart 14: Word clouds on most common words in ESG reports

(a) 2016	(b) 2017
	
(c) 2018	**(d) 2019**
	

Chart 15: Distribution of dominant topic in ESG reports



Note: This chart shows the distribution of dominant topic among the sample ESG reports, defined as the topic that represent most of the sentences in that report. Each topic is defined based on five words that receive the highest weight in each topic.

4. EMPIRICAL ANALYSIS ON THE EFFECT OF ESG DISCLOSURES ON LISTED FIRMS STOCK VALUATION

4.1 Model specification

We adopt two simple panel regression models to evaluate the impact of ESG disclosure on the price information efficiency and return volatility of listed firms. First, we consider Equation (1) to evaluate the relationship between listed firms' ESG disclosures and bid-ask spreads, a proxy for information efficiency:

$$BA_{i,t} = \alpha_{i,t} + \sum_m \beta_{i,t}^m ESG_{i,t}^m + \sum_k \gamma_{i,t}^k Control_{i,t}^k + FE_i + FE_t + \delta_{i,t} VIX_{i,t} + \varepsilon_{i,t} \quad (1)$$

$BA_{i,t}$ refers to the one-year average of bid-ask spread after the release of company i's ESG report for financial year t.[12] $ESG_{i,t}$ is a list of ESG disclosure measures we derived in the previous section. Specifically, it covers three numerical measures on the ESG report that includes i) report length (*LEN*), ii)

[12] For observations concerning 2019 financial year, BA and other related variables instead capture the information from the time of report issuance to 31 Jul 2020, effectively three months after the issuance (since most sample reports for 2019 financial year were issued during April 2020). The results are robust to excluding observations of 2019 financial year.

readability (*READ*), iii) comparability (*COM*) and iv) share of forward-looking statement (*FLS*). It also includes a dummy variable "*DOMTOPIC*" that equals one when the dominant topic of the ESG report contains at least one "environment" words in its top 5 key words (Appendix D). Meanwhile, $Control_{i,t}$ is a series of control variables that could affect firm's information efficiency (Siew et al., 2016), including i) firm size (*SIZE*), ii) leverage (*LEV*), iii) inverse of stock price (*INVPRICE*), and v) annual return volatility (*VOL*). The model also includes firm, industry and time-fixed effects. Given the different release times of firms' ESG reports, and therefore the different time period captured by $BA_{i,t}$, the time-fixed effect alone is not enough to capture all common effects with for a certain financial year. Therefore, we include the VIX index (*VIX*) as an additional control variable. Appendix B provides the definition of these variables.

Next, the following Equation (2) is considered for the relationship between a listed company's return volatility and ESG disclosure;

$$Vol_{i,t} = \alpha_{i,t} + \sum_m \beta_{i,t}^m ESG_{i,t}^m + \sum_k \gamma_{i,t}^k Control_{i,t}^k + FE_i + FE_{ind} + FE_t + \delta_{i,t} VIX_{i,t} + \varepsilon_{i,t} \quad (2)$$

$Vol_{i,t}$ refers to the one year stock return volatility after the release of the ESG report at financial year t for company i. $ESG_{i,t}$ is the same list of ESG disclosure measures included in Equation 1. Similarly, $Control_{i,t}$ contains a series of control variables that could affect return volatility, which include i) firms' size (*SIZE*), ii) profitability (*ROA*) and iii) average return in previous year (*PRERETURN*). Same as Equation 1, the model also includes firm, industry and time-fixed effect, as well as the VIX index as control variable.

