ESG-Based Market Risk Prediction and Management Using Machine Learning and Natural Language Processing

by

Yimeng Zhang

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Business and Economics Honors Program

NYU Shanghai

May 2023

Professor Marti G. Subrahmanyam	
Professor Christina Wang	Professor
Professor Wendy Jin	
Prof. Ming Liao	Thesis Adviser

Contents

Abstract

The future of finance goes hand in hand with social responsibility, environmental stewardship, and corporate ethics. In order to stay competitive, companies are increasingly disclosing more information about their environmental, social, and governance (ESG) performance. Higher ESG ratings are generally positively correlated with valuation and profitability while negatively correlated with volatility (Campagna et al., 2020). Chasing ESG has led to an increase in ESG data and rating providers, but investors are frustrated by the lack of standardized ESG data. On one hand, ESG rating providers give different metrics and parameters so that issuers are bombarded with "ESG reporting fatigue". On the other hand, the complexity of multidimensional ESG metrics leads to difficulties in direct comparison between different companies, especially those in different industries. But thankfully, emerging capabilities in Artificial Intelligence and Natural Language Processing (NLP) technology have the potential to provide new perspectives to existing data to help bridge the ESG disclosure gap. In this project, we aim to use NLP techniques to automatically extract and analyze companies' ESG initiatives and keywords directly from ESG reports as well as third-party news and articles, and to create an internal ESG score that is more accurate, comprehensive, and intuitive. Based on the internal ESG score, we studied the relationship between ESG initiatives and companies' market risk. The result shows that a better internal ESG internal score is positively correlated with a lower Value at Risk (VaR).

Key Words: *ESG, Natural Language Processing, Market Risk, Quantitative Research, Portfolio Management*

Preface

I have been always interested in doing inter-disciplinary research – solving a problem of one domain using knowledge and techniques from another. The motivation for doing this research is to combine my knowledge learned in both finance and data science in my undergraduate studies. I learned and became interested in market risk from the course of "Risk Management in Financial Institutions" taught by Professor Anthony Saunders at New York University Stern School of Business. I also learned natural language processing and machine learning models in the Machine Learning class taught by Professor Li Guo at NYU Shanghai. Natural language processing are particularly useful in analyzing unstructured data such as ESG initiatives, and ESG does play an important role in today's market risk. I would like to express my gratitude to Professor Ming Liao for introducing the concept of ESG to me and guide me through this research project. This research is both an application of data science tools to investment management, and an novel approach to analyze the quantitative aspect of ESG research.

1 Introduction

In light of recent and ever-more-frequently occurring world events, environmental, social, and governance, or ESG, is taking on an ever-greater significance. There is a global trend for responsible investing as consumers, particularly those from younger generations, are creating a demand for responsible investments. This trend was prominently illustrated when Cardano (touted as one of the most environmentally-friendly cryptocurrencies) defied a major downturn in the industry after Elon Musk tweeted concerns about the environmental impact of the crypto mining process. As well as futureproofing investment, analyzing ESG is also the socially responsible path. Fraud, corruption and climate change all pose a systemic risk for society. Companies are now facing more and more pressure from their shareholders to disclose more information about their environmental, social and governance strategies. Typically released on their websites on a yearly basis as a form of a PDF document, companies communicate on their key ESG initiatives across multiple themes such as how they value their employees, clients or customers, how they positively contribute back to society or even how they reduce (or commit to reduce) their carbon emissions. Consumed by third parties agencies (such as msci or csrhub), these reports are usually consolidated and benchmarked across industries to create ESG metrics.

1.1 The Challenges of ESG Analytics

ESG data in mainstream investing has three main challenges: most ESG data is qualitative, the landscape of corporate disclosures is incomplete and inconsistent, and disclosures are generally voluntary with sparse available data (Friede 2015, Park 2013). Many pertinent issues do not manifest in disclosures or regulatory filings and, if they do, the delays caused by reporting and publication cycles can cause relevant data to be out of date by the time it is in the public domain. There is also a significant bottleneck in assessing ESG performance due to the manual effort in continuously sourcing and validating disclosure data. This bottleneck is even more prominent when dealing with large volumes of unstructured text data, such as social media or news. As demand for ESG increases, the need for accurate and near real-time responses to ESG issues becomes clear, and the ability to detect and represent such issues through data sources beyond a company's filings is paramount. In the ever-changing investment landscape, news and social media data utilisation have become critical to ESG investment strategies. Moreover, the rapid growth and the resultant increase in awareness of the importance of ESG data has almost entirely been within the developed world. This makes it highly challenging for investors to assess ESG and invest in issuers in emerging markets such as China.

