Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Words are the New Numbers: A Newsy Coincident Index of the Business Cycle

Leif Anders Thorsrud

Words are the New Numbers: A Newsy Coincident Index of the Business Cycle

Leif Anders THORSRUD
Norges Bank and Centre for Applied Macro and Petroleum Economics, BI Norwegian Business School, 0484 Oslo, Norway (*leif.a.thorsrud@bi.no*)

I construct a daily business cycle index based on quarterly GDP growth and textual information contained in a daily business newspaper. The newspaper data are decomposed into time series representing news topics, while the business cycle index is estimated using the topics and a time-varying dynamic factor model where dynamic sparsity is enforced upon the factor loadings using a latent threshold mechanism. The resulting index classifies the phases of the business cycle with almost perfect accuracy and provides broad-based high-frequency information about the type of news that drive or reflect economic fluctuations. In out-of-sample nowcasting experiments, the model is competitive with forecast combination systems and expert judgment, and produces forecasts with predictive power for future revisions in GDP. Thus, news reduces noise. Supplementary materials for this article are available online.

KEY WORDS: Business cycles; Dynamic factor model (DFM); Latent Dirichlet allocation (LDA); Nowcasting.

1. INTRODUCTION

Policy makers and forecasters need to assess the state of the economy in real time to devise appropriate policy responses and condition on an updated information set when doing nowcasting. However, in real time, our main measure of economic activity, GDP growth, is not observed as it is compiled on a quarterly frequency and published with a considerable lag, usually many months. To mediate these caveats, various more timely indicators, like financial and labor market data, are monitored closely, and coincident indexes constructed (Stock and Watson 1989; Evans 2005; Banbura, Giannone, and Reichlin 2011).

However, common approaches for business cycle tracking and nowcasting face at least three drawbacks. First, the relationships between timely indicators and GDP growth are inherently unstable (Stock and Watson 2003), and state-of-the-art models, used at, for example, central banks, have a hard time performing well when economic conditions change rapidly. This was particularly evident around the Great Recession, when good forecasting performance perhaps mattered the most (Alessi et al. 2014).

Second, while there is an abundance of high-frequency financial data, there is limited availability of high-frequency data reflecting the broader economy. As a result, the type of data from which coincident indexes are constructed is constrained, making it difficult for the index user to get broad-based high-frequency information about the type of new information that drive or reflect economic fluctuations. For policy makers in particular, understanding why an index changes might be as important as the movement itself, as reflected in the broad coverage of various data in monetary policy reports and national budgets.

Finally, and related to the previous point, the agents in the economy likely use a plethora of high-frequency information to guide their actions. In this setting, news broadcasted through, for example, the media, might matter more than data from professional data providers because it can reach a broad population of economic agents and alleviate informational frictions (Sims 2003; Peress 2014; Larsen and Thorsrud 2017).

In this article, I propose a new coincident index of business cycles aimed at addressing the drawbacks discussed above. In the tradition of Mariano and Murasawa (2003), Aruoba, Diebold, and Scotti (2009), and Marcellino, Porqueddu, and Venditti (2016), I estimate a latent daily coincident index using a Bayesian time-varying Dynamic Factor Model (DFM) mixing daily data and quarterly GDP growth. To this, I make two contributions. First, the daily data come from a novel usage of textual information contained in a daily business newspaper, represented as 80 tone adjusted topic frequencies that vary in intensity across time. The term "Big Data" is used for textual data of this type because they are, before processing, highly unstructured and contain large amounts of words and articles (Nymand-Andersen 2016). Thus, words are the new numbers, and the name: A newsy coincident index (*NCI*). The extraction of topics is done using advances in the Natural Language Processing literature, while the tone is identified using simple dictionary-based techniques (Tetlock 2007). In turn, this innovation allows me to decompose the changes in the latent daily business cycle index into time-varying news components and say something more broadly about why (in terms of news topics) the index changes at particular points in time. My underlying 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

Journal of Business & Economic Statistics
April 2020, Vol. 38, No. 2
DOI: 10.1080/07350015.2018.1506344

extensively about developments in, for example, the oil sector, and the tone is positive, it reflects that something is happening in this sector that potentially has positive economy-wide effects.

Second, building on the latent threshold model (LTM) idea introduced by Nakajima and West (2013), and applied in a factor model setting in Zhou, Nakajima, and West (2014), the DFM is specified using a threshold mechanism for the time-varying factor loadings. The LTM mechanism has not been applied in a mixed-frequency factor model before, but enforces sparsity through dynamic variable selection, and explicitly takes into account that the relationship between the latent daily coincident index and the indicators used to derive it 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 LTM mechanism potentially captures such cases in a consistent and transparent way, irrespective of whether newspaper data or more standard high-frequency data are used to derive the index.

My main results, applying Norwegian data, show that the strategy adopted has potential. In in-sample evaluations I demonstrate that the *NCI* classifies the phases of the business cycle with almost perfect accuracy, and that it outperforms coincident indexes based on more traditional (daily and monthly) economic variables. The gain in performance is shown to be due to the combined usage of newspaper data and allowing for the LTM mechanism. Decomposing the historical evolution of the *NCI* into individual news topic contributions reveals sparsity patterns that vary substantially across time, while the sign and timing of their contribution map reasonably well with the narrative we now have about historical business cycle developments.

Next, when testing the news-based model in a series of out-of-sample forecasting experiments, I show that the model produces nowcasts that are competitive with the performance of official Norges Bank nowcasts and a state-of-the-art forecast combination system. As above, the good forecasting performance is shown to be due to the combined usage of newspaper data and the LTM mechanism, and there seems to be a tendency that the news-based model is performing especially well around economic turning points. If the statistical agency producing the output growth statistics itself had used the news-based methodology, I show that it would have resulted in a less noisy revision process. Thus, news also reduces noise.

To be clear, the main methodological contribution of this article is not the invention of new natural language processing techniques, or a new type of factor model for time series analysis, but rather combining newer developments from, and within, both of these areas in a manner that is novel for describing macroeconomic fluctuations. As such, this article speaks to three branches of the economic literature. First, in using newspaper data, the approach taken here speaks to a growing number of studies using text as data (Bholat et al. 2015; Gentzkow, Kelly, and Taddy 2017). On this point, commonly used methods in economics involve some kind of subjectively chosen keyword search. In this article, a dictionary-based technique is used together with what is called an unsupervised topic model belonging to the latent Dirichlet allocation (LDA) class (Blei, Ng, and Jordan 2003).

In general, topic models are statistical algorithms that categorize the corpus, that is, the whole collection of words and articles, into topics that best reflect the corpus's word dependencies. 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 usage in a time series context. Although topic models hardly have been applied in economics, see, for example, Hansen, McMahon, and Prat (2018) for an exception, their usage as a natural language processing tool in other disciplines have been massive. The LDA's popularity stems from its success in classifying text and articles into topics in much the same manner as humans would do (Chang et al. 2009). Accordingly, the LDA approach offers a conceptual advantage over other often applied textual data techniques because it provides interpretable output in a highly automated fashion.

Second, this article is directly related to a voluminous literature, starting with Burns and Mitchell (1946), that seeks to measure business cycles and construct coincident indexes. Influential contributions were cited above, while important features are shared with especially Balke, Fulmer, and Zhang (2017). They use customized text analytics to decompose the Federal Reserve System's *Beige Book* into time series. Based on these data they then construct a coincident index for the U.S. business cycle, and document that the textual data source contains information about current economic activity not contained in quantitative data, complementing my findings. However, the *Beige Book* is published at an irregular frequency, and not all countries have this type of information. In contrast, most countries have publicly available newspapers published daily, and the LDA method applied here is not customized to neither the specific news source nor country.

Finally, this article contributes to a large literature where the usage of factor models and mixed-frequency data have proven particularly useful for business cycle tracking and nowcasting (Stock and Watson 2002; Giannone, Reichlin, and Small 2008; Breitung and Schumacher 2008; Kuzin, Marcellino, and Schumacher 2011; Marcellino, Porqueddu, and Venditti 2016). In terms of modeling, I extend this literature by allowing for the LTM mechanism in a mixed-frequency factor model. In terms of data usage, I provide novel evidence on how information in the newspaper can be used to predict the present. A different approach, with the same objective, is offered by recent studies using Internet search volume (Choi and Varian 2012). Like when user generated search volume is used, the news topic approach captures economic agents' frame of focus, and thereby resembles some type of survey. In contrast to user generated Internet search, however, the information in the newspaper is constrained by page limits and goes through an editorial process. Good editing might amplify the signal and reduce the noise. Moreover, for obvious reasons, long time series of search volume, or other social media data, is not available, while historical newspaper content is. For estimation and testing purposes in a macroeconomic context, this is a clear advantage.

The rest of this article is organized as follows. Section 2 describes the newspaper data, the topic model, the estimated news topics, and the GDP data. The DFM is described in Section 3. Results are presented in Section 4, while Section 5 concludes.

2. DATA

The main raw data used in this analysis consist of a long sample of the entire newspaper corpus for *Dagens Næringsliv* (DN) and quarterly GDP growth in Norway. DN is the largest and most read business newspaper in Norway, and also the fourth largest newspaper irrespective of subject matter. While Norway is a small and open economy, and thereby representative of many western countries, the methodology I use for extracting news from newspaper data and linking it with macroeconomic developments 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, and then transformed into tone adjusted time series. The newspaper corpus and the topic model specification used in this article is similar to that described in Larsen and Thorsrud (in press), but the data are updated to include more recent time periods. I provide a summary of the computations below. In the interest of preserving space, technical details are relegated to online Appendix F. Quarterly GDP growth, and a preliminary analysis of its relationship with the news topics, is described in the latter part of this section.

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 June 28, 2016. This amounts to over half a million articles, well above one billion words, more than a million unique tokens, and a sample of over 10,000 days. This massive amount of data makes statistical computations challenging, but as is customary in this branch of the literature some steps are taken to clean and reduce the raw dataset before estimation. These are removing common words, surnames, reducing 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 online Appendix F.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. 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. As such, the LDA shares many features with latent (Gaussian) factor models used in conventional econometrics, but with factors (representing topics) constrained to live in the simplex and fed through a multinomial likelihood at the observation equation.

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 parameters (vectors) α and β, 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.

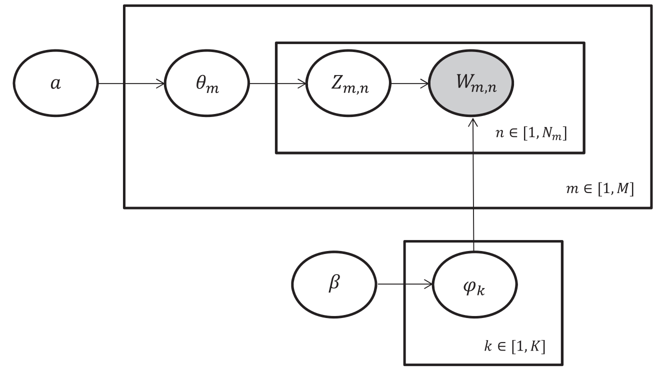

Figure 1. The LDA model visualized using plate notation.

