The *CoRisk*-Index: Mining industry-specific risks related to COVID-19 in real time

— Detailed Documentation —

Fabian Stephany[1,2], Leonie Neuhäuser[3], Niklas Stoehr,
Philipp Darius[3], Ole Teutloff[1,3], Fabian Braesemann[1,4*]

[1]Oxford Internet Institute, University of Oxford, 1 St Giles, OX1 3JS, Oxford, United Kingdom
[2]Humboldt Institute for Internet and Society Berlin, Französische Straße 9, 10117 Berlin, Germany
[3]Hertie School Berlin, Friedrichstraße 180, 10117 Berlin, Germany
[4]Saïd Business School, University of Oxford, Park End Street, OX1 1HP, Oxford, UK

*To whom correspondence should be addressed;
E-mail: fabian.braesemann@sbs.ox.ac.uk

Code and data: `http://github.com/Braesemann/CoRisk/`

Interactive dashboard: `http://oxford.berlin/corisk`

June 18, 2020

Contents

Methodological Details

1 Data collection

1.1 Automated extraction of 10-K reports

The U. S. Securities and Exchange Commission (SEC) stores all reporting in a central repository[1]. Here, users can access meta-level information, such as index files, e.g., lists of all reports issued in the second quarter of 2020 directly. Alternatively, individual filings and meta-level information can be retrieved via various statistical packages, e.g., *"edgar"* (*R*) or *"sec-edgar-downloader"* (*python*), or a freemium API[2].

However, several limitations make the use of these ready-made devices impractical for the specific research purposes of this work. Downloading all relevant reports as .txt/.html files via the ready-made packages is possible in theory but requires a lot of time and hard disk space. In addition, the .txt file contains several unwanted css/html code patterns that make the identification of corona-sentences and the counting of words from them unreliable.

Alternatively, we initially refer to the *crawler.idx* index file[3] in the SEC repository. The index file holds a full list of all reports issued in a given quarter. This document shows the meta-level

[1] `https://www.sec.gov/Archives/edgar/full-index/`

[2] `https://sec-api.io/`

[3] `https://www.sec.gov/Archives/edgar/full-index/2020/QTR2/crawler.idx`

index.htm page for each company. On this page, the most recent 10-K report is linked[4]. Unfortunately, the htm-version of each 10-K report has a cryptic file name that can not be anticipated or guessed with the knowledge of company, date or industry parameters. Hence the currently implemented scraper (*python 2.7*) pipeline of our project, first, fetches the list of recently listed 10-K reports from *crawler.idx*, secondly, constructs the meta-level index.htm for each company and finds the link to the most recent 10-K report. The algorithm than scrapes the report text before, lastly, identifying the sentences related to corona (examples of text element that contain the term 'coronavirus' are displayed in Figure S1). All sentences and their respective report properties are stored for later processing.

1.2 Collection of stock market data

The collection of stock market data follows the incentive to provide reference data for the company SEC filings. We obtained stock market data through the *Yahoo! Finance API* [1] following two successive steps: First, we extracted the CIK (central index key) identifier of each company from the SEC filings to compile a comprehensive list of all companies, their CIK and stock ticker identifiers. Using the stock tickers, we then retrieved the historic closing stock values of each company per trading day between January 1st 2018 and May 15th 2020.

1.3 Collection of unemployment data

Due to the recent outbreak of the crisis, there are not many data points of monthly industry-specific unemployment rates available, which are provided by the Bureau of Labour Statistics. The crisis had first labour market repercussions only in March; until 5th June, there are only unemployment rates for March (published in April) and April (published in May) available. Therefore, we consider a different data set to approximate the labour market repercussions of the pandemic. We use weekly unemployment initial claims data provided by the Economic Policy Institute, as shown in Figure S2.[5] The data covers the number of weekly initial claims per sector in nine weeks from 14th March to 9th May 2020 for 19 US states. Additionally, the data contains information about the total employment per industry and state.

Using this dataset, we can calculate the share of weekly unemployment initial claims per total employment (the total employment is assumed to be constant in the observation period) in the 19 US states.

[4] E.g., `https://www.sec.gov/Archives/edgar/data/1084869/0001437749-20-002005-index.htm`

[5] Source: Economic Policy Institute (2020) "Weekly UI initial claims by state and industry" (Data collected from various sources), `https://economic.github.io/ui_state_detailed/`. We are grateful to Andrew Van Dam how has pointed us to this data source.

1.4 Specification of corona keywords

Since this study examines the attention attributed to COVID-19 in the SEC filings, the discovery mechanism of relevant COVID-19 mentions is of central importance. To mitigate susceptibility to errors due to word splitting, stemming and other text preprocessing, we decided for the most simple approach based on the matching of regular expressions. We scanned the reports for the two relatively unambiguous terms "corona" and "covid", also accounting for "coronavirus" and "covid-19" without duplication. For this process, the entire text is set to lower case.

2 Methodology of the CoRisk-Index

2.1 Filing of 10-K reports to SEC

Companies with more than 10 million USD in assets or a class of equity securities that is held by more than 2,000 owners must file annual 10-K reports to the SEC, regardless of whether the securities are publicly or privately traded[6]. All 10-K reports are made publicly available by the SEC. In particular, but not exclusively, in the risk section of the report, the company lays anything that could go wrong, likely external effects, possible future failures to meet obligations, and other risks disclosed to adequately warn investors and potential investors. Companies are required to use "plain English" in describing these risk factors, avoiding overly technical jargon that would be difficult for a layperson to follow.

2.2 Comparing 10-K and 10-Q reports

In addition to annual 10-K reports, the SEC requires companies to publicly disclose their actions in a set of reporting formats (10-X). Figure S3 summarises the context of all types of 10-X reports. Apart from historical reporting standards and niche reporting categories, company disclosures with the SEC can be separated in two groups: Annual 10-K and quarterly 10-Q reporting. One of the four quarterly 10-Q reports is subsumed by the annual 10-K report.

We limit our analysis to the information contained in 10-K reporting alone for the following reason. On March 4, the SEC has explicitly advised public companies to assess what the coronavirus means for their future operations and financial results and to make appropriate disclosures to their shareholders and other members of the investment community. Furthermore, the SEC encouraged companies to delay SEC filings if necessary to develop the information required to make accurate and complete disclosures of the impact of the coronavirus on its operations and financial conditions. Specifically, the SEC issued an order stating that public companies that are unable, because of the coronavirus, to meet filing deadlines for SEC reports due to be filed March 1 to April 30, 2020, will have 45 additional days to file these reports so long as, among other things,

[6] `https://www.sec.gov/fast-answers/answers-form10khtm.html`

they file reports on Form 8-K describing the reasons why the report may not be filed on a timely basis[7]. This announcement has caused and interference with the normal reporting procedures in at least two ways. Companies delayed their quarterly reporting until the last possible date (April 30, 2020) and companies were incentivised to "talk" about Covid-19 related issues.

From our perspective, this reporting bias should manifest most strongly in 10-Q reports, as they are more susceptible to short-term changes in reporting standards. Figure S4 illustrates the delay of quarterly reporting. 10-K reports, in contrast, require a longer preparation period as they summarise the general company outlook for the reporting year. 10-K reports that had been published in March and April are most likely not influenced by the SEC's change of reporting schedule and it can be argued that they did not have any incentive to explicitly mention issues related to Covid-19, but for the actual impact the pandemic had started to cause on their business. Figure S5 depicts the effect that an inclusion of 10-Q reports would have on our analysis. Clearly, the atypical delay of quarterly reporting and an accumulated issuing of reports at the end of April 2020 would artificially skew our index values.

In summary, given the aforementioned considerations, we decide to include solemnly 10-K reports in our CoRisk analysis, as they have a long-term outlook of one year and a lower susceptibility to ad-hoc changes in reporting standards compared to 10-Q reports.

2.3 Matching industry classification systems

The SEC classifies firms into industries using an amended version of the 1987 Standard Industrial Classification (SIC).[8] The SIC is a system for classifying industries by a four-digit code. It was replaced by the North American Industry Classification System (NAICS) in 1997. The NAICS system is still in use, in its last revision from 2017, and it is being used by US government agencies. Thus, it is important to match the SIC codes used by the SEC with the NAICS system, in order to allow researcher and economists to use the CoRisk data in industry-specific applications.

To match the different classification systems, we use merging tables provided by the US Census Bureau.[9] The code (R) and the merging tables are available on GitHub.[10] After having merged the four-digit SIC codes, we use the NAICS overview provided by the US Census Bureau to identify the larger two-digit sectors, on which the CoRisk data are aggregated.[11] During merging, some industries are assigned to several categories. In order to drop duplicates, we only keep the most frequent section per SIC code. While it might be possible that some information is lost during the merging process from the 1987 four-digits SIC system to 2017 two-digit NAICS sectors, we

[7] https://www.sec.gov/news/press-release/2020-53

[8] Information about the SIC standards used by the SEC can be found here: http://www.secinfo.com/\protect\T1\textdollar/SEC/NAICS.asp.

[9] https://www.census.gov/eos/www/naics/concordances/concordances.html.

[10] https://github.com/Braesemann/CoRisk

[11] https://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2017.

assume that less categories, which are compatible with the system used in current US statistics, provide a better overview of relevant processes than an aggregation, which is too fine-grained and which consists of low sample sizes within each category.

2.4 Elements of the CoRisk-Index

The CoRisk-Index contains two statistics, (a) the text negativity, and (b) the number of 'corona'-keywords per report. Combining both measures into one index has the advantage, that the index reflects both the sentiment with which the firms report about the pandemic and the frequency they report about it.

The measure of text negativity is calculated as the share of negative words per sentence that includes at least one 'corona'-keyword. To identify negative words in SEC reports, we rely on the methodology established by Loughran and McDonald (2011) [2]. The researchers have derived a sentiment dictionary of words explicitly applicable to the analysis to SEC filings. In addition to counting the number of negative words per sentence, we count the total number of words per sentence to calculate the share of negative words per sentence. To derive the overall text negativity measure per industry, we then calculate the average share of negative words per 'corona'-sentence per day and industry.

The number of 'corona'-keywords per report measures how often firms mention these keywords. It is a straightforward measure that is meant to assess how relevant the pandemic is for businesses. To calculate the measure, we simply count the number of 'corona'-keywords per report. While the share of firms that report about 'corona' has, in general, increased significantly during the first months of 2020, there are still substantial differences between firms and industries. Figure S6 displays these differences as a dotplot (note that we have added small random noise to reduce overplotting). There is substantial variation within industries. The average differences between industries are less pronounced, but still relevant: while 50 % of the firms in Finance, Transportation & Utilities, and Mining mention 'corona' or 'covid' five times or less, the value is twice as high in Wholesale & Retail. Similar to the text negativity, the measure is then aggregated as daily averages per industry.

From these two measures, we derive the CoRisk-Index as the daily geometric mean for all eight industries and one for the total of all eight industries. In order to smooth out differences in the daily reporting patterns, we calculate a 14-day moving average (aligned right) of the daily CoRisk-Index values. For the online visualisation, index values are multiplied by 100 to display whole numbers.

Additionally, we provide two other measures of industry-specific corona-related risk awareness in the report and on the online dashboard: the share of firms mentioning 'corona' at least once in their 10-K reports (the count of a simple indicator variable, taking the value one, if a report contains at least one 'corona'-keyword and zero otherwise), and the share of topic-specific keywords per industry (details about the topics can be found in Section S3).

