

Assessing press releases as a data source for spatial research

Burcu Ozgun¹, Tom Broekel²

¹ University of Kassel, Kassel, Germany

² Stavanger University Business School, Stavanger, Norway

Received: 25 July 2021/Accepted: 31 May 2022

Abstract. Describing the distribution and development of socio-economic activities in space is frequently limited due to data availability, as official statistical data sources are often restricted to specific topics and geographical scales. Consequently, new and alternative data sources are needed as complements. This paper presents, discusses, and empirically explores an example of such an alternative data source that promises to give detailed and novel insights into regions' socio-economic structures: press releases. While press releases have seen some use in the literature, they have hardly seen any applications in regional research. Consequently, it is still largely unknown to what extent their information content gives systematic insights into regional socio-economic patterns. The paper closes this gap by assessing the degree to which the content of press releases corresponds to the socio-economic characteristics of regions, by means of an empirical analysis of German NUTS-3 regions. The results confirm that press release content varies systematically between regions, and regional differences in socio-economic characteristics, events, and activities seem to be well-presented by press releases. This suggests that press releases offer substantial potential and might be a useful complementary data source in regional studies.

1 Introduction

Information on economic activities is at the heart of the empirical literature in economic spatial sciences. Unless researchers collect their own primary data, most contemporary studies rely on official secondary data sources. While the primary data collection requires massive efforts and is simply unfeasible in many situations, the use of secondary data sources implies that researchers are restricted to look at socio-economic developments using available data. This may not always give a precise representation of the phenomenon in question. Finding new data sources that capture socio-economic processes in novel ways or from new perspectives is therefore imperative to empirical research.

This imperative motivates the present study, which presents a data source that has seen some applications in the literature, yet has not been discussed nor evaluated with respect to its usefulness for spatial economic research at the sub-national level: press releases. A press release is a written statement about a matter that might be of public interest, written by an organization related to the matter. Organizations issue press releases to inform the public about noteworthy news and recent developments. Press releases contain rich and useful textual information covering a wide range of topics in domains like political (e.g., announcements by local administrations), social (e.g.,

sport and music events), and economic (e.g., updates on new products and firms). Not surprisingly, their value as a data source has been recognized for some time (Johnson, Haythornthwaite 1989, Berger et al. 2019, von Bloh et al. 2020). However, the extent to which they contain systematic information for describing socio-economic structures at the sub-national (regional) level has not been explored so far. The present paper conducts such an assessment and discusses whether press releases represent a useful data source in this context from an empirical basis. More precisely, we argue that the content of press releases is systematically shaped by regions' socio-economic structures, which in consequence allows inferences about the latter by (empirically) observing the former.

We test this argument by assessing the spatial correspondence of information obtained from geo-located press releases of Germany's biggest press release agency to that of regions' general characteristics. At the level of 401 districts (NUTS-3 regions), the study confirms that inter-regional variations in the events reported in these press releases correlate well with general socio-economic characteristics of regions. For instance, population density, income level, and touristic potential of regions are clearly reflected in the content of press releases. As there are little to no reasons for why this correlation should be systematically different for other and more specific content, our findings substantiate the promise that press releases are a complementary data source to official statistics in spatial research at the sub-national level, which is likely to capture socio-economic patterns and processes not covered by these. An example of the latter is the study by von Bloh et al. (2020) that use press releases to capture the entrepreneurial activities and sentiments towards it in regions.

The remainder of the paper is organized as follows. Section 2 discusses why press releases might be a valuable data source in spatial sciences. Section 3 presents the empirical basis for testing press releases' correspondence to regional characteristics. It outlines the data collection and cleaning process and describes the overall structure of the data set. Section 4 discusses the employed empirical set-up of the assessment. Section 5 presents the empirical results and discusses the findings. Section 6 concludes.

2 Press releases: An underutilized data source in spatial studies

In the age of big data, empirical information has been crucial for one's understanding of the world. The analysis of documented and codified facts and narratives is a crucial task in all social sciences independent of the analysis being quantitative or qualitative in nature. Frequently, it is the availability and richness of data that shapes what researchers can investigate and what perspectives they can apply. However, equally true is that all data are "biased" and "incomplete," meaning that each data set only offers a selective and particular view of the world. In sum, high-quality and reliable science requires researchers to utilize and exploit a range of distinct data sets.

This is not any different in spatial sciences, where the interest is frequently to assess, describe, and explain structural differences at the sub-national level, i.e., variations between regions. Doing this requires empirical data that provides comparable and representative pictures of what socio-economic activities are happening within regions. Unfortunately, researchers are frequently limited to the use of official secondary data sources. While the topics covered by this type of data have been greatly expanded in size and improved in quality, they still cover only specific portions of spatial socio-economic structures. For instance, information on the average value-added, unemployment, and investments are usually reliable and available in most countries as part of official statistics. They also usually exist for multiple years and at varying spatial scales. In contrast, processes like entrepreneurship, environmental initiatives, social events, types of sports conducted, etc. are less documented in official statistics (see, e.g., Stuetzer et al. 2018).¹ Yet, such activities are not less relevant. Sport is a major economic factor and this is similarly true for social activities. They both play crucial roles in the emergence and development of social networks and local cultures that facilitate knowledge diffusion, learning, and innovation, which in turn are essential for regions' long-term economic

¹Further limitations and shortcomings of official statistics are discussed in greater length elsewhere (see, e.g., Pérez 2006, Bean 2016)

development (Hofstede et al. 2010, Gertler 1997, Saxenian 1996). Consequently, it is essential for researchers to seek and explore novel data sources that complement existing ones by either allowing for looking at the same activities and events from an alternative perspective or by capturing aspects completely outside of the scope of official statistics. These may include activities and events that are recent phenomena and hence are not yet captured by such statistics. In this case, it takes unusual approaches and greater efforts by researchers to create empirical information capturing their existence and spatial distribution.

Recent technological developments and improvements in computational methods opened new toolboxes to obtain, structure, and transform data from various alternative sources. Frequently, these sources are not originally intended to be used in this context but provide novel opportunities for spatial research nonetheless (Arribas-Bel 2014). For example, recent studies show that insights on firms' activities and their spatial interactions can be obtained from website data (Kinne, Axenbeck 2018, Kinne, Lenz 2021). Similarly, social media data, such as Twitter, allow researchers to explore topics like well-being and social attitudes. (Mitchell et al. 2013, Lansley, Longley 2016).

The present paper contributes to the literature by discussing a data source that has been around for a long time, but that has hardly seen any utilization in the spatial sciences: press releases. Press releases issued by companies (Ahern, Sosyura 2014, Henry 2008, Dahlin et al. 2006), by political parties (Senninger, Wagner 2015, Klüver, Sagarzazu 2016), by local governments (Boussalis et al. 2018), and by health organizations (Park, Reber 2010) have been valuable information sources for researchers. Although several studies indicated press releases' potential value (Feldman, Lowe 2015, von Bloh et al. 2020), they have not yet been used on a larger basis in spatial sciences.

Press releases provide important information about their issuers and their activities. In some cases, both the issuer and the event have a spatial dimension. The latter motivates the present paper, as the events reported in press releases may give insight into what is happening in specific locations. In other words, there are good reasons to suspect that information in press releases reflect regional events and activities. Given the great variety of issuers and topics, it is more than likely that they contain information on socio-economic issues.