More formally, the joint distribution of all known and hidden variables given the hyper parameters is

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

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 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 $P(\boldsymbol{W}, \boldsymbol{Z}; \alpha, \beta) = P(\boldsymbol{Z}|\boldsymbol{W}; \alpha, \beta)P(\boldsymbol{W}; \alpha, \beta)$ 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 online Appendix F.2.

The model is estimated using 7500×10 draws. The first 15,000 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. Visual inspection of changes in the model's perplexity score across Markov chain Monte Carlo (MCMC) iterations suggests that the model converges to its ergodic distribution (Heinrich 2009). Likewise, perplexity score comparisons across LDA models estimated using smaller numbers of topics indicate that 80 topics provide the best statistical decomposition of the DN corpus (Larsen and Thorsrud in press).

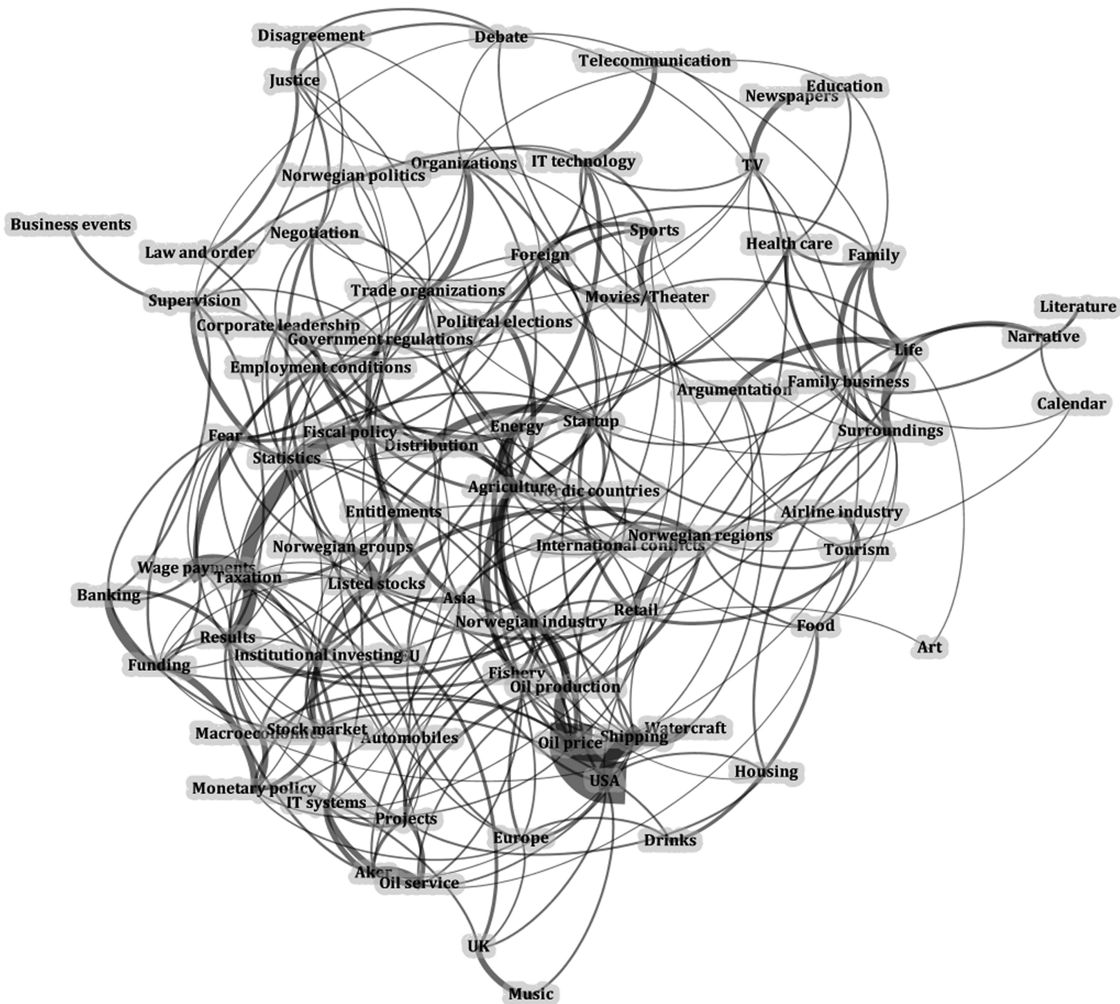

Figure 2. A network representation of the estimated news topics. The nodes in the graph represent the identified topics, and the edges illustrate the importance of the words that connect the topics. The *Fiscal policy* and *Taxation* topics, for example, share many important words and therefore have a thick edge. Since all topics share all words, only the 17 most important words in each topic are considered for visual clarity. Topics for which labeling is *Unknown*, confer Table 12 in online Appendix D, are removed from the graph.

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, see Table 12 in online Appendix D. While it is, in most cases, conceptually simple to classify them, the exact 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 is more interesting is whether the LDA decomposition gives a meaningful and easily interpretable topic classification of the DN newspaper. As illustrated in Figure 2, it does: The topic decomposition reflects how DN structures its content, with distinct sections for particular themes, and that DN is a Norwegian newspaper writing about news of particular relevance for Norway. We observe, for example, separate topics for Norway's immediate Nordic neighbors (*Nordic countries*), largest trading partners (*EU* and *Europe*), and biggest and second biggest exports (*Oil production* and *Fishing*). A richer discussion of this decomposition is provided in Larsen and Thorsrud (in press).

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 described in detail in online Appendices F.3 and F.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, that is, whether or not the news is positive or negative. This is done using an external word list and simple word counts, similar to in 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 a statistic containing the normalized difference between positive and negative words associated with a particular article. These statistics are then used to sign adjust the topic frequencies computed in

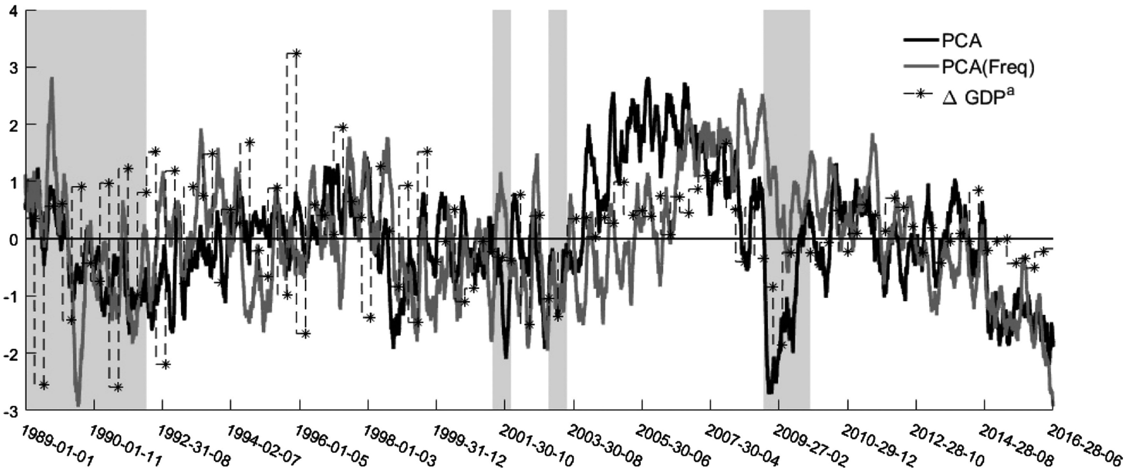

Figure 3. ΔGDP^a is the standardized first release of output growth. It is recorded at the end of each quarter, but reported on a daily basis using end-of-period values throughout the quarter. The black and gray lines are the standardized first principal component estimates of the news topic dataset and the news topic dataset without tone adjustment, respectively. Recession periods, defined by an MS-FMQ model, see Section 4.1, are illustrated using gray shading.

step one. Finally, I remove high-frequency noise from each topic time series by using a 60-day (backward looking) moving average filter, and, as is common in factor model studies (Stock and Watson 2016), standardize the resulting series. In Section 4.5, I show that the news-based methodology is robust to alternative "noise-removing" strategies.

Notice from the description above that also the tone adjustment procedure explicitly uses the output from the topic model. Still, the method used for identifying the tone of the news using dictionary-based techniques is simple, and could potentially be improved upon with more sophisticated algorithms (Pang, Lee, and Vaithyanathan 2002). While leaving such endeavors for future research, I have also tried to use the daily topic time series without the tone adjustment, see the discussion in Section 2.3 and the results in Section 4.1.

Figure 7 in online Appendix A reports 12 of the 80 topic time series. We observe that some of the topics covary, at least periodically. Overall, the average absolute value of the correlation among the topics is just 0.1, suggesting that topics are given different weight in the DN corpus across time.

2.3 Real-Time GDP and News

For estimation and evaluation I use a real-time dataset, including 68 vintages of seasonally adjusted quarterly Gross Domestic Product (GDP) growth rates, for the Norwegian mainland economy. In Norway, using GDP excluding the petroleum sector is the commonly used measure of economic activity. I follow suit here because it facilitates the formal evaluation of the *NCI* in Section 4. The real-time dataset is maintained by Norges Bank, and the vintages cover the time period 2000:Q1 to 2016:Q2. I sort these observations according to their release r, with $r = 1, \ldots, \bar{r}$. Thus, in real-time jargon, I work with the diagonals of the real-time data matrix. For future reference, I refer to these as $\Delta\text{GDP}_t^{r,k_q}$, where t is the daily time index and k_q denotes the quarterly observation interval. In other words, $\Delta\text{GDP}_t^{r,k_q}$ is observed for $t = k_q, 2k_q, \ldots$ $\bar{r} = 5$ is considered to be the "final release." Working with a higher $\bar{r}$ results in a loss of sample length, making model evaluation less informative. For time series observations prior to 2001:Q1, each r is augmented with earlier time series observations collected from the 2001:Q1 vintage. This data augmentation process might create a break in the time series. Model specifications allowing for time-varying parameters potentially adapt to such breaks. Prior to estimation, and as above, the series are standardized. To distinguish the adjusted series from the unadjusted growth rates, I label them $\Delta\text{GDP}_t^{r,a,k_q}$.

How do the news topics relate to output growth? To get a first pass impression I compute the first principal component of the news topic dataset, using all 80 topics, and label this factor *PCA*. The factor explains only roughly 15% of the overall variation among the news topics, but seems to capture important business cycle fluctuations surprisingly well, see Figure 3. For comparison, I also do the same principal component computations for the news topic dataset without tone adjustment, and label this factor *PCA(Freq)*. As seen from Figure 3, the *PCA(Freq)* also captures important business cycle variation, but, compared to the *PCA*, misses the timing of both the 2001 recession and the financial crisis in 2008. For this reason, I continue to work with the tone adjusted news topic dataset.

Although good, the *PCA* does not track output well during the early and later part of the sample. Still, it is interesting that an unsupervised principal component decomposition of newspaper topics provides information about the cyclical variations in output growth in the manner reported here. It is not only a novel finding in itself, but also motivates the usage of a more supervised factor model using this type of data.