3 Topic detection

We apply unsupervised methods for topic detection from Natural Language Processing to further analyse the reports with regard to topic related risks. Different sectors are facing different challenges, therefore companies are reporting about different corona-related risks. We aim to capture these risk topics via a keyword search on predefined topics. In order to explore possible topics, we used Latent Dirichlet Allocation (LDA) for unsupervised topic modelling, similar to Dyer et al. (2017) [3]. We only apply the topic model to corona-related paragraphs in the risk sections. We additionally examine the most frequent words and bi-grams in the documents. Using this exploratory analysis, we define a set of topics, which are specified by keywords. We then conduct a keyword search to count how much these terms are mentioned in the different industries in order to estimate the topic prevalence. The resulting topic heatmap (see Section 10) reports the share of keywords per topic per 1,000 words in corona-containing sentences for the different industries. The following sections describe the different steps of the keyword derivation and topic detection in detail.

3.1 Unsupervised topic modelling

We use unsupervised learning techniques to explore the space of topics that companies discuss when describing coronavirus-related risks. Latent Dirichlet Allocation (LDA) is a Bayesian computational linguistic technique that identifies the latent topics in a corpus of documents [4]. This statistical model falls into the category of generative probabilistic modelling: a generative process which defines a joint probability distribution over the observed random variable, i.e. the words of the documents, and the hidden random variables, i.e. the topic structure. In other words, LDA uses the probability of words that co-occur within documents to identify sets of topics and their associated words [3]. The number of topics has to be defined in advance. LDA is a frequently used technique to identify main topics in a corpus. Nevertheless, the interpretation of these topics can sometimes be difficult. We thus perform LDA for explorative purposes in our research only and apply the following steps:

Sample restriction Referring to Section 2, we filter all sentences from the risk sections that mention either "corona" and "covid", thereby also accounting for "coronavirus" and "covid-19".

Text preparation Before we train the LDA model we prepare the documents to achieve better performance of the method. We remove all common English stopwords, which are frequent words such as “is,” “the,” and “and” as well as those words which appear in at least 80% of the documents. These words are not useful in classifying topics as they are too frequent and therefore decrease performance. Moreover, we delete all words that do not occur in at least two documents.

LDA We turn the documents into numerical “Bag of words” feature vectors, disregarding word order. We then use LDA to extract the topic structure. Like any unsupervised topic model, this requires setting the number of topics a priori. We selected this key parameter based on semantic coherence, evaluating a range of two to eight topics leading to a final model of four topics. The top ten terms of each topic are displayed in Table S1.

3.2 Dictionary-based topic search

The algorithmically derived topics give a good insight into the general narratives of risks used in the documents. Nevertheless, they are hard to interpret, as early corona-related risk reports are still generic in that various risk factors are covered. Topic four, for example, provides an unspecified context with regard to the outbreak of the illness in China. Similarly, topic three covers the potential impact of the crisis in an unspecified context. In contrast, most of the business and economics related keywords appear to be covered in topic two. Moreover, the unsupervised methods are not deterministic. To ensure a robust and comparable topic identification over time, we use a dictionary-based keyword search. For this, we combine the results of the unsupervised methods with domain knowledge from economics to label the five main topics and specify defining keywords, displayed in Table S2. The most frequently mentioned bi-grams and words are considered. Using these, we can conduct dictionary-based searches in the filings. We measure the topical context of each corona-sentence by calculating the share of topic keywords relative to the word length of the sentence. This metric is later aggregated for industries and specific points in time.

3.3 Methodological limitations

Our approach is based on the risk assessments in the U. S. Securities and Exchange Commission (SEC) filing reports. Thus, the value of the approach relies crucially on company self-reporting. While the firms are unlikely to provide a risk prediction with a high forecasting accuracy, it might still be worth exploring the reports as alternative data source to measure risk perceptions. As the reports serve as legal and insurance requirements against financial risks, but also as a basis for investment decisions of investors, companies are implicitly encouraged to neither over- nor understate the risks they are facing. Nevertheless, our results are limited to this self-assessment. As many of the implications of the corona crisis are still uncertain, our approach thus reflects a way to approximate potential implications on current estimations of experts in the different sectors, represented by the companies, and does not include risks that are unforeseeable for themselves at a given point in time. Moreover, the data to assess the precision of these estimations does not exist yet, as there has not been a pandemic with comparable global economic consequences in recent time. In the future, we will evaluate our results by looking at employment data in different sectors. Moreover, the pandemic continues to spread and will soon affect all industry sectors and all countries of the world. As more and more companies will report on related risks, the count of corona mentions alone will lose its information-value as a measure to differentiate between endangered industry sectors. Then, more granular measures, such as the identified topic categories, will become more important to distinguish between different natures of risks.

The exploration of alternative data sources that are meant to complement or now-cast established economic statistics always comes with uncertainty. For example, it is not yet clear how the short-term risks described by the different industries will translate into long-term economic outcomes, such as bankruptcies. Nonetheless, we believe that the reports could be a reliable source of empirical information about the issues faced by different industries in the current situation, and they might be used to inform forecasting models on industry-specific economic effects of the crisis, as they help fill a data gap. Models that incorporate alternative data sources such as the one presented here could then be beneficial for developing economic support packages that are currently provided by governments.

All technical methods serve the higher purpose of providing timely and comprehensible insights into the industry-specific effects caused by the global outbreak of the coronavirus. To mitigate susceptibility to errors and increase reproducibility, we mostly draw from more basic technical methods. This can be seen in the discovery of corona-relevant keywords which is based on the matching of regular expressions to avoid error-prone text pre-processing. Reduction of technical complexity, however, comes at the cost of diminished modelling fidelity and potential accuracy of results. For instance, the LDA-based topic modelling approach lacks in interpretability and robustness, and we therefore only use it for exploration until now. The topics and keywords we use for estimating topic prevalence are therefore hand-coded which limits the detection of topics to predefined terms.

Supplementary Text

4 Related work

In this section, we review related work on the economic consequences of COVID-19 and studies on previous epidemics. Moreover, we discuss the assessment of economic risks via reports such as the 10-K reports required by the U. S. Securities and Exchange Commission (SEC).

4.1 Studies on the economic effects of COVID-19

Global pandemics of infectious diseases are not a new phenomenon. Throughout the last century, the world experienced several global and regional outbreaks: Most severely, millions died during the spread of the Spanish flu in 1918-1920. More recently, smaller epidemics spread around many countries, such as SARS in 2002, the swine flu in 2009 and Ebola in 2014. Nonetheless, the current coronavirus pandemic is, in recent history, unprecedented in its global social and economic consequences. Governments are taking drastic measures while researchers attempt to provide urgently needed policy advice. Much of the coronavirus-related social science and public health research focuses on disease transmission, global spread, and different interventions (see for example [5–9]). A rapidly growing body of literature investigates the (potential) economic consequences induced by the COVID-19 pandemic [10–15].[12]

Many of these studies aim to asses the potential economic consequences in presenting simulation-based macroeconomic models. For example, Dorn et al. (2020) use scenario calculations to estimate the economic costs of the pandemic for the case of Germany [15]. The authors estimate the financial consequences for the state budget and the employment effects depending on the length of the economic shutdown. Moreover, they include differential adverse effects by sectors, based on press releases and the provisional Ifo business climate index for March 2020. They conclude that the travel and restaurant industry is likely to face a complete shutdown, whereas the pharmaceutical, logistics and health sectors are likely to continue to operate at full capacity. Ludvigson et al. (2020) aim to estimate the macroeconomic consequences of the pandemic in investigating the impact of disasters in the recent U. S. history [12]. Baldwin and Mauro (2020) collect a great variety of perspectives on the potential economic implications of COVID-19 [14, 16]. Topics range from impacts on trade, economic policy measures, monetary policy, and finance to labour market effects. These contributions rely on simulations, scenarios, descriptive statistics and qualitative arguments.

Regarding the economic impact by sector, Gopinath (2020) identifies manufacturing and the services sector as disproportionately affected in China, based on the Purchasing Managers' Index [17]. Ramelli and Wagner (2020) use Google search intensity to measure attention paid to COVID-19 and stock market data to reveal the economic impact by sector [18]. Their analysis shows that the energy, retail, and transportation sector experienced the largest losses in the United States and China, whereas health care gained considerably in both countries. The analysis by Huang et

[12] At this point we want to highlight that it is not the aim of this study to provide a model of risk forecasting that is competing with established macro-economic approaches as described in this section. In contrast, the data-driven methodology presented here aims to explore an alternative data source that could help to inform such economic models. The advantage of the data we provide here is the high time resolution. More traditional sources of empirical information used to calibrate macro-economic models usually include, for example, unemployment rates. While the value of such statistics is undisputed, they are reported with a time-lag. We perceive the purpose of the index provided here to be a complementary data source that could be compared with official statistics on the economic effects of the crisis over time.

al. (2020) (also based on stock market data) confirms that the services sectors seem to be the most severely affected in China [19], and del Rio-Chanona et al. (2020) provide quantitative predictions of first-order supply and demand shocks to the U. S. economy on the level of individual industries [10].

There are first works that use data mining approaches similar to our proposed methodology. Hassan et al. (2020) use their previously introduced text-based measure for firm-level political risk [20] in the light of the COVID-19 pandemic [21]. Nevertheless, while the authors also measure risks related to Covid-19 with metrics similar to the CoRisk-Index, the researchers clearly take a different perspective on the scenario. They regard Covid-19 and the risk associated to it from a business perspective: Were businesses prepared and if so, did they manage to avoid return losses? Our work, on the other hand takes a (macro) economic perspective on the pandemic and its risks: How do risk assessments differ between industries? Is our index a forward-looking economic indicator? Does our index relate to employment developments? In contrast to Hassan et al. (2020), we do not provide extensive inferential findings (regression analyses) but supply the public and policy makers with an online index that allows in depth topical investigations in real-time.

In sum, the contributions presented in this section provide valuable insights, quantitative scenario calculations, and timely policy recommendations. Most of the macroeconomic analyses are, however, based on assumptions-driven simulations and models, as up-to-date empirical data to assess the immediate economic consequences or industry-specific risks are lacking. These studies could benefit from the alternative data set we are exploring in this study.

4.2 Historical pandemics

While research on recent epidemics are limited to simulations [22, 23], or specific sectors [24], the study of historical pandemics might provide informative empirical assessments of pandemic-related economic effects. Studies on the 1918 Spanish Flu confirm the primordial effectiveness of non-pharmaceutical interventions, e.g., social distancing, even if these come at the cost of economic slowdowns [25, 26]. Moreover, based on the historical data, researchers find a correlation between mortality rates and declines in GDP, consumption and returns on stocks [27] and an increase in poverty [28]. However due to the global scope of the crisis for the highly inter-connected world economy of 2020, insights derived from past epidemics are of limited use in identifying the various industry-related risks during the COVID-19 pandemic.

Both, the research on COVID-19 and the historical pandemics rely on stock market information to quantify the economic effects of infectious diseases. However, stock market information comes with several drawbacks. Most importantly, stock markets are prone to irrational herd behaviour and prices capture a variety of information signals into one aggregated index. Examining current stock market dynamics reveals a general economic downturn, but it does not allow to isolate the sector-specific COVID-19 risks. Therefore, we propose to use SEC reports which include risk statements. We argue that these reports represent a promising real-time measure of industry-specific business risks. Furthermore, the analysis of report statements discussing 'coronavirus' allows to isolate the business risks exclusively associated with the COVID-19 outbreak.

