CENTRE FOR APPLIED MACROECONOMICS AND COMMODITY PRICES (CAMP)

CAMP Working Paper Series
No 14/2023

Risky news and credit market sentiment

Paul Labonne and Leif Anders Thorsrud

Risky news
and credit market sentiment*

Paul Labonne[†]   Leif Anders Thorsrud[‡]

This version: December 14, 2023

Abstract

The nonlinear nexus between financial conditions indicators and the conditional distribution of GDP growth has recently been challenged. We show how one can use textual economic news combined with a shallow Neural Network to construct an alternative financial indicator based on word embeddings. By design the index associates growth-at-risk to news about credit, leverage and funding, and we document that the proposed indicator is particularly informative about the lower left tail of the GDP distribution and delivers significantly better out-of-sample density forecasts than commonly used alternatives. Speaking to theories on endogenous information choice and credit-market sentiment we further document that the news-based index likely carries information about beliefs rather than fundamentals.

JEL-codes: C53, E27, E32, E44, G12

Keywords: Forecasting, non-linearity, news, word embeddings

*Comments from Hilde C. Bjørnland, Jamie Cross, and seminar participants at BI Norwegian Business School, e61 Institute, University of Sydney, and the Dolomiti Macro Meeting improved the quality of this paper considerably. This work is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School, and we are grateful for funding from the Research Council of Norway through the project "NRC 315008: The Corona-crisis, structural change, and macroeconomic policy". The project would not have been possible without data shared to use by the *Dow Jones Newswires Archive*. We also thank Vegard H. Larsen for help in preparing some of the data.

[†]Centre for Applied Macroeconomics and Commodity Prices, BI Norwegian Business School.

[‡]Corresponding author. Centre for Applied Macroeconomics and Commodity Prices, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Email: `leif.a.thorsrud@bi.no`

1 Introduction

Macro-prudential policies, adopted by numerous institutions around the world, rely on economic models highlighting the role of the financial sector in creating and propagating vulnerabilities in the economy (Minsky, 1977, 1986; Kindleberger, 1978; Gertler and Bernanke, 1989; Kiyotaki and Moore, 1997; Bernanke et al., 1999). From a time series perspective this theory and practice have spurred a large literature documenting how financial indicators of, e.g., credit, leverage, and funding, have a strong association with the downside risk, as opposed to the central tendency, of future output growth. The work by Giglio et al. (2016) and Adrian et al. (2019) have been particularly influential, showing how financial indicators combined with quantile regressions, or related non-linear methods, can be used to produce out-of-sample growth predictions with potentially fat tails capturing growth-at-risk (GaR).[1]

However, the nonlinear relationship between financial conditions indicators and the conditional distribution of GDP growth has not been uncontested. Hasenzagl et al. (2020), henceforth HPRR, have recently convincingly argued that the financial indicators usually considered, with the National Financial Condition Index (NFCI) in the U.S. being a leading example, do not provide robust and precise advance warnings about any features of the GDP growth distribution other than the mean. In particular, they argue, the NFCI contribute little to such distributional forecasts beyond the information contained in real indicators, such as the first common factor (FMDF) extracted from the FRED-MD dataset.[2]

Motivated by this debate and its policy relevance, we show how one can use textual economic news combined with a shallow Neural Network to construct an alternative financial indicator based on word embeddings. We then evaluate this index relative to the NFCI in terms of how well it characterizes and forecasts U.S. GDP growth using a parametric skewed Student-t (Skew-t) distribution with, building on Delle Monache et al. (2021) and Labonne (2022), time-varying location, scale, and shape parameters. Finally,

[1]Following Adrian et al. (2022) GaR is typically defined as the value of GDP growth at the lower fifth percentile of the predicted growth distribution, conditional on an index of financial stress. Prasad et al. (2019) provides a description of the use of this framework at, e.g., the IMF, and Greenspan (2004) and Kilian and Manganelli (2008) shows how forecasting can be looked upon as a risk-managing exercise. See, e.g., Loria et al. (2022), Brownlees and Souza (2021), Figueres and Jarocinski (2020), and Chernis et al. (2023) for recent contributions to the GaR literature.

[2]This dataset is compiled by McCracken and Ng (2016), contains over 100 monthly economic indicators, and builds upon the seminal contribution by Stock and Watson (1989), and the literature that followed, using large datasets for macroeconomic forecasting and monitoring. Similarly, the NFCI, developed by Brave et al. (2012), is constructed as the common factor extracted from over 100 financial variables, covering four categories of data; credit quality, funding risk, nonfinancial and financial leverage.

we dissect the informational content of the derived index by linking it to shocks to expectations about the current state of the economy and popular sentiment-driven views on the credit cycle (Greenwood et al., 2016; Bordalo et al., 2018; López-Salido et al., 2017).

Our usage of economic news is motivated by the fact that it is timely, seldom subject to revisions, and might contain new information about economic fundamentals as well as fluctuations in sentiment. Similarly, word embeddings represent words in vector space and have been crucial for scientific advances and improvements on down stream tasks in the Natural Language Processing (NLP) literature over the last decade. The reason is they capture well shared context of words in the corpus, densely encode many linguistic regularities and patterns, and allow for arithmetic operations capturing associative meaning. Here we take advantage of these properties by constructing monthly word embedding matrices based on a large corpus of business news provided by the *Dow Jones Newswires Archive* (DJ) covering the sample 1985M1 to 2020M4. At each point in time (month), we then regress a word vector pointing in the growth-at-risk (recession risk) dimension in vector space on a word vector related to credit, leverage, and funding, and use the size of the association as a measure of financial conditions relevant for growth.

As seen in Figure 1, the proposed indicator, which we label the Risky News Index (RNI), tend to increase at least one year prior to economic recessions. Compared to the NFCI and FMDF the RNI is more volatile but also seems to be leading. Consistent with this, we show more formally later that the RNI Granger-causes these variables and that it is only weakly contemporaneously correlated with them. In line with these statistics, we document using the Skew-t model that the RNI performs on-par with the NFCI for one-quarter-ahead predictions, but significantly improves upon it at the one-year-ahead horizon and during recessions periods. Figure 1 provides an example, showing how downside risk predictions become much more informative when using the RNI relative to the NFCI during the Great Financial Crisis (GFC). Section 3 presents a battery of more formal tests of (density) forecast accuracy telling the same story: For the one-year-ahead horizon, and during recessions, the RNI provides significantly better predictions for the GDP growth distribution. Moreover, in-sample estimates of the time-varying moments are relatively precisely estimated, and suggest that the RNI significantly affects the evolution of the shape parameter in particular.

These predictive results are important for at least two reasons. First, the recent debate about the nonlinear nexus between financial indicators and the distribution of future GDP growth questions the justification for macro-prudential policies based on GaR predictions. In contrast, our results are well in line with a large theoretical literature linking financial conditions to adverse growth outcomes and thereby also provide a more positive view on such policies. Indeed, when performing an intrinsic evaluation of the

Figure 1. The left graph reports the (standardized) estimates of quarterly U.S. GDP growth, and the monthly NFCI, RNI, and FMDF. The FMDF is the first common factor (FMDF) extracted from the FRED-MD dataset using a principal components decomposition. The observation dates are restricted by the data availability for estimating the RNI. Shaded areas represent NBER recession dates. The right graph reports the one-year-ahead out-of-sample density predictions obtained from fitting a Skew-t model to the data using either the NFCI or the RNI as explanatory variables for the time-varying moments.

estimated embeddings we find positive evidence regarding their financial interpretation and growth association.

Second, from a practical perspective informative one-year-ahead predictions are more valuable than shorter horizon predictions because it gives policymakers time to respond proactively on budding crises. As such, although we acknowledge that most studies in this literature, including ours, are riddled by relatively few observations and events, the suggested RNI is a potentially helpful alternative to commonly used indicators. Related to this, while a large literature has highlighting the importance of "excessive" credit growth as an indicator of more long-run risk (Jordà et al., 2011, 2013; Reinhart and Rogoff, 2009; Schularick and Taylor, 2012), we also show that the RNI outperforms different credit cycle aggregates for one-year-ahead predictions.[3]

Since it is implausible to assume that the RNI contains information about future shocks, its relatively strong predictive performance must be driven by properties of the underlying news data or by the technology used to construct it. However, we document that also simpler NLP methods linking growth-at-risk to financial conditions yield forecasting gains relative to the NFCI, suggesting an important role for the news data.

To make further progress along this line of thinking we start from a view, motivated by classic accounts of financial crises, emphasizing credit market sentiment (Minsky, 1977, 1986; Kindleberger, 1978), and proceed in two complementary directions. First, we use the Structural Vector Autoregression (SVAR) framework proposed by Enders et al. (2021) to estimate aggregate shocks to beliefs and fundamentals and analyze their correlation with

[3] In concurrent work Adämmer et al. (2023) document that textual news data helps predict the left tail of a larger set of macroeconomic variables. In contrast to our study, however, they do not explore the nonlinear nexus between financial conditions and growth-at-risk.

the RNI. Second, we build on recent work by López-Salido et al. (2017) and investigate to what extent the RNI contains a predicable component related to lagged valuation indicators and expected returns on credit assets, which López-Salido et al. (2017) associate with credit market sentiment.

Our more structural analysis shows that the RNI is significantly correlated with belief shocks and not significantly correlated with shocks to fundamentals. Likewise, we find a significant correlation between lagged valuation indicators and the RNI. Thus, both of these results are supportive of the view that the RNI captures shocks to expectations and credit market sentiment. However, we also show that the predictive power of credit market sentiment derived from credit market valuation indicators is surpassed by the RNI, which suggest that it either provides a better approximation of this concept, or that media coverage about financial conditions and growth-at-risk has independent effects on economic outcomes.

While it is difficult to empirically disentangle these two explanations, it is interesting to note that the credit market sentiment view advocated in many newer studies, such as Greenwood et al. (2016) and Bordalo et al. (2018), rests on assumptions about "diagnostic expectations" and extrapolative beliefs of investors. In contrast, in general models of endogenous information choice and independent media effects, a departure from the standard rational expectations paradigm is not needed (see, e.g., Nimark and Pitschner (2019) and Chahrour et al. (2021)). To the best of our knowledge, however, nobody has derived theoretical models incorporating this type of mechanism for describing growth-at-risk. In light of our findings, this seems to be a promising avenue for future research.

The rest of this article is organized as follows: In Section 2 we describe the news data and the word embedding methodology. Section 3 introduces the Skew-t model and provides in- and out-of-sample evidence on the non-Gaussian news-GDP-growth relationship. Section 4 links the derived index to theories on endogenous information choice and credit-market sentiment, while Section 5 concludes.

2 News data and word embeddings

In the following we first present the DJ news corpus and the word embedding procedure. Then we describe how the news-based word embeddings are used to construct a measure of financial conditions relevant for growth, and, in light of the discussion in HPRR, investigate some of its statistical features relative to the NFCI and FMDF. Finally, we provide more intuition for the word embedding methodology by conducting an intrinsic evaluation of the estimated embeddings.

2.1 Data and estimation

The DJ corpus consists of roughly 25 million news articles written in English, covering the period from 1985M1 to 2020M4. The database contains a large range of Dow Jones' news services, including content from *The Wall Street Journal*. This is *Dow Jones* company's flagship publication, and also one of the largest newspapers in the U.S. in terms of circulation. This means that it has a large footprint in both the U.S. and global media landscape and that important ongoing stories and discussions are well covered by this type of news outlet.

The news corpus is cleaned prior to estimation. We remove all email and web addresses, numbers, and special characters, erase punctuation, set all letters to lowercase, and remove words containing fewer than two or more than 15 letters. These feature selection steps reduce the vocabulary size to approximately 90000 unique terms. The dimension reduction facilitates estimation and is common in the literature. Finally, the corpus is partitioned into monthly blocks of text.

The monthly data blocks are used as input in a word embedding model. The famous and widely used word2vec algorithm (Mikolov et al., 2013) is one of many algorithms used to compute such vectors and is often denoted as a skip-gram model with negative sampling. In essence, the method uses a binary classification problem, evaluating if the center word w_c is likely to show up near the target word w_j to compute the classifier weights that will be the actual word embeddings.[4]

More formally, let w_j be a word from the vocabulary V, with size $|V|$, define a context window of size m, and assume a bigram model, where the probability of the sequence depends on the pairwise probability of a word in the sequence and the word next to it, as $P(w_{c-m}, w_{c-m+1}, \ldots, w_{c+m-1}, w_{c+m}) = \prod_{j=0, j\neq m}^{2m} P(w_{c-m+j}|w_c)$. The intuition for the skip-gram model is then to maximize this probability such that a word is likely to occur near the target if its embedding is similar to the target embedding, where similarity is approximated by the dot product of the word vectors. For one target/center word pair (w_{c-m+j}, w_c), with vector representations $\boldsymbol{u}_{c-m+j}$ and $\boldsymbol{v}_c$, the likelihood is

$$L(\boldsymbol{\theta}) = log\frac{1}{1+e^{-\boldsymbol{u}'_{c-m+j}\cdot\boldsymbol{v}_c}} + \sum_{k=1}^{K} log\frac{1}{1+e^{\tilde{\boldsymbol{u}}'_k\cdot\boldsymbol{v}_c}}, \tag{1}$$

[4]The word2vec model can be trained using either the skip-gram or the continous bag-of-words (CBOW) algorithm. Skip-gram works well with small datasets, and can better represent less frequent words. We therefore use this approach here. Other widely used word embedding models include GloVe (Pennington et al., 2014) and fastText Joulin et al. (2016). Newer Transformer-based models, such as BERT (Devlin et al., 2018), explicitly learn context specific embeddings, and might potentially deliver higher quality embeddings. However, these models typically contain hundreds of millions parameters, making it unfeasible to estimate on a monthly basis. Transfer learning and fine-tuning approaches are possible, but would easily involve some type of forward-looking bias in our setting.

where $\boldsymbol{\theta}$ contains the latent word vectors, and the logistic (or sigmoid) function is used to turn the similarity measure between the word vectors for $\boldsymbol{v}_c$ and $\boldsymbol{u}_{c-m+j}$ into probabilities. The last term in (1) relates to the negative sampling part of the skip-gram model name. As running text is used as input to the model, only positive examples are present and negative examples must be generated and added to the data. These terms are commonly called noise terms ($\tilde{\boldsymbol{u}}_k$). For each target word, it is common to add K noise words.

Maximizing (1) and estimating the latent word vectors is done using a two-layered neural network. $\boldsymbol{V} \in \mathbb{R}^{n\times|V|}$ is the parameter matrix in the first layer, with column $\boldsymbol{v}_i$ the input vector representation of word w_i (word embedding). $\boldsymbol{U} \in \mathbb{R}^{n\times|V|}$ is the parameter matrix in the second layer, with row $\boldsymbol{u}_i$ the output vector representation of word w_i (context embedding). Learning proceeds as follows: For a one hot input vector $\boldsymbol{x} \in \mathbb{R}^{|V|}$ of a center word, the first layer selects $\boldsymbol{v}_c = \boldsymbol{V}\boldsymbol{x}$ and the second layer uses the sigmoid activation function on the score $\boldsymbol{z} = \boldsymbol{U}\boldsymbol{v}_c$. The predicted values are compared to the one hot vectors of the actual output, and the unknown parameters ($\boldsymbol{V}$ and $\boldsymbol{U}$) are updated using Stochastic Gradient Descent. This method is fast, efficient to train, and available in many software packages. We set the context window $m = 5$ and restrict the word embedding length to $n = 100$. The network is trained for five epochs on every monthly data partition.

