Early Warning Systems? Building Time Consistent Perception Indicators for Economic Uncertainty and Inflation Using Efficient Dynamic Modeling

Jonas Rieger, [1] **Nico Hornig,** [2] **Tobias Schmidt,** [2] **Henrik Müller** [2]

[1] Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
[2] Institute of Journalism, TU Dortmund University, 44221 Dortmund, Germany
rieger@statistik.tu-dortmund.de, nico.hornig@tu-dortmund.de,
tobias3.schmidt@tu-dortmund.de, henrik.mueller@tu-dortmund.de

Abstract

We propose a framework for the dynamic construction of newspaper-based indicators, e.g., as variables in financial forecasting. We follow the idea of the composition of the economic policy uncertainty (EPU) indicator by Baker, Bloom, and Davis (2016), develop it further and present the general conception of this enhanced version of indicators. Using topic modeling, we refine analysis possibilities so that topics within the corpus (subindices) can be analyzed in addition to the global index. By using a rolling approach (RollingLDA), little human effort and computational resources are needed for updates, which makes the framework besides nowcasting also interesting for live monitoring and change detection scenarios. We present the framework in terms of the German uncertainty perception indicator (UPI) and inflation perception indicator (IPI), but the general concept can easily be adapted to related research questions.

1 Introduction

Understanding the current overall economic situation is of great importance (not least due to the sharp rise in inflation) to both private individuals and companies themselves. When tabular information describing it such as established economic indicators (e.g., Deutsche Bundesbank 2022) is missing, either because it cannot be represented in a tabular form or because it will not be published until a later date, textual data can be and is often used to bridge this gap. Notable examples include gauging the perception of uncertainty using newspaper articles (cf. EPU by Baker, Bloom, and Davis 2016) or nowcasting economic indices such as GDP (e.g., Garnitz, Lehmann, and Wohlrabe 2019; Thorsrud 2020; Müller 2022; Shrub et al. 2022).

To model these articles, topic models (Blei 2012) are particularly suitable, as they open up the possibility of making different facets of the indicators both measurable and interpretable as subindices. If the models are to be used not only as one-time snapshot, but also in live scenarios, they should — in addition to an appropriate fitting to the data — fulfill mainly two conditions: fast computability and the possibility to update the model without rewriting past statistics, so that temporally consistent indicators result. In addition, reducing the need for human interaction may be of special interest. In the sense of updating, the model used should also be able to incorporate mutations of topics.

1.1 Related work

One of the first models that enabled temporal topics is topics over time (Wang and McCallum 2006). The model creates temporally narrow topics based on the knowledge that specific word co-occurrences appear very frequently in a short period of time. However, the fact that it considers future time points for modeling past time points makes it unsuitable for use in a monitoring context, since updating is not possible without recalculating the model. A rather similar idea implements the quite well-known continuous time dynamic topic model by Wang, Blei, and Heckerman (2008) as a continuous version of the discrete time dynamic topic model by Blei and Lafferty (2006). In both models, topic evolution is implemented gradually, and the parameters of past time points are also estimated taking into account future word prevalences. With respect to the Covid-19 outbreak in early 2020 it can be shown for an appropriate dataset that even a few months before the first occurrence of the word *covid* in the texts, the estimated probability of its occurrence increases sharply. This means that adding new data will also change the model for previous texts.

Hong et al. (2011) propose a model that incorporates temporal structures using simultaneous text streams from twitter and newspapers, but do not provide a monitoring solution in terms of updating the model. Another method to model temporal topics in the social media context offer Wang, Agichtein, and Benzi (2012) with their temporal latent Dirichlet allocation (LDA) that focuses on changes in writing of authors instead of global changes in topics. Streaming-LDA (Amoualian et al. 2016) models articles in a consecutive order, but uses information of future articles in later iterations of the Gibbs procedure as well.

Including information about future texts in the modeling is an unrealistic scenario for the monitoring context. In addition, all described models can only deal with fixed vocabularies or do not address how it could integrate new vocabularies. This lacks the desired flexibility of the model, e.g., when a model is updated in early 2020, it should be possible to include the word *covid*. However, with a variant of the online LDA, Zhai and Boyd-Graber (2013) propose a model that considers infinitely large vocabularies. Thus, it enables

This article is accepted for publication via
Third International Workshop on Modelling Uncertainty in the Financial World (MUFin'23) —
https://sites.google.com/view/w-mufin/ — held in conjunction with the 2023 AAAI.

the addition of new vocabularies and the deletion of vocabularies that are no longer used. The architecture of the model allows it to be used in a monitoring context, but the static use of the entire past for all subsequent updates prevents the creation of new topics, so that the flexibility of the model is limited.

1.2 Contribution

We propose a framework for modeling and updating newspaper based indicators in a time consistent manner. In this context, we present two example indicators in particular. We define keyword combinations that suitably shrinks the corpus into smaller corpora. Then, in our pipeline, we apply a rolling modeling approach (Rieger, Jentsch, and Rahnenführer 2021) to the filtered corpus. We present the functionality of the framework and show how human effort as well as computational expense is reduced during the update process. For this, we make use of two German example indicators, the uncertainty perception indicator (UPI) and the inflation perception indicator (IPI).

