DORTMUND CENTER
FOR DATA-BASED
MEDIA ANALYSIS

DoCMA Working Paper #9

March 2022

A German Inflation Narrative

How the Media frame Price Dynamics: Results from a RollingLDA Analysis

Henrik Müller, Tobias Schmidt, Jonas Rieger, Lena Marie Hufnagel and Nico Hornig*

Cite as:
Müller, H., Schmidt, T., Rieger, J., Hufnagel, L. M. and Hornig, N. (2022). "A German Inflation Narrative - How the Media frame Price Dynamics: Results from a RollingLDA Analysis", *DoCMA Working Paper #9*. DOI: 10.17877/de290r-22632

Version 1.0, March 2022

*Henrik Müller holds the chair of Economic Policy Journalism at TU Dortmund University and is a co-founder of Dortmund Center for data-based Media Analysis (DoCMA). Tobias Schmidt, Jonas Rieger, Lena Marie Hufnagel and Nico Hornig are researchers at TU Dortmund University and DoCMA. The authors wish to thank Profs Torsten Schmidt (RWI – Leibnitz Institute for Economic Research) and Carsten Jentsch (TU Dortmund) for valuable comments and suggestions.
Corresponding author: henrik.mueller@tu-dortmund.de

Abstract

In this paper, we present a new indicator to measure the media coverage of inflation. Our Inflation Perception Indicator (IPI) for Germany is based on a corpus of three million articles published by broadsheet newspapers between January 2001 and February 2022. It is designed to detect thematic trends, thereby providing new insights into the dynamics of inflation perception over time. These results may prove particularly valuable at the current juncture, where massive uncertainty prevails due to geopolitical conflicts and the pandemic-related supply-chain jitters. Economists inspired by Shiller (2017; 2020) have called for analyses of economic narratives to complement econometric analyses. The IPI operationalizes such an approach by isolating inflation narratives circulating in the media.

Methodically, the IPI makes use of *RollingLDA* (Rieger et al. 2021), a dynamic topic modeling approach refining the rather static original LDA (Blei et al. 2003) to allow for changes in the model's structure over time. By modeling the process of collective memory, where experiences of the past are partly overwritten and altered by new ones and partly sink into oblivion, *RollingLDA* is a potent tool to capture the evolution of economic narratives as social phenomena. In addition, it is suitable to produce stable time-series, to the effect that the IPI can be updated frequently.

Our initial results show a narrative landscape in turmoil. Never in the past two decades has there been such a broad shift in inflation perception, and therefore, possibly, in inflation expectations. Also, second-round effects, such as significant wage demands, that have not played a major role in Germany for a long time, seem to be in the making. Towards the end of the time horizon, raw material prices are high on the agenda, too, triggered by the Russian war against Ukraine and the ensuing sanctions against the aggressor. We would like to encourage researchers to use our data and are happy to share it on request.

Key words: Inflation, Expectations, Narratives, Latent Dirichlet Allocation, Covid-19, Text Mining, Computational Methods, Behavioral Economics

1 Introduction

More than 60 years ago, John Muth formulated the rational expectations hypothesis. It had pretty radical implications when first applied to macroeconomic theory (Lucas 1972; Sargent and Wallace 1975), since it asserted the ineffectiveness of large parts of the economic policy toolkit, particularly monetary policy. Still, to this day Muth's core concept underpins much of macroeconomics. Put plainly, it assumes that individuals and firms form expectations about the proximate future by making use of state-of-the-art econometric forecasts. If they ignored professional prognoses and fell prey to their own false convictions instead, they would "waste information", which would be irrational. Muth conceded that individuals made erroneous predictions, but they'd be right in aggregate. In effect, expectations were "essentially the same as the predictions of the relevant economic theory" (Muth 1961, p. 316).

The rational expectations hypothesis has been disentangled by distinguishing between knowledge and information; the former concerns an analytical representation of the economic structure (the "true model"), while the latter involves the data going into it. New information is referred to established scientific knowledge and thereby transformed into the best prediction possible at the time. One standard critique of the rational expectations approach has been that it leaves open the question how exactly people obtain the knowledge of professional forecasters (e.g. Friedman 1979). That is, it lacks a micro foundation. Carroll (2003) provided an answer to this question by formulating and testing a model where individuals read newspaper articles about economic forecasts, but "only occasionally pay attention to news reports" (Carroll 2003, p. 269). They are inattentive and consequently error-prone, a trait that generates "stickyness" in aggregate expectations. In Carroll's version, the rational expectations hypothesis remains intact for professional forecasters, but it is somewhat relaxed for laypeople who only update their views from time to time.

Plenty of economic decisions are delegated to professionals who can hardly be assumed to make systematically erroneous predictions. Union leaders, for example, who resort to "true model"-inflation forecasts when calculating their demands in upcoming wage negotiations, contribute to these forecasts actually turning out to be true. What's more, if they are able to accomplish a wage rise equal to the "rationally expected" rate of price inflation, there should be no real effect on employment or output *ceteris paribus*, which is to say that the Philips curve is vertical, just as the "New Classical" branch of macroeconomics controversially asserted. To reach this result, individual workers do not need to be rationally expecting as long as union bosses are.

However, studies based on survey data show that the full-information rational expectations (FIRE) hypothesis is "increasingly at odds" with empirical findings (Coibion et al. 2018a, p. 1447). Gathering and processing macro information is costly and time-consuming, to the effect that even individual firms' inflation expectations are rather detached from reality (Coibion et al. 2018b).

More fundamentally, the notion that there could be something like a "true model" is somewhat odd. After all, in science there can be no such thing as truth, in the sense that eternally correct descriptions of reality are set in stone, but only faulty approximations that are valid until a better explanation ensues, or, as is often the case in social sciences like economics, the phenomena to be explained change. One way of thinking about the "true model" is that it is a particular form of social convention. Using it to form expectations potentially alters behavior: if economic agents and institutions make decisions according to the model's predictions, they inadvertently reinforce its accuracy. As Muth puts it, "the way expectations are formed depends specifically on the structure of the relevant system describing the economy" (Muth 1961, p. 316). This "relevant system" is deemed to be the most accurate description of the underlying structure that economists have come up with yet, itself being the result of ongoing scientific competition. The "true model" needs to be altered or replaced when the structure of the economy changes and observable phenomena cannot be explained anymore; or when new methods, new data sources and enhanced computing power produce better results.

Of course, there are different ways of "describing the economy" (Muth 1961). Mathematical models and statistical methods are one way of doing so. By using these tools, economists cut out particular aspects of economic reality and sketch relationships between causes and consequences, which can then be quantified at sufficient significance levels. But when relevant relationships break down, or new forces come to the fore that are outside the scope of the model, more fundamental problems arise.

At the time of writing, recent inflation surprises underline the notion of Kay and King (2020), who argue that under conditions of "radical uncertainty" tried and tested economic models are prone to failure, as they have been fitted to past observations derived from an economic structure that might not exist anymore. Crises and their aftermaths are periods when economic agents and policy makers navigate in blind flight, hardly having statistically reliable gauges at their disposal. Indeed, the Covid-19 pandemic and, even more recently, the Russian war against Ukraine and its global repercussions may well prove to be mega-events that induce a sweeping transformation of the underlying structure: from labor markets to the global division of labor, from mobility patterns to saving behavior, from sped-up digitization to excessive debt burdens, from notions of equity to the desired role of government, and much more. All these developments potentially alter inflation dynamics and other aspects of economic activity. Yet, we do not know, if, how and to what degree these alterations actually come to effect. Without a "true model" at hand, rational expectations are impossible to form. Now what?

One possible candidate to fill the gap is the concept of the narrative. In some aspects, economic narratives resemble the "true model" envisioned by the rational expectations hypothesis. Narratives, too, are a way of "describing the economy". Like economic models and methods, narratives can be considered as social conventions incorporating beliefs about the workings of the economy. To be meaningful, they need to be anchored in facts, that is, in observable reality. Focusing on particular aspects and neglecting the rest, they reduce complexity and tend to state unambiguous causal relationships, hence informing expectations and economic decisions of individuals, groups and entire societies. Seen this way, narratives are a mode of sense-making in uncertain and overwhelmingly complex circumstances. Contrary to mathematically formulated economic models, they do not lend themselves to numerical precision. But narratives affect the human emotional system, which in turn may help on detecting turning points and fundamental shifts that formal models are prone to miss.

Shiller (2017), who coined the term “narrative economics”, stresses that narratives emanate from the social intercourse of people, constituting a type of group think, for instances among financial market actors, which may lead to herd behavior. Yet, economic narratives as social conventions are also conveyed via the media. Journalistic as well as social media mirror the narratives that are circulating in public at a given time, emphasizing some political priorities while downplaying others. The process of economic agenda setting (McCombs and Shaw 1972) can be interpreted as a public competition of narratives where, if the information market functions reasonably well, the most convincing “system describing the economy” (Muth 1961) prevails, until it is itself outdated and replaced. This paper aims at detecting inflation narrative dynamics in the media. To this end, we are introducing a new measure we call Inflation Perception Indicator (IPI). It is based on large corpora of newspaper articles and calculated by applying a newish topic modeling method, *RollingLDA* (Rieger et al. 2021), that enables us to build consistent time-series from the otherwise unstable Latent Dirichlet Allocation (LDA) algorithm (Blei et al. 2003). The IPI is designed to capture inflation narratives floating in the public sphere, thereby complementing established measures of inflation expectations derived from surveys and financial market data. The concept of the narrative has the potential to shed extra light on inflation dynamics, since it adds context to figures: what prompts people to believe price levels to rise at a certain speed? Which causes do they have in mind? Who do they blame? Since the price momentum is largely driven by inflation expectations, identifying narrative factors that drive these projections strikes us as a worthwhile endeavor.

This paper is organized as follows: section 2 provides an overview of the relevant literature, particularly sketching a working definition of what we actually mean by an inflation narrative. Section 3 formulates the research questions. Section 4 describes the *RollingLDA* method, the data base and model calibration, while also providing details on the process of identifying media narratives. Section 5 presents the results and section 6 concludes.

