Financial Econometrics Report: A sentiment analysis related to ESG matters based on 10K reports

BENJELLOUN Ghali, BERRONDO Alice,
LA CAVA Caterina, MECHTI Majda
Email address: ghali.benjelloun@epfl.ch, alice.berrondo@epfl.ch,
caterina.lacava@epfl.ch, majda.mechti@epfl.ch

June 11, 2023

Introduction and research question

In this study, we conducted a sentiment analysis related to ESG matters, based on 10K reports. In particular, sentiment analysis is the use of natural language processing, text analysis, computational linguistics and biometrics to identify, classify and categorize texts to understand if they have a positive/neutral/negative tone (sentiment). As our input texts, we're going to use 10K-reports which are comprehensive reports filled annually by publicly-traded companies about their financial performance. They are required by the U.S. Securities and Exchange Commission (SEC) and contain much more detail than a company's annual report, which is sent to shareholders before an annual meeting to elect company directors. EDGAR database provided by SEC allow us to download all the reports we needed [1].
10K-reports include five distinct sections:

- Business: overview of the company's main operations, including its products and services.
- Risk factors: outline any and all risks the company faces or may face in the future.
- Selected financial data: detail of specific financial information about the company.
- Management's discussion and analysis: financial condition and results of operations; gives the company an opportunity to explain its business results from the previous fiscal year.
- Financial statements: company's audited financial statements including the income statement, balance sheets, and statement of cash flows.

We could also use proxy statements, but instead we decided to proceed with 10K-reports because:

- the purpose of proxy statement is to enable shareholders to understand the agendas which will be discussed in the Annual Meeting, the compensation structure of management and related party transactions; 10K-reports are useful for the financial performance of the business and the management's view on the business and future aspects.

- proxy statements are used especially by shareholders of the company with voting rights, while 10K-reports are used also by investors, financial institutions and rating agencies.

For these reasons, we think that analyze 10K-reports allow us to extract more useful and relevant information related to ESG-topics.
What we will aim to answer in this paper is:

Has companies' awareness on ESG matters evolved with time?

In a world where ESG topics are becoming always more important, it's useful to understand if also companies in different sectors are trying to deal, face and improve their attitude towards ESG. Trying to answer to the main question, we will explain the different approaches we used, the scientific articles we were focused on and the relation between the ESG-related sentiments we found and the current ESG scores.

Recall that ESG score is a quantitative metric whose interpretation depends on the rating agencies that provide it; in fact, it can measure the degree to which a company's economic value is at risk driven by ESG factors that is the magnitude of a company's unmanaged ESG risks (the higher the riskier) or a measure of how well companies manage environmental, social and governance risks and opportunities (the higher the better). An example of the first and second type of measure is given by Sustainalytics Inc. and MSCI, respectively: the former provides ESG scores between 1 (best) and 100 (worst) while the latter gives ESG scores between 1 (worst) and 10 (best).

Review of Literature

Throughout this paper, amidst our own analysis we will conduct a review of literature on three articles relating to ESG matters as a way to draw from, understand and implement a model on given companies' reports. Whenever pertinent, the review of literature will be done immediately preceding our model construction and implementation. Here is a brief summary of the three articles used:

- Article 1 : A Sentiment Analysis of the Effect of ESG Information on Stock Price [2]. Study of the relationship between ESG information and investors' reaction by relying on a time series approach; extraction of news articles' sentiment as input for an AutoRegressive Distributed Lag model.

- Article 2 : A Dataset for Detecting Real-World Environmental Claims [3]. Introduces a newly trained model specific to the analysis of environment-related lexicography to answer the need of such a specific model type, which was lacking prior.

- Article 3 : Sentiment Analysis of ESG disclosures on Stock Market [4]. Introduces a model based on FinBERT to analyse newspaper articles and tweets revolving around four specific companies and their effects on the stock market.

Data analysis and summary statistics

We have conducted the analysis on 20 companies' 10-K reports grouped by industry. More precisely, we have selected four industries which are automotive, bank, energy and food and five companies per industry. Companies list is provided in the following:

- Automotive sector: Tesla (TSLA), Ford Motor company (F), Motorcar Parts of America Inc. (MPAA), Rivian (RIVN), Standard Motor Products (SMP).
- Bank sector: Morgan Stanley (MS), Bank of America (BAC), Citigroup Inc. (C), JPMorgan Chase & Co. (JPM), Sillicon Valley Bank (SIVBQ).
- Energy sector: Chevron Corporation (CVX), ConocoPhillips (COP), PetroGas Company (PTCO), General Electric Company (GE), Royale Energy Inc. (ROYL).
- Food sector: Kellogg Company (K), Beyond Meat Inc. (BYND), Campbell Soup Company (CPB), The Hershey Company (HSY), McDonald's Corporation (MCD).

From the three figure below (1,2,3), we can notice that banks reports tend to be larger than other sector companies' reports. This is reasonable since we chose banks with global influence which are mostly regulated and need to fill more information in their reports.

