

Müller, Henrik; Rieger, Jonas; Hornig, Nico

Working Paper

"We're rolling". Our Uncertainty Perception Indicator (UPI) in Q4 2020: introducing RollingLDA, a new method for the measurement of evolving economic narratives

DoCMA Working Paper, No. 6

Provided in Cooperation with:

Dortmund Center for Data-based Media Analysis (DoCMA), TU Dortmund University

Suggested Citation: Müller, Henrik; Rieger, Jonas; Hornig, Nico (2021) : "We're rolling". Our Uncertainty Perception Indicator (UPI) in Q4 2020: introducing RollingLDA, a new method for the measurement of evolving economic narratives, DoCMA Working Paper, No. 6, TU Dortmund University, Dortmund Center for Data-based Media Analysis (DoCMA), Dortmund, <https://doi.org/10.17877/DE290R-21974>

This Version is available at:

<https://hdl.handle.net/10419/232949>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**DORTMUND CENTER
FOR DATA-BASED
MEDIA ANALYSIS**

DoCMA Working Paper #6

March 2021

“We’re rolling”

Our Uncertainty Perception Indicator (UPI) in Q4 2020: introducing *RollingLDA*,
a New Method for the Measurement of Evolving Economic Narratives

Henrik Müller, Jonas Rieger and Nico Hornig*

Version 1.0, March 2021

Cite as:

Müller, H., Rieger, J. & Hornig, N. (2021). “‘We’re rolling’. Our Uncertainty Perception Indicator (UPI) in Q4 2020: introducing *RollingLDA*, a new Method for the Measurement of Evolving Economic Narratives.” *DoCMA Working Paper #6, Mar 2021*
<http://dx.doi.org/10.17877/DE290R-21974>

*Prof. Dr. Henrik Müller is a professor of Economic Policy Journalism at TU Dortmund University and Dortmund Center for data-based Media Analysis (DoCMA). Jonas Rieger and Nico Hornig are researchers at TU Dortmund University and DoCMA. The authors wish to thank Prof. Dr. Erich Schubert and Prof. Dr. Carsten Jentsch for helpful comments.

Abstract:

In this paper, we present a new dynamic topic modeling method to build stable models and consistent time series. We call this new method *RollingLDA*. It has the potential to overcome several difficulties researchers, who use unsupervised probabilistic topic models, have grappled with: namely the problem of arbitrary selection, which is aggravated when models are to be updated with new sequences of data.

RollingLDA is derived by combining the *LDAPrototype* approach (Rieger, Jentsch and Rahnenführer, 2020) with an implementation that uses preceding LDA results as an initialization for subsequent quarters, while allowing topics to change over time. Squaring dual-process theory, employed in Behavioral Economics (Kahneman, 2011), with the evolving theory of Economic Narratives (Shiller, 2017), *RollingLDA* is applied to the measurement of economic uncertainty.

The new version of our Uncertainty Perception Indicator (UPI), based on a newspaper corpus of 2.8 million German newspaper articles, published between 1 January 2001 and 31 December 2020, proves indeed capable of detecting an *uncertainty narrative*. The narrative, derived from the thorough quantitative-qualitative analysis of a key-topic of our model, can be interpreted as *collective memory* of past uncertainty shocks, their causes and the societal reactions to them. The uncertainty narrative can be seen as a collective intangible cultural asset (Haskel and Westlake, 2017), accumulated in the past, informing the present and potentially the future, as the story is being updated and partly overwritten by new experiences. This concept opens up a fascinating new field for future research.

We would like to encourage researchers to use our data and are happy to share it on request.

Keywords: Uncertainty, Narratives, Latent Dirichlet Allocation, Business Cycles, Covid-19, Text Mining, Computational Methods, Behavioral Economics

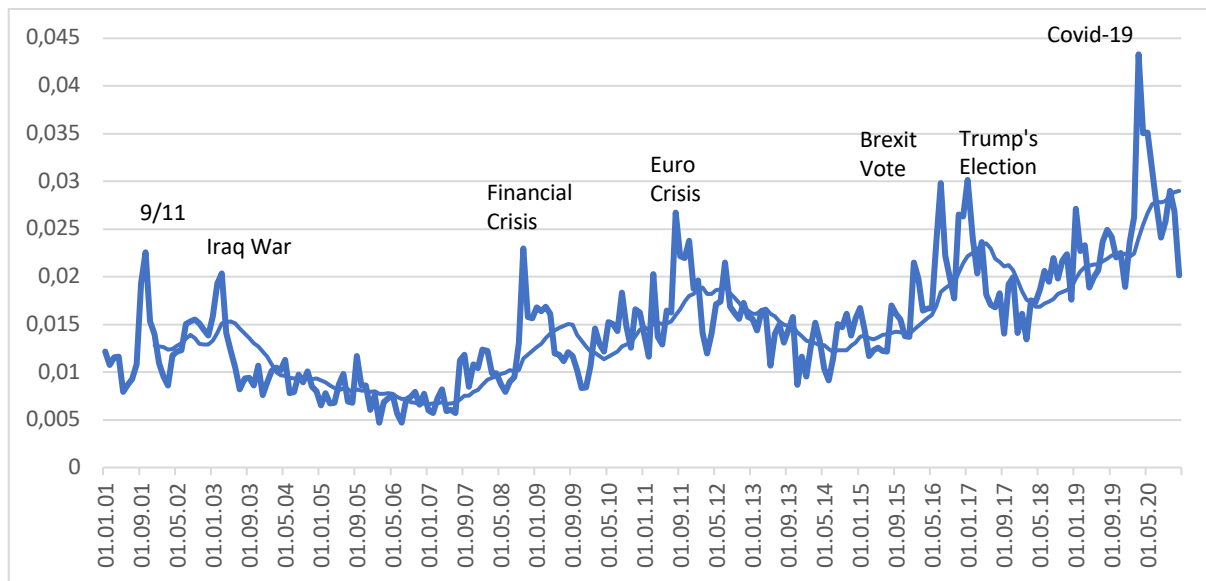
1. Introduction: Size matters

When an uncertainty shock hits, one way of dealing with it, is estimating its size. We may at first have little knowledge of the exact nature of the beast, but at least we can make educated guesses about its strength. Being able to gauge the size of a recent event means we can compare it to past uncertainty shocks and their impacts. When Covid-19 spread around the globe in the Spring of 2020, early on uncertainty indicators showed the enormity of the economic fall-out to be expected. It became obvious that a really big hit was in the making, as Covid-induced uncertainty dwarfed the related values in the wake of earlier shocks. Governments (and central banks) launched unprecedented stimulus programs, while company treasurers hoarded all the liquidity available and private households raised their savings rates dramatically. It is plausible that the immediate reactions of 2020 were informed by experiences made during the financial crisis; without the memories of 2008/09, and governments' willingness to step in, the impact of the Covid-shock might have been even graver.

Yet, not all shocks are created equal. As we argued in Müller and Hornig (2020a), three types of uncertainty can be distinguished: *market-based uncertainty* originates in the economy itself and is the result of our, at times, insufficient understanding of market mechanisms; *economic policy uncertainty* leads to economic consequence of developments in the realms of politics; *truly-exogenous uncertainty* originates outside of both politics and the economy.

Among the seven major peaks our news-based *Uncertainty Perception Indicator* (UPI) produces over the past two decades (fig. 1), all three types of uncertainty are detectable. The first two local maxima are (geo)political in nature: 9/11 (coinciding with the fallout from the bursting of the dot.com bubble, a market-based event preceding the period covered by our data) and the Iraq war. The third is associated with the financial crisis, i.e. market-based. The fourth peak is triggered by the financial stress surrounding the Euro area in 2011, which can be considered market-based as well as economic policy-related, given Eurozone member states' inability to swiftly enhance the Euro's institutional framework. Clearly economic policy-related shocks occurred in the second half of 2016, with the Brexit vote and Donald Trump's election, events that impacted the economy indirectly and structurally through their corrosive effects on international trade. The seventh peak, and the most severe one by far, is the Covid-19 shock.

Figure 1: Seven Peaks – the overall Uncertainty Perception Indicator (UPI) for Germany Q1 2001 – Q4 2020*



*Analysis corpus relative to entire corpus, monthly data and backward-looking twelve-month moving average, for methodical further details see sec. 3.

Three provisional observations should be noted:

First, outbreaks of uncertainty seem to come in waves. A major event is typically being followed by subsequent (smaller) ones. The two geopolitical shocks of the early 2000s were related to each other, the US-led war in Iraq being a direct consequence of the terror attacks of 9/11. The financial crisis, in turn, triggered the Euro crisis; the peak of Euro-related uncertainty in July 2011 was followed by subsequent repercussions resulting in local maxima in November 2011 and July 2012. The political shocks of 2016 were related to each other in so far as they were the most prominent manifestations of the simmering populist turn in Western democracies. Major uncertainty shocks seem to be cumulations of underlying tensions in the political economy. Once laid bare, these tensions tend to unload in a series of additional eruptions.

