IARIW 2022

Monday 22 - Friday 26 August

Covid resilience and digital readiness: An analysis using online company data

Oleksii Romanko
(King's College London, United Kingdom)

Mary O'Mahony
(King's College London, United Kingdom)

Paper prepared for the 37th IARIW General Conference

August 22-26, 2022

Session 4C-2, Proposal on Big Data: Economic Value, Digital Transformation, Productivity and Policy

Time: Wednesday, August 24, 2022 [16:00-17:30 CEST]

Covid resilience and digital readiness: An analysis using online company data

Oleksii Romanko and Mary O'Mahony

August 7, 2022

Abstract

This paper examines the extent to which prior digital readiness impacted on firms' resilience to the pandemic. The data are based on scraping the websites of a large number of UK firms monthly since May 2020. Textual analysis is used to divide firms into groups, based on their behaviour following the lockdowns, and these indicators are linked to the firms use of cloud services and their hiring of digital staff, as well as control variables. The paper finds that firms who invested in cloud services and hired digitally trained staff were more likely to be active after the lockdowns.

1 Introduction

The Covid-19 pandemic has had severe negative impacts on economic activity in general, but the extent of this is likely to vary enormously across firms, depending on the sector and region in which they operate. Impacts are also likely to depend on the firms' performance before the crisis, most importantly their financial position, but also their digital readiness. The aim of this research is to study the behaviour, of UK enterprises during the COVID-19 crisis. This is hampered by the lack of timely data on firm performance since the pandemic started. To address this we use innovative online sources of data in order to obtain more granular and faster indicators for private sector companies. Using web scraping and natural language processing tools, we create a companies resilience indicator dividing firms into those whose websites show some reaction to the crisis and those who were inactive. We also divide the former group by the extent of their activity, as explained below.

We use the word resilience to refer to the way companies cope, adapt and overcome Covid induced challenges. The main research aim is to evaluate if companies with better technical capabilities and technologically skilled human capital were better able to adapt and survive during and after the COVID-19 crisis. Our study aims to document the relationships between firm's characteristics, its pre-crisis digital intensity and its ability to be resilient.

Our study is based on online-generated datasets, using four sources of information. The first, our resilience measure is based on scraping the websites of approximately 130,000 UK firms monthly since May 2020. The next two are measures of digital readiness. One is an indicator of cloud technologies usage as a proxy to firm's technological readiness, using webscraping and meta-information domain name server registries (DNS). To this we add job vacancy data, also collected through web-scraping online job platforms (OJV) throughout 2019 for the UK, to study the occupational and skill patterns in hiring prior to the COVID-19 crisis, focusing in particular on digital skills. Finally we combine the above data with background information about businesses and their financial performance using the Financial Analysis Made Easy (FAME) dataset.

This paper begins with a short review of the literature. It then explains the methodology used to construct our resilience and digital readiness indicators and their merging with company accounts, including some descriptive statistics. We then set out the framework we will use to examine the relationship between digital readiness and resilience.

2 Literature review

Government restrictions during the COVID-19 lockdowns led to heterogeneous reactions of firms. Some firms reacted by taking some extraordinary and fundamental steps to sustain their revenue flows - introduce new business models, establish online delivery options, roll out new products or services etc. (Bai et al., 2020; Dingel and Neiman, 2020). Many companies went on pause, introduced salary cuts, sent their employees on furlough or went out of business. This includes many 'non essential' businesses such as cinemas and retailers (Baker et al., 2016).

The Business Impact of COVID-19 survey (ONS, BICS[1]) clearly showed that knowledge and IT intensive industries survived the first COVID-19 wave better. We hypothesize that it happened due to the better agility and faster ability to respond to changes, and better IT skills that enabled companies

[1] https://www.ons.gov.uk/economy/economicoutputandproductivity/output/datasets/businessimpactofcovid19surveybicsresult

to quickly switch to remote working procedures. Nunes and Lopes (2013) suggest that firms with established innovation processes and high economic dynamics are able to better survive crises. Jin and McElheran (2018) also evidence that usage of modern ICT technologies increases survival of firms. However innovation and R&D takes significant amount of time to lead to tangible results and there is a very high failure rate among innovative enterprises - around 90%. BeTheBusiness (BeTheBusiness) suggests that the COVID-19 pandemic had a heterogeneous effect on UK firms, driving technology adoption and innovation for some firms and delaying decision making for others.

The literature has shed light on how important different firm characteristics can contribute to firms' resilience during the crisis. First, firm's pre-crisis financial situation was shown to be positively correlated with their resilience. During economic crises firms can rapidly running out of cash, and a strong balance sheet was shown to help firms to be resilient (Ding et al., 2021). In the UK, ONS figures show that 35% of British single site businesses had cash reserves between 0-3 months across all industries by November 2020. This is quite heterogeneous across sectors with 55% of firms in accommodation and food services having cash reserves between 0-3 months, while for the wholesale and retail trade sector this was only 30%[2]. Buchheim et al. (2022) use a survey on German firms and find that not only does the pandemic amplify pre-crisis weaknesses but weaker firms appear to be harder hit initially. There are similar findings for China (Xiong et al., 2020) and for the US, where larger firms in the restaurant sector with more leverage and cash flows, were more resilient to stock declines (Song et al., 2021). The financial situation of firms allows them to cope with the costs, and money is available to invest in digital tools. Also, firms' characteristics such as their size in terms of number of employees or their age might be helpful in surviving. In contrast, Guo et al. (2020) find that smaller firms characteristics of moving fast have enabled them to react quickly to the situation. Other recent papers look at the contributing factors to firms' resilience, such as access to liquidity (Acharya and Steffen, 2020) and Bai et al. (2020) use a labour-related measure, i.e. flexible work arrangements within firms - see also Dingel and Neiman (2020).

However, another strand of the literature suggests that during economic crises, the uncertain economic environment leads firms to save more and not necessarily invest in digital transformations (Baker et al., 2016). Therefore, the digital situation of the firms pre-crisis could explain part of the resilience of firms. While already digitalised firms remain active during the pandemic, difficult

[2] Office for National Statistics – BICS wave 18

situations and closures increase in others. Consequently, the crisis may have been reinforcing the digital divide across workers, firms and industries.

There have been some attempts to use 'big data' sources to examine firms' response and their resilience to the crisis. Kinne et. al (2020) applied natural language processing techniques (supervised BERT model, as per Devlin et al. (2018)) to the web scraped pages of German websites in order to classify companies into different groups, depending on their reaction to the Covid crisis. Similar techniques were applied by Yang and Han (2020): they analysed Covid responses in the hospitality industry using the user-generated content on Twitter. Using unsupervised structural topic modelling (Roberts et al., 2013) examined adverse business reactions, mainly driven by the need to survive, adopt new technologies and new business strategies. Stephany et. al (2020) applied the Latent Dirichlet Allocation model to online risk assessment reports data in order to measure industry-specific risks due to Covid and assemble their "CoRisk-Index". To conclude, online text data and text mining methods have become increasingly useful for answering Covid-related economic questions. Most of the recent studies use a similar staged approach as in this paper, related to Automated Content Analysis methods (Hasbullah et al., 2016; Petchler and González-Bailon, 2015). Our methodology is explained in the next section.

Our aim is to contribute to the strand of literature that tries to understand why some firms have adapted their business and remained active while some have shut down and even potentially will exit the market when government supports stop. Why is the response to the crisis so heterogeneous? Archibugi et al. (2013) suggest two types of innovative behaviour which companies employ in order to survive during a crisis. The first one, following Schumpeterian-type models, is called 'technological acccumulation' and is commonly assigned to larger companies, which have resources and scale for delivering incremental innovative changes to their products or processes over the long run. Such companies tend to better survive crisis periods, and better adapt to changed circumstances. Another type of behaviour is called "creative destruction". It is inherent in small and medium size enterprises, who are able to produce fast and drastic changes to their business process in order to survive or enter new, more profitable market niches.

.

3 Methodology

3.1 Resilience indicator and automated content analysis

We build an indicator of the company resilience to the pandemic by web scraping the companies' website. It allows us to understand their response to the lockdown, changed economic environment or other important changes in the daily business routine through the analysis of companies' online posts or important updates on their websites. We call this type of analysis an Automated Field Study, to contrast with a more standard survey-based approach.

In order to collect a target dataset, we obtained a list of UK based companies from the Financial Analysis Made Easy (FAME) dataset. FAME covers the population of businesses in the UK and derives information from Companies House records. One of the benefits of using this dataset is an opportunity to collate companies' online behaviour with their financial status and other covariates such as employment, age and past business activity. The subset of companies, which have a website listed was taken from FAME since the availability of a website is a prerequisite for our web scraping exercise. We then matched these companies to an earlier exercise used to determine if firms used cloud services and for how long.

Using web scraping spiders built using python language, we collected the main page from companies' websites. We transformed and cleansed the information collected. We then use the automated content analysis methodology (Hasbullah et al., 2016; Petchler and González-Bailon, 2015). We apply an identification procedure and extract COVID-19-related keywords. Examples of keywords include common phrases used by companies to signal availability of their services to their customers (for example phrases like "continue to operate as usual", "continue to provide services", "office is open"), or to indicate the pause or closure of their business ('close our offices', 'closed due to COVID-19', 'activities are cancelled'), or do not show any activity on their website (no COVID-related words). There is a multiple step process for gathering COVID-related keywords: manual extraction of keywords, using unsupervised text modelling to search for new words, enhancing the set of keywords using Google Trends, classification of gathered COVID-19 words into categories, and building the prediction model to automatically classify unseen webpages into groups.

As a first step, we manually go through web scraped content in order to determine important keywords which appear on companies' web pages. We search for specific COVID-19-related sentences on the webpage, and collect them into a unified list of words and phrases that highlight the response to the COVID-19 disruption. Examples of phrases include those highlighting general information or concerns regarding the pandemic; statements highlighting resilient and active position; and phrases suggesting temporary or permanent closures or suspensions.

In the second step, we feed each of the previously found keywords into the Google Trends search engine in order to determine related topics, user queries and trending searches. In this way we extend the initial set of words to cover all related and important topics that could be mentioned on websites and we minimize the probability of missing an important mention on the website related to our topic of interest. Also, we expand clusters of popular words, phrases and user queries.

Inspired by the independent study of UK firms response to COVID-19 (BeTheBusiness, BeTheBusiness), we divide all companies into five groups: 'innovators' (firms that proactively react to changed circumstances, exercise innovation),'online' (firms that perform their business using online tools and solutions, e-commerce businesses or transformed firms who provide online purchase or delivery), 'stickers' (firms that choose to wait until things come back to normal or wait for better business circumstances, stick to the government guidelines or operate under reduced hours), 'informers' (firms that provide some information or news about pandemic, but their reaction is not clear) and 'inactive' (a group of firms who do not mention anything about COVID-19 disruption). We track whether companies suspend their business, reopen or experience any other changes which are communicated online by tracing updates on the main page of the website.

As a verification step, we take expert-determined sets of keywords and search for mentions of those keywords across text corpuses. We then manually verify results and keywords found in order to quality assure our keywords of interest, as well as verify the coverage and completeness of the keywords. This exercise assures that the keywords are sufficiently relevant to our topics of interest and we can fully express the true intent of the information on companies websites, using the keywords of interest. We then identify erroneous results, or missing keywords and feed them back into the keyword selection process. For every keyword, we match it to the web pages and go though a list of keywords and sentences where they appear in order to do a verification of correctness of the content in which these keywords are used. Then, we manually go through all keywords generated and separate them into four groups ('innovators','online','stickers, and 'informers'). We then based our group assignment decision on the prevalence of keywords found in one of these groups. If no keywords of interest are found on the web page, we assign the company into a fifth group, 'inactive'.

