Informational content of newspaper articles for business cycle analysis

Jasper de Winter[*] & Maurice Bun[*†] & Dorinth van Dijk[*‡]

De Nederlandsche Bank (DNB)[*]
University of Amsterdam[†]
Massachusetts Institute of Technology[‡]

Joint Research Centre European Commission, Ispra (IT), 20 June, 2019

Disclaimer: views expressed do not necessarily reflect official position of De Nederlandsche Bank

DeNederlandscheBank
EUROSYSTEEM

Introduction

Research question

- Can we use text data to nowcast GDP growth and predict turning points in the Dutch economy?

Motivation

- New source of information besides "hard" data and forecasts of analists;
- Several recent papers seem to be quite successful in forecasting economic variables using textual data

Main contributions

- Comparison of forecasting accuracy of pre-defined lexicon method versus unsupervised machine learning methods on novel database of large Dutch financial newspaper;
- Analyze effectiveness of machine learning methods to filter relevant newspaper articles;
- Analyze interrelatedness of news topics.

DeNederlandscheBank
EUROSYSTEEM

Outline presentation

What have we learned, where are we going?

- Nowcasting: forecasting q-o-q growth of Gross Domestic Product (GDP) for nearby quarters;
- Three recent De Nederlandsche Bank (DNB) papers on selection of nowcasting model and comparison to professional forecasters, i.e.:
  1. Best of popular linear nowcasting models (**part 1 presentation**)
     Jansen, Jin en de Winter, 2016, *International Journal of Forecasting*
  2. Nowcasting models in comparison with forecasts of professional analysts (**part 1 presentation**)
     Jansen en de Winter, 2018, *Oxford Bulletin of Economics and Statistics*
  3. Best specification of (dynamic) factor model for nowcasting
     Hindrayanto, Koopman, de Winter, 2016, *International Journal of Forecasting*

- Follow up: compare nowcasting models with new big-data sources: news-articles from financial newspaper. Preliminary results (**part 2 presentation**)

DeNederlandscheBank
EUROSYSTEEM

Part 1: Data sources

Data used

1. Monthly indicators
   - Real-time vintages for approx. 40 headline "market moving" indicators that are readily available to economic agents (e.g. Bańbura et al., 2013, Bańbura and Modugno, 2014);
   - Data on global economy (i.e commodity prices, semiconductor sales, Baltic Dry index), domestic economy (i.e. industrial production, consumer confidence, key indicators trading partners (i.e. import, exports);
2. Quarterly Gross Domestic Product
   - Real-time vintages for GDP;
3. Quarterly Forecast Professional Analysts
   - New data set constructed of paper copies of quarterly forecasts for G7-countries from Consensus Forecasts;
4. Daily newspaper articles
   - Exclusive access to digitalized datbase containing all articles in (only) financial newspaper in the Netherlands (Financieele Dagblad) over the period 1985–2018;

DeNederlandscheBank
EUROSYSTEEM

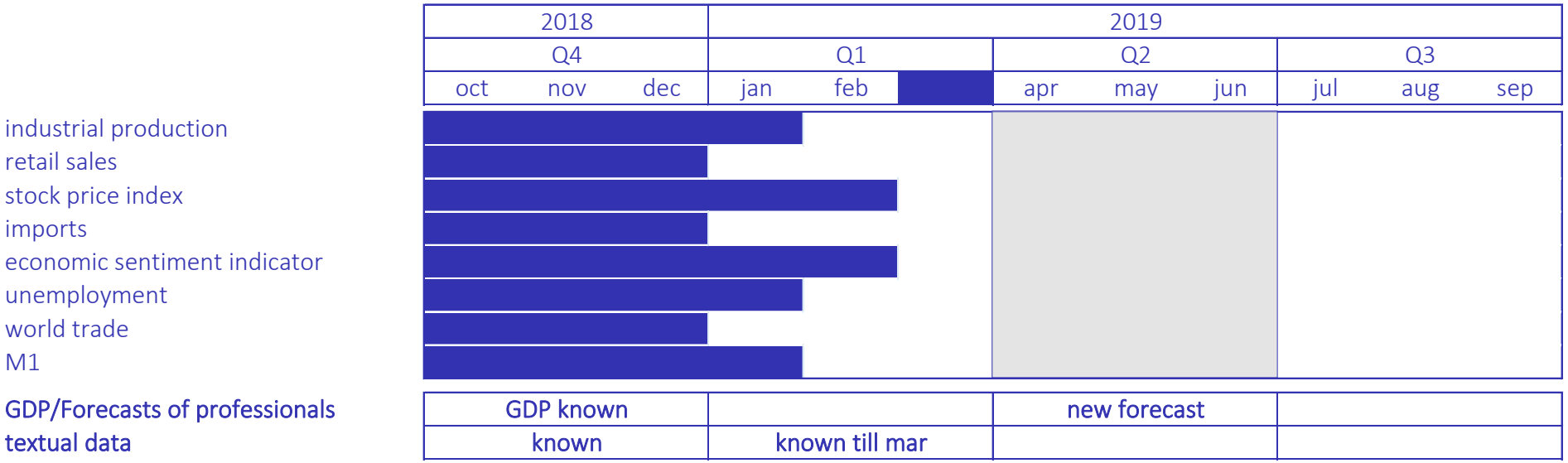
Forecasting GDP growth in the short term: three challenges

1. Large size of the monthly information set ('curse of dimensionality');
2. Indicators are observed at different frequencies
   - Daily: newspaper articles, financial markets;
   - Monthly: consumer confidence, industrial production;
   - Quarterly: gross domestic product & forecasts of professional analysts;
3. Dating of most recent observations varies per variable ('ragged edges');
   - Consumer confidence up until May 2019, *flash* June 2019;
   - Industrial production up until April 2019;
   - Consensus forecasts made at the beginning of June;
   - Newspaper articles up until June 20th.

DeNederlandscheBank
EUROSYSTEEM

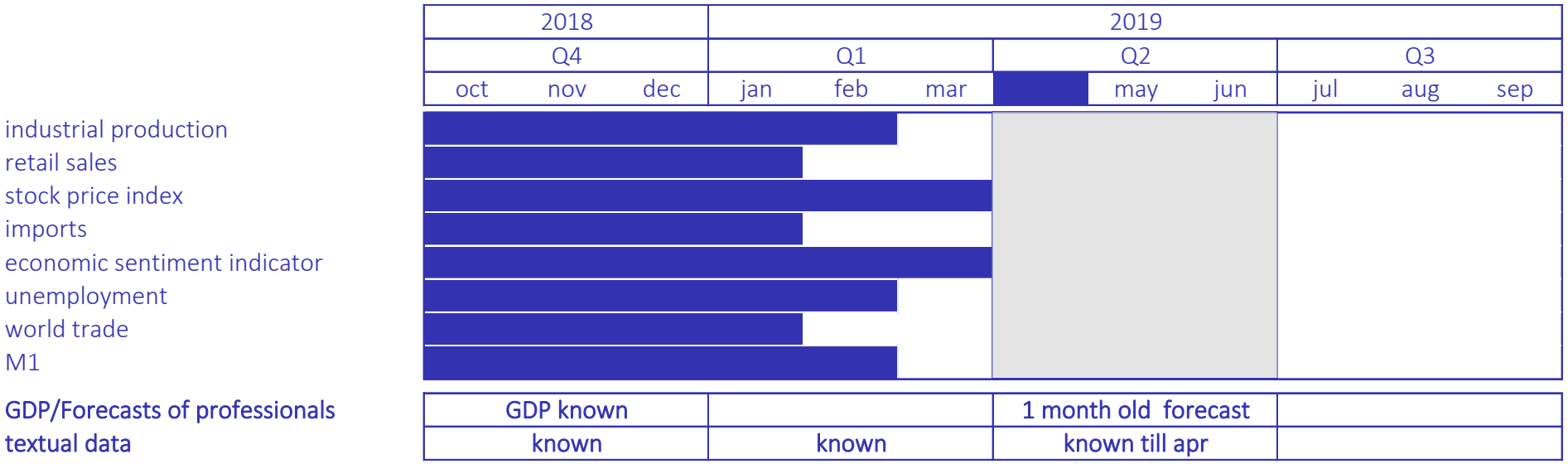
Part 1: Flow of monthly & quarterly data

