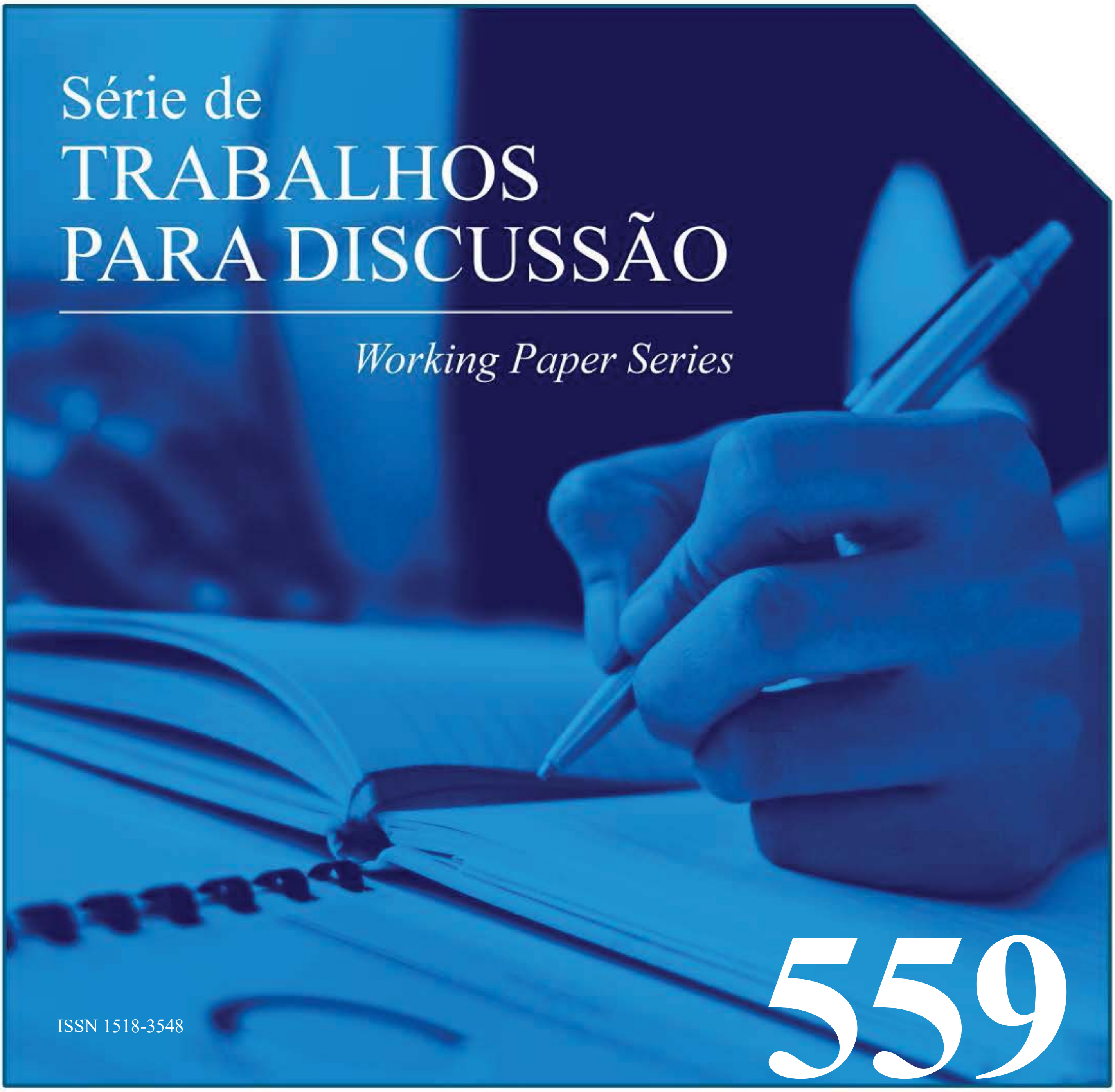

November 2021

Forecasting with VAR-teXt and DFM-teXt Models:exploring the predictive power of central bank communication

Leonardo N. Ferreira

BANCO CENTRAL DO BRASIL

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	no. 559	Novembro	2021	p. 1-41

Working Paper Series

Edited by the Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Rodrigo Barbone Gonzalez

Co-editor: José Valentim Machado Vicente

Head of the Research Department: André Minella

Deputy Governor for Economic Policy: Fabio Kanczuk

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: http//www.bcb.gov.br/?CONTACTUS

Non-technical Summary

Text has long been ubiquitous in many fields, and economics is no exception. Newspapers, press releases, statements, speeches and several other formats are used to report events that may affect the economy or even to share views on economic conditions. Recently and conveniently, the progress in machine learning has given rise to approaches that can process and summarise these large amounts of strings in a timely and tractable manner, making them amenable to the incorporation in conventional economic models.

At the same time, central bank releases have become an increasingly relevant aspect of monetary policy. In this regard, even Bernanke acknowledged: "When I was at the Federal Reserve, I occasionally observed that monetary policy is 98% talk and only 2% action". In fact, central bank releases telegraph not only likely future actions but also some central bank private information about the future state of the economy. Therefore, they can provide information beyond that contained in traditional economic models, being potentially important for forecasting.

The central objective of this paper is to investigate this potential by assessing the contribution of information extracted from monetary policy statements to the forecasting of macroeconomic and financial variables. In order to do that, text mining and conventional econometric techniques are combined in text-augmented models: the text-augmented vector autoregressive model (VAR-teXt) and the text-augmented dynamic factor model (DFM-teXt). This idea was inspired by traditional factor-augmented models, where factors help exploit additional data sources and uncover hidden patterns even without necessarily being identified as specific economic concepts.

Therefore, the contribution of this paper is twofold. The novelty in terms of methodology is to explore the complementarity between traditional econometrics and machine learning to estimate VAR-teXt and DFM-teXt models. The economic contribution is to assess whether, using the proposed models and algorithms, US central bank communication can improve forecasts. In fact, many important variables, such as inflation and interest rates, are better predicted when textual factors based on monetary policy statements are added to the model, showing it is important to use all available information regardless of its form.

Sumário Não Técnico

Textos estão presentes em muitos campos e há muito tempo, e isso não é diferente em economia. Notícias, notas à imprensa, comunicados, discursos e vários outros formatos são usados para relatar eventos que podem afetar a economia ou mesmo para compartilhar visões sobre condições econômicas. Recente e convenientemente, progressos em *machine learning* possibilitaram o processamento e a sintetização dessas grandes quantidades de caracteres de maneira rápida e relativamente fácil, o que facilita sua incorporação em modelos econômicos convencionais.

Ao mesmo tempo, a comunicação de bancos centrais tem se tornado um aspecto cada vez mais relevante da política monetária. A este respeito, até mesmo Bernanke reconheceu: "Quando eu estava no Federal Reserve, ocasionalmente observei que a política monetária é 98% discurso e apenas 2% ação". De fato, documentos produzidos por banco centrais telegrafam não apenas prováveis ações futuras, mas também sua visão sobre o futuro estado da economia. Portanto, tais comunicados podem fornecer informações além das contidas nos modelos econômicos tradicionais, sendo potencialmente importantes para previsões econômicas.

O objetivo central deste artigo é investigar esse potencial, avaliando a contribuição de informações extraídas de comunicados de política monetária para a previsão de variáveis macroeconômicas e financeiras. Para isso, *text mining* e técnicas econométricas convencionais são combinadas em modelos aumentados com texto: modelos de vetores autorregressivos aumentados com texto (VAR-teXt) e modelos de fatores dinâmicos aumentados com texto (DFM-teXt). Essa ideia foi inspirada em modelos tradicionais aumentados com fatores, em que os fatores ajudam a explorar fontes de dados adicionais e a descobrir padrões, que não precisam ser necessariamente identificados como conceitos econômicos específicos.

Desse modo, as contribuições deste artigo são as seguintes. A novidade em termos de metodologia é explorar a complementaridade entre a econometria tradicional e *machine learning* para estimar os modelos VAR-teXt e DFM-teXt. A contribuição econômica é avaliar se, usando os modelos e algoritmos propostos, a comunicação do banco central dos Estados Unidos contribui para a melhoria das previsões. Os resultados mostram que muitas variáveis relevantes, como inflação e taxas de juros, são melhor previstas quando fatores textuais baseados em comunicados de política monetária são adicionados ao modelo. Isso mostra quão importante é quantificar e usar todas as informações disponíveis, independentemente de sua forma, na análise econômica.

Forecasting with VAR-teXt and DFM-teXt models: exploring the predictive power of central bank communication

Leonardo N. Ferreira*†

Abstract

This paper explores the complementarity between traditional econometrics and machine learning and applies the resulting model – the VAR-teXt – to central bank communication. The VAR-teXt is a vector autoregressive (VAR) model augmented with information retrieved from text, turned into quantitative data via a Latent Dirichlet Allocation (LDA) model, whereby the number of topics (or textual factors) is chosen based on their predictive performance. A Markov chain Monte Carlo (MCMC) sampling algorithm for the estimation of the VAR-teXt that takes into account the fact that the textual factors are estimates is also provided. The approach is then extended to dynamic factor models (DFM) generating the DFM-teXt. Results show that textual factors based on Federal Open Market Committee (FOMC) statements are indeed useful for forecasting.

Keywords: Text Mining, MCMC Sampling, Density forecast, FOMC statements, Interest rate

JEL Classification: C8, C11, C53, D84, E50

---

*Research Department, Central Bank of Brazil and Queen Mary University of London, e-mail: leonardo.ferreira@bcb.gov.br.

†I am greatly indebted to Haroon Mumtaz for his supervision and valuable comments. I am very grateful to Andrea Carriero for insightful discussions. I also would like to thank Antonio David, Marcelo Medeiros, Nora Lamersdorf, Saulo Benchimol, Vinicius Brandi, and an anonymous referee for their suggestions as well as the participants at presentations at the Econometric Society European Meeting 2021, the 2021 Latin American Meeting of the Econometric Society, the European Seminar on Bayesian Econometrics 2021, the 2021 Africa Meeting of the Econometric Society, 25th Spring Meeting of the European Association of Young Economists, the First Annual Southern PhD Economics Conference, the 21th Brazilian Finance Meeting, Universidade Católica de Brasília, and the 5th Workshop of the Research Network of the Banco Central do Brasil for useful comments. All errors are mine.

1 Introduction

Central bank communication has become an increasingly relevant aspect of monetary policy. The days of "never explain, never excuse", reputedly Sir Montagu Norman's motto[1], Governor of the Bank of England from 1920 to 1944, were replaced with days of regular communication that takes place via a wide array of formats: statements, minutes, implementation notes, press conferences, and so on.

During crisis times, the role of communication in shaping expectations becomes even more relevant since the short-term policy rate is usually constrained by the zero lower bound (Blot and Hubert, 2018). In these times, as claimed by Gros (2018), central bank communication becomes the policy. In fact, central bank releases not only telegraph likely future actions but also some private information about the future state of the economy (Campbell et al., 2012). Hence, they can provide information beyond that contained in traditional economic models, being potentially important for forecasting.

This paper investigates this potential by assessing the contribution of information extracted from Federal Open Market Committee (FOMC) statements to the forecasting of macroeconomic and financial variables. In order to do that, text mining and conventional econometric techniques are combined in a VAR-teXt: a vector autoregressive (VAR) model augmented with exogenous variables that capture information retrieved from text.

