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

Zhou Xing*

June 10, 2020

Abstract

We adapted a context-based natural language pretraining model, BERT, to construct a measurement of firm-level risk on epidemic diseases, and focus on the impact of COVID-19 since Q4,2019 to Q1,2020. Based on the share of quarterly earnings reference call transcripts, we combined semantics and sentiment analysis to gauge pandemic-related risk faced by individual US firms. We focus on two main topics. First, the effectiveness of the BERT-based classifier in determining the relatedness of a sentence to pandemic. Second, the firm-level and sector-level heterogeneity in epidemic-related risk. The result shows that the BERT-based NLP model outperforms traditional classifier in accuracy. We obtained firm-level and sector-level risk index by applying the classifier on transcripts text, and found that companies in healthcare and consumer goods are most impacted. Sector-wise, financial industry is faced with higher risk, while technology sector is more immune to the epidemic impact.

Keywords: Natural Language Processing, Corporate Finance

*The University of Chicago, Division of Social Science. zhouxing@uchicago.edu

1 Introduction

Since World Health Organization declared a global health emergency on January 30, 2020, the disease has caused 117,052 confirmed cases and 7,270,563 confirmed cases and 411,953 deaths worldwide [1]. The outbreak spread from the Chinese city of Wuhan to more than 180 countries and territories—affecting every continent except Antarctica. Efforts to stamp out the pneumonia-like illness have led to entire nations enforcing lockdowns, widespread halts of international travel, mass layoffs and battered businesses and financial markets. Stock markets around the world crashed. After an Oval Office address by US President Trump failed to calm markets on March 11, major stock indices fell another 10 percent on the following day. Unemployment rate in the U.S reached 14.7 percent in April, 2020, worse than at any time since the Great Depression of the 1930s.

The pandemic puts business worldwide under an extreme situation different from previous outbreaks of epidemic diseases. The world has witnessed six global pandemics since the 20 century: 1918 Pandemic Influenza started from U.S., and one outbreak every decade from 1947 to 1977. More recently, We have seen SARS (severe acute respiratory syndrome) spreading to South East Asia, Canada in 2002-2003, and 2009 H1N1 Pandemic. Both COVID-19 and SARS are caused by corona viruses, but the former differentiates itself from higher intractability with lighter symptoms at the early stages, easier interperson transmission and wider range of damage on patients body including lungs, liver and kidney.[3][2] On the other hand, the impact of COVID-19 is intensified by modern business structure with disrupted international trade, broken financial support along value chains, and halted labor-intensive production. However, with different production mode and technological integration, firms are exposed to diverse level of risks in this turmoil. The need for work-from-home solutions can be realised by laptops, routers, and cloud computing solutions. IoT and internet-based logistics ensures efficiency and safety of online shopping and shipping. The direct application of mass testing with Infra-Red Imaging, AI and 5G, diagnostics with AI-augmented CT scans and remote contact-less examination enable firms to some extent to resume operation. In an effort to aid evidence-based policy responses, in this paper, we construct a time-varying, firm-level measure of exposure to epidemic diseases.

The measure we propose is based on a sentence classifier pretrained with

[1] Till June 10, 2020. source: `bloomberg.com`

a Natural Language Processing algorithm, BERT (Bidirectional Encoder Representations from Transformers), and identifies the exposure of firms to an outbreak of an epidemic disease by counting the number of pandemic-related sentences in the quarterly earnings conference call that public listed firms host with financial analysts. Previous works on firm or national level uncertainty measure adapted a simple counting of the number of times a certain topics mentioned in the text data (such as mentioning of a certain disease in , "Brexit" in Baker et al.(2016)[6].

To better analyze and predict firm-level risks, Hassen et. al(2020)[9] built a firm risk index. They analyzed the counting of the number of times the disease is mentioned in the the quarterly earnings conference call that public listed firms host with financial analysts - but did so with a relatively simple n-gram search term. However, sentences such as *"The continuous stay-at-home policy may hinder the marketing performance in Q2 2020."* involve no mentioning of disease, but are classified as highly pandemic-related. Recently, much more complex classification algorithms have emerged in natural language processing (NLP), which involves context-based semantics analysis. In this project, we apply new transformer models from the BERT-family to improve the current method of binary text classification in the context of epidemic outbreak. We find that all of our models achieve remarkable results in classifying the given newspaper data (AUC's ranging from 0.87-0.90), with RoBERTa achieving the best results compared to BERT, DistilBERT, and ALBERT (as well as the non-transformer benchmarks). This indicates that the models are well equipped to take over tasks that researchers have previously solved in less optimal ways. We then apply the trained classifier on sentences in quarterly earnings conference call transcripts and provide a measure of risk exposure with the ratio of disease-related sentences. We also extend such measurement into sector level, and find that technology, services are more resilient to the impact of epidemic disease.

