FEELING THE HEARTBEAT OF REGIONS: LOCAL NEWS AND ECONOMIC SENTIMENTS

Tom Broekel

Papers in Evolutionary Economic Geography

25.19

Utrecht University Human Geography and Planning

Feeling the Heartbeat of Regions: Local News and Economic Sentiments

Tom Broekel*
School of Business and Law
University of Stavanger
Kjell Arholms gate 35, Stavanger, Norway
tom.brokel@uis.no

Accepted for publication in:
Journal of Economic Geography

June 30, 2025

Abstract

Timely and spatially detailed indicators of regional economic activity are limited. This paper introduces the Regional Economic News Sentiments (REGENS) index, a high-frequency measure based on geocoded news headlines from 300+ German-language outlets since 2019. REGENS captures local economic sentiment, aligns with national indicators, and significantly leads regional unemployment by up to four months. While its link to GDP growth is weaker, it consistently reflects regional economic patterns. The study highlights how media signals contribute to understanding economic development and illustrates the potential of text-based indicators to sharpen the spatial and temporal resolution of regional monitoring.

Keywords news, media, sentiments, regions, regional development

JEL: R11, C55, O33

*E-mail: tom.brokel@uis.no, Web: tombroekel.de

1 Introduction

Monitoring regional economic developments with high-frequency, real-time data is essential for practitioners and scholars. Traditional economic indicators often suffer from significant delays and do not fully capture the dynamic nature of regional economies, creating the need for more responsive and granular measures (Deller, 2010). This necessity aligns with calls in economic geography to better integrate temporal information to understand regional disparities and growth dynamics (Pike et al., 2006). Moreover, economic geography has long emphasized the importance of time in understanding economic processes (Boschma & Martin, 2007; Henning, 2019). However, while the discipline has focused extensively on long-term trajectories, short-term economic fluctuations, which may shape or disrupt regional development paths, have received much less attention in empirical research.

This study introduces the Regional Economic News Sentiments (*REGENS*) index—a novel, location-specific, real-time measure of economic sentiment. Building on advances in text analysis and natural language processing, including transformer-based models, the index is constructed from systematically geocoded news content. It uses the Regional News Syndication database (RegNeS), which aggregates daily news headlines and snippets from more than 300 German-language outlets since 2019 (Ozgun & Broekel, 2021). By extracting sentiment signals from economic coverage, *REGENS* aims to capture how regional economies are portrayed in the media, and, by extension, how they are perceived with high temporal and spatial resolution.

Although news data has long been used in macroeconomics and finance to measure sentiment, expectations, and informational shocks (Cutler et al., 1988; Friedman & Schwartz, 1963; Garcia, 2013; Tetlock, 2007), its use in regional economic analysis remains limited. Recent work has begun to bridge this gap: Berle and Broekel (2025) and Ozgun and Broekel (2021) demonstrate that news sentiment and content reflect regional characteristics, while Peris et al. (2021) highlights how news flows between places mirror broader spatial interdependencies. Garz (2018) and van Binsbergen et al. (2024) show that national-level news sentiment indices can predict macroeconomic indicators such as GDP and unemployment. However, most of these contributions operate at coarse spatial scales (e.g., national or state level) and focus on where news circulates rather than what is reported about specific regions. This study contributes to closing this gap by developing a subnational fine-grained indicator based on the portrayal of regional economies in the news.

In doing so, the paper makes several contributions. Empirically, it shows that the *REGENS* index aligns with established national sentiment benchmarks and systematically tracks regional unemployment dynamics, with statistically significant associations emerging at lags of up to four months. This temporal pattern indicates that the index captures forward-looking economic signals embedded in news reports. Although the relationship with annual GDP growth is more modest, the point estimates remain directionally consistent, suggesting that the sentiment of the regional news reflects elements of the underlying economic trajectory. In sum, these results highlight the potential of *REGENS* as an early signal of changing regional economic conditions, particularly in settings where conventional statistics are released with substantial delay or lack sufficient spatial resolution.

Conceptually, the study engages with broader debates on the role of media as both a mirror and a shaper of regional perceptions and economic expectations. It connects with recent perspectives that frame media as institutional actors and emphasizes the emergence of "local narrative landscapes" (Berle & Broekel, 2025). Methodologically, the paper highlights the potential of integrating news data and natural language processing tools into regional economic analysis. The approach introduced here is scalable and adaptable, providing a blueprint for developing similar indicators for other topics such as sustainability or innovation. Over time, such indices may offer more flexible and fine-grained monitoring of regional developments than is possible with conventional data sources (Arribas-Bel, 2014; Batty, 2013; Ye & He, 2016).

The structure of the paper is as follows: Section 2 reviews the literature on news-based sentiment indices and the spatial dimensions of news data. Section 3 introduces the REGNES database and outlines the empirical framework. Section 4 details the construction of the *REGENS* index. Section 5 evaluates its national counterpart against established indicators. Section 6 investigates the regional patterns of *REGENS* and assesses its explanatory power. Finally, Section 7 summarizes the findings and outlines future research directions.

2 News, Sentiments, and Space

2.1 News as Information Source Covering Time and Space

News has been recognized as a rich and timely source of information on societal and economic dynamics for a long time. For instance, research in political science uses it to identify, measure, and analyze political conflicts and developments over time and space (Arzheimer & Bernemann, 2023; Blood & Phillips, 1995; Lheureux, 2023). Similarly, in economics, there is a long-standing interest in using news coverage to capture, nowcast, and forecast economic developments (Barsky & Sims, 2012; Beaudry et al., 2011; Iselin & Siliverstovs, 2016; Jiang et al., 2023; Nowzohour & Stracca, 2020). This research builds on the idea that economic activities are reflected in news coverage through the reporting of economic events and stories (Boydstun et al., 2018).

In this context, the *framing* of news is particularly relevant, as it plays a key role in how audiences understand and interpret issues (Vicsek, 2011). Entman (1993) conceptualized framing as the process of highlighting certain aspects of reality in communication to shape how issues are defined, interpreted, evaluated, or addressed. Framing captures how a topic is presented, including the selection of information, tone, and evaluative stance.

Among the various dimensions of framing, *sentiments* reflect the evaluative tone of coverage, and capture whether a topic is reported in a positive, neutral, or negative light. Concerning economic news, their sentiments (*economic sentiments*) describe how economic topics and events are evaluated at a given moment in time, which makes them a frequently used vehicle for tracking and monitoring the perceived state of an economy (Shapiro et al., 2022). Thereby, positive economic sentiments are generally associated with reports of favorable developments, such as increasing profits, revenues, business opportunities, or growth-supporting social, technological, and institutional changes. Conversely, negative sentiments often align with adverse evaluations of these aspects. Yet, while sentiments provide a descriptive lens on the economic situation, this description is not objective but subject to many biases. With significant heterogeneity between outlets, media tends towards negative presentation (DellaVigna & Kaplan, 2007) and selectivity, including the underrepresentation of minorities (Gentzkow & Shapiro, 2006; Khoo et al., 2012). Consequently, the mirror has some imperfections that must be considered when working with such data. This includes the fact that news in general and economic news sentiments, in particular, are not just reflections of events and processes but also shaping forces of these. For instance, (economic) news sentiments have been shown to influence perceptions, expectations, and subsequent actions of economic actors such as businesses, consumers, and policymakers (Shapiro et al., 2022) as well as to impact economic output, employment, and business activities (Acharya et al., 2017; Benhabib et al., 2016).

Nonetheless, the value of economic news sentiments as an information source remains evident, as illustrated by their central role in widely used indices such as the Daily News Sentiment Index and the Economic Policy Uncertainty Index (Baker et al., 2016; Buckman et al., 2020). These two indices also showcase the tendency of using economic news sentiment information at the country level to capture macroeconomic trends and developments (Baker et al., 2016; Tetlock, 2007). At this level, they have proven effective in predicting business cycles, stock market returns, prices, trading volumes, and even house prices (Engelsberg & Parsons, 2011; Greenwood & Shleifer, 2014; Li et al., 2014; Schumaker et al., 2012; Soo, 2018; Uhl, 2011a, 2011b). By capturing the temporal fluctuations in tone and content, news-based indicators often outperform traditional data sources, which are delayed or lack granularity.

Despite the wealth of studies leveraging news sentiments to understand economic dynamics over time, the spatial dimension remains underexplored. Most research confines spatial analyses to international comparisons, overlooking the rich potential of subnational variations. Yet, news coverage is inherently shaped by more granular geographic targeting, as local, regional, and even parts of national outlets tailor their focus to specific subnational audiences. This intrinsic link between news coverage and geographic specificity highlights the potential of economic news sentiments to capture and analyze regional economic dynamics (Ozgun & Broekel, 2021). By providing geographically targeted insights at the subnational level, news sentiment analysis addresses many of the limitations of traditional economic indicators. The first benefit is its high level of *granularity*. Given sufficient observations, news information can be linked to specific locations and aggregated to almost any desirable level. Its widespread availability further facilitates the design of tailor-made indicators for monitoring regional economic developments both within and across regions. Additionally, news sentiment analysis provides a *wide breadth* of coverage. By processing sentiment data from diverse sources, researchers can extend their analyses beyond economic topics, incorporating political, social, and technological dimensions (see, e.g., Berle & Broekel, 2025; Ozgun & Broekel, 2021, 2024). This capability supports the call for broader and more inclusive perspectives when exploring regional developments (Coenen et al., 2012).

Another critical advantage of economic news sentiments is their *timeliness*. Unlike traditional indicators, such as the unemployment rate, which is typically released monthly and lags actual economic developments by up to six months (Decressin & Fatás, 1995), or GDP figures, which are published annually with a time lag of two to three years (see, e.g., Directorate-General for Economic and Financial Affairs, European Commission, 2024), economic news sentiments tend to capture shifts in economic conditions at the national level almost instantly (Iselin & Siliverstovs, 2016; Shapiro et al., 2022). Such can provide researchers and local stakeholders, including policymakers, with faster and more actionable information. In addition, news sentiments offer unparalleled flexibility in temporal resolution, enabling observations at weekly, monthly, or even daily intervals (Ozgun & Broekel, 2024; van Binsbergen et al., 2024). Thereby, regional economic developments become traceable more continuously and with observations at time intervals aligning with the underlying economic processes.

2.2 The Spatial Dimension of News

Media outlets frequently cater to specific market segments and audiences, which are often geographically defined (Gentzkow et al., 2011; Soroka, 2012; Strömberg, 2004). Consequently, the media adapts its content to align with the preferences and concerns of these geographically distinct audiences (Gentzkow et al., 2011; Hamilton, 2011). For instance, national newspapers typically focus on issues relevant to their home country, whereas regional and local newspapers disproportionately emphasize topics and events of local significance (Hamilton, 2004). This geographic targeting not only determines the selection of stories but also influences how they are framed, ensuring that the narratives resonate with the priorities and interests of their respective audiences (Entman, 1993).

International research highlights the strong link between (national) audience characteristics and news reporting. Factors such as population size, economic performance, proximity to news agency headquarters, and distance to events help explain why some locations receive more prominent news coverage than others (Howe, 2009). Similar dynamics are observed within countries. For example, Ozgun and Broekel (2021, 2024) show that the content of regional newspapers in Germany reflects local characteristics, while Avraham and Ketter (2012) find significant differences in how central and peripheral cities are portrayed in national news.

These subnational, interregional variations in news reporting include framing in general and sentiments in particular. Local newspapers often use fewer negative terms when reporting on nearby companies, reflecting a more favorable tone compared to national outlets (Gurun & Butler, 2012). Dominick (1977) and Brooker-Gross (1983) also show that sentiments in TV reporting differ across states in the USA. Moreover, Ozgun and Broekel (2021) and Ozgun and Broekel (2024) document significant variations in the sentiments with which innovation and COVID-19 are covered across Germany's subnational regions. van Binsbergen et al. (2024) extends these findings to economic news sentiments.

While subnational variations in (economic) news sentiments stand at the forefront of these studies, the methods used to link news information to specific locations are not merely empirical decisions but carry theoretical implications. Specific methods capture different perceptions and interpretations of economic realities by different entities. For instance, Ozgun and Broekel (2021, 2024) use newspapers' distribution areas to associate news with specific locations. Similarly, van Binsbergen et al. (2024) assign geographical relevance to news using the federal state where a newspaper is published.[2] However, associating news data with readership is not necessarily the most suitable approach when constructing economic sentiment indices aiming at capturing the economic status and trends of specific locations. Theoretically, linking news to locations involves multiple dimensions of meaning: it can be based on the audience

[2]This assessment is based on the limited details provided by the authors.

(reflecting the news that the audience is exposed to), the events covered (capturing the content of the news), or the place of authorship (representing the perspective of the media institution).

The first approach links news data to the audience's location, focusing on where news is consumed (Gentzkow et al., 2019). This approach reflects the potential influence of news on its audience through the information and sentiments conveyed. Hence, it is most suitable for studying the effects of news on regional perceptions and behaviors.

The second approach involves assigning locations based on the geographic location of the events that the news describes. In this case, each news item is tagged with the location(s) of the event(s) it reports on. News geolocated in this manner provides insights into the events occurring in specific places. Thus, it is suitable when the focus is on describing what is happening in locations.

The third approach considers the location of the individuals who produce the news (writing, filming, editing). This approach focuses on the production side of news media. By adopting this approach, the perspectives on events that may occur in different places are captured. This approach is especially useful when investigating potential institutional or organizational motives behind selecting covered events or their framing (Hamilton, 2004).

Consequently, the second approach is most suitable when using news information, such as economic news sentiments, to capture and monitor the economic development of locations. The other approaches risk producing biased results, as the location of events covered in the news often differs from the location of news consumers, even for outlets focused on local news. For example, van Binsbergen et al. (2024) geolocate news based on the audience's location, producing a sentiment index that reflects the status and development of the economy across all locations mentioned in the news, rather than just a specific focal area. This approach highlights inter-regional variations in the news and sentiments to which readers are exposed, which may influence audience behavior—a finding also supported by van Binsbergen et al. (2024) in their historical analysis.

However, constructing a regional economic sentiment index in this way is less likely to offer insights into variations in the current status of regional economies, particularly when significant portions of the news cover events outside the audience's locations. Here, the second approach is more suitable, which uses the location of covered events to geolocate news. An index constructed in this way provides insights into the collective (when multiple news outlets cover the same event) information on and perception of the reported event. It captures the relevance and sentiments associated with all reported events and developments in a location, thus describing its (newsworthy) economic situation. This approach also has a practical advantage as a side effect. News consumption data is often difficult to obtain and rarely available at fine spatial resolutions. For example, van Binsbergen et al. (2024) rely on USA state-level information. In contrast, in-text location references provide precise place names, enabling geolocating events and sentiments at very granular spatial levels. Consequently, this approach will be used in this paper.

The three distinct approaches highlight the multiple dimensions characterizing the relationship between news and space, which must be carefully considered when analyzing news from a spatial perspective. News is produced by someone (third approach) to inform others (first approach) about events occurring somewhere (second approach). These dimensions collectively influence which events and locations are represented in the news and how they are framed. For instance, local news outlets tend to overemphasize local events compared to national outlets, making the content highly dependent on the type and scope of the media outlet. Additionally, the agenda-setting power of news outlets (McCombs & Shaw, 1972; Wanta et al., 2004) influences the selection of stories and their framing, reflecting the institutional, organizational, and geographic contexts of news production.

Mitigating such distortions requires leveraging news from a diverse and extensive range of outlets, especially when analyzing coverage of particular topics or regions. The greater the diversity, the more likely these various spatial, institutional, and organizational influences are accounted for.

3 Regional Economic News Data

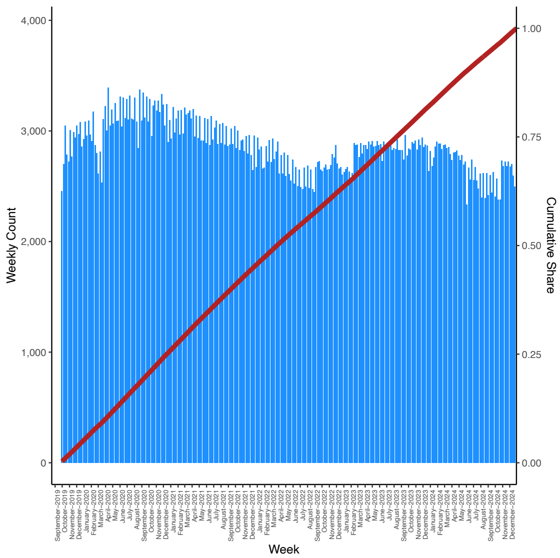
3.1 Regional News Data

Capturing geographical variations in economic news sentiment requires comprehensive data from various outlets. Large-scale initiatives, such as the Global Database of Events, Language, and Tone (GDELT), have demonstrated the feasibility of assembling multi-source, geographically diverse news datasets on an international scale (Leetaru & Schrodt, 2013). However, databases specifically designed to support representative subnational news analysis have historically been unavailable or limited to specific thematic focuses (Garz, 2018). Recent advancements, however, have addressed this gap, as evidenced by the extensive datasets employed by Ozgun and Broekel (2021, 2024) and van Binsbergen et al. (2024) to analyze subnational (regional) sentiment variations.

