// ZEW expert brief 20-05 · 04/17/2020

Jan Kinne/ Miriam Krüger/ David Lenz/ Georg Licht/ Peter Winker

Coronavirus Pandemic Affects Companies Differently

A high-frequency website analysis of companies' reactions to the coronavirus pandemic in Germany

Leibniz Association

istari.ai

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

ZEW

INHALT/CONTENT

SUMMARY

We analyze the websites of approximately 1.1 million German companies twice a week for references of the word "coronavirus" and corresponding synonyms. The identified text passages are then evaluated using AI text analysis models to determine the context of the references. This procedure allows us to assess the impact of the coronavirus pandemic on companies in Germany in a timely and comprehensive manner.

Since the first web scraping on 03/19/2020, we have been able to observe an increase in the number of affected companies by approximately one percentage point per week (from 10% to 14% on 04/15/2020). We also note significant sectoral and age- and size-related differences. Older and especially larger companies report the pandemic much more frequently. We are still investigating the influence of possibly larger digital capacities (e.g. dedicated media department) and a consequently more up-to-date website of larger companies. Regionally, we observe both an uneven distribution of the current case numbers (affected companies) at the district level, and uneven growth. An East-West divide with higher case numbers in the West of Germany and a stronger growth of case numbers in the South-West of Germany can be observed. City districts also show higher values overall than rural districts. The influence of the regional composition of the enterprise population on this spatial pattern is the subject of current analyses.

Furthermore, our experiments with advanced AI-based text analysis models (*istari.ai webAI Engine*) indicate that companies report on their websites about their specific reactions to the coronavirus pandemic. In this way, we can show that the "coronavirus" references found can be divided into different subject areas, the relevance of which varies by industry and region. Coronavirus references with a "problem" context (plant and shop closures, cancellations of events, short-time work, etc.), "adaptation" context (adaptation of business hours, new hygiene regulations, home office, etc.) and "no problem" context (no adverse effects of the pandemic) occur more frequently in certain regions.

1. Data

This study is based on 1.1 million web addresses (URLs) of companies in Germany taken from the Mannheim Company Panel (as of the end of 2019). For each of these companies, basic company characteristics such as company location, number of employees, and industry sector are also known. From previous research it is known that about half of the companies in Germany have their own website [1], with larger companies in particular having good to very good coverage (from 25 employees 84%; from 250 employees 97%). Furthermore, there are sector-specific differences (low coverage e.g. for agricultural enterprises; high coverage e.g. for mechanical engineering).

The companies' websites were queried and downloaded using an approach developed by the startup istari.ai, with a maximum of five webpages per company (a website usually consists of several (sub-)webpages). The selection of these sub-webpages is not random, but follows a simple heuristic: First, sub-webpages are selected that are probably written in German and also have the shortest URL. The latter leads to these sub-webpages with more general and up-to-date ("top-level") information being downloaded most frequently. For example, "zew.de/news" would be downloaded before "zew.de/news/2019/february/". Overall, 81% (1.11 million) of the original 1.36 million web pages were successfully accessed and downloaded. The failed attempts were due to websites being no longer up-to-date and permanently or temporarily deactivated.

2. Keyword search and text analysis

The downloaded web pages were then searched for variations of the term "coronavirus pandemic" and its synonyms (e.g. "SARS-CoV2"). In case of a hit, the affected sections of the web page (HTML markups) were marked. This simple approach allowed for a first estimation of the percentage of companies reporting about the coronavirus pandemic on their website.

The identified website sections can also be further subdivided by means of computer-based text analysis, meaning that the context of the mention of the "coronavirus" term can be recorded. For this purpose, a machine learning based text analysis model specially adapted by istari.ai was used. In a first, labor-intensive step, the content (e.g. "delays in production", "employee information", "plant closures", etc.) of a random selection of previously identified "coronavirus" web page sections was screened and manually classified. Each web page section was classified into one of the following classes:

- **Problem**: The company reports on problems related to the coronavirus pandemic. Particularly, this includes closures of stores, cancellations, and postponements of events, reports of delivery bottlenecks, short-time work, and other similar problems.

- **No Problem**: The company reports that it is not affected by the coronavirus pandemic or that it has no impact on its business.

- **Adaption**: The company reports that it is adapting to the new circumstances. This includes measures such as new hygiene regulations, changed opening hours, and home office, among others.