Taking advantage of the common estimation approach – Gibbs sampling – the transformation of text into quantitative data is done through a Latent Dirichlet Allocation (LDA) model. LDA is a probabilistic topic model developed by Blei et al. (2003) that, essentially, groups words into topics and describes documents in terms of them. As such, it helps reduce documents to low-dimensional space.

Nevertheless, departing from the common approach in economic applications, which involves selecting the number of topics based on their interpretability, and

[1] See Bernanke (2007).

consistently with the aim of this paper, the number of topics (or textual factors) is chosen based on the predictive performance of the VAR-teXt. While some topics will be meaningful, others will not. Still, the full set of time series of statement-specific topic distributions may contain variation that is worth exploring. That is why they will generally be treated as textual factors. This paper then evaluates whether such textual factors are forecast improving by comparing the performance of the VAR-teXt with that of the benchmark VAR.

This agnostic approach is inspired by factor-augmented models, where factors help exploit additional data sources and uncover hidden patterns even without necessarily being identified as specific economic concepts.[2] In fact, both being examples of unsupervised learning tasks, topic modelling and factor analysis share many features. A complementary, and probably more familiar to economists, definition of LDA, attributed to Hansen et al. (2018), highlights this connection: LDA is a Bayesian factor model for discrete data with factors representing topics.

This paper also extends previous literature by proposing a Markov chain Monte Carlo (MCMC) sampling algorithm for the estimation of VAR-teXt models that takes into account the fact that textual factors themselves are estimates. By exploring the entire distribution over textual factors, the proposed MCMC sampling algorithm properly captures the estimation uncertainty, leading to more accurate posterior predictive distributions.

Finally, a natural extension of the VAR-teXt is evaluated. With the purpose of using more conventional data in the analysis, the approach is extended to dynamic factor models (DFM). This allows for use of all the information that is easily accessible regardless of its form and gives rise to the DFM-teXt, which is simply a VAR-teXt in the (data) factors.

In summary, the contribution of this paper is twofold. In terms of methodology, the novelty is to explore the complementarity between traditional econometrics

---

[2] For instance, in the context of sufficient information in structural VARs, Forni and Gambetti (2014) suggest keeping adding principal components to the VAR until the model is no longer deficient.

and machine learning to estimate more accurately VAR-teXt and DFM-teXt models. The economic contribution is to show that, using the proposed models and algorithms, central bank communication can improve forecasts.[3]

Specifically, results show the VAR-teXt consistently outperforms the benchmark VAR in forecasting the consumer price inflation and the interest rate as measured by log-scores, and such results are robust to several alternative specifications. The DFM-teXt also performs better than the DFM, even for industrial production, as well as many other variables, but only for the 3-month ahead forecasts.

Related Literature

Several papers have explored the link between text and the economy. The closest ones can be divided into two groups. The first one explores central bank communication. Hayo and Neuenkirch (2010), for instance, interpret and sort central bank communications into three categories depending on whether they indicate likely increases (+1), decreases (-1), or no change (0) in the fed funds rate. They then use this indicator in an ordered probit model estimated to predict changes in monetary policy and show it performs very well in their out-of-sample assessment.

While such narrative approaches may capture some nuances, they are prone to subjectivity. Automated approaches, on the other hand, have two meaningful advantages: scalability to larger corpora and reduction of biases that may unfold when readers overlook patterns that do not conform to prior beliefs (Bholat et al., 2015).

Lucca and Trebbi (2009) address this concern by using an automated approach to extract information from FOMC statements in order to forecast macroeconomic variables with univariate and VAR models. Their objective is rather similar to

[3]Even though this may seem expected, there is also evidence such as Lustenberger and Rossi (2020) showing more communication may increase forecast errors.

this paper's, but they differ in that i) they construct semantic orientation scores; and ii) they focus on in-sample predictive power. Results show that changes in communication as measured by the scores help predict future policy rate.

Hansen et al. (2019) use LDA and dictionaries to measure a set of high-dimensional signals based on the Bank of England's Inflation Report. Using these signals in an elastic net regression, they find that, beyond the conventional expectations channel, signals about the expected uncertainty can have important effects along the yield curve, especially in the long run. Their analysis, however, is also in sample.

The second group gathers papers that are closer to forecasting evaluations used in practice. Models are re-estimated recursively and out-of-sample predictions are assessed with log-scores or root mean square errors (RMSE). Nonetheless, they use text from newspapers instead of from central banks. Thorsrud (2018) builds a daily coincident index of the business cycle based on textual data retrieved from a major Norwegian business newspaper. Specifically, he uses tone-adjusted topic time series in a mixed-frequency time-varying factor model and shows the resulting model performs substantially better than simple time series models.

As it is becoming standard in this new literature, topic proportions before tone adjustment are given by the average over LDA draws. In addition to this and the application, Thorsrud (2018) differs from this paper in that topics are added in the panel of variables of the observation equation and not as exogenous variables. This can create an inconsistency between the data generating process for the topics, which rules out time-series dependencies across documents, and their dynamics in the regression.

Larsen and Thorsrud (2019) use LDA on the same newspaper and augment autoregressive (AR) models with tone-adjusted topic time series (AR-X). Unlike Thorsrud (2018), however, they include these series as exogenous variables in the regression. They then compare the predictive power of the AR-X with the benchmark AR and identify which topics improve the forecasts of key economic variables. In particular, they find some topics have predictive power for asset prices, which are forward looking and supposed to reflect all publicly available information.

Kalamara et al. (2020) retrieve information from UK newspapers using several methods and incorporate the resulting measures in AR and factor models. They choose to exclude topic models to avoid identification issues that may arise when the model is re-estimated recursively. The forecasts of most macroeconomic variables are improved when news text is included in the AR model, but not in the factor model. As the papers cited above, however, they do not address the fact that the outputs of the machine learning techniques are estimates.

This paper complements this literature by evaluating the contribution of textual factors based on FOMC statements to the prediction of macroeconomic and financial variables in a recursive exercise, useful for policy makers and forecasters, while taking into account that such factors are estimates, and thus providing more accurate posterior predictive distributions. The rest of the paper is organised as follows. Section 2 describes the data. Section 3 presents the VAR-teXt model and the MCMC sampling algorithm. Section 4 evaluates the performance of this model. Section 5 presents some robustness exercises. Section 6 introduces the DFM-teXt model and Section 7 concludes.

2 Data

The traditional dataset consists of 3 macroeconomic and 1 financial variables from 1998M09, when the release of statements after scheduled meetings became more common, to 2020M02. Industrial production growth and consumer price inflation are calculated by taking the first difference of the logarithm of the corresponding indices downloaded from FRED-MD, a large macroeconomic database with 128 time series described in McCracken and Ng (2016). The shadow rate proposed by Wu and Xia (2016) is used as a measure of the stance of the Fed since the zero lower bound affected a considerable part of the sample period. The shadow rate is convenient in that it also summarises unconventional monetary policy. Finally, to capture information about future economic activity and improve the informational content of the regression, the excess bond premium (EBP), introduced by Gilchrist and Zakrajšek (2012), is also incorporated into the vector of variables. The EBP is a corporate bond credit spread purged from the default risk, a useful leading indicator.

2.1 The Corpus of FOMC Statements and the LDA

The text-augmented model also includes information retrieved from the corpus of the 169 FOMC statements that followed scheduled meetings during the period of analysis. The structure of the statements that follow scheduled meetings is fairly comparable over time, making them a fitting choice for the text analysis. The corpus was scraped from the Federal Reserve website and pre-processed before estimation. This involves conversion to lower case, removal of punctuation, white spaces, numbers and stopwords, and stemming. Stopwords are words that are not very informative, such as "a", "and" and "the". Stemming means reducing words to common linguistic roots. For instance, "increasing", "increased" and "increases" become "increas".

After this pre-processing, the corpus of 169 statements has 29,903 stems, of which 858 are unique. Nonetheless, in order to make the analysis more granular, the first step of the estimation is conducted at the level of the paragraph, meaning that there are actually 693 documents. LDA can then be used. As explained by Blei et al. (2003), LDA reduces any document to a fixed set of real-valued features on a low-dimensional latent space: the posterior Dirichlet parameters. Given such parameters, it is possible to obtain the statement-specific mixing probabilities $\tilde{\theta}_d$: the textual factors.

It is worth mentioning there are alternatives to the LDA, such as the Dynamic

Topic Model (DTM), which introduces time-series dependencies into the data generating process, and the Structural Topic Model (STM), which introduces covariates into a topic model (Blei and Lafferty, 2006; Roberts et al., 2016). Nonetheless, non-conjugacy makes sampling methods more difficult for such models, and the algorithms usually depart from Gibbs sampling, which is more familiar to economists. Therefore, the focus here will be on handling the generated regressors issue while leaving a joint solution for both problems for future research.

LDA is estimated with a collapsed Gibbs sampling algorithm. First, the estimation of word's topic assignment consists of the following steps:[4]

Step 1. Randomly allocate to each token in the corpus a topic assignment drawn uniformly from $\{1, ..., K\}$, where K denotes the number of textual factors.

Step 2. For each token, sequentially draw a new topic assignment via multinomial sampling where

$$P[q_{d,n} = k | Q^{-}, W, \alpha, \eta] \propto \frac{m^{-}_{k,v} + \eta}{\sum_v m^{-}_{k,v} + V\eta} (n^{-}_{d,k} + \alpha) \tag{1}$$

where the '-' superscript denotes counts excluding (d, n) term, with d representing documents and n their terms. $q_{d,n} = k$ denotes the topic assignment of (d, n) term and $Q = (\mathbf{q}_1, ..., \mathbf{q}_D)$ the other topic assignments. $W = (\mathbf{w}_1, ..., \mathbf{w}_D)$ is the observed data. α is the prior on the document-specific mixing probabilities and η on the topic-specific term probabilities. $m_{k,v} \equiv \sum_n \sum_d \mathbb{1}(q_{d,n} = k)\mathbb{1}(w_{d,n} = v)$ is the number of times topic k allocation variables generate term v, and $n_{d,k} \equiv \sum_n \mathbb{1}(q_{d,n} = k)$ is the number of words in document d that have topic allocation k. V is the number of unique terms.

[4]This subsection brings forward part of the estimation so as to show how documents and words are summarised into time series. This follows very closely the Online Appendix of Hansen et al. (2018). For more details see also Murphy (2012).

Step 3. Repeat Steps 1 and 2 until the required number of draws has been reached.

To fit the LDA, the number of textual factors K and hyperparameters α and η need to be fixed a priori. In applications in economics, K is usually chosen based on interpretability. In this application, however, a grid for K will be explored and the best K will be selected based on the out-of-sample forecasting performance of the VAR-teXt. This helps discipline the choice.[5] Following Griffiths and Steyvers (2004) and Hansen et al. (2018), η is set to $200/V$ and α to $50/K$.