2 Of Money and Men: Review of the Literature

Over the past two decades, the interplay between the economy and the media has attracted more and more attention by economists, political and communication scientists. Their findings can be summed up, paraphrasing Luhmann (2017, p. 9), that whatever we know about the economy, “we know through the mass media”.[1] This does not necessarily imply, though, that people’s beliefs are directly shaped by the stuff they read or watch, but rather that “the world that we have to deal with politically is out of reach, out of sight, out of mind. It has to be explored, reported, and imagined”, as Lippmann (1922, p. 18) puts it. What society as a whole is concerned with, is the “shared reality” (Donsbach 2014, p. 664) essentially constructed by the media.

[1] Actually, Luhmann was not just talking about the economy, but about “our society, even the world that we live in”.

The media's influence on people's views with respect to the economy seems to be pronounced. Since the economy, and the aggregates that constitute its description (e.g. consumer price indices, Gross Domestic Product), are somewhat abstract in nature and therefore do not lend themselves to direct personal observation, people's perceptions of the state of the economy mainly rely on the reality-reducing coverage of news media. Like Carroll (2003) Doms and Morin (2004) base their analysis on counts of newspaper articles about inflation over time and compare the results with economic outcomes. Lamla and Lein (2014, p. 62) conclude that the "full-information rational expectations model is clearly rejected by the data", but that people's personal forecasts are strongly influenced by information provided by the media. In particular, they find that intensive reporting improves the accuracy of consumers' inflation expectations. The media, in turn, tend to report more on economic news when negative developments arise, e.g. when inflation is higher or on the rise. If there is also a negativity bias in "news tone", however, they find a diminished influence on the accuracy of expectations. These findings broadly confirm Soroka (2006) who demonstrates that negativity contributes to a rather pessimistic perception of economic and social reality. Evaluations of personal economic well-being are less influenced by news reporting (Boomgaarden et al. 2011). In the same vein, Damstra and Boukes (2018) formulate a process of information transmission that travels from the real economy to news coverage to public perceptions of the past and the future state of the economy.

Conrad et al. (2021) find "major" media influences on the level of perceived past and expected future inflation. Users of traditional media, as opposed to social media, have a lower and more accurate inflation perception for the past year, as well as lower inflation expectations for the coming year. But consumers also obtain price information from personal experience, and the authors find that these direct observations have a larger effect on the expected future change in inflation. They interpret these results by turning the rational expectations hypothesis somewhat upside down, claiming that the (true) "economic model" seemed to be shaped by personal experience, while the data going into it would be provided by the media. Taken literally, this interpretation would have severe implications, as it might render much of central bank communication irrelevant.

In recent decades, monetary policy makers have greatly intensified their public relations efforts, and the media have become an important transmission channel (Berger et al. 2011). As traditional monetary policy instruments such as short-term interest rates and asset purchases have been all but exhausted, forward guidance has emerged as a policy instrument in its own right, i.e. communication has become part of the standard monetary policy toolbox. Informing the public about its views concerning the workings of the economy (the model) and its current state (the data), as well as stating an explicit medium- to long-term policy goal (two per cent annual CPI inflation), central banks intend to influence long-term yields and thereby forward-looking decisions concerning the real economy (investment). Hansen et al. (2019) find that there is indeed such a long-term effect on asset-prices, particularly by shaping perceptions of long-run uncertainty. Ter Ellen et al. (2021) consider "narrative monetary policy surprises" and come to conclude that such information shocks lead to real macroeconomic effects. The authors stress the role of the media as information intermediaries. Seen this way, complexity-reducing media coverage can be considered to be an efficient way to transmit monetary policy narratives, which, as Larsen et al. (2021) point out, are key ingredients in any interest rate decision and important for households. Thus, central bank communication prompt media coverage and, thereby, influences expectations. Households do not need to know anything about macroeconomics and the transmission channels of monetary policy, let alone the specifics of a "true model", as long as media coverage informs them of what to expect (and as long as central banks are credible and the media is trusted).

Media content has also been used in business-cycle forecasting. Lamla et al. (2020) show that changes in reporting tone, a measure of news sentiment, correlate with changes in the assessment and expectations of the business situation in the German manufacturing sector. In the same vein, Ashwin et al. (2021) construct a media-based nowcasting indicator that improves the forecasts especially in the first half of each quarter when other recent data are not available yet. Blagov et al. (2021), too, are able to improve short-term forecasts of business investment by applying news-based indicators and extract an "investment narrative" from an LDA exercise.

2.1 Narratives, Stories and Convictions

The term "narrative" has become something of a buzz word. In the economics literature a clear distinction from news, pure information, or simply something that's been reported in the media, is often missing. Shiller (2017; 2020) uses the term narrative synonymously with story, conviction or belief, making it hard to capture narratives empirically. Sometimes the term is even used with a dismissive slant, implying some kind of manipulative motivation. In an attempt to overcome the fuzziness, Roos and Reccius (2021) draw from several disciplines, most notably from literature studies, and stress that a core property of a narrative is that it constitutes causal relationships. Indeed, connecting causes and potential consequences is an essential part of forming expectations, that are in turn the basis for decision making, most notably in the mode of "thinking fast" (Kahnemann 2011). As dual-process theory purports, spontaneous decision-making relies on the human emotional system ("System 1") and makes use of narratives as short-cuts of reasoning.

Tuckett and Nikolic (2017) have developed a Conviction Narrative Theory (CNT) that stresses narrative thinking as the cognitive mode in which humans make sense of an uncertain environment. According to this approach, economic actors are able to make decisions by telling themselves and others stories about what is going on around them and which effects their actions are likely to have. "Narratives arise from the interplay between individual cognition and the social environment, with reasoners adopting a narrative to explain the available evidence; using that narrative to imagine plausible futures; and affectively evaluating those imagined futures to make a choice" (Johnson et al. 2020, p. 2).

What is true of economic agents also applies to policy makers: taking note of circulating narratives may improve their understanding of the current trajectory of the economy. At the same time, by communicating narratives of their own, policy makers strive to influence narratives prevailing in the economy, thereby steering expectations and thus economic decisions of businesses and private households in desired directions. Indeed, Kay and King (2020) have forcefully called for the analysis of narratives in times of "radical uncertainty", i.e. when probabilities cannot be assigned to potential outcomes and "true models" are nonexistent.

In this spirit, the Bank of England has implemented a multi-stage process by which its agents gather information through lengthy interviews with company executives. They explicitly ask the interviewees to formulate narratives about the state and likely future of the economy, about the current business climate they encounter and their own company's plans for the future. In subsequent stages of the process, the core content of thousands of interviews is being condensed to a single narrative that fits on just one or two pages. This current account of the economy then becomes an input for the assessment of the Bank's Monetary Policy Committee, alongside its traditional econometric analyses. As they face profound uncertainty about the current state and likely path of the economy, policy makers do what humans have done ever since: they turn to narrative accounts for sense-making and orientation. Moreover, the bank uses these written accounts of the economy to formulate its own narrative, by which it intends to alter – or reinforce – the narratives prevalent across the country (Tuckett et al. 2020).

Andre et al. (2021) conducted a comprehensive survey among US citizens in the wake of the inflation surprise of the fall of 2021, when the American consumer price index began rising a lot faster than expected. Participants were asked to give a brief written account of their reasoning what drove inflation. These findings by and large confirm the empirical results of related macroeconomic studies, but in addition allow for a couple of further interesting observations we'd like to stress: first, ordinary citizens, as well as company executives, explained the rise of inflation by pointing to certain protagonists, putting the blame on someone; only the economists surveyed were inclined to give predominantly structural explanations, such as global supply-chain disruptions. Second, whom participants blamed for the surge of inflation, depended on their political leanings; Republicans predominantly blamed the Biden administration and the Fed, Democrats tended to blame corporate greed and big business. Third, people's inflation convictions coincided with the editorial stances of the media they consumed, e.g. Republicans preferred Fox News, Democrats MSNCB and the New York Times.

3 Research Questions

In particular, our quest is guided by the following research questions:

RQ1 *How much media coverage is devoted to inflation over time?* As agenda setting and issue attention cycle theories hold, reporting intensity varies over time and is kicked off by certain key events or developments.

RQ2 *What is the geographical focus of inflation coverage?* The traditional news values of proximity and affectedness suggest that national price level developments should be dominant in national news media. However, since Germany is part of the Euro area and at the same time an open economy vis-à-vis the rest of the world, extensive coverage of inflationary developments abroad would be warranted as well.

RQ3 *Are we able to isolate inflation frames and narratives from the data?* Media frames consist of problem definitions, diagnosis, moral judgements and possible remedies. These features can be specified as follows:

RQ3.1 *Which causes of inflation are emphasized over time?* Economic theory suggests several factors that could contribute to rising price levels: lose monetary policy (e.g. money supply growth, excessively low interest rates), lose fiscal policy (e.g. excessive budget deficits, breaking of fiscal rules), rising prices of natural resources (e.g. oil, natural gas, agricultural products, metals, minerals, climate change mitigation policies), excessive wage growth, general overheating (e.g. positive output gap accompanied by rising prices).

RQ3.2 *Is current inflation, or the prospect of rising inflation, framed as benign or as a social ill?* Central banks in advanced economies have come to target a norm of annual average of two per cent CPI inflation. Undershooting this target could be seen as unduly risking deflation and hindering the adjustment of relative prices, overshooting would imply redistribution of income and wealth, either way constituting a moral judgement.

Our definition of a media narrative adds two more features: events (RQ1) and protagonists.

RQ3.3 *Which persons, institutions, or social groups are prevalent in inflation coverage?* Protagonists may involve actors that are deemed responsible for, suffering from or conveying judgements of inflation.

