

Nowcasting GDP in Real-Time with a Tone-Adjusted, Time-Varying Layered Topic Model*

Dorinth van Dijk **Jasper de Winter**

De Nederlandsche Bank
Economic Policy & Research
Econometrics & Modelling Department

IAAE 2023, June 30, 2023

* Views expressed are those of the authors and do not necessarily reflect the position of De Nederlandsche Bank.

DeNederlandscheBank

EUROSYSTEM

Idea

- Can we use **newspaper articles** to track the business cycle and nowcast GDP growth?
- Extract **topics** from newspaper articles using unsupervised machine-learning model
- Extract **sentiment** using lexicon-based method
- Combine **topics** and **sentiment** in **tone-adjusted time-varying news topics**

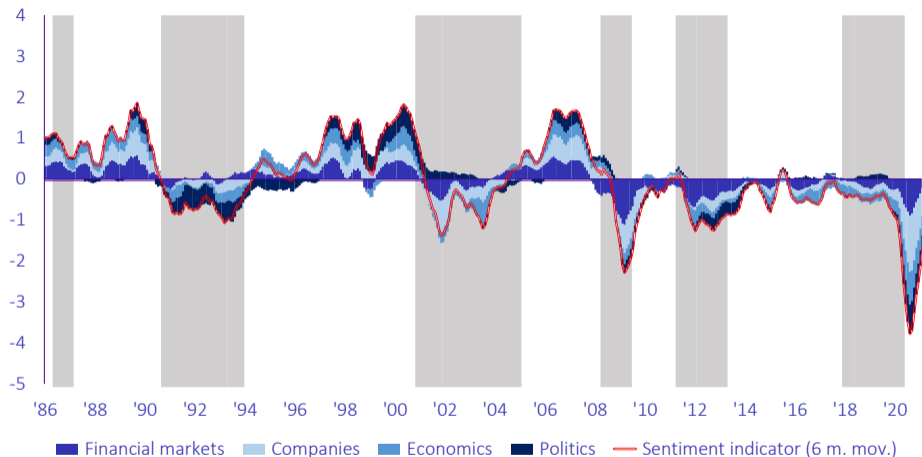
Motivation

- Understand what drives business cycle fluctuations
- Nowcast current pace of economic growth to have point of departure medium term projections

Main contributions

- Analyze unique new source of Dutch newspaper texts
- Extend tone-adjusted topicmodel (LDA) with **time-variation** and **layering** in topics
- Analyze forecasting quality of tone-adjusted time-varying news topics in nowcasting model

Main idea paper...



OECD recessions in grey

- 1 **Newspaper sentiment** is a good **coincident indicator** of the business cycle
- 2 **Tone-adjusted time varying layered topics** add “**story-telling**” layer to newspaper sentiment
- 3 **Tone-adjusted topics** embody information **not captured in other monthly indicators** during periods of high volatility
- 4 Use a **parsimonious** topic model for nowcasting purposes

- 1 Sentiment analysis
- 2 Topic Model
- 3 Nowcasting

1. Sentiment analysis

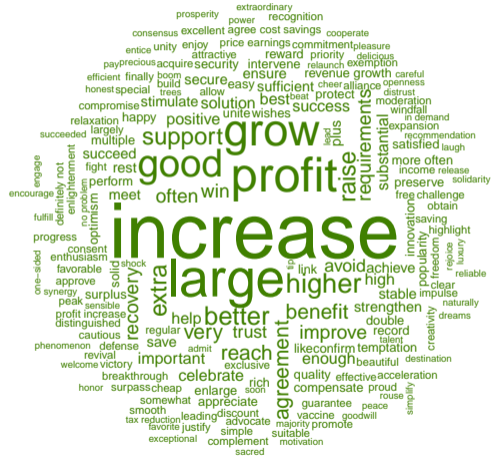
A short word on the data

- Complete **full-text archive** of Dutch “Financial Times” (Financieele Dagblad)
- Strong focus on financial markets, macro-economics and political-economic issues
- Analyzed period | 36 years | January 1st 1985–January 1st 2021
- “Standard” cleaning steps from the literature (e.g. Hansen et al., 2018 and Thorsrud, 2020)