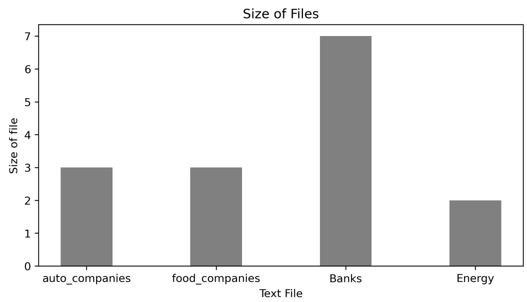

Figure 1: Size of Files (in MB)

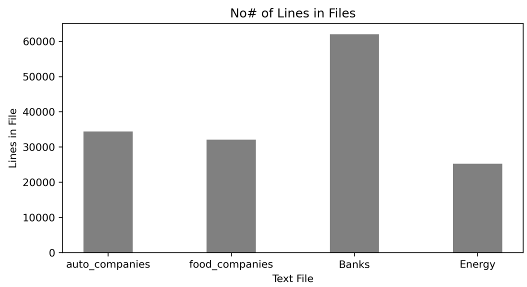

Figure 2: Nb of Lines in Files

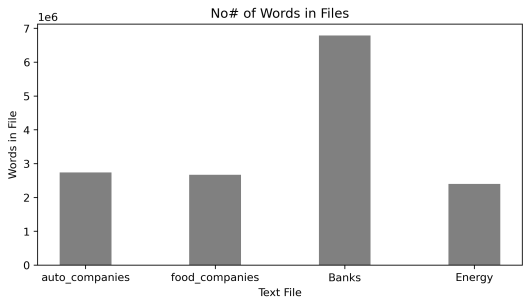

Figure 3: Nb of Words in Files

In order to have a general idea of what information the reports provide, we searched for the most repeated words in all (merged by industry) reports. We used unigrams in order to split the reports into single words. We obtained the results reported in figures 4,5,6,7.

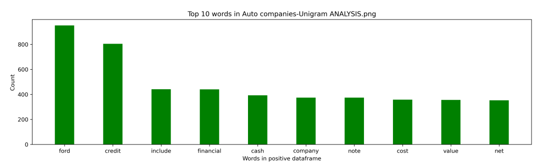

Figure 4: Top 10 words in Car companies

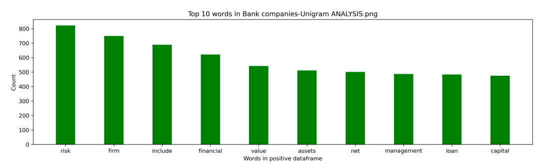

Figure 5: Top 10 words in Banks

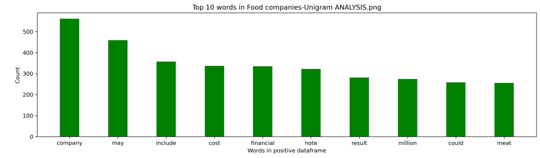

Figure 6: Top 10 words in Food companies

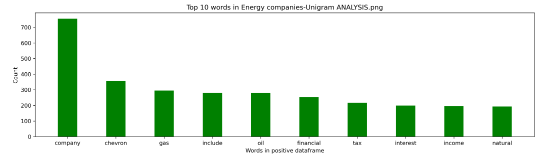

Figure 7: Top 10 words in Energy companies

As expected, most used words differ depending on the industry we look at; also, there are not ESG-related words that can be easily recognize: this is reasonable since we are analyzing 10K-reports that are not specific document on ESG like articles or news properly selected as treating ESG-arguments.

Selection of ESG sentences with a dictionary approach

"A Sentiment Analysis of the Effect of ESG Information on Stock Prices"

In this article [2], the authors set out to investigate the relationship between ESG information and investors' reaction by relying on a time series approach. They gather a large data set of news articles with ESG-relevant content from which they extract the articles' sentiment, which is then used as an input to an AutoRegressive Distributed Lag (ARDL) model to explain stock market returns for the constituents of the Dow Jones Industrial Average (DJIA). Due to lack of sufficient material, two over 30 constituents have been dropped from the study, leaving 28 constituents to be modeled.

For the period of interest, between January 2010 and December 2018, the authors take as data points news articles relevant to ESG topics as well as DJIA constituents through a filtration based on a relevance score (in percent): an article were linked to a certain company only when it shows a relevance score of more than 50% for that respective company. Furthermore, for each DJIA constituent, the authors obtain daily observations of stock prices as well as closing prices of the S&P 500 index from Yahoo Finance. The latter were then used to compute daily log-returns of the given stocks, that, then, were regressed through ordinary least squares method. For the given time period, a total of 2263 trading days were observed and to each observation the authors associated an ESG-sentiment time series extracted from the aforementioned news article data set.

