

AS6 Report Naman Kedia

Relevance vs Irrelevance Selection

Using the keyword classifier given to us we can filter out portions of the sentence that are relevant or irrelevant. To make the computation effective we did the following -

- Extract only important words by removing stopwords
- Stemming and Lemmitizer for checking variations of keywords (eg. attacking, attack, attacks)

Constructing a Measure of Cybersecurity Risk

There are 2 metrics we can construct:

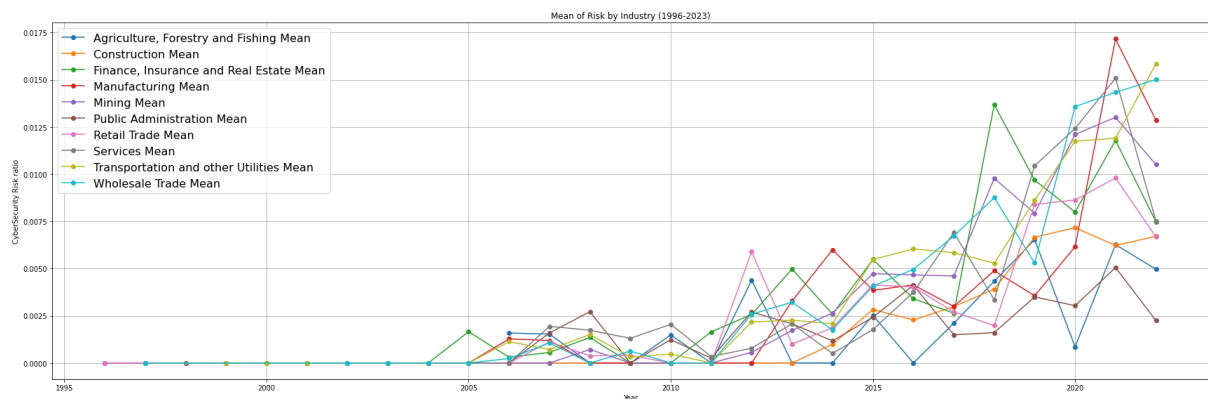
- *Cyber Security Risk Ratio (Relevant/Total) [Option 1]*
- *Absolute Cyber security risk (Relevant) [Option 2]*

We will proceed with the Cybersecurity Ratio [Option 1]. This approach provides a normalized measure, allowing for fair comparisons across firms of varying sizes and lengths of disclosures. By focusing on the proportion of relevant sentences, it highlights the relative emphasis a firm places on cybersecurity risk within its overall risk narrative. The absolute measure can be a slightly unfair metric to compare and analyse across industries as longer 10K filings could mention the risk more but that doesn't imply it is a key concern if the ratio of the risk to the total filing is relatively less.

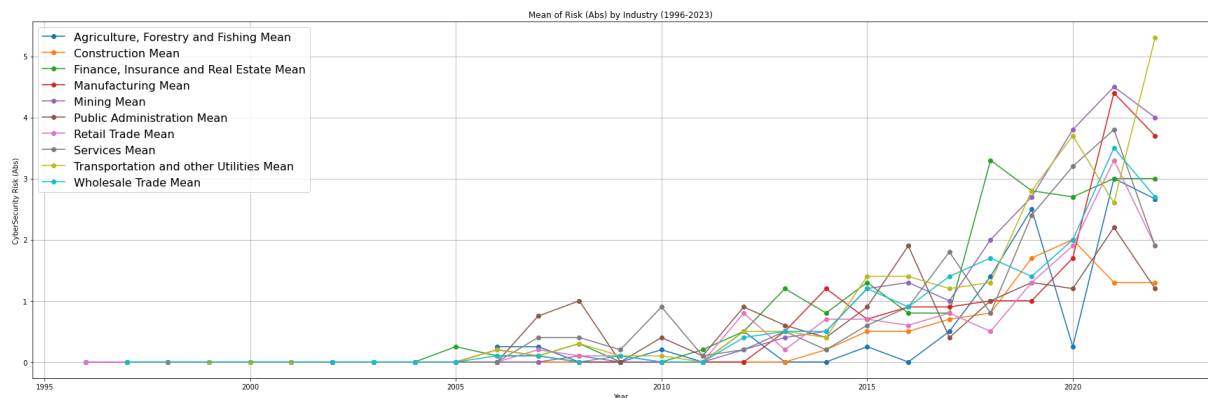
However just from a point of analysis we can see that both the metrics have been showing a consistently increasing risk across the years which points towards the higher operation risk across sectors, and also indicates the dependence of all sectors on the internet with time. It is easy to note that industries across had much lower cybersecurity risk (almost 0) almost till 2005 post which there has been a steady threat reported.