2 Literature Review

2.1 Risk Measurement

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. [10] [11] [9], 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. [12] Hassen et al. discuss risk measurement on three specific topics: political risk [10], Brexit Uncertainty [11], and epidemic diseases [9], 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)]. [4]

2.2 Text Data Analysis in Economics Research

Economic analysis based on text data and sentiment analysis is a new trend in economics literature. Dougal et al. (2012) [8] provide causal evidence of financial media influencing subsequent investor behavior and stock market returns. Soo et al. (2018) [14] 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)[5] 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.

2.3 Applications of BERT in Economics Research

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) [7], one of the most popular pre-trained language model armed with Transformer (Vaswani et al., 2017)[15]. 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). [13]. 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) [16] 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.

3 Data

Our research is based on two corpus of data. The main part of our analysis relies on **quarterly shared earnings conference call transcripts** where company executives discuss their company's financial performance. The earning calls usually involves three members of the management board, most commonly the CEO (Chief Executive Officer) , CFO (Chief Financial Officer and senior analyst or CMO (Chief Market Officer). During a conference call, investors and analysts can call in over the phone or listen online to hear a company's management comment on the financial results of a recently completed quarter. The duration of earnings conference call is on avarage approximately one hour.

We obtained earnings conference call transcripts of 3597 companies of nine industries from Q3,2019 to Q2, 2020 from *seekalpha.com.* Although related literature obtain the transcripts from Thomson One Financial and Management Database, which is of the form PDF and thus harder to parse and obtain the text, we tend to make use of the html data from websites, but the number of companies may be smaller. The industrial distribution of transcripts among industries is as follows, with classification based on the categories provided by *seekalpha.com.* [Figure 1] The main technology used in deriving transcripts is web scrapping, mostly based on selenium (webdriver) and regular expression modules on Python.

To classify the text into epidemic-disease-related and unrelated, we need a **training corpus for our classification model**. Here we make use of the reports from financial newspaper and label their content based on the keyword. Specifically, we search on Google News with keywords : *covid-19, lockdown, pandemic*, and we label the articles in search results as "epidemic-disease-related". For contrast, news before the outbreak of COVID-19 from same financial media

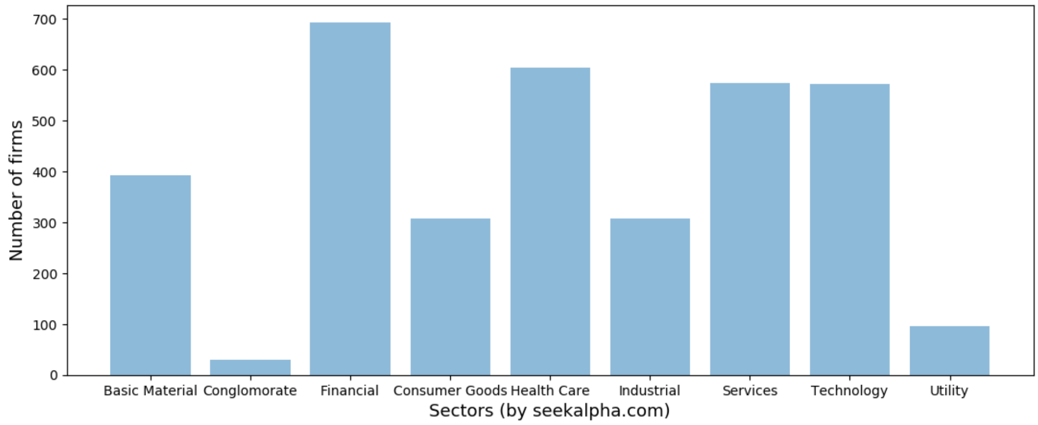

Figure 1: Distribution of firms across sectors

sources are collected, and labeled as "not-epidemic-disease-related". We concentrate on news from September 2019 to December 2019 as the control group. To maintain the business-relatedness, we focus on reports from two major business news media **Bloomberg** and **Financial Times.** We extracted 200 articles with disease-related title, such as *Biotech Tourists Drive Short-Lived Rallies in Covid-19 Stocks*[2], with average length 24.6 sentences. On the other hand, 600 disease-free articles from the same plateforms are used as contrary samples, such as news with title *Elon Musk declares plan to take Tesla private.* [3] with average length 35.8 sentences.

