Journal of Financial Stability 70 (2024) 101209

Contents lists available at ScienceDirect

Journal of Financial Stability

journal homepage: www.elsevier.com/locate/jfstabil

A thousand words tell more than just numbers: Financial crises and historical headlines☆

Kim Ristolainen [a,*], Tomi Roukka [a], Henri Nyberg [b]

[a] Department of Economics, Turku School of Economics, University of Turku, Finland
[b] Department of Mathematics and Statistics, University of Turku, Finland

ARTICLE INFO

JEL classification:
G00
G01
N01
C25
C82

Keywords:
Financial crisis
Text data
Leading indicators
Topic model

ABSTRACT

We show that financial crises are preceded by changes in specific types of narrative information contained in newspaper article titles. Our novel international dataset and the resulting empirical evidence are gathered by integrating information from a large panel of economic news articles in global newspapers between the years 1870 and 2016 with conventional macroeconomic and financial indicators. We find that the predictive information of newspaper article titles that signals coming crisis episodes is substantial over and above the macroeconomic and financial indicators. Feature contribution analysis and crisis case studies reveal that the new indicators capture more detailed, but still generalizable information on the buildup of crises.

1. Introduction

Financial crises have been of central interest since the global financial crisis of 2008 and the subsequent European sovereign debt crisis. In the recent economic literature, large historical panel datasets such as the one assembled by Jordà et al. (2017) covering more than 150 years of macroeconomic and financial data for multiple countries have generated newfound interest in examining leading indicators. In a series of papers utilizing this new rich data, it has been shown that credit expansions (Schularick and Taylor, 2012) and private sector debt (Jordà et al., 2016) have predictive power for banking crises, but a crisis preceded by lower capital ratios (Jordà et al., 2020) and large public debt is more costly in general.[1] Although many leading indicators have been found, the prediction performance of these models is still modest: A randomly selected year can be classified as a crisis (or "normal") year correctly with around 70% accuracy. Despite this (better than a random coin toss) predictive power, the models can still be seen as not accurate enough to persuade policy makers to rely solely on the signals of these models to take systematic preventive action in a timely manner.

Recent research has uncovered many new insights into multiple economic applications and issues by incorporating newspaper data in measuring relevant economic variables. Specifically, text data has been used to measure stock market sentiment (Tetlock, 2007), economic policy uncertainty (Baker et al., 2016), firm-level policy uncertainty (Hassan et al., 2019), sentiment in central bank financial stability reports (Correa et al., 2021), the relationship between news article topics and U.S. economy (Bybee et al., 2023a) and information diffusion in social media (Gorodnichenko et al., 2021). Text data from economic newspaper articles may contain much broader information

☆ We thank Alp Simsek, Janne Tukiainen, Paavo Okko, Matti Virèn, Karlo Kauko and seminar participants at the University of Turku, the Bank of Finland, 2022 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics, the macro-finance session at the 2022 annual meeting of the EEA-ESEM (Milan), the 4th conference on Non-traditional Data, Machine Learning, and Natural Language Processing in Macroeconomics (Stockholm), and the 2023 ASSA annual meeting (New Orleans) for their useful comments and discussions. We also thank the editor and the two anonymous referees for their comments that have improved the paper significantly. We gratefully acknowledge the research assistance provided by Elsi Lindell and Miko Pasanen, as well as the financial support provided by the Emil Aaltonen Foundation, OP Group Research Foundation, (Ristolainen) Yrjö Jahnsson Foundation, and (Nyberg) the Academy of Finland (grant 321968). This research is part of the COST Action project CA21163 "Text, functional and other high-dimensional data in econometrics: New models, methods, applications" (HiTEc).

* Correspondence to: Turku School of Economics, Department of Economics, University of Turku, FI-20014, Finland.
E-mail address: kim.ristolainen@utu.fi (K. Ristolainen).

[1] Recently, Greenwood et al. (2022) use the post-WWII sample of the historical banking crisis database newly assembled by Baron et al. (2020)–that starts from the year 1870 and covers 46 countries–to show that the interaction of rapid credit and asset price growth in the previous three years makes a crisis much more likely in the following years.

https://doi.org/10.1016/j.jfs.2023.101209
Received 2 December 2022; Received in revised form 29 November 2023; Accepted 6 December 2023
Available online 16 December 2023

about economic issues than just the plain numerical values of GDP growth, interest rates, unemployment, inflation, or even asset prices. Importantly, macroeconomic variables are also usually reported with a lag and typically revised multiple times in the future. In addition, asset prices may be influenced by financial market anomalies, belief formation biases, and heuristics that violate the rational expectations assumption and complicate their use as leading indicators.

In this paper, we use Blei et al. (2003)'s Latent Dirichlet Allocation topic model to assemble a large annual panel dataset containing specific time- and country-invariant economy-related topics in news article titles published in a collection of global newspapers during the period 1870–2016. The topic model is formed without any specific purpose or target, i.e. including only the topics that we might think to be related to financial crises. That is, we aim to include and distinguish all important economy related themes from this large collection of news article titles and also to quantify the popularity of these topics relative to all other economy related topics. We provide the annual frequencies and common words for each of the 110 topics of the final model for all 17 developed countries in the historical macro-dataset assembled by Jordà et al. (2017), so that the two information sources can be used together in an economic analysis. We validate the dataset (the topics and their frequencies) by showing that the majority of the topics capture a coherent economic factor that makes sense to a human (and not just to a machine) and that the frequencies of the topics capture historical events across topics, countries, and time. Although newspaper articles have increasingly been used in economic analysis in recent years, we are the first to derive topic frequency series from global newspapers for a large panel of countries. In addition, the topics that we estimate are not time or region specific, but they can be found from the whole time span of 150 years and across all 17 countries.

As a second step after first dissecting the text information accurately, we contribute to the literature that aims to understand and predict historical financial crises by adding a new, important, and previously absent element: text data. We show that throughout history, shifts in the attention to specific latent topics or narratives of newspaper articles precede financial crises. These leading indicators derived from text data outperform macroeconomic and financial data in crisis prediction and most importantly, significantly improve the predictive ability of these models even when the usual indicators are included in the model. Together, the text topic series adds around 11 to 13 percentage points to the prediction accuracy of the benchmark model (containing only macroeconomic and financial variables) for 2–4 year prediction horizons, bringing the overall classification ability to the 91%–96% range in terms of the Area Under the ROC curve. Even more impressive is that the text topic indicators alone achieve similar out-of-bag prediction accuracy as the best model with both text and macroeconomic and financial variables. The superior performance of the model using text data is not specific to crisis types or specific time periods, and it outperforms the benchmark model across time before and after the World War II as well as in out-of-sample forecasting experiment. These results are confirmed to hold in a large set of robustness checks.

The most important text-based leading indicators contain text topics that, for instance, cover issues of *periodic production numbers*, *high-level and government economists*, *financial crises*, *trade and marketing*, and *change in international trade barriers*. Separate case studies of specific crisis episodes show that the topics capture general factors like external shocks, changes in economic policies, economic atmosphere and detailed production/output information, connected to events, and at least partial causes, for the specific historical crisis episodes. Although credit growth is the most important crisis indicator in our analysis, as it is in the existing crisis literature, the total contribution of topics to the predicted probability is over 75%. This implies that despite credit growth is the key predictor, the information in news topics brings a significant improvement to the prediction accuracy. We find a large group of individual crisis episodes where the model does not rely at all on credit growth when making a prediction. The case studies reveal that in the majority of these cases credit growth played a role, but the topic indicators provide more detailed information on the credit boom and the buildup of the crisis. Although there are crises in which credit did not play a role, the credit-booms-gone-bust explanation for crises still fits to the majority of episodes. Moreover, the text topic frequencies provide information on the reasons and characteristics of the credit boom (i.e. whether the credit boom is either good or bad).

The results of this paper imply that there is much more generalizable public information than just the usual macroeconomic and financial indicators that is relevant for signaling a financial crisis in the future. From the perspective of policy makers, news article information can be used to form more accurate early warning models for financial crises. Text data can be used to measure more specific and detailed general factors that the usual economic statistics may not capture. The importance of these factors to financial crises may be of general knowledge, but incorporating them into prediction models has been difficult in previous research. The results also point out that there is still a lot that we do not understand about financial crises as our theory-implied indicators cannot explain the large difference in the prediction accuracy. Furthermore, our novel empirical evidence provides necessary "first step" evidence and encourages for further research attempts to examine, e.g., how different taken, and not taken, fiscal and monetary policy actions under different levels of crisis probabilities, now also affected by text information, may lead to utility gains (cf., e.g., Berge and Jordà (2011)) and implications to the "lean against the wind" arguments put forth by, e.g., Svensson (2017) and Richter et al. (2021).

The article is organized as follows. In Section 2, we show how we form our new text-based measures with topic modeling that incorporates historical news article data. In Section 3, we present the results for the financial crisis prediction analysis and various robustness analyses of the main results. In Section 4, we focus on the interpretation of different topic-based leading indicators and analyze the relationship between historical crisis episodes and the new leading indicators. Finally, in Section 5, we conclude.

2. Quantifying economic news articles

2.1. Historical newspaper data

We collect a large corpus of historical newspaper texts from ProQuest Historical newspapers. Our aim is to build a collection of global newspapers that would have reached readers on a global scale for the last 150 years. The newspapers covered are *The New York Times, The Washington Post, The Globe and Mail, The Times of India, The Guardian*, and *The Wall Street Journal*. We collect the titles of the 200 most relevant[2] economic articles for each newspaper-year-country combination that mention at least one economic keyword and country name.[3] The countries included are Australia, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom (UK), Italy, Japan, The Netherlands, Norway, Portugal, Sweden, and the United States (US).

[2] Relevance is defined by the ProQuest search engine.

[3] The list of 104 economic keywords was assembled with a search and error procedure in which different economy-related keywords were used to find economy-related news articles from Proquest historical newspapers. After the news articles were collected, a random manual audit was performed by reading the articles of specific decades and countries. During this procedure, texts and keywords that had selected them were removed if the article was not related to the economy. Ten percent of the different decade-country combinations were randomly selected for the audit. This process resulted in the removal of keywords based on the root word "work", because they mostly resulted in sports- and art-related articles.