Forecast, month 3 for 2019Q2



Part 1: Flow of monthly & quarterly data

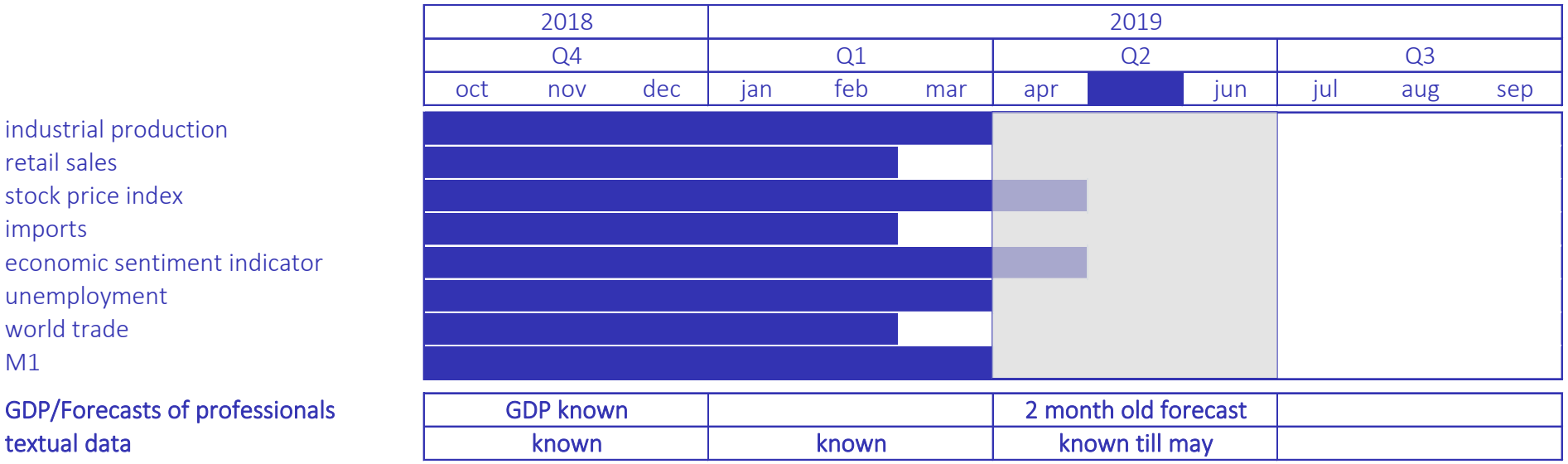
Nowcast, month 1 for 2019Q2



DeNederlandscheBank
EUROSYSTEEM

Part 1: Flow of monthly & quarterly data

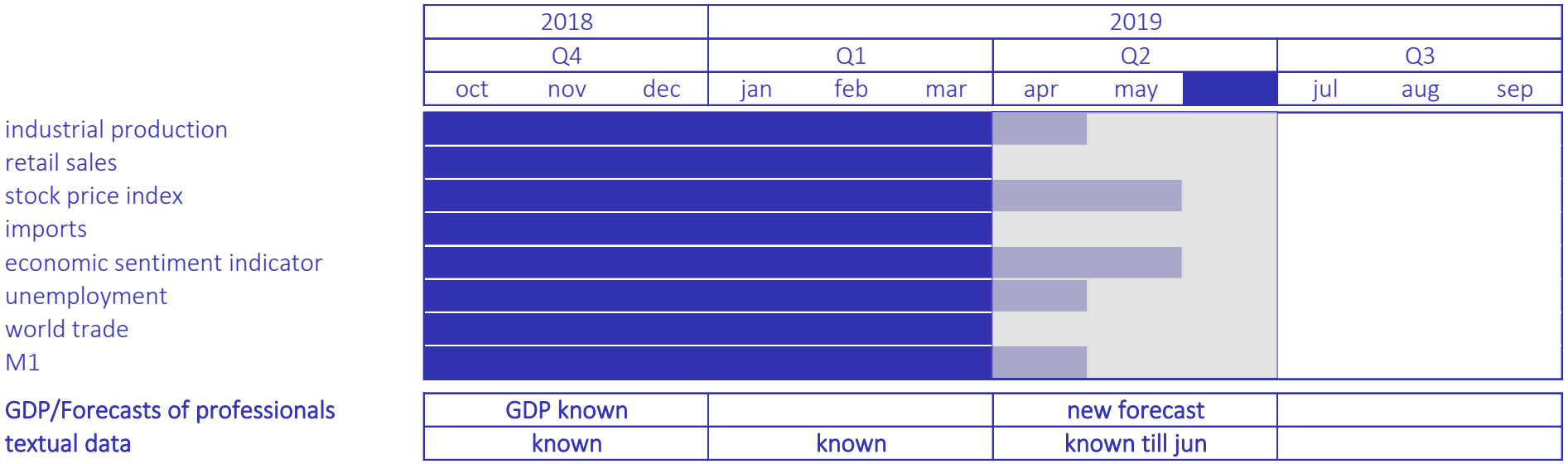
Nowcast, month 2 for 2019Q2



DeNederlandscheBank
EUROSYSTEEM

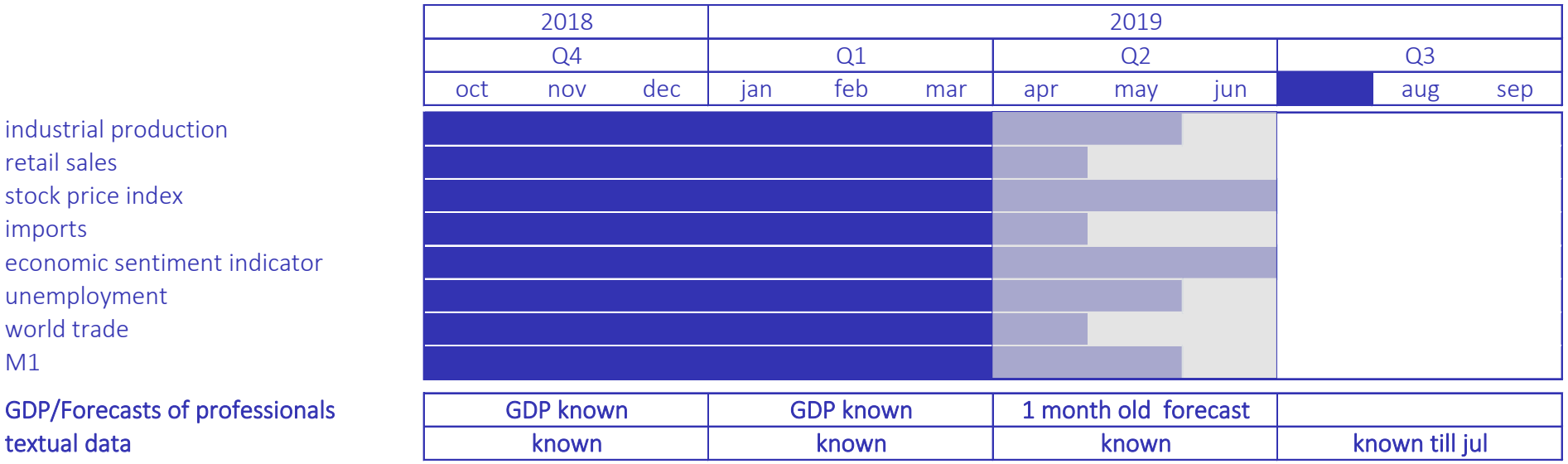
Part 1: Flow of monthly & quarterly data

Nowcast, month 3 for 2019Q2



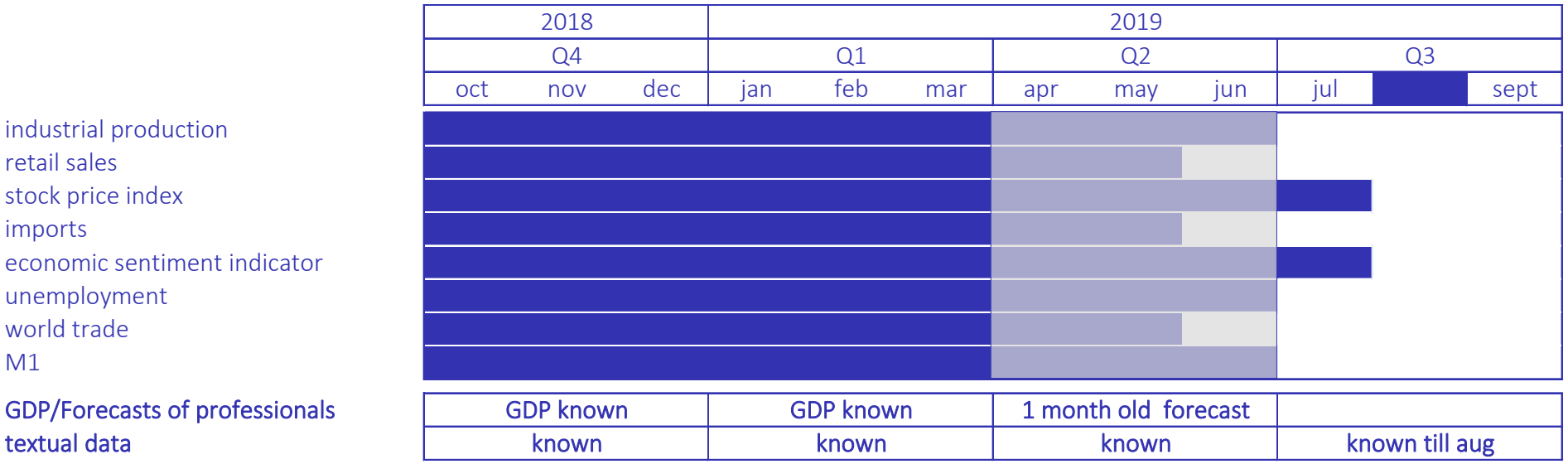
Part 1: Flow of monthly & quarterly data