4.2 Results

First, more ESG disclosures and more forward-looking contents could lead to better price information efficiency. Estimation results

for Equation 1 are given in Table 3. The first column shows that the report length (*LEN*) has a significant and negative effect on the bid-ask spread (i.e. -0.52). As *LEN is* measured by total word count in logarithm form, our estimation means that a one percent increase in report length is associated with a 0.0052 percentage point (p.p., i.e. -0.52*0.01) reduction in the bid-ask spread. Reports that are more forward-looking could also lead to a lower bid-ask spread. With an estimated coefficient of -0.03, a one p.p. increase in the share of forward-looking statements is associated with a 0.03 p.p. reduction in the bid-ask spread. As a lower bid-ask spread implies a better price efficiency, this suggests firms' stock price efficiency improves with more ESG disclosures and when the ESG report is more forward-looking.[13] For other ESG measures, we do not find a linear relationship between readability (*READ*, with an estimation coefficient of 0.01), comparability (*COM*, with an estimation coefficient of 0.66) and bid-ask spread. There is also no significant difference whether or not the dominant topic of an ESG report covers environmental issues (*DOMTOPIC*, with an estimation coefficient of -0.13).[14]

An alternative model specification is considered to investigate the potential impact of *READ* and *COM* on the bid-ask spread. Specifically, two dummy variables *D_ READ and D_COM* were considered respectively, which equal 1 if *READ* or *COM* is larger than the sample median, and 0 vice versa. Equation 1 is re-estimated and the results are given in Column 2. The significant negative impact of *LEN* and *FLS* remains robust to this alternative specification (i.e. -0.51 and -0.03 respectively). The estimated effect of readability and comparability consistently remains insignificant with the alternative

[13] The effect found here may be due to both information (i.e. more information is provided regardless of company's ESG performance) and signalling (i.e. firms with better ESG performance are willing to provide more ESG disclosure) effect. We cannot rule out the effect of the latter here, as an objective measure of ESG performance for firms across industries is not available and we cannot include such a control variable. Meanwhile, a longer report may also be subjected to the "green washing" effect, meaning a lot of ESG information (and misleading information in extremity) could be provided by a company with a poor ESG performance. Therefore, the "signalling" and "green washing" effect may counteract each other here.

[14] Based on 5 words that receive the most weights under each topic (see Appendix D), ESG reports are assigned with value 1 for variable DOMTOPIC if the dominant topic of that report is Topic #1 (containing "energy"), #2 (containing "sustainability") and #5 (containing "waste").

specification.

Table 3: Effect of ESG disclosure on bid-ask spread

Dependent variable	$BA_{i,t}$	
	(1)	(2)
LEN	**-0.52****	**-0.51****
READ	0.01	
COM	0.66	
FLS	**-0.03*****	**-0.03****
DOMTOPIC	-0.13	-0.15
D_READ		0.09
D_COM		-0.06
SIZE	**-0.71*****	**-0.71*****
LEV	0.02	0.02
INVPRICE	**0.04*****	**0.04*****
VOL	**0.02*****	**0.02*****
VIX	0.03	0.03
Observations	3619	3619
Time effect	Yes	Yes
Stock effect	Yes	Yes
Industry effect	Yes	Yes
Number of stocks	1571	1571

Second, ESG reports with better readability, more forward-looking statements and also include environmental-related information, may help reduce return volatility. Table 4 presents the estimation results of Equation 2. In particular, Column 1 shows that ESG reports whose dominant topic covers environmental issues are associated with a 2.7 p.p. reduction in annualized return volatility (i.e. -2.7 in Column 1). Other measures on ESG reports do not show a significant relationship with return volatility.

Same as the bid-ask spread case, we next transform all level variables that are not significant (i.e. *LEN*, *READ, COM* and *FLS*) into dummy variables (i.e. *D_LEN*, *D_READ, D_COM* and *D_FLS*) for further investigation. Column 2 shows the estimation results of this alternative specification. In particular, firms whose ESG reports are more readable (specifically with *READ*

larger than the sample median) will have a significantly lower return volatility (i.e. -1.43 in Column 2). This suggests that while the readability of ESG reports does not exert a continuous and linear effect on return volatility, it could result in a significant reduction in return volatility once it exceeds a certain level. The level of forward-looking statements (*FLS*) in ESG reports is also found to affect return volatility in a similar manner. (i.e. -1.29 in Column 2). Lastly, the length of the ESG report does not have a significant impact on return volatility, both in level and dummy forms, suggesting that having a longer report alone is not enough to reduce the uncertainty of ESG-related risks on investors. Taken together, ESG reports with better readability, more discussion of environmental and forward-looking issues will facilitate a better assessment of a firm's ESG-related risks. This may contribute to lower uncertainty and thus lower return volatility of the stock.[15,16]

