#2203 Bremen Papers on Economics & Innovation

Exploration or Exploitation: Innovation Behavior of SMEs and Large Firms during the COVID-19 Crisis

Jessica Birkholz*, Jarina Kühn*, Mariia Shkolnykova*

April 2022

Abstract

This paper contributes to the discussion on exploration and exploitation by analyzing the innovation behavior of SMEs and large firms during the first year of the COVID-19 pandemic in Germany. It provides a novel way to measure the type of firm innovation behavior in a dynamically changing environment. After collecting news articles about innovation activities conducted by firms, we applied text mining techniques to identify the positioning of each firm on the continuum from exploitation to exploration. The results of our analyses indicate three main dynamics: 1) all studied firms tend to conduct more explorative innovation activities during the COVID-19 crisis, 2) large and "technology-intensive" firms are more prone to perform explorative innovation activities than SMEs and firms that are not "technology-intensive", and 3) technology intensity is associated with explorative innovation behavior during the crisis. Our results suggest that considering technology intensity and the size of firms is important for designing effective policies during crises.

Keywords

COVID-19; Crisis; Innovation; SME; Text Mining; News Data; Exploration; Exploitation

JEL Classifications

O31, O33, L25

*University of Bremen,
Max-von-Laue-Straße 1,
28359 Bremen, Germany
Corresponding author: Jessica Birkholz (jessica.birkholz@uni-bremen.de)
DOI: http://dx.doi.org/10.26092/elib/1489

1. Introduction

For countries worldwide, the years 2020 and 2021 have been imprinted by a severe crisis caused by the COVID-19 pandemic. The COVID-19 shock is unique in comparison to former crises: it is described as rapid, severe, and exogenous, i.e., it hit firms independently of their past performance (O'Toole et al. 2021). Turnover and employment have been affected strongly, failure risks were increased (Kalemli-Ozcan 2020), and productivity levels decreased (Bloom 2020). Due to their vulnerable liquidity situation, SMEs are especially threatened (Nehrebecka 2021) and experience a greater negative shock than large firms through the COVID-19 crisis (Harjoto et al. 2021).

Despite the damage to the economy, crises also open windows of opportunity (e.g., Archibugi 2013). Especially innovation is a lever for firms to counteract crisis effects and entails possibilities to benefit from the impaired competition (Ba & Bai 2020). This requires managerial responses to the changing macro-environments. According to a recent study by Ebersberger & Kuckertz (2021), start-ups provided the quickest innovation response to the COVID-19 crisis. Further, Fritsch et al. (2021) show that the number of start-ups in innovative manufacturing and technology-oriented services increased during the COVID-19 crisis. What remains unclear is the kind of innovation behavior that firms pursued in this crisis. This question is relevant for policy makers to be able to design targeted support programs. Literature mainly separates innovation activities into exploration, based on diversification and the entering new markets, and exploitation, based on the reuse of prevalent knowledge (March 1991). In practice, the two strategies may be described on a continuous scale.

There is an extensive literature on how firms reacted to the Global Financial Crisis. In this context systematic differences between the innovation behavior of SMEs and large companies were identified (Ortega-Argilés et al. 2009; Noori et al. 2017). Thus, based on a survey of executives in the software industry, Latham (2009) states that SMEs follow revenue-generation strategies and exploration of new market opportunities during recession. In contrast, large firms tend to reduce costs and maintain existing behavior. Archibugi et al. (2013a), examining the data from UK Community Innovation Survey, state that new entrants, that are not afraid to look for new markets or cooperative partners, profit during the crisis. Furthermore, technology-intensive firms and firms that are not technologically advanced are found to react differently to the crisis. Technology-intensive firms are found to be able to mitigate post-crisis uncertainty in a better way than their non-technology-intensive counterparts (Nemlioglu & Mallick 2020). However, to the best of our knowledge, research in this vein is scarce for the case of the COVID-19 crisis. Based on the uniqueness of this recent crisis, the innovation behavior of different types of firms during the crisis may be distinct from former observations. Thus, this study aims at closing this gap by investigating the following research question:

RQ: How does the type of firm innovation behavior - exploration or exploitation - during the COVID-19 crisis relate to firm size and technology intensity?

In order to approach this research question, the access to recent and constantly updating data on firm innovation behavior is necessary. Conventional indicators for innovation activities are not suitable for research endeavors requiring current data. Patent-, survey-, and publication-based indicators, commonly used for investigating innovation activities, have the major limitation of coming with a time lag. Thus, in order to obtain information about the innovation behavior of firms during the COVID-19 pandemic, this paper uses a novel innovation indicator: news data, also known as daily press data. This data, broadly used in social science disciplines (e.g., Arnaldi 2008; Vargo et al. 2018), is paving the way in innovation studies. Thus, researchers use news data in order to map regional entrepreneurial activities (von Bloh et al. 2020) or to follow the regional reporting on innovation and new technologies (Ozgun & Broekel 2021).

In this paper, we propose a new way of using news data in order to assess the innovation activities in a rapidly changing environment. The news data base Factiva is used to retrieve the daily press data on innovation activities in the first year of the COVID-19 pandemic in Germany. This news data is then related to the respective firm. Main variables of interest (firm size and patenting activity) along with firm-related control variables are collected from the Orbis database and used in a regression analysis. We chose to focus on Germany, as German firm landscape is particularly dominated by mittelstand[1] and as well as it belongs to technologically advanced countries, depicting an optimal field for the main variables of interest.

In the same way, the reference data set, covering the five-year period before the pandemic, was collected. These data are used to assess, whether the innovation activities, which particular firms performed during the first year of the pandemic, can be referred to as more explorative or more exploitative in comparison to the period before the crisis. Text mining techniques are applied in order to combine the two data sets and to estimate the position of the firm on a continuum from exploitation to exploration, further referred to as 'score'. This score is used as a dependent variable in the analysis.

The study delivers three main results. First, SMEs and large firms generate different types of innovations before and during the crisis. Large firms tend to perform more explorative innovation activities than SMEs, but during the crisis SMEs re-direct their activities towards exploration in a more pronounced way than large firms. Second, knowledge accumulation conducted by large firms and, thus, Schumpeter II dynamics, result in more explorative innovation behaviors than the activities performed by SMEs.

[1] German mittelstand expresses medium-sized firms with a turnover between 1 million and 50 million Euro per year and with 10 to 499 employees, representing the largest proportion of firms in Germany. Gabler Wirtschaftslexikon, https://wirtschaftslexikon.gabler.de/definition/mittelstand-40165/version-263557.

Third, technology-intensive firms tend to perform more explorative innovation activities during crisis, but this relation does not hold for the time period before the crisis. This means that technology intensity fosters innovative resilience during crisis.

This study has several important implications. Beyond the particular analysis of SMEs, large firms, and firms with different degrees of technology intensity, we contribute to the development and application of new innovation indicators not depending on patent, publication, or survey data. This approach makes it possible to identify innovation behavior of service companies not fully covered by the traditional data sources. Apart from that, this study will inform policy makers about the trend of innovation activities of different types of firms, which serves as valuable information for the design of funding initiatives (e.g., programs promoting riskier innovation projects may be directed towards SMEs). Managers may profit from knowing main trends in innovation activities during crisis and, therefore, push forward specific areas (e.g., digitalization), where networking opportunities can be found. Finally, the study lays the empirical foundation for a long-term analysis, asking about the success of exploitative versus explorative innovation behavior.

The paper is structured as follows: Section 2 provides the theoretical considerations along with the generated hypotheses. Section 3 gives insights on the applied data and methodology. The results are depicted in section 4, which are discussed in the 5th section.

2. State of the art

2.1. Innovation behavior of different types of firms

Differences in innovation activities depend on several firm characteristics. Industry classification (e.g., Bhattacharya & Bloch 2004; Cefis & Orsenigo 2001), firm age (e.g., Balasubramanian & Lee 2008), cooperation networks (e.g., Baum et al. 2000), and regional environment (e.g., Sternberg & Arndt 2001) are some of the identified determinants. Research and development (R&D) investments and their subsequent product technology intensity, as well as the firm size are prominent factors in explaining differences in innovation activities between firms (Acs & Audretsch 1987; 1988; Archibugi et al. 2013).

Reasons for the impact of firm size on innovation activities are manifold and mostly related to structural differences. In comparison to SMEs, large firms can profit from economies of scale and scope, easier access to external finance, and the ability to diversify risks via investing in a number of R&D projects. Thus, large firms often possess greater market power, which allows them to set entry barriers into the industry (Vossen

1998; Ortega-Argilés et al. 2009; Noori et al. 2017). Moreover, because of the possibility of the internalization of costs (Hollenstein 2005), their transaction costs are lower. Large firms have the possibility to focus on knowledge management strategies through multiple management layers (Rizea et al. 2011). Furthermore, large firms can draw upon past experiences (Sanidas 2014). On average, the propensity of large firms to innovate is higher (Triguero et al. 2014), they better sustain product and process innovations, and their process innovations also have a longer lasting effect on productivity. (Rochina-Barrachina et al. 2010).

On the other side, innovation is also essential for the survival of SMEs (Castrogiovanni et al. 2012). Thereby, they promote economic development (Acs et al. 2008). Particularly through process innovations, they contribute significantly to job creation (Triguero et al. 2014). In comparison to large firms, SMEs can profit from flatter hierarchies, which lead to quicker decision-making as well as profiling and the takeover of specific market niches (Vossen 1998). As Spithoven et al. (2013) pointed out, they do have superior responsiveness to market needs (Dahl & Moreau 2002), organizational flexibility (Sivadas & Dwyer 2000), and less bureaucracy (Cassiman & Veugelers 2006). Thus, they are prone to be agents of change (Audretsch 2002).

SMEs benefit from unique cooperation patterns with, among others, science organizations and R&D laboratories, which may help to overcome resource disadvantages (Ortega-Argilés et al. 2009; Sahut & Peris-Ortiz 2014). Also, previous literature states that there is a tendency of SMEs to innovate around the firm's core technologies due to their restricted resources (Corradini et al. 2016; Antonelli & Scellato 2015; Nelson 1985; Dosi 1982), exploiting internal distinctive competencies (Corradini et al. 2016; Kogut & Zander 1992; Corradini et al. 2015). Technical specialization helps small innovators through strong focus, but technical diversification is important to better cope with changes (Corradini et al. 2016). They do have the determination and inspiration to pursue radical innovations (Falck 2009, Baumol 2002), though they are lacking the stamina to improve and extend the innovations, which large firms rather do (Falck 2009). As part of their protection strategy, SMEs are fastest to adapt to changing environments and regulations (Urbańska et al. 2021; Kortelainen et al. 2012).

Comparing SMEs and large firms' activities, there are some additional findings: Large firms are more likely to perform R&D activities and SMEs are rather averse to this activity. However, R&D-performing SMEs tend to have proportionally higher intensity of these activities (Taymaz & Üçdoğruk 2009). Moreover, there is empirical evidence that SMEs' per employee number of patents is larger than in large firms (Audretsch 2002).

These differences in innovation processes are likely to translate into different innovation behaviors of firms. One way to describe innovation behavior is in terms of exploration and exploitation. Explorative innovation behavior fosters discovering new possibilities, whereas exploitation relies on previous knowledge when innovating (March

1991). Rather than seeing exploration and exploitation as two distinct processes, March (1991) developed the idea that they are the two extremes on a continuum which he also describes as exploration-exploitation trade-off.

Entrepreneurs as new market entrants are characterized by Schumpeter (1934) as forces to kick-off economic developments by introducing radically new products and services. The innovation behavior of these actors therefore can be classified as explorative, i.e., they search for novelties. The late Schumpeter (1942) argues, that innovation has become a routine for big firms that are able to employ personal in R&D occupations. These firms show an exploitative innovation behavior, i.e., the exploitation and further development of prior knowledge (Schumpeter 1942).

2.2. Innovation behavior of different types of firms during crisis

During a crisis resources are more restricted, and therefore, it is important for firms to follow the behavior that promises the best outcome (D'Agostino & Moreno 2018). The previously mentioned characteristics of SMEs and large firms are expected to lead to differences in innovation behavior with regard to exploration and exploitation, especially during crises. Empirical evidence mostly supports this assumption. Latham (2009), following the data from a survey of software executives regarding innovative strategy during recession in the software industry, states that large firms and SMEs follow different strategies during the crisis: while SMEs look for revenue-generation, large firms tend towards cost reductions. Moreover, Archibugi et al. (2013a), using the panel data from UK Community Innovation Survey, state that while before a crisis well-established large firms are expanding with regard to innovative investments, new entrants profit during a crisis. Such firms are not afraid to look for new markets or cooperative partners, following thus an exploration strategy. On the other side, relying on the comprehensive data set of innovative performance of Latin American firms during the Global economic crisis, Paunov (2012) finds no effect of size on the probability of a firm to pursue innovation projects during a crisis.