Next, a process called querying is carried out. Querying allows the document distributions aggregated at the level of the statement to be recovered. It corresponds to running the Gibbs sampling keeping the topic-specific term probabilities φ fixed at their estimated values. This is done by collapsing paragraphs back into the statement level and sequentially sampling from:

$$P[\tilde{q}_{d,n} = k | \tilde{Q}^{-}, \tilde{W}, \alpha, \eta] \propto \hat{\varphi}_{v_{d,n}} (n^{-}_{\tilde{d},k} + \alpha) \tag{2}$$

where tilde denotes the new document level. Nonetheless, this gives estimates of each word's topic assignment since the topic proportions θ were integrated out in the derivation of the collapsed Gibbs sampling. In order to recover them, the output of interest, the final step is to compute the statement predictive distributions using:

$$\tilde{\hat{\theta}}^{k}_{\tilde{d}} = \frac{n_{\tilde{d},k} + \alpha}{\sum_{k=1}^{K} n_{\tilde{d},k} + \alpha} \tag{3}$$

Such textual factors are incorporated in the VAR as described in the next section.[6]

[5]There are alternative ways to make the selection of the number of topics more objective and disciplined such as the topic coherence proposed by Newman et al. (2010), which uses a co-occurrence measure based on pointwise mutual information over Wikipedia. Nonetheless, selecting the number of topics based on the predictive ability of the VAR-teXt is more consistent with the objective of this paper.

[6]The textual factors are transformed into first differences and standardised as in Larsen and Thorsrud (2019).

3 The VAR-teXt Model

The point of departure for the analysis is the benchmark VAR model of the form:

$$Y_t = \sum_{p=1}^{P} \beta_p Y_{t-p} + \mu + A_0 \varepsilon_t \tag{4}$$

where Y_t is the $N \times 1$ vector of standard macro variables, p denotes the lags, with $p = 1, ..., P$, and A_0 is a decomposition of the covariance matrix Σ such that $Var(u_t) = A_0 A_0' = \Sigma$.

The VAR-teXt, text-augmented VAR, is given by:

$$Y_t = \sum_{p=1}^{P} \beta_p Y_{t-p} + \phi X_{t-1} + \mu + A_0 \varepsilon_t \tag{5}$$

where X_{t-1} are the first $K - 1$ textual factors.[7]

Because traditional data is not taken into account in the estimation of the textual factors, the feedback from lagged endogenous variables to the textual factors is restricted and they are incorporated into the model as exogenous variables. Hence, the VAR-teXt model is consistent with the data generating process for the textual factors. Moreover, as in static factor models and approximate static factor models, LDA does not account for the time-series dependence in estimating the textual factors. That is why the textual factors are treated differently and appear in the equation only once. As aforementioned, these issues could be tackled with the use of alternative approaches. This, however, would impair the use of a simple MCMC sampling algorithm.

[7] $K - 1$ because, being proportions, textual factors add up to 1 and the model has an intercept.

3.1 Estimation

Following Bańbura et al. (2010), Equation (5) can be written in a more compact way, which is more convenient for the estimation:

$$Y = Z\beta + \varepsilon A_0 \tag{6}$$

where $Y = (Y_1, ..., Y_T)'$, $Z = (Z_1, ..., Z_T)'$ with $Z_t = (Y'_{t-1}, ..., Y'_{t-P}, X'_{t-1}, 1)'$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$, and $\beta = (\beta_1, ..., \beta_P, \phi, \mu)'$ is the $(NP + K) \times N$ matrix containing all the coefficients.

A prior for the VAR parameters is then introduced by augmenting the vector of variables with the following dummy observations:

$$Y_{d,1} = \begin{pmatrix} \frac{diag(\delta_1\sigma_1,...,\delta_N\sigma_N)}{\tau} \\ 0_{N(P-1)\times N} \\ \cdots\cdots\cdots \\ diag(\sigma_1, ..., \sigma_N) \\ \cdots\cdots\cdots \\ 0_{1\times K} \end{pmatrix}, \quad Z_{d,1} = \begin{pmatrix} \frac{J_p \otimes diag(\sigma_1,...,\sigma_N)}{\tau} & 0_{NP\times K} \\ \multicolumn{2}{c}{\cdots\cdots\cdots} \\ 0_{N\times NP} & 0_{N\times K} \\ \multicolumn{2}{c}{\cdots\cdots\cdots} \\ 0_{K\times NP} & diag(1_{K\times 1}\, c) \end{pmatrix} \tag{7}$$

where δ_1 to δ_N denote the prior mean for the coefficients on the first lag, τ controls the overall tightness of the prior distribution on the VAR coefficients, c is the tightness of the prior on the intercept and the exogenous variables and $J_p = diag(1, ..., P)$ to denote the lags, with $P = 13$. The prior means are chosen as the OLS estimates of the coefficients of an AR(1) regression estimated for each endogenous variable and σ_i's are set using the standard deviation of the error terms from these regressions. As is standard for US data, $\tau = 0.1$ and $c = 1/10^5$ indicating an informative prior on the lags of the endogenous variables but flat for the intercept and the exogenous variables.

As highlighted by Bańbura et al. (2010), the literature suggests the forecasting performance can be improved by adding a prior on the sum of the coefficients of the form:

$$Y_{d,2} = \begin{pmatrix} \frac{diag(\delta_1\mu_1,...,\delta_N\mu_N)}{\lambda} \end{pmatrix}, \quad Z_{d,2} = \begin{pmatrix} \frac{1_{1\times P}\otimes diag(\sigma_1,...,\sigma_N)}{\lambda} & 0_{N\times K} \end{pmatrix} \tag{8}$$

where μ_i denotes the sample means of the endogenous variables. Following Bańbura et al. (2010), the tightness of the prior of the sum of coefficients is set to $\lambda = 10\tau$, a loose prior.

These vectors are plugged into the vector with the actual observations. However, there is a difference with respect to conventional VAR-X models: as aforementioned, X assumes a different value in each draw.[8] Given the artificial data, the observables and the product of the LDA, the model is estimated using an MCMC sampler. The algorithm cycles through the following steps:

Step 1. Compute X using the draw of the LDA. Start from the statement predictive distributions as in Equation (3):

$$\tilde{\hat{\theta}}^k_{\tilde{d}} = \frac{n_{\tilde{d},k} + \alpha}{\sum_{k=1}^{K} n_{\tilde{d},k} + \alpha} \tag{9}$$

X is then given by:

$$X = [\tilde{\hat{\theta}}^1, \; ..., \; \tilde{\hat{\theta}}^{K-1}]$$

where $\tilde{\hat{\theta}}^k$ denotes the time series of $\tilde{\hat{\theta}}^k_{\tilde{d}}$.

Step 2. Draw Σ from the Inverse Wishart:

$$H(\Sigma/\beta, Y, X) = IW(S^*, T^*)$$

where the posterior scale matrix is given by $S^* = (Y^* - Z^*\beta)'(Y^* - Z^*\beta)$, with $\beta^* = (Z^{*'}Z^*)^{-1}(Z^{*'}Y^*)$, $Y^* = [Y; Y_{d,1}; Y_{d,2}]$ and $Z^* = [Z; Z_{d,1}; Z_{d,2}]$. T^* denotes the posterior degrees of freedom given by the number of rows of Y^*.

[8] This is inspired by Bernanke et al. (2005)'s Bayesian FAVAR, even though here, for the aforementioned reasons, textual factors are added to the model as exogenous variables.

Step 3. Draw β from the Normal distribution:

$$H(\beta/\Sigma, Y, X) = N(\beta^*, \Sigma \otimes (Z^{*'} Z^*)^{-1})$$

Step 4. Repeat steps 1 to 3 until the required number of draws has been reached.[9]

In this application, the algorithm is iterated 35,000 times for each data window, with the first 5,000 draws discarded as burn-in. Moreover, in order to reduce correlation across draws and increase efficiency, only every 10th draw from the Markov chain is kept. This gives 3,000 draws which are used to simulate the posterior distributions of β, Σ and X.[10]

4 Forecast Evaluation

Models are compared and selected based on their predictive performance as in Geweke and Amisano (2010) and many other applications in forecasting. In particular, the VAR-teXt is estimated with $K = 5, 10, 15, ..., 50$ and evaluated in comparison with the benchmark VAR. All the models are estimated recursively over an expanding data window.[11] As the full sample period is short, due to the availability of statements, only approximately a third of the sample period is evaluated. Starting from an initial 1998M09-2013M07 window, this results in a set

[9] This was implemented using a combination of Matlab and the 'topicmodels' package in Python (available on https://github.com/sekhansen/text-mining-tutorial).

[10] Results are similar with 6,000 draws and a sampling lag of 5.

[11] It is worth noting that topics may change over the expanding window, but this is not a concern here since $K-1$ textual factors are used in the VAR-teXt. An alternative to fully re-estimate the model every month would involve estimating the topic distributions based on a truncated corpus and using these estimates to obtain the topic time series, selecting only the interpretable topics. This, however, has the caveat of not using all the information available in the regression estimation window to estimate the topic distributions. For example, if the model is estimated for the UK and the term Brexit does not appear in the truncated corpus, it will be overlooked by the LDA even if its number of occurrences in the later period is high. Still, for robustness, this alternative will be explored in the next section.

of 77 out-of-sample forecasts.[12] The comparison is conducted based on forecast densities. Given the draws for β and Σ (and X_t in the case of the augmented model), it is straightforward to compute the 1- and 3-step ahead forecast densities by simulating Y_t forward:

$$H(\hat{Y}_{t+h}/Y_t) = \int H(\hat{Y}_{t+h}/Y_t, \Gamma) \times H(\Gamma/Y_t, \tilde{W}_t) d\Gamma \tag{10}$$

where $h = 1, 2, 3, ...$ and $\Gamma = \{\beta, \Sigma, X_t\}$.

As the proposed MCMC sampling is designed to properly capture textual factor estimation uncertainty leading to more accurate posterior predictive distributions, the analysis will focus on log-scores.[13] Log-scores are the (log) likelihood the model assigns to the actual observations Y_{t+1} given data up to t:

$$LS^i_{t,h} = \ln H(Y^i_{t+h}/Y_t) \tag{11}$$

where both Y^i_{t+h} and Y_t are actual data. Following Alessandri and Mumtaz (2017), these (log) predictive densities are estimated using kernel methods.

Table 1 reports the differences in log-scores over the entire evaluation period relative to the benchmark: positive values favour the VAR-teXt. The table also shows the p-values of Giacomini and White (2006)'s tests of unconditional and conditional predictive ability.[14] Overall, the inclusion of text improves the forecasting performance. In particular, the forecasts of the consumer price inflation and the interest rate are the ones that benefit the most from the information retrieved from the statements.

[12]Models perform poorly in forecasting the shadow rate in 2013M07 due to a spike in this time series, and the resulting relative performance at this month is such an outlier that it would determine the ranking of the average performance had it been included in the evaluation period. That is why the initial window ends in 2013M07 and not earlier in that year. Nonetheless, starting the forecast evaluation at the beginning of 2013 and dropping 2013M07 from the average gives very similar results.

[13]Point forecasts are presented in the appendix.

[14]Such p-values, however, are only indicative since the distributions of the tests are derived based on fixed rolling window estimators. The same applies to the decision rule, which will soon be described. As in Giacomini and White (2006) and Alessandri and Mumtaz (2017) the conditional test is based on the same information set used to generate the forecasts.