4 Model and Methods

The pipeline of the model is as follows: first, we adapt BERT pretraining model to the training data corpus, which gives us a classification model for sentences. Second, we label sentences in firms' earnings conference call transcripts as disease-related or disease-unrelated. Third, we perform sentiment analysis using pre-trained Natural Language Processing models on sentences labeled as disease-related. For both the number of disease-related sentences and each sentence's sentiment score, we normalized each of these two features for all firms and take the product of two features as the measure of firm-level epidemic-related risk. Fi-

[2] https://www.bloomberg.com/news/articles/2020-05-26/biotech-tourists-drive-short-lived-rallies-in-covid-19-stocks?srnd=coronavirus

[3] https://www.ft.com/content/73b700dc-9a2d-11e8-ab77-f854c65a4465

nally, we compare the risk measure for individual firms with the panel data of stock price volatility with simple OLS regression.

The key part of our model is the training of the sentence classifier. The advantages of BERT-based sentence classifier were threefold: (1) it allowed for a parallelization of tasks, (2) resulted in simpler operations, and (3) achieved better results overall. Their idea was to build a model based on attention mechanisms, which some of the CNNs and RNNs at that time used to connect their encoder and decoder.

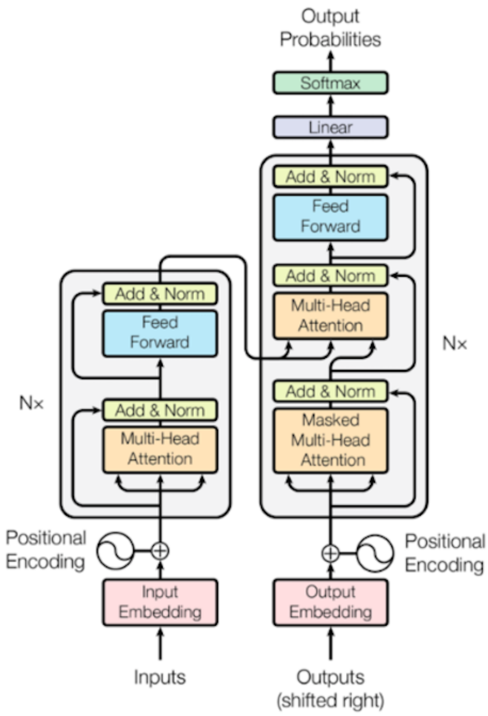

Figure 2: Transformers Architecture.

Adapted from Vaswani et al., 2017[15], Figure 2 displays a transformer. On the left side is the encoder, next to it on the right the decoder. As in preceding models, the encoder is responsible for forming continuous representations of an input sequence. The decoder in turn maps these representations back to output sequences. Both encoder and decoder consist of several layers (denoted with Nx in the graph) with two and three sub-layers each, respectively:

- **Multi-Head Attention:** here, keys, values and queries (which come from the self-attention - in this case, the previous layer's output) are linearly projected to then perform the attention function in parallel (which Vaswani et al. call the “scaled dot-product attention”). The multi-head characteristic makes it possible to use “different representation subspaces at different positions” .

- **Masked Multi-Head Attention** (only in decoder): as the first sub-layer in the decoder, this layer performs the multi-head attention on the encoder’s output. It is masked in order to prevent predictions based on information that must not be known yet at a certain position.

- **Feed Forward Network:** two linear transformations are applied to each position separately and identically.

Before information enters the layers, the positional encoding conveys information on the relative position of a token in a sequence and allows the transformer to make use of the token order. Once the layers have performed their attention function and transformations, another two transformations take place: the Linear, applying another linear transformation, and the Softmax, which transforms the output back to probabilities.