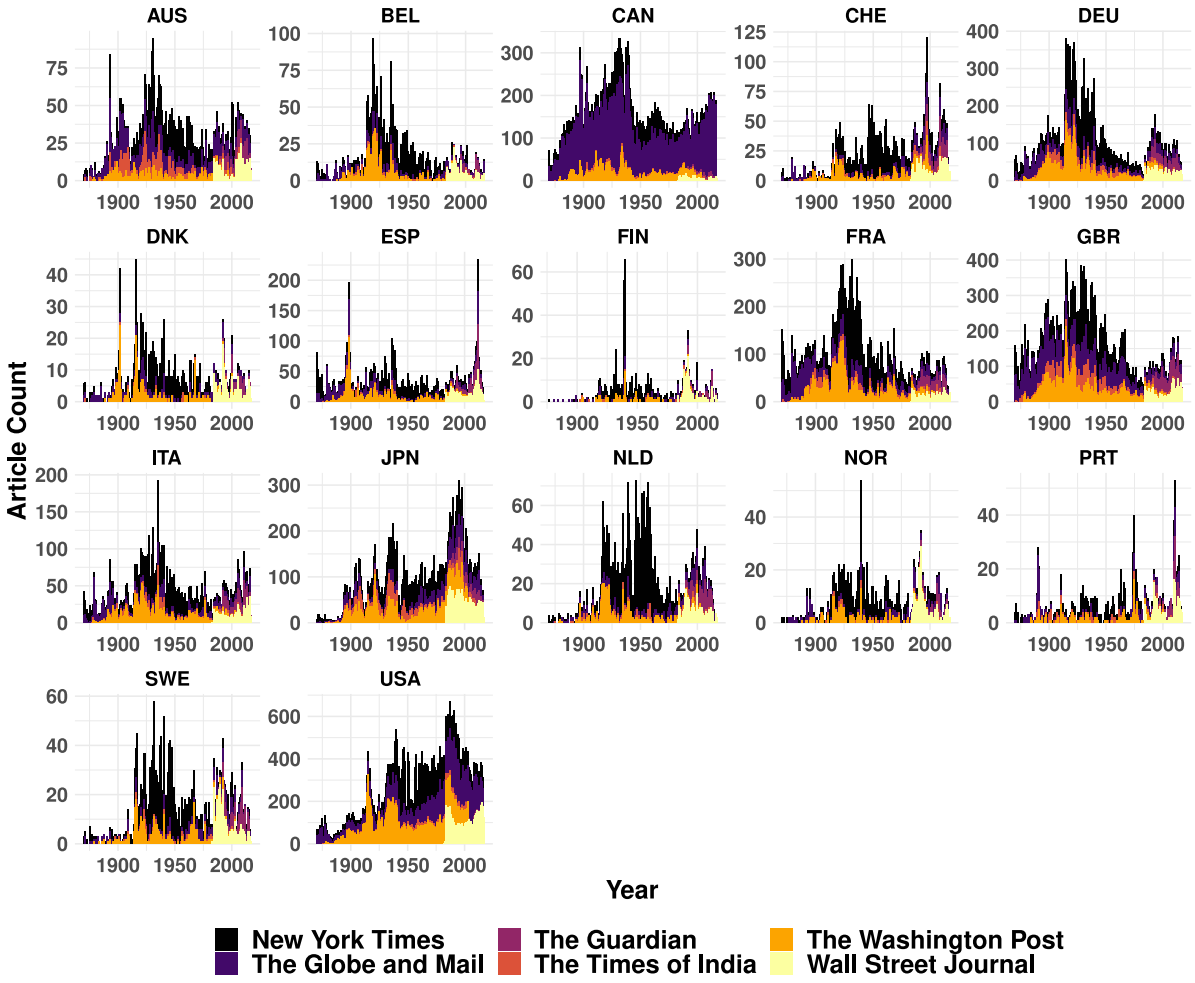

Fig. 1. Annual article counts across time, newspapers and countries.

After an extensive text audition, we decided to focus on the article titles.[4] Each article has to mention the name of the country and at least one of the economic keywords in the article title. This requirement makes it more likely that the chosen article discusses about economic topics related to that specific country, when compared to the procedure where it would be enough that the keywords are mentioned in the full text. In addition, article titles explicitly summarize the content of the news articles, whereas the full texts also include a lot more useless content. The use of titles also keeps the data collection process feasible. With this text collection procedure, the final corpus includes 171,711 economy-related article titles.

Fig. 1 displays the distribution of the newspaper article titles across countries, years, and newspapers. It is no surprise that the United States has the largest number of articles mentioning the economy–more than 40,000. Behind the US, Japan, Germany, the United Kingdom, Canada, and France form a group with around 15,000–25,000 articles. Italy, Spain, Switzerland, The Netherlands, Belgium, and Sweden are mentioned in 2300–7000 articles. The countries with fewer than 2000 total articles are Portugal, Denmark, Sweden, and Finland. We do not see the large dispersion in country articles as a problem as the number of articles reflects a country's economic importance and relevance in the world. For example, it can be seen that the number of articles mentioning Japan has increased gradually since 1950s reflecting the change in the country's role in the global markets. Another good example is the UK, for which the number of articles decreased gradually from the 1900s onward as it is losing its place as a global superpower to the US. Finally, as an example, the number of articles mentioning Finland reflects well how a small country is discussed on a global scale. The peaks for Finland occur during the Second World War and in the 1990s when this small Nordic country experienced one of the worst banking crises in history and also recovered partly due to the rise of mobile phones and Nokia. The data collection procedure is described in more detail in Appendix.

2.2. Topic modeling

We use the most commonly used topic model LDA (Blei et al., 2003) to learn latent topics in the historical newspaper corpus. LDA is an unsupervised learning model (when referring to the classification of supervised and unsupervised statistical and machine learning models) that uncovers structures in the data without having any knowledge of the data labels. In this context, we do not provide the model the information that some news are titled as, for example, sports or economic news. The LDA model assumes that each news article title is generated by a generative process where each title is possibly a combination of multiple topics (title can be represented by distribution of topics) and each topic is a distribution of words from a fixed vocabulary. The word latent refers to the latent nature of the topics in the text corpus, and the word Dirichlet refers to distributions from which the topic and word distributions are assumed to be generated.

Let us assume that there is a corpus of D news titles with N words in each title from a vocabulary of size V and a total number of K topics. For title d there is a distribution $\theta_d \sim Dirichlet(\alpha_1, \alpha_2, \ldots \alpha_K)$

[4] The manual audit is described more thoroughly in Appendix.

of topics, and for each topic k, there is a word distribution $\beta_k \sim Dirichlet(\eta_1, \eta_2, \dots \eta_V)$. The actual title d is generated so that for each word position n in title d a topic $z_{d,n}$ is generated from the topic distribution so that $z_{d,n} \sim Multinomial(\theta_d)$. Next, a word $w_{d,n}$ is generated for each word position n in title d from the word distribution of the assigned topic $z_{d,n}$ so that $w_{d,n} \sim Multinomial(\beta_{z_{d,n}})$. Smaller values of the parameter α of the Dirichlet distribution make the titles consist more of single topics, whereas larger values force titles to consist more of multiple topics.

The process can be more formally written as the following joint probability function:

$$P(\theta, \beta, Z, W) = \prod_{k=1}^{K} P(\beta_k|\eta) \prod_{d=1}^{D} P(\theta_d|\alpha) \prod_{n=1}^{N} P(z_{d,n}|\theta_d) P(w_{d,n}|\beta, z_{d,n}). \quad (1)$$

In this generative process, we know only $w_{d,n}$ and the number of topics K, as it is predetermined by the researcher. The distributions, the topic assignments, and the other parameters are unknown. We can estimate the following posterior distribution in Eq. (2) with Gibbs sampling:

$$P(\theta, \beta, Z|W) = \frac{P(\theta, \beta, Z, W)}{P(W)}. \quad (2)$$

This procedure searches for the parameter values that make the process to generate newspaper article titles that are as close to the ones that we observe in the actual text corpus.

As the number of topics K is a predetermined hyperparameter of the model, the optimal value of K has to be estimated via a grid search. We use a model's exclusivity and semantic coherence (Mimno et al., 2011) together with human judgment as criteria to select the optimal value for K. Semantic coherence is high when the most frequent words in a topic co-occur frequently. Exclusivity is high when a topic's most frequent words are not frequent in other topics. There is a trade-off between these two measures and the number of topics. When there are many topics, a model's exclusivity is often high, and its semantic coherence is often low, and vice versa when there are only a few topics. This means that often human judgement is the final criteria to select the optimal amount of topics.

Semantic coherence and exclusivity for different values for K are plotted in Fig. C.1. We chose the topic model with 110 topics as it holds a sufficiently large number of topics to describe the large news coverage of the text corpus, and the majority of the topics were identifiable at an adequate level. Based on our findings of our new title corpus and by using human judgement, 110 as the number of topics is sufficiently large to distinguish certain important themes from each other, as well as, small enough not to have too narrow, event specific, topics. Finally, we form the annual topic frequency $TF_{k,t,c}$ for a specific topic k by averaging the proportion of each topic k across the L articles that are published in year t and mention country c. More formally,

$$TF_{k,t,c} = \frac{1}{L} \sum_{l=1}^{L} \theta_{k,l,c}. \quad (3)$$

A number of authors have implemented LDA in economic research in different applications. Hansen and McMahon (2016) use topic modeling to study the effects of central bank communication on the economy. Larsen and Thorsrud (2019) and Larsen and Thorsrud (2022) show with Norwegian business newspaper data that a number of topics have predictive power on key economic variables like asset prices. Larsen et al. (2021) estimate a topic model by incorporating all news from the Dow Jones newswire archives from 1990 to 2016. The authors show that a subset of the topics predict inflation and inflation expectations. Bybee et al. (2023a) construct a topic frequency series of the *Wall Street Journal* business articles from 1984 onward and show that these topics explain fluctuations in economic variables well. In Bybee et al. (2023b), the authors use topic modeling in an asset pricing framework and extract risk factors from news narratives. Our methods and purpose differ in three important aspects from theirs. First, we construct the series from 1870 onward from articles in six global newspapers. Second, we form the series separately for each country by identifying economy-related articles mentioning specific countries. Third, we clean the text data so that we remove all time- and country-variant words from the corpus before we estimate the topic model. The purpose of using this procedure is to get general and timeless topics that can be utilized in an econometric analysis. This way, we do not get topics that discuss "World War II" or "Ronald Reagan", but general versions of these specific topics, like war and presidents. These procedures are needed to analyze the relationship between financial crises and text-based information, because crises are rare events that occur relatively seldom—every 28 years on average (Jordà et al., 2011).

2.3. Validating topic attention series

2.3.1. Identifying topic content

We validate the output of the final LDA model in terms of two different aspects. First, we want to make sure that the actual topics make sense. That is, the most frequent words for a specific topic according to the topic's word distribution β_k and the most representative titles for that specific topic form a coherent description of the issues that the topic is covering. Simply looking at the x most frequent words or most representative titles of a topic may not reveal the content of every topic exhaustively, which is why it is sufficient that at least the majority of the topics can be interpreted and not necessarily every one.[5]

Fig. 2 displays the 110 topics and Table C.1 in the Appendix shows the seven most exclusive words of each of these topics according to the frequency and exclusivity (FREX) metric (Airoldi and Bischof, 2016). In contrast to using only the word frequencies of a given topic, FREX gives a higher value for the words that are more frequent in a specific topic but less frequent in others. This way, we do not put too much emphasis on words that are very common across many topics. An example of an easily interpretable topic is topic 4 (*economy-growth-spur-signs-grows-slowing-pace*), which clearly discusses the direction and magnitude of economic growth. Topics 5 (*central-bank-governor-institution-raises-eyes-credit*) and 6 (*banks-savings-accounts-cash-lending-institutions-fail*) are likely to loosely cover issues on central banks and banking. Furthermore, topics 12 (*investors-briefs-look-wary-feel-yields-interest*), 37 (*labor-party-unions-leader-strikes-reds-vote*), and 55 (*gold-metal-standard-silver-reserves-reserve-shipment*) could be about views on financial markets, left-wing politics, labor unions and strikes, and noble metals.