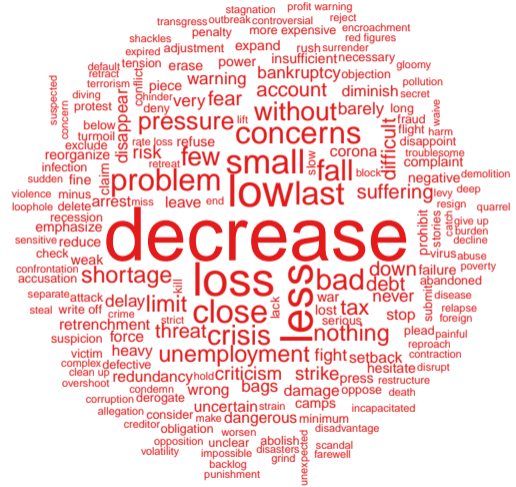
Sentiment extraction

- Customized and extended Dutch version of **Loughran and McDonald (2011)**
- Check for **double negations** i.e.: deficit decreased, unemployment decreased
- Total list: 1,532 words | Positive: 468 | Negative: 1,063
- Sentiment score **per article** (see e.g. Tetlock, 2007 and Shapiro et al., 2020):
(# positive words - # negative words)/(# words in article)

Sentiment analysis

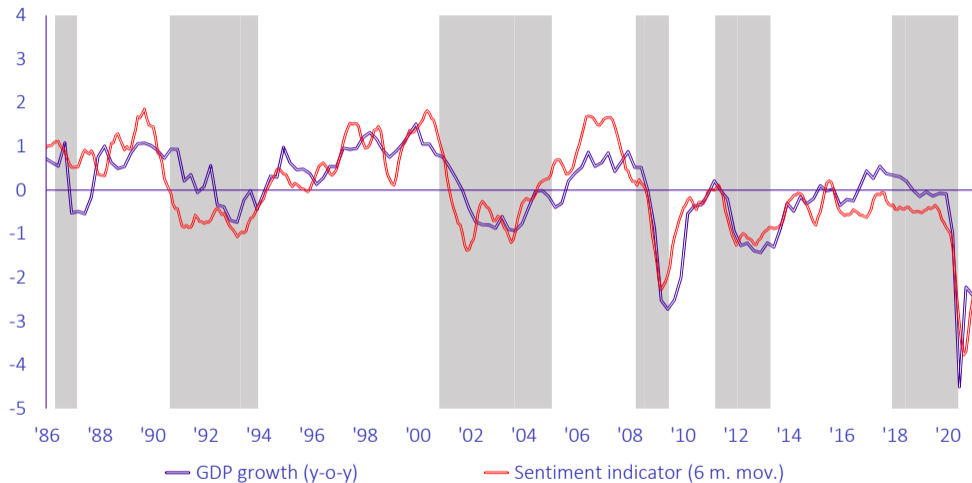


(a) Positive words



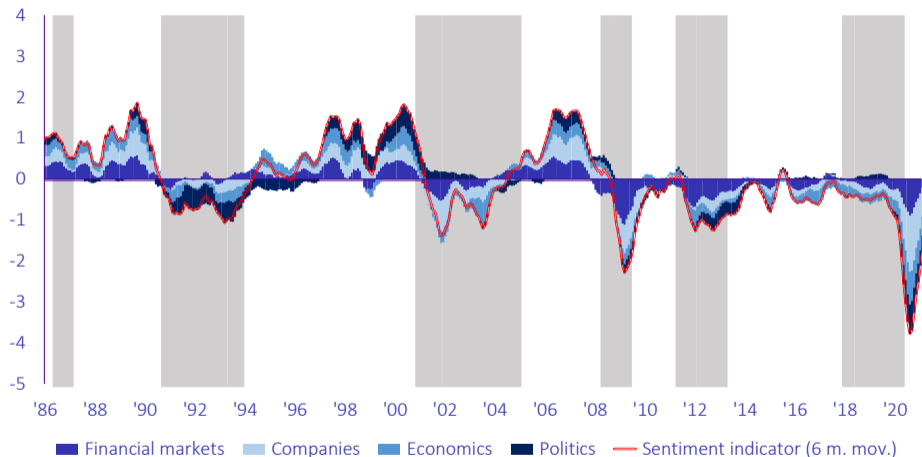
(b) Negative words

Sentiment: newspaper sentiment and GDP growth



OECD recessions in grey

... Next part of presentation: sentiment → per topic sentiment



OECD recessions in grey

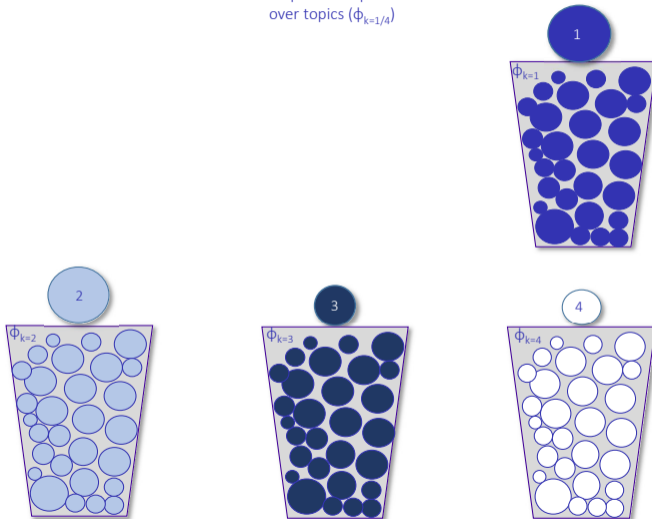
2. Topic Model

Model: intuition Latent Dirichlet Allocation



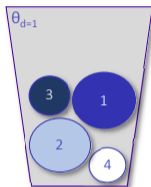
Model: intuition Latent Dirichlet Allocation

Step 1: Draw prob. distribution for words over topics ($\phi_{k=1/4}$)

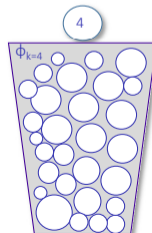
