

Nowcasting using news topics.

Big Data versus big bank*

Leif Anders Thorsrud[†]

December 21, 2016

Abstract

The agents in the economy use a plethora of high frequency information, including news media, to guide their actions and thereby shape aggregate economic fluctuations. Traditional nowcasting approaches have to a relatively little degree made use of such information. In this paper, I show how unstructured textual information in a business newspaper can be decomposed into daily news topics and used to nowcast quarterly GDP growth. Compared with a big bank of experts, here represented by official central bank nowcasts and a state-of-the-art forecast combination system, the proposed methodology performs at times up to 15 percent better, and is especially competitive around important business cycle turning points. Moreover, if the statistical agency producing the GDP statistics itself had used the news-based methodology, it would have resulted in a less noisy revision process. Thus, news reduces noise.

JEL-codes: C11, C32, E37

Keywords: Nowcasting, Dynamic Factor Model (DFM), Latent Dirichlet Allocation (LDA)

*This Working Paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. I thank Knut A. Aastveit, Hilde C. Bjørnland and Jonas Moss for valuable comments. Vegard Larsen provided helpful technical assistance, and Anne Sofie Jore was helpful in collecting data, for which I am grateful. This work is part of the research activities at the Centre for Applied Macro and Petroleum economics (CAMP) at the BI Norwegian Business School.

[†]Norges Bank and Centre for Applied Macro and Petroleum Economics, BI Norwegian Business School.
Email: `leif.a.thorsrud@bi.no`

1 Introduction

Because macroeconomic data are released with a substantial delay, predicting the present, i.e., nowcasting, is one of the primary tasks of market and policy oriented economists alike. However, producing accurate assessments of the current conditions is difficult. The difficulty is particularly pronounced for key policy variables, such as GDP growth, because the variables themselves, and the information set thought to explain them, are hampered by revisions and ragged-edge issues.¹

The nowcasting literature has addressed these difficulties using a variety of techniques, such as applying high frequency information as predictors, potentially not subject to subsequent data revisions, methods that handle mixed frequency data, and forecast combination. Accordingly, the literature on nowcasting is voluminous, and I refer to [Banbura et al. \(2011\)](#) for a relatively recent survey. Still, although recent advances have delivered promising results, the state-of-the-art models and systems, used at, e.g., central banks, have a hard time performing well when economic conditions changes rapidly. This was particularly evident around the Great Recession, see, e.g., [Alessi et al. \(2014\)](#), when good forecasting performance perhaps mattered the most.

In this paper I show how model-based nowcasting performance of quarterly GDP growth can be improved further using Big Data. That is, textual data collected from the major business newspaper in Norway.² In particular, I show that it is possible to obtain nowcasts that perform up to 50 percent better than forecasts from simple time series models, and at times up to 15 percent better than a state-of-the-art forecast combination system. If a big bank of experts, here represented by official Norges Bank nowcasts, could have utilized the methodology proposed, it would have resulted in lower forecasting errors, especially around the period of the Great Recession. Moreover, if the statistical agency producing the output growth statistics itself had utilized the news-based methodology, it would have resulted in a less noisy revision process.

How are these gains achieved? Compared with existing nowcasting approaches, the framework proposed has two important new characteristics. First, the information set I use to predict output growth is textual. It is collected from a major business newspaper, and represented as a large panel of daily tone adjusted topic frequencies that vary in intensity across time. The extraction of topics is done using advances in the natural

¹The former refers to the fact that macroeconomic data are typically heavily revised after their initial release, while the ragged edge problem is due to the asynchronous manner in which economic statistics becomes available within a given period, e.g., month or quarter.

²The term “Big Data” is used for textual data of this type because it is, before processing, highly unstructured, contains millions of words and thousands of articles. See, e.g., [Nyman-Andersen \(2016\)](#) for a more elaborate discussion about what “Big Data” constitutes.

language processing literature, while the tone is identified using simple dictionary based techniques. My hypothesis is simple: To the extent that the newspaper provides a relevant description of the economy, the more intensive a given topic is represented in the newspaper at a given point in time, the more likely it is that this topic represents something of importance for the economy's current and future needs and developments. For example, I hypothesize that when the newspaper writes extensively about developments in, e.g., the oil sector, and the tone is positive, this reflects that something is happening in this sector that potentially has positive economy-wide effects. This approach stands in stark contrast to conventional data usage, where the indicators used in the information set when doing model-based nowcasting are obtained from structured databases and professional data providers. The agents in the economy, on the other hand, likely use a plethora of high frequency information to guide their actions and thereby shape aggregate economic fluctuations. It is not a brave claim to assert that this information is highly unstructured and does not come (directly) from professional data providers, but more likely reflect information shared, generated, or filtered through a large range of channels, including media.

Second, to bridge the large panel of daily news topics to quarterly GDP growth, I use a mixed frequency, time-varying, Dynamic Factor Model (DFM). In general, the factor modeling approach permits the common dynamics of a large number of time series to be modeled in a parsimonious manner using a small number of unobserved (or latent) dynamic factors. A key property of the factor model used here, however, is that it is specified with an explicit threshold mechanism for the time-varying factor loadings. In turn, this enforces sparsity onto the system, but also takes into account that the relationship between the variables in the system might be unstable. A prime example is if a topic is associated with the stock market, where the stock market has been shown to be very informative of GDP growth in some periods, but not in others ([Stock and Watson \(2003\)](#)). The threshold mechanism potentially captures such cases in a consistent and transparent way and safeguards against over-fitting.

In using newspaper data, the approach taken here shares many features with a growing number of studies in economics using textual information. As thoughtfully described in [Bholat et al. \(2015\)](#), two often used methods in this respect are so called dictionary based and Boolean techniques. [Tetlock \(2007\)](#) is a famous example of the former. He classifies textual information using negative and positive word counts, and link the derived time series to developments in the financial market. An example of the usage of Boolean techniques is given in [Bloom \(2014\)](#), who discusses the construction of uncertainty indexes where the occurrence of words in newspapers associated with uncertainty are used to derive the indexes. In this paper, a dictionary based technique is used together with

what is called a topic model belonging to the Latent Dirichlet Allocation (LDA) class (Blei et al. (2003)). In general, topic modeling algorithms are statistical algorithms that categorizes the corpus, i.e., the whole collection of words and articles, into topics that best reflect the corpus's word dependencies. As such, each topic can be viewed upon as a word cloud, where the font size used for each word represents how likely it is to belong to this specific topic. A vast information set consisting of words and articles can thereby be summarized in a much smaller set of topics facilitating interpretation and usage in a time series context. Although topic models hardly have been applied in economics,³ their usage as a natural language processing tool in other disciplines has been massive. Their popularity, and the LDA's in particular, stems from their success in classifying text and articles into topics in much the same manner as humans would do, see Chang et al. (2009).

Compared to existing textual approaches used in economics, the LDA approach offers several conceptual advantages. In terms of a pure dictionary based approach for example, what is positive words and what negative obviously relates to an outcome. A topic does not. A topic has content in its own right. Moreover, the LDA is an automated machine learning algorithm, so subjectively choosing the words or specific categories to search for (e.g., uncertainty) is not needed. Instead, the LDA automatically delivers topics that best describe the whole corpus. This permits us to examine if textual information in the newspaper is representative for economic fluctuations, and if so, also identify the type of new information, in terms of topics, that might drive the nowcasts. As discussed in, e.g., Evans (2005) and Banbura et al. (2013), such decompositions are valuable, and particularly so for policy makers for which explaining what drives a forecast might be equally important as the forecast itself.

In sum, this article contributes to three branches of the economic literature. First, it contributes to the nowcasting literature where the usage of factor models and mixed frequency data have proven particularly useful, see, e.g., Stock and Watson (2002), Giannone et al. (2008), Breitung and Schumacher (2008), Kuzin et al. (2011), and Marcellino et al. (2013). In terms of modeling, I extend this literature by allowing for time-varying parameters and a latent threshold mechanism (Nakajima and West (2013) and Zhou et al. (2014)). In terms of data usage, I provide novel evidence on how information in the newspaper can be used to nowcast. The DFM I use was developed in Thorsrud (2016) to construct a daily news-based coincident index of the business cycle with almost perfect (in-sample) classification properties. Turning off the threshold mechanism, or estimating the model with a set of conventional high frequency business cycle indicators instead of news topics, results in much worse classification properties. Here, an extension to the

³See, e.g., Hansen et al. (2014) and Hansen and McMahon (2015) for exceptions.

model developed in [Thorsrud \(2016\)](#) is offered through the use of particle filtering techniques, making (daily) real-time updates of the model and nowcast computationally fast and efficient.

Second, this paper is related to a range of newer studies using internet search volume to predict the present. The most famous example among these is perhaps [Choi and Varian \(2012\)](#), who use Google Trends and specific search terms to construct predictors for present developments in a wide range of economic variables. Like when user generated search volume is used, the news topic approach can be said to capture economic agents' frame of focus, and thereby resemble some type of survey. In contrast to user generated internet search, however, the information in the newspaper has already gone through some type of filter. The number of pages in the newspaper are constrained, and the editors have to make choices regarding which news to report, and what to leave out. As both search and newspaper data surely are noisy "survey" measures, good editing likely amplifies the signal and reduces the noise.

Lastly, this paper speaks to a larger literature where alternative data sources and methods are used to understand economic behavior and information diffusion. See [Tetlock \(2014\)](#) and [Varian \(2014\)](#), and the references therein, for two illuminating overview articles in finance and economics in general, and [Dougal et al. \(2012\)](#) and [Peress \(2014\)](#) on the role of media, in particular. It is especially noteworthy that studies using almost exactly the same news measures as I use here finds that: unexpected news innovations lead to permanent increases in consumption and productivity, as predicted by the news driven view of the business cycle ([Larsen and Thorsrud \(2015\)](#)); news can be used to classify the phases of the business cycle with almost perfect accuracy ([Thorsrud \(2016\)](#)); news predicts intra-day returns and leads to significant continuation patterns lasting for roughly five business days ([Larsen and Thorsrud \(2016\)](#)). Coupled with the evidence brought forward here, it stands to reason that textual information generated and shared through the media is a highly important input when trying to understand how economic agents form their expectation about the future and thereby shape economic outcomes.

The rest of this paper is organized as follows. Section 2 describes the real-time GDP dataset, the newspaper data, the topic model, and the estimated news topics. The DFM is described in Section 3. Section 4 describes the nowcasting experiment. Results are presented in Section 5, while Section 6 concludes.

2 Data

The raw data used in this analysis consist of a long sample of the entire newspaper corpus for a daily business newspaper and quarterly, real-time, GDP growth for Norway. I

focus on Norway because it is a small and open economy and thereby representative of many western countries, and because small economies, like Norway, typically have only one or two business newspapers, making the choice of corpus less complicated. Here, I simply choose the corpus associated with the largest and most read business newspaper, *Dagens Næringsliv* (DN), noting that DN is also the fourth largest newspaper in Norway irrespective of subject matter. DN was founded in 1889, and has a right-wing and neo-liberal political stance. Importantly, however, the methodology for extracting news from newspaper data, and analyze whether or not it is informative for nowcasting, is general and dependent neither on the country nor newspaper used for the empirical application.

To make the textual data applicable for time series analysis, the data are first decomposed into news topics using a Latent Dirichlet Allocation (LDA) model. The news topics are then transformed into tone adjusted time series, where the tone is identified using simple dictionary based techniques, as in, e.g., [Tetlock \(2007\)](#). The newspaper corpus and the LDA specification in this paper is similar to that described in [Larsen and Thorsrud \(2015\)](#), while the way in which news topics are transformed to tone adjusted time series follows the procedures described in [Thorsrud \(2016\)](#). I provide a summary of the computations below. In the interest of preserving space, technical details are delegated to [Appendix D](#). The real-time dataset for quarterly GDP growth is described in the latter part of this section.

Two points are, however, worth noting. First, although many different topic models exist, a favorable property with the LDA is that it treats each article in the corpus as consisting of a mixture of topics, as humans typically would do (as opposed to only one).⁴ Second, identifying the tone of the news using dictionary based techniques is simple, but could potentially be improved upon with more sophisticated sentiment classification machine learning techniques, see, e.g., [Pang et al. \(2002\)](#). I leave such endeavors for future research. Still, in unreported results I experience that using news topics that are not adjusted by the tone gives much worse nowcasting properties. Thus, the combined usage of news topic and tone identification is important for the nowcasting performance.

2.1 The news corpus, the LDA and topics

The DN news corpus is extracted from Retriever’s “Atekst” database, and covers all articles published in DN from May 2 1988 to December 29 2014. In total this amounts to nearly half a million articles, well above one billion words, more than a million unique tokens, and a sample of almost ten thousand days. This massive amount of data makes statistical computations challenging, but as is customary in this branch of the literature

⁴[Blei \(2012\)](#) provides a nice layman introduction to topic modeling. More technical expositions of the LDA approach can be found in, e.g., [Blei et al. \(2003\)](#) and [Griffiths and Steyvers \(2004\)](#).

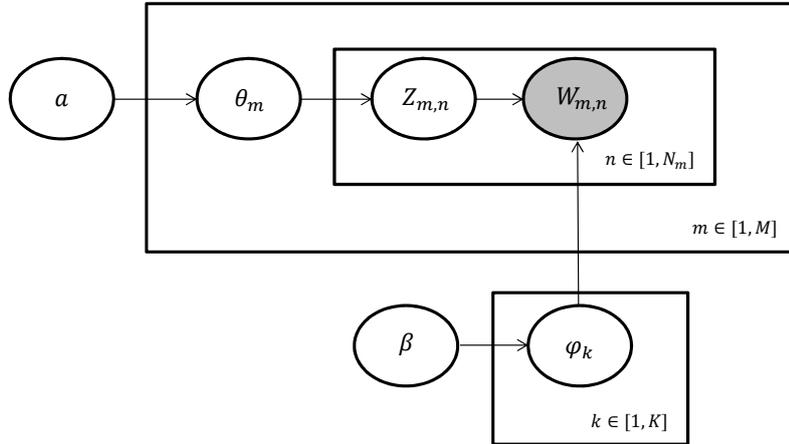


Figure 1. The LDA model visualized using plate notation.

some steps are taken to clean and reduce the raw dataset before estimation. This is done by removing common words, surnames, reduce all words to their respective word stems, and finally trimming the corpus using what is called the *term frequency - inverse document frequency*. A description of how this is done is given in Appendix D.1. I note here that around 250 000 unique tokens are kept after the filtering procedure.

The “cleaned”, but still unstructured, DN corpus is decomposed into news topics using a Latent Dirichlet Allocation (LDA) model. The LDA model is an unsupervised topic model that clusters words into topics, which are distributions over words, while at the same time classifying articles as mixtures of topics. By unsupervised learning algorithm, I mean an algorithm that can learn/discover an underlying structure in the data without the algorithm being given any labeled samples to learn from. The term “latent” is used because the words, which are the observed data, are intended to communicate a latent structure, namely the subject matter (topics) of the article. The term “Dirichlet” is used because the topic mixture is drawn from a conjugate Dirichlet prior.

Figure 1 illustrates the LDA model graphically. The outer box, or plate, represent the whole corpus as M distinct documents (articles). $N = \sum_{m=1}^M N_m$ is the total number of words in all documents, and K is the total number of latent topics. Letting bold-font variables denote the vector version of the variables, the distribution of topics for a document is given by $\boldsymbol{\theta}_m$, while the distribution of words for each topic is determined by $\boldsymbol{\varphi}_k$. Both $\boldsymbol{\theta}_m$ and $\boldsymbol{\varphi}_k$ are assumed to have conjugate Dirichlet distributions with (hyper) parameter (vectors) $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Each document consists of a repeated choice of topics $Z_{m,n}$ and words $W_{m,n}$, drawn from the Multinomial distribution using $\boldsymbol{\theta}_m$ and $\boldsymbol{\varphi}_k$. The circle associated with $W_{m,n}$ is gray colored, indicating that these are the only observable variables in the model.

At an intuitive level, the best way to understand the LDA model is likely to make a thought experiment of how the articles in the newspaper (the corpus) were generated.

1. Pick the overall theme of articles by randomly giving them a distribution over topics, i.e.: Choose $\boldsymbol{\theta}_m \sim \text{Dir}(\alpha)$, where $m \in \{1, \dots, M\}$.
2. Pick the word distribution for each topic by giving them a distribution over words, i.e.: Choose $\boldsymbol{\varphi}_k \sim \text{Dir}(\beta)$, where $k \in \{1, \dots, K\}$.
3. For each of the word positions m, n , where $n \in \{1, \dots, N_m\}$, and $m \in \{1, \dots, M\}$
 - 3.1. From the topic distribution chosen in 1., randomly pick one topic, i.e.: Choose a topic $Z_{m,n} \sim \text{Multinomial}(\boldsymbol{\theta}_m)$.
 - 3.2. Given that topic, randomly choose a word from this topic, i.e.: Choose a word $W_{m,n} \sim \text{Multinomial}(\boldsymbol{\varphi}_{z_{m,n}})$.

More formally, the total probability of a document, i.e., the joint distribution of all known and hidden variables given the hyper-parameters, is:

$$P(\mathbf{W}_m, \mathbf{Z}_m, \boldsymbol{\theta}_m, \boldsymbol{\Phi}; \alpha, \beta) = \underbrace{\prod_{n=1}^{N_m} P(W_{m,n} | \boldsymbol{\varphi}_{z_{m,n}}) P(Z_{m,n} | \boldsymbol{\theta}_m) \cdot P(\boldsymbol{\theta}_m; \alpha)}_{\text{word plate}} \cdot \underbrace{P(\boldsymbol{\Phi}; \beta)}_{\text{topic plate}} \quad (1)$$

document plate (1 document)

where $\boldsymbol{\Phi} = \{\boldsymbol{\varphi}_k\}_{k=1}^K$ is a $(K \times V)$ matrix, and V is the size of the vocabulary. The two first factors in (1) correspond to the word plate in Figure 1, the three first factors to the document plate, and the last factor to the topic plate. Different solution algorithms exist for solving the LDA model. I follow [Griffiths and Steyvers \(2004\)](#), and do not treat $\boldsymbol{\theta}_m$ and $\boldsymbol{\varphi}_k$ as parameters to be estimated, but instead integrate them out of (1). Considering the corpus as a whole, this results in an expression for $P(\mathbf{W}, \mathbf{Z}; \alpha, \beta) = P(\mathbf{Z} | \mathbf{W}; \alpha, \beta) P(\mathbf{W}; \alpha, \beta)$ which can be solved using Gibbs simulations. Estimates of $\boldsymbol{\theta}_m$ and $\boldsymbol{\varphi}_k$ can subsequently be obtained from the posterior distribution. Further technical details, and a short description of estimation and prior specification, are described in [Appendix D.2](#).

The model is estimated using 7500×10 draws. The first 15000 draws of the sampler are disregarded, and only every 10th draw of the remaining simulations are recorded and used for inference. $K = 80$ topics are classified. Marginal likelihood comparisons across LDA models estimated using smaller numbers of topics, see [Larsen and Thorsrud \(2015\)](#), indicate that 80 topics provide the best statistical decomposition of the DN corpus.