Table 1: Log-scores: VAR-teXt versus benchmark VAR

	1M				3M			
K	y	π	r	s	y	π	r	s
5	0.48	2.27	2.01	1.21	1.14	3.51	2.21	1.81
	(0.493)	*(0.000)*	*(0.028)*	*(0.214)*	*(0.186)*	*(0.000)*	*(0.050)*	*(0.009)*
	(0.517)	*(0.002)*	*(0.014)*	*(0.538)*	*(0.006)*	*(0.000)*	*(0.012)*	*(0.043)*
10	-0.10	3.32	2.63	0.64	1.35	5.72	2.74	0.74
	(0.937)	*(0.001)*	*(0.022)*	*(0.407)*	*(0.239)*	*(0.000)*	*(0.011)*	*(0.323)*
	(0.987)	*(0.006)*	*(0.072)*	*(0.353)*	*(0.136)*	*(0.000)*	*(0.000)*	*(0.482)*
15	-1.63	1.98	3.37	1.01	0.22	4.59	3.39	1.61
	(0.297)	*(0.055)*	*(0.027)*	*(0.281)*	*(0.876)*	*(0.000)*	*(0.011)*	*(0.116)*
	(0.512)	*(0.178)*	*(0.100)*	*(0.425)*	*(0.300)*	*(0.000)*	*(0.000)*	*(0.000)*
20	-0.67	2.22	3.41	0.44	1.63	5.96	4.05	2.26
	(0.601)	*(0.054)*	*(0.004)*	*(0.682)*	*(0.275)*	*(0.000)*	*(0.006)*	*(0.010)*
	(0.851)	*(0.170)*	*(0.016)*	*(0.550)*	*(0.048)*	*(0.000)*	*(0.000)*	*(0.000)*
25	-0.01	2.25	2.35	0.25	1.11	6.50	4.38	3.10
	(0.993)	*(0.138)*	*(0.063)*	*(0.822)*	*(0.559)*	*(0.000)*	*(0.004)*	*(0.003)*
	(0.977)	*(0.213)*	*(0.063)*	*(0.979)*	*(0.058)*	*(0.000)*	*(0.000)*	*(0.001)*
30	-0.41	1.64	1.93	-0.96	1.53	7.55	4.55	1.81
	(0.783)	*(0.327)*	*(0.189)*	*(0.547)*	*(0.356)*	*(0.000)*	*(0.005)*	*(0.082)*
	(0.393)	*(0.543)*	*(0.339)*	*(0.812)*	*(0.116)*	*(0.000)*	*(0.001)*	*(0.167)*
35	0.44	2.17	1.30	-1.60	1.16	7.88	4.27	2.85
	(0.784)	*(0.241)*	*(0.438)*	*(0.465)*	*(0.509)*	*(0.000)*	*(0.011)*	*(0.010)*
	(0.328)	*(0.558)*	*(0.701)*	*(0.647)*	*(0.214)*	*(0.000)*	*(0.007)*	*(0.044)*
40	-0.56	1.39	1.43	-2.76	0.85	8.08	4.73	2.23
	(0.755)	*(0.563)*	*(0.462)*	*(0.318)*	*(0.662)*	*(0.000)*	*(0.007)*	*(0.075)*
	(0.683)	*(0.857)*	*(0.682)*	*(0.536)*	*(0.414)*	*(0.000)*	*(0.003)*	*(0.037)*
45	-1.13	1.51	0.74	-2.39	1.26	10.17	4.85	2.89
	(0.615)	*(0.523)*	*(0.725)*	*(0.266)*	*(0.583)*	*(0.000)*	*(0.012)*	*(0.028)*
	(0.656)	*(0.883)*	*(0.866)*	*(0.528)*	*(0.391)*	*(0.000)*	*(0.003)*	*(0.077)*
50	-0.64	1.47	0.11	-2.69	2.04	10.03	3.46	2.58
	(0.794)	*(0.543)*	*(0.964)*	*(0.328)*	*(0.362)*	*(0.000)*	*(0.119)*	*(0.079)*
	(0.462)	*(0.905)*	*(0.949)*	*(0.598)*	*(0.509)*	*(0.000)*	*(0.035)*	*(0.157)*

Notes: The table shows the average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The p-values of Giacomini and White (2006)'s test of unconditional (conditional) predictive ability are in the first (second) parenthesis. The variables are industrial production growth (y), inflation (π), shadow rate (r), and EBP (s).

The improvement for $h = 1$ is only marginal, suggesting it may take some time for the economy to react in line with the literature of monetary policy transmission. Moreover, approximately half of the meetings are not followed by a meeting in the next month, so changes in the statements that may indicate changes in interest rate will take more time to be fulfilled. If one is mostly interested in forecasting the interest rate, $K = 20$ gives the largest difference in log-scores: 3.41%. However, if the focus is on inflation, the model with $K = 10$ is the best one (+ 3.32%). For industrial production and the excess bond premium, it is hard to distinguish between the augmented and the benchmark models.

The improvement is stronger for the 3-month ahead forecasts, a horizon that always comprises the next FOMC meeting. For this horizon, the observations and the forecasts of the variables in first differences (y and π) are cumulated, so the performance is evaluated based on the industrial production growth and the inflation over the following quarter. The model with the largest gains is the one with K=45.[15] In this model, there is an improvement of 10.17% for inflation, 4.85% for the shadow rate and 2.89% for the excess bond premium. For industrial production, there is no significant difference.

Since the predictive power of textual factors and, consequently, models can change, henceforth, the forecast evaluation will be conducted over time. Figure 1 shows the model selection over time using the cumulative difference in log-scores between the augmented model and the benchmark model with $K = 45$.

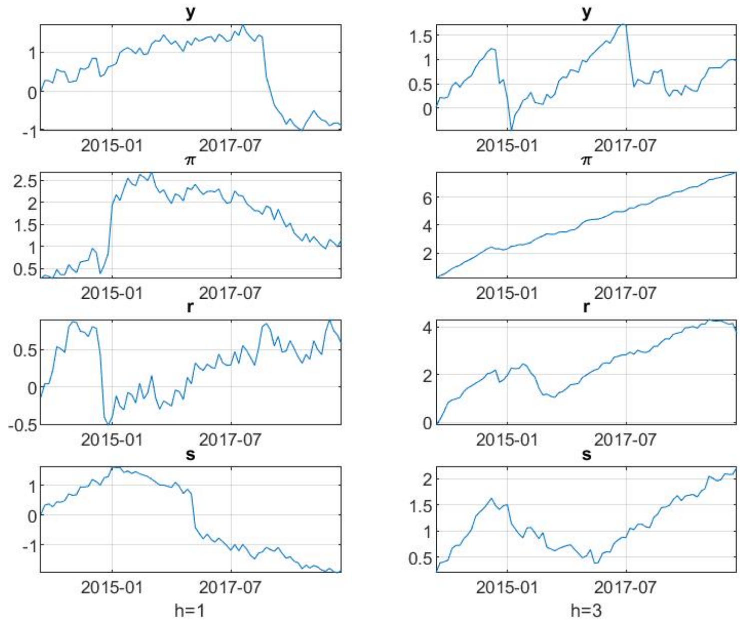

Figure 1: Cumulative log-score difference over time

Notes: The lines show the cumulative difference in log-scores between the VAR-text and the VAR models for horizons 1 and 3.

For $h = 1$, the plots are more erratic and confirm the average performance displayed in the table. For $h = 3$, the plot of industrial production confirms that

[15]While some textual factors are interpretable topics, most of them are not.

the textual factors do not contribute to its forecasts. Inflation exhibits an upward trend indicating that the evidence in favour of the augmented model is gradually built over time but with some reversals in the relative performance as indicated by the changes in the slope. As for the shadow rate, there is also an upward trend, especially from 2016 on. Finally, the behaviour over time of the forecasts of the excess bond premium is more volatile. Overall, ex-post, it is possible to associate some of the changes in the slope with the statements. At time t, however, it would have been hard to anticipate the effects of particular changes in the statements on the performance.

The decision rule proposed by Giacomini and White (2006) provides an alternative way to evaluate the forecasting performance over time. It uses current information to select at every t the model that is expected to work better in $t + h$. Figure 2 shows that, using the log-scores up to time t as the decision criterion, the VAR-teXt is predicted to perform better in the future to forecast inflation and the interest rate apart from few troughs at the beginning of the evaluation period.

In particular, and consistent with the previous table and graph, the augmented model is predicted to perform around 10% better in relation to the benchmark in terms of the log-scores of inflation for $h = 3$ throughout almost the entire evaluation period, sometimes even more than 15%. For inflation, the augmented model is expected to perform around 5% better most of the time, and sometimes more than 10%. Regarding $h = 1$, there is also a consistent, although lower, gain in the use of the VAR-teXt for inflation and, mainly, the interest rate. On the other hand, the other variables display an erratic behaviour for both horizons.

It is worth noting at this point that the first 2 years of the evaluation period are within the ZLB period. This does not invalidate the previous exercise because textual factors can capture unconventional monetary policy through pattern changes in the statements and the shadow rate also summarises this unconventional policy. Nevertheless, the results are also presented next for the post-ZLB period. This reduces the set of forecasts but it is a useful exercise since it allows for the reaction

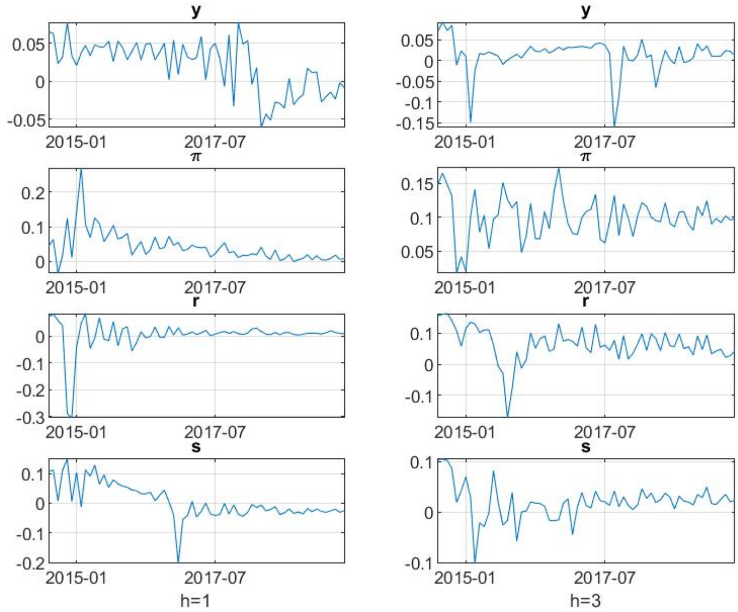

Figure 2: Giacomini and White(2006)'s decision rule

Notes: Positive values indicate that the VAR-teXt is expected to work better in the future and should be selected.

of the Federal Reserve to economic and financial conditions to be properly captured during the ZLB period and for the evaluation of the forecast of the actual fed funds rate, which replaces the shadow rate after 2015M12 in Wu and Xia (2016)'s time series.[16]

Table 2 reports the relative log-scores averaged over 2015M12-2020M02. For $h = 3$, gains are significant for both the inflation and the interest rate for all K. In particular, the forecast of the interest rate improves 5.28% when K=45 and 5.33% when K=40, suggesting the model is better at forecasting the fed funds rate after the ZLB than the shadow rate during it. There are also relevant gains for the inflation forecasts: 9.41% when K=45 and 9.58% when K=50. The VAR-teXt also performs better in forecasting the excess bond premium for all $K's$, even though the magnitude of the improvement is lower.

Overall, using this approach, text improves the forecast of the consumer price

[16]Remember that the shadow rate is given by the minimum value between the fed funds rate and the product of a shadow rate term structure model, so in normal times it is simply the fed funds rate.