Second, after an uncertainty shock has hit, a phase of relative calm sets in. Once the geopolitical shocks of 2001/03 were digested, several years of low perceived uncertainty followed, an era dubbed “the great moderation”. Also, after the years of financial market turmoil several years of stability ensued. Even after Trump took office and the Brexit talks between the UK and the EU stalled, a rather relaxed mood prevailed for a while. Why? We’ve come up with three preliminary explanations: a) the initial uncertainty shock disappears as societies learn more about the respective phenomenon and adapt to it (we’ll get back to this theme in sec. 2); b) the media toned down their coverage of current events deliberately, since they felt that the public – and maybe journalists themselves – got tired of reporting in constant panic mode; c) the media did not detect simmering problems or did not take them seriously. (While b) could be dismissed given the properties of news as an economic product (Müller and Hornig, 2020b), c) would constitute a failure of journalism to act as an early warning system.)

Third, uncertainty perception follows a rising trend. Since 2008 the overall volume of uncertainty-related reporting has increased. This rise has come in several steps, with each major peak leaving behind subsequently higher levels of background noise, as the 12-month moving average graph in fig. 1 shows. Again, there are two preliminary explanations for this phenomenon: a) the world has indeed become more unstable, associated with the volatility of financial market-driven capitalism and the decline of US-led hegemonic stability, both epitomized in the US-centered financial crisis of 2008; b) the media may have taken to framing all kinds of developments in terms of attention-grabbing uncertainty, in this case, the long-term rise would be the manifestation of a just another media fad. (We tend to embrace a) and dismiss b), since the traditional quality media in our corpus bank on a reputation of level-headedness.)

These ad-hoc observations are intended to provide some food for thought at the outset of our quest. Yet, sophisticated text-mining techniques allow us to dig deeper into the data. That's the purpose of this paper: we explore the role of narratives in the perception of uncertainty (sec. 2) and we introduce a new dynamic topic modeling method, *RollingLDA*, that allows us to build consistent time-series of evolving media narratives (sec. 3). The results are discussed in sec. 4, while sec. 5 draws some conclusions and highlights possible further directions of research.

2. Uncertainty Narratives

Humans have the awe-inspiring ability to adapt to all kinds of changes. After all, it is this trait that enabled our ancestors to inhabit the entire planet, bar Antarctica. Coming out of Africa, mankind was able to survive and strive in very different climate zones and landscapes, flora and fauna, through ice ages and warm periods, earthquakes, droughts and floods, wars, revolutions and epidemics.

When the unexpected hits, humans have a three-step strategy to deal with the unknown: *halt, cooperate, reorganize*. The first reaction to radically changing circumstances is to just stop. Encountering some unidentifiable threat, we refrain from going forward, literally halting in our step. If we are sure that we are not mistaken and the threat has not vanished, we turn to the people in our immediate community and try to cope through ad-hoc cooperation; as social beings we instantaneously sense that our chances to survive increase if we are acting collectively. Threats that prove to be persistent call for organizational changes in our community, e.g. the diffusion of new social norms, an altered division of labor, now including specialists to manage a particular threat, the installation of new institutions, or the overhaul of existing ones, and so forth. When the dust settles, humans adapt and take the new environment for granted. What at first appeared to be outrageously uncertain, becomes predictable over time, with the associated detriments fading. As societies acquire knowledge about the circumstances they are in, uncertainty gradually diminishes to manageable risk, until the next shock rattles them.

The impressive human track-record through the ages could provide us with some confidence whenever an uncertainty crisis occurs. Yet, the basic behavioral pattern has changed very little: initially, we react by halting in our step; we stop whatever we've been doing and try to make sense of the new environment before we proceed. Large uncertainty shocks induce sudden stops. That's why they are of immense interest to economists. Investment, consumption, hiring or lending decisions are being postponed, thereby hitting the economy

as a whole, possibly altering the future path of expansion (Bloom, 2009; Bloom, Bond and Van Reenen, 2007). What matters to the economy, though, is not just the initial threat itself, but its *perception*, which is what our Uncertainty Perception Indicator (UPI) strives to measure. Its aim is twofold: gauging the sheer *size of a shock* (see introduction) and decoding the underlying *uncertainty narrative*.

Cognitive shortcuts

The concept of the narrative is central to the analysis of uncertainty since it is the cognitive mode in which humans process unknown circumstances. “Fast thinking”, to quote Daniel Kahneman (2011), tends to be anchored in stories. As *dual-process theory* holds, stories serve as cognitive short-cuts that enable humans to react swiftly to new challenges by harking back to the past. Known patterns are applied to current unknowns to navigate complex environments. Narratives sort past events into plausible causal relationships, categorize groups of protagonists and attach simple binary judgements of the good/bad-friend/foe type to them.

With stories in mind, we are able to take rapid, potentially life-saving decisions. Even without precise and comprehensive knowledge of the circumstances reasonable reactions are possible. After all, this is exactly what the term “uncertainty” implies: a lack of knowledge to the effect that we are unable to make reliable forecasts about the immediate future. In such a situation, a story we have in mind may be our best shot at making sense of our environment. Narratives are central to processing uncertainty shocks.

Social dimensions of Stories

Narratives that are shared by large groups of people have a social component: they facilitate a common perception of the state of the world in complexity-reduced form, possibly even providing groups of people with a sense of collective purpose, i.e. enabling them to cooperate through the accumulation of social capital (Putnam, 1993) and providing the social glue needed to cooperate in times of danger. Some narratives may also inform people of (in)appropriate reactions to challenging circumstances, thereby supporting the spreading of social norms. The perception of economic uncertainty as a social phenomenon is transmitted through shared stories.

The spreading of these stories can be likened to the dispersion of infectious diseases (Shiller 2017, p.17): a particular narrative is not ubiquitously present in an instant, but it needs time to take hold, before gradually petering out and being overlaid by subsequent stories. However, a once established narrative can be reactivated by new events; the telling of the tale resumes. There is a connection here to the Issue Attention Cycles studied in communication science (Downs, 1972; Miltner and Waldherr 2013): when a cycle is over, it leaves behind a slightly elevated level of news activity (phase 5 in Downs’ taxonomy), a background noise that can be interpreted as collective memory, to which media refer whenever events of apparent resemblance strike. Memories tend to fade over time, and they may be altered by new experiences – humans are prone to hindsight bias. Therefore, when we strive to measure the impact of current events on public perception, these alterations need to be considered.

Measuring narratives

There are plenty of narratives floating around in public spheres, many unrelated to economic uncertainty. How can researchers actually capture these narratives? In liberal societies the prevailing narratives are not preformulated by top-down authoritarian propaganda, but are formed by the interplay of many, often contradictory, voices in vast media spheres connected by a host of channels. Even though social media have opened up the market place for narratives, traditional journalistic media, e.g. newspapers and their digital offspring, still tend to reflect properly what's going on in the public sphere due to interlinkages between traditional and newish media (von Nordheim, 2019; von Nordheim et al. 2018a, b). For all their flaws and biases, major newspapers are still suitable points of measurement when one wants to know what's driving the general public (Müller and Hornig, 2020b, pp. 7-9).

In this paper, as in three preceding ones (Müller and Hornig, 2020 a, b; Müller, Hornig and Rieger, 2021), we focus on measuring uncertainty by isolating patterns in media coverage of economic uncertainty. Inspired by the ground-breaking work of Baker, Bloom and Davis (2016) that introduced the Economic Policy Uncertainty Index (EPU), our UPI is based on a content analysis of major (German) newspapers. But while the EPU is focused solely on the *size* of uncertainty shocks, our approach also seeks to unveil *uncertainty narratives* by applying a topic model-based thematic deconstruction of uncertainty-related news. The main contribution of this paper is a new method, a dynamic version of Latent Dirichlet Allocation (LDA) we call *RollingLDA* (see next section), intended to improve the modeling of the properties of narratives.

Still, what do we actually expect to find? As proposed earlier (Müller et al., 2018, p. 558), the concept of the *media narrative* is related to that of the *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. We have augmented this approach by adding two more elements. According to our definition, 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 and may change over time; and f) events, that are chronologically integrated and that are (often) assumed to constitute causal relationships. To put it metaphorically: a frame is to a narrative what a still photo is to a movie.

LDA in its original form 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. Over longer time-horizons, however, the correspondence between research object and method is less obvious. After all, what we are interested in is detecting the *evolution* of thematic trends. LDA, in contrast, assumes the structural stability of topics over time, which collides with the more fluid nature of narratives, that may change over time, and may even be altered in hindsight. In the next section we present a new dynamic version of LDA that allows topic-structures to change over time by modeling *fading collective memory*, as newer versions of stories overwrite older ones. New data are fitted to a topic model that is calculated based on a rolling window of past observations. The past informs the human perception of the present, but older memories sink into oblivion.

In earlier attempts we showed that recalculating the entire UPI model at each observation point, e.g. each quarter, does not produce consistent time-series; the resulting models bore

some resemblance, but were hardly comparable. Applying strategies such as “zooming” (i.e. variations of the number of topics, parameter K of the LDA) and “prototyping” (see sec. 3) didn’t lead to the desired stability over time either. Holding the topic structure constant and simply fitting new data to the old model (Müller, Hornig and Rieger, 2021) may produce initially plausible results. Over time, though, the build-in inflexibility collides with reality, i.e. the models’ topic structure doesn’t fit the ever-changing *real* narrative structure anymore. In contrast, *RollingLDA* allows for the evolution of topics, that can be interpreted as narratives. Put metaphorically, this is not a series of snapshots anymore, but a proper motion picture. *We are rolling*, as a director on a film set would announce.