To compute the sentiment that each individual news article conveys, they rely on a dictionary approach developed by Loughran and McDonald (LM) [5], trained on financial texts. While the dictionary by LM (2011) seems appropriate to capture the sentiment of news content particularly relevant to investors, there is a shortcoming when applying the dictionary to data set used to capture ESG-related tone. If a text combines positive finance-related and negative ESG-specific topics, the sentiment index might be biased in favor of the positive financial news at the expense of adequately representing the ESG-related information. To avoid that, the authors followed Nasukawa and Yi (2003) [6] and conduct a domain-oriented sentiment analysis: they rely on the word list by Mysková and Hájek (2018) [7], which comprises words related to sustainability topics, based on sustainable development glossaries provided by the United Nations and the Environmental Protection Agency, amongst others, and evaluate their ESG news articles only around sentences that contain at least one of the words that are on the lists. More precisely, to evaluate the sentiment in only the ESG-related part of the articles, they computed, (for each article) what they refer as "sentiment polarity":

$$\text{polarity} = \frac{\#\text{positive words} - \#\text{negative words}}{\#\text{positive words} + \#\text{negative words}}$$

where #positive words/#negative words refers to the number of words the LM list classify with a positive/negative tone; polarity is between [-1,1].

Polarity approach

In this section we replicate the paper "A Sentiment Analysis of the Effect of ESG Information on Stock Prices" [2] just mentioned above.
To analyze only the main corpus of 10K reports, that is the part that does not contain financial statements, at first, we split each report in sentences and we cleaned them removing numbers, links, punctuation and stop words with the nltk library in python. Then, we used the lists of words provided by "Sustainability and Corporate Social Responsibility in the Text of Annual Reports - The Case of the IT Services Industry" [7] to select the ESG-related sentences from our cleaned reports: this article gives three word lists to evaluate the frequency of comments that are related respectively to economic, environmental and social sustainability. The number of words in the list are 104, 177, 174, respectively. Both single words and word phrases were considered in the word lists.
We merged the three lists in a unique one to then extract from the cleaned reports only the sentences that had

at least one of the words in the unique list; we will call the selected sentences "ESG-related sentences". The percentage of ESG-related sentences extracted from the cleaned reports goes from a minimum of 3.4% (ConocoPhillips) to a maximum of 16.6% (Kellogg Company); unfortunately, these number are low, but reasonable. For each ESG-related report, we computed the polarity, defined in [2].
Polarity lies between -1 and 1 and can be a measure of the sentiment of the ESG-related reports: when is negative, the report has a negative ESG-related sentiment while when is positive the report has a positive ESG-related sentiment. The worse polarity is for JPMorgan Chase & Co.-2019 (-0.84) while the best is for McDonald's Corporation-2019 (0.12); however, McDonald's it's the only company that has a positive polarity over four years of time. This method gives to each company and each year except McDonald's a negative sentiment.
We think that this result was driven by the fact that the number of negative tone words in the Loughran and McDonald list is about six time the number of the positive ones (2345 vs 347) [8]. To understand better which positive and negative words have driven this result, we displayed the ones that were repeated most both on the ESG-related report and the LM list, for each company. The results is given in the appendix 13. Unfortunately, we can see that there are not lot of words related to ESG, especially related to environmental topics; despite that, companies like PetroGas, Bank of America and Silicon Valley Bank have less positive tone words than negative ones. A particular case is for Bank of America which has only one positive sentiment word ("perfect"). We find inconsistent and unreliable use this approach, thus, in the following, we illustrate other different methods.

Vader approach

Starting from the ESG-related reports of above, we applied Vader algorithm [9]: this a rule-based algorithm already pre-trained on social media texts, but also applicable to other domains. The output of Vader is, for example:

$$\text{'pos'} : 0.746, \ \text{'compound'} : 0.8316, \ \text{'neu'} : 0.254, \ \text{'neg'} : 0.0$$

where 'pos' , 'neg' and 'neu' scores are ratios for proportions of text that fall in each category; these are the most useful metrics if you want to analyze the context and presentation of how sentiment is conveyed or embedded in rhetoric for a given sentence. While, the compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence; it classifies the text as:

- negative if $-1 \leq$compound score≤ -0.05
- neutral if $-0.05 <$compound score< 0.05
- positive if $0.05 \leq$compound score≤ 1

We applied this algorithm at sentence level (first method) instead of doing that for each entire ESG-related report (second method): our choice was driven by the fact that the second method gave us very inconclusive results and also Vader's demo reports that the algorithm works better at a sentence level. In fact, trying to apply this method at the entire ESG-related reports, we got a compound score equal to one almost for all reports; this means that Vader classifies them as all positive sentiment reports and there are no differences between reports' compound scores (more details in "ResultsVADER" folder).
Proceeding with the first method, we have a compound score for each sentence in each of the ESG-related report; thus, we made the average of these compound scores to have, in the end, an average compound score (ACS) for each ESG-related report. The average compound scores are between 0.094 (PetroGas-2020) and 0.635 (Bank of America-2019), then, since they are all positive and bigger than 0.05, Vader classifies them as all with a positive sentiment.
To understand the evolution of the average compound score over the four past years (from 2019 to 2022), we plotted these scores for each company. More precisely, in figure 8 are reported the evolution of each company in its industry. For the energy and automotive sector, we could not include in the plots Royale Energy Inc. and Rivian respectively due to lack of data (we did not have all the four reports).