3. THE DYNAMIC FACTOR MODEL

To estimate a coincident index of business cycles using the joint informational content in quarterly output growth and higher frequency variables, I build on Mariano and Murasawa (2003), Aruoba, Diebold, and Scotti (2009), and Marcellino, Porqueddu, and Venditti (2016), and develop a mixed-frequency time-varying dynamic factor model (DFM).

Following Harvey (1990), and letting bold-font letters denote vectors and bold-font capital letters matrices, the DFM containing quarterly, monthly and daily variables, can be written as (abusing the notation used in Section 2.1):

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{a}_t + \boldsymbol{e}_t \tag{2a}$$

$$\boldsymbol{a}_t = \boldsymbol{F}_t \boldsymbol{a}_{t-1} + \boldsymbol{R}_t \boldsymbol{\Sigma}_t \boldsymbol{\omega}_t \tag{2b}$$

$$\boldsymbol{e}_t = \boldsymbol{P} \boldsymbol{e}_{t-1} + \boldsymbol{u}_t \tag{2c}$$

with

$$\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{y}_t^{k_q} \\ \boldsymbol{y}_t^{k_m} \\ \boldsymbol{y}_t^{d} \end{pmatrix} \quad \boldsymbol{Z}_t = \begin{pmatrix} \bar{\boldsymbol{Z}}^{k_q} & 0 & 0 \\ 0 & \bar{\boldsymbol{Z}}^{k_m} & 0 \\ 0 & 0 & \boldsymbol{z}_t^d \end{pmatrix}$$

$$\boldsymbol{a}_t = \begin{pmatrix} \boldsymbol{a}_t^{k_q} \\ \boldsymbol{a}_t^{k_m} \\ a_t^{d} \end{pmatrix} \qquad \boldsymbol{e}_t = \begin{pmatrix} \boldsymbol{e}_{1,t}^{k_q} \\ \boldsymbol{e}_{2,t}^{k_m} \\ \boldsymbol{e}_{3,t}^{d} \end{pmatrix}$$

$$\boldsymbol{F}_t = \begin{pmatrix} \boldsymbol{\Upsilon}_t^{k_q} & 0 & -\boldsymbol{\pi}_t^{k_q}\Phi \\ 0 & \boldsymbol{\Upsilon}_t^{k_m} & -\boldsymbol{\pi}_t^{k_m}\Phi \\ 0 & 0 & \Phi \end{pmatrix} \quad \boldsymbol{R}_t = \begin{pmatrix} 1 & 0 & -\boldsymbol{\pi}_t^{k_q} \\ 0 & 1 & -\boldsymbol{\pi}_t^{k_m} \\ 0 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \boldsymbol{\sigma}_{t,\omega_q} & 0 & 0 \\ 0 & \boldsymbol{\sigma}_{t,\omega_m} & 0 \\ 0 & 0 & \sigma_{t,\omega_d} \end{pmatrix}$$

$$\boldsymbol{\omega}_t = \begin{pmatrix} \boldsymbol{\omega}_{t,q} \\ \boldsymbol{\omega}_{t,m} \\ \omega_{t,d} \end{pmatrix} \quad \boldsymbol{P} = \begin{pmatrix} \boldsymbol{\Phi}^{k_q} & 0 & 0 \\ 0 & \boldsymbol{\Phi}^{k_m} & 0 \\ 0 & 0 & \boldsymbol{\Phi}^{d} \end{pmatrix} \quad \boldsymbol{u}_t = \begin{pmatrix} \boldsymbol{u}_t^{k_q} \\ \boldsymbol{u}_t^{k_m} \\ \boldsymbol{u}_t^{d} \end{pmatrix},$$

where t is the daily time index, k_q, k_m, and d denote the quarterly, monthly, and daily observation intervals, respectively, and the model has been written with simple autoregressive time series processes of order one for notational simplicity.

Equation (2a) is the observation equation of the system. $\boldsymbol{y}_t^{k_q}$, $\boldsymbol{y}_t^{k_m}$, $\boldsymbol{y}_t^{d}$, are $N_q \times 1$, $N_m \times 1$, and $N_d \times 1$ vectors of quarterly, monthly, and daily variables, respectively, with $N = N_q + N_m + N_d$. $\boldsymbol{Z}_t$ is an $N \times N_a$ matrix with dynamic factor loadings linking the variables in $\boldsymbol{y}_t$ to the latent dynamic factors in $\boldsymbol{a}_t$. The time series processes for the time-varying elements in $\boldsymbol{Z}_t$ are modeled following the Latent Threshold Model (LTM) idea by Nakajima and West (2013), and described in greater detail below. The vector $\boldsymbol{e}_t$ contains the idiosyncratic errors. It is assumed that these evolve as independent AR(p) processes given by (2c), where $\boldsymbol{u}_t \sim \text{iid}N(0, \boldsymbol{U})$, and $\boldsymbol{U}$, $\boldsymbol{\Phi}^{k_q}$, $\boldsymbol{\Phi}^{k_m}$, and $\boldsymbol{\Phi}^{d}$ are diagonal matrices.

Equation (2b) is the transition equation of the system. The common factors follow a VAR(h) process, where Φ determines the time dependence. $\boldsymbol{\omega}_t \sim \text{iid}N(0, \boldsymbol{I})$ and $\boldsymbol{\Sigma}_t$ is a diagonal matrix with $\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t' = \boldsymbol{\Omega}_t$, allowing for stochastic volatility. Marcellino, Porqueddu, and Venditti (2016) were the first to introduce this feature into mixed-frequency models, finding that it leads to an improvement in point (and density) forecast accuracy. In this article, it is included to also capture the obvious changes in output growth volatility seen in Figure 3 between the first and latter part of the sample. The individual elements in $\boldsymbol{\Sigma}_t$ are assumed to follow random walk processes:

$$\log(\boldsymbol{\sigma}_{t,\omega.}) = \log(\boldsymbol{\sigma}_{t-1,\omega.}) + \boldsymbol{b}_{t,\cdot} \quad \boldsymbol{b}_{t,\cdot} \sim \text{iid}N(0, \boldsymbol{B}.), \tag{3}$$

where $\boldsymbol{B}.$ is a diagonal matrix.

The last element in $\boldsymbol{a}_t$, the scalar a_t^d, is interpreted as the latent common daily business cycle index. The other elements in $\boldsymbol{a}_t$, and in $\boldsymbol{F}_t$ and $\boldsymbol{R}_t$, contain cumulator variables used to handle the mixed-frequency property of the model, an issue I turn to next.

As is common in mixed-frequency models, lower frequency variables are treated as daily series with missing observations (Foroni and Marcellino 2013). For a generic variable y_t^k, time aggregation from higher to lower frequency is restricted as follows:

$$\begin{aligned} y_t^k &= \log(\upsilon_{1,t}^k) - \log(\upsilon_{1,t-k}^k) \\ &\approx \log\left(\sum_{i=0}^{k-1} \upsilon_{1,t-i}\right) - \log\left(\sum_{i=k}^{2k-1} \upsilon_{1,t-i}\right) \\ &\approx \sum_{i=0}^{k-1} \log(\upsilon_{1,t-i}) - \sum_{i=k}^{2k-1} \log(\upsilon_{1,t-i}) \\ &= \sum_{i=0}^{2k-2} \omega_i^k y_{1,t-i}, \quad t = k, 2k, \ldots, \end{aligned} \tag{4}$$

where y_t^k is the observed low-frequency growth rate, υ_t^k its level, and $\omega_i^k = i + 1$ for $i = 0, \ldots, k-1$; $\omega_i^k = 2k - i - 1$ for $i = k, \ldots, 2k-2$; and $\omega_i^k = 0$ otherwise. Imposing a common factor structure for y_t^k, it follows from (4) that at the observation interval:

$$y_t^k = \sum_{i=0}^{2k-2} \omega_i^k y_{1t-i} = \sum_{i=0}^{2k-2} \omega_i^k (z a_{t-i}^d + e_{t-i}). \tag{5}$$

A caveat with the model formulation in (5) is that it increases the number of state variables in the system considerably. For example, when aggregation is from daily to quarterly frequency, the number of elements in the state vector exceed 180, posing significant challenges for estimation. To limit the size of the state vector, temporal aggregation is handled using a double cumulator variable approach as in Banbura et al. (2013). The temporal aggregator variables are recursively updated such that at the end of each respective period we have

$$a_t^k = \sum_{i=0}^{2k-2} \omega_i^k a_{t-i}, \quad t = k, 2k, \ldots \tag{6}$$

As shown in online Appendix G, these recursions can be computed with the help of only two additional state variables and selection and weight matrices. In (2) this is reflected in the partition $\boldsymbol{a}_t^k = (\, a_t^k \;\; \bar{a}_t^k \,)'$, the selection matrix $\boldsymbol{\Upsilon}_t^k$, and the vector $\boldsymbol{\pi}_t^k$ which contains the aggregation weights ω_i^k. Accordingly, $\bar{\boldsymbol{Z}}^k = (\, z^k \;\; 0 \,)$. Notice here that the factor loadings are static. Allowing for time-varying loadings for the low-frequency variables will be in conflict with the aggregation scheme in (5) and (6).

The time aggregation structure of the model, given by Equation (5), introduces moving average terms into the idiosyncratic errors for the monthly and quarterly variables. In the case of only one monthly and quarterly variable, this is captured by the

$\boldsymbol{R}_t \boldsymbol{\Sigma}_t \boldsymbol{\omega}_t$ term in (2b). However, allowing for such time series patterns, I find that the model becomes substantially more difficult to estimate. For this reason, I follow the specification adopted in Banbura et al. (2013), and assume iid errors at the monthly and quarterly observation intervals. This amounts to restricting $\boldsymbol{R}_t = [-\boldsymbol{\pi}_t^{k_q} \ -\boldsymbol{\pi}_t^{k_m} \ 1]'$, $\boldsymbol{\Sigma}_t = \sigma_{t,\omega_d}$, $\boldsymbol{\omega}_t = \omega_{t,d}$, and $\boldsymbol{\Phi}^{k_q} = \boldsymbol{\Phi}^{k_m} = 0$.

Finally, I globally identify the sign and size of the latent factor by restricting the factor loading for the first element among the N_d variables to equal 1 for all time periods using the daily *Monetary policy* news topic as the normalizing variable. Bai and Wang (2014) and Bai and Wang (2015) showed that this restriction uniquely identifies the factor and the loadings, but leaves the transition equation dynamics completely unrestricted. In online Appendix B, I show that the model estimates are robust to using alternative news topics as the normalizing variable.

3.1 Enforcing Sparsity

Following the Latent Threshold Model (LTM) idea introduced by Nakajima and West (2013), and applied in a factor model setting in Zhou, Nakajima, and West (2014), dynamic sparsity is enforced through the time-varying factor loadings using a latent threshold mechanism. For one particular element in the $\boldsymbol{z}_t^d$ vector, $z_{i,t}$, the LTM structure can be written as

$$z_{i,t} = z_{i,t}^* \varsigma_{i,t} \quad \varsigma_{i,t} = I(|z_{i,t}^*| \geq d_i), \tag{7}$$

where

$$z_{i,t}^* = z_{i,t-1}^* + w_{i,t} \tag{8}$$

with $w_{i,t} \sim \text{iid}N(0, \sigma_{i,w}^2)$, and $\boldsymbol{w}_t \sim \text{iid}N(0, \boldsymbol{W})$ where $\boldsymbol{W}$ is a diagonal matrix. In (7) $\varsigma_{i,t}$ 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 $\varsigma_{i,t} = 0$.