3. Towards consistent time-series derived from topic models

Data

The UPI 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 31 December 2020 are considered. The entire corpus has a size of 2.8 million texts. In a first step, the corpus 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 is produced by applying a rather open query (Müller and Hornig, 2020a):

„unsicherheit“ OR „unsicher“ OR „unsicherheiten“

AND „wirtschaftlich“ OR „wirtschaft“

The subsequent analysis corpus has a size of 37,067 articles.

Topic Modeling: selection and stabilization

As noted above, we make use of the topic modeling method Latent Dirichlet Allocation, albeit in a modified way. The original LDA method (Blei et al., 2003) has the far-reaching disadvantage, that the random initialization of the Gibbs sampler (Griffiths and Steyvers, 2004) can result in fundamentally different models, when run several times on the same data with identical parameter sets. This property of probabilistic topic models has been a nuisance to researchers and has limited the scientific reproducibility of results.

To overcome the problem of arbitrary selection, different strategies have been suggested, all of whom have their shortcomings.¹ For the UPI model we use the selection method *LDAPrototype* (Rieger, Koppers, Jentsch and Rahnenführer, 2020). At several stages of the

¹ Nguyen et al. (2014) have proposed *average LDA* runs. Following this approach, however, no unique assignments for individual tokens of the documents can be obtained. A different method of objective and automated selection is *perplexity optimization* (Grün and Hornik, 2011). However, Chang et al. (2009) were able to show that selection mechanisms aiming at optimizing likelihood-based measures do not correspond to the human perception of a well-adapted model of text data. In fact, the authors observe a negative correlation and propose a so-called intruder procedure based on human coding to obtain results that correlate most to human perceptions. The corresponding methodology is implemented in the software package *tosca* (Koppers et al., 2020), but this method requires large human resources and lacks automation capabilities.

process, a prototypical LDA is being chosen from a set of LDA models. The method solves the problem of arbitrary selection and thus improves the reliability of findings (Rieger, Rahnenführer and Jentsch, 2020). Prototyping is achieved following a typical statistical approach: for a given combination of parameters a number of models is calculated (usually about 100), from which the particular model is determined that is most similar to all the other models in the set. For this purpose, pairwise model similarities are calculated using the S-CLOP measure (Similarity of Multiple Sets by Clustering with Local Pruning). These similarities are determined measuring the deviation from strictly matched topics in the resulting local clusters, which are created based on a hierarchical clustering result of topics using pairwise topic similarities of two LDA results considered. These deviations are computed for all possible pairs of LDA models. 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).

From Seeding to Rolling

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”, an approach we already utilized for the UPI in Müller, Hornig and Rieger (2021).

Here, we refine the initialization approach by implementing it on a rolling basis. 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 10,806 texts with an average of 251 tokens from a 44,942 words-vocabulary for the first modeling period. These texts from the first five-year-period are modeled using the *ldaPrototype* procedure as described. In earlier versions of the UPI (Müller and Hornig, 2020b; Müller, Hornig and Rieger, 2021) we identified a number of $K = 14$ topics as the most appropriate for our purpose. Accordingly, here we model $K = 14$ topics and choose as Dirichlet parameters $\alpha = \eta = 1/K$ as common, while the Gibbs sampler iterates 200 times over the dataset.

In a *second modeling step* we consider the articles from the subsequent first quarter of 2006, i.e. the 303 articles published between 1 January and 31 March 2006. By applying the “seeding” procedure described above, we model the topic assignments to these 303 articles. However, we only use the last three quarters as *memory*, i.e. we initialize the model with the 1222 articles from April to December 2005. The vocabulary is extended by words that occur more than five times in the new 303 articles and that were not included in the vocabulary before. Employing this procedure, we add 73 words in the first quarter of 2006. The topic assignments of the new articles are initialized randomly and the Gibbs sampler iterates over each of the new articles again 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 quarter of 2006 on a rolling basis for all subsequent quarters, so that we finally obtain assignments to the 14 topics for the entire analysis corpus with an average of 301 tokens over a vocabulary of 52,771 different words. Modeling of newly occurring articles, for example from the first quarter of 2021, can then be performed analogously. We refer to this approach of dynamic modeling as *RollingLDA*.

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

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.² This measure is useful for assessing whether a topic remains sufficiently stable over time to be treated as consistent. In addition, the measure can be used to identify events. For this, the aggregation of word frequencies over a smaller unit of time, such as months, is useful. The quarter-to-quarter similarity in our K=14 model is rather high and stable for all the topics (see appendix). All described steps of preprocessing, modeling and evaluation can be found as R codes in the GitHub repository <https://github.com/JonasRieger/upi>.

² In a first analysis, the use of other common measures in the context of topic similarities such as Jaccard coefficient (Jaccard, 1912), Jensen-Shannon divergence (Lin, 1991), or rank biased overlap (Webber, 2010) has been shown to be less useful. More detailed studies to assess the variation of a topic over time are outside the scope of this paper, but will be addressed in further research by DoCMA.

4. Results: of fear, shocks and awe

The refined method to calculate the UPI yields most of the topics already found in earlier versions, although in slightly altered ways. Table 1 provides an overview of the model.

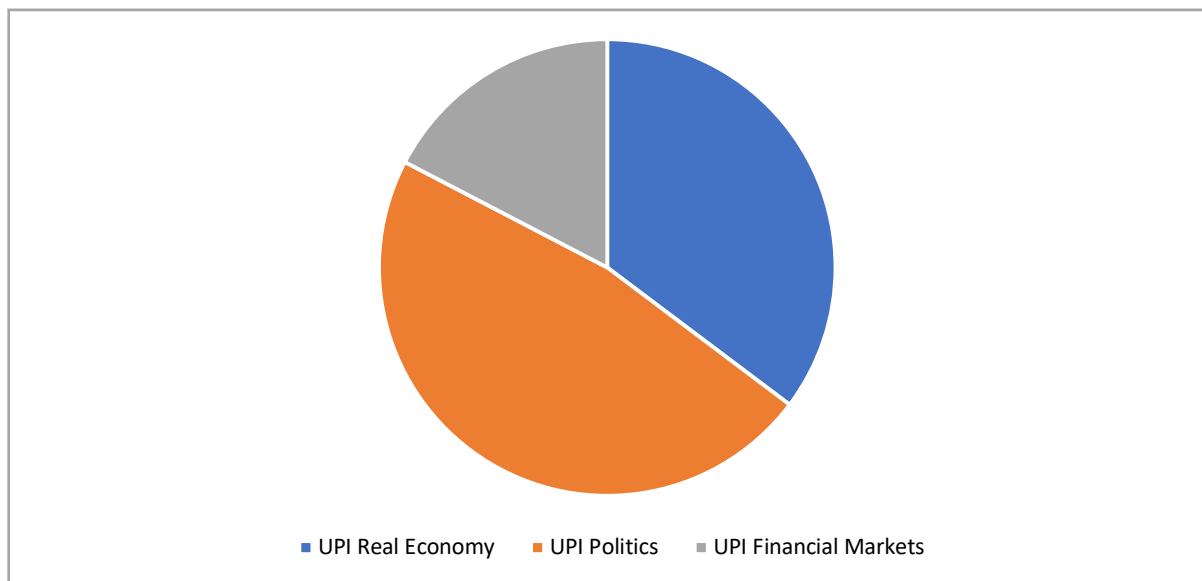
Table 1: Overview of Topics and Labels (RollingLDA, prototyped, Q4-2020, K=14)

| Topic No. | Label | Share in analysis corpus (per cent) | Content | Part of Uncertainty Factor... |
|-----------|------------------------------------|-------------------------------------|---|-------------------------------|
| 1 | Corporate Culture | 6.6 | Trust, Technology, Entrepreneurship, Digitalization, Knowledge, Workplace, Career | UPI Real Economy |
| 2 | EU Conflicts | 5.4 | Brexit, Greece debt, democratic values, Russia, Turkey etc. | UPI Politics |
| 3 | Energy & Climate Change Mitigation | 4.5 | Energy market developments, transition to sustainables etc., Fukushima disaster (2011) as focal event | UPI Real Economy |
| 4 | Companies & Markets | 7.3 | Developments at quoted international corporates | UPI Real Economy |
| 5 | Geopolitics | 6 | Geopolitical tensions | UPI Politics |
| 6 | Society | 11.3 | Debates on Capitalism, Globalization, Democracy, Populism, Immigration, national identity | UPI Politics |
| 7 | Financial Markets I | 6.1 | Trouble concerning financial institutions (banks, insurance...), retail investor aspects | UPI Financial Markets |
| 8 | German Politics I | 7 | Structural Reforms, Labor Markets, Welfare State | UPI Politics |
| 9 | Miscellaneous | 10.1 | Diverse | – |
| 10 | Financial Markets II | 9.5 | Retail Investor perspective | UPI Financial Markets |
| 11 | Leisure & Hospitality | 4.4 | Entertainment, arts, sports, travel, Corona-related peak | UPI Real Economy |
| 12 | Central banks | 7.3 | ECB, Fed etc. actions against crises | UPI Politics |
| 13 | German Economy | 8.9 | Business cycle developments, forecasts, surveys | UPI Real Economy |
| 14 | German Politics II | 5.6 | Parties and governments | UPI Politics |

We combine 7 and 10, 8 and 14, due to their thematic proximity.