- **Information**: The company generally reports on the coronavirus pandemic. This includes general information such as the current spread as well as symptoms of the disease, news about coronavirus, and the announcement of official regulations.

- **Unclear**: This group includes texts that cannot be clearly assigned because they are either artifacts or misclassifications, or because it is not clear what the context of the "coronavirus" designation is.

These labelled data were then used as training data for a transfer learning based language model [2]. During this training, the model "learns" to recognize differences between the texts and is then able to automatically classify each of the identified web page sections as one of the previously defined groups. This enables an evaluation of the content of each identified website section, which provides an up-to-date picture of the impact of the coronavirus pandemic on companies in Germany.

3. Results

3.1. Affected companies

Figure 1 shows the proportion of companies that have at least one reference to the coronavirus pandemic on their website. An increasing trend can be observed, which weakens over the four weeks shown and then remains at a level of about 14%.

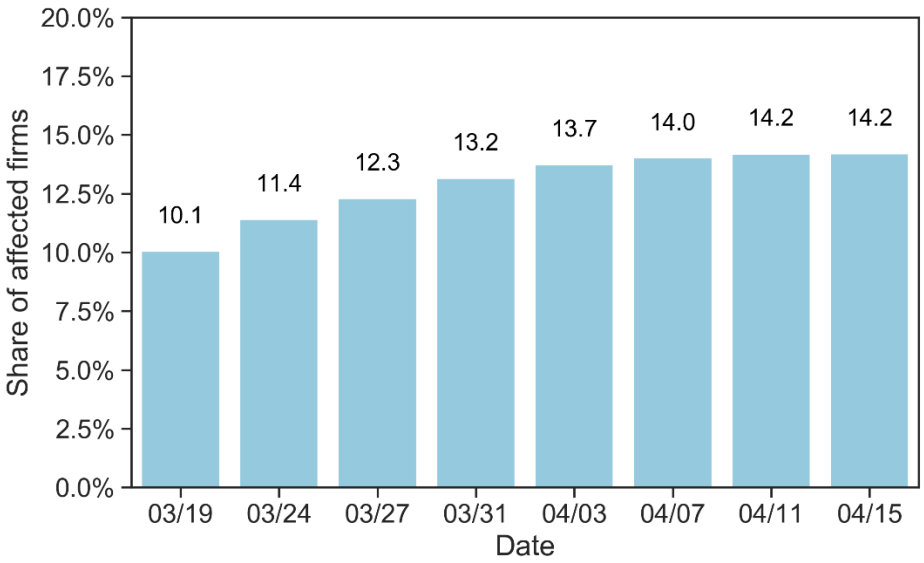

Figure 1: Share of companies with at least one "coronavirus" web reference.

Figure 2 shows that larger (and older) companies with more than 250 employees in particular are reporting on the coronavirus pandemic and have also seen the highest increase in recent weeks. To what extent this reflects the actual impact on the companies, or rather their larger digital capacities (e.g. dedicated media department) and a consequently more up-to-date website, is still under investigation.