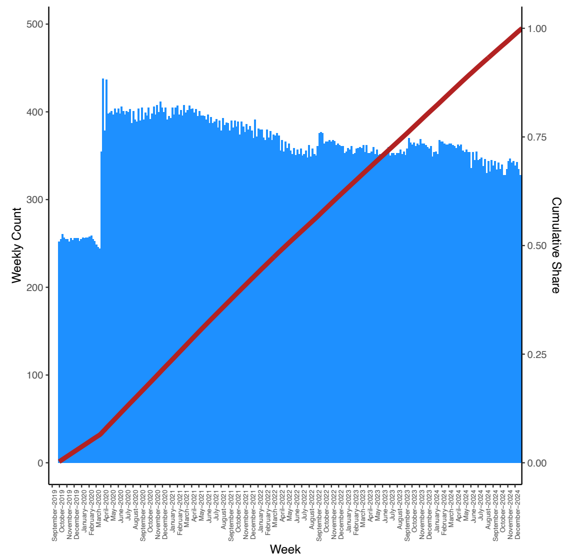
This study builds on the works of Ozgun and Broekel (2021, 2024) by utilizing the Regional News Syndication *RegNeS* database. The database is a daily collection of German-language news, which extracts information from RSS feeds of professional news outlets from all over the German-speaking world (Germany, Austria, Liechtenstein, Luxembourg, Switzerland) multiple times a day. Private blogs or special-interest news channels (public administration, sports clubs, etc.) are not covered. The data exclusively features outlets publishing daily or weekly. The basis of the news extraction is RSS-feed addresses that have been manually collected and that are constantly updated. Since its start in 2019, it has experienced changes in the number of working RSS-Feeds and covered news outlets (see Figures 1a and 1b), it has grown in coverage for most of its time (Figure 1c).[3] At the end of 2024, the database included over twenty-five million news items obtained from over four hundred outlets. This makes it one of the largest available databases of German news.

There are more RSS feeds than news outlets because most outlets operate multiple feeds, which focus on different geographical areas and topics. The present paper considers only news from German media outlets and the period between November 2019 and the end of December 2024. While most covered news outlets are newspapers, from

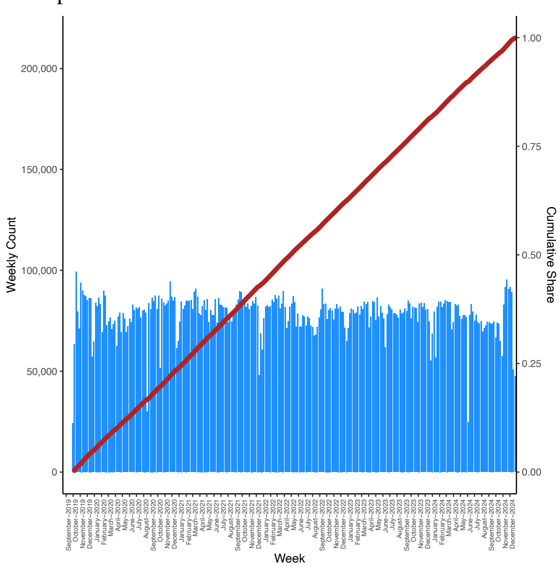
[3] The temporal variance in coverage is primarily due to fluctuations in the capacities of the involved scholars to clean and maintain the feed database. The noticeable dips in news numbers in some weeks were caused by technical issues in the news extraction procedure.



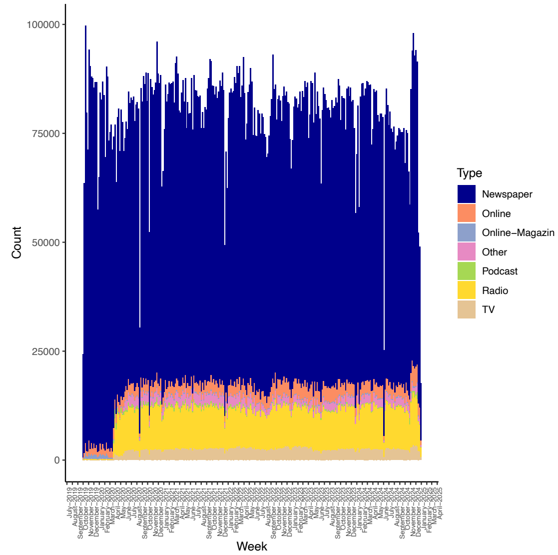
(a) Weekly number and cumulative share of incoming RSS feeds processed over time.



(b) Weekly number and cumulative share of distinct media outlets contributing articles.



(c) Weekly number and cumulative share of news articles processed.



(d) Weekly count of news articles by media outlet type (e.g., newspaper, online, radio).

Figure 1: Descriptives of the RegNes news data processing pipeline by week.

2020 onward, the database includes a noticeable number of TV- and radio stations, online sources, and some online magazines, see Figure 1d. The present paper considers all of these media types. Potential differences between these types must be accounted for in the empirical approach.

To reduce noise in the data, news from outlets with fewer than 100 news items. The filtering leaves $21,531,561$ news items obtained from $4,957$ RSS feeds operated by 495 media outlets.

Theoretically, a media outlet can contribute a maximum of $1,917$ days to the data considered in this paper. The empirical mean is $1,543$ days, implying that the average outlet is present in the database for about three-quarters of the considered time period. However, the distribution is strongly skewed towards outlets that are included the entirety of the time (see Figure 2).

The distribution of news items across media outlets is moderately skewed (see Figure 3a), with the Gini coefficient being close to 0.672. About 50% of news outlets produce less than 5% of the news items. 75% of the media outlets account for about 25% of the news.

The database does not feature full-texts; only the headlines and text snippets are extracted from the RSS feeds. This marks a significant difference from other news-based indicators that tend to utilize full-texts (Iselin & Siliverstovs, 2016). Naturally, it implies less available information, which is generally undesirable and represents a limitation of the underlying database. However, research highlights the significant role of news headlines in shaping reader perception and attention (Barros et al., 2021; Djurayev, 2020; Hagar et al., 2022; Robertson et al., 2023). Headlines also guide the reader's understanding of the text and usually serve as a summary (Andrew, 2007). While there can be a certain disconnect between the emphasis and tone of a headline and the actual content of the news story (Andrew, 2007), they are generally seen as adequate representations of the full text (Barros et al., 2021; Djurayev, 2020; Kochetkova et al., 2018). Consequently, they are helpful in capturing core information and sentiments of news. In addition to headlines, *RegNeS* also contains news snippets, which exceed the average length of the combined (headline plus snippet) news text to about 33 words[4], whereby the distribution is bimodal (Figure 3b).[5] This distribution highlights the difference between news items without snippets (first peak) and most items featuring snippets (second peak). Snippets provide a more extended summary of the news texts, increasing the empirical data's representativeness of the unobserved full texts.

Three core tasks must be done to construct a regional economic news sentiment index. Firstly, sentiments have to be quantified. Secondly, geographic locations must be identified in the texts. Thirdly, a classification has to be conducted to identify news items related to economic topics. These steps are presented in the following subsections.

3.2 Quantification of Sentiments

Many methods exist for quantifying the sentiments of texts, each with specific strengths and weaknesses. Traditionally, dictionary-based approaches dominated the field (Stone et al., 1966), characterized by their reliance on predefined lists of words classified as positive, negative, or neutral. However, these methods have notable limitations, particularly in handling context, irony, and negations (Liu, 2012). Following the trend in natural language processing to move from rule-based to learning-based approaches (Cambria & White, 2014), more advanced methods have become available that provide sentiments with greater precision (Poria et al., 2014). In particular, deep learning has significantly advanced sentiment analysis by allowing for the automatic learning of semantic features from data, thereby capturing the context around words and phrases to interpret sentiment more accurately (Zhang et al., 2018).

[4]This number is obtained after an extensive cleaning including the removal of special characters, HTML tags, links, etc.

[5]A very small number of news texts have more than 200 words. While they are considered in all calculations, they are excluded from the visualization in Figure 3b.

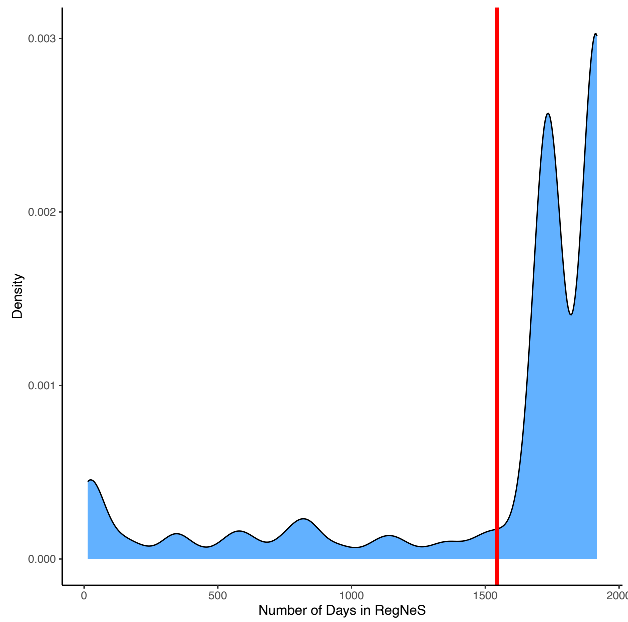
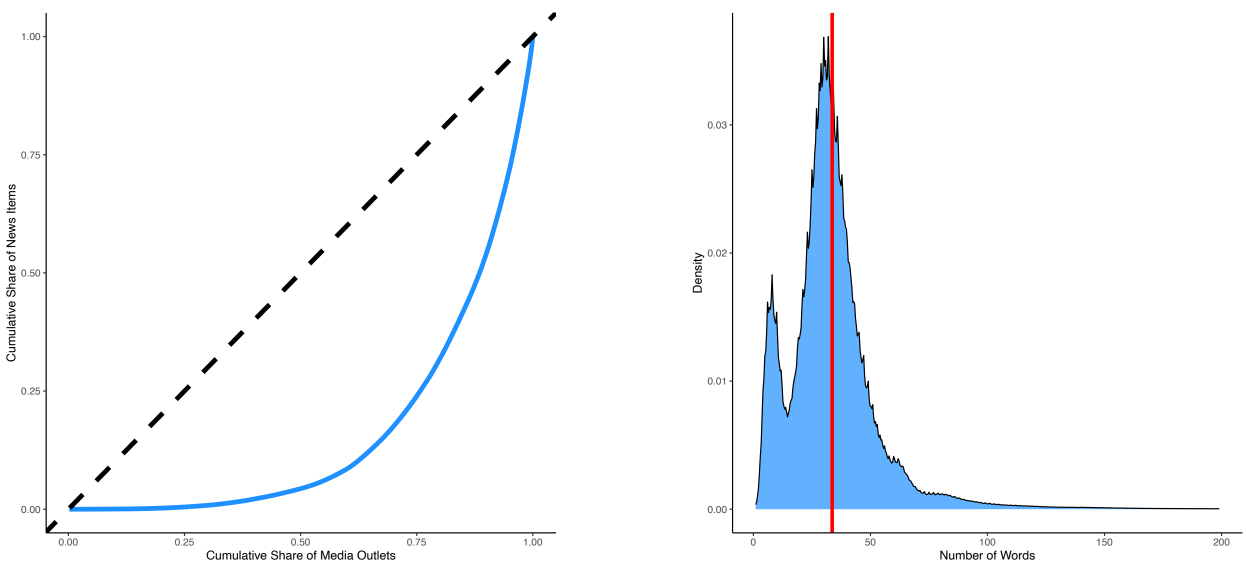


Figure 2: Kernel density of the number of days each media outlet is present in the RegNeS database. The red line marks the average.



(a) Lorenz curve showing the concentration of news volume across media outlets.

(b) Distribution of news item length (word count) with a red line marking the average.

Figure 3: Descriptive statistics on the concentration and length of news articles in the RegNeS database.

Following this literature, I employ a multilingual-uncased-sentiment model to assess the sentiments of news texts. More precisely, I use the *nlptown/bert-base-multilingual-uncased-sentiment* model, which is based on the Bidirectional Encoder Representations from Transformers (BERT) architecture (Peirsman, 2020). Such models find increasing applications in quantifying news sentiments (Rozado et al., 2022). The model was trained using thousands of product reviews across multiple languages, including 150k English, 137k German, and 80k Dutch reviews. The model shows notable accuracy in its predictions, with an exact match accuracy ranging from 57% to 67% across different languages and a higher off-by-1 accuracy (predicting the sentiment within one star of the human reviewer's rating) from 93% to 95%, making it a frequent choice in the field of sentiment analysis (Lheureux, 2023).

The BERT-base-multilingual-uncased-sentiment model assigns sentiment scores on a scale from 1 to 5. This scale can be interpreted as follows: scores of 1 and 2 are categorized as negative, three as neutral, and 4 and 5 as positive sentiments (Sasmaz & Tek, 2021). As a result, each document is assigned a vector reflecting the document's probability of belonging to the corresponding sentiment class. Based on this vector and the numerical values of (1 to 5) assigned to each class, the probability-weighted average of the numerical class values is calculated for each news item (see Equation 1). It is denoted as S_i.

$$S_i = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 + 5 \cdot p_5 \tag{1}$$

The procedure successfully assigned sentiment values to nearly all of the 21,531,561 news items, except $8,111$ items, which were excluded from the dataset. Figure 4 visualizes the distribution of sentiment values, which are bimodal with elevations on both sides of the mean of 2.93.

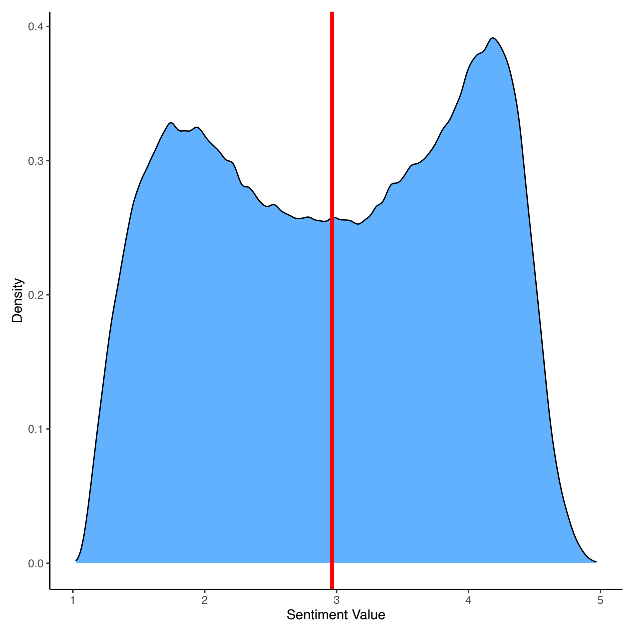


Figure 4: Kernel density estimate of raw sentiment values across all news items. The red line indicates the mean sentiment.

In the Appendix, Table 5 lists a sample of economic news that obtained a very high ranking of sentiments. In contrast, Table 6 presents a sample of news with very low sentiments. The difference is apparent, as the positive news covers primarily recognitions of business excellence. In contrast, the negative news items feature insolvencies, damages, and garbage problems.

While this model performs well in distinguishing between positive and negative sentiments in economic news, its training on product reviews introduces a noticeable bias toward language and tone typical of consumer feedback. This is evident in some of the highly positive examples in Table 5, where five-star style expressions align more with evaluative product language than with typical journalistic reporting. Although this model outperformed several alternatives tested in this study, future research should explore models trained on news-specific corpora to reduce such domain-related biases.

3.3 Geolocating News

In light of the discussion in Subsection 2.2, it is essential to identify locations mentioned in the news text. To do this, I rely on the *OpenGeoDB* (http://opengeodb.- giswiki.org/wiki/OpenGeoDB) database. It features more than $7,400$ unique German location names assigned to geo-coordinates. German location names frequently include punctuation and stop words, have multiple meanings, and the same names are used for different places. Standard named-entity recognition tools and simple string-matching approaches struggle with this. Therefore, I employ a multi-step process. Initially, locations outside Germany were excluded from the OpenGeoDB, and all remaining names were standardized. The standardization involved transforming abbreviations and capitalizing certain words. [6] In the second step, relevant strings in the news texts underwent similar transformations for consistency.

The third and core step of the process involved string matching against the OpenGeoDB. The matching was enhanced by a context-specific and iteratively compiled list of $1,280$ location *qualifiers* to resolve ambiguities.[7] More precisely, to uniquely identify a specific location in a text, it must contain the search string (usually a location's name as included in OpenGeoDB) and additional qualifiers. For instance, distinguishing between the two locations of *Freiberg* - the city district in Stuttgart and the one in Saxony - requires additional tokens like *Neckar*, *Stuttgart*, *Sachsen*, or *Universität* to appear in the text. The list of qualifiers was developed manually and sequentially. In some cases, a default location is assigned unless qualifiers indicate otherwise. In the previous case of *Freiberg*, the town in Saxony has $41,800$ inhabitants, with Stuttgart's district only accounting for $7,100$. Clearly, the first can be considered more "news-prone" and thus is chosen as the default. It means that news texts containing the token *Freiberg* are always assigned to the location in Saxony unless a qualifier token indicates a reference to Stuttgart's district. Identification of locations of similar news prominence necessitates at least one qualifier token.

The quality of the matching was iteratively evaluated and improved by examining the association between the number of news items and the population of each location. Specifically, all statistically significant (at the level of 95%) deviations from the log-log regression line were checked manually for potential locational misidentification, and further qualifiers were developed when necessary. The final log-log linear regression obtains a coefficient for population on news of $population = 1.099^{***}$ and an R^2 of 0.351. While there is a substantial deviation from the trend line for smaller cities,

[6] Note that in German, the first letter of location names is always uppercase.
[7] My deepest gratitude belongs to my student assistant for her incredible efforts and help in this endeavor.

the association becomes more pronounced for larger regions, with the goodness of fit reaching almost 0.6 for locations with more than $200,000$ inhabitants (see Figure 5).