Table 4: Estimated effect of ESG disclosure on return volatility

Dependent variable	$Vol_{i,t}$	
	(1)	(2)
LEN	-0.38	
READ	-0.02	
COM	-9.34	
FLS	6.40	
DOMTOPIC	**-2.70****	**-2.76****
D_LEN		0.41
D_READ		**-1.43***
D_COM		-1.16
D_FLS		**-1.29***
SIZE	**-5.05*****	**-5.07****
ROA	**-0.31****	**-0.32****
PRERETURN	0.64	0.56

[15] Like the impact of the ESG report length on the bid-ask spread we found earlier, the effect documented here may also be attributed to the fact that firms with a better ESG performance could disclose ESG-information in a more direct and readable manner. They would also be more willing to disclose information about the future. As documented in Barko et al (2018), Bialkowski and Starks (2018), firms with a better ESG performance or engagement may also result in a lower price volatility, as investors would be more certain of these firms' fundamental values.

[16] The argument is also consistent with the findings by Dhailwal et al. (2012), which show the availability of non-financial disclosure (proxied by the issuance of corporate social responsibility report in their study) is able to reduce analyst forecast error on future's earnings per share.

VIX	**0.33****	**0.31****
Observations	3457	3457
Time effect	Yes	Yes
Stock effect	Yes	Yes
Industry effect	Yes	Yes
Number of stocks	1484	1484

5. CONCLUSION AND IMPLICATION

The COVID-19 pandemic has prompted a renewed focus on ESG, acting as a wake-up call for investors to prioritize a more sustainable approach to investment (JP Morgan, 2020). The successful implementation of ESG investing requires comprehensive information on listed firms' ESG policies and exposures. HK, among other stock market regulators, has intensified efforts to improve listed firms' disclosures on ESG-related matters. To broaden the understanding of potential channels through which ESG factors may affect firms' stock valuation, this study analyses the annual ESG-related textual disclosure of firms listed in HK since the introduction of mandatory disclosure requirements by the HKEX in 2016.

Using a range of computer-based textual analysis techniques, this study first examined the key attributes of ESG disclosures of these firms and found four key observations: (i) the length of ESG disclosures has broadly increased since 2016, with firms in sectors more exposed to environmental issues, such as utility and energy, tending to disclose more; (ii) the importance of environmental issues has increased ESG disclosures, although social and governance-related disclosures continue to dominate; (iii) forward-looking information accounted for an important part of ESG disclosures; and finally (iv) the comparability of ESG reports among firms has increased over time, while there was little improvements in reports' readability.

This study further examined how the key attributes of ESG disclosure would affect the bid-ask spread and return volatility of stocks.

Empirical findings suggest that firms can benefit from ESG disclosure, as it improves informational efficiency and reduces the uncertainty of stock valuation. Specifically, firms with more ESG disclosure and more forward-looking information are found to have a lower bid-ask spread of the stock price. In addition, firms that disclose more environmental-related and forward-looking information can reduce return volatility. Together, these findings show that the exposure of firms to ESG risks is one important source of uncertainty in their stock valuations, and such uncertainty can be reduced effectively by their ESG disclosure. The fact that firms can benefit from being more transparent in ESG issues will provide a strong support to regulators in their ongoing efforts to improve firms' ESG disclosures.