We built a manual labelling dashboard in order to go through a random sample of websites gathered. The dashboard is built using Python programming language and interactive HTML widgets (ipywidgets). The dashboard allows us to view both the historical snapshot and live version of the website, read the extracted text from the website, spot keywords of interest and count them, extract the sentences where the keyword is used, apply the automatic labelling model and manually assign the final group label, type of the website or input any relevant comments. Based on keywords found and expert judgment we do a manual labelling of websites into resilience cohorts.

During the manual labeling verification step we have studied the coverage issue and distribution of companies groups, based on the labelling insights. About 7% of manually screened websites are 'innovators', which mean that they took an active position during the pandemic, introduced new online services or continued their business with new security measures. About 15% of websites were classified as 'stickers', and have temporarily suspended their business or stopped providing some of services. Additionally, around 7% of the websites were impossible to reach at the time of our manual review, as they were sold or liquidated. This likely corresponds to the closure of the business.

More the a half of websites fall into the category of inactive. It is worth saying a few words about the inactive label, that was assigned to the companies' websites that did not contain any relevant COVID-related updates.Most of inactive websites constitute a simple 'online business card' type of the website, which is a simple form of the website, created with the intent of showcasing existence and services of a company. Thus, such websites are less likely to post timely updates about the business situation or communicate with clientele.

As a result of the labelling step, we have built a labelled dataset for around 2000 websites. We used this dataset to build an automated labelling process using several approaches explained in the Appendix E. This then enabled us to classify firms into the five groups.

3.2 Digital readiness indicators

Our research plans to combine the above measures with those relating to measures of technological readiness. Due to the complexity of digital readiness, we consider different dimensions as it is difficult to capture using a single indicator. Calvino et al. (2018) put forward two components in their measure of digital transformation of sectors: the technological component and the human capital component.

First, our measure of investment in ICT technologies is constructed using the firm level cloud usage built by Romanko (2021), using webscraping and meta-information domain name server registries (DNS) that provides information about companies' use of modern ICT technologies. The indicator is built by assigning 1 to companies that employ cloud-related vendors for their website hosting, and 0 otherwise. Note we know the date firms adopted these technologies so we can also use this information to construct a measure of the length of time that firms have been using cloud services prior to the pandemic. Cloud indices were parsed for 88% of companies from the original sample. Empty values for the cloud variable could appear for several reasons. For example, the website was not yet established at specific year or it was non-functioning at the time of dns parsing.

Second, our measure of digital human capital is constructed using the demand for IT specialists and workers with digital skills. Our measure relies on the digital skills firms have in their labour force. Some specific digital skills enable both to create the technologies and also to use them appropriately in order to make the firm digital ready. Using a natural language processing model we extracted skills mentioned inside job postings collected during several scraping cycles in 2019. We used a set of digital skills defined by Dice Skills Center [3], which broadly coicides with technical skills classification by O*NET[4]but has a larger set of unique technical skills defined (11 thousand versus 8 thousand in O*NET).The technical skills indicator equals one if the company advertised jobs containing technical skills, and zero otherwise. As a result, we create a firm-level measure of hiring digital skills during the pre-crisis period. Details of the approach are given in Romanko et al. (2022).

3.3 Control variables

The recent literature highlights the importance of including measures of the financial viability of a firm pre-crisis. We extracted a number of financial indicators using the Financial Analysis Made Easy (FAME) dataset. This derives information from web scraping companies reports filed to the Companies House in the UK. We use a subset of firms that have a website. Although this can potentially lead to sampling bias, there are strong incentives for companies to maintain their online

[3]https://insights.dice.com/employer-resource-center/introducing-dice-skills-center/

[4]https://www.onetonline.org/find/descriptor/browse/2.B/2.B.3

presence to keep in touch with customers by providing timely updates about their business, new products and services.

We remove public companies from the analysis (charitable organisations, companies limited by guarantee, public and public AIM, industrial/provident companies), as they have a different nature of business and business cycles, compared to private companies. As such, we proceed with the classical analysis of the privately owned enterprises. This adds more consistency and reduces heterogeneity with regard to companies' responses to COVID disruption.

Bakhshi and Mateos-Garcia (2012) suggest that the information about companies with less than 50 employees is very limited, and even many medium size firms have missing data - see Romanko (2022) for details on the reporting by companies in FAME. Although we would have liked to include control variables on profitability and productivity for our resilience cohorts, these variables are often missing. Therefore, we proceed with a limited subset of financial variables, such as fixed assets, total assets, shareholders funds as well as number of employees as a measure of firm size. As a result the final dataset consists of 34,964 firms. We also use FAME to add controls for age of the company and information on their industry and regional locations. Note that although FAME and the Cloud variable were easily matched, there were some issues matching with the skill data - details are discussed in Romanko (2022).

We also utilize the internet broadband speed data, collected and provided by Offcom [5], the UK regulator of broadband, TV, home phone and mobile services. In particular, Offcom provides reports on postcode level broadband speeds, as well as yearly reports about region-level broadband statistics. We incorporate 2017 postcode level statistics into our dataset. We match 85% of the data by exact postcode match and 15% by using aggregated average 3-digit postcode values. It is important to include this variable, as availability and speed of broadband connections is a prerequisite for many (if not most) cloud adoption cases.

4 Dataset and descriptives

The dataset contains 34,950 companies and spans across the years 2014-2021, with the resilience measures covering 2020-2021 and the other variables measured pre-pandemic. The dataset contains financial information about the firm (fixed assets, ratio of shareholders capital to total assets),

[5] https://www.ofcom.org.uk/

number of employees, company age, legal type, broadband connection speed, hiring indicators, and cloud usage information.

Our resilience measure (outcome variable) is a categorical variable which has 5 levels: innovator, online, sticker, informer, inactive. Each company's response to the COVID disruption is observed from March 2020 till January 2022 but we aggregate the measure using the maximum number of periods in our base estimations. For example, if company A was an inactive for 5 months and sticker for 7 months, the final aggregated status will be sticker. We also consider the dynamic pattern across the lockdown periods as an extension.

After finalizing the group assignment for all companies for every period, we obtained timeseries that show the dynamics of firms' behaviour (Figure 1). For illustrative purposes we do not include the inactive firms as these represent about 60% of our sample.

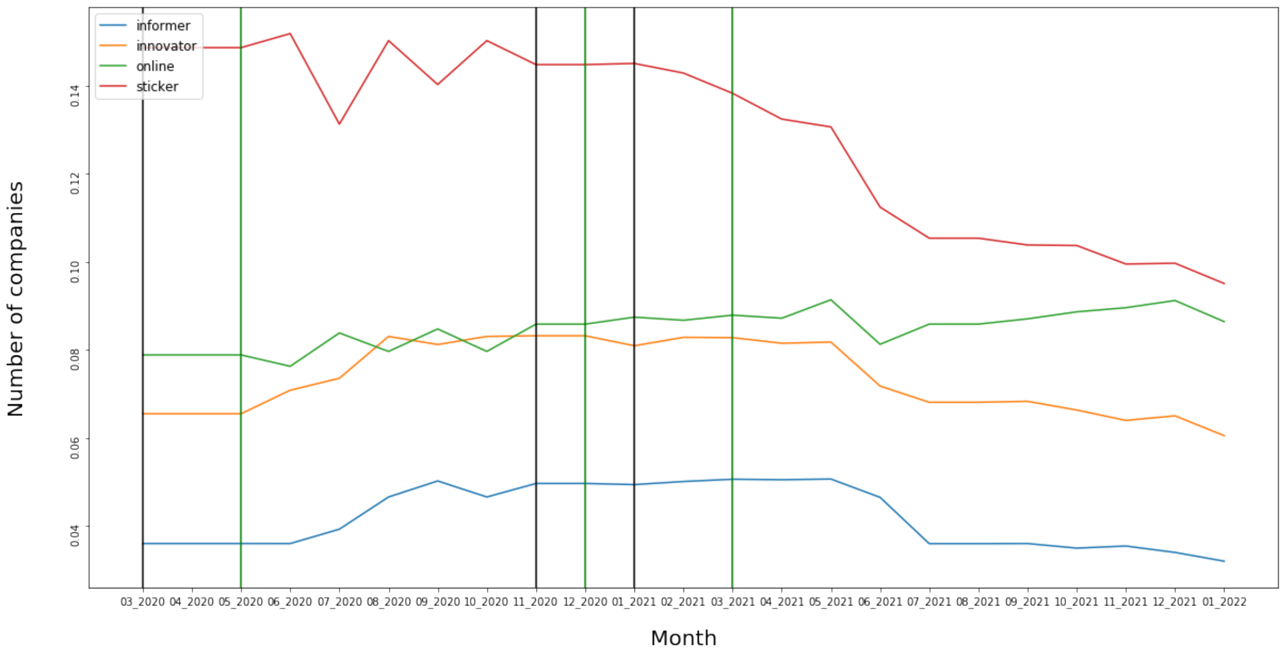

Figure 1: Time series chart for the UK enterprises from the following groups: innovators, informers, inactive. Years 2020-2021. Source: own compilation, lockdown timelines are taken from Institute for Government Analysis [6]

The black vertical lines represent dates when lockdown measures were implemented in the UK, and the green vertical lines show when restrictions were eased. We can see that after the first lockdown and subsequent easing of restrictions, the number of companies that signal activity increased. There is an increase in the percentage of online, innovator and informer companies after the first easing period. The number of online was steadily increasing throughout the whole period, and the number of stickers was rapidly decreasing during and after the second and third lockdown periods(from 14% to almost 9%). The number of informers and innovators were experiencing a slight lagged uplift during and after COVID lockdowns, which highlights the wave of response until every business got accommodated to the "new normal".

According to the UK Innovation Survey[7], there are were about 38% of innovation active companies in 2018. As we use a quite specific definition of innovator, percentages do not precisely match (however if we aggregate companies who were giving updates and informing the customers through the website, we would get to about 40% level). The wave of innovative and informative responses that initially increased but than gradually decreased, could be related to generally decreasing online consumer demand. It is worth noting though, that the post-COVID online shopping queries have almost doubled in comparison to the pre-COVID period.

We next show histograms for the characteristics of companies that have been classified into different cohorts (innovators, online, stickers, informers, inactive). The average percentage of companies that adopted cloud computing related technologies is generally higher for online and innovator groups (Figure 2). There are also relatively high cloud adoption rates in the stickers group and could be due to the fact that this group have the highest amount of fixed capital in our sample, which is correlated with cloud adoption.