Looking to the graphs, it's difficult to establish an overall trend across industries, however it's worthy analyze the behavior of the average compound score within each sector.

- energy: ConocoPhillips increased its average compound score following a upward trend over 2019/22. PetroGas is at the bottom of the plot.
- car: marked downtrend for Tesla and Motorcar Parts of America Inc.; even if the other two companies experienced an increasing, it's not so significative.
- bank: stable average compound score across time except for JPMorgan Chase & Co.; peculiar the case of Bank of America which is positioned at the top of the graph.

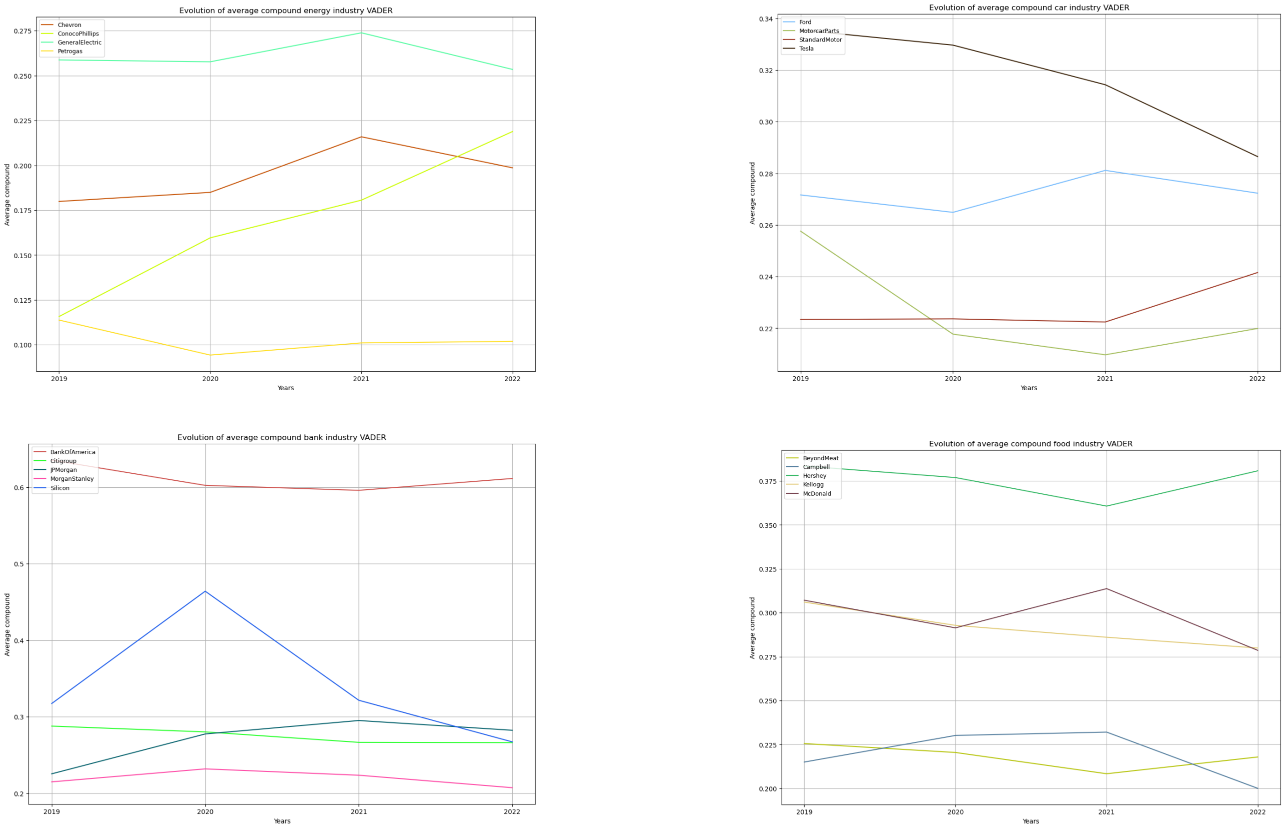

Figure 8: Evolution of VADER sentiment score by industry

- food: it seems that all the companies have decreased their average compound score during the analyzed period.

To understand the quantitative differences of scores during time, we computed first the average change of the average compound score between 2019 and 2022 for each company:

$$\text{for every c in companies: } \text{change}_{c,19,22} = \frac{\text{ACS}_{c,22} - \text{ACS}_{c,19}}{\text{ACS}_{c,19}}$$

and then, we also made an average of these changes within the same industry:

$$\text{for every i in industry: } \text{average\_change}_i = \frac{1}{N_i} \sum_{\text{c in i}} \text{change}_{c,19,22}$$

where N_i is the number of companies in industry i.
In "ResultsVADER" folder you can see the average change per company (change_per_company.png), while in table 1 there is the average change industry by industry.