Now the LDA estimation procedure does not give the topics any name or label. To do so, labels are subjectively given to each topic based on the most important words associated with each topic. As shown in [Table 7](#) in [Appendix A](#), which lists all the estimated topics together with the most important words associated with each topic, it is, in most cases, conceptually simple to classify them. I note, however, that the labeling plays no material role in the experiment, it just serves as a convenient way of referring to the different topics (instead of using, e.g., topic numbers or long lists of words). What

2.2 News topics as tone adjusted time series

Given knowledge of the topics (and their distributions), the topic decompositions are translated into tone adjusted time series. To do this, I proceed in three steps, which are described in greater detail in Appendix D.4. In short, I first collapse all the articles in the newspaper for a particular day into one document, and then compute, using the estimated word distribution for each topic, the topic frequencies for this newly formed document. This yields a set of K daily time series. Then, for each day and topic, I find the article that is best explained by each topic, and from that identify the tone of the topic, i.e., whether or not the news is positive or negative. This is done using an external word list and simple word counts, as in, e.g., Tetlock (2007). The word list used here takes as a starting point the classification of positive/negative words defined by the *Harvard IV-4 Psychological Dictionary*, and then translates the words to Norwegian. For each day, the count procedure delivers two statistics, containing the number of positive and negative words associated with a particular article. These statistics are then normalized such that each observation reflects the fraction of positive and negative words that day, and are then used to sign adjust the topic frequencies computed in step one. Finally, I remove high frequency noise from each topic time series by using a (backward looking) moving average filter. As is common in factor model studies, see, e.g., Stock and Watson (2012), I also eliminate very low frequency variation, i.e., changes in the local mean, by removing a simple linear trend and standardize the data.

Figure 6, in Appendix A, reports six of the topic time series, and illustrates how the different steps described above affect the data. The gray bars show the data as topic frequencies across time, i.e., as constructed in the first step described above. As is clearly visible in the graphs, these measures are very noisy. Applying the subsequent transformations changes the intensity measures into sign identified measures and removes much of the most high frequency movements in the series. From Figure 6 we also observe that topics covary, at least periodically. The maximum (minimum) correlation across all topics is 0.57 (-0.40). However, overall, the average absolute value of the correlation among the topics is just 0.1, suggesting that different topics are given different weights in the DN corpus across time.

2.3 Real-time GDP

In a real-time out-of-sample forecasting experiment it is important to use data that were actually available on the date of the forecast origin. While the newspaper data are not revised, GDP is. For this reason a real-time dataset for Gross Domestic Product (GDP) is used. The raw data include 64 vintages of GDP for the Norwegian mainland economy,

covering the time period 2001:Q1 to 2015:Q4.⁵ Moreover, each vintage contains time-series observations starting in 1978:Q1.

To facilitate usage in the nowcasting experiment, see Section 4, I sort these real-time observations according to their release r , with $r = 1, \dots, \bar{r}$. Thus, in real-time jargon, I work with the diagonals of the real-time dataset. For each r , the sample is truncated such that each r covers the same sample. $\bar{r} = 5$ is considered to be the “final release”. Although this is somewhat arbitrary, working with a higher \bar{r} results in a loss of sample length, making an evaluation of the model’s nowcasting performance relative to benchmarks less informative. The truncation process yields \bar{r} time series with a length of 56 observations each, for the sample 2001:Q1 to 20014:Q4. For time series observations prior to 2001:Q1, each r is augmented with earlier time series observations collected from the 2001:Q1 vintage.⁶

Finally, the raw release series are transformed to quarterly growth rates. For future reference, I refer to these as $\Delta GDP_{t_q}^r$, where t_q is the quarterly time indicator. Prior to DFM estimation, and as in [Stock and Watson \(2012\)](#), the local mean of these growth rates is removed using a linear time trend, and the series are standardized. To distinguish these series from the unadjusted growth rates, I label them $\Delta GDP_{t_q}^{r,a}$.

3 The Dynamic Factor Model

To map the large panel of daily news topic to quarterly GDP growth, and produce nowcasts, I use the mixed frequency, time-varying Dynamic Factor Model (DFM) developed in [Thorsrud \(2016\)](#). Compared with related time-varying factor models in the literature,⁷ this model offers two extensions: First, sparsity is enforced upon the system through the time-varying factor loadings using a latent threshold mechanism. Second, since the variables in the system are observed at different frequency intervals, cumulator variables are used to ensure consistency in the aggregation from higher to lower frequencies and make estimation feasible. A short description of the model is provided below. A more technical description is given in Appendix E. I refer to [Thorsrud \(2016\)](#) for results showing the advantages of this modeling strategy relative to existing factor modeling approaches.

⁵The real-time dataset is maintained by Norges Bank. I thank Anne Sofie Jore for making this data available to me. In Norway, GDP excluding the petroleum sector is the commonly used measure of economic activity. I follow suit because it facilitates the formal evaluation of the nowcasts in Section 5.

⁶Generally, this data augmentation process might create a break in the time series. Here it is needed to be able to conduct the nowcasting experiment and simply to have enough observations to estimate the time series models. However, as explained in the next section, the model used contains time-varying parameters, which potentially adapt to such breaks.

⁷See, e.g., [Lopes and Carvalho \(2007\)](#), [Del Negro and Otrok \(2008\)](#), [Marcellino et al. \(2013\)](#), [Ellis et al. \(2014\)](#), and [Bjørnland and Thorsrud \(2015\)](#).

Measured at the highest frequency among the set of mixed frequency observables, the DFM can be written as:

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t \quad (2a)$$

$$\mathbf{a}_t = \mathbf{F}_1 \mathbf{a}_{t-1} + \cdots + \mathbf{F}_h \mathbf{a}_{t-h} + \boldsymbol{\omega}_t \quad (2b)$$

$$\mathbf{e}_t = \mathbf{P}_1 \mathbf{e}_{t-1} + \cdots + \mathbf{P}_p \mathbf{e}_{t-p} + \mathbf{u}_t \quad (2c)$$

Equation (2a) is the observation equation of the system. \mathbf{y}_t is a $N \times 1$ vector of observable and unobservable variables assumed to be stationary with zero mean, decomposed as follows:

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_{1,t}^* \\ \mathbf{y}_{2,t} \end{pmatrix} \quad (3)$$

where $\mathbf{y}_{1,t}^*$ is a $N_q \times 1$ vector of unobserved daily output growth rates, mapping into quarterly output growth rates as explained below, and $\mathbf{y}_{2,t}$ is a $N_d \times 1$ vector of daily newspaper topic variables, described in Section 2.2. $N = N_q + N_d$, and \mathbf{Z}_t is a $N \times q$ matrix with dynamic factor loadings linking the variables in \mathbf{y}_t to the latent dynamic factors in \mathbf{a}_t . The factors follow a VAR(h) process given by the transition equation in (2b), where $\boldsymbol{\omega}_t \sim i.i.d.N(0, \boldsymbol{\Omega})$. Finally, equation (2c) describes the time series process for the $N \times 1$ vector of idiosyncratic errors \mathbf{e}_t . It is assumed that these evolve as independent AR(p) processes with $\mathbf{u}_t \sim i.i.d.N(0, \mathbf{U})$. Thus, \mathbf{P} and \mathbf{U} are diagonal matrices, and \mathbf{u}_t and $\boldsymbol{\omega}_t$ are independent. In the specification used here $q = 1$. Therefore, \mathbf{a}_t is a scalar, and can be interpreted as a latent daily coincident index of the business cycle. Moreover, I set $h = 10$ and $p = 1$, and as discussed below $N_q = 1$ and $N_d = 20$.

The model's only time-varying parameters are the factor loadings. Following the Latent Threshold Model (LTM) idea introduced by Nakajima and West (2013), and applied in a DFM setting in Zhou et al. (2014), sparsity is enforced onto the system through these using a latent threshold mechanism. For example, for one particular element in the \mathbf{Z}_t matrix, $z_{i,t}$, the LTM structure can be written as:

$$z_{i,t} = z_{i,t}^* \varsigma_{i,t} \quad \varsigma_{i,t} = I(|z_{i,t}^*| \geq d_i) \quad (4)$$

where

$$z_{i,t}^* = z_{i,t-1}^* + w_{i,t} \quad (5)$$

with $w_{i,t} \sim i.i.d.N(0, \sigma_{i,w}^2)$. In (4) $\varsigma_{i,t}$ is a zero one variable, whose value depends on the indicator function $I(|z_{i,t}^*| \geq d_i)$. If $|z_{i,t}^*|$ is above the threshold value d_i , then $\varsigma_{i,t} = 1$, otherwise $\varsigma_{i,t} = 0$. Accordingly, the LTM works as a dynamic variable selection mechanism. Letting $\mathbf{w}_t \sim i.i.d.N(0, \mathbf{W})$, it is assumed that \mathbf{w}_t is independent of both \mathbf{u}_t and $\boldsymbol{\omega}_t$, and that \mathbf{W} is a diagonal matrix.

Due to the mixed frequency property of the data, the \mathbf{y}_t vector in equation (2a) contains both observable and unobservable variables. Thus, the model as formulated in (2) cannot be estimated. However, following Harvey (1990), and since $y_{1,t}^*$ is a flow measure, the model can be reformulated such that observed quarterly series are treated as daily observations with missing observations. To this end, the \mathbf{y}_t vector is decomposed as in equation (3). Assuming further that the quarterly variable is observed on the last day of each quarter, we can define:

$$\tilde{y}_{1,t} = \begin{cases} \sum_{j=0}^m y_{1,t-j}^* & \text{if } \tilde{y}_{1,t} \text{ is observed} \\ NA & \text{otherwise} \end{cases} \quad (6)$$

where $\tilde{y}_{1,t}$ is treated as the intra-period sum of the corresponding daily values, and m denotes the number of days since the last observation period. Because quarters have uneven number of days, $\tilde{y}_{1,t}$ is observed on an irregular basis, and m will vary depending on the relevant quarter and year. This variation is however known and easily incorporated into the model structure.

Given (6), temporal aggregation can be handled by introducing a cumulator variable of the form:

$$C_{1,t} = \beta_{t,q} C_{1,t-1} + y_{1,t}^* \quad (7)$$

where $\beta_{t,q}$ is an indicator variable defined as:

$$\beta_{t,q} = \begin{cases} 0 & \text{if } t \text{ is the first day of the period} \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

and $y_{1,t}^*$ maps to the latent factor, a_t , from equation (2b). Thus, $\tilde{y}_{1,t} = C_{1,t}$ whenever $\tilde{y}_{1,t}$ is observed, and treated as a missing observation in all other periods. Because of the usage of the cumulator variable in (7), one additional state variable is introduced to the system. Importantly, however, the system will now be possible to estimate using standard filtering techniques handling missing observations. Details are given in Appendix E.

As is common for all factor models, the factors and factor loadings in (2) are not identified without restrictions. To separately identify the factors and the loadings, the following identification restrictions on \mathbf{Z}_t in (2a) are enforced:

$$\mathbf{Z}_t = \begin{bmatrix} \tilde{\mathbf{z}}_t \\ \hat{\mathbf{z}}_t \end{bmatrix}, \quad \text{for } t = 0, 1, \dots, T \quad (9)$$

Here, $\tilde{\mathbf{z}}_t$ is a $q \times q$ identity matrix for all t , and $\hat{\mathbf{z}}_t$ is left unrestricted. Bai and Ng (2013) and Bai and Wang (2014) show that these restrictions uniquely identify the dynamic factors and the loadings, but leave the VAR(h) dynamics for the factors completely unrestricted.

The DFM is estimated by decomposing the problem of drawing from the joint posterior of the parameters of interest into a set of much simpler ones using Gibbs simulations. The Gibbs simulation employed here, together with the prior specifications, are described in greater detail in Appendix E. The results reported in this paper are all based on 9000 iterations of the Gibbs sampler. The first 6000 are discarded and only every sixth of the remaining is used for inference.

To ease the computational burden, I truncate the news topic dataset to include only 20 of the 80 topics. Thus, the number of daily observables $N_d = 20$. The truncation, and choice of topics to include, is based on the findings reported in Thorsrud (2016), and is examined in greater depth there. Table 7 in Appendix C highlights the included topics. Without going into detail, I note that the topics do for the most part reflect topics one would expect to be important for business cycles in general, and for business cycles in Norway in particular. Examples of the former are the *Monetary policy*, *Fiscal policy*, *Wage payments/Bonuses*, *Stock market*, and *Retail* topics, while the *Oil service* topic is an example of the latter.⁸ Still, although most topics are easily interpretable, some topics either have labels that are less informative, or reflect surprising categories. An example is the *Life* topic. That said, such exotic or less informative named topics, are the exception rather than the rule.

The number of quarterly observables $N_q = 1$. Instead of using the “final” vintage of GDP growth to estimate the model, I use the first release series, i.e., $\Delta GDP_{t_q}^{1,a}$, as this series is not subsequently revised. Accordingly, as described in, e.g., Croushore (2006), data revisions will not affect the parameter estimates of the model nor lead to a change in the model itself (such as the number of lags). In turn, both of these properties facilitate real-time updating of the model when conducting the nowcasting experiment, confer Section 4. It follows from the discussion of equations (6) to (8) that $\tilde{y}_{1,t} = C_{1,t} = \Delta GDP_{t_q}^{1,a}$ whenever $\Delta GDP_{t_q}^{1,a}$ is observed.

4 Nowcasting

To produce nowcasts of GDP growth I take advantage of the fact that the daily news topics are always available, while quarterly GDP growth is published with a substantial lag. Because the state-space system in (2) is non-linear this conditioning can not be implemented using the Kalman Filter. Instead, I keep the model’s hyper-parameters fixed at some initially estimated values, and create updated estimates of the latent states

⁸Norway is a major petroleum exporter, and close to 50 percent of its export revenues are linked to oil and gas. See Bjørnland and Thorsrud (2015), and the references therein, for a more detailed analysis of the strong linkages between the oil sector and the rest of the mainland economy.

and thus nowcasts of GDP, conditional on knowing the daily observables using particle filter methods.

Generally, particle filters are part of what’s called Sequential Monte Carlo (SMC) methods, and many different types of particle filters exist. The one used here, and accommodated to the model in (2), is a so-called mixture auxiliary particle filter, see [Michael K. Pitt \(1999\)](#), [Chen and Liu \(2000\)](#) and [Doucet et al. \(2001\)](#). A detailed description of the algorithm is provided in Appendix H. Here I note that the mixture term is used because the filter exploits that some parts of the model can be solved analytically, which always is an advantage as it reduces Monte Carlo variation.

An alternative strategy would have been to re-estimate the model at every point in time when a real-time nowcast is desired. This will work since the model is designed to handle missing values by definition. However, due to the MCMC procedure used to estimate the model, this will be a very time consuming and computational demanding exercise. While a full re-estimation of the system takes many hours (or days), constructing updated estimates of the latent states using the particle filter takes minutes and also allows for parallelization. Of course, to the extent that the model’s hyper-parameters change through time, keeping them constant might degrade the model’s forecasting performance. Still, because of the large improvement in computation time, this is the solution chosen in the experiments conducted here.

4.1 Rescaling and accounting for drift

The DMF model gives estimates, and thus forecasts, of the adjusted output growth measure $\Delta GDP_{t_q}^{1,a}$. What we are interested in, however, are predictions of actual growth, $\Delta GDP_{t_q}^1$. In other words, a rescaling is needed. Moreover, while we might be interested in how the model predicts the first release of actual GDP growth, we might also be interested in how the model predicts later releases, cf. Section 4.2. Insofar as the initial GDP release is an efficient estimate of later releases, i.e., that data revisions are due to new information obtained by the statistical agency after the time of the first release, a model that does a good job of predicting the initial release will also do a good job of predicting later releases. However, there is substantial evidence showing that initial GDP releases are not efficient estimates of later (or the “final”) releases, see, e.g., [Faust et al. \(2005\)](#) for an application to G-7 countries.⁹ In such cases, it might be better to predict the release of interest directly.

To address these issues, I estimate a simple time-varying (*TVP*) rescaling model.

⁹The forecast rationality arguments used here are due to [Mankiw et al. \(1984\)](#), who introduced the distinction between news (not news topics) and noise in the revision process for real-time macroeconomic data.

First, assume that we are standing at time T_q and have produced DFM estimates and predictions of $\Delta GDP_{t_q}^{1,a}$ for $t_q = 1, \dots, T_q$ using the methodology described above. Then, to map $\Delta GDP_{t_q}^{1,a}$ to $\Delta GDP_{t_q}^r$, the following equation is estimated for each r :

$$\Delta GDP_{t_q-r}^r = \alpha_{t_q-r}^r + \beta_{t_q-r}^r \Delta GDP_{t_q-r}^{1,a} + e_{t_q-r}^r \quad (10)$$

where the r indexing addresses our interest in how the model predicts different data releases, and the subtraction of r in the time indexes is done to take into account that GDP growth is not available at time T_q , because it is published with a lag. To the extent that GDP growth inhabits a low frequency change in the mean or variance, this will be captured by the time-varying parameters α and β , which are assumed to follow independent random walk processes. The posterior estimates of the parameters are obtained by applying Carter and Kohn’s multimove Gibbs sampling approach (Carter and Kohn (1994)),¹⁰ and predictions of $GDP_{T_q}^r$ are formed simply as:

$$\Delta G\hat{D}P_{T_q}^r = \hat{\alpha}_{T_q}^r + \hat{\beta}_{T_q}^r \Delta G\hat{D}P_{T_q}^{1,a} \quad (11)$$

Apart from allowing for time-varying parameters and using various data releases when estimating (10), the steps taken here to bridge the DFM estimates with actual GDP growth are similar to those used in, e.g., Giannone et al. (2008), and many other factor modeling applications involving mixed frequency data.¹¹

4.2 The nowcasting experiment

To assess the DFM’s nowcasting performance, I run a real-time out-of-sample forecasting experiment. The informational assumptions are as follows: At each forecast round, I assume that a full quarter of daily information is available. The first release of GDP growth, on the other hand, is only available for the previous quarter. In Norway, the National Accounts Statistics for the previous quarter are usually published in the middle of the second month of the current quarter. As such, these informational assumptions are realistic.

The DFM is first estimated using the sample 1 June 1988 to 31 December 2003, and the hyper-parameters are saved for later usage. Then, for each quarter from 2004:Q1

¹⁰The general algorithm is described in Appendix F. Since this is a relatively straight forward application of a time-varying parameter model, I do not expand on the technical details. In terms of priors, α_0 and β_0 is set to 0, and the variance of the error term associated with the law of motion for the parameters is assumed a priori to be 0.01. For both processes I apply a roughly 10 percent weight on these beliefs relative to the data.

¹¹I have also estimated the rescaling model as a simple static regression (*OLS*). Compared to the *TVP* case, the results are slightly, but not significantly, worse. Interestingly, however, there are some important differences across time. See Appendix B for details.

to 2014:Q4, I produce a nowcast of *GDP* growth, yielding a sample of 40 real-time nowcasts.¹² For each forecast round, I use the particle filter, together with the stored hyper-parameters, to construct updated state estimates, see Appendix H, and then the rescaling methodology described in Section 4.1 to predict actual GDP growth. For the models used in the rescaling step, a recursive estimation window is used, starting in 1988:Q2 for all forecast rounds. I note that the word distributions, estimated using the LDA model and described in Section 2.1, are kept fixed throughout the analysis. In principle, these should also be updated in real time, but re-estimating these distributions at each forecast round would be computationally very demanding since simulating the posterior distribution takes a long time to complete. However, in unreported results I have compared the implied news topic time series using word distributions estimated over the full sample with those obtained when truncating the corpus sample to end in 31 December 2000, finding that the correlations are very close to one.¹³ This suggests that the output from the LDA model is not very sensitive to the sample split used in the nowcasting experiment, and that keeping the word distributions fixed is an innocuous simplification.