In summary, the literature shows that the innovative strategies pursued by large firms and SMEs generally differ. This can be observed especially during a crisis: while large firms tend to take a "waiting" position, slowly following existing innovative projects by exploiting benefits from size and scope, SMEs tend to look for a niche to develop and to explore new markets and cooperation opportunities. SMEs, though reacting fast to changes, are threatened through the impaired access to credits (Gourinchas et al. 2021), being further amplified through the high-risk indicators (Nehrebecka 2021). Against the backdrop of these considerations, the first hypothesis can be derived:

Hypothesis 1: During the COVID-19 crisis, SMEs perform explorative innovation activities while large firms perform exploitative innovation activities.

Of further interest are also the differences in industries along technology-related factors. Established research describes factors such as the richness of technological opportunities, the cumulativeness of the knowledge base, or the means to benefit economically from innovation (Dachs et al. 2017; Cohen 2010; Dosi & Nelson 2010; Marsili 2001). Dachs et al. (2017) found compensation and displacement effects to increase with the technology intensity of the sector and the industry. Pöschl et al. (2016) investigated productivity effects in technology-intensive regimes and found a positive effect from innovation in knowledge-intensive, high-technology business services as well as in manufacturing spillovers. Nemlioglu & Mallick (2020) state that R&D and patenting can help firms to cope with their post-crisisperiod uncertainties.

When investigating which firm characteristics support a crisis-persistent innovation behavior, research provides evidence that firms with an in-house R&D section are more likely to engage in innovation during crisis (Archibugi et al. 2013). It can be assumed, that they belong to the group of technology-intensive firms, when committed to an in-house R&D section (Archibugi et al. 2013). Those firms are also considered to be rather explorative. On the other hand, Kitching et al. (2009) found that firms, when threatened, such as in times of crisis, change their investment strategies in response to the changing macro-environment. As innovation is risky and costly, they are prone to focus on exploitative, incremental, and cost-reducing innovation activities (Kitching et al. 2009). This behavioral pattern is more probable for non-technology-intensive firms. Thus, the assumption we draw for technology-intensive versus non-technology-intensive firms is as follows:

Hypothesis 2: Technology-intensive firms perform explorative innovation activities while non-technology-intensive firms perform exploitative innovation activities during the COVID-19 crisis.

3. Data and Methodology

3.1. Research design

In order to perform the analysis, we applied a four-step procedure including: 1) data set creation, 2) data preparation, 3) descriptive analysis, and 4) score estimation and econometric analysis (see Table 1 Research design).

The first step involved the creation of two data sets: The first consists of articles related to firms' innovation activities during and before the COVID-19 pandemic, the second data set serves to create dictionaries of innovation behavior in terms of exploration and exploitation. In order to obtain the data, we used the news database Factiva, which is run by Dow Jones and includes more than 30,000 daily press sources sampled in nearly every country[2]. Factiva permits searching for content in the text material using keywords and combinations of them. The firm population for the firm-level data sets was then manually derived from the articles.

Second, data cleansing of all extracted data sets was performed. This procedure included stop words and numbers removal, spell check, and stemming. For the exploration and exploitation data set, news corpus data cleansing and unigrams extraction was completed.

In the third step, data sets were analyzed with the help of descriptive methods. Firm-level and regional variables were explored and a preliminary descriptive analysis was executed. The term frequency–inverse document frequency *(tf-idf)* indicator was calculated for exploration and exploitation data set and the difference was used to assign each term to the exploration or the exploitation dictionary.

The fourth step encompasses an econometric analysis by means of Ordinary Least Squares (OLS) regression. The dependent variable in this regression is based on the cosine similarity score between the firm-related news corpus and the dictionary. It reflects the score of each firm on the continuous scale from exploitation to exploration. Mathematically it is expressed as the difference between cosine similarity scores for the exploitation and exploration dictionaries for each individual firm. The main independent variables are of a binary nature. They relate to the two main hypotheses and capture the firm's size and its patenting activity. Depicted steps are further elaborated in the next subsections[3].

Table 1: Research design

	Data set creation	Data preparation	Descriptive analysis	Econometric analysis
Innovative firms	Extraction of articles & firms	Data cleansing	Examination of firm & regional characteristics	Score estimation and econometric analysis
Dictionaries	Extraction of articles	Data cleansing extraction of unigrams	Creation of dictionaries	

[2] According to information from Factiva Customer Service.

[3] The analysis is performed in RStudio open source software (accessible under: https://www.rstudio.com/).

3.2. Firm-level data processing

3.2.1. Data set creation

In order to identify innovating firms, we first searched for articles covering innovation activities. To detect innovation activities related to the COVID-19 crisis, the search was limited to the keywords *Corona* or *Covid**[4] and *innov** in the title or first paragraph of the article. Further, only news entries of German publishing houses were considered. To make sure that innovation activities appeared during the COVID-19 crisis, the search was limited to the time period from 01.12.2019 to 10.01.2021. This time period starts with the approximate first notion of COVID-19 in the media[5] and ends with the designated end of the lockdown restrictions that was estimated at the time of the data collection. Additionally, this period covers the whole year, which allows to account for seasonal fluctuations. As a result, 2,960 news entries of innovation activities were identified.

These news articles were later manually checked to identify the mentioned firms. Here, state and research institutions, registered associations, and charity foundations were not considered, as the primary interest of the paper is the investigation of firm innovation activities. Furthermore, only German firms were included. In case of multinational enterprises, the German office was considered. As a result, 993 firms were identified. These firms constitute the data set, which is used for the further analysis.

Additionally, it was necessary to secure that the firms' innovation behavior was influenced by the COVID-19 crisis. For that, the news articles regarding innovation activities of the same firms were collected for seasonally the same period before the COVID-19 pandemic (from 01.12.2017 to 10.01.2019). These articles were obtained with the help of free-text search using the keywords *innov** and the firm's name in the title or first paragraph. For consistency reasons, only the German publishing houses were considered. As a result, 5,407 news entries were identified, which included innovation activities of 329 firms from the data set. This constitutes one third of the identified innovative firms during the COVID-19 crisis. For the other firms, no news articles within the pre-COVID-19 period could be identified.

3.2.2. Data preparation

In order to work with the collected news corpus, several data cleansing exercises were performed. The aim was to standardize the text and remove words and word

[4] With * reflecting any number of symbols, which end the word.
[5] This was checked in the media coverage.

constructions, which may bias the results of the analysis. Thus, the following stages of data cleansing were executed:

- Tokenization: At this step, the news corpus, attached to each of the firms, was divided into individual words (or tokens), which presented the units of further analysis. Then, it was transitioned in word-per-row format (Silge & Robinson 2017).
- Removal of stop words and numbers: For this purpose, a German dictionary of stop words was used[6] and data-specific stop words were removed.
- Spell check: Here, the presence of each word in the German dictionary[7] was controlled. In this step the words, incorrectly read by software along with authors' neologisms could be ruled out. Such words can otherwise complicate the analysis, as they are hard to interpret.
- Stemming: In this step we unified words, having similar meaning. This was done via word stems. For example, the words 'system', 'systems', 'systemic' were unified to 'system'. Thus, unbiased word counts could be obtained[8].

3.2.3. Descriptive analysis

At this stage the firm-level data set was extended with variables, relevant for the regression analysis. Overview and description of the variables can be found in Table 2 Explanatory and control variables[9].

The main explanatory variable, *SIZE*, indicates, whether the firm is an SME or a large corporation. Here, the size classification of Orbis was adopted. This classification is based on three indicators: operating revenue, total assets, and employees. According to Orbis, a large firm has at least 150 employees and total assets of 20 Million Euro or more. The very large firms category according to Orbis include companies that have at least 1000 employees as well as total assets of at least 200 Million Euro or operating revenue of at least 100 million Euro. This definition however differs from the one of the

[6] Available under https://rdrr.io/cran/lsa/man/stopwords.html [Date of the last access: 18.02.2022].

[7] Available under https://github.com/titoBouzout/Dictionaries/blob/master/German_de_DE.aff [Date of the last access: 18.02.2022].

[8] An alternative to stemming would be lemmatization. Lemmatization uses the dictionary form of each word in order to unify the terms. We prefer to use stemming for our analysis, as lemmatization presents the lexicon of a stated point of time and not all words can be lemmatized (Zeroual & Lakhouaja 2017).

[9] All variables were derived from the Orbis database from Bureau van Dijk.

European Commission. According to them, the SME boundary lies at 249 employees and 43 Million Euro. Hence, we also included firms that are considered large according to Orbis into the SME category for our analysis. Thus, large corporate groups were separated from larger individual ventures.

Control variables include regional and individual characteristics of the firms. *NACE* stands for level one codes of the Statistical Classification of Economic Activities in the European Community and, thus, presents the industry a firm is related to. The NUTS variable encompasses the federal state, where the firm is located. The *AGE* variable relates to the age of the firm from founding date until 2021. The variable *PATENTS* shows whether the firm has at least one patent, distinguishing between patenting and non-patenting firms. The two last variables relate to the firm's corporate structure: *HOLDING* refers to whether a firm is the parent company of a corporate group and *CORPGROUP* reflects whether the firm is part of a corporate group, i.e., whether there is at least one more enterprise belonging to the same corporate group.

Table 2: Explanatory and control variables

Variable	Description	Values
SIZE	Whether the firm is considered SME (1) or large corporation (0)	Binary
NACE	NACE Rev. 2 industry classification	Sections A to U
NUTS	NUTS 1 regional classification	German federal states coded
AGE	Firm age in years in 2021	Non-negative integers
PATENTS	Whether a firm has at least one patent (1) or no patents (0)	Binary
HOLDING	Whether a firm is a headquarter of a corporate group (1) or not (0)	Binary
CORPGROUP	Whether the firm is a part of the corporate group (1) or not (0)	Binary

3.3. Exploration and exploitation dictionaries

3.3.1. Data set creation

The creation of the dictionaries started with the identification of keywords, which could be associated with exploration and exploitation. For that the following initial definition of March (1991) was considered:

"Exploration includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation. Exploitation includes such things as refinement, choice, production, efficiency, selection, implementation, execution." (March 1991)

Thus, an exploration strategy can be characterized by terms such as *search, variation, risk taking, experimentation, play, flexibility, discovery,* while an exploitation strategy can be characterized by terms such as *refinement, choice, production, efficiency, selection, implementation, execution.* These keywords were translated into German[10] and were used for the free-text search. A news article was considered eligible for one of the dictionaries, if at least one of the keywords together with the stem *innov** were present in the title or first paragraph of the article. Further, to limit the results to the German economy, only news entries from German publishing houses were considered. The observation period covers five years from 10.01.2016 to 10.01.2021. As the result of the search, 17,691 articles on exploitation and 15,298 articles regarding exploration were obtained.

3.3.2. Data preparation

For the case of the exploration and exploitation news corpus, the same data cleansing procedure was applied as for the firm-level news. It included tokenization, removal of stop words and numbers, spell check, and stemming. The analysis was performed on the single-word (unigram) level. As dictionaries are usually composed of single words, the choice of the unigram level is obvious. Apart from that, there is often a strong correlation between uni- and bigram indicators (Braga et al. 2009) with unigrams being easier to interpret. Thus, at this step, two tables were created (exploration and exploitation news corpus), with each row representing one word with the reference to the specific document.

3.3.3. Descriptive analysis

After the news corpus for both the exploration and exploitation data set had been cleaned, the dictionaries, which describe exploration and exploitation, were created. For that, it was necessary to distinguish the words connected to exploration from the ones related to exploitation within the observed news corpus. This is done with the help of the term frequency–inverse document frequency *(tf-idf)* indicator. *Tf-idf* helps to identify important words for each document by ruling out the words, which are common to the whole news corpus (Silge & Robinson 2017). Mathematically, it contains two parts: term

[10] For exploration: *Suche, Variation, Risikobereitschaft, Experimentierfreude, Spiel, Flexibilität, Entdeckung*; for exploitation: *Verfeinerung, Wahl, Produktion, Effizienz, Auswahl, Implementierung, Ausführung,* respectively.

frequency *(tf)* reflects how often a term appears in the document; inverse document frequency *(idf)* reflects the number of documents in the corpus, which contain the term (Silge & Robinson 2017).

$$tf-idf = tf \cdot idf = tf \cdot \ln\left(\frac{n_{docs}}{n_{docs\,with\,term}}\right) \quad (1)$$

Thus, *Tf-idf* measures the weight of a term for a specific document in the news corpus. In order to calculate the weight of each term over all documents in the news corpus, the average *tf-idf* indicator is calculated for each term

$$tf-idf_{av} = \frac{\sum_{d=1}^{n} tf-idf \cdot l_d}{\sum_{d=1}^{m} l_d} =, \frac{\sum_{d=1}^{n} \frac{n_d}{l_d} \cdot l_d}{\sum_{d=1}^{m} l_d} \quad (2)$$

where *d* – document, n – total number of documents, – number of times a term appears in a document, – length of each document.