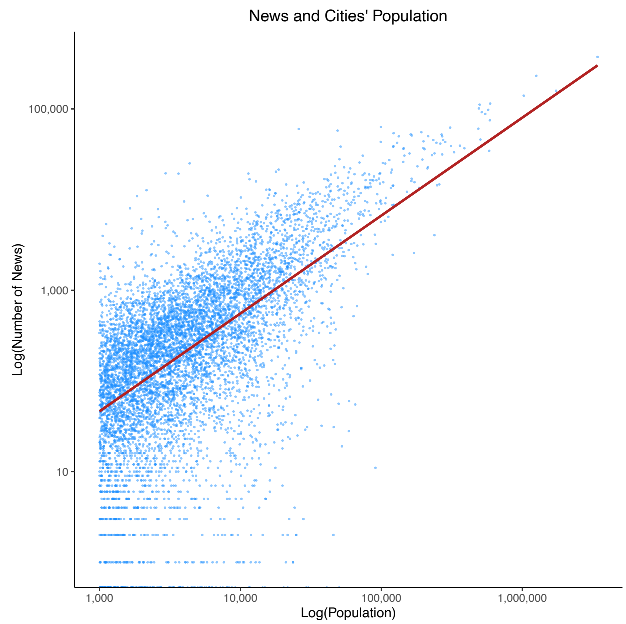


Figure 5: Log-log scatterplot showing the relationship between city population and the number of news articles tagged with that location. The red line indicates fitted regression.

In total, about 45.0% of all news items are associated with at least one German location. [8] Of all news items with at least one location, the average item features 1.24 locations. To construct a regional news indicator and allow for matching with secondary data, the location information is projected to the German districts (NUTS3) level, which will be used as *regions* in the following.[9]

3.4 Classifying Economic News

The next step in building a regional economic news sentiment index is identifying economic news. Various approaches have been employed in the past. For instance, van Binsbergen et al. (2024) use an automated procedure to generate a dictionary of key terms and phrases related to economic contexts, while Möhrle et al. (2022) identify companies in news texts to classify news.

In contrast, this study benefits from a pre-labeled corpus of news texts. Specifically, the RegNeS database contains a substantial number of news items already labeled by the originating news outlets. These labels are diverse, including location names, individuals' names, companies, and generic topic categories. However, some labels are particularly useful for distinguishing economic from non-economic news. In the first step, a binary classification was created by grouping all economy-related labels under the label "economy". The economy-related labels include "wirtschaft"

[8]The total share of news containing locational information will increase when locations outside Germany are included.

[9]Notably, while news could be matched to the districts of *Baden-Baden* and *Weiden in der Oberpfalz*, their multi-meaning names resulted in too many false-positive identifications. To ensure the reliability of the analysis, these districts are therefore excluded.

(economy), “ökonomie” (economy), “streik” (strike), “markt” (market), and “industrie” (industry). Noticeably, “wirtschaft” dominates these labels, accounting for over 50% of the news items within this category. All remaining labels were grouped under *other*, which includes categories such as “politik” (policy), “sport” (sport), “sonstiges” (miscellaneous), “leute” (people), “polizei” (police), and “kultur” (culture).

This binary-labeled dataset consists of $4,237,026$ items and gives a representative subsample of the full data with contributions from over 290 distinct news outlets. 15.22% $(644,920)$ items are categorized as *economy*-related, while the remaining items are classified as *other*. This makes it ideal to train a topic-classification model, which was done in the second step of the process.

Various binary classification methods were considered. Among these were Long Short-Term Memory (LSTM) Network models with Cased and Uncased Inputs, SetFit for Few-Shot Text Classification, and approaches based on DistilBERT sentence transformer models.[10]

Several setups were tested to determine the optimal classification method, as summarized in Table 7 in the Appendix. A Contrastive Learning approach, leveraging a fine-tuned DistilBERT model with $200,000$ contrastive pairs, emerged as the best-performing method and was therefore chosen. The training was done using an 80% random sample of the pre-classified observations. The evaluation used the remaining 20% (out-of-sample) observations. The evaluation considered standard metrics such as precision, recall, F1 score, and AUC-ROC (Fawcett, 2006; Powers, 2011).

Contrastive Learning optimizes the embedding space by positioning embeddings of semantically similar news items closer together while pushing embeddings of dissimilar items further apart (Gao et al., 2021; Reimers & Gurevych, 2019). The embeddings generated by the fine-tuned sentence-transformer DistilBERT model (*distilbert-base-german-cased*) were subsequently used for classification. A lasso-regularization regression model was fitted using a cross-validation approach on the training data to identify the optimal lambda and classification threshold. Lasso regression is particularly effective for identifying sparse and interpretable relationships between features while minimizing overfitting (Tibshirani, 1996).

The optimal prediction threshold (maximizing the F1 score) was identified as 0.43, resulting in an F1 score of 0.83 (see Table 7 in the Appendix). Table 1 presents the confusion matrix at this threshold. While approximately $43,000$ mis-classifications occurred, these represent slightly more than 5% of all test observations. Furthermore, the shares of false positives and false negatives are nearly identical, demonstrating that the model effectively handles the class imbalance in the dataset.

Table 1: Confusion matrix of the binary classification model used to distinguish economic from non-economic news articles. Rows represent model predictions and columns indicate true labels, with class 1 referring to economic news.

		Reference	
		0	1
Prediction	0	695,606	21,231
	1	22,816	107,753

[10]See for a brief description of these approaches Appendix 9.

To further check the goodness of the classification, a manual analysis of a random sample of false-negative and false-positive cases was done. These comparisons did not reveal any noticeable biases or patterns. Given the heterogeneity and breadth of the topic of "economic news", these results indicate a high level of performance. When applied to the full dataset using the optimal threshold of 0.43, the model classified 8.3% of all news items into the category *economy*.

Several news items are not unique to one news outlet. This can be because news outlets belong to the same mother outlet and share certain news sections, or the news item is acquired from the same third party. Lacking more information on this issue, I kept all of these instances as independent observations in the data set, resulting in a final data set of $9,385,631$ news items with sentiments, topic classification, and at least one identified location in Germany. Of these, $776,169$ observations are classified as being *economy*-related.

Figure 6 visualizes the complete data processing pipeline.

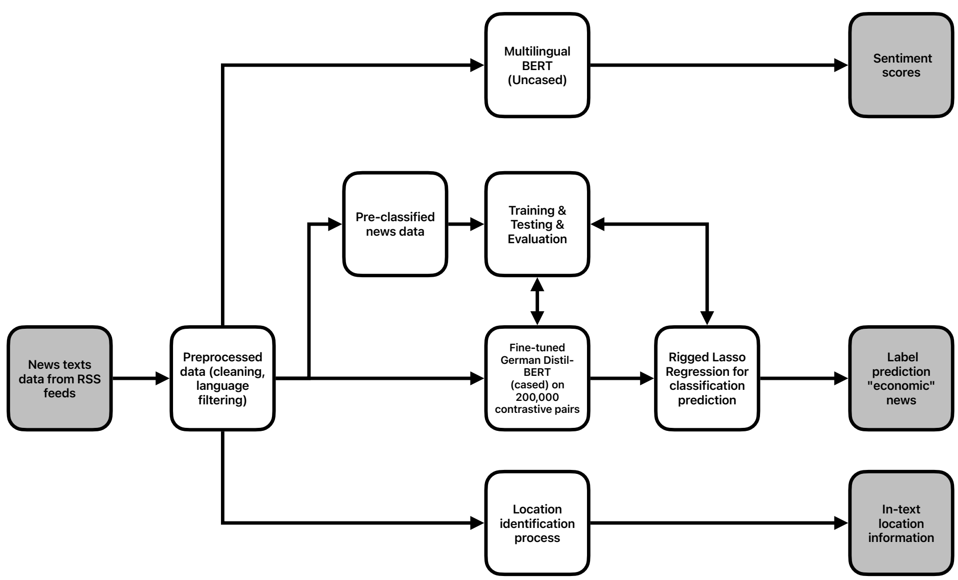


Figure 6: Overview of the data processing pipeline, including sentiment scoring, classification of economic news, and geographic assignment.

3.5 The Relevance of News Outlets, Regions, and Time for Economic News Sentiments

With the available data, I investigate the extent to which sentiment differences in news articles can be attributed to the news outlets themselves. I utilized a series of generalized linear models to isolate the impact of news outlets' reporting styles on sentiments from regional characteristics and temporal dynamics. Each model iteration includes a unique set of dummy variables for news outlets, regions, or time periods. By comparing the total variance in sentiments against the residual variance of each model, I evaluated the degree to which sentiments are differentiated across the dimensions represented by these dummies.

The dummy variable for news outlets accounts for 5.1% of the total sentiment variance of the $776,169$ economic news items, signifying that news outlets exhibit distinct sentiment expressions in their economic reporting. This proportion is almost twice as large as the regional dimension, constituting approximately 2.7% of the total variance. The temporal dimension is relatively minor, contributing about half a percent. Collectively, these three dimensions account for about 6.7% of the total variance. For comparison, the total variance contributions for all news types (including economic news) are 4.3% for news outlets, 2.5% for regions, and 0.8% for time periods, culminating in 5.9% overall. Hence, within the economic news subset, these dimensions substantially shape sentiment variance more than the broader news spectrum.

Although the combined and individual contributions seem modest, it is crucial to acknowledge the vast diversity within news concerning topics, events, and the particular writing styles of reporters, which are likely to predominate the overall variance. Remarkably, the regional dimension accounts for more than half of the variance in economic news sentiments compared to news outlets. Given the substantial academic focus on the latter (Rozado & Al-Gharbi, 2022; Rozado et al., 2022), the relative oversight of the subnational/regional dimension is surprising.

4 Building the Regional Economic News Sentiment Index

4.1 Constructing Regional and National News Sentiment Indices

To aggregate the news data, which includes sentiments, locations, and the classification as economic or non-economic topics, into a coherent regional economic sentiment index, I adopt the methodologies proposed by Shapiro et al. (2022) and Möhrle et al. (2022). These researchers employ regression analysis to mitigate unwanted biases in economic news sentiments, particularly the *composition effect* arising from the evolving contributions of news outlets over time. Additionally, newspapers exhibit systematic variations in their tone, framing, and sentiment (refer to Section 3.5). To address these variations, a regression is used that relates the sentiment values of news items to temporal factors and media outlet identifiers:

$$S_i = \theta_t \cdot \text{month}_i + \beta_j \cdot \text{outlet}_i + \epsilon_i. \tag{2}$$

Here, S_i denotes the sentiment values of news items i, with θ_t and β_j representing the coefficients corresponding to time t and news outlet j, respectively. The error component is ϵ_i. Given the extensive number of dummy variables (490 news outlets $\times$ 62 months), the analysis simplifies the estimation by treating news outlets as fixed effects.

While previous studies estimate this regression using the full sample of available data, I depart from this approach and adopt a 12-month moving window.[11] Specifically, for each focal month t, the regression is estimated using data from the current month and the preceding 11 months. From each window, only the coefficient corresponding to month t

[11] The 12-month moving window provides a good compromise: it spans a complete annual cycle, helping to control for time-fixed and outlet-specific seasonal reporting patterns, while still preserving within-year variation in sentiment.

is retained and interpreted as the sentiment index value for that month. This modification addresses a key limitation in the original framework: full-sample estimation relies on future observations, rendering real-time monthly updates impractical and conceptually inconsistent. By contrast, the moving-window setup ensures that sentiment in a given month is evaluated solely against information available up to that point, preserving the time-consistent logic of the index.

The regression is applied to the subset of economic news. The coefficient θ_t encapsulates the temporal aspect of news sentiments, reflecting the overarching time-related trends that influence news sentiments across all outlets for each period t, independent of the data's composition concerning news outlets. The regression is calculated without an intercept. Consequently, the θ_t coefficients can be interpreted as the mean sentiment (adjusted for news outlet influences) for each month-year relative to a zero baseline. In essence, the normalized θ_t coefficients constitute a national economic news sentiment index, hereafter referred to as the *National Economic News Sentiment Index (NENS)*.

Nonetheless, as identified by van Binsbergen et al. (2024), economic news sentiments are influenced by temporal factors that affect general news sentiments. To accommodate this, they devise a net index by orthogonalizing their economic sentiment index against a sentiment index encompassing all news items not classified as economy-related. Emulating this concept, I first construct a sentiment index for all non-economic news analogous to the construction of *NENS*. In addition, I incorporate the orthogonalization technique into the economic news sentiment index by amending Equation 2 as follows:

$$S_i = \theta_t \cdot \text{econ}_i \cdot \text{month}_t + \beta_j \cdot \text{outlet}_i + \epsilon_i. \tag{3}$$

This regression is applied to all news items, economic and non-economic. Here, econ_i is a binary variable assigned a value of one for entries identified as economy-related news and zero otherwise. The variable month_t represents a set of binary variables for each month (technically a month-year combination). The coefficient θ_t denotes the interaction effect between time and news type (economic vs. other), serving as an index of economic news sentiment for that particular month and providing a comparative measure against non-economic news. It quantifies how sentiments in economic news diverge from those in other news categories for each month. Note that in this specification, only the interaction terms are included, and the main effects of the variables are excluded from the model. The normalized coefficients of θ_t are referred to as the *Net National Economic News Sentiment Index (NET_NENS)* in the following.

Expanding upon Equation 2, the *Regional Economic News Sentiment Index (REGENS)* is formulated to capture both regional and temporal variations in economic news sentiment. In this specification, the time-specific dummies (month_t) are replaced with an interaction between month_t and region-specific dummies (region_r), which is further interacted with the economic classification indicator (econ_i). The resulting coefficient $\theta_{r,t}$ reflects the combined effect of regional and temporal variations in economic news sentiment across all media outlets for each time period t. Similar to the previous model, this regression is executed without an intercept and the main effects of the interaction variables.

$$S_i = \theta_{r,t,econ} \cdot \text{econ}_i \cdot \text{month}_t \cdot \text{region}_r + \beta_j \cdot \text{outlet}_i + \epsilon_i. \tag{4}$$

Theoretically, normalized $\theta_{r,t,econ}$ coefficients represent the net economic news sentiment index at the regional level, *NET_REGENS*. Unfortunately, including an interaction term with econ_i in this way is not feasible due to the high dimensionality involved in the regression calculation. Consequently, I adopt the approach by van Binsbergen et al. (2024) and split the data into economic and non-economic news and calculate for each a regression without the interaction with econ_i:

$$S_{i,econ} = \theta_{r,t,econ} \cdot \text{month}_t \cdot \text{region}_r + \beta_j \cdot \text{outlet}_i + \epsilon_{i,econ}. \tag{5}$$

$$S_{i,other} = \theta_{r,t,other} \cdot \text{month}_t \cdot \text{region}_r + \beta_j \cdot \text{outlet}_i + \epsilon_{i,other}. \tag{6}$$

Subsequently, I regress the $\theta_{r,t,econ}$ on $\theta_{r,t,other}$ and extract the residuals:

$$\theta_{r,t,econ} = \theta_{r,t,other} + \epsilon_{r,t}. \tag{7}$$

The residuals $\epsilon_{r,t}$ are normalized and represent $\text{NET\_REGENS}_{r,t}$, the net regional economic news sentiment index, which controls for inter-outlet variance and the region-month-specific sentiments of non-economic news.

The primary focus of this paper is the development and analysis of the **Regional Economic News Sentiment Index (REGENS)** and its net version (*NET_REGENS*). However, validating these regional indices presents a methodological challenge, as suitable benchmarks or high-frequency indicators are not available at the regional level. To address this, I introduce two national-level counterparts, **NENS** and **NET_NENS**, constructed using the same methodological framework. These serve a supportive role: they provide a reference point to evaluate whether the sentiment indices capture plausible and interpretable economic patterns. Although the national economy is not the analytical focus of the study, this initial validation step at the national level is critical for demonstrating the internal consistency and general behavior of the approach before its application at the regional scale.

Accordingly, I first present the national-level results to examine how *NENS* and *NET_NENS* co-move with conventional indicators such as the IFO, ZEW, and unemployment rates. This serves as a methodological prelude to the main contribution: the analysis of regional economic sentiment through *REGENS*.

5 National News Index and Validation

5.1 Description of the National News Index

The National Economic News Sentiments (*NENS*) index, its equivalent for non-economic news (*NONS*), and the net version *NET_NENS* are visualized in Figure 7. *NENS* and *NONS* share a cyclic, slightly upward-moving development that overlaps to a moderate degree (correlation $r = 0.65^{***}$). Accordingly, sentiment patterns of economic news share some variance with non-economic news sentiments, implying that both seem to be influenced by similar macroeconomic trends and cycles. For instance, both indicators were impacted by COVID-19 waves and are subject to typical intra-year cycles. The variance of *NET_NENS* reflects the divergence between the two when these macro trends and cycles are controlled for.

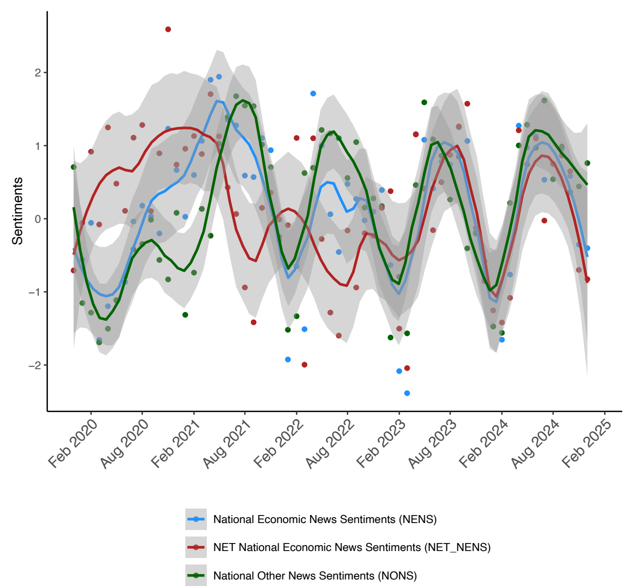


Figure 7: Time series of national-level economic (*NENS*), net economic (*NET_NENS*), and other news sentiment (*NONS*) indices. Smoothed trends with 95% confidence intervals.