The nowcasts are evaluated using two simple statistics, namely the Root Mean Squared Forecast Errors (RMSFE), and the bias. A key issue in this exercise is the choice of a benchmark for the “actual” measure of GDP growth. In principle, one could argue that the nowcasts should be as accurate as possible when evaluated against the most revised release, i.e., \bar{r} , as this release presumably contains more information and is closer to the true underlying growth rate. In practice however, the forecaster is often evaluated in real time, i.e., against the first release. Thus, depending on the forecaster’s loss function, it is not clear a priori which release one should evaluate the nowcast against (Croushore (2006)). For this reason, when computing the test statistics, $r = 1, \dots, \bar{r}$ different $GDP_{t_q}^r$ releases are used as the “true” value. Together with the estimation of \bar{r} nowcasts, see Section 4.1, this permits a $\bar{r} \times \bar{r}$ evaluation of the predictive performance and answering the question: Which data release is the best to use to form nowcasts when evaluated against release r ?

¹²The sample split is chosen because it gives the model a substantial number of observations to learn the hyper-parameter distributions prior to doing the nowcasting experiment, and because the out-of-sample period corresponds to the sample period available for the benchmarks models, see Section 4.3 and Appendix C.

¹³As mentioned in Appendix D.2, one caveat with a comparison like this is lack of identifiability, meaning that *topic 1* in one sample might not equal *topic 1* in another sample. Still, creating a (informal) mapping between the word distributions for different samples based on the most important words that belong to each distribution is possible, although somewhat cumbersome.

4.3 Benchmark forecasts

Four benchmark forecasts are used and compared with those from the DFM, news-based, predictions. The first benchmark forecast is obtained from a simple $AR(1)$, estimated on $GDP_{t_q}^1$. For the $AR(1)$, a recursive estimation window is applied, and the estimation sample begins in 1988:Q2 irrespective of forecast origin. The second benchmark is simply taken to be the mean of $GDP_{t_q}^1$, computed using the previous 5 quarters at each forecast round. I denote this forecast RM .

Arguable, these benchmarks are very simplistic. Therefore, I also compare the nowcasts from the DFM to nowcasts produced in real time by Norges Bank. Two different types of Norges Bank nowcasts are compared. The first type is the official Norges Bank nowcasts (NB), while the second type is predictions from Norges Bank’s model-based nowcasting system (System for Averaging Models, SAM).¹⁴

The Norges Bank nowcasts are interesting benchmarks for two reasons. First, the NB predictions are subject to judgment, and potentially incorporate both hard economic and more anecdotal information. As discussed in Alessi et al. (2014), expert judgment potentially plays an important role when producing forecasts in most institutions, and might be particularly important when economic conditions change rapidly, as they did around the financial crisis. Second, SAM predictions are produced using a state-of-the-art forecast combination framework. In contrast to the NB predictions, the SAM forecasts are purely model based. Aastveit et al. (2014) uses, more or less, the same system to nowcast U.S. GDP growth, and show that the nowcasts produced by the forecast combination scheme is superior to a simple model selection strategy and that the combined forecast always performs well relative to the individual models entertained in the system. In general, a large forecasting literature has documented the potential benefits of using forecast combination techniques relative to using predictions generated by individual models, see Timmermann (2006) for an extensive survey. Thus, in a pure forecasting horse-race, it is difficult to envision a better model-based competitor than SAM .

A detailed description of how the SAM system works and performs can be found in Bjørnland et al. (2011), Bjørnland et al. (2012), and Aastveit et al. (2014). A brief description of the SAM system is provided in Appendix C. Importantly, although the SAM system incorporates information from hundreds of individual models and economic time series (including forward-looking variables such as surveys, financial market variables, and commodity prices), none of the individual models entertains textual information.

¹⁴I thank Anne Sofie Jore at Norges Bank for making these data available to me. See Appendix C for a description of how the two Norges Bank nowcasts are compiled to match the timing assumptions for the nowcasting experiment.

5 Results

I start this section by reporting how the estimated daily coincident index evolves when estimated in real time, see Figure 3a. The red line in the figure shows the index as estimated at the last time period in the sample. The blue lines show the index as estimated at each point in time according to the updating schedule used in the nowcasting experiment. Apart from the fact that the index seems to pick up important business cycle phases in the Norwegian economy, as documented thoroughly in Thorsrud (2016), one other fact stands out.¹⁵ The index estimates do not change markedly across vintages.

The index's stability across vintages is partly due to how the updating of the model is done when doing the real-time nowcasting experiment, where the hyper-parameters are kept constant. It is also due to the use of the first releases of output growth when estimating the model. Vintages of data are revised, but first releases are not. In any case, to the extent that the index is used for business cycle classification, as in Thorsrud (2016), small real-time revisions of the index are desirable.

Figure 3b reports the actual realizations of GDP growth together with the nowcasts produced using the news-based models. For the realizations I report the full range (min and max) of outcomes at each point in time, for all \bar{r} releases, as a gray colored area. Likewise, for the predictions I report the prediction range (min and max), where the different predictions are determined by the release used for estimating the rescaling model. As clearly seen in the graphs, data revisions are substantial. For example, around March 2007 the different data releases indicated that output growth could be anything between roughly 1.25 to 2 percent. Despite these big revisions, visual inspection alone suggests that the news-based models perform well in predicting the current state of the economy. It is particularly interesting that all models catch the downturn associated with the Great Recession particularly well. However, the spread in the predictions are relatively large, possibly reflecting the flexibility offered by the time-varying *TVP* class.

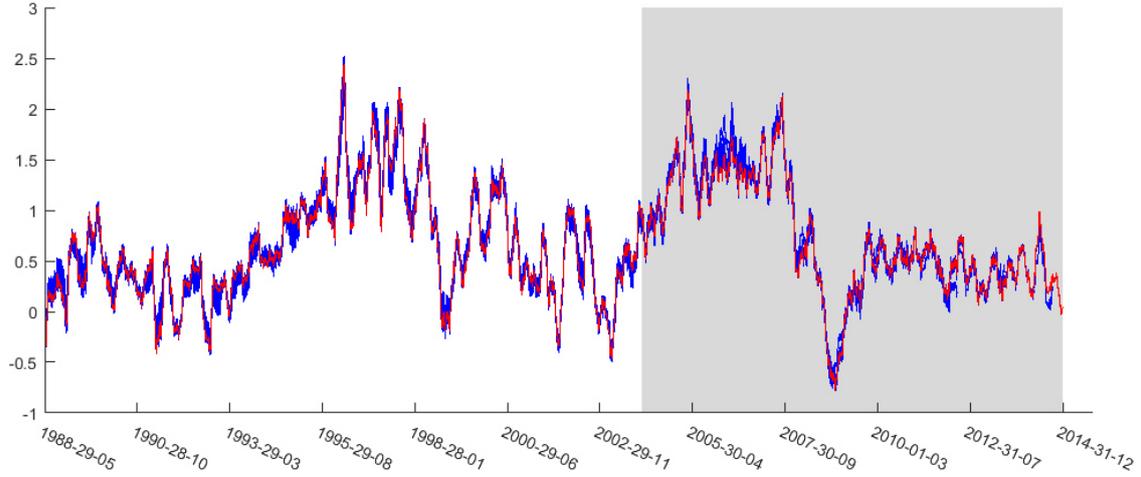
5.1 News versus benchmarks

Tables 1 and 2 summarize the first main finding of this article, namely that nowcasts produced by the news-based model framework are highly competitive in terms of RMSFE and bias compared with the benchmark forecasts.

I start by looking at the RMSFE comparisons, reported in Table 1, and the results

¹⁵The index captures the low growth period in the early 1990s, the boom and subsequent bust around the turn of the century, and finally the high growth period leading up to the Great Recession. I note that the downturn in the economy following the Norwegian banking crisis in the late 1980s was longer lasting than the downturn following the global financial crisis in 2008.

(a) The coincident index (in real-time)



(b) Nowcasts: News and actual

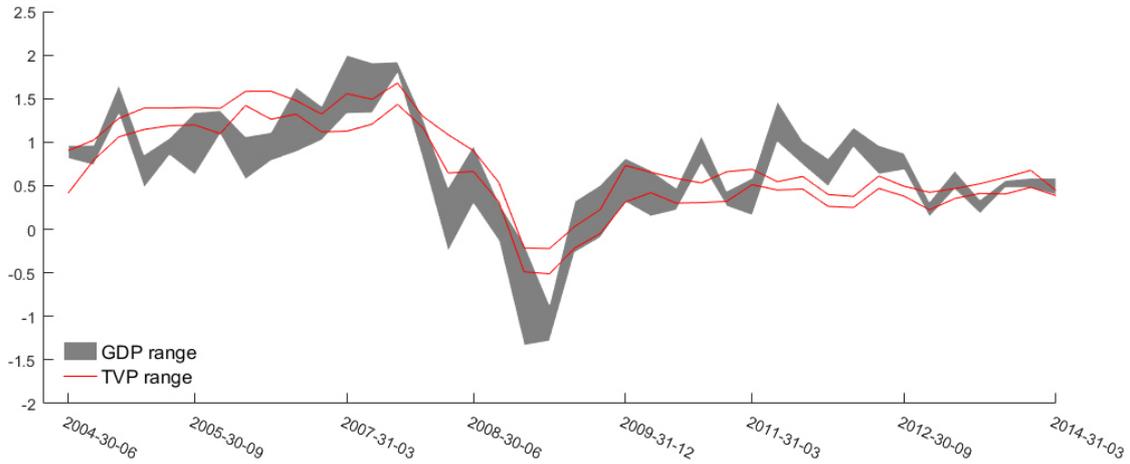


Figure 3. Figure 3a reports the real-time estimates of the daily news index. The red solid line reports the news index as estimated using the full sample. The blue line reports the index as estimated at each point in time according to updating schedule defined by the nowcasting experiment. The out-of-sample period is marked by the gray area. Figure 3b reports the prediction range (min and max) for the news-based models, where the different predictions are determined by the release used for estimation ($r = 1, \dots, \bar{r}$). The gray area is the range (min and max) of outcomes at each point in time, for releases $r = 1, \dots, \bar{r}$.

presented under the *Estimation release* heading, i.e., the internal news-based evaluation across estimation and evaluation combinations. We observe that it is generally easier to obtain lower RMSFE values when the nowcasts are evaluated against the first release. If later data releases contain new information not available at the forecast origin, the general increase in RMSFE across evaluation releases in Table 1 is a natural outcome. The difference in terms of which release to estimate the *TVP* model class on varies too, but to a smaller degree. On average across evaluations, however, there seems to be a tendency that the nowcasts are more accurate when the *TVP* model is estimated on either the first or second release. In particular, if the fifth release is treated as the “final” release, the best performance in terms of RMSFE is obtained when constructing the nowcasts using

Table 1. Nowcasting RMSFE. Each column represents a different model, and each row reports the RMSFE when evaluated against a different release. The models 1 - 5 under the *Estimation release* heading refer to the news-based model where forecasts are rescaled using the *TVP* model class, and estimated on releases 1 - 5. The remaining columns refer to the benchmarks models described in Section 4.3. The numbers in parentheses report forecasting performance relative to the news-based model estimated on the second release. Likewise, tests for significant difference in forecasting performance are computed relative to this model, using the DM test with HAC corrected standard errors (Diebold and Mariano (1995)). *, **, *** denote the 10, 5, and 1 percent significance level, respectively.

		Estimation release					AR	RM	NB	SAM
Evaluation release		1	2	3	4	5				
	1	0.372 (1.025)	0.381 (1.000)	0.391 (0.975)	0.392 (0.973)	0.363 (1.050)	0.569* (0.670)	0.442 (0.864)	0.331 (1.152)	0.372 (1.026)
	2	0.420 (1.028)	0.431 (1.000)	0.444 (0.972)	0.450 (0.959)	0.423 (1.021)	0.679** (0.635)	0.549 (0.786)	0.403 (1.069)	0.468 (0.923)
	3	0.404 (1.030)	0.416 (1.000)	0.431 (0.967)	0.435 (0.957)	0.399 (1.043)	0.676** (0.616)	0.574* (0.726)	0.414 (1.007)	0.452 (0.922)
	4	0.422 (1.030)	0.434 (1.000)	0.452 (0.961)	0.450 (0.964)	0.425 (1.023)	0.701** (0.619)	0.613* (0.709)	0.391 (1.109)	0.454 (0.957)
	5	0.406 (0.978)	0.397 (1.000)	0.424 (0.938)	0.424 (0.937)	0.413 (0.962)	0.738*** (0.538)	0.576* (0.690)	0.393 (1.013)	0.403 (0.987)

the second release.

Compared to the two simple benchmark forecasts, *AR* and *RM*, the news-based model produces superior nowcasts. Using the second release estimation column as a reference, we see from Table 1 that irrespective of which release the nowcasts are evaluated against, the news-based model is between 16 to 46 percent better. In most cases the difference in forecasting performance is also significant, and widens with the age of the release used for evaluation, suggesting that the news-based predictions contain more information about the “true” underlying growth rate in the economy.

Compared to the more sophisticated benchmarks, *NB* and *SAM*, and continuing using the news-based model estimated on the second release as a reference forecast, we see that the news-based performance is sometimes better, sometimes worse, but on average very similar.¹⁶ There is a tendency that the two benchmarks performs relatively better when evaluated against the first release, but when using the fifth release to evaluate the predic-

¹⁶Because of the data availability for the Norges Bank predictions, see Appendix C, only 31 nowcasts are available for evaluation for these two benchmarks. In contrast, for the news-based models, and the two simpler benchmarks, 40 nowcasts are evaluated.

Table 2. Nowcasting bias. A positive number indicates that the nowcasts are on average too high relative to the outcome. See Table 2 for further details.

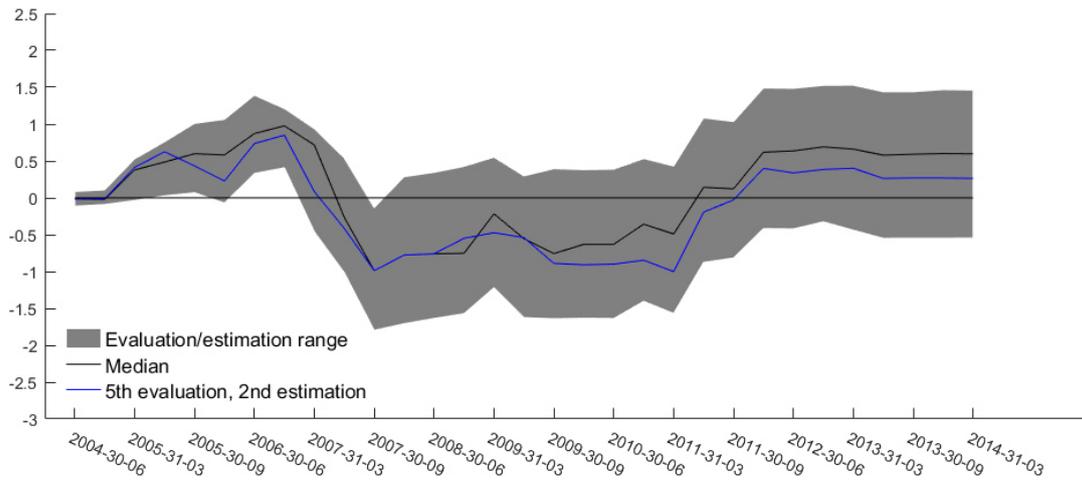
		Estimation release					AR	RM	NB	SAM
		1	2	3	4	5				
Eval. release	1	0.093	0.029	0.021	0.007	0.008	-0.181	0.011	-0.048	0.009
	2	0.089	0.025	0.017	0.003	0.004	-0.185	0.007	-0.060	-0.002
	3	0.112	0.049	0.040	0.027	0.028	-0.161	0.030	-0.039	0.019
	4	0.131	0.067	0.058	0.045	0.046	-0.143	0.049	-0.012	0.046
	5	0.078	0.014	0.005	-0.008	-0.007	-0.196	-0.004	-0.043	0.014

tions, there are hardly any differences. If anything, the *NB* predictions are slightly better than the news-based predictions, which again are slightly better than the *SAM* predictions. In contrast to above, the difference in nowcasting performance is not significant for any of the comparisons. Still, it is rather remarkable that one single DFM, containing only quarterly output growth and daily data based on a textual decomposition of a business newspaper, can produce such competitive nowcasts compared with the *NB* and *SAM* benchmarks. After all, as described in Section 4.3, these benchmarks incorporates an impressive amount of hard economic information, different models specifications and classes, and expert judgment.

Table 2 reports the biases of the nowcasts. Starting again with the internal news-based evaluation, it is interesting to note that there is a tendency for the bias to be smaller in absolute value whenever a later release is used for estimation and evaluation. For the reference model considered above, the bias is 0.014. Comparing this number with the ones achieved for the *AR*, we observe that it is always smaller in absolute value. In fact, the *AR* has a large negative bias (meaning that it consistently predicts lower growth than realized) across all evaluations. The *RM* benchmark, however, is much better, also compared with the news-based reference model. The two last columns in Table 2 report the biases for the *NB* and *SAM* benchmarks. The *NB* predictions are on average too low, while the *SAM* predictions are too high. Compared with the news-based biases, and in particular those for the reference model, we observe that they are generally higher (in absolute value), especially when evaluated against the fifth release.

The results reported thus far represent averages across the evaluation sample. The second main finding of this article is summarized in Figure 4, which reports the cumulative difference in squared prediction errors between the news-based factor model and the two benchmarks *NB* and *SAM*. Each figure shows the time path for this relative performance measure when the best performing news-based model is used, i.e., the one with the lowest RMSFE in Table 1, and also the range of outcomes when the whole battery of estimation

(a) News relative to *NB*



(b) News relative to *SAM*

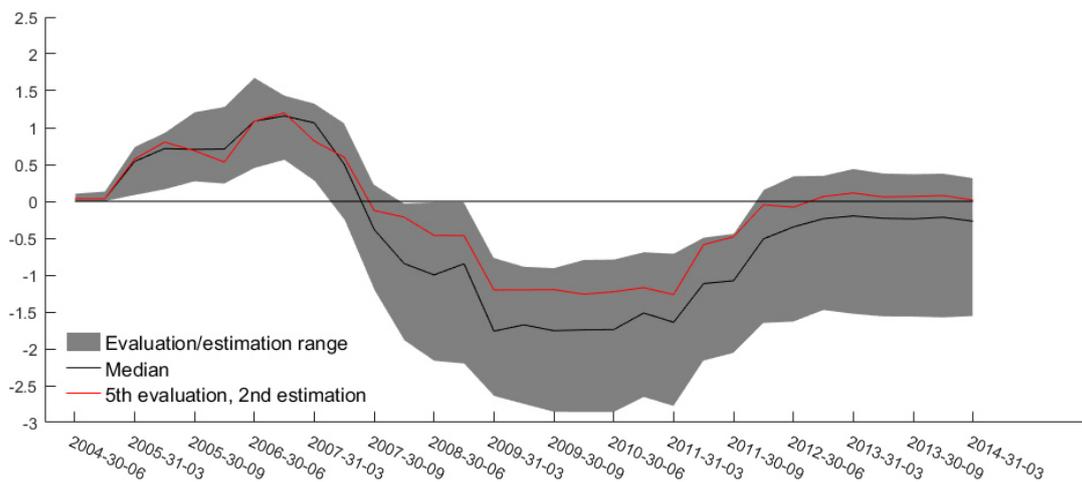


Figure 4. Cumulative difference in squared prediction error between the *NB* (*SAM*) benchmark and the news-based models. The gray shaded area shows the full range of outcomes according to this metric when each outcome in the $r \times r$ evaluation matrix is taken into account. The black solid line is the median of this range. The blue (red) solid line reports the cumulative error difference path when the news-based model with the lowest RMSFE in the estimation evaluation space is used, i.e., the one estimated on the second release and evaluated against the fifth release.

and evaluation combinations are used.¹⁷ Two interesting facts stand out: First, when the business cycle turned heading for the Great Recession, the news-based model starts to improve relative to the two benchmarks, and already in mid-2007, the news-based model is better in an absolute sense. Although the Great Recession had a very long-lasting effect on many economies, this was not the case in Norway, cf. Figure 3b. Already

¹⁷For a given evaluation, irrespective of which estimation release that is used for the news-based model, it is compared to the same benchmark. In producing the cumulative results in the figures, the time observations for which no *NB* and *SAM* forecast error exists are excluded. For readability, and given the relative poor performance of the other two simple benchmarks (*AR* and *RM*), I do not report these results in the graphs.

in 2009 the Norwegian economy was well into its recovery phase. In the comparisons reported in Figure 4, we see that this period is associated with a further improvement of the nowcasts produced by the news-based model relative to the *SAM* projections, but a small deterioration relative to the *NB* projections. Still, the news-based nowcasts continue to be better (in an absolute sense) than the two benchmarks until 2011, and particularly so when compared against the *SAM* nowcasts. Thus, although the results reported in Table 1 suggested that the three nowcasts on average were equally good (in terms of RMSFE), there seems to be a substantial difference in their ability to capture turning points. In particular, if RMSFE scores had been computed up until 2011, the news-based reference model would have performed roughly 15 percent better than the two Norges Bank benchmarks. Second, something happens around early 2011, which is picked up by the two benchmarks, but not the news-based model. Accordingly, the relative performance of the latter worsens until 2012, before it stabilizes around zero, indicating equal predictive ability.