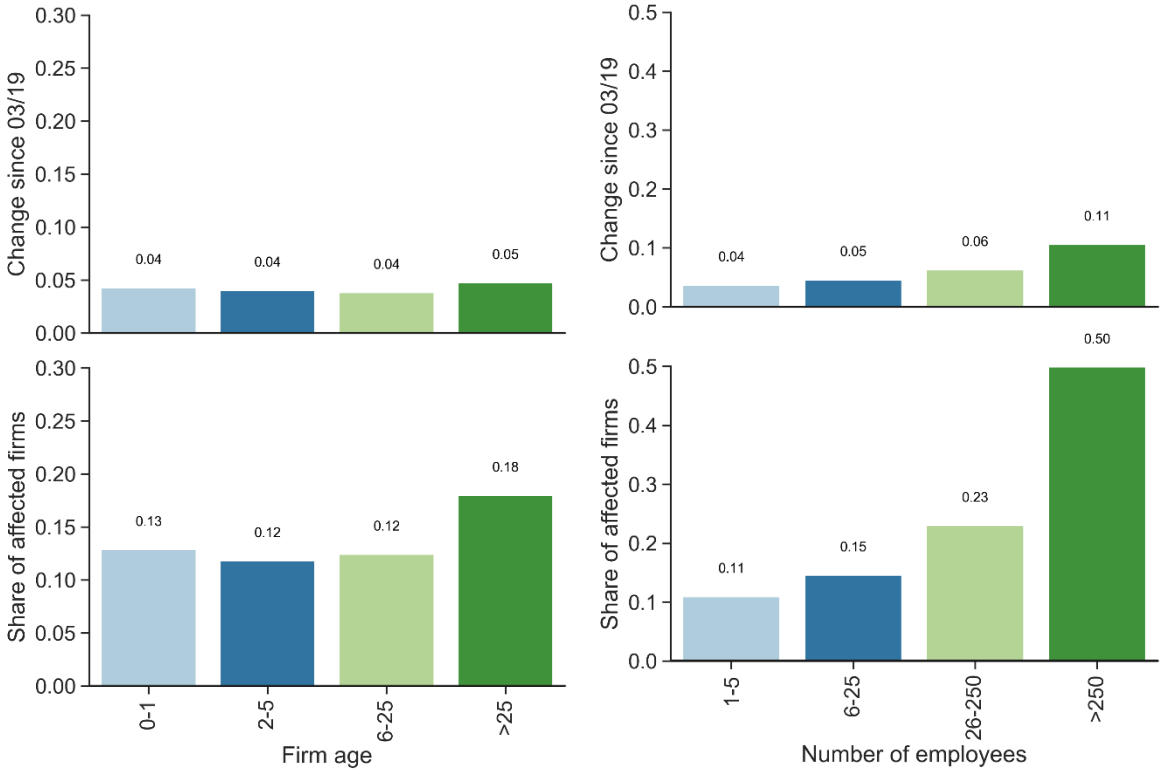

Figure 2: Share of companies with at least one "Coronavirus" web reference by company age and size.

Figure 3 shows that there are also large sectoral differences. Public administration in particular, but also associations, entertainment companies, healthcare, and educational and training institutions report on the coronavirus pandemic. In contrast, relatively small proportions of reports are to be found for the construction industry, wholesale, agriculture, and logistics companies, and the manufacturing industry.

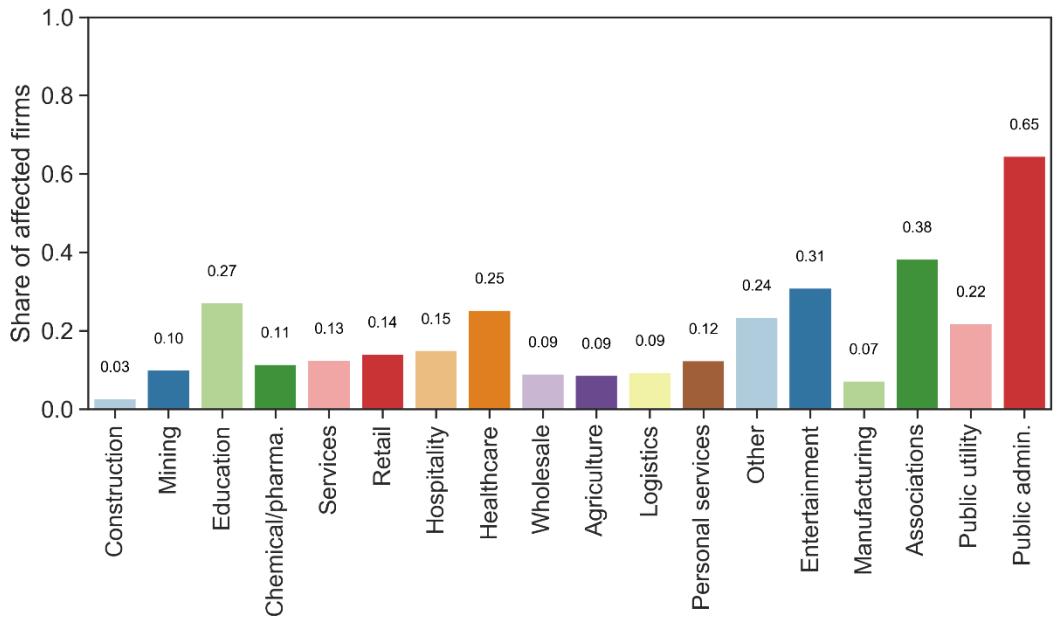

Figure 3: Share of companies with at least one "coronavirus" web reference by sector.