4 Data and Method

4.1 Data

LDA produces what can be called mean macro-content analysis. We are not looking for extreme occurrences on the edges of a polarized media sphere, but at some kind of average thematic media coverage on inflationary developments in mainstream media. The IPI is based on a corpus of three leading nation-wide German newspapers: Süddeutsche Zeitung (center left), Die Welt (center right) and Handelsblatt (business). The data was obtained from LexisNexis and from the publishing houses. Articles published between 1 January 2001 and 28 February 2022 were considered. In a first step, the corpus ($n = 2,866,214$) is cleaned. For example, all words are converted to lower case and umlauts are resolved. Afterwards, we delete an extended selection of stop words that do not contribute to the generation of topics or that might even involve noise. Following these preprocessing steps, an issue-specific analysis corpus ($n = 50,495$) was produced as a sub-set of the entire newspaper corpus.

4.2 Rolling LDA

Classic LDA (Blei et al. 2003) is well-suited for the identification of media frames (di Maggio et al. 2013). Frame being an inherently static concept and LDA being a static method, they fit together well over limited time-horizons and for thematically limited text corpora. For longer time horizons, however, the correspondence between research object and method is less obvious. After all, we are interested in detecting the development of thematic trends - and in monitoring potential changes in consecutive updates of the model. Classical LDA, in contrast, builds upon the assumption of structural stable topics over time, which contradicts the more fluid nature of narratives, which can change over time and even be altered in hindsight. Rieger et al. (2021) construct a dynamic version of LDA, *RollingLDA*, that allows topic structures to change over time by modeling the fading of collective memory as newer versions of narratives overwrite older ones. New data are fitted to a topic model calculated based on a rolling window of previous observations.

This approach has two major advantages. First, it solves the problem of arbitrariness that plagues the classical LDA method, which generates fundamentally different models at each run due to the random initialization of the Gibbs sampler (Griffiths and Steyvers 2004), even when exactly identical data and parameter settings are used, which runs counter to the scientific requirement of reproducibility. Second, *RollingLDA* allows us to produce a consistent, updatable time series of the Inflation Perception Indicator. More specifically, we use a combination of the selection method LDAPrototype (Rieger et al. submitted) and *RollingLDA*. At the start of the modeling process, a prototypical LDA is selected. This method solves the mentioned problem of arbitrary selection and thus improves the reliability of the results (Rieger et al. 2020). Prototyping follows a typical statistical approach: for a given parameter combination, several models are computed (usually around 100). The similarities of two LDA models are calculated based on the deviation of strictly topic matching. The LDA that has the highest average similarity to all others is selected as the prototypical LDA. The methodology is implemented in the corresponding R package ldaPrototype (Rieger 2020).

In addition to the LDAPrototype method for initial estimates of the model, we employ an implementation of LDA that uses preceding LDA results as an initialization for subsequent quarters. We modify an existing implementation of LDA (Chang 2015) by iterating the collapsed Gibbs sampler over the new data only: the topic assignments of all the previously modeled articles remain constant and we obtain assignments to the existing topics solely for all new articles. The process of fitting new data to a predefined topic model is known as “seeding”. The described procedure as a combination of a prototyping and rolling approach is implemented in the R package rollinglda (Rieger 2021).

After a thorough content analysis (described in more detail below) we chose the model’s parameter $K = 10$ topics and accordingly as Dirichlet parameters $\alpha = \eta = 1/K$, while the Gibbs sampler iterates over the dataset 200 times.

The first modeling step is limited to all the articles published between 1 January 2001 and 31 December 2005. Using a rather low threshold, we determine the vocabulary for this initial modeling: all the words that occur more than five times in this time interval are considered. This procedure removes the long tail of very infrequently occurring words that provide very little information. The result is a sub-corpus of 13,895 texts with an average of 239 tokens from a vocabulary of size 42,834 for the first modeling period. These texts from the first five-year-period are modeled using the LDAPrototype procedure as described.

In a second modeling step we consider the articles from the subsequent first month of 2006, i.e. the 193 articles published between 1 and 31 January 2006. By applying the seeding procedure described above, we model the topic assignments to these 193 articles. However, we only use the last three months as memory, i.e. we initialize the model with the 863 articles from October to December 2005. The vocabulary is extended by words that occur more than five times in the new 193 articles and that were not included in the vocabulary before. Employing this procedure, we add 11 words in the first month of 2006. The topic assignments of the new articles are initialized randomly and the Gibbs sampler iterates over each of the new articles another 200 times, while the topic assignments of all articles acting as initializing memory remain constant.

We apply the model updating procedure described for the first month of 2006 on a rolling basis for all subsequent months, so that we finally obtain assignments to the 10 topics for the entire analysis corpus with an average of 280 tokens over a vocabulary of 48,487 different words. Modeling of newly occurring articles, for example from the first month of 2022, can then be performed analogously.

The initial modeling by the LDAPrototype approach ensures the reliability of the method, while the restriction to three months as memory opens the possibility for the appearance of new topics or the mutation of existing ones. This parameter can be varied. However, three months are intuitive from the point of view that the memory spans one quarter. A larger number of months, i.e. a longer memory, could lead to very inflexible models, a reduction to fewer months to more flexible, but also to rapidly changing topics.

4.3 Calibration

To check the memory parameter, we recommend looking at the self-similarity of the topics over time. Since we allow topic structures to change, we have to make sure that topics remain stable over time to a degree that comparability is ensured, i.e. that they actually deal with similar content. Certain actors may change, new terms may be coined, some words may fade from vocabulary while others become fashionable; nonetheless, a topic should contain articles about similar issues over the entire time horizon. Here, we use cosine similarity to calculate the similarity of the word frequency vector of each topic from the current quarter to the previous one. The month-to-month similarity in our $K = 10$ model is rather high and stable for all the topics (see appendix).

To ensure the best possible validity of the IPI, we first exclude all articles that do not provide any information on inflation reporting – which applies to the majority of the texts in our corpus. Our goal is to consider all articles that actually mention inflation while also including articles that address this topic without using the term “inflation”. Furthermore, we aim at calibrating our indicator so that it responds early to changes in reporting. To achieve this, we tested various search terms to filter the dataset for relevant content. In a first step, we applied a rough search pattern that filters out all articles that do not cover the economy or prices at all. Based on these articles, we drew a random sample of 300 texts and coded each text as relevant (1) or not relevant (0). Afterwards, we applied all of our potential search terms to our sample. Based on these results we calculated their "recall" and "precision" values, respectively, as proposed by Stryker et al. (2006), who find a trade-off between the two conflicting goals of getting all the relevant articles and getting as little irrelevant ones as possible. Following this concept, a search term should provide the best compromise between a high recall and a high precision value. However, since our aim is to provide a rather sensitive indicator, we prefer a high recall over a high precision value.

The final search term in our analysis is of the form *inflation* OR teuerung OR geldentwertung* OR preissteigerung**[2]. It comes with a recall value of 0.808 and a precision value of 0.576. The selection of its individual components is based on both intuition and statistics. The inclusion of the word "inflation" is (intuitively argued) self-evident. The synonyms used, in turn, go back to their respective cosine similarities to the word "inflation".

To calculate the cosine similarities of the words in our corpus, we followed a procedure proposed by Moody (2017). First, we draw a random sub-corpus of size 50,000 from our corpus, which we tokenized and cleaned from stop-words. Based on this sub-corpus, we computed the normalized skipgram probabilities of each possible word tuple. That is, for each word in the sub-corpus, we compute the probability with which it appears near every other word. Knowing the probability of a word appearing at all, we can thus determine which words appear next to each other more often than would be expected. In our case, "nearby" means "within a word window of size 4".

Logarithmizing these values results in a pointwise mutual information (PMI) matrix where a row represents a word x, a column represents a word y, and the corresponding value represents the probability that x and y occur nearby in our texts. Since this matrix has as many rows and columns as we have words in our sub-corpus dictionary, it is of the size 101,972x101,972. This dimension (and the fact that most of the entries are zeroes) leads us to use sparse array data structures to represent the PMI matrix and to using single value decomposition (SVD) to reduce the dimensionality of our matrix. This step effectively decomposes our large and sparse matrix into two small matrices, where each row (column) represents one word. The number of columns (rows) reflects the dimension k of the resulting word vectors. Our analysis leads to vectors of the size $k = 300$.

Based on these word vectors, we determine the cosine similarity of individual words and detect those words that are most similar to our target word "inflation". This approach has the advantage that we take the peculiarities of journalistic writing and the actual word distributions within our articles into account. This makes our embeddings much more accurate than embeddings that were computed and trained on (e.g.) a Wikipedia

[2] Note that all words are written in lowercase (as is our whole corpus) Furthermore, "teuerung" is included as a word, not a pattern, which keeps us from getting false positives due to the words "besteuerung" or "steuerung".

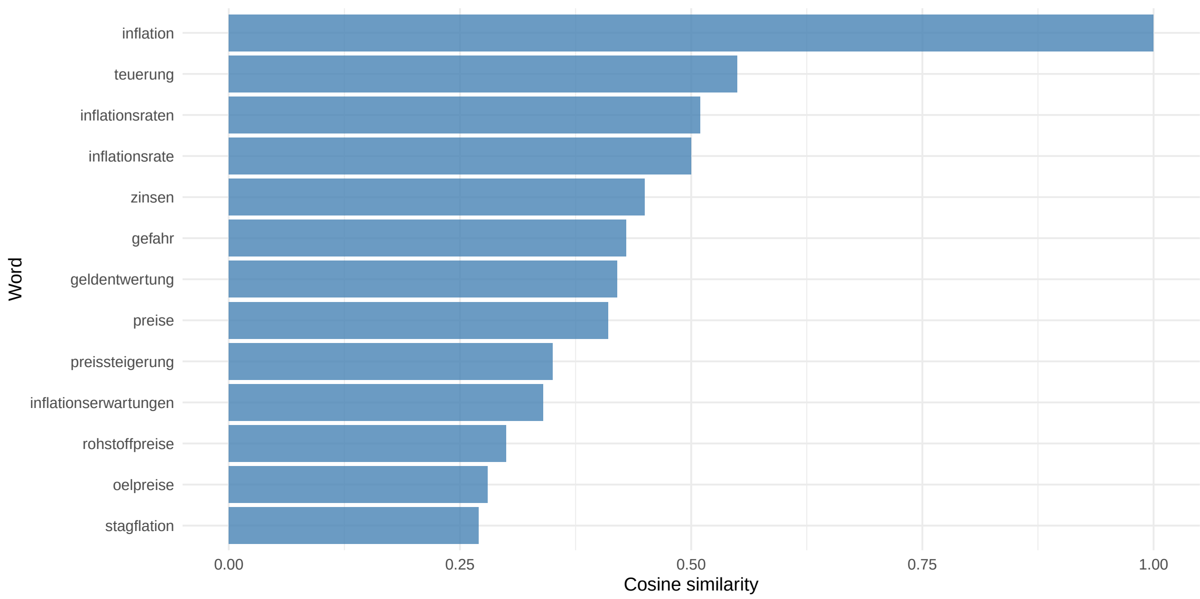

Figure 1: Nearest neighbours to word "Inflation". Source: Authors' calculations.

dataset.[3]

The words that, according to our corpus, have the highest similarity to our target word are relatives of the word "Inflation" (such as "inflationserwartung" or "inflationsrate"). Then "teuerung", "geldentwertung" and "preissteigerung" are those terms that show the greatest similarity to "Inflation" (see Figure 1). After comparing several search queries based on these words, we find the query presented above to be most valid in terms of recall and precision. Our further analysis is based on this keyword.

