# Business News and Business Cycles

*- Bybee et al. (2021)*

Discussant: Tao Wang (JHU)

Macro reading group at University of Pennsylvania

March 24, 2022

# Roadmap

Overview

The model

# Overview

# Text as data

Traditional tools of textual analysis used in economics/finance

- Words counts or its variants: i.e. Baker, Bloom, and Davis, 2016, Caldara and Iacoviello, 2022, etc.
- Sentiment analysis, Tetlock, 2007, Soo, 2015, Cookson and Niessner, 2020

# Text as data

Traditional tools of textual analysis used in economics/finance

- Words counts or its variants: i.e. Baker, Bloom, and Davis, 2016, Caldara and Iacoviello, 2022, etc.
- Sentiment analysis, Tetlock, 2007, Soo, 2015, Cookson and Niessner, 2020

Topic modeling in Natural Language Processing (NLP) can be used to answer the following questions

- What are the topics that are discussed in the news?
- How frequently is each topic covered in the news?
- How does the news interact with the macroeconomy and financial markets?

# This paper

Part 1: Topic modeling

1. Data: 800,000 Wall Street Journal articles for 1984–2017
2. Basic textual processing procedures
3. Latent Dirichlet Allocation (LDA)
4. Outputs of topic models

Part 2: Economic implications

1. correlation with economic activities
2. augment macro VAR with news attention
3. narrative retrieval

# Outputs

- Taxonomy of news topics
- Each article’s topic weights
- Topic-specific attention over time

# Roadmap

Overview

The model

# The model

# How to represent text data

- The bag-of-words approach
- $w$: the article-term matrix, sized of $T \times V$
- significantly lower yet still very high-dimension
- Very sparse

| | Word 1 | Word 2 | ... | Word $V$ |
|---|---|---|---|---|
| Article 1 | 1 | 0 | 0 | 0 |
| Article 2 | 0 | 0 | 0 | 0 |
| ... | 0 | 0 | 1 | 0 |
| Article T | 3 | 0 | 1 | 1 |

... Word can be n-gram

# LDA model (Blei, Ng, and Jordan, 2003)

- A further dimension reduction of $w$ via
  - explicit parametric assumptions on the distribution of word counts
    - each topic is a distribution of word counts
  - a factor structure on word counts
    - each article consists of a mixture of topics

$$\underbrace{w_t}^{V\times 1} \sim \text{Mult}\left(\underbrace{\Phi'}^{V\times K}, \underbrace{\theta_t}^{K\times 1}, \underbrace{N_t}\right)$$

word dist of article $t$ | topic-word dist | topic dist | Nb of words

- Dimension reduction via $K << V$
  - $T \times V \rightarrow T \times K$ (the size of $\theta$) + $K \times V$ (the size of $\Phi$)

# More intuition

*How to "write" an article $t$ with $N_t$ words*

- For the first word $x_{1,t}$, draw a topic from $\theta_t$

$$z_{t,i} \sim \text{Mult}(\theta_t, 1)$$

- Then draw a word from $\Phi_k$

$$x_{t,i} \sim \text{Mult}(\phi_{z_{t,i}}, 1)$$

- Repeat these two steps for $N_t$ times

# What is missing?

- Words are independently distributed given topic
- Sequence of the words does not matter
- Context-free, beyond those captured by n-grams
  - i.e. Federal Reserve

# Estimation

- In theory, MLE can be used, but computationally unrealistic
- In practice, Bayesian methods
  - Open-source package exists, for instance gensim and Nltk

# Potential extensions in economic applications

- From topics to narratives
  - Narrative as a semantic structure, i.e. *who* does *what* to *whom* (Ash, Gauthier, and Widmer, 2021)
  - Narrative as a causal model, a la Directed Acyclic Graph (DAG)

# References I


Ash, Elliott, Germain Gauthier, and Philine Widmer (2021). "Text semantics capture political and economic narratives". *arXiv preprint arXiv:2108.01720*.

Baker, Scott R, Nicholas Bloom, and Steven J Davis (2016). "Measuring economic policy uncertainty". *The quarterly journal of economics* 131.4, pp. 1593–1636.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". *Journal of machine Learning research* 3.Jan, pp. 993–1022.

Bybee, Leland et al. (2021). *Business news and business cycles*. Tech. rep. National Bureau of Economic Research.

Caldara, Dario and Matteo Iacoviello (2022). "Measuring Geopolitical Risk". *American Economic Review*.

# References II


Cookson, J Anthony and Marina Niessner (2020). "Why don't we agree? Evidence from a social network of investors". *The Journal of Finance* 75.1, pp. 173–228.

Soo, Cindy (2015). "Quantifying animal spirits: news media and sentiment in the housing market". *Ross School of Business Paper* 1200.

Tetlock, Paul C (2007). "Giving content to investor sentiment: The role of media in the stock market". *The Journal of finance* 62.3, pp. 1139–1168.