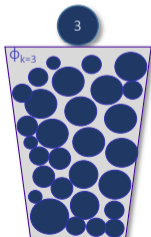
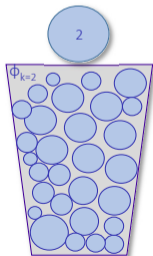
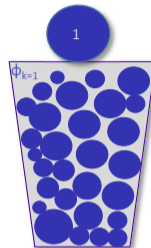


Model: intuition Latent Dirichlet Allocation

Step 2: Draw prob. distribution
for topics over document for document
 d_1 ($\theta_{d=1}$)



Step 1: Draw prob. distribution for words
over topics ($\phi_{k=1/4}$)

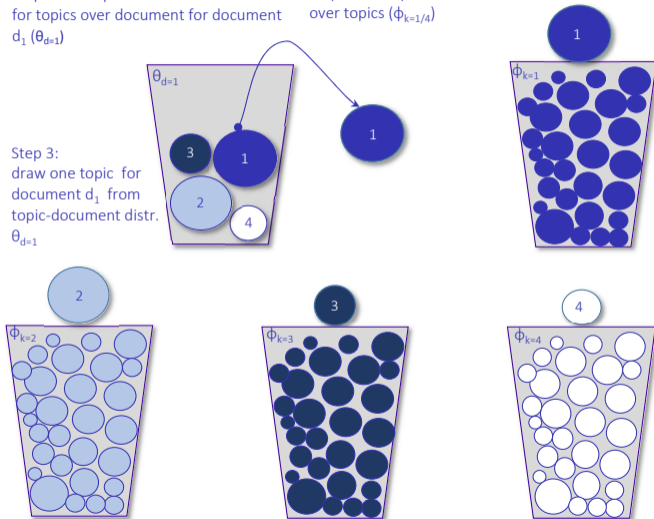


Model: intuition Latent Dirichlet Allocation

Step 2: Draw prob. distribution for topics over document for document d_1 ($\theta_{d=1}$)

Step 1: Draw prob. distribution for words over topics ($\phi_{k=1/4}$)

Step 3: draw one topic for document d_1 from topic-document distr. $\theta_{d=1}$



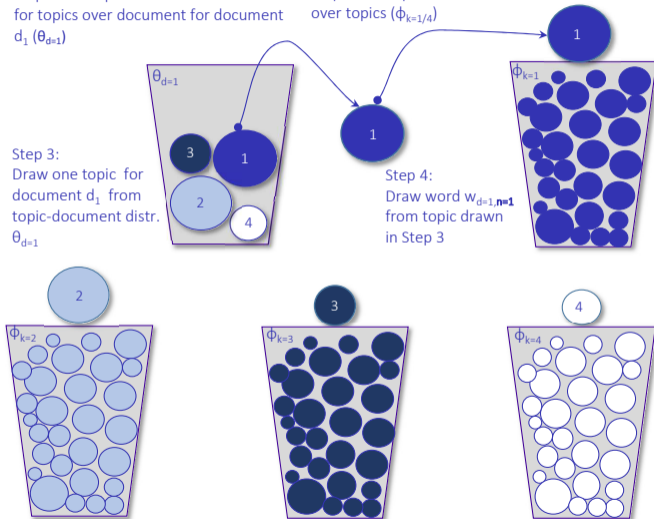
Model: intuition Latent Dirichlet Allocation

Step 2: Draw prob. distribution for topics over document for document d_1 ($\theta_{d=1}$)

Step 1: Draw prob. distribution for words over topics ($\phi_{k=1/4}$)

Step 3: Draw one topic for document d_1 from topic-document distr. $\theta_{d=1}$