NENS clearly shows patterns of news' seasonality. All months with the most negative economic news fall into the winter (March 2020, December 2021, January 2022, March 2023, February 2024). The most positive sentiments are either before or after the summer months, with 2020 being the exception (November 2020, May 2021, April 2022, September 2023, April 2024). Despite the relatively high correlation of *NENS* and *NONS*, the cyclic pattern of the sentiment index is still visible even when considering the net news sentiments (*NET_NENS*).

In contrast to the admittedly much longer time series of van Binsbergen et al. (2024), no evidence for an increasing negativity bias is observed. The averages of economic and non-economic sentiments (not controlling for news outlet effects) are relatively stable and close to the value of 3, which is larger than the technical average of 2.5 the employed BERT-model would deliver for neutral news, see Table 8 in the Appendix.

5.2 Validation Benchmarks

The *Ifo Business Climate Indices* stand out as the most prominent indicators of the German business climate. These indices are derived from approximately $9,000$ monthly surveys of companies in the manufacturing, construction, wholesaling, and retailing sectors. Respondents evaluate their present business conditions and their six-month outlook. The *Ifo Institute for Economic Research* compiles these assessments into indices reflecting the overall business climate (*IFO_CLIMAT*), current business situations (*IFO_SIT*), and future expectations (*IFO_EXP*). These indices represent normalized aggregates of the positive and negative responses and serve as composite indicators (Sauer, 2023). Known for their predictive accuracy regarding economic turning points, these indices have been extensively utilized in scholarly research (Demmelhuber et al., 2023; Englmaier & Fackler, 2022; Lehmann, 2023) and are available monthly from 1991 to the present.

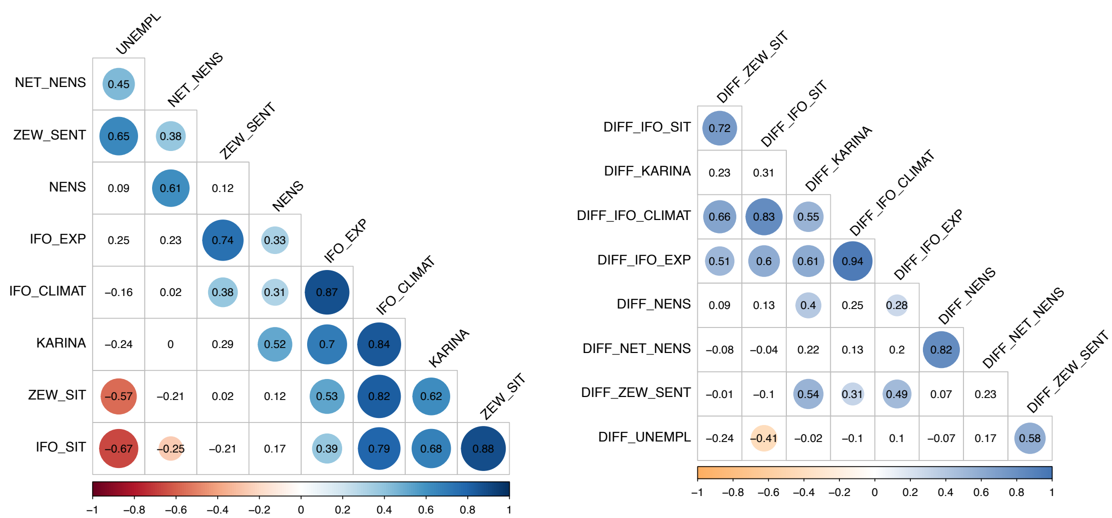
The *Leibniz Centre for European Economic Research (ZEW)* also offers sentiment indices constructed from the evaluations of financial experts in banking, insurance, and corporate finance. These experts assess the current economic climate and forecast six-month economic trends, including inflation, interest rates, stock markets, and exchange rates. The resulting indices, *ZEW Indicator of the Economic Situation* (*ZEW SIT*) and *ZEW Indicator of Economic Sentiments* (*ZEW SENT*), mirror the *Ifo Business Climate Index* in purpose and methodology. Their utility has been confirmed in various contexts (Hüfner & Schröder, 2002; Lehmann & Reif, 2021), with data available monthly since 1991.

KARINA, another benchmark index for *NENS*, is a collaborative effort between the *KfW* and the *Ifo Institute for Economic Research*. This news-based sentiment index leverages multiple news sources processed by an AI platform designed to pinpoint credit risks within news articles and press releases. The AI system identifies articles with mentioned companies and financial institutions, extracts critical risk keywords, evaluates the relevance to the associated entity, and, crucially, gauges the sentiment of the content. Sentiment analysis occurs at the sentence level, focusing on those proximate to company mentions. The resulting sentiment scores range from -10 (negative) to $+10$ (positive), with zero denoting neutrality. These scores offer a granular view of company news sentiment, which can be aggregated by industry or across the economy. The index incorporates $324,000$ news articles about $10,000$ companies from 550 sources, spanning January 2020 to December 2022. Monthly index values are calculated using the above-mentioned regression method to adjust for company-specific and source-related effects.

Lastly, while not a sentiment or composite business climate indicator in the narrow sense, the monthly unemployment rate is included as an essential benchmark for evaluating the *REGENS* index. It represents the only economic indicator available at both the required spatial and temporal resolution—monthly and at the NUTS-3 level—for Germany. In contrast, national-level sentiment indicators such as the *ZEW Indicator of Economic Sentiments* (*ZEW SENT*) or the forward-looking component of the Ifo Business Climate Index (*IFO_EXP*) cannot be disaggregated regionally and are therefore less suitable for comparison. Importantly, the goal of this paper is not to measure sentiment for its own sake, but to approximate the real-time economic situation across specific locations. In this context, unemployment is a pragmatic and policy-relevant reference point for assessing whether regional news sentiment patterns align with observable short-term economic developments. The variable used ($UNEMPL$) measures the proportion of unemployed individuals in the total labor force and is published monthly by the German Federal Statistical Office.

5.3 Validation of National Index

This section provides a descriptive validation of the national economic news sentiment index (*NENS*) by examining its co-movement with established economic indicators. The purpose is not to assess causality or forecast economic outcomes, but to evaluate the extent to which *NENS* tracks recognized measures of economic conditions over time.



(a) Correlations of indices in levels.

(b) Correlations of first-differenced indices.

Figure 8: Correlation of *NENS* and *NET_NENS* with Economic Indicators.

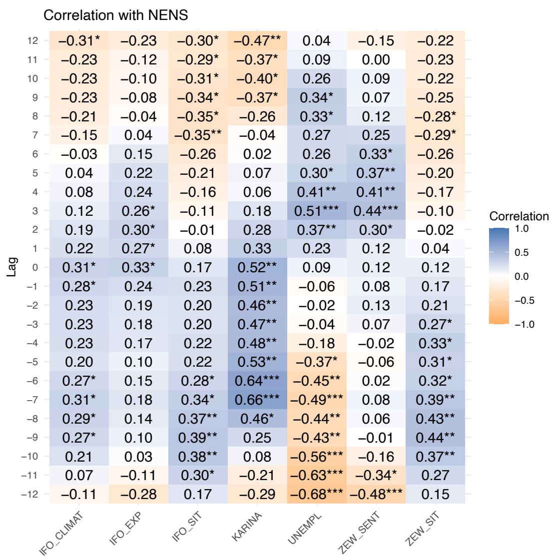
Note: Only statistically significant correlations at the 5% level are color-shaded. White cells indicate non-significance.

Given this exploratory aim, formal stationarity tests (e.g., the Augmented Dickey-Fuller test) are not reported. However, first-differencing is applied where appropriate to emphasize short-term dynamics. The data series are not deseasonalized to remain consistent with those from the *IFO* and *ZEW*.

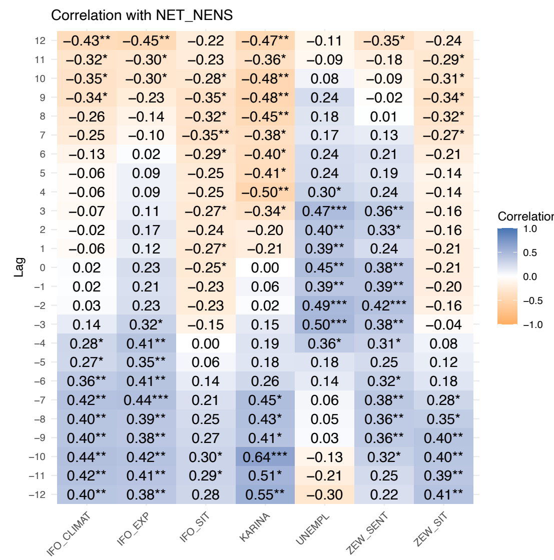
Figure 8a displays the correlations between *NENS* and selected economic indicators in levels, while Figure 8b shows correlations based on first differences to capture dynamic relationships. Overall, the results indicate limited alignment, though several notable exceptions emerge.

The strongest observed correlation is between *NENS* and the news-based sentiment index *KARINA* ($r = 0.52^{**}$), reflecting their conceptual similarity, as both are derived from news sentiment. Another notable correlation appears with the forward-looking *IFO_EXP*, where a positive relationship ($r = 0.33^{*}$) suggests that higher sentiment levels are associated with a more positive business outlook. In contrast, correlations with current-situation indices such as *IFO_SIT* and *ZEW_SIT* show considerably weaker associations.

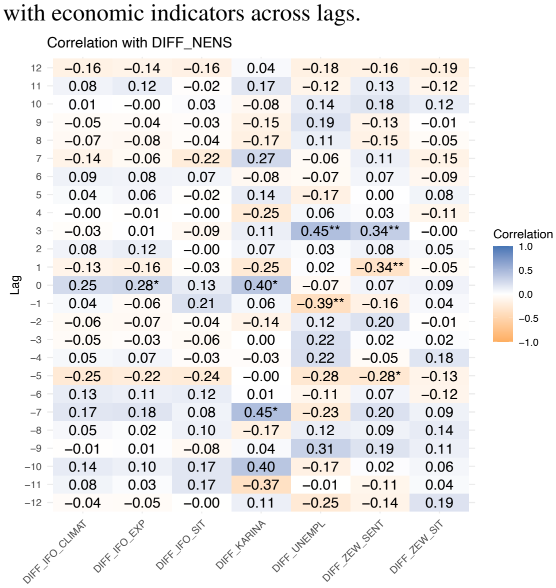
This pattern persists in the first-difference analysis (Figure 8b), where changes in *NENS* overlap significantly with *DIFF_KARINA* and *DIFF_IFO_EXP*. In contrast, *DIFF_NET_NENS* exhibits no strong relationships with any national indicators, suggesting that the netting-out of general sentiment isolates content less aligned with conventional benchmarks.



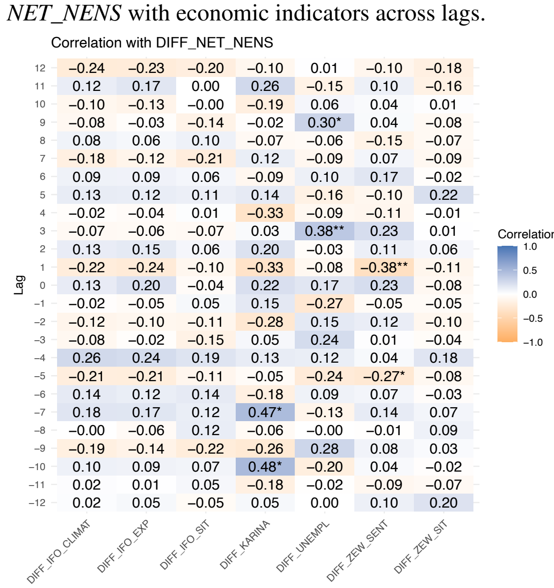
(a) Correlation of national economic news sentiments *NENS* with economic indicators across lags.



(b) Correlation of net national economic news sentiments *NET_NENS* with economic indicators across lags.



(c) Correlation of first-differenced *NENS* (*DIFF_NENS*) across lags.



(d) Correlation of first-differenced *NET_NENS* (*DIFF_NET_NENS*) across lags.

Figure 9: Lagged correlation analyses of national economic sentiment indices with selected macroeconomic indicators. *Note:*: Levels of statistical significance: ($p < 0.1^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$). A colored background highlights correlations significant at $p < 0.05$ or below.

The correlation results reveal meaningful lead-lag dynamics between sentiment indices and economic indicators when incorporating time lags. Figure 9 depicts correlations for lags ranging from -12 to $+12$ months. Negative lags indicate that the sentiment index precedes the economic indicator (i.e., leads), while positive lags imply that it follows. First-difference versions of the indices are included to capture short-term dynamics and remove shared low-frequency trends.

NENS exhibits its strongest correlations with indicators such as *IFO_CLIMAT*, *IFO_SIT*, and *ZEW_SIT* at negative lags between -3 and -6 months, suggesting that *NENS* may lead these indicators. The correlation with *IFO_EXP* peaks at a lag of 0, indicating a contemporaneous relationship. Notably, *NENS* also displays strongly negative correlations with *UNEMPL* and *ZEW_SIT* at longer negative lags (especially around -10 to -12 months), reinforcing the impression of a forward-looking pattern.

The correlation structure changes after controlling for general news sentiment using *NET_NENS*. *NET_NENS* exhibits stronger and more consistent positive correlations with *IFO_CLIMAT*, *IFO_SIT*, and *ZEW_SIT*, particularly at shorter negative lags (around -2 to -5 months), suggesting improved alignment with near-term economic expectations. Correlations with *KARINA* remain pronounced, peaking at -2 to -3 months, indicating that *NET_NENS* continues to lead shifts in public economic sentiment. However, positive correlations with *UNEMPL* at zero and positive lags complicate this interpretation, as it may suggest that positive economic sentiment often follows periods of elevated unemployment.

The first-difference analysis (Figures 9c and 9d) provides additional insights into the dynamic behavior of the sentiment indices. *DIFF_NENS* shows a modest but visible contemporaneous correlation (lag 0) with *DIFF_IFO_EXP* and with *DIFF_KARINA*, suggesting that monthly changes in economic sentiment align with shifts in expectations and public economic perceptions. A particularly noteworthy finding is the negative correlation with *DIFF_UNEMPL* at lag -1, indicating that rising sentiment tends to be followed by improvements in labor market conditions—specifically, a decline in unemployment. However, a positive correlation at a forward lag of $+3$ partially offsets this pattern, potentially capturing a rebound effect or delayed sentiment response to cyclical labor market adjustments.

By contrast, though some structure remains, *DIFF_NET_NENS* exhibits generally weaker correlations with economic indicators. Positive associations with *DIFF_KARINA* are observed at negative lags (especially at -7 and -10), indicating that even after controlling for general sentiment, the index continues to anticipate shifts in business sentiment. The only other notable correlations are observed for *DIFF_ZEW_SENT* and *DIFF_UNEMPL*. In the case of *DIFF_ZEW_SENT*, a noticeable negative correlation appears at a short negative lag (-5) and again at a positive lag ($+1$), suggesting a more complex temporal relationship with investor expectations, possibly reflecting delayed or bidirectional adjustments between media sentiment and expert forecasts. For *DIFF_UNEMPL*, a positive correlation at lag $+3$ suggests that improvements in the labor market are followed by modest declines in sentiments. While this may appear counterintuitive, the pattern is neither intense nor persistent and is preceded by a weak and statistically insignificant positive correlation at lag 0. This points to a tenuous and possibly coincidental relationship rather than a robust temporal dynamic.

In summary, both *NENS* and *NET_NENS* align strongly with *KARINA*, confirming their shared focus on capturing news-based economic sentiment. *NENS* exhibits meaningful correlations with current-situation indicators such as

IFO_SIT and *ZEW_SIT*, particularly at negative lags, suggesting it contains forward-looking informational content. It also shows robust contemporaneous alignment with *IFO_EXP*, reinforcing its suitability as a real-time barometer of economic expectations. Moreover, *NENS* correlates strongly and negatively with the unemployment rate in levels (e.g., $r = -0.68^{***}$ at 10–12 month lags), implying that periods of higher sentiment tend to precede labor market improvements. In contrast, first-differenced correlations are weaker and temporally inconsistent, indicating a potentially more complex or non-linear short-term dynamic.

These results highlight that the undifferenced *NENS* index offers the clearest and most interpretable signal in bivariate settings. Its timely and substantial correlations with survey-based indicators and labor market outcomes support its use as a complementary, sentiment-based monitoring tool at the national level. For more elaborate analytical settings, especially where causal inference is of interest, the orthogonalized version, *NET_NENS*, is conceptually preferable. By abstracting from general media sentiment, it isolates the component specific to economic news and thus aligns better with causal identification strategies (van Binsbergen et al., 2024). At the regional level, however, the choice between the two becomes practically irrelevant: *REGENS* and *NET_REGENS* are nearly indistinguishable, with a correlation of $r = 0.967^{***}$ and highly similar associations with external indicators. For this reason, and to streamline the presentation, the analysis uses the orthogonalized version only, referred to simply as *REGENS*.