Table 2: Forecast evaluation after 2015M12: VAR-teXt versus benchmark VAR

	1M				3M			
K	y	π	r	s	y	π	r	s
5	0.08	1.79	2.43	-0.17	2.92	3.90	2.10	1.60
	(0.922)	*(0.026)*	*(0.014)*	*(0.891)*	*(0.001)*	*(0.000)*	*(0.019)*	*(0.027)*
	(0.428)	*(0.037)*	*(0.047)*	*(0.837)*	*(0.002)*	*(0.000)*	*(0.066)*	*(0.059)*
10	-0.54	2.27	2.57	-0.19	2.36	5.53	2.76	0.60
	(0.725)	*(0.021)*	*(0.017)*	*(0.838)*	*(0.113)*	*(0.000)*	*(0.013)*	*(0.473)*
	(0.898)	*(0.041)*	*(0.069)*	*(0.428)*	*(0.063)*	*(0.000)*	*(0.005)*	*(0.194)*
15	-2.71	0.43	4.17	0.14	1.20	4.93	4.30	1.91
	(0.195)	*(0.637)*	*(0.001)*	*(0.916)*	*(0.436)*	*(0.000)*	*(0.000)*	*(0.092)*
	(0.369)	*(0.969)*	*(0.004)*	*(0.990)*	*(0.358)*	*(0.000)*	*(0.001)*	*(0.016)*
20	-0.90	0.86	4.09	-1.22	2.60	5.72	5.08	2.27
	(0.556)	*(0.419)*	*(0.001)*	*(0.386)*	*(0.168)*	*(0.000)*	*(0.000)*	*(0.014)*
	(0.742)	*(0.523)*	*(0.004)*	*(0.638)*	*(0.058)*	*(0.000)*	*(0.000)*	*(0.013)*
25	-1.46	0.13	2.40	-1.69	2.16	6.64	5.44	3.09
	(0.419)	*(0.919)*	*(0.059)*	*(0.241)*	*(0.364)*	*(0.000)*	*(0.000)*	*(0.008)*
	(0.554)	*(0.117)*	*(0.211)*	*(0.457)*	*(0.035)*	*(0.000)*	*(0.000)*	*(0.001)*
30	-2.36	-1.07	2.22	-3.87	2.49	7.31	5.10	1.47
	(0.190)	*(0.478)*	*(0.153)*	*(0.087)*	*(0.196)*	*(0.000)*	*(0.001)*	*(0.207)*
	(0.252)	*(0.071)*	*(0.245)*	*(0.232)*	*(0.056)*	*(0.000)*	*(0.002)*	*(0.097)*
35	-1.73	-0.97	1.87	-5.16	1.02	7.38	4.27	2.18
	(0.386)	*(0.576)*	*(0.263)*	*(0.104)*	*(0.647)*	*(0.000)*	*(0.004)*	*(0.079)*
	(0.323)	*(0.359)*	*(0.243)*	*(0.275)*	*(0.332)*	*(0.000)*	*(0.009)*	*(0.025)*
40	-3.04	-3.13	1.70	-6.42	1.00	7.59	5.33	2.20
	(0.179)	*(0.115)*	*(0.426)*	*(0.117)*	*(0.679)*	*(0.000)*	*(0.002)*	*(0.130)*
	(0.268)	*(0.048)*	*(0.423)*	*(0.268)*	*(0.533)*	*(0.000)*	*(0.007)*	*(0.066)*
45	-3.74	-2.72	1.33	-6.38	1.84	9.41	5.28	2.55
	(0.215)	*(0.209)*	*(0.548)*	*(0.034)*	*(0.479)*	*(0.000)*	*(0.004)*	*(0.080)*
	(0.329)	*(0.032)*	*(0.244)*	*(0.090)*	*(0.377)*	*(0.000)*	*(0.014)*	*(0.054)*
50	-3.74	-2.59	1.30	-6.72	1.67	9.58	5.09	2.26
	(0.265)	*(0.267)*	*(0.588)*	*(0.086)*	*(0.502)*	*(0.000)*	*(0.007)*	*(0.167)*
	(0.342)	*(0.039)*	*(0.291)*	*(0.219*	*(0.472)*	*(0.000)*	*(0.027)*	*(0.035)*

Notes: The table shows the average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2015M12–2020M02. The p-values of Giacomini and White (2006)'s test of unconditional (conditional) predictive ability are in the first (second) parenthesis. The variables are industrial production growth (y), inflation (π), shadow rate (r), and EBP (s).

inflation and the fed funds rate, and sometimes of the excess bond premium, as summarised by the average performance and further detailed in the graphs.

5 Discussion

This section raises some issues found during the analysis.[17] First, the shadow rate is not the only possible choice of policy indicator. Other candidates are the 1-year government bond rate and the fed funds rate. The 1-year rate has been commonly

[17]For more details, the reader is referred to the appendix.

used by papers that include the ZLB, such as Gertler and Karadi (2015) and Jarociński and Karadi (2020), and the fed funds rate is ultimately the variable of interest although the lack of substantial variation during the ZLB may compromise estimation.

Table 3 shows the main results hold when these variables replace the shadow rate in the model.[18] Results with the alternative policy indicators are as good as the benchmark results. In fact, if one is mostly interested in forecasting the excess bond premium, one of these models should be selected.

Table 3: Forecast evaluation: alternative policy indicators

		1M				3M			
r	K	y	π	r	s	y	π	r	s
GS1	45	-1.57	0.99	2.03	-1.92	1.33	9.48	2.69	3.92
		(0.529)	*(0.668)*	*(0.323)*	*(0.355)*	*(0.490)*	*(0.000)*	*(0.037)*	*(0.001)*
		(0.704)	*(0.893)*	*(0.014)*	*(0.520)*	*(0.428)*	*(0.000)*	*(0.028)*	*(0.002)*
FF	45	-0.95	-0.01	1.03	-2.81	2.09	9.72	5.96	3.85
		(0.688)	*(0.997)*	*(0.595)*	*(0.196)*	*(0.295)*	*(0.000)*	*(0.000)*	*(0.001)*
		(0.783)	*(0.991)*	*(0.072)*	*(0.389)*	*(0.209)*	*(0.000)*	*(0.000)*	*(0.003)*

Notes: The table shows the average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The p-values of Giacomini and White (2006)'s test of unconditional (conditional) predictive ability are in the first (second) parenthesis. The variables are industrial production growth (y), inflation (π), policy indicator (r), and EBP (s).

Furthermore, in order to check whether only some interpretable topics are driving the results while the other textual factors are just introducing noise, topic distributions are now estimated based on a truncated corpus (1998M09-2012M12) and these estimates are used to obtain the topic time series as it is customary in text mining applications in forecasting. The point of departure is Hansen and McMahon (2016), who examine the causal effects of FOMC statements using 15 topics. Their sample period is very similar, and this choice leads indeed to a good fit in terms of interpretability. For details, see the appendix. Five topics related to the economic situation and prospects are selected. Table 4 reports the log-scores of this VAR-teXt with labelled topics relative to the benchmark VAR. The smaller

[18] Note that the benchmark models to which the new VAR-teXts are compared are different.

gains imply this model underperforms the benchmark VAR-teXt.[19]

Table 4: Forecast evaluation: Labelled VAR-teXt versus benchmark VAR

	1M				3M			
K	y	π	r	s	y	π	r	s
15	0.81	2.35	1.32	3.32	0.58	3.24	1.84	3.03
	(0.442)	*(0.014)*	*(0.306)*	*(0.003)*	*(0.642)*	*(0.000)*	*(0.108)*	*(0.000)*
	(0.375)	*(0.006)*	*(0.586)*	*(0.007)*	*(0.049)*	*(0.000)*	*(0.201)*	*(0.002)*

Notes: The table shows the average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and 3-month ahead forecasts over 2013M08–2020M02. The p-values of Giacomini and White (2006)'s test of unconditional (conditional) predictive ability are in the first (second) parenthesis. The variables are industrial production growth (y), inflation (π), shadow rate (r), and EBP (s).

As pointed out before, the text mining literature has advanced considerably in recent years, and a combination of topic models and dictionary methods could also be explored. However, the main takeaway from the results is that even a standard LDA can produce textual factors that are useful for forecasting. In this respect, it is reassuring to confirm the results are not driven only by the interpretable textual factors.

6 The DFM-teXt Model

The DFM-teXt model is a natural extension of the VAR-teXt since the transition equation of dynamic factor models is simply a VAR in the factors, and observables are expressed as a weighted average of factors via the observation equation. The idea behind the DFM-teXt model is to use all available information regardless of its form: quantitative data or text. The point of departure for the analysis is a DFM model of the form:

$$W_{i,t} = \Gamma_i F_t + e_{i,t} \tag{12}$$

$$F_t = \sum_{p=1}^{P} \phi_p F_{t-p} + \mu + A_0 \varepsilon_t \tag{13}$$

[19] 15 is also the number of topics in the grid that gives the lowest perplexity. Therefore, the goodness-of-fit of the LDA does not seem correlated with the forecasting performance of the VAR-teXt, and criteria should be used according to the objective of the model.

where $W_t = (W_{1,t}, ..., W_{N_W,t})$ is a panel of N_W variables, $F_t = (F_t^1, ..., F_t^R)$ denotes the R latent factors, $W_{i,t}$ is related to the factors via the factor loadings Γ_i, $e_{i,t}$ is the *i.i.d.* idiosyncratic component in the observation equation, and ε_t is the error term in the transition equation. Given the transition equation (13) and Γ_i from the observation equation (12), it is possible to forecast all the variables in the panel.

As in the VAR-teXt, the DFM is augmented with textual factors:

$$F_t = \sum_{p=1}^{P} \phi_p F_{t-p} + \phi_X X_{t-1} + \mu + A_0 \varepsilon_t$$

The algorithm for DFM-teXt is presented in the appendix and is similar to the algorithm for VAR-teXt, but, being the DFM-teXt a state-space model, one has to carry out additional steps to sample factor loadings and factors. The panel of variables is composed of the 128 series downloaded from the FRED-MD dataset, the excess bond premium and the shadow rate. Following Bai and Ng (2002)'s criteria, 7 factors are used.[20] Furthermore, a flat prior is used in the estimation of the observation equation and the same prior specification of the VAR is imposed for the transition equation: 13 lags and dummy observations. The DFM-teXt is estimated with K=20 and K=45, and then evaluated as the VAR-teXt was.[21]

Table 5 shows that, in contrast with Kalamara et al. (2020), the added value of text does not degrade in the DFM, although the gains are concentrated in the 3-month ahead forecasts. In particular, even the forecast of industrial production is enhanced using this model. The augmented model continues to be very good at forecasting inflation, but its ability to forecast the interest rate decreases. This is related to the fact that the interest rate enters the DFM in first differences. The appendix reports the results for the other variables. In summary, for $K = 45$, 104 variables present a positive average difference in predictive log-scores, of which 53 (47) are significant at the 10% level according to the unconditional (conditional) test.

[20]Even though the estimation is fully Bayesian, this criterion is chosen because it is less computationally intensive.

[21]As each point in the grid takes weeks to run, only two points are explored.

