FOMC uncertainty and monetary shock identification

Péter Horváth

September 11, 2024

Abstract

This paper introduces a novel set of indices using Natural Language Processing (NLP) techniques applied to Transcripts and Staff Reports of the Federal Open Market Committee (FOMC) meetings. These indices are designed to quantify uncertainty in monetary policy decisions and are applied in two key ways: first, to augment the standard Structural Vector Autoregression (SVAR) framework, and second, within an Instrumental Variables SVAR (IV-SVAR) framework to identify interest rate shocks. Evidence from the SVAR analysis indicates that these policy uncertainty measures are significant drivers of economic activity while exerting minimal impact on financial markets or prices. However, IV-SVAR results reveal that most of these indices are weak instruments, failing to fully resolve the price puzzle.

1 Introduction

Motiváció/literature szinten bizonytalan vagyok, hogy mihez kéne tartozni. Leginkább NLP? Uncertainty? Forward Guidance (proxy)? Identifikáció?

A results preview-nak pedig empírikus eredmények híjján még nem álltam neki.

2 Measuring uncertainty disturbances of the FOMC

This section details the process of turning textual data of FOMC meetings into numerical measures. All publications of the FOMC can be retrieved from https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm - hosting documents released in the latest 5 years, and https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm - hosting documents from all preceding years dating back to 1939. I use Transcripts and Staff Reports[1] for the purposes of this paper. Transcripts provide a complete, word-for-word accounting on beliefs about the economy and motivation of policy choices behind policy decisions. Staff Reports contain the experts' analysis on the real economy and financial markets, and as such, serve as both a basis information set for each FOMC meeting, as well as information available to the public.

First, I use data from Transcripts to identify key economic concepts that are discussed during the meetings of the FOMC, and measure the share of time spent on discussing them, for which the LDA topic model of Blei et al. (2003) is an appropriate tool. Using an LDA model to identify topics in FOMC Transcripts has been done in the work of Hansen et al. (2018), which I closely follow, however, I make some changes in the text processing as well as the LDA algorithm used that I felt necessary in order to better pinpoint individual topics discussed. Using an LDA over a traditional clustering algorithm instead of assigning a categorical label to each observation in the text, the LDA assigns a distribution, and as such is more accurate in tracking the share of time spent covering each topic.

2.1 Preprocessing the textual data

Similar to Hansen et al. (2018), I retrieve the PDF of Transcripts from the above mentioned websites using a webscraper, then mine the raw text from each file. I remove headers and footers using regular expressions, then proceed to split the text int paragraphs, and construct a corpus where each paragraph is considered as a document (observation) of the corpus. While the original work uses individual interjections opposed to paragraphs, this simplification is of no concern for the purposes of this paper, as I do not aim to track on the individual decision maker level, but over the decision making body as a whole. Moreover, the paragraph split can be carried out on the Staff Reports as well.

Next,the documents are part-of-speech tagged using the UDPIPE parser of Wijffels (2021) and tokenized

---

[1]These include the Redbooks, Tealbooks, Greenbooks, Bluebooks and Beigebooks.

into unigrams, bigrams and trigrams[2]. As described in Justeson and Katz (1995), specific word patterns[3] are kept that likely correspond to distinct word collocations. Each word of each token is then converted to lowercase, stopwords[4] are removed and words are stemmed to their linguistic roots, so that words with the same meaning conjugated differently are accounted for with the same token. The stemmed words need not necessarily be traditional English words, as for example "inflation" is stemmed to "inflat," or "uncertainty" to "uncertainti."

2.2 Vocabulary selection

In selecting the appropriate vocabulary for the LDA model, I make several alterations compared to the original work of Hansen et al. (2018). Firstly, as opposed to using a minimum term-frequency approach to filter out infrequently used tokens, I use a maximum inverse-document-frequency approach.[5]. I keep unigrams with an IDF of at most 7.6, bigrams with and IDF of at most 8.3 and trigrams with an IDF of at most 9[6]. This change helps account for the time difference - and as such additional transcripts used - between the two papers.

