Summary Report

Data

As the aim of the project is to analyze the evolution of corporate executives' sentiment, attitude, and stance towards China by constructing a time-series representation, it is crucial to have a way to sample relatively frequently in a consistent format. For this reason, we move away from annual shareholder letters and instead use quarterly earnings call transcripts.

Analysis Method

To analyze the sentiment of the executives, we use Natural Language Processing methods in financial literature, rather than using LLMs which is more of a blackbox and does not yet have a robust academic track record in financial literature yet (even though I am sure it would perform very well). The method of interest was employed in *Firm-Level Political Risk: Measurement and Effects* (Hassan, 2019 Quarterly Journal of Economics) in which the authors define sentiment as

$$Sentiment_{it} = \frac{\sum_{b}^{B_{it}} S(b)}{B_{it}}$$

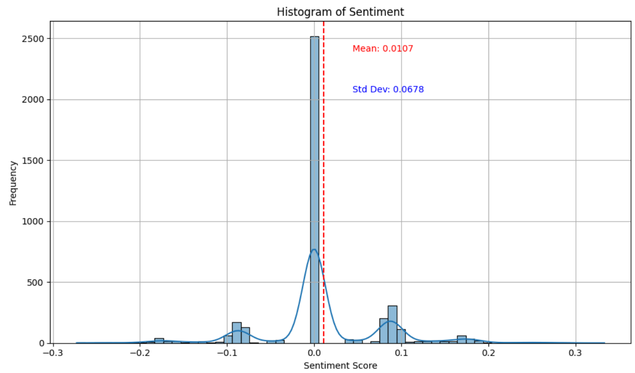
of the context window, where B_{it} is the total number of bigrams in the context window, $S(b)$ is a function that assigns a value of +1 if a bigram is associated with positive sentiment and -1 if associated with negative sentiment, and 0 otherwise. We follow the proposed method by choosing a context window "within the set of 10 words surrounding a synonym … on either side" (Hassan, 2019) for `["China", "Chinese"]` and a list of Chinese cities. The python notebook for the sentiment analysis can be found in the Github repo as `Overseas_Project_Week1_textual_analysis_for_sentiment_towards_China.ipynb`.

Furthermore, $S(b)$ utilizes a sentiment dictionary from *When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks* (Loughran and McDonald, 2011 Journal of Finance). The sentiment dictionary `Loughran-McDonald_MasterDictionary_1993-2024.csv`, which is utilized in our textual sentiment analysis, can also be found in the Github repo.

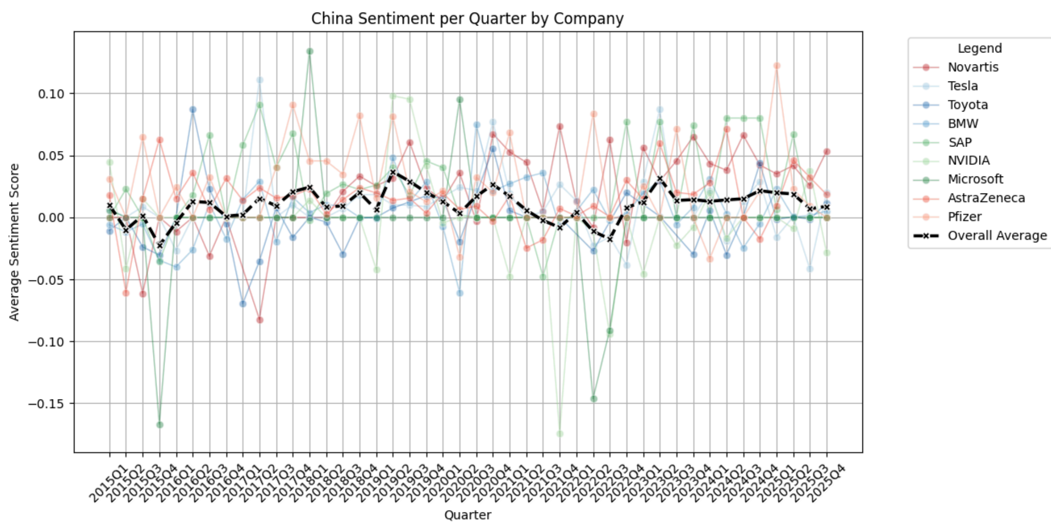
The two papers referenced are both published in established journals and have 1500+ and 7500+ citations, respectively. Additionally, the sentiment analysis method used in the paper also analyzes earnings calls, so using this method is very appropriate.

Results

Looking at the histogram to check the sentiment distribution, we notice that the average leans slightly positive (mean = 0.01) and the shape of the distribution shows up in lumps, likely because the method we used only looks at +10/-10 word context. The positive tendency of executive commentaries is not unexpected as executives will generally want to avoid saying anything particularly negative about any particular country unless absolutely necessary.



When plotting individual companies, no trends are immediately evident. Looking at the overall average, we notice that there is a downward shift towards negative sentiment around 2020 due to the pandemic, and a bounceback into positive sentiment as mutations of coronavirus-19 became less lethal and China began to loosen lockdown restrictions.