BERT (Bidirectional Encoder Representations from Transformers) was published in 2018 by Devlin et al. [7] from Google and performed so well that - within a year - it inspired a whole model-family to develop. BERT built on the original transformer idea, but used a slightly changed architecture, different training, and (as a result) increased size.

- **Architecture:** BERT’s architecture is largely based on the original transformer but is larger (even in its base version) with more layers, larger feed-forward networks and more attention heads.

- **Training:** The real innovation comes in the bidirectional training that BERT performs. There are two pre-training tasks: masked language modeling (MLM) and next sentence prediction (NSP). These are performed on the pre-training data of the BookCorpus (800 million words) and English Wikipedia (2500 million words).

After understanding the transformer-based models in theoretical context, we apply them to our classification problem: to determine if there is epidemic-disease-related uncertainty in conference call transcript with the aim of extending and improving the current classification methodologies. For this purpose, we use the package **Simple Transformers**, which was built upon the Transformers package (made by HuggingFace). Simple Transformers supports binary classification, multiclass classification and multilabel classification.

5 Model Implementation and Results

5.1 Data Precrocessing

The original dataset we used was labeled and consisted of four columns:

1. Date (e.g. May 1, 2020)
2. Title (e.g. *New Jersey Reports More New Deaths Than N.Y. for Second Day*)
3. Article content
4. Label - a binary indicator showing if the article is related to epidemic disease or not.

The entire dataset contains 800 articles and is imbalanced, with about 25% of the articles being marked (1), related to epidemic disease.All articles are written completely in English. The average length of the newspaper articles is 534 words (after cleaning).

Overall, after removing punctuations, stop words and fixing the (i'd, she'd, etc), the following top words appear in **disease-related** news is summarized in Figure 3. And in the control group, we can see the most frequent words from Figure 4.

For data cleaning, We applied the standard text cleaning procedures to the articles. All text was made lower case, punctuation removed, HTML links/email addresses removed. We also removed marks such as double quotes but kept the commas and periods.

The data imbalance in the original dataset was 75% labeled as "0" and 25% as "1". However, instead of generating synthetic samples to regenerate the data, we opted to rebalance the dataset. We balanced the training and validation set on almost 80-20 ratio of 0:1. This was to ensure that the models will have enough samples to learn. However, for the test set we maintained the original imbalance ratio which was 75:25.

Figure 3: Top words in disease-related news

Figure 4: Top words in disease-unrelated news

5.2 Model Parameters Tuning

At this stage we have obtained initial result for the classification model. BERT-based models come with a set of hyper-parameters which need to be adjusted properly for the task. For our binary classification approach, we replaced the hyper-parameters' default values for the most impactful ones:

- **Number of training epochs:** The epochs were increased from 1 to a number as high as 10. However, increasing the number of epochs also increased the running time of the model intensively. At the end, 4 epochs turned out to be the most optimal number, balancing between running time and giving good results.
- **Batch size:** The batch size was increased from 2 till 16, and batch size of 8 gave good results. With a batch size of 4, we observed a certain overfitting.
- **Maximum Sequence Level:** Although the model can handle a maximum of 512 tokens, we started off first with 100 tokens, and then opted for 200 tokens. The average length of the articles in the entire dataset was 534, which implies that setting this hyper-parameter to 512 could have been an ideal choice. However, owing to certain computational difficulties it was limited to 200.
- **Learning Rate:** 4e-5. Changing this learning rate made the model run either faster or slower, results were optimal for this chosen learning rate.

5.3 Classifier Performance

We compare the performance of our classifier to two non-transformer based models as well. We have used a small bidirectional neural network and a SVM classifier. We also ran other BERT-family models (RoBERTa, DistilBERT and ALBERT) with the same hyper-parameters for comparison and further improvement in the implementation of BERT in our model.

To gauge the performance of each model, we adapted the following metrics:

- **Area Under Curve (AUC)** is a measure that takes into account the true and false positives. It measures the area under the (ROC) curve, which plots the false positive rate (specificity) against the true positive rate (sensitivity). A perfect model would have an AUC of 1, a randomly assigning model would be at 0.5.

- **F1-Score (F1)** balances between precision (= True positives /(True positives + False positives)) and recall (= True positives /(True positives + False negatives)) and ranges from 0 to 1 - the higher the better.