Another challenge is the complexity of a multidimensional metrics of ESG data. It does not make perfect sense to compare a company's anti-corruption policies with its commitment to sustainability, or to compare assessment of fraud prevention strategies from one region to another, even though each of them might contribute exactly the same to the overall ESG score. Such external ratings do not give us a three-dimensional picture in such a nuanced field, meaning that a change in the external rating does not explicitly reveal information to our understanding about this company. Quality and traceability are also challenges for ESG data. Analysts need access to raw data to get a clear and honest understanding of a target company's performance and commitment to best practices and addressing key issues. In summary, lack of regulation along with the complexity of ESG makes the comparison nearly impossible. But thankfully, emerging capabilities in artificial intelligence and Natural Language Processing (NLP) technology have the potential to provide new data points and new perspectives to existing data to help bridge the ESG disclosure gap.

1.2 Use of Unstructured Data and NLP

Investors can use AI technologies to collect and analyze unstructured data to add breadth and depth to ESG assessment. News articles, project disclosures to multilateral development banks (MDBs), sustainability reports and bond prospectuses are all listed as underused sources of unstructured data. Social media and review data can also support metrics such as employee satisfaction and worker rights. Historically assessing long-form text has been a time and resource-intensive process. However, innovations in NLP have revolutionized how unstructured text can be collected, analyzed and interpreted. NLP tools can analyze unstructured data rapidly and on a massive scale. Unstructured text can be taken from public sources, eliminating ESG reporting as an absolute requirement, and enabling a more thorough analysis of ESG performance within emerging markets. By analyzing text from media and other document types not specifically created for ESG reporting, there's a good chance that there will be less likelihood that the data has been manipulated or skewed to give a favourable view of the issuer.

1.3 Research Question and Significance

In this research, we aim to solve the problem of non-standardized, unreliable external ESG rating by extracting companies' key ESG initiatives directly from their ESG reports and related third-party articles based on natural language processing and machine learning techniques. We want to quantify the ESG-related measures taken by companies and whether those actions contribute to mitigate their market risk in a given period of time after their announcement. Even though the ESG rating providers are developing rapidly in China, it is still highly challenging for investors to assess ESG and invest in issuers with accurate rating and information. Thus, it is more interesting to use novel approach to analyze ESG reports in the emerging market such as China. Therefore, this research is based on reports and articles of 100 Chinese public companies ranging from eight different sectors, from which we create internal ESG scores and analyze their relationship with market returns and volatility.

The core of this research is based on the idea of post-announcement drift phenomenon, which has drawn great research interest and became well-studied in the last thirty years in financial economics and accounting research (Fama, 1998, Griffin et al., 2010, Hung et al., 2015). They are primarily concerned with the post-earnings-announcement drift (PEAD), referring to the phenomenon that stock prices tend to continue to drift upward (downward) following earnings announcements when the quarterly earnings were above (below) expectations. Meursault et al. studied the Post-Earnings-Announcement Drift using text, which takes the unstructured texual information and identified which groups of words are usually associated with upward post-earning adjustment (2021). All of the previous research are primarily concerned with finance-related reports and data. The post-ESG report-announcement drift has not been studied yet to the best of our knowledge. Inspired by previous research, our research will focus on the change of market returns after the announcement of companies' ESG reports as well as third-party articles to answer the question of 1) whether ESG announcement has an impact on companies' market returns in a given period of time 2) whether there are certain initiatives/actions companies take that lead to more significant change in market returns than others. Specifically, we first parse ESG-reports into sentences and individual words. Secondly, we created a term-frequency matrix for each company that measures how many times a word appear in this report or article. Then, we use this term-frequency matrix as features and the change of post-announcement excess returns (in categories) as target to train a mapping between terms and returns using machine learning models such as Logistic

Regression and XGBoost.