4.3 Assessing economic risks via business reports

Since the great recession hit the world economy in 2008, risk has been a crucial topic in governance and finance. While risk assessments of the financial system led to diverse measures to make the world economy less vulnerable to economic shocks originating in the financial sector, a health crisis, such as the current pandemic, poses different risks to the economy. While government measures against the spread of the disease hinder the population from working and consuming, which results in businesses interrupting production, many economies face demand and supply shocks at the same time. In particular, as different industries rely on distinct input factor compositions and supply chains, the sectors of the economy react differently to shocks [29]. Regarding the COVID-19 crisis, it might be expected that sectors whose operations are connected to global supply chains could publicly report corona-related risk earlier than others. These sectors are also highly connected and interdependent within the national economy and risk might spread between sectors [30, 31].

Most risk assessment approaches focus on quantitative probability-based methods and financial data [32, 33]. The data published in such quantified risk assessments are often made available retrospectively, which makes a real-time evaluation of risks and economic outlooks difficult. In contrast to such assessments, we investigate the annual 10-K reports filed to the U. S. Securities and Exchange Commission (SEC), which provide verbal corporate risk disclosure and financial statements. Besides, SEC filings are imperative for legal and insurance requirements and they need to contain the most relevant risks to protect the company from legal liabilities. Furthermore, the reports inform investment decisions and risk governance at the same time and, thus, companies, as rational agents, are likely to communicate moderate risk assessments [34]. In fact, prior work has underlined the forward-looking nature of the reports, since they allowed a more effective prediction of volatility on stock market returns than the compared approaches [35]. Correspondingly, we expect the 10-K reports to also provide forward-looking information on risk assessments during the observed time period and in particular on rising business risk factors in relation to the spread of COVID-19.

5 Correlation with stock market data

The value of the CoRisk-Index is on providing a fine-grained overview of the industry- and topic-specific risks and repercussions of the corona crisis, not in predicting fluctuations at the financial markets. Nonetheless, it can be useful to investigate the relation between the corona-related sentiment of the SEC filings and the reactions of the stock markets to the spreading pandemic. As it is shown in the main report, the overall text negativity, as measured by the share of negative words in the corona-sentences seems to be correlated with the S&P Global 1200 index, a measure of global stock markets. The overall text negativity in corona-sentences of 10-K reports has already risen prior to the stock market crash on 20th February. This seems to hold also with regards to the stocks of the individual firms reporting to the SEC.

In Figure S7 we show the average stock market development[13] of firms reporting to the SEC (upper panel 'Diff') that are in our dataset and the share of negative words in 10-K reports per industry (lower panel 'negativity'). In all eight industries, there is a significant correlation between the text negativity and the average stock market development. In some industries, such as Finance and Manufacturing, which are both sectors that file many reports to the SEC, the text negativity preempts stock market developments by weeks. After an immediate steep rise in negativity and a drop in stock markets, the recovery of the stock markets is accompanied by a decreasing text negativity.

In summary, the figure shows that the corona-related text negativity extracted from 10-K reports is correlated with short-term reaction at the financial markets, which are foreshadowing the deep economic crisis.

6 Representativeness

The aim of the CoRisk-Index is to provide a real-time approximation of economic risks in different industries. It is important that the index reflects a large part of the respective industry and that it is not biased towards a very unpresentative sample of firms. As discussed in Section S2, firms that have more than 10 mio. USD in assets or more than 2,000 owners need to report to the SEC. As a consequence, it will be most likely larger firms that are well represented in the data. On the other hand, the number of employees per firms follows a fat-tailed distribution. Thus, the firms reporting to SEC will represent a large share of the overall US economy in terms of employees. As the average firm size differs between industries, the index might be biased to represent those industries better that are home to many large firms.

To reflect the differences in coverage between industries and to avoid excessive extrapolation to those industries that are not well represented, we use the COMPUSTAT database (made available via Wharton Research Data Services[14].) to obtain employee count data from the reporting firms. Through that database, we could obtain the employee count of 4700 of 6400 reporting firms in 2020. These 4700 firms have a total of 44 million employees in 13 sectors of the US economy.

[13] Calculated as the average value of all stocks per industry per day in USD.

[14] https://wrds-web.wharton.upenn.edu/wrds/support/Data/

The total economy has 150 million employees in these 13 sectors.[15] Thus, the reporting firms represent around one third of all employees in the US economy. However, not all sectors are well represented (see Figure S8). in eight sectors, the firms filing to the SEC, for which we could obtain data, represent at least 22 % of all US employees in that sector. Therefore, we limit the CoRisk-Index to these eight industries.

The CoRisk-Index reports corona-specific risks on a daily basis. This time series nature makes the index prone to potential biases, if the industries tend to have vastly changing reporting schedules during the calendar year. We investigate the seasonal reporting pattern in Figure S9. The figure shows the share of reports filed to the SEC per quarter in the period 2014 to 2018. A clear seasonal pattern can be identified in the first quarter of each year. However, this seasonality appears to be driven by Finance. This sector reports substantially more in the first quarter. Finance has reported relatively low corona-related risks in the first quarter, despite the high number of reports. Thus, the CoRisk-Index does not seem to be biased towards those industries that file more reports in a given quarter. The other sectors do not show substantial seasonal patterns. Thus, we conclude that the results presented in the report are not substantially biased due to seasonal reporting patterns.

7 Correlation with unemployment data

To validate the economic meaningfulness of the CoRisk-Index, we provide a comparison with short-term unemployment data in the report. Unfortunately, the comparison with macroeconomic data are limited in the early phase of the crisis, as not much data are available and as backward looking economic indicators, such as the number of bankruptcies, usually react with a time-lag to economic crises.

Nonetheless, we aim to provide such a comparison in using weekly unemployment initial claim data from 19 US states (as described in section S1.3).

In order to compare the CoRisk-Index data with unemployment initial claims, we first have to de-trend the data. Both the CoRisk-Index and the total share of unemployment initial claims are driven by a general upwards trends throughout the past few months (Figure S10). Comparing these variables reveals spurious correlations (Figure S11). In order to investigate the relations between the CoRisk-Index and unemployment data that are less overshadowed by the overall time trend, the variables need to be de-trended. To do so we calculate (a) the number of weekly unemployment initial claims as share of the overall employment per industry (this is held constant throughout the the observation period) and (b) the week-to-week changes in the CoRisk-Index (CoRisk in week $t-1$ divided by CoRisk in week $t-2$). Both the variable *CoRisk-Diff* and its one week lag *CoRisk-Diff-lag* are correlated with the weekly unemployment initial claims, but the

[15] `https://www.statista.com/statistics/200143/employment-in-selected-us-industries/`.

lagged *CoRisk-Diff* variable exhibits stronger correlation (Figure 3 of the report). Despite the low number of observations per industry, the R^2 is high in many industries, potentially indicating a strong and forward-looking correlation between the CoRisk-Index and relevant macroeconomic variables. In weeks of substantial growth of the CoRisk-Index, the unemployment initial claims in many industries show high growth afterwards, and both variables did not grow much in more recent weeks.

While this comparison is only a first step into the the investigation of relationships between the corona-related information captured in 10-K reports and macroeconomic data, the results provide promising first insights into that relation. We would expect that reduced unemployment rates (when we will be able to obtain more data points of monthly unemployment rates from the Bureau of Labour Statistics) would coincide with lowering CoRisk values in the future. Over the course of the crisis, when more unemployment and other macroeconomic data become available, we will provide interactive comparisons between the CoRisk-Index and such data on the online dashboard.

8 Historical robustness

The findings provided in the report rely on the assumption that the firm reports filed to the SEC actually contain information that is reflecting actual economic circumstances, and not just artefacts of specific wording, maybe due to trends, herding or currently relevant 'hot' topics. In order to investigate this, we have examined historical SEC filings data [2][16] and calculated the share of negative words (one of the two key components of the CoRisk-Index) for these filings in two ways.

First, we calculated the share of negative words per report in all 152,694 reports that have been filed between 2000 and 2018. To compare this historical text negativity with macroeconomic data, we aggregated the per-report negativity score for all reports filed in one quarter and compared this to US quarterly unemployment rates and quarter-to-quarter changes in GDP.[17] Figure S12 shows the results. During the two recessions (grey bars) that happened in the period from 2000 to 2018, the GDP (upper panel) dropped significantly and the unemployment rate increased (central panel).[18] Correlated with these overall macro-economic developments, the share of negative words in 10-K reports (lower panel) increased in these periods. In periods of economic recovery (lowering unemployment rates in 2004 to 2008 and in 2010 to 2012), the overall text negativity decreased.

This observation provides supportive evidence that textual data extracted from SEC filings contains information that is correlated with the changing real-world economic circumstances. How-

[16] `https://sraf.nd.edu/crsp-flat-file/`

[17] The macroeconomic data has been collected from: `https://research.stlouisfed.org/econ/mccracken/fred-databases/`.

[18] Recession periods from `https://www.nber.org/cycles.html`.

ever, the overall text negativity is a highly aggregated measure, not comparable to the fine-grained data on specific topics that covered by the text analysis of sentences mentioning specific keywords, such as Covid-19. Moreover, during the course of the past 20 years, the length of 10-K reports has increased substantial, which might influence text mining measures on the report level.

Therefore, we provide a second historical robustness check. Figure S13 displays the share of negative words in 'china'-sentences (that is all sentences in 10-K reports that mention the keyword-token 'china') in 2018, the time of the 'US-China Trade War'. During that period, US President Donald Trump announced a number of tariffs particular targeted at Chinese exports to the United States. Similar policies were introduced by the Chinese government. These governmental interventions and protectionist policies could potentially have harmed globally densely connected US companies. One could assume that firms might have mentioned risks related to the US-China trade war in their SEC filings. To investigate this assumption, we replicated the text negativity analysis for all 'china'-sentences in the SEC reports filed in 2018 and plotted it against some relevant events during the trade war.[19] Protectionist events, i. e. the introduction of tariffs by the US government, are highlighted as red bars, while events that might indicate a release of the political tensions, i. .e. trade talks or the consultation of the WTO, are highlighted as green bars in the figure. The solid line shows the 14-day moving average of the text negativity and the dotted line shows a trend.

In all cases highlighted in the figure, the real-world events were reflected shortly afterwards in the sentiment of the 'china'-sentences in the 10-K reports. Whenever new tariffs were announced by the US administration, the share of negative words increased in subsequently filed SEC reports. After positive events, the negativity dropped.

These findings exemplify that 10-K reports contain economically relevant information, which is reflecting overall economic circumstances, in particular with regards to specific events. We conclude that the rising frequency of 'corona' mentions in the 2020 10-K reports and the differences between industries are not statistical artefacts, but instead represent actual time- and industry-specific economic risk perceptions with regards to the corona crisis.

9 Future Work

The CoRisk-Index is an ongoing research project. Its value does not come from the presentation of findings in the report alone, but from the continued tracking of industry-specific risks during the whole course of the corona crisis. Consequently, the CoRisk-Index will be available online for the months and, eventually, years to come. In that period, the index provides a microscopic view into the day-to-day dynamics of the economic crisis.

[19] https://en.wikipedia.org/wiki/China%E2%80%93United_States_trade_war.