In general, the LTM framework is useful for models where the researcher wants to introduce dynamic sparsity. For example, as shown in Zhou, Nakajima, and West (2014), allowing for such mechanism uniformly improves out-of-sample predictions in a high-dimensional portfolio analysis due to the parsimony it induces. Here, the LTM concept serves one additional purpose. If estimating constant factor loadings, the researcher might conclude that a given topic has no relationship with a_t^d, that is, that z_i^d equals zero for all time periods, simply because, on average, periods with a positive $\boldsymbol{z}_i^d$ cancels with periods with a negative $\boldsymbol{z}_i^d$. The threshold mechanism potentially captures such cases in a consistent and transparent way, and controls for the fact that the relationship between the news topics and output growth might be unstable, confer the discussion in Section 1. A related concept in this respect is the spike-and-slab prior used in a high-dimensional factor model by Scott and Varian (2013). In contrast to this approach, however, the LTM mechanism allows for dynamic variable selection.

3.2 Model and Prior Specifications

Apart from the usage of newspaper data, the mixed-frequency property, and the LTM mechanism used for the factor loadings, the time-varying DFM described above is fairly standard (Lopes and Carvalho 2007; Del Negro and Otrok 2008; Ellis, Mumtaz, and Zabczyk 2014; Bjørnland and Thorsrud in press). In the interest of brevity, estimation details are relegated to online Appendix H. I shortly note that the DFM is estimated by decomposing the problem of drawing from the joint posterior into a set of much simpler ones using MCMC simulations. The full-sample-based results are all based on 50,000 iterations. The first 10,000 are discarded and only every 10th of the remaining are used for inference.

To implement the MCMC algorithm, prior specifications for the initial state variables $\boldsymbol{a}_0$, $\boldsymbol{Z}_0$, $\boldsymbol{\Sigma}_0$, and for the hyper-parameters $\boldsymbol{B}, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{F}_t, \boldsymbol{P}$, and $\boldsymbol{d}$ are needed. The prior specifications used for the initial states take the following form: $\boldsymbol{a}_0 \sim N(0, I \cdot 10)$, $\boldsymbol{Z}_0 \sim N(0, I)$, and $\boldsymbol{\Sigma}_0 \sim N(1, I)$. The priors for the hyper-parameters $\mathit{\Phi}$ and Φ, which are part of the $\boldsymbol{F}_t$ and $\boldsymbol{P}$ matrices, respectively, are set to $\mathit{\Phi} \sim N(0, I)$ and $\Phi_i \sim N(0, 0.5)$. For the constant parameters in $\boldsymbol{Z}_t$, that is, $\boldsymbol{Z}^k$, I assume for each element i that $z_i^k \sim N(1, 1)$. The priors for $\boldsymbol{B}$, $\boldsymbol{U}$, and $\boldsymbol{W}$ 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: $\sigma_{b_d}^2 \sim \text{IG}(T^{b_d}, \kappa_{b_d}^2)$ with $T^{b_d} = T \cdot 0.1$ and $\kappa_{b_d} = 0.01$; $\sigma_{i,u}^2 \sim \text{IG}(T^u, \kappa_u^2)$ with $T^u = T \cdot 0.5$ and $\kappa_u = 0.3$; $\sigma_{i,w}^2 \sim \text{IG}(T^w, \kappa_w^2)$ with $T^w = T \cdot 1$ and $\kappa_w = 0.003$, where T is the sample size. As the full sample contains over 10,000 observations, these priors are informative for the variance terms associated with the time-varying factor loadings, but less so for the other parameters. Finally, to draw the latent threshold, $\boldsymbol{d}$, a K parameter needs to be defined. K controls our prior belief concerning the marginal sparsity probability. A neutral prior will support a range of sparsity values to allow the data to inform on relevant values. For stationary processes, Nakajima and West (2013) gave some advice in terms of setting K using the parameters' marginal distribution. Here the parameters follow simpler random walk processes which are nonstationary, and do not have a marginal distribution. I set $K = 0.4$, but have experimented with estimating the model using different values for K, finding that higher values, coupled with the rather tight priors for the variance of the factor loadings, result in an unreasonably large degree of sparsity. An alternative approach would be to treat K as a tuning parameter, and more formally set its value based on a specific loss function.

In the proposed model, labeled *NCI*, I only include $\Delta\text{GDP}_t^{1,a,k_q}$ and all the daily news topics, that is, $N_q = 1$ and $N_d = 80$. The MCMC simulations are initialized using simple ordinary least-square (OLS) estimates obtained using the first principal component of the news topics as a measure of the daily business cycle index. To investigate the potential gains or losses of the LTM mechanism, the stochastic volatility component, and the usage of the newspaper data, I also estimate the system in (2) using four alternative specifications, summarized by the model names NCI^{notvp}, NCI^{nosw}, $NCI^{notvpsw}$, and *CI*. In the alternative NCI^{notvp}, I turn off the time-varying parameters associated with the daily factor loadings, but keep the information set as above. In this case $z_t^d = z^d$ for all time periods, with the prior assumption $z_i^d \sim N(0, 1)$. In the alternative NCI^{nosw}, I turn off the stochastic volatility component, and draw the constant variance from an Inverse-Gamma prior. The $NCI^{notvpsw}$ model is estimated without allowing for any time-varying parameters, while the *CI* specification resembles a more conventional coincident index model in terms of data usage. In particular, only monthly and daily hard economic indicators are used as observables together with $\Delta\mathrm{GDP}_t^{1,a,k_q}$. Among the observable variables used are commonly applied business cycle indicators like the difference between long- and short-run interest rates (*Spread*), the return on the Oslo Stock Exchange (*OSEBX*), and labor market conditions (*LF*). A more detailed description of the data used in the *CI* model is given in online Appendix D. In total $N_q = 1$, $N_m = 16$, and $N_d = 4$ for this model. For all four model specifications I allow for one lag in the equation for the idiosyncratic errors ($p = 1$), and up to ten lags for the latent common business cycle index ($h = 10$). The (full) estimation sample is January 1, 1989 to June 28, 2016, yielding 10,041, 329, 109, daily, monthly, and quarterly observations, respectively.

4. A NEWSY COINCIDENT INDEX OF THE BUSINESS CYCLE

The full-sample-based estimate of the *NCI* is illustrated in Figure 4. As clearly seen in the figure, the index tracks the general economic fluctuations closely. For example, compared to the simple PCA estimates reported in Figure 3, the *NCI* provides a better fit: It captures the recession 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. Note, however, that in Norway, the downturn in the economy following the Norwegian banking crisis in the late 1980s was just as severe as the downturn following the global financial crisis in 2008. Figure 8 in online Appendix A reports the *NCI* together with the alternative indexes $NCI^{not\upsilon psw}$ and *CI*. Again, simple visual inspection suggests that the *NCI* tracks the overall state of the economy better than the alternatives.

The time-varying changes in the variance of the *NCI* errors are illustrated in Figure 9 in online Appendix A. Unexpectedly, the model picks up a substantially higher variance in the first part of the sample relative to in the latter part, although the financial crisis period is associated with increased uncertainty. Convergence statistics indicating that the MCMC algorithm has reached the ergodic distribution are discussed in online Appendix C.

In the following I first formally evaluate the model's in-sample classification properties. Then I illustrate how movements in the *NCI* can be decomposed into news topic contributions and how the sparsity structure of the model changes significantly across time. Finally, the model is tested in an out-of-sample nowcasting experiment.

4.1 In-Sample Evaluation

Like in Travis and Jordà (2011), and in the tradition of Burns and Mitchell (1946), I categorize aggregate economic activity into phases of expansions and contractions and evaluate the index's ability to classify such phases using Receiver Operating Characteristic (ROC) curves and area under the curve (AUROC) statistics.

In contrast to in, for example, the U.S., which has an official business cycle dating committee (NBER), no formal dating exists for Norway. For this reason, I use four different business cycle chronologies developed by Aastveit, Jore, and Ravazzolo (2016) for the Norwegian economy as measures of the "truth." Each chronology is constructed using different methodologies to extract the unobserved phases: uni- and multivariate Bry-Boschan approaches (BB-GDP and BB-ISD), a univariate Markow-switching model (MS-GDP), and a Markov-Switching factor model (MS-FMQ). Since all these methods provide a quarterly classification of the business cycle phases, daily classifications are obtained by assuming that the economy remains in the same phase on each day within the quarterly classification periods. Details about the chronologies are summarized in Table 11 in online Appendix D.

I also compare the *NCI*'s performance against the four model-based alternatives described at the end of Section 3.2 ($NCI^{not\upsilon p}$,

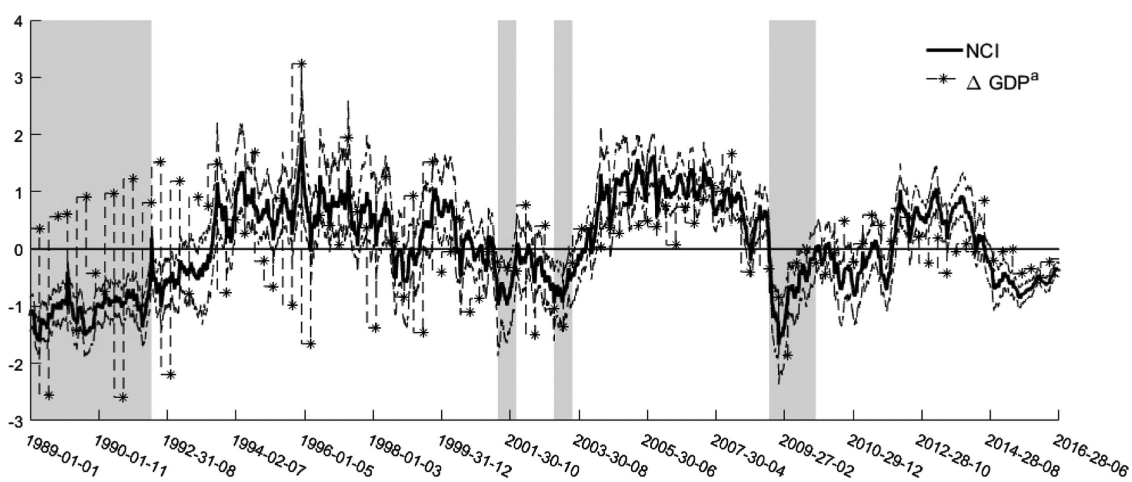

Figure 4. $\Delta\mathrm{GDP}^a$ is the standardized first release of output growth. It is recorded at the end of each quarter, but reported on a daily basis using end-of-period values throughout the quarter. *NCI* is the standardized measure of the daily coincident index. The dotted black lines are 68% probability bands. The solid black line is the median estimate. Recession periods, defined by an MS-FMQ model, see Section 4.1, are illustrated using gray shading.