The results show that 1) key words such as "sustainability" and "diverse" have higher level of significance than others 2) the term-frequency matrix has great power of predicting the direction of excess returns after announcement of ESG-related reports or articles 3) companies in different industries have different focus and priority in terms of ESG strategies 4) constructed internal ESG score has great ability in stratifying stocks, and higher-ESG internal score portfolio is statistically associated with lower VaR. Such results will provide insights to companies to adjust their ESG strategies or initiatives to potentially bump up their market value. More interestingly, we will be able to construct an internal ESG score given any set of ESG-related reports, which is helpful for ESG funds to construct good-performant portfolios with higher returns and lower volatility.

The remainder of the paper is organized as follows. Section 2 summarizes the literature. The data and methodology are discussed in Section 3 and Section 4. Section 5 reports the results, and Section 6 concludes.

2 Literature Review

Disclosure is critical in the investment field because it helps outsiders to analyze, evaluate, and predict companies' performance, which directly or indirectly influences companies' market value – their risk and returns. An interesting and typical phenomenon about financial disclosure is the "post-earnings-announcement drift" (PEAD). It refers to an anomaly in financial markets, describing the drift of a firm's stock price in the direction of the firm's earnings surprise for an extended period of time. Contrary to what the efficient market hypothesis predicts, an earnings surprise does not lead to a full, instantaneous adjustment of stock prices, but to a slow, predictable drift (Fink, 2021). The phenomenon has been described at length for decades. Numerous studies have investigated the drift's origins and properties, covering drivers such as insufficient risk adjustment of returns, trading frictions, or behavioral explanations(Fink, 2021). Before 2021, research surrounding the post-earnings-announcement phenomenon is primarily concerned about the structured, numerical financial data which is directly related to companies' value and stock price. Investors and analysts can do valuation based on numerical data. However, financial disclosure not only discloses quantitative information, it also reveals qualitative information that implicitly reveals some more important updates or changes in terms of companies' performance or management.

Meursault et al. moved beyond numerical and quantitative data in analyzing the PEAD, instead they explored text-based empirical model to show that the calls' news content is about details behind the earnings number and the fundamentals of the firm (2021). They proposed a new numerical earnings surprise measure based on the text of earnings calls without explicitly incorporating the earnings number. They created a standardized unexpected earnings call text (SUE.txt) which generates a text-based post-earnings-announcement drift (PEAD.txt) larger than the classic PEAD. The magnitude of PEAD.txt is considerable even in recent years when the classic PEAD is close to zero. This measure, labeled SUE.txt, is calculated using output from a prediction model based on a regularized logistic text regression that extracts "good news" and "bad news" from earnings call text using natural language processing (bag of words). The prediction model is trained using past earnings calls and associated one-day abnormal returns; its parameters are dynamically calibrated.

To properly realize the potential of news data, millions of articles need to be processed daily, and one must look towards the power and capability of Machine Learning ("ML"). Latest advances in Natural Language Processing ("NLP") increase/strengthen our ability to process unstructured text data. Moving away from pre-determined text/keyword ontologies of the past (Lee et al., 2009), advances in the field of deep learning have pushed the state-of-the-art towards Transformer-based architectures such as BERT (Devlin, 2018). The key advantage here is leveraging context in decision making. Language is complex – for example, homographs exist, words whose meaning is entirely dependent on context. Without contextual understanding, false positives are likely, and many prominent classical methods are known to fall into this trap. Such approaches have focused on words and the frequency of their occurrence, with words weighted by how often they appear. For example, if a corpus of articles frequently mentions the word 'exploitation', such techniques can systematically discount its relevance. Similarly, identifying the difference between the word 'carbon' in the context of greenhouse gas emissions or when discussing carbon allotropes is critical in understanding the text in question. In other words, "context is king".