Step 4: Draw word $w_{d=1,n=1}$ from topic drawn in Step 3



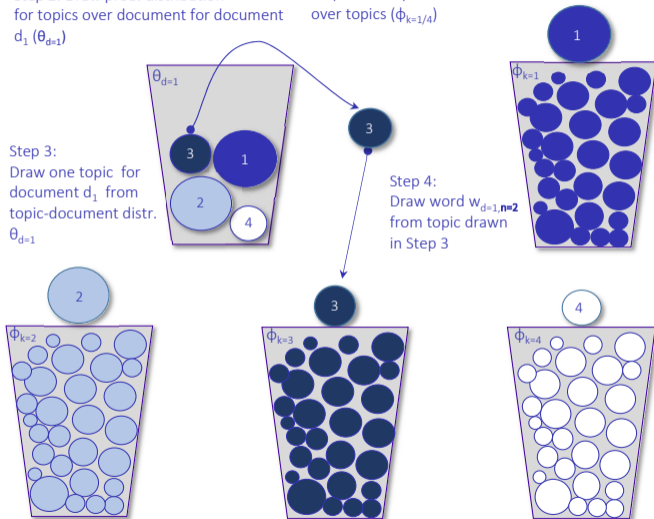
Model: intuition Latent Dirichlet Allocation

Step 2: Draw prob. distribution for topics over document for document d_1 ($\theta_{d=1}$)

Step 1: Draw prob. distribution for words over topics ($\phi_{k=1/4}$)

Step 3: Draw one topic for document d_1 from topic-document distr. $\theta_{d=1}$

Step 4: Draw word $w_{d=1, n=2}$ from topic drawn in Step 3



Repeat steps 3 and 4 until document d_1 is filled with N_1 words. **Repeat** Steps 2/4 for $d=2 \dots d=D$

Bayesian inference of model parameters

- Based on DGP just described, derive **joint distribution** of the document-topic distributions θ_d the topic-word distribution ϕ_k and the allocation of words $w_{d,n}$ to topics k in all documents.
- Bayesian inference via **Gibbs sampling** feasible but quite costly computationally, i.e.: $Pr(\phi, \theta, \mathbf{x} | \mathbf{w}, \alpha, \beta)$;
- **Collapsed Gibbs sampling** reduces computations to $Pr(\mathbf{x} | \mathbf{w}, \alpha, \beta)$
- **Essence Collapsed Gibbs sampler (CGS)**: With each pass of the CGS assign each word based on balance between how **likely** a word is for a topic and the **dominance** of a topic in a document based on the assignment of all **other words** to topics, i.e.:

$$Pr(x_{i,n} = K | \mathbf{x}_{-i}, w_i, d_i, \cdot) \propto \text{“likeliness”} \times \text{“dominance”}$$

Intuition layering



Intuition layering

- **Imposed hierarchy**, different from hierarchical topic models where hierarchy is based on correlations (Griffiths et al., 2003).
- In this paper
 - Estimate 4 topics in first layer with (“largest doll”)
 - Estimate 4 topics in second layer for **each of the 4 topic** in the first layer (“second largest doll”)
 - Estimate 4 topics in third layer for **each of the 16 topic** in the second layer (“third largest doll”)
 - **Estimation**: Use **words** assigned to topic predictive document-topic distribution ($\hat{\theta}_d$) of topic model in previous layer in slice t . Use random initialization using the Dirichlet priors for ($\hat{\theta}_d$) and ($\hat{\phi}_k$)

