A Deep Learning Framework for Climate Responsible Investment

Tian Guo[*] **Nicolas Jamet** **Valentin Betrix** **Louis-Alexandre Piquet**

Emmanuel Hauptmann

Systematic Equity Research
RAM Active Investments

Abstract

Incorporating climate considerations into portfolio analysis and systematic investments has drawn numerous attention recently. It is motivated by the pursuit of sustainable investing for a low-carbon transition. In this paper, we propose to integrate both structured and unstructured climate related data into quantitative investing for stock markets, e.g. carbon emission scores and climate events from news flows. We develop a deep learning framework to consume these data for assessing climate related opportunity and risk of stocks in the investing universe. Experimental evaluation on real data demonstrates the low-carbon intensity of the constructed portfolio as well as decent investing return.

1 Introduction

Over the last decade, awareness surrounding climate change issues and its profound impact has been growing rapidly. Despite the emergency issue of the global warming, we continue to see number of skepticism around sustainable investments. In this paper, we aim to demonstrate how investment solutions with bias towards climate and environmental factors represents direct response to this emergency without being detrimental to an investment long-term performance.

In the quantitative investing area, the recent proliferation of climate and environmental related data reported by companies and third party entities provide the possibility of incorporating climate issues into investment decision making. In 2016, the number of companies issuing sustainability reports has increased to over 80% of SP 500 companies. Environmental metrics in these reports cover different aspects of the firm's interaction with the climate and environment, such as its CO2 emissions, its approach to the climate change transition or its broad strategy in the use of natural resources. Meanwhile, the fast development of machine learning and data mining techniques enables to mine climate related data from a variety of open data sources, e.g. news, social media, etc. [11].

A desirable investment strategy for climate good is expected to finance the companies that are best positioned within their industries to help achieve a sustainable climate transition, while not sacrificing long-term investing return. Despite the increasingly available data, it is still challenging to quantitatively integrate climate related data into investment strategies and to build systematic investing for climate good. On one hand, it is necessary to design climate related scores bearing the relation to the investment performance of the corresponding companies, such that a quantitative investing strategy including these scores can be tuned to bias the companies good for climate, while gaining return [3, 8]. On the other hand, the data reported by companies or third parties often has the delay in reflecting the corresponding events in real world. In order to promptly respond to cli-

[*]Correspondence to `tig@ram-ai.com`

mate related events happening to companies in the investment process, it is also beneficial to have an online process to assess the impact or risk of these events.

Related Work. In the machine learning and data mining domain, previous studies focus on extracting various features, e.g. sentiments, economic metrics, etc, for market predicting [12, 5, 14]. Some recent work exploited the transformer based language models in Forex, financial sentiment analysis, and transcripts [2, 1, 16, 15, 7, 6, 10]. Our work focuses on an end to end framework of building investing portfolio for climate good, by leveraging structured and unstructured climate related data.

Contributions. In this paper, we present a systematic and explicit inclusion of climate opportunities and risks into our quantitative engines, by taking advantage of deep learning and natural language processing techniques. Specifically, the contribution of this paper is as follows:

- We propose the deep learning framework ClimateQuant, which is designed to consume structured and unstructured climate related data to model the relation to stock behaviors.
- For structured data, we present the carbon emission footprint and its relation to investing performance.
- For unstructured data, we focus on online extracting climate related events from news flow and modeling the impact of these events on the risk (or volatility) of corresponding stocks.
- We report quantitative assessment to demonstrate that the investing strategy of Climate-Quant leads to less lower carbon intensity, while retaining decent return performance.

2 ClimateQuant Framework

In this section, we give an overview of the ClimateQuant framework and then go through the spectrum of structured and unstructured climate related data we use in this paper. Our framework is flexible to digest more alternative data, depending on the availability.

Overview. Fig. 1 illustrates our deep learning framework consuming both structured and unstructured data to generate investing decisions. For the structured data, we add climate related factors into existing ones to enhance the climate consideration. Then, in order to capture more timely climate related events and their impact on the performance of companies, in parallel we have the pipeline processing news flow.