Secondly, I leverage unigrams to further reduce the variation in the vocabulary. I manually review the list of unique entries, and filter for terms that are clearly related to economic concepts narrowing them from over 4000 to just 410. As all bigrams and trigrams contain at least one noun, this is further leveraged by filtering for tokens that contain at least one noun from the narrowed list of unigrams, narrowing the total count of unique entries from 30304 to 16178.

Thirdly, I re-calculate the TF, IDF and TF-IDF scores for each unique token and rank each group in descending order. This step helps further reduce the vocabulary by discarding words with low added information. As seen in Figure 1, I discard bigram tokens ranking lower than 7500, and trigram tokens ranking lower than 5000. Additionally, I keep only the top 200 unigrams. These relatively strict filtering criteria are implemented in order to reduce the sparsity in the resulting Document-Term Matrix, which is later fed into the LDA model.

2.3 LDA model

The standard LDA model has three parameters: the number of topics, K, the Dirichlet distribution prior on the per-document topic distribution, α and the Dirichlet distribution prior on the per-topic term distribution, η. In setting the latter two hyperparameters, I follow Hansen et al. (2018) and set them at $\alpha = 50/k$ and $\eta = 0.025$. The low η value promotes a sparse term distribution, as such topics are characterized

---

[2]Tokens consisting of one word, two words and three words respectively

[3]For bigrams this includes: adjective-noun and noun-noun; for trigrams this includes: adjective-adjective-noun, adjective-noun-noun, noun-adjective-noun, noun-noun-noun and noun-preposition-noun. Additionally only noun unigrams are retained.

[4]Words such as "the," "that," "or," "of."

[5]Term-frequency (TF) refers to the count of a token within the corpus, and inverse-document-frequency (IDF) refers to the share of documents a token appears in. I calculate TF as $tf_v = 1 + log(n_v)$, where n_v is the absolute frequency of term v, in the corpus. IDF is calculated as $idf_v = log(\frac{D}{D_v})$, where D_v is the number of documents word v appears in and D is the total number of documents in the corpus. Finally, TF-IDF scores can be calculated as the product of these two scores.

[6]These refer to a minimum of 0.05%, 0.025% and 0.0125% share of documents the token appears in respectively.
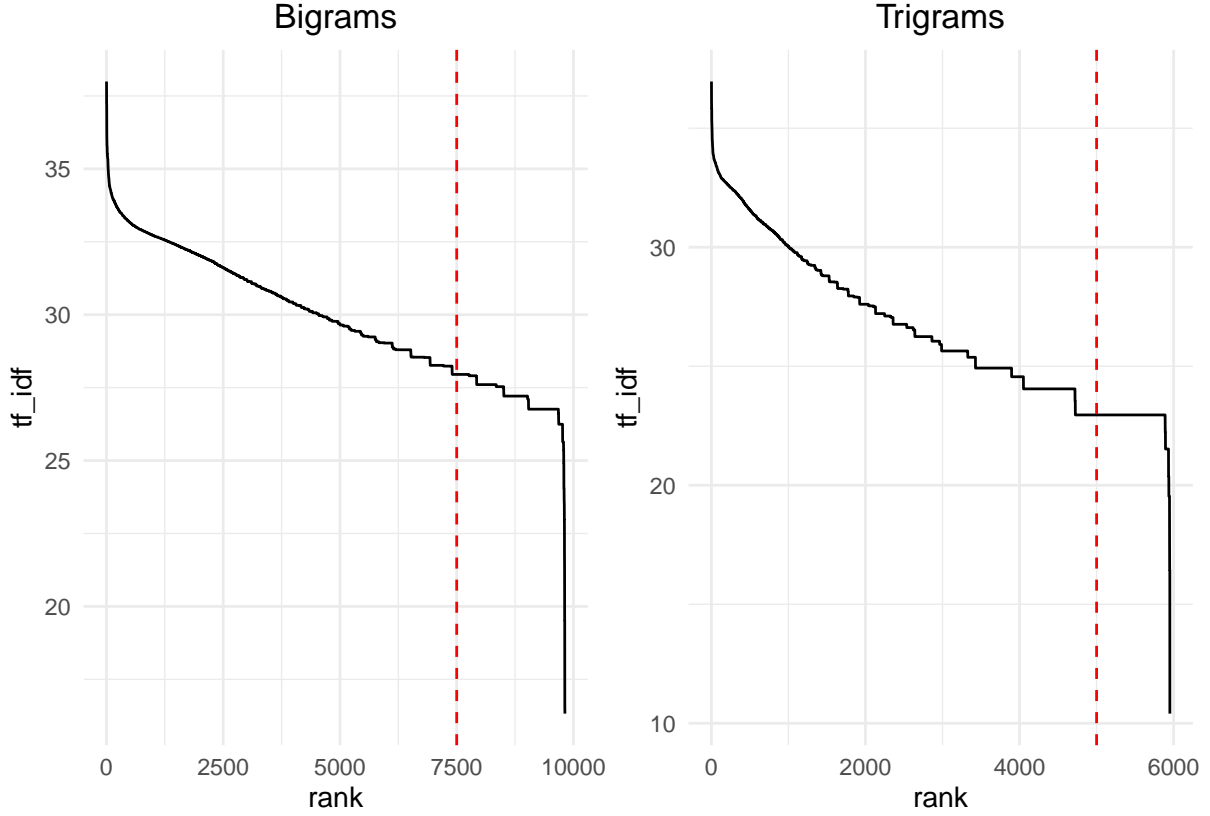
Figure 1: TF-IDF ranking of bigrams and trigrams