Model: Extension layering

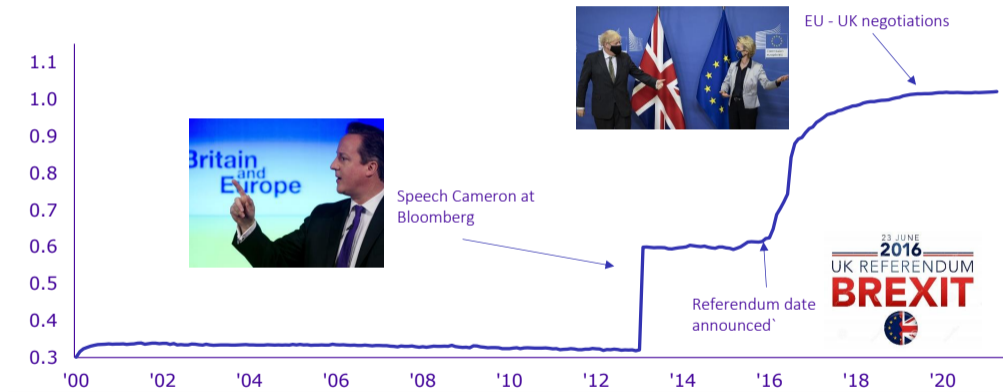
Layer 1	Financial Markets	Firms	Economics	Politics
Layer 2	Markets	Infrastructure	Elections	Parliament
Layer 3	1. Raw materials 2. Exchanges 3. International 4. Monetary Policy	17. Chemical & Pharma 18. Indices 19. Mobility 20. Company Results	33. Elections 34. Eastern Europe 35. Africa & Asia 36. United States	49. Politics 50. Budgetary policy 51. Cabinets 52. Ministries
Layer 2	Financials	Multinationals	Indicators	National
Layer 3	5. Corporate Finance 6. Financials 7. Banks 8. Insurance	21. Telecom 22. Customers 23. Big Tech 24. Media	37. International 38. Europe 39. Trading Partners 40. Fiscal Policy	53. Justice 54. Pensions & Healthcare 55. Supervision 56. Education & Research
Layer 2	News	Construction & Energy	Raw Materials	Lower Government
Layer 3	9. Emissions 10. Takeovers 11. Trade 12. Insurers	25. Construction 26. Logistics 27. Energy 28. Industry	41. Asia 42. Oil & gas 43. Conflicts 44. Emerging Markets	57. Housing 58. Public-private 59. Agriculture & Fishery 60. Transport
Layer 2	Fin. Indices	Demography	European Union	Social Partners
Layer 3	13. Stock Markets 14. Euronext 15. Analysts 16. Results	29. Retail 30. Bankruptcies 31. Listed 32. International	45. Germany 46. European Union 47. Italy & Spain 48. France	61. Wage Negotiations 62. Labor Market 63. Entrepreneurs 64. Social Security & Pensions

Time variation enables trends in topics

- Standard LDA models will underestimate trends because the model is estimated over the whole time period 1985 – 2020. E.g. word “Brexit” has very low “likeliness” when measured over total sample but very high “likeliness” when measured in last time-slice
- **Difference with other dynamic topic models:** We do **not** impose word-dynamics as in dynamic topic models but let the model decide (see e.g. Blei and Lafferty, 2006 and Bitterman and Rieger ,2022 for other approaches)
- **Fixed vocabulary** of 2,153 words. time-invariant vocabulary based on total sample, so words can slowly gain importance in topics by appearing in **time slices** (15 year estimation window)
- **Estimation:** Use vocabulary of topic predictive document-topic distribution ($\hat{\theta}_d$) and topic-word distribution ($\hat{\phi}_k$) in time slice t to initialize collapsed Gibbs sampler in time slice $t + 1$, repeat for time slice $[t + 2, \dots, T]$
- **Stability of topic model is crucial:** We check/correct topic stability with Cosine Distance (Newman et al., 2010 and Aletras & Stevenson, 2014)

Model: Extension time-variation

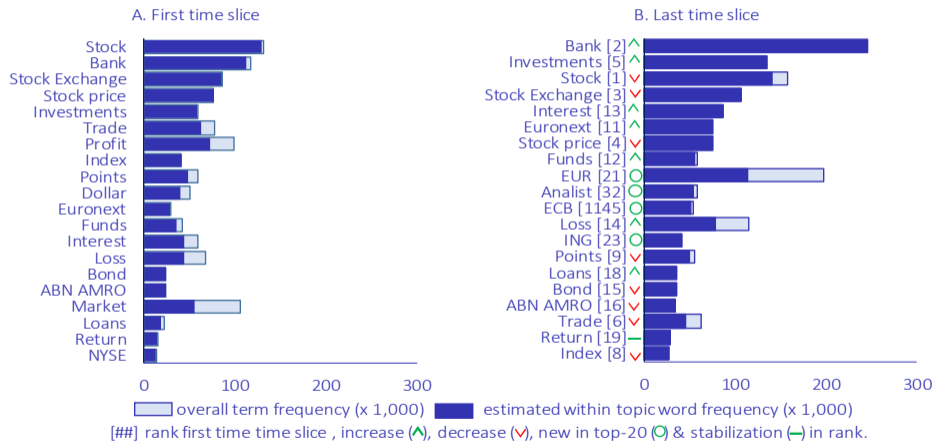
Time variation in relevance "Brexit", within topic "Economics"