90 per cent of the corpus could be interpreted in terms of our research interests. As in earlier exercises, we sort the model's topics according the three *Uncertainty Factors*: politics, real economy and financial markets. Roughly half of the articles in the analysis corpus were identified as having politics-related uncertainty as their primary subject, in line with journalism's overall focus on political developments. About a third of the corpus deals with uncertainty in the context of businesses and markets, while a sixth is financial markets-centered (fig 2).

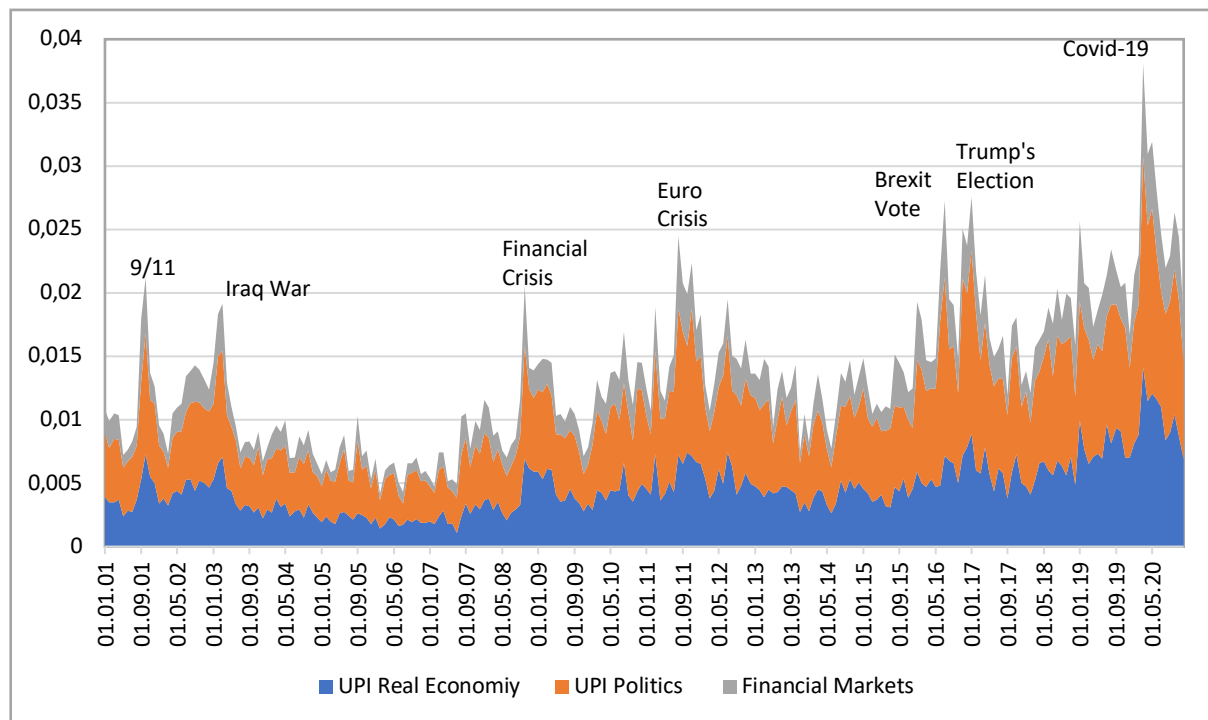
Figure 2: UPI by Uncertainty Factors – shares of analysis corpus



The categorization of individual topics was motivated by the actors detectable in the topics. Since narratives are built around protagonists, we let institutions and persons guide our categorization efforts. Still, some differences to earlier models are noteworthy.

For instance, topic 1 (“Corporate Culture”) bears some resemblance with “Human Resources”, a topic found in Müller and Hornig (2020b), but also with “Society” in the present version. While the latter topic deals with general political debates, “Corporate Culture” is about the actions – and reactions – of companies. That’s why we grouped them into different Uncertainty Factors. Another border-line example is “Central Banks”, that was sorted into UPI Politics because central banks are the main actors in this topic. However, many of the articles in this cluster deal with the consequences of central bank actions for financial markets. One could argue that central banks are *re-actors* who set their policies according to developments in financial markets. Thus, with some justification “Central Banks” could have been sorted into UPI Financial Markets as well. We chose not to, for good reasons we think, but there’s a degree of arbitration, implying that the frequencies of the UPI’s Uncertainty Factors should be interpreted with caution. Fig. 3 allows a first glimpse at the underlying uncertainty scenery. (Contrary to earlier papers, where we showed frequencies relative to the analysis corpus, here we depict individual topics and Uncertainty Factors relative to the entire corpus.)

Figure 3: UPI by Uncertainty Factors, frequency over time*



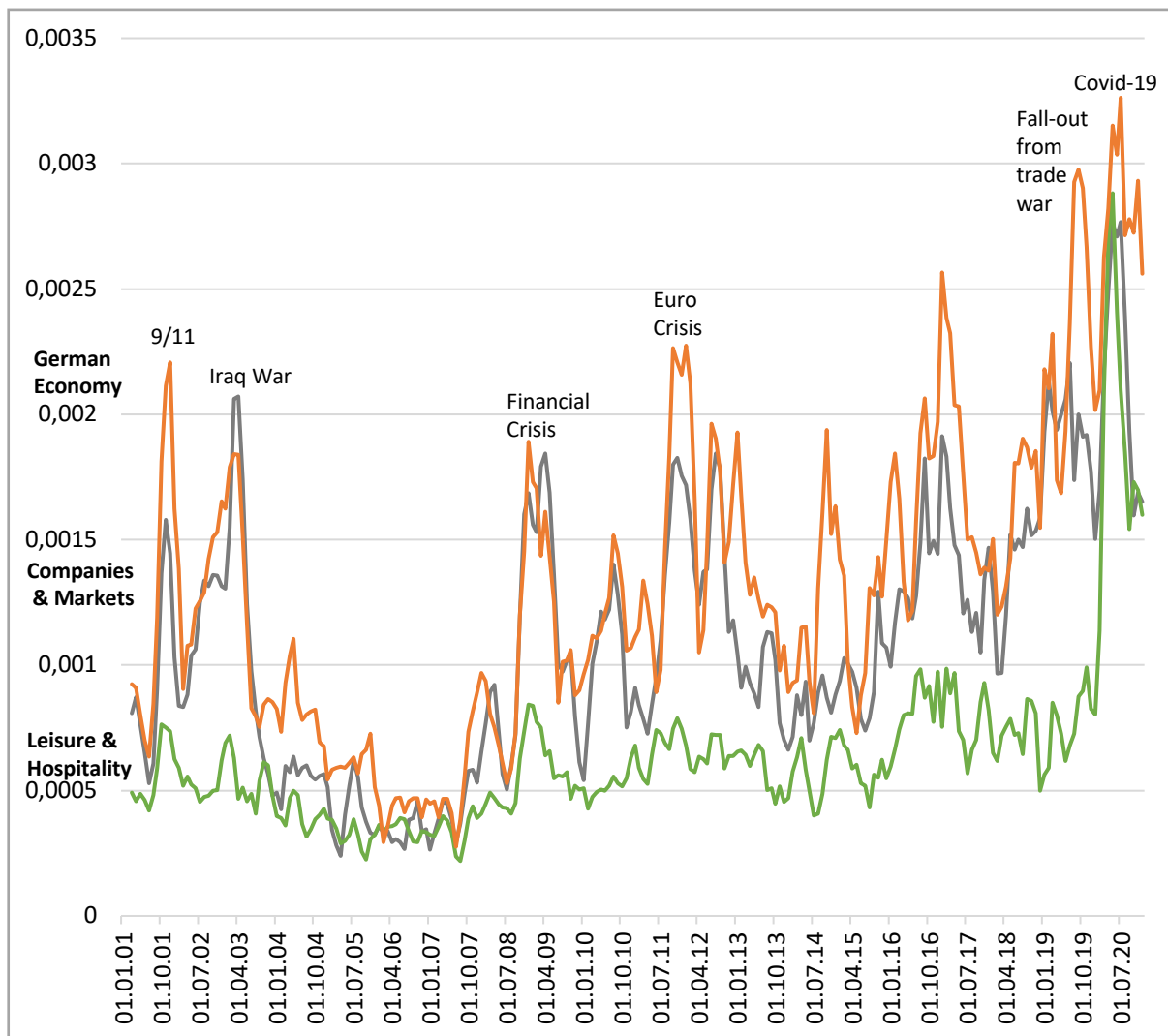
*shares of corpus

Compared to earlier versions of the UPI, a high degree of synchronization is visible: spikes around the *seven peaks* (see introduction) are driven by all three Uncertainty Factors, though individual topics differ considerably in terms of their frequency (see below). Note that the secular rise of uncertainty since the middle of the 2010s is associated with a more prominent presence of real economy issues.