by a limited set of prominent words, while values of $\alpha > 1$ promote a more uniform per-document topic distribution. Selecting the optimal number of topics K however requires more consideration. An LDA model can be evaluated along Perplexity and Coherence scores[7], as well as the authors subjective evaluation on how coherent the estimated topics are. Similar to the findings of Hansen et al. (2018), I also found that a larger number of topics continuously improves Perplexity, however, Coherence scores, as well as subjective judgement would dictate to set K at a lower number of topics. As the primary goal of an LDA is to extract easily interpretable topics from a body of text, I argue for using fewer topics.

Additionally, the selection of topics can be further guided by the dual mandate of the Federal Reserve (i.e. maximum employment and stable prices), as well as byproduct goals such as promoting growth or maintaining stability in banking and financial markets. For these reasons, I set $K = 19$ and switch from the traditional unsupervised LDA model to a semi-supervised LDA using seed words. The seed words set predefined prominent words for each topic, which also helps labeling the topics. The list of topics and seed words (tokens) I use for each topic can be seen in the list below and Figure 2 shows the wordcoulds generated for each topic.

---

[7]Perplexity is calculated as Perplexity $= \exp\left(-\frac{1}{N}\sum_{d=1}^{D}\log P(v_d)\right)$, where N is the total number of words, D is the total number of documents and $P(v_d)$ is the probability of terms in document d. A lower Perplexity score is associated with better model performance. Coherence measures the the co-occurence of top words of each topic estimated by the model. Coherence[8] score is calculated as Coherence $= \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\log\frac{D(v_i,v_j)+\epsilon}{D(v_i)}$, where $D(v_i,v_j)$ is the number of documents containing both terms v_i and v_j, and $D(v_i)$ is the number of documents containing term v_i. A higher Coherence score is associated with better interpretability.