Overall, this study complements the existing manual analysis of compliance of ESG reports issued by HK-listed firms. In particular, the study shows glimpses on how technology can help the analysis of textual ESG reports, an unstructured yet important source of ESG information to stakeholders, in a systematic manner. In addition, the quantifications on textual contents allow us to understand better on how information published on ESG reports can be channelled to the stock market. That said, there are plenty of areas which merit further exploration with technology, in particular a deeper content analysis of issues such as i) the identification and quantification of the materiality assessments[17] in ESG reports, ii) the separation of high quality reports from those following the "tick-the-box" approach and iii) dividing the disclosure into environmental, social and governance-related elements for further analysis.[18] These are left for future research.

[17] Materiality assessment refers to the identification of ESG issues determined by the board that are sufficiently important to investors and other stakeholders, which are instrumental in setting the tone of a company's overall thinking and approach to ESG issues (HKEX, 2019).

[18] These analyses would require more "intelligence" in the sense that computer programs need to be trained for performing the tasks accurately. Substantial human efforts are required in feeding these "intelligence" into the program.

REFERENCES

Barko, T., Cremers, M. and Renneboog, L. (2018): "Shareholder Engagement on Environmental, Social, and Governance Performance", available at https://ecgi.global/sites/default/files/working_papers/documents/finalbarkocremersrenneboog.pdf.

Bialkowski, J. and Starks, L.J. (2016): "SRI Funds: Investor Demand, Exogenous Shocks and ESG Profiles", *Working Papers in Economics*, 16/11, University of Canterbury, Department of Economics and Finance.

CFA (2019): "Environmental, Social and Governance issues in Investing"

Dhaliwal, D., Radhakrishnan, S., Tsang, A. and Yang, Y.G. (2012): "Nonfinancial Disclosure and Analyst Forecast Accuracy: International Evidence on Corporate Social Responsibility Disclosure", *The Accounting Review*, 87(3), 723-759

Dyer, T., Lang, M. and Stice-Lawrence, L. (2017): "The evolution of 10-k textual disclosure: Evidence from Latent Dirichlet Allocation", Journal of Accounting and Economics, 64(2), 221-245

Goloshchapova, I., Poon, S.H., Pritchard, M. and Reed, P. (2019): "Corporate social responsibility reports: topic analysis and big data approach", The European Journal of Finance, 25(17), 1637-1654

Jenkins, R. (2020, Jun 1): "COVID-19: Flows into ESG funds continue during 1Q20," *Refinitiv*, Retrieved from https://www.refinitiv.com/perspectives/market-insights/covid-19-flows-into-esg-funds-continue-during-1q20/

Lang, M. and Stice-Lawrence, L. (2015): "Textual Analysis and international financial reporting: Large sample evidence", *Journal of Accounting and Economics*, 60(2), 110-135

Li, F. (2010): "Textual Analysis of Corporate Disclosure: A Survey of the Literature", *Journal of Accounting Literature*, 29, 143-165

MSCI (2019): “The MSCI Principles of Sustainable Investing”

Siew, R.Y.J., Balatbat, M.C.A. and Carmichael, D.G. (2016): “The impact of ESG disclosures and institutional ownership on market information asymmetry”, Asia-Pacific Journal of Accounting & Economics, 23(4), 432-448

Tao, J., Deokar, A. and Deshmukh, A. (2018): “Analysing forward-looking statements in initial public offering prospectuses: a text analytics approach”, *Journal of Business Analytics*, 1(1), 54-70

Task Force on Climate-related financial Disclosures (2016): “Phase I Report of the Task Force on Climate-Related Financial Disclosures”

Task Force on Climate-related financial Disclosures (2017): “Why and How Investors Use ESG Information: Evidence from a Global Survey”

The Stock Exchange of Hong Kong (2019): “Analysis of Environmental, Social and Governance Practice Disclosure in 2018”

The Stock Exchange of Hong Kong (2020): “How to prepare an ESG report – A step by step guide to ESG reporting”

Appendix A: Technical details of the textual analysis

1. Converting textual contents into numerical attributes

   a. ESG report length

      The first attribute we consider is the length of the ESG disclosure. All else equal, we expect a longer ESG report to be more informative. The length of disclosure is measured by total word count after removing the stop words. Stop words are the commonly used words that do not add much meaning to a sentence, such as "a", "an" and "the". We remove these words from the length count to better reflect the information content of the ESG reports extracted.