Figure 4 shows that all sectors have experienced a similar increase in "coronavirus" references over the past four weeks. Already at the beginning of our measurements, however, some sectors, such as public administration, started at a very high level already and thus show rather low relative growth. Other sectors, such as wholesale trade, start at a very low level but then show very high relative growth. The number of affected wholesale companies therefore almost doubled over the period shown.

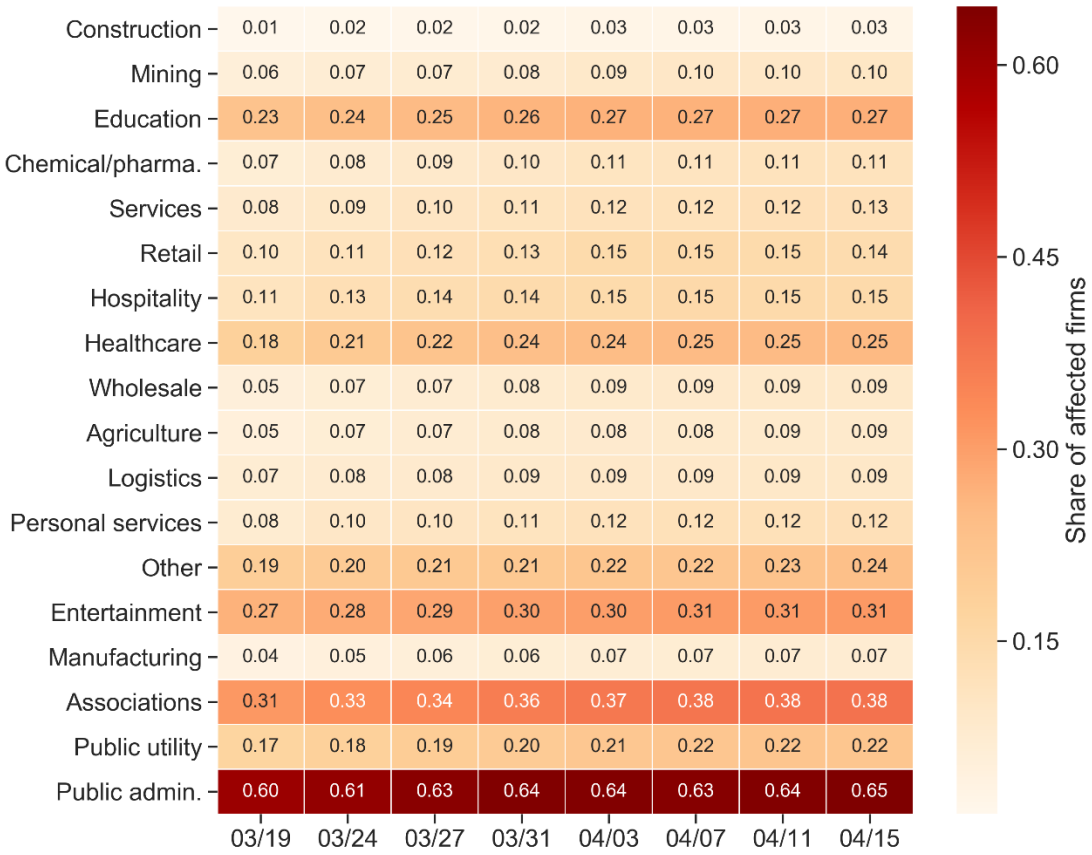

Figure 4: Shares and change companies with at least one "coronavirus" web reference by sector.

Figure 5 illustrates an east-west divide with regard to the proportion of regionally affected companies in Germany (left map). The shares of affected enterprises in urban districts are usually higher than in rural districts, though this difference is particularly striking in the eastern part of Germany. Regional differences can also be seen in the change over the past few weeks (right map). For example, particularly high growth rates can be observed in the western states of Baden-Württemberg, Hesse, Rhineland-Palatinate, North Rhine-Westphalia, and Lower Saxony.