Table 5: Forecast evaluation: DFM-teXt versus benchmark DFM

	1M				3M			
K	y	π	r	s	y	π	r	s
20	0.77	1.15	-0.33	0.28	3.82	4.53	1.24	2.50
	(0.538)	*(0.178)*	*(0.698)*	*(0.777)*	*(0.007)*	*(0.000)*	*(0.290)*	*(0.006)*
	(0.741)	*(0.447)*	*(0.524)*	*(0.662)*	*(0.012)*	*(0.000)*	*(0.363)*	*(0.007)*
45	-0.55	3.08	-1.65	-2.97	4.07	10.29	-0.78	4.23
	(0.693)	*(0.022)*	*(0.118)*	*(0.031)*	*(0.002)*	*(0.000)*	*(0.622)*	*(0.000)*
	(0.949)	*(0.039)*	*(0.089)*	*(0.028)*	*(0.008)*	*(0.000)*	*(0.570)*	*(0.000)*

Notes: The table shows the average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The p-values of Giacomini and White (2006)'s test of unconditional (conditional) predictive ability are in the first (second) parenthesis. The variables are industrial production growth (y), inflation (π), shadow rate (r), and EBP (s).

The forecasts of only 6 (4) variables become significantly worse when the model is augmented with text. The numbers for $K = 20$ are very similar.

7 Conclusion

This paper has explored the complementarity between traditional econometrics and machine learning and applied the resulting models – the VAR-teXt and the DFM-teXt – to central bank communication. It is becoming a consensus that text can provide information beyond that contained in traditional economic models, being potentially important for forecasting. This paper adds to this literature.

In order to deal with the fact that the LDA outputs are estimates, an MCMC sampling algorithm that tackles the generated regressors issue is also provided. Moreover, unlike previous economic applications, the number of topics is selected based on the predictive performance, which is consistent with the objective of this paper and helps discipline the choice.

Results show that textual factors based on FOMC statements are indeed useful for forecasting in the small-scale VAR and in the DFM-teXt, especially at the 3-month horizon as it seems that the information retrieved from FOMC statements takes some time to affect the forecasts. Specifically, the VAR-teXt outperforms the benchmark VAR in forecasting the consumer price inflation and the interest rate, and this holds under various specifications. As for the DFM-teXt, gains are more general, despite also being even more concentrated in the 3-month horizon. Thus, like factors in factor-augmented models, textual factors can increase the forecasting performance of VAR models even without necessarily having a clear meaning. As a consequence, this approach also favours replicability, since the choice of the number of textual factors is data-driven and does not rely on researchers' interpretability.

Another clear advantage of an automated procedure such as this is scalability, so it is easy to apply it to datasets containing many more documents and words. This approach can also be easily extended to incorporate more than one corpus. Provided that the aggregate number of textual factors is low, just add them as exogenous variables in the regression; otherwise, one has only to run the LDA for each corpus and extract principal components from the pooled textual factor estimates in each draw. Then add the principal components, rather than the textual factors, as exogenous variables at each iteration of the VAR.

In this vein, a straightforward extension is to apply the model to other types of central bank communication, such as minutes and speeches. As text is ubiquitous in many other branches of economics, there are also many potential applications outside monetary policy. Equipped with a supercomputer, one could also depart from previous literature in setting the values of the hyperparameters and explore a grid for α, η and K, selecting the triplet with the best out-of-sample performance. In terms of methodology, an avenue for future work is to make the estimation of textual factors also depend on traditional data.

References

Alessandri, P. and Mumtaz, H. (2017). Financial conditions and density forecasts for US output and inflation. *Review of Economic Dynamics*, 24:66–78.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Bernanke, B. (2007). Federal reserve communications. Speech at the Cato Institute 25th Annual Monetary Conference, Washington, DC, November 14.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Bholat, D., Hans, S., Santos, P., and Schonhardt-Bailey, C. (2015). *Text mining for central banks*. Number 33 in Handbooks. Centre for Central Banking Studies, Bank of England.

Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine learning*, pages 113–120.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Blot, C. and Hubert, P. (2018). Central bank communication during normal and crisis times. *European Parliament Directorate-General for Internal Policies, Monetary Dialogue*.

Campbell, J. R., Evans, C. L., Fisher, J. D., Justiniano, A., Calomiris, C. W., and Woodford, M. (2012). Macroeconomic effects of federal reserve forward guidance [with comments and discussion]. *Brookings Papers on Economic Activity*, pages 1–80.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Forni, M. and Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics*, 66:124–136.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2):216–230.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235.

Gros, D. (2018). When communication becomes the policy. *European Parliament Directorate-General for Internal Policies, Monetary Dialogue.*

Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99:S114–S133.

Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and deliberation within the FOMC: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870.

Hansen, S., McMahon, M., and Tong, M. (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics*, 108:185–202.

Hayo, B. and Neuenkirch, M. (2010). Do federal reserve communications help predict federal funds target rate decisions? *Journal of Macroeconomics*, 32(4):1014–1024.

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., and Kapadia, S. (2020). Making text count: economic forecasting using newspaper text.

Larsen, V. H. and Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1):203–218.

Lucca, D. O. and Trebbi, F. (2009). Measuring central bank communication: an automated approach with application to FOMC statements. Technical report, National Bureau of Economic Research.

Lustenberger, T. and Rossi, E. (2020). Does central bank transparency and communication affect financial and macroeconomic forecasts? *International Journal of Central Banking*.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Newman, D., Lau, J. H., Grieser, K., and Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, pages 100–108.

Roberts, M. E., Stewart, B. M., and Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515):988–1003.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, pages 1–17.

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

Appendix

A. Point forecasts

Point forecasts are given by the root mean square errors (RMSE) based on the arithmetic mean of the draws of the simulated forecasts:

$$RMSE^i_{t,h} = \sqrt{(\hat{Y}^i_{t+h}(M) - Y^i_{t+h})^2} \tag{1}$$

where $\hat{Y}^i_{t+h}(M)$ denotes the average over the forecast density produced by model M for variable i and Y^i_{t+h} is the actual data.

The table below reports the ratios of the RMSE: values lower than 1 favour the VAR-teXt.

Table 1: RMSE: VAR-teXt versus benchmark VAR

	1M				3M			
K	y	π	r	s	y	π	r	s
5	1.01	1.00	1.03	0.99	1.00	1.00	1.01	1.00
	(0.289)	*(0.686)*	*(0.289)*	*(0.495)*	*(0.266)*	*(0.966)*	*(0.468)*	*(0.630)*
	(0.222)	*(0.902)*	*(0.466)*	*(0.368)*	*(0.249)*	*(0.863)*	*(0.114)*	*(0.780)*
10	1.02	0.99	1.00	1.00	1.01	0.99	0.99	1.00
	(0.144)	*(0.642)*	*(0.819)*	*(0.729)*	*(0.115)*	*(0.084)*	*(0.455)*	*(0.850)*
	(0.304)	*(0.346)*	*(0.875)*	*(0.895)*	*(0.240)*	*(0.075)*	*(0.006)*	*(0.673)*
15	1.02	0.99	1.01	0.99	1.01	0.99	0.98	1.00
	(0.162)	*(0.313)*	*(0.660)*	*(0.732)*	*(0.044)*	*(0.198)*	*(0.248)*	*(0.873)*
	(0.311)	*(0.650)*	*(0.420)*	*(0.953)*	*(0.086)*	*(0.421)*	*(0.006)*	*(0.967)*
20	1.02	0.98	1.00	0.99	1.01	0.99	0.98	1.00
	(0.297)	*(0.284)*	*(0.824)*	*(0.736)*	*(0.226)*	*(0.044)*	*(0.185)*	*(0.896)*
	(0.478)	*(0.355)*	*(0.236)*	*(0.938)*	*(0.288)*	*(0.060)*	*(0.006)*	*(0.982)*
25	1.01	0.98	1.00	0.98	1.00	0.99	0.98	0.99
	(0.590)	*(0.162)*	*(0.976)*	*(0.271)*	*(0.521)*	*(0.055)*	*(0.230)*	*(0.629)*
	(0.826)	*(0.328)*	*(0.324)*	*(0.555)*	*(0.548)*	*(0.177)*	*(0.016)*	*(0.916)*
30	1.02	0.98	1.01	0.99	1.01	0.98	0.98	0.99
	(0.433)	*(0.269)*	*(0.737)*	*(0.808)*	*(0.359)*	*(0.048)*	*(0.273)*	*(0.531)*
	(0.701)	*(0.516)*	*(0.476)*	*(0.926)*	*(0.472)*	*(0.169)*	*(0.023)*	*(0.423)*
35	1.01	0.97	1.01	0.99	1.01	0.98	0.98	1.00
	(0.552)	*(0.089)*	*(0.552)*	*(0.749)*	*(0.538)*	*(0.023)*	*(0.269)*	*(0.845)*
	(0.736)	*(0.225)*	*(0.467)*	*(0.913)*	*(0.242)*	*(0.049)*	*(0.060)*	*(0.754)*
40	1.02	0.97	1.01	1.00	1.01	0.98	0.97	0.99
	(0.515)	*(0.181)*	*(0.764)*	*(0.865)*	*(0.406)*	*(0.032)*	*(0.225)*	*(0.637)*
	(0.773)	*(0.208)*	*(0.549)*	*(0.973)*	*(0.508)*	*(0.153)*	*(0.029)*	*(0.439)*
45	1.02	0.96	1.01	0.98	1.01	0.98	0.98	0.99
	(0.470)	*(0.138)*	*(0.658)*	*(0.561)*	*(0.465)*	*(0.020)*	*(0.347)*	*(0.472)*
	(0.708)	*(0.064)*	*(0.603)*	*(0.801)*	*(0.436)*	*(0.105)*	*(0.033)*	*(0.571)*
50	1.01	0.97	1.02	0.99	1.01	0.98	0.98	0.99
	(0.654)	*(0.217)*	*(0.438)*	*(0.818)*	*(0.589)*	*(0.042)*	*(0.493)*	*(0.490)*
	(0.784)	*(0.124)*	*(0.990)*	*(0.901)*	*(0.572)*	*(0.183)*	*(0.107)*	*(0.384)*

Notes: The table shows the ratio of the average RMSEs, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The p-values of Giacomini and White (2006)'s test of unconditional (conditional) predictive ability are in the first (second) parenthesis. The variables are industrial production growth (y), inflation (π), shadow rate (r), and EBP (s).

According to the table, it is hard to distinguish between the augmented and the benchmark models in terms of point forecasts, apart from the 3-month ahead forecast of inflation and interest rate. The magnitude of the improvement is also lower: 2% for the former and 3% for the latter when $K = 40$.

B. Alternative Specifications

i. 1-year rate

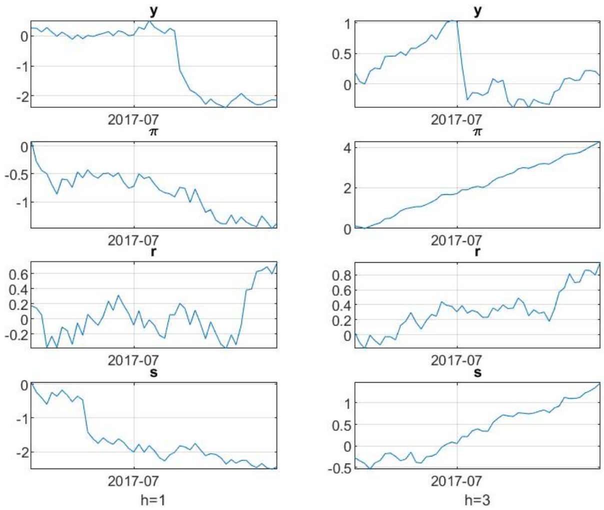

Figure 1: Cumulative log-score difference over time

Notes: The lines show the cumulative difference in log-scores between the VAR-text and the VAR models for horizons 1 and 3.