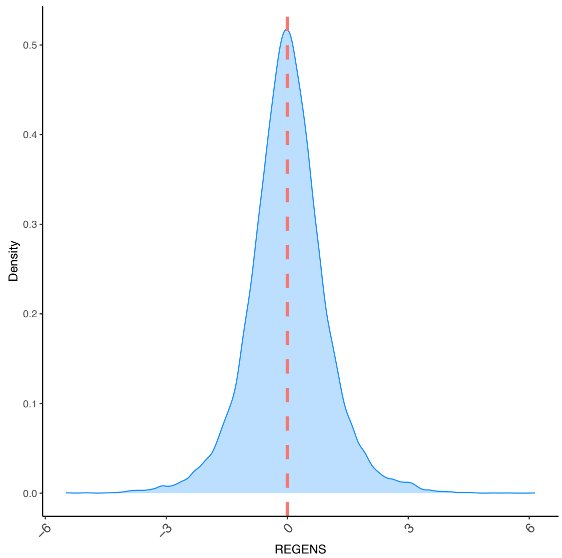
6 Regional Economic News Sentiment Index

6.1 General Descriptives of *REGENS*

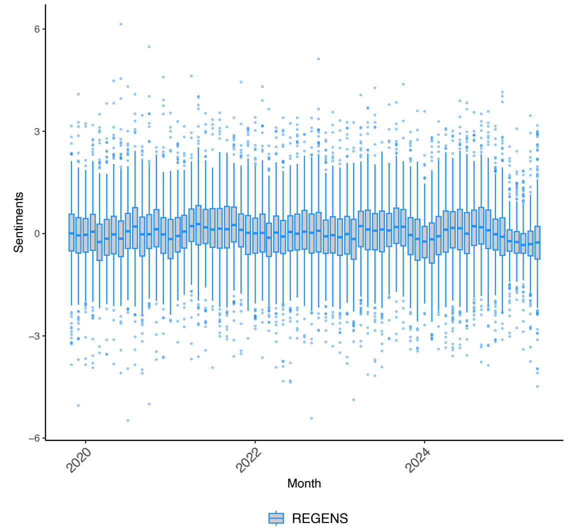
Initial observations of the Regional Economic News Sentiment (*REGENS*) are presented in Figure 10. The density plot in sub-figure 10a shows an expected distribution centered around zero. While the distribution is highly symmetric, it exhibits a subtle peak at sentiment values around 2.7.

The temporal evolution of *REGENS* is illustrated in Figures 10b and 10c. The index shows notable variation in both its mean and dispersion over time. While its temporal variance increasingly mirrors that of the national economic news sentiment index (*NET_NENS*), clear differences remain. For instance, a general sentiment decline is evident between 2021 and 2023 in the national index, which is potentially related to COVID-19 (Ozgun & Broekel, 2024). Yet, it is far less pronounced in the regional average. This likely reflects the averaging of more heterogeneous and locally framed coverage across regions, which dilutes the intensity of nationwide shocks captured more strongly in national-level reporting.

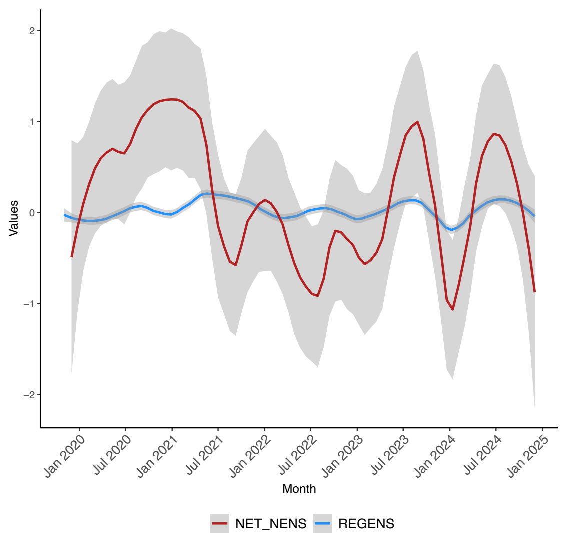
Figure 10d illustrates the interplay between the temporal and spatial dimensions of *REGENS*. Regions are ordered along the y-axis according to the similarity of their monthly sentiment trajectories, revealing several distinct groupings in the heatmap. At the top, one group is marked by predominantly positive (blue) sentiment values. At the bottom, a larger set of regions exhibits persistently negative (orange) sentiment scores. In between, a third category includes regions with more volatile sentiment patterns, alternating between positive and negative values over time. These groupings indicate that *REGENS* follows structured spatial and temporal dynamics rather than emerging from random variation.



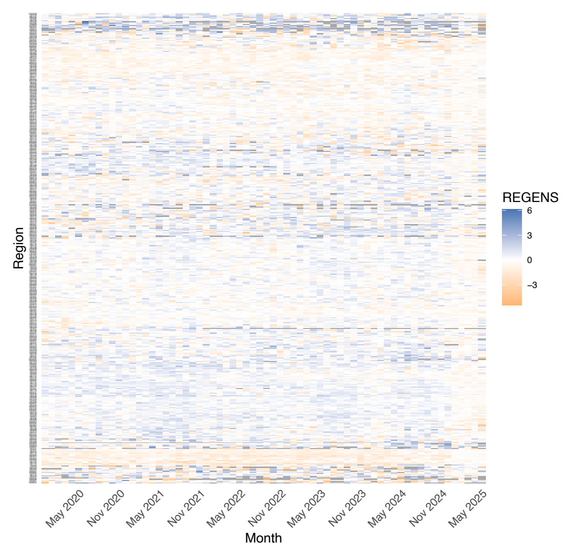
(a) Density of *REGENS*, centered around zero.



(b) Temporal Distribution of *REGENS* with boxplots representing the monthly dispersion of sentiment values.



(c) Comparison of *REGENS* and *NET_NENS*. The lines represent the means, and the grey areas the interquartile band.



(d) *REGENS* values across NUTS-3 regions and Time as heatmap.

Figure 10: Descriptive analysis of the regional economic news sentiment index (*REGENS*) across space and time.

The subsequent analysis will explore the spatial component of these variations in more detail.

6.2 Spatial Analysis of Regional Economic News Sentiment

Figure 11 visualizes the spatial distribution of the Regional Economic News Sentiment (*REGENS*) index, presenting average annual values across six years. Echoing the temporal-spatial patterns seen in Figure 10d, the maps reveal systematic geographic variation. Southwestern regions consistently exhibit more negative sentiment scores, particularly in 2020 and 2022, while Southeastern regions tend to display more positive values, most notably in 2021 and 2024. Spatial clustering is evident: areas with high (or low) sentiment are typically surrounded by similarly scoring neighbors.

Nevertheless, the maps also expose local anomalies. For example, the city of Rostock in the Northeast consistently shows more positive sentiment than its surrounding district ("Landkreis Rostock"), a pattern not mirrored in other urban centers such as Dresden or Leipzig. These discrepancies underscore the limitations of visual inspection alone and point to the need for more rigorous, quantitative analysis.

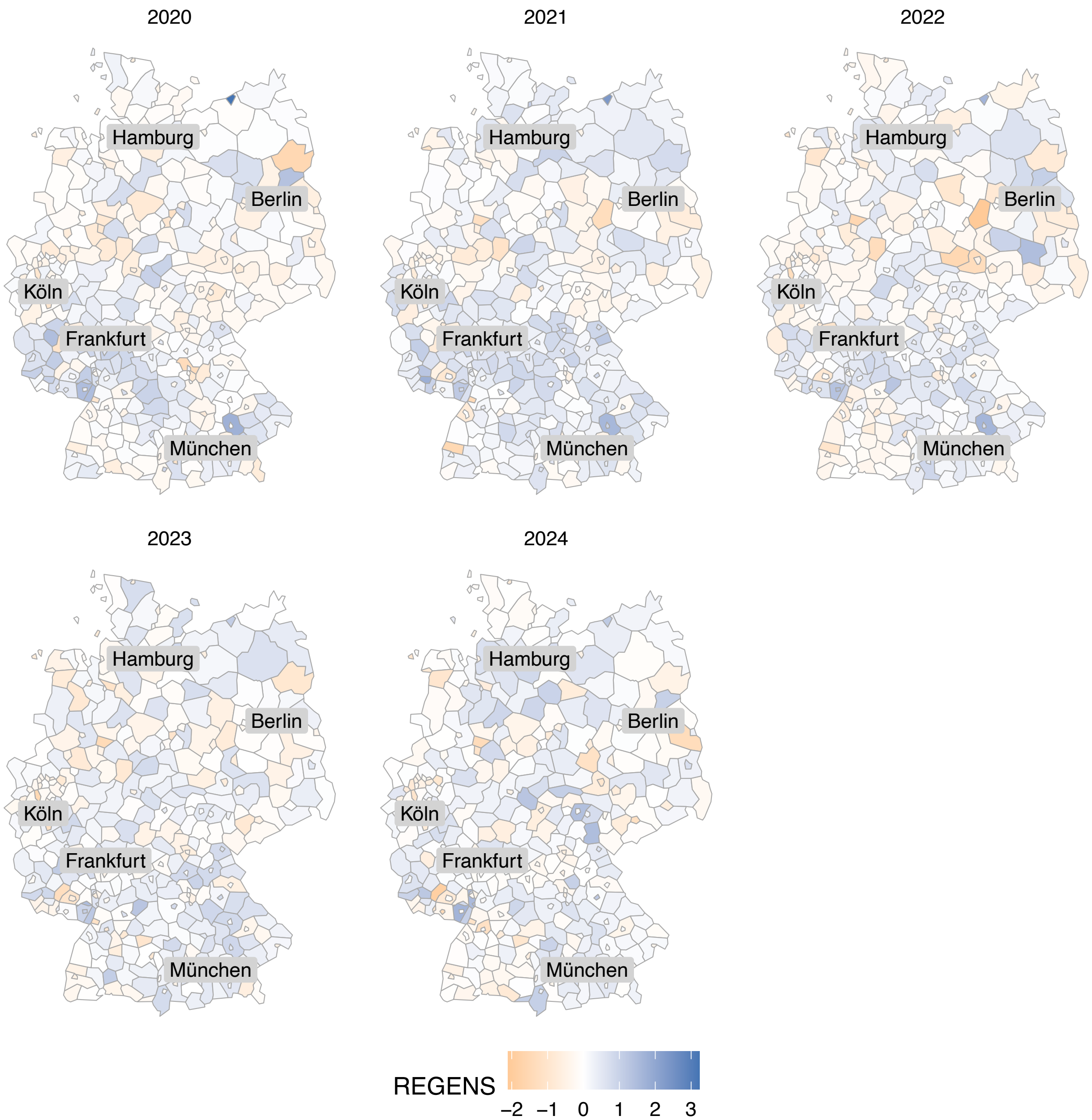


Figure 11: Annual averages of regional economic news sentiment (*REGENS*) across German NUTS-3 regions, 2020–2024. Shades of blue indicate above-average (more positive) sentiment, while shades of orange denote below-average (more negative) sentiment; white areas represent values close to zero.

To investigate the geographical characteristics of *REGENS*, a correlation analysis (Figure 12) and a cross-sectional spatial regression analysis were performed. All variables were averaged across the analyzed years, and except for *REGENS*, they were log-transformed.[12]

The correlation analysis reveals that regions with higher population density (POP_DEN), younger populations (AGE), higher GDP (GDP), better digital infrastructure (DSL), and more educational opportunities (HIGHER_EDU) are associated with more frequent reporting (NUM_NEWS) and greater diversity of news outlets (NUM_OUTLETS) for economic topics. This aligns with prior research on German local news (Ozgun & Broekel, 2021). However, with the exception of GDP growth (*GDP_G*), which shows no significant association, nearly all spatial characteristics correlate negatively with *REGENS*.

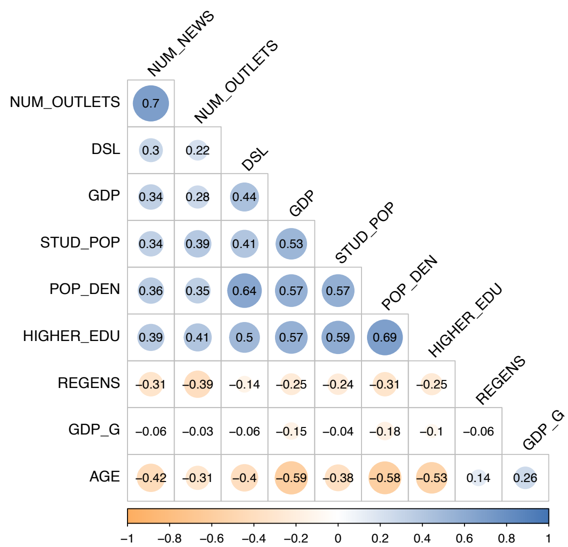


Figure 12: Correlation matrix showing the relationships between average regional economic news sentiment (*REGENS_mean*) and regional characteristics across NUTS-3 areas. Only statistically significant correlations at the 5% level are colored; non-significant cells are left white.

The results of the multivariate regression analysis are presented in Table 2. Spatial dependencies were accounted for using a $k(7)$-nearest neighbor, row-normalized, binary spatial weights matrix. The significant and positive *lambda* values (spatial error parameter) indicate that spatial autocorrelation characterizes the index, necessitating the use of a spatial autoregressive model (SAC). However, the insignificant coefficient for *rho* suggests that clustering of similar sentiment values in nearby locations, as indicated by the map, is primarily driven by shared underlying regional characteristics, such as population density and GDP.

The regression results reveal a consistently negative relationship between the number of news outlets (*NUM_OUTLETS*) and *REGENS*, suggesting that greater media diversity may be associated with more critical or less positive overall

[12] Descriptive statistics of the variables are presented in Table 9 in the Appendix.

Dependent Variable: REGENS		
	SAC Model	LAG Model 1
rho	-0.202	-0.182
	(0.199)	(0.201)
(Intercept)	4.223	4.252
	(2.740)	(2.733)
NUM_NEWS	-0.030	-0.029
	(0.026)	(0.026)
NUM_OUTLETS	-0.213***	-0.214***
	(0.050)	(0.050)
POP_DEN	-0.106***	-0.108***
	(0.027)	(0.027)
EAST	-0.222**	-0.202*
	(0.086)	(0.087)
GDP	-0.184**	-0.185**
	(0.066)	(0.066)
DSL	0.291*	0.297*
	(0.147)	(0.148)
HIGHER_EDU	0.014	0.016
	(0.072)	(0.071)
AGE	-0.471	-0.481
	(0.632)	(0.629)
STUD_POP	0.020+	0.019+
	(0.011)	(0.011)
lambda	0.533***	0.518***
	(0.133)	(0.138)
GDP_G		-1.927
		(1.924)
AIC	221.2	222.2
BIC	273.1	278.1
RMSE	0.30	0.30

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Spatial regression results for the cross-sectional determinants of regional economic news sentiment (*REGENS*) across NUTS-3 regions.

sentiment. Population density (*POP_DEN*) likewise exhibits a negative association, indicating that economic news originating from densely populated urban areas tends to be more negative than that from rural regions. This pattern holds even when controlling for economic wealth, as GDP correlates negatively with *REGENS*. Additionally, regions located in East Germany (*EAST*) display significantly lower economic news sentiment compared to their Western counterparts. In contrast, regions with higher broadband coverage (*DSL*) tend to report more positive sentiments. Other spatial characteristics, such as the number of higher education students per capita (*STUD_POP*), are only weakly significant (p<0.1), but suggest that educational environments may be associated with slightly more positive sentiment levels.

The negative association between population density and economic news sentiment is somewhat unexpected and merits further investigation. Urban areas often grapple with complex socio-economic challenges, such as high living costs, congestion, and social inequalities, which may foster more critical or problem-focused news coverage. Moreover, the competitive dynamics of urban media markets could incentivize outlets to emphasize conflict or negativity to capture audience attention. It is also plausible that the construction of *REGENS*, particularly the orthogonalization against non-economic news sentiment, contributes to these spatial patterns. Urban regions may generate a higher volume of non-economic news, particularly in domains like culture or education, which tends to be framed positively. Removing these general sentiments through orthogonalization could, in turn, suppress the net economic sentiment signal in urban contexts. However, these interpretations remain speculative and require more targeted investigation, which lies beyond the scope of the present analysis.

6.3 *REGENS* and Regional Economic Development

As previously noted, the availability of sub-annual regional data is limited, posing challenges for validating the Regional Economic News Sentiment index (*REGENS*) at the regional scale. Exceptions are the monthly unemployment rates published for NUTS3 regions in Germany by the Federal Employment Agency (2025). Although unemployment rates are generally considered lagging economic indicators (Blanchard & Katz, 1992), they remain closely aligned with current economic conditions and are frequently referenced in the media.

To assess the relationship between regional monthly unemployment rates and *REGENS*, this study adopts the methodology of van Binsbergen et al. (2024) and employs a straightforward linear predictive model. Specifically, the first differences of monthly unemployment rates are regressed on their own lags and various lags of *REGENS*, denoted as *l(DIFF_REGENS, x)*.[13] The regressions include controls for month, year, and region-specific fixed effects, and standard errors are clustered at the regional level (NUTS_ID).[14]

It is important to emphasize that this analysis does not aim to establish a causal relationship between *REGENS* and regional unemployment rates. Instead, the focus is on examining how much the index aligns with unemployment rate dynamics to assess its potential for capturing regional economic trends.

Table 3 presents the results, which reveal significantly positive coefficients for one-month lagged unemployment rates (*l(DIFF_UNEMPL, 1)*) and negative coefficients for two-month lags (*l(DIFF_UNEMPL, 2)*), reflecting the well-documented auto-regressive nature of unemployment rates (Greenlaw et al., 2018).