Press releases are public statements created by governments, companies, research organizations and the like, which are delivered to inform the media and targeted audiences. In other words, press releases are *told only to be retold* (Jacobs 1999). By being one of the most important public relation tools, press releases provide journalists with crucial *raw materials*, as they are delivered regularly, reliably, and contain usable information (Walters, Walters 1992, Hong 2008, Turk 1986, Aronoff 1976). Accordingly, mass media and these information sources are in a so-called symbiotic relationship, due to reciprocity of interest and economic necessity (Herman, Chomsky 2010). Not surprisingly, studies show that the reliance of journalists on public relations statements and press releases is extensive (Erjavec 2005, Lewis et al. 2008, O'Neill, O'Connor 2008, Reich 2010, Macnamara 2014). Crucially, press releases are not news. News is a commodity that is shaped by the dynamics of supply and demand (Hamilton 2004). Since all events, people, regions, and discussions cannot be covered daily by the news media, all news outlets, e.g., local newspapers, broadcast television networks, or international news agencies, have to make decisions of what to report and what not to. The primary determinant of which events are reported is their newsworthiness (Shoemaker 2006). Newsworthiness is related to both the inherent characteristics of an event and journalists' assessments (Staab 1990, Kepplinger, Ehmig 2006, Galtung, Ruge 1965, Harcup, O'Neill 2017, Eilders 2006). From a set of potential newsworthy events, media outlets filter what is perceived to be the most newsworthy for their own audience (Boukes, Vliegthart 2020). Even when a story is considered newsworthy, it is not automatically published. For example, on a slow news day, minor events can become news that wouldn't have been reported on a news-rich day (Gans 2004).

On the other hand, press releases are subject to just a fraction of this selection process. They do not have to be considered relevant or newsworthy for anyone but the issuer. At least in the short run, it is the issuer's urge to supply information rather than their

desire to satisfy demand. Nevertheless, it can be expected that the issuer will invest the efforts of producing a press release only if they expect some demand for it from the press or another kind of audience. In any case, press releases do not go through a similar multi-stage selection, filtering, and editorial processes as news, which is known to induce all sorts of biases (Robinson, Sparkes 1976, Kariel, Rosenvall 1984, Chang et al. 1987, Wu 2003, 2007, Himmelboim et al. 2010, Segev 2015). Consequently, in contrast to news, they are much less subject to these biases built into the journalistic system. However, that doesn't mean that they are unbiased. In many instances, an individual press release is more subjective than a news piece that has made it through a more rigorous journalistic process. Crucially, their information content is limited to what the issuer wants to share, and they tend to be written in a positive tone (Maat, de Jong 2013). Not surprisingly, it is known that for press releases to become news, they need to be rewritten to match the (hopefully more objective and rigorous) editorial and stylistic requirements (Walters et al. 1994).

While the subjectivity and a potential positivity bias will translate to the regional level, when press release information is aggregated, at least the degree of subjectivity is likely to "average out" with larger numbers of (heterogeneous) issuers in a location. Indeed, this number will in many cases be larger than the number of news outlets in a region, which frequently are a much smaller and more homogeneous (in terms of education, qualification, and motivation) group, i.e., journalists. Consequently, in many circumstances, and conditional on a large number of issuers, press releases might be a more direct and less-biased information source than news. We also argue that the positivity bias is less of a problem when looking at press releases from a spatial perspective because it would need to have a distinct and systematic inter-regional variance to impact cross-regional analyses. This logic similarly applies to the sentiments of press releases, which are likely to be very positive in general, yet, it is only their inter-regional variance that matters in regional analyses. Nevertheless, they are still subject to specific selection processes implying that they are not useful information sources in all circumstances. They are less likely to give insights into negative events.

While press releases are different from news, they have a similar potential to identify regional particularities and differences. The heterogeneity of issuers and events they deem to be press release-worthy is substantial. The low effort of writing a press release makes it quite likely that events with a relatively minor level of (social) significance will show up in press releases. The non-restrictiveness of what they can inform about gives more room to heterogeneous events and activities being featured. On top of that, it is their textual nature that gives broader possibilities to extract information. In any case, the frequency with which events in specific locations are covered in press releases may give further insights into the observed or unobserved spatial distribution of the underlying events.

The purpose of this study is to unveil the potential of press releases in reflecting regional characteristics based on aggregate numbers of events reported. For this, we adopt a straightforward approach: We expect that specific categories of events found in press releases are related to specific regional characteristics. For instance, events related to economic activities are more likely to occur in regions with plenty of economic actors (banks, insurance); locations attractive to tourists are likely to feature many entertainment events, and so on. Consequently, we seek to assess the empirical fit between the frequency of events of a certain kind reported in press releases in a specific location, and general characteristics of regions obtained from official statistics that should relate to them.

3 Data

We work with a collection of press releases brought together by the *Presseportal*, a subsidiary of DPA (Deutsche Presse Agentur). DPA is Germany's leading press agency (Kleinstauber, Thomass 2007) and the *Presseportal* is the largest and most popular press release portal in Germany.² Their services are used by more than 12,000 organizations, which are represented with their own newsrooms. The portal attracts around 20 million

²<https://www.presseportal.de>

visitors per month ([Presseportal 2021](#)). The agency operates without a clear thematic focus (except for a geographical one on Germany) and provides only minimal professional journalistic editing and selection process. This implies that the press release collection covers a wider variety of topics and events. For example, companies share their financial reports, new product announcements, innovations, new store openings; universities, research institutes, and medical associations share recent findings; local governments share new developments in their regions; local media outlets share recent local events and incidents; sports magazines share information on games and scores; political parties share statements made by their members; cultural venues share upcoming event information. In sum, the portal provides a collection of events and developments which are found worth sharing by their creators. Accordingly, we can expect this data to carry a wealth of up-to-date information on what these organizations do in specific places.

We accessed and downloaded press release data from *Presseportal* for four years (from May 2016 to May 2020). Releases were retrieved daily by using the portal's API service. In total, we downloaded 140,833 press releases. For all press releases, we recorded the title, full text, date, location, name of the issuing organization, and keywords. While this text data is generally of high quality, several steps were necessary before it could be analyzed.

First, some press releases include TV program information published by TV broadcasters. These releases usually review the content of movies and TV series implying that in most cases, they represent fictional information. Consequently, we removed them from our data.

Second, the location given in the press releases refers to the location of its issuer, which might be distinct from the location of content, i.e., where the event described in the release takes place. For example, a press release about an upcoming play in Hamburg might be assigned to Berlin as the location because the entertainment company issuing the release is located there. In the context of the present paper, we are more interested in where the event takes place. For this reason, we applied a two-step procedure to geo-locate the press release content. In the first step, we identified the names of locations mentioned in the press releases using an adapted string-matching approach based on an extensive list of location names in Germany. Given that the location names in Germany can be a combination of multiple words and some names may refer to multiple locations, we had to develop a somewhat more sophisticated string-matching approach including intensive cleaning and the use of additional information, to identify the correct locations. When present in the text, we assigned multiple locations to an individual press release. It implies that press releases are counted multiple times when the press releases are distributed across regions. The multiple counting is reasonable, as multiple locations being named in a press release is most likely due to the underlying event having place in multiple locations or multiple events being referred to. With this approach, we were able to assign 55,776 press releases to at least one location in Germany. In the second step, we assigned all press releases that did not contain location information, to the location of the issuer, as given in the original database. After the two steps, 99,984 press releases were assigned to one or more locations in Germany.³ The assigned locations were later aggregated to the NUTS-3 level. More than 80% of all press releases were assigned to just one NUTS-3 region. Figure 1 gives a first impression of the spatial distribution of press releases.

