Quantitative finance for TM

Daniel Celeny, Loïc Maréchal

June 22, 2023

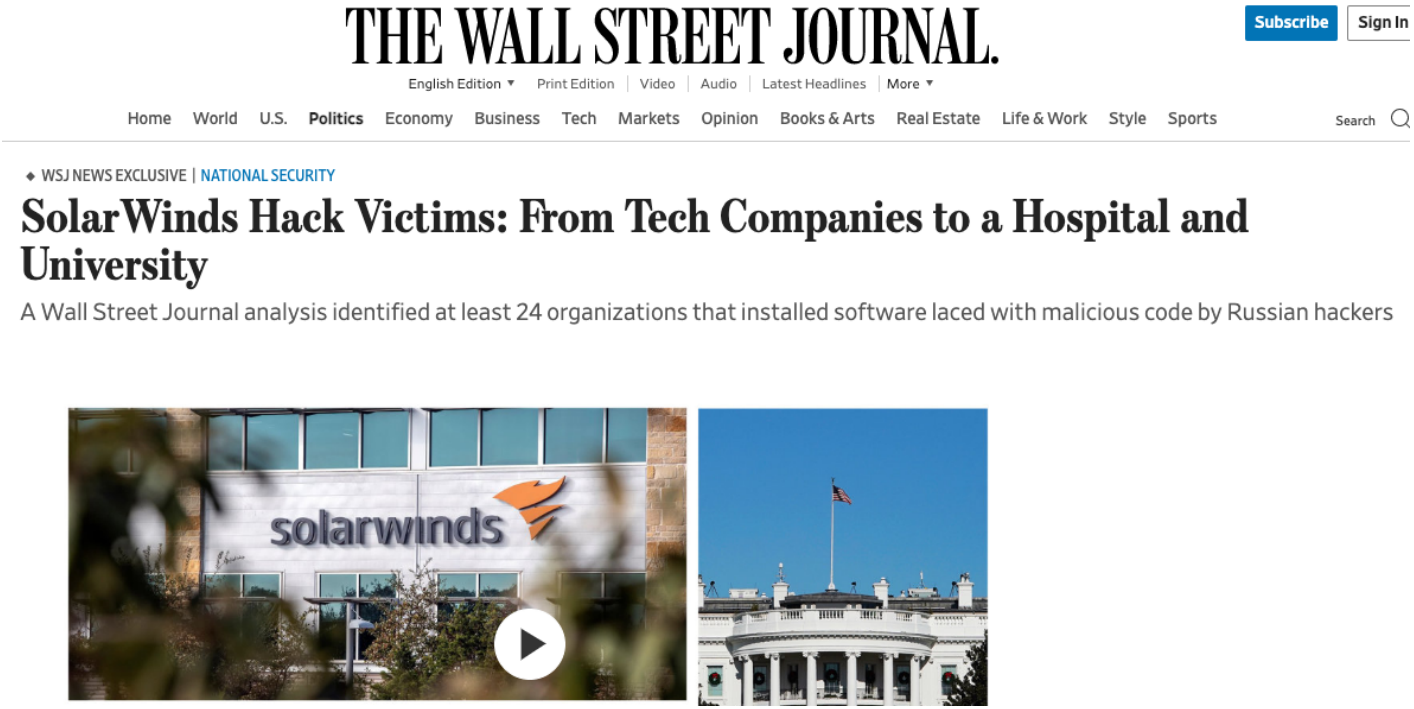
Risk and return

- Price of an investment: sum of all expected (futures) discounted cash flows
- Same cash flow expectation can result in different prices as investors discount stocks with high "risk"
- Hence, for a similar investment, more risk implies larger returns
- Relevant for investors, but also for researchers as we directly observe prices and returns

What is risk?

- Any characteristic that makes an investor dislike (discount) an investment
- Total SD, market risk, pricing factors, and "factor zoo"
- Equivalence between risk premium affecting the price (market capitalization of firms) and the insurance premium
- Estimating the risk premium gives us the amount of cyber security or cyber insurance that a firm or an economy as a whole is willing to invest in

Does cyber risk matter? - Indirect evidence



- December 13, 2020: government agencies and large companies attacked through SolarWinds' Orion software. The company stated in a SEC filing that 18,000 of Orion customers were affected

- The cross-sectional variation in stock prices around the SolarWinds event informs us about the exposure of each firm to Orion Software

