



ITF Seminar 1



ESG 研究中心
RESEARCH CENTRE FOR ESG
香港恒生大學
THE HANG SENG UNIVERSITY
OF HONG KONG



Funding organisation

 創新科技署

Innovation and Technology Commission

SEMINAR 1

TECHNICAL ISSUES AND BUSINESS OPPORTUNITY OF BIG DATA APPLICATIONS ON REFLECTING CORPORATE ESG PERFORMANCE

14th August, 2023

Copyright by

Prof Louis Cheng

Special thanks to

Dr Maureen Chan, Mr Taiming Cao and Ms Kirsten Tsui

Project Title	Fostering Big Data Applications and Innovation to Reflect Corporate ESG Performance through Media and Social Listening (GSP/033/22)
Funding Body	Innovation and Technology Fund
Project Organiser	Research Centre for ESG (CESG), HSUHK
Supporting Organisations	Excellent Global and Master Insight
Seminar 1: Technical Issues and Business Opportunity of Big Data Applications on Reflecting Corporate ESG Performance	
Date and Time	14th August 2023, 2:30 pm to 5:30 pm
Location (Hybrid)	Unit 818, 8/F, Excellent Global Business Centre, China Insurance Group Building, 141 Des Voeux Rd Central, Central
Presenters	Prof Louis Cheng, HSUHK Dr Liane Lee, HSUHK Ms Vivi Hu and Mr David Cheng, YoujiVest Dr Jeff Jianfu Shen, PolyU

DISCLAIMER

This report is prepared by The Research Centre for ESG at the Hang Seng University of Hong Kong. This report and the results should be used only as a reference in understanding the ESG and big data. It is not an offer to sell or solicitation for an offer to buy any financial products and services and they should not be considered as investment advice. The Research Centre for ESG, Excellent Global, Master Insight, the research team, and the consultation participants accept no responsibility or liability for the accuracy and completeness of such information in the report. Any opinions, findings, conclusions or recommendations expressed in this material/event (or by members of the project team) do not reflect the views of the Government of the Hong Kong Special Administrative Region, the Innovation and Technology Commission or the Vetting Committee of the General Support Programme of the Innovation and Technology Fund.



PROJECT OVERVIEW

On 14th August 2023, the Research Centre for ESG (CESG), HSUHK organised the first seminar on ESG learning and development. The seminar, which was conducted with a hybrid mode, attracted over 60 participants. The seminar was the first of the four events that will be organised in 2023 to 2024. This series of events aim to promote big data analytics through a technology-driven ESG sentiment indicator in this ITF project. It is to adopt social listening measures for listed companies to generate ESG sentiment with the expectation of supporting Hong Kong society by enhancing the competitive edge of industries.



To start with, Seminar 1 “Technical Issues and Business Opportunity of Big Data Applications on Reflecting Corporate ESG Performance” aims to explain the importance of AI tools to evaluate the ESG performance of corporations. Representatives from YoujiVest introduced their ESG sentiment data used for this project. The seminar featured some initial output through a Python-based three-dimensional visualisation and a joint research project on the effect of ESG disagreements towards the ESG performance of corporations.



SEMINAR 1 REPORT

INTRODUCTION

Prof Louis Cheng gave an overview of the seminar objectives, expressed his gratitude toward supporting organisations (Excellent Global for venue provision and Master Insight for the media coverage), and introduced the presenters. He emphasised that the seminar was interactive and participants could share their expectations of ESG areas that can be explored in future. It is envisaged that more understanding and availability of ESG performances will help improve the financial performance of companies and foster the decision-making of investors

TOPIC 1: INTRODUCING THE BASIC COMPONENTS TO CONSTRUCT THE ESG INTELLIGENCE DATA

ESG DATA BY DR LIANE LEE

Dr Liane Lee presented on how the media view FTSE (Financial Times Stock Exchange) 100 companies' performance about SDG goal 13, one of the United Nations (UN) 17 SDG goals. Data are available for the total 17 SDG goals and 167 subgoals. The objective was to evaluate if the information in the reports was correctly reflected by the media, which involved a reality check whether the companies are engaging in greenwashing. A paper done by Aurora University Network was used for presenting the evaluation method. It started with the keyword search about climate actions. From applying the Python coding for extracting the dataset from RavenPack, an article was generated about oil spill. The thorough keyword search resulted in an extraction of a number of articles.