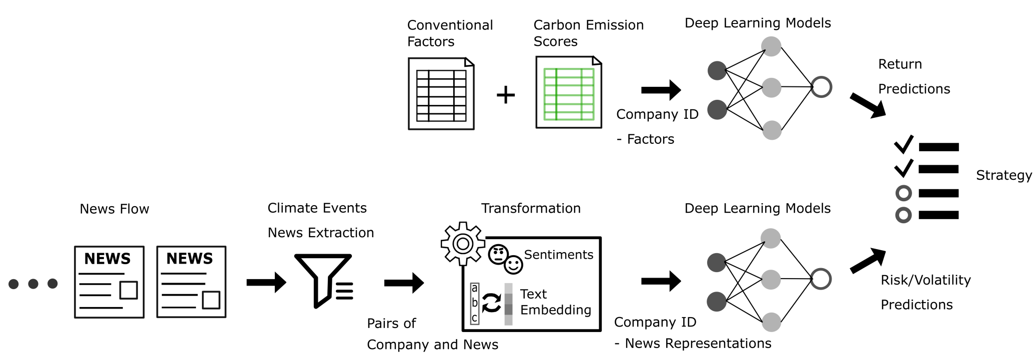

Figure 1: Illustration of ClimateQuant Framework.

Structured Data: Carbon Emission Score. Structured data refers to numerical variables reported by companies or third parties. Among several climate related factors, in this paper, we mainly introduce the carbon footprint metric, as the primary focus of climate concerns. According to the Climate Change Report (2014) [9], the continued emission of greenhouse gases will cause further warming and long-lasting changes in all components of the climate system, increasing the likelihood of severe, pervasive and irreversible impacts for people and ecosystems.

A carbon footprint is the total set of greenhouse gases (Carbon Dioxide, Methane etc.) a company directly or indirectly releases. Two scopes of emissions are took into account: (1) direct emissions from sources that are owned or controlled by the firm; (2) indirect emissions that are caused by the company through the consumption of imported heat, electricity, cooling, or steam, etc. Furthermore, in order to differentiate between efficient use of resources and company scale, carbon emissions level is adjusted via dividing by the company's revenue, resulting into the carbon intensity ratio.

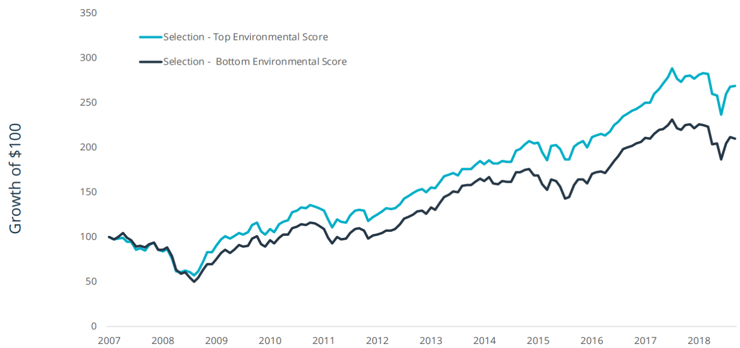

Figure 2: Cumulative Simulated Performance of Top v.s. Bottom Carbon Intensity Ratio.

As for the relation of the carbon score to investment return, our hypothesis is that owing to the management's commitment to lowering its carbon impact or its ability to optimize operational processes, carbon emissions utilized in an efficient way can result in improved operational performance. In turn, production efficiency will translate into lower costs for the same level of top line revenues, which can lead directly to high/ stable profit margins and high/sustainable bottom line earnings. To illustrate this relation, we charted the cumulative backtested performance of the top deciles of stocks (in a global developed universe) ranked by carbon intensity ratio vs the bottom decile in Fig 2.

Unstructured Data: Climate Events from News Flow. In our framework, unstructured data is collected by processing online news flow to extract climate related events or accidents w.r.t. the companies in our investing universe. These news are used to assess the future risk of the corresponding company's stock return given the events. Although this type of data is complementary to the structured carbon emission score by providing timely event information integration, it is challenging to process and model it, due to the nature of categorical symbols of news text. To this end, we resort to natural language processing techniques (NLP) to build the pipeline from climate news to risk prediction, as is shown in Fig. 1.