Cyber risk measure

- Researchers use direct measurements (*e.g.* market capitalization the book value of equity) or extract scores from text
- In particular, compulsory filings required by regulators (SEC in the US) give consistent text corpus for each listed firm
- Cyber risk premium corresponds to the price of a cyber insurance contract, or acquisition of a cyber security product that a company is willing to subscribe to make this risk disappear

Current state of the literature on cyber risk

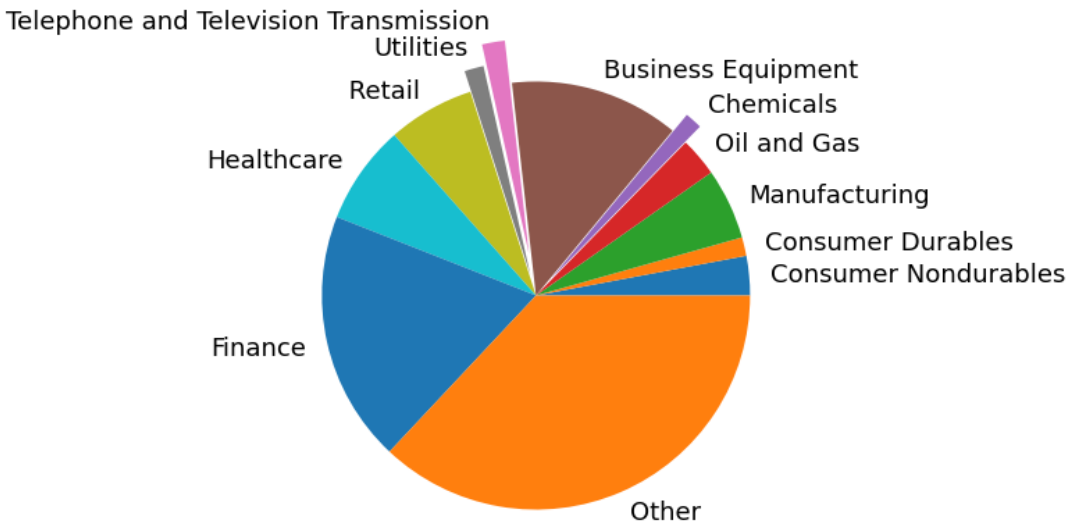
- Jamilov, Rey, and Tahoun (2021)
- Florackis, Louca, Michaely, and Weber (2022)

Data

- SEC-EDGAR-10Ks
  - A financial filing submitted by public companies to the U.S. Securities and Exchange Commission (SEC) annually
  - Includes information such as the company's financial statements, risk factors, and executive compensation
- CRSP-COMPUSTAT
  - Financial information databases, providing data on publicly traded companies, such as stock price information and key financial ratios
- MITRE ATT&CK
  - A knowledge base of cyber security adversary tactics and techniques based on real-world observations
  - The knowledge base is a breakdown and classification of offensively oriented actions that can be used against particular platforms, such as Windows

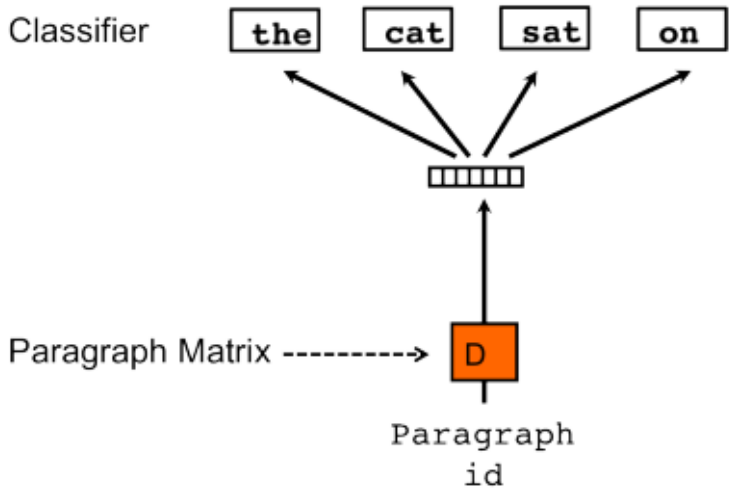
Overview of the data

- January 2007-December 2022 period
- 7059 firms encompassing various industries
- 60470 10-K statements
- 785 MITRE technique descriptions