The average *tf-idf* indicator was calculated for each term in the exploration and exploitation data set. In order to assign each term to either the exploration or exploitation dictionary, the difference between in both news corpuses was calculated. Whenever for a term was higher in the exploitation corpus, the term was assigned to the exploitation dictionary. Otherwise, the term was assigned to the exploration dictionary. For example, for the stemmed term *'qualitatsgaranti'* (English: 'quality guarantee') for exploitation was equal to 0.00000653 whereas for exploitation it was equal to 0.00000519. Thus, the term was assigned to exploitation.

As a result, the exploration dictionary contains 86,388 terms and the exploitation dictionary contains 41,839 terms. Table 3 Top terms of exploration and exploitation dictionaries presents the top ten terms with the highest differences for both dictionaries. It can be observed, that while the exploitative dictionary contains many financial terms and names of big enterprises, the explorative dictionary has a broader range, including industries with transformation potential, apps or places, where new products can be presented.

Table 3: Top terms of exploration and exploitation dictionaries

Explorative dictionary		Exploitative dictionary	
Term	**$tf-idf_{av}$ difference**	**Term**	**$tf-idf_{av}$ difference**
Wohngrupp *(English: residential group)*	0.0000839	Blindtext *(English: blank text)*	0.0001692
Troponin	0.0000809	Jahresabschluss *(English: annual financial statement)*	0.0001241
Sokrat *(English: Sokrates)*	0.0000598	Vermögensgegenstand *(English: asset)*	0.0001025
Edelstahlwerk *(English: stainless steel plant)*	0.0000515	Axon	0.0001007
Prospekt (as adjective prospektiv, English:	0.0000502	Huawei	0.0000794

prospective)			
Pegelalarm (Water Information and Flood Warning app)	0.0000499	Produktkampagn (Produktkampagne, *English: product campaign*)	0.0000764
Spielwarenmess (Spielwarenmesse, *English: toy fair*)	0.0000435	Tableu	0.0000716
Bürgschaftsbank *(English: guarantee bank)*	0.0000416	Telefunk (Telefunken Enterprise)	0.0000666
Horregion (social program of Hanover, related to hearing)	0.0000408	Restlaufzeit (English: remaining term)	0.0000658
Transpond (Transponder)	0.0000407	Rückstell (Rückstellung, *English: provision*)	0.0000581

3.4. Score estimation and economic analysis

The final step of data processing is the calculation of the score, which reflects whether the firm-related news corpus has rather exploration or exploitation character. This score thus positions a firm on the continuum from exploitation to exploration and is further used as a dependent variable in the regression.

In order to obtain the score, the *td-idf* indicator was also calculated for each word of each firm's news corpus. Next, cosine similarity as an acknowledged measure of textual relatedness (Li & Han 2013; Muflikhah & Baharudin 2009; Feng 2020) was calculated. It measures the similarity between firm's news corpus and each of the dictionaries according to the following formula:

$$\boldsymbol{SCORE} = \frac{\sum_{j=1}^{m} \boldsymbol{firm_j} \cdot \boldsymbol{dict_j}}{\sqrt{\sum_{j=1}^{m} \boldsymbol{firm_j^2}}\sqrt{\sum_{j=1}^{m} \boldsymbol{dict_j^2}}}, \quad (3)$$

where $\boldsymbol{firm}$ – *tf-idf* in firm data set, $\boldsymbol{dict} – \boldsymbol{tf} - \boldsymbol{idf_{av}}$ in dictionary data set, *m* – number of terms.

Cosine similarity helps to identify how closely related the firm news corpus is to each of the dictionaries, with higher values of cosine similarity reflecting closer relation.

Consequently, the difference between cosine similarity scores for the exploitation and exploration dictionaries was calculated. This difference represents the dependent variable used in the analysis. It shows the position of each firm on the continuum from exploitation to exploration with values above zero meaning exploitative and values below zero explorative reporting about the particular firm. This value was calculated for the data sets for both periods, during and before the pandemic.

Finally, the OLS regression analysis is performed with *SCORE* as dependent and *SIZE* and *PATENTS* as the main independent variables:

$$SCORE = \beta_0 + \beta_1 \cdot SIZE + \beta_2 \cdot PATENTS + \beta \cdot CONTROLS + \varepsilon \quad (4)$$

OLS appeared to be suitable for the purpose of the analysis. The assumption check revealed no violations of main OLS assumptions (see Appendix). Several specifications of the model were calculated and presented: (1) using the full firm-level data set for the score during the pandemic, with control variables included; (2) using the full firm-level data set for the score during the pandemic to test the impact of firm size, with control variables excluded; (3) using the full firm-level data set for the score during the pandemic to test the impact of patenting, with control variables excluded; (4) using the partial firm-level data set, consisting of firms, which were found in the news both before and during the pandemic, with the score before the pandemic and control variables included; (5) using the partial firm-level data set with the score before the pandemic to test the impact of firm size, with control variables excluded; (6) using the partial firm-level data set with the score before the pandemic to test the impact of patenting, with control variables excluded; (7) using the partial firm-level data set, consisting of firms, which were found in the news both before and during the pandemic, with the score during the pandemic and control variables included; (8) using the partial firm-level data set with the score during the pandemic to test the impact of firm size, with control variables excluded; (9) using the partial firm-level data set with the score during the pandemic to test the impact of patenting, with control variables excluded.

4. Descriptive Results

In order to check whether all variables can be included into the analysis, the variance inflation factor (VIF) was initially examined. None of the variables achieved the critical value of ten with the average VIF of 1.725 for the data set during the pandemic and 1.579 before the pandemic. The individual VIF values range from 1.252 to 2.293 during the pandemic and from 1.204 to 4.817 before the pandemic. Thus, no variable was excluded from the analysis because of multicollinearity.

Table 4 Descriptive statistics presents descriptive statistics for the dependent, explanatory, and control variables[11] during and before the pandemic. It can be stated, that on average firms show explorative behavior both during and before the pandemic (as the *SCORE* variable has a negative mean value). Furthermore, larger and older firms, which more often are patenting and belong to a corporate group, were seen in the news corpus both during and before the pandemic. While during the pandemic around 2/3 of firms, for which reporting about innovation activities occurred, before the pandemic

[11] Excluding dummies.

only 1/3 of the firms in the sample were SMEs. Additionally, almost half of the firms reported about during the pandemic were patenting and led a corporate group. Stand-alone enterprises were seldomly seen in the data set both before and during the pandemic.

Table 4: Descriptive statistics

	Obs	Mean	St. dev.	Min	Median	Max
During pandemic						
SCORE (full data set)	993	-0.016	0.014	-0.069	-0.016	0.033
SCORE (partial data set)	328	-0.018	0.015	-0.069	-0.017	0.033
SIZE	993	0.669	0.471	0.000	1.000	0.000
NACE	978	Dummy				
NUTS	993	Dummy				
AGE	980	27.680	32.533	0.000	18.000	302.000
PATENTS	993	0.444	0.497	0.000	0.000	1.000
HOLDING	993	0.430	0.495	0.000	0.000	1.000
CORPGROUP	988	0.751	0.433	0.000	1.000	1.000
Before pandemic						
SCORE	328	-0.007	0.021	-0.058	-0.009	0.094
SIZE	328	0.338	0.474	0.000	0.000	1.000
NACE	324	Dummy				
NUTS	323	Dummy				
AGE	323	40.240	40.373	1.000	24.000	236.000
PATENTS	328	0.634	0.482	0.000	1.000	1.000
HOLDING	328	0.427	0.495	0.000	0.000	1.000
CORPGROUP	327	0.872	0.335	0.000	1.000	1.000

Table 5 SCORE distribution for groups of interest additionally shows the distribution of the variable *SCORE* for subsamples containing SMEs and large firms as well as for patenting and non-patenting firms, reflecting the hypotheses. The statistic is presented for all three data sets - all firms vs. the ones that were reported upon before and during the pandemic. Across all subsamples, firms on average tend to explorative innovation behavior, with exploration scores being higher for large and for patenting firms. Additionally, independent of the size and patenting activities, firms became more explorative during the pandemic. The most pronounced difference towards exploration can be seen for SMEs: while before the pandemic the average of the score was equal to -0.002, it became -0.014 during the pandemic.

Table 5: SCORE distribution for groups of interest

	Obs	Mean	St. dev.	Min	Median	Max
Full during pandemic						
SCORE SMEs	664	-0.014	0.013	-0.054	-0.016	0.033
SCORE Large	329	-0.019	0.015	-0.069	-0.017	0.022
SCORE Patenting	441	-0.018	0.015	-0.069	-0.017	0.020
SCORE Non-patenting	552	-0.014	0.013	-0.043	-0.014	0.033
Part before pandemic						
SCORE SMEs	111	-0.002	0.027	-0.040	-0.008	0.094
SCORE Large	217	-0.010	0.017	-0.058	-0.010	0.051
SCORE Patenting	208	-0.008	0.020	-0.058	-0.009	0.091
SCORE Non-patenting	120	-0.006	0.022	-0.046	-0.010	0.094
Part during pandemic						
SCORE SMEs	111	-0.014	0.014	-0.054	-0.016	0.033
SCORE Large	217	-0.020	0.015	-0.069	-0.018	0.022
SCORE Patenting	208	-0.020	0.016	-0.069	-0.018	0.020
SCORE Non-patenting	120	-0.014	0.014	-0.042	-0.016	0.033

Investigation of the industrial *(NACE)* and the geographical *(NUTS)* distribution of the firms is presented in Table 6 Industrial distribution of the firms and in the figure A11: Geographical distribution of the firms in the appendix. Almost 30% of the firms from the data set belong to manufacturing followed by 17.5% of firms from professional, scientific and technical activities, and almost 15% of firms are involved in the information and communication (IT) industry. It should be noted, that the proportion of IT firms increased during the pandemic, as they account for only 10.9% of the data set before the pandemic (whereas other top categories retained their proportion). In contrast, the proportion of wholesale and retail trade firms decreased significantly: from 13.7% before the pandemic to 11.9% during the pandemic. These trends show the necessity of digitalization and introduction of IT-related innovation, which became especially crucial during the first wave of the COVID-19 pandemic. On the other hand, retail industry was impaired by the lockdown and had to rely on the IT and digital innovations in order to resume performing their activities (e.g., online delivery, creative website services).

Table 6: Industrial distribution of the firms

NACE Section	Number of firms during the pandemic	Number of firms before the pandemic
A - Agriculture, Forestry and Fishing	1	1
B – Mining and Quarrying	3	0
C – Manufacturing	277	100
D - Electricity, Gas, Steam and Air Conditioning Supply	14	8
F – Construction	27	3
G - Wholesale and Retail Trade	118	45
H - Transporting and Storage	22	13
I - Accommodation and Food Service Activities	15	2
J- Information and Communication	146	36
K - Financial and Insurance Activities	60	22
L - Real Estate Activities	19	5
M - Professional, Scientific and Technical Activities	174	61
N - Administrative and Support Service Activities	41	12
O - Public Administration and Defense, Compulsory Social Security	6	4
P – Education	5	2
Q - Human Health and Social Work Activities	12	2
R - Arts, Entertainment and Recreation	6	3
S - Other Service Activities	32	5
Not available	15	5

The geographical distribution reflects the same trend during and before the pandemic, with most of the firms located in South and West Germany (particularly, federal states of Bavaria, Baden-Württemberg, and North Rhine-Westphalia) and less firms located in East Germany as well as in smaller German federal states, like Bremen and Saarland. Additionally, Hamburg was present with a higher fraction of firms before the pandemic, whereas Lower Saxony showed a higher fraction of firms during the pandemic. The graphical representation of the geographical distribution can be found in the appendix.

5. Econometric Results

To test the hypotheses of this study, OLS regressions are conducted. The results of these analyses are summarized in Table 7.[12]

The results show positive coefficients for *SIZE*, which predict a more exploitative behavior of SMEs in all specifications apart from model (7) where the coefficient is not significant. This finding contradicts former expectations that SMEs follow an explorative approach during the first year of the COVID-19 crisis. However, the size of the coefficients is smaller in the specifications (1), (2), and (8) compared to the models (4) and (5). This provides a hint that SMEs are oriented more towards exploration during crisis than they are in non-crisis times. Nevertheless, based on these findings, the hypothesis 1, stating that SMEs act explorative and large firms exploitative in their way to innovate during a crisis, is rejected.

The analyses show negative coefficients for *PATENTS* in the models (1), (3), (7) and (9). In the models (4) and (6) the relation is not significant. Patenting by firms relates to a more explorative innovation behavior during the crisis. Because patents require a significant degree of novelty to be successfully approved by the patent offices, this result confirms that firms that have experience with patenting follow an explorative approach. However, before the crisis, the relationship between patenting and the type of innovation activities is not significant. It appears as if a crisis brings along new conditions that open up new combinations of resources and new possibilities of fields of application, which requires an explorative discovery of solutions. Thus, the hypothesis 2 of this study, stating that patenting and, thus, technology-intensive firms are able to make use of new conditions during crisis, is confirmed.