Specifically, the news flow in our framework is generic financial streaming news provided by a finance data vendor. Each piece of news is associated with the company ID it refers to. We developed the in-house extraction process by making use of a climate event vocabulary defined by our domain experts. This vocabulary includes the keywords, e.g. air pollution, environmental penalty, etc, and can be extended with ease. This extraction process yields pairs of a company and the associated list of climate event news. For instance, Fig. 3 shows the top climate topics in news in one week period.

Next, since news textual data is unstructured and infeasible for quantitative models to process, the subsequent transformation step aims to transform news text into quantitative representations by neural language model and sentiment analysis [13, 1, 4]. Language models or text embedding, a powerful NLP technique, is used to transform symbol represented text into numerical dense vectors preserving the semantic relatedness of text in the numerical space [13, 4]. The sentiment analyzer extracts the sentiment score of news content (i.e. a real value between -1 and +1). Eventually, the derived numerical representation of a piece of news is a vector including both the text embedding and sentiment score. This representation is fed into the risk (e.g. volatility) predicting neural network, which was trained as a supervised regression problem using a large collection of historical news and the corresponding stock risk values.

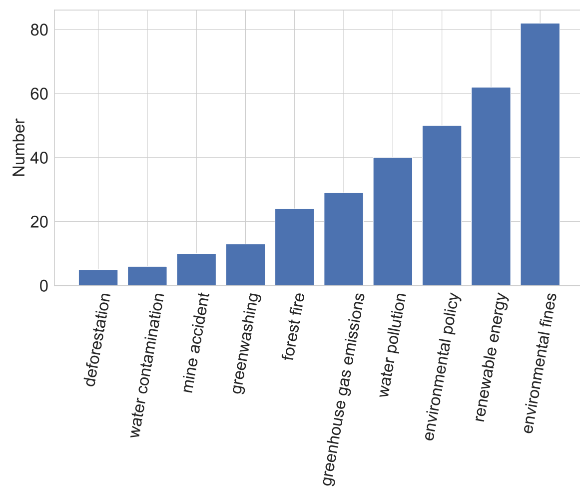

Figure 3: Top-10 climate events in the news from an example period April 06, 2020 to April 12, 2020.

3 Evaluation

In this part, we report some evaluation performance of the portfolio built by our ClimateQuant framework. Specifically, after the deep learning models processing structured and unstructured data are trained on historical data in the time range from 2000 to 2014, we test the framework using the data from January 2014 to June 2020, which is never seen by the model in the training phase. In this testing period, we build the portfolio by choosing stocks with high return and low risk predictions. The portfolio is held for one week and updated weekly. The investing universe in the testing period is mainly in the EU and US equity markets.

Through the following figures, we mainly show that the stocks selected by our framework give rise to substantial lower carbon intensity, while retaining long-term return performance. As is shown in Fig. 4 (a), compared to the benchmark index (i.e. MSCI World Total Return Index), our strategy favors more the stocks of companies with less carbon intensity. Fig. 4 (b) shows the corresponding cumulative return curve in the testing period.

To assess the effectiveness of our risk predictions, we split the stocks based on quintiles of volatility predictions and calculate the average return of each quintile with the 1 week holding period. As is shown in Fig. 4 (c), the portfolio built with the highest predicted risk names (Qt 4 portfolio) exhibits significantly lower returns.

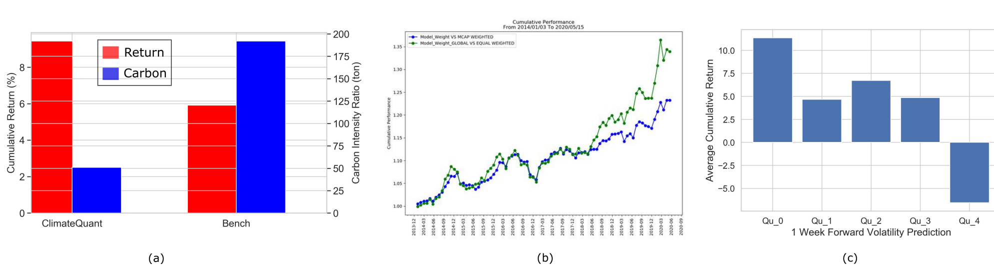

Figure 4: (a) Overall cumulative return and carbon intensity of the ClimateQuant portfolio and benchmark index. (b) Cumulative return curve over the backtest period. (c) Quintile return based on 1 forward volatility predictions.