The map reveals a strong imbalance in the geographic distribution of press releases. Of 401 districts, only one either had not been covered by press releases, or our approach could not identify any location in the region. 197 districts had been covered in less than fifty press releases, implying that they have been reported on about less than once a month, on average. The most frequently covered region was Berlin; it was mentioned 26,559 times. One-third of all press releases referred to four (of 401) regions; namely Berlin, Dusseldorf, Munich, and Hamburg, which are among the most populous German cities. Other regions shown in dark colors on the map are Cologne, Frankfurt, and Stuttgart,

³The remaining press releases were either issued by organizations outside of Germany or did not contain any location name in Germany. Further details on the geolocation procedure can be found in Appendix A.

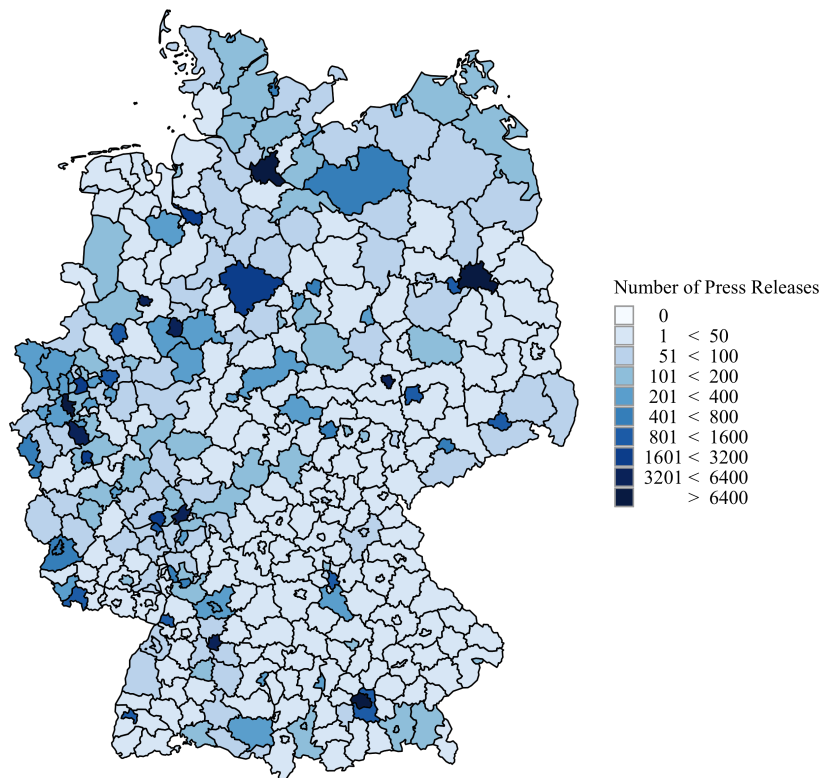


Figure 1: Spatial distribution of press releases

which are also among the largest cities in Germany. This observation does not come as a surprise. Districts with a greater population are usually the ones where larger numbers of events are taking place, and where more issuers are located. Other than this, we do not observe any strong spatial pattern such as spatial clustering, or differences between federal states. Spatial autocorrelation also appears to be largely absent.⁴

Third, there were 12,327 distinct organizations issuing press releases. Unfortunately, for these organizations, we do not have background information other than their (often incomplete) names. To get at least some idea about who they are, we manually classified all issuers with more than 100 releases, which in sum contribute about 50% of all press releases. The majority of them are firms (32%), non-profit associations⁵ (25%), and newspapers (25%).⁶ Future research might be able to match these names to register data, to learn more about their areas of activity, size, and type.

Fourth, press releases in the data set cover a wide variety of topics. Fortunately, the press release portal associates each press release with one or multiple keywords. In total, there are 283 distinct keywords in the dataset. On average, an article is associated with three keywords. The keywords' frequency distribution is very skewed. One-third of the keywords do not show up in more than ten press releases, i.e., they appear to be very case-specific. In contrast, some keywords are used very frequently, suggesting that the press releases seem to be simultaneously assigned to general and specific keywords. Manual inspection of a randomly selected set of press releases confirmed a good fit between the press releases' content and the assigned keywords. Figure 2 shows the frequency of the most common 20 keywords. Clearly, words associated with politics and the economic situation dominate. However, words related to tourism, health, entertainment, and sports themes are also relatively common.

Figure 3 visualizes the most frequent keyword for each NUTS-3 region to give an idea

⁴A Moran's I test for spatial autocorrelation confirms this with a test statistic -0.01 and p-value 0.67 .

⁵As indicated by an "e.V." being part of their name, which applies to all formally registered non-profit organizations / associations in Germany.

⁶The coverage rate and selection of news articles issued as press releases are unknown at this stage.

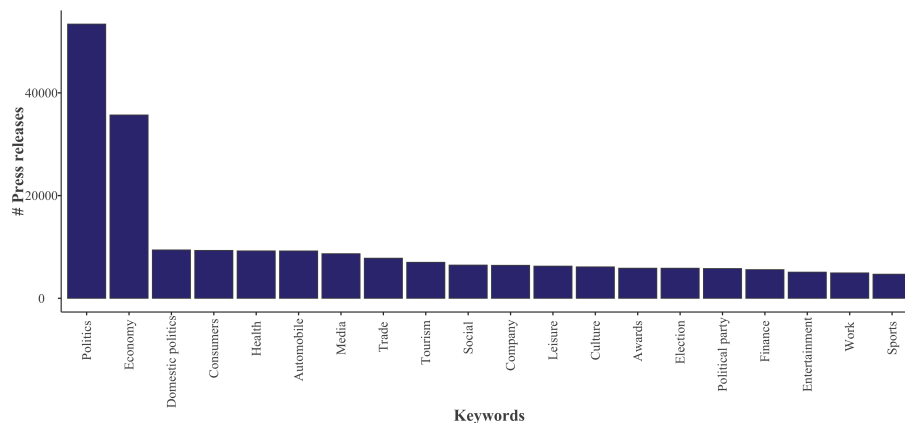


Figure 2: Most common keywords and frequency in press releases

about the keywords' distribution in space. Of 283 keywords, 13 turn out dominant in at least one district, with *economy* and *politics* being dominant in 184 and 132 districts, respectively. The distribution of general keywords does not contain insights into inter-regional differences. However, looking at less frequent keywords gives the first impression of a potential fit between press releases' content and regional characteristics. For example, the majority of press releases referring to the city of *Mönchengladbach* are categorized to the keyword *Sports*. This reflects that the city is strongly associated with its soccer team *Borussia Mönchengladbach*, which plays in the Bundesliga. Another illustrative example is *Saarlouis* for which the most frequent keyword is *Automobile*. Correspondingly, the largest employer in this town is the Ford Motor Company's body and assembly facility. These examples show how the portrayal of regions in press releases varies and how significant events and sectors shape the content of press releases linked to these places. In the following, we will explore this link in a more systematic way.⁷

4 Empirical approach

Our analysis seeks to assess the degree to which the content of press releases corresponds to the socio-economic characteristics of regions. Empirically, we test if regional characteristics explain the frequency with which events are associated with corresponding regions. In our baseline model, the total number of press releases associated with each region, *PRESS*, is the dependent variable, which is related to a set of regional socio-economic factors to assess which of these relate to the quantities of press releases. Afterwards, we estimate four more separate models in which the dependent variables correspond to the number of press releases representing specific topics (i.e., keywords). While the set of explanatory variables, i.e., the set of regional socio-economic factors remains the same in each model, it is usually one of them that is contextually most closely related and hence is expected to show the strongest empirical association with the dependent variable.