2.2 Constructing the RNI

To construct a news-based measure relating growth-at-risk to financial conditions, we use the linguistic regularities and patterns encoded in the estimated word embeddings together with arithmetic operations. We first compute two aggregate embeddings aimed at capturing "growth-at-risk" and "financial conditions" as

$$\begin{aligned} \textbf{growth-at-risk}_t &= \boldsymbol{recession}_t + \boldsymbol{risk}_t \\ \textbf{financial conditions}_t &= \boldsymbol{credit}_t + \boldsymbol{leverage}_t + \boldsymbol{funding}_t, \end{aligned} \tag{2}$$

where the individual terms on the right hand side of each quality sign are the word embedding associated with that particular word in a given month t.

The motivation for the **financial conditions**$_t$ aggregate is simply taken from the GaR literature, where financial conditions is linked to exactly fluctuations in credit, leverage, and funding. Indeed, the commonly used NFCI is a weighted average of financial indicators reflecting these three categorizes (Brave et al., 2012). Therefore, by adding these terms together in (2) we obtain an aggregate vector aimed at pointing in the relevant financial dimension in the high-dimensional vector space.

The motivation for the **growth-at-risk**$_t$ aggregate is somewhat more subtle. Since the term "growth" together with "risk" can be used in both a positive (upside risk) and negative (downside risk) context, but the GaR concept is explicitly defined to be associated with the left tail of the growth distribution (Adrian et al., 2022), it is misleading to simply add $\boldsymbol{growth}_t$ to $\boldsymbol{risk}_t$ to capture growth-at-risk. Instead, we use the term "recession", which has a solely negative connotation.

Next, to capture the monthly association between these two aggregates, i.e., the RNI, we solve

$$RNI_t \equiv \hat{\beta}_t = \arg\min S(\beta_t) \quad S(\beta_t) = \|\textbf{growth-at-risk}_t - \textbf{financial cond.}_t \times \beta_t\|^2 ,$$

where an increase in RNI_t implies a stronger association between how the news media writes about growth-at-risk and financial conditions. Although $\hat{\beta}_t$ is estimated using the OLS estimator on each monthly data block, the subscript t is used to highlight that this relationship potentially changes through time.[5]

In Section 2.4 we provide further intuition for the word embedding approach. Here we emphasize two additional aspects of the procedure. First, changes in RNI_t across time can be due to either high-frequency changes in how the media relates growth-at-risk to financial conditions, more persistent changes in how this relationship is focused upon, or noise and breaks in the news coverage and style. To isolate the two former components we apply simple (backward-looking) moving average filters which first normalizes each observation with the mean and standard deviation of the last five years of raw data and then smooths the resulting series by the trailing six month average.[6] Second, to take into account estimation uncertainty of the RNI we conduct subsampling. The approach is described in Appendix A.1. However, as seen in Figure A.1, in Appendix A, the index is very precisely estimated, and going forward we only use the mean estimate.

[5]The studies by Kapfhammer et al. (2020) and Kozlowski et al. (2019) entertain the same type of idea, and successfully show how time-variation in these types of associations can be informative about climate change transition risk and changes in cultural associations, respectively.

[6]Similar types of normalization procedures are commonly applied in the news-based finance and economics literature (Tetlock et al., 2008; Baker et al., 2016; Caldara and Iacoviello, 2022). The length of the window used to remove the slow-moving trend component is motivated by the cycle lengths typically assumed in the literature studying debt and credit cycles, see, e.g., Hamilton and Leff (2020) and Drehmann and Yetman (2021) for recent applications and discussions regarding filtering methodology. Admittedly, the length of the smoothing window is somewhat arbitrary. One could treat these choices as tuning parameters for minimizing loss on the downstream task of predicting the conditional GDP distribution. We have not explored this alternative and leave that for future research. Figure A.2, in Appendix A, shows that the out-of-sample forecasting results obtained in Section 3 are fairly robust to using either the raw version of the RNI or a version without any smoothing.

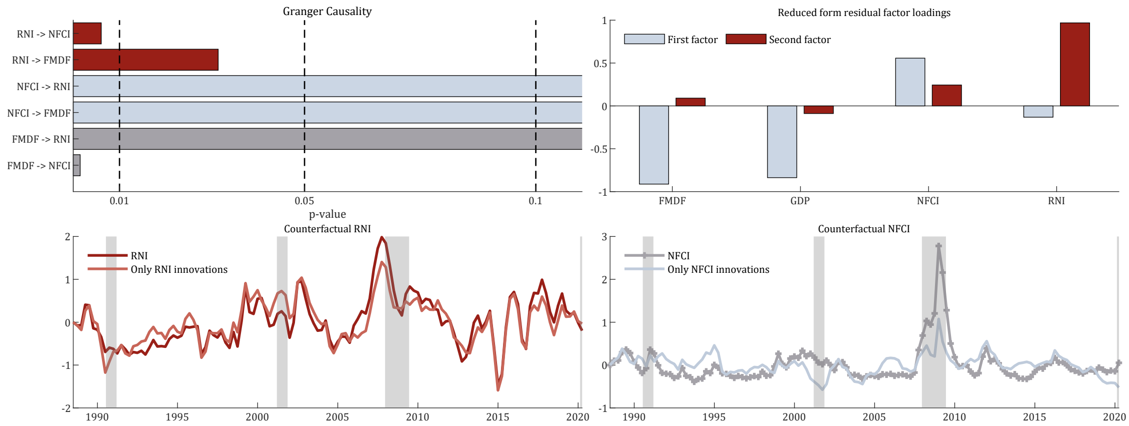


Figure 2. The upper left graph reports Granger-causality test statistics. A four variable quarterly Vector Autoregression (VAR) model, including GDP growth, the NFCI, RNI, and FMDF, is fitted to the sample 1988Q2 to 2020Q1 using four quarterly lags. The Granger-causality results are qualitatively similar when estimating a monthly VAR using only the three monthly variables. The upper right graph reports the factor loadings associated with the two first principal components of the residuals from the VAR used to calculate the Granger causality statistics. Using the same VAR, the two lower graphs report the historical shock decomposition of the RNI and NFCI under the assumption that the two variables are only driven by their own unexpected innovations. Orthogonal innovations are obtained by assuming a simple recursive identification scheme and ordering either the RNI or NFCI last in the system.

2.3 Statistical properties

The mean estimate of the RNI is plotted together with the NFCI, FMDF, and quarterly GDP growth, in Figure 1 in Section 1. The RNI tend to reach high levels at least one year prior to crisis episodes, but it is also somewhat volatile, at least compared to the NFCI. Moreover, while the FMDF and NFCI are highly correlated, particularly around crisis episodes, the RNI seems to be leading. In terms of explaining the conditional mean, this is more formally confirmed by the Granger-causality statistics presented in the upper left part of Figure 2. The FMDF Granger-causes the NFCI but the NFCI does not Granger-cause the FMDF. The RNI, however, Granger-causes the other indicators whereas neither of the other series Granger-causes the RNI.

The results presented in the upper right part of Figure 2 speaks to the critical review by HPRR, who convincingly show that the information contained in the NFCI is encompassed by the information already contained in real indicators, such as the FMDF. In particular, when calculating the two first principal components of the reduced form residuals from the VAR used to calculate the Granger causality statistics, we find that the FMDF, NFCI, and GDP, load strongly on the first component whereas the RNI loads strongly on the second component. In other words, the NFCI and the FMDF seem to contain similar contemporaneous information, as argued in HPRR, whereas the RNI is fairly exogenous to contemporaneous innovations in both the FMDF and NFCI.

An alternative illustration of this finding can be obtained by assuming a simple recursive identification structure to orthogonalize the VAR innovations, ordering either the RNI or the NFCI last in the system, and then computing the historical shock decompositions. This is done in the lower graphs in Figure 2, and as seen there, most of the variation in the RNI is explained by its own innovations. In contrast, and in line with the argumentation in HPRR, the historical evolution of the NFCI is at times heavily affected by innovations to the other variables in the system.

2.4 Understanding the RNI in vector space

Since word embeddings represent words as vectors, distance metrics, such as cosine similarity, are commonly used to perform intrinsic evaluation of the estimated embeddings. The word scatter plot in Figure 3, produced using the t-SNE algorithm (Van der Maaten and Hinton, 2008), provides an example of this type of evaluation, where words colored red and green report the most similar words to the two aggregated vectors from equation (2). For ease of interpretability we focus on one particular year, 2007, and the average embeddings across months for this year. Figures 4 and A.7, in Appendix A, reports the 30 most related words in a more readable format and across six different years. For comparison, we also plot in Figure 3 the most similar words to three alternative aggregate embeddings defined such that $\textbf{monetary policy}_t = \boldsymbol{monetary}_t + \boldsymbol{policy}_t$, $\boldsymbol{pandemic}_t = \boldsymbol{virus}_t + \boldsymbol{pandemic}_t$, and $\textbf{policy uncertainty}_t = \boldsymbol{economic}_t + \boldsymbol{policy}_t + \boldsymbol{uncertainty}_t$. These alternative concepts are distinct from the financial stress concept typically used in the GaR literature, but might all be important for understanding recession periods in our sample.[7]

While one should be cautious about drawing strong conclusions regarding potential cluster sizes and distances in this type of plot, the analysis delivers at least three relevant insights. First, words that are closely related to the words used to define the different concepts tend to cluster together. For example, both the "Pandemic" and "Financial conditions" terms form distinct clusters, signaling that these are very different concepts. The "Monetary policy" and "Policy uncertainty" terms, on the other hand, overlap to some extent because they are more related. Thus, although the concepts above share many words, they form distinct clusters in Figure 3 because the aggregated embeddings capture associative meaning and shared context.[8]

Second, as seen from Figures 4 and A.7, the most similar words to the aggregate

[7] We do not claim that these alternative constructions are the best possible approximations to the different concepts. They are mainly used for illustrative purposes.

[8] The "purity" of the clusters visualized in Figure 3 are not a result of the t-SNE algorithm. Using K-means clustering to estimate four clusters from the combined embedding matrix shows that the unique concept terms in Figure 3 typically are allocated to distinct clusters (Table A.7 in Appendix A).

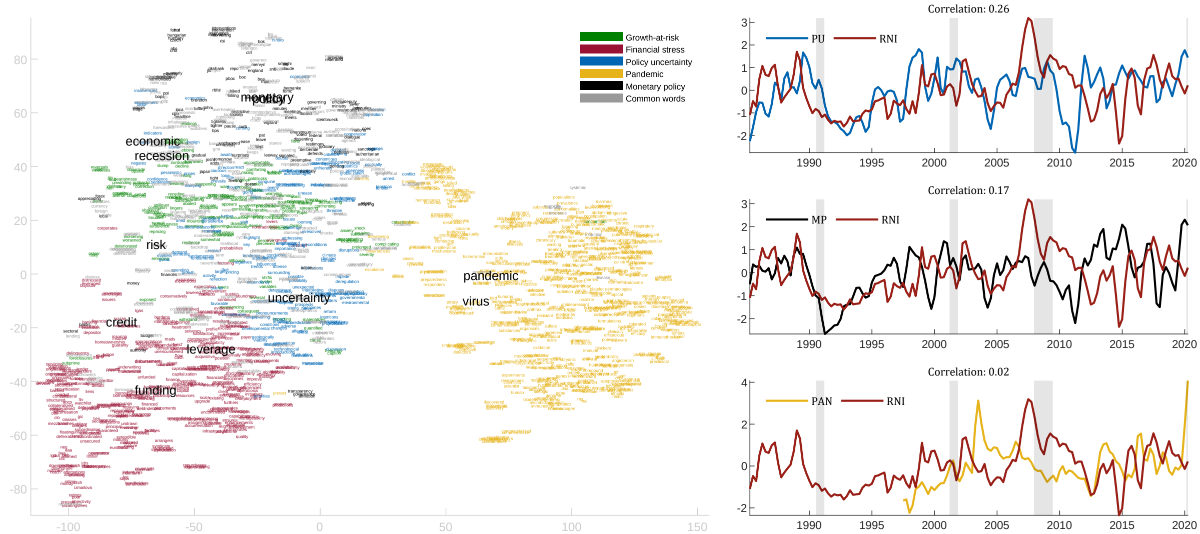


Figure 3. The left graph reports a scatter plot of the words related to the five concepts listed in the legend. See the text and Appendix A.3 for a description of how the graph is constructed. The right graph reports the RNI together with the implied time-series for the three alternative concepts. The alternative time-series are computed using the same methodology as described in Section 2.2, but now replacing the **financial conditions**$_t$ vector by one of the alternatives defined in the text.

embeddings in (2) are fairly stable across time. Moreover, these wordclouds suggest that the aggregated vectors indeed point in the intended directions in vector space. E.g., among the most similar embeddings to the **growth-at-risk**$_t$ vector in 2006 include words such as "vulnerable", "slowdown", and "downturn".

Third, the results reported in the Figure 3 might suggest that "Growth-at-risk" is more related to "Monetary policy" and "Policy uncertainty" than "Financial conditions". This can be true. A number of studies have highlighted the role of (systematic) monetary policy and (general) uncertainty for growth (see, e.g., Berger et al. (2020), Caldara et al. (2016), Jurado et al. (2015), Brunnermeier et al. (2021)). However, the question we address is not how to best predict the conditional distribution of GDP growth, but, in light of theories linking financial conditions to growth vulnerabilities and the critical review by HPRR, whether an alternative measure of financial conditions adds value relative commonly used indicators, such as the NFCI.