2 Rolling modeling approach

The model idea is based on classical LDA (Blei, Ng, and Jordan 2003) estimated via collapsed Gibbs sampling (Griffiths and Steyvers 2004). However, there is the issue that repeated runs on the same data lead to strongly different results — especially due to different initializations. If the user is not interested in the concrete assignments of the words from the texts to the topics, the possibility of averaging nested and lagged Gibbs iterations (Nguyen, Boyd-Graber, and Resnik 2014) is an option. Moreover, this methodology still relies on a single starting initialization, so it is not guaranteed that a suitable representative of all possible solutions will be found in this way.

Optimization of evaluation measures such as perplexity (Grün and Hornik 2011) or NPMI (Lau, Newman, and Baldwin 2014) is another possibility. However, Chang et al. (2009) show that the optimization of likelihood-based measures such as perplexity is negatively correlated with human perception for well-separated topics and Doogan and Buntine (2021) show that coherence based measures such as NPMI can be unreliable for specific tasks. Accordingly, appropriate procedures for automated topic model evaluation are rather ongoing research (Hoyle et al. 2021) (again), so that we instead rely on the use of human-in-the-loop for some decisions on model parameters. However, overall Harrando, Lisena, and Troncy (2021) were able to show that LDA provides robustly consistent performance regarding established (but to be further reviewed) model metrics, which justifies the use of a rather traditional technique like LDA in comparison to transformer based methods such as BERT (Devlin et al. 2019).

Therefore we use the method LDAPrototype (Rieger, Jentsch, and Rahnenführer 2022) for a reliable initialization of our rolling approach. The method selects from a set of candidates the model that represents the medoid, i.e., the model that has the highest average similarity to all other candidates. In our analysis, we modeled 100 candidate LDAs for the selection, which is also the default value of the method. The authors show that for larger datasets even 50 candidates are sufficient, while for smaller datesets they recommend to select the medoid from 100 models.

We use the method on the entire dataset up to a user-defined date (parameter `init`), up to which a sufficient large number of articles have been published, so that a reliable initial modeling of the topics can be ensured. From this date on, modeling is done in minibatches (`chunks`) of user-defined size or periods, respectively. All articles from a predefined period (`memory`) right before this minibatch are used to initialize the topic distributions for the modeling of the minibatch. The model is then updated adding the topic assignments of the modeled texts of the given minibatch. If new words are observed during the modeling of new articles, which were not included in the vocabulary before this minibatch, they will be added to the model if they exceed certain thresholds (`vocab.abs`, `vocab.rel`). The combined methodology as a rolling approach is referred to as RollingLDA (Rieger, Jentsch, and Rahnenführer 2021) and is implemented in the R package rollinglda.

3 UPI and IPI as example applications

In the following, we demonstrate how the described conceptual framework is practically applied by using the UPI (cf. Müller et al. 2022) and IPI (cf. Müller, Rieger, and Hornig 2022) as examples. The idea of the two indicators is to measure the German coverage of economic uncertainty or inflation over time and to decompose it thematically.

Both indices are inspired by the concept of the well-known EPU (economic policy uncertainty, cf. Baker, Bloom, and Davis 2016). That is, from a set of media coverage that is intended to be as representative as possible for the media landscape of interest, articles on economic uncertainty on the one hand and on inflation on the other hand are identified and organized in separate corpora. Then, we apply topic modeling techniques on these corpora. The soft-clustering into these topics within the already restricted corpora is a substantial refinement compared to the EPU. Thus, in addition to the share of a thematically narrowed corpus with respect to the overall corpus (global index), the shares of the (sub)topics (subindices) can be determined and analyzed over time.

The results and analysis scripts can be accessed at https://github.com/JonasRieger/upi for the UPI and at https://github.com/JonasRieger/ipi for the IPI. The data is continuously updated on a monthly basis. Results regarding specific calculations for the paper are available at https://github.com/JonasRieger/mufin23.

3.1 Data and corpus filtering

Both indices, UPI and IPI, are based on a German database corpus of all newspaper articles published in *Die Welt*, *Handelsblatt* and *Süddeutsche Zeitung* since 01/01/2001. By the end of November, this corpus contains 2.9 million texts or 565 billion words. There are several reasons to favor smaller and denser corpora against bigger and more general ones. One reason is faster computability and the responsible use of resources (Strubell, Ganesh, and McCallum 2019). In addition, a larger corpus implies the need for a higher number of topics to obtain a similar granularity of relevant topics compared to a more specialized corpus. This means that more human effort is required to determine the relevant topics and their labels. In order to limit the required amount of human-in-the-loop, it seems useful to reduce the corpus in such a way that irrelevant articles are removed. This means that the precision with respect to the relevance of the articles is maximized while — at best — maintaining a recall close to the 100% by an appropriate keyword filter.