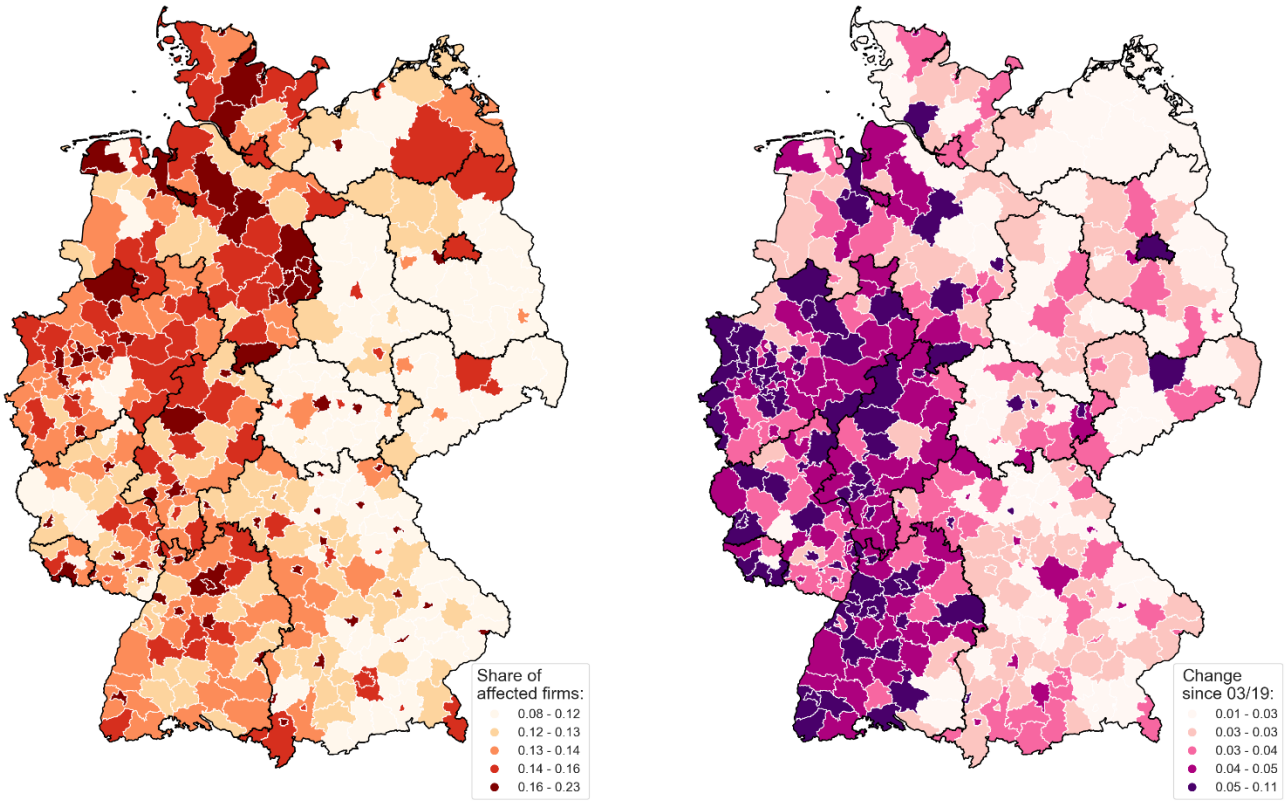

Figure 5: Share (left) and change (right) of companies with at least one "coronavirus" web reference by districts.

3.2. Type of affectedness

Figure 6 shows the most common words of the "coronavirus" reference classes we defined ("Problem", "No Problem", "Adaptation" and "Information", see section 2) as word clouds. The basis for this are the approximately 460,000 "coronavirus" references (as of 03/24) from the company websites identified and classified with the istari.ai text analysis model.