That said, one cannot rule out that news coverage of, e.g., monetary policy or (general) uncertainty, feeds into the RNI which in turn predicts GaR. If this was the case, however, we would expect to see a strong correlation between movements in the RNI and time series analogs of the alternative concepts. The right graph in Figure 3 suggests that this is not the case. Although we observe co-movement in certain time periods, the overall correlation is relatively low. For the "Pandemic" time series this finding is particularly clear. It starts late because of missing data, spikes around the SARS and COVID-19 pandemics in 2003 and 2019, but has very little overall correlation with the RNI. Moreover,

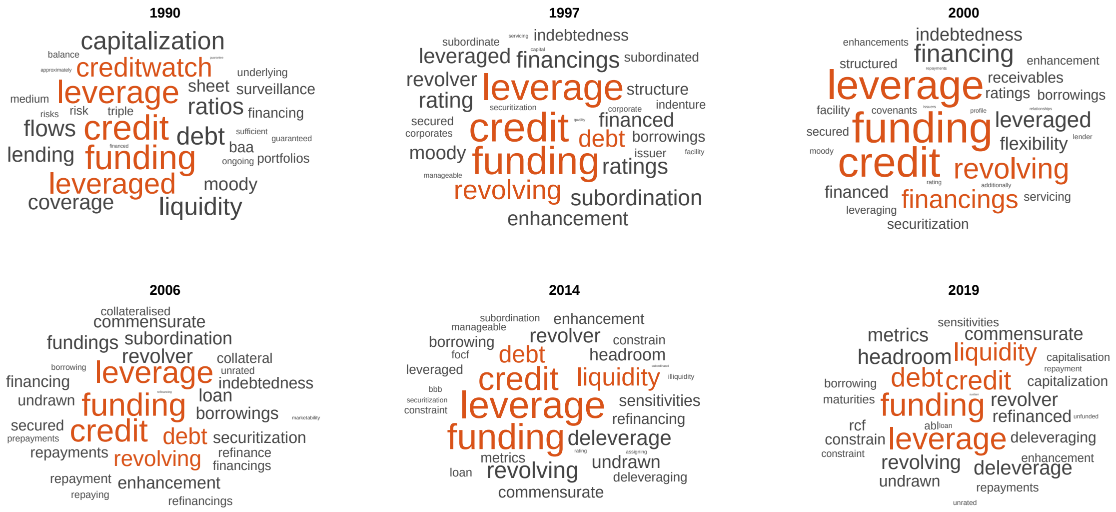


Figure 4. The figure reports a wordcloud of the 30 most similar words to the **financial conditions**$_t$ vector at six different points in time. Similarity is measured using the cosine similarity metric. Vectors are averaged across months of the year prior to computing their distance.

when doing the out-of-sample forecasting experiment, described in Section 3.2, using the alternative "Monetary policy" and "Policy uncertainty" indicators, we find that they are outperformed by the Skew-t(RNI) model (Tables A.2 and A.3 in Appendix A).

3 GDP and risky news

For characterizing and predicting the conditional GDP distribution we use a parametric Skew-t distribution with time-varying location, scale, and shape parameters. This type of model is often used in the recent GaR literature, where quarterly real GDP growth (Δy_t) can be written as

$$\Delta y_t = \mu_t + v_t, \quad v_t \sim Skt(0, \sigma_t, \alpha_t, \nu), \tag{3}$$

and the Skew-t specification follows Arellano-Valle et al. (2005). I.e., the location (μ_t), scale (σ_t), shape (α_t), and degrees-of-freedom (ν) parameters define the distribution's conditional mean, variance, asymmetry (skewness), and the tickness of the tails.

Here, following recent research by Delle Monache et al. (2021) and Labonne (2022), the time-varying parameters are assumed to evolve according to score-driven processes containing a persistent trend component and a transitory component.[9] Using link func-

[9]Score driven methods, introduced by Creal et al. (2013) and Harvey (2013), provide an efficient way of modeling dynamic distributions by using the conditional score of the likelihood as the driving force for the model's time-varying parameters. The likelihood of score driven models are generally available in closed form because they are observation-driven, making maximum likelihood an efficient estimation approach (Blasques et al., 2022). Koopman et al. (2016) show that they provide comparable forecasting performances to nonlinear non-Gaussian state space models. See Delle Monache and Petrella (2017), Gorgi et al. (2019) and Delle-Monache et al. (2020) for recent forecasting applications.

tions, such that $\gamma = log(\sigma)$ and $\varrho = arcthan(\alpha)$, and letting $u_t \in \{\mu_t, \gamma_t, \varrho_t\}$, the sum of the two components is $u_t = \tau_{u,t} + \upsilon_{u,t}$ with

$$\tau_{u,t} = \tau_{u,t-1} + \varsigma_u s_{u,t} \quad \upsilon_{u,t} = \phi_u \upsilon_{u,t-1} + \beta_u x_{t-1} + \kappa_u s_{u,t}, \tag{4}$$

where x is an exogenous predictor, e.g., the quarterly RNI or NFCI, and ς_u and κ_u denote the gain parameters giving weight to the time-varying scores $s_{u,t}$ of the predictive likelihood. Thus, parameter updates are defined by the steepest ascent direction for improving the model's local fit, and the trajectories of the time-varying parameters are perfectly predictable given past information.

The model's unknown parameters are estimated using maximum likelihood (ML) where the log density at time t takes the form

$$\log p(\Delta y_t|Y_{t-1}) = -\log C(\eta) - \log\sigma_t^2 - \frac{\eta+1}{2\eta}\log\Big[1 + \frac{\eta(y_t - \mu_t)^2}{(1 - sgn(y_t - \mu_t)\alpha_t)^2\sigma_t^2}\Big], \tag{5}$$

with $\sigma_t > 0$, $\alpha_t = [-1, 1]$, $\eta = 1/\nu$ with $\nu > 1$, and $C(\nu) = \Gamma((\nu+1)/2)/(\sqrt{\nu\pi}\Gamma(\nu/2))$ where $\Gamma(.)$ is the Gamma function. The distribution is skewed towards positive values if $\alpha_t < 0$ and towards negative values if $\alpha_t > 0$. When the tail parameters are constrained to be very large and the shape parameter set to zero the distribution is equivalent to a normal distribution.

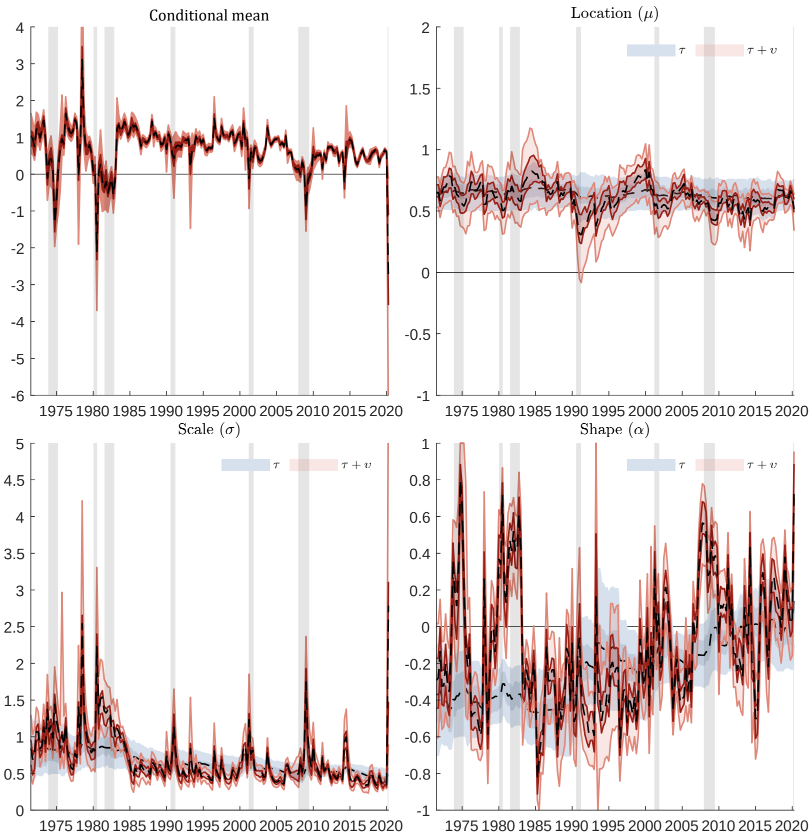
As shown in Delle Monache et al. (2021), using the NFCI and the model specification in (3) and (4) produce significantly better out-of-sample predictions for U.S. GDP growth than those obtained in the original work by Adrian et al. (2019) and even the more critical review by HPRR. Indeed, this specification nests the one used in HPRR, where the evolution of the time-varying moments are driven only by the variation in the exogenous variables. Accordingly, using (3) and (4) together with the NFCI is a natural and strong benchmark for assessing the potential value added of the RNI.[10]

In the following we present in- and out-of-sample results using the Skew-t model in conjuncture with either the RNI or the NFCI. For completeness, we also present results documenting how the RNI performs relative to different credit aggregates often used in the growth vulnerability literature focusing on more long-run growth outcomes. We then discuss how the word embedding approach adds value relative to related NLP approaches in this context.

3.1 In-sample evidence

Figure 5 summarizes how the Skew-t(RNI) model characterizes GDP growth in-sample. To be able to compare our estimates with findings in the existing literature, such as Adrian

[10] In the interest of preserving space we refer to Delle Monache et al. (2021) and Labonne (2022) for detailed expositions of this model class and specification. We note here that we depart (slightly) from the modeling choices used in Delle Monache et al. (2021) along two dimensions. These are described in Appendix A.2.



Parameter estimates

	Skew-t(RNI)	Skew-t(NFCI)
ϕ_μ	0.912 (0.556, 1.211)	0.001 (-29.662, 24.896)
ϕ_γ	0.509 (0.166, 0.798)	0.564 (0.352, 0.741)
ϕ_ϱ	0.115 (-1.477, 1.452)	0.001 (-23.970, 20.118)
ς_μ	0.000 (-0.000, 0.000)	0.000 (-0.000, 0.000)
κ_μ	0.041 (-0.002, 0.078)	0.000 (-0.001, 0.001)
ς_γ	0.030 (-0.006, 0.060)	0.000 (-0.000, 0.000)
κ_γ	0.243 (0.123, 0.345)	0.117 (0.016, 0.202)
ς_ϱ	0.031 (0.003, 0.055)	0.038 (0.016, 0.056)
κ_ϱ	0.136 (0.042, 0.216)	0.117 (0.016, 0.202)
β_μ	-0.000 (-0.000, 0.000)	-0.198 (-0.430, -0.004)
β_γ	0.087 (-0.001, 0.161)	0.113 (0.026, 0.186)
β_ϱ	0.293 (0.121, 0.437)	0.253 (0.089, 0.391)
T	196	196
LL	-187,532	-180.138

Figure 5. The graphs report the estimated time-varying moments, together with 50% and 95% confidence bands, obtained from the Skew-t(RNI) model. The table reports the parameter estimates from both the Skew-t(RNI) and Skew-t(NFCI) models. Values in parenthesis are 95% confidence bands.

et al. (2019), HPRR, and Delle Monache et al. (2021), the model is estimated using data from 1973:Q3 to 2020:Q2. Since the RNI is only available from 1985:Q1, we construct synthetic RNI observations using its correlation with the NFCI, which is available for the whole sample. We do so by first regressing the RNI on eight leads and lags of the NFCI, and then use the resulting parameter estimates together with the NFCI to construct the values for the missing RNI history.[11]

As seen in the figure, the time-varying moments are rather precisely estimated and vary significantly over time. Economic expansions tend to be characterized by positive skewness, which is equivalent to a negative shape parameter in the figure, and there is a positive correlation between the mean and variance of the conditional GDP distribution during such expansions. In contrast, during recessions, the correlation between the mean

[11] Figure A.3, in Appendix A, shows that our main conclusions regarding the forecasting performance of the Skew-t(RNI) relative to the Skew-t(NFCI) model, reported in Section 3.2, remain robust to starting estimation in 1985. We do, however, find that the in-sample characteristics of GDP growth using the Skew-t models deteriorate when a shorter sample is used. Thus, the high variance and crisis episodes in the 1970s and 1980s seem important when fitting the Skew-t model to the data.

and variance turns negative. While these results are not new, and very similar to those from the Skew-t(NFCI) model in Delle Monache et al. (2021), they continue to hold when using the RNI as the exogenous variable in the model.[12]

The contribution of the RNI index to these time-varying patterns is illustrated in the table to the right in Figure 5. An increase in the RNI does not significantly affect the location of the GDP distribution, has a borderline significant effect on its variance, but significantly increases the asymmetry of the distribution. Moreover, the persistence of the transitory component is large and significant for the location, but insignificant for the shape, while the gain estimates are not significant for the location processes, but generally significant for both the scale and shape processes. One implication of this, as also seen in Figure 5, is that the trend component of the location process is close to constant, whereas the trend components of the shape and scale processes gradually declines and increases, respectively, over the sample.

The second column of parameter estimates reported in Figure 5 are obtained when re-estimating the model using the NFCI as an exogenous variable. In general, the parameter estimates are very similar to those obtained with the Skew-t(RNI) model, with two important exceptions. The NFCI contributes more towards describing the location of the GDP distribution and the persistence of the transitory component for this moment is small and insignificant. From a likelihood perspective, the Skew-t(NFCI) model is actually preferred relative to the Skew-t(RNI) model. However, the differences are small, and we fail to reject the null hypothesis of equal fit when conducting a likelihood ratio rest.

In relation to the study by HPRR, an important argument in their analysis is that the time-varying moments of the conditional GDP distribution are very imprecisely estimated and that the NFCI contributes mostly towards explaining the conditional mean of the distribution. In line with the results reported in Delle Monache et al. (2021), neither of these conclusion are supported by the Skew-t(RNI) model used here. In particular, Figure 5 illustrates well how the conditional mean varies significantly across time, and that most of this variation stems from changes in the shape of the distribution and not the location.

3.2 Out-of-sample evidence

To evaluate the out-of-sample performance of the Skew-t(RNI) model we perform a quasi-real-time forecasting experiment, and compare its performance with the Skew-t(NFCI) model. The models are first estimated from 1973:Q2 to 1993:Q1, and predictions are constructed for horizons one-quarter- to one-year-ahead. This process is then repeated

[12]See, e.g., Smith and Vahey (2016), Salgado et al. (2019), Carriero et al. (2020), and Fernández-Villaverde and Guerrón-Quintana (2020) for related findings about how the correlations between higher order moments vary with the business cycle.

Table 1. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(NFCI) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(NFCI) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	-0.25 (0.46)	1.01 (0.53)	1.18 (0.30)	-0.08 (0.49)	0.93 (0.53)	0.91 (0.50)
Rec.	-1.07 (0.17)	1.12 (0.38)	1.42 (0.53)	-0.69 (0.01)	0.90 (0.27)	0.96 (0.22)
GFC	-1.64 (0.19)	1.12 (0.71)	1.48 (0.66)	-1.32 (0.00)	0.66 (0.00)	0.53 (0.00)

for the remaining sample using an expanding estimation window.[13]

For each quarterly forecasting vintage we assume that both the RNI and NFCI are available (at the last day of the quarter). Thus, one-quarter-ahead predictions are obtained directly from the model and the implied parameter estimates. For longer horizons forecasts we make the same assumption as in Delle Monache et al. (2021) and condition the predictions on the last available predictor values, i.e., assume their future values are approximated well by a random walk prediction. For multi-step ahead forecasting horizons the time-varying parameters are subject to additional uncertainty coming from the unobserved scores. To integrate this uncertainty into our forecasts we use the "bootcasting" algorithm of Koopman et al. (2017) which consists of sampling unobserved (scaled) scores from their in-sample realizations (see also Delle Monache et al., 2021).