After reading the data of 10 years (2013-2022) for FTSE 100 companies, half of a million articles were located. From the data output, companies could be listed from top to bottom in terms of weighted sentiment. With the sentiment data, one can realise the relative ESG performance of the companies and their corresponding industries. The information generated can be an objective reference for making investment decisions by fund managers. The sentiment value ranged from +1 to -1, and with "0" as the neutral.

The weighted sentiment of the top performer on the list was 1.498. This implied that the media was feeling really good about the particular company in achieving SDG goal 13. By drawing a bar and line graph, one could visualise the upward trend of its ESG performance and the increasing number of articles generated about the company.



To explore factors for the good sentiment data through word cloud generation, the keyword “Partnership” and other more prominent factors were found. The partners of the top performing company probably helped reduce carbon emission at work actively, supplying guidance and handling other logistic issues. The company was putting in genuine efforts to do well in ESG in the past 10 years continuously. From reality check, there was no information regarding disasters, major mistakes and crises that happened in the company over the 10 years, showing that the company was doing well consistently.

For the bottom ESG performing company, the weighted sentiment was -0.064. This indicated that most of the ESG sentiment (-0.064) was interpreted by the media as negative. From the word cloud, the unfavourable factors for its poor sentiment performance could be located. It might involve a lot of climate issues which were damaging the environment. Accordingly, their ESG reports also reflected the unstable situation, for example, the change of CEO and the big drop of their share prices (down 20%), indicating that they were not doing well.

The implications of the data are as follows:

1. Both the top and bottom performers for SDG 13 were energy companies. It implies that the company of the same nature can either do well or poorly if they do not pay enough attention to putting in ESG efforts.
2. For the top performer, it was very active and could also draw the attention of the media to focus on their ESG activities on SDG 13. Therefore, they got strong sentiment values, showed no identified crises and also got public endorsement of their ESG efforts.
3. The factors leading to the different levels of ESG performance also reflected truthfully the good and poor ESG performance of the companies. Some of the problems found were the government interventions and the spillover issues. It will take years for the company's reputation to reestablish once damage has been done. Therefore, it is critical for companies to pay attention to SDG 13 and tackle problems heads-on.





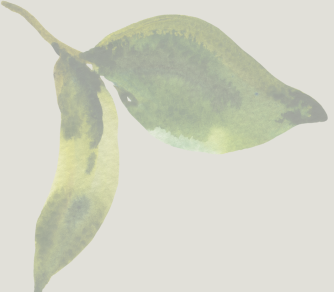
SENTIMENT DATA BY MS VIVI HU AND MR DAVID CHENG

Ms Vivi Hu, CEO of YoujiVest, discussed the methodology for the compilation of the sentiment data and the collaboration details with HSUHK. Vivi shared the process of sentiment analysis from data collection through the use of technologies to the generation of the ESG data finally. The output could be treated as an alternative data to indicate the ESG performance of companies. This helped check against the data disclosed by the company itself which might be subject to the suspected greenwashing effect as positive news (good disclosed data) was the focus of reporting. As in the previous five years of research experience, Vivi mentioned that there appeared no correlation between the NLP (Natural Language Processing) event-driven data from the positive sentiment data and the business returns.

YoujiVest has got a lot of good quality data from different sources like mainstream news, social media, regulatory bodies, NGO independent reports etc. Together with the application of many deep learning technologies like RNNs (Recurrent Neural Networks) or CNNs (Convolutional Neural Networks) and also many new models like LIAMA (Large Language Model Meta AI) and ChatGPT, those transformer architecture-based models help to produce easy-to-understand quality output.