An article is kept in the respective subcorpus if the associated text satisfies the keyword combination. For the UPI, we use the keyword combination

$$\text{*unsicher* AND *wirtschaft*,} \qquad \text{(UPI)}$$

where both words (uncertain, economic) are searched as pattern search, i.e., subwords of tokens are also matched. The combination is chosen broader than the popular EPU indicator (Baker, Bloom, and Davis 2016) to ensure a higher recall of the relevant articles. In comparison, the EPU is build using the keyword combination

$$\begin{aligned} &\{\text{wirtschaft OR wirtschaftlich}\} && \text{(EPU)}\\ \text{AND } &\{\text{steuer OR wirtschaftspolitik} &&\\ &\text{OR regulierung OR regulierungs} &&\\ &\text{OR ausgaben OR bundesbank} &&\\ &\text{OR ezb OR zentralbank OR haushalt} &&\\ &\text{OR defizit OR haushaltsdefizit}\} && \text{(EPU)}\\ \text{AND } &\{\text{unsicher OR unsicherheit}\} && \text{(EPU)} \end{aligned}$$

utilizing the archives of articles from *Handelsblatt* and *Frankfurter Allgemeine Zeitung*. Here, all words are searched as actual tokens, which means that there is no hit for matched patterns as subwords of tokens.

For the IPI, we use the combination of inflation synonyms

$$\begin{aligned} &\text{*inflation* OR *geldentwertung*} &&\\ &\text{OR teuerung OR *preissteigerung*.} && \text{(IPI)} \end{aligned}$$

All searchword combinations are case insensitive. For the IPI, "teuerung" is searched as the actual token to avoid incorrect matches by, e.g., "besteuerung" (taxation). This results in 40 210 texts for the UPI and 56 085 texts for the IPI, which are preprocessed using common steps like use of lowercase, number removal, conservative stopword removal (cf. Schofield, Magnusson, and Mimno 2017), but no stemming (cf. Schofield and Mimno 2016).

3.2 Precision and recall

To evaluate the keyword combinations, we draw samples of articles. Specifically, to estimate the recall, we draw 1800 of a total of 2.8 million (0.7‰) articles from the outer set of articles that are not covered by the UPI or IPI searchword combination. For precision (and recall) estimation, we draw 100 articles each from the UPI (2.5‰) and IPI (1.9‰) corpus. All articles are labeled by a human coder whether they are relevant to the UPI/IPI corpus or not. From this, the precision estimators can be calculated straight forward, while the recall can be estimated using weighted proportions. This results in the 95% confidence intervals given in Table 1.

	UPI		**IPI**	
Bound	Precision	Recall	Precision	Recall
lower	0.5354	0.2232	0.7564	0.7774
upper	0.7246	0.4071	0.9036	0.9259

Table 1: Precision and recall estimation (95% confidence interval) for the relevance of articles matched by the UPI and IPI searchword combinations.

3.3 Parameter setting

To model the texts, we use the RollingLDA method described in Section 2. In addition to the parameters explained there, the method requires a specification of the parameters of basic LDA, namely the number of modeled topics K, as well as the Dirichlet priors of the topic and word distributions α and η. To determine an appropriate number of topics for the two corpora, we tried $K = 5, 6, \ldots, 20$ with $\alpha = \eta = 1/K$ for both. Given the top words and top texts per topic and based on human judgment, we selected the most adequate model for the purpose of thematically differentiated economic indicators for each dataset. For the UPI, this procedure results in the parameters $K = 14$ and

- `init = "2005-12-31"`,
- `chunks = "month"`,
- `memory = "3 quarter"`.

The IPI dataset is modeled with $K = 10$ and with six months instead of three quarters for `memory`. For all other parameters, the default values were used; in particular we choose `vocab.abs = 5` and `vocab.rel = 0`, so that words that occur more than five times in a chunk are taken into account for the current (and all further) modeling steps. The modeling schemes are shown graphically in Figure 1 and 2. The numbers in the memory batches (blue) indicate the rolling ID of the corresponding minibatch. Monthly updates of the model take only a few minutes due to the reduction to relevant articles and can be analyzed timely after the publication of the last texts of the month.

3.4 Findings

From the RollingLDA models, we obtain 14 (UPI) and ten (IPI) topics characterized by their word distributions as well as topic assignments for the modeled articles. In addition, we can calculate the proportion of each topic compared to the overall corpus, and thus visualize their courses over time. In Figure 3, the temporal courses of all UPI and IPI topics are shown. The labels of the topics were determined based on top words and top texts using human labeling once when the model was built from scratch and are also checked manually for validity after each update. In all charts, the period up to the end of 2005 is shaded gray, as the RollingLDA starts at the beginning of 2006, cf. Section 3.3.

Topics as subindices It appears that each model contains one topic *Miscellaneous*, which results from less than 100% precision in the corpora (cf. Table 1) and which is formed by the model out of irrelevant paragraphs of articles. The curves in Figure 3 display the topics' monthly share of the

This article is accepted for publication via
Third International Workshop on Modelling Uncertainty in the Financial World (MUFin'23) —
https://sites.google.com/view/w-mufin/ — held in conjunction with the 2023 AAAI.