To evaluate the predictions we use three different scoring metrics; the log-score (LS), the Continuously Ranked Probability Score (CRPS), and Root Mean Forecasting Errors (RMSE). To test for significant differences in predictive performance we use the conditional Giacomini and White (2006) test. We also evaluate the calibration of the predictive densities using the Probability Integral Transforms (PITs) and the test suggested by Rossi and Sekhposyan (2019).

Figure 6 and Table 1 summarize our main results. The figure reports the PITs and the cumulative differences in LS, CRPS, and squared forecast errors between the Skew-t(RNI) and Skew-t(NFCI) models for one-quarter- and one-year-ahead predictions. All relative scores are normalized such that an increase implies a relative improvement of the Skew-t(NFCI) model. The Skew-t(NFCI) model is better on the one-quarter-ahead horizon for CRPS and RMSE loss, while the Skew-t(RNI) model is better according to

[13] The RNI is not subject to historical revisions as new data vintages are made available, while both the GDP statistics and the NFCI are. For this reason we conduct a quasi-real-time experiment using the final vintages (of GDP and the NFCI). Amburgey and McCracken (2023) show that using the real-time estimates of the NFCI rather than the final vintage actually performs better in terms of predicting GaR. Table A.6, in Appendix A, shows that re-doing the experiment using real-time vintages of both GDP and NFCI does not alter our main conclusions.

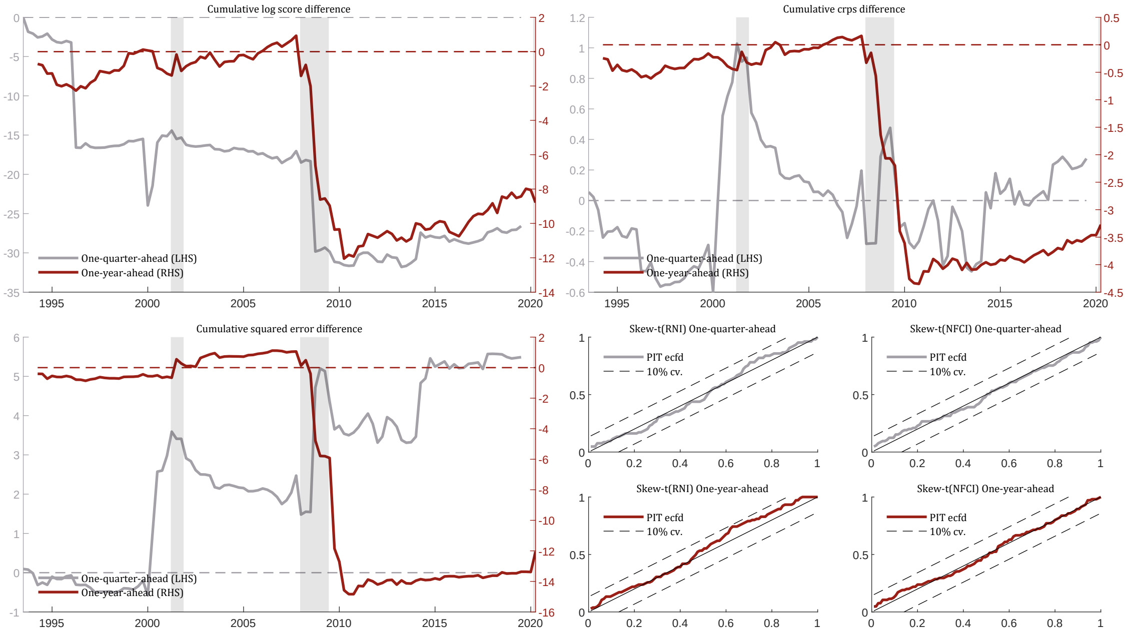


Figure 6. The figure reports cumulative differences in LS, CRPS, and squared forecast errors between the Skew-t(RNI) and Skew-t(NFCI) models for one-quarter- and one-year-ahead predictions. All relative scores are normalized such that an increase implies a relative improvement of the Skew-t(NFCI) model. The lower right quadrant reports the empirical cumulative distribution function of the PITs together with the Kolmogorov–Smirnov class of test statistic proposed by Rossi and Sekhposyan (2019).

LS. However, none of these differences are significant when considering the whole out-of-sample period, or when focusing on recession periods or the GFC in particular (Table 1). In contrast, at the one-year-ahead horizon the Skew-t(RNI) model clearly outperforms the Skew-t(NFCI) model. For the RMSE loss the performance gain is roughly 10% when considering the full sample. Most of this gain is driven by recession periods and the GFC, and when focusing only on such episodes the differences in predictive performance are also generally significant at the 5% level.[14]

The lower right quadrant of Figure 6 reports whether the conditional predictive densities are correctly specified using the PITs and the Kolmogorov–Smirnov class of test statistic proposed by Rossi and Sekhposyan (2019). Under the null hypothesis the predictive density is correctly specified. In this case the empirical cumulative distribution function of the PITs should follow closely the 45 degree line. As seen in the figure, there are signs that the Skew-t(RNI) model has too many realizations falling in the center of the distribution at the one-year-ahead horizon (the slope in the center is steep relative to the 45 degree line).However, for both models and forecasting horizons, we can not reject the null hypothesis of correctly specified predictive densities.

The strong performance of the Skew-t(RNI) specification, particularly during reces-

[14]For completeness, Table A.1, in Appendix A, shows that these results are qualitatively similar if we instead compare the Skew-t(RNI) model to a Skew-t(FMDF) model.

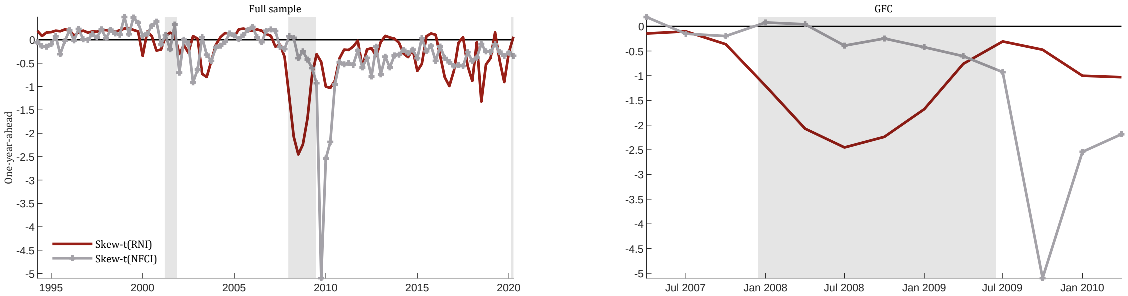


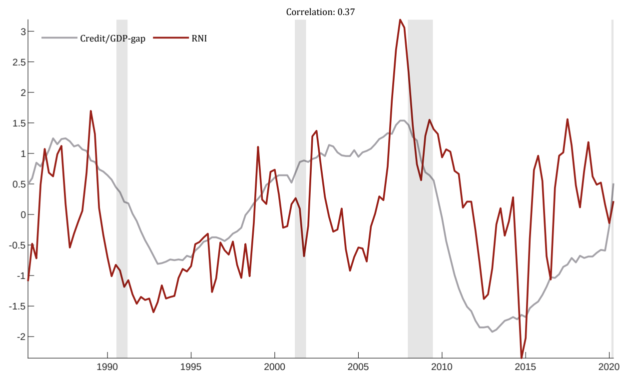
Figure 7. The figure reports the one-year-ahead GaR predictions from the Skew-t(RNI) and Skew-t(NFCI) models. The figure to the left shows the predicted GaR for the whole out-of-sample evaluation periods while the figure to the right focuses on the GFC period.

sions and the GFC episode, is not only statistically significant but also economically relevant. As seen from the density plots in Figure 1 in Section 1, in every quarter of 2008 the Skew-t(RNI) model predicted considerably larger down-side risks than the Skew-t(NFCI) model. This stands in sharp contrast to similar plots in HPRR, where there are essentially no differences between predictions based on either real or financial variables. An alternative illustration of this is given in Figure 7, which reports the predicted GaR across the whole out-of-sample evaluation period and the GFC in particular. For example, one year prior to 2008:Q3, the Skew-t(RNI) model predicted that growth at the lower fifth percentile of the growth distribution would be -2.5 whereas the Skew-t(NFCI) prediction hardly had turned negative.

3.3 Credit aggregates and the RNI

It is well known in the macro-finance literature that different financial variables have heterogeneous dynamics along the business cycle. The buildup of (growth) vulnerabilities are typically associated with excessive leverage and credit growth (e.g., increasing credit gaps) while the outbreak of financial distress is foremost associated with external financial premiums (e.g., widening credit spreads) (Gertler et al., 2020). Now, the NFCI is typically seen as a financial distress indicator, but a prominent set of work has documented how vulnerability indicators have strong predictive power, especially for more long-run growth outcomes (Jordà et al., 2011, 2013; Reinhart and Rogoff, 2009; Schularick and Taylor, 2012).

Figure 8 reports the RNI together with a widely used measure in this literature, namely the credit-to-GDP gap (CGG) produced by the Bank for International Settlements (BIS). Although the correlation between these two series is far from insignificant, and they share important low-frequency characteristics, the results presented in the table in Figure 8 suggest that the RNI outperforms the CGG for the out-of-sample growth predictions considered here. That is, re-doing the out-of-sample analysis with a Skew-t(CGG) model



	One-quarter-ahead		
	LS	CRPS	RMSE
Full	-0.22 (0.17)	0.93 (0.42)	0.96 (0.48)
Rec.	0.04 (0.20)	1.02 (0.23)	1.12 (0.54)
GFC	-0.18 (0.26)	0.95 (0.22)	1.05 (0.50)
	One-year-ahead		
Full	-0.17 (0.24)	0.92 (0.19)	0.92 (0.23)
Rec.	-0.78 (0.06)	0.90 (0.00)	0.95 (0.00)
GFC	-1.35 (0.00)	0.70 (0.00)	0.61 (0.00)

Figure 8. The left graph reports the quarterly CGG and RNI. The table reports the forecasting performance of the Skew-t(RNI) relative to the Skew-t(CGG) model. See Table 1 for further details.

we find that the Skew-t(RNI) model performs significantly better at the one-year-ahead horizon.

Figure A.9, in Appendix A, reports the correlation between the RNI and five other credit-to-GDP aggregates often used in the literature, and documents that it varies from 0.22 (for credit to households from all sectors) to 0.61 (for credit to the private non-financial sector from banks). As we show below, however, irrespective of which measure we use, the qualitative conclusions regarding the value added of the RNI remain robust.[15]

3.4 Is it the method or the data?

The word embedding methodology we propose is supervised along two dimensions. First, the RNI results from minimizing the errors from projecting a financial conditions vector on a growth-at-risk vector. Second, the keywords used are subjectivity chosen (motivated by economic theory linking growth-at-risk to financial conditions).

The second supervised feature makes the methodology related to two alternative approaches often used in economics, namely count- and Boolean-search-based methods. Count-based methods derive indexes of the subject of interest by simply counting the frequency of terms in the corpus related to a set of predefined keywords. Boolean-search-based methods share this count feature, but focus on the co-occurence of terms and register counts only when a set of logical conditions are met, e.g., if an article contains the words "recession" *and* "risk".

Count- and Boolean-search-based methods do, however, not share the first supervised feature of the word embedding methodology. Therefore, it is far from guaranteed that a

[15]From a purely statistical point of view, one reason for the value added of the RNI in terms of capturing business cycle fluctuations is that it contains more high frequency variation compared to the various credit-to-GDP gaps. See Figure A.8 in Appendix A.

high count actually reflects a real relationship between words or simply reflect the chosen unit of observation, e.g., sentence, article, or day. I.e., if conducting Boolean-search at the article level with a given set of keywords, the end results might potentially be very few counts simply because the likelihood of all words occurring in the same article is low. Word embeddings, on the other hand, are trained to capture shared context between words, implying that words never occurring in the same unit of observation (e.g., article) might still stand close to each other in vector space because they share context.[16] Moreover, since word embedding algorithms represent words as vectors, standard regression analysis can be used to measure the degree of association between different concepts, which is exactly what we do.

It is of course an empirical question whether the extra complexity introduced by the word embedding methodology adds value. For this reason we re-do the out-of-sample experiment for both a Skew-t(Count) and Skew-t(Bol) model. The alternative "Count" and "Bol" indicators are constructed using standard procedures and the same keywords defined in Section 2.2. In the interest of preserving space, a detailed description of how this is done is relegated to Appendix A.4. Figure A.4, in Appendix A, reports the alternative indexes together with the RNI, and Tables A.4 and A.5 report the out-of-sample results. In short, both the "Count" and "Bol" indicators tend to spike around recession periods and is at times highly correlated with the RNI. In terms of out-of-sample performance we find no significant differences at the one-quarter ahead horizon. However, as before, the Skew-t(RNI) model tend to provide significantly better predictions than both the Skew-t(Count) and Skew-t(Bol) models at the one-year-ahead horizon, especially when considering recession periods and the GFC.

Still, although the Skew-t(RNI) model outperforms the two news-based alternatives, Figures A.5 and A.6, in Appendix A, illustrate that using the alternative news-based indicators deliver predictions that for some horizons and score metrics are competitive, or even better, than those produced using the Skew-t(NFCI) model. This suggests that the informational content of the news plays an important role. In the next section we put further light on the mechanisms that might give rise to this finding.

[16] This property stems from the fact that the word embedding algorithm itself can be labeled semi-supervised because running text is used as implicit supervised training when estimating the embeddings. This avoids the need for any hand-labeled supervision signal and makes the methodology flexible and user-friendly in many contexts.

4 Why is the RNI so informative?

From a theoretical perspective there are at least two views linking financial conditions to economic downturns. In the classical theories motivating much of the growth vulnerability literature, borrowers and lenders are typically seen as fully rational but subject to various forms of credit limits or collateral constraints (Gertler and Bernanke, 1989; Kiyotaki and Moore, 1997; Bernanke et al., 1999). In turn, these financial market frictions play a central role in propagating and amplifying shocks to the economy, making economic downturns preceded by excessive credit growth and leverage especially severe. In essence, credit booms go bust when unexpected (exogenous) shocks happen.

An alternative literature emphasizes the role of sentiment for the credit cycle. This view draws on recent work in behavioral finance and on classic accounts of financial crises by Minsky (1977, 1986) and Kindleberger (1978), and asserts that predictable time variation in the sentiment of credit-market investors, due to, e.g., "diagnostic expectations" and extrapolative beliefs of investors, is an important determinant of the credit cycle (Greenwood et al., 2016; Bordalo et al., 2018). The empirical work by López-Salido et al. (2017), henceforth denoted LSZ, connects this to the business cycle and associates credit booms not with balance-sheet measures of credit growth but with proxies for the expected returns on credit assets. They then show that the economy performs poorly following periods when these proxies are unusually low by historical standards, i.e., when buoyant sentiment unwinds. In contrast to the balance-sheet-based view, no exogenous shock is needed. Once asset prices are significantly elevated an economic correction is closer at hand because overly optimistic investors will be disappointed going forward.