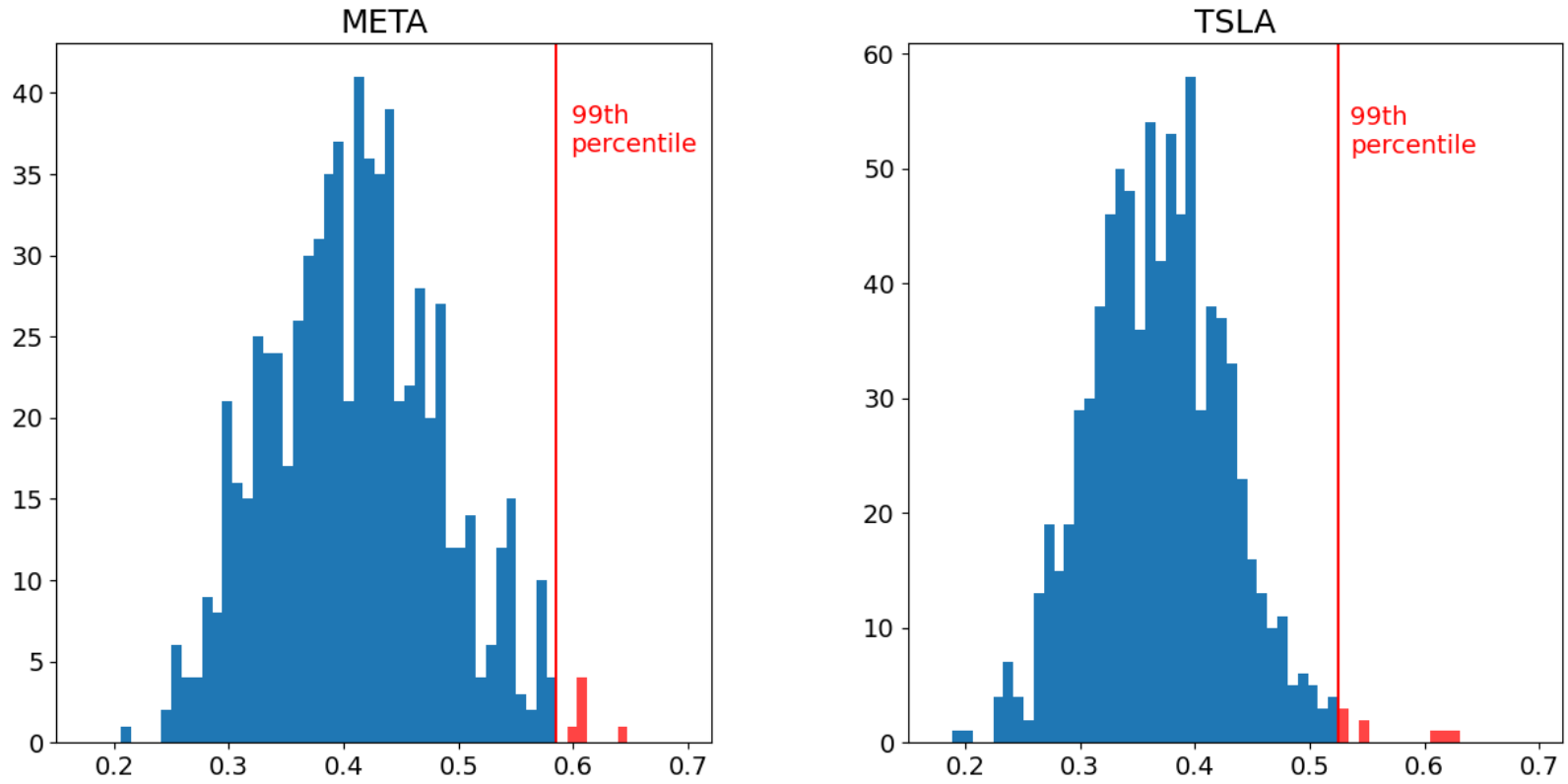
Methodology

- Doc2Vec (Le and Mikolov (2014))
  - Distributed bag of words (DBOW) model with vectors of size 200
    - During training, sample a window of words from the paragraph
    - Classification task: predict the words from the window given the paragraph vector
  - Trained using paragraphs from 10-Ks filed in 2007 + MITRE
  - More than 1.7 million training paragraphs, 44 words per paragraph on average