   b. Readability

      A readability score is an objective measure of the complexity of text. It is important such that the information can be effectively communicated to the public. Readability of ESG report is measured by Flesch Reading Ease (FRE) in this study, one of the commonly used readability score in literature. It is also one of the readability measure advocated by former SEC chairman Christopher Cox on the plain English motive for listed firms textual disclosure.[19] Developed by Rudolph Flesch (an Austrian-American readability expert) in 1948, FRE is calculated using the following formula;

$$FRE = 206.823 - 1.015 * Average\ \ sentence\ \ length - 84.6 * Average\ \#\ syllables\ of\ \ words$$

      In short, the FRE measures readability based on the average sentence length and average number of syllables per word. The core idea is that longer sentences and longer words are harder to read, and therefore a

[19] "Just as the Black-Scholes model is a commonplace when it comes to compliance with the stock option compensation rules, we may soon be looking to the Gunning-Fog and Flesch-Kincaid models to judge the level of compliance with the plain English rules."

lower FRE. Table A.1 displays the classification of FRE score with respective readability level.

Table A.1: Classification of FRE scores

FRE score range	Readability level
90-100	very easy to read, easily understood by an average 11-year-old student
80-90	easy to read
70-80	fairly easy to read
60-70	easily understood by 13- to 15-year-old students
50-60	fairly difficult to read
30-50	difficult to read, best understood by college graduates
0-30	very difficult to read, best understood by university graduates

c. Comparability

It is widely documented the comparability of ESG reports across firms presents a challenge for market participants to incorporate ESG information for investment decisions (TCFD, 2017). The reason is intuitive: if the ESG-information reported in one report is very different to the others, investors would have difficulty in making fair comparison on the ESG-risk borne by firms.

We follow Lang and Stice-Lawerence (2015) and use cosine similarity as measure of comparability.[20] It measures similarity of two textual documents by comparing the relative word frequencies between them. In particular, a document can be represented by a term-frequency vector which stores the frequency of a series of words or phrases.[21] Then, the cosine similarity is captured by dot product of two term-frequency vectors

[20] The Financial Account Standards Board defines comparability as not only like things look alike, but different things should also look different. While comparability in this regard is difficult to be captured completely, Lang and Stice-Lawerence (2015) argued it is more sensible to focus on the former, i.e. comparing "like" things because it is easier to execute.

[21] Raw text is pre-processed with the most common words and punctuation removed, tokenization, and stemming before the term frequency vector is generated.

before scaled by the product of two vector lengths, as represented by the Equation A.1 (where A and B denotes any two given documents);

$$Similarity_{A,B} = \frac{A \cdot B}{|A||B|} \quad \text{(A.1)}$$

Cosine similarity is analogous to the correlation measure in numerical data analysis in two ways. First, it is also bounded between 0 and 1 where documents with identical proportions of words have the cosine similarity equals 1 (and 0 vice versa). Second, it is also a pair-wise measure which we can compute cosine similarity of any 2 reports. Given the pair-wise property, for each ESG report we need to first calculate the cosine similarity with each of its peer ESG report, before taking the median value of the set of cosine similarity as our final measure of comparability. Two peer groups are adopted, including i) all ESG reports issued in the same financial year and ii) all reports issued by firms from the same industry. This allows us to capture the comparability of a particular ESG report from both time and industry dimension.