In order to maintain and increase the relevance of the real-time data we provide on the dashboard, we plan a number of extensions of the index.

Adjusting the keywords As the crisis unfolds, it transforms in reach and scope. During that process, industries are affected differenently, which is captured by the CoRisk-Index. To keep the topic structure, which is represented by the index, relevant, we will include novel keyword categories to capture the potentially different implications of individual aspects of the pandemic. For example, we plan to create one category dealing with employee welfare and one with digital solutions that might become more relevant for specific industries to deal with the novel regulations to reduce the spread of the virus.

Extension to other economies The CoRisk-Index is currently limited in scope to the United States, as it appears to be the only country that provides company risk reports in a freely available and accessible format. We have looked for similar reports in other countries (UK, Germany, Austria, Finland), but have so far not been able to obtain reports that allow for a structured provision of real-time information as we have presented it for the case of the U. S. If we were not able to collect business risk reports from other countries, we would propose an alternative approach for the extending the risk assessment internationally. Its results will be included in the dashboard later in 2020. We will collect news data about the companies reporting to the SEC, using a News API (Eventregistry). With that additional data set we aim to train a statistical model that represents the firm-specific CoRisk measures in terms of news article data. If it proves possible to establish a robust relationship between news mentions (for example, the sentiment of news articles about companies that are concerned with the corona crisis) and corona-specific risk measures from the SEC filings, we would use this model to extrapolate the risk index to other countries. In any case, the connection with news articles (as an additional source of non-traditional data about businesses) will allow to provide more firm-specific insights into the industry-specific repercussions of the corona crisis.

An outlook to repercussions on digitalisation The adjustments of the work routine made by many due to the stay-at-home policies will most likely have a long-lasting effect on the penetration of digital technologies in the organisation of work. For example, video calls have largely replaced face-to-face meetings. This has not only resulted in sharply rising stock market prices of video communication companies, such as Zoom Video Communications, but it has already sparked a spatial reorganisation of work. Over the course of the crisis, which might involves multiple periods of release and restriction [7], such digital technologies are likely to become more relevant for firms affected by the repercussions of the Covid-19 pandemic. We plan to add a section to the dashboard that discusses some of the potential effects of the crisis for technology adoption. In particular, we plan to include new topic-specific keywords on digital technologies and remote working (see above), and to track the potentially changed spatial contribution patterns on platforms of the digital knowledge economy, such as Stack Overflow, GitHub, or Wikipedia.

Other Supplementary Material

10 Online dashboard

All raw data and findings are published on an online dashboard. It is being made available on `http://oxford.berlin/corisk`. Figure S14 shows a screenshot of the main panel of the dashboard. It provides an interactive visualisation of the main industry-specific statistics described in the report: (a) the CoRisk-Index, (b) the share of negative words per corona sentence in the panel 'Text Sentiment', (c) the share of reporting firms per industry in the panel 'Industry View', and (d) the share of topic-speicifc keywords per corona-sentence in the panel 'Topic Heatmap'. Additionally, we have established an easy-to-use filter and download option (Right panel: 'Customise Date Range', 'Select Industry', and 'Download CoRisk Data'). It allows researchers, journalists, and the general public to get full access to both the daily aggregated index data and the raw data on the level of individual reports.

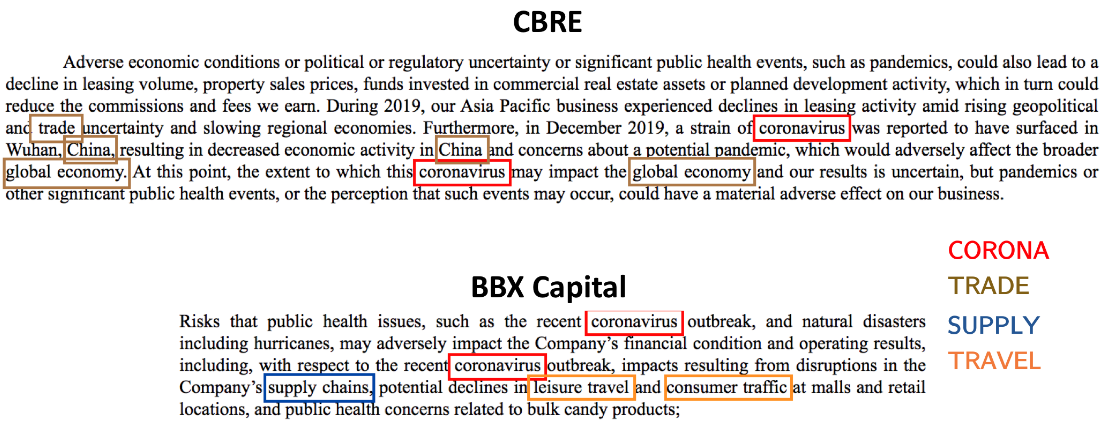
11 Software code and raw data

In addition to the online dashboard, which allows researchers and the public interested in the CoRisk-Index to gain detailed insights and to download the aggregated CoRisk data, we moreover publish all raw data and the software code that is being used to generate the main findings described in the report. Code and data will ease reproducibility and adoption to other use cases. The software code is available on a GitHub repository (`http://github.com/Braesemann/CoRisk`) The repository is also linked to on the website.

Figure S15 shows the folder structure of the GitHub repository. The File `Paper_Figures.R` contains all the code to reproduce the main figures of the paper (with `Paper_Figures_Data.zip`). To reproduce the CoRisk-Index, the files in the folder `CoRisk-Index-Code` may be used, together with the data in `CoRisk-Index-Data`. This folder also contains the raw data on the report level: `Raw data (report level).zip`, which can be used for more fine-grained industry- and company-specific analyses of corona-related business expectations not covered in this study.

In order to map the 1987 SIC codes used by the SEC with the current 2017 NAICS system, one can use `SIC to NAICS.R` and the merging tables from the folder `SICtoNAICS`. Code and data to reproduce figures from the supplement are available in the folders `Supplement-Code` and `Supplement-Data`.

Figures



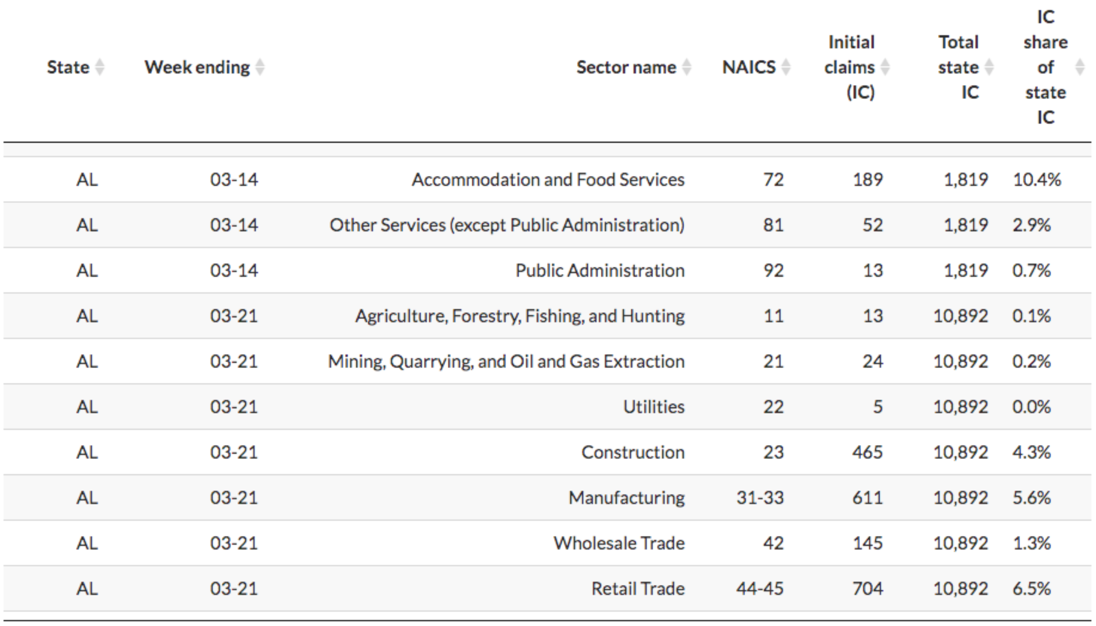
CBRE

Adverse economic conditions or political or regulatory uncertainty or significant public health events, such as pandemics, could also lead to a decline in leasing volume, property sales prices, funds invested in commercial real estate assets or planned development activity, which in turn could reduce the commissions and fees we earn. During 2019, our Asia Pacific business experienced declines in leasing activity amid rising geopolitical and trade uncertainty and slowing regional economies. Furthermore, in December 2019, a strain of coronavirus was reported to have surfaced in Wuhan, China, resulting in decreased economic activity in China and concerns about a potential pandemic, which would adversely affect the broader global economy. At this point, the extent to which this coronavirus may impact the global economy and our results is uncertain, but pandemics or other significant public health events, or the perception that such events may occur, could have a material adverse effect on our business.

BBX Capital

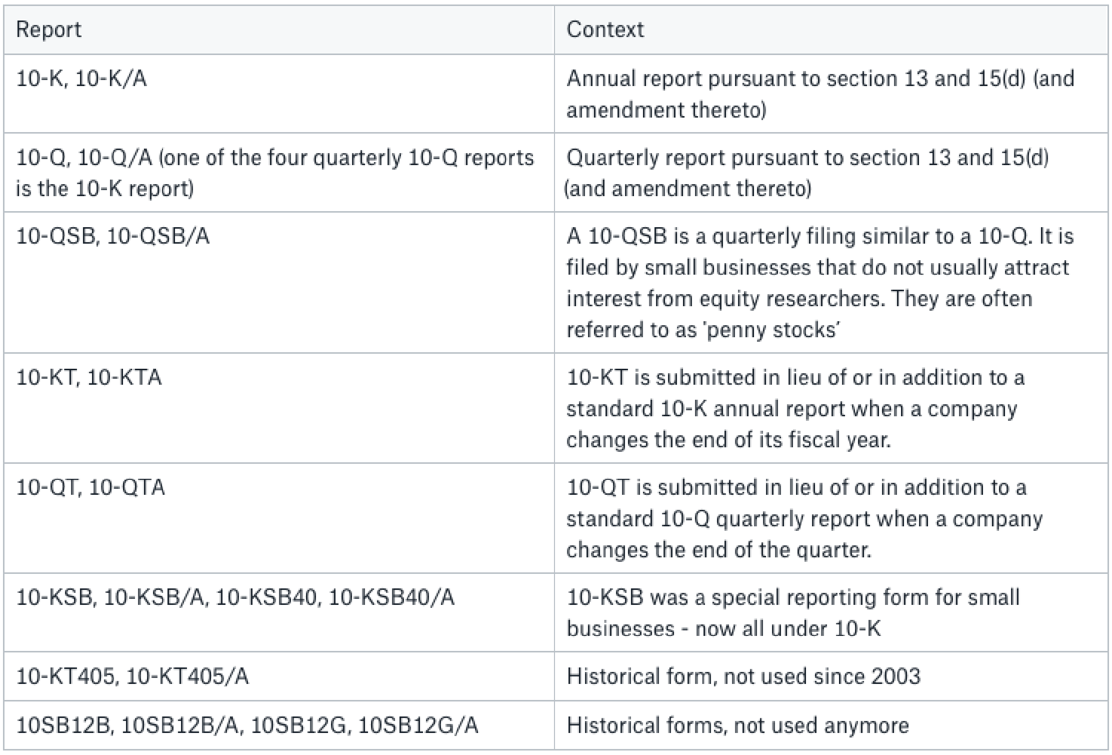
Risks that public health issues, such as the recent coronavirus outbreak, and natural disasters including hurricanes, may adversely impact the Company's financial condition and operating results, including, with respect to the recent coronavirus outbreak, impacts resulting from disruptions in the Company's supply chains, potential declines in leisure travel and consumer traffic at malls and retail locations, and public health concerns related to bulk candy products;

Fig. S 1 Examples of paragraphs in 10-K reports that mention the term 'coronavirus' and other business specific terms.