Across the investment community, researchers and engineers are using machine learning in new and disruptive ways, analysing linguistic information from content, using ESG and sentiment data to determine a company's commitment to ESG, and evaluating the impact of this commitment on stakeholders. Sokolov et al. show how BERT can be used as a classifier to aid in ESG Scoring, with aggregation approaches used on the output to construct a score. Such scores allow investors to recognize and understand what drives high and low ESG performance among their holdings, informing their approach for engagement (2021). For instance, reflecting the impact of "trashgate" in their decision-making process for HM. These also supplement brand and reputational risk management with a specific focus on sustainability issues and controversies.

In this research, we would like to examine the PEAD effect of ESG reports on companies' market returns using natural language processing and machine learning. This study is different from previous research in two ways: 1) First, to the best of our knowledge, it is the first research that examines the PEAD effect using the ESG reports; 2) It integrates natural language processing techniques that captures the different aspects of ESG initiatives including comprehensiveness, maturity, and change. This allows us to gain insights into the implicit impact of ESG on companies' market risk.

In addition, this paper uses the concept of word impact proposed by Yano et al. to study the explicit or implicit relationship between companies' description about their ESG initiatives and (2012). Different from the importance of traditional features, the importance of words can not only reflect the importance of each keyword to the final prediction result of the model, but also reflect the impact direction of the word on the final result: for example, under ideal conditions, words such as "rising" "should have a greater positive impact on the results, while words such as "down" should have a greater negative impact on the results. The calculation of word importance consists of two parts: regression coefficient and word frequency. The regression coefficient reflects the direction and intensity of each appearance of a single word on the final result, while the word frequency reflects the number of occurrences of each keyword in the text. Therefore, word impact is defined as the product of the difference between the "high" and "low" classification coefficients of the logistic regression model and the term frequency. We will show a list of positive words with their term frequency and word impact as part of the results in the Result section.

3 Data

3.1 Data Source And Description

We collected ESG-related reports and articles of 100 Chinese public companies. Typically released on their websites on a yearly basis as a form of a PDF document, companies communicate on their key ESG initiatives across multiple themes such as how they value their employees, clients or customers, how they positively contribute back to society or even how they reduce (or commit to reduce) their carbon emissions. Consumed by third parties agencies (such as msci or csrhub), these reports are usually consolidated and benchmarked across industries to create ESG metrics. Here is a paragraph from Alibaba ESG report as an example of our primary data source (2022):

> We also believe that the boundaries of social responsibility should be broader. To begin with, we are committed to creating an equal, inclusive, and dignified working environment for our employees, providing them with growth opportunities as well as fair and reasonable compensation and benefits. We are leveraging technology to support the high-quality development of small businesses and the creation of inclusive and flexible employment opportunities at scale while enabling more consumers to access a sustainable lifestyle. We are contributing to building a digital rural economy and improving community resilience to combat pandemics and natural disasters. We are promoting a culture of participatory philanthropy, which means charitable actions and social impact efforts can be influential in addressing sustainable development challenges beyond business goals.

To examine whether and how different industries value different aspects of ESG, we collected data from companies ranging across eight different sectors and are almost equally distributed, including technology, finance, healthcare, manufacturing, consumer services, infrastructure, cycle, and comprehensive. Because we intend to use bag of words/term frequency as the fundamental data after transformation, and we find that companies' ESG reports differ greatly from each other in terms of length and number of pages/words. We would like to control for the length of each report to avoid redundancy. Therefore, we mainly use the "executive summary" or "action plan" part of the report as our primary data source instead of the whole report. This summary is long enough to cover all the critical actions companies take regarding ESG but short enough to be more conveniently processed.