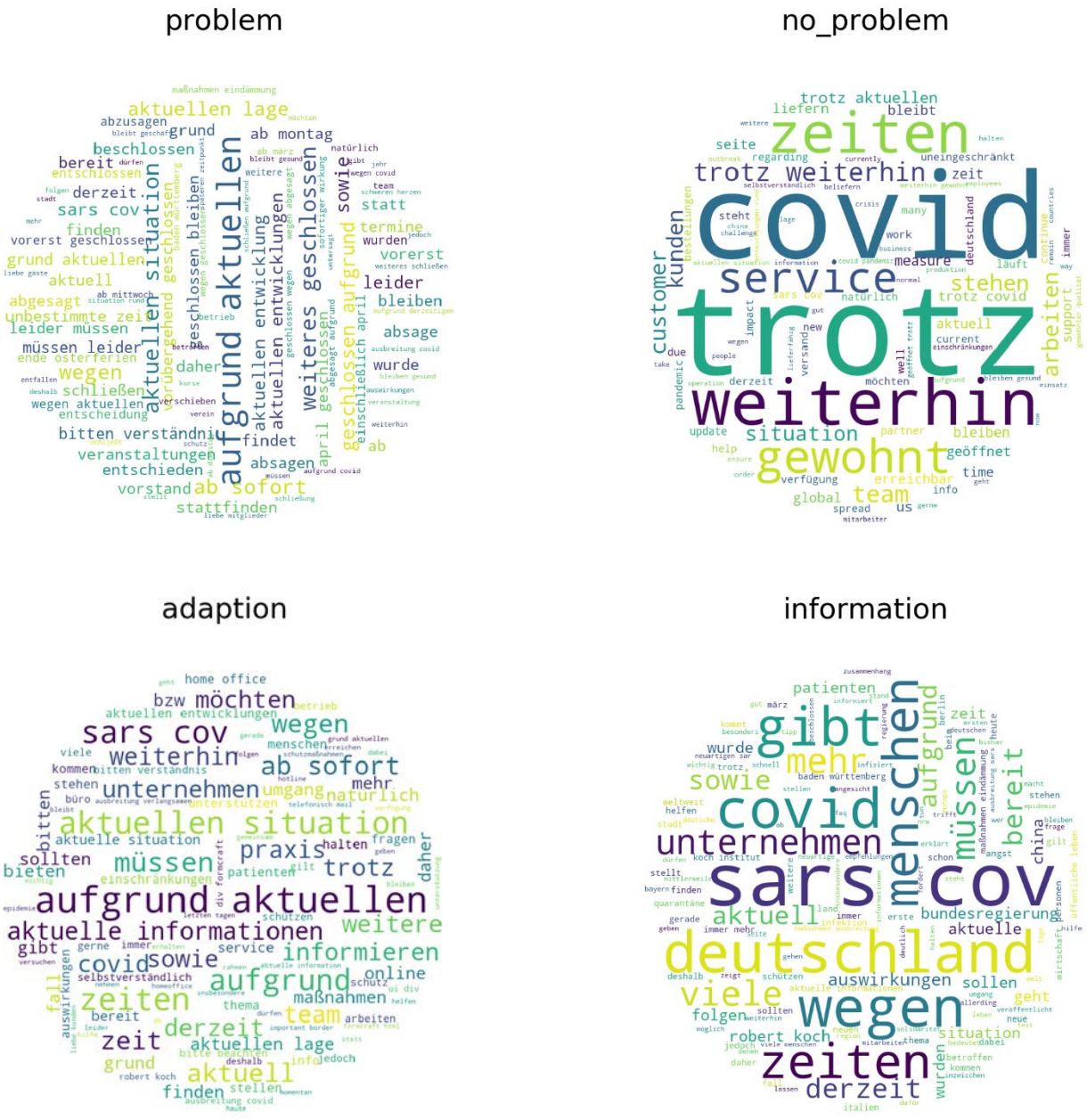

Figure 6: Word clouds for four "coronavirus" reference classes.

Figure 6 shows that companies with at least one "coronavirus" reference are affected to varying degrees by the individual reference classes, depending on the sector. For example, public administration very often informs on "coronavirus"-specific issues, while the health care sector mainly reports on adjustments, and the entertainment industry in particular reports on more serious problems. The chemical and pharmaceutical, wholesale, and (non-personal) services sectors appear to have no or only minor problems. It should be noted here that a company can be affected by several reference classes if several "coronavirus" references are found on its website. For example, a company may be affected by both "Problem" and "No problem". An example of this would be gardening centers that have to close down their (now widely run) on-site gastronomy but are able to continue their core business (sale of plants and gardening supplies), for the most part without restrictions.