Not all topics are as interpretable as others from only looking at the seven most exclusive words, which is why we also look at the most representative titles of each topic to identify the content of the topic. By viewing the titles that have the greatest share of a given topic in the title-topic distribution we can attain better insight of the actual content of that particular topic, and hence it also helps labeling the topics. With this procedure, we are able to label more than 82% of the topics clearly, and only one topic remains completely unnamed. Recall that the corpus used in the model estimation is cleaned from the impact of time- and country-variant words and names, which makes the interpretation task a little harder. This implies that a topic that discusses war, for example, cannot have words like "Vietnam" or "Iraq" included in it. These kinds of context words would make the identification of topics easier, but they would also make topics time, country and event specific. This would make the frequency series possibly increasing only during a single period or in a single country, which we do not want as the topic frequency series should be as common as possible in the sense that their frequencies could be monitored all the time, and they could possibly be relevant in any country or year.

[5] We mark the topics that we could not straightforwardly label as a certain topic with a star in Table C.1.

Fig. 2. Wordcloud of topic labels in the dataset. The size of each label is proportional to its frequency in the corpus.

2.3.2. Association with historical events

Second, we want to make sure that the time series of the topic frequencies captures major historical events related to issues that the topic is supposed to approximate. Next, we look at examples of the topic frequencies TF derived from the topic distributions θ_d. Figs. 3(a)–3(c) plot the frequency of topic 37 labeled *labor, unions and strikes* for the US, the UK and France from 1870 onward. In the United States, the frequency of this topic spikes in 1879, when Knights of Labor started to flourish, in 1886 when the American Federation of Labor (AFL) was founded, during the Great Steel Strike of 1919, and in 1942 when massive labor-intense war material production started. In the United Kingdom, the series captures the great losses and wins of the Labor Party and the mine strike in 1974. In France, the series is elevated during major strikes in 1974, 1984, 1997, 1998, and 2016. It peaks in 1942–1944 during the so-called *service du travail obligatoire*, when thousands of French workers were sent to Nazi Germany for hard labor.

Figs. 4(a)–4(c) plot the frequencies of topic 28 for the United States, Finland, and Japan. The most common exclusive words for this topic are *banking-system-giant-scandal-beats-sector-crises*, which point to banking system difficulties. Thus, the topic was labeled as a *banking crises* topic. For the US, the series spikes during the panics of 1873, 1893, and 1907 and the global financial crisis (GFC) of 2008. In Finland, the series co-moves with similar events, such as the difficulties of commercial banks in 1928–1934, the Finnish banking crisis in the 1990s, and the GFC of 2008. The series is also elevated in 1998 and 2002, which were the years when Finland was accepted into the European Monetary Union (EMU) and when the country started using the euro as currency. Similarly, in Japan, the series rises when the yen was established in 1872, during Japan's financial panic in 1891, the Showa financial crisis of 1927, and most of the 1990s when Japan experienced a severe financial crisis and a recession. Moreover, in 2017 the IMF released a report discussing how the population aging affects Japan's banking sector. The report stated several future threats concerning the Japanese banking sector, especially the small commercial banks, due to declining growth and interest rates implying increasing pressure on the profits of banks. This visual evidence indicates that the topic frequency series captures relevant and significant historical events that are related to the topic.[6]

3. Predicting financial crises

The financial crisis literature has proposed a number of potential crisis indicators, many of which have been empirically confirmed to have predictive power regarding crises. We analyze how these commonly used and accepted financial crisis indicators perform when compared to our newspaper topic frequency variables.

As shown in the previous section, the topic frequencies relate to different economic variables and capture significant economic events. That is why it is possible that these variables have additional, more detailed information on the same issues that the usual leading indicators are supposed to capture, such as asset price and credit booms, and thus improve prediction accuracy. Furthermore, the topics are likely to have information about a much wider set of issues for which we have not had numerical measures (data) previously, which can also improve prediction accuracy. Shiller (2017) argues that many historical crisis episodes

[6] More examples of topic frequency series matched with historical events can be found in the personal website of the corresponding author.

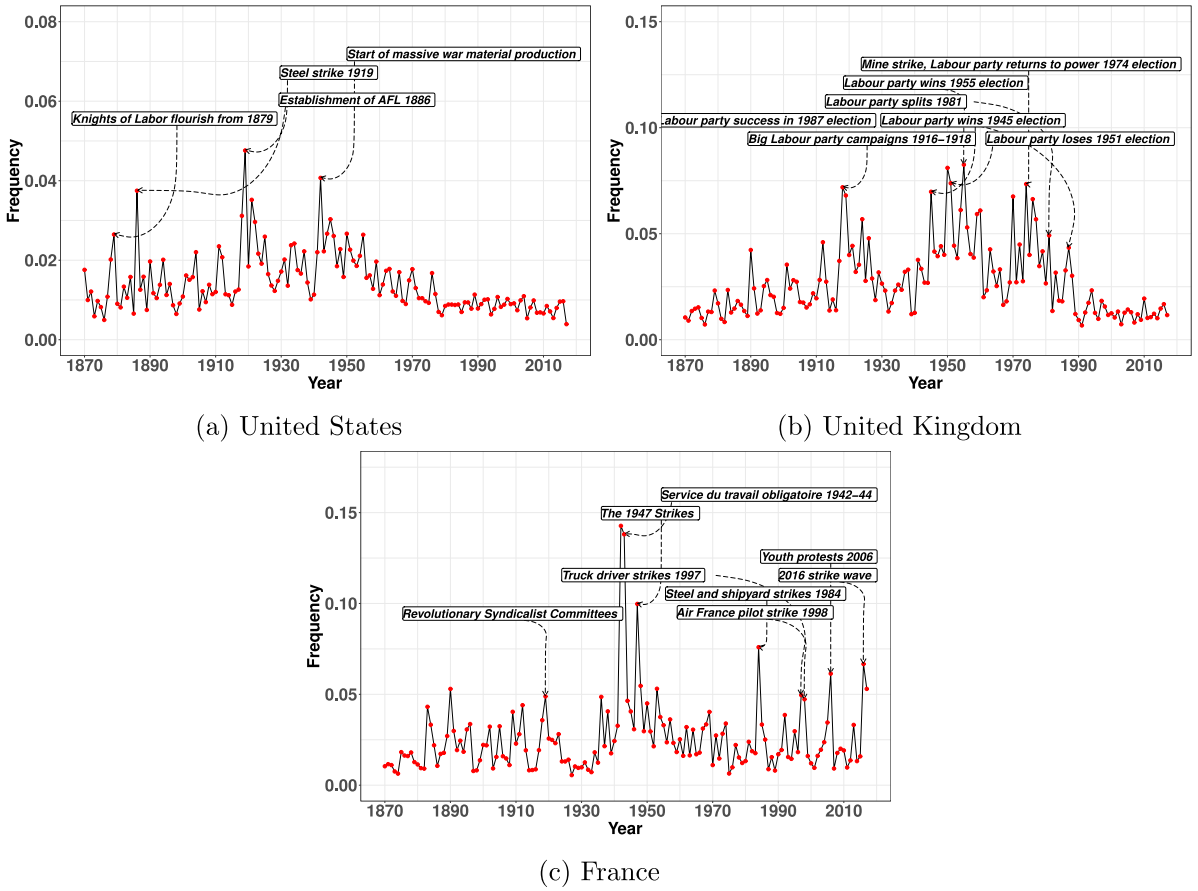

Fig. 3. The frequency of topic 37 *Labor, unions and strikes* and historical events in selected countries.

have been caused, at least partially, by narratives. News articles capture and create general narratives that either receive attention or do not. This attention might last for a long time, fade away, or resurface at a point in time. Our 110 individual topic frequency series can be seen to measure the attention to specific general narratives that are related to the economy, as we do not want to limit the topic model to extract only topics that might be related to financial crises. Therefore, we are able to obtain more detailed information about the relative importance and the coverage of all of the economy-related topics, and thus, the themes that we ex ante do not necessarily associate with financial crises.

3.1. Text-augmented crisis prediction model

The financial crisis prediction or the so-called early warning system (EWS) literature has two branches. The first focuses on improving prediction accuracy via new econometric and machine learning methods, whereas the other focuses on finding new important indicators that can deepen our understanding of this phenomenon. The focus has increasingly been on the latter group of studies that utilize the new long macrohistory datasets collected during the past decade (Reinhart and Rogoff, 2009; Jordà et al., 2017; Baron et al., 2020), whose time coverage spans all way back to the 19th century. This focus on historical data enables these studies to make more general analyses of potential crisis indicators and economic mechanisms behind crises relative to the previously used datasets that cover more countries, but only from the 1970s onward.

Many studies have analyzed the role of credit (Schularick and Taylor, 2012; Jordà et al., 2013; Richter et al., 2021), debt (Jordà et al., 2016) and asset prices (Jordà et al., 2015a,b) on the likelihood of a crisis. The most robust finding is that credit growth possesses superior crisis prediction ability among individual indicators. Most recent advances utilize historical bank balance sheet information (Jordà et al., 2020) and so called red zones of credit and asset price growth (Greenwood et al., 2022) in the prediction model to achieve even better accuracy. As we want to compare the crisis prediction ability of text data/economic news data to the state-of-the-art in this literature, we incorporate the most robust indicators mentioned earlier in a benchmark model, either in the main analysis or as a robustness check depending on the time coverage of data for a specific indicator candidate.

The EWS literature focusing on new indicators relies mainly on utilizing panel logistic regressions that estimate linear decision boundaries in the crisis indicator space to classify crisis episodes from non-crisis years. However, recent studies (Ward, 2017) have shown that machine learning techniques are capable of achieving far better accuracy in this prediction task. To analyze the predictive power of topic frequencies for financial crises, we follow Ward (2017) and estimate a random forest classification model with a binary financial crisis variable as the dependent variable and lagged three-year moving averages of potential crisis indicators as explanatory variables.[7] We estimate the model separately for different pre-crisis horizons of one, two, three, and four years (e.g., the binary pre-crisis dependent variable $Y_i^{(h)}$ is equal to 1, if a crisis occurs within the next h years and it is equal to 0 otherwise) so that we can analyze whether economic news data capture the buildup or the onset of the crisis. In addition, this enables us to compare our

[7] Jordà et al. (2016) use five-year moving averages of the main explanatory variables. Greenwood et al. (2022) use three-year changes, and Schularick and Taylor (2012) use the five lags of the explanatory variables.