2. Content analysis

In addition to characterise an ESG report with some general properties, computer-based textual analysis also allows us to perform certain content analysis on ESG reports. We first apply two popular methods, namely word cloud analysis and topic modelling, to diagnosis topics that are commonly discussed in the ESG reports. As an important purpose of ESG disclosure is to inform stakeholders on the future ESG-risks, we also adopt a dictionary-based approach to assess how forward-looking these ESG reports are. Brief descriptions of the three approaches are given below.

a. Forward-looking content

The impact of ESG issues on financial risks can be minimized if there is

a strategic approach that is forward-looking and reflects the long-term interests of the company (HKEX, 2020). It is therefore useful to measure the amount of "forward-looking" information presented in an ESG report is. In this study, we employ the dictionary approach and identify forward-looking sentences as those containing "forward-looking" words.[22] We then calculate share of forward-looking sentences (to the total number of sentences in the report) to measure the relative "forward-looking-ness" among ESG reports.

b. Word cloud analysis

Word cloud is a popular visualization tool in textual analysis. A word cloud displays the list of words that appear most frequently in the text. These words appear in varying sizes with the size representing the relative frequency of appearance. By pooling a series of ESG reports for word cloud analysis, we can identify the most discussed words, commonly across these reports. In addition to removing stop words that do not have any meaning, we also remove words that are expected to appear in the text but could not relevant as a topic, such that we could obtain as much information as possible from the word cloud analysis.[23]

c. Topic modelling

Topic modelling is a statistical method for identifying topics and other hidden patterns from a collection of textual object. It is a type of

[22] Our dictionary-based method combines the "Keyword matching" and "Linguistic patterns" approach in Tao and Deokar (2018). In particular, a sentence is tagged as forward-looking if it either i) contains any of the following words that indicate future, which include "will", "future", "aim", "anticipate", "assume", "commit", "estimate", "expect", "forecast", "foresee", "hope", "intend", "plan", "project", "seek", "predict", "worry", "optimistic", "pessimistic" and "target" or ii) contains combinations of adjectives such as "next", "subsequent", "following", "upcoming", "incoming", "coming", "succeeding", and "carryforward" and time indicators such as "month", "quarter", "year", "fiscal", "taxable", and "period".

[23] This includes, but not limited to, "Environment, Social and Governance" (or "ESG" in short), name of places such as "Hong Kong", "China" and company names. To identify such list, we first extract the list of most common words before filtering the words manually.

unsupervised machine learning algorithm as all topics are observed from the documents, rather than with prior knowledge in other rule or dictionary-based methods.

There are many topic modelling algorithms and in this study we adopt Latent Dirichlet Allocation (LDA) algorithm which is commonly used in analysing firms' disclosures (Dyer et al., 2017 and Goloshchapova et al., 2019)). LDA posits each text object is a mixture of a small number of topics, and each topic is represented by a set of words with certain probabilities of co-occurrence.[24] The process is akin to principal component analysis (PCA) in numerical data analysis in two ways. First, both method identify "latent" patterns (topic in LDA and principal component in PCA respectively) purely from the underlying data. Second, the latent pattern under both methods are represented by a weight sum of underlying data (words in LDA and data series in PCA) with the weight reflecting the importance of underlying data to that latent pattern. While we could identify as many topics as we want with LDA, some topics may actually be redundant as they could overlap with others. The optimal number of topics can be decided using the coherence score. it measures how topics are different from each other by looking at the extent of overlapping in key words. A higher coherence score would mean less redundancy in the topics extracted.

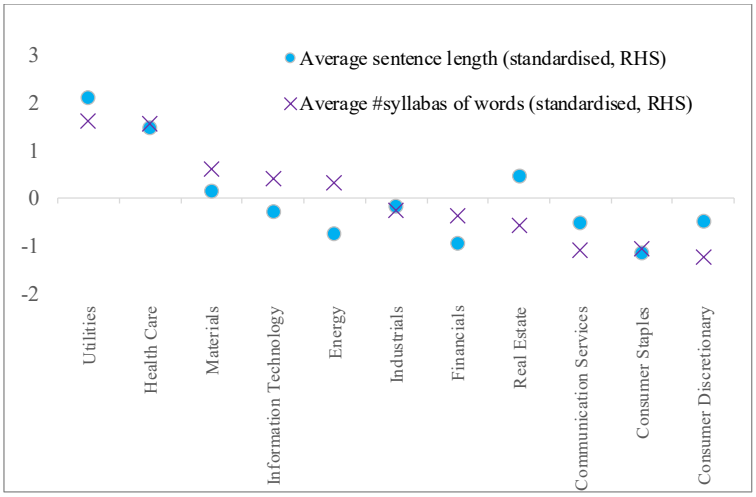
We apply LDA to identify the dominant topic for each ESG report. To achieve this, we first input our sample ESG report to the LDA algorithm for topics extraction. Then, for each ESG report, the algorithm assigns a topic to each sentence, which is the one having highest probability of matching the words contained in that sentence. After assigning topic to each sentence, the dominant topic of an ESG report could then be identified as the one that is assigned to most sentences in the report.