4 Conclusion and Future Work

In this paper, we focus on the problem of systematic investing in stocks beneficial for low-carbon transition as well as generating return. We propose a deep learning framework to integrate climate related structured and unstructured data sources into quantitative portfolio construction. Experimental evaluation shows our framework exploiting carbon emission factors and climate events from news flow gives rise to desirable portfolio.

We believe the future work lies in the data and model aspects. On the data side, abundant unstructured climate and financial text data needs more tailored information extracting and representing techniques to serve the downstream models and decision makings. On the model side, the complex interaction between climate and finance are motivating the development of more advanced models and learning algorithms.

References

[1] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.

[2] Deli Chen, Keiko Harimoto, Ruihan Bao, Qi Su, Xu Sun, et al. Group, extract and aggregate: Summarizing a large amount of finance news for forex movement prediction. *arXiv preprint arXiv:1910.05032*, 2019.

[3] Indrani De and Michelle R Clayman. The benefits of socially responsible investing: An active manager's perspective. *The Journal of Investing*, 24(4):49–72, 2015.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[5] Tian Guo, Albert Bifet, and Nino Antulov-Fantulin. Bitcoin volatility forecasting with a glimpse into buy and sell orders. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 989–994. IEEE, 2018.

[6] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 261–269, 2018.

[7] Qikai Liu, Xiang Cheng, Sen Su, and Shuguang Zhu. Hierarchical complementary attention network for predicting stock price movements with news. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1603–1606, 2018.

[8] Zoltán Nagy, Altaf Kassam, and Linda-Eling Lee. Can esg add alpha? an analysis of esg tilt and momentum strategies. *The Journal of Investing*, 25(2):113–124, 2016.

[9] Rajendra K Pachauri, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, Purnamita Dasgupta, et al. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc, 2014.

[10] Yu Qin and Yi Yang. What you say and how you say it matters: Predicting financial risk using verbal and vocal cues. In *57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, page 390, 2019.

[11] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.

[12] Robert P Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2):1–19, 2009.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[14] Yaowei Wang, Qing Li, Zhexue Huang, and Junjie Li. Ean: Event attention network for stock price trend prediction based on sentimental embedding. In *Proceedings of the 10th ACM Conference on Web Science*, pages 311–320, 2019.

[15] Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 1970–1979, 2018.

[16] Linyi Yang, Tin Lok James Ng, Barry Smyth, and Riuhai Dong. Html: Hierarchical transformer-based multi-task learning for volatility prediction. In *Proceedings of The Web Conference 2020*, pages 441–451, 2020.

Disclaimer

This document is not intended for persons who are citizens of, domiciled or resident in, or entities registered in a country or a jurisdiction in which its distribution, publication, provision or use would violate current laws and regulations.

This publication has been prepared for general guidance on matters of interest only, and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this publication, and, to the extent permitted by law, RAM Active Investments SA (RAM) does not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it. Furthermore, the information, opinions and estimates in this document reflect an evaluation as of the date of initial publication and may be changed without notice. Past performance must not be considered an indicator or guarantee of future performance, and the addressees of this document are fully responsible for any investments they make. The content of this document is confidential and can only be read and/or used by its addressee. RAM is not liable for the use, transmission or exploitation of the content of this document. Therefore, any form of reproduction, copying, disclosure, modification and/or publication of the content is under the sole liability of the addressee of this document, and no liability whatsoever will be incurred by RAM. The addressee of this document agrees to comply with the applicable laws and regulations in the jurisdictions where they use the information reproduced in this document. This document is issued by RAM. This publication and its content may be cited provided that the source is indicated.