Backcast, month 1 for 2019Q2



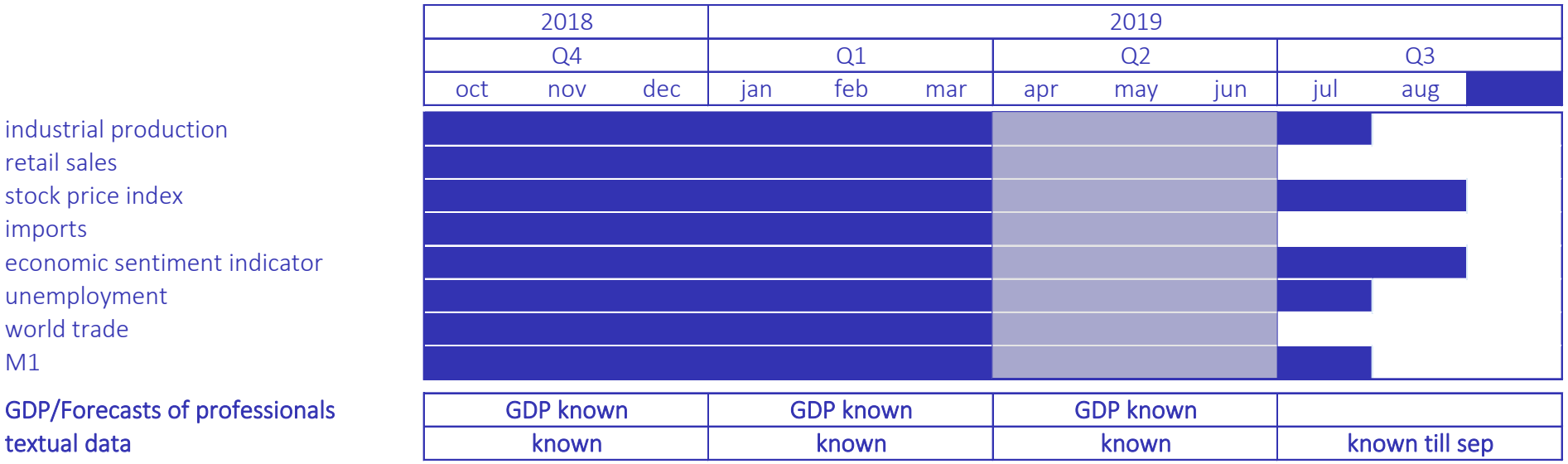
Part 1: Flow of monthly & quarterly data

Backcast, month 2 for 2019Q2



Part 1: Flow of monthly & quarterly data

Backcast, month 3 for 2019Q2



DeNederlandscheBank
EUROSYSTEEM

Summing up: timing of forecasts for GDP growth 2019Q2

Forecast type		Month
Forecast	3	March
Nowcast	1	April
	2	May
	3	June
Backcast	1	July
	2	August

DeNederlandscheBank
EUROSYSTEEM

Part 1: Which model is best in "nowcasting"?

Best nowcasting model

Horse race between suit of currently popular linear nowcasting models over the period 1996–2011, for the EA and it's five largest countries

Pseudo real-time analysis

- Models re-estimated each month taking into account flow of information;
- Root Mean Squared Forecast Error is measure of forecast accuracy;

Econometric models

- Quarterly model for GDP growth: aggregate all indicators to quarterly level and then estimate model;

  Bridge Equations: BEQ (Baffigi et al., 2004, Kitchen and Monaco, 2003)

  Quarterly Vector Autoregressive Model: QVAR (Camba-Mendez et al, 2001)

  Bayesian Vector Autoregressive Model: BVAR (Bańbura et al, 2010)

- Mixed-Frequency models: mix daily, monthly and quarterly information the estimated model;

  Dynamic Factor model: DFM (Bańbura and Rünstler, 2011)

  Mixed-Frequency Vector Autoregressive Model: MF-VAR (Kuzin et al., 2011)

  Mixed-Data Sampling Regression Model: MIDAS (Ghysels et al., 2007)

DeNederlandscheBank
EUROSYSTEEM

Part 1: Which model is best in "nowcasting"?

Main takeaways model forecasting horse-race

(Jansen,Jin, Winter de, 2016)

1. Useful to use monthly indicators, especially for nowcasting and backcasting & during volatile times;

DeNederlandscheBank
EUROSYSTEEM

Part 1: Which model is best in "nowcasting"?

Main takeaways model forecasting horse-race

(Jansen,Jin, Winter de, 2016)

1. Useful to use monthly indicators, especially for nowcasting and backcasting & during volatile times;
2. Extracting factors is a better strategy than averaging single-indicator models (aggregating versus pooling);

DeNederlandscheBank
EUROSYSTEEM

Part 1: Which model is best in "nowcasting"?

Main takeaways model forecasting horse-race

(Jansen,Jin, Winter de, 2016)

1. Useful to use monthly indicators, especially for nowcasting and backcasting & during volatile times;
2. Extracting factors is a better strategy than averaging single-indicator models (aggregating versus pooling);
3. Adding autoregressive terms to models does not matter much in terms of forecasting accuracy;

DeNederlandscheBank
EUROSYSTEEM

Part 1: Which model is best in "nowcasting"?

Main takeaways model forecasting horse-race

(Jansen,Jin, Winter de, 2016)

1. Useful to use monthly indicators, especially for nowcasting and backcasting & during volatile times;
2. Extracting factors is a better strategy than averaging single-indicator models (aggregating versus pooling);
3. Adding auto-regressive terms to models does not matter much in terms of forecasting accuracy;
4. Dynamic factor model is best model overall dues to it's ability to incorporate multiple factors;

DeNederlandscheBank
EUROSYSTEEM

Main takeaways model forecasting horse-race

(Jansen,Jin, Winter de, 2016)

1. Useful to use monthly indicators, especially for nowcasting and backcasting & during volatile times;
2. Extracting factors is a better strategy than averaging single-indicator models (aggregating versus pooling);
3. Adding auto-regressive terms to models does not matter much in terms of forecasting accuracy;
4. Dynamic factor model is best model overall dues to it's ability to incorporate multiple factors;
5. The cost of applying a sub-optimal model is highest during volatile times (e.g financial crisis);

DeNederlandscheBank
EUROSYSTEEM

Main takeaways model forecasting horse-race

(Jansen,Jin, Winter de, 2016)

1. Useful to use monthly indicators, especially for nowcasting and backcasting & during volatile times;
2. Extracting factors is a better strategy than averaging single-indicator models (aggregating versus pooling);
3. Adding auto-regressive terms to models does not matter much in terms of forecasting accuracy;
4. Dynamic factor model is best model overall dues to it's ability to incorporate multiple factors;
5. The cost of applying a sub-optimal model is highest during volatile times (e.g financial crisis);
6. The information content of different model types overlaps to a large extent: averaging is not very helpful.

DeNederlandscheBank
EUROSYSTEEM

Part 1: Mechanical models versus professional analysts

"Men versus machine"

Horse race between best nowcasting model (dynamic factor model) and professional analysts over the period 1999–2013, for the G-7 countries

Real-time analysis

- Models re-estimated each month taking into account flow of information and data-revisions;
- Root Mean Squared Forecast Error is measure of forecast accuracy;

Graphical presentation outcomes

- Averages for G-7 countries, measures of differences between countries;
- Equality of forecasting performance in "economic" and "statistical" terms: 10% rule and Diebold Mariano (1995) test.
- Forecast accuracy against the *flash* GDP release;

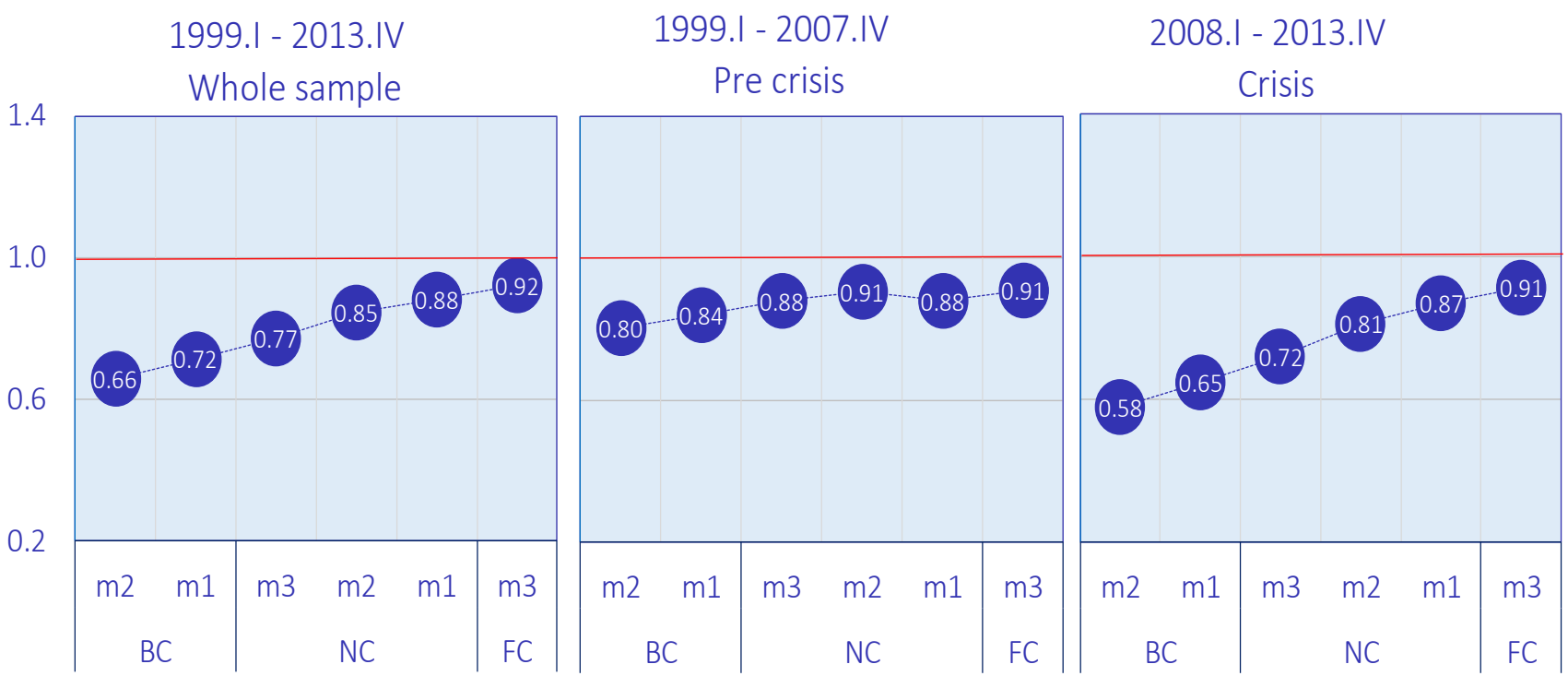
DeNederlandscheBank
EUROSYSTEEM

Remember: timing of forecasts for GDP growth 2019Q2

Forecast type		Month	
Forecast	3	March	FC M3
Nowcast	1	April	NC M1
	2	May	NC M2
	3	June	NC M3
Backcast	1	July	BC M1
	2	August	BC M2