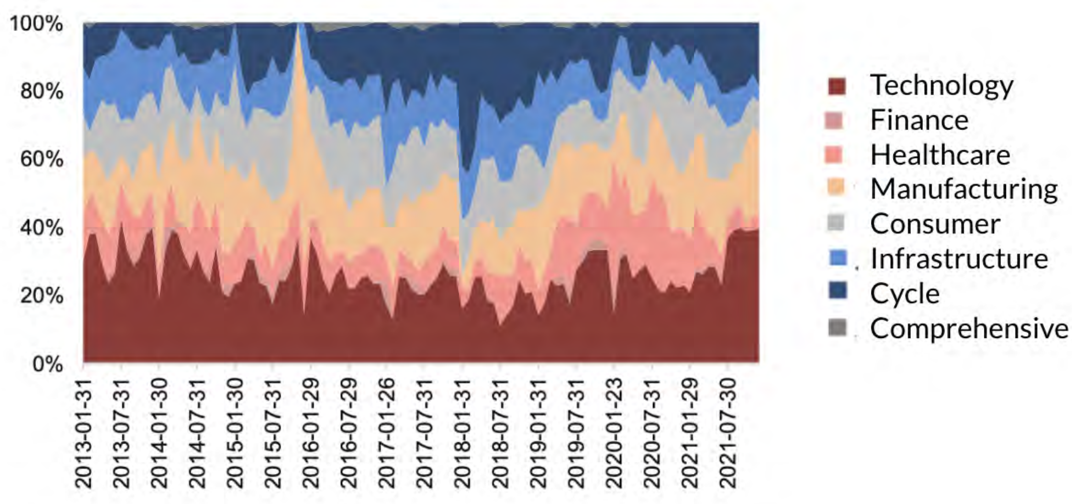

Figure 1: Data: ESG Report Industry Distributions

In addition to the texts from reports, we also collected their release dates, company names, stock indices, and the change in returns within five trading dates from Wind. We assume that investors would react to ESG reports within 5 trading days, a longer reaction period window compared with a one-day window used by convention if given financial disclosure, as we believe that the ESG reports do not have the same time effectiveness as financial disclosure. Note that the number of days is a hyper-parameter that is subject to change during model training. With comparison between other numbers, this would be the optimal choice as a reaction period window. In terms of our prediction target, the absolute change of return itself is not optimal if there is systematic change in the whole market. Therefore, we use the access return – the absolute return minus the return of CSI 500 for China-listed companies or the HSI Index for Hong Kong-listed companies as our preliminary target.

After the preliminary data cleaning process, our cleaned-data would be organized as the following chart:

Security Index	Security Name	Release Date	Executive Summary	Abnormal Return
9988.HK	Alibaba	2022-08-29	Enabling a sustainable digital life ...	-2.12%
9988.HK	Alibaba	2022-09-05	Alibaba reduces carbon footprint by 620,000 tons...	+0.65%
2318.HK	PNGAY	2023-03-15	Green financial ... development	+1.6%
...	...	...	...	...

3.2 Data Pre-processing

With the organized text data, we use the "textparser" package in Python to parse the texts into paragraphs, and tokenize them into individual terms that can be further analyzed. This allows us to annotate the grammatical category of individual words and only keep the most essential ones including normal nouns, proper nouns, verb, gerund, adjectives, and adverbs. Then, we use the CountVectorizer in Python Scikit-learn to automatically calculate the frequency of each word's appearance in each text. We combine all the reports or articles for companies from the same industry together, and select the top 50 most frequent terms respectively. This process allows us to create a term frequency matrix as features in training our machine learning model. In terms of the target variable, it makes more sense to use price movement class – a categorical variable rather than absolute value. According to previous research, we can divide the stocks into three different classes: high, middle, and low (based on which 33% percentile of their access returns each individual company belongs to). The following table shows the complete clean version of our dataset.

	diversity	sustainability	emission	change	...	improve	change of return
sample 1	0	1	1	1	0	0	high
sample 2	1	2	5	0	0	0	high
...	...	...	...	...	...	...	...
sample n	0	0	1	0	0	0	low

4 Methodology

With clean data, we are able to build machine learning models based on the term-frequency matrix. Given that this is a classification task, we would like to use Logistic Regrssion as our baseline model and XGBoost Classifier as our major model. Both models are widely applied and relatively interpretable compared with other more complicated models such as deep learning.