The size of regions is approximated by their population, which is denoted by *POP*. We capture urbanization with the variable *POPDENS*, which is the number of resident individuals per square kilometer. The economic success of regions is approximated by the gross domestic product per capita (*GDPC*).⁸ All three variables capture the overall frequency with which events take place in regions, in general. That is, larger and more economically successful regions are likely to experience more "press-release" worthy events.

A variable that refers to a more specific regional characteristic is tourism (*TOUR*), which shows the touristic attractiveness of regions, measured by the number of overnight stays in tourist facilities per inhabitant. Straightforwardly, we expect the frequencies of press releases related to tourism and leisure to be empirically associated with this variable.

⁷The anonymized version of the dataset is available at: <https://github.com/burcuozgun/grpr>.

⁸All data are obtained from the statistical office of the European Union, i.e., Eurostat and *INKAR*, the interactive online atlas of the Federal Institute for Building, Urban Affairs and Spatial Development of Germany (<https://www.inkar.de>).

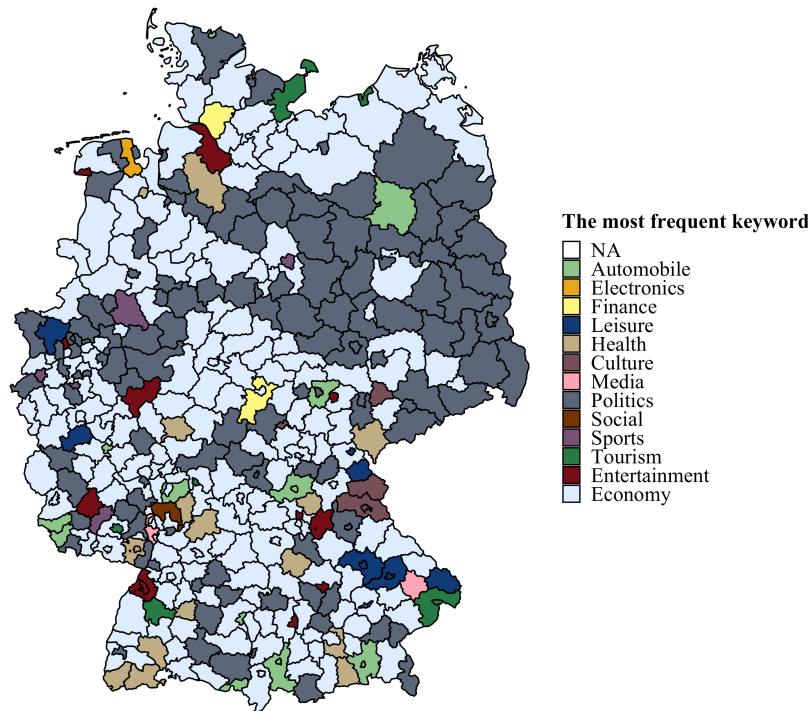


Figure 3: The most frequent keyword in each region

Peoples' interest in politics is captured by voter turnout in regions (*VOTE*). It is defined as the percentage of eligible voters that participated in the federal elections in 2017.⁹ As for tourism, this variable represents a specific set of activities and consequently, should be mirrored by the frequency of political events mentioned in press releases. We also test for correspondence of the intensity of regional innovation activities and these types of events being mentioned in press releases. Innovation activities of regions are approximated by the number of patent applications (*PAT*).

To isolate the relationships between these socio-economic characteristics and press release frequencies, we consider several control variables. The first one is the dummy variable *EAST*, which has a value of one for regions in former East Germany or zero otherwise. Given the peculiar history of the two parts of Germany, these differences might be reflected in the general practice of using press releases. While we don't have direct evidence, systematic differences in journalistic activities are still known to exist (Haller 2012) and it is likely that they extend to the practice of press releases.

The German Press Agency (DPA), to which the portal providing the utilized database belongs, has 54 editorial desks distributed across Germany. We suspect that districts in which an editorial desk exists might be over-represented in press releases because they may attract more (local) issuers to utilize the portal's services. In addition, the locations of editorial desks might not be chosen randomly, but rather be proximate to important press release sources such as political agencies or newspapers. In either case, the existence of editorial desks can cause a potential bias that needs to be controlled for. We include the dummy variable *EDESK*, which is one for districts in which the DPA has an editorial desk or zero otherwise. Lastly, we include the number of press release issuing organizations (*ISSUERS*) to correct for some regions having larger numbers due to more organizations being active issuers. Descriptive statistics of these variables are presented in Tables 1 and 2.¹⁰ The bivariate correlations are reported in Table B.1 in Appendix B.

Our dependent variable, the number of press releases (in general, or concerning a

⁹Ideally we would like to have a more direct measure of political engagement, like the number of people being members of political parties. However, such data is currently missing.

¹⁰The total population count refers to the year 2020; the gross domestic per capita and population density refer to the year 2019; voter turnout is recorded for the federal elections in 2017; and the information on touristic overnight stays is the most recent one available, which is 2015.

Table 1: Descriptive statistics

	PRESS	ISSUERS	POP	POPDEN	GDPC	VOTE	TOUR	PAT
#Obs	401	401	401	401	401	401	398	399
#Null	1	1	0	0	0	0	0	0
#NA	0	0	0	0	0	0	3	2
Min	0	0	3,4193	36	0.02	63.10	0.40	0.58
Max	26,881	2,696	3,669,491	4,767.30	0.18	84.10	43.60	824.38
Mean	340.10	71.86	207,398.28	546.39	0.04	75.08	5.48	56.87
Median	49.00	27.00	154,899	200.60	0.03	75.30	3.40	31.28
Std.dev	1632.02	203.06	245,162.41	723.15	0.02	3.80	6.28	82.69

Table 2: Descriptive statistics for categorical variables

Variable	Levels	# Obs	%
EAST	1	77	19.2
	0	324	80.8
	All	401	100.0
EDESC	1	54	13.5
	0	347	86.5
	All	401	100.0

particular keyword), is a count variable. This suggests the use of generalized linear models. Since the variable is characterized by an over-dispersion, we employ a negative binomial distribution.¹¹ We do not detect any issues of spatial autocorrelation in our models.¹² In the regression analyses, we include all variables representing shares or rates in their original form. Variables measuring absolute numbers are log-transformed to reduce the effect of potential outliers.

We run separate models for different categories of press releases. These categories are selected among the most frequent keywords, based on the availability of secondary data at the district level for which a fit with the keyword can be expected. The subset of press releases related to economic events is given by the dependent variable *ECON*. It denotes the number of press releases assigned to keyword *Economy*. The second category considered is politics (dependent variable: *POL*), for which press releases are considered that are associated with the keyword *Politics*. In both cases, we expect the population variables and income level to show a significant relation. The third subset is the press releases assigned to the keyword *Tourism*, (dependent variable *TOURM*), which is hypothesized to correlate with the touristic attractiveness of regions. The same is true for the fourth subset that comprises press releases associated with the keyword *Leisure*, (dependent variable *LEISR*). We expect districts with a larger population, higher population density, and larger tourist facilities to host a larger number of events and thus be mentioned with leisure-related press releases. Lastly, in the fifth subset, we consider press releases related to technological events. The dependent variable *TECH* denotes the number of press releases associated with the keyword *Technology*. We expect a larger number of technology-related press releases in regions where gross domestic product per capita and patent applications are higher, as these are commonly associated with the frequency of science and technology events (Balland et al. 2020).¹³

¹¹A χ^2 test based on the models' residual deviance indicates the Poisson distribution does not fit. This is further substantiated with a likelihood ratio test for over-dispersion.

¹²Lagrange multiplier test statistic for the error model is 1.820 with a p-value of 0.177 and the test statistic for the spatial autocorrelation model is 0.407 with a p-value of 0.523.