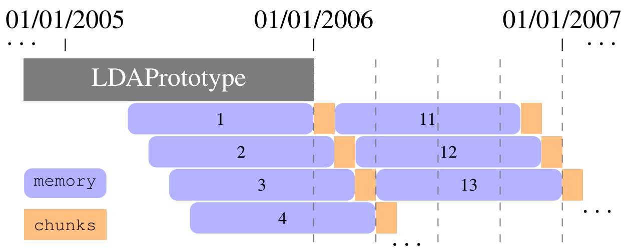

Figure 1: Modeling scheme for the UPI.

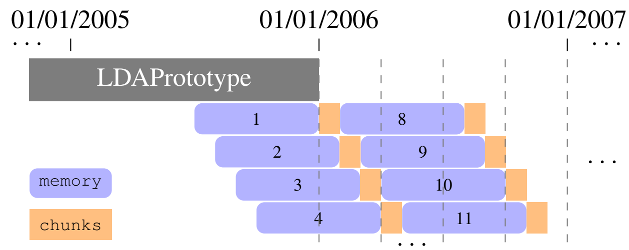

Figure 2: Modeling scheme for the IPI.

total coverage of the newspapers *Die Welt*, *Handelsblatt* and *Süddeutsche Zeitung*. The sum of all thematically relevant topic shares results in the global UPI or IPI. The advantage of our approach compared to that of Baker, Bloom, and Davis (2016) is that not only a single indicator is created, but the split into topics enables the interpretation for the causes of changes in the global index. In addition to the single courses and the overall sum, we classify each topic of the UPI to be part of one of the three subindices *Real Economy*, *Politics* or *Financial Markets*. For the IPI, we define the three classes *Causal*, *Consequences* and *Other*. The classification as an inflation-influencing or inflation-influenced topic is not as clear-cut as the classification of the subindices of the UPI, but it provides a first orientation for deeper insights.

Brexit, Covid-19, RU-UA war There are plausible patterns to known events. The clearest peak is marked by the discussion about the leaving of Great Britain from the European Union (Brexit). This is apparent in 2016 in the topic *EU Conflicts* from the UPI subindex *Politics*. However, overall, the Covid-19 pandemic has the largest global effect on the UPI so far. At the beginning of 2020, many topics rise sharply. The most notable topics are *Leisure & Hospitality (Real Economy)* and *Society (Politics)*. At the beginning of 2022 a change in coverage due to the war in the Ukraine can be observed. Here, the greatest (short-term) effect is (plausibly) observed for *Geopolitics* but the topics *German Economy* and *Energy & Climate* also show (rather long lasting) irregularities. For the IPI all nine curves reached their all-time high within the last three months. The *Financial Markets (Consequences)* topic already shows notable peaks at the end of 2020. This is followed by a sudden increase in the share of reporting in all IPI topics.

Inflation For July 2022, a slight increase of the IPI can be seen compared to June, while since then the indicator settles at a high level. The UPI also continues to indicate high prevailing economic uncertainty in the considered newspapers. The topic *News* (cf. IPI) surpasses the other topics (except for *Financial Markets*) in July, i.e., a lot of content about inflation during this time are classical (breaking) news. In contrast, the more concrete topics *Central Banks* and *Raw Materials* (short-term peak in June 2022) lose some ground, while the topic *German Politics* rises. Since then, it remains consistently at a high until the end of November 2022. It can be seen that topics related to the consequences of inflation decrease, while mainly topics related to regulations (cf. *German Politics*, *Eurozone*) remain present. At the beginning of September, e.g., the German government agreed on a package of measures to ease the financial burden on the population. This shows that individual curves and (sub-)indicators may have the potential for early warning systems, or to be included in nowcasting scenarios to predict financial indicators. In principle, however, the indicators represent a retrospective view of the reporting of the respective months. The extent to which trends in each of the curves have predictive value can be examined in further research by integrating the indicators into state-of-the-art nowcasting frameworks.

3.5 Features

An important feature of the rolling framework is the possibility to include new vocabularies in the modeling. In Table 2 the size of the vocabulary for both indices is specified depending on the month of update. Vocabularies, once included, are not removed, so the vocabulary may only grow. It becomes apparent that a relatively high number of new words are added to the IPI vocabulary in July 2022. For example, there are eight words containing the pattern "gas" as well as the words "gesetzespaket" (legislative package) and "grundbedarf" (basic needs). For the UPI, in July 2022 even eleven words containing "gas", such as "gaslieferstopp" (gas supply stop), "gaskrise" (gas crisis), "gasmangel" (gas shortage), but also the prominent new disease of "affenpocken" (monkeypox) are added.

Static modeling as an alternative to the presented dynamic

Month	UPI	IPI
January	53 358	48 594
February	53 413	48 688
March	53 484	48 805
April	53 548	48 949
May	53 585	49 122
June	53 662	49 397
July	53 736	49 630
August	53 810	49 844
September	53 874	50 099
October	53 916	50 281
November	53 956	50 445

Table 2: Number of vocabulary considered for UPI and IPI in the months of 2022.