[24] Put it another way, LDA assigns for each word a certain probability to appear in a particular topic A particular word may occur in several topics with a different probability. Words without special relevance will have roughly even probability to appear across topics.

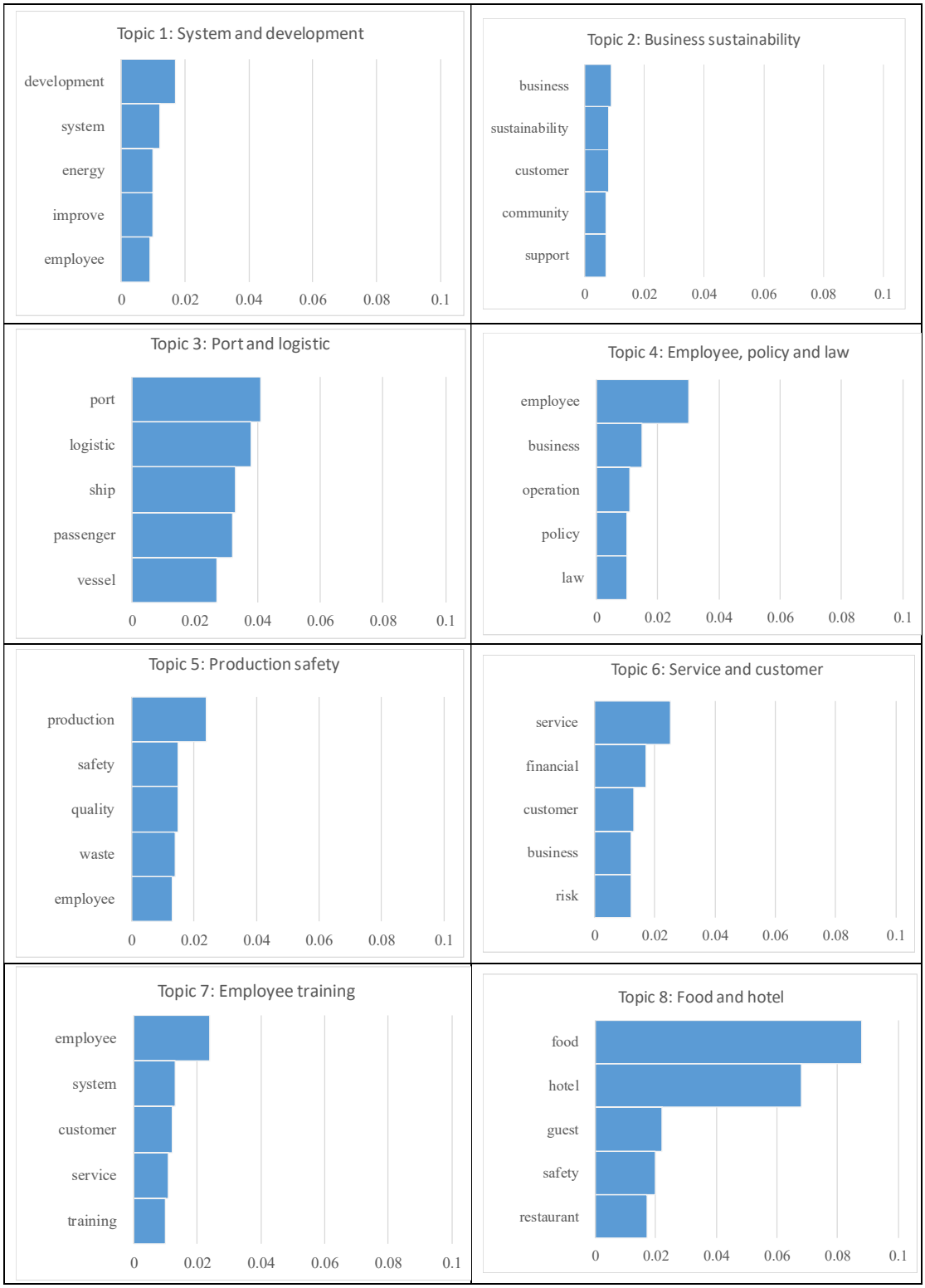
Appendix B: Data variables in empirical analysis