The resulting corpus comprises 50,495 articles, 21,318 of which are from Handelsblatt, 14,670 from Süddeutsche Zeitung and 14,507 from Die Welt. LDA requires the choice of parameter K, that is, the number of topics the algorithm is set to produce. This is a critical part of the analysis. An inadequate value of K results in topics that are thematically indistinguishable and therefore not applicable to the research questions. Instead, K should be set at a value where topics are formed that are coherent in the sense that they are separately interpretable by human researchers according to their research questions. Setting K arbitrarily, or according to some mathematical optimization approach, runs the risk of producing irrelevant result (Chang et al. 2009; Hoyle et al. 2021).

We produced four LDA models on the sub-corpus, with K set to 6, 8, 10, and 12 respectively, and resorted to an "eyeballing" procedure: three coders labelled the topics of each model independently, making use of the most characteristic word ("top words") and articles ("top articles") as well as each topic's frequency distribution over time. A value of $K = 10$ was found to be the most appropriate in terms of our research interest.

[3] Note that this procedure does not include a neural network or deep learning techniques to train our word embeddings. This makes our vectors less stable compared to static word embeddings generated with word2vec (Mikolov et al. 2013) or Glove (Pennington et al. 2014) and definitely less powerful compared to contextualized word embeddings from BERT (Devlin et al. 2019). However, since our goal is not to generate state-of-the-art word embeddings for the word inflation but to simply find good synonyms based on our dataset, our analysis is completely sufficient.

4.4 Identification of narratives

In our view, a media narrative bares considerable resemblance to a media frame. Following Entman (1993), a media frame contains four elements: a) a *problem definition*, b) a *problem diagnosis*, c) a *moral judgement*, and d) *possible remedies*. Müller et al. (2018) augment this approach by adding two more elements. Thus, a media narrative comprises a frame, or several ones, plus e) one or several *protagonists* – persons, institutions, or social groupings (nations, classes, etc.) – whose relationships are often antagonistic (villains and victims) and may change over time; and f) *events*, that are chronologically integrated and that are (often) assumed to constitute causal relationships. While frame is an inherently static concept, narrative is dynamic in nature, evolving over time, introducing new protagonists and dismissing old ones, as bygone aspects fade to oblivion.

In Müller et al. (2018; 2021) and Blagov et al. (2021), this open working definition has been put to work with regard to the measurement of economic uncertainty and the corporate investment climate respectively. We build on this earlier work and strive to capture inflation narratives in German mainstream media. LDA provides a promising tool for this endeavor, because the topics generated during the process can be interpreted as frames under certain circumstances (DiMaggio et al. 2013; Matthes and Kohring 2008). Considering our above mentioned definition of a narrative, we take this one step further and investigate, whether or not a topic reflects not only a frame, but a media narrative.

To do so, we conducted an in-depth qualitative analysis of the top-texts in each topic (i.e. the LDA-assigned articles that are most characteristic of a topic's content). To be precise, 100 top-texts were sampled for each topic of the nine relevant topics. Each article was read closely, bearing in mind the six constitutive elements of a media narrative. In this context, linguistic elements such as the choice of words, syntactical patterns or roles of actors (Lams 2016), can be of aid as well, because they may indicate the presence of a frame in an article – a condition that must be met before being able to formulate narratives. Following the principles of discourse analysis, it is at first analyzed how the different linguistic elements interact with each other in order to evaluate whether or not an article sustains a certain ideological message or societal belief. In a second step, these messages and beliefs are transferred into an overarching narrative.

5 Results: Villains and Victims

Figure 2 displays the overall results for the entire sub-corpus (without LDA-guided thematic decomposition) and monthly CPI inflation data (y-o-y) for Germany. In the 2000s, an initial period of relative calm is visible where inflation oscillates around two per cent and inflation is only a minor issue of interest. This changes in 2008 when the boom preceding the Great Financial Crisis (GFC) drives up inflation rates above three per cent annually. Inflation coverage reaches a local peak in July, in parallel with actual price developments. After that, inflationary pressures recede in an environment of plummeting financial markets, turning media attention away from inflation. In 2011, somewhat similar developments occur: inflation and its coverage are both on the rise, until the sovereign

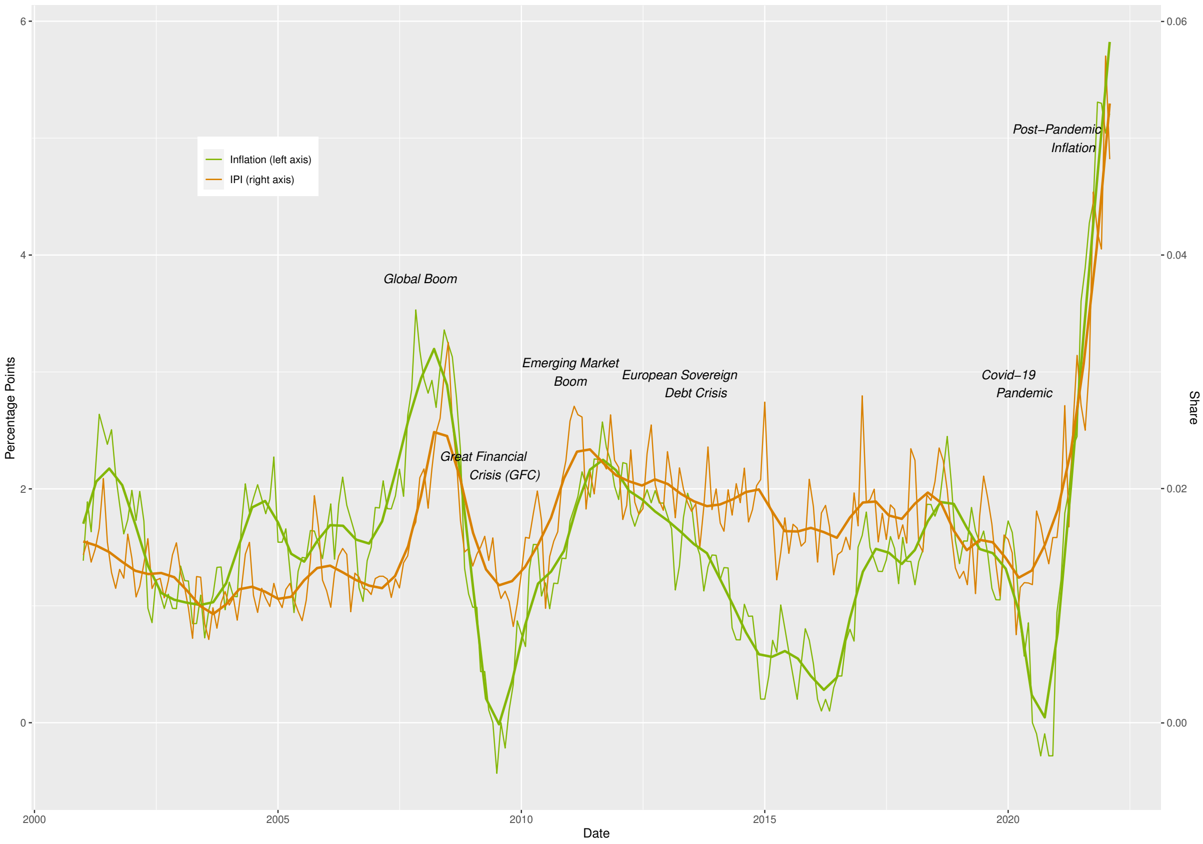

Figure 2: Inflation (CPI for Germany, y-o-y percentage change) and overall IPI (rhs) (share of analysis corpus relative to entire corpus). Source: authors' calculations, Deutsche Bundesbank

debt crisis rattles the Eurozone to an extent that leads to a gradual and prolonged decline of inflation rates. This time, however, inflation coverage does not revert to low pre-GFC levels, but stays at somewhat elevated levels even as CPI inflation falls towards zero (see Figure 2), a period where price dynamics and their coverage seem somewhat out of sync. The Covid-19 pandemic from 2020 prompts an unprecedented slump in economic activity and a subsequent elimination of inflationary pressures; media attention duly declines. The post-pandemic recovery causes inflation readings to shoot up unexpectedly which is causing considerable media coverage broadly in line with actual developments.

Turning to RQ1, the IPI curve progression indeed resembles the predictions of communication theory: media attention is largely driven by actual developments, with distinct waves of coverage being kicked-off by key-events (inflation spikes), subsiding as public attentiveness is re-directed to other issues. However, there is an exception: between 2012 and 2015 inflation coverage remains at elevated levels, even though inflation rates decline considerably. This effect can be likened to phase 5 in the Downs' (1972) taxonomy, where issue attention stays above pre-cycle levels after a wave of coverage, a phenomenon that can be attributed to collective memory and the news value of consonance, i.e. new information is referred to dominant existing frames.