In principle, the RNI can speak to both of these views and capture changes in fundamental information as well as sentiment and beliefs. Therefore, to better understand the informational content of the RNI and the underlying mechanism that gives rise to it's strong predictive performance, we proceed in two complementary directions. First, we build on Enders et al. (2021), henceforth EKM, and estimate time series reflecting unexpected fundamental information and changes in beliefs, and then correlate these time series with the RNI. Second, we use credit market valuation indicators and the methodology proposed in LSZ to construct credit market sentiment approximations for the quarterly data and sample considered here, and then relate these to the business cycle and the RNI.

4.1 SVAR evidence

EKM propose a simple, but powerful, SVAR methodology for identifying aggregate shocks to expectations (belief shocks) from other fundamental shocks. Their basic premise is that

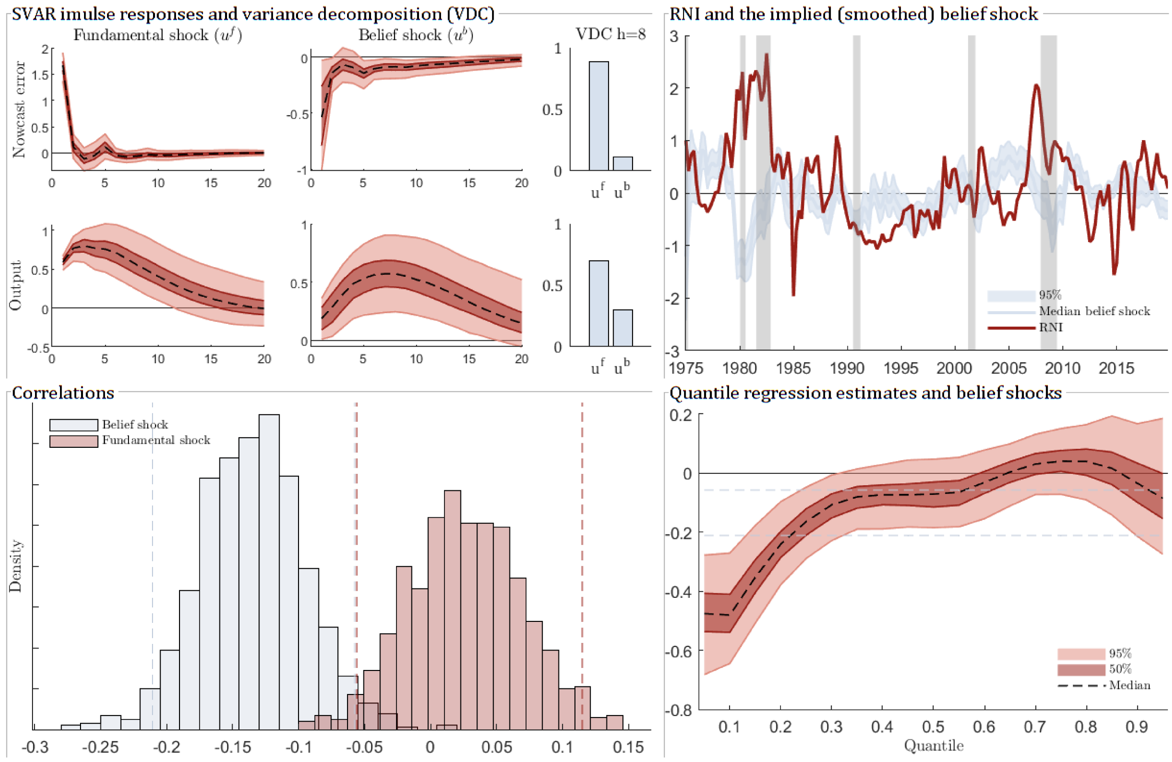


Figure 9. The upper left graphs report the impulse responses, medians with posterior uncertainty, to a one standard deviation fundamental or belief shock. The y-axis denotes the response (in percent) and the x-axis the response horizon. The corresponding variance decompositions are reported only for one horizon. The upper right graph reports the RNI together with the posterior distribution of u_t^b. For visual clarity, the belief shocks are smoothed using a three quarter moving average filter. The lower left graph reports the correlation coefficient between the RNI and (every draw) of either the (not smoothed) u_t^b or u_t^f. The broken vertical lines are 95 percent posterior intervals. The lower right panel shows the coefficient estimates for RNI_t from a standard quantile regression model using the draws of u_t^b as dependent variable.

any fluctuations in the (ex-post available) nowcast error that cannot be explained by new fundamental information has to be attributed to changes in beliefs. And, in turn, such belief shocks move output and the nowcast error in opposite directions. I.e., waves of optimism (or pessimism) have causal effects on economic outcomes. Below we replicate their analysis for the data and sample considered here.

More formally, let $E(\Delta y_t|I_t)$ denote the (median) nowcast of real GDP growth obtained from the Survey of Professional Forecasters (SPF) and define $ne_t = \Delta y_t - E(\Delta y_t|I_t)$ as the nowcast error. Then, following EKM, the VAR we consider can be written as:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim i.i.d.N(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), \tag{6}$$

where $\boldsymbol{y}_t = (ne_t, y_t)'$ is a vector of the endogenous variables, $\boldsymbol{x}_t = (1, tr, \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}')'$ a vector containing a constant, a quadratic time trend, and $p = 4$ lags of $\boldsymbol{y}$, $\boldsymbol{\Phi}$ is a matrix of coefficients, and $\boldsymbol{\varepsilon}_t$ a vector of reduced form errors. The relationship between the reduced form errors and the structural shocks is given by $\boldsymbol{\varepsilon}_t = \boldsymbol{A}\boldsymbol{u}_t$. Because $(\boldsymbol{A}\boldsymbol{u}_t)(\boldsymbol{A}\boldsymbol{u}_t)' = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, identification is achieved by assuming that $E(\boldsymbol{u}_t\boldsymbol{u}_t') = \boldsymbol{I}$ and by restricting the elements of $\boldsymbol{A}$ such that the following relationship holds:

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} = \begin{bmatrix} + & - \\ + & + \end{bmatrix} \begin{bmatrix} u_t^f \\ u_t^b \end{bmatrix} \tag{7}$$

As in EKM, u_t^f is interpreted as a fundamental shock moving the nowcast error and output in the same direction. In contrast, a positive belief shock u_t^b leads to a negative nowcast error. But, because expectations are positive, output rises although not as much as expected.

For inference, we consider the sample 1975:Q1 to 2019:Q4, and estimate the model assuming a normal prior distribution for $\boldsymbol{\Phi}$ and an inverse Wishart distribution for $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. A total of 100 thousand draws are obtained from the conditional posterior distributions via Gibbs sampling, and thousand draws are retained for inference after thinning the chain. The prior means and covariances are set according to the popular Minnesota scheme (Litterman, 1986), and we use the algorithm proposed in Rubio-Ramirez et al. (2009) for sign identification (of $\boldsymbol{A}$).

Figure 9 reports the results. The upper left quadrant shows the impulse responses and the variance decompositions (for one horizon) from the SVAR analysis. Both in qualitative and quantitative terms these results are very similar to those reported in EKM: Following a positive fundamental shock we observe delayed updating via the nowcast error response and a fairly persistent, but gradually trend-reverting, output response. In contrast, a positive belief shock which drives down the nowcast error leads to a more u-shaped output response. I.e., the initial response is positive but considerably smaller than the peak response occurring after roughly eight quarters.

The results reported in the upper right graph are new to this study, and illustrates the RNI together with the estimated belief shock, i.e., the posterior estimates of $\hat{u}_t^b$. The lower left graph reports a histogram of the correlation between these two series, when every draw of $\hat{u}_t^b$ is considered. As clearly seen in these graphs, the RNI has a significant and negative correlation with a positive belief shock. In contrast, when computing the correlation between all the draws of $\hat{u}_t^f$ and the RNI, we obtain a much weaker and insignificant correlation. Finally, in the lower right graph we use a simple quantile regression to evaluate the relationship between the belief shock and the RNI, and find that the RNI is in particular negatively correlated with the lower tail of the belief shock.

4.2 Valuation indicators evidence

LSZ provides an alternative framework for identifying credit-market sentiment more directly from market prices and proxies for expected returns. The econometric framework used is the following two-step regression specification

$$\Delta_4 s_t = \boldsymbol{\theta}' \boldsymbol{z}_{t-p_1} + \upsilon_{1,t} \tag{8a}$$

$$\Delta_4 y_t = \beta_1 \Delta_4 \hat{s}_t + \boldsymbol{\gamma}' \boldsymbol{x}_{t-p_2} + \upsilon_{2,t}. \tag{8b}$$

In the first-step auxiliary regression (8a) the year-on-year change in Moody's Baa-Treasury credit spread ($\Delta_4 s_t$) is regressed on a vector $\boldsymbol{z}_{t-p_1}$ containing the lagged high-yield bond issuance (HYS_{t-p}), expressed as a share of total bond issuance, and the level of the credit spread (s_{t-p}). As documented in Greenwood and Hanson (2013), these valuation indicators are strongly associated with negative expected returns at horizons ranging from one up to three-years-ahead. In the second step regression (8b) the fitted values from the auxiliary regression ($\Delta_4 \hat{s}_t$) are used to predict business cycle fluctuations conditional on $\boldsymbol{x}_{t-p_2}$. Here, $\Delta_4 y_t$ is computed as year-on-year changes in current quarter (log) GDP and $\boldsymbol{x}_{t-p_2}$ contains lags of $\Delta_4 y_t$ and the CGG.[17]

The first row in the upper left panel in Figure 10 replicates the analysis in LSZ using our sample and data frequency, and three different lag structures for the valuation indicators and the CGG. As in LSZ, lagged growth is kept fixed at $t-4$ in the second stage regression. Although somewhat imprecisely estimated, the qualitative conclusions from their study hold: Following a period of aggressive credit market pricing, i.e., an elevated level of the high-yield share and narrow (below average) spreads, the component of credit-spread changes that is driven by a reversal of prior positive sentiment ($\Delta_4 \hat{s}_t$) has a negative impact on economic activity.

To analyze to what extent the RNI also contains this sentiment component, we follow the intuition in LSZ and "clean" it for news about fundamentals by re-estimating the model in (8) using the RNI as dependent variable in the first-step auxiliary regression. The second row in the upper left panel of Figure 10 reports the results. Consistent with the sentiment-based credit market view, the RNI contains a predictable component positively related to lagged values of high-yield bond issuance and negatively related to lagged credit spreads (at least for lag structures larger than two years), and this predicable component has a negative impact on the business cycle. In terms of significance, we actually find that the HYS is a more accurate predictor of the RNI than of changes in the credit spread while lagged levels of the credit spread tend to matter more for $\Delta_4 s_t$.

The sentiment estimates ($\Delta_4 \hat{s}_{t,p=.}$) are reported in the upper right panel in Figure 10. To not clutter the graph we do not report the corresponding $R\hat{N}I_{t,p=.}$ estimates, but

[17] LSZ consider yearly data for the period 1926 to 2015. Our sample starts in 1985, for which working with annual data yields less than 40 observations. We therefore replicate their main analysis using quarterly observations. We use year-on-year changes to capture business cycle fluctuations and to resemble the annual changes analyzed in LSZ. Similarly, (8) is estimated using nonlinear least squares (NLLS) to take into account the generated-regressor nature of expected returns.

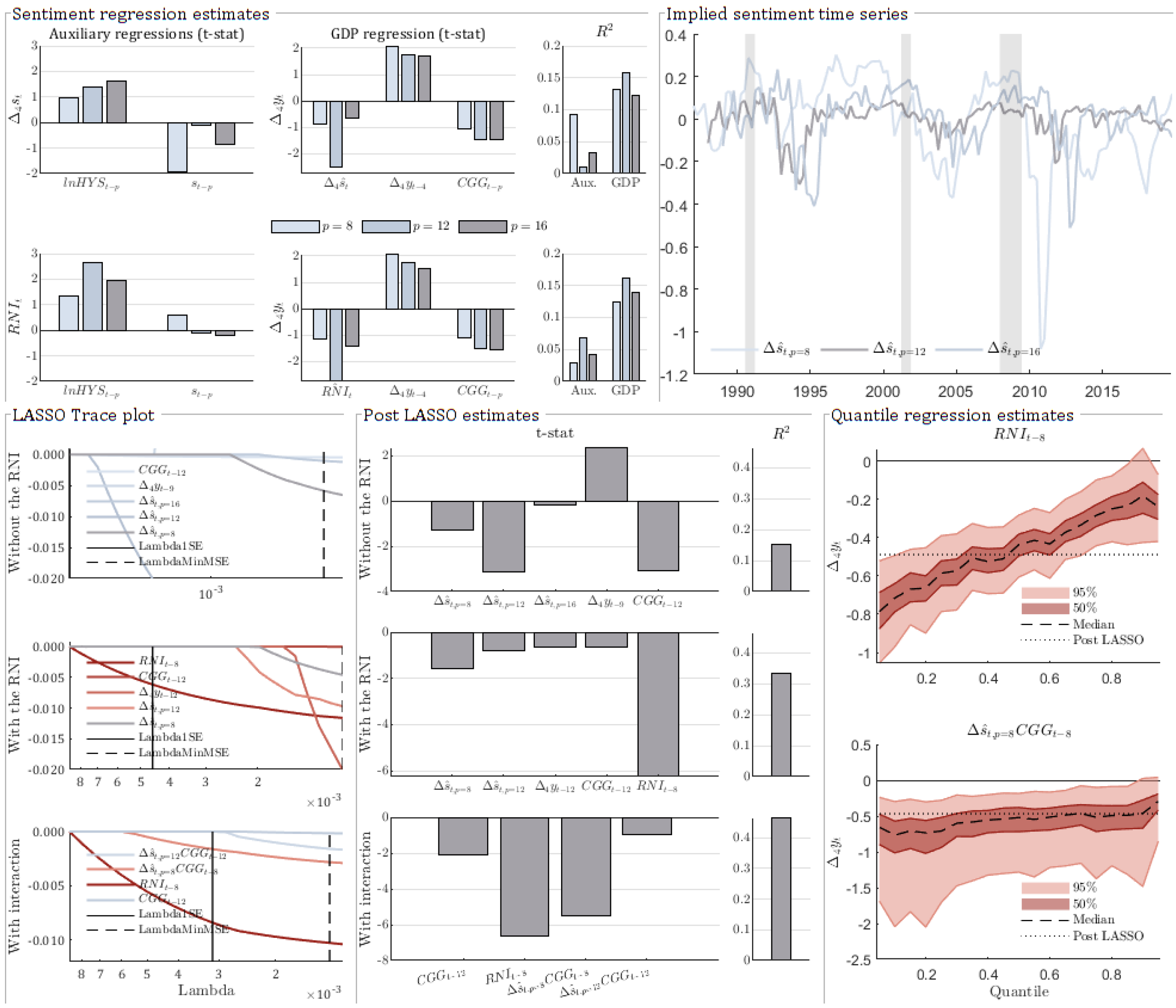


Figure 10. The upper left panel reports t-statistics from estimating (8) using the NLLS estimator. Predictor names are reported on the x-axis and the dependent variable on the y-axis. The different bars for each predictor correspond to a specific lag structure. The upper right panel reports the $\Delta_4\hat{s}_{t,p=\cdot}$ estimates associated with the first-step auxiliary regression (8a). The lower left and middle panels report the LASSO trace plot and post LASSO OLS t-statistics obtained from running three regularized regression experiments for $\Delta_4 y_t$. In all regressions we include the $\Delta_4\hat{s}_{t,p=\cdot}$ approximations and 8 to 12 quarterly lags of the CGG_t and $\Delta_4 y_t$ as predictors. The predictor set is then subsequently augmented with 8 to 12 quarterly lags of the RNI_t and the interaction terms described in Section 4.3. All t-statistics are based on standard errors computed according to Newey and West (1987). The lower right panel shows the coefficient estimates for RNI_{t-8} and $\Delta_4\hat{s}_{t,p=8}CGG_{t-8}$, with bootstrapped confidence bands, from a standard quantile regression model. See the text for details.

note that the correlations between $\Delta_4\hat{s}_{t,p=\cdot}$ and $R\hat{N}I_{t,p=\cdot}$ are 0.94 and 0.99 for $p = 12$ and $p = 16$, and -0.12 for $p = 8$. I.e., for the lag structure where the two sentiment approximations have a significant negative effect on the business cycle, $\Delta_4\hat{s}_{t,p=12}$ and $R\hat{N}I_{t,p=12}$, they are close to perfectly correlated.