The regression analyses also provided insights into the role of the control variables in predicting the type of innovation activities before and during the COVID-19 crisis. Firms that were classified as a headquarter were more likely to perform explorative innovation activities than their non-headquarter counterparts in the first model specification. This finding could be partially explained by the practice of many company groups to attribute innovations to the headquarter and not to the subsidiaries. In the analyzed news corpus, this could be reflected in the way that if it is written about an innovation of a company group, the subsidiary is not named explicitly but it is implicitly attributed to the headquarter.

[12] For the whole regression tables see the appendix.

Although the interpretability of coefficients for control variables in regressions is limited (Hünermund & Louw 2020), the results suggest that the regional environment could be a valuable avenue for future research, while belonging to a certain industry does not appear to be significant for the innovation behavior. The finding that the regional environment matters, corresponds to research in the vein of economic geography showing that regions exhibit a distinct innovation culture (Garretsen et al. 2019; Obschonka et al. 2013; 2015; Rentfrow et al. 2008). In contrast, belonging to a certain industry is not related to either explorative or exploitative innovation activities. This finding highlights that technology-intensive firms tend to perform explorative innovation activities independently of their industry classification. Thus, in our case, industry classification is not a good measure for technology intensity if a more specific variable is included in the model.

Table 7: Regression results

	Full data set during crisis			Partial data set before crisis			Partial data set during crisis		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SIZE	0.004*** (0.001)	0.005*** (0.001)		0.011*** (0.003)	0.008*** (0.002)		0.003 (0.003)	0.006*** (0.002)	
AGE	0.00002 (0.00002)			-0.00003 (0.00003)			0.00003 (0.00002)		
PATENTS	-0.003** (0.001)		-0.004*** (0.001)	0.0002 (0.003)		-0.002 (0.002)	-0.005** (0.002)		-0.006*** (0.002)
HOLDING	-0.003** (0.001)			-0.001 (0.003)			-0.004 (0.002)		
CORPGROUP	-0.001 (0.002)			-0.004 (0.005)			-0.003 (0.004)		
NACE dummies	Yes	No	No	Yes	No	No	Yes	No	No
NUTS dummies	Yes (sig)	No	No	Yes (sig)	No	No	Yes (sig)	No	No
Observations	958	993	993	316	328	328	316	328	328
Adjusted R^2	0.084	0.024	0.016	0.099	0.034	0.0001	0.060	0.029	0.036
F Statistic	3.376*** (df = 37; 920)	25.558*** (df = 1; 991)	17.315*** (df = 1; 991)	1.963*** (df = 36; 279)	12.338*** (df = 1; 326)	1.020 (df = 1; 326)	1.560** (df = 36; 279)	10.888*** (df = 1; 326)	13.335*** (df = 1; 326)

6. Discussion

As former research showed, economic crises lead to considerable shifts in innovation activities of different actors in the economy (Archibugi et al. 2013). In this vein, the size and former innovation experience of firms are found to impact the innovation activities during crises significantly. However, the COVID-19 crisis differs from former economic crises as the global spread of the virus represents a purely exogenous shock, independent from the economic system, over an unknown period of time (James 2020; Ratten 2020). Therefore, this study aims to reassess the relations between firm size, technology intensity, and the type of innovation activities conducted before and during the crisis. The study delivers three main results that come along with implications for policy-makers in times of economic crises.

First, the results provide evidence that SMEs and large firms generate different types of innovations before and during the crisis. This finding corresponds to former research that provided empirical evidence for differences in innovation behavior before and during crises (Archibugi et al. 2013). Thus, for designing impactful innovation policies to counteract the effects of economic downturns during crises, the dynamics of the innovation activities of SMEs and large firms are crucial to understand. Policies need to take SMEs and large firms separately into account rather than to follow a one-size-fits-all approach.

Second, it was shown that SMEs perform more exploitative innovation behavior than large firms both before as well as during the crisis, which led to the rejection of the hypothesis 1 of this study. Although both firm categories, SMEs and large firms, overall engage more in explorative innovation activities. In the case of SMEs, the difference of the exploration value before and during the crisis is greater than this difference for large firms. That means, in comparison to the period before the crisis, more SMEs engaged in explorative innovation during the crisis. For large firms this ratio did not change as much. However, large firms make up the greater share of companies that pursue explorative innovation activities during the COVID-19 crisis. This finding delivers evidence for Schumpeter Mark II dynamics in our analyzed time and geographical frame. As Schumpeter (1942) pointed out, large firms profit from knowledge accumulation in a path-dependent manner which converts innovation activities to a routine. This leads to the assumption that large firms are able to innovate continuously unaffected by economic shocks (Friz & Günther 2021). Our results confirm this notion. Consequently, for exploring solutions for current and future crises, the results of the study suggest that innovation policies should not only concentrate on SMEs, but also include large firms.

Third, firms with former patenting activities, reflecting technology-intensive firms, innovate exploratively during the crisis, confirming hypothesis 2 of this study. However, this relation only appears to be significant during the crisis and not before the crisis. This finding highlights that technology intensity enables firms to react to a dynamically changing environment by making use of new possibilities. Therefore, it is beneficial to support innovation activities in non-crisis times even though their impact on the type of innovations is limited. Technology intensity in non-crisis times strengthens the resilience of explorative innovation activities during crises. In light of great societal challenges, as the climate crisis, this is a crucial point to consider when designing innovation policies.

Additionally, the geographical environment matters for the type of innovation activities performed by incumbent firms. This result corresponds to findings in the vein of economic geography that find evidence for an only slowly changing local institutional setting that governs regional innovation activities (e.g., Urbano & Alvarez 2014; Feldman 2014; Huggins & Thompson 2020). Therefore, regional policy approaches are important to steer innovation activities of firms. Thus, it is suggested to include the distinct conditions and trajectories of regions into the design of impactful policies.

A review of supportive policy measures in the European Union conducted by the European Investment Bank states that political support was granted to firms that showed the most severe losses in sales, which was more pronounced in the category of small firms (Harasztosi et al. 2022). Thus, the policies were considering indirectly the size of firms. Although the support measures are found to increase the willingness of the supported firms to invest in digital technology (Harasztosi et al. 2022), the policies did not directly consider technology intensity as a criterion for support allocation. Based on the results of our study, future policies could take into consideration the innovation activities of firms when allocating resources. Furthermore, the conditions for explorative innovation activities by large firms during crisis should be fostered.

Nevertheless, the scope of this study is limited, paving the way for future research. First, the news data applied in this study covers only inventions and companies which are of public interest and may underestimate innovation activities that can be classified as process innovations or are done by small firms. However, news data enables the access to various kinds of innovation activities in contrast to specialized data sets as patent data which is only covering technological innovations. Further, the innovation activities of large companies might receive more media coverage than the ones of SMEs, as the immediate impact of large firms is probable to be higher in a region. This may lead to the underrepresentation of some minor innovation activities; however, the most important innovation activities are reported on anyways. Additionally, limitations were also found in the stemming procedure, as the algorithm for German stemming cannot perfectly assign stems to words. However, stemming allows for the unification of most of the terms (Birkholz et al. 2021).

Second, connected to the finding that headquarters have a significant impact on innovation activities, it can be the case that names of subsidiaries were abbreviated to the firm name of the headquarter. This could potentially have led to an overproportioned attribution of innovation activities to headquarters instead of the subsidiaries.

Third, changes over time in firm size, location, and industry specifications were not taken into account in this study. Future research could investigate the growth and shrinkage of firms in relation to the innovation activities. However, for the scope of this study this effect is unlikely to impact the results as size was measured as a binary variable (SME or large firm) and, due to the rather short time period under investigation, switching of categories is unlikely.

Fourth, this study proposes a new way to classify innovation activities with no time lag that is able to deliver real-time insights. It may be a beneficial avenue for future research to examine how this measurement of explorative and exploitative innovation activities could be complemented by further indicators and validated by comparing this classification with commonly used ones, like for example patent class recombination.

Last, the methodological approach could be conveyed to study further time periods to analyze the innovation activities of SMEs and large firms as well as those of other actors in different pandemic situations and other crises.

References

Acs, Z. J., & Audretsch, D. B. (1987). Innovation, market structure, and firm size. The review of Economics and Statistics, 567-574.

Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: an empirical analysis. The American economic review, 678-690.

Acs, Z., Desai, S., & Hessels, J. (2008). Entrepreneurship, economic development and institutions. Small Business Economics, 31(3), 219–234. https://doi.org/10.1007/s11187-008-9135-9.

Antonelli, C., & Scellato, G. (2015). Firms size and directed technological change. Small Business Economics, 44(1), 207–218. https://doi.org10.1007/s11187-014-9593-1.

Archibugi, D., Filippetti, A., & Frenz, M. (2013). The impact of the economic crisis on innovation: Evidence from Europe. Technological Forecasting and Social Change, 80(7), 1247-1260. https://doi.org/10.1016/j.techfore.2013.05.005.

Archibugi, D., Filippetti, A., & Frenz, M. (2013). Economic crisis and innovation: is destruction prevailing over accumulation? Research Policy, 42(2), 303-314. https://doi.org/10.1016/j.respol.2012.07.002.

Arnaldi, S. (2008). Converging technologies in the Italian daily press 2002–2006: preliminary results of an ongoing research project. Innovation: The European Journal of Social Science Research, 21(1), 87-94. https://doi.org/10.1080/13511610802002304.

Audretsch, D. B. (2002). The dynamic role of small firms: Evidence from the US. Small Business Economics, 18(1), 13-40. https://doi.org/10.1023/A:1015105222884.

Ba, S., & Bai, H. (2020). COVID-19 Pandemic as an Accelerator of Economic Transition and Financial Innovation in China. Journal of Chinese Economic and Business Studies, 18(4), 341–348. https://doi.org/10.1080/14765284.2020.1855394.

Balasubramanian, N., & Lee, J. (2008). Firm age and innovation. Industrial and Corporate Change, 17(5), 1019-1047. https://doi.org/10.1093/icc/dtn028.

Baum, J. A., Calabrese, T., & Silverman, B. S. (2000). Don't go it alone: Alliance network composition and startups' performance in Canadian biotechnology. Strategic management journal, 21(3), 267-294. https://doi.org/10.1002/(SICI)1097-0266(200003)21:3<267::AID-SMJ89>3.0.CO;2-8.

Baumol, W., (2002). Entrepreneurship, innovation and growth: the David-Goliath symbiosis. The Journal of Entrepreneurial Finance and Business Ventures, 7(2), 1-10.

Bhattacharya, M, & Bloch, H. (2004). Determinants of Innovation. Small Business Economics, 22, 155-162. https://doi.org/10.1023/B:SBEJ.0000014453.94445.de.

Birkholz, J., Günther, J., & Sholnykova, M. (2021). Using Topic Modelling in Innovation Studies: The Case of a Small Innovation System under Conditions of Pandemic Related Change. Bremen Papers on Economics & Innovation, 2101. https://doi.org/10.26092/elib/451.

Bloom, N., Bunn, P., Mizen, P., Smietanka, P., & Thwaites, G. (2020). The Impact of Covid-19 on Productivity. NBERWorking Papers 28233, National Bureau of Economic Research, Inc.

Braga, I. A., Monard, M. C., & Matsubara, E. T. (2009). Combining unigrams and bigrams in semi-supervised text classification. Proceedings of Progress in Artificial Intelligence, 14th Portuguese Conference on Artificial Intelligence (EPIA 2009), Aveiro, 489-500.

Cassiman, B., & Veugelers, R. (2006). In search of complementarity in innovation strategy: Internal R&D and external technology acquisition. Management Science, 52(1), 68–82. https://doi.org/10.1287/mnsc.1050.0470.

Castrogiovanni, G. J., Domenech, J., & Mas-Verdú, F. (2012). Variations in SME Characteristics and the Use of Service Intermediaries for R&D. Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 29(2), 154-164. https://doi.org/10.1002/cjas.231.

Cefis, E., & Orsenigo, L. (2001). The persistence of innovative activities: A cross-countries and cross-sectors comparative analysis. Research Policy, 30(7), 1139-1158. https://doi.org/10.1016/S0048-7333(00)00139-6.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. Handbook of the Economics of Innovation, 1, 129-213. https://doi.org/10.1016/S0169-7218(10)01004-X.

Corradini, C., Battisti, G., & Demirel, P. (2015). Serial innovators in the UK: Does size matter? Industrial and Corporate Change, 25(1), 23-47. https://doi.org/10.1093/icc/dtu040.

Corradini, C., Demirel, P., & Battisti, G. (2016). Technological diversification within UK's small serial innovators. Small Business Economics, 47(1), 163-177. https://doi.org/10.1007/s11187-015-9698-1.

D'Agostino, L., & Moreno, R. (2018). Exploration during turbulent times: an analysis of the relation between cooperation in innovation activities and radical innovation performance during the economic crisis. Industrial and Corporate Change, 27(2), 387-412. https://doi.org/10.1093/icc/dtx035.