Regarding regional economic news sentiments, the findings indicate a negative correlation between changes in economic sentiment and unemployment rates. While the non-lagged version of *DIFF_REGENS* is not independently significant in Models 2 and 3, its first lag (*DIFF_REGENS, 1*) becomes significant in all subsequent specifications. Furthermore, when multiple lags are included, the non-lagged *DIFF_REGENS* also attains statistical significance (Models 4 to 8). Across all specifications, the coefficients are negative, suggesting that improvements in regional economic sentiment tend to be followed by declines in unemployment rates. Depending on the model, this relationship persists for up to 5 months. These findings are, by and large, consistent with the regressions using variables in levels (Table 12 in the Appendix), underscoring the robustness of this relationship. However, when considering levels, not all considered lags of *REGENS* become significant simultaneously.

[13]Results for the variables in levels are provided in Table 12 in the Appendix.
[14]Descriptive statistics for the variables are presented in Table 10 in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
l(DIFF_UNEMPL, 1)	0.298***		0.298***	0.298***	0.265***	0.291***	0.291***	0.291***	0.260***
	(0.015)		(0.015)	(0.015)	(0.018)	(0.016)	(0.016)	(0.016)	(0.017)
l(DIFF_UNEMPL, 2)	−0.084***		−0.084***	−0.083***	−0.120***	−0.078***	−0.078***	−0.078***	−0.124***
	(0.011)		(0.011)	(0.011)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
(Intercept)		0.019***							
		(0.000)							
DIFF_REGENS		0.000	−0.001	−0.004**	−0.005***	−0.003*	−0.003*	−0.003*	
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
l(DIFF_REGENS, 1)				−0.006***	−0.007***	−0.004**	−0.004**	−0.004**	−0.003*
				(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
l(DIFF_REGENS, 2)				−0.003*	−0.005**	−0.004**	−0.004**	−0.004**	
				(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	
l(DIFF_REGENS, 3)					−0.002	−0.004*	−0.004*	−0.004*	
					(0.002)	(0.002)	(0.002)	(0.002)	
l(DIFF_REGENS, 4)					−0.003+	−0.004**	−0.004**	−0.004**	
					(0.001)	(0.001)	(0.001)	(0.001)	
l(DIFF_REGENS, 5)						−0.003*	−0.003*	−0.003*	
						(0.001)	(0.001)	(0.001)	
Num.Obs.	22 549	23 830	22 549	22 549	21 475	20 978	20 978	20 978	22 549
R2	0.417	0.000	0.417	0.417	0.445	0.518	0.518	0.518	0.440
R2 Adj.	0.406	0.000	0.406	0.406	0.434	0.508	0.508	0.508	0.430
R2 Within	0.084		0.084	0.084	0.071	0.102	0.102	0.102	0.070
R2 Within Adj.	0.083		0.083	0.084	0.071	0.102	0.102	0.102	0.069
AIC	−13 075.8	151.8	−13 074.5	−13 088.4	−13 228.6	−18 518.8	−18 518.8	−18 518.8	−13 991.9
BIC	−9770.2	168.0	−9760.8	−9758.7	−9879.2	−15 171.3	−15 171.3	−15 171.3	−10 646.1
RMSE	0.18	0.24	0.18	0.18	0.17	0.15	0.15	0.15	0.17
Std.Errors	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID
FE: month	X		X	X	X	X	X	X	X
FE: NUTS_ID	X		X	X	X	X	X	X	X
FE: year					X	X	X	X	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 3: Panel regression results examining the relationship between changes in monthly regional economic news sentiment (*DIFF_REGENS*) and changes in the unemployment rate at the NUTS-3 level.

Including additional lags of the sentiment index improves the goodness of fit ($R^2\ Within\ Adj.$) marginally. However, sentiment variables generally exhibit limited explanatory power in the growth rate and levels models. For instance, the largest coefficient for *DIFF_REGENS, 1* in the unemployment growth model (Model 5) is approximately -0.007. Given a standard deviation of 1.21 for the *DIFF_REGENS* index, this implies that a one-standard-deviation increase in economic sentiment from the prior month is associated with a 0.86 percentage point decrease in the monthly change of the unemployment rate. In the levels model and a coefficient of -0.005, the effect is about (-0.48 percentage points) on the unemployment rate level (REGENS with no lag in model 7, Table 12 in the Appendix.). Hence, while statistically significant, the effect size is minor.

Nonetheless, these results demonstrate that *REGENS* has some mirroring capability. Increases in the *REGENS* index are statistically likely to precede decreases in regional unemployment rates by up to five months. This aligns with unemployment being a lagging economic indicator (Decressin & Fatás, 1995) and findings for the national-level news-based sentiment measures (Tetlock, 2007; Thorsrud, 2017). However, the index's limited explanatory and predictive power highlights the need for further research to refine and enhance its *REGENS*.

6.4 Relation with Regional GDP

In a second setup, the monitoring and predictive properties of *REGENS* are examined concerning regional GDP. Unlike unemployment rates, GDP figures are only available as annual data for German NUTS3 regions. At the time of writing, the most recent official statistics cover up to the year 2022, which limits the temporal overlap with *REGENS* for conducting a robust empirical analysis. Consequently, this study utilizes a mixture of actual regional GDP per capita (in constant prices) information and predictions of these figures (2023, 2024, and 2025) provided by Directorate-General for Economic and Financial Affairs, European Commission (2024). The GDP per capita predictions are NUTS3-level estimates based on larger NUTS2 regions. Therefore, the analysis should be seen as a complement to the main investigation using the unemployment rate.

The setup of the empirical approach is almost identical to the one with the unemployment rate as the explanatory variable. In this model, the annual growth rate of regional GDP (difference in logs) is regressed against its lagged values and the first differences of the annually averaged monthly *REGENS* values.[15] The model further incorporates different combinations of region- and year-fixed effects, and one model without any fixed effects (Model 2). Standard errors clustered at the regional level.

The regression results are displayed in Table 4.[16] Lagged GDP Growth ((*l(DIFF_GDP, 1)*) is consistently negative and highly significant across all models, indicating the known inverse relationship between the current and previous year's GDP growth rates.

Contemporary changes in regional economic news sentiment (*DIFF_REGENS*) achieve solid significance levels ($p < 0.05$ and $p < 0.001$, respectively) when excluding year-fixed effects and including both the first and second lags (Model 5). This significance vanishes when the lags are removed (Model 8), which may suggest that more complex dynamics characterize the relationship. Alternatively, excluding the year 2020, which is necessary when including the two-year lag, is decisive for identifying the signal. The models specified in levels (see Table 13) point to a modest but significant positive association between *REGENS* and GDP, particularly for the one-year lagged value in Model 3 ($p < 0.05$). In both specifications, however, effect sizes remain small. In the growth model, a one-standard-deviation increase in *DIFF_REGENS* corresponds to a mere 0.17 percentage point increase in the GDP growth rate, an order of magnitude smaller than the magnitude observed by van Binsbergen et al. (2024) for the United States. In the levels model, the corresponding effect is similarly limited: a one-standard-deviation increase in lagged *REGENS* translates into only a 0.036 percentage point increase in GDP. These findings indicate that, while statistically significant associations exist under certain model specifications, their substantive economic impact at the regional level is minimal.

In summary, while regressions on annual GDP (both as growth rates and in levels) suggest some explanatory power of *REGENS*, this relationship is indistinguishable from general annual fluctuations captured by year-fixed effects. In other words, the cross-regional variance of regions' average monthly *REGENS* values within individual years does not appear to reflect regional GDP differences in levels or growth.

Several potential explanations exist for this observation. For instance, it is questionable whether average monthly economic news sentiments validly represent sentiments that fluctuate over shorter intervals. This raises the broader issue of how to aggregate continuous temporal developments, such as news sentiments, over specific time frames. Additionally, using projected GDP values at a relatively small spatial scale (districts) may hinder identifying an empirical relationship with economic news sentiments. Therefore, while this analysis does not reliably reveal an

[15]Due to the limited number of periods available, the two-year lag of *DIFF_GDP* is not included in the model.
[16]The descriptives of the involved variables are provided in Table 11 in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
l(DIFF_GDP, 1)	−0.230***			−0.280***	−0.049**	−0.042***	−0.230***	−0.230***
	(0.020)			(0.015)	(0.015)	(0.012)	(0.020)	(0.020)
(Intercept)		0.003***						
		(0.000)						
DIFF_REGENS		0.001	0.001	0.003***	0.003***	0.000		0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		(0.001)
l(DIFF_REGENS, 1)					0.000	0.000	0.001	
					(0.001)	(0.001)	(0.001)	
l(DIFF_REGENS, 2)					0.002*	−0.001		
					(0.001)	(0.001)		
Num.Obs.	1993	2392	2392	1993	1594	1594	1993	1993
R2	0.413	0.000	0.085	0.290	0.190	0.534	0.413	0.413
R2 Adj.	0.264	0.000	−0.098	0.112	−0.083	0.376	0.264	0.263
R2 Within	0.079		0.001	0.175	0.022	0.010	0.079	0.079
R2 Within Adj.	0.078		0.000	0.174	0.019	0.006	0.078	0.078
AIC	−9597.2	−9999.6	−9414.8	−9226.1	−8611.7	−9487.0	−9596.4	−9595.3
BIC	−7335.8	−9988.1	−7102.9	−6981.5	−6446.0	−7305.2	−7329.4	−7328.4
RMSE	0.02	0.03	0.03	0.02	0.01	0.01	0.02	0.02
Std.Errors	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID
FE: NUTS_ID	X		X	X	X	X	X	X
FE: year	X					X	X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Panel regression results examining the relationship between changes in annual regional economic news sentiment (*DIFF_REGENS*) and regional GDP growth rates at the NUTS-3 level.

empirical relationship between *REGENS* and regional GDP, future studies that address these limitations may yield different results.

7 Conclusion

This study introduces the Regional Economic News Sentiments (*REGENS*) index, a new tool that leverages data from over 300 German-language news outlets, utilizing advanced text analysis methods such as transformer models. This index offers novel insights into economic news sentiments and their development over time at the regional level in Germany.[17] The study makes several significant contributions to the literature.

Firstly, the *REGENS* index addresses a critical gap in spatial studies, where news data and sentiment indices are not yet widespread. These indices offer unique features, including real-time availability, scalability, and coverage of specific and general topics. These can outperform and complement most traditional economic indicators, particularly in reflecting immediate economic conditions without the typical delays associated with surveys and similar methods. This paper strengthens existing efforts to establish this data source in the literature (Garz, 2018; Ozgun & Broekel, 2021, 2024).

Secondly, by demonstrating the (albeit still limited) potential to monitor economic developments and topic-specific sentiments in nearly real-time, this study challenges the conventional time frames typically used in regional economic studies, which often rely on annual data. This raises important questions about the appropriate temporal granularity for studying regional economies and whether shorter periods are overlooked due to data limitations or the inherent time required for micro- and meso-level processes to manifest and impact regional outcomes. It also touches on how these (now observable) short-term dynamics translate into the long-term developments that are the focus of many studies in economic geography. Addressing these questions requires developing a comprehensive theoretical framework,

[17]Updated values of *REGENS* are available at: www.tombroekel.de/regens.

collecting high-frequency (sentiment) data over an extended period, and employing methods capable of simultaneously modeling short- and long-term processes. This necessitates expanding the methodological toolbox of Regional Science, Economic Geography, and related fields by integrating finance and data science approaches, such as time series analysis and advanced text analysis. For example, wavelet gain analysis, a method primarily used in macroeconomics and finance to study high-frequency and cyclic data, is a promising methodological approach in this context (see, e.g., Broekel & Klarl, 2024).

Thirdly, and related to the previous point, using news and other real-time data allows for temporal flexibility, enabling scholars to choose time frames, from days to years, that best suit their research questions. This capability necessitates precisely defining and conceptualizing 'timing' and 'temporal sequences.' While with a different focus, works in the field of time geography may provide important ideas and concepts in such an endeavor (Hägerstrand, 1970).

Fourthly, this paper enhances research on news sentiment indicators by discussing different methods for assigning news to locations. While assignments based on readership, authorship, or content typically align at the national level, distinguishing these factors becomes crucial at the subnational level. While standard in studies on news flows (Segev, 2015), this adds a fascinating research direction, including the changing shape and content of locational information fields (Peris et al., 2021).

Fifthly, the reliance on off-the-shelf methods, including pre-trained language models, has kept computational demands relatively manageable in the present study. This approach, along with the use of pre-labeled data and a binary classification methodology, may serve as an educational illustration for developing indicators tailored to other topics.

Sixthly, this study underscores the need to better understand the media's role as a shaping force in regional economic development, including its contribution to the creation, diffusion, and adaptation of narratives. While the media's influence on economic perceptions and behaviors is well-documented at the national level, its impact at the subnational level remains underexplored. Questions related to the consolidation of media landscapes and shifts in information consumption due to the rise of social media are particularly relevant and warrant further investigation. The recently introduced concept of *local narrative landscapes* (Berle & Broekel, 2025) also offers a promising avenue for integrating news sentiments into the study of narratives across space.

Lastly, this study contributes to media studies by further promoting a new database of news headlines and snippets (*REGNES*) that includes a broad range of continuously contributing news outlets. Although currently limited to five years, it compensates with its spatial representativeness, addressing a common limitation in many media studies regarding temporal and spatial representativeness (Reinemann et al., 2024).

However, the methodological approach used in this study to construct the regional economic news index and the underlying database have some limitations, notably the absence of full-text articles. While research supports the significance of headlines and snippets (Barros et al., 2021; Djurayev, 2020; Hagar et al., 2022; Robertson et al., 2023), full texts would enrich the information content, improve location identification, enhance topic classifications, and sentiment quantification. Additionally, while reliance on pre-classified texts simplifies the process, the quality of this data conditions the identification of economic news, and cleaner training data could further enhance the employed topic classifier's discriminative power. The analysis also does not differentiate between the relative significance of events. Consequently, it treats all economic events equally. Future research could enhance the precision of the index by introducing event-specific weighting, accounting for the varying economic and social importance of different types of events featured in the news. Moreover, the data is limited to Germany and the period from 2019 to 2024. It remains to be demonstrated whether the observed features of the news-sentiment index are replicable in other countries and contexts.

A related limitation is that the sentiment index might inherently capture sudden economic shocks more effectively than gradual changes due to the nature of news reporting. The observed (weak) correlations between economic news sentiments and annual GDP growth suggest that the index might be able to reflect both cyclical fluctuations and longer-term trends. Nonetheless, the potential underreporting of incremental changes remains a limitation, and future research with extended datasets or qualitative analyses of news events could provide a deeper understanding of how these dynamics influence regional sentiment measurements. In this context, it is also important to point out that despite a growing number of studies (see, e.g., Ozgun & Broekel, 2021, 2024), there is still a limited understanding of what factors and mechanisms shape regional (economic) news sentiments.

Looking ahead, the methodological framework of *REGENS* could provide adaptable monitoring capabilities for various non-economic themes, such as sustainability transitions and social innovations, areas traditionally facing challenges related to limited data availability, timeliness, and geographic specificity. Recent studies (Abbasiharofteh et al., 2023; Berle & Broekel, 2025; Kriesch & Losacker, 2024) demonstrate that alternative data sources and methods, including news data, can offer timely insights into public perception and media representations of critical issues. These insights can potentially inform regional policy responses to sustainable practices and technological innovations. Furthermore, regional economic news sentiment indices could be extended to regions beyond Germany, shedding light on how subnational variations interact with and are shaped by national differences.

References

Abbasiharofteh, M., Krüger, M., Kinne, J., Lenz, D., & Resch, B. (2023). The digital layer: alternative data for regional and innovation studies. *Spatial Economic Analysis*, *18*(4), 507–529. https://doi.org/10.1080/17421772.2023.2193222

Acharya, S., Benhabib, J., & Huo, Z. (2017). The Anatomy of Sentiment-Driven Fluctuations. *NBER Working Paper*. https://doi.org/10.3386/w23136

Andrew, B. C. (2007). Media-generated Shortcuts: Do Newspaper Headlines Present Another Roadblock for Low-information Rationality? *Harvard International Journal of Press/Politics*, *12*(2), 24–43. https://doi.org/10.1177/1081180X07299795

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*, 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012

Arzheimer, K., & Bernemann, T. (2023). 'Place' does matter for populist radical right sentiment, but how? Evidence from Germany. *European Political Science Review*, 1–20. https://doi.org/10.1017/S1755773923000279

Avraham, E., & Ketter, E. (2012). *Media strategies for marketing places in crisis: Improving the image of cities, countries and tourist destinations*. https://doi.org/10.4324/9780080557076

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Barros, C., Vicente, M., & Lloret, E. (2021). To what extent does content selection affect surface realization in the context of headline generation? *Computer Speech & Language*, *67*, 101179. https://doi.org/10.1016/j.csl.2020.101179

Barsky, R. B., & Sims, E. R. (2012). Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence. *American Economic Review*, *102*(4), 1343–1377. https://doi.org/10.1257/aer.102.4.1343

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, *3*(3), 274–279. https://doi.org/10.1177/2043820613513390

Beaudry, P., Nam, D., & Wang, J. (2011). Do Mood Swings Drive Business Cycles and is it Rational? *NBER Working Paper*. https://doi.org/10.3386/w17651

Benhabib, J., Liu, X., & Wang, P. (2016). Sentiments, financial markets, and macroeconomic fluctuations. *Journal of Financial Economics*, *120*(2), 420–443. https://doi.org/10.1016/j.jfineco.2016.01.008

Berle, E. C., & Broekel, T. (2025). Spinning stories: Wind turbines and local narrative landscapes in germany. *Technological Forecasting and Social Change*, *211*. https://doi.org/10.1016/j.techfore.2024.123892

Blanchard, O. J., & Katz, L. F. (1992). Regional evolutions. *Brookings Papers on Economic Activity*, *1992*(1), 1–75. https://doi.org/10.2307/2534556

Blood, D. J., & Phillips, P. C. B. (1995). Recession headline news, consumer sentiments, the state of the economy and presidential popularity: a time series analysis1989–1993. *International Journal of Public Opinion Research*, *7*(1), 2–22. https://doi.org/10.1093/ijpor/7.1.2

Boschma, R., & Martin, R. (2007). Editorial: Constructing an evolutionary economic geography. *Journal of Economic Geography*, *7*, 537–548. https://doi.org/doi:10.1093/jeg/lbm021

Boydstun, A. E., Highton, B., & Linn, S. (2018). Assessing the relationship between economic news coverage and mass economic attitudes. *Political Research Quarterly*, *71*, 989–1000. https://doi.org/10.1177/1065912918775248

Broekel, T., & Klarl, T. (2024). The long-term evolution of technological complexity and its relationship with economic growth. *Papers in Evolutionary Economic Geography*.