¹³The correlations between the count of press releases in the given categories and regional socio-economic variables are given in Table C.1 in Appendix C.

Table 3: Determinants of visibility in press releases

	PRESS (1)	ECON (2)	POL (3)	TOURM (4)	LEISR (5)	TECH (6)	TECH (7)	TECH (8)
EDESK	0.851*** (0.109)	0.617*** (0.126)	1.571*** (0.182)	0.002 (0.239)	0.164 (0.156)	-0.443* (0.213)	-0.456* (0.213)	-0.456* (0.217)
log(ISSUERS)	1.044*** (0.045)	0.950*** (0.054)	1.394*** (0.078)	1.472*** (0.109)	1.120*** (0.072)	0.952*** (0.108)	0.977*** (0.108)	1.029*** (0.106)
log(POP)	-0.138 (0.090)	0.173 (0.107)	-0.439** (0.152)	-0.497* (0.205)	-0.251 (0.137)	0.335 (0.205)	0.193 (0.191)	0.041 (0.175)
log(POPDEN)	0.132** (0.042)	0.099* (0.050)	-0.003 (0.072)	0.133 (0.098)	0.136* (0.065)	0.119 (0.100)	0.051 (0.094)	0.128 (0.088)
log(GDPC)	0.304* (0.124)	0.295* (0.145)	-0.011 (0.209)	0.321 (0.279)	0.237 (0.185)	0.605* (0.257)	0.559* (0.255)	
VOTE	-0.031** (0.012)	-0.010 (0.014)	-0.109*** (0.020)	-0.036 (0.027)	-0.022 (0.018)	0.056* (0.026)		
TOUR	-0.002 (0.005)	-0.016* (0.006)	-0.016 (0.009)	0.081*** (0.011)	0.025** (0.008)	-0.006 (0.014)	-0.005 (0.014)	-0.0001 (0.013)
log(PAT)	0.044 (0.057)	-0.048 (0.068)	0.114 (0.096)	0.086 (0.132)	0.101 (0.089)	0.119 (0.137)	0.287** (0.110)	0.371*** (0.104)
EAST	-0.129 (0.098)	-0.343** (0.117)	-0.067 (0.164)	-0.186 (0.224)	-0.019 (0.149)	0.589** (0.216)	0.549* (0.215)	0.465* (0.218)
Constant	4.720*** (1.366)	-0.930 (1.620)	10.934*** (2.315)	4.501 (3.139)	1.881 (2.098)	-10.934*** (3.203)	-5.422** (1.847)	-6.381*** (1.842)
Observations	396	396	396	396	396	396	396	396
R ² McFadden	0.208	0.225	0.199	0.172	0.206	0.266	0.262	0.259
R ² (Linear)	0.307	0.312	0.206	0.413	0.446	0.302	0.300	0.298
Max VIF	5.38	5.45	5.46	5.54	5.59	6.01	4.23	3.42
Log Lik.	-1,994.679	-1,536.445	-1,531.040	-1,006.969	-1,059.938	-522.546	-524.914	-527.220
θ	3.122***	2.463***	1.148***	0.733***	1.852***	1.773***	1.753***	1.647***
AIC	4,009.357	3,092.890	3,082.080	2,033.939	2,139.877	1,065.093	1,067.828	1,070.439

Note: Numbers in parentheses are standard errors of coefficients.

*p<0.05; **p<0.01; ***p<0.001

5 Results and Discussion

The regression results are reported in Table 3. Model (1) of Table 3 shows our baseline model that relates the total number of press releases to all regional characteristics. Two of our control variables are strong predictors of the number of press releases. Regions where editorial desks of the press agency (DPA) are located (EDESK) are more likely to appear in press releases. Apparently, they either attract additional press release activities or are in regions where there are more events to report. Since we control for the number of press release issuing organizations (ISSUERS), which also obtains a significantly positive coefficient, EDESK does not capture the presence of many press release issuing organizations. This suggests that editorial desks rather motivate proximate organizations to be more active in this respect. In any case, the findings suggest that the location of editorial desks and the number of press release issuing organizations in the regions as well as the underlying processes need to be taken into consideration when working with press release data in a spatial context.

In addition, we find regions with higher population density, i.e., urban regions, to be mentioned more in press releases. Accordingly, it is urbanization that translates into more events taking place and eventually getting featured in press releases. Crucially, this effect materializes despite the consideration of the number of press release issuers (ISSUERS) and the total population (POP) in the models. It implies that it is not a size, but a (social) density effect that becomes visible in the number of press releases. GDPC also obtains a significantly positive coefficient. It confirms our predictions that more prosperous regions are more likely to host a larger number of events. From this general analysis, it is unclear to what extent the availability of greater economic resources to initiate and organize events drives this finding. We must leave this to future research.

We observe a statistically significant and negative relationship between press release frequency and voter turnout (VOTE), which does not meet our expectations, as we expected voter turnout to be positively correlated to the number of political events taking place in a region which should translate into more press releases. As the subsequent analysis will show, this argumentation seems to be far from reality.

In addition to the variables for which we find a statistically significant relationship, it

is interesting to note that there do not seem to be significant differences between former East and West German regions. Accordingly, variations in the frequency of press releases at the sub-national level do not seem to be strongly impacted by the former division of Germany, at least not when controlling for the other variables.

In sum, our baseline model highlights that there is a strong regional imbalance in the frequency of press releases, which is systematically related to regional characteristics, most notably to the number of press release issuing organizations, presence of the editorial desks, and the degree of urbanization. Consequently, these variables should be included in analyses at this level when using press release information.

Columns (2) to (8) in Table 3 present the findings for the estimations using the frequency of press releases associated with specific keywords as the dependent variable. Other than that, their specifications are identical to that of the baseline model. Note that in these models, the number of press release issuing organizations serves two purposes. First, it controls for inter-regional variation in precisely this number (see discussion above). Second, it accounts for the general tendency of regions to be mentioned in press releases because the variable is highly correlated to the total number of press releases. When including both PRESS and ISSUERS, multicollinearity distorts the estimations ($VIF > 10$). Therefore, we only consider one of the two variables in the models reported here.

The first category of press releases whose spatial distribution is explored is related to the economy (ECON). Besides the control variables discussed above, the model in column (2) shows that economy-related press releases are mostly concentrated in regions with a higher per capita income. While the level of statistical significance is relatively low, it can still be seen as confirmation of our expectations that more economy-related press releases are issued in regions that are economically more successful. The slightly significantly positive relationship with urbanization (POPDEN) adds to this, as more urbanized regions also tend to be more prosperous or at least host a larger number of diverse economic activities. A similar rationale may explain the somewhat stronger significantly negative coefficient of EAST. Given that regions in the Eastern part of the country still lag their Western counterparts in terms of economic prosperity, they appear to be less likely to host economic events that find their expression in press releases. When accounting for inter-regional differences in economic prosperity and urbanization, tourism frequently tends to be over-represented in peripheral and less economically dynamic regions. Consequently, we find a negative relationship between the corresponding variable TOUR and press releases dealing with economic topics. In sum, for this category, our models confirm the idea that press releases' content reflects important economic characteristics of regions. However, while the directions of the coefficients are in full support, a caveat is the relatively low levels of significance of the observed relations.