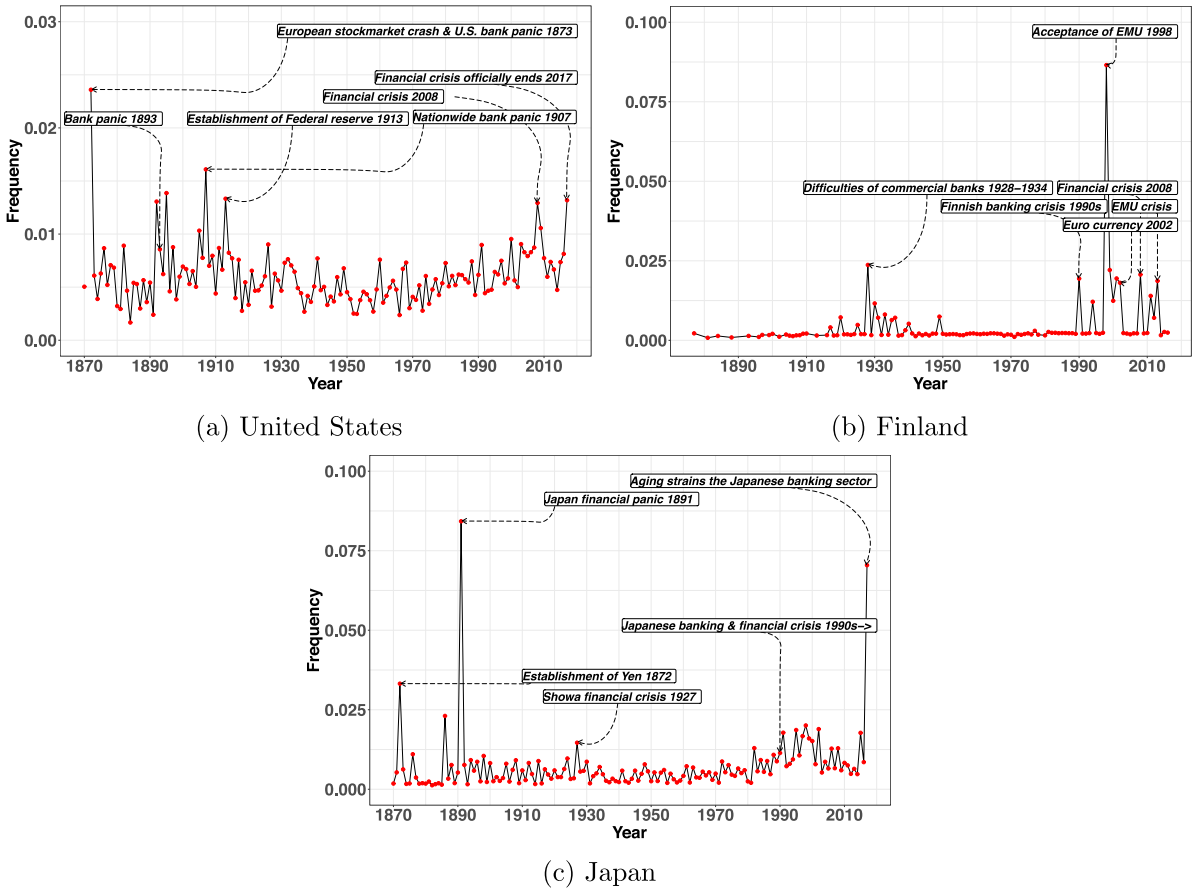

Fig. 4. The frequency of topic 28 *Banking crises* and historical events in selected countries.

prediction results to those of Ward (2017) and Greenwood et al. (2022) who have achieved the best prediction accuracy currently.[8]

A single classification tree is a set of M non-overlapping terminal regions R_m that are formed with discrete rules for the indicator values in a tree-shaped form that aim to separate the pre-crisis observations from the non-crisis periods as accurately as possible. An example of a discrete rule set in a crisis context is that if credit growth is higher than 10% and asset price growth is higher than 15%, then the probability of a crisis is 75%. A classification tree is estimated with recursive partition in the following way. In each partition step $s = 1, \ldots, S$, an indicator j and a partition threshold t_j^s for that indicator are selected so that the discrimination between the two classes is maximized for the two new subregions. The discrimination between the two classes or the impurity of the new regions can be measured with the Gini impurity measure $G = -2p_a^2 + 2p_a$ that has a value between 0 (only same-class observations in the subset) and 0.5 (equal numbers of both class observations in the subset). Here p_a is the share of pre-crisis observations in region R_a. The information gain I from a partition can be measured as the difference between the Gini impurity of the parent region and the average of the two child regions. More formally,

$$I_{ab} = G_{a\cup b} - 0.5(G_a + G_b). \tag{4}$$

The optimal predictor and the partition threshold are selected as those that maximize the information gain measure at that partition step. This procedure is repeated until the final regions include observations of the same class or when a minimum number of observations can be found in the terminal regions. The prediction of the final classification model for observation i is based on the vector of predictive variables X_i. Below $p_m^{(h)}$ is the share of pre-crisis observations $Y_i^{(h)}$ (i.e., is the probability estimate of a crisis) in the terminal region R_m, where observation X_i belongs according to the discrete rule set. More formally, for the tree T,

$$T(X_i) = \sum_{m=1}^{M} p_m^{(h)} I(X_i \in R_m). \tag{5}$$

A random forest (Breiman, 2001) is an ensemble of B classification (or regression) trees where multiple classification trees are estimated separately with a bootstrap subsample of the original dataset and then used to predict or "vote" whether an observation is a pre-crisis period. The share of votes for a pre-crisis period out of all individual classification trees is considered the random forest's prediction of the probability of a pre-crisis:

$$RF_{X_i} = \frac{1}{B} \sum_{b=1}^{B} T_b(X_i). \tag{6}$$

The subsample from which each individual tree training dataset is bootstrapped with replacement is usually 67% of the original sample. The remaining 33% of the observations that are not used in estimating the individual tree are used to obtain a so-called out-of-bag (OOB) estimate of the prediction error during the estimation process. The random forest performs better than a single classification tree because it reduces the variance part of the mean squared error relative to the error variance of an individual tree. This is because the lower bound of the error variance of the random forest (an ensemble of individual trees) depends on the product of the correlation between individual trees and the variance of the error of an individual tree. This correlation between individual trees is made smaller in the random forest first by

[8] Ward (2017) uses pre-crisis horizons of one, two, and three years, whereas (Greenwood et al., 2022) estimate a four-year pre-crisis horizon model, in addition to the shorter horizon models.

estimating each individual tree with a different bootstrap subset of the original data sample and second, by choosing a random subset of the predictors in the dataset in each partition step s of an individual tree T_b.[9]

When estimating the random forest early warning model, we use data for the crisis indicator variable from Baron et al. (2020), who improved previous historical crisis chronologies by including crisis episodes without panics by investigating decreases in bank equity. The potential predictive indicators in the model include change in credit to GDP, return on equity, real GDP growth, inflation, changes in imports, exports, the current account, and the exchange rate between the domestic currency and the U.S. dollar. We also include the change in house prices, bank leverage, the bank non-core funding ratio, and bank liquidity, and the so-called red zones of credit and asset price growth as separate robustness tests, because including these variables decreases the sample size significantly. In addition to these macroeconomic and financial variables from the historical databases of Jordà et al. (2017) and Baron et al. (2020), the topic frequencies that were estimated in the previous section are included as predictors in the model as lagged three-year moving averages.[10] The main data used in the estimation include observations from 17 developed countries for the period 1870–2016 and a total of 72 separate crisis observations.

To compare the prediction ability of the usual crisis indicators and the topic frequencies, we estimate separately a benchmark prediction model composed of the common macroeconomic and financial crisis indicators and then different models that include topic frequency variables with or without the usual macroeconomic and financial indicators. Models that include only topic-based variables can be used to assess how these variables perform on their own. Specifications that include the common crisis indicators can be used to assess whether newspaper topics add to the predictive ability of the model, or in other words, include additional useful information about financial crisis probability in the usual indicators above.

Throughout the analysis, and following the common practice in closely related crisis prediction studies, we assess the prediction performance of each model with the area under the receiver operating characteristic (ROC) curve (AUC). As a brief introduction to the AUC, we predict a crisis when the fitted crisis probability exceeds a probability threshold $g \in (0, 1)$. The ROC curve plots each combination of a true positive rate (sensitivity, crisis years labeled correctly) and a false positive rate (i.e., 1-specificity, "normal" years incorrectly labeled as crises) when varying threshold g. Finally, the AUC, the area under the ROC curve, has values between 0 and 1, where 0.5 means that the prediction accuracy is equal to a random coin toss, and a value of 1 equals a perfect classifier. For a detailed introduction to the ROC curve, see, for example, Berge and Jordà (2011). We estimate each model's AUC with OOB data to obtain as realistic a view as possible on the prediction ability of the model.

3.2. The main prediction results

The OOB data ROC curves for the benchmark prediction model (with the macroeconomic and financial variables only), a model with only topic frequencies (news data), and a model specification with both indicator groups included as predictive variables are plotted in Fig. 5 and reported in more detail in Table 1 for different pre-crisis horizons. The results of the benchmark prediction model are very similar to what Ward (2017) reports with similar data and potential crisis indicators. At a one-year prediction horizon, the AUC value is only 0.626, but with longer horizons of two, three, and four-years AUC values are in the 0.794—0.838 interval with only macroeconomic and financial variables. This is in line with the results of Ward (2017) that the random forest is able to achieve significantly more accurate results than previous studies utilizing the logistic regression framework. However, when topic frequency data are incorporated into the model, the prediction accuracy measured by the AUC value improves around 11 to 13 percentage points, making these early warning models achieve close to perfect classification ability with the AUC values spanning between 0.915 and 0.958 for the longer horizons. These AUC values are even higher than those that Ward (2017) achieves with a larger set of 76 macroeconomic and financial variables. This is clearest with the longer prediction horizons (e.g., three years) where the model with topic frequencies and variables of the benchmark model achieves an AUC value of 0.949, which is almost 7 percentage points higher than the corresponding value from the results of Ward (2017).

Table 1
Random forest based financial crisis prediction results with different pre-crisis horizons.

Horizon (h)	Model	AUC	Difference	p-value
1 year	Benchmark	0.626		
1 year	Benchmark+Text	0.608	−0.019	0.744
1 year	Text only	0.591	−0.035	0.839
2 years	Benchmark	0.794		
2 years	Benchmark+Text	0.915	0.122	0.000
2 years	Text only	0.911	0.117	0.000
3 years	Benchmark	0.838		
3 years	Benchmark+Text	0.949	0.111	0.000
3 years	Text only	0.945	0.107	0.000
4 years	Benchmark	0.827		
4 years	Benchmark+Text	0.958	0.131	0.000
4 years	Text only	0.950	0.123	0.000

Notes: In this table, in addition to the AUC statistics based on out-of-bag (OOB) data in the third column, the statistical difference in the AUC of the benchmark model (macroeconomic and financial predictors only) is tested, with the p-values using the (DeLong et al., 1988) test. The dependent variable is equal to one either the 1, 2, 3, or 4 years prior crisis. The random forest was estimated with 1585 annual observations and 72 crisis episodes, 5000 trees, and $\sqrt{p}$ variables randomly sampled as candidates at each split.