4.1 Logistic Regression

Logistic regression is a type of generalized linear model used to solve problems related to "classification" (Berkson, 1944). Its loss function is:

$$C(\vec{w}) = \sum_{i=1}^{n} \log(exp(-y_i(X_i\vec{w}_c)) + 1)$$

In this paper, the elastic network (elasticnet) regularization is used to constrain the complexity of the logistic regression model (Zou, 2005). The overall loss function of the model is:

$$C(\vec{w}) = \frac{1-p}{2}\vec{w}^T\vec{w} + p||\vec{w}||_1 + \lambda\sum_{i=1}^{n} \log(exp(-y_i(X_i\vec{w}_c)) + 1)$$

where $||\vec{w}||$ represents the norm one of the vector . The parameter p is the allocation ratio between L1 regularization and L2 regularization, and the parameter is the reciprocal of the regularization strength coefficient, that is, when is small, the overall regularization strength is greater. During training, we set the ratio p between L1 and L2 regularization to 0.5; the parameter uses the form of grid search and 5-fold cross-validation, and selects the λ with the highest average AUC in the verification set as the final parameter.

model	hyper-parameter	range
Logistic Regression	regularization term	[0.00001, 0.00003, 0.00006, 0.00008, 0.0001, 0.0003, 0.0006, 0.0008, 0.001, 0.003, 0.006, 0.008, 0.01]

At the same time, since the problem dealt with in this paper is a multi-classification problem, we use the OvR (one-vs-rest) strategy, that is, for the classification decision of the Kth class, we take all samples of this class as positive examples, and all other samples as negative examples. For example, we do binary logistic regression to get the classification model of the Kth class. The classification model of other categories follows the same process.

4.2 XGBoost

Extreme Gradient Boosting (XGBoost) is a Boosting integration algorithm, which is a strong learner that combines multiple weak learners (such as decision trees) in series by iterating between weak learners so that the loss function is continuously reduced (Chen & Guestrin, 2016).

Consistent with the general regression model, we conduct a grid search on all hyperparameter groups of the XGBoost classifier, and use a 5-fold crossover test to select a group of hyperparameters with the highest average AUC in the validation set as the final hyperparameters of the model. The hyperparameter settings are shown in the table below.

model	hyper-parameter	range
XGBoost	learning rate	[0.025, 0.05, 0.075]
	max depth	[3,5]
	sub sample	[0.8,0.85,0.9,0.95]

4.3 ESG Internal Score Calculation

After the training process, we did out-of-sample testing. We test the model in newly input ESG reports and predict their stock movements after announcements. The prediction is $p_c(x)$, from which we calculate its log-odds value $\mathrm{L}_c(x)$:

$$L_{c\in h,m,l}(x) = log\frac{p_c(x)}{1-p_c(x)}$$

$$\text{Internal ESG Score}{=}L_h(x) - L_l(x)$$

Among them, $c \in \{h, m, l\}$ are three category labels, which represent up, middle and down respectively. We calculate the difference between the log-odds values of its up and down categories as our final internal ESG score.

5 Results

After training and testing the machine learning models, first we are able to identify the "word impact" of each term to find out which term or initiative is valued more in certain industry. As we know that the word impact is calculated by :

$$I_w(x) = (\beta_x^{high} - \beta_x^{low})\frac{1}{N}\sum_{i=1}^{N} c_{ix}$$

where $I_w(x)$ is the importance of word x, β_x^{high} and β_x^{high} are the importance of words x in "high" and "low" respectively in the fitted logistic regression model. N is the total number of samples, and c_{ix} is the logarithmic word frequency of word x in the i-th sample. We choose the last training period as an example, and we are able to create two tables containing the top 15 positive impact words as follows:

	log (term frequency)	$\beta_{up} - \beta_{down}$	word impact
increase	10.3	0.06	0.136
sustainability	8.34	0.3	0.097
diversity	9.75	0.07	0.094
inclusive	10.05	0.04	0.082
value	9.95	0.05	0.075
woman	10.01	0.04	0.072
support	7.98	0.29	0.067
opportunity	10.14	0.02	0.049
climate	8.88	0.09	0.049
change	9.11	0.05	0.038
solution	8.2	0.12	0.035
provide	9.17	0.05	0.035
higher	9.49	0.03	0.029
achieve	7.81	0.14	0.028
governance	8.9	0.05	0.027

From the table, we are able to identify that "increase, sustainability, diversity, inclusive, value, and women" have high correlation with higher market returns, which means these aspects are valued more than others from investors' perspectives.