Decomposing the IPI sub-corpus as described in the method section yields ten clearly distinguishable topics, nine of which, almost 90 per cent of the analysis corpus, are interpretable in terms of our research interest. Moreover, the value of LDA parameter K appears to be well-chosen, for many aspects that were touched upon in our research questions play a prominent role in different topics. Several topics have a clear geographical focus, namely 3 and 10 deal with Emerging Markets, 4 with the Eurozone, and 9 with Germany. Others fit into our thematic framework of potential causes (1, 4, 10) and consequences (2, 3, 5, 9) of inflation. Protagonists can also be detected, ranging from those potentially responsible (central bankers, governments, unions) to those adversely affected (savers, workers) – hence, there's villains and victims. Table 1 provides an overview of the model's topics.

Table 1: Overview of the LDA model's topics, January 2001 to February 2022

No.	Label	Share	Content	Protagonists	Key Events
1	Central Banks	14.1	Speculation about policy measures of advanced economies' central banks (rate changes, asset purchases/sales etc.)	ECB, Fed, Draghi, Trichet, Bernanke, Bundesbank, Weidmann, Lagarde, Yellen	7-08 (rate hike), 11-11 (rate drop), 1-15 (before QE), 7-19 (disc. of QE restart), 21/22 (inflation surprise)
2	News	6.5	News briefs covering data releases (CPI, market rates, oil price etc.)	Consumers, renters, statisticians	inflation spikes (7-08, 2-11, 12-21), QE (1-15)
3	Emerging Markets	6.5	Inflationary developments in Argentina, Turkey, Venezuela, Iran ...	Government, Argentina, Turkey, president, Kirchner, Chavez, Iran, Mugabe	EM booms (7-08, 1-11), Iran uprising (1-18), inflation spike (12-21)
4	Eurozone	12.4	Policy discussions and developments in other Eurozone countries and at EU level (fiscal stance, Stability and Growth Pact etc.)	Germany, EU, Greece, France, states, government, citizens	Sov. Debt crisis (11-11), Greek stand-off (1-15), post- covid inflation surprise (10-21)
5	Private Invest-ment	10.8	How to cope with low real yields, "private saver" perspective, focus on inflation hedges (Gold, Real Estate...)	customers, funds, banks, insurers	Boom (7-08), EZ uncertainty (sev. peaks 3-11 – 11-13), post-cov-19 inflation surprise (12-21)
6	Misc.	11.4	Diverse		
7	Financial Markets	12.5	Financial Market developments and reactions to inflation risks	Investors, analysts, USA, Fed, experts, traders	Nat. election (10-05), EM Booms (7-08, 3-11), Inaug. Trump (1-17), trade war (2-18), post-covid rally (3-21)

8	Companies	4.8	Developments in certain companies and sectors in Germany, focus on shareholder meetings and earnings calls, many calendar previews	Company, Corporation, Berlin, Paris, Lufthansa Stuttgart	Post-covid 19 inflation surprise (12-21)
9	German Politics	6.7	Collective bargaining, social, tax, fiscal policies – reactions to inflation	SPD, government, CDU, trade unions, employers, workers	Inflation surprises in 7-08, 11-11 and 11-21)
10	Raw Materi-als	8.9	Inflationary developments in EMs with particular focus on raw material demand and prices (gold, oil, copper, wheat...)	China, Russia, Turkey, India, Investors, Venezuela	Price hikes doe to trade tensions (most pronounced spike: 8-18), also inflation surprise 11-21

RQ2 can be answered right away. A large share of reporting on inflation – topics in sum comprising about two thirds of the analysis corpus – is international or European in scope. Only topics 5, 8 and 9 have a predominantely national focus. Given the international integration of product and financial markets and the European integration of monetary (and to some degree fiscal) policy, the breadth of coverage seems adequate.

5.1 Perceived Causes of inflation

Turning to RQ3.1 we focus on the media discourse about potential causes of inflation. Topics 1 and 4 capture monetary and fiscal policies, topics 9 and 10 wage negotiations and raw material prices respectively. Figure 3 shows that the four topics are moving largely in parallel. The situation in 2021/22 is unique in so far as all four topics shoot up to unprecedented levels.

The topics' frequencies reveal that the heightened awareness of inflation in German media after the GFC is driven in particular by an enduring preoccupation with the ECB. In addition, fiscal policy and business-cycle developments across the Eurozone get a lot more attention than during the days of the "great moderation" before the 2008 crash.[4] Raw material prices are only a temporary source of concern, most notably in 2008, in 2018/19 following the deterioration of trade relations under the Trump administration, and in 2021 due to post-Covid global supply-chain issues. Wages and social policies in Germany barely contribute to inflation concerns.

To get a glimpse of the framing of the topics, one way is a qualitative analysis of the top articles (see section 4 for details):

[4] The rising trends in both topics' shares can also be interpreted as the German media paying more attention to European developments and institutions, which could indicate the emergence of a European public sphere, and thus a positive development.

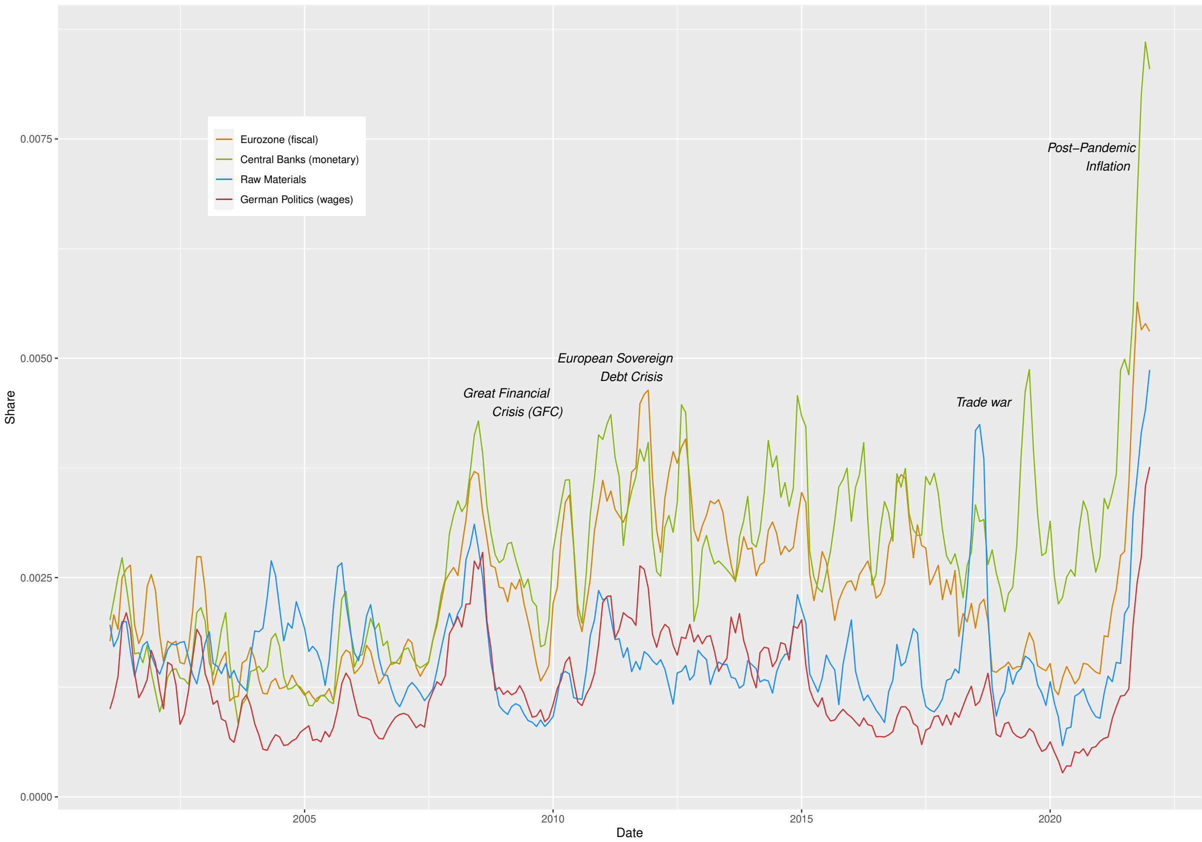

Figure 3: "Causal" IPI topics (three-month moving averages). Source: authors' calculations

Topic 1: Central banks. The top articles deal mostly with changes of the monetary policy stance. The ECB and its chiefs are the main protagonists; the Fed and other advanced-country central banks play minor roles. While the articles are mostly worded matter-of-factly, a certain bias towards a toughening of monetary policy can be detected, particularly when the ECB is concerned. Examples of headlines read: "ECB: Rates to be raised if inflationary pressures ensue", "Weber: ECB must react to inflation risk", "German inflation above ECB target", "Dangerous debate about higher inflation", "Tentative shift away from bond purchases", "ECB council member warns of inflation".[5]

Topic 4: Eurozone. Fiscal policies of Eurozone countries, the recurring efforts to reform its framework (the stability and growth pact), and the enlargement of the Euro area are major themes of the top articles. The connect to inflation is rather indirect via potentially detrimental effects of high accumulated debt levels on the conduct of monetary policy. Before the current bout of inflation two events drove this topic's dynamics: the financial crisis of 2008 and the Eurozone sovereign debt crisis peaking in 2011. Characteristic headlines read: "Financial Crisis: Greek

[5] "EZB: Zinsen werden bei Inflationsdruck erhöht", Handelsblatt, 9/23/2005; "Weber: EZB muß auf Inflationsgefahr reagieren", Welt, 5/11/2006; "Deutsche Inflation übersteigt EZB-Ziel", Handelsblatt, 5/2/2017; "Gefährliche Debatte um höhere Inflation", Welt, 6/19/2018; "Behutsame Abkehr von Anleihekäufen", Süddeutsche Zeitung, 9/21/2021; "EZB-Ratsmitglied warnt vor Inflation", Handelsblatt, 12/7/2021

bankruptcy affects Germany, too", "New Ideas for the Eurozone", "Our Currency's stability is severely threatened".[6]

Topic 9: German Politics. The focus here is on collective bargaining and social spending. These developments can be attributed to second round effects, that have the immediate implication of alleviating the consequences of inflation for wage earners and recipients of government transfers, but may prompt companies to embark on further rounds of price hikes. In times of higher inflation, such as 2008, substantial wage demands made the headlines, such as: "Eight per cent wage increase".[7] Since strikes are rare events in Germany and the unions have a reputation of taking macroeconomic side-effects into account when formulating their demands, the topic has not ranked high in the past. Towards the end of the time horizon, though, it has shot up considerably. If inflationary pressures remain on the agenda, topic 9 deserves closer monitoring.