Interestingly, with a short prediction horizon of one year, the topic frequency does not improve prediction accuracy over that achieved with conventional crisis indicators. These results imply that news data have relevant information on the buildup rather than the onset of a crisis. The fact that the AUC values are from OOB data rather than too optimistic in-sample results that are often reported in the literature makes these results even more impressive. In addition, it seems that these improved results can be achieved with only topic frequency data, without including any macroeconomic or financial indicators in the model. The ROC curves in Fig. 5 reveal that the text-augmented prediction models stochastically dominate the benchmark prediction model, with all longer prediction horizons. This implies that no matter what the policy makers' preferences are regarding Type I (fail to predict a crisis observation as a crisis) and Type II (predict a normal period as a crisis period) errors, one would also be better off using the text-augmented models.

Another way of comparing the predictive power of different models is to compare the actual OOB crisis probabilities that the models produce for the different country-years in the dataset. As prediction ability in this context means how well the probability estimates can be used to distinguish crisis years from normal years, it is enough that the model gives a 2% crisis probability for crisis years and a 1% for normal years. In this case, the model can be seen to classify crisis years from normal years perfectly, and it will achieve an AUC of 1.

[9] For a more detailed introduction to random forest estimation, see Ward (2017) and Hastie et al. (2009). As a comparison with the conventional logit model used in past related research, our notation would imply $\log\left(\frac{Pr[Y_i^{(h)}=1|X_i]}{Pr[Y_i^{(h)}=0|X_i]}\right) = \beta^{(h)} X_i$.

[10] We excluded topic 110 from the model selection, as we were not able to clearly label it. In addition, we exclude topics 9, 12, 51, 74, 90, and 94, as the Augmented Dickey–Fuller test could not reject the null that a unit root is present in these time series.

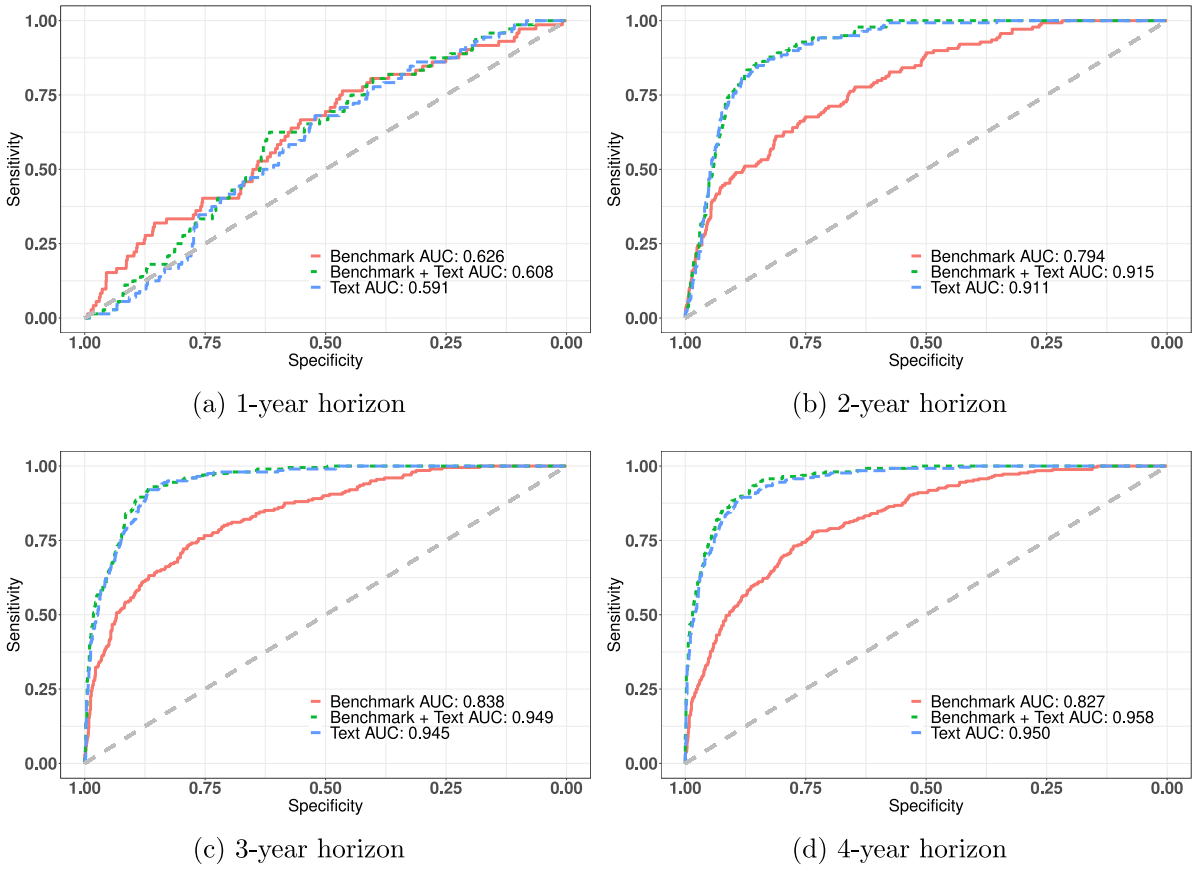

Fig. 5. Out-of-bag ROC curves for random forest financial crisis prediction models including macroeconomic and financial indicators with and without topic frequencies.

However, in practice, when policy makers use these models to assess whether preventive actions are necessary, it would be desirable that the difference between these crisis probabilities is as large as possible, to enhance the interpretability and information signals for a policy maker to trust the models. From this point of view, the topic-based model is better than the benchmark prediction model, as the mean probability given by the model during normal periods is 12 percentage points lower than that given during pre-crisis periods for the latter model, whereas the same difference is 15 percentage points for the former model. In addition, the crisis probabilities are much more volatile during pre-crisis periods for the benchmark model (15% standard deviation) when compared with the text-augmented model (10% standard deviation).

Fig. 6 shows the difference in the average predicted crisis probability between pre-crisis and normal years of the economy for the two prediction models across crisis episodes. Very often, both models–with and without text–show a heightened crisis probability when a crisis occurs relative to the probabilities that the models predict for normal years. However, there are some cases when a crisis was "hidden" from the benchmark model to some extent, and the text-augmented model more or less saw the crisis coming. These episodes include the crises in Australia (1893), Finland (1931), France (1937), Great Britain (1890 and 1973), Norway (1898), Portugal (1931), Sweden (1878), and United States (1907 and 1990). In these cases, the benchmark model predicted a smaller crisis probability relative to the average predicted crisis probability for normal years. The opposite when the variables of the benchmark model "see" the crisis and text data fail does not occur. However, both models failed to give a higher crisis probability relative to normal years during the crises in Japan in 1923 and in United States in 1884.[11] In summary, the text-augmented model is a more accurate predictor of crisis than the one that relies solely on the "usual suspects". Although both models perform well and give an increased predicted probability for the majority of the crisis episode, the fact that the given probabilities are much more stable[12] and reliable for the text-augmented model is very important for practical users and policy makers.

In the next section, we perform an extensive number of robustness checks on this main finding on a substantial improvement in prediction accuracy by using the information contained in the text data (Section 3.3). Furthermore, in Section 4, we make the results more interpretable regarding the specific indicators and their contributions to different and specific crisis episodes.

3.3. Robustness analyses

We assess the robustness of the improvement in prediction accuracy due to the use of topic frequency data by running a large number of robustness checks. These checks are related to crisis definitions, additional crisis indicators in the benchmark prediction model specification, the country and time sample, and different formations of the topic frequency series.

3.3.1. Different crisis definitions and the impact of major crises

In the following set of robustness tests, we use the crisis definitions by Reinhart and Rogoff (2011) and Jordà et al. (2017) instead of the

[11] The predicted crisis probabilities for the entire sample are visualized in more detail in Figs. C.2 and C.3 in Appendix.

[12] In situations where two prediction models show the same average difference in predicted crisis probabilities between crisis and normal periods, users would generally prefer the model with less volatile predictions. This preference is due to the clearer interpretations provided by more stable probability estimates.

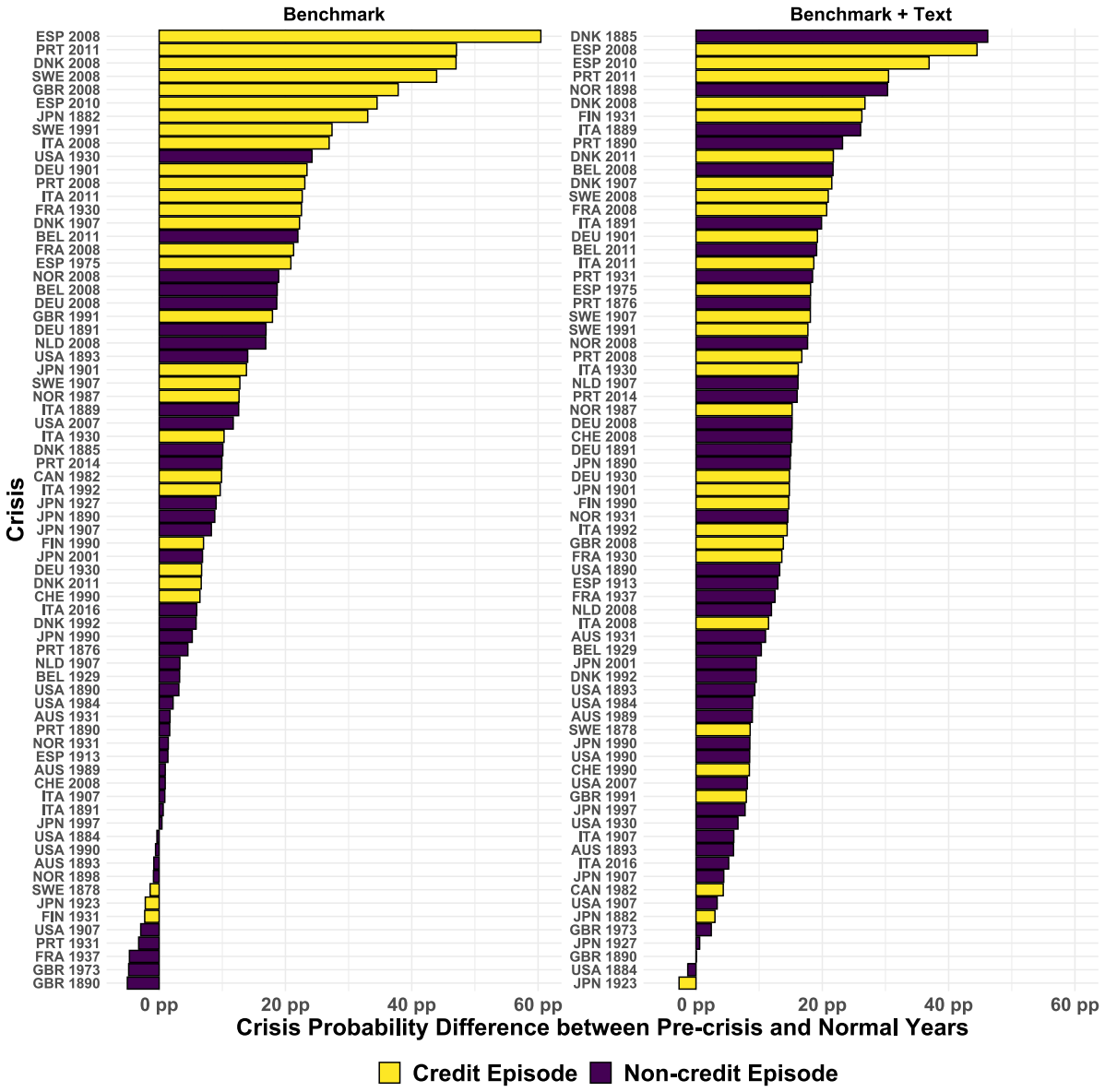

Fig. 6. Difference in average predicted crisis probability between pre-crisis and normal periods across models and crisis episodes. The predicted probabilities are from random forest crisis prediction models that are estimated with a pre-crisis horizon of two years.

definition by Baron et al. (2020). The timing and even the occurrence of a financial crisis can differ between these different datasets, as crises are defined in varying ways and most commonly with narrative information.

