Assignment 1: Measuring Climate Risk

Leonardo Ciotti & Dominik Damast

Deadline: 30 September 2025, 23:59

General Information

Submission. Submit your results and answers to the assignment's questions in a report PDF. Additionally, you will have to submit the code you used to generate these results. For each group only one member of the group needs to submit the files. For the data analysis, please use Python. Name all files: *Assignment1_Group#*, adding the respective file ending.

Grading. The assignment will be worth 10% of your final grade. The assignment will be graded based on three components:

- *Functioning of the code.* The sole requirement is that the code runs without errors on our machines and produces the results in your report. Include all necessary libraries/modules, give instructions on packages to install if necessary, and indicate path variables that must be changed when running the code on another machine. The Python file in the assignment folder already provides code for the path variables.
- *Quality of the analysis.* We will evaluate the quality of the analysis based on the correctness and clarity of the results, e.g., whether the figures and regression coefficients are correct and the appropriateness of the methods used.
- *Clarity of the presentation.* We will evaluate the clarity of the presentation of the results. Describe your results clearly and exhaustively, consistent with potential visualizations and calculations. Comprehensively label all figures and tables. To answer the questions, do not use bullet points. **Be concise in your answers.**

Assignment

Objective. This assignment aims to familiarize you with two examples of NLP approaches to climate risk measures. In this assignment, you will have to analyze the data to understand better the time series of climate risks and the difficulties in choosing the right measures.

Data. For this assignment, please download the firm-quarter level risk measures from the following two papers:

- Sautner et al. (2023): https://osf.io/fd6jq/files/osfstorage (file "firmquarter_score")
- Li et al. (2024): https://www.corporateclimaterisk.com/ (the data labeled "original data" in the "Data" section)

Additionally, there is data on the Wall Street Journal (WSJ) climate change attention index from Engle et al. (2020) provided on the course's website (file "EGLKS_data.xlsx").

1 NLP Measures of Climate Risk [60 Points]

Consider the climate risk measures from Sautner et al. (2023). Focus your analysis **on U.S. public firms**, that is, firms headquartered in the United States. Multiply all measures by the factor 1000 for better readability in graphs. To visualize data in Python, you can use the `matplotlib` or `seaborn` library.

(a) Describe the main idea and methodology of NLP-based climate risk measures (max. 5 sentences). [10 Points]

(b) Calculate the number of zero values of the climate change exposure measure (*cc_expo_ew*) in Q1 2002 and Q4 2017. Why are there many values equal to zero? Did the frequency of zero values change over time? Give an intuition for the results (max. 5 sentences). (**Hint:** Choose the same scale for the x-axis and the y-axis in both plots.) [10 Points]

(c) Plot the mean climate risk exposure (*cc_expo_ew*) of the top-100 exposed firms (by quarter) between 2002 Q1 and 2017 Q4. How did the climate risk exposure change over time? Give an intuition for the results (max. 5 sentences). [10 Points]

Now, additionally download the file *EGLKS_data.xlsx* from the course website. The file contains the Wall Street Journal (WSJ) climate change attention index from Engle et al. (2020). A description of the WSJ climate change attention index can be found in Engle et al. (2020), pp. 1191-1196.

(d) Add to your plot from exercise (c), the WSJ climate change attention index (*wsj*) between 2002 Q1 and 2017 Q4. Take the average over a quarter of the WSJ index each time. [10 Points]

(e) Calculate for each component of the climate risk exposure measure, that is, the regulatory exposure, *rg_expo_ew*, the opportunities exposure, *op_expo_ew*, and the physical risk exposure, *ph_expo_ew*, the correlation between the top-100 exposed firms' average exposure and the WSJ climate change attention index (*wsj*) over time. Is the correlation positive or negative? Is it high or low? Give an intuition for the results (max. 5 sentences). [20 Points]

2 Aggregate Confusion in NLP Measures? [30 Points]

Match the data from Li et al. (2024) to the U.S. public firms in the Sautner et al. (2023) dataset using firms' *gvkey*. Consider only firms that are in **both** datasets.

In the following analysis, consider a snapshot of the data **in 2017 Q4.**

(a) Split the firms in the matched dataset into deciles based on their transition risk exposure measure from Li et al. (2024) (*tran_risk_w_std*). Do the same for the regulatory risk (*rg_expo_ew*) exposure measure from Sautner et al. (2023). Plot a histogram for the difference in deciles between *tran_risk_w_std* and *rg_expo_ew*. What does the histogram tell you about the agreement between the two measures? Give an intuition for the results (max. 5 sentences). [10 Points]

(b) Repeat the previous exercise, but now for the opportunities exposure measure from Sautner et al. (2023) (*op_expo_ew*). What does the histogram tell you about the agreement between the two measures? Give an intuition for the results (max. 5 sentences). [10 Points]

(c) Provide at least one reason for potential differences in the measures of Li et al. (2024) and Sautner et al. (2023). [10 Points]

3 References

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging Climate Change News. *Review of Financial Studies*, 33(3):1184–1216.

Li, Q., Shan, H., Tang, Y., and Yao, V. (2024). Corporate Climate Risk: Measurements and Responses. *Review of Financial Studies*, 37(6):1778–1830.

Sautner, Z., Van Lent, L., Vilkov, G., and Zhang, R. (2023). Firm-Level Climate Change Exposure. *Journal of Finance*, 78(3):1449–1498.