Higher score is more relevant, lambda = 0.6

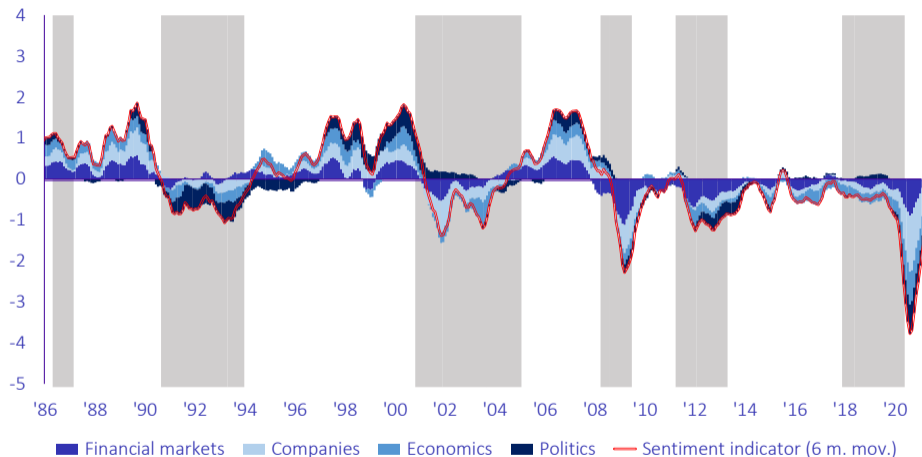
Model: Extension time-variation

Time variation in relevance top-20 words, within topic “Financial Markets”



▶ Other topics

Model: newspaper sentiment per topic



OECD recessions in grey

3. Nowcasting

Nowcasting horse-race

- **Pseudo real-time exercise:** using short-term forecasts of dynamic factor model **with(out) tone-adjusted topics** to see if newspaper sentiment adds to forecast accuracy;
- **Estimation Quasi maximum-likelihood** in a two-step procedure (Doz et al, 2012), taking into account **differences in frequencies** (GDP: quarterly, indicators: monthly) and **publication delays**. All series start in 1996M1
- **Evaluation:** 2003Q1–2020Q3 | 70 monthly macro-economic indicators | (max) 64 newspaper sentiment indicators
- Difference in forecasting accuracy are **meaningful** if two measures are satisfied (Jansen et al., 2018, Hindrayanto et al., 2016): **Statistical** significance (one sided Diebold Mariano, 1995) and **Economic** significance (RMSFE model/RMSFE benchmark ≤ 0.95)

Total sample

- Tone-adjusted topics **increase the forecasting accuracy** of the dynamic factor model
- Particularly when it pertains to the **current or previous quarter**

Total sample

- Tone-adjusted topics **increase the forecasting accuracy** of the dynamic factor model
- Particularly when it pertains to the **current or previous quarter**

Outcomes highly sensitive to crisis quarters (financial crisis & COVID crisis)

- Excluding **crisis quarters** strongly reduces the gain in forecasting accuracy
- **Flip-side**: topics **really** help during periods of crises, when models matters most

Outcome nowcasting exercise in a nutshell

Total sample

- Tone-adjusted topics **increase the forecasting accuracy** of the dynamic factor model
- Particularly when it pertains to the **current or previous quarter**

Outcomes highly sensitive to crisis quarters (financial crisis & COVID crisis)

- Excluding **crisis quarters** strongly reduces the gain in forecasting accuracy
- **Flip-side**: topics **do reduce forecasting error** during periods of crises, when models matters most

Plain-vanilla topic model vs time variation & layering

- Plain-vanilla topic model is **hard to beat** in terms of nowcasting accuracy
- Therefore, **time variation and layering seem mainly important for story-telling**

▶ Details

Four main takeaways from our research

- 1 **Newspaper sentiment** is a good **coincident indicator** of the business cycle
- 2 **Tone-adjusted time varying layered topics** add “**story-telling**” layer to newspaper sentiment
- 3 **Tone-adjusted topics** embody information **not captured in other monthly indicators** during periods of high volatility
- 4 Use a **parsimonious** topic model for nowcasting purposes

Thank you for your attention!