UPI Real Economy

Five out of 13 interpretable topics of our model are sorted into the real economy category. We find the usual “German Economy” topic that covers uncertainty with respect to movements in the business cycle. The articles in this topic are mostly triggered by releases of new data and forecasts. Developments in individual “Companies & Markets” move pretty much in synch with reporting on the economy as a whole. The trade war, kicked off in 2018 by the Trump administration, leaves a mark on Germany’s large manufacturing sector that was already stagnating when tariffs and embargoes caused additional headaches (fig. 4). A topic that was non-existent in earlier UPI models has been labeled “Leisure & Hospitality”, a category in which reporting on the business of entertainment, sports, arts, travel, and tourism, is combined, including also some education issues, the common thread being the personal contact of larger groups of people. The topic’s frequency moves in line with the other business cycle-sensitive topics, although with less pronounced peaks. The Covid restrictions of 2020 sent this topic to an alltime high.

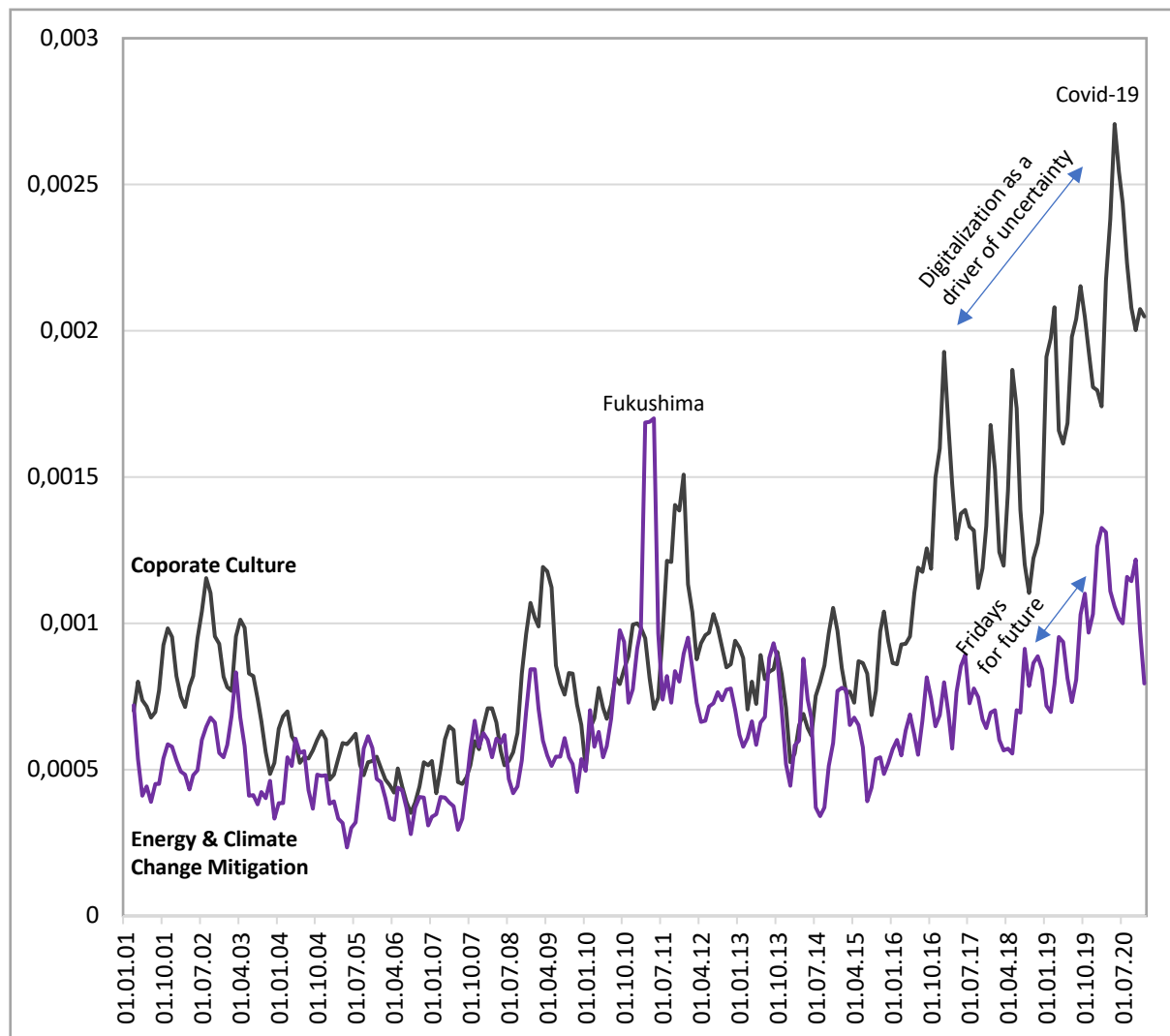
Figure 4: UPI Real Economy, business cycle-sensitive topics*



*shares of corpus, backward-looking three-month moving average

The UPI Real Economy also includes structural topics, that are somewhat detached from business cycle developments. “Corporate Culture” is characterized by a series of peaks, some of them triggered by corporate scandals coming to light in the aftermath of financial excesses. Since the mid-2010s the topic is on a secular rise, driven by the challenges posed by digitalization and (foreign) digital giants such as Google, Apple or Amazon, the final peak being triggered by the Corona pandemic. “Energy & Climate Change Mitigation” is a familiar uncertainty-related topic; as in earlier versions of the UPI model, the disaster at the nuclear plant at Fukushima in 2011 stands out as this topic’s central event. The Fridays for Future protests cause an increase towards the end of the decade. Contrary to earlier versions of the model, the fallout from Covid-19 does not have a particularly influential part.

Figure 5: UPI Real Economy, structural topics*

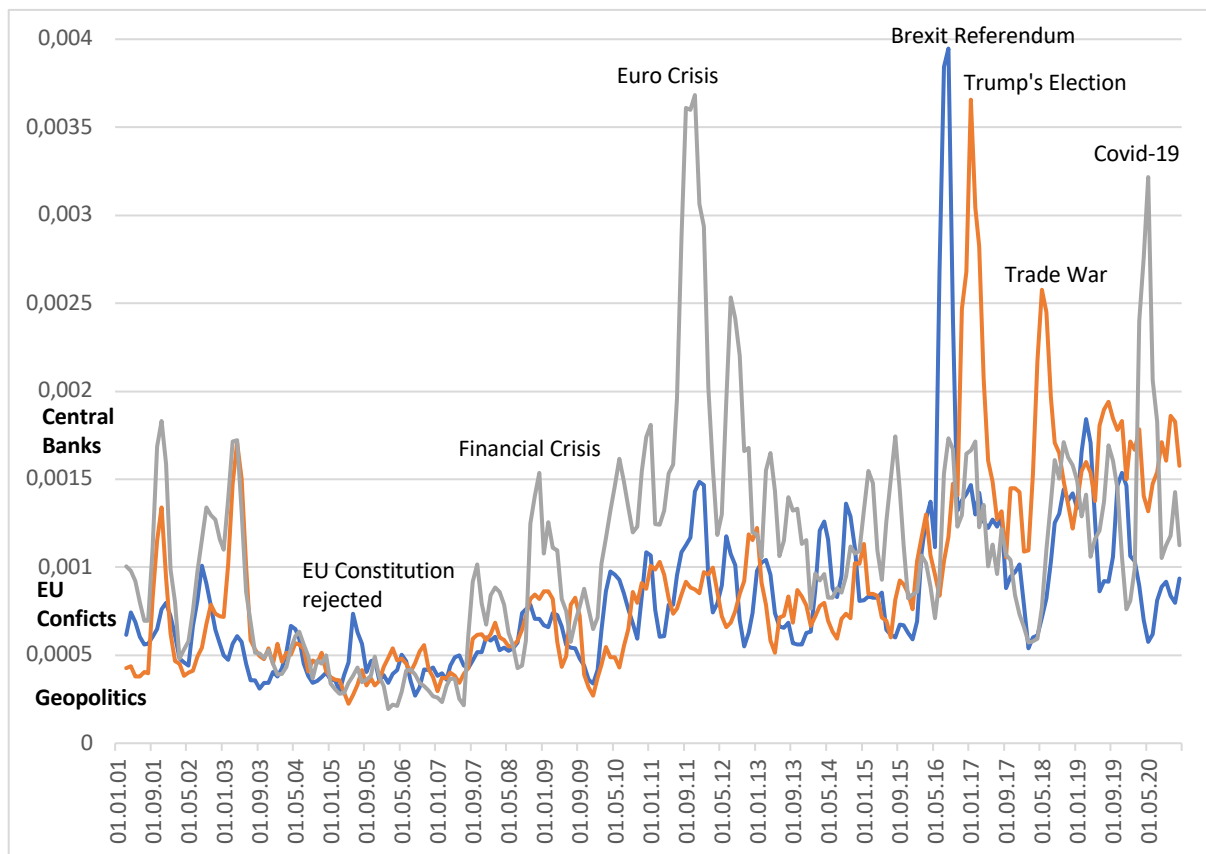


*shares of corpus, backward-looking three-month moving average

UPI Politics

The main factors contributing to political uncertainty over the period considered are international and European developments. Again, after 2008 the swings in uncertainty perception have become more pronounced, with the Euro crisis, the Brexit referendum, the election of Donald Trump to the US presidency, the trade war and the Covid-19 as major events driving the dynamics of the three topics. We subsumed “Central Banks” under European topics, since its main actor is the ECB. Unconventional, and therefore initially hardly predictable, reactions by the central bank to detrimental developments cause short-term additional uncertainty. This effect is most pronounced during the Euro crisis and the pandemic, when the ECB was the first line of defense, stabilizing the financial sector and the real economy. The model distinguishes precisely between conflicts within the EU and in the wider realm of geopolitics (fig. 6).