Topic 10: Raw Materials. Top articles deal with price developments, but also with implications for exporting countries, such as Iran, Venezuela and Turkey. Hence, this topic partly overlaps with topic 3. Characteristic headlines read: "Boom in Asia clears oil markets", "Crop shortfalls ignite speculation", "Oil pushes up Gold".[8] The topic has shot up in early 2022, when one the major global raw materials exporters, Russia, was about to invade Ukraine, a big wheat producer, and severe Western sanctions against the aggressor became an issue as a consequence.

Summing up the *causes of inflation* addressed in German coverage, we find that the blame is mainly put on the central bank. Fiscal policy plays its part, too, particularly the Eurozone's weak fiscal framework, but to a lesser extent. Over the past two decades second-round effects in the form of rising wages and social benefits were not much of an issue in inflation reporting, though this seems to be changing. Towards the end of the time horizon, raw material prices are high on the agenda, as Russia was about to go to war against Ukraine, and actually did so on 24 February 2022.

5.2 Perceived Consequences of Inflation

With respect to RQ3.2 concerning *problem definitions* and *moral judgements* – is inflation good or bad? – we turn to topics mainly dealing with the consequences of inflation: topics 5, 7 and 8 in particular, topic 2 can also be attributed to this question. As already noted above, very low inflation rates may be considered problematic, if there is a risk to slide into deflationary area. Hence, some inflation may actually be portrayed as a good thing, as central banks and many economists have stressed in the past two decades. Here, we investigate whether these professional views gained traction in the media.

[6] "Finanzkrise: Der Bankrott der Griechen streift auch Deutschland", Welt, 12/10/2009; "Neue Ideen für die Eurozone", Süddeutsche Zeitung, 6/4/2004; "Stabilität unserer Währung ist schwer gefährdet", Welt 7/9/2005

[7] "Acht Prozent mehr Lohn", Handelsblatt, 9/9/2008

[8] "Asienboom fegt den Ölmarkt leer", Handelsblatt, 9/29/2004; "Ernteausfälle heizen Spekulation an", Handelsblatt, 3/7/2006; "Öl treibt Gold", Handelsblatt, 11/30/2005

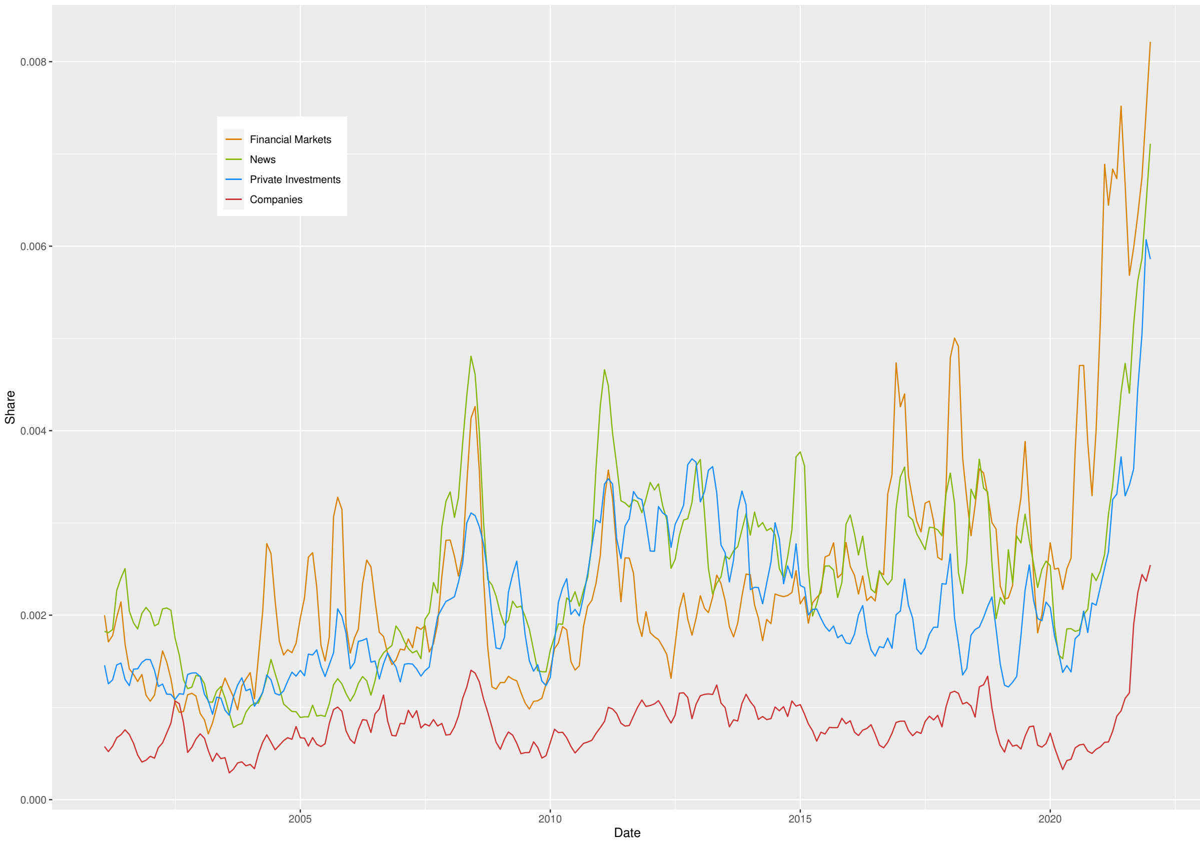

Figure 4: "Consequences" IPI topics. Source: Authors' calculations

Figure 4 shows three peaks associated with higher-than-usual CPI inflation in Germany, in 2008, 2011 and 2021. Additionally, financial market coverage is also driven by inflation concerns in 2017 and 2018, when the phasing-out of the ECB's asset purchase program was conditional on inflation developments, so analysts and investors watched the associated developments closely.

The topics' content can be summarized as follows:

> **Topic 5: Private Investment.** The consequences of inflation for middle-income savers and consumers are addressed. Low inflation is seen as good (for consumers), very low (real) rates as bad (for savers). Articles address the need to take some risk in order to get positive returns on investments as well as the need to save, even though returns are sparse. Typical headlines read: "Bundesbank: saving pays off despite mini rates", "Savers' fears", "Zero per cent on savings".[9]
>
> **Topic 7: Financial Markets.** The articles share an unambiguously negative framing of inflation, the common thread being that rising inflation (readings and prospects) fuel fears of monetary tightening, thereby depressing asset prices. Characteristic headlines read: "Inflation fears depress sentiment", "Global bourses: cheap

[9] "Bundesbank: Sparen lohnt sich trotz Mini-Zinsen", Welt 25.10.2015; "Die Angst der Sparer". 10/14/2011, Handelsblatt; "Null Prozent auf das Ersparte", Süddeutsche Zeitung, 8/12/2014

oil provides push", "Gold becomes more popular with investors".[10]

Topic 8: Companies. This small topic consists mainly of short article that announce the publication of company results in the coming days. The impact of inflation on companies is seen as rather neutral.

Topic 2: News. As the neutral label indicates, this topic contains new information about price developments at different levels of the economy (consumers, producers etc.) and in different countries. Declining as well as rising price levels are reported. The top articles suggest matter-of-factly, balanced reporting. Typical headlines read: "Producer prices rise more slowly", "Italy: Inflation rate drops slightly", "Wholesale goods only slightly more expensive".[11]

Overall, the consequences of inflation are portrayed as negative. Positive aspects of (some) inflation are not detectable in the media considered.

5.3 A German inflation narrative

The LDA-aided deconstruction of our newspaper corpus has yielded topics whose frequencies move broadly in parallel, as they are mainly driven by the same events and developments. It is therefore straightforward to consider these topics as different strands of a single narrative. We return to RQ3, which asks whether we can detect a German inflation narrative. Taking account of Entman's (1993) definition of a frame and the definition of a narrative from Müller et al. (2018), we name the six relevant aspects:

(a) *Problem definition*: inflation is problematic, particularly for savers and consumers.

(b) *Problem diagnosis* (RQ3.1): inflation is driven by rising raw material prices (short-term), longer-term inflationary developments are caused by monetary and fiscal policy. Over our time-horizon, second-round effects encompassing wages and social benefits have not attracted much attention.

(c) *Moral judgement* (RQ3.2): inflation is bad, though the low rates during the period of time considered here have not been much of a problem. However, the expropriation of savers, due to extreme low or negative real yields, is to be condemned.

(d) *Possible remedies*: tighter monetary and fiscal policies – even if that hurts financial markets in the short-run.

(e) *Protagonists* (RQ3.3):

[10] "Inflationsängste drücken Stimmung", Süddeutsche Zeitung, 2/22/2001; "Weltboersen: Billigeres Oel gibt Auftrieb, Süddeutsche Zeitung", 3/22/2008; "Gold bei Anlegern immer beliebter", Handelsblatt, 7/4/2016

[11] "Produzentenpreise steigen langsamer", Süddeutsche Zeitung, 1/10/2001, "Italien: Inflationsrate geht leicht zurück", Handelsblatt 12/22/2003; "Großhandelswaren nur etwas teurer", Süddeutsche Zeitung 11/11/2003

- villains: central banks, finance ministries, raw material-exporting countries

- victims: savers, investors, workers

(f) *Events* (RQ3.1): three major events can be detected: 2008, associated with higher import prices and substantial wage demands; 2011, associated with higher import prices and substantial loosening of monetary conditions by the ECB; 2021, associated with post-Covid supply-constrains, expensive raw materials, ultra-loose monetary and fiscal policies, potentially even rising wages and benefits.