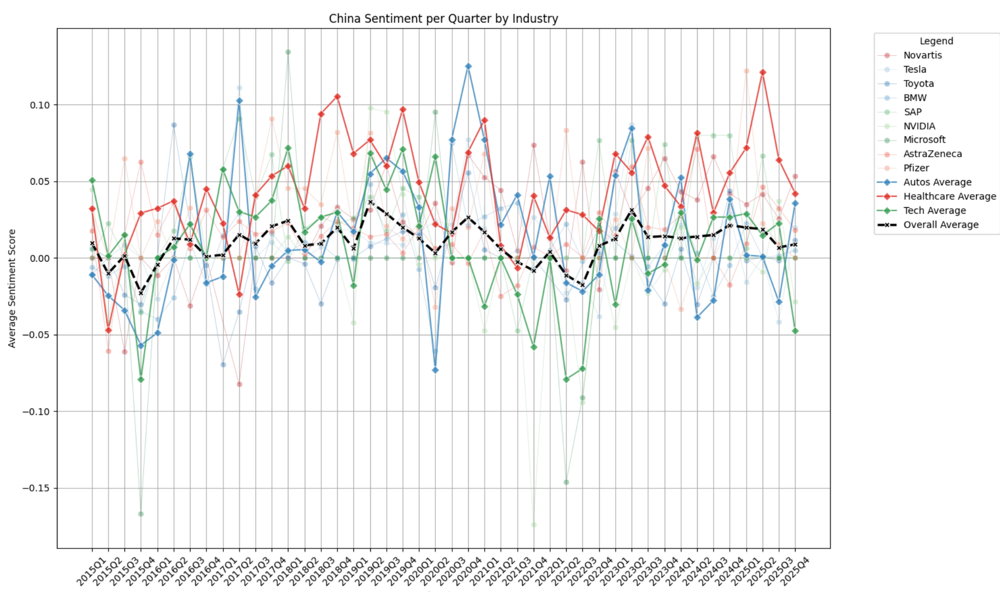
When the sentiment scores are aggregated on the industry level, we can see a clearer trend.

Looking at tech companies, their sentiment towards China remained positive until 2020 as most tech companies cited “software revenue growth” in China (SAP, 2017Q2) and “China became one of the largest gaming markets” (NVIDIA, 2018Q1), particularly “tremendous growth of Minecraft expansion” was seen (Microsoft, 2018Q1). However, tech companies admitted that things were at the “outset of the pandemic, pretty difficult” (SAP, 2021Q3) as “ impact of shutdowns in China” (Microsoft, 2022Q2) were felt, while regulation from China became harsher as some “transaction has been under review by China Antitrust Authority” (NVIDIA, 2021Q4). Such negative sentiment continued after things bounced back after lockdowns were lifted as the Biden administration continued the export restrictions that was originally imposed by the previous Trump administration such that “restrictions prohibiting the sale of our Data Center GPUs to China” (NVIDIA, 2022Q4).

Healthcare, on the other hand, does not seem to be negatively impacted by the pandemic and continues to have a positive sentiment towards China throughout the decade window. Comments such as

“China being a very positive growth driver” (Pfizer, 2016Q1) and “China remains very strong” (AstraZeneca, 2017Q2) seem to suggest that China was constantly a strong market for healthcare. Even during the pandemic, some healthcare companies had access to “China, which is already covered through a separate BioNTech collaboration” (Pfizer, 2020Q2), but the positive sentiment was due to undisrupted business throughout the pandemic thanks to “the loyal customer and the patients” (AstraZeneca, 2020Q3). Even after the pandemic, the outlook remained positive as healthcare companies “expect China growth to accelerate” (Novartis, 2023Q1).

Lastly, autos do not seem to have a clear trend aside from the sharp drop at the onset of the pandemic and the equally abrupt comeback once remote work became a thing. We can see Elon Musk being strategic and hopeful about breaking into the China market “long-term we do want to succeed in China” (Tesla, 2015Q1). Companies were seen to be “strengthening our local production at Shenyang” (BMW, 2015Q1) and putting an emphasis on EVs as they have “been successfully in getting EV plate exemption everywhere” (Tesla, 2015Q3) and were seeing “good performance of the Chinese business” (Toyota, 2017Q3). However, once the pandemic hit, “weaker demand in China, resulting from the Coronavirus and closure of many dealerships worldwide” (Toyota, 2020Q2). But it soon became clear that “successful performance was obviously the handling of the pandemic in China” (BMW, 2021Q1).



Conclusion

This NLP method can be thought of sampling positive words and negative words defined from the aforementioned sentiment dictionary. In essence, the method relies on the Law of Large Numbers to kick in as sample sizes increase. In this case, this is attained with approximately 4000 samples. However, inspecting the data line by line, it is clear that there are a lot of erroneous entries in the raw data.

Even though LLMs may be a blackbox, it has proven to be robust, so utilizing LLMs for contextual sentiment analysis looks to be the way to go. For this time, I wanted to try rudimentary NLP techniques to see how effective this particular method was used in a published paper in an established journal. However, it is clear that the method is not robust enough, and I am honestly surprised the paper has 1500 citations. Perhaps using just the transformer layer (like BERT) instead of a full LLM model might be a natural next step.