In addition, we are also able to classify all the stocks based on the internal ESG factor constructed in the previous section, based on which we created two portfolios with the first one containing top 20% score in the total stocks while the other containing the bottom 20%. We want to examine further how these two groups are different from each other in terms of risk and volatility. Therefore, we collected the portfolio historical returns respectively and got the following chart.

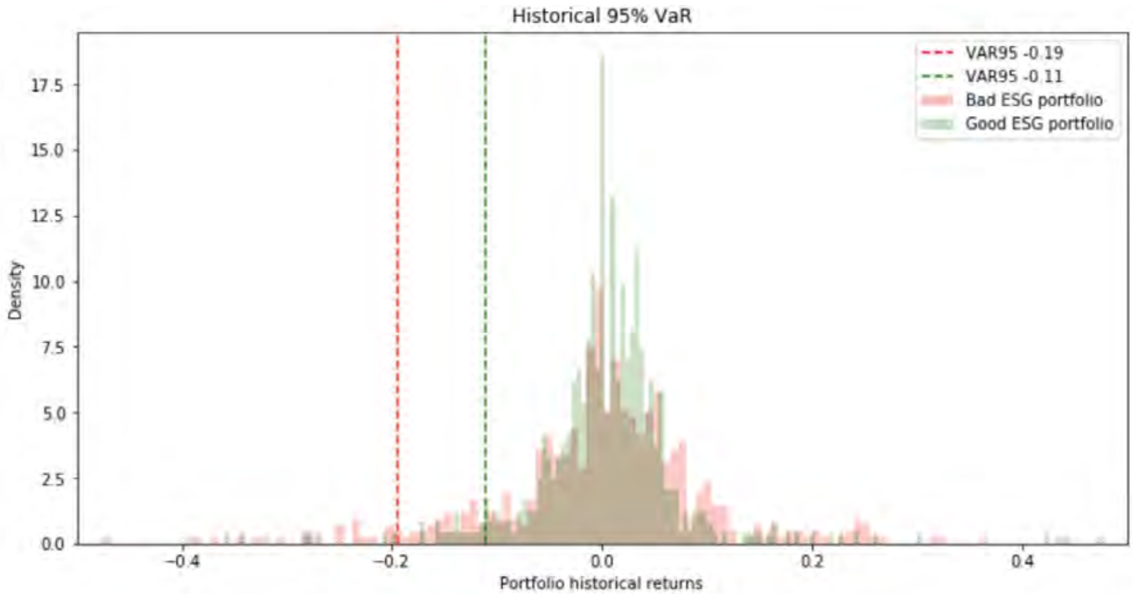

Figure 2: Good and Bad ESG Portfolio Historical Returns Distribution

This figure shows that good and bad internal ESG score portfolios are statistically significantly different in terms of their historical returns distribution. The bad ESG portfolio has longer tail, implying higher volatility in a given period of time. Moreover, the bad ESG score portfolio is left skewed that refers to lower/negative returns. Specifically, the 95% VaR for the good and bad ESG portfolios are -0.13 and -0.2 at a statistically significant level. These results show that ESG initiatives has correlation with market risk and returns, and the constructed internal ESG score has the ability to distinguish or stratify good or bad stocks.

6 Conclusion

This research successfully identifies the fact that some key ESG initiatives/concepts are valued more by investors compared with others by quantifying companies' text-form initiatives using natural language processing and machine learning. With quantitative metrics such as word impact, we found that on an aggregate level investors put more emphasis on the value of people such as "diverse" and "inclusive" than other key words. Companies in different industries have different focus and priority in terms of ESG strategies. For example, the technology industry values more "governance" while "sustainability" is more important for manufacturing industry, which perfectly matches our intuition as well. In addition, we construct an internal ESG score based on the machine learning model, which mitigates the problem of less standardized, less organized, less intuitive issues of external ESG ratings in China. Based on our internal ESG score, we are able to select/stratify stocks, constructing portfolios containing stocks with the best potential of ESG. Specifically, the 95% VaR for the good and bad ESG portfolios are -0.13 and -0.2 at a statistically significant level. These results show that "good" ESG initiatives has positive correlation with lower market risk and higher returns. Our model has great power of predicting the direction of excess returns after announcement of ESG-related reports or articles. It provides insights to companies to adjust their ESG strategies or initiatives to potentially bump up their market value. In addition, this is also helpful for ESG funds to construct good-performant portfolios with higher returns and lower volatility.