- Topic 1 - Inflation: 'term inflat expect,' 'pce inflat,' 'core cpi,' 'price pressur,' 'inflationari pressur'
- Topic 2 - GDP growth: 'real gdp growth,' 'nomin gdp,' 'output growth,' 'output gap,' 'potenti output'
- Topic 3 - Labor Market: 'employ growth,' 'unemploy gap,' 'labor market condit,' 'wage growth,' 'job growth'
- Topic 4 - Monetary Policy: 'rate cut,' 'monetari base,' 'asset purchas,' 'polici stanc,' 'forward guidanc'
- Topic 5 - Fiscal Policy: 'fiscal stimulu,' 'govern spend,' 'budget deficit,' 'tax cut,' 'govern debt'
- Topic 6 - Bank system: 'bank system,' 'deposit rate,' 'consum debt,' 'financi system,' 'busi loan'
- Topic 7 - Yield Curve: 'yield curv,' 'credit spread,' 'term premium,' 'treasuri secur,' 'treasuri yield'
- Topic 8 - Housing Market: 'commerci real estat,' 'mortgag rate,' 'properti,' 'hous price,' 'hous market'
- Topic 9 - Commodity Markets: 'crude oil,' 'food price,' 'steel,' 'commod price,' 'natur ga'
- Topic 10 - Trade: 'trade deficit,' 'import price,' 'net export,' 'current account,' 'tariff'
- Topic 11 - Exchange rates: 'foreign currenc,' 'exchang rate,' 'yen,' 'euro area,' 'forward rate'
- Topic 12 - Recessions: 'busi cycl,' 'financi crisi,' 'bubbl,' 'neg effect,' 'consider uncertainti'
- Topic 13 - Money Market: 'monei market,' 'monei suppli,' 'credit market,' 'short term rate,' 'repo'
- Topic 14 - Open Market Operations: 'monetari stimulu,' 'size balanc sheet,' 'mb,' 'cd,' 'monetari aggreg'
- Topic 15 - Financial Stability: 'default,' 'stress test,' 'financi stress,' 'financi system,' 'buffer'
- Topic 16 - Financial Markets: 'risk premium,' 'equiti price,' 'market volatil,' 'corpor bond,' 'stock market'
- Topic 17 - Consumer Confidence: 'consum sentiment,' 'retail sale,' 'econom outlook,' 'market expect,' 'person incom'
- Topic 18 - Policy Rate: 'fed fund rate,' 'natur rate,' 'real rate,' 'real interest,' 'term structur'
- Topic 19 - Domestic Supply: 'labor cost,' 'inventori invest,' 'product growth,' 'industri product,' 'capac util'

2.4 Topic distribution and uncertainty

To track the share of each topic over time, I use the LDA to estimate the posterior per document topic distribution on the Transcript of each meeting. I then repeat the preprocessing and vocabulary filtering steps outlined above using Staff Reports from each meeting, and use the LDA estimated on Transcripts to calculate the per document topic distributions on them as well. The resulting time series can be seen in Figure 3. The share of topics stays relatively stable around 5-10 percent in both Transcripts and Staff Reports, however, the former showing larger variance.

The share of discussion around topics follows real world events and series with relatively good accuracy, as for example the topic shares of inflation and policy rate closely coincide with the evolution of the

Figure 2: Wordclouds

Feredal Funds Rate; the time spent discussing asset purchases increases as interest rates were close to the Zero-Lower-Bound and Quantitative Easing measures had to be implemented;

3 Empirical Applications

Applying the measures derived above, I two schools of thought that are worth investigating. Firstly, these measures could be taken as measures of economic uncertainty, similar to the indices created by Baker et al. (2016) or Caldara and Iacoviello (2022). Modern macroeconomic literature has relied on using such indices to augment the standard SVAR framework to identify how uncertainty shocks. Results form such estimations point clearly to economic uncertainty being a key driver of aggregate demand and economic activity through influencing investment decisions. (See e.g.: Bloom (2009), Colombo (2013), Leduc and Liu (2016), Husted et al. (2020).) On the other hand, by the construction of these new measures, they are designed to reflect the beliefs of policymakers relative to an available information set. As such, they could be a useful tool in identifying interest rate shocks with an external instrument.