DeNederlandscheBank
EUROSYSTEEM

Dynamic factor model vs. Random Walk

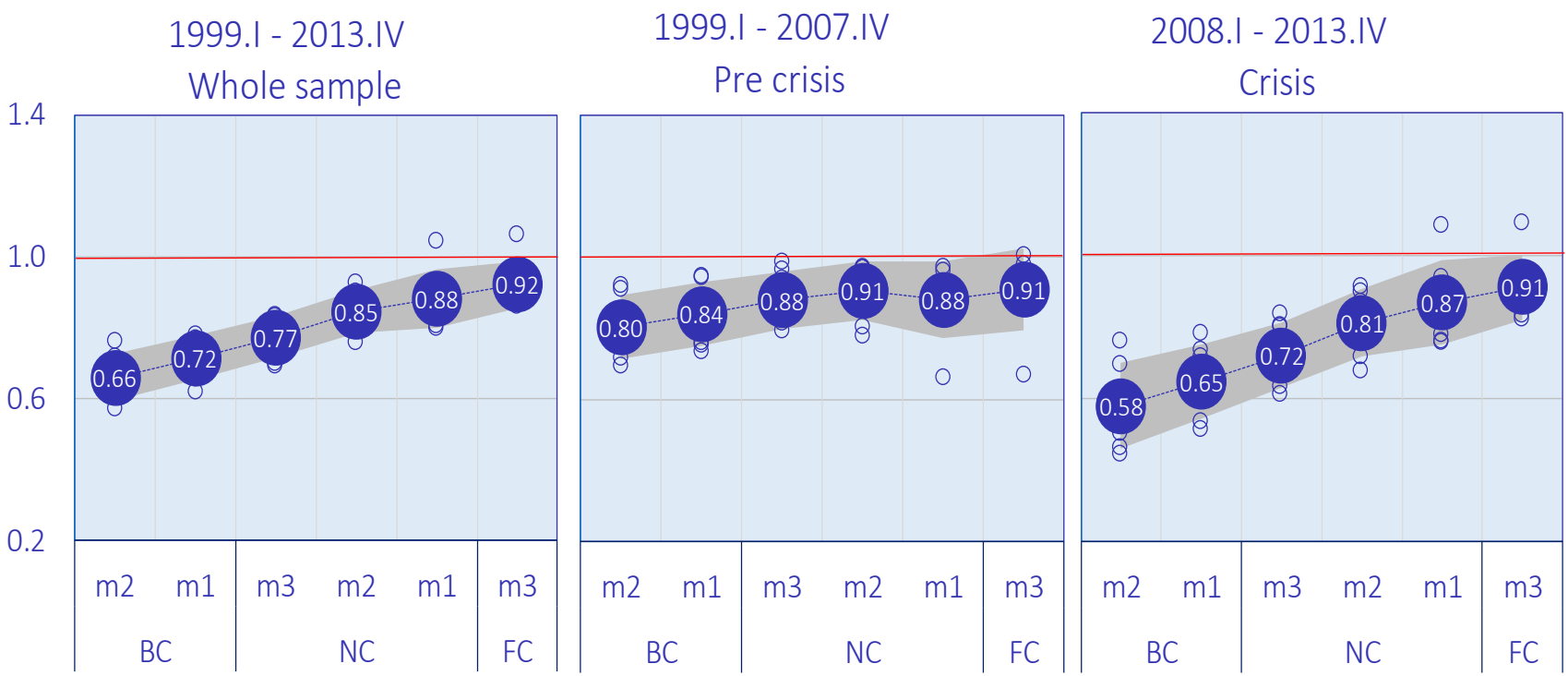


Source: DNB.
Note: RMSFE Dynamic Factor Model divided by RMSFE Random Walk; average G-7; BC = Backcast, NC=Nowcast, FC=Forecast.

1. Incorporating monthly information pays off (all rRMSFES < 1);
2. rRMSFE post-crisis $<<$ pre-crisis;
3. DFM's relatively strength is now- and backcasting.

DeNederlandscheBank
EUROSYSTEEM

Dynamic factor model vs. Random Walk

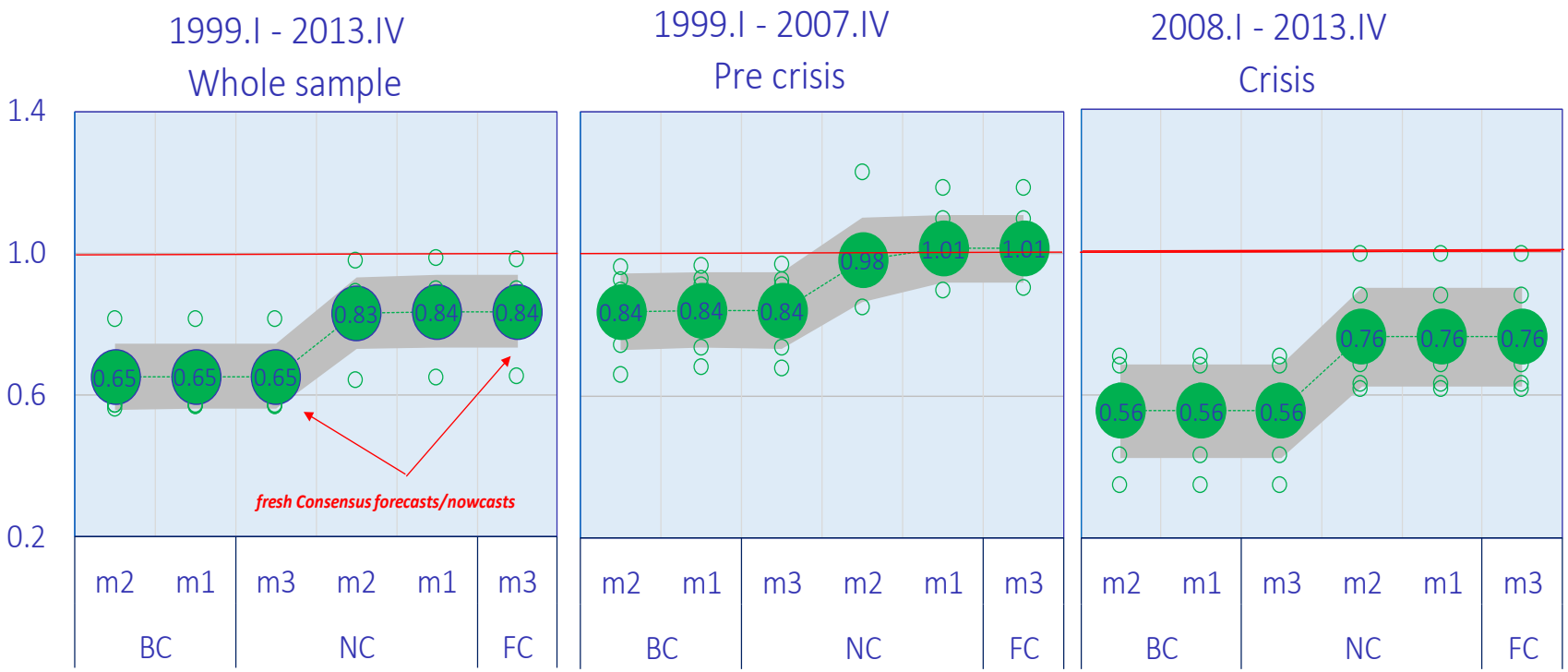


Source: DNB.
Note: RMSFE Dynamic Factor Model divided by RMSFE Random Walk; BC = Backcast, NC=Nowcast, FC=Forecast.

1. Country variability (grey area): +/- 1 standard deviation; dots are G7-countries
2. All countries relatively close together;