Industry	change (%)
Automotive	-5.22
Banking	-1.70
Energy	21.76
Food	-5.78

Table 1: Average change per industry, Vader approach

"A Dataset for Detecting Real-World Environmental Claims"

The thesis of this article [3] revolves around the need to introduce a new sentence-based ESG-sentiment model, trained to detect environmental claims made by listed companies, using an expert-annotated data set of 3000 words and phrases. To achieve this, the authors train the model on sentences between 10 and 40 words. About 60% of the data set was used to train the model, while the remaining 40% was used as test set.
The analysis was done in two steps: first, the BERT model was trained to detect environmental claims and an error analysis was made to take into account limiting factors such as data set size. Then, the model was tested on listed companies' earning calls to have the empirical results.
The authors go on to compare five baseline models and four transformer models to determine accuracy and

consistency between different methods and observe that some baseline models fail to reach specific standards. The same approach was used on a Pledge Detection data set and seemed to perform slightly better.
The authors used these different tests as justification for the necessity of creating a new sentiment analysis model trained specifically to detect environmental claims on companies' earning calls, as a step towards detection of potential green washing activity.

A first BERT based approach

We implemented the pre-trained BERT customed in the previously explained article [3] on the 10K-reports. We used it to compute the percentage of sentences that have "ecological commitments": this gave an idea of the importance of ecology for our selected companies. We ran this algorithm sentence per sentence, as it was trained on a sentence level also in the article.
In figure 9 is reported the number of 10K-reports with a percentage of ecological commitments between different values.

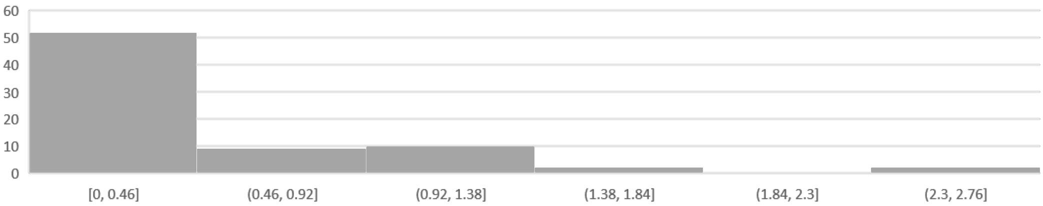

Figure 9: Number of 10K reports with a percentage of ecological commitments between [x1, x2]

25 out of 77 reports have 0 sentences labelled as ecological commitments (around 1/3), 80% of the reports have less than 1% and none of the report have more that 3%. Companies also tend to do worst through time: if some companies, as for example Kellogg Company, never had any positively classified sentences, other, as for example Tesla, fell from 1.43% (one of the best score) to 0 between 2019 and 2022. More generally, nearly all companies had a drop in 2020.
This results seem coherent. First, companies are moving with the previous Vader results, were we could observe a negative slope on ESG-related sentiment for a lot of companies. Moreover, the fall in 2020 could be related to COVID-19, a period when companies did not prioritize ecology in their 10K-reports and as their major issues. Finally, a low percentage of ecological commitments is logical given the 10K format, which is not an ESG statement and mostly talk about financial performance.

However, it can be surprising for the common thinking that companies as Tesla have none ecological statements; it may actually be due to the over enthusiastic tone of this company which has in its reports sentences like: "Our mission to accelerate the world's transition to sustainable energy, engineering expertise, vertically integrated business model and focus on user experience differentiate us from other companies." (Tesla 10K, 2022) that can be potentially classified as green washing.

"Sentiment Analysis of ESG disclosures on Stock Market"

This article [4] focuses on the sentiment analysis of ESG scores of four prominent companies in their respective sectors: Amazon in the e-commerce sector, Tesla in the automobile sector, HSBC Bank in the banking sector and Goldman Sachs in the finance sector. These companies were chosen after having taken into consideration the fact that ESG sentiment analysis is a fairly recent way of correlating stock market prices to the prominence of Environmental, Social and Governance importance for companies.

This model is based on newspaper articles and real-time Twitter tweets: at each of them will be attributed a sentiment value among $\{-1, 0, 1\}$, where the values correspond to a negative, neutral and positive sentiment respectively. For each data point, the sentiment value is multiplied with a sentiment score in $[0, 1]$. The sentiment score is the output of a pre-trained python dictionary FinBERT model applied to finance-related material which returns a sentiment index (or compound) in $[-1, 1]$ for each data point.

The study shows that Goldman Sachs, Amazon and Tesla had a decreasing sequence of sentiment index respectively equal to $0.918, 0.875, 0.639$; on the contrary, HSBC Bank had a negative sentiment index close to -0.95. This allowed the authors to conclude that the stock prices of the positive sentiment-indexed companies are positively correlated to ESG news published about them, while HSBC Bank's stock prices are inversely correlated, given the negative sentiment index.