Methodology

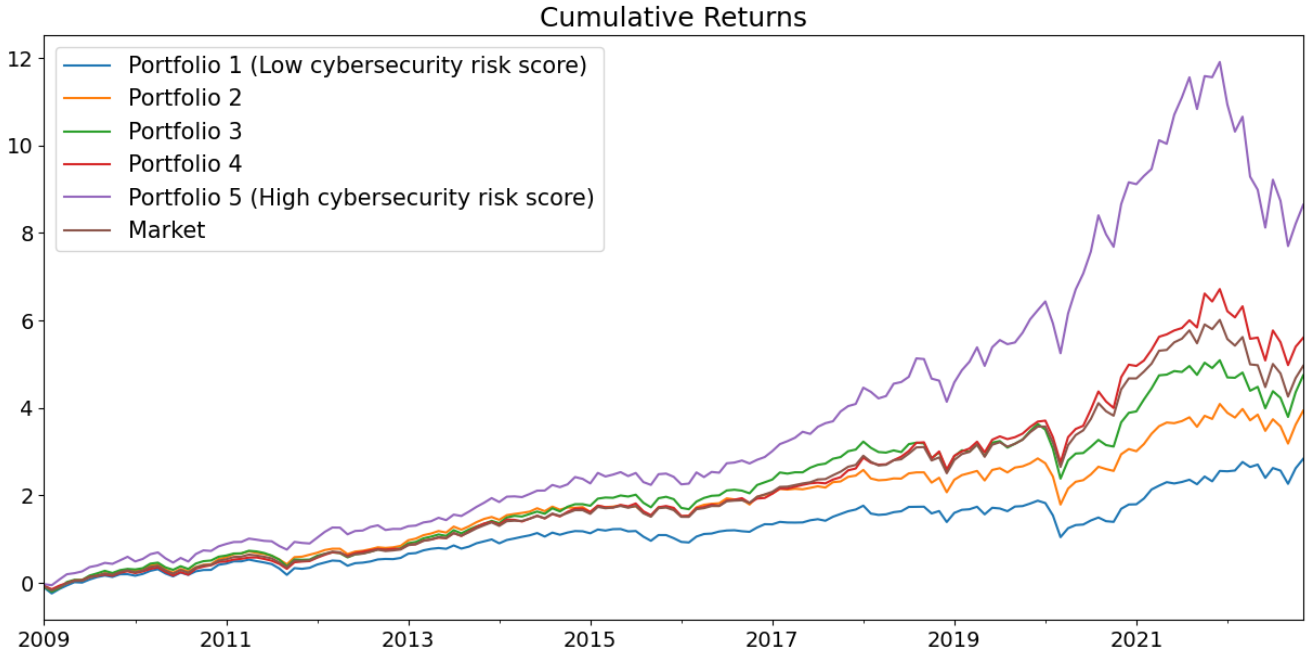
- Cyber security score
  - Paragraph score = maximum cosine similarity with MITRE descriptions
  - 10-K score = average score of the 1% of its highest-scoring paragraphs