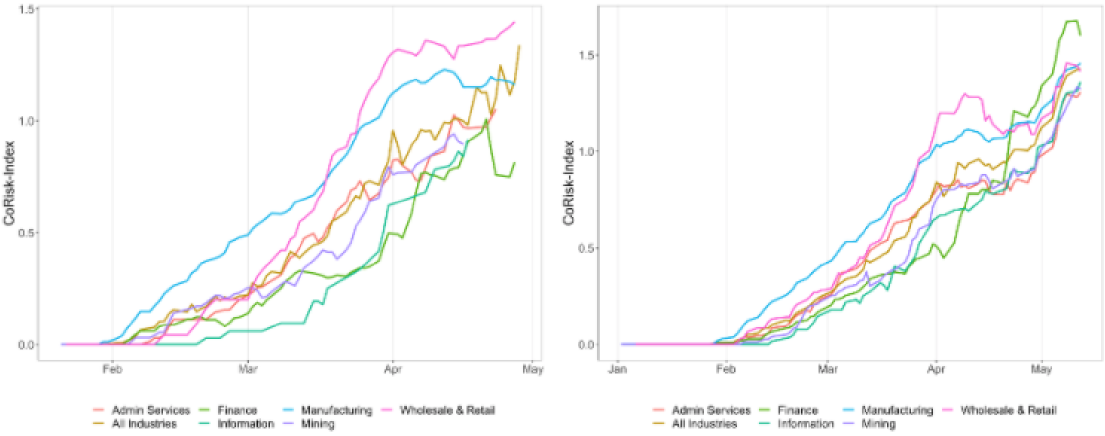
State	Week ending	Sector name	NAICS	Initial claims (IC)	Total state IC	IC share of state IC
AL	03-14	Accommodation and Food Services	72	189	1,819	10.4%
AL	03-14	Other Services (except Public Administration)	81	52	1,819	2.9%
AL	03-14	Public Administration	92	13	1,819	0.7%
AL	03-21	Agriculture, Forestry, Fishing, and Hunting	11	13	10,892	0.1%
AL	03-21	Mining, Quarrying, and Oil and Gas Extraction	21	24	10,892	0.2%
AL	03-21	Utilities	22	5	10,892	0.0%
AL	03-21	Construction	23	465	10,892	4.3%
AL	03-21	Manufacturing	31-33	611	10,892	5.6%
AL	03-21	Wholesale Trade	42	145	10,892	1.3%
AL	03-21	Retail Trade	44-45	704	10,892	6.5%

Fig. S 2 The structure of the weekly unemployment initial claims data provided by the Economic Policy Institute. The table contains unemployment initial claims data from 19 US states per week and sector (including NAICS code).



Report	Context
10-K, 10-K/A	Annual report pursuant to section 13 and 15(d) (and amendment thereto)
10-Q, 10-Q/A (one of the four quarterly 10-Q reports is the 10-K report)	Quarterly report pursuant to section 13 and 15(d) (and amendment thereto)
10-QSB, 10-QSB/A	A 10-QSB is a quarterly filing similar to a 10-Q. It is filed by small businesses that do not usually attract interest from equity researchers. They are often referred to as 'penny stocks'
10-KT, 10-KTA	10-KT is submitted in lieu of or in addition to a standard 10-K annual report when a company changes the end of its fiscal year.
10-QT, 10-QTA	10-QT is submitted in lieu of or in addition to a standard 10-Q quarterly report when a company changes the end of the quarter.
10-KSB, 10-KSB/A, 10-KSB40, 10-KSB40/A	10-KSB was a special reporting form for small businesses - now all under 10-K
10-KT405, 10-KT405/A	Historical form, not used since 2003
10SB12B, 10SB12B/A, 10SB12G, 10SB12G/A	Historical forms, not used anymore

Fig. S 3 Overview of SEC report types and the context in which they are used.



CoRisk Index using only 10-K reporting

CoRisk Index using only 10-K AND 10-Q reporting

Fig. S 4 The CoRisk-Index based on 10-K reports (left panel) and on 10-K and 10-Q reports (right panel). The indices show some similarity, however the index on the right is biased by the very high number of 10-Q reports being filed in May.

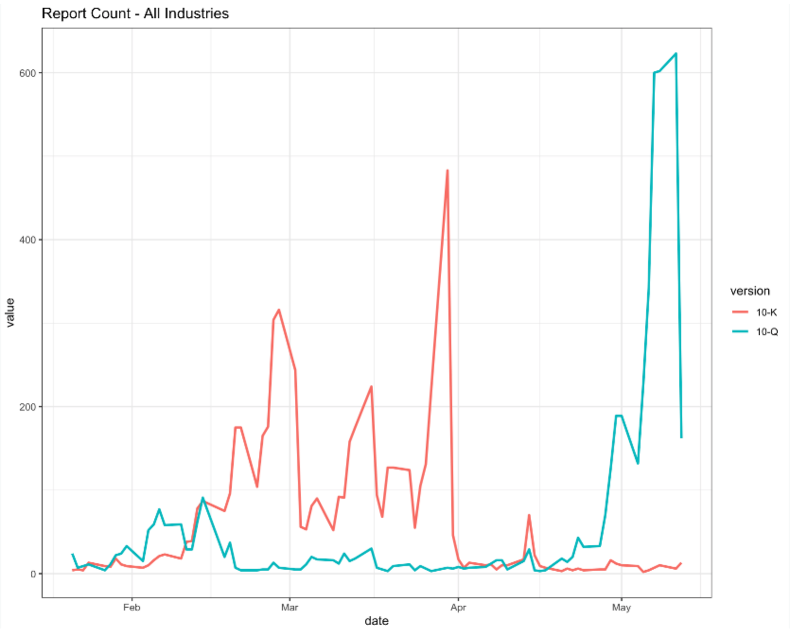


Fig. S 5 Number of reports filed to the SEC per day January to May 2020. The 10-Q reports show a very different dynamic from the 10-K reports. While 10-K reports are filed throughout the first quarter (and to a lower extend in the second quarter), the number of filed 10-Q reports is vastly skewed towards May.

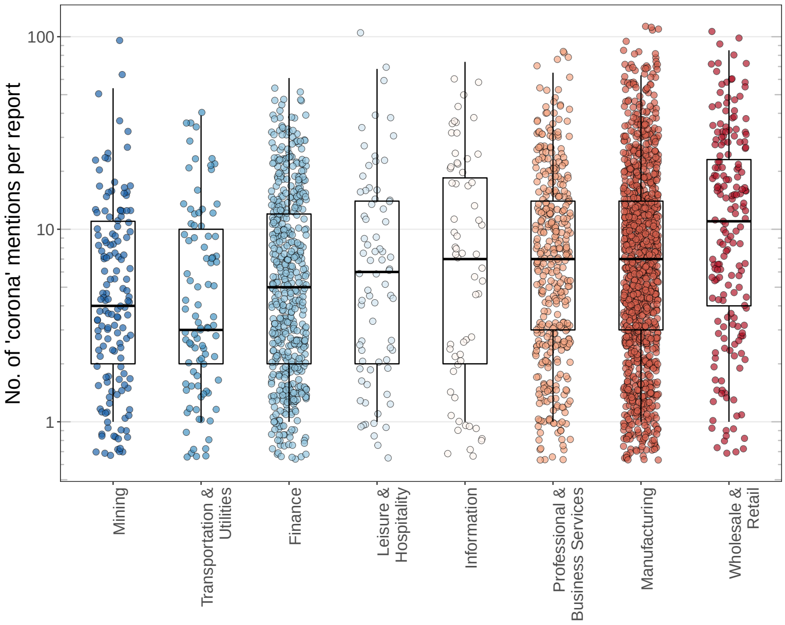


Fig. S 6 Number of 'corona' keywords in reports per industry. There is a substantial variation within industries. The average differences between industries are less pronounced, but still relevant: while 50 % of the firms in Finance, Transportation & Utilities, and Mining mention 'corona' or 'covid' five times or less, the value is twice as high in Wholesale & Retail.