6 Conclusion

We may not possess the "true model" that the rational expectations hypothesis envisioned, but we were able to isolate a collective inflation narrative for Germany and thereby capture convictions about the causes and consequences of inflation, that may in turn contribute to the formation of economic expectations. To this end, we developed the Inflation Perception Indicator, a new measure to capture convictions about the causes and the consequences of inflation. Our initial results show a narrative landscape in turmoil. Never in the past two decades, that are the scope of this paper, has there been such a broad shift in inflation perception, and therefore, possibly, in inflation expectations. Second-round effects, such as significant wage demands, were not much of an issue in inflation reporting in the past, but this seems to be changing. Towards the end of the time horizon, raw material prices are high on the agenda, too, triggered by the prospects of Russian war against Ukraine.

Whether the IPI and its thematic components lend themselves to enhance inflation-related forecasts of macro variables, asset prices or survey-based expectations is a question we left unanswered in this paper. This is an issue for future research.

In any case, our indicator provides an impression of the degree to which inflation is becoming entrenched in the public's perception. It may thus be of interest to monetary policy makers who have tended to complain about an alleged tendency in the German media to exaggerate inflation risks (e.g. Schnabel 2021). The IPI distinguishes between different factors that influence inflation dynamics, such as raw-material prices, monetary and fiscal policy and collective bargaining. Since it is based on an analysis of leading broadsheet newspaper corpora, it permits a macro view of public – or rather: publicized – opinion: an average mainstream perception of inflation. That's what we've been looking for: a common explanatory thread, not the competition of different ones. But the latter would indeed be an interesting field of research, too. Looking for the extreme views on inflation on the edges of the public sphere could shed light on the political dynamics in polarized societies.

In this paper we solely considered German newspapers. Possibly, we have produced a distinctly German view on inflation stemming from the preoccupations of a nation that suffered two monetary breakdowns and ensuing currency reforms in the 20th century. The ongoing coverage of the ECB and its expansionary policy stance in the 2010s, over a period when actual inflation rates were very low, may mirror these earlier historic experiences. IPIs conducted for other nations may show quite different patterns. Comparing country specifics certainly looks like a fertile field of research.

The IPI is based on a corpus of roughly three million articles published by broadsheet newspapers Süddeutsche Zeitung, Die Welt and Handelsblatt between January 2001 and February 2022. Methodically, the IPI makes use of *RollingLDA* (Rieger et al. 2021), a dynamic topic modeling approach refining the rather static original LDA (Blei et al. 2003) to allow for changes in the model's structure over time. By modeling the process of collective memory, where experiences of the past are partly overwritten and altered by new ones and partly sink into oblivion, *RollingLDA* is a potent tool to capture the evolution of economic narratives as social phenomena. What's more, it is suitable to produce stable time-series to the effect that the IPI can be updated frequently. Future research may combine this approach with other computational methods such as entity recognition and sentiment analysis.

References

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart (2021). *Inflation Narratives*. ECONtribute Discussion Papers Series 127. University of Bonn and University of Cologne, Germany. URL: https://ideas.repec.org/p/ajk/ajkdps/127.html.

Ashwin, Julian, Eleni Kalamara, and Lorena Saiz (2021). *Nowcasting euro area GDP with news sentiment: a tale of two crises*. ECB Working Paper Series 2616.

Berger, Helge, Micheal Ehrmann, and Marcel Fratzscher (2011). "Monetary Policy in the Media". In: *Journal of Money, Credit and Banking* 43.4, pp. 689–709. URL: http://www.jstor.org/stable/20870072.

Blagov, Boris, Henrik Müller, Carsten Jentsch, and Torsten Schmidt (2021). *The Investment Narrative - Improving Private Investment Forecasts with Media Data*. Ruhr Economic Papers 922. DOI: 10.4419/96973067.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3, pp. 993–1022. DOI: 10.1162/jmlr.2003.3.4-5.993.

Boomgaarden, Hajo G., Joost van Spanje, Rens Vliegenthart, and Claes H. de Vreese (2011). "Covering the crisis: media coverage of the economic crisis and citizens' economic expectations". In: *Acta Politica* 46, pp. 353–379. DOI: 10.1057/ap.2011.18.

Carroll, Christopher D. (2003). "Macroeconomic Expectations of Households and Professional Forecasters". In: *The Quarterly Journal of Economics* 118.1, pp. 269–298. DOI: 10.1162/00335530360535207.

Chang, Jonathan (2015). *lda: Collapsed Gibbs Sampling Methods for Topic Models*. R package version 1.4.2. URL: https://CRAN.R-project.org/package=lda.

Chang, Jonathan, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei (2009). "Reading Tea Leaves: How Humans Interpret Topic Models". In: *NIPS: Advances in Neural Information Processing Systems*. Curran Associates Inc., pp. 288–296. URL: https://papers.nips.cc/paper/2009/hash/f92586a25bb3145facd64ab20fd554ff-Abstract.html.

Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar (2018a). "The Formation of Expectations, Inflation, and the Phillips Curve". In: *Journal of Economic Literature* 56.4, pp. 1447–1491. DOI: 10.1257/jel.20171300.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar (2018b). "How Do Firms Form Their Expectations? New Survey Evidence". In: *American Economic Review* 108.9, pp. 2671–2713. DOI: 10.1257/aer.20151299.

Conrad, Christian, Zeno Enders, and Alexander Glas (2021). *The role of information and experience for households' inflation expectations*. Deutsche Bundesbank Discussion Paper 07/2021. DOI: 10.18452/21833.

Damstra, Alyt and Mark Boukes (2018). "The Economy, the News, and the Public: A Longitudinal Study of the Impact of Economic News on Economic Evaluations and Expectations". In: *Communication Research* 48.1, pp. 26–50. DOI: 10.1177/0093650217750971.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 NAACL-Conference, Volume 1 (Long and Short Papers)*. ACL, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

DiMaggio, Paul, Manish Nag, and David Blei (2013). "Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding”. In: *Poetics* 41.6, pp. 570–606. DOI: 10.1016/j.poetic.2013.08.004.

Doms, Mark and Norman Morin (2004). *Consumer sentiment, the economy, and the news media.* Finance and Economics Discussion Series. URL: https://www.federalreserve.gov/pubs/feds/2004/200451/200451pap.pdf.

Donsbach, Wolfgang (2014). “Journalism as the new knowledge profession and consequences for journalism education”. In: *Journalism* 15.6, pp. 661–677. DOI: 10.1177/1464884913491347.

Downs, Anthony (1972). “Up and Down with Ecology - the Issue-Attention Cycle”. In: *The Public Interest* 28. URL: https://www.proquest.com/magazines/up-down-with-ecology-issue-attention-cycle/docview/1298108041.

Entman, Robert M. (1993). “Framing: Toward Clarification of a Fractured Paradigm”. In: *Journal of Communication* 43.4, pp. 51–58. DOI: 10.1111/j.1460-2466.1993.tb01304.x.

Friedman, Benjamin M. (1979). “Optimal expectations and the extreme information assumptions of ‘rational expectations’ macromodels”. In: *Journal of Monetary Economics* 5.1, pp. 23–41. DOI: https://doi.org/10.1016/0304-3932(79)90022-9.

Griffiths, Thomas L. and Mark Steyvers (2004). “Finding scientific topics”. In: *Proceedings of the National Academy of Sciences* 101.suppl 1, pp. 5228–5235. ISSN: 0027-8424. DOI: 10.1073/pnas.0307752101.

Hansen, Stephen, Michael McMahon, and Matthew Tong (2019). “The long-run information effect of central bank communication”. In: *Journal of Monetary Economics* 108, pp. 185–202. DOI: 10.1016/j.jmoneco.2019.09.002.

Hoyle, Alexander, Pranav Goel, Andrew Hian-Cheong, Denis Peskov, Jordan Lee Boyd-Graber, and Philip Resnik (2021). “Is Automated Topic Model Evaluation Broken? The Incoherence of Coherence”. In: *NeurIPS: Advances in Neural Information Processing Systems.* URL: https://openreview.net/forum?id=tjdHCnPqoo.

Johnson, Samuel G B, Avri Bilovich, and David Tuckett (2020). *Conviction Narrative Theory: A Theory of Choice Under Radical Uncertainty.* DOI: 10.31234/osf.io/urc96.

Kahnemann, Daniel (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux.

Kay, J. A. and Mervyn A. King (2020). *Radical Uncertainty - Decision-Making Beyond the Numbers.* London: The Bridge Street Press.

Lamla, Michael J. and Sarah M. Lein (2014). “The role of media for consumers’ inflation expectation formation”. In: *Journal of Economic Behavior & Organization* 106, pp. 62–77. DOI: 10.1016/j.jebo.2014.05.004.

Lamla, Michael J., Sarah M. Lein, and Jan-Egbert Sturm (2020). “Media reporting and business cycles: empirical evidence based on news data”. In: *Empirical Economics* 59.3, pp. 1085–1105. DOI: 10.1007/s00181-019-01713-5.

Lams, Lutgard (2016). “China: Economic magnet or rival? Framing of China in the Dutch- and French-language elite press in Belgium and the Netherlands”. In: *International Communication Gazette* 78.1–2, pp. 137–156. DOI: 10.1177/1748048515618117.

Larsen, Vegard H., Leif Anders Thorsrud, and Julia Zhulanova (2021). “News-driven inflation expectations and information rigidities”. In: *Journal of Monetary Economics* 117, pp. 507–520. DOI: 10.1016/j.jmoneco.2020.03.004.

Lippmann, Walter (1922). *Public Opinion.* Harcourt, Brace and Company. URL: https://archive.org/details/publicopinion00lippgoog/mode/1up.