For both possible applications, I use the same set of variables to estimate a medium-scale VAR model, similar to the one seen in Brunnermeier et al. (2021)[9]. These variables are the Federal Funds Rate (R) to measure interest rates; PCE inflation rate (P); a chain-type industrial production index (IP) to measure

---

[9]Due to a lack of availability, spot commodity prices are omitted, additionally a standard bond spread is usend in place of the Gilchrist and Zakrajšek (2012) spread.
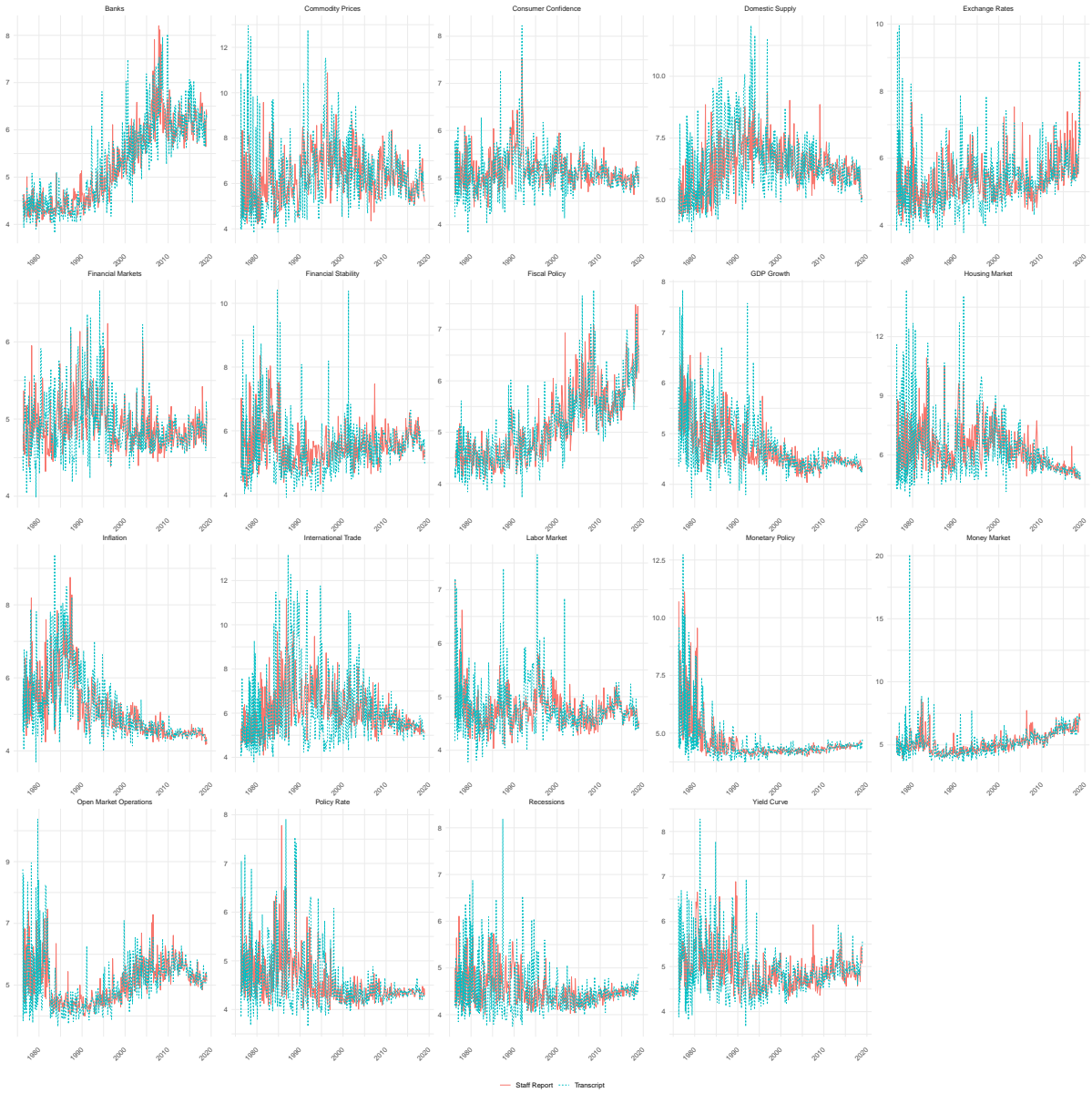
Figure 3: Distribution of Topics by publication type