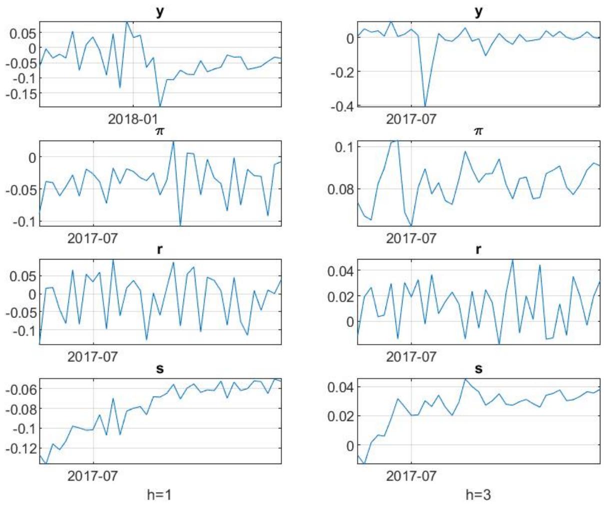

Figure 2: Giacomini and White(2006)'s decision rule

Notes: Positive values indicate the VAR-teXt is expected to work better in the future and should be selected.

ii. Fed funds

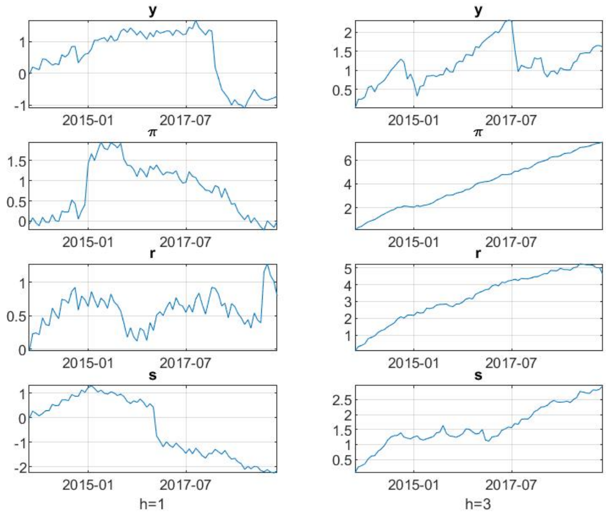

Figure 3: Cumulative log-score difference over time

Notes: The lines show the cumulative difference in log-scores between the VAR-text and the VAR models for horizons 1 and 3.

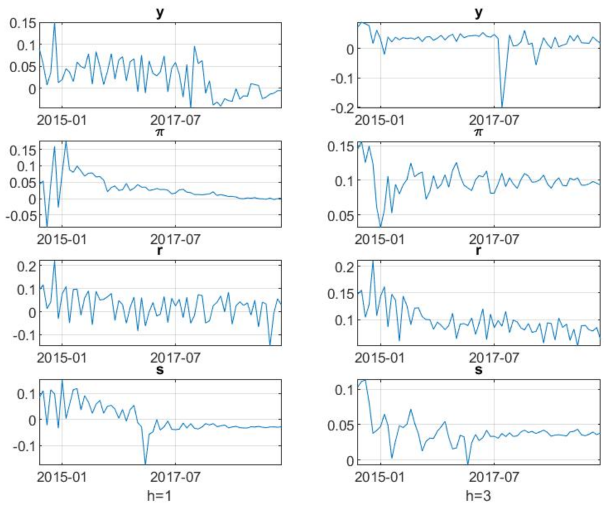

Figure 4: Giacomini and White(2006)'s decision rule

Notes: Positive values indicate the VAR-teXt is expected to work better in the future and should be selected.

C. Labelled Topics

5 topics are selected: topic 3 (economic conditions and growth prospect), topics 8 and 9 (inflation), topic 10 (economic outlook) and topic 13 (interest rates).

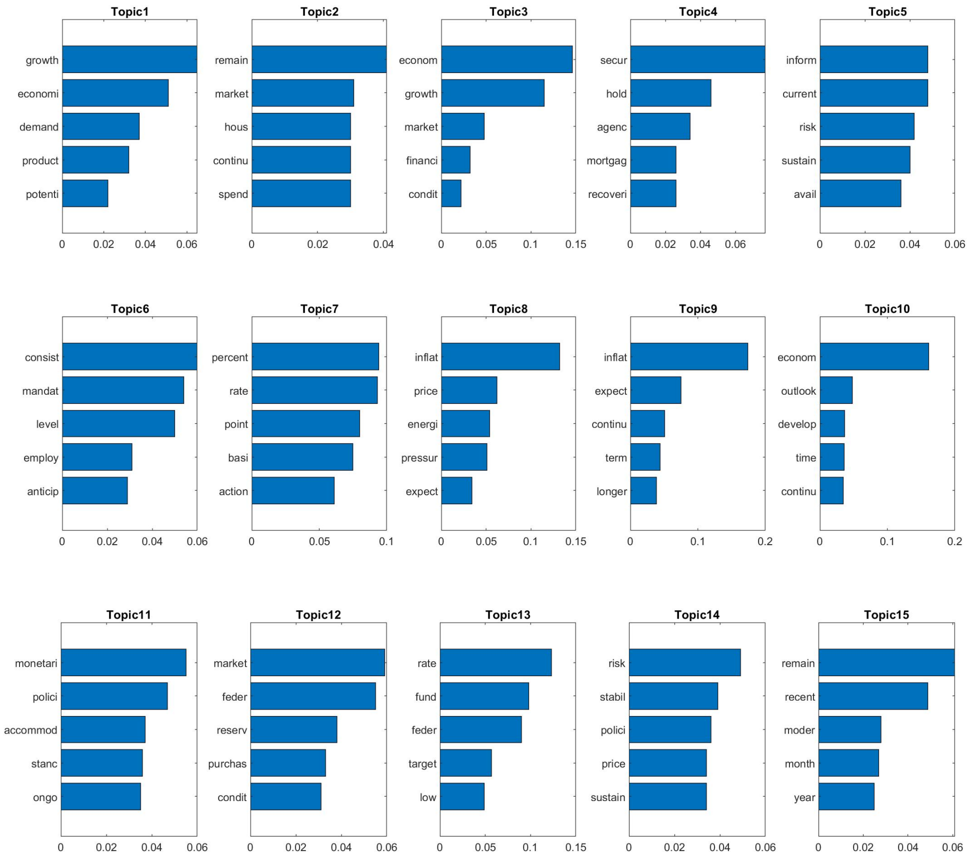

Figure 5: Relative frequency of the top 5 words for K=15

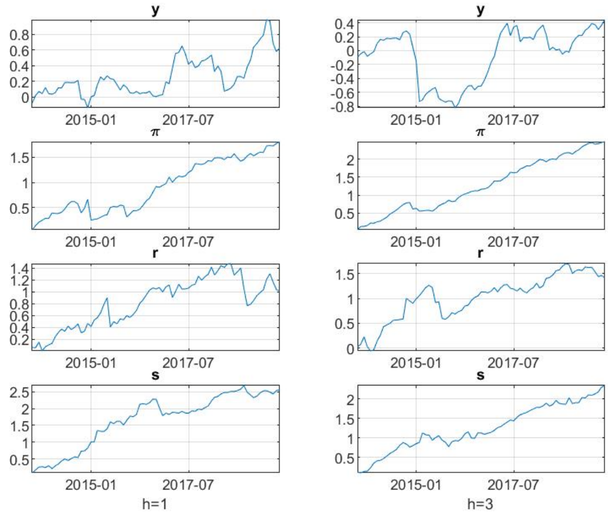

Figure 6: Cumulative log-score difference over time

Notes: The lines show the cumulative difference in log-scores between the VAR-text and the VAR models for horizons 1 and 3.

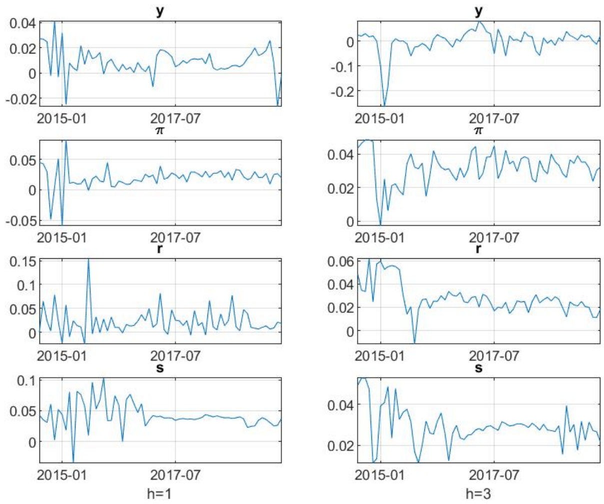

Figure 7: Giacomini and White(2006)'s decision rule

Notes: Positive values indicate the VAR-teXt is expected to work better in the future and should be selected.

D. Algorithm for DFM-teXt

The additional steps in comparison with the algorithm for the VAR-teXt are highlighted:

Step 1. Substep LDA as in Subsection 2.1.

Step 2. Sample factor loadings and the variance of the idiosyncratic components. Conditional on the factors and Ω_{ii}, the factor loadings are sampled from their normal conditional distributions:

$$\Gamma_i \sim N(\bar{\Gamma}_i, \Omega_{ii}\bar{M}_i^{-1})$$

where $\bar{\Gamma}_i$ represents the OLS estimate and $\bar{M}_i = (F'_{i,t}F_{i,t})$.

Conditional on the factors and the factor loadings, the variance of the error term of the observation equation is sampled from:

$$\Omega_{ii} \sim IG\left(\frac{T}{2}, \frac{e'_F e_F}{2}\right) \tag{2}$$

where $e_F = (W_{i,t} - \Gamma_i F_t)$.

Step 3. Sample VAR-X coefficients and covariance as in Subsection 3.1.

Step 4. Prepare matrices for the state-space treating textual factors as exogenous observables and sample factors via the Carter and Kohn (1994) algorithm.

For details see Bernanke et al. (2005). The only modification is that, as the textual factors enter in the model as exogenous variables, they are treated as part of the intercept in the Kalman filter.

Step 5. Repeat steps 1 to 4 until the required number of draws has been reached.

Because estimating the DFM-teXt recursively takes much longer, the algorithm is iterated only 20,000 times for each data window. With a burn-in of 5,000 draws and a sampling lag of 5, this results in 3,000 draws.

E. DFM-teXt Results

Variables	1M	GW-UC	GW-C	3M	GW-UC	GW-C
Real Personal Income	0.57	0.303	0.462	1.07	0.088	0.290
RPI ex Transfers	-0.49	0.417	0.103	1.67	0.009	0.002
Real Consumption	-0.46	0.447	0.501	-0.23	0.812	0.850
Real Sales	-0.55	0.579	0.352	2.39	0.014	0.038

Notes: The table shows average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The GW-UC and GW-C columns display the p-values of Giacomini and White (2006)'s test of unconditional and conditional predictive ability respectively. The variables are described as in the FRED-MD database in addition to the excess bond premium and the shadow rate.