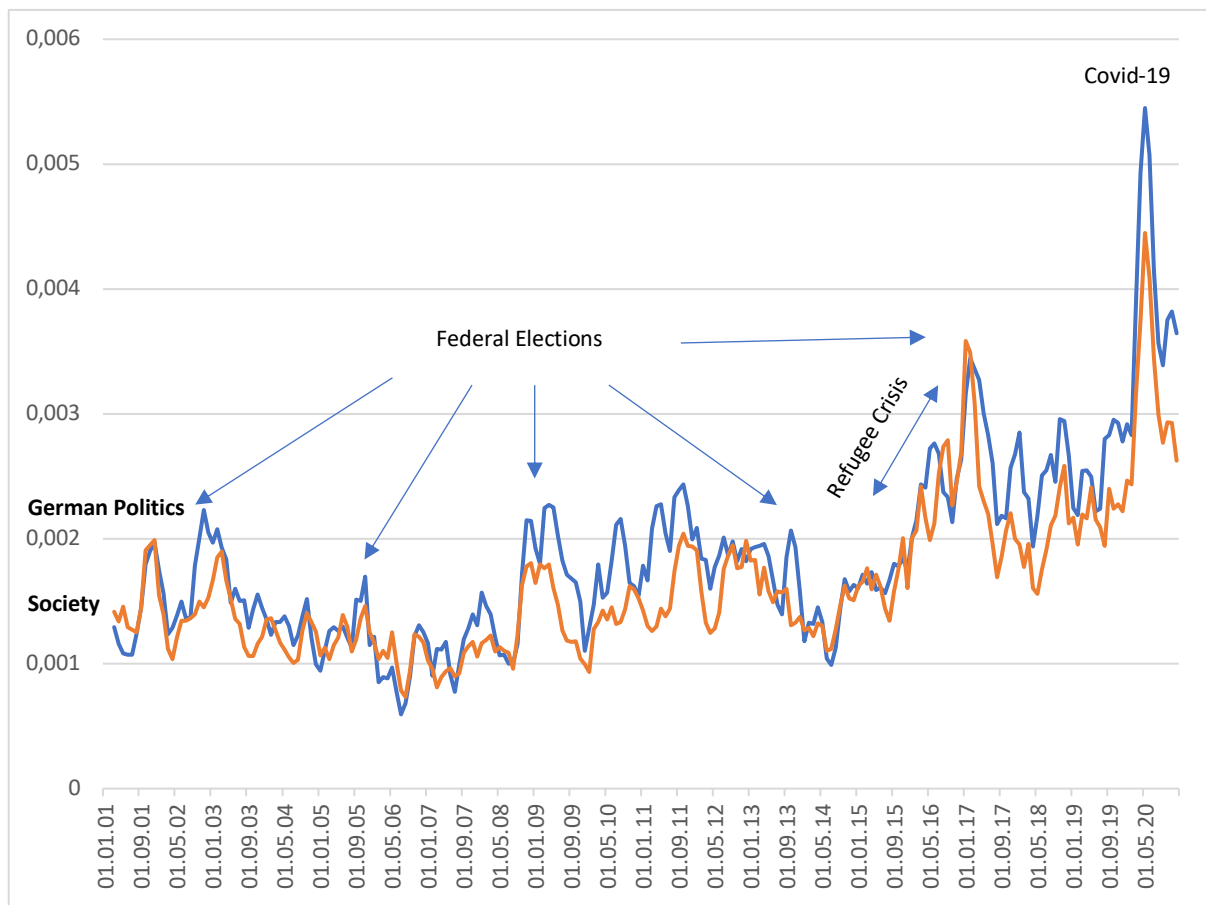
Figure 6: UPI Politics – International and European Topics



*Backward-looking three-month moving average

Political uncertainty at the domestic level is captured by “German Politics” and “Society”. These topics move closely synchronized. As in earlier models, a longer period of relative calm is followed by a considerable rise of both topics starting in the mid-2010s. The refugee crisis, in combination with the populist surge in other countries as well as at home, culminated in a rapid escalation in the run-up to the federal elections of 2017. Later, the pandemic naturally pushed the national fever chart to unprecedented levels (fig. 7). While “German Politics” contains stories about traditional political actors, such as governments and political parties, “Society” gets closer to the concept of the narrative. Dealing with questions of identity and ideology, it allows a glimpse of the nation’s prevailing ideas, images and ailments.

Figure 7: UPI Politics – National Topics

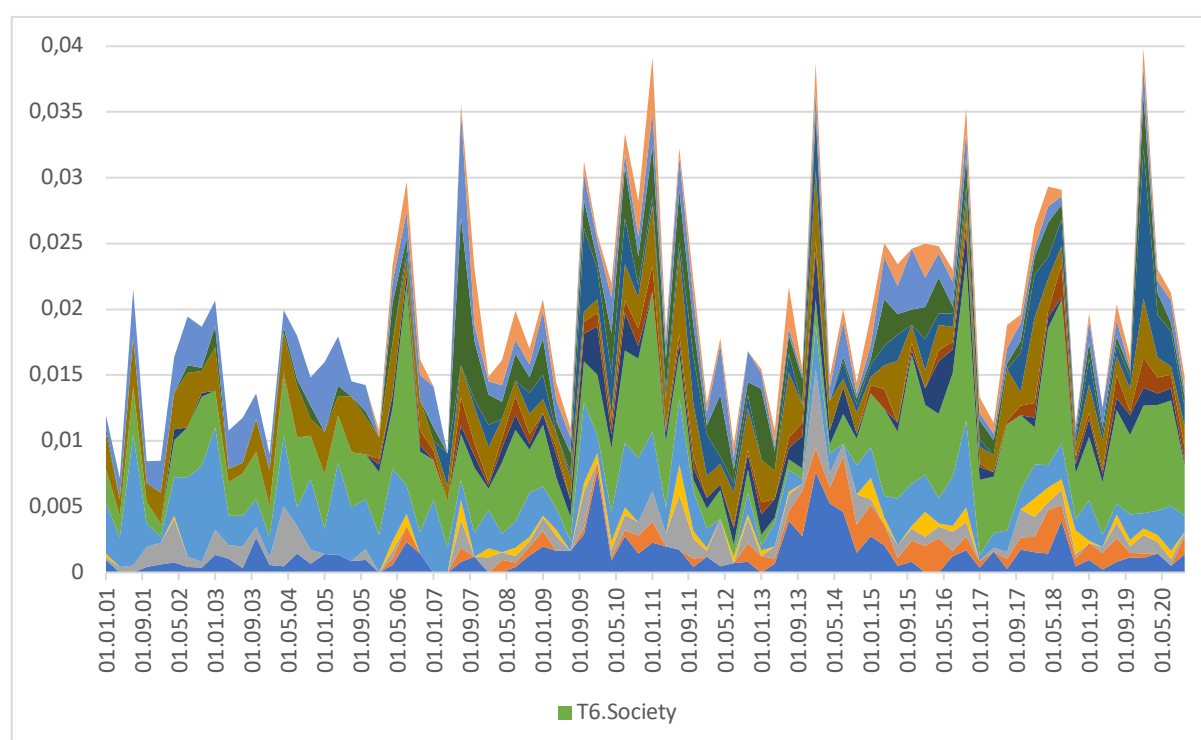


*Backward-looking three-month moving average

An Uncertainty Narrative

To capture emotional aspects of the content, we looked for mentions of the charged word “Angst”, an exaggeration of uncertainty. Fig. 8 shows the results for our model’s topics: this *fear gauge* can be interpreted as a measure of public anxiety. It is most pronounced in the topic “Society” (green segments).

Figure 8: Fear Gauge – Salience of the “Angst” in UPI topics*



*share of articles in UPI topics comprising the word “angst”

“Society” deals with the self-images Germans harbor: how they think they are doing, what they believe in, what they worry about and how they consider to alleviate their collective problems. In terms of issues, the topic deals with globalization, diversity, equity, justice and freedom. It is not just about economic uncertainty, but about the collective insecurity uncertainty causes. Since we strive to isolate a national *uncertainty narrative*, this appears to be a promising area to search.

We get back to our definition from sec. 2, where we proposed that a *media narrative* contains the following elements:

- a) a problem definition,
- b) a problem diagnosis,
- c) a moral judgement,
- d) possible remedies
- e) protagonists
- f) a series of events

Events are associated with peaks in the frequency curve. *Protagonists* can be derived from the lists of top words LDA produces. Here, we use top words that are calculated both for the topic over the entire time-span and for the topic’s quarter-by-quarter manifestations. The first four elements (Entman’s features of a media frame) can be obtained from a qualitative analysis of the top words and top articles LDA produces.