To summarize the findings reported in this section, it is clear that the news-based model on average can produce real-time nowcasts with an RMSFE and bias as low as those produced by expert judgment (*NB*) and a state-of-the-art forecast combination framework (*SAM*). When looking more closely on how the nowcast performance varies across the business cycle, the good performance of the news-based model seems to be associated with its ability to capture turning points in the economy better than the benchmarks.

5.2 Could news add value for policy makers?

The correlations between the nowcasts produced by the news-based model and the two benchmarks *NB* and *SAM* are high and vary around roughly 0.7. Still, the time-varying relative performance documented in Figure 4 suggests that if the central bank could have exploited the news-based forecasts during the evaluation sample looked at here, news might have added value. To investigate this hypothesis more thoroughly, I follow [Romer and Romer \(2008\)](#), and run a number of regressions based on the following equation:

$$\Delta GDP_{t_q}^5 = \alpha + \beta^{NB} NB_{t_q} + \beta^{SAM} SAM_{t_q} + \beta^{News} News_{t_q} + e_{t_q} \quad (12)$$

Here, NB_{t_q} , SAM_{t_q} , and $News_{t_q}$ are the nowcasts for period t_q produced by *NB*, *SAM*, and the news-based model, respectively. I use the fifth release of actual GDP growth as dependent variable because this is the most revised release, while the news-based predictions are those produced by the model estimated using the second release (the reference model used above). The primary object of interest is to investigate if β^{News} is

Table 3. Ordinary least squares estimates - original sample. The dependent variable in each regression is $\Delta GDP_{t,q}^5$. R^2 denotes the adjusted coefficient of determination. Standard errors (in parenthesis) and test statistics are derived using a residual bootstrap. *, **, ***, denote the 10, 5, and 1 percent significance level, respectively.

α	β^{NB}	β^{SAM}	β^{News}	R^2	N
-0.218 (0.139)	0.403** (0.165)	0.578 (0.347)	0.330 (0.231)	0.657	31
-0.253 (0.153)	0.501*** (0.164)	0.869** (0.322)		0.637	31
-0.041 (0.077)	0.589*** (0.165)		0.492** (0.215)	0.638	31
-0.262* (0.152)		0.913** (0.349)	0.427* (0.238)	0.630	31

significantly different from zero: Conditional on the actual nowcasts produced by Norges Bank, could news-based predictions have added value?

The results reported in Table 3 suggest that the answer to this question is yes. First, when all three nowcasts are projected on the actual outturn, only the *NB* predictions seem to be significant. However, given the relatively short sample available for estimation, and the high correlation between the predictors, these estimates are likely not very trustworthy. As seen in rows two to four of the table, when dropping one of the three regressors, the two remaining always become significant. In particular, we see that the news-based predictions can add value in terms of predicting output growth, even after conditioning on either the *NB* or *SAM* nowcasts.

One caveat with the results presented in Table 3 is that relatively few observations are available for estimation. As described in Appendix C, this is due to missing Norges Bank data points affecting the sample up until 2012. However, since the results reported in Figure 4 suggest that the news-based nowcasts perhaps could have added more value during the middle part of the sample, it is unfortunate that this part of the sample is somewhat down-weighted in the regressions. To address this issue, I do the following: Instead of dropping the observations for which no *NB* and *SAM* nowcasts exist, interpolated values are used to fill in the missing observations. The interpolation is linear, and is performed between the two closest neighboring data points to the missing observations. Arguably, this procedure is ad-hoc, and constructs nowcast series for the two benchmarks that actually do not exist. Still, it permits running regressions utilizing the full sample of available news-based nowcasts.¹⁸

¹⁸I have experimented with other ways of handling the missing observations, such as using the *RM* predictions, and using the two-step ahead predictions from the *NB* and *SAM* benchmarks. Qualitatively, none of these alternative methods changes the results.

Table 4. Ordinary least squares estimates - forecast efficiency test. The dependent variable is $R_{t_q} = \Delta GDP_{t_q}^5 - \Delta GDP_{t_q}^1$. Standard errors (in parenthesis) and test statistics are derived using a residual bootstrap. *, **, ***, denote the 10, 5, and 1 percent significance level, respectively.

α	β^{GDP}	β^{News}	R^2	N
-0.166 (0.110)	-0.210 (0.133)	0.45*** (0.145)	0.230	40

The results of the alternative experiment are reported in Table 6, in Appendix A: Irrespective of how the regression is specified, the news-based predictions are now always significant at either the five or one percent level. Thus, taken together the results in Tables 3 and 6 clearly suggest that if the central bank could have exploited the news-based forecasts, it would have added value.

Could the news-based predictions also be informative for the statistical agency producing the GDP statistics? The results reported in Figure 3b suggested that data revisions are large. As explained in Section 4.1, if these revisions are unpredictable, each new release of GDP contains new information obtained by the statistical agency after the time of the first release. Conversely, if the revisions are predictable, the revisions are said to contain noise (Mankiw and Shapiro (1986)). In this latter case, noise reduction could be possible using information that could have been available to the statistical agency when publishing their initial release. A standard way to distinguish between these two views is to use forecast efficiency tests (Mincer and Zarnowitz (1969)). Let $R_{t_q} = \Delta GDP_{t_q}^5 - \Delta GDP_{t_q}^1$, then, the following regression is estimated:

$$R_{t_q} = \alpha + \beta^{GDP} \Delta GDP_{t_q}^1 + \beta^{News} News_{t_q} + e_{t_q} \quad (13)$$

Under the news view (not news topics), revisions must be mean zero and the coefficients in (13) should not be significant; under noise, the revisions need not be mean zero and the coefficients might be significant. Here it is particularly interesting if β^{News} is significantly different from zero? If so, it means that the statistical agency could have used the nowcasts produced by the news-based model to improve their own first release of GDP (given the maintained assumption that the fifth release is closer to the unknown true state of the economy than the first release).

The results reported in Table 4 indicate that the news-based predictions could be informative also for the statistical agency producing GDP statistics. In particular, the β^{News} coefficient is highly significant and positive. In contrast, neither the constant nor the β^{GDP} coefficients are significantly different from zero. In unreported results, when running the regression without the news-based predictions as an extra explanatory variable, this holds through: The null hypothesis of forecast rationality cannot be rejected

Table 5. News based information flow and RMSFE. T is the last day of the quarter T_q . GDP for the previous quarter (T_{q-1}) is released around $T - 40$. The nowcasts constructed for $T - d \leq T - 40$ are two-step ahead predictions (relative to the quarter T_q), and the predictions constructed for $T + d > T$ are termed backcasts. The news-based predictions are constructed using the *TVP* model estimated on the second release. All evaluations are done with respect to the fifth GDP release. Thus, by construction, the accuracy of the nowcast produced using time T information is identical to that reported in the last row and second column of Table 1.

Type	Nowcast for T_q				Backcast for T_q		
	$T_q - 2$		$T_q - 1$		$T_q - 1$		
Quarterly information							
Daily information	$T - 80$	$T - 60$	$T - 40$	$T - 20$	T	$T + 10$	$T + 30$
RMSFE	0.460	0.424	0.425	0.426	0.397	0.387	0.386
Cumulative imp. (%)		-7.87	-7.59	-7.31	-14.09	-16.55	-16.85

when conditioning on only the first release itself. However, when including the news-based predictions, forecast rationality is rejected. One potential reason for this finding could be that economic agents respond to the information contained in the daily newspaper, and thereby endogenously shape the within quarter and subsequent outcomes. In fact, such an interpretation is very consistent with that postulated in the news driven business cycle view, see, e.g., [Larsen and Thorsrud \(2015\)](#) and the references therein.

5.3 High frequency updating and the output news feed

Two important features of newer nowcasting methods are their ability to provide the forecast user with within the quarter updates as more information becomes available, and also a decomposition of how this information affects the nowcast updates, see, e.g., [Giannone et al. \(2008\)](#) and [Banbura et al. \(2013\)](#). Both of these features carry through to the news-based methodology proposed here. First, newspaper news is available at a daily frequency, making high frequency updates of the forecasts throughout the quarter a simple matter of on-line inference. Second, given the large set of news topics that enter into the news-based DFM, it is possible to decompose any high frequency changes in the nowcast of output growth into news topic contributions. For policy makers in particular, as reflected in the broad coverage of various financial and macroeconomic data in monetary policy reports and national budgets, explaining what drives a forecast might be equally important as the forecast itself.

Table 5 illustrates how the information flow affects the accuracy of the nowcast (and backcast) for a generic quarter T_q . The results are obtained by running an experiment similar to that described in Section 4.2, but now assuming that daily newspaper infor-

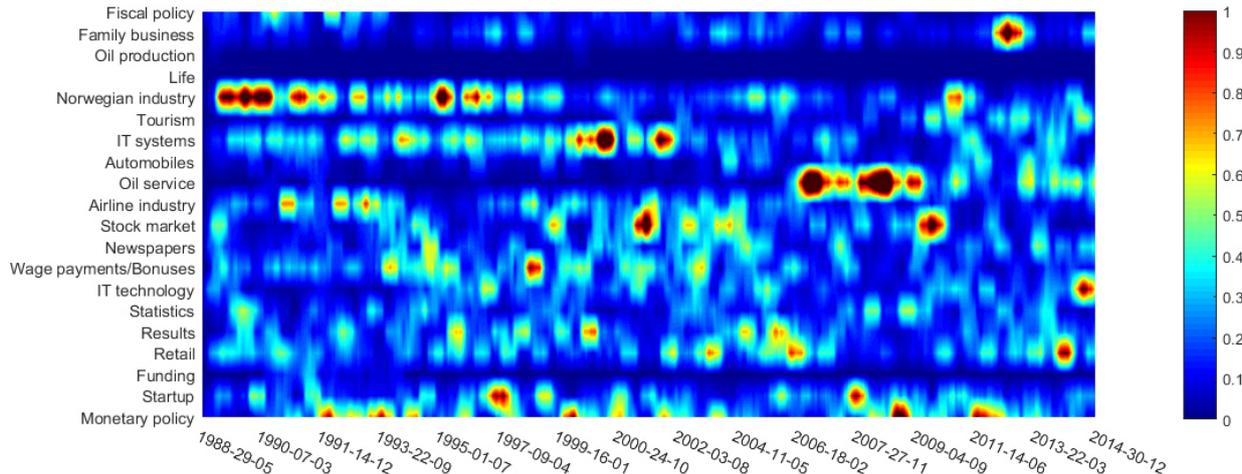


Figure 5. News topics and their (median) contribution to the latent factor estimate, and thus GDP forecast, through time. The forecast decomposition illustrated here is similar to the ones used in the related factor modeling literature, see, e.g., [Banbura et al. \(2013\)](#). Technically, the decomposition is constructed using the Kalman Filter iterations and decomposing the state evolution at each updating step into news contributions, see [Appendix G](#). For readability, the contributions are normalized to sum to one at each time period, and smoothed using a 90-day (backward looking) moving average filter. A higher value indicates a higher weight.

mation is available only at roughly 20-day windows between $T - 80$ and $T + 30$, where T is the last day of the quarter T_q . Since GDP for the previous quarter (T_{q-1}) is released around $T - 40$, the nowcasts constructed for $T - d \leq T - 40$ will be two-step ahead predictions (relative to the quarter T_q), while the predictions for $T - d > T - 40$ will be one-step ahead predictions. Finally, the predictions constructed for $T + d > T$ will be termed backcasts because the daily information set used to construct the predictions are extended into the subsequent quarter ($T_q + 1$).

As seen in [Table 5](#), more information reduces the news-based forecast errors. For example, comparing the accuracy of the nowcasts made using $T - 80$ information to that using $T + 30$ information (backcast), we observe a roughly 16 percent improvement. Perhaps somewhat surprising, however, the largest gain in relative forecasting performance is not obtained when GDP for the previous quarter is released (after $T - 40$), but earlier in the quarter, when daily newspaper information starts to accumulate. Comparing the results in [Table 5](#) with those in [Table 1](#), we see that already early in the quarter the news-based model produces nowcasts that are almost as accurate as those produced by the Norges Bank benchmarks at the end of the quarter (at time T). When utilizing $T + d$ information, i.e., backcasting, the accuracy of the news-based predictions become marginally better.

The second point addressed above is exemplified in [Figure 5](#), which shows that there is a substantial variation in terms of which news topics contribute to the nowcasts. For example, during the 1990s, the *Norwegian industry* topic seems to contain important

information, while the *Oil Service* topic becomes particularly important during the mid-2000s. Other topics, like *IT systems*, *Stock market*, and *Monetary policy* varies much more in their relevance, but all seem to contain important news at different points in time. Conversely, and perhaps somewhat surprising, the *Oil production* topic does not seem to add any useful information at any point in time.¹⁹

6 Conclusion

The agents in the economy use a plethora of high frequency information, including media, to guide their actions and thereby shape aggregate economic fluctuations. In this paper I show how unstructured textual data collected from a major business newspaper can be used to improve nowcasting performance with regards to quarterly GDP growth, on average, and around turning points. In particular, by decomposing the textual data into daily news topics, and using a mixed frequency time-varying Dynamic Factor Model, I show that it is possible to obtain nowcasts that perform up to 50 percent better than forecasts from simple time series models, and at times up to 15 percent better than a state-of-the-art nowcasting system. It is noteworthy that, if a big bank of experts, here represented by official central bank nowcasts, could have utilized the methodology proposed, it would have resulted in lower forecasting errors, especially around the period of the Great Recession. Moreover, if the statistical agency producing the output growth statistics itself had utilized the methodology proposed, it would have resulted in a less noisy revision process. Thus, news reduces noise.

The (macro)economic literature utilizing textual information and alternative (Big) data sources is fast growing, but still in its early stages. Key to the results obtained in this analysis is the use of a Latent Dirichlet Allocation (LDA) model to decompose the unstructured text into news topics. Compared with existing textual approaches used in economics and finance, such as word counting, sentiment classification, and collecting web search terms and trends, the approach taken here is undoubtedly highly data and computationally intensive. However, one could argue that the LDA approach offers several advantages in terms of automation and interpretation. Still, to what extent these advantages also make the topic model a preferred tool to work with when unstructured textual data are used to nowcast macroeconomic variables is an open question worthy future research.

¹⁹The observant reader, comparing these results with the ones presented in [Thorsrud \(2016\)](#), will notice that the news decomposition differs. This is natural, since the news-based factor model used here is estimated on the first release of output growth, and not the “final” vintage, as in [Thorsrud \(2016\)](#).

References

- Aastveit, K. A., K. R. Gerdrup, and A. S. Jore (2011). Short-term forecasting of GDP and inflation in real-time: Norges Bank's system for averaging models. Staff Memo 2011/9, Norges Bank.
- Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics* 32(1), 48–68.
- Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014). Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences. *Journal of Business & Economic Statistics* 32(4), 483–500.
- Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of Euro area GDP growth. *The Econometrics Journal* 14(1), 25–44.
- Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the Euro area GDP. *International Journal of Forecasting* 20(3), 447 – 460.
- Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics* 176(1), 18 – 29.
- Bai, J. and P. Wang (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics* 178(2), 794–804.
- Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. Working Paper Series 1564, European Central Bank.
- Banbura, M., D. Giannone, and L. Reichlin (2011). *Nowcasting*. The Oxford Handbook of Economic Forecasting (Oxford Handbooks in Economics). New York: Oxford University Press.
- Bholat, D., S. Hansen, P. Santos, and C. Schonhardt-Bailey (2015). Text mining for central banks: Handbook. *Centre for Central Banking Studies* 33, pp. 1–19.
- Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2011). Weights and pools for a Norwegian density combination. *The North American Journal of Economics and Finance* 22(1), 61–76.
- Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2012). Does forecast combination improve Norges Bank inflation forecasts? *Oxford Bulletin of Economics and Statistics* 74(2), 163–179.

- Bjørnland, H. C. and L. A. Thorsrud (2015). Commodity prices and fiscal policy design: Procyclical despite a rule. Working Papers 0033, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM* 55, 77–84.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3, 993–1022.
- Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives* 28(2), 153–76.
- Breitung, J. and C. Schumacher (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting* 24(3), 386–398.
- Carter, C. K. and R. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika* 81(3), 541–553.
- Chang, J., S. Gerrish, C. Wang, J. L. Boyd-graber, and D. M. Blei (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Eds.), *Advances in Neural Information Processing Systems* 22, pp. 288–296. Curran Associates, Inc.
- Chen, R. and J. S. Liu (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62(3), 493–508.
- Choi, H. and H. Varian (2012). Predicting the present with Google trends. *Economic Record* 88(s1), 2–9.
- Creal, D. (2009). A survey of sequential Monte Carlo methods for economics and finance. Serie Research Memoranda 0018, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.
- Croushore, D. (2006). Chapter 17 forecasting with real-time macroeconomic data. Volume 1 of *Handbook of Economic Forecasting*, pp. 961 – 982. Elsevier.
- Del Negro, M. and C. Otrok (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Reports 326, Federal Reserve Bank of New York.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–63.

- Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.
- Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons (2012). Journalists and the stock market. *Review of Financial Studies* 25(3), 639–679.
- Ellis, C., H. Mumtaz, and P. Zabczyk (2014). What Lies Beneath? A Time-varying FAVAR Model for the UK Transmission Mechanism. *Economic Journal* 0(576), 668–699.
- Evans, M. D. D. (2005). Where Are We Now? Real-Time Estimates of the Macroeconomy. *International Journal of Central Banking* 1(2).
- Faust, J., J. H. Rogers, and J. H. Wright (2005). News and Noise in G-7 GDP Announcements. *Journal of Money, Credit and Banking* 37(3), 403–19.
- Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* 55(4), 665–676.
- Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America* 101(Suppl 1), 5228–5235.
- Hansen, S. and M. McMahon (2015). Shocking Language: Understanding the macroeconomic effects of central bank communication. Discussion Papers 1537, Centre for Macroeconomics (CFM).
- Hansen, S., M. McMahon, and A. Prat (2014). Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. CEP Discussion Papers 1276, Centre for Economic Performance, LSE.
- Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Books. Cambridge University Press.
- Heinrich, G. (2009). Parameter estimation for text analysis. Technical report, Fraunhofer IGD.
- Kuzin, V., M. Marcellino, and C. Schumacher (2011). Midas vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting* 27(2), 529–542.
- Larsen, V. H. and L. A. Thorsrud (2015). The Value of News. Working Papers 0034, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.