Dachs, B., Hud, M., Koehler, C., & Peters, B. (2017). Innovation, creative destruction and structural change: firm-level evidence from European countries. Industry and Innovation, 24(4), 346-381. https://doi.org/10.1080/13662716.2016.1261695.

Dahl, D. W., & Moreau, C. P. (2002). The influence and value of analogical thinking during new product ideation. Journal of Marketing Research, 39(1), 47–60. https://doi.org/10.1509/jmkr.39.1.47.18930.

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. Research Policy, 11, 147–162. https://doi.org/10.1016/0048-7333(82)90016-6.

Dosi, G., & Nelson, R. R. (2010). Chapter 3 – Technical Change and Industrial Dynamics as Evolutionary Processes. Handbook of the Economics of Innovation, 1, 51-127. https://doi.org/10.1016/S0169-7218(10)01003-8.

Ebersberger, B., & Kuckertz, A. (2021). Hop to it! The impact of organization type on innovation response time to the COVID-19 crisis. Journal of Business Research, 124, 126-135. https://doi.org/10.1016/j.jbusres.2020.11.051.

Falck, O. (2009). Routinization of innovation in German manufacturing: the David–Goliath symbiosis revisited. Industrial and Corporate Change, 18(3), 497-506. https://doi.org/10.1093/icc/dtn018.

Feldman, M. P. (2014). The Character of Innovative Places: Entrepreneurial Strategy, Economic Development, and Prosperity. Small Business Economics, 43(1), 9–20. https://doi.org/10.1007/s11187-014-9574-4.

Feng, S. (2020). The proximity of ideas: An analysis of patent text using machine learning. PLOS ONE, 15(7), e0234880. https://doi.org/10.1371/journal.pone.0234880.

Fritsch M., Greve M., Wyrwich M. (2021). The COVID-19 Pandemic and Entrepreneurship in Germany. Foresight and STI Governance, 15(4), 42–51. https://doi.org/10.17323/2500-2597.2021.4.42.51.

Friz, K., & Günther, J. (2021). Innovation and economic crisis in transition economies. Eurasian Business Review, 11(4), 537-563. https://doi.org/10.1007/s40821-021-00192-y.

Gabler Wirtschaftslexikon (2018). Mittelstand. https://wirtschaftslexikon.gabler.de/definition/mittelstand-40165/version-263557. Last accessed on 31.03.2021.

Garretsen, H., Stoker, J. I., Soudis, D., Martin, R., & Rentfrow, J. (2019). The Relevance of Personality Traits for Urban Economic Growth: Making Space for Psychological Factors. Journal of Economic Geography, 19(3), 541–565. https://doi.org/10.1093/jeg/lby025.

Gourinchas, P.-O., Kalemli-Özcan, Ş., Penciakova, V., & Sander, N. (2021). COVID-19 and Small- and Medium-Sized Enterprises: A 2021 'Time Bomb'? AEA Papers & Proceedings, 111, 282–286. https://doi.org 10.1257/pandp.20211109.

Harasztosi, P., Maurin, L., Pál, R., Revoltella, D., & van der Wielen, W. (2022). Firm-level policy support during the crisis: So far, so good? EIB Working Paper, 2022/01. https://doi.org/10.2867/24308.

Harjoto, M. A., Rossi, F., & Paglia, J. K. (2021). COVID-19: Stock Market Reactions to the Shock and the Stimulus. Applied Economics Letters, 28(10), 795–801. https://doi.org/10.1080/13504851.2020.1781767.

Hollenstein, H. (2005). Determinants of International Activities: Are SMEs Different? Small Business Economics, 24(5), 431–450. https://doi.org/10.1007/s11187-005-6455-x.

Huggins, R., & Thompson, P. (2020). Human Agency, Entrepreneurship and Regional Development: A Behavioural Perspective on Economic Evolution and Innovative Transformation. Entrepreneurship and Regional Development, 32(7–8), 573–589. https://doi.org/10.1080/08985626.2019.1687758.

Hünermund, P., & Louw, B. (2020). On the nuisance of control variables in regression analysis. arXiv preprint arXiv:2005.10314.

James, H. (2020). Seven Transformative Crises from European Revolution to Corona: Globalization and State Capacity. Financial History Review, 27(2), 139–159. https://doi.org/10.1017/S0968565020000098.

Kalemli-Ozcan, S., Gourinchas, P.-O., Penciakova, V., & Sander, N. (2020). COVID-19 and SME Failures. IMFWorking Papers 2020/207, International Monetary Fund.

Kitching, J., Blackburn, R., Smallbone, D., & Dixon, S. (2009). Business strategies and performance during difficult economic conditions.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. Organization Science, 3, 393–397. https://doi.org/10.1287/orsc.3.3.383.

Kortelainen, J.T., Ratinen, M., Linnanen, L. 2012. Attitudes Towards Energy Efficiency and Renewable Energy in European Small and Medium-Sized Enterprises, in Vazquez-Brust, D.A. and Sarkis, J. (eds) Green Growth: Managing the Transition to a Sustainable Economy: Learning By Doing in East Asia and Europe. Dordrecht: Springer Netherlands, 229-243. https://doi.org/10.1007/978-94-007-4417-2_12.

Latham, S. (2009). Contrasting Strategic Response to Economic Recession in Start-Up versus Established Software Firms. Journal of Small Business Management, 47(2), 180-201. https://doi.org/10.1111/j.1540-627X.2009.00267.x.

Li, B., & Han, L. (2013). Distance weighted cosine similarity measure for text classification. International conference on intelligent data engineering and automated learning, Springer, Berlin, Heidelberg, 8206, 611-618. https://doi.org/10.1007/978-3-642-41278-3_74.

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization science, 2(1), 71-87. https://doi.org/10.1287/orsc.2.1.71.

Marsili, O. (2001). The Anatomy and Evolution of Industries. Books. Edward Elgar Publishing, 2272.

Muflikhah, L., & Baharudin, B. (2009). Document clustering using concept space and cosine similarity measurement. 2009 International conference on computer technology and development, 1, 58-62. https://doi.org/10.1109/ICCTD.2009.206.

Nehrebecka, N. (2021). COVID-19: Stress-Testing Non-financial Companies: A Macroprudential Perspective: The Experience of Poland. Eurasian Economic Review, 11(2), 283–319. https://doi.org/10.1007/s40822-020-00163-0.

Nelson, R. R. (1985). An evolutionary theory of economic change. Harvard University Press.

Nemlioglu, I., & Mallick, S. K. (2020). Do innovation-intensive firms mitigate their valuation uncertainty during bad times? Journal of Economic Behavior & Organization, 177, 913-940. https://doi.org/10.1016/j.jebo.2020.06.004.

Noori, J., Nasrabadi, M. B., Yazdi, N., & Babakhan, A. R. (2017). Innovative performance of Iranian knowledge-based firms: Large firms or SMEs? Technological Forecasting and Social Change, 122, 179-185. https://doi.org/10.1016/j.techfore.2016.04.025.

O'Toole, C., McCann, F., Lawless, M., Kren, J., & McQuinn, J. (2021). New Survey Evidence on COVID-19 and Irish SMEs: Measuring the Impact and Policy Response. The Economic and Social Review, 33.

Obschonka, M., Schmitt-Rodermund, E., Silbereisen, R. K., Gosling, S. D., & Potter, J. (2013). The regional distribution and correlates of an entrepreneurship-prone personality profile in the United States, Germany, and the United Kingdom: A socioecological perspective. Journal of Personality and Social Psychology, 105(1), 104–122. https://doi.org/10.1037/a0032275.

Obschonka, M., Stuetzer, M., Gosling, S. D., Rentfrow, P. J., Lamb, M. E., Potter, J., & Audretsch, D. B. (2015). Entrepreneurial Regions: Do Macro-Psychological Cultural Characteristics of Regions Help Solve the "Knowledge Paradox" of Economics? PLOS ONE, 10(6), e0129332. https://doi.org/10.1371/journal.pone.0129332.

Ortega-Argilés, R., Vivarelli, M., & Voigt, P. (2009). R&D in SMEs: a paradox? Small Business Economics, 33(1), 3-11. https://doi.org/10.1007/s11187-009-9187-5.

Ozgun, B., & Broekel, T. (2021). The geography of innovation and technology news-An empirical study of the German news media. Technological Forecasting and Social Change, 167, 120692. https://doi.org/10.1016/j.techfore.2021.120692.

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A Theory of the Emergence, Persistence, and Expression of Geographic Variation in Psychological Characteristics. Perspectives on Psychological Science, 3(5), 339–369. https://doi.org/10.1111/j.1745-6924.2008.00084.x.

Rizea, I. C., Parpandel, D. E., Caldararu, A. E., & Radu, A. L. (2011). Current Approaches Regarding the Knowledge Management Impact on SMEs Performance. Economia: Seria Management, 14(2), 563–568.

Rochina-Barrachina, M. E., Mañez, J. A., & Sanchis-Llopis, J. A. (2010). Process innovations and firm productivity growth. Small Business Economics, 34, 147-166. https://doi.org/10.1007/s11187-008-9110-5.

Sahut, J. M., & Peris-Ortiz, M. (2014). Small business, innovation, and entrepreneurship. Small Business Economics, 42(4), 663-668. https://doi.org/10.1007/s11187-013-9521-9.

Sanidas, E. (2014). How SMEs or Larger Firms and Industries' Productivity Respond to Technology: A Panel Data Study. SPOUDAI Journal of Economics and Development, 64(1), 16–28.

Schumpeter, J.A. (1934). The Theory of Economic Development. Harvard University Press, Cambridge.

Schumpeter, J. A. (1942). Socialism, capitalism and democracy. Harper and Brothers.

Silge, J., & Robinson, D. (2017): Text mining with R: A tidy approach. Sebastopol, CA.

Sivadas, E., & Dwyer, F. R. (2000). An examination of organizational factors influencing new product success in internal and alliance-based processes. Journal of Marketing, 64(1), 31–49. https://doi.org/10.1509/jmkg.64.1.31.17985.

Spithoven, A., Vanhaverbeke, W., & Roijakkers, N. (2013). Open innovation practices in SMEs and large enterprises. Small Business Economics, 41(3), 537–562. https://doi.org/10.1007/s11187-012-9453-12.

Sternberg, R., & Arndt, O. (2001). The Firm or the Region: What Determines the Innovation Behavior of European Firms? Economic Geography, 77(4), 364-382. https://doi.org/10.1111/j.1944-8287.2001.tb00170.x.

Taymaz, E., & Üçdoğruk, Y. (2009). Overcoming the double hurdles to investing in technology: R&D activities of small firms in developing countries. Small Business Economics, 33(1), 109–128. https://doi.org/10.1007/s11187-009-9181-y.

Triguero, A., Córcoles, D., & Cuerva, M. C. (2014). Persistence of innovation and firm's growth: Evidence from a panel of SME and large Spanish manufacturing firms. Small Business Economics, 43(4), 787–804. https://doi.org/10.1007/s11187-014-9562-9.

Urbano, D., & Alvarez, C. (2014). Institutional Dimensions and Entrepreneurial Activity: An International Study. Small Business Economics, 42(4), 703–716. https://doi.org/10.1007/s11187-013-9523-7.

Urbańska, K., Parkitna, A., & Kubicka, J. (2021). Success of SMEs in the Era of Pandemics. European Research Studies, 24(2), 407-423. https://doi.org/10.35808/ersj/2134.