Lucas, Robert E (1972). “Expectations and the neutrality of money”. In: *Journal of Economic Theory* 4.2, pp. 103–124. DOI: 10.1016/0022-0531(72)90142-1.

Luhmann, Niklas (2017). *Die Realität der Massenmedien*. German. 5th ed. Wiesbaden: Springer VS. DOI: 10.1007/978-3-658-17738-6.

Matthes, Jörg and Matthias Kohring (2008). "The Content Analysis of Media Frames: Toward Improving Reliability and Validity". In: *Journal of Communication* 58.2, pp. 258–279. DOI: 10.1111/j.1460-2466.2008.00384.x.

McCombs, Maxwell E. and Donald L. Shaw (1972). "The Agenda-Setting Function of Mass Media". In: *The Public Opinion Quarterly* 36.2, pp. 176–187. URL: http://www.jstor.org/stable/2747787.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781.

Moody, Chris (2017). *Stop Using word2vec*. URL: https://multithreaded.stitchfix.com/blog/2017/10/18/stop-using-word2vec/ (visited on 03/02/2022).

Müller, Henrik, Gerret von Nordheim, Karin Boczek, Lars Koppers, and Jörg Rahnenführer (2018). "Der Wert der Worte – Wie digitale Methoden helfen, Kommunikations- und Wirtschaftswissenschaft zu verknüpfen". German. In: *Publizistik* 63.4, pp. 557–582. DOI: 10.1007/s11616-018-0461-x.

Müller, Henrik, Jonas Rieger, and Nico Hornig (2021). "We're rolling – Our Uncertainty Perception Indicator (UPI) in Q4 2020: introducing RollingLDA, a New Method for the Measurement of Evolving Economic Narratives". In: *DoCMA Working Paper #6*. DOI: 10.17877/DE290R-21974.

Muth, John F. (1961). "Rational Expectations and the Theory of Price Movements". In: *Econometrica* 29.3, pp. 315–335.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (2014). "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 EMNLP-Conference*. ACL, pp. 1532–1543. DOI: 10.3115/v1/D14-1162.

Rieger, Jonas (2020). "ldaPrototype: A method in R to get a Prototype of multiple Latent Dirichlet Allocations". In: *Journal of Open Source Software* 5.51, p. 2181. DOI: 10.21105/joss.02181.

Rieger, Jonas (2021). *rollinglda: Construct Consistent Time Series from Textual Data*. R package version 0.1.0. DOI: 10.5281/zenodo.5266717. URL: https://github.com/JonasRieger/rollinglda.

Rieger, Jonas, Carsten Jentsch, and Jörg Rahnenführer (2021). "RollingLDA: An Update Algorithm of Latent Dirichlet Allocation to Construct Consistent Time Series from Textual Data". In: *Findings Proceedings of the 2021 EMNLP-Conference*. ACL, pp. 2337–2347. DOI: 10.18653/v1/2021.findings-emnlp.201.

Rieger, Jonas, Carsten Jentsch, and Jörg Rahnenführer (submitted). "LDAPrototype: A Model Selection Algorithm to Improve Reliability of Latent Dirichlet Allocation". In: *Data Mining and Knowledge Discovery.*

Rieger, Jonas, Jörg Rahnenführer, and Carsten Jentsch (2020). "Improving Latent Dirichlet Allocation: On Reliability of the Novel Method LDAPrototype". In: *Natural Language Processing and Information Systems, NLDB 2020*. Vol. 12089. LNCS. Springer, pp. 118–125. DOI: 10.1007/978-3-030-51310-8_11.

Roos, Michael and Matthias Reccius (2021). *Narratives in Economics*. Ruhr Economic Papers 922. DOI: 10.4419/96973068.

Sargent, Thomas J. and Neil Wallace (1975). ""Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule". In: *Journal of Political Economy* 83.2, pp. 241–254. DOI: 10.1086/260321.

Schnabel, Isabel (2021). *New narratives on monetary policy – the spectre of inflation.* URL: https://www.ecb.europa.eu/press/key/date/2021/html/ecb.sp210913~031462fe79.en.html (visited on 03/02/2022).

Shiller, Robert J. (2017). "Narrative Economics". In: *American Economic Review* 107.4, pp. 967–1004. DOI: 10.1257/aer.107.4.967.

Shiller, Robert J. (2020). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events.* Princeton University Press. DOI: 10.1515/9780691212074.

Soroka, Stuart N. (2006). "Good News and Bad News: Asymmetric Responses to Economic Information". In: *The Journal of Politics* 68.2, pp. 372–385. DOI: 10.1111/j.1468-2508.2006.00413.x.

Stryker, Jo Ellen, Ricardo J. Wray, Robert C. Hornik, and Itzik Yanovitzky (2006). "Validation of Database Search Terms for Content Analysis: The Case of Cancer News Coverage". In: *Journalism & Mass Communication Quarterly* 83.2, pp. 413–430. DOI: 10.1177/107769900608300212.

Ter Ellen, Saskia, Vegard H. Larsen, and Leif Anders Thorsrud (2021). "Narrative Monetary Policy Surprises and the Media". In: *Journal of Money, Credit and Banking.* DOI: 10.1111/jmcb.12868.

Tuckett, David, Douglas Holmes, Alice Pearson, and Graeme Chaplin (2020). *Monetary Policy and the Management of Uncertainty: A Narrative Approach.* Bank of England Working Paper 870. DOI: 10.2139/ssrn.3627721.

Tuckett, David and Milena Nikolic (2017). "The role of conviction and narrative in decision-making under radical uncertainty". In: *Theory & Psychology* 27.4, pp. 501–523. DOI: 10.1177/0959354317713158.

Appendix

Figure 5 shows the cosine similarity of all ten topics to earlier or later points in time of the same topic. The month-to-previous-month similarity is plotted in black, month-to-first-month in red, month-to-last-month in blue and the quarter-to-previous-quarter similarity in gray. Topic 9 shows a noticeable change in wording during the period from 2015 to 2020; for topic 3 the short-term similarity decreases with time; topic 10 shows a similar, though not as pronounced, behavior as topic 3, and topic 8 shows a relatively large diachronic change in word usage. Topics 1, 2, 4, 5, 6 and 7 are comparatively stable over the whole observation period.

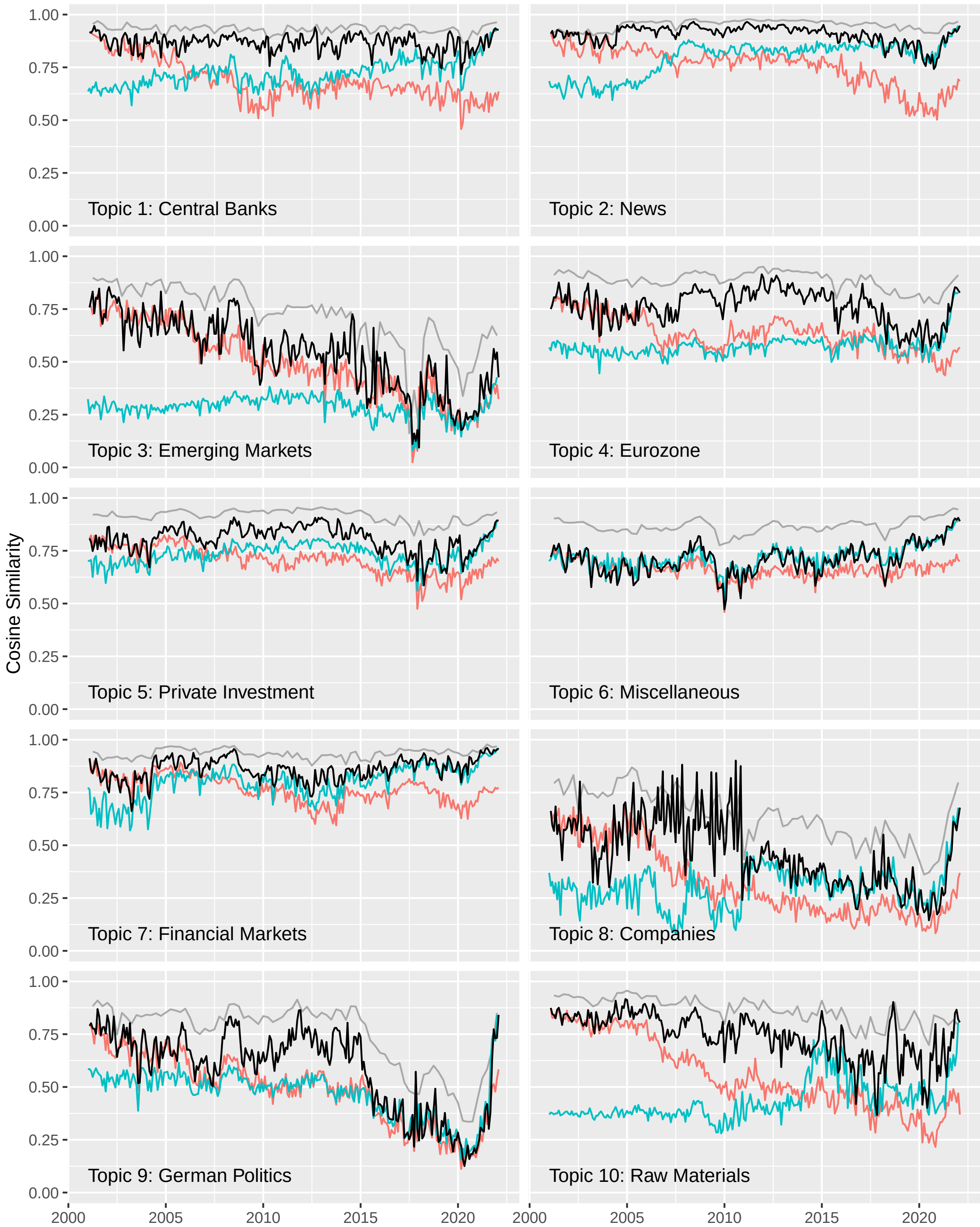

Figure 5: Cosine similarity of topics across different time points. Source: Authors' calculations