In Western society, it is easy to find good global providers like RepRisk which does a good job in sentiment data analysis, but the Chinese language is usually not available. To meet the market need, YoujiVest focuses on using Chinese language news to understand the data from different sources and this helps solve the understanding problem when Chinese language has been used. Therefore, by using technologies and overcoming the language barrier, YoujiVest can obtain a lot of news information daily and to provide a very high frequency of ESG data ratings on a weekly basis. Investors do not need to rely solely on those company ESG reports or simply the annual filings. For example, YoujiVest can easily tell the fraud of a company, say Luckin Coffee, and follow up with early investigations, much earlier before something happened by using the sentiment analysis like CERS (ESG Risk Scores). All the efforts taken can lead to support the alpha generation of companies.

Mr David Cheng, Head of data of YoujiVest, explained how YoujiVest customised the users' institutional needs, taking the collaboration with HSUHK for a research project as an example. The timeframe of the project is around two years of the news data. Even for this short time period when compared to the 10 years' time period, YoujiVest had already got like more than 1 million (1171747) of mainstream news and



more than (10 million of social media news) (1113143) for the 800 companies located in Prof Cheng i-Score universe. The project goal was to provide a higher frequency view to the i-Score with ESG Risk Score.

Many of the current fundamental ESG data providers will dig into the company's financial reports and their own reporting. They would claim what they have done during the year in the ESG region but the data availability would be relatively low both for the institutional message to trade or for researchers to study, making it hard to compare the price movements with the ESG scores. To solve these problems, YoujiVest can provide a weekly or even higher frequency to the scores that Prof Cheng's project can use to compare between a company's financial behaviours and net profit movements as well as its ESG behaviours. The companies are behaving quite differently between the social media and the mainstream because social media's news sentiments are quite volatile.

To carry out the joint project with HSUHK, the large language models added help improve full news article understanding and sentiment judgement. The machines are able to read the whole sentence or even paragraphs instead of just labelling the words. YoujiVest also added several new sources of news information

in addition to existing sources. It also introduced more than 10,000 keywords into the data collection and processing system. With the expansion of this system, YoujiVest is able to exclude all the news that was not relevant to the ESG topics and is capable of separating the positive and negative sentiment analysis.

Vivi then talked about the recently launched Climate Risk Model in response to the need of academic research of pollutant discharge data and the carbon emission data. Vivi focused on elaborating the pollutant discharge data since carbon emission data has already been a popular topic for some time. The Chinese government monitors all the companies involved with heavy pollution business. Pollutant Discharge data is a kind of alternative data. It is real-time, showing the intensity of the pollution discharge. It is reliable and objective for understanding the pollution situations of the factories. From the alpha perspective, to some extent, this alternative data is a cross-reference to the company disclosure in terms of the environmental impact on Mainland China.



P.6 | ITF Seminar 1

© Prof Louis Cheng





TOPIC 2: DEMONSTRATION OF THE INITIAL FEATURES OF THE ESG INTELLIGENCE DATASET

DATA VISUALISATION BY PROF LOUIS CHENG

Prof Cheng talked about the combination of the performance data and sentiment data. Performance data are collected from the annual report, wrapping information and evaluation by the analysts with relevance to some corresponding regulations and stakeholders' evaluation. This is the basic annual renewal data to be called upon as performance data. Regarding sentiment data, it reacts very fast to things happening in the company. They do not react to things that happened last year as shown in the annual report. The annual report of the past 12 months and to include the number of 3 to 6 months to prepare the data, the data reported from the latest report will be 15 to 18 months old data.

The sentiment data is reacting to the news. If something happens to a company, say HSBC yesterday, the news sentiment platform will capture the news about the events the very next day. That information, in theory, will be captured by Ravenpack, YoujiVest, within the week and reflected in their sentiment indicators. With the available data, Prof Cheng used 3D-plot visualisation system, which contains listed companies' divergence-adjusted ESG rating performance, news ESG sentiment performance, news ESG sentiment

performance and ESG risk performance, to provide the cutting-edge solution of listed companies' full-picture of ESG performance.