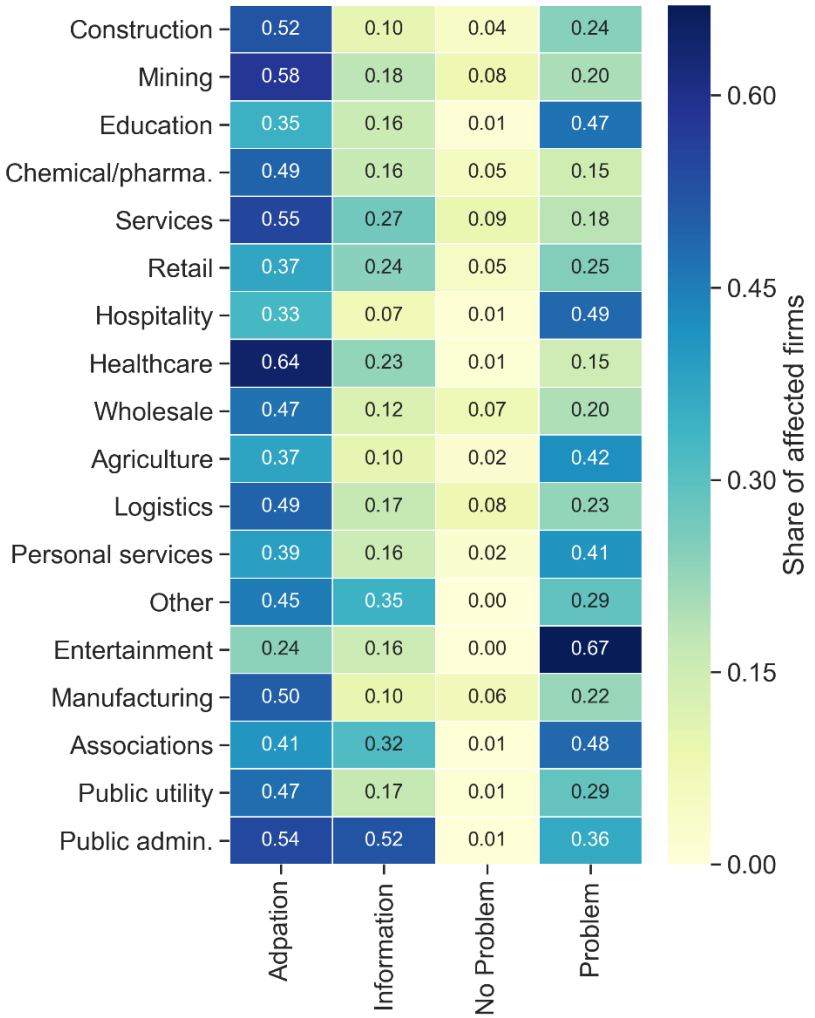

Figure 7: Share of companies with at least one "coronavirus" web reference, assigned to the respective class, by industry.

Figure 8 shows that these "coronavirus" reference classes vary from region to region and are in some cases clustered (statistically significant). This is particularly noticeable in the opposing classes "Problem" and "No Problem". The eastern German states in particular show several statistically significant (99% confidence) clusters of districts with low "Problem" (blue hatched coldspots) or high "No Problem" (red hatched hotspots) shares of affected companies. In these regions, companies are also more frequently reporting on adaptations to the coronavirus pandemic, as can be seen from the "Adaptation" hotspots (red hatched).