The crisis chronology of Baron et al. (2020) used in the main analysis is the most extensive to date as it also uses information about episodes of large bank equity declines to find crises that may not have been associated with a panic. The authors also use all narrative information-based earlier crisis chronologies to update their crisis chronology with the information from bank equity declines. As Baron et al. (2020) include new episodes and exclude episodes that they label "spurious" crises, the chronologies differ considerably. In a vast survey of the crisis literature, as well as different crisis chronologies, Sufi and Taylor (2021) conclude that the use of equity decline data with narrative crisis information is feasible.

The results of these different crisis definitions are combined in Table 2. In addition to different definitions, we want to ensure that the results are not specific to any subset of crises due to the different ways of defining these events. Columns (3) and (4) in Panel A of Table 2 show that this is clearly not the case, as the results remain seemingly the same as in the main analysis in Section 3.1. In fact, the difference between the benchmark and text-augmented models becomes even larger with the other crisis definitions. The difference is especially large with the Reinhart and Rogoff (2011) crisis definition—around 15 percentage points with all longer horizons.

Another important point of view is that the attention to different topics might capture information related only to crises that were related to a panic. This would imply that the attention paid to these topics affects public beliefs directly or indirectly through factors that are discussed in some time period, and finally tip the scale toward a state in which a panic starts easily. We repeat the analysis so that crises with only panics are included in the dataset and present the results in column (4) in Panel B. Again, the results are the same and confirm that the results are not specific only to crises with or without panics.

Another robustness perspective is based on the fact that crises are rare events, as they occur, on average, every 28 years, according to Jordà et al. (2011). Furthermore, around half of the crisis episodes occur in clusters at the same time in different countries. Examples of these clusters include the global financial crisis episodes in the 1930s and 2008. Therefore, we estimate the models again so that we exclude

Table 2
Out-of-bag robustness analyses for random forest based financial crisis prediction results with different pre-crisis horizons.

Panel A						
Horizon (h)	Model	RR	JST	Bank	House	GFC
2 years	Benchmark	0.743	0.797	0.859	0.814	0.784
2 years	Benchmark + Text	0.903***	0.923***	0.928***	0.922***	0.923***
2 years	Text	0.890***	0.910***	0.920***	0.918***	0.922***
3 years	Benchmark	0.784	0.810	0.907	0.862	0.829
3 years	Benchmark + Text	0.938***	0.949***	0.955***	0.956***	0.951***
3 years	Text	0.930***	0.941***	0.948***	0.947***	0.948***
4 years	Benchmark	0.790	0.819	0.910	0.859	0.838
4 years	Benchmark + Text	0.944***	0.951***	0.965***	0.966***	0.961***
4 years	Text	0.939***	0.945***	0.956***	0.960***	0.955***
Obs		1679	1679	1462	1283	1431
Crisis		75	57	65	55	52
Panel B						
Horizon (h)	Model	GD	Panics	Anglo	Pre-WWII	Post-WWII
2 years	Benchmark	0.817	0.808	0.815	0.702	0.873
2 years	Benchmark + Text	0.915***	0.928***	0.920***	0.842***	0.955***
2 years	Text	0.907***	0.917***	0.920***	0.831***	0.947***
3 years	Benchmark	0.859	0.851	0.853	0.774	0.908
3 years	Benchmark + Text	0.944***	0.951***	0.956***	0.907***	0.968***
3 years	Text	0.935***	0.938***	0.948***	0.896***	0.958***
4 years	Benchmark	0.842	0.840	0.844	0.764	0.886
4 years	Benchmark + Text	0.954***	0.956***	0.963***	0.922***	0.972***
4 years	Text	0.949***	0.946***	0.959***	0.909***	0.966***
Obs		1486	1540	1265	540	1045
Crisis		63	59	59	36	36

Notes: *RR*: We use the crisis definition by Reinhart and Rogoff (2011). *JST*: We use the crisis definition by Jordà et al. (2017). *Bank:* We include bank capital, loans-to-deposits and non-core funding ratio in the analysis. *House prices*: We include the lagged three-year moving average of the change in the house price index as a potential crisis indicator. *GFC*: We exclude the global financial crisis from the analysis by including only years before 2005. *GD*: We exclude the years 1929–1936 from the analysis. *Panics*: We include only the crisis episodes that co-occurred with a panic. Information on panics taken from (Baron et al., 2020). *Anglo:* We exclude the Anglo-saxon home countries of the newspapers from which the topic frequencies have been derived. *Pre-WWII:* Time coverage limited to years prior WWII. *Post-WWII:* Time coverage limited to years after WWII.

these crisis years and the surrounding years to ensure that the results are not driven by these two specific worldwide episodes. The results shown in Panel A (column (7)) and Panel B (column (3)) in Table 2 confirm that this is not the case, as the AUC values, their differences, and the corresponding p-values are very similar to the cases above that include these crisis periods. These results imply that the topic-based leading indicators capture general factors that are not specific to these major episodes.

3.3.2. Additional macroeconomic and financial control variables

Next, we include new potential common or promising additional macroeconomic and financial indicators in the benchmark model, in addition to the variables that were included in the main analysis. We include the changes in house prices[13] and bank balance sheet variables (bank capital, loans to deposits, and non-core funding ratios) used by Jordà et al. (2020). In one specification, we also include the so-called red zones for credit and asset price growth in business and the household sector by Greenwood et al. (2022).[14] Columns (5) and (6) of Panel A in Table 2 show that the classification ability of the models that include these additional indicators in the benchmark model is very similar to that in the main analysis. The differences in the AUCs are all statistically significant at the 1% significance level, favoring the use of topics as predictors. In addition, the current state-of-the-art indicator by Greenwood et al. (2022), the credit and asset price growth red zone (Table 3 columns (5) and (6)), does not remove the highly significant difference in prediction ability between the models.

3.3.3. Exclusion of some text information

In Table 3, we also report the results for two model specifications, in which we estimate the topic frequencies in alternative ways. It is possible that the news article titles at the very end of the year might hold very specific information about an unfolding crisis that has been officially dated in the crisis chronology for the next year, although the event was already sort of common knowledge at the end of the previous year. To ensure that our results are not driven by this type of endogeneity issue, we use annual topic frequencies that are composed of each year's news titles that were not published in the last quarter. Again, in this case, the major results and conclusions hold.

Next, we exclude all country-year observations for which there were fewer than 10 article titles. In this way, we can check that the topic frequencies are useful when they are composed of a large number of article titles and when there are only a few article titles responsible for the change in topic frequency. The results show that the improvements in prediction accuracy are still significant in all model specifications with topic frequencies when observations with fewer than 10 article titles are removed.[15]

[13] Specifically, we use the change in the house price index taken from the Jordà et al. (2017) dataset and include it in the model as a lagged three-year moving average.

[14] According to Greenwood et al. (2022) the periods when the three-year growth in house prices and household credit to GDP are in the top quintile and tercile, the household sector credit market is in an overheating state, and this predicts crises. Similarly, the business sector credit markets are in an overheating state or red zone when the three-year growth of credit to businesses to GDP and stock prices are in the top quintile and tercile. Greenwood et al. (2022) show that if a country experiences overheating in both sectors, a crisis is much more likely in the following years. We replicate these variables as closely as we can and include indicator variables for the red zones and their interaction term as the benchmark model.

[15] This excludes more than 20% of the data.

Table 3
Out-of-bag robustness analyses for random forest based financial crisis prediction results with different pre-crisis horizons continued.

Horizon (h)	Model	Few art.	Exc. Q4	RZ	RZT	Sent.	Sent.Δ
2 years	Benchmark	0.788	0.786	0.864	0.808	0.791	0.796
2 years	Benchmark+Text	0.896***	0.919***	0.938***	0.925***	0.914***	0.909***
2 years	Text	0.882***	0.917***	0.937***	0.929***	0.912***	0.905***
3 years	Benchmark	0.836	0.832	0.868	0.835	0.838	0.837
3 years	Benchmark+Text	0.936***	0.954***	0.960***	0.951***	0.949***	0.948***
3 years	Text	0.925***	0.947***	0.952***	0.954***	0.942***	0.939***
4 years	Benchmark	0.828	0.826	0.866	0.842	0.834	0.843
4 years	Benchmark+Text	0.951***	0.964***	0.967***	0.962***	0.957***	0.957***
4 years	Text	0.942***	0.958***	0.963***	0.964***	0.950***	0.950***
Obs		1174	1574	885	1395	1585	1507
Crisis		56	71	36	56	72	70

Notes: *Few art.*: We exclude the country-years for which there was less than 10 articles. *Exc. Q4*: We reform the annual topic frequencies so that news published in the last quarter are excluded. *RZ*: We include the credit and asset price growth red-zones (Greenwood et al., 2022) in the benchmark model. *RZT*: The credit and asset price growth red-zones are constructed using total credit instead of business and household credit separately. *Sent.*: We include the average sentiment in the news titles as a predictor variable. *Sent.Δ*: We include the change in average sentiment in the news titles as a predictor variable.

3.3.4. Country and time coverage

In Panel B (column (5)) of Table 2, we exclude the Anglo-Saxon countries (Canada, Great Britain, and the US) from the analysis, as they are the home countries of the global newspapers from whose news titles the topic frequencies are estimated. The news coverage might vary across these newspapers depending on their home country, which could make the prediction accuracy better for the subsample countries. However, the difference in prediction accuracy remains at around 10 to 12 percentage points.

Jordà and Taylor (2012) argue with convincing empirical evidence that there have been two eras of finance capitalism, in which the first period ending with World War II can be characterized as a time when *money and credit were volatile but over the long run they maintained a roughly stable relationship to each other and relative to the size of the economy as measured by GDP*, whereas the authors describe the second period starting after World War II as *when loans and assets both embarked on a long, strong secular uptrend relative to broad money, and … profound structural shifts in the relationship between credit, money, and output*. As the nature of financial crises may have changed over time, we estimate the same prediction models separately for the pre- and post-World War II samples. The statistically significant difference between the text-augmented and benchmark prediction models remains in both samples, but the gain from using topic frequencies in the model is much larger in the earlier sample, in which the AUC difference spans between 13.3 and 15.8 percentage points relative to the 6 to 8.6 percentage points of the post-World War II sample model.