Table 1. Receiver operating characteristics and area under the curve (AUROC) statistics. By definition the AUROC cannot exceed 1, perfect classification, or be lower than 0.5. I compute the AUROC score nonparametrically using the algorithm described in Travis and Jordà (2011)

		Reference chronologies			
	Model	BB-GDP	MS-GDP	BB-ISD	MS-FMQ
	NCI	0.825	0.885	0.837	0.938
Relative AUROC	NCI/NCI^{notvp}	1.122	1.108	1.061	1.076
	NCI/NCI^{nosw}	0.999	1.001	1.051	1.001
	$NCI/NCI^{notvpsw}$	1.248	1.175	1.119	1.201
	NCI/CI	1.261	1.151	1.040	1.257
	NCI/PCA	1.352	1.262	1.215	1.325
	$NCI/PCA(Freq)$	2.687	2.025	2.067	1.825
	$NCI/Spread$	1.316	1.178	1.086	1.283
	$NCI/OSEBX$	1.528	1.528	1.616	1.763
	NCI/LF	1.297	1.446	1.426	1.457

NCI^{nosw}, $NCI^{notvpsw}$, and *CI*), and five alternative variables. The five alternative variables include commonly applied economic indicators like the difference between long- and short-run interest rates (*Spread*), the return on the Oslo Stock Exchange (*OSEBX*), and labor market conditions (*LF*), as well as the *PCA* and *PCA(Freq)* news-based indicators presented in Section 2.3.

Focusing on AUROC statistics, Table 1 summarizes the in-sample classification scores. Figure 10 in online Appendix A reports the associated ROC curves for the *NCI*, $NCI^{notvpsw}$, *CI*, and *PCA* models. Irrespective of which reference cycle that is used, the *NCI* is able to provide a good classification of the Norwegian business cycle. Evaluated against the chronology preferred by Aastveit, Jore, and Ravazzolo (2016), namely, the MS-FMQ, the *NCI*'s classification power is almost perfect with an AUROC score of 0.938. Compared with existing results, this score is very competitive. In Aastveit, Jore, and Ravazzolo (2016), for example, the quarterly BB-GDP model has an AUROC of 0.93 when evaluated against the MS-FMQ reference chronology. Likewise, using U.S. data, and comparing various leading indicators and coincident indexes, Travis and Jordà (2011) showed that the best performing index is the one developed by Aruoba, Diebold, and Scotti (2009). This index receives an AUROC of 0.96 when the NBER business cycle chronology is used as a reference cycle. While these results are strong, although not perfect, the *NCI* might provide an estimate of the economy's phases that is closer to the unknown truth than any of the other reference cycles I use to evaluate it. In addition, the reference chronologies are defined based on quarterly data, while the *NCI* is scored based on its daily classification power.

Allowing for the LTM mechanism is important for the index's classification power. In terms of AUROC the NCI^{notvp} model performs between 6% and 12% worse than the *NCI* model, while turning off the stochastic volatility component (NCI^{nosw}) does not alter the results relative to the *NCI* by much. However, estimating the news-based model without allowing for any time-varying parameters ($NCI^{notvpsw}$) results in substantially lower scores. This suggests that some type of time variation in the model's signal-to-noise ratio is needed to obtain strong results. Likewise, we see from the comparison against the *CI* model that the inclusion of news topic variables in the DFM generally improves classification power substantially. Thus, it is the combined usage of daily news topic variables and allowing for time-varying parameters that enhances the *NCI*'s classification power relative to the alternatives.

Finally, we see from Table 1 that all model estimates tend to classify the business cycle phases better than any single indicator alone. The *Spread* has a better performance than the other indicators, but it is still up to 31% worse than the *NCI*. At the same time, the *PCA* indicator is usually better than the *LF* indicator, while the *OSEBX* and *PCA(Freq)* indicators are the worst performing, by far, across all reference chronologies.

In sum, the *NCI* classifies the phases of the business cycle very well, and better than the alternatives considered here. The good classification properties are due to both the usage of daily newspaper topics and the LTM mechanism.

4.2 News and Index Decompositions

An advantage with the *NCI* is that it can provide the index user with broad-based information about the type of news contributing to the index's fluctuations at a daily frequency. Technically, this is done using Kalman filter iterations and decomposing the state evolution at each updating step into news contributions using the Kalman Gain and the recursive nature of the filter, see online Appendix I and Banbura et al. (2013). Figure 5 provides an illustration of how news surprises affect the updated index estimates across time. Three distinct results stand out.

First, the degree of sparsity enforced on the factor loading space changes considerably across time. For example, toward the latter part of the sample, few factor loadings have a high probability of being zero. Toward the end of the 1990s, on the other hand, the degree of sparsity is much larger, with only a few factor loadings being above (in absolute value) their respective threshold.

Second, the topics that frequently contribute to the index movements 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*, *Funding*, *Retail*, *Airline industry*, and *Automobiles* topics, while the *Oil production* and *Oil service* topics are examples

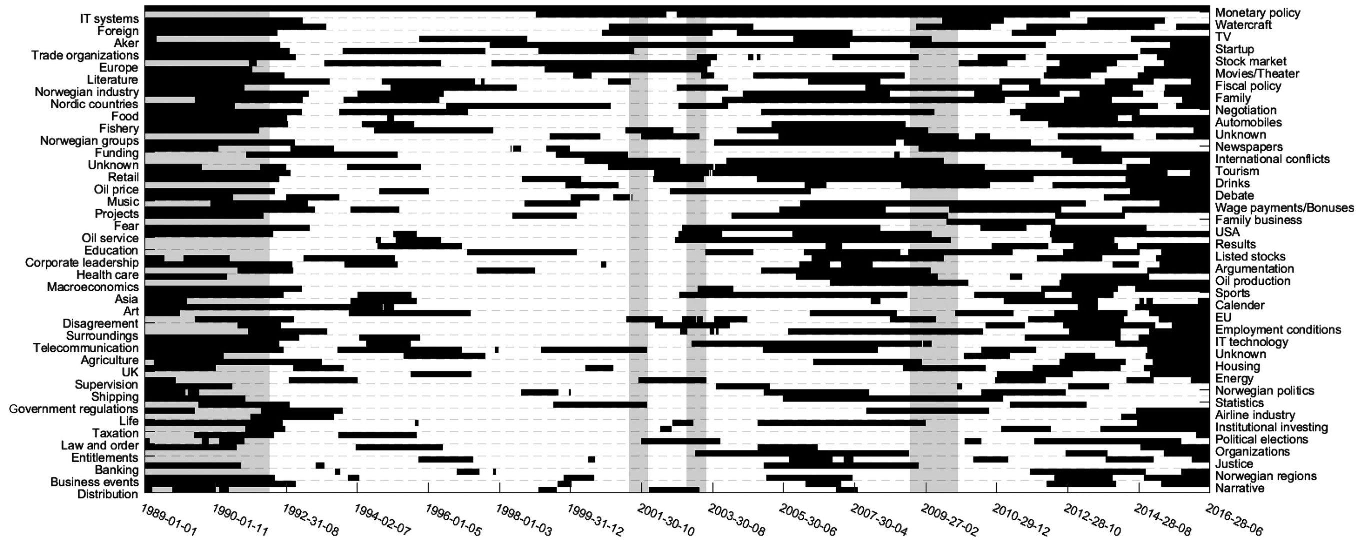

Figure 5. News topics and their contribution to *NCI* estimates across time. The reported decompositions are based on running the Kalman Filter using the posterior median estimates of the hyper-parameters and the time-varying factor loadings (at each time t). In the interest of readability, the topic names are reported on two y-axes with two-step increments. For example, the *Monetary policy* topic is associated with the first row (from above) in the figure, while the *IT systems* topic is associated with the second row (from above). White areas illustrate the time-varying sparsity patterns. Recession periods, defined by an MS-FMQ model, see Section 4.1, are illustrated using gray shading.

of the latter. Still, although most topics are easily interpretable and provide information about what is important for the current state of the economy, 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. It is also the case that a newspaper article is a mixture of topics. To the extent that different topics, meaningful or not from an economic point of view, stand close to each other in the decomposition of the corpus (see Figure 2), they might covary and therefore both add value in terms of reflecting the current state of the economy.

Third, the timing of when specific topics become important, either positively or negatively, resonates reasonably well with what we now know about the economic developments the last two decades. At the risk of cherry picking, but without dredging too deep into the historical narrative of the Norwegian business cycle, I give three examples illustrated in the colored version of Figure 5, see Figure 11 in online Appendix A: It is by now well recognized that the extraordinary boom in the Norwegian economy during the mid-2000s was highly oil-driven. The large positive contribution from the *Oil service* and *Wage payments/Bonuses* topics reflect this. We also see that topics like *Fear* and *Funding*, newspaper topics associated with uncertainty and credit and loans, contributed especially negatively during the Great Recession period, while topics like *IT systems* and *Startup* were influential around the turn of the century.

4.3 Out-of-Sample Evaluation

An important use case for daily business cycle indicators is short-term forecasting. I assess the *NCI*'s forecasting performance by running an out-of-sample nowcasting experiment. The DFM is first estimated using the sample June 1, 1988 to December 31, 2003. Then, for each quarter from 2004:Q1 to 2016:Q2, I recursively update the model to produce forecasts for evaluation. The sample split is chosen to give the model a substantial number of observations to learn the hyper-parameter distributions prior to doing the nowcasting experiment. The informational assumptions are as follows: For a generic quarter T_q, the DFM is updated seven times, assuming that daily newspaper information 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 $T-40$ in Norway, the nowcasts constructed for $T-d \leqslant T-40$ 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. The predictions constructed for $T+d > T$ are labeled backcasts because the daily information set used to construct the predictions are extended into the subsequent quarter (T_q+1).

Following, for example, Giannone, Reichlin, and Small (2008), and letting $\Delta\text{GDP}^1_{t_q}$ denote output growth measured at the quarterly time interval, a rescaled forecast for nonadjusted GDP growth for quarter T_q is, at each forecast origin, constructed from the simple projection:

$$\Delta\widehat{\text{GDP}}^1_{T_q} = \hat{\alpha} + \hat{\beta}\hat{y}^{k_q}_{I(T)}, \tag{9}$$

where $\hat{y}^{k_q}_{I(T)}$ is the implied quarterly growth rate of $\Delta\text{GDP}^{1,a}_{T_q}$ based on the information set $I(T) = \{T-80, \ldots, T+30\}$, and $\hat{\alpha}$ and $\hat{\beta}$ are (recursively estimated) in-sample OLS estimates. Notice here that an estimate of $\hat{y}^{k_q}_{I(T)}$ can always be made available from the DFM by forecasting a sufficient number of daily periods forward.

A full reestimation of the DFM takes many hours (or days), while the nonlinear property of the model prevents the usage of standard Kalman Filtering techniques for online updating. For these reasons, I develop a mixture auxiliary particle filter, allowing for computational parallelization, to update the latent state variables ($\boldsymbol{a}_t$, $\boldsymbol{Z}_t$, $\boldsymbol{\Sigma}_t$) conditional on the hyper parameters ($\boldsymbol{B}, \boldsymbol{U}$, $\boldsymbol{W}$, $\boldsymbol{F}_t$, $\boldsymbol{P}$, and $\boldsymbol{d}$) and the data. A description of the algorithm, building on Pitt and Shephard (1999), Chen and Liu (2000), and Doucet, de Freitas, and Gordon (2001), is provided in online Appendix K. This filter is then applied on all within quarter updates. A full MCMC reestimation of the model is only done at the end of a generic quarter, that is, at day T.