[7]https://www.gov.uk/government/statistics/uk-innovation-survey-2019-main-report

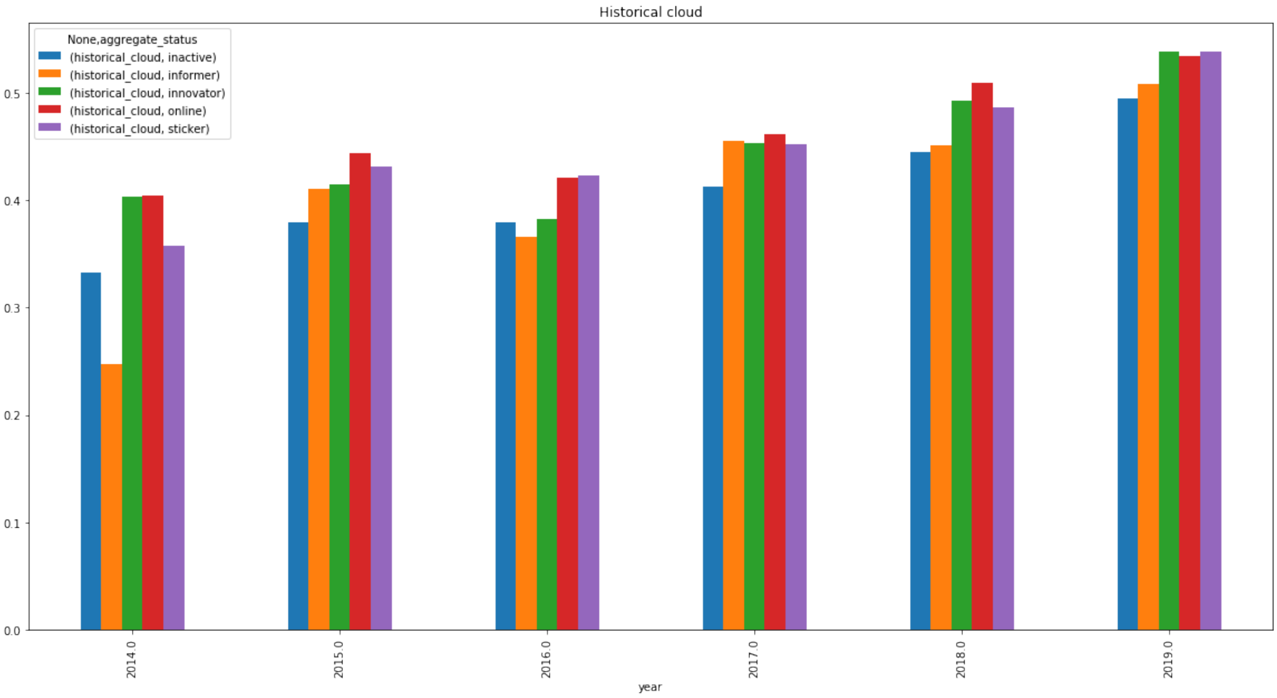

Figure 2: Average cloud adoption rates by resilience status, yearly

Firms that hire more technological workers are more likely to fall into informer, innovator or sticker groups (Figure 3). Surprisingly the online group shows relatively low rates on average.

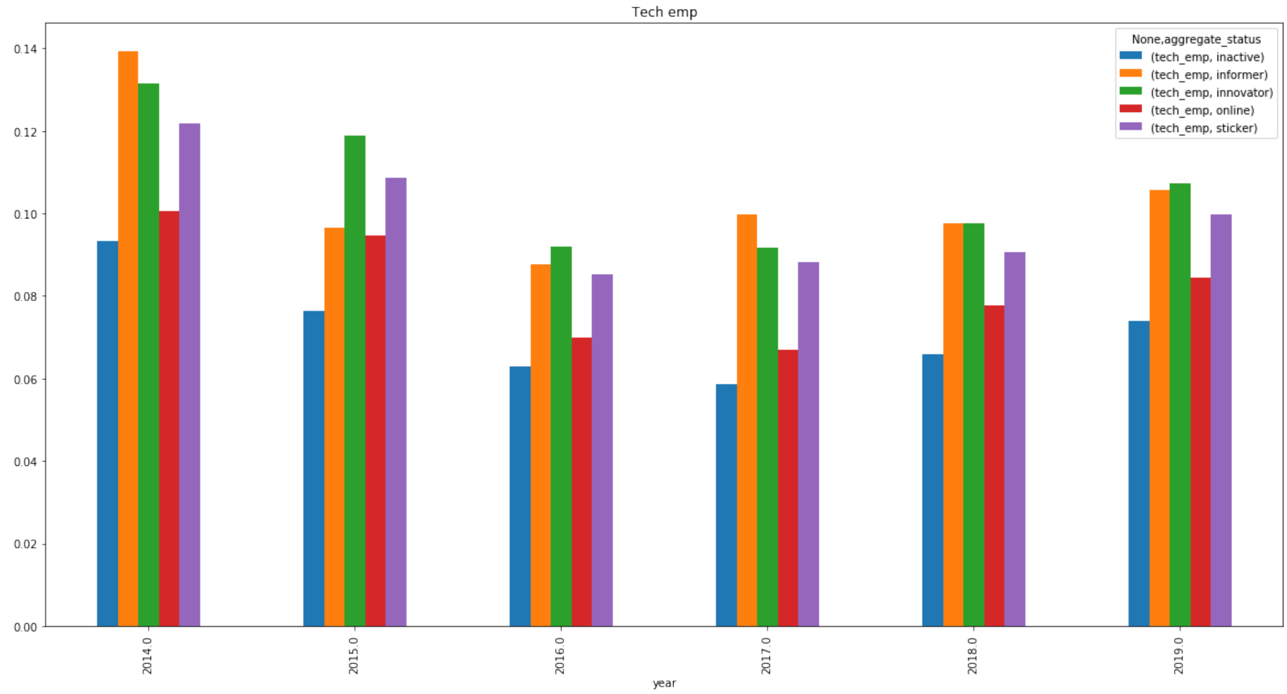

Figure 3: Average tech employment rates by resilience status, yearly

Descriptive statistics for the variables used in our study are show in Table 1.

Table 1: Statistical properties of variables used in analysis

Stats	Description	N	Mean	p50	SD	p5	p95
h	years of cloud experience	34,950	0.20	0.00	0.65	0.00	1.00
c	binary indicator, if company uses cloud (1) or not (0)	34,950	0.32	0.00	0.47	0.00	1.00
d	binary indicator, if company advertises technical (digital) jobs (1) or not (0)	34,950	0.02	0.00	0.16	0.00	0.00
r	digital readiness binary indicator, if company advertises technical jobs and uses cloud technologies (1) or not (0)	34,950	0.02	0.00	0.14	0.00	0.00
k	fixed assets per employee, 5 years average, logged	34,950	2.27	2.07	1.52	0.22	4.87
l	number of employees, 5 years average, logged	34,950	3.17	3.00	1.45	1.10	5.77
f	average internet download speed (Mb/s), logged	34,950	3.41	3.56	0.71	1.95	4.29
g	company age	34,950	17.28	14.00	13.52	4.00	42.00
s	shareholders funds to total assets ratio, 5 years average	50048	-8.51	0.38	1068.64	-0.67	0.87
postcode_1l	first letter of the postcode	34,950	-	-	-	-	-
sic07_2d	2 digit sic code	34,950	-	-	-	-	-
leg_form	legal form of the company (ltd, llp, public, unircorporated, etc.)	34,950	-	-	-	-	-

5 Modelling resilience outcomes

5.1 Estimation framework

Our main hypothesis is that, conditional on financial viability and other controls, technological readiness is associated with greater ability of the firm to proactively react to the pandemic. Specifically, we regress a firm's resilience indicator on digital readiness indicators and a number of control variables.

For the simplicity of interpretation we model the outcome using the logistic regression for every individual level (one versus the remainder regressions). Later we examine multinomial regressions. The estimation model is given by:

$$y_{ik} = a_0 + a_1 h_{ik} + a_2 c_{ik} + a_3 d_{ik} + a_4 c * d_{ik} + a_5 k_{ik} + a_6 l_{ik} + a_7 f_{ik} + a_8 g_{ik} + a_9 s_{ik} + a_1 0 \sigma_i + a_1 1 \delta_i + e_{ik}$$

where y_{ik} is a dichotonomous indicator for every status group k (innovator, online, sticker, informer, inactive) for every company i, h_{ik} is a cloud experience measured in years, $_{ik}$ - the cloud usage dummy variable, d_{ik} - indicator of technical(digital) hiring, $c * d_{ik}$ - digital readiness indicator, interaction between cloud usage and technical hiring, k_{ik}- company fixed assets (5 years average,in log values), l_{ik} - number of employees in company (in log values), f_{ik} - broadband speed (download speed) in the area (in log values, 2017 measure), g_{ik} - company age (in years since registration date), s_{ik} - shareholders funds divided by total assets of the company (average over 5 years), σ_i - regional dummies, δ_i - industry dummies, e_{ik} - error term.

5.2 Basic Results

Table 2 shows the first set of results. We can see that years of cloud experience significantly increases the probability of being an innovator, online, or sticker while being insignificant for the informers and negative for inactive. Companies that used cloud technologies have higher probability of being out of the inactive group. The high positive and significant coefficient for the innovators group supports our hypothesis that cloud usage goes hand in hand with higher innovation. These results align with our expectations as the usage of cloud technologies increases productivity and decreases risks of default as explained in Romanko (2022).

We also include dummies for companies that use cloud and hire technical employees. We can see that companies who just hire employees with technical skills are less prone to be inactive, however have higher probability to be stickers. This fact aligns with the effect of number of employees on the company status - the more employees the company have, the harder it is to switch and adjust to changing circumstances. Thus, large firms are more eager to stay aligned with government guidelines, layoff employees or put them on furlough.

Finally, we can see that companies that both hire employees with technical capabilities and use cloud technologies are able to achieve higher and statistically significant uplift in the probability of being an innovator, while reducing the odds of being inactive.

We control for company's capital and labour. As the regressions show, firms with higher fixed capital have significantly higher chances to become stickers. We also noticed that older companies are less likely to be inactive, although the coefficient is insignificant. Companies with higher share of own capital are also more likely to be stickers or innovators.

Table 2: Resilience modelling (ordinary logit)

	if_innovator	if_online	if_sticker	if_informer	if_inactive
h	0.038*	0.051**	0.049**	-0.021	-0.026
	(0.019)	(0.019)	(0.018)	(0.026)	(0.018)
c	0.248***	0.243***	0.233***	0.125**	0.024
	(0.031)	(0.031)	(0.029)	(0.044)	(0.029)
d	0.111	0.003	0.112	0.079	-0.182*
	(0.081)	(0.089)	(0.074)	(0.107)	(0.073)
c*d	0.202**	0.079	0.191**	0.069	-0.161*
	(0.077)	(0.084)	(0.073)	(0.105)	(0.072)
k	0.008	0.012	0.034***	0.022	0.006
	(0.010)	(0.010)	(0.009)	(0.013)	(0.009)
l	0.241***	0.142***	0.237***	0.181***	-0.115***
	(0.011)	(0.011)	(0.010)	(0.015)	(0.010)
f	0.023	0.022	0.026	0.013	-0.007
	(0.019)	(0.019)	(0.018)	(0.027)	(0.018)
g	0.001	0.000	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
s	0.012	0.002	0.012	0.004	-0.004
	(0.011)	(0.004)	(0.009)	(0.010)	(0.004)
constant	-2.446***	-1.299***	-2.089***	-2.576***	1.985***
	(0.226)	(0.210)	(0.211)	(0.296)	(0.207)
nobs	34945	34959	34960	34905	34964
R-sqr	0.054	0.067	0.071	0.055	0.029
BIC	34580	34191	39712	21236	39569

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

However, as we can see from looking at the constant term, they are different for each regression.

All constant coefficients are negative, except for the coefficient for the inactive group. This is the result of different distributions of zeros and ones in the groups. Since the inactive group is the most prevalent, logistic regression for this group has another baseline, as it is taken into account that being 'inactive' is the most probable outcome by default. On top of that, all regressions presented model the outcome as 'one versus the rest', thus, coefficients display the aggregate probability of being in the specific group and does not help to see the relative differences between the groups.