Our results for politics-related press releases (Model 3) are probably the most surprising. While the control variables are again in line with the expectations and particularly the presence of editorial desks correlates with the spatial distribution, our core explanatory variable for this set of press releases (VOTE) shows a significantly negative relationship with their frequency. This is opposite to our expectations, as high voter turnout (VOTE) was seen as an indication of democratic engagement and consequently, of the frequency of political events taking place. However, the literature on voter turnout suggests that other determinants, such as labor market conditions, domestic political cleavages, the local age composition, and regional attachment are more important determinants in this context (Fiorino et al. 2019, Henderson, McEwen 2010). Consequently, our primary indicator might not seem to correlate to the number of political events that took place, which explains our findings for this variable. This is supported by the size of regions (POP) being observed to have a significantly negative relationship with the number of political press releases as well. Less populous regions are usually rather rural, which in Germany tends to correlate with an on average older population. Older populations are politically more interested and might hence induce comparatively more political events to take place in a locality. These are subsequently mentioned in press releases. Clearly, future research will have to work with alternative indicators to assess this relationship more precisely.

In contrast to the findings for political press releases, our findings concerning tourism

(Model 4) are in line with our expectations. Regions that host a larger number of tourists per inhabitant are associated with a larger number of tourism-related press releases. Clearly, a stronger touristic orientation implies more touristic events taking place which shapes the press releases associated with that region. Apart from tourism, it is the size of the region that is found to have a slightly significant negative relation with TOURISM. Given that tourism matters a lot for smaller and rural regions, this finding strengthens the close link between press releases and touristic activities in the regions.

Model (5) relating the frequency of leisure-related press releases to regional characteristics is basically a mirror of Model (4) concerning tourism. Apparently, the use of the keyword *leisure* is very similar to that of *tourism*, which suggests that they refer to the same overarching category of press releases, which captures touristic and leisure activities.

The last category of press releases of which regional distribution is studied is technology. Here, if we use the same model as for the other categories, we find few and primarily rather weak relations. Technology-related press releases are positively related to GDP per capita, which is likely to be driven by more technologically active and advanced organizations being in regions with higher levels of economic income. However, the link is not very strong. We also observe a positive relationship with VOTE suggesting that regions with higher voter turnout are more likely to be found in press releases about technologies, for which we do not have an explanation. Crucially, our primary indicator to assess the correspondence between regional characteristics and press releases, the number of patents (PAT), is insignificant. However, the results change when excluding GDPC and VOTE variables. In Models 7 and 8, we remove these variables from the analysis. Now, the number of patents (PAT) obtains a strongly significantly positive coefficient, which is very much in line with our expectations. Consequently, it is the consideration of VOTE and GDPC that “hides” the correspondence of press release information with the intensity of innovation activities in regions.

In addition, the analysis of technology-related press releases highlights the link between press releases and the content of newspapers. The coefficient of EAST is significantly positive, signaling that press releases mentioning regions in East Germany tend to feature information on technologies more frequently than those in the Western part. While in general, innovation activities are still less frequent in the former East than in the Western part, the finding matches that of [Ozgun, Broekel \(2021\)](#), who observe the same positive relationship for the content of (regional) newspapers.

Considering our analysis, we find the assessment of press releases as a source of regional data rather promising. For four out of five studied topic categories, we find a strong correspondence between common indicators of regional characteristics and the content of press releases. In particular, for touristic and leisure activities the link appears to be substantial. The spatial distribution of economic activities also seems to be reflected in the content of press releases. The correspondence of press release content and the intensity of innovation activities (as measured by patents) only becomes observable when not controlling for regions’ economic situations and voter turnout. However, it must be pointed out that patents are just one indicator for specific kinds of innovations (those that can be patented) and that NUTS-3 regions (used in our analysis) are generally perceived to be not ideal for capturing their spatial distribution ([Brenner, Greif 2006](#)). Within the framework of the applied approach and data, we did not manage to validate press releases to follow the same spatial distribution as common statistical indicators in the case of political events. It is likely that our employed indicator is insufficient and hence the finding shouldn’t be over-interpreted.

6 Conclusion

As in most social sciences, regional scientists and economic geographers are constantly seeking to advance their available databases. Given the fast-changing and complex nature of contemporary developments, standard official datasets may not include the information required to track and analyze these. In the present study, we discuss press releases as an alternative data source in this context. Despite their use in other literature streams, so far, they have been rarely exploited by spatial scientists. Press releases offer a number

of attractive features, such as being generally widely available, containing rich textual information that includes aspects frequently not covered by official statistics, and, most importantly, being geo-locatable.

Given these promises of press releases, the aim of this paper was to assess the usefulness of press releases as a data source in spatial economic research from an empirical perspective. We argue that press releases seem to be an attractive data source which, however, requires several crucial aspects to be considered. This includes a certain positivity bias, as organizations are less likely to issue press releases on negative events. A more important aspect from a spatial perspective is that press release information is greatly shaped by the distribution and characteristics of the underlying population of press release issuing organizations. For instance, locations with many organizations active in the sports sector are more likely to show up in press releases on sports. Consequently, press releases will primarily reflect the population of issuing organizations. Yet, this is not always a negative thing, as researchers might be explicitly interested in this population which may otherwise be difficult to be observed. Sticking to the example of sports, for countries in which a comprehensive list of sport-organizations at small geographical units does not exist, press releases with reference to sports activities might give a reasonable approximation of their spatial distribution. Although they cannot be used to infer their absolute number or general importance, identifying regions with comparative strength in these activities is feasible.

To be used in such tasks or similar ones, it is important to know if press releases, when aggregated to the regional level, reflect the socio-economic characteristics of regions. To gain insights into this, we empirically assessed the correspondence of a range of regional characteristics and associated events in press releases. To achieve this, we used a novel data set on press releases published by the German news agency DPA covering all 401 German districts (NUTS-3 regions). The analyses revealed that the overall frequency of press releases and that of specific types of events mentioned therein align reasonably well. In particular, aggregated counts of press releases on touristic activities are strongly correlated to the presence of tourists in regions. A similarly strong link was observed for technology press releases and patents when not controlling for regions' economic prosperity and voter turnout. Although weaker, we also detect a relationship between regional economic development and economy-related press releases. Our analysis was unable to confirm a correspondence between press releases on political topics and political engagement (voter turnout) at the regional level, which we suspect is primarily due to voter turnout not being a good proxy for the number of political events taking place in a region.

In sum, in many instances, regional differences in events and activities seem to be well-presented by press releases. This makes press release data a useful data source in spatial studies. However, there are important limitations to press release data that apply to all press releases in general and some that apply to the specific database on press releases used in this study. The latter's relevance in other step-ups depends on how much the data is like the one used in this study. As for the first group of limitations, as pointed out, press releases suffer from a positivity bias and they are shaped by the underlying population of press release issuing organizations. Concerning the latter, the motivations and abilities to produce press releases, as well as the styles thereof, are likely to differ substantially between organizations. So far, it is unknown to what extent these may differ between locations. In addition, press releases are always highly subjective texts, written with a specific purpose, which, unfortunately, is not information to spatial scientists. These may introduce unknown distortions. There is also no information on which events are covered by press releases and which are not. Consequently, as with any other data source, press releases represent incomplete information which must be considered in their interpretation.

As for the data we have used in our study, all these limitations apply. In addition, however, it includes very little information on the issuing organizations. We do not know what fraction of all press releases in a region are covered in the database. For about half of the press releases, no location information was given in the text, which makes assigning them to the correct region in which the event took place very difficult. Even when location

information was included, we faced several challenges. For instance, there are still several ambiguous location names. A prominent example is “Essen”. It frequently appears in press releases and may refer to, in this case, the city of Essen or to “food”. While manual inspections confirmed a high discriminating power of our geo-location procedure and we are confident about its proper working, there might still be cases in which the location assignment was incorrect. For disambiguation of location names, future studies can use natural language processing tools. In terms of classifying the press releases by topics, we relied on keywords assigned to press releases. Neither do we know who assigned those keywords (either the issuing organization, or the database administrators), nor do we know how reliable they are. Future research may shed light on this by using text classification or clustering techniques. We tried to overcome many of these limitations and in this regard, our empirical study hopefully provides valuable insights for the applied researchers working with press release data and facing similar issues. We also point out a couple of aspects that may help the specification of empirical spatial models based on press release data. For instance, when using their spatial variation, it is essential to consider the location of editorial desks and the number of issuing organizations.