Brooker-Gross, S. R. (1983). Spatial Aspects of Newsworthiness. *Geografiska Annaler. Series B, Human Geography*, *65*(1), 1–9. https://doi.org/10.2307/490840

Buckman, S. R., Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). News Sentiment in the Time of COVID-19. *FRBSF Economic Letter*, *08*, 1–5. https://www.frbsf.org/economic-

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. *IEEE Computational Intelligence Magazine*, *9*(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227

Coenen, L., Benneworth, P., & Truffer, B. (2012). Toward a spatial perspective on sustainability transitions. *Research Policy*, *41*(6), 968–979. https://doi.org/10.1016/j.respol.2012.02.014

Cutler, D., Poterba, J., & Summers, L. (1988, March). *What Moves Stock Prices?* (Tech. rep.). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w2538

Decressin, J., & Fatás, A. (1995). Regional labor market dynamics in Europe. *European Economic Review*, *39*(9), 1627–1655. https://doi.org/10.1016/0014-2921(94)00102-2

DellaVigna, S., & Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, *122*, 1187–1234. https://doi.org/10.1162/qjec.122.3.1187

Deller, S. C. (2010). Rural economic development: Theory and practice. *Economic Development Quarterly*, *24*(2), 151–164. https://doi.org/10.17848/9781441615050

Demmelhuber, K., Sauer, S., & Wohlrabe, K. (2023). Beyond the Business Climate: Supplementary Questions in the ifo Business Survey. *Jahrbucher fur Nationalokonomie und Statistik*, *243*(2), 169–182. https://doi.org/10.1515/jbnst-2022-0032

Directorate-General for Economic and Financial Affairs, European Commission. (2024). European economic forecast. autumn 2024. *Publications Office of the European Union*, *296*.

Djurayev, M. (2020). HEADLINE IS AN IMPORTANT ELEMENT OF NEWSPAPER TEXTS. *Theoretical & Applied Science*, *89*(09), 421–424. https://doi.org/10.15863/TAS.2020.09.89.55

Dominick, J. (1977). Geographic bias in network TV news. *Journal of Communication*, *27*(March), 94–99.

Engelsberg, J. E., & Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. *The Journal of Finance*, *66*(1), 67–97. https://doi.org/10.1111/j.1540-6261.2010.01626.x

Englmaier, F., & Fackler, T. (2022). Was sagen metadaten des ifo geschäftsklimaindex über die digitalisierung deutscher unternehmen ? *ifo Schnelldienst*, *9*, 31–34.

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Federal Employment Agency. (2025). Regional unemployment rates - Long-term series. https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Zeitreihen/Lange-Zeitreihen-Nav.html

Friedman, M., & Schwartz, A. J. (1963). *A Monetary History of the United States, 1867-1960*. Princeton University Press.

Gao, T., Yao, X., & Chen, D. (2021). Simcse: Simple contrastive learning of sentence embeddings, 6894–6910. https://github.com/princeton-nlp/SimCSE.

Garcia, D. (2013). Sentiment during Recessions. *The Journal of Finance*, *68*(3), 1267–1300.

Garz, M. (2018). Effects of unemployment news on economic perceptions – Evidence from German Federal States. *Regional Science and Urban Economics*, *68*, 172–190. https://doi.org/10.1016/j.regsciurbeco.2017.11.006

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gentzkow, M., & Shapiro, J. M. (2006). Media Bias and Reputation. *Journal of Political Economy*, *114*(2), 280–316. https://doi.org/10.1086/499414

Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*, *101*(7), 2980–3018. https://doi.org/10.1257/aer.101.7.2980

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Greenlaw, D., Hamilton, J., Harris, E., & West, K. (2018). A Skeptical View of the Impact of the Fed's Balance Sheet. *NBER Working Paper*. https://doi.org/10.3386/w24687

Greenwood, R., & Shleifer, A. (2014). Expectations of Returns and Expected Returns. *Review of Financial Studies*, *27*(3), 714–746. https://doi.org/10.1093/rfs/hht082

Gurun, U. G., & Butler, A. W. (2012). Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value. *The Journal of Finance*, *67*(2), 561–598. https://doi.org/10.1111/j.1540-6261.2012.01725.x

Hagar, N., Diakopoulos, N., & DeWilde, B. (2022). Anticipating Attention: On the Predictability of News Headline Tests. *Digital Journalism*, *10*(4), 647–668. https://doi.org/10.1080/21670811.2021.1984266

Hägerstrand, T. (1970). What about people in Regional Science? *Papers of the Regional Science Association*, *24*(1), 6–21. https://doi.org/10.1007/BF01936872

Hamilton, J. T. (2004, October). *All the News That's Fit to Sell: How the Market Transforms Information into News*. Princeton University Press. https://doi.org/10.2307/j.ctt7smgs

Hamilton, J. T. (2011, October). *All the News That's Fit to Sell*. Princeton University Press. https://doi.org/10.2307/j.ctt7smgs

Henning, M. (2019). Time should tell (more): evolutionary economic geography and the challenge of history. *Regional Studies*, *53*(4), 602–613. https://doi.org/10.1080/00343404.2018.1515481

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Howe, P. D. (2009). Newsworthy Spaces : The Semantic Geographies of Local News. *Phoenix Usa*, *4*(March 2009), 43–61.

Hüfner, F. P., & Schröder, M. (2002). Forecasting Economic Activity in Germany – How Useful are Sentiment Indicators? *ZEW Discussion Paper*, *56*(02).

Iselin, D., & Siliverstovs, B. (2016). Using newspapers for tracking the business cycle: a comparative study for Germany and Switzerland. *Applied Economics*, *48*(12), 1103–1118. https://doi.org/10.1080/00036846.2015.1093085

Jiang, C., Zhou, Y., & Chen, B. (2023). Mining semantic features in patent text for financial distress prediction. *Technological Forecasting and Social Change*, *190*(May 2022), 122450. https://doi.org/10.1016/j.techfore.2023.122450

Khoo, C. S.-g., Nourbakhsh, A., & Na, J.-C. (2012). Sentiment Analysis of Online News Text : A Case Study of Appraisal Theory. *Online Information Review*, *36*(6).

Kochetkova, N., Pronoza, E., & Yagunova, E. (2018). News Headline as a Form of News Text Compression. https://doi.org/10.1007/978-3-030-01159-8_13

Kriesch, L., & Losacker, S. (2024). Bioeconomy firms and where to find them. *Region*, *11*(1), 55–78. https://doi.org/10.18335/region.v11i1.523

Leetaru, K., & Schrodt, P. A. (2013). GDELT: Global Data on Events, Location and Tone, 1979-2012. *Annual Meeting of the International Studies Association*, (April), 1979–2012. http://data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf

Lehmann, R. (2023). *The Forecasting Power of the ifo Business Survey* (Vol. 19). Springer International Publishing. https://doi.org/10.1007/s41549-022-00079-5

Lehmann, R., & Reif, M. (2021). Predicting the German Economy: Headline Survey Indices Under Test. *Journal of Business Cycle Research*, *17*(2), 215–232. https://doi.org/10.1007/s41549-021-00055-5

Lheureux, Y. (2023). Predictive insights: leveraging Twitter sentiments and machine learning for environmental, social and governance controversy prediction. *Journal of Computational Social Science*, (0123456789). https://doi.org/10.1007/s42001-023-00228-5

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, *278*(4), 826–840. https://doi.org/10.1016/j.ins.2014.03.096

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, *36*(2), 176. https://doi.org/10.1086/267990

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. http://arxiv.org/abs/1301.3781

Möhrle, S., Löffler, C., Altmeyer, M., & Den, I. (2022). Ein nachrichtenbasierter sentiment- index für die konjunkturanalyse. *Ifo Schnelldienst*, *9*, 43–46.

Nowzohour, L., & Stracca, L. (2020). MORE THAN A FEELING: CONFIDENCE, UNCERTAINTY, AND MACROECONOMIC FLUCTUATIONS. *Journal of Economic Surveys*, *34*(4), 691–726. https://doi.org/10.1111/joes.12354

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*(February), 120692. https://doi.org/10.1016/j.techfore.2021.120692

Ozgun, B., & Broekel, T. (2024). Saved by the news? COVID-19 in German news and its relationship with regional mobility behaviour. *Regional Studies*, *58*(2), 365–380. https://doi.org/10.1080/00343404.2022.2145012

Peirsman, Y. (2020). Nlptown/bert-base-multilingual-uncased-sentiment. https://doi.org/https://doi.org/10.57967/hf/1515

Peris, A., Meijers, E., & van Ham, M. (2021). Information diffusion between Dutch cities: Revisiting Zipf and Pred using a computational social science approach. *Computers, Environment and Urban Systems*, *85*(November), 101565. https://doi.org/10.1016/j.compenvurbsys.2020.101565

Pike, A., Rodríguez-Pose, A., & Tomaney, J. (2006). *Local and regional development*. Routledge.

Poria, S., Cambria, E., Winterstein, G., & Huang, G.-B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, *69*, 45–63. https://doi.org/10.1016/j.knosys.2014.05.005

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, *2*(1).

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks, 3982–3992. https://github.com/UKPLab/

Reinemann, C., Maurer, M., Kruschinski, S., & Jost, P. (2024). The Quality of COVID-19 Coverage: Investigating Relevance and Viewpoint Diversity in German Mainstream and Alternative Media. *Journalism Studies*, 1–22. https://doi.org/10.1080/1461670X.2024.2326642

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, *7*(5), 812–822. https://doi.org/10.1038/s41562-023-01538-4

Rozado, D., & Al-Gharbi, M. (2022). Using word embeddings to probe sentiment associations of politically loaded terms in news and opinion articles from news media outlets. *Journal of Computational Social Science*, *5*(1), 427–448. https://doi.org/10.1007/s42001-021-00130-y

Rozado, D., Hughes, R., & Halberstadt, J. (2022). Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models. *PLoS ONE*, *17*(10 October), 1–14. https://doi.org/10.1371/journal.pone.0276367

Sasmaz, E., & Tek, F. B. (2021). Tweet Sentiment Analysis for Cryptocurrencies. *Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021*, (September), 613–618. https://doi.org/10.1109/UBMK52708.2021.9558914

Sauer, S. (2023). Concept of the ifo Business Survey. In S. Sauer, M. Schasching, & K. Wohlrabe (Eds.), *Handbook of ifo surveys*. Ifo Leibniz Institute for Economic Research.

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, *53*(3), 458–464. https://doi.org/10.1016/j.dss.2012.03.001

Segev, E. (2015). Visible and invisible countries: News flow theory revised. *Journalism*, *16*(3), 412–428. https://doi.org/10.1177/1464884914521579

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, *228*(2), 221–243. https://doi.org/10.1016/j.jeconom.2020.07.053

Soo, C. K. (2018). Quantifying Sentiment with News Media across Local Housing Markets. *The Review of Financial Studies*, *31*(10), 3689–3719. https://doi.org/10.1093/rfs/hhy036

Soroka, S. N. (2012). The Gatekeeping Function: Distributions of Information in Media and the Real World. *The Journal of Politics*, *74*(2), 514–528. https://doi.org/10.1017/S002238161100171X

Stone, D., Dunphy, D. C., & Smith, M. S. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* MIT Press, Cambridge, Massachusetts, London.

Strömberg, D. (2004). Mass media competition, political competition, and public policy. *Review of Economic Studies*, *71*(1), 265–284. https://doi.org/10.1111/0034-6527.00284

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thorsrud, L. A. (2017). *Words are the New Numbers: A Newsy Coincident Index of Business Cycles*. https://doi.org/10.2139/ssrn.2901452

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288.

Tieleman, T. (2012). Lecture 6.5 rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, *4*, 26.

Tunstall, L., Reimers, N., Jo, U. E. S., Bates, L., Korat, D., Wasserblat, M., & Pereg, O. (2022). Efficient few-shot learning without prompts. http://arxiv.org/abs/2209.11055

Uhl, M. W. (2011a). Nowcasting private consumption with TV sentiment. *KOF Working Papers*, *293*. https://doi.org/https://doi.org/10.3929/ethz-a-006804308

Uhl, M. W. (2011b). Reuters Sentiment and Stock Returns. *KOF Working Papers*, *288*. https://doi.org/http://dx.doi.org/10.3929/ethz-a-006620590

van Binsbergen, J. H., Bryzgalova, S., Mukhopadhyay, M., & Sharma, V. (2024). Almost (200) years of news-based economic sentiment. *NBER Working Paper*.

Vicsek, L. (2011). Costs and benefits of stem cell research and treatment: Media presentation and audience understanding in hungary. *Science Communication*, *33*, 309–340. https://doi.org/10.1177/1075547010389820

Wanta, W., Golan, G., & Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication ...*, *81*(2), 364–377. https://doi.org/10.1177/107769900408100209

Ye, X., & He, C. (2016). The new data landscape for regional and urban analysis. *GeoJournal*, *81*, 811–815. https://doi.org/10.1007/s10708-016-9737-8

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, *8*(4). https://doi.org/10.1002/widm.1253

8 Appendix A: Further Tables

News Text	News Text (English)	Sentiment (BERT)
Einer der 20 Besten. Heribert Hedrich ist in München mit dem Staatsehrenpreis des Bäckerhandwerks 2024 ausgezeichnet worden. Damit gehört seine Bäckerei mit Filialen in Hausen und Winkels zu den 20 besten.	One of the 20 best. Heribert Hedrich was awarded the State Honorary Prize for the Bakery Trade 2024 in Munich. This makes his bakery, with branches in Hausen and Winkels, one of the 20 best.	4.95
Fünf Sterne für die Kramerei am Kreisel. Die Kramerei am Kreisel in Dorfen hat bei einem bundesweiten Dorfladen-Wettbewerb überzeugt und darf sich jetzt 5-Sterne-Dorfladen nennen.	Five stars for the Kramerei am Kreisel. The Kramerei am Kreisel in Dorfen impressed in a nationwide village shop competition and can now call itself a 5-star village shop.	4.95
Bester Brunnenbauer der Nation. Der beste Brunnenbauer Deutschlands kommt aus Delbrück. Luis Bökamp holte nach dem Kammer- und Landessieg im Leistungswettbewerb des Deutschen Handwerks auch den Sieg auf Bundesebene.	Best well builder in the nation. Germany's best well builder comes from Delbrück. After winning the chamber and state competition in the German Skilled Trades Performance Competition, Luis Bökamp also secured victory at the national level.	4.93
Fünf-Sterne-Makler: Vier der besten Makler kommen aus Bayreuth. Er gilt schon was in der Branche der jährliche bundesweite Immobilienmakler-Kompass des Wirtschaftsmagazins Capital . In Bayreuth haben dabei gleich mehrere Vermittler sehr gut abgeschnitten mit einem Sieger.	Five-star real estate agents: Four of the best agents come from Bayreuth. The annual nationwide Real Estate Agent Compass by the business magazine *Capital* is highly regarded in the industry. Several agents from Bayreuth performed exceptionally well, with one emerging as the winner.	4.92
Fünf-Sterne-Restaurant in Wedel: Mit Schnitzel an die Spitze. Nachfolger von traditionsreichem Restaurant Hohmann's startet bei Google durch. Was das kleine Gasthaus von Yvonne so besonders macht.	Five-star restaurant in Wedel: To the top with schnitzel. The successor to the traditional Hohmann's restaurant is making a name for itself on Google. What makes Yvonne's small inn so special?	4.92
Fünf-Sterne-Bau mit Blick auf den Hafen ist bald fertig. Schifffahrts- und Großhandelskaufleute werden in Bremen nach den Ferien an einer nagelneuen Berufsschule unterrichtet. Die Branche ist zufrieden anders als das Handwerk das nach Bremen-Nord ziehen soll.	Five-star building with a view of the harbor is nearing completion. Shipping and wholesale merchants in Bremen will be taught in a brand-new vocational school after the holidays. The industry is satisfied, unlike the trades sector, which is being relocated to Bremen-Nord.	4.91
Der Beste seiner Art in Deutschland. Der Raiffeisen-Markt in Altenberge hat unter den rund 420 in Deutschland befindlichen Raiffeisen-Märkten den ersten Platz in puncto Qualität Kompetenz und Service erreicht.	The best of its kind in Germany. The Raiffeisen Market in Altenberge has achieved first place in terms of quality, competence, and service among the approximately 420 Raiffeisen Markets in Germany.	4.91
Herausragend Biobetrieb Ritzleben ausgezeichnet. Das Bundesministerium für Landwirtschaft zeichnet jedes Jahr drei herausragende Bio-Betriebe von über 35.000 Bio-Höfen in ganz Deutschland aus. In diesem Jahr hat es der Biohof Ritzleben aufs Treppchen geschafft.	Outstanding: Organic farm Ritzleben honored. The Federal Ministry of Agriculture annually honors three outstanding organic farms out of over 35,000 organic farms across Germany. This year, the Ritzleben organic farm made it to the podium.	4.91
Beste Kräfte des Handwerks im Kreis Paderborn ausgezeichnet. Die Handwerkskammer OWL zu Bielefeld hat jetzt die Besten des Praktischen Leistungswettbewerbs des Deutschen Handwerks aus dem gesamten Kammerbezirk geehrt. Auch viele Ausgezeichnete aus dem Kreis Paderborn sind dabei.	Best skilled tradespeople in the Paderborn district honored. The Chamber of Skilled Trades OWL in Bielefeld has now recognized the best participants in the Practical Performance Competition of the German Skilled Trades from across the district. Many awardees from the Paderborn district are also included.	4.91
Hightech-Champion aus Aachen: Pulsar Photonics punktet mit Tempo und Präzision weltweit. Außergewöhnlich und erfolgreich: Pulsar Photonics zählt zu den am stärksten wachsenden Unternehmen Europas. Ihre Maschinen und Optikmodule zur Laser-Mikrobearbeitung stehen in immer mehr Laboratorien genauso wie in der industriellen Produktion.	High-tech champion from Aachen: Pulsar Photonics scores with speed and precision worldwide. Exceptional and successful: Pulsar Photonics is among the fastest-growing companies in Europe. Their machines and optical modules for laser micromachining are increasingly found in laboratories as well as industrial production.	4.90