3.3.5. Newspaper sentiment vs. topic attention

A possible concern related to the prediction process is that the topic frequencies might capture only the general economic sentiment contained in the economic news article titles and that the improved crisis prediction performance is mostly due to the addition of this sentiment measure (indirectly measured) to the prediction models. This would imply that the variables measuring the attention to specific general narratives (topics) are not likely to be important for crisis prediction performance. Instead, adding a news title sentiment to a prediction model would be sufficient to improve the crisis prediction ability.

We test this possibility by re-estimating the prediction models so that the specification that includes macroeconomic and financial variables includes a measure of the average newspaper title sentiment during that year for titles that mention a specific country. As a sentiment measure, we exploit a commonly used procedure that utilizes the number of positive and negative words to form a polarity score for a title.[16]

The results in Table 3 columns (7) and (8) reveal that adding this sentiment score as a possible crisis indicator variable to the benchmark model does not improve crisis prediction ability of a model as the AUC value is basically identical to the benchmark model with only macroeconomic and financial variables. In addition, a model with topic frequencies has statistically higher prediction power over these models as the AUC values are significantly higher after adding these to the benchmark model or when they are used without a sentiment measure and macroeconomic- and financial variables. These results hold in both cases where the sentiment measure is included in levels or in differences (lagged 3-year moving averages). These results seem to confirm the finding that the improved prediction ability is driven by shifts in attention to specific topics, or narratives, and not just by the changes in the sentiment of the newspaper titles.

Overall, this extensive set of robustness tests[17] confirms that the topic-based prediction models significantly outperform the benchmark prediction model, which contains only the usual macroeconomic and financial leading indicators. In the next section, we examine the interpretation of these topics more deeply and how they might contribute to different crisis episodes.

4. Credit booms gone bust?

As we have shown that the topic frequency series have very useful information about an upcoming financial crisis, we next assess which topics are the most important and why. In the random forest modeling

[16] It is common to count only the positive and negative words of a text and then subtract the sum of the latter from the sum of the former to comprise a very simple measure of sentiment. Normalization is often performed by dividing this measure by the total number of words in the text. However, this kind of naive way of calculating the sentiment score does not take into account so-called valence shifter or amplifying/de-amplifying words. A previous word or group of words can switch the sentiment of a word to the opposite direction. For example, the word *good* on its own is clearly a positive word, but if it is preceded by the word *not*, then its sentiment turns negative. Amplifying words like the word *very* increase the sentiment of a word. Again, the word *good* on its own has a positive sentiment, but if it is preceded by the word *very*, then it becomes even more positive. We use the R package qdap (Rinker, 2020) to calculate a sentiment score that takes these factors into account. The algorithm counts the polarity of a single word by taking the context, for example, the surrounding words, into account and then adds all individual word polarity scores together and divides this sum by the square root of the total number of words in the newspaper title. This results in an unbounded polarity score, in which more negative values correspond to more negative sentiment for a newspaper title and vice versa for positive polarity score values.

[17] We included the robustness analysis with out-of-sample predictions in the Appendix A.1.

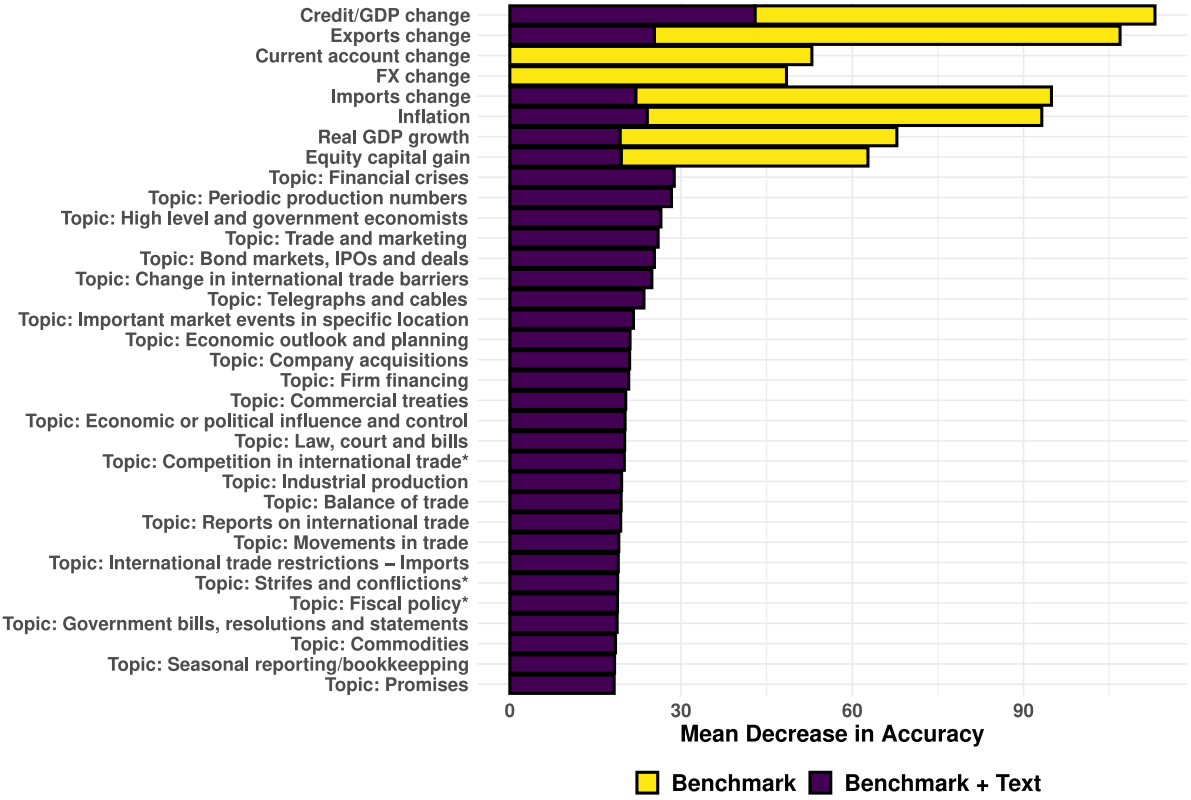

Fig. 7. Feature importance (OOB) for random forest financial crisis prediction models including macroeconomic and financial indicators with and without topic frequencies. Both models are estimated with a pre-crisis horizon of two years.

framework, it is of common practice to calculate the so-called feature importance of each variable in the prediction model

$$DecAcc_{b,l} = [Y^{(h)} - E_b(Y^{(h)}|X_{perm(l)})] - [Y^{(h)} - E_b(Y^{(h)}|X)]. \tag{7}$$

Feature importance is calculated by permuting the values of a single predictor variable l in the OOB sample and collecting the OOB error rate after permuting. The increase in the OOB error rate after permuting the predictor's values is considered the predictor's importance to the individual tree b. This procedure is repeated separately for each variable and for each tree, after which this decrease in accuracy is averaged across all B trees for each predictor

$$MeanDecAcc_l = \frac{1}{B}\sum_{b=1}^{B} DecAcc_{b,l}. \tag{8}$$

The average decrease in accuracy is normalized by dividing it by the standard deviation of the decrease in accuracy across all trees. Feature importance reveals how much a prediction model relies on the values of a specific variable when making predictions.

Fig. 7 reports the mean decrease in accuracy for the top 40 indicators of the text-augmented model and all variables in the benchmark prediction model. It is evident that the change in credit to GDP is the most important variable for the text-augmented model and the third most important variable for the benchmark model after the changes in exports and imports. This is in line with the recent financial crisis literature (Jordà and Taylor, 2012; Jordà et al., 2013; Greenwood et al., 2022) advocating the so-called *credit booms gone bust* explanation for a crisis. However, the text-augmented prediction model puts at least as much and even more weight on certain topics relative to the other macroeconomic and financial indicators found in the benchmark prediction model. In addition, the not-much-smaller feature importance of a large group of indicators in a model that gives very accurate predictions out of bag and out of sample implies that although credit plays a large role in the majority of the crisis episodes, a crisis is most likely a combination of multiple factors. As these factors are quantifiable and contribute clearly to the buildup, and not the onset of the crisis, the story clearly is not that in all cases credit growth alone sows the seeds of a crisis, and then the other factors trigger the crisis. If, in general, the credit boom gone bust is a suitable explanation for crises, our results indicate that other factors measured by topic frequencies provide more detailed information about the credit boom nature—whether it ends ups in a crisis or not.

To examine this hypothesis of *more than just credit booms gone bust* in more detail, we calculate the contribution of each feature to each observation separately. In the random forest modeling framework, it is common practice to estimate this contribution by examining at the path of observation X_i to the predicted value in the classification tree. Throughout the tree, from the root to the terminal regions, there are decision rules that depend on features (crisis indicators). Recall the formal definition of classification trees prediction in Eq. ((5)). This prediction can be decomposed to a bias term and the contributions of each feature j in the following way:

$$f_b(X_i) = bias_b + \sum_{j=1}^{J} contribution_b(X_i, j). \tag{9}$$

Bias is the mean value of the dependent variable in a regression tree or the class rate in a classification tree. The contribution of each feature in a given decision rule point in the tree is either the gain or loss in the prediction value from that decision rule from the prediction value of the parent decision rule:

$$contribution_{RF}(X_i, j) = \frac{1}{B}\sum_{b=1}^{B} contribution_b(X_i, j). \tag{10}$$

In a random forest with many classification trees, the contribution of a feature is the average of the contributions in the individual classification trees (Eq. ((10))).

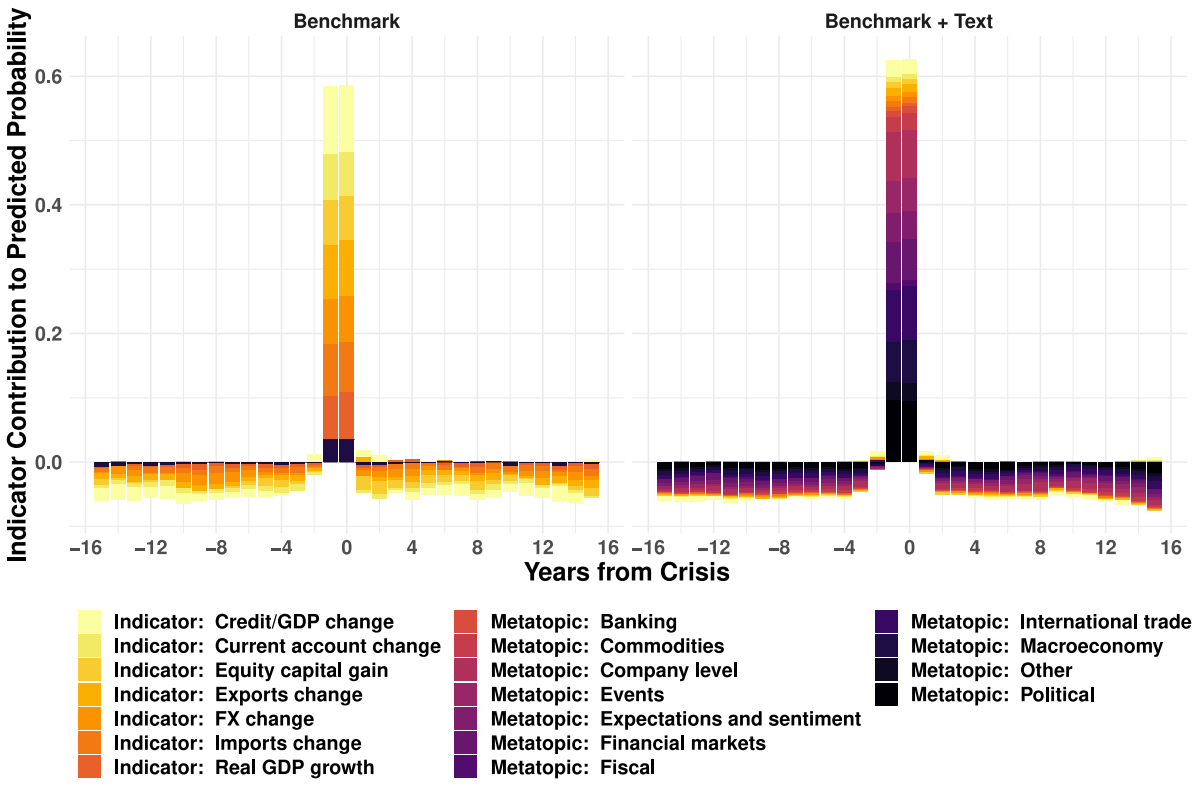

Fig. 8. Average feature contribution to predicted probability 15 years around crisis in random forest crisis prediction models including macroeconomic and financial indicators with and without topic frequencies. Both models are estimated with a pre-crisis horizon of two years.