To avoid potential look-ahead biases when using the (full-sample-based) news topic estimates, the news corpus is truncated to end on the last day of 2003, and 80 new news topics are estimated using the LDA model. Then, using the estimated topic distributions from the truncated corpus, the period 2004–2016 is classified, and topic time series for the whole sample period (1989–2016) are constructed as described in Section 2. The out-of-sample classification is done following Heinrich (2009) and Hansen, McMahon, and Prat (2018), where a procedure for querying documents outside the set on which the LDA is estimated is implemented. This corresponds to using the same Gibbs simulations described in online Appendix F.2, but with the difference that the sampler is run with the estimated word distributions (from the training sample) held constant. Updating the topic model recursively is infeasible, as one estimation round takes roughly 1 week using our infrastructure. Thus, at the end of the evaluation period, we are essentially using topic distributions based on 12-year old estimates. As mentioned in online Appendix F.2, another caveat with reestimating the LDA recursively is the lack of identifiability. That is, topic estimates cannot be combined across samples for an analysis that relies on the content of specific topics.

I focus on point forecast evaluation using Root Mean Squared Forecast Error (RMSFE) statistics. 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, I report and evaluate the nowcasts against both the first and fifth release of GDP growth, that is, $r = 1$ and $r = \bar{r}$. Nowcast evaluations against other releases are qualitatively similar, and I show robustness to using the final vintage as a measure of "true" growth.

The accuracy of the *NCI*-based predictions are compared to eight different benchmarks. The four first benchmark models are the same as those described in the latter part of Section 3.2, that is, NCI^{notvp}, NCI^{nosw}, $NCI^{notvpsw}$, and *CI*. These models are updated as described above. For the *CI* model I assume that the monthly variables become available at the last day of each month. I also compare the *NCI*-based predictions to two more simplistic models, namely, an *AR* of order one, and a constant growth model (*RW*). Finally, I compare the *NCI*-based nowcasts to the official Norges Bank nowcasts (*NB*) and predictions from Norges Bank's model-based nowcasting system (System for Averaging Models, *SAM*).

The Norges Bank nowcasts are interesting benchmarks for two reasons. First, the *NB* predictions are subject to judgment, and potentially incorporate both hard and soft (news) economic information. Second, in contrast to the *NB* predictions, the *SAM* predictions are purely model-based and produced using a state-of-the-art forecast combination system. In line with a large forecasting literature documenting the potential benefits of using forecast combination techniques (Timmermann 2006), Aastveit et al. (2014) used the same system to nowcast U.S. GDP growth, and document superior performance relative to a simple model selection strategy. Thus, in a pure forecasting horse-race, it is difficult to envision a better model-based competitor than *SAM*. A brief description of the *SAM* system is provided in online Appendix E together with a description of how the two Norges Bank nowcasts are compiled to match the timing assumptions of the nowcasting experiment.

Table 2 reports the RMSFE statistics. Irrespective of whether the forecasts are evaluated against the first (Panel A) or fifth release (Panel B), the *NCI* forecasts generally improve as news accumulate toward the end of the quarter. The biggest gains are typically obtained when we have a full quarter of the daily newspaper topics. For example, at time T, the improvement in RMSFE relative to at time $T - 80$ is 7% when evaluated against the fifth data release.

As in the in-sample evaluation, it is the combined usage of time-varying parameters and the news topic variables that adds value. The NCI^{notvp}, $NCI^{notvpsw}$, and *CI* models perform, at times, substantially worse than the *NCI* model. The performance of the $NCI^{notvpsw}$ model is especially bad, echoing the in-sample results, and suggesting that some type of time variation in the signal-to-noise ratio is beneficial when working with a model or data of this type. When evaluated against the first data release, the differences in RMSFE are also generally significant. Likewise, turning off the stochastic volatility component (NCI^{nosw}) leads to a deterioration in forecasting performance. Among the simpler benchmark models, the *RW* performance is substantially worse than the news-based model irrespective of informational assumptions. The *AR*, on the other hand, is performing relatively good at the two-step ahead horizon, but deteriorates dramatically when the predictions are one-step ahead projections.

Compared to the more sophisticated benchmarks, *NB* and *SAM*, we see that the news-based predictions are sometimes better and sometimes worse. For example, when evaluated against the first (fifth) release, the news-based model is between 11 and 12 (1–4)% better (worse) than *SAM* and *NB* at time $T + 30$. Although the differences in performance are not significant for any of the comparisons, it is noticeable that one single news-based model can produce competitive predictions relative to *NB* and *SAM*.

The results reported thus far represent averages across the evaluation sample. Figure 6 reports the cumulative difference in squared prediction errors between the *NCI* and the two best-performing benchmarks *NB* and *SAM*. Looking first at the *NCI* relative to *SAM*, we see that when the business cycle turned heading for the Great Recession, the news-based model starts to improve. This improvement continues into the recovery face of the recession, but then levels off and worsens, irrespective of which data release that is used for evaluation. In contrast, when comparing the *NCI* to the *NB* predictions, containing both hard and soft information, the news-based model experience a more or less steady fall. However, following the large drop in oil prices and subsequent slower growth in the Norwegian economy toward the end of the evaluation period, the news-based model improves upon the *NB* predictions. Thus, there seems to be a small tendency that the news-based model is relatively good at capturing economic turning points. One reason for this might be the timeliness of news data. Another reason might be

Table 2. 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 $T-d \leqslant T-40$ are two-step ahead predictions (relative to the quarter T_q), and the predictions constructed for $T+d > T$ are termed backcasts. Tests for significant differences in forecasting performance are done using the Diebold–Mariano test statistic (Diebold and Mariano 1995). $*$, $**$, and $***$ denote the 10%, 5%, and 1% significance level, respectively. The test statistics are based on 50 out-of-sample forecast errors. Gray colored areas indicate that the alternative model is only updated at the beginning of the period. For example, the AR model can only be updated at time $T-80$ and $T-20$. See online Appendix E for a description of the timing assumptions regarding the *NB* and *SAM* predictions. Due to the data availability, the total number of out-of-sample observations equal 40 for the *NB* and *SAM* benchmarks

Type	Nowcast for T_q					Backcast for T_q	
Quarterly info.	T_q-2			T_q-1			
Daily info.	$T-80$	$T-60$	$T-40$	$T-20$	T	$T+10$	$T+30$
Panel A: First release							
RMSFE	0.384	0.400	0.422	0.387	0.377	0.379	0.378
Cum. imp. (%)		4.139	9.711	0.708	−1.886	−1.275	−1.669
Relative RMSFE							
NCI^{notvp}/NCI	1.767***	1.750***	1.559**	1.640***	1.350**	1.346**	1.347**
NCI^{nosw}/NCI	1.127**	1.074*	1.082*	1.106	1.072**	1.072**	1.076**
$NCI^{notvpsw}/NCI$	2.303***	2.240***	2.096***	2.362***	1.739***	1.714***	1.720***
CI/NCI	1.198*	1.167	1.070	1.330**	1.176*	1.171*	1.202*
AR/NCI	1.186	1.139	1.083	1.577**	1.619*	1.609*	1.615*
RW/NCI	2.261***	2.171***	2.064***	2.112***	2.168***	2.154***	2.163***
SAM/NCI	1.097	1.040	0.965	1.072	0.990	1.103	1.110
NB/NCI					1.000	1.114	1.121
Panel B: Fifth release							
RMSFE	0.537	0.538	0.547	0.520	0.497	0.502	0.500
Cum. imp. (%)		0.304	1.958	−3.153	−7.304	−6.472	−6.801
Relative RMSFE							
NCI^{notvp}/NCI	1.287**	1.324**	1.220	1.243**	1.107	1.105	1.104
NCI^{nosw}/NCI	1.083**	1.076***	1.092**	1.106**	1.074***	1.074***	1.077***
$NCI^{notvpsw}/NCI$	1.747***	1.764***	1.704***	1.867***	1.427***	1.407***	1.411***
CI/NCI	1.113	1.106	1.072	1.241**	1.080	1.076	1.102
AR/NCI	1.142	1.139	1.121	1.491***	1.557**	1.543***	1.549***
RW/NCI	1.696***	1.692***	1.665***	1.553***	1.622***	1.608***	1.614***
SAM/NCI	1.020	0.974	0.934	1.008	1.000	0.991	0.996
NB/NCI					0.961	0.953	0.958

(a) NCI relative to SAM **(b)** NCI relative to NB

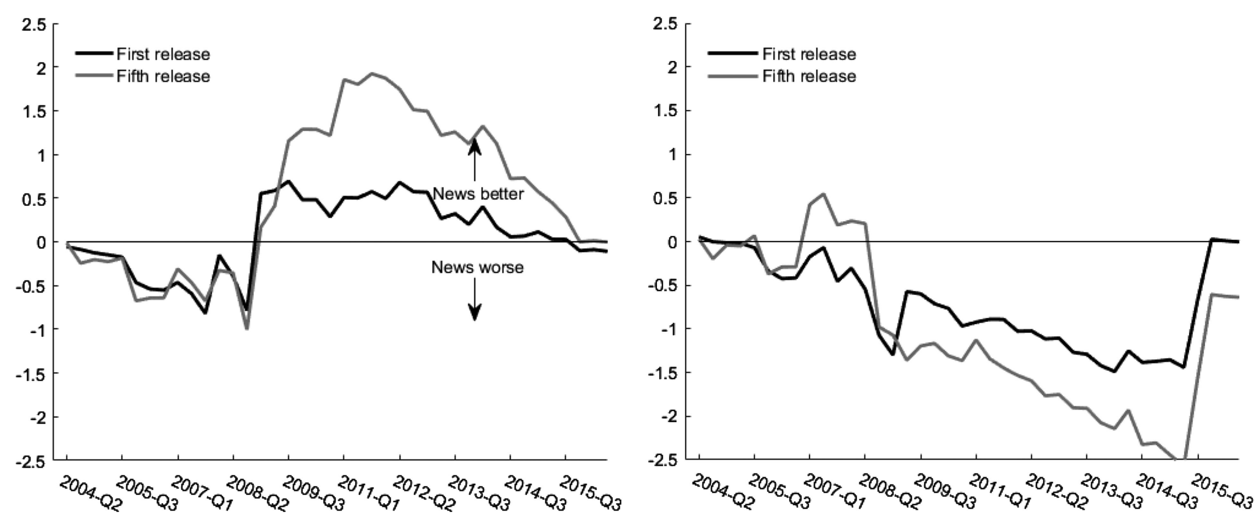

Figure 6. Cumulative difference in squared prediction errors between the *NB* and *SAM* benchmarks and the news-based *NCI* model. With reference to Table 2, all the predictions are generated at time T.