DeNederlandscheBank
EUROSYSTEEM

Consensus Forecasts vs. Random Walk

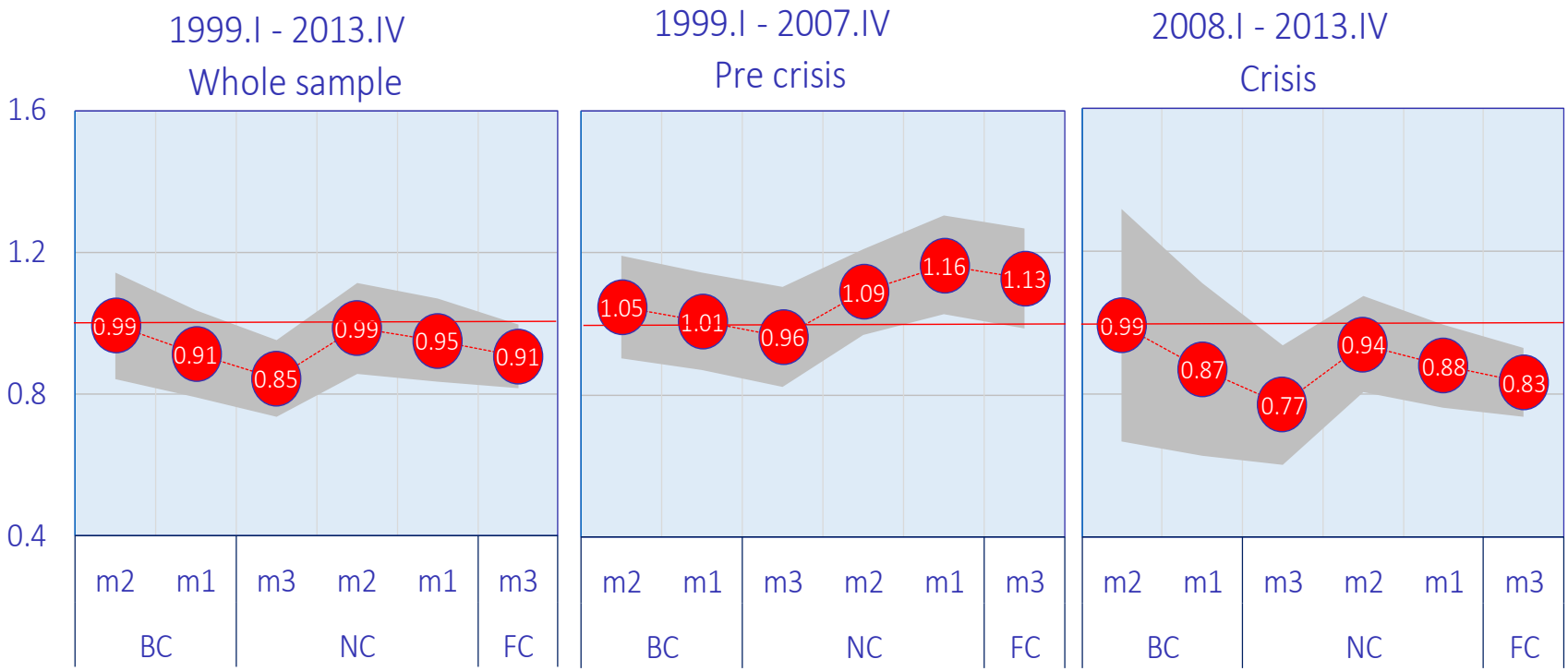


Source: DNB.
Note: RMSFE Quarterly Consensus Forecasts divided by RMSFE Random Walk; BC = Backcast, NC=Nowcast, FC=Forecast.

1. Very steep learning curves
2. Relative forecasting advantage N3>>F3

DeNederlandscheBank
EUROSYSTEEM

Consensus Forecasts vs. Dynamic Factor Model



Source: DNB.
Note: RMSFE Quarterly Consensus Forecasts divided by Dynamic Factor Model; BC = Backcast, NC=Nowcast, FC=Forecast.

1. Fresh CF nowcasts always better than DFM → driven by crisis;
2. DFM catches up in between fresh CF's → especially during crisis;

DeNederlandscheBank
EUROSYSTEEM

Part 1: Mechanical models versus professional analysts

What have we learned?

- Predictive power subjective Consensus forecasts has improved relative to predictions from the dynamic factor model over time;
- Consensus forecasts improve after the crisis, making them a tough competitor to the DFM;
- Relative forecasting CF current quarter (some information);

Combination of DFM and professional forecasters

- Enhances forecasting accuracy of the DFM, even when CF are somewhat dated;
- Analysts "bring something new to the table";

Follow up: what about the informational content of news articles?

- Professional forecasters seem to give "different" information
- How about news media/journalists? $\rightarrow$ **Part 2**

DeNederlandscheBank
EUROSYSTEEM

Part 2: (Some) Recent papers using news media

Lexicon method

- Forecast change in financial markets (Tetlock, 2007; Garcia, 2013);
- Create sentiment score for a press article based on (weighted) frequency count of pre-defined dictionary with "positive" and "negative" keywords;
- Either generic (Harvard IV dictionary as in e.g. Thorsrud, 2016) or specific for the study area (Loughran & McDonald's financial dictionary, 2011 and Baker et al. sentiment uncertainty dictionary, 2016)

DeNederlandscheBank
EUROSYSTEEM

Part 2: (Some) Recent papers using news media

Lexicon method

- Forecast change in financial markets (Tetlocjk, 2007; Garcia, 2013);
- Create sentiment score for a press article based on (weighted) frequency count of pre-defined dictionary with "positive" and "negative" keywords;
- Either generic (Harvard IV dictionary as in e.g. Thorsrud, 2016) or specific for the study area (Loughran & McDonald's financial dictionary, 2011 and Baker et al. sentiment uncertainty dictionary, 2016)

Unsupervised machine learning methods

- Change in frequency of subjects detected with Latent Dirichlet Allocation (LDA)
- Test if topics have predictive power for GDP and other macro-economic variables (Larsen & Thorsrud, JoE, 2019);

DeNederlandscheBank
EUROSYSTEEM

Part 2: (Some) Recent papers using news media

Lexicon method

- Forecast change in financial markets (Tetlocjk, 2007; Garcia, 2013);
- Create sentiment score for a press article based on (weighted) frequency count of pre-defined dictionary with "positive" and "negative" keywords;
- Either generic (Harvard IV dictionary as in e.g. Thorsrud, 2016) or specific for the study area (Loughran & McDonald's financial dictionary, 2011 and Baker et al. sentiment uncertainty dictionary, 2016)

Unsupervised machine learning methods

- Change in frequency of subjects detected with Latent Dirichlet Allocation (LDA)
- Test if topics have predictive power for GDP and other macro-economic variables (Larsen & Thorsrud, JoE, 2019);

Combination unsupervised machine learning and lexicon method

- LDA with pre-defined sentiment list to get tone-adjusted topic $\rightarrow$ extract factors $\rightarrow$ estimate dynamic factor model (Thorsrud, 2016a en 2016b) and compare to mechanical nowcasts and nowcasts of central bank;
- LDA with pre-defined sentiment list to get tone-adjusted topic $\rightarrow$ panel data regression of topics on assets returns (Thorsrud, 2016c)

DeNederlandscheBank
EUROSYSTEEM

Part 2: Can textual data be helpful for nowcasting?

Our paper

- Joined topic-sentiment index of newspaper-articles & analyze if it improves forecasts accuracy of mechanical models & professional analysts;

Main contributions to the literature

- Comparison of forecasting accuracy of pre-defined lexicon method versus unsupervised machine learning methods on novel database of large Dutch financial newspaper;
- Analyze effectiveness of machine learning methods to filter relevant newspaper articles;
- Analyze interrelatedness of news topics.

DeNederlandscheBank
EUROSYSTEEM

Part 2: Can textual data be helpful for nowcasting?

Financieele Dagblad

- Largest and only daily financial newspaper in the Netherlands;
- $\pm$ 100,000 subscriptions; mostly firms, government and universities
- Pilot project with Financieele Dagblad: exclusive to De Nederlandsche Bank;

DeNederlandscheBank
EUROSYSTEEM

Part 2: Can textual data be helpful for nowcasting?

Raw database

- All articles in Financieele Dagblad in the period January 1st 1985 - December 31st 2018;
- $\pm$ 1 million articles;

Restrict database: keep only the articles relevant for forecasting business cycle/GDP growth

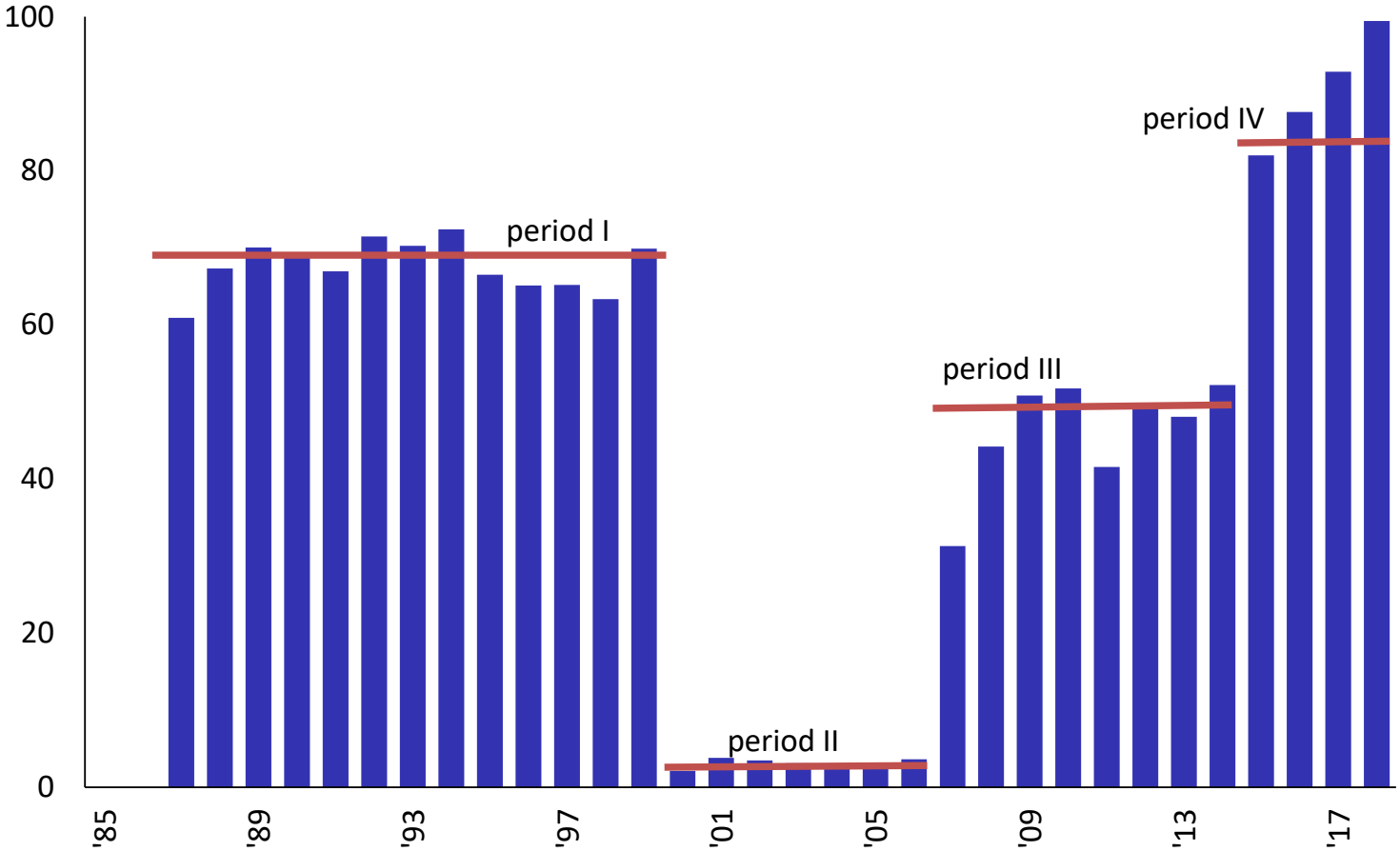
- Remove information based on pre-defined newspaper categories ,i.e. fd-private, personal finance, selections, weekend specials, background pieces, reader's letters, photo pages etc.
- All in all 350 categories removed;