- Larsen, V. H. and L. A. Thorsrud (2016). Asset prices, news topics, and the media effect. Mimeo, Norges Bank.
- Lopes, H. F. and C. M. Carvalho (2007). Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes. *Journal of Statistical Planning and Inference* (137), 3082–3091.
- Loughran, T. and B. McDonald (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* 66(1), 35–65.
- Mankiw, N., D. E. Runkle, and M. D. Shapiro (1984). Are preliminary announcements of the money stock rational forecasts? *Journal of Monetary Economics* 14(1), 15 – 27.
- Mankiw, N. G. and M. D. Shapiro (1986). News or Noise? An Analysis of GNP Revisions. NBER Working Papers 1939, National Bureau of Economic Research, Inc.
- Marcellino, M., M. Porqueddu, and F. Venditti (2013). Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility. Economic working papers 896, Bank of Italy, Economic Research and International Relations Area.
- Michael K. Pitt, N. S. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 94(446), 590–599.
- Mincer, J. A. and V. Zarnowitz (1969). The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER Chapters, pp. 3–46. National Bureau of Economic Research, Inc.
- Nakajima, J. and M. West (2013). Bayesian Analysis of Latent Threshold Dynamic Models. *Journal of Business & Economic Statistics* 31(2), 151–164.
- Nyman-Andersen, P. (2016). Big data: The hunt for timely insights and decision certainty. IFC Working Paper 14, Bank for International Settlements.
- Pang, B., L. Lee, and S. Vaithyanathan (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, Stroudsburg, PA, USA, pp. 79–86. Association for Computational Linguistics.
- Peress, J. (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *The Journal of Finance* 69(5), 2007–2043.
- Romer, C. D. and D. H. Romer (2008). The FOMC versus the staff: Where can monetary policymakers add value? *American Economic Review* 98(2), 230–35.

- Stock, J. and M. Watson (2012). Disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity Spring 2012*, 81–135.
- Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20(2), 147–62.
- Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41(3), 788–829.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3), 1139–1168.
- Tetlock, P. C. (2014). Information Transmission in Finance. *Annual Review of Financial Economics* 6(1), 365–384.
- Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008). More Than Words: Quantifying Language to Measure Firms’ Fundamentals. *Journal of Finance* 63(3), 1437–1467.
- Thorsrud, L. A. (2016). Words are the new numbers: A newsy coincident index of business cycles. Working Papers 44, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.
- Timmermann, A. (2006). *Forecast Combinations*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 4, pp. 135–196. Elsevier.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2), 3–28.
- Zhou, X., J. Nakajima, and M. West (2014). Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *International Journal of Forecasting* 30, 963–980.

Appendices

Appendix A Additional results

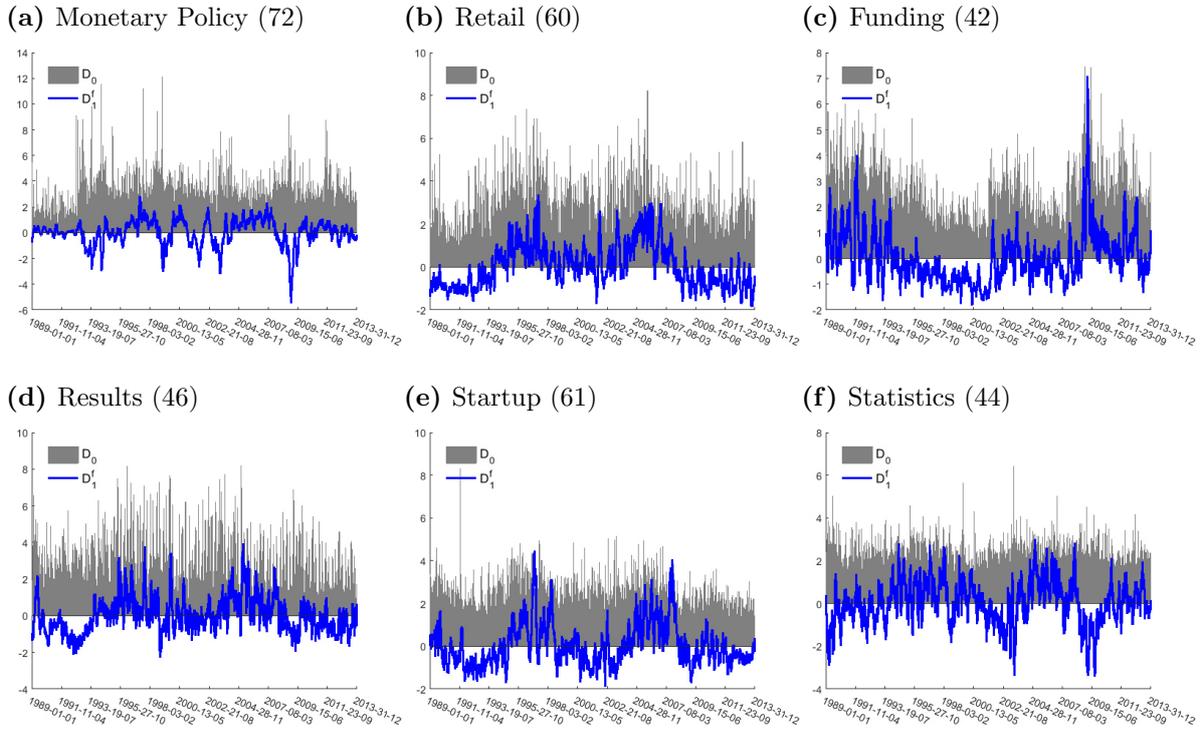


Figure 6. Individual news topics (topic numbers, cf. Table 7 in Appendix A, in parenthesis). The gray bars and blue lines report topic time series from step one and three of the data processing, respectively. See the text for details.

Table 6. Ordinary least squares estimates - interpolated sample. The dependent variable in each regression is $\Delta GDP_{t_a}^5$. R^2 denotes the adjusted coefficient of determination. Standard errors (in parenthesis) and test statistics are derived using a residual bootstrap. *, **, ***, denote the 10, 5, and 1 percent significance level, respectively.

α	β^{NB}	β^{SAM}	β^{News}	R^2	N
-0.214 (0.142)	0.525*** (0.180)	0.275 (0.319)	0.509** (0.195)	0.698	40
-0.281 (0.179)	0.727*** (0.214)	0.686** (0.328)		0.639	40
-0.124 (0.098)	0.617*** (0.163)		0.576*** (0.172)	0.700	40
-0.266 (0.161)		0.687** (0.324)	0.657*** (0.220)	0.655	40

Table 7. Estimated topics and labeling. The topics are labeled based on the meaning of the most important words, see the text for details. The “Corr” column reports the topics’ correlation with linearly interpolated daily *GDP*^a, see [Thorsrud \(2016\)](#) for details. The topics included in the DMF model, see, Section 3, are marked with bold. The words are translated from Norwegian to English using Google Translate.

Topic	Label	Corr	First words
Topic 0	Calender	0.03	january, march, october, september, november, february
Topic 1	Family business	0.18	family, foundation, name, dad, son, fortune, brothers
Topic 2	Institutional investing	0.10	fund, investments, investor, return, risk, capital
Topic 3	Justice	0.04	lawyer, judge, appeal, damages, claim, supreme court
Topic 4	Surroundings	0.18	city, water, meter, man, mountain, old, outside, nature
Topic 5	Housing	0.14	housing, property, properties, apartment, square meter
Topic 6	Movies/Theater	0.08	movie, cinema, series, game, producer, prize, audience
Topic 7	Argumentation	0.11	word, besides, interesting, i.e., in fact, sure, otherwise
Topic 8	Unknown	0.09	road, top, easy, hard, lift, faith, outside, struggle, fast
Topic 9	Agriculture	0.03	industry, support, farmers, export, production, agriculture
Topic 10	Automobiles	0.18	car, model, engine, drive, volvo, ford, møller, toyota
Topic 11	USA	0.09	new york, dollar, wall street, president, usa, obama, bush
Topic 12	Banking	0.00	dnb nor, savings bank, loss, brokerage firm, kreditkassen
Topic 13	Corporate leadership	0.05	position, chairman, ceo, president, elected, board member
Topic 14	Negotiation	0.04	solution, negotiation, agreement, alternative, part, process
Topic 15	Newspapers	0.22	newspaper, media, schibsted, dagbladet, journalist, vg
Topic 16	Health care	0.00	hospital, doctor, health, patient, treatment, medication
Topic 17	IT systems	0.17	it, system, data, defense, siem, contract, tanberg, deliver

Continued on next page

Table 7 – continued from previous page

Topic	Label	Corr	First words
Topic 18	Stock market	0.23	stock exchange, fell, increased, quote, stock market
Topic 19	Macroeconomics	0.07	economy, budget, low, unemployment, high, increase
Topic 20	Oil production	0.18	statoil, oil, field, gas, oil company, hydro, shelf, stavanger
Topic 21	Wage payments	0.26	income, circa, cost, earn, yearly, cover, payed, salary
Topic 22	Norwegian regions	0.17	trondheim, llc, north, stavanger, tromsø, local, municipality
Topic 23	Family	0.04	woman, child, people, young, man, parents, home, family
Topic 24	Taxation	0.03	tax, charge, revenue, proposal, remove, wealth tax, scheme
Topic 25	EU	0.04	eu, eea, commission, european, brussel, membership, no
Topic 26	Norwegian industry	0.20	hydro, forest, factory, production, elkem, industry, produce
Topic 27	Unknown	0.07	man, he, friend, smile, clock, evening, head, never, office
Topic 28	Norwegian groups	0.09	orkla, storebrand, merger, bid, shareholder, acquisitions
Topic 29	UK	0.06	british, london, great britain, the, of, pound, england
Topic 30	Narrative	0.03	took, did, later, never, gave, stand, happened, him, began
Topic 31	Shipping	0.10	ship, shipping, dollar, shipowner, wilhelmsen, fleet, proud
Topic 32	Projects	0.10	project, nsb, development, fornebu, entrepreneurship
Topic 33	Oil price	0.11	dollar, oil price, barrel, oil, demand, level, opec, high
Topic 34	Sports	0.00	olympics, club, football, match, play, lillehammer, sponsor
Topic 35	Organizations	0.10	leader, create, organization, challenge, contribute, expertise
Topic 36	Drinks	0.13	wine, italy, taste, drinks, italian, fresh, fruit, beer, bottle
Topic 37	Nordic countries	0.04	swedish, sweden, danish, denmark, nordic, stockholm
Topic 38	Airline industry	0.21	sas, fly, airline,norwegian, braathens, airport, travel

Continued on next page

Table 7 – continued from previous page

Topic	Label	Corr	First words
Topic 39	Entitlements	0.02	municipality, public, private, sector, pension, scheme
Topic 40	Employment conditions	0.08	cut, workplace, measures, salary, labor, working, employ
Topic 41	Norwegian politics	0.05	høyere, party, ap, labor party, stoltenberg, parlament, frp
Topic 42	Funding	0.31	loan, competition, creditor, loss, bankruptcy, leverage
Topic 43	Literature	0.01	book, books, read, publisher, read, author, novel, wrote
Topic 44	Statistics	0.27	count, increase, investigate, share, average, decrease
Topic 45	Watercraft	0.01	ship, boat, harbor, strait, shipowner, on board, color
Topic 46	Results	0.31	quarter, surplus, deficit, tax, group, operating profit, third
Topic 47	TV	0.12	tv, nrk, channel, radio, digital, program, media
Topic 48	International conflicts	0.10	war, africa, irak, south, un, army, conflict, troops, attack
Topic 49	Political elections	0.02	election, party, power, politics, vote, politician, support
Topic 50	Music	0.09	the, music, record, of, in, artist, and, play, cd, band, song
Topic 51	Oil service	0.19	rig, dollar, contract, option, offshore, drilling, seadrill
Topic 52	Tourism	0.21	hotel, rom, travel, visit, stordalen, tourist, guest
Topic 53	Unknown	0.16	no, ting, think, good, always, pretty, actually, never
Topic 54	Aker	0.11	aker, kværner, røkke, contract, shipyard, maritime
Topic 55	Fishery	0.16	fish, salmon, seafood, norway, tons, nourishment, marine
Topic 56	Europe	0.08	german, russia, germany, russian, west, east, french, france
Topic 57	Law and order	0.06	police, finance guards, aiming, illegal, investigation
Topic 58	Business events	0.00	week, financial, previous, friday, wednesday, tdn, monday
Topic 59	Supervision	0.10	report, information, financial supervision, enlightenment

Continued on next page

Table 7 – continued from previous page

Topic	Label	Corr	First words
Topic 60	Retail	0.31	shop, brand, steen, rema, reitan, as, group, ica, coop
Topic 61	Startup	0.28	bet, cooperation, establish, product, party, group
Topic 62	Food	0.19	food, restaurant, salt, nok, pepper, eat, table, waiter
Topic 63	Listed stocks	0.11	shareholder, issue, investor, holding, stock exchange listing
Topic 64	Asia	0.09	china, asia, chinese, india, hong kong, south, authorities
Topic 65	Art	0.09	picture, art, exhibition, gallery, artist, museum, munch
Topic 66	Disagreement	0.08	criticism, express, asserting, fault, react, should, alleging
Topic 67	Debate	0.15	degree, debate, context, unequal, actually, analysis
Topic 68	Life	0.18	man, history, dead, him, one, live, church, words, strokes
Topic 69	Distribution	0.18	customer, post, product, offers, service, industry, firm
Topic 70	Telecommunication	0.08	telenor, mobile, netcom, hermansen, telia, nokia, ericsson
Topic 71	IT technology	0.21	internet, net, pc, microsoft, technology, services, apple
Topic 72	Monetary policy	0.33	interest rate, central bank, euro, german, inflation, point
Topic 73	Education	0.04	school, university, student, research, professor, education
Topic 74	Government regulations	0.03	rules, authorities, competition, regulations, bans
Topic 75	Trade organizations	0.16	lo, nho, members, forbund, strike, organization, payroll
Topic 76	Fear	0.04	fear, emergency, hit, severe, financial crisis, scared
Topic 77	Fiscal policy	0.19	suggestions, parliamentary, ministry, selection, minister
Topic 78	Energy	0.05	energy, emissions, statkraft, industry, environment
Topic 79	Foreign	0.07	foreign, abroad, japan, japanese, immigration, games

Appendix B OLS versus TVP

As described in Section 4.1 of the main paper, I use a time-varying parameter model to rescale the news-based forecasts. As a robustness check, I have also done the rescaling using a static model, estimated using simple ordinary least squares (*OLS*). A formal evaluation of the nowcasts produced by the two model classes *TVP* and *OLS* are illustrated in Figures 7a and 7b, which report relative RMSFE and bias of each estimation and evaluation combination considered. Two impressions stand out: First, the performance of the two model classes is very similar, and we do not observe any significant difference between the two in terms of RMSFE across evaluation vintages. At best (worst), the *TVP* class is roughly 7 percent better (worse) than the *OLS* class. Still, for a given evaluation release, it is almost always the case that the model with the lowest RMSFE belongs to the *TVP* class. Across all the 25 different evaluation and estimation combinations, the *TVP* class has a lower RMSFE in 60 percent of the cases. Second, there is a substantial difference in the biases. In fact, the nowcasts produced by the *OLS* class are on average always too low, resulting in a negative bias. In contrast, the nowcasts produced by the *TVP* class are on average almost always too high, resulting in a positive bias (results not shown). However, comparing the absolute value of the biases across model classes, I find that in 85 percent of the cases, the nowcasts produced by the *TVP* model class is smaller (closer to zero), cf. Figure 7b.

Figure 8 provides a comparison of the two model class predictions across time, and reports the cumulative difference in squared prediction error between the *TVP* and *OLS* model class. At each point in time, a value above zero indicates that the *OLS* class is better, and vice versa. As seen in Figure 8, during the first part of the evaluation sample, the *OLS* class performs slightly better than the *TVP* class, almost irrespective of which release is used for evaluation and estimation. Around the Great Recession, the gap between the two classes widens further. Only well into the recovery phase of the recession does the *OLS* class start to deteriorate relative to the *TVP* class, and at the end of the sample, the difference between the two is basically zero. It is tempting to attribute the relatively good performance of the *TVP* class in the latter part of the sample to the mediocre growth rates that many developed countries have experienced in the aftermath of the Great Recession. To the extent that these represent a low frequency change in the trend growth rate, this would potentially be picked up much faster in the *TVP* models than in the *OLS* models. If one focuses on the best performing model in terms of RMSFE across the whole sample, it turns out that this model belongs to the *TVP* class when estimated on the second release and evaluated against the fifth release. See the red line in the figure.

I draw two main conclusions from the evaluations conducted above. First, although

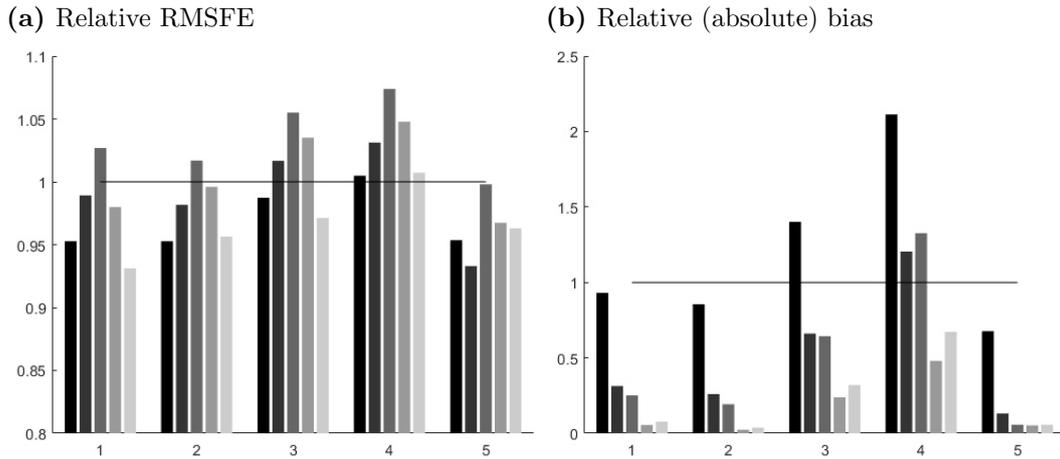


Figure 7. Relative RMSFE and bias (in absolute value). A value below one indicates that the *TVP* class has a lower RMSFE or bias (in absolute value) than the *OLS* class. The evaluations are done for the \bar{r} different releases, as indicated on the x-axis. For each evaluation, \bar{r} different releases are used to estimate the models and construct nowcasts, as indicated by the \bar{r} bars for each evaluation ticker.