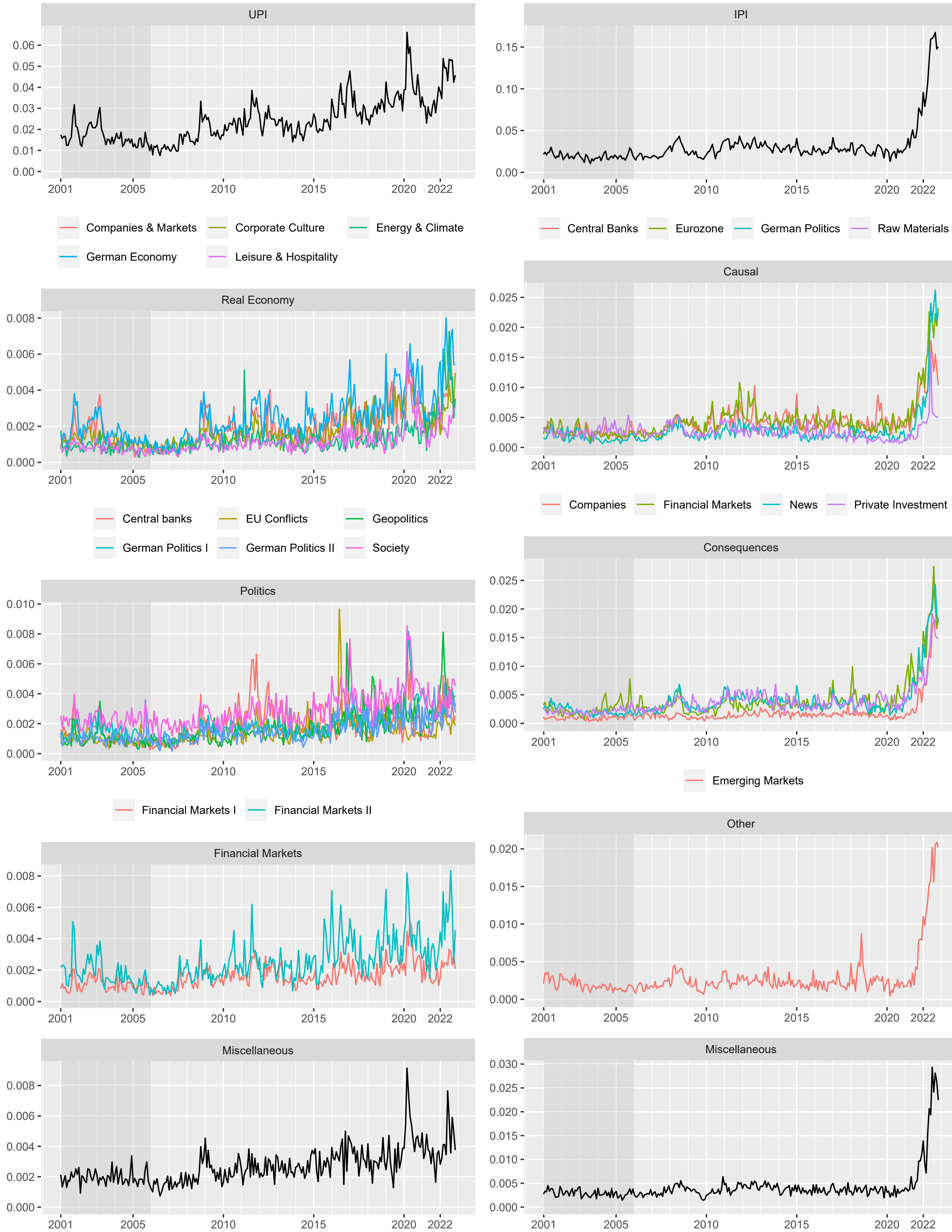

Figure 3: Temporal course of the 13 UPI topics and the nine IPI topics. The UPI and IPI indices are defined as the sum of all respective topics except *Miscellaneous*. All topics are associated with one of the three subindices *Real Economy*, *Politics* and *Financial Markets* for the UPI, or classified roughly as to be *Causal*, *Consequences* or *Other* for the IPI. Each curve shows the monthly share of the total coverage of the newspapers *Die Welt*, *Handelsblatt* and *Süddeutsche Zeitung* from 01/01/2001 until 11/30/2022.

This article is accepted for publication via
Third International Workshop on Modelling Uncertainty in the Financial World (MUFin'23) —
https://sites.google.com/view/w-mufin/ — held in conjunction with the 2023 AAAI.