Restrict database: remove and transform words

- Transform to lowercase letter, remove HTML-tags, punctuation and numbers;
- Spell check not necessary (fully digitized, no OCR);

DeNederlandscheBank
EUROSYSTEEM

First look at the data: topic tags in Fd database



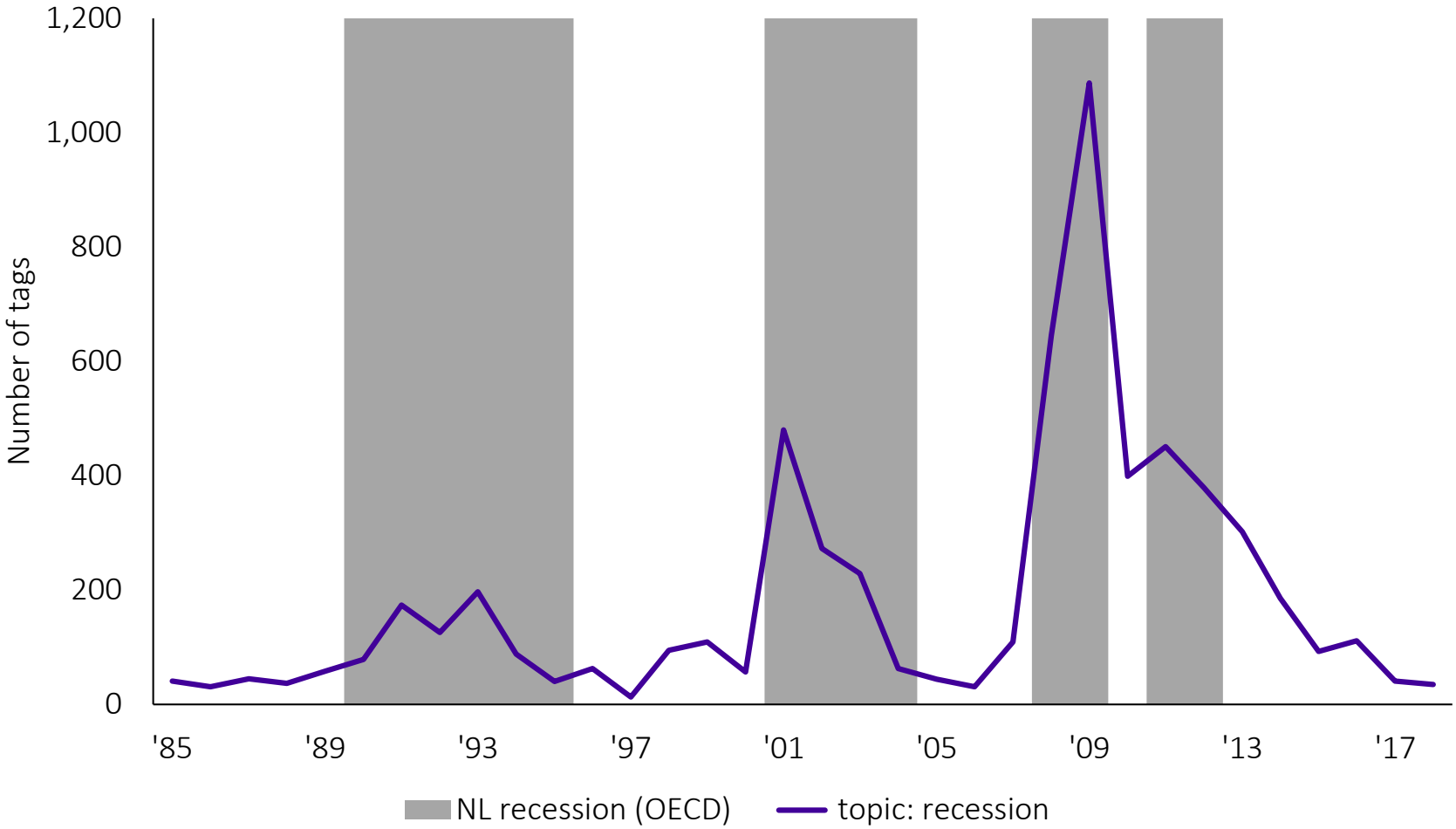
- 3,255 hand tagged topics; not all articles tagged;
- $<<$ 5% of articles tagged in period '00–'06;
- Re-tag with simple rule: occurence of tag in article;
- Topics = 3 tags with highest frequency;

DeNederlandscheBank
EUROSYSTEEM

Topic tags in Fd database after re-tagging

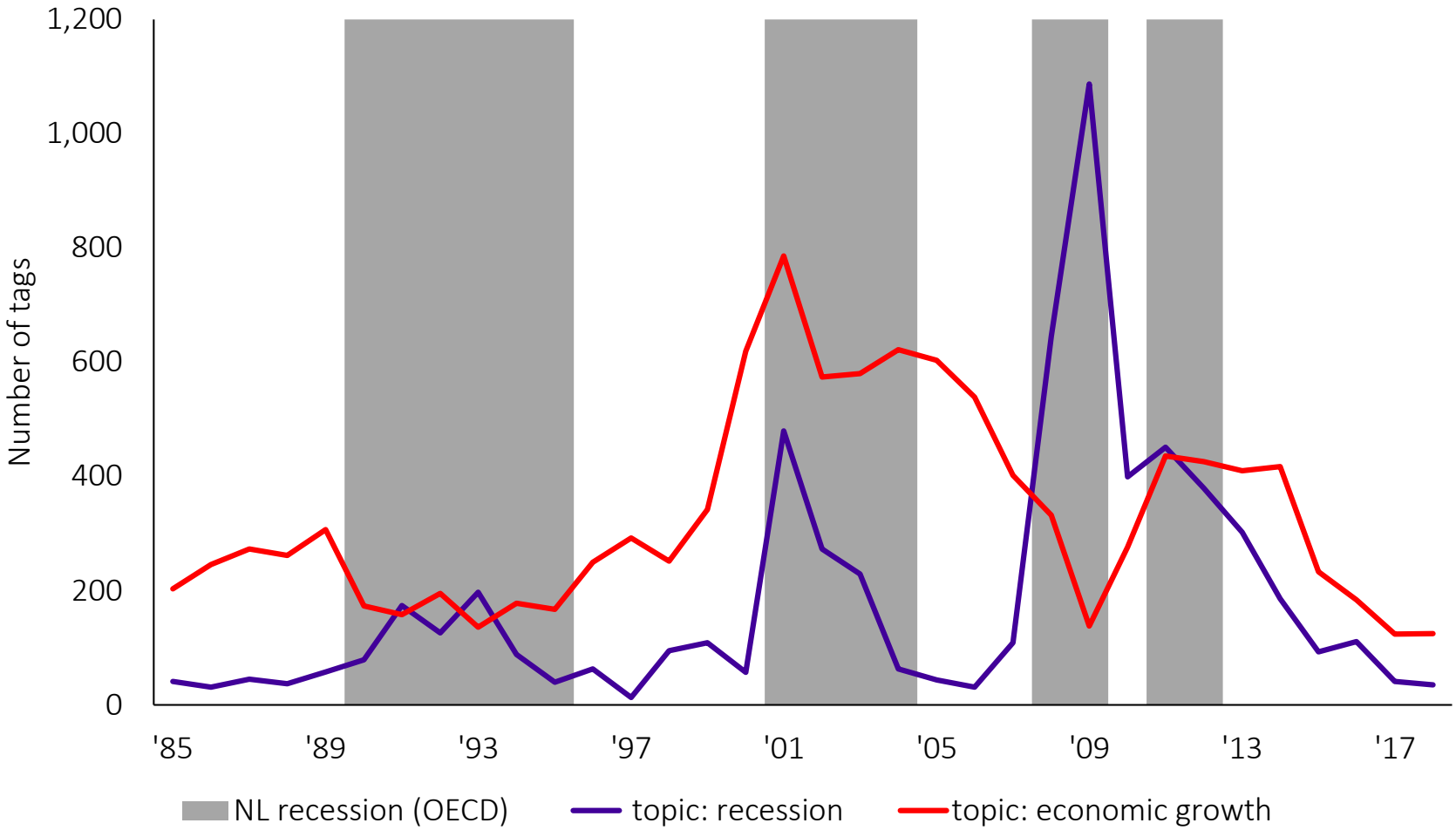
DeNederlandscheBank
EUROSYSTEEM

Count of the topic recession



DeNederlandscheBank
EUROSYSTEEM

Counts of the topic recession and economic growth



DeNederlandscheBank
EUROSYSTEEM

Part 2: Tone adjusted content of articles

Lexicon based sentiment scoring

- Sentiment score per article: freq. count of pre-defined dictionary with "positive" and "negative" keywords;
- Daily/Monthly/Quarterly score: average daily/monthly/quarterly article score.