Due to these facts, we cannot compare the coefficients between groups to assess whether a given regressor has a higher impact on the specific outcome. For this purpose, we need to utilize a multinomial logit regression, which is the model typically used for modelling multiple outcomes and the modelling is performed simultaneously, provides more efficient estimates and lower standard errors (Agresti, 2002). Discussion of the multinominal model is presented in the Appendix A.

The next table presents the results of the multinomial logit regression performed on categorized outcomes and all previously used explanatory variables (Table 3).

Table 3: Resilience modelling (multinomial logit)

	inactive	informer	innovator	online	sticker
h	0.000	-0.066	0.000	0.041	0.061**
	(.)	(0.043)	(0.028)	(0.026)	(0.021)
c	0.000	0.064	0.235***	0.215***	-0.059
	(.)	(0.066)	(0.045)	(0.041)	(0.034)
d	0.000	0.082	0.311**	-0.072	0.169*
	(.)	(0.164)	(0.110)	(0.120)	(0.084)
c*d	0.000	0.203	0.285*	0.136	0.151
	(.)	(0.158)	(0.111)	(0.111)	(0.084)
k	0.000	0.007	-0.011	-0.008	0.003
	(.)	(0.020)	(0.014)	(0.013)	(0.010)
l	0.000	0.243***	0.232***	0.097***	0.162***
	(.)	(0.022)	(0.015)	(0.015)	(0.012)
f	0.000	0.066	-0.014	0.064**	-0.027
	(.)	(0.040)	(0.028)	(0.025)	(0.021)
g	0.000	-0.003	-0.001	0.005***	-0.003*
	(.)	(0.002)	(0.001)	(0.001)	(0.001)
s	0.000	0.004	0.014	0.000	0.005
	(.)	(0.014)	(0.017)	(0.003)	(0.007)
constant	0.000	-2.472***	-2.400***	-2.577***	-0.969***
	(.)	(0.199)	(0.148)	(0.164)	(0.106)
nobs	34951				
R-sqr	0.016				
BIC	81871				

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

As the inactive group is taken as a baseline for comparison, we can compare results between the groups. Table 3 provides useful insights on the difference in coefficients. We see that usage of the cloud significantly increases the odds of being an innovator or online while hiring activities alone increase chances for company to become either innovators or stickers. Yet, the combination of hiring of the technologically skilled personnel and cloud activities have a significantly positive impact on chances of being an innovator. Of the control variables, size of company, measured by number of employees, has a significant effect for all four groups. Older companies were better able to respond to the crisis by being 'online' and less likely to be stickers.

5.3 Estimations by company size and industry

In this section we present results obtained by dividing by firm size and industry. The results suggest that these division highlight heterogeneity across firms and provide more granular insights (Tables 4 and 5). More detailed tables of results along with all controls are shown in the Appendix B.

Cloud technologies increase the probability of being an innovator across all firm size bands, while years of cloud experience significantly matter for small and medium enterprises (Table 5). Combined hiring and technological vigour becomes the factor of huge significance for medium and large firms. Monetary endowment also appears to play an important role for large firms. Amount of employees is an important determinant of reaction, as it was highlighted before.

Table 4: Resilience modelling by firm size

	micro	*small*	*medium*	*large*
if_innovator				
h	0.042	0.079*	0.076*	−0.015
c	0.266***	0.241***	0.192**	0.261*
d	−0.262	0.205	0.107	0.161
c*d	0.181	0.048	0.306*	0.411*
if_online				
h	−0.052	0.071*	0.086*	0.086*
c	0.348***	0.210***	0.202**	0.136
d	−0.412	0.014	−0.041	0.278
c*d	−0.007	−0.060	0.083	0.173
if_sticker				
h	0.075	0.084**	0.046	0.003
c	0.260***	0.169***	0.253***	0.344***
d	0.016	0.301*	−0.097	0.103
c*d	0.357	0.335**	0.191	0.145
if_informer				
h	−0.120	0.010	−0.005	−0.029
c	0.267**	0.045	0.114	0.130
d	0.138	0.044	0.152	0.142
c*d	0.051	0.238	−0.024	0.182
if_inactive				
h	−0.075	0.004	−0.031	0.012
c	0.121*	0.012	−0.027	−0.103
d	0.038	−0.288*	−0.300*	0.048
c*d	0.627*	−0.232	−0.297*	−0.261

The distinctive difference in the 'innovators' group is a significantly positive impact of combination of labour and technologies for medium and large firms. We can see no significant impact of cloud or tech skills on the probability of being an informer (except cloud for micro firms). These results seem to be similar to the 'stickers' group. Finally, we obtain important results for the 'inactive' group. While cloud usage seems not to be a relevant factor, tech hiring and combination of hiring and technologies gives lower probability of being 'inactive' for small and medium firms.

From the exercises completed above it seems that firms' reaction to the pandemic was primarily going along the style which Schumpeter explained as continuous innovation. Firms of larger size had better chances of surviving the crisis. A combination of hiring activities and usage of modern technologies were displaying the highest benefit for medium and large enterprises. Cloud experience increased the chance of being an innovator, online or sticker, while decreasing the probability of being inactive. The amount of experience using the cloud was found to be significant for small and medium sized companies. In comparison, the cloud usage indicator was found to be significant for most companies sizes, showing a positive effect on probabilities of being in four categories but a negative contribution to odds of being inactive (while being not significant).

Our final set of results covers logit regressions within different industries (Table 4). We divide all industrial activities into several groups: manufacturing, mining, fishing and agriculture (SIC codes 1 to 33, we name the group "Manufacturing"); retail, food and accommodation (SIC codes 45 to 47, 55 to 56, "Retail"); ICT and professional services (SIC 58 to 84, "ICT"); education and health (85 to 88, "Education and Healthcare"); construction and supplies (35 to 43, "Construction"); arts and entertainment (SIC 90 to 99, "Entertainment"). In the Appendix C we show results for each industry and each resilience group.

Table 5: Resilience modelling by industry: manufacturing

	(1)	(2)	(3)	(4)	(5)
	if_innovator	if_online	if_sticker	if_informer	if_inactive
Manufacturing					
h	0.033	0.098**	0.024	−0.050	0.017
c	0.120*	0.232***	0.107*	−0.000	−0.039
d	−0.690***	0.148	0.269*	−0.380*	0.356**
c*d	−0.191	0.477***	0.116	0.161	−0.026
Retail					
h	−0.067*	0.053	0.023	0.045	−0.052
c	0.242***	0.264***	0.250***	−0.025	−0.014
d	0.229*	0.209*	−0.198*	0.547***	−0.429***
c*d	0.075	0.499***	−0.421**	0.407*	−0.244*
ICT					
h	0.028	0.001	0.029	−0.060*	−0.016
c	0.193***	0.162***	0.197***	0.156**	−0.071*
d	0.024	0.298***	0.050	0.054	0.174*
c*d	0.431***	0.370***	0.044	−0.135	0.192*
Education and Healthcare					
h	−0.009	0.025	0.071*	−0.036	0.033
c	0.109	0.300***	0.141**	0.193**	0.019
d	0.186*	0.087	0.153	0.253*	−0.301***
c*d	0.151	0.630***	0.086	0.250*	−0.129
Entertainment					
h	0.072	−0.017	0.088*	−0.029	−0.051
c	0.097	0.276***	0.123	0.015	0.236**
d	0.258*	0.091	−0.155	0.012	−0.032
c*d	0.092	−0.085	0.206	0.142	0.292
Construction					
h	0.060	−0.015	−0.034	−0.205*	−0.089
c	0.165	0.128	0.209**	−0.024	0.160*
d	0.024	−0.270	0.140	0.243	−0.229
c*d	−0.005	−0.149	0.495**	0.254	−0.074

The results are more heterogeneous by industry group. Nevertheless, in most sectors the probabilities of being an innovator or online are positively related, and the probability of being inactive negatively related, to the digital readiness indicators. There are some exception to these general patterns. For example, in the manufacturing group hiring technical specialists decreases odds of being an innovator but increases chances for belonging to stickers or inactive group and usage of both hiring and technologies only significantly increases the chances of being online.

5.4 The dynamic impact of the pandemic

As another exercise, we divided the general resilience outcome into 5 outcomes aggregated separately by the periods of lockdown: 1st lockdown (Mar2020-June 2020 - 1lk), lockdown easing (June 2020 – September 2020 – 1rel), 2nd lockdown (November 2020 -2lk), lockdown easing (December 2020 -2rel), 3rd lockdown (January 2021- March 2021 - 3lk), and final periods (April 2021 – January 2022-3rel). We run simple logit regressions and estimate effects per every lockdown and per every period of the easing of restrictions (see Table 6). More detailed estimations are provided in Appendix D.

Table 6: Modelling resilience outcome during every lockdown and lockdown easing period

	1lk	1rel	2lk	2rel	3lk	3rel
innovators						
h	-0.024	0.007	0.009	0.009	-0.021	-0.007
c	0.300***	0.208***	0.332*	0.332*	0.240***	0.232***
d	0.048	0.082	-0.017	-0.017	0.098	0.117
c*d	0.224	0.164	0.444	0.444	0.205	0.248*
online						
h	0.079*	0.065*	-0.093	-0.093	0.064*	-0.004
c	0.154*	0.087	0.346*	0.346*	0.152*	0.216***
d	0.129	0.147	0.250	0.250	-0.010	-0.009
c*d	0.016	-0.131	-0.289	-0.289	0.178	0.352**
stickers						
h	0.060**	0.038	0.078*	0.078*	0.047*	0.022
c	-0.150***	-0.061	-0.058	-0.058	-0.069	-0.191***
d	0.011	-0.010	-0.024	-0.024	0.026	-0.037
c*d	-0.034	-0.010	0.091	0.091	0.011	-0.099
informers						
h	-0.020	-0.097*	0.006	0.006	-0.018	-0.065
c	0.244*	-0.021	-0.078	-0.078	-0.037	-0.038
d	0.160	-0.134	0.583	0.583	-0.138	-0.026
c*d	0.056	0.079	-0.122	-0.122	0.022	0.021
inactive						
h	-0.077***	-0.050*	-0.111**	-0.111**	-0.063**	-0.001
c	-0.035	-0.041	-0.036	-0.036	-0.064	-0.027
d	-0.078	-0.035	-0.094	-0.094	-0.023	-0.011
c*d	-0.062	-0.041	-0.185	-0.185	-0.176*	-0.175*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry

We can see that additional years of cloud experience have a significantly positive effect on probability of being online during the first and 3rd lockdown. It is likely that the short period of the 2nd lockdown and the following easing is the reason why these periods' estimates are mostly insignificant. Hiring only activities bring positive effect to the odds of being online but the effect is not statistically significant. All other coefficients are within the expected magnitude and sign. There is no observable significant differences in the effect through the pandemic, however we might notice that the mix of the cloud technologies usage and hiring of technologically skilled employees helps to gain significant and positive effect during the 3rd lockdown and the following periods of easing of the restrictions.

We witness that years of cloud experience per is not associated with being an innovator. However, the fact that companies use cloud technology provides a significant uplift to their chances of being in this category. Usage of technologically skilled workforce does provide a positive effect, however we failed to find the statistical significance. Using both technological and skilled force potential provides positive effect and we find it to be significant in the long run (in periods of easing after the 3rd lockdown).