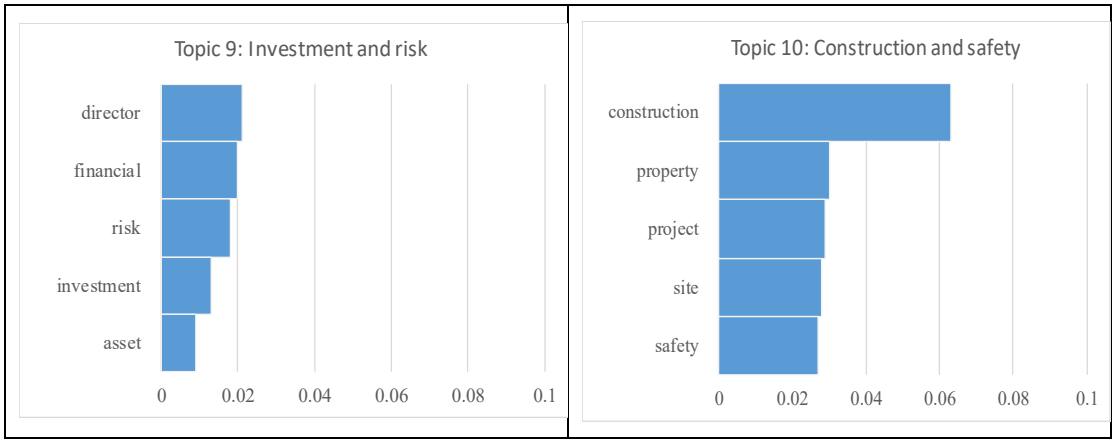
Variable	Definition	Source
BA	1 year average of bid-ask spread after the issuance of the ESG report	S&P Capital IQ and HKMA Staff calculation
VOL	1 year return volatility after the issuance of ESG report	S&P Capital IQ and HKMA Staff calculation
LEN	ESG report length (total non-stop words count, in logarithm form)	HKMA Staff calculation
READ	Flesh Reading Ease score of the ESG report	HKMA Staff calculation
FLS	Percentage of forward-looking sentences in the ESG report	HKMA Staff calculation
SIZE	Total asset (in logarithm form) at the end of previous financial year	S&P Capital IQ
LEV	Debt to equity ratio in previous financial year	S&P Capital IQ
INVPRICE	Inverse of 1 year average stock price after the issuance of ESG report	S&P Capital IQ
PRERETURN	1 year average stock return prior to the issuance of ESG report	S&P Capital IQ and HKMA Staff calculation
ROA	Return on assets in previous financial year	S&P Capital IQ
VIX	1 year average value of VIX index after the issuance of the ESG report	Bloomberg

Appendix C: Breakdown of readability measure by industry



Appendix D: Top 5 keywords for topics identified by topic modelling




Topic 9: Investment and risk
director
financial
risk
investment
asset
0
0.02
0.04
0.06
0.08
0.1
Topic 10: Construction and safety
construction
property
project
site
safety
0
0.02
0.04
0.06
0.08
0.1