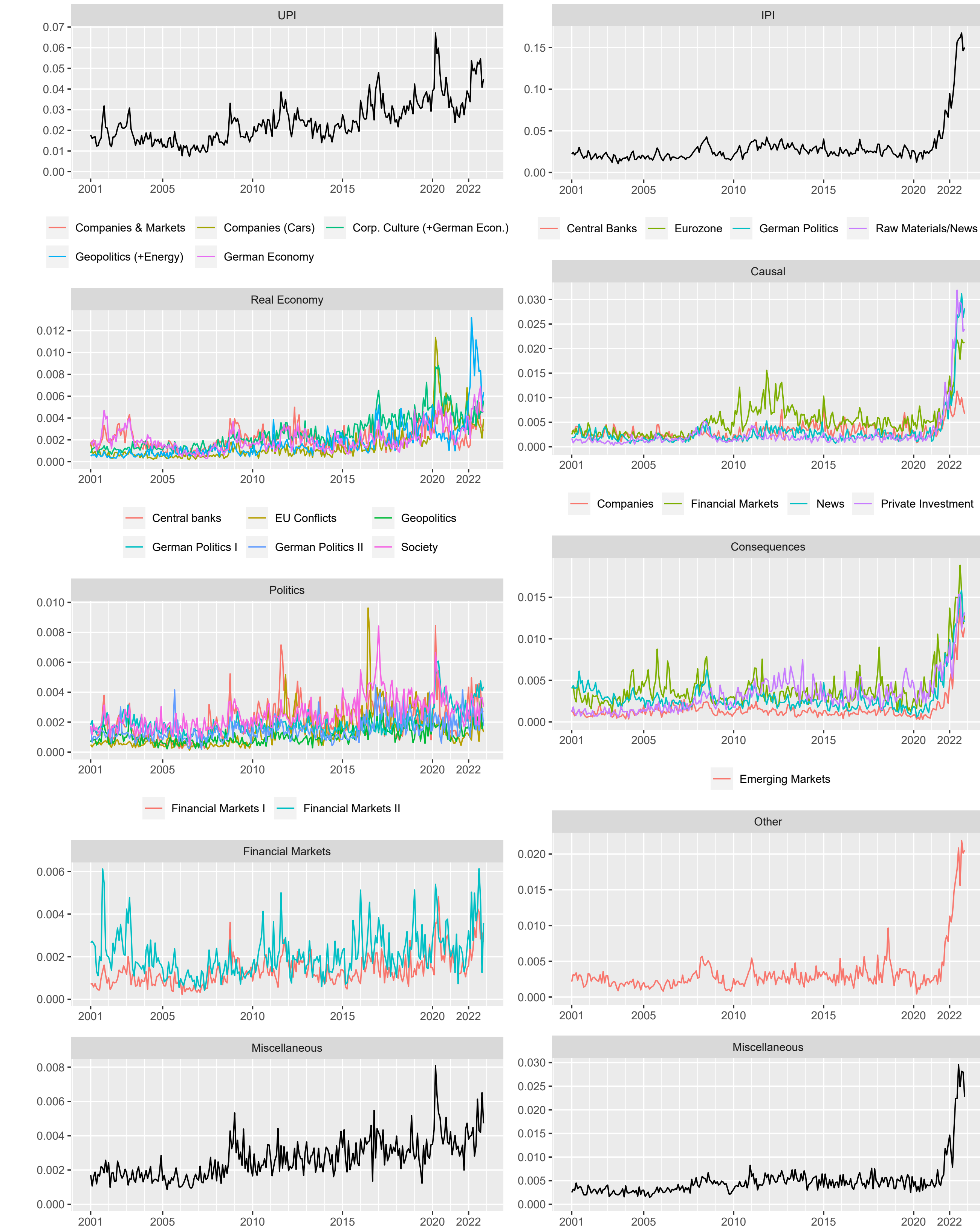

Figure 4: Temporal course of the topics resulting from snapshot modeling. The topic labels were matched as closely as possible with those from Figure 3.

This article is accepted for publication via
Third International Workshop on Modelling Uncertainty in the Financial World (MUFin'23) —
https://sites.google.com/view/w-mufin/ — held in conjunction with the 2023 AAAI.

framework only allows snapshot analyses. In this case, the vocabulary would also change with each update, depending on the selected preprocessing steps. However, this potential inclusion of a word in the vocabulary would also affect previous articles' assignments. Likewise, all articles would have to be remodeled with each update, which leads to high computational costs. A comparison with previous models is not possible without further techniques, e.g., topic matching. Furthermore, it is unclear how time series of the resulting models could be analyzed consistently.

As comparison, we apply a static snapshot modeling approach using LDAPrototype (Rieger, Jentsch, and Rahnenführer 2022) to select the most reliable of 100 candidate models with 14 (UPI) or ten (IPI) topics, respectively. All other parameters are chosen as package defaults (Rieger 2020). Figure 4 shows the temporal courses of all modeled topics of the static approach until the end of November 2022 are shown. The topic structure of the static model show a closer relationship to events, whereas RollingLDA models the topics at a higher level and shows a weaker co-occurrence of abruptly increasing topics' shares with the appearance of events. Instead, RollingLDA's topics mutate in these cases. These findings show that recalculating of static LDA at for new incoming data would lead — beside high computational costs — to enormous human effort, since the topic granularity is strongly affected by newly occurring events, so that matching topics across models would be costly. The rolling variant makes it possible to integrate events into existing topics so that they are automatically assigned to a topic.

4 Conclusion and limitations

We presented an example application of a framework to create newspaper based indicators dynamically and time consistently. Further analyses are required to assess the predictive power, the extent to which these are suitable for the nowcasting of several different important economic indicators. This may also answer the specific question of whether the UPI and IPI can serve as early warning systems for various facets of the economic landscape. At least, these indicators already represent the coverage of the corresponding (sub-)topics. At this point, it should be noted, however, that increased coverage is not (necessarily) equivalent to prevailing uncertainty or inflation, respectively. In general, when interpreting UPI and IPI curves, the various dependencies should be kept in mind, namely, to what extent an increase in a curve was caused by an exogenous event and to what extent this increase in the public's attention may result in a further increase.