economic activity; M1 monetary aggregate (M) to measure shifts in money supply; two loan volume aggregates: household loans (HL) measured as the sum of consumer and real estate loans, and business loans (BL); and lastly three financial distress measures, which are a bond spread (BS) measured as the difference between AAA and BAA rated loans, TED spread (TED) measured as the difference between 3-month Eurodollar deposit rates and 3-month treasury yields, as well as a term spread (YC) measured as the difference between yields of 10-year and 3-month treasury securities.[10]

3.1 Augmenting the standard SVAR framework

3.2 IV-SVAR application

4 Conclusions

5 Egyéb konsziderációk (Jegyzet és ötlet outlet - to be removed)

Aruoba and Drechsel (2024) egy újabb NBER-es publikáció ami text analysis alapú sokk identifikációt csinál, viszont ők csak a ColorBook-okat használják. A monetáris sokk "döntéshozási" szemponjából viszont fontos lehet a Transcriptekkel összevetni - vagyis Elemzői Vélekedés vs. Döntéshozói Vélekedés (sentiment) eltérése lehet egy kvázi sokk (vagy legalább is az identifikáció alapja).

Albrizio et al. (2023) vizsgál egy "attention to FED" hatást vállalati earningscall-okban. Ebből szép eredményeik vannak (segíti a transzmissziót, előrelátóbbak/kevésbé kitettek a sokkoknak, jobbak az inflációs várakozásaik). Erre esetleg lehetne építeni - szintén sentiment elven, FED Vélekedés vs. Szakmai News Article (jobb híjján Google News) Vélekedése.

---

[10] All data are available in the Federal Reserve Bank of St. Louis database with the following codes: R = FEDFUND; P = PCEPI; IP = INDPRO; M = M1SL; BL = BUSLOANS; HL = CONSUMER + REALLN; TED = IR3TED01USM156N - TB3MS; BS = AAA - BAA; YC = GS10 - TB3MS.

6 References

Albrizio, Silvia, Allan Dizioli, and Pedro Vitale Simon. 2023. *Mining the Gap: Extracting Firms' Inflation Expectations from Earnings Calls*. International Monetary Fund.

Aruoba, S Borağan, and Thomas Drechsel. 2024. "Identifying Monetary Policy Shocks: A Natural Language Approach." National Bureau of Economic Research.

Baker, Scott R, Nicholas Bloom, and Steven J Davis. 2016. "Measuring Economic Policy Uncertainty." *The Quarterly Journal of Economics* 131 (4): 1593–1636.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77 (3): 623–85.

Brunnermeier, Markus, Darius Palia, Karthik A Sastry, and Christopher A Sims. 2021. "Feedbacks: Financial Markets and Economic Activity." *American Economic Review* 111 (6): 1845–79.

Caldara, Dario, and Matteo Iacoviello. 2022. "Measuring Geopolitical Risk." *American Economic Review* 112 (4): 1194–1225.

Colombo, Valentina. 2013. "Economic Policy Uncertainty in the US: Does It Matter for the Euro Area?" *Economics Letters* 121 (1): 39–42.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–1720.

Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." *The Quarterly Journal of Economics* 133 (2): 801–70.

Husted, Lucas, John Rogers, and Bo Sun. 2020. "Monetary Policy Uncertainty." *Journal of Monetary Economics* 115: 20–36.

Justeson, John S, and Slava M Katz. 1995. "Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text." *Natural Language Engineering* 1 (1): 9–27.

Leduc, Sylvain, and Zheng Liu. 2016. "Uncertainty Shocks Are Aggregate Demand Shocks." *Journal of Monetary Economics* 82: 20–35.

Wijffels, Jan. 2021. "Udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the'UDPipe' 'NLP'toolkit. 2020." *R Package Version 0.8* 8.