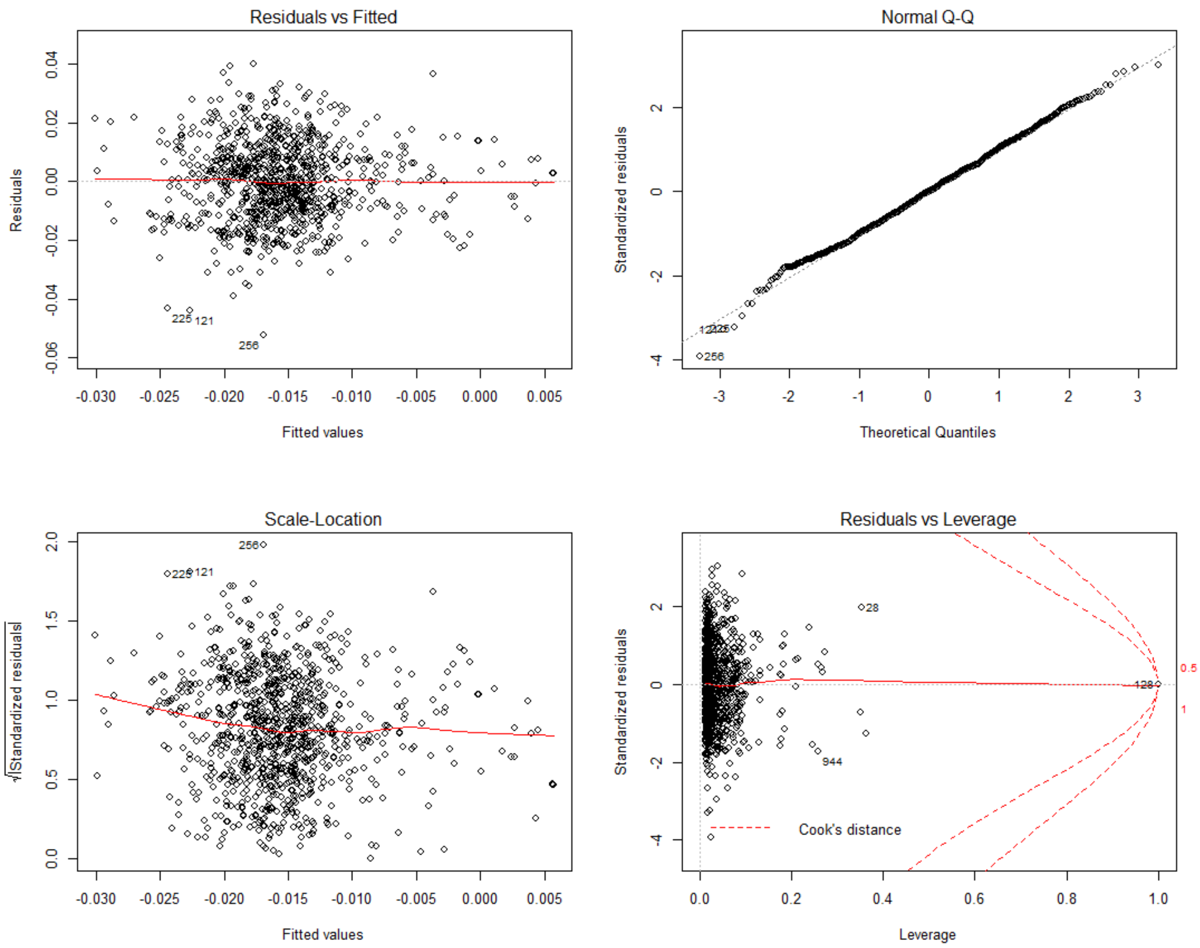
von Bloh, J., Broekel, T., Özgun, B., & Sternberg, R. (2020). New (s) data for entrepreneurship research? An innovative approach to use big data on media coverage. Small Business Economics, 55(3), 673-694. https://doi.org/10.1007/s11187-019-00209-x.

Vossen, R. W. (1998). Relative strengths and weaknesses of small firms in innovation. International Small Business Journal, 16(3), 88-94. https://doi.org/10.1177/0266242698163005.

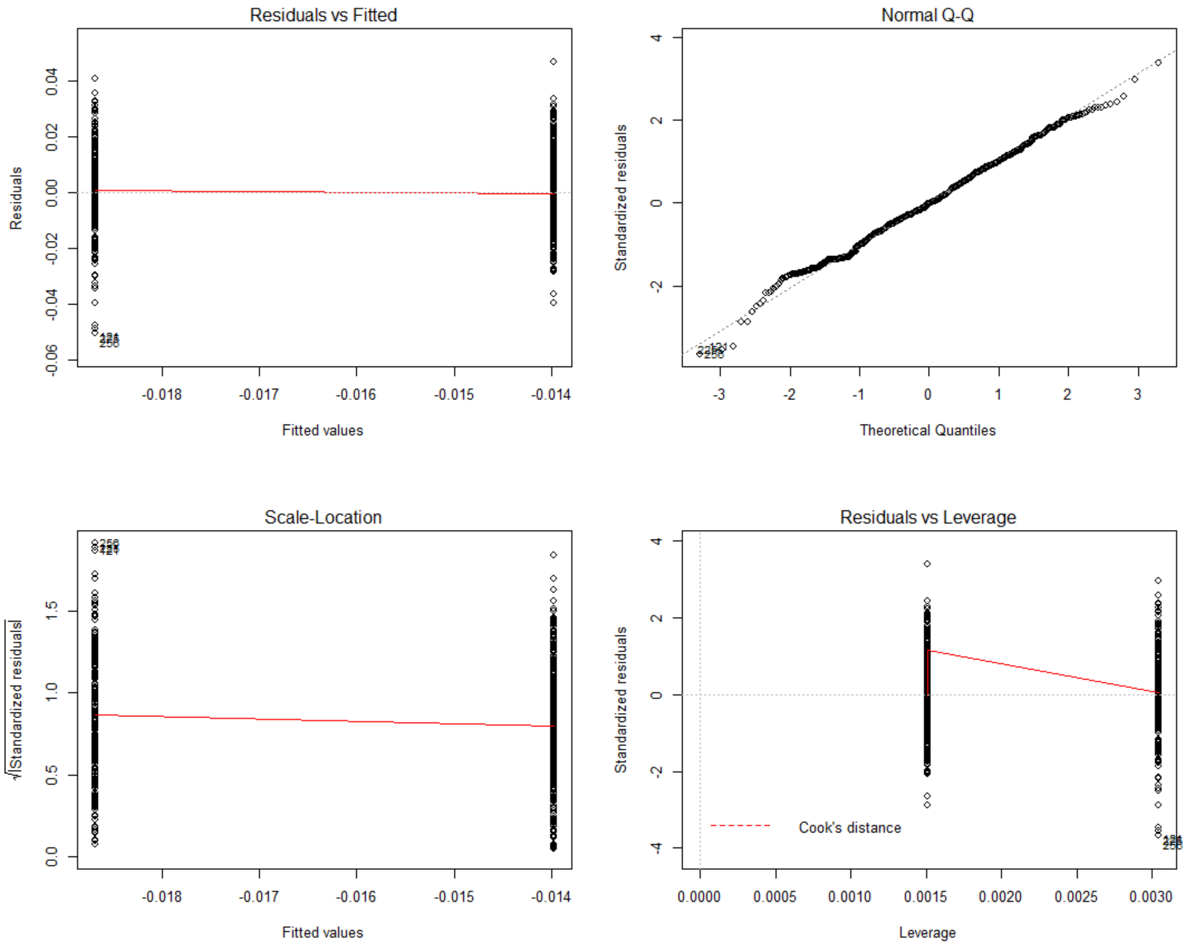
Zeroual, I., & Lakhouaja, A. (2017). Arabic information retrieval: Stemming or lemmatization? 2017 Intelligent Systems and Computer Vision (ISCV), IEEE, 1-6. https://doi.org/10.1109/ISACV.2017.8054932.

Appendix

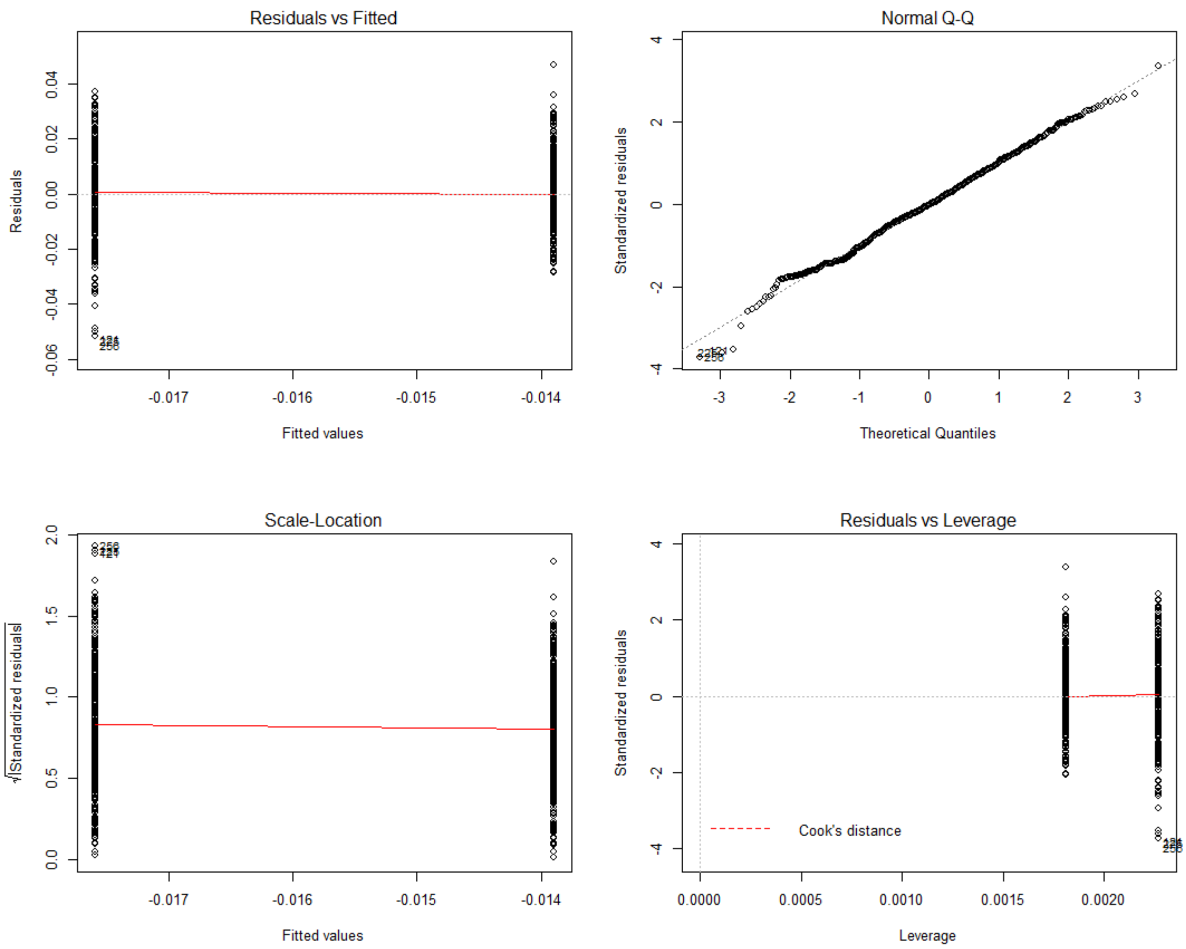
A1: Assumptions check for the specifications full data set during the crisis with controls



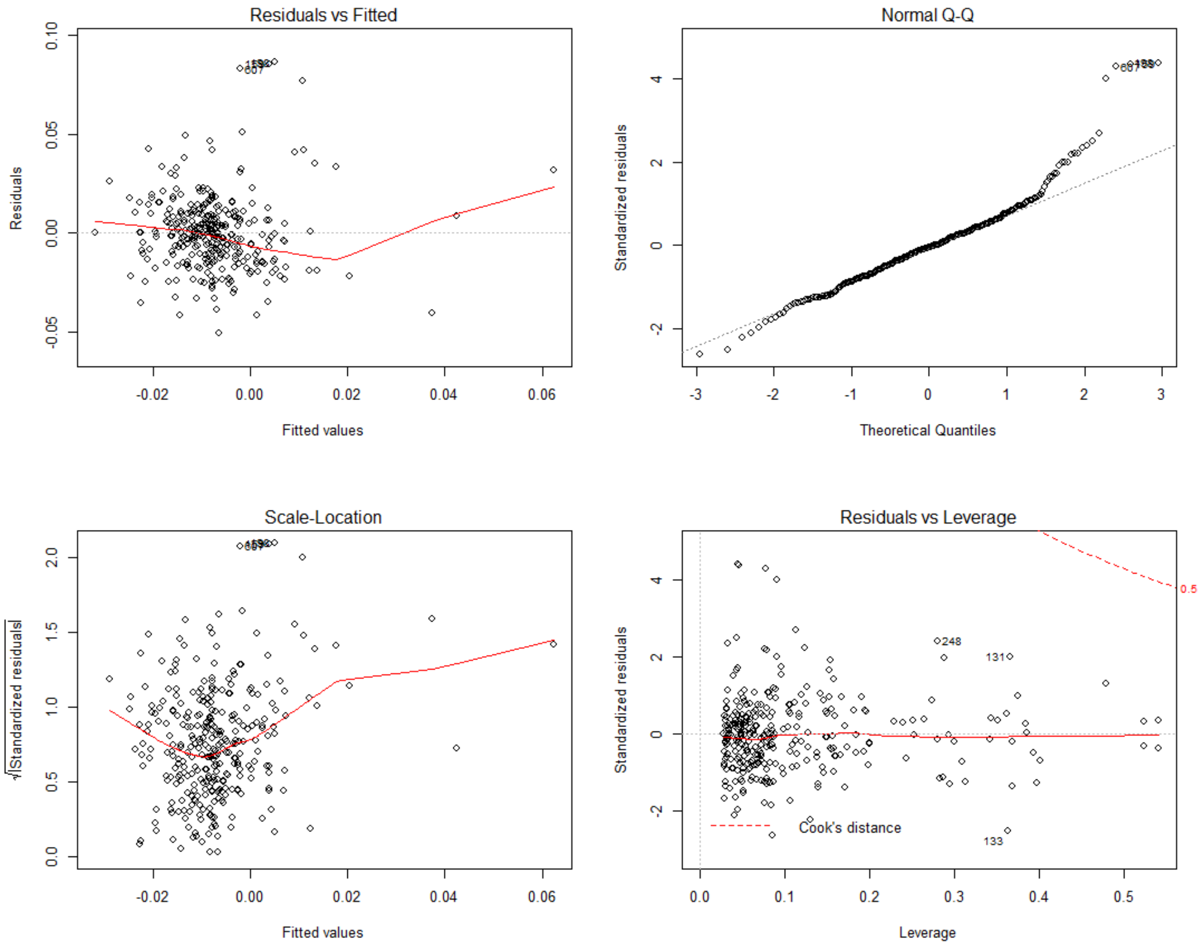
A2: Assumptions check for the specifications full data set during the crisis without controls