However, there are some limitations in the analysis: in particular, the sample size of tweets and news articles could skew conclusions if the data set is too small or if one doubts the legitimacy of sources or their credibility. With a wide range of data points, one could find a more accurate measure of ESG sentiment score.

A second BERT based approach: FinBert for sentiment

We then implemented a FinBert (a Bert pre-trained on financial data) fine-tuned for sentiment analysis: finbert-tone [10]. We selected ESG-related sentences in the same way we did for Vader and polarity approaches. For each sentence, the model returns a single label $\in$ {-1, 0, 1} for negative, neutral or positive sentiment. We then computed a global score for each ESG-related report as

$$score = \frac{\Sigma_S(\text{sentence S label})}{\text{total number ESG sentences}}$$

This score is thus on the same scale of Vader's. We can observe similarities (really low score of Petrogas, ...) and differences (no so big increase of ConocoPhillips, ...).
Unlike the Vader approach, we did not try to run FinBert on the whole text, since BERT model have a 512 tokens limit: basically, it is impossible to run a BERT based model on texts of more than 512 words. However, there are some techniques to adapt the input in order to avoid this issue. The most used one is the sliding windows technique: it consists in selecting 512 long "slices" of text with 50 tokens of overlaps. It did not seem pertinent here, as the ESG-related selected sentences are selected from the whole text: two sentences following each other did not actually followed each other in the 10K-report. Running a sliding-window BERT on this set of ESG sentences would thus give a "false context" to FinBert.
In figure 10 is reported the evolution of the BERT score for each company.

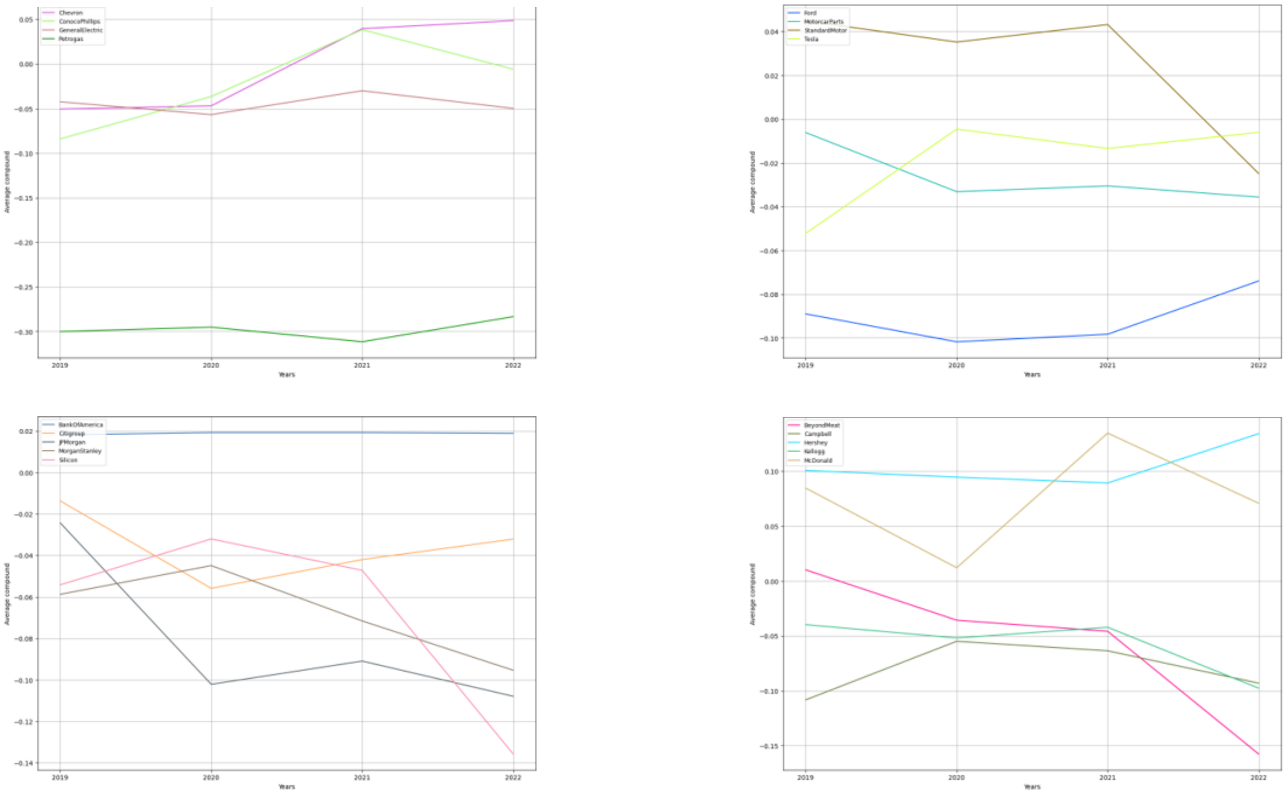

Figure 10: Evolution of FinBert sentiment score by industry

Using BERT to select ESG sentences?