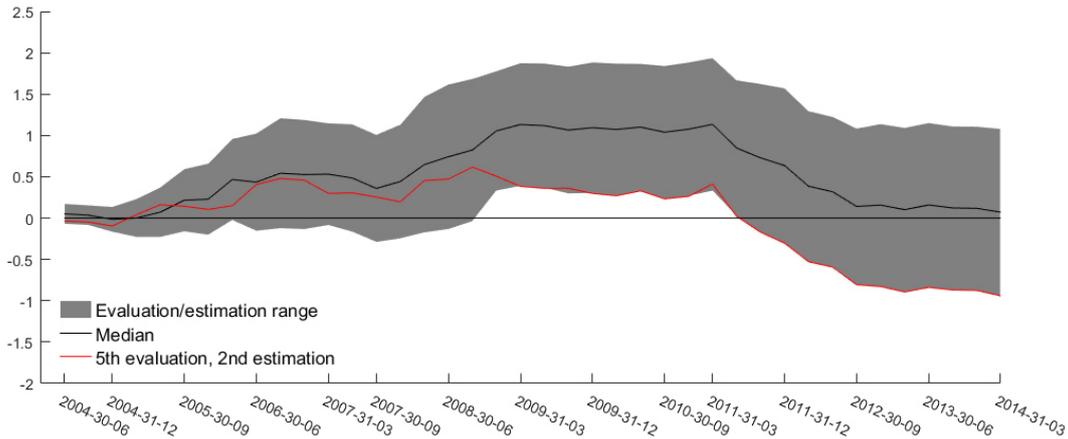


Figure 8. Cumulative difference in squared prediction error between the *TVP* and *OLS* model classes. The gray shaded area shows the full range of outcomes according to this metric when each outcome in the $r \times r$ evaluation matrix is taken into account - containing a total of 25 cumulative error differences. The black solid line is the median of this range. The red solid line reports the path for the evaluation and estimation combination that produces the lowest RMSFE across the whole sample.

the differences are small, both the RMSFE and bias comparisons indicate that the best performing specification allows for time-varying parameters. Second, the variation in performance across time between the two model classes suggests that a well devised forecast combination scheme might improve upon only using one of them. To keep the analysis focused on the main contribution of this paper, I do not follow up on this insight here, but leave such extensions for future investigation.

Appendix C The Norges Bank benchmarks

Starting in 2009, both the *NB* and *SAM* predictions are published in Norges Bank's Monetary Policy Report. Prior to 2009, however, only the official Norges Bank predictions (*NB*) were published. Still, real-time predictions from the *SAM* system exist for earlier periods. In total I have access to 31 real-time predictions from Norges Bank, covering the sample 2003:Q4-2014:Q1.

One caveat with using the Norges Bank forecasts as benchmark models is that up until 2012 the *NB* and *SAM* forecasts were only produced three times every year; in March, June, and October. During this period I assign the March (1-step) predictions from Norges Bank to the fourth quarter the previous year, the June (1-step) predictions to the second quarter of the same year, and the October (1-step) predictions to the third quarter. Thus, from 2003 to 2012, Norges Bank's predictions for the fourth quarter will be backcasts, while the other predictions (published in June and October) will be nowcasts. Compared to the informational assumptions used when running the news-based model, see Section 4.2, the Norges Bank benchmarks will potentially have a three-month informational advantage for the fourth quarter predictions, a two-month disadvantage for the third quarter predictions, and roughly similar information available for the second quarter nowcast.

The latest documentation of the *SAM* system used for policy at Norges Bank can be found in [Aastveit et al. \(2011\)](#). Here I briefly note that the system contains almost 500 individual models. These models belong to three different model classes, known to be particularly useful for nowcasting: Bridge regressions, Mixed frequency vector autoregressions, and Dynamic Factor Models, see, e.g., [Baffigi et al. \(2004\)](#), [Angelini et al. \(2011\)](#), [Kuzin et al. \(2011\)](#), [Breitung and Schumacher \(2008\)](#), and [Giannone et al. \(2008\)](#). For all model classes the system is set up such that monthly information can be used together with quarterly information when making predictions for output growth. Each individual model produces, in real-time, both a density and point forecast. For each forecast round, these individual predictions are combined in two steps using a linear opinion pool. In the first step, models belonging to the same class are combined, and in the second step, the combined forecasts from step one are combined to produce the finalized prediction. At each combination step a larger weight is put on models that have historically performed well, where the weights are either computed using historical logarithmic scores or mean squared prediction errors.

Appendix D The textual data

D.1 Filtering the news corpus

To clean the raw textual dataset a stop-word list is first employed. This is a list of common words not expected to have any information relating to the subject of an article. Examples of such words are *the*, *is*, *are*, and *this*. The most common Norwegian surnames and given names are also removed. In total the stop-word list together with the list of common surnames and given names removed roughly 1800 unique tokens from the corpus. Next, an algorithm known as stemming is run. The objective of this algorithm is to reduce all words to their respective word stems. A word stem is the part of a word that is common to all of its inflections. An example is the word *effective* whose stem is *effect*. Finally, a measure called *tf-idf*, which stands for term frequency - inverse document frequency, is calculated. This measures how important all the words in the complete corpus are in explaining single articles. The more often a word occurs in an article, the higher the *tf-idf* score of that word. On the other hand, if the word is common to all articles, meaning the word has a high frequency in the whole corpus, the lower that word's *tf-idf* score will be. Around 250 000 of the stems with the highest *tf-idf* score are kept, and used as the final corpus.

D.2 LDA estimation and specification

The LDA model was developed in [Blei et al. \(2003\)](#). Here the estimation algorithm described in [Griffiths and Steyvers \(2004\)](#) is implemented. First, recall that the corpus consists of M distinct documents. $N = \sum_{m=1}^M N_m$ is the total number of words in all documents, K is the total number of latent topics, and V is the size of the vocabulary. Each document consists of a repeated choice of topics $Z_{m,n}$ and words $W_{m,n}$. Let t be a term in V , and denote $P(t|z = k)$, the mixture component, one for each topic, by $\Phi = \{\varphi_k\}_{k=1}^K$. Finally, let $P(z|m)$ define the topic mixture proportion for document m , with one proportion for each document $\Theta = \{\theta_m\}_{m=1}^M$. The goal of the algorithm is then to approximate the distribution:

$$P(\mathbf{Z}|\mathbf{W}; \alpha, \beta) = \frac{P(\mathbf{W}, \mathbf{Z}; \alpha, \beta)}{P(\mathbf{W}; \alpha, \beta)} \quad (14)$$

using Gibbs simulations, where α and β are the (hyper) parameters controlling the prior conjugate Dirichlet distributions for θ_m and φ_k , respectively. A very good explanation for how this method works is found in [Heinrich \(2009\)](#). The description below provides a brief summary only.

With the above definitions, the total probability of the model can be written as:

$$P(\mathbf{W}, \mathbf{Z}, \Theta, \Phi; \alpha, \beta) = \prod_{k=1}^K P(\varphi_k; \beta) \prod_{m=1}^M P(\theta_m; \alpha) \prod_{t=1}^N P(z_{m,t} | \theta_m) P(w_{m,t} | \varphi_{z_{m,t}}) \quad (15)$$

Integrating out the parameters φ and θ :

$$\begin{aligned} P(\mathbf{Z}, \mathbf{W}; \alpha, \beta) &= \int_{\Theta} \int_{\Phi} P(\mathbf{W}, \mathbf{Z}, \Theta, \Phi; \alpha, \beta) d\Phi d\Theta \\ &= \int_{\Phi} \prod_{k=1}^K P(\varphi_k; \beta) \prod_{m=1}^M \prod_{t=1}^N P(w_{m,t} | \varphi_{z_{m,t}}) d\Phi \int_{\Theta} \prod_{m=1}^M P(\theta_m; \alpha) \prod_{t=1}^N P(z_{m,t} | \theta_m) d\Theta \end{aligned} \quad (16)$$

In (16), the terms inside the first integral do not include a θ term, and the terms inside the second integral do not include a φ term. Accordingly, the two terms can be solved separately. Exploiting the properties of the conjugate Dirichlet distribution it can be shown that:

$$\int_{\Theta} \prod_{m=1}^M P(\theta_m; \alpha) \prod_{t=1}^N P(z_{m,t} | \theta_m) d\Theta = \frac{\Gamma(\sum_{k=1}^K \alpha_k) \prod_{k=1}^K \Gamma(n_m^{(k)} + \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k) \Gamma(\sum_{k=1}^K n_m^{(k)} + \alpha_k)} \quad (17)$$

and

$$\int_{\Phi} \prod_{k=1}^K P(\varphi_k; \beta) \prod_{m=1}^M \prod_{t=1}^N P(w_{m,t} | \varphi_{z_{m,t}}) d\Phi = \prod_{k=1}^K \frac{\Gamma(\sum_{t=1}^V \beta_t) \prod_{t=1}^V \Gamma(n_k^{(t)} + \beta_t)}{\prod_{t=1}^V \Gamma(\beta_t) \Gamma(\sum_{t=1}^V n_k^{(t)} + \beta_t)} \quad (18)$$

where $n_m^{(k)}$ denotes the number of word tokens in the m^{th} document assigned to the k^{th} topic, and $n_k^{(t)}$ is the number of times the t^{th} term in the vocabulary has been assigned to the k^{th} topic.

Since $P(\mathbf{W}; \alpha, \beta)$, in (14), is invariable for any of \mathbf{Z} , the conditional distribution $P(\mathbf{Z} | \mathbf{W}; \alpha, \beta)$ can be derived from $P(\mathbf{W}, \mathbf{Z}; \alpha, \beta)$ directly using Gibbs simulation and the conditional probability:

$$P(Z_{(m,n)} | \mathbf{Z}_{-(m,n)}, \mathbf{W}; \alpha, \beta) = \frac{P(Z_{(m,n)}, \mathbf{Z}_{-(m,n)}, \mathbf{W}; \alpha, \beta)}{P(\mathbf{Z}_{-(m,n)}, \mathbf{W}; \alpha, \beta)} \quad (19)$$

where $Z_{(m,n)}$ denotes the hidden variable of the n^{th} word token in the m^{th} document, and $\mathbf{Z}_{-(m,n)}$ denotes all Z s but $Z_{(m,n)}$. Denoting the index of a word token by $i = (m, n)$, and using the expressions in (17) and (18), cancellation of terms (and some extra manipulations exploiting the properties of the gamma function) yields:

$$P(Z_i = k | \mathbf{Z}_{-(i)}, \mathbf{W}; \alpha, \beta) \propto (n_{m,-i}^{(k)} + \alpha_k) \frac{n_{k,-i}^{(t)} + \beta_t}{\sum_{t=1}^V n_{k,-i}^{(t)} + \beta_t} \quad (20)$$

where the counts $n_{\cdot,-i}^{(\cdot)}$ indicate that token i is excluded from the corresponding document or topic. Thus, sampling topic indexes using equation (20) for each word in a document

and across documents until convergence allows us to approximate the posterior distribution given by (14). As noted in Heinrich (2009), the procedure itself uses only five larger data structures; the count variables $n_m^{(k)}$ and $n_k^{(t)}$, which have dimension $M \times K$ and $K \times V$, respectively, their row sums n_m and n_k , as well as the state variable $z_{m,n}$ with dimension W .

With one simulated sample of the posterior distribution for $P(\mathbf{Z}|\mathbf{W}; \alpha, \beta)$, φ and θ can be estimated from:

$$\hat{\varphi}_{k,t} = \frac{n_k^{(t)} + \beta_t}{\sum_{t=1}^V n_k^{(t)} + \beta_r} \quad (21)$$

and

$$\hat{\theta}_{m,k} = \frac{n_m^{(k)} + \alpha_k}{\sum_{k=1}^K n_m^{(k)} + \alpha_k} \quad (22)$$

In the analysis of the main paper the average of the estimated $\hat{\theta}$ and $\hat{\varphi}$ from the 10 last samples of the stored Gibbs simulations are used to construct the daily news topic frequencies.²⁰ In un-reported experiments, the topic extraction results reported in Section 2.1 do not change much when choosing other samples for inference, for example using the last sample only.

Before estimation three parameters need to be pre-defined: the number of topics and the two parameter vectors of the Dirichlet priors, α and β . Here, symmetric Dirichlet priors, with α and β each having a single value, are used. In turn, these are defined as a function of the number of topics and unique words:

$$\alpha = \frac{50}{K}, \quad \text{and} \quad \beta = \frac{200}{N}$$

The choice of K is discussed in Section 2.1. In general, lower (higher) values for α and β will result in more (less) decisive topic associations. The values for the Dirichlet hyper-parameters also reflect a clear compromise between having few topics per document and having few words per topic. In essence, the prior specification used here is the same as the one advocated by Griffiths and Steyvers (2004).

D.3 Estimating daily topic frequencies

Using the posterior estimates from the LDA model, the frequency with which each topic is represented in the newspaper for a specific day is computed. This is done by first collapsing all the articles in the newspaper for one specific day into one document. Following Heinrich (2009) and Hansen et al. (2014), a procedure for querying documents outside the set on

²⁰Because of lack of identifiability, the estimates of $\hat{\theta}$ and $\hat{\varphi}$ can not be combined across samples for an analysis that relies on the content of specific topics. However, statistics insensitive to permutation of the underlying topics can be computed by aggregating across samples, see Griffiths and Steyvers (2004).

which the LDA is estimated is then implemented. In short, this corresponds to using the same Gibbs simulations as described above, but with the difference that the sampler is run with the estimated parameters $\Phi = \{\varphi_k\}_{k=1}^K$ and hyper-parameter α held constant.

Denote by \tilde{W} the vector of words in the newly formed document. Topic assignments, \tilde{Z} , for this document can then be estimated by first initializing the algorithm by randomly assigning topics to words and then performing a number of Gibbs iterations using:

$$P(\tilde{Z}_i = k \mid \tilde{\mathbf{Z}}_{-(i)}, \tilde{\mathbf{W}}; \alpha, \beta) \propto (n_{\tilde{m},-i}^{(k)} + \alpha_k) \hat{\varphi}_{k,t} \quad (23)$$

Since $\hat{\varphi}_{k,t}$ does not need to be estimated when sampling from (23), fewer iterations are needed to form the topic assignment index for the new document than when learning both the topic and word distributions. Here 2000 iterations are performed, and only the average of every 10th draw is used for the final inference. After sampling, the topic distribution can be estimated as before:

$$\tilde{\theta}_{\tilde{m},k} = \frac{n_{\tilde{m}}^{(k)} + \alpha_k}{\sum_{k=1}^K n_{\tilde{m}}^{(k)} + \alpha_k} \quad (24)$$

D.4 News Topics as time series

Given knowledge of the topics (and their distributions), the topic decompositions are translated into time series. To do this, I proceed in three steps:

Step 1. For each day, the frequency with which each topic is represented in the newspaper that day is calculated. This is done by collapsing all the articles in the newspaper for a particular day into one document, and then computing, using the estimated word distribution for each topic, the topic frequencies for this newly formed document. See Appendix D.3 for details. By construction, across all topics, this number will sum to one for any given day. On average, across the whole sample, each topic will have a more or less equal probability of being represented in the newspaper. Across shorter time periods, i.e., days, the variation can be substantial. I define this as the D_0 dataset, which will be a $T^d \times K$ matrix, where T^d is the number of days in the sample and K the number of topics.²¹

Step 2. Since the time series objects constructed in *Step 1* will be intensity measures, i.e., reflecting how much DN writes about a given topic at a specific point in time, their tone is not identified. That is, whether the news is positive or negative. To mediate this, a sign identified dataset based on the number of positive relative to negative words

²¹Since DN is not published on Sundays, but economic activity also takes place on Sundays, missing observations are filled by simple linear interpolation. Note also that the construction described in *Step 1* does not mean that only one topic is used as representative for a given day. For such an assumption mixture models other than the LDA would have been more appropriate.

in the text is constructed. In particular, for each day t , all N_t^a newspaper articles that day, and each news topic in D_0 , the article that news topic k describes the best is found. Given knowledge of this topic article mapping, positive/negative words in the articles are identified using an external word list and simple word counts. The word list used here takes as a starting point the classification of positive/negative words defined by the *Harvard IV-4 Psychological Dictionary*. As this dictionary contains English words only, it must be translated to Norwegian. The translated set of words consists of 40 positive and 39 negative Norwegian words, which is somewhat different from the *Harvard IV-4 Psychological Dictionary* both in terms of numbers and exact meaning.²²

The count procedure delivers two statistics for each article, containing the number of positive and negative words. These statistics are then normalized such that each article observation reflects the fraction of positive and negative words, i.e.:

$$Post_{t,n^a} = \frac{\#positivewords}{\#totalwords} \quad Neg_{t,n^a} = \frac{\#negativewords}{\#totalwords} \quad (25)$$

The overall mood of article n^a , for $n^a = 1, \dots, N_t^a$ at day t , is defined as:

$$S_{t,n^a} = Post_{t,n^a} - Neg_{t,n^a} \quad (26)$$

Using the S_{t,n^a} statistic and the topic article mapping described above, the sign of each topic in D_0 is adjusted accordingly as:

$$D_{t,1} = S_{t,n^a} D_{t,\tilde{k},0}$$

where \tilde{k} reflects that article n^a is mapped to topic k .

Step 3. To remove daily noise from the topic time series in the D_1 dataset, each topic time series is filtered using a 60 day (backward looking) moving average filter. As is common in factor model studies, see, e.g., [Stock and Watson \(2012\)](#), I also eliminate very low frequency variation, i.e., changes in the local mean, by removing a simple linear trend and standardize the data. For future reference I label this dataset D_1^f .

Appendix E The Dynamic Factor model, estimation, and prediction

For estimation the Dynamic Factor Model described in Section 3 is re-written to incorporate the latent threshold mechanism for the time-varying factor loadings and the mixed

²² The translated word list can be obtained upon request. Counting the number of positive and negative words in a given text using the *Harvard IV-4 Psychological Dictionary* is a standard methodology in this branch of the literature, see, e.g., [Tetlock et al. \(2008\)](#). In finance, [Loughran and McDonald \(2011\)](#) among others, show that word lists developed for other disciplines mis-classify common words in financial text, and suggest an alternative (English language) list. I leave it for future research to investigate if this also holds for macroeconomic applications and other languages than English.

frequency variables. For notational simplicity, I assume in the following that the number of lags in the transition equations for the dynamic factors and idiosyncratic errors are $h = 1$ and $p = 1$, respectively. Moreover, I describe a model structure which includes both one quarterly and monthly variable, i.e., $N^q = 1$ and $N^m = 1$, in addition to a $N^d \times 1$ vector of daily observables. Accordingly, following [Harvey \(1990\)](#), the system used for estimation can be written in matrix form as:

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t \quad (27a)$$

$$\mathbf{a}_t = \mathbf{F}_t \mathbf{a}_{t-1} + \mathbf{R} \boldsymbol{\omega}_t \quad (27b)$$

$$\mathbf{e}_t = \mathbf{P} \mathbf{e}_{t-1} + \mathbf{u}_t \quad (27c)$$

where

$$\mathbf{y}_t = \begin{bmatrix} y_{1,t}^q \\ y_{1,t}^m \\ y_{1,t}^d \\ y_{2,t}^d \\ \vdots \\ y_{N^d,t}^d \end{bmatrix} \quad \mathbf{Z}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z_{1,t} \\ 0 & 0 & z_{2,t} \\ \vdots & \vdots & \vdots \\ 0 & 0 & z_{N^d,t} \end{bmatrix} \quad \mathbf{a}_t = \begin{bmatrix} C_{t,q} \\ C_{t,m} \\ a_{t,d} \end{bmatrix} \quad \mathbf{e}_t = \begin{bmatrix} 0 \\ 0 \\ e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{N^d,t} \end{bmatrix} \quad \mathbf{F}_t = \begin{bmatrix} \beta_{t,q} & 0 & z_q \Phi \\ 0 & \beta_{t,m} & z_m \Phi \\ 0 & 0 & \Phi \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & z_q \\ 0 & 1 & z_m \\ 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{\omega}_t = \begin{bmatrix} \omega_{t,q} \\ \omega_{t,m} \\ \omega_{t,d} \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \Phi_1 & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \Phi_{N^d} \end{bmatrix} \quad \mathbf{u}_t = \begin{bmatrix} 0 \\ 0 \\ u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{N^d,t} \end{bmatrix}$$

Here, q , m , or d superscripts denote that the variable is observed on a quarterly, monthly or daily frequency, respectively, while q , m , or d subscripts denote that the parameter or variable is associated with quarterly, monthly or daily variables, respectively. $C_{t,q}$ and $C_{t,m}$ are the quarterly and monthly cumulator variables, and $a_{t,d}$ is the daily coincident index. $\beta_{t,q}$ and $\beta_{t,m}$ are indicator variables, associated with quarterly and monthly variables, respectively.