Given the increasing demands for new and rich data sets in regional science, we believe that press releases offer substantial potential and may be very useful in many contexts. For instance, when looking at specific topics, press releases may allow for identifying regional economic specializations that remain hidden in official statistics. Our study is an early step in this direction and might help researchers assess if press releases are useful for their work.

References

- Ahern KR, Sosyura D (2014) Who writes the news? Corporate press releases during merger negotiations. *The Journal of Finance* 69: 241–291. [CrossRef](#)
- Aronoff CE (1976) Predictors of success in placing releases in newspapers. *Public Relations Review* 2: 43–57. [CrossRef](#)
- Arribas-Bel D (2014) Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography* 49: 45–53. [CrossRef](#)
- Balland PA, Jara-Figueroa C, Petralia SG, Steijn MP, Rigby DL, Hidalgo CA (2020) Complex economic activities concentrate in large cities. *Nature Human Behaviour* 4: 248–254. [CrossRef](#)
- Bean C (2016) Independent review of uk economic statistics – Professor Sir Charles Bean. *Independent Review of UK Economic Statistics*
- Berger J, Humphreys A, Ludwig S, Moe WW, Netzer O, Schweidel DA (2019) Uniting the tribes: Using text for marketing insight. *Journal of Marketing*: 1–25. [CrossRef](#)
- Boukes M, Vliegthart R (2020) A general pattern in the construction of economic newsworthiness? Analyzing news factors in popular, quality, regional, and financial newspapers. *Journalism* 21: 279–300. [CrossRef](#)
- Boussalis C, Coan TG, Holman MR (2018) Climate change communication from cities in the USA. *Climatic change* 149: 173–187. [CrossRef](#)
- Brenner T, Greif S (2006) The dependence of innovativeness on the local firm population – An empirical study of German patents. *Industry and Innovation* 13: 21–39. [CrossRef](#)
- Chang TK, Shoemaker PJ, Brendlinger N (1987) Determinants of international news coverage in the US media. *Communication Research* 14: 396–414. [CrossRef](#)
- Dahlin P, Fors J, Öberg C (2006) Press releases, annual reports and newspaper articles – Using alternative data sources for studies on business network dynamics. 22nd annual IMP conference, Milan

- Elders C (2006) News factors and news decisions. Theoretical and methodological advances in Germany. *Communications* 31: 5–24. [CrossRef](#)
- Erjavec K (2005) Hybrid public relations news discourse. *European Journal of Communication* 20: 155–179. [CrossRef](#)
- Feldman M, Lowe N (2015) Triangulating regional economies: Realizing the promise of digital data. *Research Policy* 44: 1785–1793. [CrossRef](#)
- Fiorino N, Pontarollo N, Ricciuti R (2019) Supranational, national and local dimensions of voter turnout in European parliament elections. *JCMS: Journal of Common Market Studies* 57: 877–893. [CrossRef](#)
- Galtung J, Ruge MH (1965) The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research* 2: 64–90. [CrossRef](#)
- Gans HJ (2004) *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time*. Northwestern University Press
- Gertler MS (1997) The invention of regional culture. In: Lee R, Wills J (eds), *Geographies of Economies*. Edward Arnold, London, 47–58. [CrossRef](#)
- Haller M (2012) Lokaljournalismus in den neuen Bundesländern. <http://www.bpb.de/gesellschaft/medien-und-sport/lokaljournalismus/151237/lokaljournalismus-in-den-neuen-bundeslaendern?p=all>
- Hamilton J (2004) *All the news that's fit to sell: How the market transforms information into news*. Princeton University Press, Princeton. [CrossRef](#)
- Harcup T, O'Neill D (2017) What is news? News values revisited (again). *Journalism Studies* 18: 1470–1488. [CrossRef](#)
- Henderson A, McEwen N (2010) A comparative analysis of voter turnout in regional elections. *Electoral Studies* 29: 405–416. [CrossRef](#)
- Henry E (2008) Are investors influenced by how earnings press releases are written? *The Journal of Business Communication* (1973) 45: 363–407. [CrossRef](#)
- Herman ES, Chomsky N (2010) *Manufacturing consent: The political economy of the mass media*. Random House
- Himelboim I, Chang TK, McCreery S (2010) International network of foreign news coverage: Old global hierarchies in a new online world. *Journalism & Mass Communication Quarterly* 87: 297–314. [CrossRef](#)
- Hofstede G, Garibaldi de Hilal AV, Malvezzi S, Tanure B, Vinken H (2010) Comparing regional cultures within a country: Lessons from Brazil. *Journal of Cross-Cultural Psychology* 41: 336–352. [CrossRef](#)
- Hong SY (2008) The relationship between newsworthiness and publication of news releases in the media. *Public Relations Review* 34: 297–299. [CrossRef](#)
- Jacobs G (1999) *Preformulating the news: An analysis of the metapragmatics of press releases*, Volume 60. John Benjamins Publishing. [CrossRef](#)
- Johnson K, Haythornthwaite J (1989) Press releases: A neglected source of information. *Aslib proceedings* 3. [CrossRef](#)
- Kariel HG, Rosenvall LA (1984) Factors influencing international news flow. *Journalism Quarterly* 61: 509–666. [CrossRef](#)
- Kepplinger HM, Ehmig SC (2006) Predicting news decisions. An empirical test of the two-component theory of news selection. *Communications* 31: 25–43. [CrossRef](#)

- Kinne J, Axenbeck J (2018) Web mining of firm websites: A framework for web scraping and a pilot study for Germany. *ZEW-Centre for European Economic Research Discussion Paper* (18-033)
- Kinne J, Lenz D (2021) Predicting innovative firms using web mining and deep learning. *Plos one* 16: e0249071. [CrossRef](#)
- Kleinsteuber HJ, Thomass B (2007) The German media landscape. In: Terzis G (ed), *European Media Governance*. Intellect, Bristol, 111–123
- Klüver H, Sagarzazu I (2016) Setting the agenda or responding to voters? Political parties, voters and issue attention. *West European Politics* 39: 380–398. [CrossRef](#)
- Lansley G, Longley PA (2016) The geography of Twitter topics in London. *Computers, Environment and Urban Systems* 58: 85–96. [CrossRef](#)
- Lewis J, Williams A, Franklin B (2008) A compromised fourth estate? UK news journalism, public relations and news sources. *Journalism studies* 9: 1–20. [CrossRef](#)
- Maat HP, de Jong C (2013) How newspaper journalists reframe product press release information. *Journalism* 14: 348–371. [CrossRef](#)
- Macnamara J (2014) Journalism-PR relations revisited: The good news, the bad news, and insights into tomorrow's news. *Public Relations Review* 40: 739–750. [CrossRef](#)
- Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM (2013) The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one* 8: e64417. [CrossRef](#)
- O'Neill D, O'Connor C (2008) The passive journalist: How sources dominate local news. *Journalism Practice* 2: 487–500. [CrossRef](#)
- 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. [CrossRef](#)
- Park H, Reber BH (2010) Using public relations to promote health: A framing analysis of public relations strategies among health associations. *Journal of Health Communication* 15: 39–54. [CrossRef](#)
- Pérez AA (2006) La Estadística Pública como apoyo a la investigación. *Investigaciones Regionales - Journal of Regional Research* 2006: 173–205
- Presseportal (2021) Was ist Presseportal.de? <https://www.presseportal.de/about>
- Reich Z (2010) Measuring the impact of PR on published news in increasingly fragmented news environments: A multifaceted approach. *Journalism Studies* 11: 799–816. [CrossRef](#)
- Robinson GJ, Sparkes VM (1976) International news in the Canadian and American press: A comparative news flow study. *Gazette* 22: 203–218. [CrossRef](#)
- Saxenian A (1996) *Regional advantage – Cluture and Competition in Silicon Valley and Route 128*. Cambridge, Mass. [CrossRef](#)
- Segev E (2015) Visible and invisible countries: News flow theory revised. *Journalism* 16: 412–428. [CrossRef](#)
- Senninger R, Wagner M (2015) Political parties and the EU in national election campaigns: Who talks about Europe, and how? *JCMS: Journal of Common Market Studies* 53: 1336–1351. [CrossRef](#)
- Shoemaker PJ (2006) News and newsworthiness: A commentary. *Communications* 31: 105–111. [CrossRef](#)