Variables	1M	GW-UC	GW-C	3M	GW-UC	GW-C
Retail Sales	-0.09	0.932	0.275	1.11	0.324	0.581
Industrial Production	-0.55	0.693	0.949	4.07	0.002	0.008
IP Products	-0.46	0.768	0.546	4.50	0.001	0.004
IP Finished Goods	-1.8	0.336	0.599	3.59	0.010	0.015
IP Consumer Goods	-0.05	0.983	0.85	3.38	0.010	0.022
IP Cons Durables	0.98	0.573	0.789	4.35	0.000	0.001
IP Cons Nondurables	-0.73	0.57	0.828	1.51	0.092	0.286
IP Bus Equipment	-0.14	0.91	0.862	1.84	0.161	0.033
IP Materials	0.22	0.847	0.839	3.81	0.002	0.003
IP Durable Materials	-0.58	0.67	0.905	5.33	0.000	0.001
IP Nondurable Materials	-0.37	0.738	0.914	1.98	0.035	0.113
IP Manufacturing	0.13	0.932	0.638	5.80	0.000	0.001
IP Residential Utilities	2.1	0.026	0.059	0.25	0.751	0.654
IP Fuels	-0.92	0.16	0.342	0.47	0.406	0.714
Cap Util: Manufacturing	0.1	0.947	0.498	5.33	0.001	0.003
Help Wanted Index	0.61	0.738	0.706	-0.66	0.490	0.486
Help to Unemployment Ratio	0.54	0.669	0.396	-1.89	0.136	0.147
Civilian Labor Force	-0.08	0.95	0.333	0.40	0.582	0.712
Civilian Employment	-0.15	0.903	0.457	-0.01	0.990	0.871
Unemployment Rate	1.08	0.41	0.375	1.99	0.155	0.242
U Mean Duration	5.26	0.237	0.389	-0.18	0.893	0.792
Unemployed <5 weeks	0.04	0.969	0.326	0.53	0.373	0.589
Unemployed 5-14 weeks	0.1	0.909	0.745	1.36	0.043	0.102
Unemployed >15 weeks	1.02	0.426	0.731	1.34	0.256	0.484
Unemployed 15-26 weeks	-0.05	0.958	0.679	0.22	0.817	0.936
Unemployed >27 weeks	0.89	0.376	0.673	0.24	0.825	0.977
Initial Claims	0.41	0.748	0.855	1.09	0.193	0.424
Non Farm Payroll Employment	-1.7	0.239	0.475	2.07	0.197	0.464
NFP Goods	-0.33	0.835	0.406	1.61	0.307	0.530
NFP Mining	-4	0.417	0.435	-46.95	0.305	0.560
NFP Construction	-1.62	0.172	0.323	-1.35	0.373	0.303
NFP Manufacturing	0.04	0.981	0.384	3.24	0.033	0.085
NFP Durables	0.78	0.619	0.185	5.31	0.000	0.002
NFP Nondurables	-2.64	0.167	0.204	-4.36	0.052	0.122
NFP Services	-0.9	0.445	0.097	2.62	0.165	0.349
NFP TT&U	-0.74	0.633	0.523	2.80	0.138	0.345
NFP Wholesale Trade	1.81	0.316	0.537	2.95	0.018	0.068
NFP Retail Trade	0.01	0.995	0.83	2.76	0.286	0.492
NFP Financial	-1.42	0.17	0.347	-2.83	0.061	0.146
NFP Government	0.41	0.484	0.823	1.57	0.011	0.026
Average Weekly Hours Goods	-6.08	0.001	0.001	-8.72	0.000	0.000
Overtime Weekly Hours Mfg	-2.34	0.044	0.097	0.42	0.653	0.467

Notes: The table shows average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The GW-UC and GW-C columns display the p-values of Giacomini and White (2006)'s test of unconditional and conditional predictive ability respectively. The variables are described as in the FRED-MD database in addition to the excess bond premium and the shadow rate.

Variables	1M	GW-UC	GW-C	3M	GW-UC	GW-C
Average Weekly Hours Mfg	-6.71	0	0	-9.96	0.000	0.000
Housing Starts	-0.14	0.905	0.902	3.42	0.018	0.002
HS Northeast	-18.57	0.324	0.377	-20.49	0.280	0.403
HS Midwest	-19.36	0.304	0.483	6.30	0.119	0.148
HS South	-1.54	0.106	0.198	1.20	0.355	0.284
HS West	-0.55	0.646	0.51	0.37	0.795	1.000
Building Permits	-0.43	0.782	0.006	5.12	0.001	0.000
BP Northeast	-39.3	0.159	0.356	-51.18	0.291	0.377
BP Midwest	-0.28	0.834	0.721	3.06	0.049	0.066
BP South	-2.8	0.05	0.121	2.05	0.132	0.036
BP West	2.9	0.052	0.133	1.00	0.395	0.602
Orders Consumer Goods	0	0.999	0.046	1.70	0.127	0.309
Orders Durable Goods	9.53	0.234	0.287	1.82	0.010	0.047
Orders Nondefense Capital Goods	0.06	0.998	0.337	-1.70	0.368	0.653
Unfilled Orders Durable Goods	-3.93	0.462	0.456	0.76	0.624	0.335
Business Inventories	1.46	0.164	0.265	3.75	0.024	0.072
Inventories to Sales Ratio	-0.71	0.493	0.802	1.72	0.089	0.182
M1 Money Stock	0.9	0.215	0.415	1.91	0.001	0.000
M2 Money Stock	1.02	0.203	0.319	0.74	0.202	0.537
M2 Real	-0.18	0.848	0.968	0.34	0.710	0.875
Monetary Base	0	0.998	0.587	2.09	0.003	0.013
Total Reserves	-0.3	0.584	0.315	0.49	0.323	0.520
Nonborrowed Reserves	-1.99	0.318	0.564	-0.34	0.781	0.996
Business Loans	-0.21	0.681	0.886	0.54	0.345	0.594
Real Estate Loans	0.96	0.03	0.079	0.23	0.628	0.850
Total Nonrevolving Credit	15.65	0.467	0.366	1.04	0.070	0.026
Credit to Income Ratio	5.23	0.33	0.495	0.47	0.405	0.456
S&P 500	-6.04	0.033	0.12	1.45	0.554	0.838
S&P Industrials	-6.04	0.03	0.106	2.34	0.353	0.636
S&P Dividend Yield	-6.08	0.019	0.075	1.85	0.439	0.747
S&P PE Ratio	-2.93	0.046	0.154	2.34	0.074	0.134
Effective FFR	-2.46	0.074	0.049	1.05	0.492	0.627
3M Commercial Paper	-0.26	0.777	0.131	0.34	0.766	0.858
3M T-Bill	-1.68	0.205	0.01	1.60	0.241	0.526
6M T-Bill	-2.22	0.099	0.049	0.51	0.721	0.238
1Y T-Bond	-2.03	0.107	0.067	-0.03	0.981	0.062
5Y T-Bond	-1.32	0.145	0.337	0.78	0.471	0.611
10Y T-Bond	-0.95	0.224	0.268	0.15	0.858	0.882
Aaa Corporate Bond Yield	-0.27	0.657	0.908	-0.16	0.829	0.911
Baa Corporate Bond Yield	-0.38	0.649	0.671	1.02	0.352	0.376
CP-FFR Spread	-0.47	0.589	0.889	3.58	0.000	0.000
3M-FFR Spread	-2.6	0.077	0.188	4.11	0.007	0.002

Notes: The table shows average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The GW-UC and GW-C columns display the p-values of Giacomini and White (2006)'s test of unconditional and conditional predictive ability respectively. The variables are described as in the FRED-MD database in addition to the excess bond premium and the shadow rate.

Variables	1M	GW-UC	GW-C	3M	GW-UC	GW-C
6M-FFR Spread	-2.23	0.134	0.288	3.28	0.015	0.031
1Y-FFR Spread	-3.49	0.022	0.063	2.84	0.035	0.121
5Y-FFR Spread	-4.33	0.028	0.018	-0.83	0.723	0.318
10Y-FFR Spread	-5.47	0.016	0.049	-3.11	0.299	0.326
Aaa-FFR Spread	-13.34	0	0.002	-7.90	0.052	0.120
Baa-FFR Spread	-6.37	0.001	0.006	-4.44	0.082	0.127
Trade Weighted Exchange Rate	-0.09	0.917	0.796	0.58	0.464	0.482
FX Rate CHF	0.04	0.941	0.357	0.95	0.035	0.106
FX Rate JPY	-0.65	0.452	0.623	-0.61	0.541	0.654
FX Rate GBP	0.89	0.522	0.551	1.69	0.036	0.080
FX Rate CAD	-1.16	0.361	0.64	-0.66	0.615	0.885
PPI Final Goods	1.65	0.122	0.314	6.90	0.000	0.000
PPI Consumer Goods	1.21	0.225	0.531	7.61	0.000	0.000
PPI Intermediate Material	2.8	0.006	0.017	7.54	0.000	0.000
PPI Crude Material	-0.06	0.947	0.919	4.27	0.000	0.000
Crude Oil WTI Price	1.08	0.305	0.217	3.57	0.000	0.000
PPI Commodities	-0.79	0.181	0.3	0.75	0.172	0.312
CPI All	3.08	0.022	0.039	10.29	0.000	0.000
CPI Apparel	-6.19	0.285	0.28	0.71	0.349	0.345
CPI Transport	2.89	0.032	0.11	10.56	0.000	0.000
CPI Medical	3.38	0.077	0.228	1.18	0.099	0.128
CPI Commodities	2.72	0.033	0.07	10.42	0.000	0.000
CPI Durables	1.47	0.457	0.339	0.98	0.108	0.146
CPI Services	0.29	0.686	0.056	1.72	0.003	0.015
CPI ex Food	3.01	0.025	0.037	10.51	0.000	0.000
CPI ex Shelter	2.54	0.048	0.068	10.33	0.000	0.000
CPI ex Medical	2.91	0.023	0.041	10.42	0.000	0.000
PCE Deflator	1.89	0.114	0.101	9.10	0.000	0.000
PCE Durables	1.54	0.16	0.382	0.92	0.077	0.130
PCE Nondurables	2.66	0.039	0.079	9.95	0.000	0.000
PCE Services	0.04	0.961	0.884	1.18	0.011	0.037
Average Earnings Goods	-2.42	0.003	0.006	0.49	0.373	0.587
Average Earnings Construction	-27.18	0.139	0.229	-0.22	0.651	0.093
Average Earnings Manufacturing	-0.45	0.514	0.866	0.38	0.485	0.716
Consumer Sentiment	-1.64	0.101	0.2	1.08	0.223	0.350
MZM Money Stock	-0.89	0.228	0.172	2.10	0.000	0.002
CL Motor Vehicles	1.25	0.185	0.36	0.02	0.965	0.419
Consumer Loans	-26.53	0.61	0.367	-0.44	0.724	0.563
Securities in Bank Credit	-0.32	0.527	0.275	0.87	0.102	0.062
VXO	-4.64	0.001	0.005	0.10	0.937	1.000
GZ Excess Bond Premium	-2.97	0.031	0.028	4.23	0.000	0.000
Shadow Rate	-1.65	0.118	0.089	-0.78	0.622	0.570

Notes: The table shows average difference in predictive log-scores multiplied by 100, relative to the benchmark for the 1-month and the 3-month ahead forecasts over 2013M08–2020M02. The GW-UC and GW-C columns display the p-values of Giacomini and White (2006)'s test of unconditional and conditional predictive ability respectively. The variables are described as in the FRED-MD database in addition to the excess bond premium and the shadow rate.