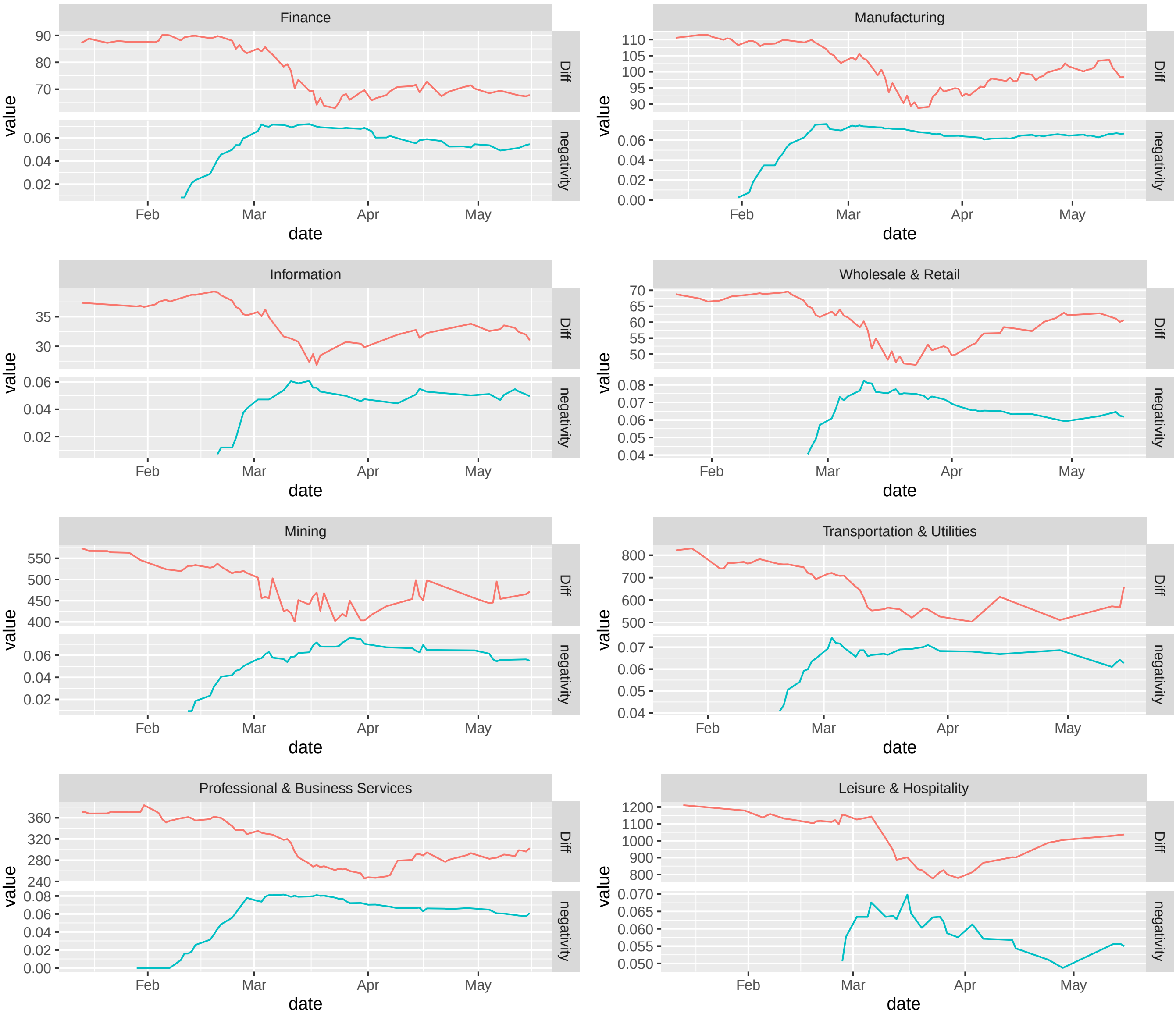
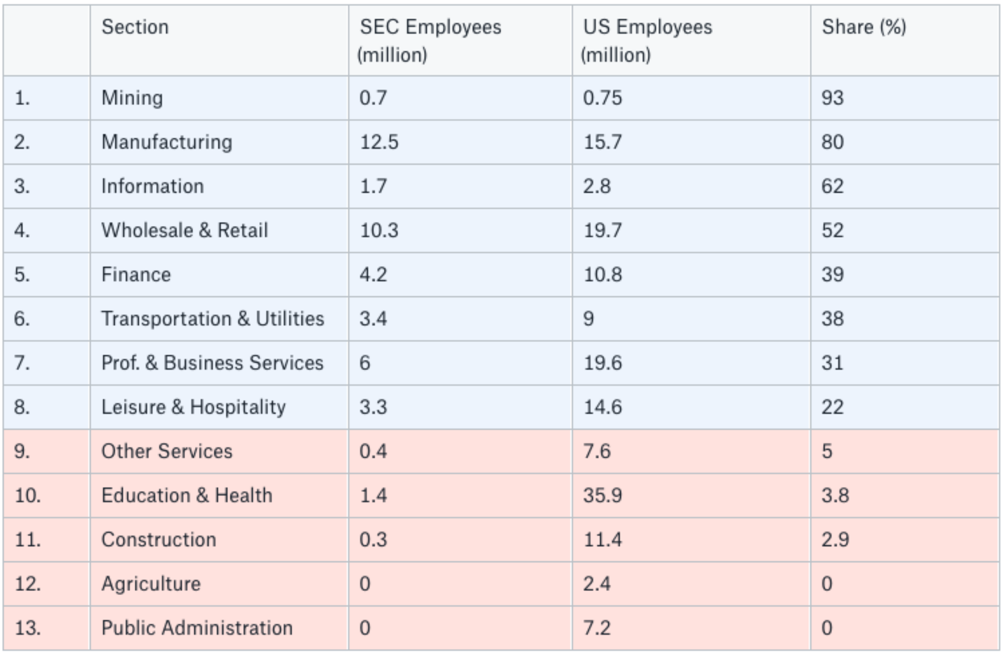


Fig. S 7 Average stock market development (upper panel 'Diff') and share of negative words in 10-K reports (lower panel 'negativity') per industry. The average text sentiment is correlated with stock market developments in all eight industries. In some industries, such as Finance and Manufacturing, the text negativity rises before the stock markets plummet.



	Section	SEC Employees (million)	US Employees (million)	Share (%)
1.	Mining	0.7	0.75	93
2.	Manufacturing	12.5	15.7	80
3.	Information	1.7	2.8	62
4.	Wholesale & Retail	10.3	19.7	52
5.	Finance	4.2	10.8	39
6.	Transportation & Utilities	3.4	9	38
7.	Prof. & Business Services	6	19.6	31
8.	Leisure & Hospitality	3.3	14.6	22
9.	Other Services	0.4	7.6	5
10.	Education & Health	1.4	35.9	3.8
11.	Construction	0.3	11.4	2.9
12.	Agriculture	0	2.4	0
13.	Public Administration	0	7.2	0

Fig. S 8 Number of employees working in different sectors of the US economy and share of employees working in firms that report to the SEC. In eight out of 13 sectors, the firms that file to SEC represent at least 22 % of all employees in that sector.

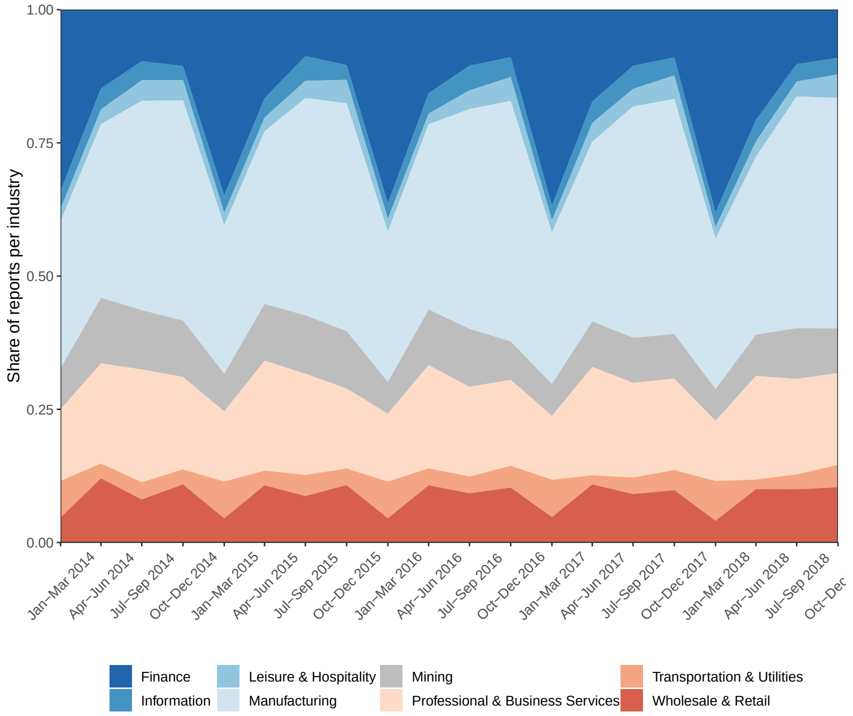


Fig. S 9 Share of reports per quarter of the eight different industries. There is some seasonality, which is mainly driven by Finance's large number of reports in the first quarter. The other industries do not show substantial variation within a calendar year.

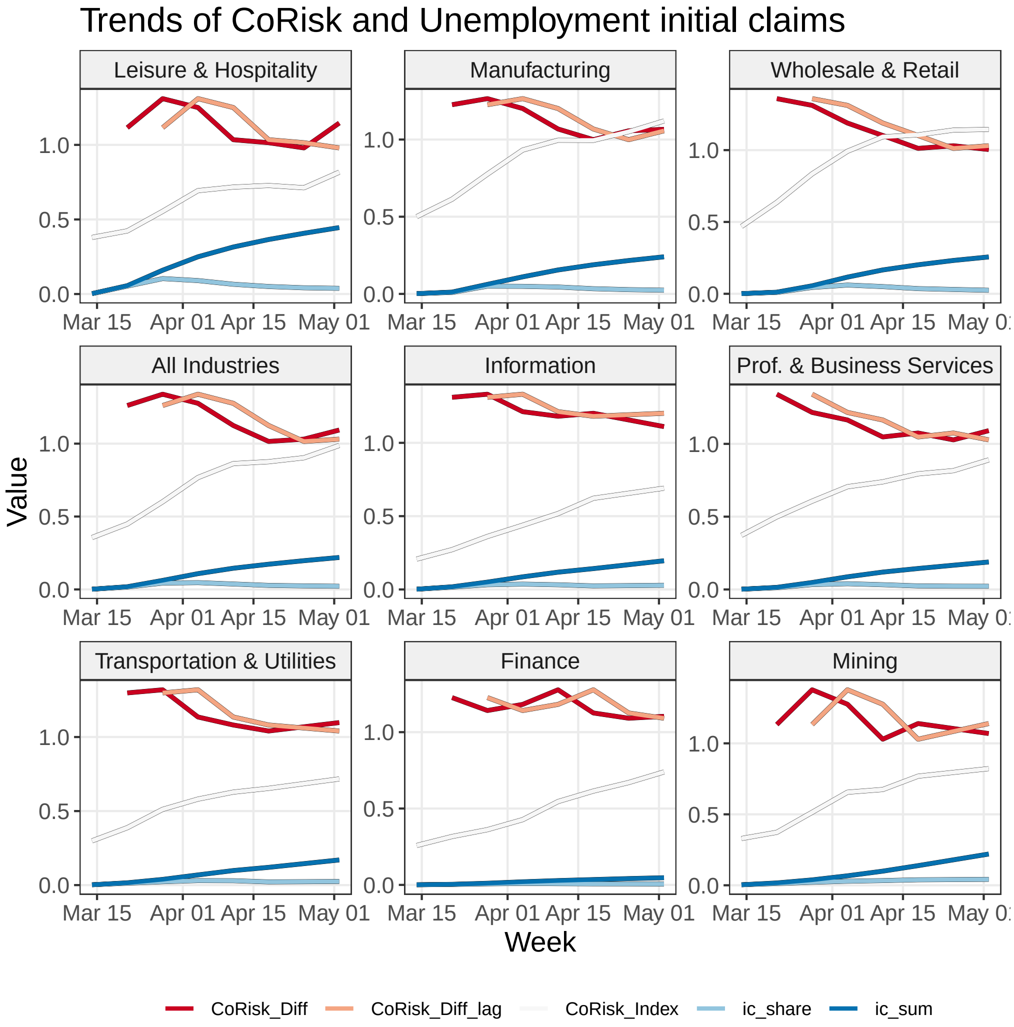


Fig. S 10 Trends of the CoRisk-Index and unemployment initial claims per industry. The CoRisk-Index (white line) and the sum of unemployment initial claims are both driven by a substantial trend component. The weekly differences in the CoRisk-Index (red line), its' lag (orange line), and the weekly unemployment initial cliams show a more stationary behaviour.