`j.m.de.winter@dnb.nl`

Algorithm LDA collapsed Gibbs sampler

Draws from the posterior distribution $Pr(x|w)$ are obtained by sampling from

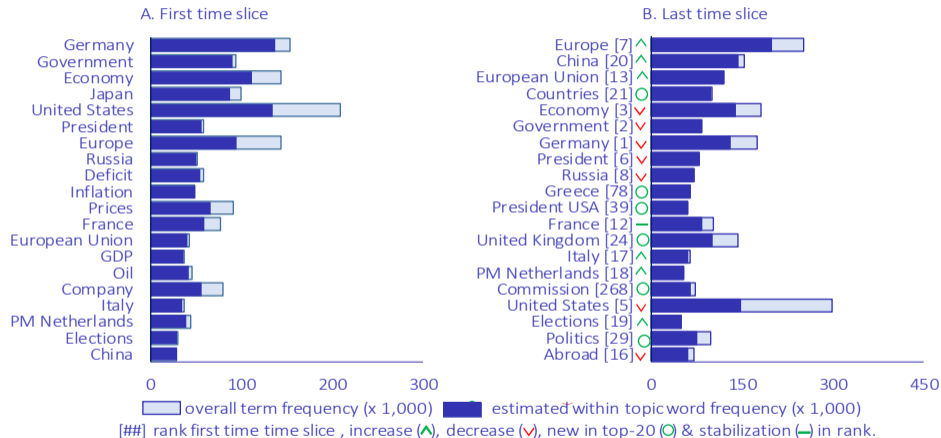
$$Pr(x_i = K | \mathbf{x}_{-i}, w_i, d_i, \cdot) \propto \underbrace{\frac{n_{-i,K}^{(j)} + \beta}{n_{-i,K}^{(\cdot)} + V\beta}}_{\text{"likeliness"}} \times \underbrace{\frac{n_{-i,K}^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + k\alpha}}_{\text{"dominance"}}$$

- V number of words in the vocabulary;
- (j) indicates w_i is equal to the j th term in the vocabulary, $j = [1 \dots, V]$;
- $n_{-i,K}^{(j)}$ freq. of the j th term assignment to topic K without the i th word;
- d_i document in the corpus to which word w_i belongs;
- \mathbf{x}_{-i} vector of current topic membership of all words without the i th word w_i ;
- (\cdot) summation over index;
- α prior of the Dirichlet distribution of the topic-word distribution (ϕ_k);
- β prior of the Dirichlet distribution of the document-topic distribution (θ_d).

One draw for all words in the corpus equals one iteration of the Gibbs sampler. Based on topic-assignments you can calculate estimated predictive document-topic distribution ($\hat{\theta}_d$) and topic-word distribution ($\hat{\phi}_k$).

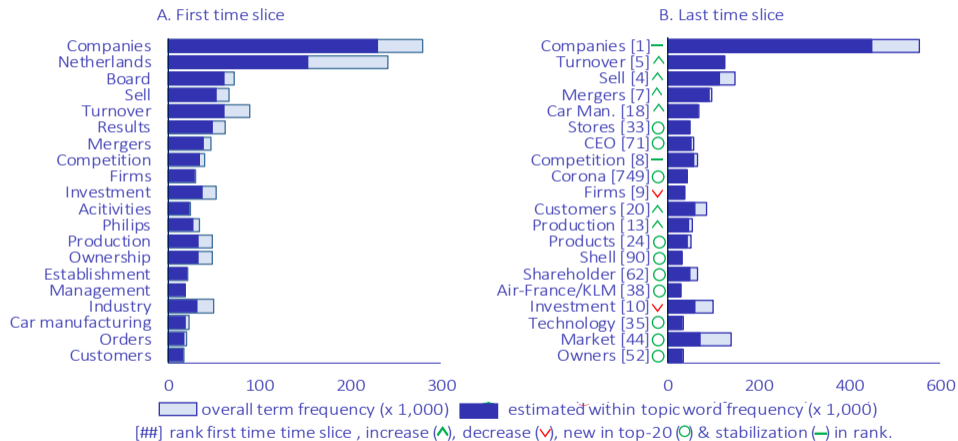
Model: Extension time-variation

Time variation in relevance top-20 words, within topic "Economics"



Model: Extension time-variation

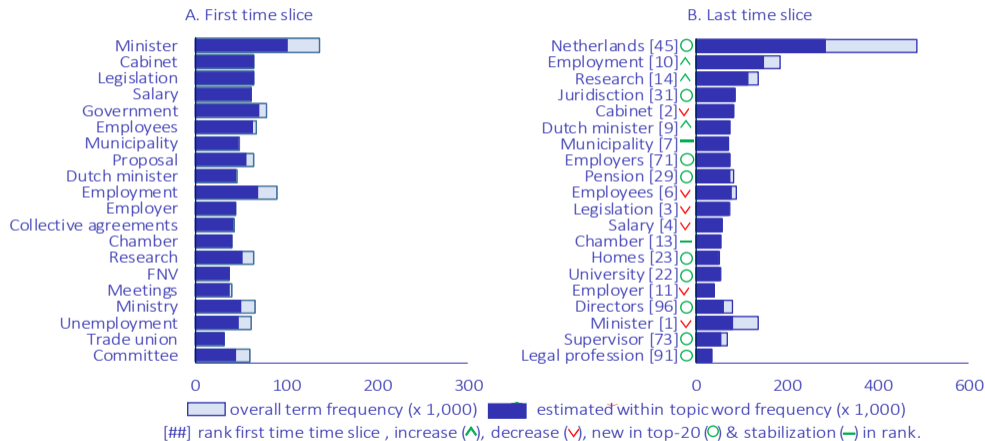
Time variation in relevance top-20 words, within topic “Firms”