The topic’s (translated) top words read:

“humans, world, society, politics, today, being, Germany, freedom, Europe, social, globalization, democracy, state, future, asking, knowing”

Quarter-by-quarter comparisons deviate somewhat from the topic's mean top words. For instance, in the wake of the financial crisis terms like "capitalism", "market economy", "trust" and "responsibility" rank high. In 2015/16, the term "refugee" ("Flüchtling") is among the most often used words.

(Translated) headlines of top articles read:

The utter entanglements of the Climate worriers

Süddeutsche Zeitung, 4 Dec 2009

Long live liberalism!

Süddeutsche Zeitung, 1 Nov 2018

The AfD will probably further gain support

Süddeutsche Zeitung, 16 Mar 2018

Globalization as a scapegoat

Handelsblatt, 16 Mar 2018

The End of the Middle Class

Süddeutsche Zeitung, 2 Nov 2018

The unsatisfiable desire for Security

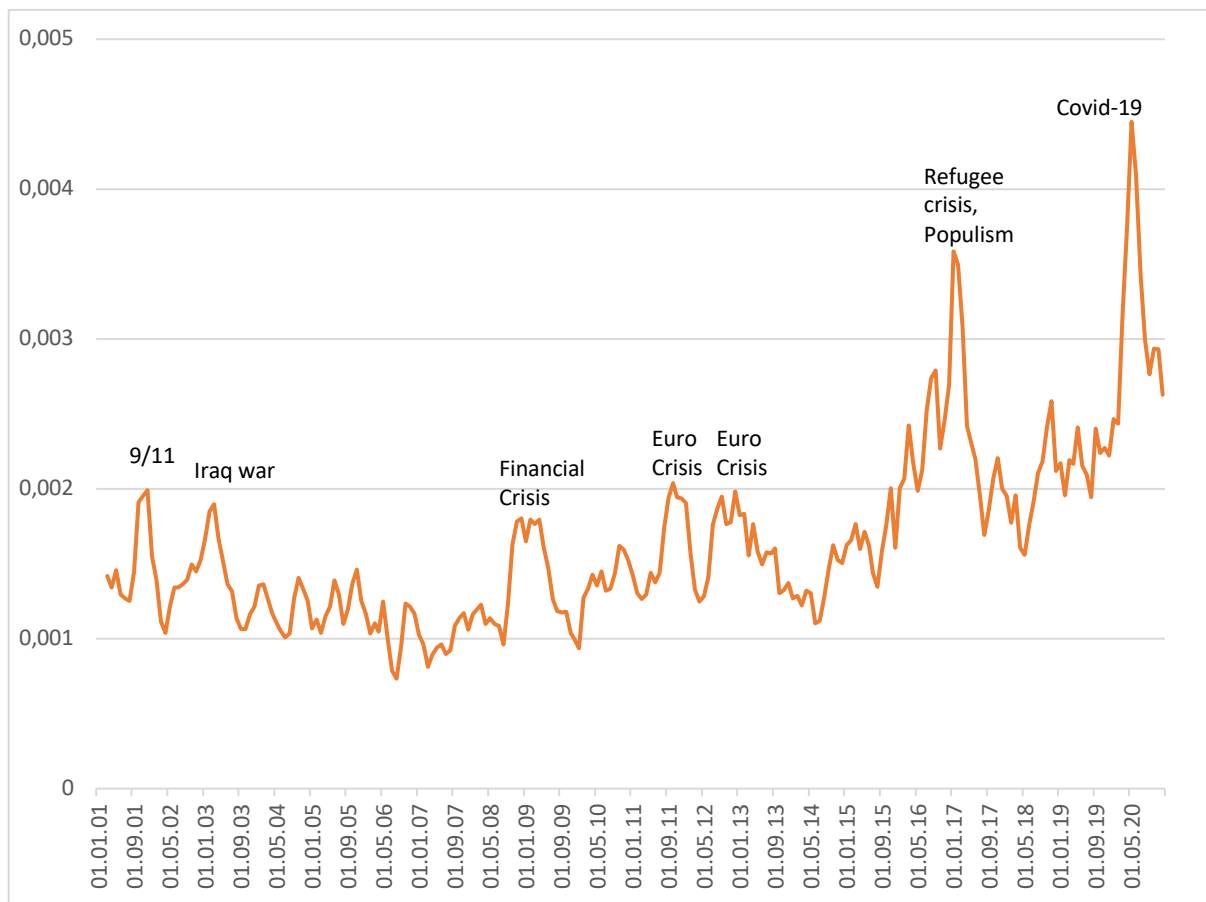
Die Welt, 2 Aug 2003

Germany's best years are yet to come

Süddeutsche Zeitung, 26 Feb 2009

Events: Peak 1 is associated with 9/11 and the vulnerability of Western societies. Peak 2 is associated with the Iraq war and the transatlantic tensions that accompanied it. Peaks 3, 4 and 5 are centered on the financial and the ensuing Euro crisis. Peak 6 is triggered by the influx of refugees into Germany in 2015/16 and the populist surge that followed. Peak 7 is Covid-related (fig 9).

Figure 9: Frequency of Topic “Society”



*share in entire corpus, backward-looking three-month moving average

Protagonists: Since this is not a genuinely political topic, but a societal one where deeper, structural issues are raised, it is not primarily about individual or institutional actors but about abstractions, such as Germany, Europe, State, and Globalization.

Problem definition: Globalization poses a challenge to Germany's society and economy. It comes in different forms and shapes: terror, unstable international alliances, financial market turmoil, migration, pandemics.

Problem diagnosis: Germany is ill-prepared to face the challenges.

Moral judgement: Ill-preparedness results in suffering and heightened uncertainty, nationally and globally.

Possible remedies: Depending on the slant of individual publications a wide range is put forward, from enhancing competitiveness and stricter immigration enforcement to more international engagement.

Summing up our quantitative-qualitative analysis, we subsume the evidence and formulate the *uncertainty narrative* as follows:

Globalization in its different manifestations poses a threat to German society and its economy. The problems are becoming more pressing as its effects move closer to home; migration and pandemics are more immediate threats, and therefore more closely watched by the media (fig. 9), than international terrorism or armed conflicts elsewhere. Germany needs to prepare for this disorderly world, but there is no consensus (yet) how to tackle the challenges.

As time proceeds, other aspects of globalization may well become more prominent in the *uncertainty narrative*. Climate change is an obvious candidate. It already ranks rather high among the topic's 2020 quarter-by-quarter top words.

Traces of Corona

In earlier papers we found the pandemic mainly attached to one distinct topic. Only in the “seeded” UPI version in Müller et al. (2021) the Corona issue appeared in a range of issues. In the *RollingLDA* model discussed here, we find Corona where such an overwhelmingly big story belongs: everywhere. Counting mentions of Corona, and related words such as Covid or pandemic, show that it is present in each and every topic (table 2). Given the pandemic's impact on everyday life, politics, the economy, European and international relations, and many more aspects, these results appear to be appropriate.

Table 2: Mentions of Corona (and related words) in the topics, number of mentions

| Month/Topic No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|-----------------|----|----|----|----|----|-----|----|----|----|-----|-----|-----|-----|----|
| 2020-01-01 | 0 | 0 | 0 | 1 | 4 | 16 | 1 | 0 | 5 | 10 | 2 | 2 | 10 | 0 |
| 2020-02-01 | 6 | 3 | 3 | 13 | 27 | 15 | 11 | 8 | 8 | 92 | 40 | 34 | 57 | 3 |
| 2020-03-01 | 56 | 16 | 17 | 67 | 32 | 127 | 71 | 63 | 72 | 135 | 144 | 156 | 88 | 43 |
| 2020-04-01 | 84 | 11 | 15 | 86 | 29 | 144 | 58 | 96 | 73 | 83 | 71 | 78 | 68 | 28 |
| 2020-05-01 | 70 | 12 | 4 | 85 | 24 | 129 | 91 | 73 | 44 | 61 | 83 | 74 | 106 | 26 |
| 2020-06-01 | 53 | 28 | 21 | 64 | 18 | 153 | 36 | 64 | 49 | 46 | 58 | 34 | 70 | 13 |
| 2020-07-01 | 51 | 15 | 5 | 56 | 11 | 68 | 26 | 47 | 42 | 55 | 45 | 41 | 110 | 10 |
| 2020-08-01 | 36 | 12 | 8 | 31 | 34 | 75 | 34 | 51 | 27 | 48 | 45 | 22 | 63 | 26 |
| 2020-09-01 | 39 | 9 | 6 | 32 | 17 | 111 | 20 | 49 | 19 | 47 | 39 | 24 | 96 | 20 |
| 2020-10-01 | 38 | 6 | 10 | 37 | 66 | 92 | 36 | 47 | 47 | 56 | 65 | 36 | 146 | 56 |
| 2020-11-01 | 36 | 16 | 8 | 48 | 8 | 93 | 41 | 56 | 43 | 75 | 41 | 33 | 93 | 14 |
| 2020-12-01 | 54 | 19 | 5 | 39 | 6 | 84 | 14 | 42 | 48 | 36 | 24 | 12 | 88 | 15 |

5. Conclusions

In this paper we presented a new dynamic topic modeling approach we call *RollingLDA*. Contrary to the now well-established static LDA, it allows for alterations in topics over time, a technique that seems capable of capturing the fluid nature of *media narratives* and collective memory. From a technical point of view, this approach has two major advantage: it produces consistent time-series over long periods, in our example 20 years, and it markedly reduces the arbitrariness of traditional LDA, a probabilistic approach often haunted by erratically changing results with each run. The latter achievement is particularly due to the combination of the *RollingLDA* method with *LDAPrototype* (Rieger, Koppers, Jentsch and Rahnenführer 2020) at various stages of the process.