We then thought about the limitation of the dictionary-based approach used to select ESG-related sentences. As this approach was used in the whole document (except the first BERT based approach), should it create any bias, this bias would be shared in all our methods.
Another solution would be to select sentences using FinBert fine tuned to recognised ESG sentences (FinBert-ESG [11]). In theory, it could be more precise, as it could take into account the whole meaning of the sentences and not just the presence or absence of words that are classified as ESG-related. We could also run it using a sliding-window method, to take into account the context of the sentences. Thus, a group of two sentences "Renewable energy is good. It should be used more." would be classified as ESG-related, while the second sentences would not be on a dictionary based approach.

However, the sliding window method raises another issue for us: what we wanted to extract from the 10K-reports is the sentiment only on ESG-related subject; our method to compute sentiment were all trained on financial news and are likely to be really sensible about good or bad financial announcement (more than good or bad ESG announcement). Then, we need to be really sure that we run the sentiment analysis only on ESG related part of the text. With a 512 windows, the chances to have both ESG and non-ESG information increase a lot. Then, a "bad" financial news followed with a "good" ESG news would probably be classified as an ESG-related topic with "bad" sentiment. This risk seems high, since when we looked at the 10K-reports, the sentences that are selected from them based on dictionary-approach seemed scattered in a lot of the document part and most of them are not regroup in ESG-linked paragraphs. We thus chose not to implement a sliding window technique, as we are afraid that the induced final bias will be bigger than the gains.

However, we still tried to run an ESG selection of sentences based on BERT, but sentence per sentence. We noticed on the 2022 10K-report of Ford that 429 sentenced were selected over 3885, i.e. 11%. This is the same percentage that we found with the dictionary approach and the sentences seemed to be the same. Given the length of the 10K report, we deemed the BERT selection output too long (around 25minutes/report) to be implemented on all our 72 10K-reports, especially given the similarity of result observed on the report used for test.

Regression

Our aim now was to compare the sentiment we obtained using the Vader and FinBERT algorithms with the ESG scores that are actually attributed to the companies we studied. We decided to conduct multiple regressions using both the average compound score provided by the approach that used Vader and the score provided by the approach that used FinBERT. In order to ensure that our results are comparable with the actual ESG score, we had to verify that the ESG metric we want to use is one that has been computed across industries. We, therefore, chose to use the MSCI Weighted average ESG Score which indicates how well the index constituents manage their most material ESG risks; this score ranges from 10 (best) to 0 (worst) [12].

The MSCI Weighted average ESG Score is computed in the following way :

$$\text{ESG SCORE} = \sum_{i=1}^{n}(Weight_i \times INDUSTRY\_ADJUSTED\_SCORE_i)/\sum_{i=1}^{n}(Weight_i)$$

where

- i = index security with ESG score available
- $Weight_i$ = closing index weight for security i
- $INDUSTRY\_ADJUSTED\_SCORE_i$ = final IndustryAdjusted Company Score: this score is calculated by normalizing the Weighted Average Key Issue Score to the Industry peer set, adjusted to reflect any Ratings Review Committee overrides.

In the following there are the different regression we have carried out:

- with Vader sentiment:
  - Bank of America :
    - X = [average_compound_2019, average_compound_2020, average_compound_2021]
    - Y = [ESG_score_MSCI_2019, ESG_score_MSCI_2020, ESG_score_MSCI_2021]

    We regressed Y on X and obtained the following results:

	Coefficients	Standard errors	P> |t|
x1	25.6131	0.626	0.001
R^2	0.999		

Table 2: Bank of America: Regression of ESG MSCI score on FinBERT sentiment

We also did the same type of regression with the following companies: Hershey, Tesla, and General Electric. For the three of them, we also obtained positive coefficients with high significance. We can, therefore, conclude that for these companies, the Vader sentiment seems to match the ESG score over the years.

	Coefficients	Standard errors	P> |t|
x1	-0.4944	1.679	0.773
const	5.0079	0.508	0.000
R^2	0.006		

Table 3: Bank of America: Regression of ESG MSCI score on FinBERT sentiment

- All companies (with an available ESG score in MSCI) in 2019:
  - X = [average_compound_2019] (for all companies)
  - Y = [ESG_score_MSCI_2019]
  We regressed Y on X and obtained the following results:

  We also did the same type of regression over 2020 and 2021. However, in all the three years, we do not have significance. We can not arrive at the same conclusion as before.

- Similarly with BERT sentiment:
  - Bank of America:
    - X = [average_compound_2019, average_compound_2020, average_compound_2021]
    - Y = [ESG_scores_MSCI_2019, ESG_scores_MSCI_2020, ESG_scores_MSCI_2021]
    We regressed Y on X and obtained the following results:

	Coefficients	Standard errors	P> |t|
x1	4.5475	0.182	0.002
R^2	0.997		

Table 4: Bank of America: Regression of ESG MSCI score on FinBERT sentiment

We also, just as with the Vader approach, did the same type of regression with the following companies: Hershey, Tesla, and General Electric. For the three of them, we also obtained positive coefficients with high significance. We can, therefore, conclude that for these companies, the BERT sentiment seems to match the ESG score over the years.