The time-varying factor loadings are modeled as random walks following the Latent Threshold Model (LTM) idea introduced by [Nakajima and West \(2013\)](#). For example, for one particular element in the \mathbf{Z}_t matrix, $z_{i,t}$, the LTM structure can be written as:

$$z_{i,t} = z_{i,t}^* \varsigma_{i,t} \quad \varsigma_{i,t} = I(|z_{i,t}^*| \geq d_i) \quad (28)$$

where

$$z_{i,t}^* = z_{i,t-1}^* + w_{i,t} \quad (29)$$

with $w_{i,t} \sim i.i.d.N(0, \sigma_{i,w}^2)$. In (28) $\varsigma_{i,t}$ is a zero one variable, whose value depends on the indicator function $I(|z_{i,t}^*| \geq d_i)$. If $|z_{i,t}^*|$ is above the threshold value d_i , then $\varsigma_{i,t} = 1$, otherwise $\varsigma_{i,t} = 0$.

The vectors of error terms, \mathbf{v}_t , \mathbf{u}_t , and \mathbf{w}_t are independent:

$$\begin{bmatrix} \boldsymbol{\omega}_t \\ \mathbf{u}_t \\ \mathbf{w}_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Omega} & 0 & 0 \\ 0 & \mathbf{U} & 0 \\ 0 & 0 & \mathbf{W} \end{bmatrix} \right)$$

and both $\boldsymbol{\Omega}$, \mathbf{U} , and \mathbf{W} are diagonal matrices:

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_{\omega_q}^2 & 0 & 0 \\ 0 & \sigma_{\omega_m}^2 & 0 \\ 0 & 0 & \sigma_{\omega_d}^2 \end{bmatrix} \quad \mathbf{U} = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \sigma_{1,u}^2 & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_{N^d,u}^2 \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \sigma_{1,w}^2 & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_{N^d,w}^2 \end{bmatrix}$$

I note here that by restricting the error matrix $\boldsymbol{\Omega}$ to be non-singular, the model specification basically assumes that the quarterly and monthly variables contain some measurement error relative to the latent daily business cycle factor. Accordingly, identification of the latent factors, $C_{t,q}$, $C_{t,m}$, and $a_{t,d}$, is obtained by restricting the upper 3×3 block of the time-varying factor loadings matrix \mathbf{Z}_t to be an identity matrix. Thus, $z_{1,t} = 1$ for all t , and $\sigma_{1,w}^2 = 0$.

The model's hyper-parameters are $\boldsymbol{\Omega}$, \mathbf{U} , \mathbf{W} , \mathbf{F}_t , \mathbf{P} , and \mathbf{d} . Inside \mathbf{F}_t , the indicator variables $\beta_{t,q}$ and $\beta_{t,m}$ are time-varying, but their evolution is deterministic and need not be estimated. Thus, the only time-varying parameters to be estimated within the model are those in \mathbf{Z}_t , which together with \mathbf{a}_t , are the model's unobserved state variables.²³

Estimation consists of sequentially drawing the model's unobserved state variables and hyper-parameters utilizing 4 blocks until convergence is achieved. In essence, each block involves exploiting the state space nature of the model using the Kalman Filter and the simulation smoother suggested by Carter and Kohn (1994), coupled with a Metropolis-Hastings step to simulate the time-varying loadings. Below I describe each block in greater detail. For future reference and notational simplicity it will prove useful to define the following: $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_T]'$, $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_T]'$, $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_T]'$, $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_T]'$, $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_T]'$, and $\mathbf{Q} = \mathbf{R}\boldsymbol{\Omega}\mathbf{R}'$.

²³Note that, in principle, the z_q , and z_m hyper-parameters could have been made time-varying. However, I experience that estimation of the model then becomes much more sensitive to the prior specification, and have therefore chosen to treat them as constant.

E.1 Block 1: $\mathbf{A}|Y, \mathbf{Z}, \mathbf{E}, \mathbf{F}, \mathbf{P}, \mathbf{U}, \mathbf{Q}$

Equations (27a) and (27b) constitute a state space system we can use to draw the unobserved state \mathbf{a}_t using the Carter and Kohn's multimove Gibbs sampling approach. However, to do so we need to make the errors in the observation equation conditionally i.i.d. Given knowledge of equation (27c), we can define $\mathbf{P}(L) = (\mathbf{I} - \mathbf{P}L)$ and pre-multiply equation (27a) by $\mathbf{P}(L)$ to obtain the system:

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{Z}}_t \mathbf{a}_t + \mathbf{u}_t \sim N(0, \mathbf{U}) \quad (30a)$$

$$\mathbf{a}_t = \mathbf{F}_t \mathbf{a}_{t-1} + \mathbf{R} \boldsymbol{\omega}_t \sim N(0, \mathbf{Q}) \quad (30b)$$

where $\tilde{\mathbf{y}}_t = \mathbf{P}(L) \mathbf{y}_t$ and $\tilde{\mathbf{Z}}_t = \mathbf{P}(L) \mathbf{Z}_t$.

Since all hyper-parameters and state variables, less \mathbf{A} , are known (or conditionally known), we can use the equations in (30) together with Carter and Kohn's multimove Gibbs sampling approach, see Appendix (F), to sample \mathbf{a}_t from:

$$\mathbf{a}_T | \cdots \sim N(\mathbf{a}_{T|T}, \mathbf{P}_{T|T}^a) \quad t = T \quad (31a)$$

$$\mathbf{a}_t | \cdots \sim N(\mathbf{a}_{t|t, a_{t+1}}, \mathbf{P}_{t|t, a_{t+1}}^a) \quad t = T-1, T-2, \dots, 1 \quad (31b)$$

to get \mathbf{A} . Note here that the Kalman Filter can be run straightforwardly despite the fact that the $\tilde{\mathbf{y}}_t$ vector contains missing values, see Harvey (1990) for details.

E.2 Block 2: $\mathbf{Z}, \mathbf{U}, \mathbf{W}, d|Y, \mathbf{A}, \mathbf{E}, \mathbf{P}$

Conditionally on \mathbf{A} the errors in (27a) are independent across the N variables in \mathbf{y}_t . Moreover, we have assumed that the covariance matrix \mathbf{W} associated with the time-varying factor loadings in equation (29) is diagonal. Consequently, one can draw \mathbf{Z} one equation at a time. As above, we deal with the fact that the errors in the observation equation are not conditionally i.i.d. by applying the quasi differencing operator, $\mathbf{P}(L)$, to each equation. Thus, for each $i = 2, \dots, N^d$, we obtain the following Gaussian system:

$$\tilde{\mathbf{y}}_{i,t}^d = \tilde{a}_{t,d} z_{i,t} + u_{i,t} \quad (32a)$$

$$z_{i,t} = z_{i,t}^* \varsigma_{i,t} \quad \varsigma_{i,t} = I(|z_{i,t}^*| \geq d_i) \quad (32b)$$

$$z_{i,t}^* = z_{i,t-1}^* + w_{i,t} \quad (32c)$$

where $\tilde{\mathbf{y}}_{i,t}^d = (\mathbf{I} - \Phi_i L) \mathbf{y}_{i,t}^d$, and $\tilde{a}_{t,d} = (\mathbf{I} - \Phi_i L) a_{t,d}$.

To simulate from the conditional posterior of $z_{i,t}^*$ and d_i in (32), the procedure outlined in Nakajima and West (2013) is followed. That is, conditional on all the data and hyper-parameters, we draw the conditional posterior of $z_{i,t}^*$ sequentially for $t = 1 : T$ using a Metropolis-Hastings (MH) sampler. As described in Nakajima and West (2013), the MH

proposals come from a non-thresholded version of the model specific to each time t , as follows: Fixing $\varsigma_{i,t} = 1$, take proposal distribution $N(z_{i,t}^* | m_t, M_t)$ where:

$$M_t^{-1} = \sigma_{i,u}^{-2} \tilde{a}_{t,d} \tilde{a}_{t,d} + \sigma_{i,w}^{-2} (I + 1) \quad (33a)$$

$$m_t = M_t [\sigma_{i,u}^{-2} \tilde{a}_{t,d} \tilde{y}_{i,t}^d + \sigma_{i,w}^{-2} \{(z_{i,t-1}^* + z_{i,t+1}^*) + (I - 1) z_{i,0}^*\}] \quad (33b)$$

for $t = 2 : T - 1$. For $t = 1$ and $t = T$, a slight modification is needed. Details can be found in Nakajima and West (2013). The candidate is accepted with probability:

$$\alpha(z_{i,t}^*, z_{i,t}^{p*}) = \min \left\{ 1, \frac{N(\tilde{y}_{i,t}^d | \tilde{a}_{t,d} z_{i,t}^{p*}, \sigma_{i,u}^2) N(z_{i,t}^* | m_t, M_t)}{N(\tilde{y}_{i,t}^d | \tilde{a}_{t,d} z_{i,t}, \sigma_{i,u}^2) N(z_{i,t}^{p*} | m_t, M_t)} \right\} \quad (34)$$

where $z_{i,t} = z_{i,t}^* \varsigma_{i,t}$ is the current state, and $z_{i,t}^p = z_{i,t}^{p*} \varsigma_{i,t}$ is the candidate.

The independent latent thresholds in d_i can then be sampled conditional on the data and the hyper-parameters. For this, a direct MH algorithm is employed. Let $d_{i,-j} = d_{i,0:s} \setminus d_{i,j}$. A candidate is drawn from the current conditional prior, $d_{i,j}^p \sim U(0, |\beta_0| + K)$, where K is described below, and accepted with probability:

$$\alpha(d_{i,j}, d_{i,j}^p) = \min \left\{ 1, \prod_{t=1}^T \frac{N(\tilde{y}_{i,t}^d | \tilde{a}_{t,d} z_{i,t}^p, \sigma_{i,u}^2)}{N(\tilde{y}_{i,t}^d | \tilde{a}_{t,d} z_{i,t}, \sigma_{i,u}^2)} \right\} \quad (35)$$

where $z_{i,t}$ is the state based on the current thresholds $(d_{i,j}, d_{i,-j})$, and $z_{i,t}^p$ the candidate based on $(d_{i,j}^p, d_{i,-j})$.

Lastly, conditional on the data, the hyper parameters and the time-varying parameters, we can sample the elements of \mathbf{U} and \mathbf{W} using the inverse Gamma distributions. Let letters denoted with an underscore reflect the prior, then:

$$\sigma_{i,u}^2 | \dots \sim IG(\bar{v}^u, \bar{\sigma}_{i,u}^2) \quad (36)$$

where $\bar{v}^u = T + \underline{T}^u$ and $\bar{\sigma}_{i,u}^2 = [\sigma_{i,u}^2 \underline{T}^u + \sum_{t=1}^T (\tilde{y}_{i,t}^d - \tilde{a}_{t,d} z_{i,t})' (\tilde{y}_{i,t}^d - \tilde{a}_{t,d} z_{i,t})] / \bar{v}^u$, and:

$$\sigma_{i,w}^2 | \dots \sim IG(\bar{v}^w, \bar{\sigma}_{i,w}^2) \quad (37)$$

where $\bar{v}^w = T + \underline{T}^w$ and $\bar{\sigma}_{i,w}^2 = [\sigma_{i,w}^2 \underline{T}^w + \sum_{t=1}^T (z_{i,t}^* - z_{i,t-1}^*)' (z_{i,t}^* - z_{i,t-1}^*)] / \bar{v}^w$. Note here that for $\sigma_{i,u}^2$, the simulations are done for $i = 1, \dots, N^d$, while for $\sigma_{i,w}^2$ they are only done for $i = 2, \dots, N^d$ because $z_{1,t} = 1$ for all t by restriction.

E.3 Block 3: $\mathbf{F}, \mathbf{\Omega} | \mathbf{A}$

Conditional on \mathbf{A} , the transition equation in (27b) is independent of the rest of the system. While the first and second equations of (27b) do depend on the estimates of Φ , the part of the transition equation associated with $a_{t,d}$ is independent of the rest of the components in (27b). Accordingly, Φ and $\sigma_{\omega_d}^2$, the element in the lower right corner of $\mathbf{\Omega}$, can first be

simulated independently from the rest of the parameters in (27b), and then $\sigma_{\omega_m}^2$, $\sigma_{\omega_q}^2$, z_m , and z_q can be simulated conditionally on Φ and $\sigma_{\omega_d}^2$.

To simulate Φ and $\sigma_{\omega_d}^2$, we employ the independent Normal-Gamma prior. Accordingly, continuing with letting letters denoted with an underscore reflect the prior, the conditional posterior of Φ is:

$$\Phi | \dots \sim N(\bar{\Phi}, \bar{V}^\Phi)_{I[s(\Phi)]} \quad (38)$$

with

$$\bar{V}^\Phi = (\underline{V}^{\Phi^{-1}} + \sum_{t=1}^T a'_{t-1,d} \sigma_{\omega_d}^{-2} a_{t-1,d})^{-1} \quad (39)$$

$$\bar{\Phi} = \bar{V}^\Phi (\underline{V}^{\Phi^{-1}} \underline{\Phi} + \sum_{t=1}^T a'_{t-1,d} \sigma_{\omega_d}^{-2} a_{t,d}) \quad (40)$$

and $I[s(\Phi)]$ is an indicator function used to denote that the roots of Φ lie outside the unit circle. Further, the conditional posterior of $\sigma_{\omega_d}^2$ is:

$$\sigma_{\omega_d}^2 | \dots \sim IG(\bar{v}^{\omega_d}, \bar{\sigma}_{\omega_d}^2) \quad (41)$$

with $\bar{v}^{\omega_d} = T + \underline{T}^{\omega_d}$, and $\bar{\sigma}_{\omega_d}^2 = [\underline{\sigma}_{\omega_d}^2 \underline{T}^{\omega_d} + \sum_{t=1}^T (a_t - a_{t-1}\Phi)'(a_t - a_{t-1}\Phi)]/\bar{v}^{\omega_d}$.

Once Φ and $\sigma_{\omega_d}^2$ are drawn, we can construct, for $j = \{q, m\}$:

$$C_{t,j} - \beta_{t,j} C_{t-1,j} \equiv C_{t,j}^* = z_j a_{t,d} + \omega_{t,j} \quad (42)$$

and draw from the conditional posterior of z_j and $\sigma_{\omega_j}^2$. Using again the independent Normal-Gamma prior:

$$z_j | \dots \sim N(\bar{z}_j, \bar{V}^{z_j}) \quad (43)$$

with

$$\bar{V}^{z_j} = (\underline{V}^{z_j^{-1}} + \sum_{t=1}^T a'_{t,d} \sigma_{\omega_j}^{-2} a_{t,d})^{-1} \quad (44)$$

$$\bar{z}_j = \bar{V}^{z_j} (\underline{V}^{z_j^{-1}} \underline{z}_j + \sum_{t=1}^T a'_{t,d} \sigma_{\omega_j}^{-2} C_{t,j}^*) \quad (45)$$

Finally, the conditional posterior of $\sigma_{\omega_j}^2$ is:

$$\sigma_{\omega_j}^2 | \dots \sim IG(\bar{v}^{\omega_j}, \bar{\sigma}_{\omega_j}^2) \quad (46)$$

with $\bar{v}^{\omega_j} = T + \underline{T}^{\omega_j}$, and $\bar{\sigma}_{\omega_j}^2 = [\underline{\sigma}_{\omega_j}^2 \underline{T}^{\omega_j} + \sum_{t=1}^T (C_{t,j}^* - a_{t,d} z_j)'(C_{t,j}^* - a_{t,d} z_j)]/\bar{v}^{\omega_j}$.

E.4 Block 4: $E|Y, A, Z$ and $P|E, U$

For each observation of the N^d daily variables we have that:

$$e_{i,t} = y_{i,t} - z_{i,t} a_{t,d} \quad (47)$$

Thus, conditional on \mathbf{Y} , \mathbf{A} and \mathbf{Z} , \mathbf{E} is observable. As above, since \mathbf{E} is independent across the N^d equations, we can sample the elements of \mathbf{P} in (27c) one equation at the time. As this is done in the same manner as in equations (38) to (40) of Block 3 (with the obvious change of notation), I do not repeat the computations here.

E.5 Prior specification

To implement the MCMC algorithm, and estimate the model, prior specifications for the initial state variables \mathbf{a}_0 , \mathbf{Z}_0 , and for the hyper-parameters $\mathbf{\Omega}$, \mathbf{U} , \mathbf{W} , \mathbf{F}_t , \mathbf{P} , and \mathbf{d} are needed. The prior specifications used for the initial states take the following form: $\mathbf{a}_0 \sim N(0, I \cdot 100)$, and $\mathbf{Z}_0 \sim N(0, I \cdot 100)$. The priors for the hyper-parameters $\mathbf{\Phi}$ and $\mathbf{\Phi}$, which are part of the \mathbf{F}_t and \mathbf{P} matrices, respectively, are set to:

$$\begin{aligned} \underline{\mathbf{\Phi}} &\sim N(\hat{\mathbf{\Phi}}_{OLS}, V(\hat{\mathbf{\Phi}}_{OLS})) \\ \underline{\Phi}_i &\sim N(0, 0.5) \quad \text{for } i = 1, \dots, N^d \end{aligned}$$

where $\hat{\mathbf{\Phi}}_{OLS}$ are the OLS estimates of an AR(h) using the first principal component of the daily news dataset as dependent variable. $V(\hat{\mathbf{\Phi}}_{OLS})$ is a diagonal matrix where the non-zero entries are the variance terms associated with the $\hat{\mathbf{\Phi}}_{OLS}$ elements. To draw z_q and z_m , which are part of the \mathbf{F}_t matrix, I use: $z_j \sim N(1, 1)$ for $j = \{q, m\}$.

The priors for the hyper-parameters $\mathbf{\Omega}$, \mathbf{U} , and \mathbf{W} , are all from the Inverse-Gamma distribution, where the first element in each prior distribution is the shape parameter, and the second the scale parameter: $\sigma_{i,w}^2 \sim IG(\underline{T}^w, \kappa_w^2)$ where $\underline{T}^w = 8000$ and $\kappa_w = 0.003$ for $i = 2, \dots, N^d$; $\sigma_{i,u}^2 \sim IG(\underline{T}^u, \kappa_u^2)$ where $\underline{T}^u = 100$ and $\kappa_u = 0.1$ for all $i = 1, \dots, N^d$; $\sigma_{\omega_j}^2 \sim IG(\underline{T}^{\omega_j}, \kappa_{\omega_j}^2)$ where $\underline{T}^{\omega_j} = 1000$ for $j = \{q, m, d\}$, and $\kappa_{\omega_q} = 0.003$, and $\kappa_{\omega_m} = \kappa_{\omega_d} = 0.1$. In sum, as the full sample size $T = 9713$ observations, these priors are very informative for the variance terms associated with the time-varying factor loadings, but less so for the other parameters. Note, however, that the prior variance associated with the quarterly cumulator variable error term, $\sigma_{\omega_q}^2$, is assumed to be considerably lower than the other variance terms.