The table 6 shows a negative impact of all cloud and hiring related variables on the probability of being inactive, providing some evidence that access to technology and labour helps to decrease inactivity and risk of suspending the business. In particular the interactive term is significantly negative in the final period for this group, suggesting that firms that were not technologically ready before the pandemic had a much higher probability of not being active by the end of the pandemic.

6 Conclusion

In this paper we use new sources of web-scraped data to investigate if firms' resilience to the Covid pandemic can be linked to digital readiness before the crisis. Our results suggest that the use of cloud services, hiring of technically skilled workers, or both, contributes to explaining the probability of actively responding to the crisis versus being inactive. Our rich data allows us to divide the former into a number of groups. The results are suggestive that the most innovative firms benefited from being digitally ready, as did firms whose business was primarily online. The dynamic analysis is suggestive that the main differences between the active and inactive groups manifest in the longer term, from the third lockdown. Results vary by firm size and industry, with small and medium firms and those in the ICT professionals and retail and hospitality sectors closest to the average.

The results shown, based on logit regressions, can only be seen as suggesting an association due to possible endogeneity and selection biases. Although our research design, that examines the impact of pre-crisis investments on post-crisis behaviours, excludes reverse causality as a source of endogeneity bias, other sources may still persist such as omitted variable bias. We have tried to address this by including many relevant controls as well as region and industry dummies. The many choices we had to make in extracting data from websites as well as combining and merging data might have resulted in some sample selection - e.g. very small firms are not available in FAME and the personal services sector is underrepresented. Therefore we need to be cautious in our conclusions. Nevertheless, using 'big data' sources did allow us to consider a larger sample of firms than has been possible to date using survey data. And by grouping firms according to their post Covid behaviours, we free ourselves from the reliance on measures of performance such as productivity and profitability that tend to be only available for the largest and most productive firms in company accounts databases.

References

Acharya, V. V. and S. Steffen (2020). The risk of being a fallen angel and the corporate dash for cash in the midst of covid. The Review of Corporate Finance Studies 9(3), 430–471.

Agresti, A. (2002). Categorical data analysis. john wiley & sons. Inc., Publication 15, 273.

Archibugi, D., A. Filippetti, and M. Frenz (2013). The impact of the economic crisis on innovation: Evidence from europe. Technological Forecasting and Social Change 80(7), 1247–1260.

Bai, J., E. Brynjolfsson, W. Jin, S. Steffen, and C. Wan (2020). The future of work: Work from home preparedness and firm resilience.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. The quarterly journal of economics 131(4), 1593–1636.

Bakhshi, H. and J. Mateos-Garcia (2012). Rise of the datavores: How uk businesses analyse and use online data. 19(6).

Bayaga, A. (2010). Multinomial logistic regression: Usage and application in risk analysis. Journal of applied quantitative methods 5(2).

Benson, A. R., R. Kumar, and A. Tomkins (2016). On the relevance of irrelevant alternatives. In Proceedings of the 25th International Conference on World Wide Web, pp. 963–973.

BeTheBusiness. A third; a third; a third: Different firm positions on recovery.

Buchheim, L., C. Krolage, and S. Link (2022). Sudden stop: When did firms anticipate the potential consequences of covid-19? German Economic Review 23(1), 79–119.

Calvino, F., C. Criscuolo, L. Marcolin, and M. Squicciarini (2018). A taxonomy of digital intensive sectors.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ding, W., R. Levine, C. Lin, and W. Xie (2021). Corporate immunity to the covid-19 pandemic. Journal of Financial Economics 141(2), 802–830.

Dingel, J. I. and B. Neiman (2020). How many jobs can be done at home? Journal of Public Economics 189, 104235.

Guo, H., Z. Yang, R. Huang, and A. Guo (2020). The digitalization and public crisis responses of small and medium enterprises: Implications from a covid-19 survey. Frontiers of Business Research in China 14(1), 1–25.

Hasbullah, S. S., D. Maynard, R. Z. W. Chik, F. Mohd, and M. Noor (2016). Automated content analysis: A sentiment analysis on malaysian government social media. In Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication, pp. 1–6.

Jin, W. and K. McElheran (2018, 01). Economies before scale: Survival and performance of young plants in the age of cloud computing. SSRN Electronic Journal.

Nunes, S. and R. Lopes (2013). Economic crisis and the firms innovation process.

Petchler, R. and S. González-Bailon (2015). Automated content analysis of online political communication. In Handbook of digital politics. Edward Elgar Publishing.

Roberts, M. E., B. M. Stewart, D. Tingley, E. M. Airoldi, et al. (2013). The structural topic model and applied social science. In Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, Volume 4, pp. 1–20. Harrahs and Harveys, Lake Tahoe.

Romanko, O. (2021, Aug). Digital opportunity: How cloud computing changes the shape of the uk economy. 36th IARIW Virtual General Conference proceedings.

Romanko, O., M. O'Mahony, et al. (2022). The use of online job sites for measuring skills and labour market trends: A review. Economic Statistics Centre of Excellence (ESCoE) Technical Reports (ESCOE-TR-19).

Song, H. J., J. Yeon, and S. Lee (2021). Impact of the covid-19 pandemic: Evidence from the us restaurant industry. International Journal of Hospitality Management 92, 102702.

Xiong, H., Z. Wu, F. Hou, and J. Zhang (2020). Which firm-specific characteristics affect the market reaction of chinese listed companies to the covid-19 pandemic? Emerging Markets Finance and Trade 56(10), 2231–2242.

A Multinomial logistic regression

. The polychotomous(multinomial) logistic regresson is designed to model the data with more than two categorical outcomes with no natural ordering. The model is an extension of the classical logit regressions framework.

The probability of the company to pursue one of the covid response strategies is modelled as:

$$Prob(Y_i = j) = \frac{\exp \beta_j x_i}{\sum_{k=0}^{4} \exp \beta_k x_i}$$

, where X is a vector of known company individual characteristics, β is a vector of weights, x is a vector of model inputs.

Due to the identification problem (any $B^* = B + q$ will result in the same probability, as q will cancel out in numerator and denominator), we set $B_0 = 0$ as a baseline for comparison (This is ok as probabilities sum up to 1 so we only need J parameter vectors to find J+1 probabilities). This can be shown by careful calculations. If we are to model individual logistic regressions (A,B,C) for every class with comparison to a baseline class (D), than standard binary logistic regression formula will be

$$log(P(A)/P(D) = b_{01} + b_{11}x_1 + ..b_{n1}x_n$$

or

$$\begin{cases} P(A) = P(D) \exp^{b_{10}+b_{11}x_1+..b_{n1}x_n} \\ P(B) = P(D) \exp^{b_{20}+b_{21}x_1+..b_{2n}x_n} \\ P(C) = P(D) \exp^{b_{30}+b_{31}x_1+..b_{3n}x_n} \end{cases} \tag{1}$$

Given that

$$P(A) + P(B) + P(C) + P(D) = 1$$

then

$$p(D) \exp^{b_{01}+b_{11}x_1+..b_{n1}x_n} + P(D) \exp^{b_{20}+b_{21}x_1+..b_{2n}x_n} + P(D) \exp^{b_{30}+b_{31}x1+..b_{3n}x_n} + P(D) = 1$$

and

$$P(D) = \frac{1}{1 + \exp^{b_{01}+b_{11}x_1+..b_{n1}x_n} + \exp^{b_{20}+b_{21}x1+..b_{2n}} + \exp^{b_{30}+b_{31}x_1+..b_{3n}x_n}}$$

So the final formula is:

$$Prob(Y_i = j|x_i) = \begin{cases} \frac{1}{1+\sum_{k=1}^{4} \exp \beta_k x_i}, if j = 1 \\ \frac{\exp \beta_j x_i}{1+\sum_{k=1}^{4} \exp \beta_k x_i}, if j > 1 \end{cases} \tag{2}$$

where β is a vector of model parameters for model k and x is a vector of explanatory variables.

In order to be able to compare groups between each other, we set the largest group as a baseline for comparison. We set inactive group as the baseline for comparison. We can arrive at relative risk ratio[8] as

$$\frac{Pr(y = j)}{Pr(y = 1)} = \exp \beta_2 x$$

or

$$log(\frac{Pr(y = j)}{y = 1}) = \beta_2 x$$

Note that $Pr(y = 1)$ is a baseline probability. We would note that the multinomial logit model allows direct comparison of the probabilities of different outcomes:

$$\frac{Pr(y = j)}{Pr(y = 1)} = \exp^{\beta_j x},$$

$$\frac{Pr(y = k)}{Pr(y = 1)} = \exp^{\beta_k x},$$

so

$$\frac{Pr(y = k)}{Pr(y = j)} = \exp^{(\beta_k - \beta_j)x}$$

, which means that we can directly compare coefficients between the groups (although one need to still account for the intercept which is different in every model, see Bayaga (2010)). Other benefits of the model are:

- the model is robust to the violation of multivariate normality and similar covariance matrices in every group
- model statistics are better interpretable
- multiple linear regression neither assumes linear relationship between target and explanatory variables nor does it assume normality of error terms citepbayaga2010multinomial

[8]https://www.stata.com/manuals/rmlogit.pdf

Although, model still relies on some of the assumptions, there are a few simple steps to verify them. For instance, there is an assumption called independence from irrelevant alternatives - odds ratios of the two selected outcomes of comparison are independent from all other alternatives (in other words, outcomes are free from unobserved impacts, see(Benson et al., 2016)). This assumption can be verified using the Hausman-McFadden test. The Hausman-Mcfadden test suggests that the IIA assumption is not violated in the results presented above. (Table 7).

Table 7: Hausman-Mcfadden test for multinomial logit model

	chi2	df	P>chi2
inactive	0.915	30	1.000
informer	3.648	30	1.000
innovator	0.892	30	1.000
online	8.893	30	1.000
sticker	12.814	30	0.997

B Detailed estimations by company size

Table 8: Resilience modelling by firm size: innovator

	micro	small	medium	large
if_innovator				
h	0.042	0.079*	0.076*	-0.015
	(0.056)	(0.034)	(0.033)	(0.040)
c	0.266***	0.241***	0.192**	0.261*
	(0.065)	(0.048)	(0.066)	(0.110)
d	-0.262	0.205	0.107	0.161
	(0.257)	(0.131)	(0.148)	(0.193)
c*d	0.181	0.048	0.306*	0.411*
	(0.262)	(0.149)	(0.133)	(0.165)
k	0.018	0.010	-0.024	0.091**
	(0.020)	(0.016)	(0.022)	(0.031)
l	0.470***	0.220***	0.095	0.244***
	(0.061)	(0.048)	(0.062)	(0.031)
f	-0.053	0.000	-0.052	0.057
	(0.043)	(0.030)	(0.039)	(0.061)
g	-0.005	-0.000	0.002	0.003
	(0.003)	(0.002)	(0.002)	(0.002)
s	0.009	-0.007	0.034	0.098
	(0.013)	(0.020)	(0.047)	(0.084)
constant	-3.381***	-1.734***	-1.991***	-3.460***
	(0.808)	(0.402)	(0.486)	(0.589)
nobs	10122	14750	7045	2895
R-sqr	0.060	0.046	0.037	0.082
BIC	8676	15126	8664	4128

