A BERT-based Firm-level Risk Measurement on Epidemic Diseases

Zhou Xing

April 2020

This paper builds on several strands of literature. The idea of constructing a measure of firm-level risk on a specific domain based on conference call text data is based on the series of literature by T.A. Hassen et al. [7] [8] [6], with observation that these calls are a venue in which senior management has to respond directly to questions from market participants about the firm's prospects. Not only are these disclosures therefore timely, but as they consists of a management presentation and, importantly, a Q&A session, they also require management to comment on matters they might not otherwise have voluntarily offered. [9] Hassen et al. discuss risk measurement on three specific topics: political risk [7], Brexit Uncertainty [8], and epidemic diseases [6], which is closely related to our paper. The metric for risk measurement is consistent in this series, which is ultimately based on a simple count of word combinations in earnings call transcripts to measure a given firms exposure to risk, and a simple count of words associated with positive or negative sentiment in earnings call transcripts to measure a firm's expectation of future performance. Our paper is also related to literature that combines economics and epidemiology. Examples include analyses of how private vaccination incentives asset epidemic dynamics and optimal public-health policy with an application of the SIR model to high-frequency data on viral diseases in France [Adda (2016)]. [2]

Economic analysis based on text data and sentiment analysis is a new trend in economics literature. Dougal et al. (2012) [5] provide causal evidence of financial media influencing subsequent investor behavior and stock market returns. Soo et al. (2018) [11] develop measurement of housing sentiment for 34 cities across the United States by quantifying the qualitative tone of local housing news. Atalay et al.(2019)[3] construct data set from the text content of approximately 7.8 million job ads appearing in three major metropolitan newspapers to describe the evolution of work from 1950 to 2000 in the United States. The volumn, timeliness and accessibility of textual data in media, ads and social media platform enable social scientists to capture market sentiment and observe economic trends in diverse niche domains. However, these works applied the most standard methodology employed by this literature, which uses a dictionary-based method to quantify the raw frequency of positive and negative words in a text. To do so, these papers typically identify words as positive or negative based on an external word list. Such method is likely to bring bias without considering the context of sentences. For example, *"This is the best laptop bag ever. It is so good that within two months of use, it is worthy of being used as a grocery bag."* [1]. The innate sarcasm in the review is evident as the user isn't happy with the quality of the bag. However, as the sentence contains words like 'best', 'good' and 'worthy', the review can easily be mistaken to be positive.

NLP (Natural Language Processing) models are powerful tools to address such problems in sentiment analysis. In particular, a pre-trained NLP model can be implemented in our task instead of building a new model from beginning. A bit of fine-tuning of this model will save computational resources and time. In this paper, we aim at leveraging the modeling power of BERT (Devlin et al., 2019) [4], one of the most popular pre-trained language model armed with Transformer (Vaswani et al., 2017)[12]. BERT model has been applied in multiple specific tasks in state-of-art NLP researches. For example, the BERT-enhanced aspect-based sentiment analysis (ABSA) model, which is employed to perform fine-grained sentiment polarity on specific aspects of interest and

allows producers to gain a granular understanding of the users' requirements for specific aspects of their products or services, achieves F1 scores 88.0 on SentiHood dataset and 92.9 on SemEval-2014 dataset, beating former RNN (Recurrent neural network), LSTM(Long short-term memory model) and GRU (Gated Recurrent Unit). [10]. However, the applications of BERT model in sentiment analysis in economics research are rare. Hiew et al.(2019) construct a textual-based sentiment index by adopting BERT to posts that are published on the Chinese social media, which represents the first attempt in the literature to apply this state-of-the-art learning model to the financial sentiment extraction. Their analysis focuses mainly on the individual stock level, by investigating three actively-trading listed companies in Hong Kong Stock Exchange (HKSE) in a pilot study, namely, Tencent (0700.HK), CCB (0939.HK), and Ping An (2318.HK), which all possess a sufficient exposure on *Weibo.com*. Through combining the BERT-based sentiment index with other two types of sentiment indices from the option-implied information and PCA on market data for the above three stocks, they claim to provide a deeper and more general financial sentiment analysis. More specifically, the BERT-based sentiment reflects more about individual investors' opinion, whereas the option-implied one followed by Han (2008) represents more about the institutions' attitude. Zhao et al. (2020) [13] has propose a sentiment analysis and key entity detection approach in finance, which is applied in online financial text mining and public opinion analysis in social media. They consider key entity detection as a sentence matching or Machine Reading Comprehension (MRC) task in different granularity and mainly focus on negative sentimental information. Godbole et al. (2020) apply new transformer models from the BERT-family to improve the current method of binary text classification in the context of economic policy uncertainty and find that all models achieve remarkable results in classifying the given newspaper data.

References

[1] A comment on a laptop.

[2] Jérôme Adda. Economic activity and the spread of viral diseases: Evidence from high frequency data. *The Quarterly Journal of Economics*, 131(2):891–941, 2016.

[3] Enghin Atalay, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. The evolution of work in the united states. *American Economic Journal: Applied Economics*, 12(2):1–34, 2020.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[5] Casey Dougal, Joseph Engelberg, Diego Garcia, and Christopher A Parsons. Journalists and the stock market. *The Review of Financial Studies*, 25(3):639–679, 2012.

[6] Tarek A Hassan, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. Firm-level political risk: Measurement and effects. Working Paper 24029, National Bureau of Economic Research, November 2017.

[7] Tarek Alexander Hassan, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. Firm-level exposure to epidemic diseases: Covid-19, sars, and h1n1. Working Paper 26971, National Bureau of Economic Research, April 2020.

[8] Tarek Alexander Hassan, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. The global impact of brexit uncertainty. Working Paper 26609, National Bureau of Economic Research, January 2020.

[9] Stephan Hollander, Maarten Pronk, and Erik Roelofsen. Does silence speak? an empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, 48(3):531–563, 2010.

[10] Xinlong Li, Xingyu Fu, Guangluan Xu, Yang Yang, Jiuniu Wang, Li Jin, Qing Liu, and Tianyuan Xiang. Enhancing bert representation with context-aware embedding for aspect-based sentiment analysis. *IEEE Access*, PP:1–1, 03 2020.

[11] Cindy K Soo. Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies*, 31(10):3689–3719, 2018.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017.

[13] Lingyun Zhao, Lin Li, and Xinhao Zheng. A bert based sentiment analysis and key entity detection approach for online financial texts, 01 2020.