- **Matthew's Correlation Coefficient (MCC)** uses all four measures - true positives, true negatives, false positives and false negatives - in its calculation. The score goes from -1 (everything is wrongly classified) to 1 (everything is correctly classified).

The following result table was calculated based on the standard binary cut-off of 0.5. If the probability (converted from the log-odds output of the model) is greater than 0.5, the article is labeled as related to epidemic disease and marked as "1", otherwise it is marked as "0". This cutoff of 0.5 was maintained across all models to provide a standard comparison.

Evaluation	BERT	RoBERTa	DistilBERT	ALBERT	Bidirectional NN	SVM
AUC	0.8354	0.8732	0.8523	0.8713	0.8034	0.8234
F1-Score	0.4122	0.4346	0.4256	0.4623	0.3743	0.3812
MCC	0.3945	0.4273	0.4023	0.4380	0.2919	0.4128
Runtimes	15min	15min	13min	11min	20min	7min

Figure 5: Evalutions of models

From Figure 5 we can see that the BERT-family models all perform better than the benchmark models. While BERT's, DistilBERT's and ALBERT's performance does not differ much from each other across all of our metrics (AUC of around 0.87), RoBERTa largely outperforms them. Their runtimes were similar around 18 minutes, with DistilBERT and ALBERT being slightly faster, and the SVM taking by far the least time with only 5 minutes runtime.

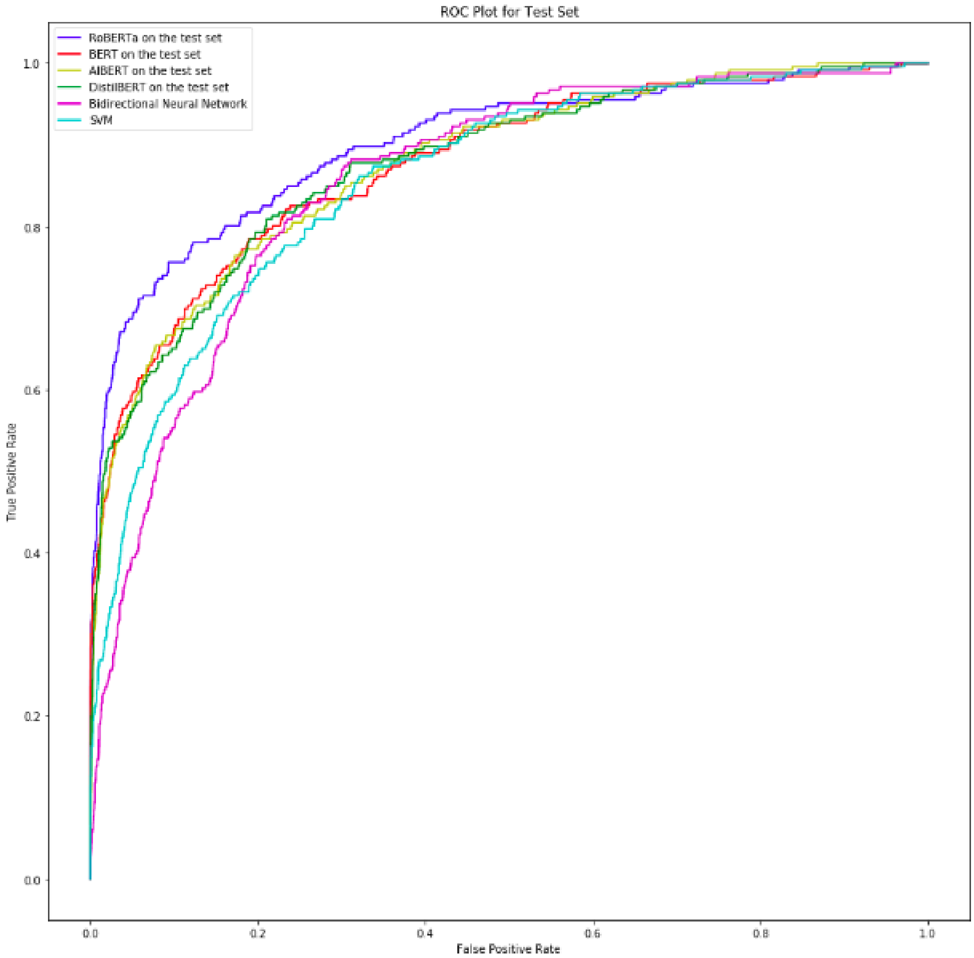

Figure 6: Comparison of ROC among models

5.4 Measure of Firm-Level Risk

We pre-cleaned the quarterly earnings reference call transcripts in the same manner as the news data, and applied the trained classifier **RoBERTa** model to parsed sentences of transcript text. For the measurement of exposure to epidemic-related risk, we design the follwing algorithm:

- Firm-level risk: We denote the total number of sentences in transcript of firm i at time t as $S_{i,j}$, and the number of sentences labeled as epidemic-disease-related as $D_{i,t}$, then we take the measure of related risk as

$$Risk_{i,t} = \frac{D_{i,t}}{S_{i,t}}$$

- Sector-level risk: We sum up the nominator and denominator across all listed firms in a certain sector J, and take the division as risk measure:

$$Risk_{J,t} = \frac{\sum_{i \in J} D_{i,t}}{\sum_{i \in J} S_{i,t}}$$

We selected the top 20 firms with highest Risk Index and sector-wise ranking as follows:

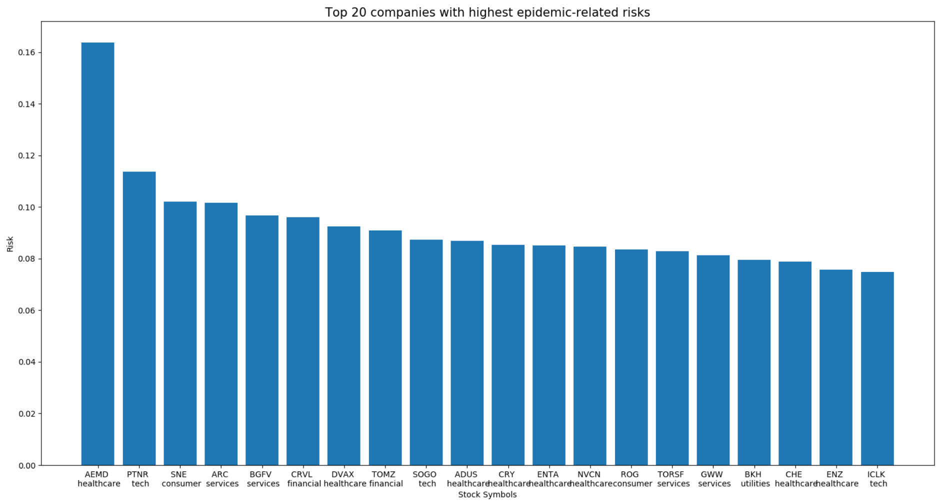

Figure 7: Top 20 firms with highest Risk Index

We can see that among the top 20 firms, companies in healthcare and technology sectors take the majority, which is well understandable, since their business are directly involved in testing, diagnosis and treatment of the epidemic disease. For example, Aethlon Medical, Inc., (NASDAQ:AEMD) is a medical technology company that focuses on addressing unmet needs in health and biodefense worldwide. Aethlon Medical Inc. is known for its Aethlon Hemopurifier, a clinical-stage immunotherapeutic device that removes exosomes and life-threatening viruses from the human circulatory system. Aethlon's Hemopurifier therapy works through a unique blood purification process. The device is a first-in-class therapeutic technology receiving two FDA Breakthrough designations, through multiple therapeutic targets: viral disease and cancer, which is closely related with the treatment of Covid-19. Another example is ARC Document Solutions, Inc.,(NASDAQ:ARC) a reprographics company, which provides document solutions worldwide. It offers managed print services, an onsite service that