Descriptive Analysis

Mean Risk Ratio against years (1996-2023) [Industry Wise]



Mean Absolute Risk against years (1996-2023) [Industry Wise]



We should note that the 10K filings since 1996-2003 were very limited in the dataset and therefore we see descriptive/noteworthy graph trends only post 2004

AS6 Report Naman Kedia

High Risk Ratio:

- Finance and Insurance: Exhibits the highest cybersecurity risk ratio. (eg. Banks, Insurance firms)
- Services, Transport Industry: Follow closely behind in risk exposure. (eg. Software, Air transportation)
- These industries heavily utilize technology (e.g., cloud computing, banking servers, air traffic, personal data), necessitating stringent cybersecurity measures. Their heightened risk awareness is reflected in their 10-K filings.
- We can also see that the max value lies in the service industry. This too backs the argument that the IT sector is highly risk prone to cyberthreat. The high skewness seen in this could be one of the reasons that the sample had some outliers as well.

Low Risk Ratio:

- Agricultural Industry: Shows the lowest cybersecurity risk ratio.
- Construction Industry: Also ranks low in risk exposure.
- These traditional sectors rely more on manual labour and have lower exposure to cyber threats. Although they are adopting technology, their relative risk remains significantly lower compared to tech-intensive industries.

	N	mean	std_dev	min	max	1%	5%	25%	50%	75%	95%	99%	skewness	kurtosis
industry														
Agriculture, Forestry and Fishing	62.0	0.002304	0.003511	0.0	0.011218	0.0	0.0	0.0	0.0	0.004656	0.010977	0.011203	1.319423	0.621444
Construction	170.0	0.002338	0.004811	0.0	0.027972	0.0	0.0	0.0	0.0	0.002776	0.014504	0.019531	2.684829	7.872704
Finance, Insurance and Real Estate	185.0	0.004151	0.007885	0.0	0.058442	0.0	0.0	0.0	0.0	0.005319	0.020490	0.029321	3.190970	14.366936
Manufacturing	219.0	0.003074	0.007306	0.0	0.054487	0.0	0.0	0.0	0.0	0.001742	0.020273	0.028064	3.324466	14.002481
Mining	178.0	0.004099	0.007830	0.0	0.049296	0.0	0.0	0.0	0.0	0.005508	0.020326	0.035644	3.026179	11.526359
Public Administration	152.0	0.002164	0.003787	0.0	0.023211	0.0	0.0	0.0	0.0	0.003094	0.009204	0.015340	2.519707	7.855384
Retail Trade	182.0	0.003132	0.006703	0.0	0.048077	0.0	0.0	0.0	0.0	0.003964	0.018670	0.031003	3.364513	14.539731
Services	193.0	0.003728	0.009124	0.0	0.082090	0.0	0.0	0.0	0.0	0.004274	0.012589	0.050628	5.408264	37.167950
Transportation and other Utilities	178.0	0.004572	0.009064	0.0	0.065574	0.0	0.0	0.0	0.0	0.006231	0.021812	0.039215	3.544799	16.526066
Wholesale Trade	177.0	0.004643	0.008893	0.0	0.053097	0.0	0.0	0.0	0.0	0.006289	0.022381	0.043419	2.735728	9.033300

Portfolio Analysis

Only two quintiles were created instead of five. This suggests that there might not be enough unique values in the cybersecurity risk measure to form five distinct groups. This potentially means that -

- the cybersecurity risk ratios are very similar or clustered around certain values, making it difficult to split them into more groups.
- there may not be enough data points to create five meaningful quintiles.