To analyze the incremental signal strengths of credit market sentiment, the credit-to-GDP gap and the RNI at business cycle frequency in a more general manner, and at the same time be relatively agnostic about the exact lag structures, we proceed by running simple LASSO regressions including as predictors the three different $\Delta_4\hat{s}_{t,p=\cdot}$ approximations and 8 to 12 quarterly lags of the RNI_t, CGG_t, and $\Delta_4 y_t$.

The results from the regularized regression experiment are reported in the lower left and middle panels of Figure 10. If we allow for a flexible lag structure, but do not include the RNI in the LASSO regression, none of the predictors are selected for the most sparse model specification (Lambda1SE). If the regularization parameter is set to a value that minimized the cross-validated MSE (LambdaMinMSE), both the sentiment approximations and the credit-to-GDP gap are selected. According to the post LASSO OLS regressions, the parameter estimates are also significant for $\Delta_4\hat{s}_{t,p=12}$ and CGG_{t-12}. However, if lags of the RNI are included in the LASSO even the most sparse model specification includes it, and although $\Delta_4\hat{s}_{t,p=\cdot}$ and CGG_{t-12} are also chosen in the LambdaMinMSE specification, these predictors are not significant. In contrast, the parameter estimate for RNI_{t-8} is highly significant, and including it in the regularized regression experiment increases the R^2 from roughly 0.15 to 0.33.[18]

The last row in the lower left and middle panels of Figure 10 extends the predictor set further by allowing for an interactive regression specification, where credit market sentiment (the trigger) is interacted with the CGG (the vulnerability). Consistent with the findings reported in Kirti (2018), Krishnamurthy and Muir (2017) and López-Salido et al. (2017), the interaction term is a strong and statistically significant predictor, at least for the two-year lag specification. Including the interaction term in the regression also increases the R^2, but it does not make the RNI redundant. In fact, the RNI is still a highly significant predictor of the business cycle.

Figure A.9, in Appendix A, documents that these findings are not driven by the particular credit aggregate we control for. The figure reports the result from the same type of regularization experiment as above, but now increasing the predictor set with the five additional credit-to-GDP aggregates. As seen in the graphs, the strong predictive power of the RNI remains robust to also this variable augmentation exercise.

Finally, the lower right panel in Figure 10 connects these results to our earlier focus on the non-Gaussian properties of the conditional distribution of GDP growth. The graph reports the parameter estimates for the RNI and interaction term from a simple quantile regression model including RNI_{t-8} and $\Delta_4\hat{s}_{t,p=8}CGG_{t-8}$ and CGG_{t-12} as predictors for $\Delta_4 y_t$.[19] In line with our earlier results presented in Section 3, the RNI has a significant

[18] In neither of these regressions do we take into account the additional uncertainty related to the generated regressor issue. The t-statistics should be interpreted with caution.

[19] A residual bootstrap is used to construct confidence bands and take into account the additional uncertainty generated from the first-stage regressions (to construct $\Delta_4\hat{s}_{t,p=8}$). Because the R^2 in this first stage regression is rather low, and thereby generates a lot of variability, we follow López-Salido et al. (2017) and include the term spread in the first-stage regression to sharpen the inference. Our qualitative conclusions

negative effect on GDP at the business cycle frequency, and the effects seem to be especially negative at the lower left tail of the distribution. In contrast, the we find little support for a nonlinear relationship between the interaction term and the business cycle.

4.3 An (media) integrated view?

In sum, the analysis in the preceding sections provides positive evidence for the role of expectation shocks and the sentiment driven view of the credit cycle. More than this, however, the strong performance of the RNI relative to the sentiment variable(s) derived exclusively from valuation indicators, and the fact that the RNI is constructed from media coverage data alone, also suggest a role for independent media effects. In particular, although the identifying restrictions in (7) hold across a broad class of models with different, but related, information structures (Angeletos and La'O, 2010; Barsky and Sims, 2012; Blanchard et al., 2013; Angeletos and La'O, 2013), the significant correlations between belief shocks and the RNI suggest a more specific mechanism, where media coverage on financial conditions and growth-at-risk has adverse effects on expectations about the current state of the economy.

In terms of independent media effects, similar findings were documented in Chahrour et al. (2021) when analyzing how the relative representativeness of news coverage relates to belief shocks. As also stated in their analysis, a reverse causation argument, where belief shocks affect media coverage, is hard to square with the evidence put forward above. In particular, given the RNI's insignificant correlation with fundamental shocks, this would require the existence of fluctuations in beliefs that are correlated across individuals, but unrelated to either economic fundamentals or reports. And, it would require that news media can distinguish between, e.g., recessions driven by fundamentals and those driven by beliefs. We do not find these arguments very plausible.

Related to this, it is interesting to note that the credit market sentiment view advocated in newer studies such as Greenwood et al. (2016) and Bordalo et al. (2018) rests on assumptions about "diagnostic expectations" and extrapolative beliefs of investors. In contrast, in general models of endogenous information choice and independent media effects, a departure from the standard rational expectations paradigm is not needed. For example, the mechanism proposed in Nimark and Pitschner (2019), and Chahrour et al. (2021) in particular, only requires that agents in the economy rely on news media to monitor the economy on their behalf and to report the most newsworthy developments. Since even accurate reports provide only a partial picture of the economy, and agents receive the same partial information via news media, aggregate fluctuations can emerge

are not affected by this, but the parameter estimate for $\Delta_4 \hat{s}_{t,p=8} CGG_{t-8}$ becomes insignificant at all quantiles if not expanding the information set in the first-stage regression.

because media functions as a coordination device for the economy (beliefs). To the best of our knowledge, however, nobody has derived theoretical models incorporating this type of mechanism for the growth-credit relationship. In light of our findings, this seems to be a promising avenue for future research.[20]

5 Conclusion

A large theoretical literature emphasizes the importance of financial conditions for understanding adverse growth outcomes and business cycle fluctuations. However, the potentially nonlinear nexus between financial condition indicators and the conditional distribution of GDP growth has recently been challenged on the grounds that such indicators contribute little to distributional forecasts of GDP growth beyond the information contained in real indicators.

In this article we show how one can use textual economic news combined with a shallow Neural Network to construct an alternative financial conditions indicator based on word embeddings. By design the index, which we label the Risky News Index (RNI), associates growth-at-risk to news about credit, leverage and funding. The index can be updated in real-time, is not revised, and delivers a timely signal about financial conditions relevant for growth-at-risk.

We document using quarterly U.S. GDP growth and a parametric Student-t model that the in-sample estimates of the time-varying moments of the conditional GDP distribution are relatively precisely estimated and that the RNI significantly affects the evolution of the distribution's shape in particular. In an extensive out-of-sample density forecast analysis we further show that using the RNI outperforms commonly used financial condition indicators and factors summarizing the developments in real indicators. The performance of the RNI is particularly strong at the one-year-ahead forecasting horizon and during the Great Financial Crisis. For this horizon and episode, growth-at-risk predictions using conventional predictors are relatively uninformative whereas predictions from the RNI suggest a severe recession.

Finally, we link the derived index to structural shocks to fundamentals and beliefs, as well as credit market valuation indicators, and document that it is significantly correlated with shocks to expectations and expected returns. However, the predictive power of the RNI surpasses that of sentiment derived from credit market valuation indicators, providing positive evidence in favor of classical accounts of financial crisis in combination with theories on endogenous information choice and independent media effects.

[20] Of course, this does not rule out that also other types of (news-driven) mechanisms might be at play, such as, e.g., the narratives in Shiller (2017), Andre et al. (2021) and Nyman et al. (2021).

References

Adämmer, P., J. Prüser, and R. Schüssler (2023). Forecasting macroeconomic tail risk in real time: Do textual data add value? *arXiv preprint arXiv:2302.13999*.

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Adrian, T., F. Grinberg, N. Liang, S. Malik, and J. Yu (2022, July). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics 14*(3), 283–323.

Amburgey, A. J. and M. W. McCracken (2023). On the real-time predictive content of financial condition indices for growth. *Journal of Applied Econometrics 38*(2), 137–163.

Andre, P., I. Haaland, C. Roth, and J. Wohlfart (2021). Narratives about the macroeconomy.

Angeletos, G.-M. and J. La'O (2010). Noisy Business Cycles. In *NBER Macroeconomics Annual 2009, Volume 24*, NBER Chapters, pp. 319–378. National Bureau of Economic Research, Inc.

Angeletos, G.-M. and J. La'O (2013). Sentiments. *Econometrica 81*(2), 739–779.

Arellano-Valle, R. B., H. W. Gómez, and F. A. Quintana (2005, February). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference 128*(2), 427–443.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Barsky, R. B. and E. R. Sims (2012). Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence. *American Economic Review 102*(4), 1343–77.

Berger, D., I. Dew-Becker, and S. Giglio (2020). Uncertainty shocks as second-moment news shocks. *The Review of Economic Studies 87*(1), 40–76.

Bernanke, B. S., M. Gertler, and S. Gilchrist (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics 1*, 1341–1393.

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni (2013). News, Noise, and Fluctuations: An Empirical Exploration. *American Economic Review 103*(7), 3045–70.

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022, April). Maximum likelihood estimation for score-driven models. *Journal of Econometrics 227*(2), 325–346.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *The Journal of Finance 73*(1), 199–227.

Brave, S., R. A. Butters, et al. (2012). Diagnosing the financial system: Financial conditions and financial stress. *International Journal of Central Banking 8*(2), 191–239.

Brownlees, C. and A. B. Souza (2021). Backtesting global growth-at-risk. *Journal of Monetary Economics 118*, 312–330.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021). Feedbacks: financial markets and economic activity. *American Economic Review 111*(6), 1845–79.

Caldara, D., C. Fuentes-Albero, S. Gilchrist, and E. Zakrajšek (2016). The macroeconomic impact of financial and uncertainty shocks. *European Economic Review 88*, 185–207.

Caldara, D. and M. Iacoviello (2022). Measuring geopolitical risk. *American Economic Review 112*(4), 1194–1225.

Carriero, A., T. E. Clark, and M. G. Marcellino (2020). Capturing macroeconomic tail risks with bayesian vector autoregressions.

Chahrour, R., K. Nimark, and S. Pitschner (2021, December). Sectoral media focus and aggregate fluctuations. *American Economic Review 111*(12), 3872–3922.

Chernis, T., P. J. Coe, and S. P. Vahey (2023). Reassessing the dependence between economic growth and financial conditions since 1973. *Journal of Applied Econometrics 38*(2), 260–267.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics 28*(5), 777–795. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.1279.

Delle-Monache, D., A. De-Polis, and I. Petrella (2020). Modelling and Forecasting Macroeconomic Downside Risk. *EMF Research Papers*. Number: 34 Publisher: Economic Modelling and Forecasting Group.

Delle Monache, D., A. De Polis, and I. Petrella (2021). Modeling and forecasting macroeconomic downside risk. *Bank of Italy Temi di Discussione (Working Paper) No 1324*.

Delle Monache, D. and I. Petrella (2017). Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting 33*(2), 482–501. Publisher: Elsevier.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Drehmann, M. and J. Yetman (2021). Which credit gap is better at predicting financial crises? a comparison of univariate filters. *International Journal of Central Banking*.

Enders, Z., M. Kleemann, and G. J. Muller (2021, 11). Growth Expectations, Undue Optimism, and Short-Run Fluctuations. *The Review of Economics and Statistics 103*(5), 905–921.

Fernández-Villaverde, J. and P. A. Guerrón-Quintana (2020). Uncertainty shocks and business cycle research. *Review of economic dynamics 37*, S118–S146.

Figueres, J. M. and M. Jarocinski (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters 191*, 109126.

Gertler, M. and B. Bernanke (1989). Agency costs, net worth and business fluctuations. *American Economic Review 79*, 14–31.

Gertler, M., N. Kiyotaki, and A. Prestipino (2020). Credit booms, financial crises, and macroprudential policy. *Review of Economic Dynamics 37*, S8–S33.

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

Giglio, S., B. Kelly, and S. Pruitt (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics 119*(3), 457–471.

Gorgi, P., S. J. Koopman, and M. Li (2019, October). Forecasting economic time series using score-driven dynamic models with mixed-data sampling. *International Journal of Forecasting 35*(4), 1735–1747.

Greenspan, A. (2004). Risk and uncertainty in monetary policy. *American Economic Review 94*(2), 33–40.

Greenwood, R. and S. G. Hanson (2013, 04). Issuer Quality and Corporate Bond Returns. *The Review of Financial Studies 26*(6), 1483–1525.

Greenwood, R. M., S. G. Hanson, and L. J. Jin (2016). A model of credit market sentiment. *Harvard Business School working paper series# 17-015*.

Hamilton, J. D. and D. Leff (2020). Measuring the credit gap. Technical report.

Hamilton, W. L., J. Leskovec, and D. Jurafsky (2016). Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge: Cambridge University Press.

Hasenzagl, T., M. Plagborg-Møller, L. Reichlin, and G. Ricco (2020). When is growth at risk? *Brookings Papers on Economic Activity Spring*, 167–229.

Jordà, Ò., M. Schularick, and A. M. Taylor (2011). Financial crises, credit booms, and external imbalances: 140 years of lessons. *IMF Economic Review 59*(2), 340–378.

Jordà, Ò., M. Schularick, and A. M. Taylor (2013). When credit bites back. *Journal of money, credit and banking 45*(s2), 3–28.

Joulin, A., E. Grave, P. Bojanowski, and T. Mikolov (2016). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kapfhammer, F., V. Larsen, and L. A. Thorsrud (2020). Climate risk and commodity currencies.