Table 5: Examples of Economic News with High Sentiments

News Text	News Text (English)	Sentiment (BERT)
November-Lockdown: Das ist eine reine Katastrophe. Sie haben Konzepte erstellt und investiert - doch nun müssen die Gastronomen im Dillgebiet wegen Corona erneut schließen. Wir haben stellvertretend mit dreien von ihnen gesprochen.	November lockdown: This is a total disaster. They created plans and invested – but now the restaurateurs in the Dill region have to close again due to COVID-19. We spoke to three of them as representatives.	1.05
Schrott Schmidt in Thedinghausen schließt. Thedinghausen Die Gemeinde Thedinghausen wird um ein Stück Ortsgeschichte ärmer. Schrott Schmidt wie der Betrieb in Dibbersen im Volksmund heißt schließt Ende März seine Pforten. Es handelt sich darauf wird Wert gelegt bei der Schließung nicht um eine Insolvenz sondern um eine normale Abwicklung.	Schrott Schmidt in Thedinghausen closes. Thedinghausen: The community of Thedinghausen loses a piece of local history. Schrott Schmidt, as the business in Dibbersen is popularly called, will close its doors at the end of March. Emphasis is placed on the fact that this closure is not due to insolvency but a normal liquidation.	1.07
300 Kilo pro Einwohner Soviel Müll produzieren die Steinburger Das Unternehmen Abfall Steinburg holt fast 40 000 Tonnen Abfall pro Jahr ab Das sind fast 300 Kilogramm pro Einwohner	300 kilos per resident: That's how much waste the residents of Steinburg produce. The company Abfall Steinburg collects nearly 40,000 tons of waste per year, which equates to almost 300 kilograms per resident.	1.09
Katastrophe für Käufer: Gleich mehrere Nürnberger Bauträger insolvent - mit drastischen Folgen. Nürnberg - Mehrere Tochterunternehmen der Project-Immobilien-Gruppe sind insolvent. Betroffen sind auch prominente Wohnanlagen in Nürnberg. Das raten Experten Menschen die dort bereits eine Wohnung gekauft - und Geld angezahlt - haben.	Disaster for buyers: Several real estate developers in Nuremberg are insolvent – with drastic consequences. Nuremberg: Multiple subsidiaries of the Project Real Estate Group are insolvent. Prominent residential complexes in Nuremberg are also affected. Experts give advice to people who have already purchased apartments there and made down payments.	1.10
Insolvenz droht: Medizinische Versorgungszentren in der Ortenau machen Millionen-Verlust. Die Medizinischen Versorgungszentren im Kreis sind höchst defizitär. Das Landratsamt spricht sogar von einer drohenden Pleite. Betroffen sind elf Standorte darunter zwei in Lahr und Hunderte Mitarbeiter.	Insolvency looms: Medical care centers in Ortenau are incurring millions in losses. The medical care centers in the district are highly deficit-ridden. The district office even speaks of an impending bankruptcy. Eleven locations, including two in Lahr, and hundreds of employees are affected.	1.10
Flughafen Hamburg: Beschädigte Kabel verursachen Totalausfall am Airport. 70 Minuten lang ohne Strom. Die Folge: keine Starts und Landungen in Hamburg. 13 Flüge wurden gestrichen fünf umgeleitet.	Hamburg Airport: Damaged cables cause a complete blackout at the airport. 70 minutes without power. The result: no take-offs or landings in Hamburg. Thirteen flights were canceled, five diverted.	1.10
Rosenheimer Wohnmobil-Anbieter FlexiCamper pleite - Katastrophe für Hunderte Kunden. Der Insolvenzverwalter spricht von einer Katastrophe geprellte Kunden von Abzocke : Der Wohnmobil-Anbieter FlexiCamper aus Rosenheim ist pleite. Und wohl weit über hundert Wohnmobil-Fans stehen nach Anzahlungen von zehntausenden Euro vor dem Totalschaden. Was nun?	Rosenheim camper van provider FlexiCamper bankrupt – a disaster for hundreds of customers. The insolvency administrator speaks of a disaster and cheated customers of fraud. Camper van provider FlexiCamper from Rosenheim is bankrupt, and likely well over a hundred camper fans are facing total losses after deposits of tens of thousands of euros. What now?	1.10
Gezielt ausgeführter Angriff : Cyberattacke legt komplette Fritzmeier Group lahm. Die Fritzmeier Group die unter anderem ein Werk in Bruckmühl unterhält ist Opfer eines Cyberangriffs geworden. Das bestätigte ein Unternehmenssprecher am Mittwochnachmittag 18. Januar . Das Unternehmen hat Strafanzeige gestellt.	Targeted attack: Cyberattack cripples the entire Fritzmeier Group. The Fritzmeier Group, which maintains a facility in Bruckmühl, has fallen victim to a cyberattack. A company spokesperson confirmed this on Wednesday afternoon, January 18. The company has filed a criminal complaint.	1.10
Eine Katastrophe : Galeria-Konzern erzürnt Beschäftigte in Karlsruhe Pforzheim und Offenburg. Der angeschlagene Galeria-Konzern hat seinen Zeitplan für die Verkündung von Filialschließungen abermals verschoben. Auch in Karlsruhe Pforzheim und Offenburg geht das Bangen weiter. Einem Arbeitnehmervertreter platzt jetzt der Kragen.	A disaster: Galeria Group angers employees in Karlsruhe, Pforzheim, and Offenburg. The struggling Galeria Group has once again postponed its timeline for announcing branch closures. The anxiety continues in Karlsruhe, Pforzheim, and Offenburg. A workers' representative has now lost their temper.	1.10
Streik am Montag legt VGN in Nürnberg komplett lahm: Alle Fahrzeuge bleiben stehen. Der große Streik trifft am Montag 27. März auch Nürnberg. Der VGN hat keine Notpläne für Bus U-Bahn und Tram vorgesehen. Das bedeutet Totalausfall.	Strike on Monday brings VGN in Nuremberg to a complete standstill: All vehicles remain stationary. The major strike will also affect Nuremberg on Monday, March 27. VGN has no contingency plans for buses, subways, or trams. This means a complete shutdown.	1.10

Table 6: Examples of Economic News with Low Sentiments

Model	Threshold	Precision	Recall	F1 Score
LSTM (Cased & Punctuation)	0.40	0.79	0.79	0.79
LSTM (Lowercase & No Punctuation)	0.40	0.79	0.79	0.79
Few-Shot Learning (SetFit, Cased)	0.90	0.54	0.80	0.65
DistilBERT (Fixed Embeddings, Lasso)	0.75	0.72	0.78	0.75
Contrastive Learning (Fine-Tuned DistilBERT, Lasso, 400k Samples)	0.44	0.79	0.84	0.81
Contrastive Learning (Fine-Tuned DistilBERT, Lasso, 200k Samples)	0.43	0.80	0.86	0.83

Table 7: Performance metrics (precision, recall, and F1 score) for different text classification models applied to economic news detection.

Year	ECON	OTHER
2019	3.13	3.10
2020	2.94	2.89
2021	3.02	2.92
2022	3.01	2.99
2023	3.02	3.03
2024	3.02	3.03

Table 8: Annual average sentiment values for economic (*ECON*) and non-economic (*OTHER*) news.

REGENS	1	399.00	-0.02	0.37	-0.03	-0.02	0.32	-1.32	1.43	2.75	0.12	1.28	0.02
NUM_NEWS	2	399.00	8.05	0.91	8.07	8.09	0.80	4.76	11.11	6.36	-0.34	0.81	0.05
NUM_OUTLETS	3	399.00	4.84	0.47	4.91	4.87	0.48	3.21	5.85	2.64	-0.57	0.02	0.02
POP_DEN	4	399.00	5.62	1.11	5.30	5.55	1.02	3.60	8.47	4.87	0.57	-0.76	0.06
EAST	5	399.00	0.14	0.34	0.00	0.05	0.00	0.00	1.00	1.00	2.12	2.52	0.02
GDP	6	399.00	10.43	0.33	10.39	10.40	0.27	9.66	11.92	2.26	1.12	1.89	0.02
GDP_G	7	399.00	0.00	0.01	0.00	0.00	0.01	-0.04	0.09	0.13	2.75	30.61	0.00
DSL	8	399.00	4.45	0.14	4.49	4.47	0.13	3.63	4.61	0.98	-1.82	5.76	0.01
HIGHER_EDU	9	399.00	2.66	0.37	2.57	2.63	0.33	2.01	3.85	1.84	0.84	0.14	0.02
AGE	10	399.00	3.82	0.04	3.82	3.82	0.04	3.72	3.95	0.23	0.30	-0.06	0.00
STUD_POP	11	399.00	1.99	1.95	1.96	1.86	2.90	0.00	6.22	6.22	0.27	-1.50	0.10

Table 9: Descriptive statistics for regional economic news sentiment (*REGENS*) and regional characteristics at the NUTS-3 level.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
REGENS	1	24282.00	0.02	0.97	0.03	0.02	0.81	-5.48	6.14	11.63	0.01	1.75	0.01
UNEMPL	2	24229.00	5.26	2.22	4.90	5.04	2.08	1.30	16.20	14.90	1.02	1.18	0.01
DIFF_REGENS	3	23883.00	0.00	1.21	0.01	0.00	0.92	-7.53	8.37	15.90	-0.01	3.37	0.01
DIFF_UNEMPL	4	23830.00	0.02	0.24	0.00	-0.00	0.15	-1.60	2.10	3.70	0.98	3.63	0.00

Table 10: Descriptive statistics for monthly regional economic news sentiment (*REGENS*) and unemployment at the NUTS-3 level.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
REGENS	1	2784.00	-0.02	0.60	-0.02	-0.02	0.48	-5.04	4.09	9.13	-0.26	5.56	0.01
log_GDP	2	2784.00	10.43	0.33	10.39	10.40	0.27	9.64	12.12	2.48	1.12	1.94	0.01
DIFF_REGENS	3	2385.00	-0.03	0.61	-0.02	-0.03	0.46	-3.47	6.49	9.97	0.93	10.39	0.01
DIFF_GDP	4	2385.00	0.00	0.03	0.00	0.00	0.01	-0.19	0.57	0.76	2.50	58.91	0.00

Table 11: Descriptive statistics for annual regional economic news sentiment (*REGENS*) and GDP at the NUTS-3 regional level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
l(UNEMPL, 1)	1.157***		1.157***	1.156***	1.141***	1.118***	0.831***	1.118***	1.141***
	(0.013)		(0.013)	(0.013)	(0.013)	(0.014)	(0.006)	(0.014)	(0.013)
l(UNEMPL, 2)	−0.316***		−0.316***	−0.315***	−0.327***	−0.323***		−0.323***	−0.327***
	(0.017)		(0.017)	(0.017)	(0.017)	(0.017)		(0.017)	(0.016)
(Intercept)		5.270***							
		(0.108)							
REGENS		−0.312***	−0.005**	−0.004**	−0.004*	−0.004**	−0.005**	−0.004**	−0.004*
		(0.043)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
l(REGENS, 1)				−0.004**	−0.003*	−0.002	−0.004**	−0.002	
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
l(REGENS, 2)				0.001	0.001	0.002	0.001	0.002	
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
l(REGENS, 3)					0.001	0.001	0.002	0.001	
					(0.001)	(0.001)	(0.001)	(0.001)	
l(REGENS, 4)					−0.003+	−0.003*	−0.002	−0.003*	−0.003*
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
l(REGENS, 5)						−0.001	0.000	−0.001	
						(0.002)	(0.002)	(0.002)	
Num.Obs.	22 939	24 229	22 939	22 939	21 856	21 357	21 358	21 357	21 986
R2	0.994	0.018	0.994	0.994	0.995	0.995	0.994	0.995	0.995
R2 Adj.	0.994	0.018	0.994	0.994	0.995	0.995	0.994	0.995	0.995
R2 Within	0.807		0.807	0.807	0.791	0.773	0.744	0.773	0.791
R2 Within Adj.	0.807		0.807	0.807	0.791	0.773	0.744	0.773	0.791
AIC	−16 399.7	106 970.8	−16 410.3	−16 414.9	−16 534.1	−16 507.9	−13 997.3	−16 507.9	−16 631.9
BIC	−13 054.8	106 987.0	−13 057.3	−13 045.8	−13 177.4	−13 152.9	−10 650.2	−13 152.9	−13 288.7
RMSE	0.17	2.20	0.17	0.17	0.16	0.16	0.17	0.16	0.16
Std.Errors	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID
FE: NUTS_ID	X		X	X	X	X	X	X	X
FE: month	X		X	X	X	X	X	X	X
FE: year	X		X	X	X	X	X	X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Panel regression results examining the relationship between monthly regional economic news sentiment (*REGENS*) and the unemployment rate across NUTS-3 regions.

	(1)	(2)	(3)	(4)	(5)	(6)
REGENS	0.001	−0.002	0.001	0.001		0.002
	(0.001)	(0.002)	(0.001)	(0.001)		(0.001)
l(REGENS, 1)			0.002*	0.002+	0.000	0.000
			(0.001)	(0.001)	(0.001)	(0.001)
l(REGENS, 2)			−0.001	−0.001		
			(0.001)	(0.001)		
l(log_GDP, 1)				0.186**	0.398***	0.398***
				(0.072)	(0.032)	(0.032)
Num.Obs.	2791	2791	1993	1993	2392	2392
R2	0.995	0.992	0.999	0.999	0.998	0.998
R2 Adj.	0.994	0.991	0.999	0.999	0.997	0.997
R2 Within	0.000	0.001	0.007	0.114	0.265	0.266
R2 Within Adj.	0.000	0.001	0.005	0.112	0.264	0.265
AIC	−12 386.4	−11 011.6	−11 593.7	−11 819.9	−12 310.1	−12 310.9
BIC	−9977.2	−8637.9	−9321.2	−9541.7	−9963.5	−9958.5
RMSE	0.02	0.03	0.01	0.01	0.02	0.02
Std.Errors	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID	by: NUTS_ID
FE: NUTS_ID	X	X	X	X	X	X
FE: year	X		X	X	X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Panel regression results examining the relationship between annual regional economic news sentiment (*REGENS*) and the level of regional GDP across NUTS-3 regions.

9 Appendix B: Further Classification Models

LSTM (Cased & Lowercase Variants)

The Long Short-Term Memory (LSTM) models utilize a deep learning architecture tailored for natural language processing tasks. For text preprocessing, the data was tokenized with a vocabulary size of 30,000 unique words, and input sequences were padded to a uniform length based on the 95th percentile of sequence lengths. The model architecture included an embedding layer for word representations (Mikolov et al., 2013), an LSTM layer to capture contextual dependencies (Hochreiter & Schmidhuber, 1997), and dense layers with dropout for classification. The classifier employed a softmax activation function and was optimized using categorical cross-entropy loss with RMSprop (Goodfellow et al., 2016; Tieleman, 2012). Class weights were adjusted to mitigate the imbalance in the dataset. An early stopping mechanism based on validation loss was incorporated to prevent overfitting. Two variants of this approach were tested: one preserving original casing and punctuation and another using lowercase, punctuation-free inputs.

Few-Shot Learning (SetFit)

The SetFit approach leverages few-shot learning for binary text classification, using a small balanced subset of training data to fine-tune a sentence transformer model (Tunstall et al., 2022). Text data was preprocessed to ensure non-null, valid inputs, and subsequently converted into a Hugging Face dataset format. The model, based on the pre-trained *distiluse-base-multilingual-cased-v2*, was fine-tuned using contrastive learning, followed by a linear classification head. A key aspect of this method involved sampling an equal number of examples per class to address data imbalance, enabling effective learning with a minimal amount of labeled data. Model evaluation was performed on a holdout test set, demonstrating its potential for resource-constrained settings.

DistilBERT with Fixed Embeddings

The DistilBERT-based fixed embeddings approach used pre-trained sentence transformer embeddings as input to a linear classifier. Each news item was converted into a 768-dimensional dense vector using the sentence-transformers/all-mpnet-base-v2 model (Reimers & Gurevych, 2019). These embeddings were then processed with Lasso regularization to identify the most predictive features and mitigate overfitting. This method benefits from leveraging rich contextual representations from a pre-trained model while maintaining computational efficiency during classification.