In our analysis we can infer that most of the cybersecurity values found are 0. This makes it tough to make 5 quintiles. Based on this limited distribution we may draw the following conclusions

Equal Weighted Portfolio

Negative Risk Premium:

- The returns are close to zero for both quintiles. Just slightly on the negative side.
- We may very loosely conclude that stocks with higher cybersecurity risk underperform, possibly indicating market penalization or overpricing of high-risk stocks, however they mostly do not impact the stock markets significantly.

Value Weighted Portfolio

Negative Risk Premium:

- The returns are close to zero for both quintiles. Just slightly on the negative side.

AS6 Report Naman Kedia

- Smaller absolute difference than equal weighted, meaning on normalisation on firm size we can conclude more strongly that cybersecurity threats are not directly affecting returns.
- We may very loosely conclude that stocks with higher cybersecurity risk underperform, possibly indicating market penalization or overpricing of high-risk stocks, however they mostly do not impact the stock markets significantly.

```
Average Risk by Quintile:
quintile
(-0.001, 0.00587]    0.000703
(0.00587, 0.0821]   0.014784
Name: cybersecurity_risk_ratio, dtype: float64
```

```
Equal weighted Portfolio distributions:
Average Excess Return by Quintile:
quintile
(-0.001, 0.00587]    0.018145
(0.00587, 0.0821]   0.016604
Name: returns_excess, dtype: float64
Difference in Excess Return (High Risk - Low Risk): -0.001540917201965964
```

```
Value weighted Portfolio distributions:
Average Excess Return by Quintile:
quintile
(-0.001, 0.00587]    5.348635e-08
(0.00587, 0.0821]   -3.395542e-08
Name: return_weighted, dtype: float64
Difference in Excess Return Value Weighted (High Risk - Low Risk): -8.744177579592994e-08
```

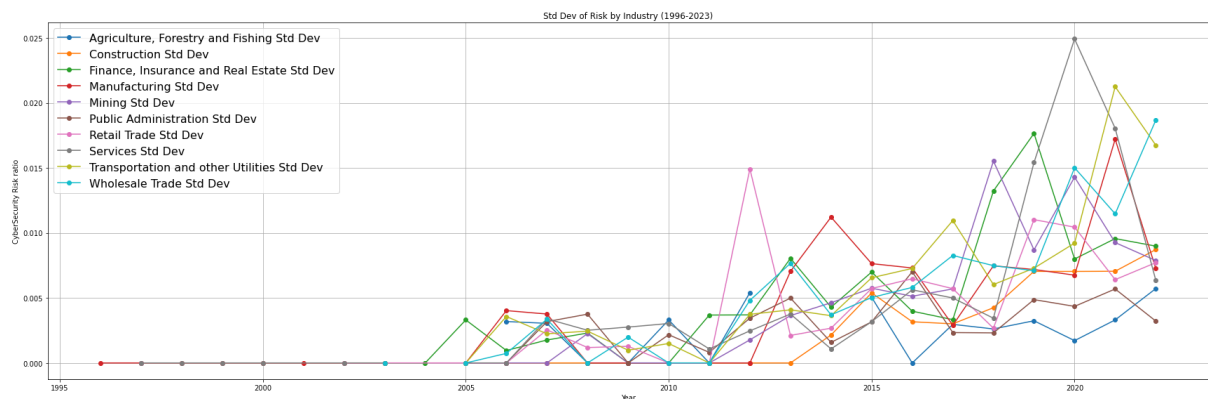
To avoid the situation of not having 5 quintiles- we can try a few things like:

- Taking a larger sample space to have more data with non-zero cybersecurity risk
- Put all 0 values in 1 quintile and make 4 quintiles on the rest of the data.

***** END OF REPORT *****

Other Graphs

Std Deviation



References

- ChatGPT used for generating a few visualisation portions and some keyword selection metrics