We applied these methodological enhancements to the measurement of economic uncertainty in a large media corpus. The Uncertainty Perception Indicator (UPI) for Germany, presented in this paper, is based on a corpus of 2.8 million articles published in three leading German newspapers between 1 Jan 2001 and 31 Dec 2020. The results are quite promising: we were able to deconstruct the evolving composition of perceived economic uncertainty over time.

Furthermore, we squared the concept of *economic narratives* with the media-based measurement of economic uncertainty. Since uncertainty causes strong emotions related to fear, it is a condition closely linked to system-1 thinking known from dual-process theory. This mode of “thinking fast” is often based on narratives. Easy-to-grasp stories provide a short-hand enabling humans to take quick decisions in dangerous circumstances, of which they have little knowledge. *Uncertainty narratives* may therefore be considered a key to understand the societal reaction to being exposed to new unknowns. In this paper, we took up a definition of a media narrative we proposed before (Müller et al., 2018) and employed it in the analysis of the results of the *RollingLDA* UPI model. Indeed, we were able to isolate an *uncertainty narrative* for Germany. This story has risen considerably in prevalence since the mid-2010s, and it is bound to inform the reactions to future uncertainty shocks. It is part of the *collective memory* that a nation can draw from when unexpected developments pose new challenges.

The uncertainty narrative can be interpreted as a collective intangible asset (Westlake and Haskel, 2017), accumulated in the past, informing the present and potentially the future, as the story is being updated and partly overwritten by new experiences; it can also be a liability, if it leads societies to replicate errors or if wrong turns of the past become a hinderance today. Studying the properties of different uncertainty narratives may prove a fascinating field of future research.

Further research at DoCMA will focus on these cultural economic aspects as well as on the enhancement of our dynamic topic modeling approach. For instance, assessing the variations of topics over time in detail, which was outside the scope of this paper, will happen in due course. Comparing uncertainty narratives in different, but economically and politically interconnected countries would be an interesting exercise. We are currently working on French, Spanish, US and Chinese versions of the UPI.

A deeper look into certain points in the data might also hint at how media deal with longer phases of uncertainty. As mentioned at the beginning, we see phases of relative calm after

events like the stalled Brexit talks or the start of Donald Trump's presidency. What causes these patterns? A fertile field of research could open up here.

Data Transparency

We feel that the time-series produced in this paper are reliable and stable to a degree that they are useable for researchers in related disciplines, such as economics, political science, sociology or psychology. We encourage such undertakings and are willing to share the data on request.

Acknowledgements

The authors gratefully acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359.

6. References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bloom, Nicholas. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3), 623–685.
<https://doi.org/10.3982/ECTA6248>
- Bloom, Nick, Bond, S., & Van Reenen, J. (2007). Uncertainty and Investment Dynamics. *The Review of Economic Studies*, 74(2), 391–415. <https://doi.org/10.1111/j.1467-937X.2007.00426.x>
- Brandt, R. (2021). *Economic Policy Uncertainty Index*. DoCMA Working Paper #4,
<https://doi.org/10.17877/DE290R-21922>
- Chang, J. (2015). *lda: Collapsed Gibbs Sampling Methods for Topic Models* (R package version 1.4.2). <https://cran.r-project.org/web/packages/lda/index.html>
- Chang, J., Boyd-Graber, J. L., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems* 22 (pp. 288–296). <http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf>
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41(6), 570–606.
<https://doi.org/10.1016/j.poetic.2013.08.004>
- Downs, A. (1972). Up and Down with Ecology. The “Issue-Attention Cycle.” *The Public Interest*, 28, 38–50.
- 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>

- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Supplement 1), 5228–5235.
<https://doi.org/10.1073/pnas.0307752101>
- Grün, B., & Hornik, K. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(1), 1–30. <https://doi.org/10.18637/jss.v040.i13>
- Haskel J., & Westlake S. (2017). *Capitalism without Capital: The Rise of the Intangible Economy Hardcover*. Princeton University Press
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin UK.
- Koppers, L., Rieger, J., Boczek, K., & Nordheim, G. von. (2020). *tosca: Tools for Statistical Content Analysis* (R package version 0.2-0). <https://doi.org/10.5281/zenodo.3591068>
- LexisNexis (2021). Nexis: LexisNexis Academic & Library Solutions. URL:
<https://www.nexis.com> and <https://www.lexisnexis.de/>
- Miltner, P., & Waldherr, A. (2013). Themenzyklus der Kriegsberichterstattung – ein Phasenmodell. *Publizistik*, 58(3), 267–287. <https://doi.org/10.1007/s11616-013-0180-2>
- Müller, H., & Hornig, N. (2020a). *Expecting the Unexpected: A new Uncertainty Perception Indicator (UPI) – concept and first results*. DoCMA Working Paper #1.
<http://dx.doi.org/10.17877/DE290R-21089>
- Müller, H., & Hornig, N. (2020b). *I heard the News today, oh Boy*. DoCMA Working Paper #2
<https://doi.org/10.17877/DE290R-21669>
- Müller, H., Hornig, N., & Rieger, J. (2021). *“For the times they are a-changin’”*. DoCMA Working Paper #3. <https://doi.org/10.17877/DE290R-21878>
- Müller, H., von Nordheim, G., Boczek, K., Koppers, L., & Rahnenführer, J. (2018). Der Wert der Worte – Wie digitale Methoden helfen, Kommunikations- und

Wirtschaftswissenschaft zu verknüpfen. *Publizistik*, 63(4), 557–582.

<https://doi.org/10.1007/s11616-018-0461-x>

Nguyen, V.-A., Boyd-Graber, J., & Resnik, P. (2014). Sometimes Average is Best: The Importance of Averaging for Prediction using MCMC Inference in Topic Modeling.

Proceedings of the 2014 Conference on Empirical Methods in Natural Language

Processing (EMNLP), 1752–1757. <https://doi.org/10.3115/v1/D14-1182>

Putnam, R. (1993). The Prosperous Community: Social Capital and Public Life. *The American Prospect*, 13, 35–42.

Rieger, J. (2020). IdaPrototype: A method in R to get a Prototype of multiple Latent Dirichlet Allocations. *Journal of Open Source Software*, 5(51), 2181.

<https://doi.org/10.21105/joss.02181>

Rieger, J., Rahnenführer, J. & Jentsch, C. (2020). Improving Latent Dirichlet Allocation: On Reliability of the Novel Method LDAPrototype.. *Natural Language Processing and Information Systems, NLDB 2020*. (pp. 118–125). Springer International Publishing.

https://doi.org/10.1007/978-3-030-51310-8_11

Rieger, J., Koppers, L., Jentsch, C., & Rahnenführer, J. (2020). Improving Reliability of Latent Dirichlet Allocation by Assessing Its Stability Using Clustering Techniques on Replicated Runs. *ArXiv:2003.04980 [Cs, Stat]*. <http://arxiv.org/abs/2003.04980>

Shiller, R. J. (2017). Narrative Economics. *American Economic Review*, 107(4), 967–1004.

<https://doi.org/10.1257/aer.107.4.967>

von Nordheim, G. (2019). *Journalism in the Age of Singularization* [Cumulative Doctoral Thesis]. Department of Cultural Studies, TU Dortmund.

von Nordheim, G., Boczek, K., & Koppers, L. (2018a). Sourcing the Sources. *Digital Journalism*, 6 (7), 807–828. <https://doi.org/10.1080/21670811.2018.1490658>

von Nordheim, G., Boczek, K., Koppers, L., & Erdmann, E. (2018b). Digital Traces in Context.

Reuniting a Divided Public? Tracing the TTIP Debate on Twitter and in Traditional Media. *International Journal of Communication*, 12, 22.

Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, 28(4), 1–38.

<https://doi.org/10.1145/1852102.1852106>

7. Appendix

The graphs in fig. 10 shows the self-similarity of the topics over time using cosine similarity. The quarter-to-quarter similarities are shown in black, the month-to-month similarities in gray. In addition, the quarter-to-quarter similarity of the respective topic at the first time point (2001 Q1) to all following quarters is shown in blue, as well as the corresponding similarities of the last quarter (2020 Q4) to all preceding quarters in orange. The black, blue and orange curves are useful for assessing the stability of the topic over the observation period, while the gray curve is useful for identifying individual events within a topic.

Figure 10: Self-Similarity of Topics over time* (Cosine Similarity)