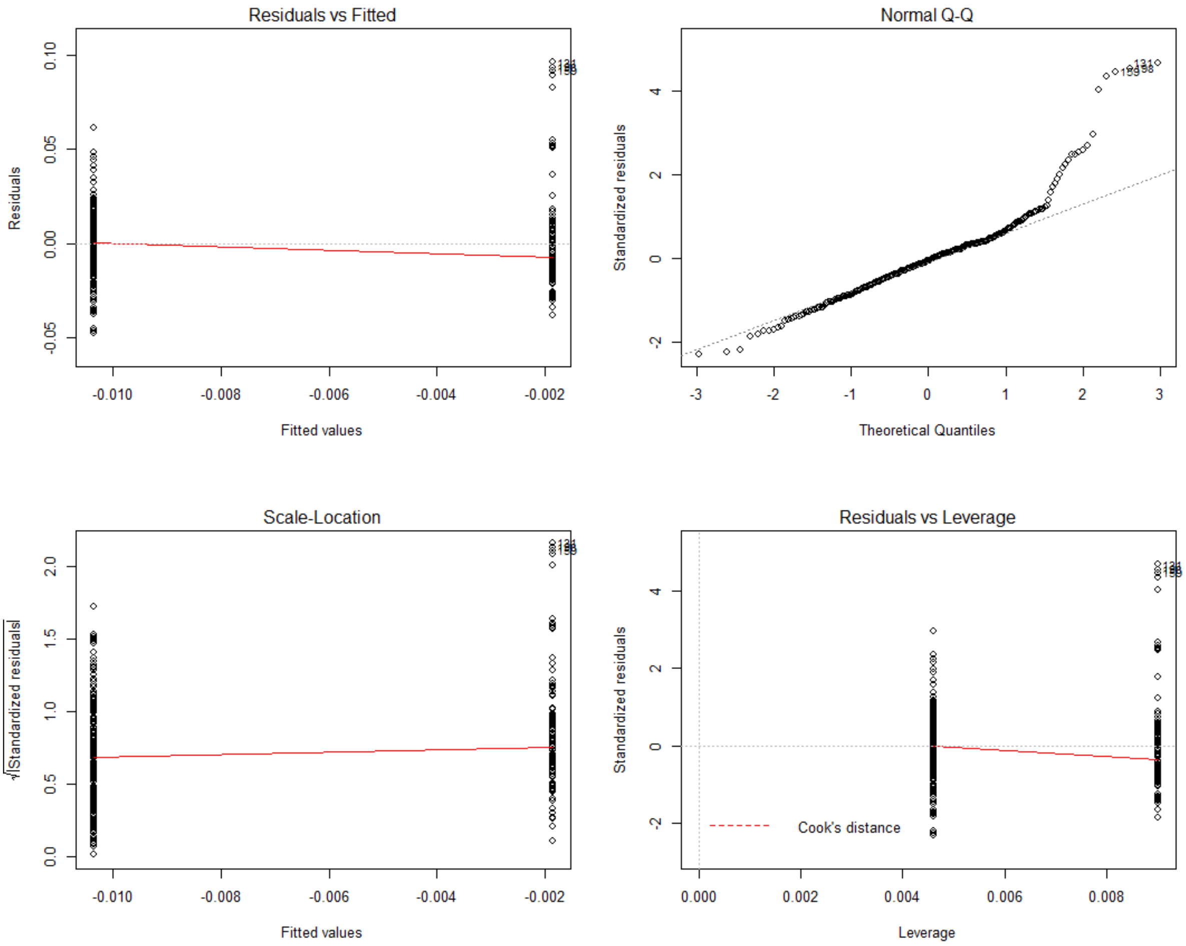
A3: Assumptions check for the specifications full data set during the crisis including patenting without controls



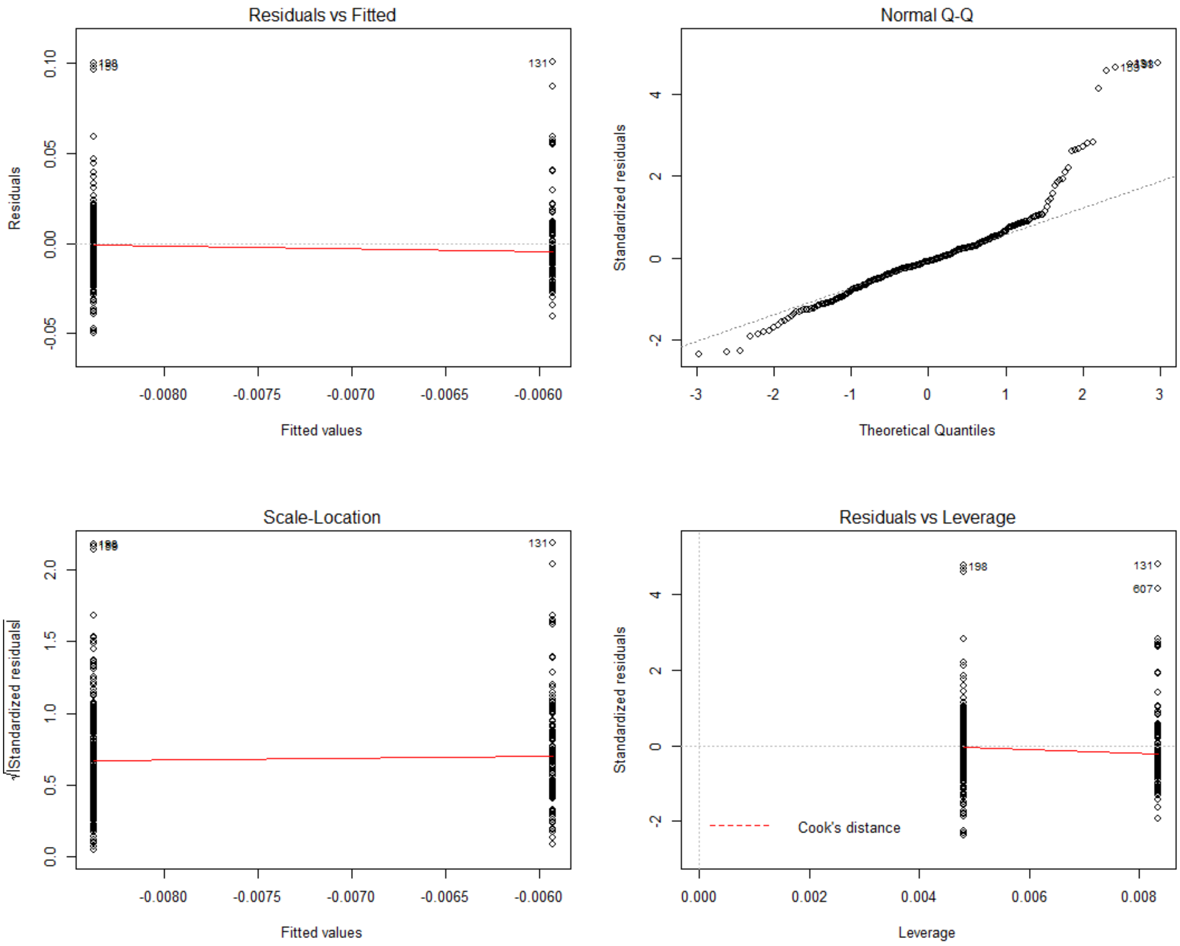
A4: Assumptions check for the specifications partial data set before the crisis with controls



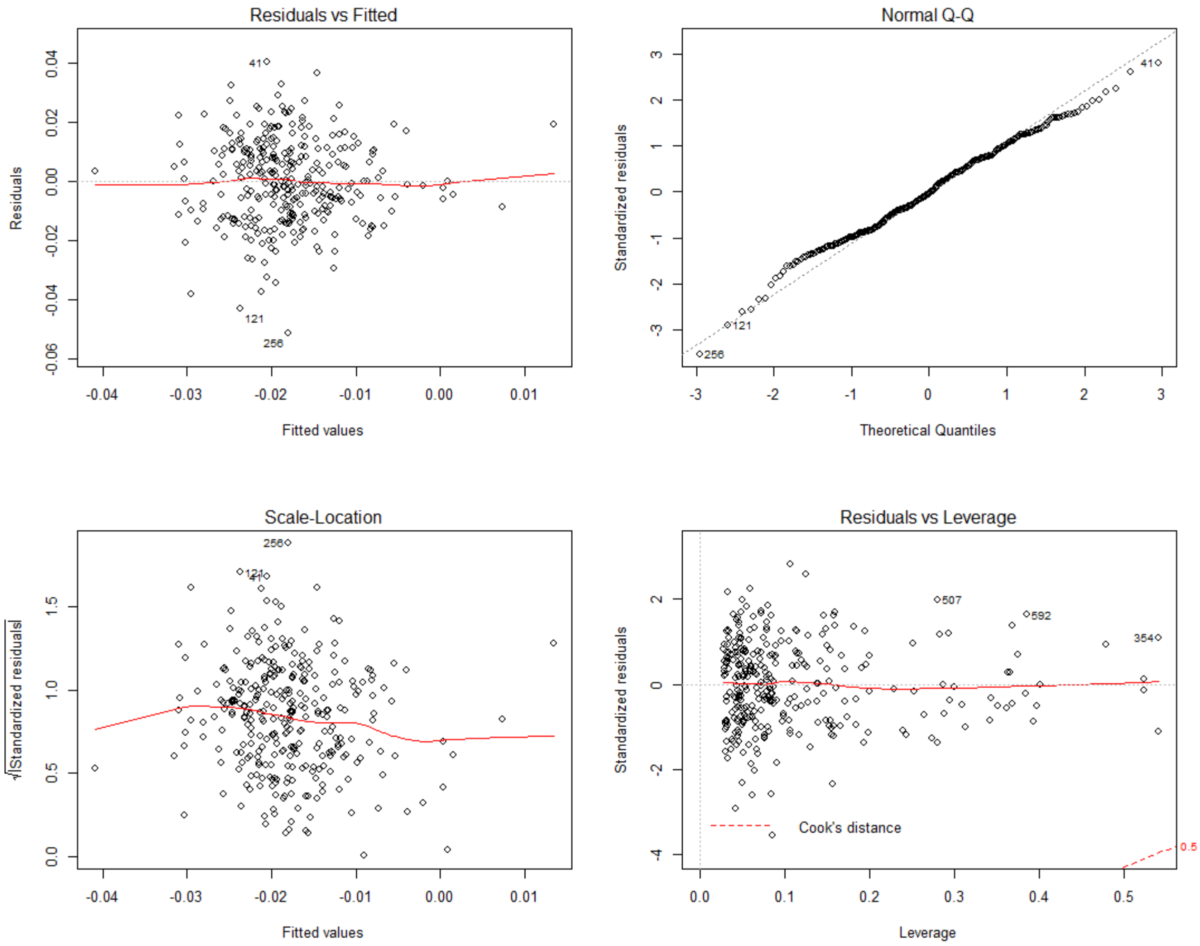
A5: Assumptions check for the specifications partial data set before the crisis without controls