Table 9: Resilience modelling by firm size: online

	micro	small	medium	large
if_online				
h	-0.052	0.071*	0.086*	0.086*
	(0.054)	(0.035)	(0.035)	(0.042)
c	0.348***	0.210***	0.202**	0.136
	(0.061)	(0.048)	(0.069)	(0.116)
d	-0.412	0.014	-0.041	0.278
	(0.267)	(0.147)	(0.167)	(0.209)
c*d	-0.007	-0.060	0.083	0.173
	(0.267)	(0.161)	(0.150)	(0.181)
k	0.005	0.011	0.018	0.061
	(0.019)	(0.017)	(0.023)	(0.032)
l	0.304***	0.162***	-0.000	0.098**
	(0.056)	(0.048)	(0.066)	(0.030)
f	0.027	-0.031	-0.049	-0.045
	(0.040)	(0.030)	(0.041)	(0.063)
g	-0.003	-0.003	0.001	0.003
	(0.003)	(0.002)	(0.002)	(0.002)
s	0.002	0.033	-0.028	-0.073
	(0.004)	(0.027)	(0.047)	(0.047)
constant	-1.300*	-0.754*	-0.888	-1.993**
	(0.557)	(0.363)	(0.502)	(0.614)
nobs	10129	14758	7073	2918
R-sqr	0.062	0.071	0.082	0.096
BIC	9529	14908	8015	3873

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Resilience modelling by firm size: sticker

	micro	small	medium	large
if_sticker				
h	0.075	0.084**	0.046	0.003
	(0.048)	(0.032)	(0.032)	(0.039)
c	0.260***	0.169***	0.253***	0.344***
	(0.057)	(0.044)	(0.062)	(0.104)
d	0.016	0.301*	-0.097	0.103
	(0.204)	(0.117)	(0.141)	(0.183)
c*d	0.357	0.335**	0.191	0.145
	(0.220)	(0.128)	(0.128)	(0.164)
k	0.052**	0.006	0.049*	0.071*
	(0.017)	(0.015)	(0.020)	(0.029)
l	0.451***	0.220***	0.119*	0.253***
	(0.053)	(0.044)	(0.058)	(0.028)
f	-0.081*	-0.014	0.001	-0.021
	(0.037)	(0.027)	(0.037)	(0.058)
g	-0.006	0.003	0.003	0.002
	(0.003)	(0.002)	(0.002)	(0.002)
s	0.010	0.008	-0.031	0.213*
	(0.011)	(0.020)	(0.042)	(0.091)
constant	-1.505**	-1.736***	-1.972***	-2.715***
	(0.577)	(0.386)	(0.468)	(0.564)
nobs	10137	14776	7075	2909
R-sqr	0.065	0.062	0.064	0.106
BIC	10724	17234	9473	4383

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

Table 11: Resilience modelling by firm size: informer

	micro	small	medium	large
if_informer				
h	-0.120	0.010	-0.005	-0.029
	(0.087)	(0.051)	(0.047)	(0.053)
c	0.267**	0.045	0.114	0.130
	(0.091)	(0.070)	(0.089)	(0.145)
d	0.138	0.044	0.152	0.142
	(0.307)	(0.184)	(0.189)	(0.245)
c*d	0.051	0.238	-0.024	0.182
	(0.383)	(0.191)	(0.187)	(0.218)
k	0.020	0.011	0.066*	-0.029
	(0.027)	(0.023)	(0.029)	(0.042)
l	0.379***	0.192**	0.209*	0.104**
	(0.083)	(0.068)	(0.082)	(0.040)
f	0.010	-0.003	-0.137**	0.175*
	(0.060)	(0.043)	(0.053)	(0.084)
g	-0.008	0.002	0.001	-0.000
	(0.005)	(0.003)	(0.002)	(0.003)
s	0.002	0.006	0.033	0.006
	(0.006)	(0.035)	(0.064)	(0.050)
constant	-0.172	-3.146***	-3.672***	-2.066**
	(1.361)	(0.706)	(0.684)	(0.738)
nobs	9869	14705	7036	2829
R-sqr	0.072	0.050	0.069	0.071
BIC	5386	8987	5777	2937

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

Table 12: Resilience modelling by firm size: inactive

	micro	small	medium	large
if_inactive				
h	-0.075	0.004	-0.031	0.012
	(0.050)	(0.034)	(0.034)	(0.039)
c	0.121*	0.012	-0.027	-0.103
	(0.056)	(0.045)	(0.065)	(0.104)
d	0.038	-0.288*	-0.300*	0.048
	(0.193)	(0.117)	(0.136)	(0.188)
c*d	0.627*	-0.232	-0.297*	-0.261
	(0.257)	(0.131)	(0.129)	(0.161)
k	-0.003	0.015	0.021	0.011
	(0.016)	(0.015)	(0.022)	(0.029)
l	-0.115*	-0.205***	-0.213***	-0.098***
	(0.049)	(0.044)	(0.060)	(0.028)
f	0.012	0.021	0.006	-0.042
	(0.035)	(0.027)	(0.039)	(0.059)
g	-0.000	0.001	-0.003	-0.003
	(0.003)	(0.002)	(0.002)	(0.002)
s	-0.002	-0.047	0.040	-0.107
	(0.004)	(0.028)	(0.043)	(0.082)
constant	2.660***	1.466***	2.632***	2.966***
	(0.644)	(0.354)	(0.464)	(0.610)
nobs	10133	14772	7071	2925
R-sqr	0.027	0.030	0.040	0.054
BIC	11523	16919	8959	4395

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

C Detailed estimations by industry

Table 13: Resilience modelling by industry: manufacturing

	(1)	(2)	(3)	(4)	(5)
	if_innovator	if_online	if_sticker	if_informer	if_inactive
	b/se	b/se	b/se	b/se	b/se
h	0.033	0.098**	0.024	-0.050	0.017
	(0.034)	(0.031)	(0.032)	(0.050)	(0.033)
c	0.120*	0.232***	0.107*	-0.000	-0.039
	(0.054)	(0.050)	(0.051)	(0.078)	(0.050)
d	-0.690***	0.148	0.269*	-0.380*	0.356**
	(0.143)	(0.109)	(0.106)	(0.179)	(0.119)
c*d	-0.191	0.477***	0.116	0.161	-0.026
	(0.146)	(0.123)	(0.131)	(0.181)	(0.132)
k	-0.071***	-0.013	0.004	0.022	0.029
	(0.021)	(0.019)	(0.019)	(0.029)	(0.019)
l	0.242***	0.128***	0.263***	0.347***	-0.125***
	(0.019)	(0.018)	(0.018)	(0.026)	(0.018)
f	0.041	0.052	-0.013	-0.037	-0.043
	(0.031)	(0.028)	(0.029)	(0.045)	(0.028)
g	0.004***	0.002	0.005***	-0.003	-0.003*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
s	0.134*	-0.020	0.076	0.078	-0.132**
	(0.059)	(0.017)	(0.049)	(0.083)	(0.051)
constant	-1.652**	-1.215*	-2.707***	-1.543**	1.124*
	(0.545)	(0.482)	(0.507)	(0.567)	(0.485)
nobs	12622	12640	12638	12517	12646
R-sqr	0.044	0.036	0.056	0.065	0.025
BIC	12568	14289	13666	7485	14234

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table 14: Resilience modelling by industry: construction

	(1) if_innovator b/se	(2) if_online b/se	(3) if_sticker b/se	(4) if_informer b/se	(5) if_inactive b/se
h	0.060	-0.015	-0.034	-0.205*	-0.089
	(0.062)	(0.073)	(0.057)	(0.098)	(0.059)
c	0.165	0.128	0.209**	-0.024	0.160*
	(0.085)	(0.096)	(0.074)	(0.109)	(0.077)
d	0.024	-0.270	0.140	0.243	-0.229
	(0.162)	(0.196)	(0.145)	(0.188)	(0.142)
c*d	-0.005	-0.149	0.495**	0.254	-0.074
	(0.217)	(0.251)	(0.183)	(0.265)	(0.193)
k	-0.052*	0.051	0.072**	-0.017	0.032
	(0.026)	(0.028)	(0.022)	(0.031)	(0.023)
l	0.382***	0.238***	0.353***	0.233***	-0.042
	(0.027)	(0.030)	(0.024)	(0.032)	(0.024)
f	-0.052	0.023	0.005	-0.219***	-0.018
	(0.048)	(0.055)	(0.042)	(0.057)	(0.042)
g	0.003	0.003	-0.004	0.007*	-0.006*
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
s	-0.008	-0.063*	-0.029	-0.066*	0.044
	(0.039)	(0.027)	(0.025)	(0.026)	(0.027)
constant	-1.921**	-0.508	-2.135***	-6.877***	0.435
	(0.698)	(0.651)	(0.582)	(1.558)	(0.591)
nobs	6922	6922	6926	6891	6926
R-sqr	0.084	0.057	0.069	0.057	0.032
BIC	5769	4851	7062	4267	7005

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Resilience modelling by industry: entertainment

	(1)	(2)	(3)	(4)	(5)
	if_innovator	if_online	if_sticker	if_informer	if_inactive
	b/se	b/se	b/se	b/se	b/se
h	0.072	-0.017	0.088*	-0.029	-0.051
	(0.044)	(0.049)	(0.043)	(0.055)	(0.045)
c	0.097	0.276***	0.123	0.015	0.236**
	(0.075)	(0.082)	(0.069)	(0.092)	(0.075)
d	0.258*	0.091	-0.155	0.012	-0.032
	(0.129)	(0.145)	(0.126)	(0.167)	(0.126)
c*d	0.092	-0.085	0.206	0.142	0.292
	(0.161)	(0.188)	(0.156)	(0.205)	(0.162)
k	0.024	0.041	0.062***	-0.024	0.068***
	(0.020)	(0.022)	(0.019)	(0.024)	(0.020)
l	0.158***	0.150***	0.199***	0.065*	-0.129***
	(0.023)	(0.025)	(0.021)	(0.028)	(0.022)
f	-0.009	-0.150**	0.030	-0.104	-0.012
	(0.048)	(0.051)	(0.044)	(0.057)	(0.047)
g	-0.007**	0.000	-0.004	-0.011***	0.004
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
s	0.032	0.003	0.018	0.041	0.020
	(0.029)	(0.018)	(0.021)	(0.038)	(0.016)
constant	-3.303***	-1.923***	-0.433	-0.892	2.097***
	(0.630)	(0.484)	(0.316)	(0.474)	(0.384)
nobs	5769	5776	5763	5742	5771
R-sqr	0.080	0.052	0.103	0.073	0.041
BIC	6700	5821	7462	4974	6893

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Resilience modelling by industry: ICT

	(1)	(2)	(3)	(4)	(5)
	if_innovator	if_online	if_sticker	if_informer	if_inactive
	b/se	b/se	b/se	b/se	b/se
h	0.028	0.001	0.029	-0.060*	-0.016
	(0.019)	(0.020)	(0.018)	(0.028)	(0.019)
c	0.193***	0.162***	0.197***	0.156**	-0.071*
	(0.034)	(0.034)	(0.032)	(0.048)	(0.033)
d	0.024	0.298***	0.050	0.054	0.174*
	(0.076)	(0.085)	(0.071)	(0.096)	(0.070)
c*d	0.431***	0.370***	0.044	-0.135	0.192*
	(0.083)	(0.100)	(0.083)	(0.123)	(0.082)
k	0.006	0.028**	0.029**	-0.012	0.019*
	(0.010)	(0.010)	(0.009)	(0.013)	(0.009)
l	0.290***	0.156***	0.278***	0.182***	-0.163***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.010)
f	0.036	0.009	0.044*	-0.035	-0.008
	(0.021)	(0.021)	(0.019)	(0.029)	(0.020)
g	0.002	-0.002	0.002*	0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
s	0.004	0.010	0.022*	0.003	-0.006
	(0.005)	(0.006)	(0.009)	(0.005)	(0.005)
constant	-1.044***	-1.572***	-0.914***	-1.645***	2.117***
	(0.201)	(0.213)	(0.195)	(0.255)	(0.216)
nobs	27991	27996	27998	27996	28001
R-sqr	0.064	0.041	0.076	0.056	0.045
BIC	29308	29011	33133	18047	31605