No standard Lexicon based sentiment list for the Netherlands: create one ...

- Translated Loughran & McDonald (2011) list (Google Translate, DeepL) $\rightarrow$ **1,672 words**;
- Take polarity scores (between -1 and 1) from VU University developed generic list $\rightarrow$ **4,634 words**;
- Manually check for overlap and delete non-domain specific terms and contradictions $\rightarrow$ **5,632 words**.

Before scoring clean some more

- Stem verbs: Pattern package $\rightarrow$ **20,061 verbs** and their stem;
- Stem words: R Hunspell package for Dutch;
- Remove stopwords: List from several packages + manually added words $\rightarrow$ **500 words**.

DeNederlandscheBank
EUROSYSTEEM

Part 2: Tone adjusted content of articles

Lexicon sentiment index: frequency terms

DeNederlandscheBank
EUROSYSTEEM

Part 2: Tone adjusted content of articles

Lexicon sentiment index: frequency positive terms

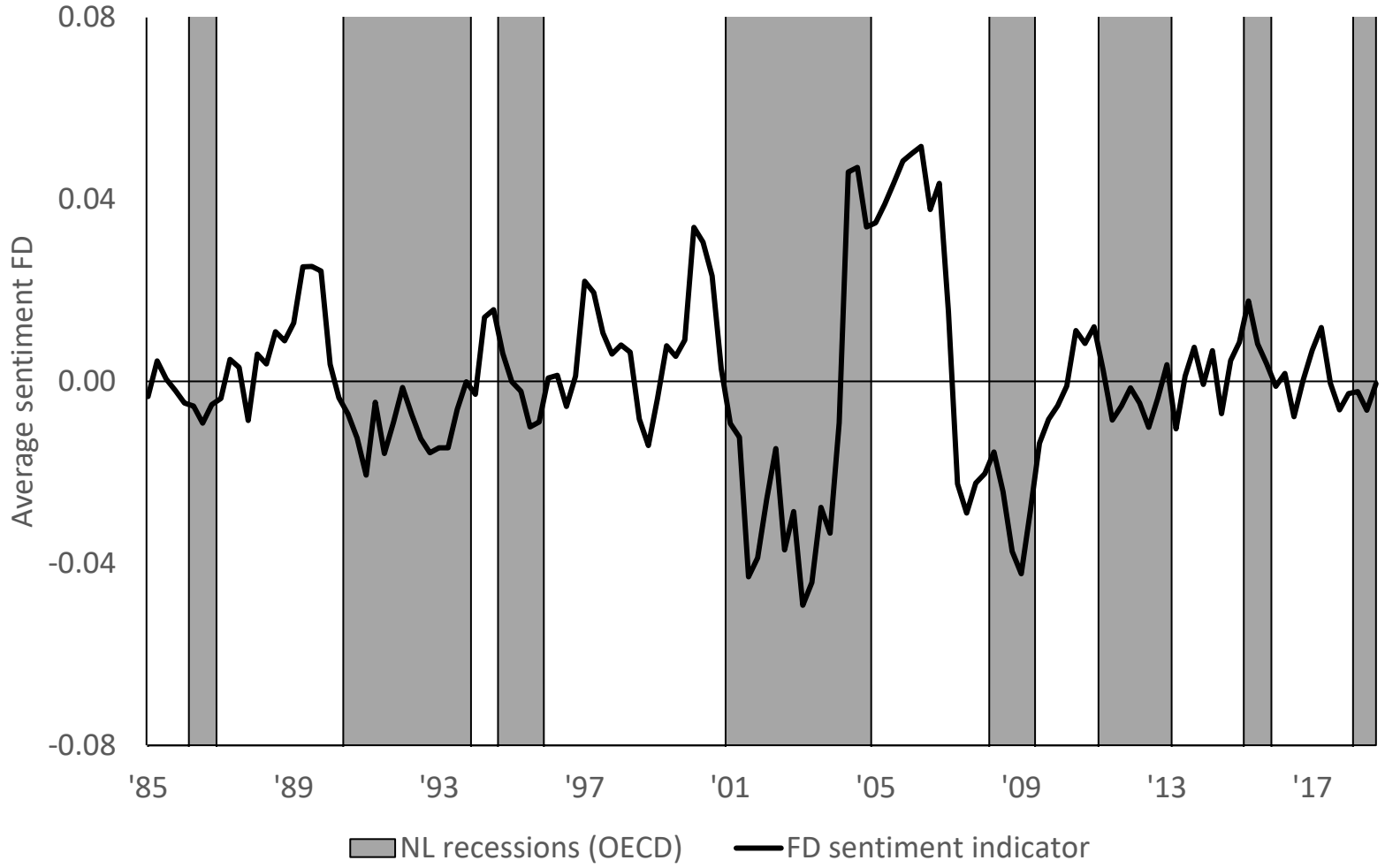
DeNederlandscheBank
EUROSYSTEEM

Part 2: Tone adjusted content of articles

Lexicon sentiment index: frequency negative terms

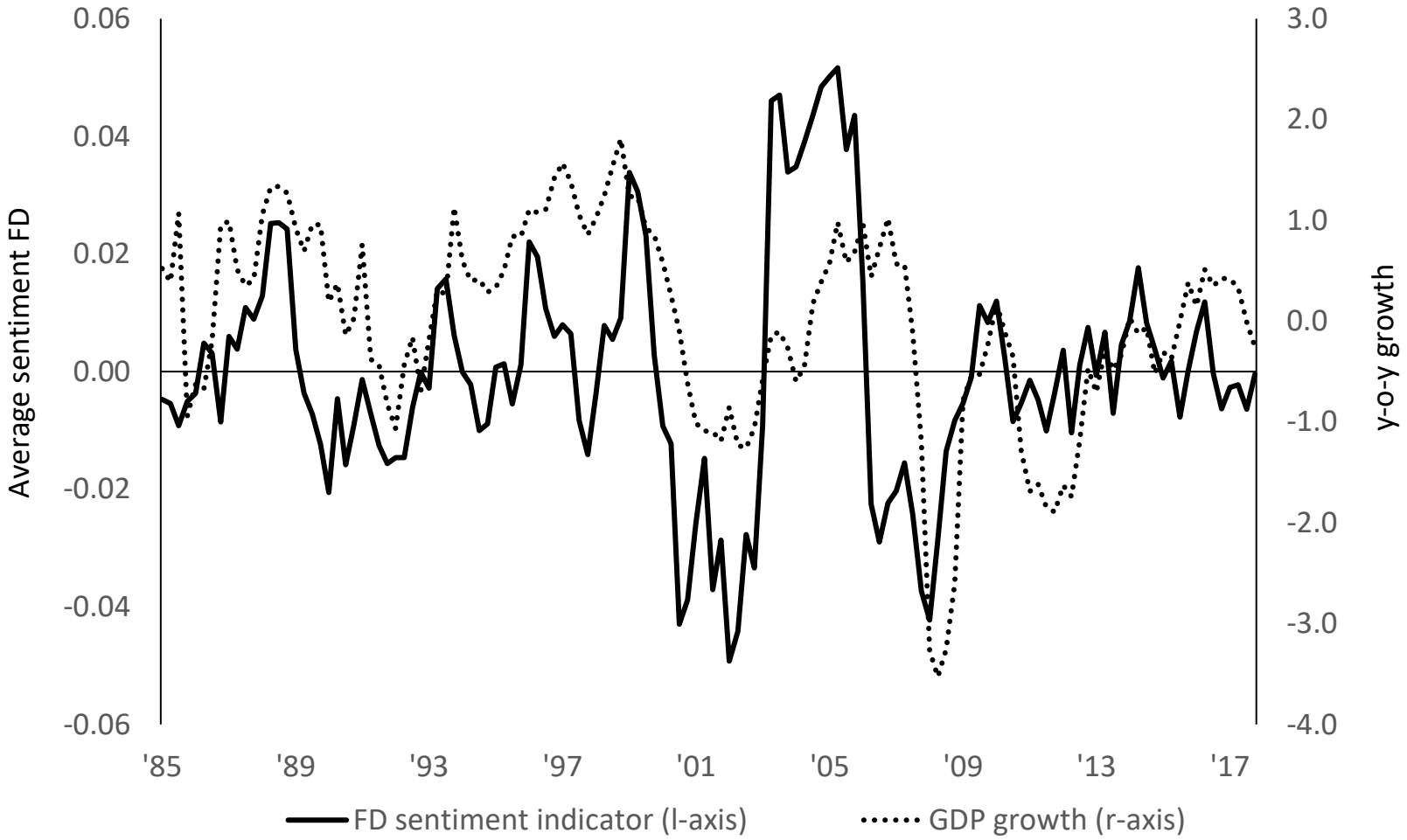
DeNederlandscheBank
EUROSYSTEEM

Lexicon sentiment indicator FD & NL recessions



DeNederlandscheBank
EUROSYSTEEM

Lexicon sentiment indicator FD & NL GDP



DeNederlandscheBank
EUROSYSTEEM

Hierarchical trend model

- Sentiment index is estimated using a Hierarchical Trend Model (HTM), taken from housing market research (Francke and Vos, 2004)
- Idea: housing market prices are sum of country + region + municipality effects (hierarchy).
- First test: total sentiment is modeled as local level model and NL,DE,FR, UK and US are modeled as a random walk:

$$y_t = \mathbf{i}\mu_t + D_t\kappa_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2 \mathbf{I}) \quad (1)$$

$$\mu_{t+1} = \mu_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2 \mathbf{I}) \quad (2)$$