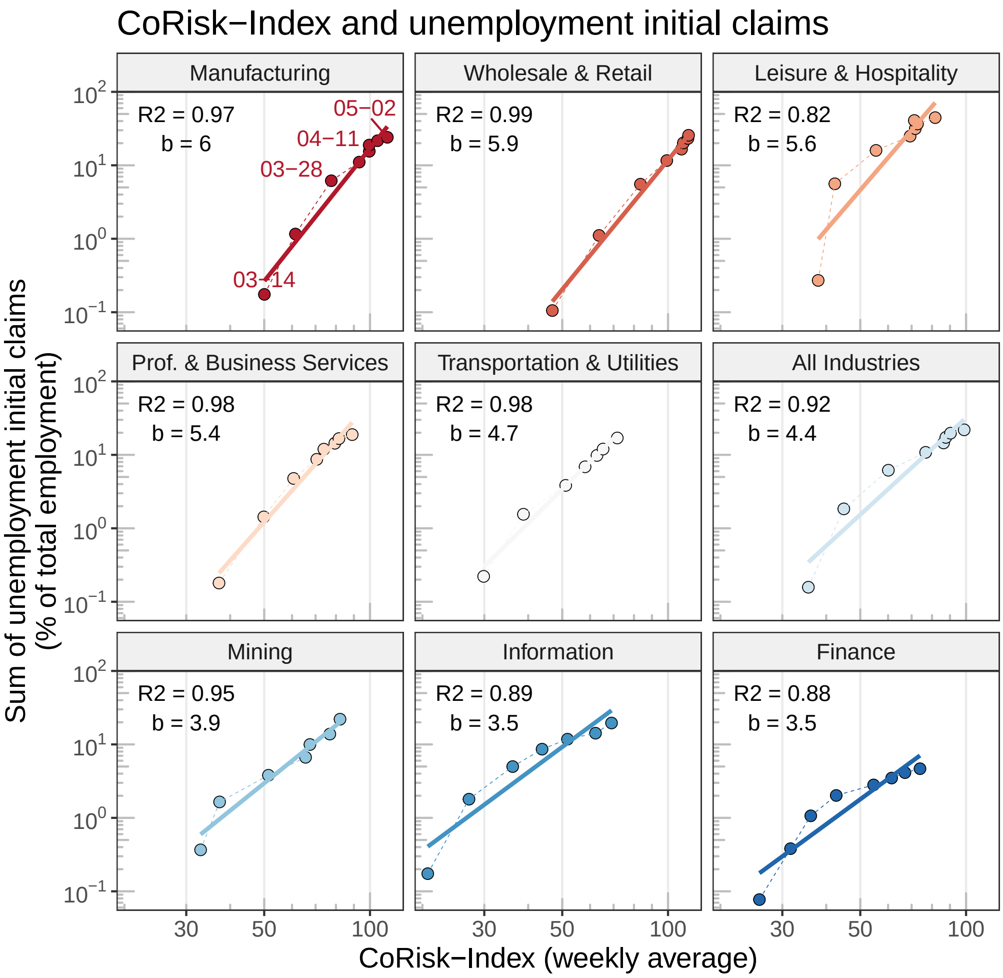


Fig. S 11 Spurious correlations between the CoRisk-Index and the total number of unemployment initial claims per industry. The overall trend overshadows any relation between both variables.

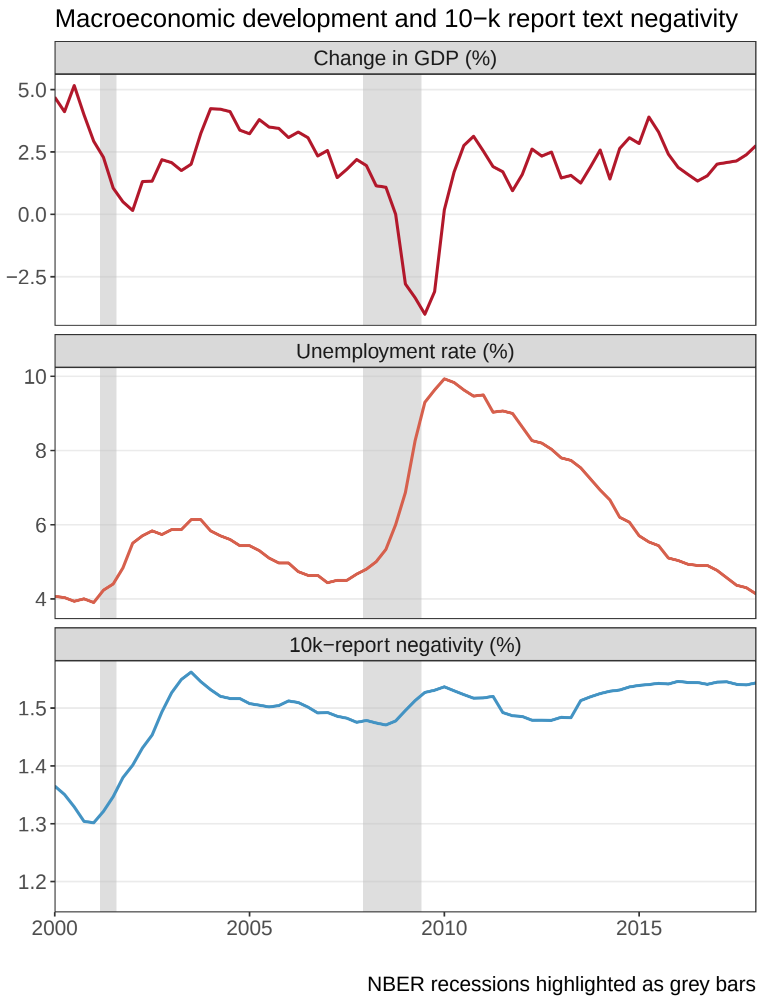


Fig. S 12 Comparison of quarterly GDP change (upper panel), unemployment rates (central panel), and share of negative words in 10-K reports (lower panel). The figure exemplifies that the sentiment of 10-K reports tend to correlate with overall macroeconomic developments.

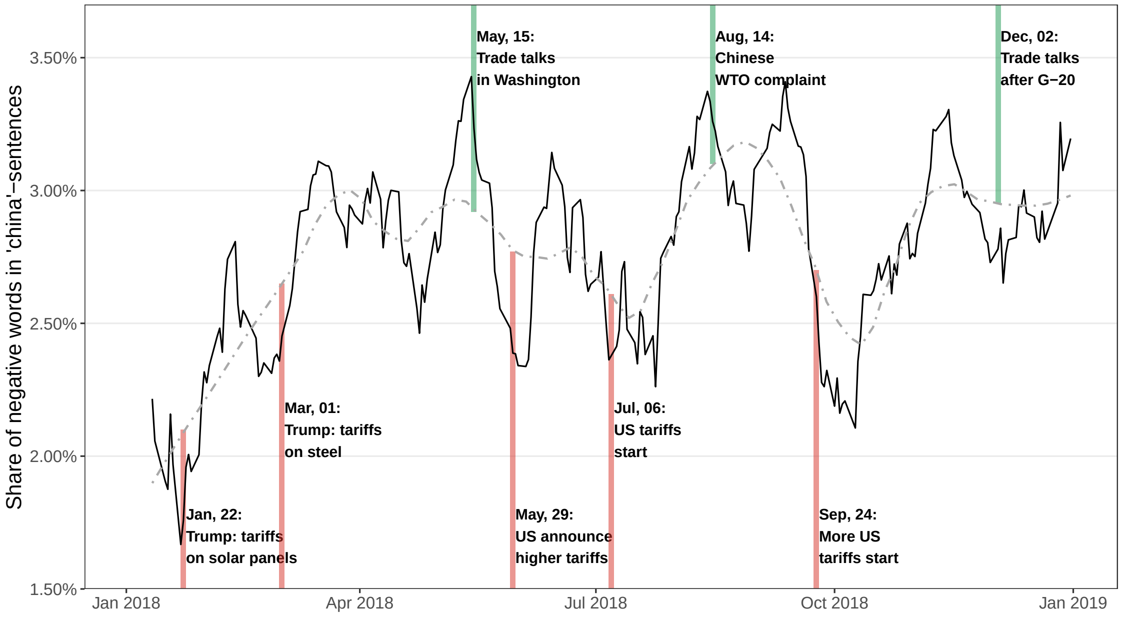
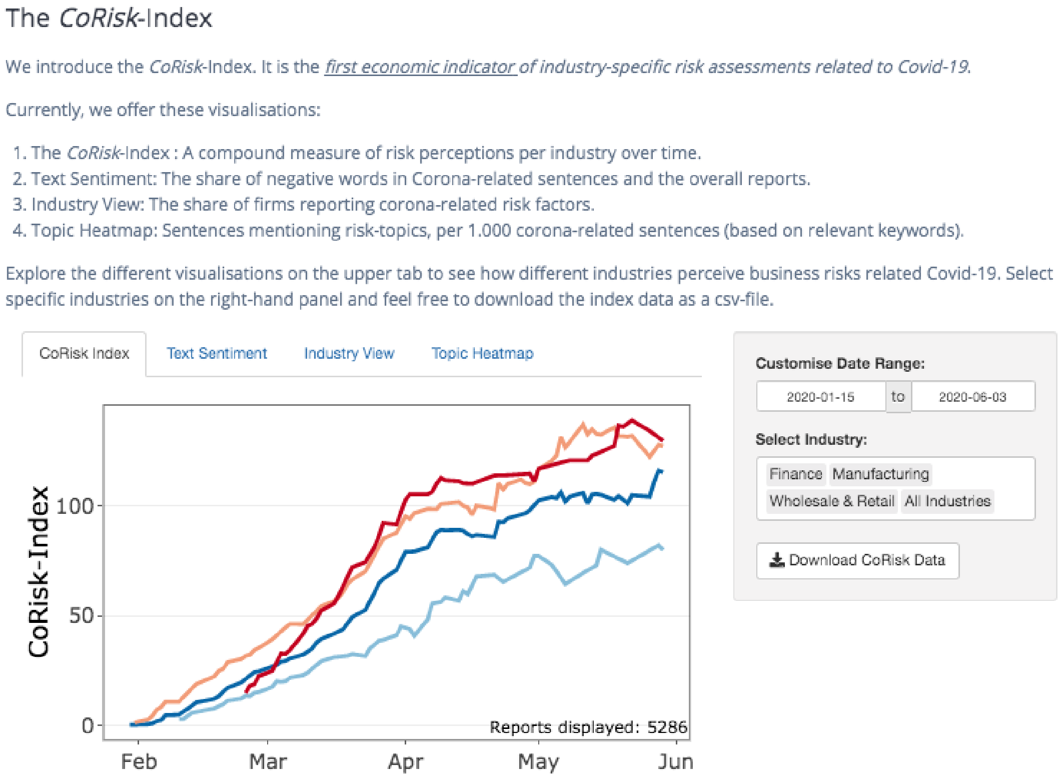


Fig. S 13 Share of negative words (14-days average, solid line) in 'china'-sentences of SEC reports during the 2018 'US-China Trade War'. The sentiment in sentences that relate to China reflects major developments during the crisis (red and green bars). Shortly after the announcement of US-tariffs, the text negativity increased, and it tended to decrease after releases.