Finally, to draw the latent threshold, \mathbf{d} , using the algorithm described in Appendix E.2, the K parameter needs to be defined. K controls our prior belief concerning the marginal sparsity probability. For example, assuming that a time-varying parameter follows $B_t \sim N(0, v^2)$, and marginalizing over B_t , it can be shown that $Pr(|B_t| = 0) = 2\Phi(\frac{d}{v}) - 1$, where Φ is the standard normal CDF. Defining $K = \frac{d}{v}$ as the standardized scaling parameter with respect to the threshold, it can be seen that $K = 3$ implies a marginal sparsity probability exceeding 0.99. As described in Nakajima and West (2013), a neutral prior will support a range of sparsity values in order to allow the data to inform on relevant values, and

they suggest that setting $K = 3$ is a reasonable choice.²⁴ However, in contrast to in Nakajima and West (2013), where the time-varying parameters follows AR(1) dynamics, the time-varying factor loadings in (32) follows independent random walk processes. The random walk is non-stationary, and does not have a marginal distribution. Thorsrud (2016) experiments with estimating the model using different values for K , finding that higher values for K , coupled with the rather tight priors for the variance of the factor loadings, results in worse model performance (in terms of ROC and AUROC scoring). Accordingly, $K = 0.05$, in the estimations conducted in this analysis.

Appendix F The Carter and Kohn algorithm

Consider a generic state space system, written in companion form, and described by:

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{u}_t \sim N(0, \mathbf{U}) \quad (49a)$$

$$\mathbf{a}_t = \mathbf{F} \mathbf{a}_{t-1} + \mathbf{R} \boldsymbol{\omega}_t \sim N(0, \mathbf{Q}) \quad (49b)$$

where we assume that the hyper-parameters $\theta = \{\mathbf{U}, \mathbf{F}, \mathbf{R}, \mathbf{Q}\}$, and \mathbf{Z}_t are known, and we wish to estimate the latent state \mathbf{a}_t for all $t = 1, \dots, T$. To do so, we can apply Carter and Kohn's multimove Gibbs sampling approach (see Carter and Kohn (1994)).

First, because the state space model given in equation (49) is linear and (conditionally) Gaussian, the distribution of \mathbf{a}_t given \mathbf{Y} and that of \mathbf{a}_t given \mathbf{a}_{t+1} and \mathbf{Y} for $t = T - 1, \dots, 1$ are also Gaussian:

$$\mathbf{a}_T | \mathbf{Y} \sim N(\mathbf{a}_{T|T}, \mathbf{P}_{T|T}), \quad t = T \quad (50a)$$

$$\mathbf{a}_t | \mathbf{Y}, \mathbf{a}_{t+1} \sim N(\mathbf{a}_{t|t, a_{t+1}}, \mathbf{P}_{t|t, a_{t+1}}), \quad t = T - 1, T - 2, \dots, 1 \quad (50b)$$

where

$$\mathbf{a}_{T|T} = E(\mathbf{a}_T | \mathbf{Y}) \quad (51a)$$

$$\mathbf{P}_{T|T} = Cov(\mathbf{a}_T | \mathbf{Y}) \quad (51b)$$

$$\mathbf{a}_{t|t, a_{t+1}} = E(\mathbf{a}_t | \mathbf{Y}, \mathbf{a}_{t+1}) = E(\mathbf{a}_t | \mathbf{a}_{t|t}, \mathbf{a}_{t|t+1}) \quad (51c)$$

$$\mathbf{P}_{t|t, a_{t+1}} = Cov(\mathbf{a}_t | \mathbf{Y}, \mathbf{a}_{t+1}) = Cov(\mathbf{a}_t | \mathbf{a}_{t|t}, \mathbf{a}_{t|t+1}) \quad (51d)$$

Given $\mathbf{a}_{0|0}$ and $\mathbf{P}_{0|0}$, the unknown states $\mathbf{a}_{T|T}$ and $\mathbf{P}_{T|T}$ needed to draw from (50a)

²⁴Note that when combined with the priors over the other hyper-parameters in the model, the implied marginal prior for each threshold will not be uniform, see Nakajima and West (2013) for details.

can be estimated from the (conditionally) Gaussian Kalman Filter as:

$$\mathbf{a}_{t|t-1} = \mathbf{F}\mathbf{a}_{t-1|t-1} \quad (52a)$$

$$\mathbf{P}_{t|t-1} = \mathbf{F}\mathbf{P}_{t-1|t-1}\mathbf{F}' + \mathbf{Q} \quad (52b)$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{Z}_t'(\mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t' + \mathbf{U})^{-1} \quad (52c)$$

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1}) \quad (52d)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{Z}_t\mathbf{P}_{t|t-1} \quad (52e)$$

That is, at $t = T$, equation 52d and 52e above, together with equation 50a, can be used to draw $\mathbf{a}_{T|T}$. Moreover, $\mathbf{a}_{t|t,a_{t+1}}$ for $t = T - 1, T - 2, \dots, 1$ can also be simulated based on 50b, where $\mathbf{a}_{t|t,a_{t+1}}$ and $\mathbf{P}_{t|t,a_{t+1}}$ are generated from the following updating equations:

$$\mathbf{a}_{t|t,a_{t+1}} = \mathbf{a}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'(\mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q})^{-1}(\mathbf{a}_{t+1} - \mathbf{F}\mathbf{a}_{t|t}) \quad (53a)$$

$$\mathbf{P}_{t|t,a_{t+1}} = \mathbf{P}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'(\mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q})^{-1}\mathbf{F}\mathbf{P}_{t|t} \quad (53b)$$

Appendix G Forecast error decompositions

When computing the news topic contributions in Figure 5, I decompose the state vector into a history of forecast error contributions. For simplicity I use the notation introduced in Appendix F to describe how this is done. At each time interval t , the forecast error in predicting \mathbf{y}_t is given by $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1}$. In computing $\mathbf{a}_{t|t}$, equation (52d) above, the Kalman gain \mathbf{K}_t is used to weight each forecast error when computing the updated state estimate. If the predictions of the i th observable at time t are perfect, $v_{i,t} = 0$ and this observation does not contribute to potential updates from $\mathbf{a}_{t|t-1}$ to $\mathbf{a}_{t|t}$. If the predictions of the i th observable at time t are not perfect, $v_{i,t} \neq 0$, the observation will influence the updated state estimate as long as it is given weight through the \mathbf{K}_t matrix. As the updating equation in 52d has a recursive structure, the time evolution of $\mathbf{a}_{t|t}$ can easily be decomposed into a set of weighted forecast error contributions, resulting in the decompositions shown in Figure 5.

For simplicity, the decompositions reported in Figure 5 are based on running the Kalman Filter using the posterior median estimates of the hyper-parameters and the time-varying factor loadings (at each time t), after the model has been estimated on the whole sample. As seen in Figure 3a already, the real-time estimates of the model's states do not seem to change considerably across time. Accordingly, there should not be large differences in a forecast decomposition performed in real-time versus the one illustrated here.

Appendix H A Mixture Auxiliary Particle Filter approach

As alluded to in Section 4, a so called mixture auxiliary particle filter is used to update the latent state estimates at each forecast round. The particle filter approach is adopted because it allows a much faster update of the states, and thus also the construction of a nowcast, than a full re-estimation of the model would allow for. Below I first give a brief description of the particle filtering methodology in general, and then describe the algorithm adopted here. See Creal (2009), and the references therein, for a useful introduction to particle filter methods used in economics, and Michael K. Pitt (1999), Chen and Liu (2000) and Doucet et al. (2001) for a more detailed description of the mixture auxiliary particle filter.

Abusing notation introduced earlier, but for notational simplicity, consider again a non-linear state space system of the type described in Appendix F, with known hyper-parameters θ , unknown state variable a_t and observable data given by y_t :

$$y_t = Z_t(a_{2,t})a_{1,t} + u_t \sim N(0, U) \quad (54a)$$

$$a_{1,t} = F a_{1,t-1} + R \omega_t \sim N(0, Q) \quad (54b)$$

$$a_{2,t} = a_{2,t-1} + w_t \sim N(0, W) \quad (54c)$$

where the state a_t is now explicitly decomposed into two blocks; $a_t = (a_{1,t}, a_{2,t})'$, and Z is a function of $a_{2,t}$. Given a system like (54), the goal is to estimate the joint smoothing distribution of the latent state, given by:

$$p(a_{0:t}|y_{1:t}; \theta) = \frac{p(y_{1:t}, a_{0:t}; \theta)}{p(y_{1:t}; \theta)} = \frac{p(y_t|a_t; \theta)p(a_t|a_{t-1}; \theta)}{p(y_t|y_{1:t-1}; \theta)} p(a_{0:t-1}|y_{1:t-1}; \theta) \quad (55)$$

However, solving (55) analytically is difficult due to the assumed non-linearity of the system. This motivates the use of SMC methods, such as the particle filter. Instead of solving (55) directly, these methods utilize the recursive structure of the joint smoothing distribution, as highlighted by the last equality sign in (55), and two of its marginal distributions, namely the predictive distribution $p(a_t|y_{1:t-1}; \theta)$ and the filtering distribution given by:

$$p(a_t|y_{1:t}; \theta) = \frac{p(y_t, a_t|y_{1:t-1}; \theta)}{p(y_t|y_{1:t-1}; \theta)} = \frac{p(y_t|a_t; \theta)p(a_t|y_{1:t-1}; \theta)}{p(y_t|y_{1:t-1}; \theta)} \quad (56)$$

Here, $p(y_t, a_t|y_{1:t-1}; \theta) = p(y_t|a_t; \theta)p(a_t|y_{1:t-1}; \theta)$ is the joint distribution of the data and the state variables, and $p(y_t|y_{1:t-1}; \theta)$ is the contribution to the likelihood function (or the normalizing constant). Accordingly, to sequentially solve 55, starting from an initial draw of the state, the last period's filtering distribution is projected forward using the predictive distribution and then updated using the filtering distribution. These iterations continue until the end of the sample.

Difficulty arises because the contribution to the likelihood ($p(y_t|y_{t-1}; \theta)$) involves integrals:

$$p(y_t|y_{1:t-1}; \theta) = \int p(y_t|a_t; \theta)p(a_t|y_{1:t-1}; \theta)da_t \quad (57)$$

that typically cannot be calculated analytically (except in special cases, like when the linear Kalman Filter is applicable). Still, this integral can be approximated using Monte Carlo integration. Instead of drawing the entire joint smoothing distribution $p(a_{0:t}|y_{1:t}; \theta)$ from standard Monte Carlo simulation, the general idea is to employ importance sampling where draws are taken from an appropriate importance distribution $q(a_{0:t}|y_{1:t}; \psi)$ that approximates the target distribution $p(a_{0:t}|y_{1:t}; \theta)$. The draws are then reweighted to correct for the fact that they are drawn from the wrong distribution:

$$w_t = \frac{p(a_{0:t}|y_{1:t}; \theta)}{q(a_{0:t}|y_{1:t}; \psi)} \quad (58)$$

While computing (58) at each iteration t would be highly computationally demanding, the importance distribution can be factorized such that it takes on a recursive structure $q(a_{0:t}|y_{1:t}; \psi) = q(a_t|a_{1:t-1}, y_{1:t}; \psi)q(a_{1:t-1}|a_{1:t-1}, y_{1:t-1}; \psi)$. Exploiting this factorization, and the last part of (55), it can be shown that w_t can be written as:

$$w_t \propto w_{t-1} \frac{p(y_t|a_t; \theta)p(a_t|a_{t-1}; \theta)}{q(a_t|a_{1:t-1}, y_{1:t}; \psi)} \quad (59)$$

Here, the densities in the numerator are determined by the state-space model, and the ratio of the densities on the right-hand side of (59) are called the incremental importance weight. Thus, given Monte Carlo draws of $\{a_{1:t}^i, w_t^i\}_{i=1}^N$ an approximation to the function $f(a_{1:t})$ can be obtained as:

$$\sum_{i=1}^N f(a_{1:t}^i) \hat{w}_t^i \approx E_q[f(a_{1:t})] \quad (60)$$

where $\hat{w}_t = \frac{w_t}{\sum_{i=1}^N w_t^i}$ are self normalizing importance weights. The empirical distribution function determined by the particles is given by:

$$\hat{p}(a_{1:t}|y_{1:t}; \theta) = \sum_{i=1}^N \hat{w}_t^i \delta_{a_{1:t}^i} \approx p(a_{1:t}|y_{1:t}; \theta) \quad (61)$$

where $\delta_{a_{1:t}^i}$ is a Dirac measure located at $a_{1:t}$.

Although the steps described above seem relatively straight forward to implement, two issues regarding implementation turn out to be crucial for the filter's success. First, as is well known, particle filter methods in general often suffer from weight degeneracy problems. As the number of iterations increases, all the probability mass will eventually be allocated to only one particle. For this reason, different re-sampling procedures have been proposed and used within the filter, after computing the importance weights. In

practical terms, this means that particles are replicated in proportion to their normalized importance weights. The re-sampling procedure used in the particle filter algorithm employed here utilizes so called systematic re-sampling. However, while the re-sampling step partly alleviates the weight degeneracy problems, it comes with its own weaknesses. Especially, the re-sampling step reduces the number of particles representing the earlier parts of the joint distribution, making the approximation of $p(a_{0:t}|y_{1:t}; \theta)$ poor. Still, the filter's approximation to $p(a_{t-k:t}|y_{1:t}; \theta)$, for small k , can be good. See Creal (2009) for an illuminating example of these effects.

Second, partly due to the importance sampling step, which creates options regarding importance distribution, many different particle filters have been proposed. The specific particle filter used here is called, as mentioned, a mixture auxiliary particle filter. The auxiliary term is used because the importance distribution is sampled in two steps, where some auxiliary variables are drawn in the first step. The point of this two-step sampling is that the first step uses information in the current observables to find particles within the existing set in order to form a better importance distribution. The mixture term is used because some parts of the model can be solved analytically, which always is an advantage because it reduces Monte Carlo variation. To see more clearly how this latter property is utilized, consider again the system in (54), where the state a_t is decomposed into two blocks; $a_t = (a_{1,t}, a_{2,t})'$. Thus, the filtering distribution can be written as $p(a_{1,t}, a_{2,t}|y_{1:t}; \theta) = p(a_{1,t}|a_{2,t}, y_{1:t}; \theta)p(a_{2,t}|y_{1:t}; \theta)$. Here, particles only need to be simulated randomly for $p(a_{2,t}|y_{1:t}; \theta)$, because $p(a_{1,t}|a_{2,t}, y_{1:t}; \theta)$ can be evaluated analytically given draws from $p(a_{2,t}|y_{1:t}; \theta)$.

Building on the discussion and notation used above, the general form of the particle filter algorithm used to construct updated state estimates at each forecast round is as follows:

1. At time t , i.e., the first day of a given quarter, draw $a_{1,t}^i$ and $a_{2,t}^i$, for $i = 1, \dots, N$ from their posterior estimates obtained from $t - 1$. Set $h = 0$, and set $w_h^i = \frac{1}{N}$.
2. Set $h = h + 1$.
 - (a) For $i = 1, \dots, N$
 - i. Run the prediction step of the Kalman Filter to obtain $a_{1,t+h|t-1+h}^i$ and $P_{1,t+h|t-1+h}^i$ (see (52a) and (52b))
 - ii. Calculate importance weights using the Kalman Filter prediction error decomposition, i.e., $w_h^i \propto N(v_{t+h}^i, V_{t+h}^i)$, where:

$$v_{t+h}^i = y_{t+h} - Z_t(a_{2,t+h|t-1+h}^i)a_{1,t+h|t-1+h}^i$$

$$V_{t+h}^i = U + Z_t(a_{2,t+h|t-1+h}^i)P_{1,t+h|t-1+h}^i Z_t(a_{2,t+h|t-1+h}^i)'$$

- (b) Normalize the importance weights as $\hat{w}_h^i = \frac{w_h^i}{\sum_{j=1}^N w_h^j}$.
 - (c) Re-sample N particles $\{a_{1,t+h-1|t-1+h}^i, P_{1,t+h-1|t-1+h}^i, a_{2,t+h-1|t+h-1}^i\}_{i=1}^N$ with probabilities $\{\hat{w}_h^i\}_{i=1}^N$, and for $i = 1, \dots, N$ set $w_h^i = \frac{1}{N}$.
 - (d) For $i = 1, \dots, N$
 - i. Draw $a_{2,t+h}^i \sim N(a_{2,t+h-1}^i, W)$.
 - ii. Set $a_{1,t+h|t-1+h}^i = a_{1,t+h-1|t-1+h}^i$ and $P_{1,t+h|t-1+h}^i = P_{1,t+h-1|t-1+h}^i$, and run the Kalman Filter conditional on $a_{2,t+h}^i$ to obtain $a_{1,t+h|t+h}^i$ and $P_{1,t+h|t+h}^i$ (see (52c) to (52e)).
 - (e) Compute the filtered estimate: $E[f(a_{j,t+h}^i)] \approx \sum_{i=1}^N \hat{w}_h^i f(a_{j,t+h}^i)$ for $j = 1, \dots, 2$.
3. Stop if $h = \bar{h}$, where $t + \bar{h}$ is the last day of a given quarter, else, return to 2.

Relating this to the notation used for the state-space model (2) in Section 3, it is easily seen that $a_{2,t}$ corresponds to the time-varying factor loadings, while $a_{1,t}$ relates to the daily factors. It therefore follows by (6) that the algorithm described above yields a nowcast of $\tilde{y}_{1,t+\bar{h}}$.

Two points that are important for the DFM used in the main paper, but not described in the algorithm above, is how the latent threshold mechanism is handled, and how the autoregressive structure of the idiosyncratic errors is handled. The first point relates to step 2.d.i of the algorithm. Here, in the actual implementation, the time-varying factor loadings are simulated from their time series process, but the draws that in absolute value are smaller than the estimated threshold, are set to zero, cf. equations (4) and (5). The treatment of the autoregressive structure of the idiosyncratic errors affects steps 2.a.ii and 2.d.ii of the algorithm. Here, the y_{t+h} , $Z_t(a_{2,t+h|t-1+h}^i)$, and $a_{2,t+h}^i$ terms need to be adjusted to appropriately take into account the autocorrelation in the idiosyncratic errors. In the practical implementation of the algorithm, this is done in the same manner as for the MCMC procedure, see Appendix E.1, using the quasi difference operator.

Finally, although the algorithm described above works well for one single forecasting round, it is my experience that it becomes less accurate across forecasting rounds. This is due to the weaknesses of re-sampling, which makes the particle filter's approximation to $p(a_{0:t}|y_{1:t}; \theta)$ poor (as described above). In particular, if the estimated latent states (and their distributions) are saved and subsequently used as starting values for the next forecast round, as in Step 1 of the algorithm, the approximations of historical latent state distributions become centered on only a few draws after relatively few forecasting rounds. To mediate this issue, at each forecast round I use the updated state estimates from the particle filter as starting values for the MCMC algorithm described in Appendix E, and construct updated and smoothed state estimates accordingly. Importantly, however, compared with a full re-estimation of the model using the MCMC approach, a substantially

smaller number of draws are needed to construct a good approximation to the posterior distributions because the MCMC algorithm is initiated using the posterior estimates of the hyper-parameters and particle filtered updates of the state estimates.