4.1 Increasing the keywords' recall

Excluding the word "inflationaer" in the search term would reduce the incorrect inclusion of articles for the IPI, which is already low, marginally. While the precision of the corpora is not such a big issue, one possible refinement of the indicators involves the recall (in particular) of the UPI keyword combination. During the process of labeling, we found some potentially useful substitutes for various keywords of the UPI and IPI which might increase the recall of the keyword combination while preserving a sufficiently high precision:

wirtschaft: aktienkurse, anleger, bank, boerse, finanzmarkt, finanzmaerkte, finanzsektor

unsicher: angst, fuerchte, gefahr, geruecht, misstrauen, nervoes, nervositaet, instabil, risiko, scheu, sorge, spekulation, spekulieren, stresstest, turbulen, umstritten, unberechenbar, unklar, unruhig

inflation: preisexplosion, stagflation

This probably results in an increased recall and a slightly lower precision. Care should be taken that by adding more articles the precision does not suffer too much: first, that the corpus does not become too large, and second, that the thematic narrowness of the corpus, with all the advantages described in Section 3.1, is preserved. It is likely that a slight decrease in precision, i.e., slightly more noise in the corpus, would lead RollingLDA to assign the corresponding articles to the *Miscellaneous* topic, which can easily be left out for further analysis.

As a side note, our analysis shows that the keyword combination used by Baker, Bloom, and Davis (2016, cf. EPU) is also too narrow. After all, the documents, from which the EPU is calculated, are a subset of the documents covered by our keyword.

5 Outlook

The presented framework allows for several further research directions. At first, it can be easily adapted to create time consistent time series for a variety of applications and research questions. For example, instead of newspaper articles, the analysis could also be performed for collections of tweets. Second, the modeled topics can be used to create a (qualitative or quantitative) early warning system for the respective topics in the corpus. For the (quantitative) monitoring of topic frequencies, classical change detection algorithms can be used (e.g., Bose and Mukherjee 2021; Fryzlewicz 2014), while changes in the content of topics can be identified by methods that use intra-topic similarities (e.g., Rieger et al. 2022; Keane, Yee, and Zhou 2015). Then, if the change detection method identifies a change, it can be evaluated by experts using additional information (e.g., top topic texts, emerging topic terms) for further analysis. Third, the splitting of the index into topical perceptions allows the integration of the subindices into state-of-the art downstream tasks such as nowcasting GDP or other economic indicators to bridge the lag of publication between established predictors using the information from text data. In addition, sentiment or stance detection methods, e.g., pre-trained adapters for transformer models (Pfeiffer et al. 2020), can be used off-the-shelf without fine-tuning to extract the articles' views on uncertainty or inflation, respectively. This would allow for explicitly distinguish between the mentions "geringe inflation" (low inflation) vs. "starke inflation" (strong inflation) in the subsequent analysis.

This article is accepted for publication via
Third International Workshop on Modelling Uncertainty in the Financial World (MUFin'23) —
https://sites.google.com/view/w-mufin/ — held in conjunction with the 2023 AAAI.

Acknowledgements

The present study is part of a project of the Dortmund Center for data-based Media Analysis (DoCMA, https://docma.tu-dortmund.de/). In addition, 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.

References

Amoualian, H.; Clausel, M.; Gaussier, E.; and Amini, M.-R. 2016. Streaming-LDA: A Copula-Based Approach to Modeling Topic Dependencies in Document Streams. In *Proceedings of the 22nd SIGKDD-Conference*, 695–704. ACM.

Baker, S. R.; Bloom, N.; and Davis, S. J. 2016. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4): 1593–1636.

Blei, D. M. 2012. Probabilistic Topic Models. *Communications of the ACM*, 55(4): 77–84.

Blei, D. M.; and Lafferty, J. D. 2006. Dynamic Topic Models. In *Proceedings of the 23rd ICML-Conference*, 113–120. ACM.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993–1022.

Bose, A.; and Mukherjee, S. S. 2021. Changepoint Analysis of Topic Proportions in Temporal Text Data. *arXiv*.

Chang, J.; Boyd-Graber, J.; Gerrish, S.; Wang, C.; and Blei, D. M. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. In *NIPS: Advances in Neural Information Processing Systems*, volume 22, 288–296. Curran Associates Inc.

Deutsche Bundesbank. 2022. Time series databases. Accessed 08/01/2022.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 NAACL-Conference, Volume 1 (Long and Short Papers)*, 4171–4186. ACL.

Doogan, C.; and Buntine, W. 2021. Topic Model or Topic Twaddle? Re-evaluating Semantic Interpretability Measures. In *Proceedings of the 2021 NAACL-Conference*, 3824–3848. ACL.

Fryzlewicz, P. 2014. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6): 2243–2281.

Garnitz, J.; Lehmann, R.; and Wohlrabe, K. 2019. Forecasting GDP all over the world using leading indicators based on comprehensive survey data. *Applied Economics*, 51(54): 5802–5816.

