Mining for Uncertainty
Finding Economic Trends in large Newspaper Corpora
“I heard the News today, oh Boy”
An updated Version of our Uncertainty Perception Indicator (UPI) – and some general thoughts on news-based economic indicators

Henrik Müller and Nico Hornig*
Economic policy uncertainty

Market-based uncertainty

- Savings rates
- Investment
- Asset prices
- Capital flows
- Rates of return
- Labour markets

Potential Growth

- Distribution of income and wealth
- Consumption patterns
- Production conditions
- Trade

Truly exogenous economic uncertainty

Source: Müller and Hornig (2020a)
How to Gauge Uncertainty?

- Since uncertainty is, by definition, an *exogenous* phenomenon economic variables, financial market data, or survey data will not do the job.

- **News-based indicators** have been applied, with the Economic Policy Uncertainty Index (EPU) being the most popular (and widely cited) example (Baker et al. 2016).
  - Rationale: Newspapers report all the news early on, particularly about politics.

- The EPU is basically a count of articles that contain a query of search terms related to uncertainty, the economy and certain areas of economic policy (fiscal, monetary, regulation, taxation).

- Researchers make *ex ante assumptions* about the sources of uncertainty, i.e. economic policy in a narrow sense.
  - *Good* because it enables Baker et al. to build a large set of national and regional indicators over long periods of time; *bad* because it filters for known unknowns, sources of uncertainty that we have experienced in the past, but not the new stuff that’s happening

- Therefore, the EPU only responds when these policy areas are effected by an uncertainty shock, that is, *rather late.*

Prof. Dr. Henrik Müller
EPU vs. UPI

- The Uncertainty Perception Indicator (UPI) uses a more open query (no specific policy areas, just uncertainty and the economy).

- Different newspapers: EPU for Germany is based on FAZ and Handelsblatt, the UPI on Süddeutsche, Handelsblatt, Welt: 2.9 million articles, applying query yields 37,000

- A topic-modelling approach (LDA, Blei et al. 2003) is used to differentiate between various sources of uncertainty.
  - We ran models for $K$-values of 6, 8, 10, 12, 14. With 14 = best results.
  - Closely related topics may be merged.

- The results of the UPI exercise highlight not only the size of shocks but also their origins as well as the interactions between different kinds of uncertainty.
Shocking Times

German EPU (Index, rhs) vs. UPI (share in corpus), 2000-2020 Q2
(source: policyuncertainty.org, Müller and Hornig, 2020b)
Results

Topic Analysis
# see Müller and Hornig (2020b, p. 13). We combine 1 and 14, 2 and 12, 7 and 13 due to their proximity. For topics highlighted in gray equivalents in Müller and Hornig (2020) can be traced.
Causes and Consequences?

UPI (shares in analysis corpus, three-month mov. av.)
(source: Müller and Hornig 2020b)
Political Uncertainties, national topics

UPI Politics (shares in analysis corpus, three-month mov. av., source: Müller and Hornig 2020b)
Political Uncertainties, international and European topics

UPI Politics (shares in analysis corpus, three-month mov. av., source: Müller and Hornig 2020b)
Where is Covid?

UPI Real Economy (shares in analysis corpus, three-month mov. av.; source: Müller and Hornig 2020b)
Our Query Seems to Work

- Comparted to the EPU our more open search terms seem to be **more sensitive** to new stuff that’s happening.

- In Feb. 2020 the UPI with the broader query shows hikes in uncertainty **earlier**.

- Still, as we have seen, the UPI is strongly driven by **political developments**, too.
  - Could be due to reporting patterns by newspapers that tend to focus on politics and the quarrels among its proponents
  - But at the same time, a lot has been happening in the political sphere over the period (international landscape, populism, Brexit, trade wars) – rise in UPI politics is plausible.

- Furthermore, additional issues become apparent, like environment, energy and climate.
Methodology

- LDA turns **content into data**.
  - Elegant way to **measure narratives** (Shiller 2017)

- Topic modelling works for uncertainty analysis (as already shown in Larsen 2017, Müller et al. 2018, Müller 2020).
  - Need to enhance our corpus by adding further papers and media.

- LDA sorts corpora into thematic clusters in an **unsupervised** fashion. This allows for a more open search.
  - Assumptions at the outset of the analysis can be relaxed (broader query viable).

- In our analysis we calculated four models for different values of $K$. Results for $K=10$ and 12 are broadly similar, but $K=12$ better suited for our purposes (Müller and Hornig 2020).

- LDA yields topics that are **clear-cut** and **interpretable**, driven by **specific events**.
Open Questions

- Will we be able to use LDA to build stable time-series?
- Will apparent causalities pass causality tests?
- Will our Uncertainty Factors and/or individual topics work as predictors of economic variables?
- Will our results be replicable for other countries?
- So, a lot of work in progress...
References


References


Vielen Dank für Ihr Interesse!