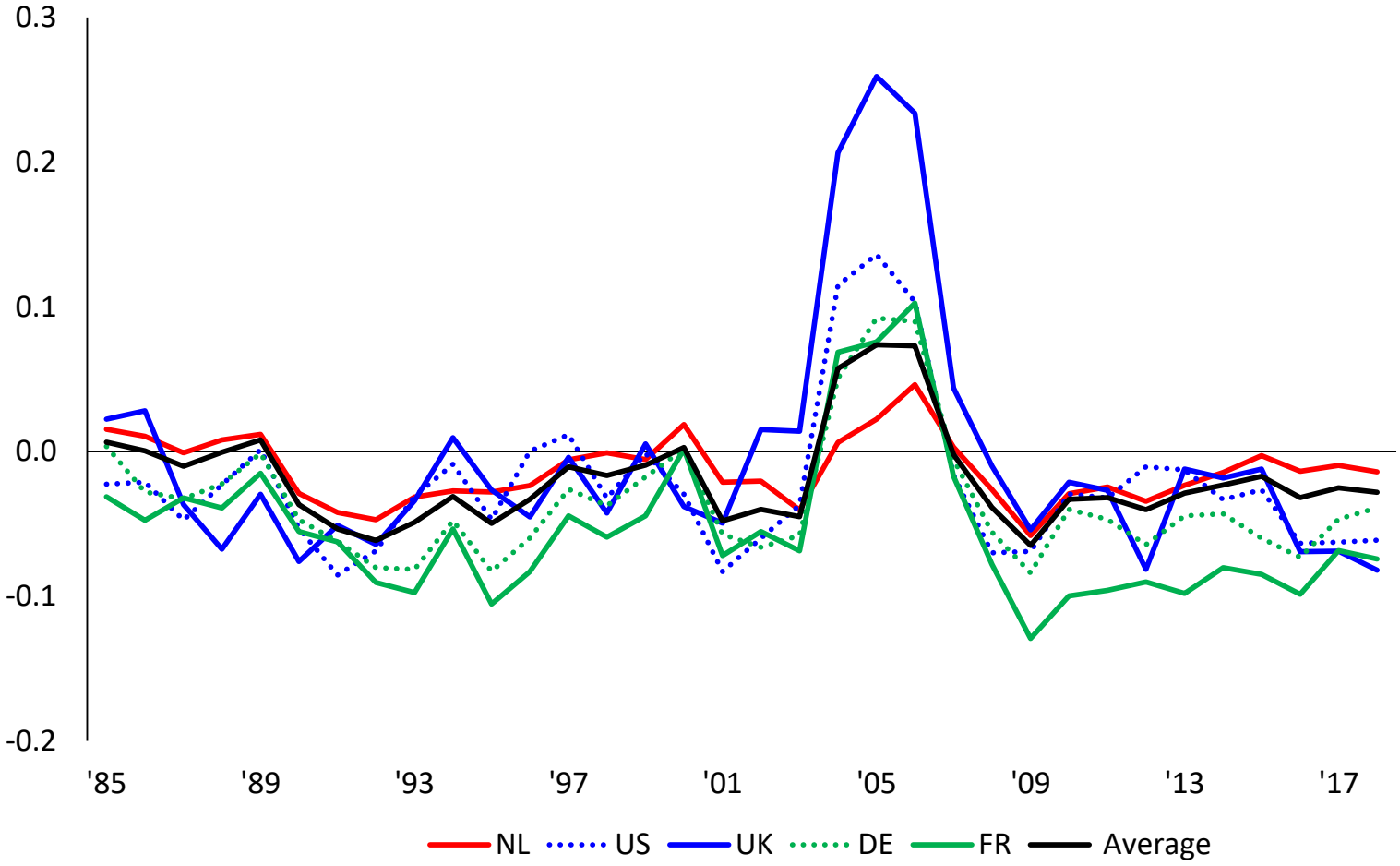
$$\kappa_{t+1} = \kappa_t + \zeta_t, \zeta_t \sim N(0, \sigma_\zeta^2 \mathbf{I}) \quad (3)$$

Sentiment of article i is a function of newspaper trend μ_t and topic-trend κ.

DeNederlandscheBank
EUROSYSTEEM

Part 2: Give meaning to sentiment scores using pre-defined topics

Lexicon sentiment indicator FD in a Hierarchical Trend Model



Possible extensions . . .

- Regions (North America, Asia, Europe) & countries;
- Different weights for countries or regions;
- Seperate trend and cycle.

DeNederlandscheBank
EUROSYSTEEM

Machine-learning alternatives

Alternative to Lexicon method

- Naïve Bayes Classifier: "spam" filter (Kotsiantsis et al., 2006)
- Naïve Bayes Classifier: positive/negative sentiment
- Idea: Frequency of words in "spam" / "non-spam", "positive/negative", e.g. apearance of *employment* and *GDP* will increase probabiliy of belonging to relevant article.

Alternative to using pre-defined categories

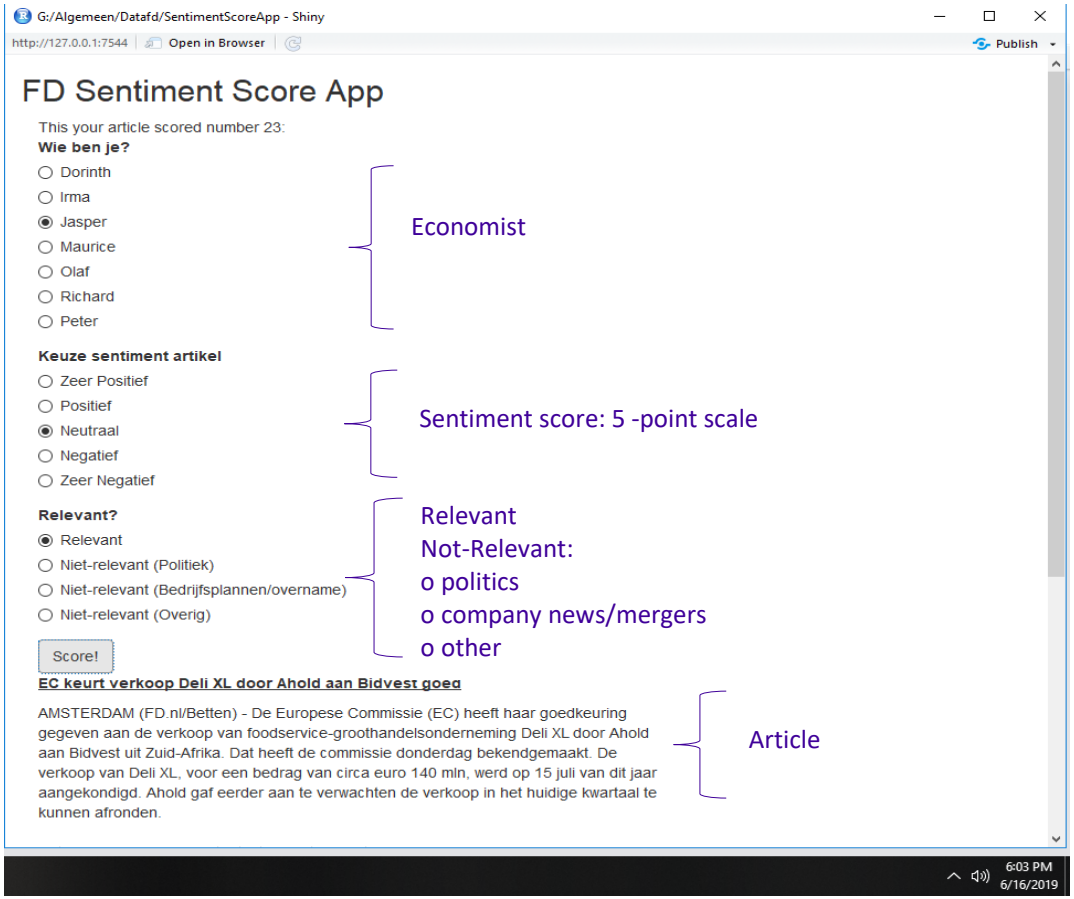
- Latent Dirichlet Allocation topic model (Hansen et al., 2018, Thorsrud, 2019)
- Idea: each article is distribution of topics, and each topics is distribution of words. Algorithm estimates probability of word belonging to a topic and topic to article.
- Need to shrink database (800,000 articles too much ...)

Work in progress: necessary to create test/training set

- We build a "scoring" app (in R) that takes random draws from articles-database:
  - Score sentiment on 5-point scale;
  - Score relevance (relevant, politics, company news/mergers, other);

DeNederlandscheBank
EUROSYSTEEM

FD sentiment score app



- First idea: score sentiment for Naïve Bayes Classifier. . .
- Only 10–20% of articles were judged as relevant . . .
- Idea of relevance and “spam” filter . . .
- Currently 12,000 articles scored . . .

DeNederlandscheBank
EUROSYSTEEM

On the agenda

On the agenda for this year

- Better train Bayes naive classifier; currently training set of 12,000 articles. Seems small compared to 800,000 articles $\rightarrow$ strive for 25,000;
- Additional fine-tuning of lexicon based sentiment list;
- Formal testing of value added textual data in nowcasting exercise;
- Formal comparison of lexicon-based method and machine-learning method;

Summing up

- Nowcasting using mechanical models is useful, espacially pertaining to the curent (and previous) quarter;
- Using/combining with forecasts of professional analysts helpful especiall in volatile times;
- Jury still out on value added of textual data, but first results are promising.

DeNederlandscheBank
EUROSYSTEEM

Thank you for your attention!

J.M.de.Winter@dnb.nl

DeNederlandscheBank
EUROSYSTEEM