Kilian, L. and S. Manganelli (2008). The central banker as a risk manager: Estimating the federal reserve's preferences under greenspan. *Journal of Money, Credit and Banking 40*(6), 1103–1129.

Kindleberger, C. P. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*. Basic Books.

Kirti, D. (2018). *Lending standards and output growth*. International Monetary Fund.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of political economy 105*(2), 211–248.

Koopman, S., A. Lucas, and M. Zamojski (2017). Dynamic term structure models with score-driven time-varying parameters: estimation and forecasting. *NBP Working Papers*.

Koopman, S. J., A. Lucas, and M. Scharth (2016, March). Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models. *The Review of Economics and Statistics 98*(1), 97–110.

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review 84*(5), 905–949.

Krishnamurthy, A. and T. Muir (2017). How credit cycles across a financial crisis. Technical report, National Bureau of Economic Research.

Labonne, P. (2022). Asymmetric Uncertainty: Nowcasting Using Skewness in Real-time Data.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics 4*(1), 25–38.

López-Salido, D., J. C. Stein, and E. Zakrajšek (2017). Credit-market sentiment and the business cycle. *The Quarterly Journal of Economics 132*(3), 1373–1426.

Loria, F., C. Matthes, and D. Zhang (2022). Assessing macroeconomic tail risk. *Available at SSRN 4002665*.

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

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Minsky, H. P. (1977). The financial instability hypothesis: An interpretation of keynes and an alternative to "standard" theory. *Challenge 20*(1), 20–27.

Minsky, H. P. (1986). *Stabilizing an unstable economy.* Yale University Press.

Newey, W. K. and K. D. West (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Nimark, K. P. and S. Pitschner (2019). News media and delegated information choice. *Journal of Economic Theory 181*, 160 – 196.

Nyman, R., S. Kapadia, and D. Tuckett (2021). News and narratives in financial systems: Exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control 127*, 104119.

Pennington, J., R. Socher, and C. Manning (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1532–1543. Association for Computational Linguistics.

Politis, D. N. and J. P. Romano (1994, 12). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics 22*(4), 2031–2050.

Prasad, M. A., S. Elekdag, M. P. Jeasakul, R. Lafarguette, M. A. Alter, A. X. Feng, and C. Wang (2019). *Growth at risk: Concept and application in IMF country surveillance.* International Monetary Fund.

Reinhart, C. M. and K. S. Rogoff (2009). The aftermath of financial crises. *American Economic Review 99*(2), 466–472.

Rossi, B. and T. Sekhposyan (2019). Alternative tests for correct specification of conditional predictive densities. *208*(2), 638–657.

Rubio-Ramirez, J. F., D. F. Waggoner, and T. Zha (2009). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies* (4), 1–34.

Salgado, S., F. Guvenen, and N. Bloom (2019). Skewed business cycles. Technical report, National Bureau of Economic Research.

Schularick, M. and A. M. Taylor (2012). Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review 102*(2), 1029–1061.

Shiller, R. J. (2017). Narrative economics. *American Economic Review 107*(4), 967–1004.

Smith, M. S. and S. P. Vahey (2016). Asymmetric forecast densities for u.s. macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence. *Journal of Business & Economic Statistics 34*(3), 416–434.

Stock, J. H. and M. W. Watson (1989). New indexes of coincident and leading economic indicators. In O. J. Blanchard and F. Stanley (Eds.), *NBER Macroeconomics Annual*, NBER Chapters, pp. 351–394. Cambridge, MA: The MIT Press.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *Journal of Finance 63*(3), 1437–1467.

Van der Maaten, L. and G. Hinton (2008). Visualizing data using t-sne. *Journal of machine learning research 9*(11).

Appendices for online publication

Appendix A Additional results

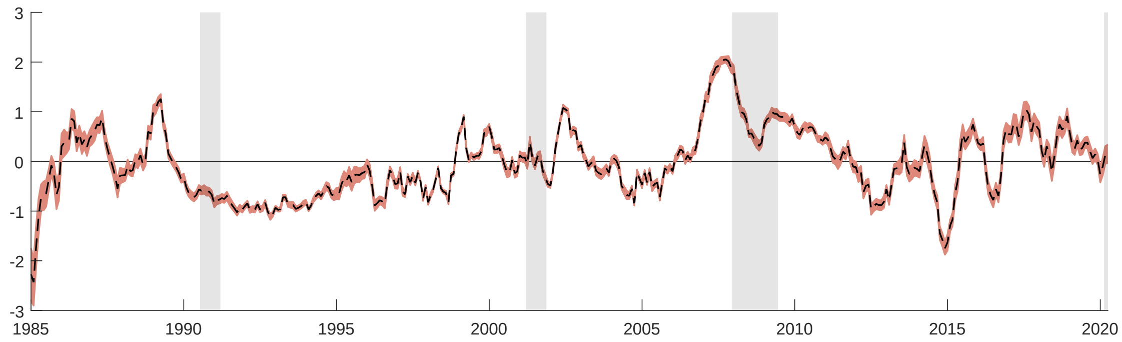


Figure A.1. The graph reports the monthly RNI estimate together with 90% confidence intervals computed as described in Appendix A.1. Shaded areas represent NBER recession dates.

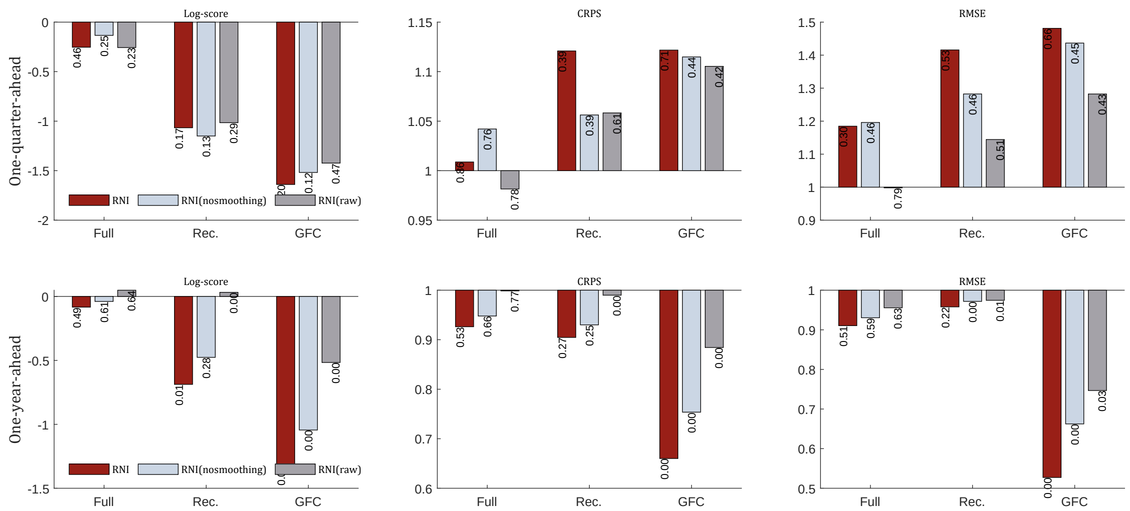


Figure A.2. The figure reports the average forecast metrics of the Skew-t(RNI), Skew-t(RNI(Raw)), and Skew-t(RNI(nosmoothing)) models relative to the Skew-t(NFCI) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(NFCI) benchmark. The p-value for the Giacomini and White (2006) test are printed vertically. The a-xis denote whether the evaluation is conducted using the whole out-of-sample periods (Full), for recession periods (Rec.), or only the Great Financial Crisis (GCF). See Section 3.2 for details about the out-of-sample forecasting experiment.

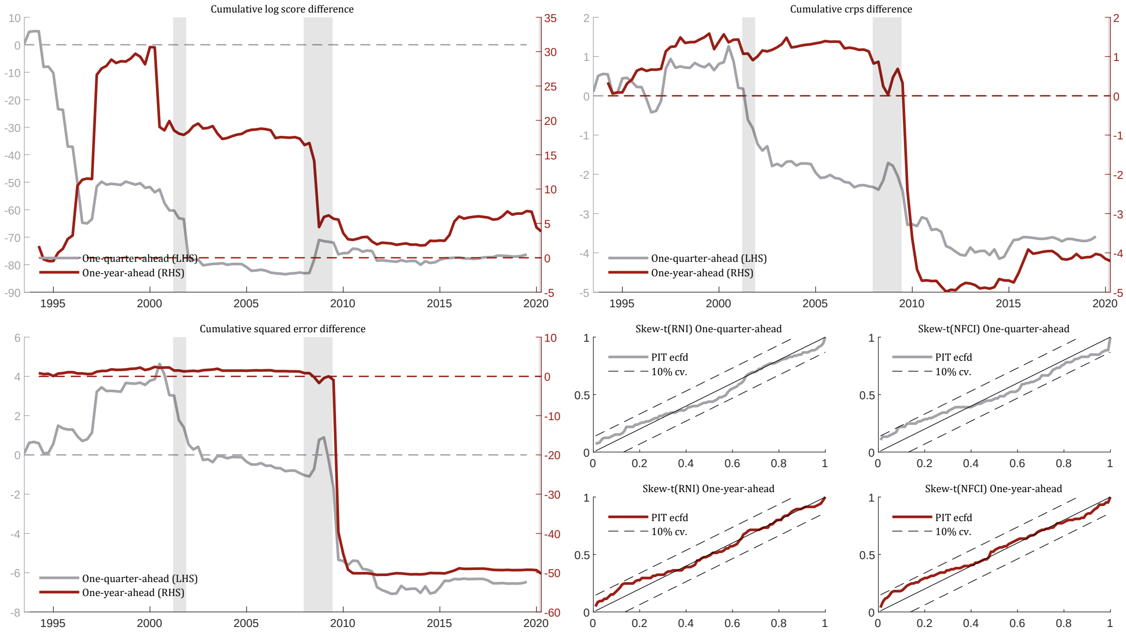


Figure A.3. The Skew-t(RNI) and Skew-t(NFI) models are estimated on data starting in 1985. The figure reports the PITs and the cumulative differences in LS, CRPS, and squared forecast errors between the Skew-t(RNI) and Skew-t(NFCI) models for one-quarter- and one-year-ahead predictions. All relative scores are normalized such that an increase implies a relative improvement of the Skew-t(NFCI) model.

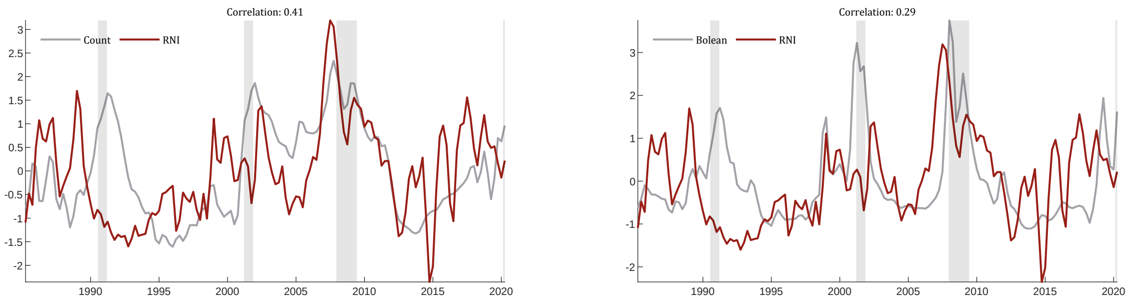


Figure A.4. The figures report the quarterly RNI together with alternative indexes created using either count- or Boolean-based methods. See Appendix A.4 for details about their construction.

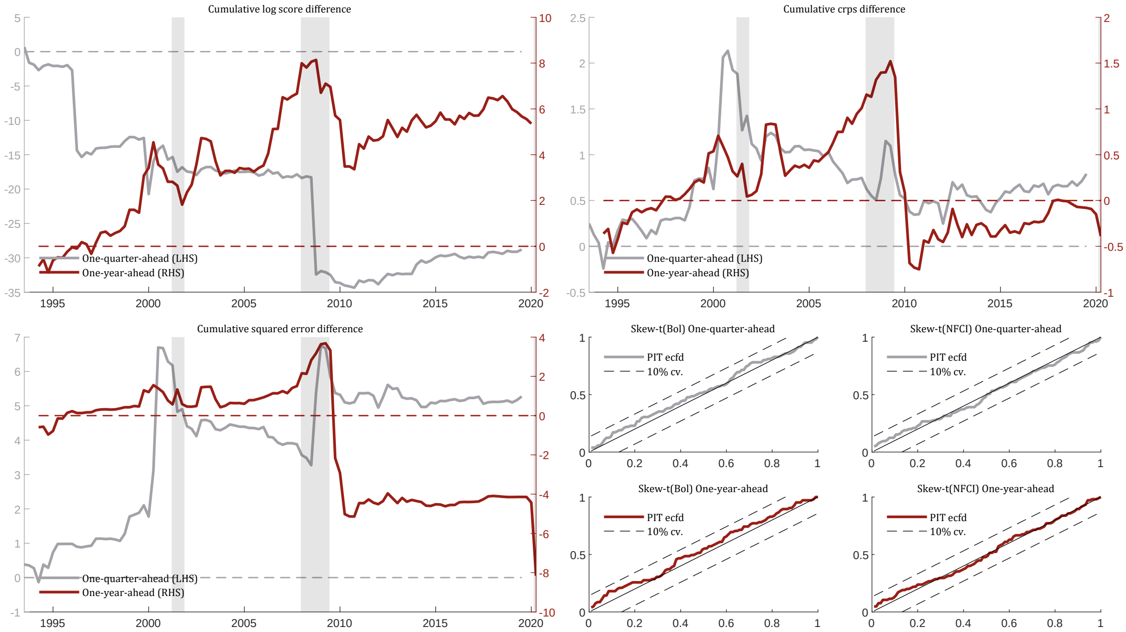


Figure A.5. The Skew-t(BoI)) model is estimated on data starting in 1985. The figure reports the PITs and the cumulative differences in LS, CRPS, and squared forecast errors between the Skew-t(NFCI) and Skew-t(BoI) models for one-quarter- and one-year-ahead predictions. All relative scores are normalized such that an increase implies a relative improvement of the Skew-t(NFCI) model.