The *CoRisk*-Index

We introduce the *CoRisk*-Index. It is the *first economic indicator of industry-specific risk assessments related to Covid-19.*

Currently, we offer these visualisations:

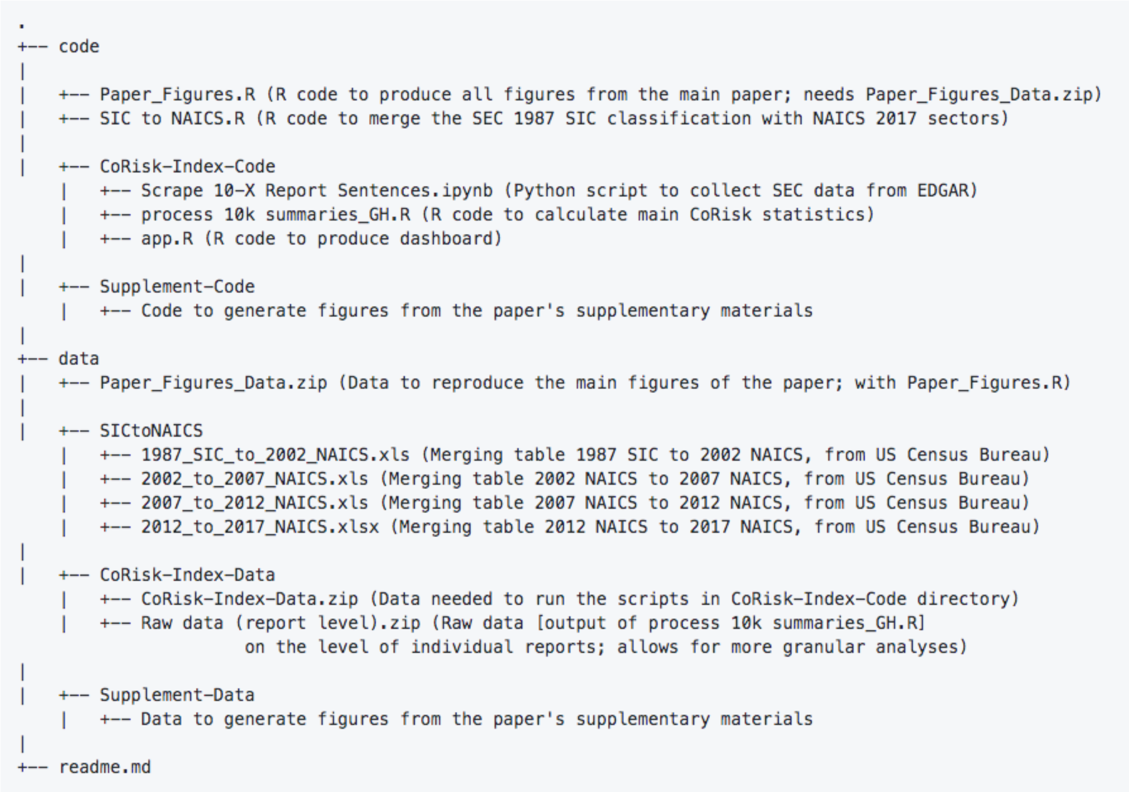
1. The *CoRisk*-Index : A compound measure of risk perceptions per industry over time.
2. Text Sentiment: The share of negative words in Corona-related sentences and the overall reports.
3. Industry View: The share of firms reporting corona-related risk factors.
4. Topic Heatmap: Sentences mentioning risk-topics, per 1.000 corona-related sentences (based on relevant keywords).

Explore the different visualisations on the upper tab to see how different industries perceive business risks related Covid-19. Select specific industries on the right-hand panel and feel free to download the index data as a csv-file.

Customise Date Range: 2020-01-15 to 2020-06-03

Select Industry: Finance, Manufacturing, Wholesale & Retail, All Industries

Fig. S 14 An overview of the online dashboard, available at `http://oxford.berlin/corisk`. The dashboard allows users to explore the CoRisk-Index, a number of other measures, to customise the date range, and to download the CoRisk data.



```
.
+-- code
|
|    +-- Paper_Figures.R (R code to produce all figures from the main paper; needs Paper_Figures_Data.zip)
|    +-- SIC to NAICS.R (R code to merge the SEC 1987 SIC classification with NAICS 2017 sectors)
|
|    +-- CoRisk-Index-Code
     |    +-- Scrape 10-X Report Sentences.ipynb (Python script to collect SEC data from EDGAR)
     |    +-- process 10k summaries_GH.R (R code to calculate main CoRisk statistics)
     |    +-- app.R (R code to produce dashboard)
|
|    +-- Supplement-Code
     |    +-- Code to generate figures from the paper's supplementary materials
|
+-- data
|    +-- Paper_Figures_Data.zip (Data to reproduce the main figures of the paper; with Paper_Figures.R)
|
|    +-- SICtoNAICS
     |    +-- 1987_SIC_to_2002_NAICS.xls (Merging table 1987 SIC to 2002 NAICS, from US Census Bureau)
     |    +-- 2002_to_2007_NAICS.xls (Merging table 2002 NAICS to 2007 NAICS, from US Census Bureau)
     |    +-- 2007_to_2012_NAICS.xls (Merging table 2007 NAICS to 2012 NAICS, from US Census Bureau)
     |    +-- 2012_to_2017_NAICS.xlsx (Merging table 2012 NAICS to 2017 NAICS, from US Census Bureau)
|
|    +-- CoRisk-Index-Data
     |    +-- CoRisk-Index-Data.zip (Data needed to run the scripts in CoRisk-Index-Code directory)
     |    +-- Raw data (report level).zip (Raw data [output of process 10k summaries_GH.R]
                        on the level of individual reports; allows for more granular analyses)
|
|    +-- Supplement-Data
     |    +-- Data to generate figures from the paper's supplementary materials
|
+-- readme.md
```

Fig. S 15 An overview of the folder structure of the material published on the GitHub repository (`https://github.com/Braesemann/CoRisk`). Software code and data published on the repository allow users to replicate the CoRisk-Index and the figures presented in the main report.

Tables

Topic number	Top 10 words
Topic 1:	impact, extent, including, outbreak, uncertain, future, highly, results, developments, depend
Topic 2:	operations, including, health, outbreak, business, supply, products, economic, public, result
Topic 3:	outbreak, spread, countries, impact, including, china, business, potential, economic, government
Topic 4:	china, outbreak, novel, covid, adversely, wuhan, strain, business, december, recent

Tab. S 1 Results of an LDA topic modelling with four topics. The algorithmically derived topics give a good insight into the general narratives of risks used in the documents.

Topic	Keywords
Production	business operation, business disruption, product, work stoppage, labor disruption, labor, work, manufacturing operation, labor shortage, employee productivity, product development, business activity
Supply	manufacturing facility, manufacture facility, contract manufacturer, service provider, logistic provider, supply disruption, party manufacturer, supply disruption, facility, supply, transportation delay, delivery delay, supplier, business partner, supply chain, material shortage
Travel	air travel, travel, travel restriction, airline industry, travel disruption
Demand	store closure, distribution channel, market condition, consumer spend, market acceptance, consumer confidence, consumer demand, consume, store, customer, store traffic
Finance	operating result, cash flow, stock price, estate value, credit availability, performance problem

Tab. S 2 Keywords for five topical domains. We combine the results of the unsupervised methods with domain knowledge from economics to label the five main topics and specify defining keywords.

- Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., & Dighe, A. (2020). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand

References

[1] R. Aroussi, yfinance: Yahoo! Finance market data downloader (2019).

[2] T. Loughran, B. McDonald, *The Journal of Finance* **66**, 35 (2011). Publisher: Wiley Online Library.

[3] T. Dyer, M. Lang, L. Stice-Lawrence, *Journal of Accounting and Economics* **64**, 221 (2017).

[4] D. M. Blei, *Communications of the ACM* **55**, 77 (2012). Publisher: ACM New York, NY, USA.

[5] J. T. Wu, *et al.*, *Nature Medicine* pp. 1–5 (2020). Publisher: Nature Publishing Group.

[6] I. I. Bogoch, *et al.*, *Journal of Travel Medicine* **27** (2020). Publisher: Oxford Academic.

[7] N. M. Ferguson, *et al.*, *London: Imperial College COVID-19 Response Team, March* **16** (2020).

[8] M. W. Fong, *et al.*, *Emerging infectious diseases* **26** (2020).

[9] C. Nicolaides, D. Avraam, L. Cueto-Felgueroso, M. C. González, R. Juanes, *Risk Analysis* **n/a** (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.13438.

[10] R. M. del Rio-Chanona, P. Mealy, A. Pichler, F. Lafond, D. Farmer, *arXiv preprint arXiv:2004.06759* (2020).

[11] L.-P. Béland, *et al.*, The Short-Term Economic Consequences of COVID-19: Exposure to Disease, Remote Work and Government Response, *Tech. rep.*, Global Labor Organization (GLO) (2020).

[12] S. C. Ludvigson, S. Ma, S. Ng, Covid19 and the Macroeconomic Effects of Costly Disasters, *Tech. rep.*, National Bureau of Economic Research (2020).

[13] D. Lewis, K. Mertens, J. H. Stock, US Economic Activity During the Early Weeks of the SARS-Cov-2 Outbreak, *Tech. rep.*, National Bureau of Economic Research (2020).

[14] R. Baldwin, B. W. d. Mauro, *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes* (CEPR Press, 2020).

[15] F. Dorn, *et al.*, *ifo Institut* **ifo Schnelldienst Vorabdruck** (2020).

[16] R. Baldwin, B. W. d. Mauro, *Economics in the Time of COVID-19* (CEPR Press, 2020).

[17] G. Gopinath, *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes* (CEPR Press, 2020).

[18] S. Ramelli, A. Wagner, *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes* (CEPR Press, 2020).

[19] Y. Huang, C. Lin, P. Wang, Z. Xu, *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes* (CEPR Press, 2020).

[20] T. A. Hassan, S. Hollander, L. van Lent, A. Tahoun, Firm-Level Political Risk: Measurement and Effects, *Working Paper 24029*, National Bureau of Economic Research (2017). Series: Working Paper Series.

[21] T. A. Hassan, S. Hollander, L. van Lent, A. Tahoun, Firm-level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1, *Tech. rep.*, National Bureau of Economic Research (2020).

[22] M. R. Keogh-Brown, R. D. Smith, J. W. Edmunds, P. Beutels, *The European Journal of Health Economics* **11**, 543 (2010).

[23] B. Buetre, Y. Kim, Q. T. Tran, D. Gunasekera, *Australian Commodities: Forecasts and Issues* **13**, 351 (2006). Publisher: Australian Bureau of Agricultural and Resource Economics.

[24] D. Rassy, R. D. Smith, *Health Economics* **22**, 824 (2013). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.2862.

[25] R. J. Hatchett, C. E. Mecher, M. Lipsitch, *Proceedings of the National Academy of Sciences* **104**, 7582 (2007). Publisher: National Acad Sciences.

[26] S. S. Morse, *Proceedings of the National Academy of Sciences* **104**, 7313 (2007). Publisher: National Acad Sciences.

[27] R. J. Barro, J. F. Ursua, J. Weng, The Coronavirus and the Great Influenza Epidemic - Lessons from the 'Spanish Flu' for the Coronavirus's Potential Effects on Mortality and Economic Activity, *SSRN Scholarly Paper ID 3556305*, Social Science Research Network, Rochester, NY (2020).

[28] M. Karlsson, T. Nilsson, S. Pichler, *Journal of Health Economics* **36**, 1 (2014).

[29] D. Bogataj, M. Bogataj, *International Journal of Production Economics* **108**, 291 (2007).

[30] E. Atalay, A. Hortacsu, J. Roberts, C. Syverson, *Proceedings of the National Academy of Sciences* **108**, 5199 (2011). Publisher: National Acad Sciences.

[31] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, A. Tahbaz-Salehi, *Econometrica* **80**, 1977 (2012). Publisher: Wiley Online Library.

[32] T. Aven, *Reliability Engineering & System Safety* **99**, 33 (2012).

[33] T. Aven, O. Renn, *Journal of Risk Research* **0**, 1 (2019). Publisher: Routledge _eprint: https://doi.org/10.1080/13669877.2019.1569099.

[34] L. D. Richman, *et al.*, *Journal of Investment Compliance* **20**, 1 (2019). Publisher: Emerald Publishing Limited.

[35] S. Kogan, D. Levin, B. R. Routledge, J. S. Sagi, N. A. Smith, *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics on - NAACL '09* (Association for Computational Linguistics, Boulder, Colorado, 2009), p. 272.