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 17: Resilience modelling by industry: retail

	(1)	(2)	(3)	(4)	(5)
	if_innovator	if_online	if_sticker	if_informer	if_inactive
	b/se	b/se	b/se	b/se	b/se
main					
h	-0.067*	0.053	0.023	0.045	-0.052
	(0.033)	(0.030)	(0.031)	(0.053)	(0.030)
c	0.242***	0.264***	0.250***	-0.025	-0.014
	(0.050)	(0.046)	(0.047)	(0.085)	(0.046)
d	0.229*	0.209*	-0.198*	0.547***	-0.429***
	(0.090)	(0.087)	(0.092)	(0.129)	(0.084)
c*d	0.075	0.499***	-0.421**	0.407*	-0.244*
	(0.123)	(0.115)	(0.128)	(0.180)	(0.114)
k	-0.004	-0.048**	0.039*	0.018	0.019
	(0.016)	(0.015)	(0.015)	(0.026)	(0.015)
l	0.277***	0.184***	0.173***	0.035	-0.099***
	(0.015)	(0.014)	(0.014)	(0.024)	(0.013)
f	-0.044	-0.017	-0.014	0.022	-0.002
	(0.028)	(0.026)	(0.027)	(0.047)	(0.026)
g	-0.003	0.004**	0.004***	-0.003	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
s	0.006	0.047*	0.008	0.013	-0.073**
	(0.014)	(0.023)	(0.014)	(0.030)	(0.026)
constant	-2.182***	-2.476***	-1.204***	-1.641***	1.446***
	(0.341)	(0.345)	(0.302)	(0.409)	(0.299)
nobs	13228	13243	13242	13177	13243
R-sqr	0.051	0.055	0.067	0.043	0.032
BIC	14683	16626	15687	7028	16699

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Resilience modelling by industry: health

	if_innovator	if_online	if_sticker	if_informer	if_inactive
h	-0.009	0.025	0.071*	-0.036	0.033
	(0.032)	(0.041)	(0.031)	(0.038)	(0.033)
c	0.109	0.300***	0.141**	0.193**	0.019
	(0.057)	(0.081)	(0.053)	(0.070)	(0.056)
d	0.186*	0.087	0.153	0.253*	-0.301***
	(0.082)	(0.125)	(0.078)	(0.099)	(0.078)
c*d	0.151	0.630***	0.086	0.250*	-0.129
	(0.106)	(0.138)	(0.102)	(0.127)	(0.103)
k	0.085***	0.115***	0.073***	0.089***	-0.066***
	(0.014)	(0.020)	(0.014)	(0.018)	(0.014)
l	0.083***	0.015	0.210***	0.073***	-0.125***
	(0.017)	(0.025)	(0.016)	(0.022)	(0.017)
f	0.006	0.147**	0.028	0.094*	-0.024
	(0.035)	(0.049)	(0.032)	(0.043)	(0.034)
g	-0.001	0.005*	-0.001	-0.003	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
s	-0.016	-0.022	0.002	0.003	0.012
	(0.019)	(0.022)	(0.019)	(0.028)	(0.019)
constant	-2.019***	-1.437***	-0.091	-2.590***	2.050***
	(0.319)	(0.370)	(0.249)	(0.438)	(0.287)
nobs	10503	10511	10503	10503	10511
R-sqr	0.065	0.063	0.084	0.076	0.029
BIC	12340	7025	13569	8900	12656

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D Estimations by lockdown periods

Table 19: Modelling resilience outcome during every lockdown and lockdown easing period:innovators

	innov_1lk	innov_1rel	innov_2lk	innov_2rel	innov_3lk	innov_3rel
h	-0.024	0.007	0.009	0.009	-0.021	-0.007
	(0.031)	(0.029)	(0.073)	(0.073)	(0.030)	(0.031)
c	0.300***	0.208***	0.332*	0.332*	0.240***	0.232***
	(0.067)	(0.062)	(0.168)	(0.168)	(0.063)	(0.066)
d	0.048	0.082	-0.017	-0.017	0.098	0.117
	(0.159)	(0.142)	(0.410)	(0.410)	(0.142)	(0.148)
c*d	0.224	0.164	0.444	0.444	0.205	0.248*
	(0.130)	(0.119)	(0.298)	(0.298)	(0.120)	(0.124)
k	-0.006	-0.006	-0.073	-0.073	-0.013	-0.006
	(0.021)	(0.019)	(0.053)	(0.053)	(0.020)	(0.020)
l	0.191***	0.193***	0.213***	0.213***	0.212***	0.182***
	(0.021)	(0.019)	(0.050)	(0.050)	(0.019)	(0.020)
f	0.056	-0.002	-0.010	-0.010	-0.050	-0.042
	(0.040)	(0.037)	(0.099)	(0.099)	(0.038)	(0.040)
g	0.000	0.000	0.001	0.001	-0.002	0.000
	(0.002)	(0.002)	(0.005)	(0.005)	(0.002)	(0.002)
s	0.068	0.027	0.251	0.251	0.011	0.002
	(0.043)	(0.029)	(0.169)	(0.169)	(0.023)	(0.010)
constant	-3.429***	-3.553***	-7.946***	-7.946***	-3.898***	-3.858***
	(0.458)	(0.429)	(1.555)	(1.555)	(0.447)	(0.469)
nobs	16525	16607	15193	15193	16605	16599
R-sqr	0.043	0.036	0.064	0.064	0.037	0.036
BIC	10344	11618	2925	2925	11396	10617

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

Table 20: Modelling resilience outcome during every lockdown and lockdown easing period:online

	online_1lk	online_1rel	online_2lk	online_2rel	online_3lk	online_3rel
h	0.079*	0.065*	-0.093	-0.093	0.064*	-0.004
	(0.031)	(0.030)	(0.078)	(0.078)	(0.029)	(0.029)
c	0.154*	0.087	0.346*	0.346*	0.152*	0.216***
	(0.068)	(0.064)	(0.146)	(0.146)	(0.063)	(0.061)
d	0.129	0.147	0.250	0.250	-0.010	-0.009
	(0.173)	(0.158)	(0.360)	(0.360)	(0.165)	(0.158)
c*d	0.016	-0.131	-0.289	-0.289	0.178	0.352**
	(0.150)	(0.145)	(0.387)	(0.387)	(0.131)	(0.122)
k	-0.026	-0.019	0.003	0.003	-0.005	-0.018
	(0.022)	(0.021)	(0.048)	(0.048)	(0.020)	(0.020)
l	0.051*	0.079***	0.088	0.088	0.061**	0.078***
	(0.022)	(0.021)	(0.048)	(0.048)	(0.020)	(0.020)
f	0.062	0.062	0.155	0.155	0.055	0.048
	(0.040)	(0.038)	(0.084)	(0.084)	(0.037)	(0.036)
g	0.001	-0.001	-0.003	-0.003	0.000	0.000
	(0.002)	(0.002)	(0.005)	(0.005)	(0.002)	(0.002)
s	0.023	0.068	0.187	0.187	0.037	0.046
	(0.023)	(0.037)	(0.141)	(0.141)	(0.028)	(0.029)
constant	-2.879***	-2.742***	-4.075***	-4.075***	-1.997***	-1.799***
	(0.490)	(0.458)	(0.998)	(0.998)	(0.409)	(0.386)
nobs	16600	16613	15256	15256	16568	16580
R-sqr	0.069	0.065	0.065	0.065	0.062	0.067
BIC	9720	10594	3291	3291	10919	11511

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

Table 21: Modelling resilience outcome during every lockdown and lockdown easing period:stickers

	sticker_1lk	sticker_1rel	sticker_2lk	sticker_2rel	sticker_3lk	sticker_3rel
h	0.060**	0.038	0.078*	0.078*	0.047*	0.022
	(0.020)	(0.020)	(0.033)	(0.033)	(0.020)	(0.024)
c	-0.150***	-0.061	-0.058	-0.058	-0.069	-0.191***
	(0.040)	(0.041)	(0.058)	(0.058)	(0.041)	(0.047)
d	0.011	-0.010	-0.024	-0.024	0.026	-0.037
	(0.093)	(0.096)	(0.144)	(0.144)	(0.095)	(0.106)
c*d	-0.034	-0.010	0.091	0.091	0.011	-0.099
	(0.083)	(0.084)	(0.135)	(0.135)	(0.084)	(0.094)
k	0.031*	0.010	0.007	0.007	0.011	-0.006
	(0.012)	(0.012)	(0.019)	(0.019)	(0.012)	(0.014)
l	0.069***	0.066***	0.019	0.019	0.081***	0.090***
	(0.013)	(0.013)	(0.020)	(0.020)	(0.013)	(0.015)
f	-0.008	0.009	-0.069	-0.069	-0.004	0.016
	(0.025)	(0.025)	(0.036)	(0.036)	(0.025)	(0.029)
g	-0.000	-0.002	-0.000	-0.000	-0.001	-0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
s	0.005	-0.010	-0.024	-0.024	-0.008	0.001
	(0.009)	(0.007)	(0.022)	(0.022)	(0.005)	(0.007)
constant	-1.480***	-1.252***	2.124***	2.124***	-1.557***	-2.737***
	(0.280)	(0.278)	(0.389)	(0.389)	(0.281)	(0.360)
nobs	16633	16638	16551	16551	16630	16638
R-sqr	0.029	0.026	0.017	0.017	0.033	0.034
BIC	21137	20700	12082	12082	20611	16939

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

Table 22: Modelling resilience outcome during every lockdown and lockdown easing period

	informer_1lk	informer_1rel	informer_2lk	informer_2rel	informer_3lk	informer_3rel
	b/se	b/se	b/se	b/se	b/se	b/se
main						
h	-0.020	-0.097*	0.006	0.006	-0.018	-0.065
	(0.047)	(0.047)	(0.113)	(0.113)	(0.045)	(0.049)
c	0.244*	-0.021	-0.078	-0.078	-0.037	-0.038
	(0.098)	(0.089)	(0.233)	(0.233)	(0.091)	(0.094)
d	0.160	-0.134	0.583	0.583	-0.138	-0.026
	(0.214)	(0.203)	(0.404)	(0.404)	(0.209)	(0.204)
c*d	0.056	0.079	-0.122	-0.122	0.022	0.021
	(0.192)	(0.171)	(0.468)	(0.468)	(0.174)	(0.180)
k	-0.035	-0.016	0.003	0.003	-0.027	0.008
	(0.030)	(0.027)	(0.071)	(0.071)	(0.028)	(0.029)
l	0.163***	0.143***	0.222**	0.222**	0.133***	0.146***
	(0.030)	(0.028)	(0.068)	(0.068)	(0.028)	(0.029)
f	0.065	0.110*	-0.199	-0.199	0.107*	0.134*
	(0.058)	(0.053)	(0.147)	(0.147)	(0.054)	(0.056)
g	0.001	0.005	-0.002	-0.002	0.001	-0.002
	(0.003)	(0.003)	(0.007)	(0.007)	(0.003)	(0.003)
s	0.078	0.077	0.002	0.002	0.019	-0.002
	(0.068)	(0.062)	(0.026)	(0.026)	(0.037)	(0.008)
constant	-3.585***	-2.890***	-6.477***	-6.477***	-2.820***	-3.454***
	(0.645)	(0.547)	(1.501)	(1.501)	(0.554)	(0.663)
nobs	16390	16413	14767	14767	16512	16445
R-sqr	0.071	0.059	0.076	0.076	0.052	0.060
BIC	5987	6728	1928	1928	6661	6273