  - All companies (with an available ESG score in MSCI) in 2019:
    - X = [average_compound_2019] (for all companies)
    - Y = [ESG_scores_MSCI_2019]
    We regressed Y on X and obtained the following results:

	Coefficients	Standard errors	P> |t|
x1	-2.9666	3.100	0.355
const	4.7956	0.187	0.000
R^2	0.061		

Table 5: Bank of America: Regression of ESG MSCI score on FinBERT sentiment

We also did the same type of regression over 2020 and 2021. However, in all the three years, we do not have significance. We can not arrive at the same conclusion as before.

Conclusion

A comparison between FinBert and Vader

The polarity approach returns results that are greatly different from Vader and FinBert ones, while between these last two methods there are both divergences and similarities. In fact, on first sight, company by company, Vader and FinBert results, vary a lot (figure 11): except for Petrogas that stays last with both clasements, we can not even classify the same companies in "good", "medium" and "bad" ones.

Vader- classed by best score	FinBERT- classed by best score
BankofAmerica	Hershey
Hershey	McDonalds
Royal_Energy	Chevron
Tesla	BankofAmerica
JPMorgan	Conoco
Kellogg	Tesla
McDonalds	Standard_Motor_Products
Ford	Motorcar_Parts_America
Silicon	CityGroup
CityGroup	Silicon
General_Electric	General_Electric
Standard_Motor_Products	Ford
MS	Campbell
Conoco	MS
Beyond_Meat	Royal_Energy
Rivian	Kellogg
Motorcar_Parts_America	JPMorgan
Campbell	Rivian
Chevron	Beyond_Meat
Petrogas	Petrogas

Figure 11: 2022' classements of companies per sentiment score via FinBERT and VADER

But then, on the other hand, the evolution industry per industry behave the same. Despite the difference in magnitude, the energy industry is the only one to have a positive average change, while the food industry behave the worst. Automotive and banking sector are similar with a medium average change (table 6). This works in favor of this two approaches, that are coherent despite the two different mean of computation.

Vader

Industry	change (%)
Automotive	-5.22
Banking	-1.70
Energy	21.76
Food	-5.78

Bert

Industry	change (%)
Automotive	-132
Banking	-137
Energy	69
Food	-349

Table 6: Average change per industry comparison

In conclusion, we tried three methods to evaluate ESG-related sentiment on 10K-reports over time and one specific to ecology commitments. Over the three ESG methods, two approaches (Vader and FinBert) allows to compute coherently the industry evolution over time of ESG sentiment in 10K-reports. Thus, both seem worthy candidate to evaluate the ESG sentiment evolution, at least for a set of reports in the same industry. Finally, with these two approaches, we found, for certain companies, a positive and significant correlation with the ESG scores provided by MSCI. Thereby, we would keep the Vader and FinBert based computation of sentiment as our best methods to measure ESG sentiment evolution over time.

Appendix

Figure 12: Words clouds polarity approach (energy and automotive sector)

Figure 13: Words clouds polarity approach (banking and food sector)

References

[1] Edgar database. [Online]. Available: https://www.sec.gov/edgar/search/#/q=reports&dateRange=all&filter_forms=10-K

[2] "Sustainable news – a sentiment analysis of the effect of esg information on stock prices," 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3809657

[3] A dataset for detecting real-world environmental claims. [Online]. Available: https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/568978/CLE_WP_2022_07.pdf?sequence=1

[4] (2022) Sentiment analysis of esg disclosures on stock market. [Online]. Available: https://www.researchgate.net/publication/364126184_Sentiment_Analysis_of_ESG_disclosures_on_Stock_Market

[5] (2010) When is a liability not a liability? textual analysis, dictionaries, and 10-ks. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1331573

[6] (2003) Sentiment analysis: Capturing favorability using natural language processing. [Online]. Available: https://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=1830388

[7] (2018) Sustainability and corporate social responsibility in the text of annual reports—the case of the it services industry. [Online]. Available: https://www.mdpi.com/2071-1050/10/11/4119

[8] Loughran-mcdonald master dictionary w/ sentiment word lists. [Online]. Available: https://sraf.nd.edu/loughranmcdonald-master-dictionary/

[9] "Hutto, c.j. gilbert, e.e. (2014). vader: A parsimonious rule-based model for sentiment analysis of social media text. eighth international conference on weblogs and social media (icwsm-14)." 2014. [Online]. Available: https://github.com/cjhutto/vaderSentiment

[10] Finbert-tone. [Online]. Available: https://huggingface.co/yiyanghkust/finbert-tone

[11] Finbert-esg. [Online]. Available: https://huggingface.co/yiyanghkust/finbert-esg

[12] Msci esg scores. [Online]. Available: https://www.msci.com/documents/10199/1283513/MSCI_ESG_Metrics_Calc_Methodology_Dec2020.pdf/92a299cb-0dbc-63ba-debb-e821bd2e2b08