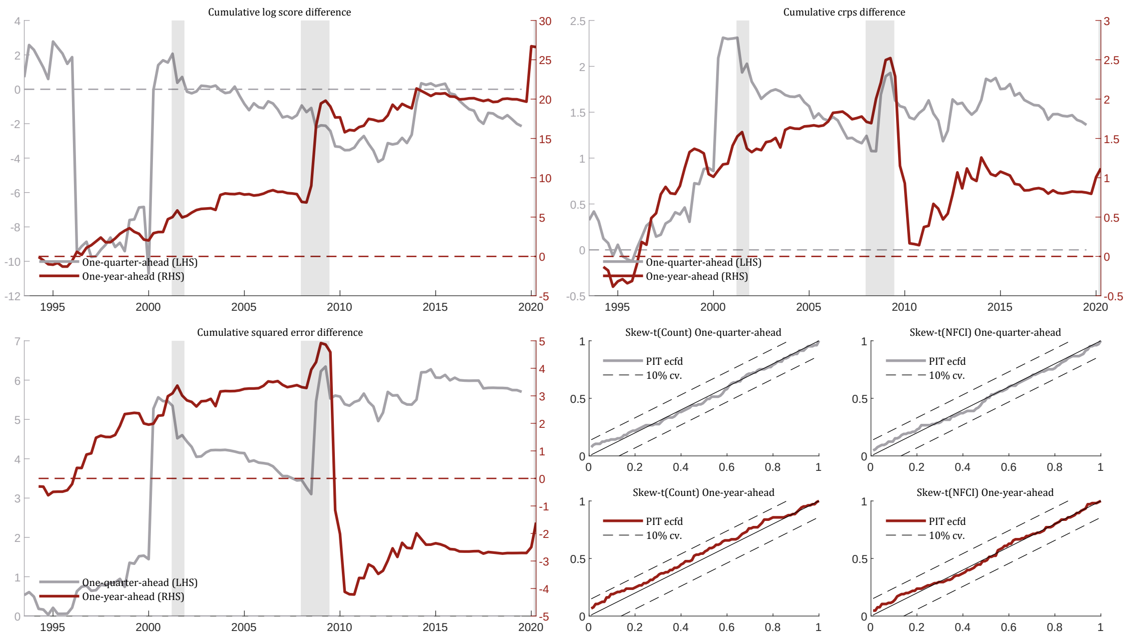


Figure A.6. The Skew-t(Count) model is estimated on data starting in 1985. The figure reports the PITs and the cumulative differences in LS, CRPS, and squared forecast errors between the Skew-t(NFCI) and Skew-t(Count) models for one-quarter- and one-year-ahead predictions. All relative scores are normalized such that an increase implies a relative improvement of the Skew-t(NFCI) model.

Figure A.7. The figure reports a wordcloud of the 30 most similar words to the **growth-at-risk**$_t$ vector at six different points in time. Similarity is measured using the cosine similarity metric. Vectors are averaged across months of the year prior to computing their distance.

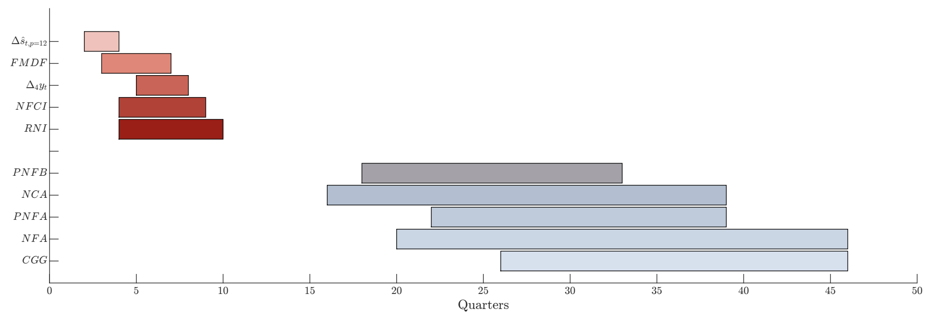


Figure A.8. The figure reports the estimated half life based on the impulse response function from univariate AR models. Lag specification is done using AIC. Filled areas correspond to 90% confidence bands. The listed variables are; credit market sentiment ()$\Delta\hat{s}_{t,p=12}$); the FRED-MD macroeconomic factor (FMDF), year-on-year growth in GDP ($\Delta_4 y_t$); the NFCI; the RNI; credit to the Non-financial sector from all sectors (NFA); credit to Households and NPISHs from all sectors (HA); credit to Non-financial corporations from all sectors (NCA); credit to Private non-financial sector from all sectors (PNFA); credit to Private non-financial sector from Banks (PNFB). The credit aggregates are all obtained from the BIS, measured relative to GDP, and computed as five-year changes. The exception is the CGG which is computed as a gap variable directly by the BIS.

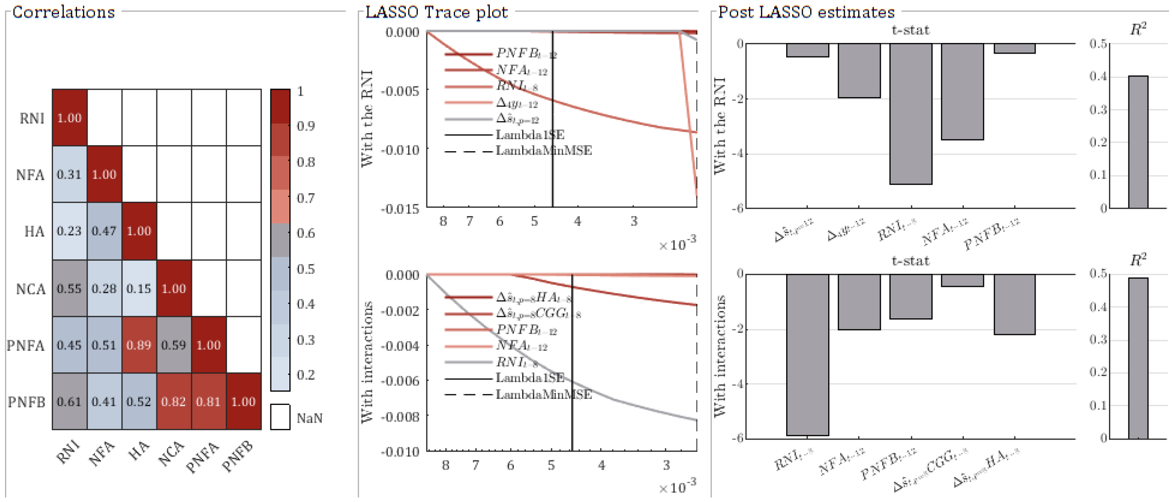


Figure A.9. The left panel reports the correlation between the RNI and five different balance-sheet measures often used in the literature. The additional measures are all obtained from the BIS, measured relative to GDP, computed as five-year changes, and include credit: to the Non-financial sector from all sectors (NFA); to Households and NPISHs from all sectors (HA); to Non-financial corporations from all sectors (NCA); to Private non-financial sector from all sectors (PNFA); to Private non-financial sector from Banks (PNFB). The middle and right panels report the LASSO trace plot and post LASSO OLS t-statistics obtained from two regularized regression experiments for $\Delta_4 y_t$. In all regressions we include the $\Delta_4 \hat{s}_{t,p=\cdot}$ approximations and 8 to 12 quarterly lags of the CGG_t, RNI_t, $\Delta_4 y_t$, and the five balance-sheet measures described above as predictors. The predictor set is then subsequently augmented with terms interacting the sentiment indicators with the balance-sheet measures. All t-statistics are based on standard errors computed according to Newey and West (1987).

Table A.1. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(FMDF) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(FMDF) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	0.19 (0.16)	1.13 (0.07)	1.39 (0.05)	-0.10 (0.75)	0.99 (0.81)	0.96 (0.66)
Rec.	0.53 (0.36)	1.32 (0.31)	1.65 (0.24)	-1.27 (0.01)	0.93 (0.04)	0.98 (0.48)
GFC	0.27 (0.63)	1.11 (0.56)	1.30 (0.67)	-1.96 (0.00)	0.70 (0.00)	0.59 (0.00)

Table A.2. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(EPU) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(EPU) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	0.21 (0.54)	0.95 (0.66)	0.84 (0.33)	-0.22 (0.53)	0.94 (0.05)	0.95 (0.05)
Rec.	-0.28 (0.55)	0.73 (0.31)	0.63 (0.35)	-1.22 (0.00)	0.92 (0.11)	0.97 (0.00)
GFC	-0.24 (0.64)	0.80 (0.32)	0.71 (0.37)	-2.41 (0.00)	0.75 (0.00)	0.73 (0.00)

Table A.3. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(MP) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(MP) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	-0.86 (0.07)	0.86 (0.03)	0.75 (0.08)	-0.00 (0.55)	0.94 (0.27)	0.95 (0.29)
Rec.	-1.38 (0.54)	0.59 (0.11)	0.50 (0.09)	-1.32 (0.00)	0.88 (0.06)	0.95 (0.01)
GFC	-0.82 (0.64)	0.72 (0.14)	0.61 (0.16)	-2.08 (0.00)	0.74 (0.00)	0.69 (0.00)

Table A.4. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(Bol) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(Bol) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	0.02 (0.31)	0.98 (0.91)	1.01 (0.98)	-0.13 (0.31)	0.93 (0.46)	0.97 (0.41)
Rec.	0.48 (0.25)	1.19 (0.14)	1.28 (0.20)	-0.56 (0.00)	0.91 (0.04)	0.98 (0.00)
GFC	0.34 (0.17)	1.03 (0.09)	1.04 (0.23)	-1.39 (0.00)	0.61 (0.00)	0.46 (0.00)

Table A.5. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(Count) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(Count) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	-0.23 (0.22)	0.97 (0.79)	0.99 (0.81)	-0.34 (0.19)	0.90 (0.21)	0.92 (0.35)
Rec.	-0.94 (0.25)	1.04 (0.33)	1.17 (0.34)	-2.15 (0.00)	0.85 (0.01)	0.93 (0.12)
GFC	-1.58 (0.17)	0.95 (0.38)	1.03 (0.30)	-3.00 (0.00)	0.59 (0.00)	0.48 (0.00)

Table A.6. The table reports the average forecast metrics of the Skew-t(RNI) model relative to the Skew-t(NFCIreal) model. We use ratios for the RMSE and CRSP, and differences for the LS. Ratios smaller than 1, and negative values of the LS differences indicate that Skew-t(RNI) model performs better than the Skew-t(NFCIreal) benchmark. The p-value for the Giacomini and White (2006) test are in parentheses.

	One-quarter-ahead			One-year-ahead		
	LS	CRPS	RMSE	LS	CRPS	RMSE
Full	-0.06 (0.89)	1.06 (0.45)	1.23 (0.23)	-0.13 (0.20)	0.92 (0.38)	0.91 (0.31)
Rec.	-1.28 (0.22)	1.20 (0.28)	1.40 (0.43)	-0.90 (0.08)	0.94 (0.17)	0.97 (0.00)
GFC	0.50 (0.34)	1.28 (0.45)	1.54 (0.48)	-0.77 (0.00)	0.82 (0.00)	0.72 (0.37)

Table A.7. K-means clustering. The K-means algorithm is used to estimate five clusters based on the combined embedding matrix visualized in Figure 3. The table reports how the fraction of unique terms for the five concepts listed in the first row are allocated across the five estimated clusters.

Cluster	"Growth-at-risk"	"Financial stress"	"Policy uncertainty"	"Monetary policy"	"Pandemic"
1	0.01	0.00	0.00	0.00	0.92
2	0.00	0.00	0.06	0.75	0.00
3	0.15	0.26	0.61	0.13	0.08
4	0.00	0.72	0.00	0.02	0.00
5	0.82	0.02	0.31	0.10	0.00

A.1 Subsampling

To construct confidence intervals for the RNI_t estimates, we follow Kozlowski et al. (2019) and conduct subsampling (Politis and Romano, 1994). For the 90% confidence interval, the corpus is randomly partitioned into 20 subcorpora, and the word2vec algorithm is run to produce the word embedding matrix for each data partition. The error of the projection statistic RNI_t for each subsample s is $e^s = \sqrt{\tau_s}(RNI_t^s - RNI_t)$, where τ_s and RNI_t^s are the number of texts and the solution to (2.2), respectively, in subsample s. Then, the 90% confidence interval spans the 5th and 95th percentile variances, defined by $RNI_t + \frac{e^{s(19)}}{\sqrt{\tau}}$ and $RNI_t - \frac{e^{s(2)}}{\sqrt{\tau}}$, where $e^{s(2)}$ and $e^{s(19)}$ denote the 2nd and 19th order statistic associated with the lower and upper bounds of the confidence interval.

A.2 Skew-t specification

The model specification described in Section 3 builds on work by Delle Monache et al. (2021). We depart (slightly) from their modeling choices along two dimensions. First, they assume ARX(2) processes for $\upsilon_{u,t}$. However, the second lag coefficient is never significant, and is thus dropped here to favor a more parsimonious model structure. Second, they estimate their model using Bayesian methods and an adaptive Random-Walk Metropolis-Hastings algorithm. Relative to a ML approach, this has the advantage that informative priors can be used to discipline the model. A disadvantage is that it becomes substantially more time consuming to estimate the model. With our ML implementation, one estimation run takes less than five minutes on a standard laptop.

A.3 Constructing the text scatter plot

The text scatter plot in Figure 3 is constructed as follows: For each month in the year 2007 the 3000 most similar words to different concepts are extracted (based on cosine similarity scores). Then, we focus on the intersection of this set and compute the average word embedding for each word (across months) in the retained set. Since independent applications of the word2vec algorithm might result in arbitrary orthogonal transformations, we follow, e.g., Hamilton et al. (2016), and use orthogonal Procrustes to align the word embeddings before averaging.

The two-dimensional visualization of the high-dimensional embeddings relies on the t-SNE algorithm (Van der Maaten and Hinton, 2008) applied on the embedding matrix containing the unique terms associated with each concept as well as the common terms shared by two or more of them. The colors reflect terms unique to one concept. Common terms are gray. The t-SNE algorithm is implemented by setting the perplexity to 10, reduce the original dimension of the embedding space to 50 using PCA prior to estimation, and allow for up to 5000 optimization iterations. These choices are common in the literature.

A.4 Constructing alternative count- and Boolean-based indexes

To construct the two alternative count- and Boolean-based indexes we follow conventional practices in the literature, but adapted to the current setting by searching for the words used to define the RNI.

Thus, for the count measure we first simply count the number of occurrences of the words "recession", "risk", "credit", "leverage", and "funding" in any given month, and then sum these counts normalized by the total number of words that month. For the Boolean approach we operate at the article level. Since all the five individual terms are unlikely to feature in the same article, we restrict our search and count procedure to three subcategories counting articles containing the words $recession\&risk\&credit$, $recession\&risk\&leverage$, and $recession\&risk\&funding$. These counts are then normalized by the number of articles, summed to monthly frequency, and averaged to get one index.

Changes in the count- and Boolean-based indexes can be due to high-frequency changes in how the media focuses upon growth-at-risk and financial conditions, more persistent changes in how this relationship is focused upon, or noise and breaks in the news coverage and style. To isolate the former component we apply the same filtering methodology as we do for the RNI. A simple (backward-looking) moving average filter normalizes each observation with the mean and standard deviation of the last five years of raw data, and then we smooth the resulting series by the trailing six month average.

Centre for Applied Macroeconomics and Commodity Prices (CAMP)
will bring together economists working on applied macroeconomic issues, with special emphasis on petroleum economics.

BI Norwegian Business School
Centre for Applied Macro - Petroleum economics (CAMP)
N-0442 Oslo

www.bi.no/camp

CAMP Working Paper Series
ISSN: 1892-2198