-
- Staab JF (1990) The role of news factors in news selection: A theoretical reconsideration. *European Journal of Communication* 5: 423–443. [CrossRef](#)
- Stuetzer M, Audretsch DB, Obschonka M, Gosling SD, Rentfrow PJ, Potter J (2018) Entrepreneurship culture, knowledge spillovers and the growth of regions. *Regional Studies* 52: 608–618. [CrossRef](#)
- Turk JV (1986) Public relations' influence on the news. *Newspaper Research Journal* 7: 15–27. [CrossRef](#)
- von Bloh J, Broekel T, Ozgun B, Sternberg R (2020) New(s) data for entrepreneurship research? An innovative approach to use Big Data on media coverage. *Small Business Economics* 55: 673–694. [CrossRef](#)
- Walters LM, Walters TN (1992) Environment of confidence: Daily newspaper use of press releases. *Public Relations Review* 18: 31–46. [CrossRef](#)
- Walters TN, Walters LM, Starr DP (1994) After the highwayman: Syntax and successful placement of press releases in newspapers. *Public Relations Review* 20: 345–356. [CrossRef](#)
- Wu HD (2003) Homogeneity around the world? Comparing the systemic determinants of international news flow between developed and developing countries. *Gazette (Leiden, Netherlands)* 65: 9–24
- Wu HD (2007) A brave new world for international news? Exploring the determinants of the coverage of foreign news on US websites. *International Communication Gazette* 69: 539–551. [CrossRef](#)

Appendices

A Geo-locating the press releases

To geo-locate the press releases, we rely on the OpenGeo-database, which contains the names and geo-coordinates of more than 12,000 places in Germany (<http://opengeodb-giswiki.org/wiki/OpenGeoDB>). Since German location names can include punctuation and stop words, and some location names are words with other meanings, we cannot apply standard text preprocessing and straightforward string-matching techniques. The procedure we apply is summarized in the following.

1. We remove numbers and all punctuation marks from the text, except for periods, commas, parentheses, and dashes, as these are part of some location names. The punctuation marks that are not removed are standardized, e.g., *hyphen*, *en dash*, and *em dash* all become short dash.
2. We transform the abbreviations in location names to their full words so that they can be matched with the original names. For example, *a.M.* is transformed into *am Main*; *a.d.R.* into *an der Ruhr*; and *St.* into *Sankt*.
3. Although press releases are in German, in some cases, English names of German places are used. We translate these into their German originals. For example, *Munich* is translated into *München* and *Nuremberg* into *Nürnberg*.
4. In German, the first letter of location names are normally in capitals, however, sometimes this is not done when articles and prepositions are part of the name. To differentiate location names from other words, such as verbs, we capitalize all the stop words and prepositions that are used in location names. For example, *an der* becomes *An Der*; *vor der* becomes *Vor Der*. Consequently, all words in location names, if mentioned in the press release, begin with a capital letter. After this step, we clean the corpus by removing words that start with lower case letters. This ensures the removal of verbs, and irrelevant prepositions and articles. This step also decreases the size of text data considerably so that pattern matching becomes feasible.
5. Using string-matching we assign latitude and longitude from the OpenGeo-database to press releases. In case multiple location names are identified in the text, we assign all locations mentioned to the respective press release.
6. Since some location names in Germany have special letters that do not exist in all alphabets (ä, ö, ü, and ß) and writing the words including these letters by using only the English alphabet is a common practice, we repeat the previous step once more, by replacing the special letters in the list of location names. For example, *ä* is replaced with *ae*; *ß* with *ss* and so on. At the end of these steps, 51,861 press releases are assigned to a location in Germany.
7. In some cases, the places' full names are not mentioned in the press releases. Although this creates ambiguity in terms of text analysis, it is usually clear for readers from the content of the release which location the text refers to. A well-known example of such is *Frankfurt*. There are two prominent locations in Germany containing *Frankfurt* in their names: (*Frankfurt am Main* and *Frankfurt an der Oder*). Hence, a reference to just *Frankfurt* in the text is insufficient to establish a unique match. Lacking any further information to solve this ambiguity, we match in this case the location with the larger population. We assume that it is more likely that the smaller (and lesser known location) is referred with its full name, i.e. *Frankfurt an der Oder*. Manual checks confirm the appropriateness of this assumption. This allows to assign locations to 3,915 additional press releases.
8. In addition to potential location names in the releases' main text, the database offers location information for all press releases for their issuing organization. If a location is mentioned in the article (and it is identified through in the previous

steps), we assume that it is a more accurate description of the event's or issue's location than the location of the issuer. However, in the lack of the first, it is still the best available information. Consequently, for all locations where the previous steps didn't result in at least one match, the location of the issuer is used to geo-locate the press releases. For these locations, we primarily use an API access to Google Maps to obtain their geo-locations. At the end of this final stage, 99,984 press releases are assigned to a location in Germany.

B Correlation matrix

Table B.1: Correlation matrix

	PRESS	POP	POPDEN	GDPG	EDESK	VOTE	TOUR	EAST	PAT
POP	0.85***								
POPDEN	0.54***	0.50***							
GDPG	0.21***	0.14**	0.48***						
EDESK	0.39***	0.35***	0.50***	0.30***					
VOTE	0.01	0.09	-0.15*	0.01	-0.06				
TOUR	0.02	-0.03	-0.15*	-0.05	-0.02	0.03			
EAST	0.06	0.01	-0.13*	-0.26***	0.07	-0.45***	0.02		
PAT	0.59***	0.77***	0.49***	0.33***	0.27***	0.32***	-0.07	-0.19**	
ISSUERS	0.92***	0.89***	0.62***	0.25***	0.46***	0.02	0.03	0.04	0.69***

C Correlation between press release frequency in selected topics and regional variables

Table C.1: Correlation between frequency of press releases in economy, politics, tourism, leisure, and technology topics, and regional socio-economic variables

	POP	POPDEN	GDPG	VOTE	TOUR	PAT	EAST
ECON	0.86***	0.56***	0.22***	0.01	0.03	0.63***	0.04
POL	0.72***	0.45***	0.15*	-0.01	-0.01	0.45***	0.08
TOURM	0.82***	0.61***	0.29***	0.01	0.01	0.67***	0.00
LEISR	0.87***	0.61***	0.28***	0.02	0.02	0.67***	0.02
TECH	0.85***	0.55***	0.22***	0.03	0.03	0.71***	0.05