One limitation of this research is that though we use third-party ESG-related articles that discuss certain companies' key ESG strategies, a large portion of our data set still heavily relies on companies' own ESG reports. Without due diligence or actual numbers, these reports alone might be misleading. However, we assume that investors react to what they are presented with instead of "fact", it still makes sense if we use those reports as our data set. Still, it would be better if we could collect more third-party articles or equity research reports that are more objective.

References

[1] Alibaba ESG Report. ESG - Alibaba Group. (n.d.). https://www.alibabagroup.com/en-US/esg

[2] Berkson, J. (1944). Application to the logistic function to bio-assay. Journal of the American Statistical Association, 39(227), 357. https://doi.org/10.2307/2280041

[3] Campagna, A., Spellman, G. K., amp; Mishra, S. (2020, January 14). ESG matters. The Harvard Law School Forum on Corporate Governance. https://corpgov.law.harvard.edu/2020/01/14/esg-matters/

[4] Chen, T., amp; Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/2939672.2939785

[5] Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance1the comments of Brad Barber, David Hirshleifer, S.P. Kothari, Owen Lamont, Mark Mitchell, Hersh Shefrin, Robert Shiller, Rex Sinquefield, Richard Thaler, Theo Vermaelen, Robert Vishny, Ivo Welch, and a referee have been helpful. Kenneth French and Jay Ritter get special thanks.1. Journal of Financial Economics, 49(3), 283–306. https://doi.org/10.1016/s0304-405x(98)00026-9

[6] Fink, J. (2021). A review of the post-earnings-announcement drift. Journal of Behavioral and Experimental Finance, 29, 100446. https://doi.org/10.1016/j.jbef.2020.100446

[7] Griffin, J. M., Kelly, P. J., amp; Nardari, F. (2010). Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies, 23(8), 3225–3277. https://doi.org/10.1093/rfs/hhq044

[8] Henriksson, R., J. Livnat, P. Pfeifer, and M. Stump. 2019. "Integrating ESG in Portfolio Construction." The Journal of Portfolio Management 45 (4): 67–81.

[9] Hung, M., Li, X., amp; Wang, S. (2014). Post-earnings-announcement drift in global markets: Evidence from an information shock. The Review of Financial Studies, 28(4), 1242–1283. https://doi.org/10.1093/rfs/hhu092

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

[11] Lee Y. H., W. J. Tsao, and T. H. Chu. "Use of Ontology to Support Concept-Based Text Categorization." In Designing E-Business Systems. Markets, Services, and Networks, edited by C. Weinhardt, S. Luckner, and J. Stößer, 201-213. WEB 2008. Lecture Notes in Business Information Processing, vol 22. Berlin, Heidelberg: Springer. 2009.

[12] Meursault, V., Liang, P. J., Routledge, B. R., amp; Scanlon, M. (2021). PEAD.txt: Post-earnings-announcement drift using text. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3778798

[13] Park, Andrew Ravenel, Curtis. (2013). Integrating Sustainability Into Capital Markets: Bloomberg LP And ESG's Quantitative Legitimacy. Journal of Applied Corporate Finance

[14] Tae Yano, Noah A. Smith, and John D. Wilkerson. 2012. Textual Predictors of Bill Survival in Congressional Committees. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 793–802, Montréal, Canada. Association for Computational Linguistics.

[15] Zou, H., Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67(2), 301–320. http://www.jstor.org/stable/3647580

Acknowledgement

I would like to express my special gratitude for Professor Ming Liao who inspired me to do ESG-related research, sharing with me important background information about ESG and helping me find out my trajectory of this research. I also want to thank Professor Wendy Jin and Professor Dan Wang for checking-in with support and making sure I am on the right track of conducting research. It has been a great journey working with all the professors throughout the year.