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

Table 23: Modelling resilience outcome during every lockdown and lockdown easing period:inactive

	inactive_1lk	inactive_1rel	inactive_2lk	inactive_2rel	inactive_3lk	inactive_3rel
h	-0.077***	-0.050*	-0.111**	-0.111**	-0.063**	-0.001
	(0.020)	(0.020)	(0.043)	(0.043)	(0.020)	(0.019)
c	-0.035	-0.041	-0.036	-0.036	-0.064	-0.027
	(0.037)	(0.037)	(0.070)	(0.070)	(0.037)	(0.038)
d	-0.078	-0.035	-0.094	-0.094	-0.023	-0.011
	(0.091)	(0.091)	(0.179)	(0.179)	(0.091)	(0.090)
c*d	-0.062	-0.041	-0.185	-0.185	-0.176*	-0.175*
	(0.081)	(0.081)	(0.172)	(0.172)	(0.082)	(0.079)
k	-0.014	0.001	0.004	0.004	0.000	0.011
	(0.012)	(0.012)	(0.023)	(0.023)	(0.012)	(0.012)
l	-0.172***	-0.195***	-0.126***	-0.126***	-0.208***	-0.183***
	(0.013)	(0.013)	(0.025)	(0.025)	(0.013)	(0.013)
f	-0.036	-0.043	0.088*	0.088*	-0.011	-0.033
	(0.023)	(0.023)	(0.043)	(0.043)	(0.023)	(0.023)
g	-0.000	0.001	0.001	0.001	0.001	0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
s	-0.019	0.003	0.015	0.015	0.003	-0.005
	(0.011)	(0.004)	(0.020)	(0.020)	(0.004)	(0.007)
constant	0.575*	0.401	-2.254***	-2.254***	0.604*	1.147***
	(0.250)	(0.249)	(0.474)	(0.474)	(0.250)	(0.252)
nobs	16640	16635	16511	16511	16620	16637
R-sqr	0.049	0.046	0.029	0.029	0.051	0.042
BIC	23026	23043	9194	9194	22862	22888

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we control for differences by region and industry by including dummies

E Automated content analysis models

In order to build the baseline model and help to explain the decision of forthcoming models and control their quality, we utilized explainable machine learning approach. We have built a manual keyword-based model that uses simple tree based logic and lists of manually identified keywords to classify input texts into the one of predefined groups (Figure 4). Lists of keywords can be found in the Appendix 1. The model achieved 70% accuracy rate (I will write more details here).

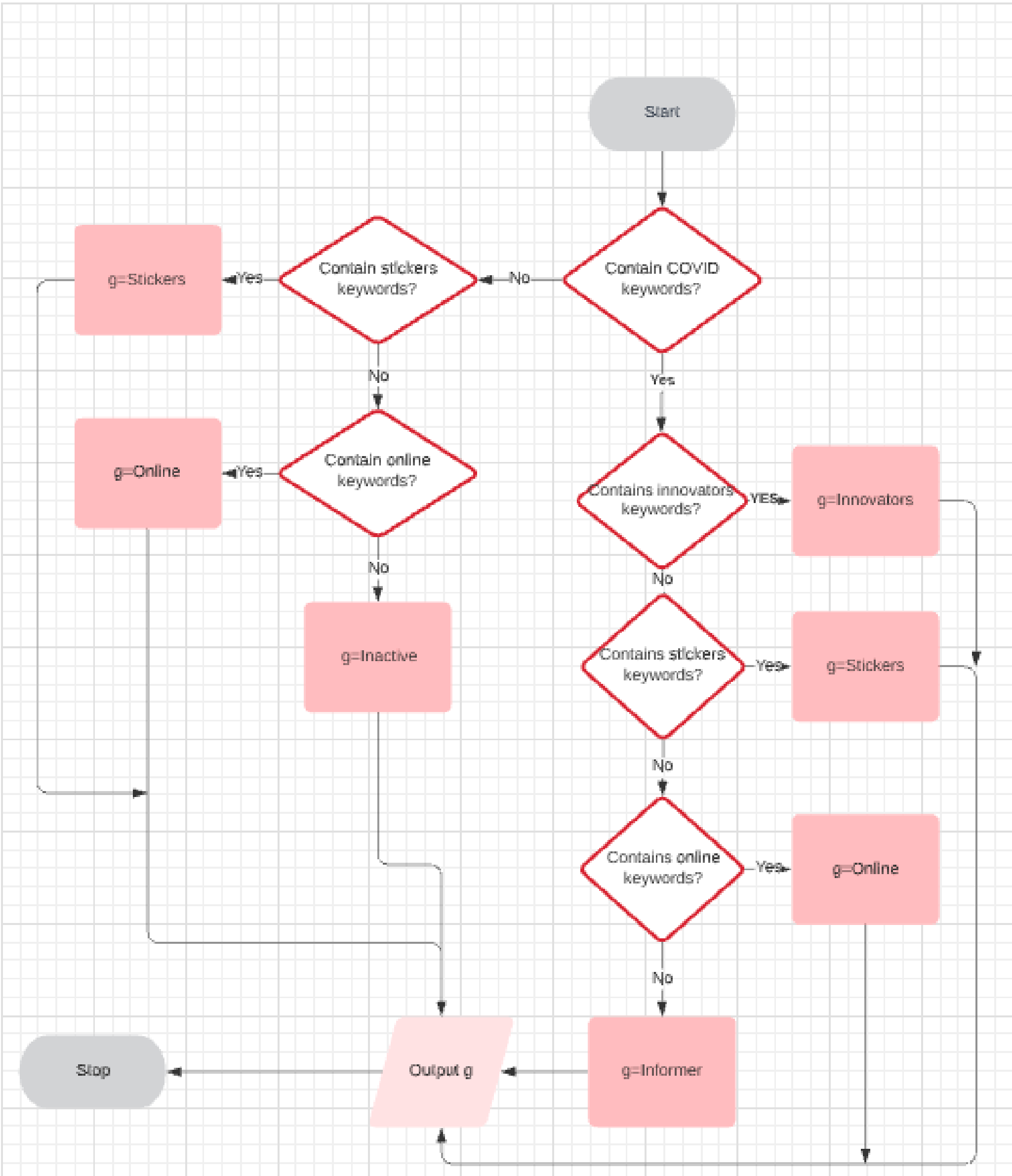

Figure 4: Keyword based classification algorithm

As a next step, we tried unsupervised modelling approach, Latent Dirichlet Allocation model in order to automatically find and cluster important words in 'supertopics'. According to the general LDA approach we think about webpages (or documents) as a specific distribution of topics (or themes – like COVID, medicine, business, etc.). Each topic is then determined by the distribution of words that form the topic. The goal of the algorithm is to find a specific set and distribution of words and cluster them into topics in order to accurately divide the corpus of texts (websites) into distinguishable groups ('supertopics'). For example, the group of texts talking about COVID problems, group trying to showcase their active and resilient position, and group that suspends its business. The idea of LDA model is based on prior and posterior words-topics distribution, so it is Bayesian model in its nature. We have run classical LDA model with predefined number of topics to search for (we have a number that minimizes the metric (??)). While the LDA model was unable to clearly identify COVID-related topics, it clearly helped to separate online-related keywords (see table).

The classical problem of automated content analysis is that generated supertopics can be far from researcher's interest and generally user needs to perform a post-classification of topics in belief that the topics of interest would be identified. Thus, to overcome a classic problem of automated content analysis we modify priors of LDA model based on keywords of interest that we have found during previous research steps. We used keyATM and Guided LDA models that allow to modify bayesian priors of topic-keyword distribution, by manually assigning some predefined words to the topics. We created initial four groups of keywords, and run the model using same total number of topics, except 4 of 7 topics were pre-filled with keywords of interest. We then used the model to automatically score unseen texts and assign them to a specific topic. The accuracy of the model based on the subset of test data is . While the model provided good basic results and helped to identify some important features of the text corpus we are working with, there were some model limitations that we have diagnosed.

First of all, the LDA-based algorithms rely on words distributions, and they do not account for context of keywords nor their interrelationship. Thus, model was unable to pay attention to the specific important phrases, since it worked with individual words only. As a result, low accuracy can be explained with inability to pick up whole phrazes instead of individual words ('we are open now' phraze has much more relevance then individual words 'we','open','now' that can be spread all over the text document with different contexts (like we hope to open next year)).

The second limitation of LDA-type of models is that they do not take into account the general context of sentences. Models just treat any text as a list of words, that are not connected contextually. However, without the context it would be hard to understand the full meaning of messages that businesses put on their websites. For example, the phraze 'we offer free deliveries. Buy online' would not be surprising to see on the ecommerse website, but it would be an outstanding case for a local restaurant[9] or handmade crafts shop. Links to online videos, photos, stories and online meeting sessions would be an inseparable part of social media platform, but it would be surprising to see some of these elements on the website of a community church[10] or local rabbit charity [11]. As a result, context makes difference and phrazes of current research interest would have totally different meaning given the nature and a context of the business.

The third limitation of the algorithm is limited classification power due to the limited training data setting. Classical LDA algorithms are unsupervised models, which utilize statistical approaches to fit models to the data. The consequences of the unsupervised approach make it hard to build the model that perfectly aligns your needs. As in our example with classical LDA model, we could find only 'online' topic that corresponded to our research interest. Other topics were not precisely identified. On the other hand, classical topic modelling techniques, like TF-IDF factorization with simple classification models built on top require substantial amount of training data, in order to achieve good classification accuracy.

Given the main limitations of previous models, we decided to proceed with modern NLP modelling approach called Bidirectional Encoder Representations from Transformers (BERT). The model was first introduced by Google in 2018 and had changed the world of text modelling since then. The model is based on deep neural network architecture, with a special 'recurrent' type of neural layers called attention layer. Main achievements of the model is that it is able to incorporate the context of surrounding sentences in the text, when working with the current sentence (that is why model is 'bidirectional'). Apart from the content, the model is able to 'understand' the text semantically, as the model is trained using 'masked language modelling' (where the model reconstructs missing words in the sentences with deliberatly 'masked' words) and the next sentence prediction task

[9] https://www.copleysfood.co.uk/

[10] https://www.hopecommunitychurch.co.uk/

[11] https://www.hopperhaven.org.uk/

(model predicts whether a given sentence is a continuation of the previous sentence). The training is performed on a billion-sized text corpus, in order to learn majority of semantic and contextual patterns. As a result of such training, the ability of the model to 'understand' the text is used for various tasks through a transfer learning. The transfer learning is a process of transforming the original model for the area-specific tasks, including classification and prediction tasks, question answering, named entity recognition and many other. Usage of models, trained on large amount of data, allows to significantly decrease the number of training data required to adjust the model to a new task.

We then based our group assignment decision on the prevalence of keywords found in one of these groups. If no keywords of interest are found on the webpage, we will assign the company into the fifth group, 'inactive'. In the fifth step we streamline the group assignment and identification stage by utilizing supervised natural language processing model named BERT [12] to automatically assign companies to one of the predefined groups. The final accuracy of the classification is 83%.

[12] https://arxiv.org/abs/1810.04805