The researcher built the i-Score divergence adjusted ESG rating system which eliminated the bias raised by the different ESG rating's ESG issues' scope, ESG rating analysts' regional and cultural characteristics and weighting gap on different ESG issues to capture the divergence effect and to make comprehensive integration of effects of local and international key ESG factors in the i-Score system.

The 3D-plot visualisation system used the i-Score performance of HK listed companies as the z-axis, the RavenPack sentiment score based on different keywords as the x-axis, and the YoujiVest positive or negative ESG sentiment score as the y-axis. The system processed ESG information for 511 HK listed companies as the i-Score universe and 334 HK listed companies as the peers (market cap > HKD2 billion at the end of 2022) universe.

P.7 | ITF Seminar 1

© Prof Louis Cheng





The data visualisation results show several concluding remarks:

1. Health care sector of HK listed companies generally have higher-than-average performance in media news ESG sentiments, and the average performance of HK listed companies in utilities and real estate sectors have lower-than-average performance in media news ESG sentiments.
 2. The i-Score ESG rating system covers almost all large-cap HK listed companies (market cap greater than HKD 100 billion), and the i-Score portfolio companies represented more than 95% of the Hong Kong stock market capitalisation.
 3. Large-cap companies' performance in media news ESG sentiments are more condense in the positive zone than mid-cap and small-cap companies. (mid-cap entities are market cap lies in between HKD 10 billion to HKD 100 billion, and small-cap entities are market cap smaller than HKD 10 billion.)
 4. There is an obvious positive correlation between corporate performance in environmental media news sentiment and corporate performance in social media news sentiment.
 5. There were more green washing activities which existed in social media news of listed companies than those of the mainstream news of listed companies.
- 





RESEARCH APPLICATION BY DR JEFF JIANFU SHEN

Dr Jianfu Shen first introduced the definition of ESG disagreement or ESG divergence, and the importance of ESG disagreement. The ESG divergence of a company was caused by the different opinions raised by multiple of rating agencies. Facing ESG divergence, it is difficult for investors to evaluate the performance of companies, funds and portfolios. The companies involved would decrease their incentives to improve their ESG performance and markets would likely price firms' ESG performance ex post.

Second, Dr Shen summarised the sources of ESG disagreement in both academic field and practical field. Different agencies might use different measurement metrics, scope of attributes that are covered by a particular rating agency, or different indicators for evaluating an attribute. Accordingly, the divergence might be big between local and international rating agencies. It was indicated that disagreement correlated with a firm's financial and accounting characteristics and several US research studies had resulted in ESG divergence and ESG risk. This requires further effort to explore ESG disagreement.

Third, Dr Shen collected ESG rating data of H-share listed companies from three ESG rating agencies from 2009 to 2018, and collected ESG rating data of A-share listed companies from seven ESG rating agencies from 2009 to 2020.

He then calculated the percentile rank of each stock in different agency's rating universe, computed the pairwise rating uncertainty as the sample standard deviation of the ranks provided by the two different raters in the pair, and obtained average pairwise rating uncertainty to be the proxy for ESG disagreement. It was found that the divergence between domestic institutions and foreign institutions far surpassed the divergence among domestic institutions and the divergence among foreign institutions.

Finally, Dr Shen used all A-share listed companies in CSI300 with two or more ESG ratings and collected these companies' cumulative abnormal return as dependent variable, and different ESG divergence as independent variable, and together with several financial indicators and operational indicators as control factors, to explore whether ESG disagreement could affect the investors' perceptions on the role of ESG in dealing with crisis.

CONCLUSION

Prof Cheng announced the Top 100 Scorecards through the website bigdata-esg.com. Visitors can learn of the methodology for computing the Top 100 i-Composite Scorecard v1.0 on the website. New versions will be released in the future when there are new groups of companies that use new methodologies for working out the Scorecards.