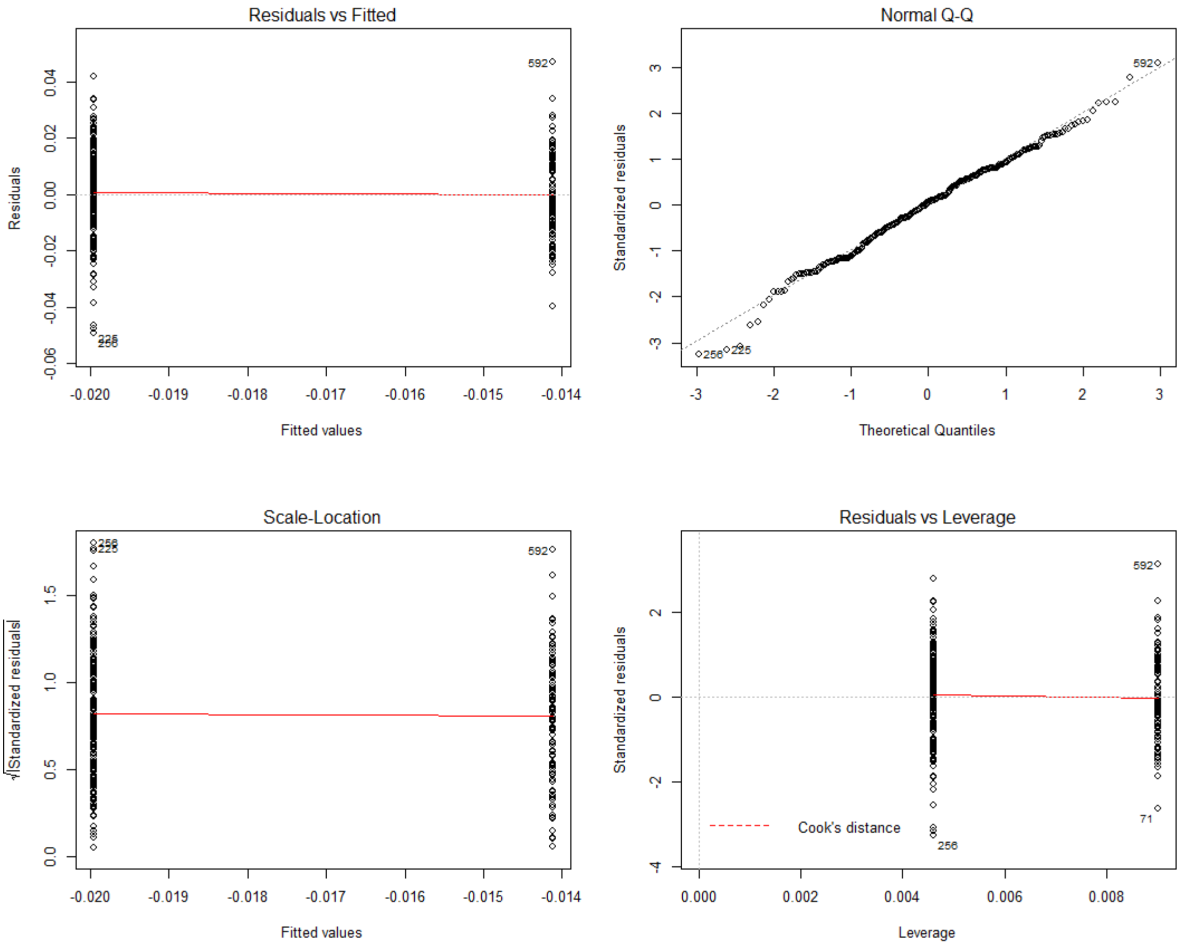
A6: Assumptions check for the specifications partial data set before the crisis including patenting without controls



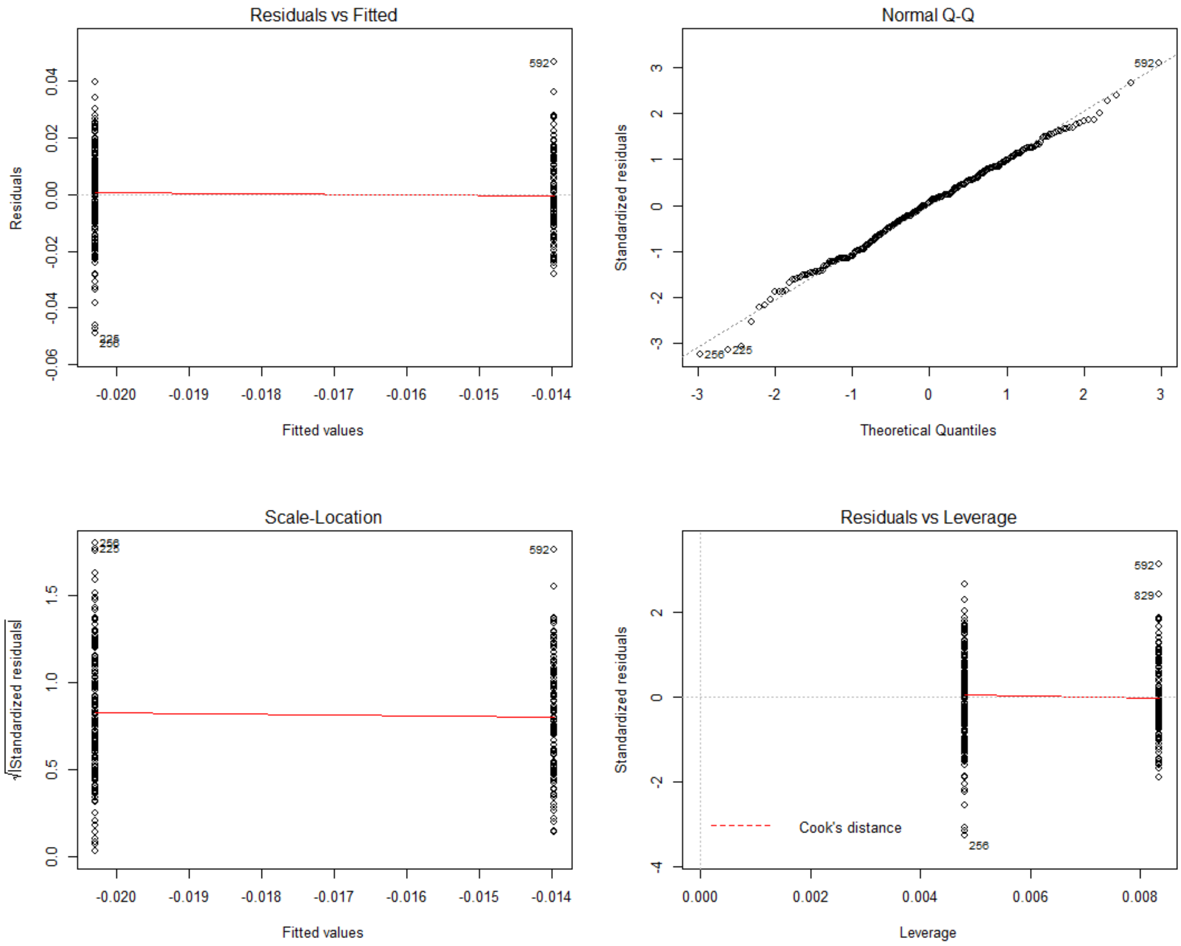
A7: Assumptions check for the specifications partial data set during the crisis with controls



A8: Assumptions check for the specifications partial data set during the crisis without controls



A9: Assumptions check for the specifications partial data set during the crisis including patenting without controls



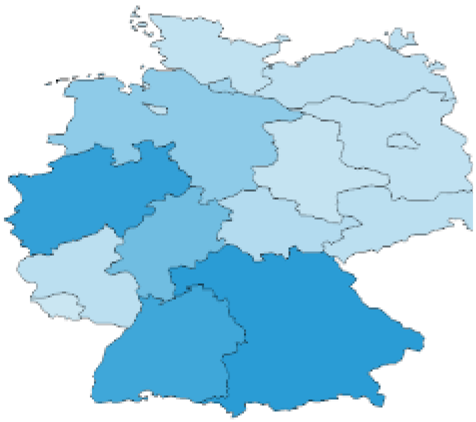
A10: Full regression results

	Full data set during crisis			Partial data set before crisis			Partial data set during crisis		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SIZE	0.004*** (0.001)	0.005*** (0.001)		0.011*** (0.003)	0.008*** (0.002)		0.003 (0.003)	0.006*** (0.002)	
AGE	0.00002 (0.00002)			-0.00003 (0.00003)			0.00003 (0.00002)		
PATENTS	-0.003** (0.001)		-0.004*** (0.001)	0.0002 (0.003)		-0.002 (0.002)	-0.005** (0.002)		-0.006*** (0.002)
HOLDING	-0.003** (0.001)			-0.001 (0.003)			-0.004 (0.002)		
CORPGROUP	-0.001 (0.002)			-0.004 (0.005)			-0.003 (0.004)		
NACE Section B	-0.015 (0.016)								
NACE Section C	-0.001 (0.014)			0.014 (0.021)			0.002 (0.016)		
NACE Section D	-0.010 (0.014)			0.014 (0.022)			-0.004 (0.017)		
NACE Section F	-0.001 (0.014)			0.001 (0.024)			0.006 (0.018)		
NACE Section G	-0.004 (0.014)			0.013 (0.021)			0.003 (0.016)		
NACE Section H	-0.005 (0.014)			0.012 (0.022)			0.001 (0.016)		
NACE Section I	0.003 (0.014)			0.023 (0.026)			0.001 (0.019)		

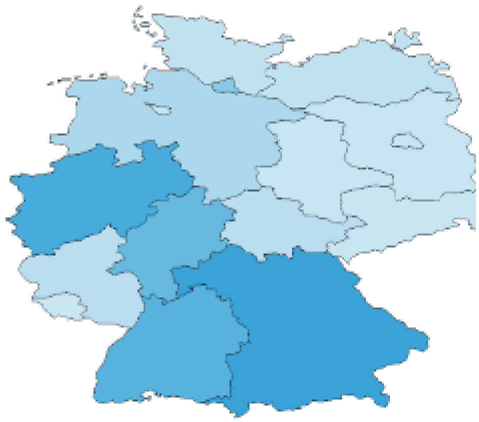
NACE Section J	**-0.006 (0.014)**			**0.0003 (0.021)**			**-0.003 (0.016)**		
NACE Section K	-0.002 (0.014)			0.002 (0.021)			0.002 (0.016)		
NACE Section L	-0.005 (0.014)			0.011 (0.024)			0.009 (0.018)		
NACE Section M	-0.005 (0.014)			0.007 (0.021)			0.001 (0.016)		
NACE Section N	-0.003 (0.014)			0.008 (0.022)			-0.004 (0.016)		
NACE Section O	-0.008 (0.015)			0.012 (0.023)			-0.007 (0.017)		
NACE Section P	0.003 (0.015)			-0.002 (0.029)			-0.013 (0.022)		
NACE Section Q	-0.008 (0.014)			0.0002 (0.026)			0.003 (0.019)		
NACE Section R	0.003 (0.015)			0.002 (0.024)			0.009 (0.018)		
NACE Section S	-0.003 (0.014)			-0.0003 (0.023)			0.007 (0.017)		
NUTS1 DE2	0.001 (0.001)			0.002 (0.004)			0.001 (0.003)		
NUTS1 DE3	0.0001 (0.002)			-0.005 (0.006)			-0.001 (0.004)		
NUTS1 DE4	0.012*** (0.004)			0.022** (0.011)			0.022*** (0.008)		
NUTS1 DE5	0.002 (0.004)			0.014 (0.011)			0.011 (0.008)		

NUTS1 DE6	**0.004*** (0.002)**			**0.002 (0.005)**			**0.008** (0.004)**		
NUTS1 DE7	0.0003 (0.002)			0.006 (0.004)			0.002 (0.003)		
NUTS1 DE8	0.0003 (0.002)			-0.004 (0.009)			0.011 (0.007)		
NUTS1 DE9	0.002 (0.002)			0.001 (0.006)			0.005 (0.005)		
NUTS1 DEA	0.001 (0.002)			0.002 (0.004)			0.003 (0.003)		
NUTS1 DEB	0.003 (0.003)			0.012 (0.008)			0.002 (0.006)		
NUTS1 DEC	0.022*** (0.006)			-0.021 (0.024)			-0.004 (0.018)		
NUTS1 DED	0.009*** (0.003)			0.060*** (0.012)			-0.010 (0.009)		
NUTS1 DEE	0.017*** (0.004)			0.011 (0.011)			0.009 (0.008)		
NUTS1 DEF	0.002 (0.004)			-0.001 (0.009)			0.010 (0.007)		
NUTS1 DEG	-0.004 (0.003)			-0.006 (0.008)			-0.0002 (0.006)		
Constant	-0.014 (0.014)	-0.019*** (0.001)	-0.014*** (0.001)	-0.017 (0.021)	-0.010*** (0.001)	-0.006*** (0.002)	-0.016 (0.016)	-0.020*** (0.001)	-0.014*** (0.001)
Observations	958	993	993	316	328	328	316	328	328
Adjusted R^2	0.084	0.024	0.016	0.099	0.034	0.0001	0.060	0.029	0.036
F Statistic	3.376*** (df = 37; 920)	25.558*** (df = 1; 991)	17.315*** (df = 1; 991)	1.963*** (df = 36; 279)	12.338*** (df = 1; 326)	1.020 (df = 1; 326)	1.560** (df = 36; 279)	10.888*** (df = 1; 326)	13.335*** (df = 1; 326)

A11: Geographical distribution of the firms



During the pandemic



Before the pandemic

Imprint

Bremen Papers on Economics & Innovation

Published by
University of Bremen, Faculty of Business Studies & Economics,
Institute for Economic Research and Policy (ierp)
Max-von-Laue-Straße 1, 28359 Bremen, Germany

Bremen Papers on Economics & Innovation #2203

Responsible Editor: Prof. Dr. Jutta Günther

ISSN 2629-3994