Go Back

Model: Extension time-variation

Time variation in relevance top-20 words, within topic "Politics"



Go Back

Nowcasting horse-race: design

- GDP released with 45-day lag | first release Dutch GDP growth first quarter 2023 was released on May 15th
- We analyze forecast quality for 8 different forecasting horizons

Example: Naming GDP forecasts first quarter 2023

Nr.	Forecast type	Month	Forecast made on the 1 st of
1	Forecast	M1	October
2		M2	November
3		M3	December
4	Nowcast	M1	January
5		M2	February
6		M3	March
7	Backcast	M1	April
8		M2	May

◀ Go back

Nowcasting GDP growth with(out) newspaper sentiment

Total sample

- **Economic** significance (relative RMSFE ≤ 0.95), **Statistical** significance (DM-test)
- Topics increases the forecasting accuracy of the dynamic factor model
- Particularly when it pertains to current or previous quarter
- Parsimonious configuration topic model more accurate (4×4)

		DFM	DFM with tone-adjusted topics		
topicmodel type		-	TVL-LDA	TVL-LDA	TV-LDA
# topics per layer (X)		-	4 X 4 X 4	4 X 4	4
		(absolute RMSFE)	(relative RMSFE against DFM)		
Backcast	M2	0.78	0.98	0.86 *	0.94
	M1	0.95	1.11	0.95 *	0.99
Nowcast	M3	1.31	1.03	0.93 *	0.98 *
	M2	1.47	1.05	1.00	0.99 *
	M1	1.71	1.04	1.01	1.00 **
Forecast	M3	1.73	1.01	1.00 *	1.00 *
	M2	1.65	1.00	1.00	1.00
	M1	1.58	1.01	1.00	1.00

RMSFE: Root Mean Squared Forecast Error, 2003Q3-2020Q3.

RMSFE ≥ 1.05
0.95 < RMSFE < 1.05
RMSFE ≤ 0.95
 * DM-test significance

Nowcasting GDP growth with(out) newspaper sentiment

Excluding financial crisis & COVID crisis

- Some advantage but either **not** statistically or economically significant
- Value added LDA mainly attributable to crisis, when its critical

		DFM	DFM with tone-adjusted topics		
topicmodel type		-	TVL-LDA	TVL-LDA	TV-LDA
# topics per layer (X)		-	4 X 4 X 4	4 X 4	4
		(absolute RMSFE)	(relative RMSFE against DFM)		
Backcast	M2	0.53	0.93	1.01 *	0.98
	M1	0.54	0.94	1.00 *	0.99
Nowcast	M3	0.55	0.96	1.00 *	0.99 *
	M2	0.56	0.97	1.00	0.99 *
	M1	0.57	1.00	0.99	0.99 **
Forecast	M3	0.57	1.01	0.97 *	1.00 *
	M2	0.59	1.04	0.99	1.00
	M1	0.59	1.08	0.97	1.00

RMSFE: Root Mean Squared Forecast Error, 2003Q3-2020Q3, excluding 2009Q1, 2020Q2 and 2020Q3.

RMSFE ≥ 1.05
0.95 < RMSFE < 1.05
RMSFE ≤ 0.95
 * DM-test significance

Nowcasting GDP growth with(out) newspaper sentiment

Plain-vanilla vs time variation & layering

- DFM with 16 tone-adjusted topics in different configurations
- No clear advantage besides nowcasts of TVL-LDA, but not statistically significant

		DFM with tone-adjusted topics			
topicmodel type		LDA	TVL-LDA	TV-LDA	L-LDA
# topics per layer (X)		16	4 X 4	16	4 X 4
		(absolute RMSFE)	(relative RMSFE against LDA)		
Backcast	M2	0.66	1.02	1.00	1.03
	M1	0.91	0.98	0.99	1.02
Nowcast	M3	1.31	0.93	0.97	0.96
	M2	1.55	0.94	0.98	0.97
	M1	1.80	0.96	0.98	0.97
Forecast	M3	1.72	1.00	1.00	1.00
	M2	1.65	1.00	1.00	1.00
	M1	1.59	1.00	1.00	1.00

RMSFE: Root Mean Squared Forecast Error, 2003Q3-2020Q3.

■ RMSFE \geq 1.05 ■ $0.95 <$ RMSFE $<$ 1.05 ■ RMSFE \leq 0.95 * DM-test significance