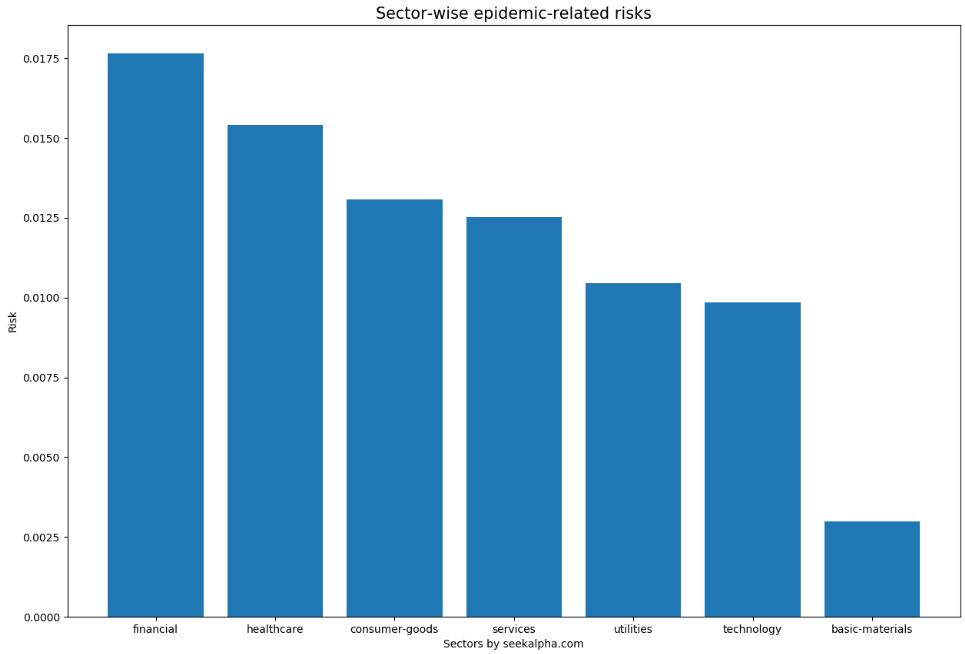

Figure 8: Sector-wise ranking in terms of Risk Index

places, manages, and optimizes print and imaging equipment in customers' offices, job sites, and other facilities; construction document and information management services. The lockdown and work-from-home policy make onsite services in professional locations difficult to deliver, and we also witness the stock price of ARC drop from \$1.34 on Mar, 04, 2020 to \$0.54 on Mar, 17, 2020, which shows great extent of exposure to risk imposed by covid-19.

From the perspective of sector-level risk measurement, we observed that financial sector is exposed to highest risk, which can be partly explained by the fact that average length of transcripts in this sector is shorter (257.89 sentences, compared to the total average 371.38 sentences), which makes the ratio of pandemic-related sentences higher. On the other hand, broken financial support along value chains and bad debt brought by economy halt is a possible cause. An interesting finding is that technology sector is not as much exposed as other sectors. Companies such as Amazon, Google, Facebook and Apple have become an essential part of making it easier to shelter in place or to track the virus — thanks a bunch for food delivery, contact-tracing apps. Stay-at-home policies may even be profitable for tech-companies: it found that people were participating in activities online more than before, including 18 percent who said they are doing more virtual consultations with their doctors and 14 percent who say they are having groceries delivered more frequently. [4] Even though every tech company has thrived as a result of the pandemic: Uber and Lyft have flailed as people stay home and avoid sharing close quarters with strangers in cars, while Airbnb has laid off a quarter of its workforce as travel has dried up, tech-companies are shown more immune to the pandemic impact thanks to the fact that most virtual digital services can be realized without interperson contact, and work-from-home is easier to implement with less loss of efficiency in this industry.

6 Conclusion

We adapted a context-based natural language pretraining model, BERT, to construct a measurement of firm-level risk on epidemic diseases, and focus on the impact of COVID-19 since Q4,2019 to Q1,2020. Based on the share of quarterly earnings reference call transcripts, we combine semantics and sentiment analysis to gauge risk faced by individual US firms, and their risk awareness. We focus on two main topics. First, the effectiveness of the BERT-based classifier in determining the relatedness of a sentence to pandemics. Second, the firm-level and sector-levl heterogeneity in epidemic-related risk. The result shows that the BERT-based NLP model outperforms traditional classifier in accuracy. We obtained firm-level and sector-level risk index by applying the clssifier on transcripts text, and found that companies in healthcare and consumer goods are most impacted. Sector-wise, financial industry is faced with higher risk, while techonology sector is more immune to the epidemic impact.

[4] https://thehill.com/policy/technology/496712-tech-firms-emerge-as-big-winners-in-new-covid-19-economy

References

[1] A comment on a laptop.

[2] Covid-19: Damage found in multiple organ systems.

[3] Covid-19 vs. sars: How do they differ?

[4] 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.

[5] 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.

[6] Scott Baker, Nicholas Bloom, and Steven Davis. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

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