Table 3. Nowcasts regressed on outcomes and revisions. *SAM* and *NCI* predictions are constructed either early within the quarter (Early), or late in the quarter (Late). Standard errors in parenthesis. ∗, ∗∗, ∗ ∗ ∗, denote the 10%, 5%, and 1% significance level, respectively

	$\Delta GDP^1_{t_q}$		$\Delta GDP^5_{t_q}$		$\Delta GDP^5_{t_q} - \Delta GDP^1_{t_q}$			
	I	II	III	IV	V	VI	VII	VIII
β^{NB}	0.350**	0.374***	0.345**	0.563***			−0.005	0.058
	(0.161)	(0.097)	(0.144)	(0.123)			(0.096)	(0.090)
β^{SAM}_{Early}	0.443		0.869***				0.426*	
	(0.270)		(0.285)				(0.226)	
β^{SAM}_{Late}		0.323*		0.487*				−0.051
		(0.190)		(0.257)				(0.175)
β^{NCI}_{Early}	0.723**		0.843**		0.312		0.121	
	(0.281)		(0.338)		(0.246)		(0.275)	
β^{NCI}_{Late}		0.474***		0.455***		0.311***		0.311***
		(0.126)		(0.109)		(0.071)		(0.086)
α	−0.375*	−0.155	−0.788***	−0.391**	−0.241	−0.249***	−0.414**	−0.251**
	(0.210)	(0.120)	(0.220)	(0.148)	(0.177)	(0.066)	(0.163)	(0.101)
R^2	0.544	0.684	0.602	0.663	0.040	0.264	0.097	0.231
N	40	40	40	40	40	40	40	40

that news coverage becomes more homogenous around major events, and thereby increasing the correlation among economic agents' actions (Nimark and Pitschner 2016).

Table 6 and Figure 12 in online Appendix A report similar results to those above, but uses the final vintage of GDP growth as a measure of the "truth." Two findings stand out relative to those above. First, the performance of the *NCI* model is on par with both the *SAM* and *NB* predictions for all the nowcasts, but deteriorates for the backcasts, perhaps indicating the forward-looking nature of the newspaper data. Second, as seen from Figure 12, if nowcasting evaluation had started just before the financial crisis in 2008, and not in 2004, the news-based model would have outperformed both benchmark models.

I report the cumulative difference in squared prediction errors between the other benchmark models and the *NCI* in Figure 13 in online Appendix A. Although the exact results are somewhat dependent on the release/vintage/model used for evaluation, the *NCI* generally outperforms the other factor-based benchmarks up until 2007, but then actually experience a fall in relative forecasting performance. Evaluated against news-based models without the time-varying parameters, in particular those without the LTM mechanism, this fall is however reversed during the years following the financial crisis. Thus, the figures confirm that the good performance of the *NCI* model relative to the other benchmark models is not driven solely by the onset of the financial crisis, but also that the time-varying parameter specification of the *NCI* likely makes the model forecasts more robust to structural breaks, like the financial crisis.

Finally, Figure 14 in online Appendix A reports how the daily coincident index evolves when estimation is done recursively. With the exception of some overshooting after the financial crisis, the recursive estimates of the *NCI* index, denoted NCI^*, track the full-sample-based estimates very well. The correlation between the two is roughly 0.9 and 0.8 over the full-sample and the out-of-sample evaluation period, respectively. When evaluating how well the NCI^* classifies the phases of the business cycle (over the full-sample), its performance is basically identical to earlier results, see Table 4 in Section 4.5.

To summarize the findings reported in this section, the news-based model produces nowcasts which on average are competitive with nowcasts produced by expert judgment (*NB*) and a state-of-the-art forecast combination framework (*SAM*). The good performance seems to be partly associated with the model's ability to capture economic turning points. Turning off the time-varying parameter specification, especially the LTM mechanism, or using other data than the news topics, generally results in worse forecasting performance.

Table 4. Receiver operating characteristics and area under the curve (AUROC) statistics. See Section 4.1 and Table 1 for further details

		Reference chronologies			
	Model	BB-GDP	MS-GDP	BB-ISD	MS-FMQ
Relative AUROC	NCI/NCI^*	1.012	0.980	1.012	0.999
	NCI/NCI^{noma}	1.067	1.045	1.097	1.050
	NCI/NCI^{7ma}	1.055	1.052	1.080	1.035
	NCI/NCI^{14ma}	1.059	1.060	1.081	1.047
	NCI/NCI^{month}	1.140	1.133	1.054	1.102
	NCI/CI^{month}	1.267	1.152	1.058	1.252

4.4 Could News Add Value for Policy Makers?

The time-varying relative performance documented in Figure 6 suggests that if the central bank could have exploited the news-based forecasts during the evaluation sample looked at here, forecast errors could have been smaller. To investigate this hypothesis more thoroughly, I follow Romer and Romer (2008), and run regressions based on the following equation:

$$\Delta\mathrm{GDP}^{r}_{t_q} = \alpha + \beta^{NB} NB_{t_q} + \beta^{SAM} SAM_{t_q} + \beta^{NCI} NCI_{t_q} + e_{t_q}. \tag{10}$$

Here, NB_{t_q}, SAM_{t_q}, and NCI_{t_q} are nowcasts for quarter t_q produced by *NB*, *SAM*, and the news-based model, respectively. The *SAM* and *NCI* predictions are those produced either at time $T-80$ (*Early*) or T (*Late*) in the quarter. The primary object of interest is to investigate if β^{NCI} is significantly different from zero: Conditional on the actual nowcasts produced by Norges Bank and SAM, could news-based predictions have added value?

The results reported in columns I to IV in Table 3 suggest that the answer to this question is yes. Irrespective of when in the current quarter the news-based predictions are produced (*Early* or *Late*), the β^{NCI} coefficient is positive and significant at either the 5% or 1% level. Using the first or fifth release of GDP growth as the dependent variable, does not alter this finding.

Could the news-based predictions also be informative for the statistical agency producing the GDP statistics? If GDP 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). With $R_{t_q} = \Delta\mathrm{GDP}^5_{t_q} - \Delta\mathrm{GDP}^1_{t_q}$, the following regression is estimated:

$$R_{t_q} = \alpha + \beta^{NCI} NCI_{t_q} + e_{t_q}. \tag{11}$$

Under the news view (not news topics), revisions must be mean zero and the coefficients in (11) 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^{NCI} 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.

The results reported in column VI of Table 3 indicate that the news-based predictions could be informative also for the statistical agency producing the GDP statistics. The β^{NCI} coefficient is significant when the *Late* version of the news-based nowcast is used. The last column in Table 3 shows that this conclusion remains robust when controlling for the *NB* and *SAM* predictions as well.

Table 7 in online Appendix A replicates the regressions from Table 3, but uses the final vintage of GDP growth as the dependent variable (and in computing R_{t_q}). Although the news-based predictions do not add significantly to those by *NB* in predicting the outcomes, there is still evidence suggesting that news have predictive power for future revisions, or, in other words, reduces noise.

4.5 Alternative Data Transformations and Aggregation

When constructing the news topic time series, as described in Section 2.2, a 60-day (backward looking) moving average filter is used to remove high-frequency noise from the series. I have also estimated versions of the model where no filter and shorter 7- and 14-day moving average filters are used to smooth the topic series prior to estimation. The alternative indexes are labeled NCI^{noma}, NCI^{7ma}, and NCI^{14ma}. The two latter are highly correlated with each other and have a correlation coefficient of 0.88 with the original *NCI*, while the correlation between the NCI^{noma} and the *NCI* is 0.82. However, as seen from Figure 15 in online Appendix A, the alternative index estimates are (very) noisy. In terms of classifying the phases of the business cycle and nowcasting, this also results in a slightly worse performance, see Tables 4 and 5, respectively. (Since the alternative

Table 5. News-based information flow and RMSFE. All numbers are reported as relative RMSFE statistics. See Table 2 for further details

Type	Nowcast for T_q					Backcast for T_q	
Quarterly info.	T_q-2			T_q-1			
Daily info.	$T-80$	$T-60$	$T-40$	$T-20$	T	$T+10$	$T+30$
Panel A: First release							
NCI^{7ma}/NCI	1.218*	1.100	1.016	1.124	1.166*	1.178*	1.179*
NCI^{month}/NCI	0.873	0.839**	0.769***	0.838**	0.900	0.894	0.898
CI^{month}/NCI	1.125	1.080	1.087	1.184	1.164	1.157	1.165
Panel B: Fifth release							
NCI^{7ma}/NCI	1.125	1.128*	1.075	1.156	1.158**	1.166***	1.167***
NCI^{month}/NCI	0.846	0.844**	0.822**	0.865	0.903*	0.895	0.893*
CI^{month}/NCI	1.005	1.002	1.019	1.073	1.091	1.082	1.088
Panel C: Final vintage							
NCI^{7ma}/NCI	1.052	1.039	1.016	1.094	1.038	1.048	1.051
NCI^{month}/NCI	0.894	0.895*	0.869**	0.929	0.935**	0.849	0.915**
CI^{month}/NCI	0.952	0.953	0.929	0.993	0.945	0.858	0.925

NCI^{7ma} model is outperforming the other alternatives in-sample, and to reduce the computational burden, the nowcasting experiment is only conducted for the NCI^{7ma} model.) Thus, although the choice of using a 60-day, or a 7- and 14-day, moving average filter is somewhat arbitrary, these results illustrate that some type of smoothing of the raw series is beneficial when working with data of this type.

An alternative "noise-removing" strategy is to work with more aggregated data. To explore such effects I compute the monthly mean of the daily topic time series (without smoothing), and estimate an NCI^{month} version of the model using months as the highest observed frequency interval. While potentially removing noise from the input series prior to estimation, this approach also makes comparisons between the news-based model and alternatives using conventional monthly economic data more informative because both model types can be specified with the LTM mechanism (for all the high-frequency variables), confer the discussion related to Equations (5) and (6) in Section 3. Therefore, I also estimate a version of the original *CI* model, but now using months as the highest observed frequency interval. I denote this model CI^{month}.

The monthly coincident index estimates are plotted in Figure 16 in online Appendix A, while their relative AUROC scores are reported in Table 4. The NCI^{month} version of the model generally performs worse than the original *NCI* model. Thus, some information is lost when aggregating the news-data from a daily to monthly frequency. On the other hand, the performance of the CI^{month} model is up to 26% and 15% worse than the *NCI* and NCI^{month} models, respectively, highlighting again the potential usefulness of the news-based approach. Qualitatively, similar results are obtained when evaluating the NCI^{month} and CI^{month} models out-of-sample, see Table 5. Irrespective of informational assumptions and which release/vintage that is used for evaluation, the NCI^{month} model outperforms the alternative CI^{month} model. Interestingly, however, the nowcasting performance of the NCI^{month} model is actually (significantly) better than the *NCI* model. In the application considered here, this creates a trade-off between obtaining the best in-sample business cycle classification properties and out-of-sample forecasting performance.

5. CONCLUSION

In this article, I show how unstructured textual data collected from a major Norwegian business newspaper can be used to construct a daily coincident index of the business cycle within a mixed-frequency time-varying dynamic factor model (DFM) framework.