Distribution of paragraph scores, 10-Ks filed in 2022

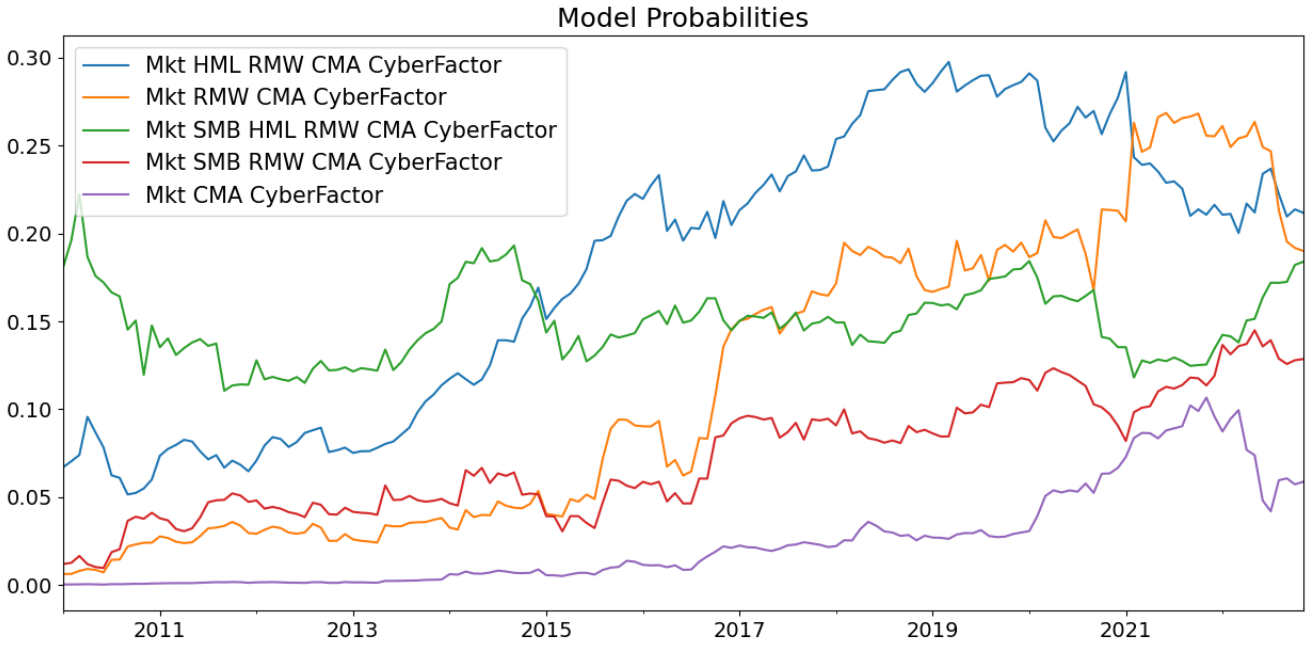
Results

- Univariate value-weighted portfolio sorts.
  - Sort firms into portfolios based on their cyber security risk score
  - Rebalance the portfolios quarterly
  - Average return and alpha increases with cyber security risk
  - Buy high cyber security risk stocks: AAR of 18.72%
  - Buy high and sell low cyber security risk stocks: AAR of 6.93%
  - Implied market price of cyber security risk $\approx$ 417 billion USD/year



Results

- Bayesian selection of factors (Barillas and Shanken (2018))
  - Compute the probability that a factor model is the best in terms of pricing returns
  - Cyber security factor = buy the high cyber security risk portfolio and sell the low cyber security risk portfolio
  - The five most likely models all contain the cyber security factor



Next steps

- Implement Instrumented Principle Component Analysis (Kelly, Pruitt, and Su (2019)), a method that makes it possible to treat cyber security as a characteristic rather than a risk factor and study the dependence of the latent risk factors on cyber security
- Among the high-scoring firms, differentiate between high-risk firms and firms that provide cyber security solutions

Appendix

- Doc2Vec
  - Baseline model from Lau and Baldwin (2016)
  - Chosen model: vector size = 200, window size = 15, min count = 5, sub-sampling = 10^{-5}, negative sampling = 5, epochs = 50
  - Pre-processing: drop common and stop words, remove punctuation and numbers, lowercase

Appendix

- META Platforms 2022:
  'Some errors, bugs, or vulnerabilities inherently may be difficult to detect and may only be discovered after the code has been released for external or internal use. For example, in September 2018, we announced our discovery of a third-party cyber-attack that exploited a vulnerability in Facebook's code to steal user access tokens and access certain profile information from user accounts on Facebook'

  'In addition, third parties may attempt to fraudulently induce employees or users to disclose information in order to gain access to our data or our users' data. Cyber-attacks continue to evolve in sophistication and volume, and inherently may be difficult to detect for long periods of time.'

Appendix

- TSLA Inc 2022:
  'While we have implemented security measures intended to prevent unauthorized access to our information technology networks, our products and their systems, malicious entities have reportedly attempted, and may attempt in the future, to gain unauthorized access to modify, alter and use such networks, products and systems to gain control of, or to change, our products' functionality, user interface and performance characteristics or to gain access to data stored in or generated by our products.'

  'We encourage reporting of potential vulnerabilities in the security of our products through our security vulnerability reporting policy, and we aim to remedy any reported and verified vulnerability. However, there can be no assurance that any vulnerabilities will not be exploited before they can be identified, or that our remediation efforts are or will be successful. Any unauthorized access to or control of our products or their systems or any loss of data could result in legal claims or government investigations.'

References I

Barillas, F., Shanken, J., 2018. Comparing asset pricing models. Journal of Finance 73, 715–754.

Florackis, C., Louca, C., Michaely, R., Weber, M., 2022. Cybersecurity risk. Review of Financial Studies 36, 351–407.

Jamilov, R., Rey, H., Tahoun, A., 2021. The anatomy of cyber risk. Available at: `https://ssrn.com/abstract=3866338`

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Lau, J. H., Baldwin, T., 2016. An empirical evaluation of doc2vec with practical insights into document embedding generation. In: *Proceedings of the 1st Workshop on Representation Learning for NLP*, Association for Computational Linguistics, pp. 78–86.

Le, Q., Mikolov, T., 2014. Distributed representations of sentences and documents. In: *Proceedings of the 31st International Conference on Machine Learning*, PMLR, vol. 32 of *Proceedings of Machine Learning Research*, pp. 1188–1196.