Griffiths, T. L.; and Steyvers, M. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1): 5228–5235.

Grün, B.; and Hornik, K. 2011. topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13): 1–30.

Harrando, I.; Lisena, P.; and Troncy, R. 2021. Apples to Apples: A Systematic Evaluation of Topic Models. In *Proceedings of the 2021 RANLP-Conference*, 483–493. INCOMA Ltd.

Hong, L.; Dom, B.; Gurumurthy, S.; and Tsioutsiouliklis, K. 2011. A Time-Dependent Topic Model for Multiple Text Streams. In *Proceedings of the 17th SIGKDD-Conference*, 832–840. ACM.

Hoyle, A.; Goel, P.; Hian-Cheong, A.; Peskov, D.; Boyd-Graber, J. L.; and Resnik, P. 2021. Is Automated Topic Model Evaluation Broken? The Incoherence of Coherence. In *NeurIPS: Advances in Neural Information Processing Systems*.

Keane, N.; Yee, C.; and Zhou, L. 2015. Using Topic Modeling and Similarity Thresholds to Detect Events. In *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, 34–42. ACL.

Lau, J. H.; Newman, D.; and Baldwin, T. 2014. Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. In *Proceedings of the 14th EACL-Conference*, 530–539. ACL.

Müller, K. 2022. German forecasters' narratives: How informative are German business cycle forecast reports? *Empirical Economics*, 62: 2373–2415.

Müller, H.; Rieger, J.; and Hornig, N. 2022. Vladimir vs. the Virus – a Tale of two Shocks. An Update on our Uncertainty Perception Indicator (UPI) to April 2022 – a Research Note. *DoCMA Working Paper*, #11.

Müller, H.; Rieger, J.; Schmidt, T.; and Hornig, N. 2022. An Increasing Sense of Urgency: The Inflation Perception Indicator (IPI) to 30 June 2022 – a Research Note. *DoCMA Working Paper*, #12.

Nguyen, V.-A.; Boyd-Graber, J.; and Resnik, P. 2014. Sometimes Average is Best: The Importance of Averaging for Prediction using MCMC Inference in Topic Modeling. In *Proceedings of the 2014 EMNLP-Conference*, 1752–1757. ACL.

Pfeiffer, J.; Rücklé, A.; Poth, C.; Kamath, A.; Vulić, I.; Ruder, S.; Cho, K.; and Gurevych, I. 2020. AdapterHub: A Framework for Adapting Transformers. In *Proceedings of the 2020 EMNLP-Conference: Systems Demonstrations*, 46–54. ACL.

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

Rieger, J.; Jentsch, C.; and Rahnenführer, J. 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*, 2337–2347. ACL.

Rieger, J.; Jentsch, C.; and Rahnenführer, J. 2022. LDAPrototype: A Model Selection Algorithm to Improve Reliability of Latent Dirichlet Allocation. *Preprint available at Research Square*.

Rieger, J.; Lange, K.-R.; Flossdorf, J.; and Jentsch, C. 2022. Dynamic change detection in topics based on rolling LDAs.

In *Proceedings of the Text2Story'22 Workshop*, volume 3117 of *CEUR-WS*, 5–13.

Schofield, A.; Magnusson, M.; and Mimno, D. 2017. Pulling Out the Stops: Rethinking Stopword Removal for Topic Models. In *Proceedings of the 15th EACL-Conference, Volume 2: Short Papers*, 432–436. ACL.

Schofield, A.; and Mimno, D. 2016. Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics*, 4: 287–300.

Shrub, Y.; Rieger, J.; Müller, H.; and Jentsch, C. 2022. Text Data Rule – Don't They? A Study on the (Additional) Information of Handelsblatt Data for Nowcasting German GDP in Comparison to Established Economic Indicators. *Ruhr Economic Papers*, #964.

Strubell, E.; Ganesh, A.; and McCallum, A. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. Florence, Italy: Association for Computational Linguistics.

Thorsrud, L. A. 2020. Words are the New Numbers: A Newsy Coincident Index of the Business Cycle. *Journal of Business & Economic Statistics*, 38(2): 393–409.

Wang, C.; Blei, D. M.; and Heckerman, D. 2008. Continuous Time Dynamic Topic Models. In *Proceedings of the 24th UAI-Conference*, 579–586. AUAI.

Wang, X.; and McCallum, A. 2006. Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends. In *Proceedings of the 12th SIGKDD-Conference*, 424–433. ACM.

Wang, Y.; Agichtein, E.; and Benzi, M. 2012. TM-LDA: Efficient Online Modeling of Latent Topic Transitions in Social Media. In *Proceedings of the 18th SIGKDD-Conference*, 123–131. ACM.

Zhai, K.; and Boyd-Graber, J. 2013. Online Latent Dirichlet Allocation with Infinite Vocabulary. In *Proceedings of the 30th ICML-Conference*, Proceedings of Machine Learning Research, 561–569. PMLR.

This article is accepted for publication via
Third International Workshop on Modelling Uncertainty in the Financial World (MUFin'23) —
https://sites.google.com/view/w-mufin/ — held in conjunction with the 2023 AAAI.