The resulting index is demonstrated to have almost perfect classification properties of the business cycle phases, and it gives the index user broad-based information about the type of news that contribute to the index fluctuations. In an out-of-sample nowcasting experiment for quarterly GDP growth, I show that the model performs substantially better than forecasts from simple time series models, and that it is competitive with a state-of-the-art forecast combination system and official Norges Bank nowcasts. Interestingly, if the statistical agency producing the output growth statistics itself had used the news-based methodology, I show that it would have resulted in a less noisy revision process. Thus, news reduces noise.

The gains in classification and predictive performance are due to the novel usage of newspaper data together with the time-varying parameter specification of the DFM. Prior to estimation, the textual data is decomposed into daily news topics using a latent Dirichlet allocation (LDA) model (Blei, Ng, and Jordan 2003). The derived topics are easy to interpret, and subsequently included in a mixed-frequency DFM. In contrast to earlier approaches in the business cycle and nowcasting literature, however, I allow for dynamic sparsity patterns in the time-varying factor loadings using a latent threshold mechanism (Nakajima and West 2013).

It is interesting to note that when using the same news topics as here, Larsen and Thorsrud (in press) and Larsen and Thorsrud (2017) found that (unexpected) news innovations are associated with persistent quarterly productivity increases and predictability of daily stock returns. Decomposing news published through a business newspaper into news topics thereby puts unstructured textual data into a format that seems highly informative for both macroeconomic developments and asset prices.

The economic literature using text as data and other alternative (Big) data sources is fast growing, but still in its early stages. An advantage of the approach taken here is that it results in easily interpretable output. Another advantage is that long historical archives of newspaper data typically exist, while long time series of high-frequency information from other sources, for example, social media or Internet search volume, are difficult to obtain. Likewise, most countries have daily (business) newspapers, but very few countries have an abundance of high-frequency economic indicators. Natural extensions to the approach taken here include: expanding the scope of the analysis to other countries than Norway; comparing the topic model approach to other Natural Language Processing techniques; and allowing for joint estimation of the news topics and the business cycle index.

SUPPLEMENTARY MATERIALS

The online supplementary materials contain Appendices A to K. Replication codes are available at *https://github.com/leifandersthorsrud/NCI*.

ACKNOWLEDGMENTS

This article does not represent the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. The author thanks Knut A. Aastveit, Hilde C. Bjørnland, Fabio Canova, Pia Glæserud, Jonas Moss, Simon von Norden, Ivan Petrella, Juan F. Rubio-Ramírez, Maximilian Rohrer, Christian Schumacher, and participants at the Joint Research Workshop of Norges Bank and Deutsche Bundesbank and the CAMP Workshop on Commodities, business cycles and monetary policy, for valuable comments. Constructive comments and suggestions from the editors and two anonymous referees have improved the quality of this article substantially. Vegard Larsen provided helpful technical assistance, and Anne Sofie Jore was helpful in collecting data, for which author is 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. This article is an updated version of the earlier working papers: "Words are the new numbers: A newsy coincident index of business cycles," and "Nowcasting using news topics. Big Data versus big bank."

[Received November 2017. Revised May 2018.]

REFERENCES

Aastveit, K. A., Gerdrup, K. R., Jore, A. S., and Thorsrud, L. A. (2014), "Nowcasting GDP in Real Time: A Density Combination Approach," *Journal of Business & Economic Statistics*, 32, 48–68. [403]

Aastveit, K. A., Jore, A. S., and Ravazzolo, F. (2016), "Identification and Real-Time Forecasting of Norwegian Business Cycles," *International Journal of Forecasting*, 32, 283–292. [400,401]

Alessi, L., Ghysels, E., Onorante, L., Peach, R., and Potter, S. (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, 483–500. [393]

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009), "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27, 417–427. [393,397,401]

Bai, J., and Wang, P. (2014), "Identification Theory for High Dimensional Static and Dynamic Factor Models," *Journal of Econometrics*, 178, 794–804. [399]

——— (2015), "Identification and Bayesian Estimation of Dynamic Factor Models," *Journal of Business & Economic Statistics*, 33, 221–240. [399]

Balke, N. S., Fulmer, M., and Zhang, R. (2017), "Incorporating the Beige Book into a Quantitative Index of Economic Activity," *Journal of Forecasting*, 36, 497–514. [394]

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013), "Nowcasting and the Real-Time Data Flow," Working Paper Series 1564, European Central Bank. [398,401]

Banbura, M., Giannone, D., and Reichlin, L. (2011), *Nowcasting, The Oxford Handbook of Economic Forecasting (Oxford Handbooks in Economics)*, New York: Oxford University Press. [393]

Bholat, D., Hansen, S., Santos, P., and Schonhardt-Bailey, C. (2015), "Text Mining for Central Banks: Handbook," *Centre for Central Banking Studies*, 33, 1–19. [394]

Bjørnland, H. C., and Thorsrud, L. A. (in press), "Commodity Prices and Fiscal Policy Design: Procyclical Despite a Rule." Journal of Applied Econometrics. [399]

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022. [394,407]

Breitung, J., and Schumacher, C. (2008), "Real-Time Forecasting of German GDP Based on a Large Factor Model with Monthly and Quarterly Data," *International Journal of Forecasting*, 24, 386–398. [394]

Burns, A. F., and Mitchell, W. C. (1946), "Measuring Business Cycles," No. 46-1 in NBER Books, National Bureau of Economic Research, Inc. [394,400]

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., and Blei, D. M. (2009), "Reading Tea Leaves: How Humans Interpret Topic Models," in *Advances in Neural Information Processing Systems*, eds. Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Cambridge, MA: The MIT Press, pp. 288–296. [394]

Chen, R., and Liu, J. S. (2000), "Mixture Kalman filters," *Journal of the Royal Statistical Society*, Series B, 62, 493–508. [403]

Choi, H., and Varian, H. (2012), "Predicting the Present with Google Trends," *Economic Record*, 88, 2–9. [394]

Croushore, D. (2006), "Forecasting with Real-Time Macroeconomic Data," in *Handbook of Economic Forecasting* (Vol. 1), eds. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, pp. 961–82. [403]

Del Negro, M., and Otrok, C. (2008), "Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles," Staff Reports 326, Federal Reserve Bank of New York. [399]

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263. [404]

Doucet, A., de Freitas, N., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag. [403]

Ellis, C., Mumtaz, H., and Zabczyk, P. (2014), "What Lies Beneath? A Time-Varying FAVAR Model for the UK Transmission Mechanism," *Economic Journal*, 124, 668–699. [399]

Evans, M. D. D. (2005), "Where Are We Now? Real-Time Estimates of the Macroeconomy," *International Journal of Central Banking*, 1, 127–175. [393]

Foroni, C., and Marcellino, M. (2013), "A Survey of Econometric Methods for Mixed-Frequency Data," Working Paper 2013/06, Norges Bank. [398]

Gentzkow, M., Kelly, B. T., and Taddy, M. (2017), "Text as Data," Working Paper 23276, National Bureau of Economic Research. [394]

Giannone, D., Reichlin, L., and Small, D. (2008), "Nowcasting: The Real-Time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55, 665–676. [394,402]

Griffiths, T. L., and Steyvers, M. (2004), "Finding Scientific Topics," *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5228–5235. [395]

Hansen, S., McMahon, M., Prat, A. (2018), "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach," *The Quarterly Journal of Economics*, 133, 801–870. [394,403]

Harvey, A. C. (1990), *Forecasting, Structural Time Series Models and the Kalman Filter* (*Cambridge Books*), Cambridge: Cambridge University Press. [398]

Heinrich, G. (2009), "Parameter Estimation for Text Analysis," Technical Report, Fraunhofer IGD. [395,403]

Kuzin, V., Marcellino, M., and Schumacher, C. (2011), "MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 27, 529–542. [394]

Larsen, V. H., and Thorsrud, L. A. (2017), "Asset Returns, News Topics, and Media Effects," Working Paper 2017/17, Norges Bank. [393,407]

——— (in press), "The Value of News for Economic Developments," *Journal of Econometrics*. [395,407]

Lopes, H. F., and Carvalho, C. M. (2007), "Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes," *Journal of Statistical Planning and Inference*, 137, 3082–3091. [399]

Mankiw, N. G., and Shapiro, M. D. (1986), "News or Noise? An Analysis of GNP Revisions," NBER Working Papers 1939, National Bureau of Economic Research, Inc. [406]

Marcellino, M., Porqueddu, M., and Venditti, F. (2016), "Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility," *Journal of Business & Economic Statistics*, 34, 118–127. [393,394,398]

Mariano, R. S., and Murasawa, Y. (2003), "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18, 427–443. [393,397]

Mincer J. A., and Zarnowitz, V. (1969), "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, ed. J. Mincer, New York: National Bureau of Economic Research, pp. 3–46. [406]

Nakajima, J., and West, M. (2013), "Bayesian Analysis of Latent Threshold Dynamic Models," *Journal of Business & Economic Statistics*, 31, 151–164. [394,398,399,407]

Nimark, K. P., and Pitschner, S. (2016), "Delegated Information Choice," CEPR Discussion Papers 11323, C.E.P.R. [405]

Nymand-Andersen, P. (2016), "Big Data: The Hunt for Timely Insights and Decision Certainty," IFC Working Paper 14, Bank for International Settlements. [393]

Pang, B., Lee, L., and Vaithyanathan, S. (2002), "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, EMNLP '02* (Vol. 10), Association for Computational Linguistics, pp. 79–86. [397]

Peress, J. (2014), "The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes," *The Journal of Finance*, 69, 2007–2043. [393]

Pitt, M. K., and Shephard, N. (1999), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599. [403]

Romer, C. D., and Romer, D. H. (2008), "The FOMC versus the Staff: Where Can Monetary Policymakers Add Value?" *American Economic Review*, 98, 230–235. [406]

Scott, S. L., and Varian, H. R. (2013), "Bayesian Variable Selection for Nowcasting Economic Time Series," NBER Working Papers 19567, National Bureau of Economic Research, Inc. [399]

Sims, C. A. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690 (Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information). [393]

Stock, J. H., and Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual*, eds. O. J. Blanchard and S. Fischer, Cambridge, MA: MIT Press, pp. 351–394. [393]

——— (2002), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162. [394]

——— (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41, 788–829. [393,394]

——— (2016), "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics* (Vol. 2), eds. J. B. Taylor and H. Uhlig. Amsterdam: Elsevier, North-Holland, pp. 415–525. [397]

Tetlock, P. C. (2007), "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *The Journal of Finance*, 62, 1139–1168. [393,396]

Timmermann, A. (2006), "Forecast Combinations," in *Handbook of Economic Forecasting* (Vol. 1), eds. G. Elliott, C. W. J. Granger, and A. Timmermann, Amsterdam: Elsevier, North-Holland, pp. 135–196. [403]

Travis, J. B., and Jordà, Ò. (2011), "Evaluating the Classification of Economic Activity into Recessions and Expansions," *American Economic Journal: Macroeconomics*, 3, 246–277. [400,401]

Zhou, X., Nakajima, J., and West, M. (2014), "Bayesian Forecasting and Portfolio Decisions Using Dynamic Dependent Sparse Factor Models," *International Journal of Forecasting*, 30, 963–980. [394,399]