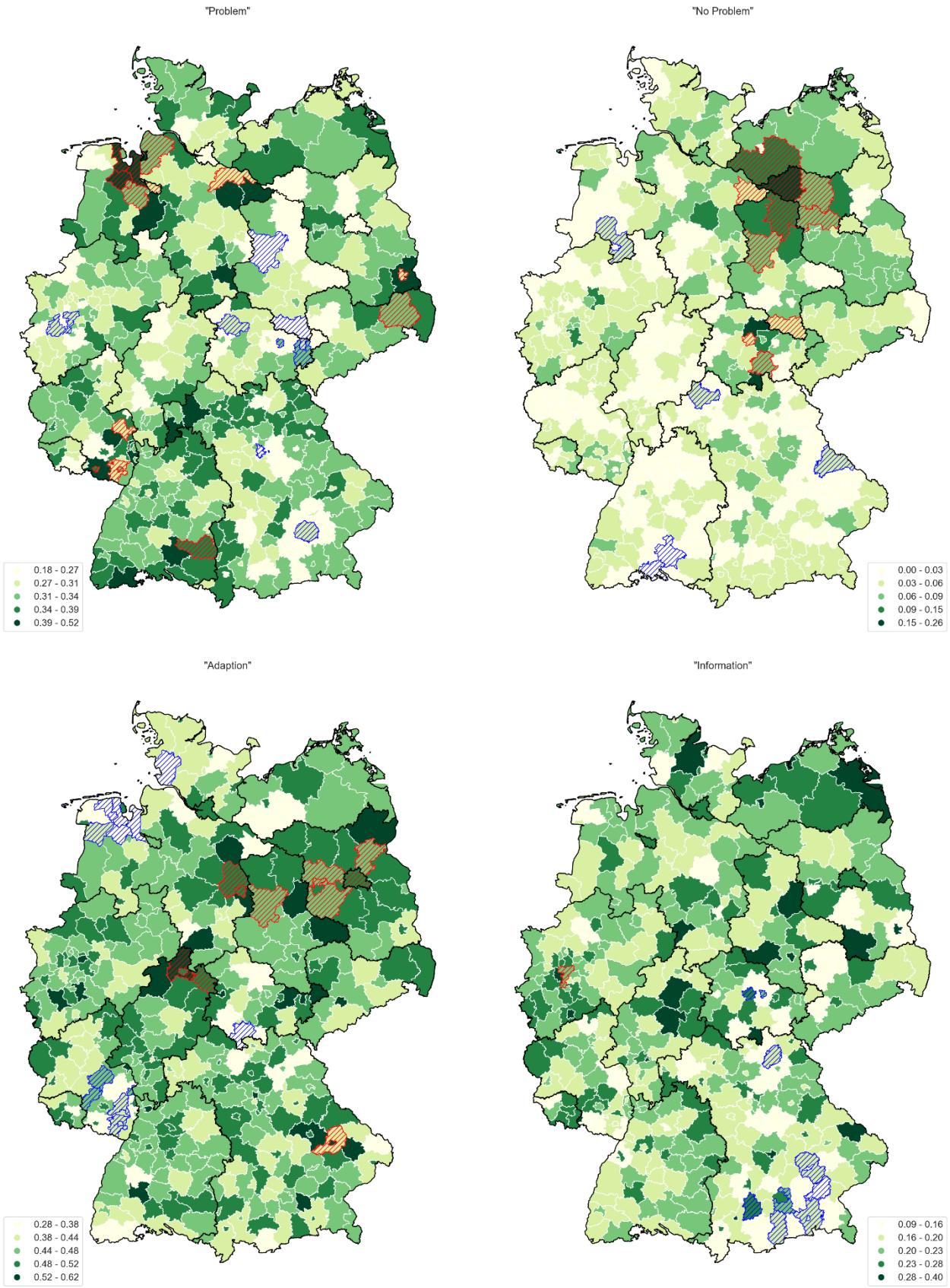

Figure 8: Share of regionally affected companies with at least one "coronavirus" web reference, assigned to the relevant class, by circle. "Hotspots" red, "coldspots" blue hatched.

Note: Hatched regions together with the adjacent (partially unhatched) regions form a hotspot or coldspot. For example, the hatched district of Erding near Munich, together with the adjacent districts of Landshut, Freising, Mühldorf, Ebersberg and Munich, form a "Problem" coldspot.

4. Conclusion

The results of this short expertise show that the effects of the coronavirus pandemic on Germany's companies can be observed daily and with high sectoral and regional resolution by means of AI-based web analysis. The insights gained from this study show the dynamic course of the pandemic and the communicated reactions of the affected companies, which in part differ significantly from region to region. In particular, a strong east-west divide was found, both in terms of the general level of affectedness (references to the pandemic on the company website) and in terms of the type of affectedness. Companies from the federal states of Saxony-Anhalt, Brandenburg, and Mecklenburg-Western Pomerania, for example, tend to report that they have adapted to the current situation or that they are less affected, while companies in western federal states tend to report more serious problems. However, there are also differences within the individual federal states. In a nationwide comparison, companies from urban areas in particular are communicating about the pandemic, although our evaluation of the content here has not revealed any clear differences about the type of affliction. For a final evaluation of the results presented, however, possible distortions that could result from regionally different industry structures and company-specific characteristics must be investigated.

The approach presented here is a first step towards a daily monitoring of the German economy and could therefore be a useful tool in steering a coordinated and evidence-based response to the current crisis.

References

[1] J. Kinne and J. Axenbeck, “Web Mining of Firm Websites : A Framework for Web Scraping and a Pilot Study for Germany,” Mannheim, 18–033, 2018.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2018.

AUTHORS

Georg Licht
ZEW – Leibniz Centre for European Economic Research Mannheim GmbH
L 7, 1
68161 Mannheim
Germany
www.zew.de
georg.licht@zew.de

Jan Kinne*
ZEW – Leibniz Centre for European Economic Research Mannheim GmbH
L 7, 1
68161 Mannheim
Germany
Phone.: +49 (0)621 1235-297
jan.kinne@zew.de

*Contact person