We compute the contribution of each feature for the observations corresponding to the 30 years around each of the 72 crisis episodes in the full sample. The average values of each indicator are plotted in Fig. 8 so that the averages of individual topics are aggregated to metatopic values.[18] A positive (negative) value indicates that the indicator, on average, increases (decreases) the predicted probability around a crisis. This breakdown of individual crisis predictions shows that the contribution of credit to GDP change is quite modest and not overwhelmingly large relative to the other indicators in the text-augmented model or in the benchmark model without topic indicators. In fact, the text-augmented prediction model relies mainly on topic indicators when making predictions around crisis episodes.

This still does not imply that credit growth is not the main reason for a crisis in the majority of cases. We want to analyze whether there is a large minority of crises in which credit is not the main source of financial instability. In Fig. 9, the contribution of the change in credit to GDP is visualized separately for individual crisis episodes. The contribution is normalized by dividing the contribution to the crisis probability by the predicted probability by the text-augmented model for that specific crisis episode. It is clear that for the majority of crisis episodes, the increase in the predicted crisis probability comes from the change in credit to GDP. However, there is a large group of episodes where credit growth plays a very small role, and the contribution is close to 0%.

Next, we separate crises into credit and non-credit related episodes depending on whether credit to GDP change is among the top five contributing indicators for a specific episode. The average absolute contribution of credit to GDP change is 3.7% among the latter crisis group whereas it is 24.8% on average for the 30 credit related episodes. In Fig. 10, we visualize the average contribution of the indicators that are among the top 15 contributors for at least one of the two crisis groups. The visualization reveals that the contribution of credit to GDP change is over three times larger relative to the second largest contributor in credit related crisis episodes. The top indicators for this crisis group include other macroeconomic and financial variables (equity capital gain, exports change, FX change and imports change) in addition to topics on politics, financial markets and the macroeconomy. The fact that the *financial crises* topic has a high contribution to the predicted crisis probability might suggests foresight. This implies that either journalists or their sources might anticipate an increased risk of crisis before it actually occurs. Interestingly, for the 42 non-credit crisis episodes, credit to GDP growth does not belong even to the top 15 contributing crisis indicators. In fact, the top 15 indicators include 14 topics and only one macroeconomic variable, the change in exports. Fig. 6 in Section 3.2 shows that the so called hidden crisis episodes for which the text-augmented model was able to predict on average higher crisis probabilities relative to normal years, whereas the benchmark model failed to do so, include both credit (2 episodes) and non-credit (8 episodes) related crises. This implies that the improvement in crisis prediction accuracy from utilizing text data is not exclusive to one crisis group. However, the topic indicators help to improve the prediction accuracy of the model especially among the non-credit related episodes.

Although there are episodes in which the credit to GDP change is contributing very little to the predicted crisis probability, it still might be that the 14 topics that contribute the most to the predicted crisis probability for the 42 non-credit episodes still capture a credit boom. Table 4 displays the contribution of these indicators before and after World War II together with their correlation with the actual credit to

[18] We manually categorize each topic into one of the following nine metatopics: Financial markets, Macroeconomy, Banking, Fiscal, Expectations and sentiment, Company level, Political, Other, International trade, Events, and Commodities.

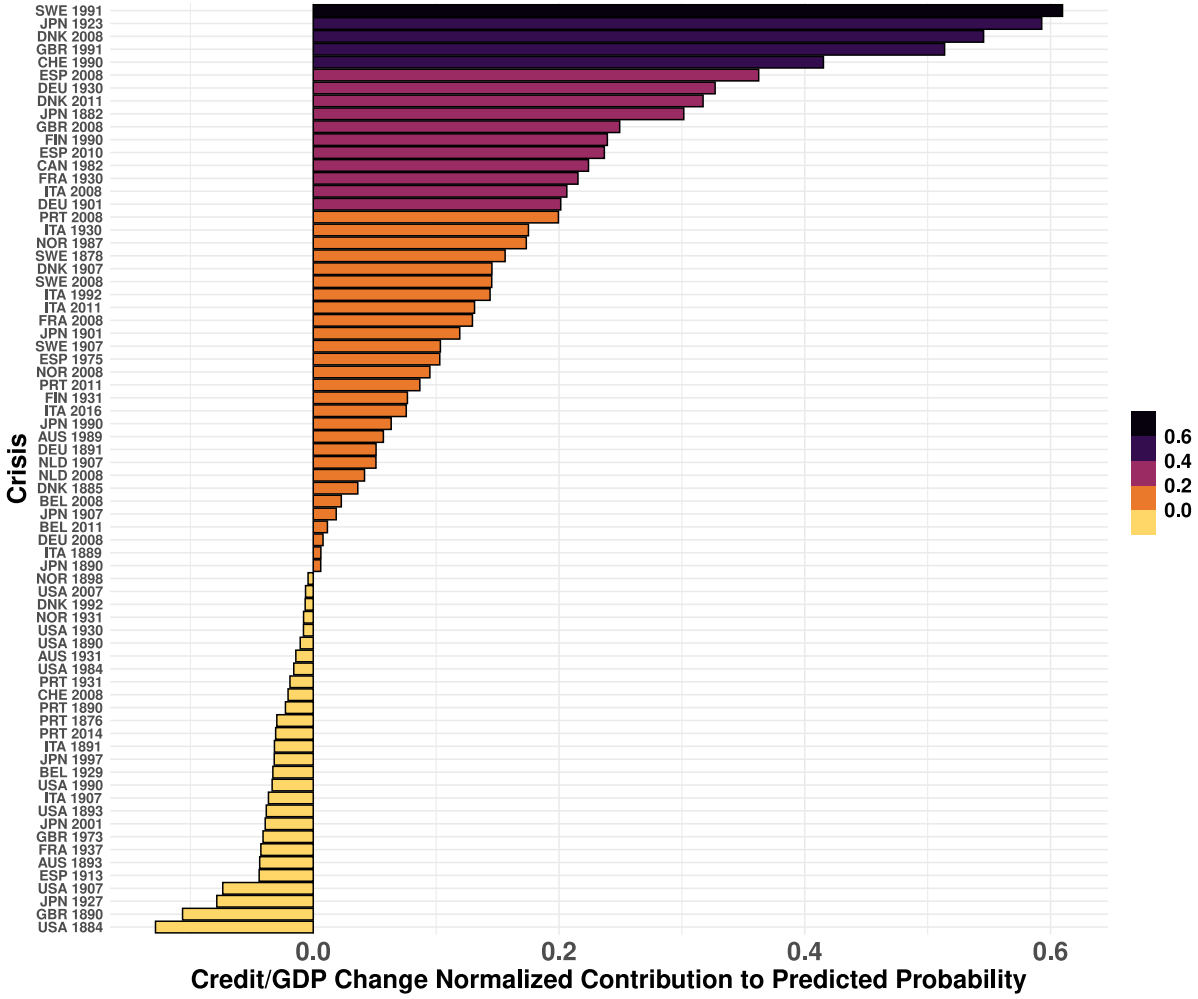

Fig. 9. The contribution of credit to GDP change to predicted crisis probability across crisis episodes. The contributions are calculated from a random forest crisis prediction model that is estimated with a pre-crisis horizon of two years and uses newsarticle topic frequencies, macroeconomic- and financial variables.

GDP change variable. The topics have a very small or close to zero correlation with credit growth before World War II. For only four of these topics, the correlation was statistically significant and in the 0.11 to 0.17 range in absolute terms. After World War II, the correlation is significant for five topics in a similar absolute range. The ratio of the indicator contributions between the two time samples (column (9) of Table 4) shows that the contributions of the top indicators for non-credit related episodes have either stayed similar or increased. The contributions of *balance of trade*, *periodic production numbers*, *exchange rate movements*, *high level and government economists*, and *financial crises* topics have multiplied since World War II and they are statistically significantly correlated with credit to GDP change. Interestingly, three out of the five indicators (including the *financial crises* topic with the second largest contribution) were not correlated with credit to GDP change prior to World War II. In addition, the combined contribution of all non-credit-correlated topics is clearly much larger than the contribution of credit growth in both time samples.

To summarize, the role of credit growth has become more important in predicting financial crises. However, there are episodes for which the contribution of credit growth is close to zero. The indicators that are relevant in predicting these episodes contain both topics that either mildly correlate with credit growth or not at all. Most importantly, the combined contribution of text topic attention is much greater than that of credit growth across all crisis episodes in the sample, and the ones in which credit does not play a role at all, non-credit-related topics contribute heavily to the predicted crisis probability.

The feature importance and contribution analysis indicates that credit growth is still a significant predictor and part of the underlying economic mechanism creating vast difficulties in the financial sector. However, our results show that although credit growth together with the usual macroeconomic and financial indicators of crisis can achieve good results in prediction accuracy with state-of-the-art prediction methods like the random forest, significant improvements in prediction accuracy can be achieved by using topic frequency time series data from economic news titles. In addition, we performed case studies for the crisis episodes where the prediction model does not emphasize the role of credit growth. These case studies can be found in the Appendix A.2 and they reveal that in most cases the *credit boom gone bust* is more or less a suitable explanation for the majority of the cases. However, as the text-augmented prediction model performs better relative to the benchmark model, the news topics likely capture more detailed information about the developments in a country that indicate that the credit boom is a "bad" credit boom. This deduction echoes the fact that not all credit booms end up in crisis, and more detailed information is needed to assess the likelihood of a crisis accurately. In addition, some crises are not preceded by a credit boom, but they are an end result of many factors that contribute to a rise in financial instability.

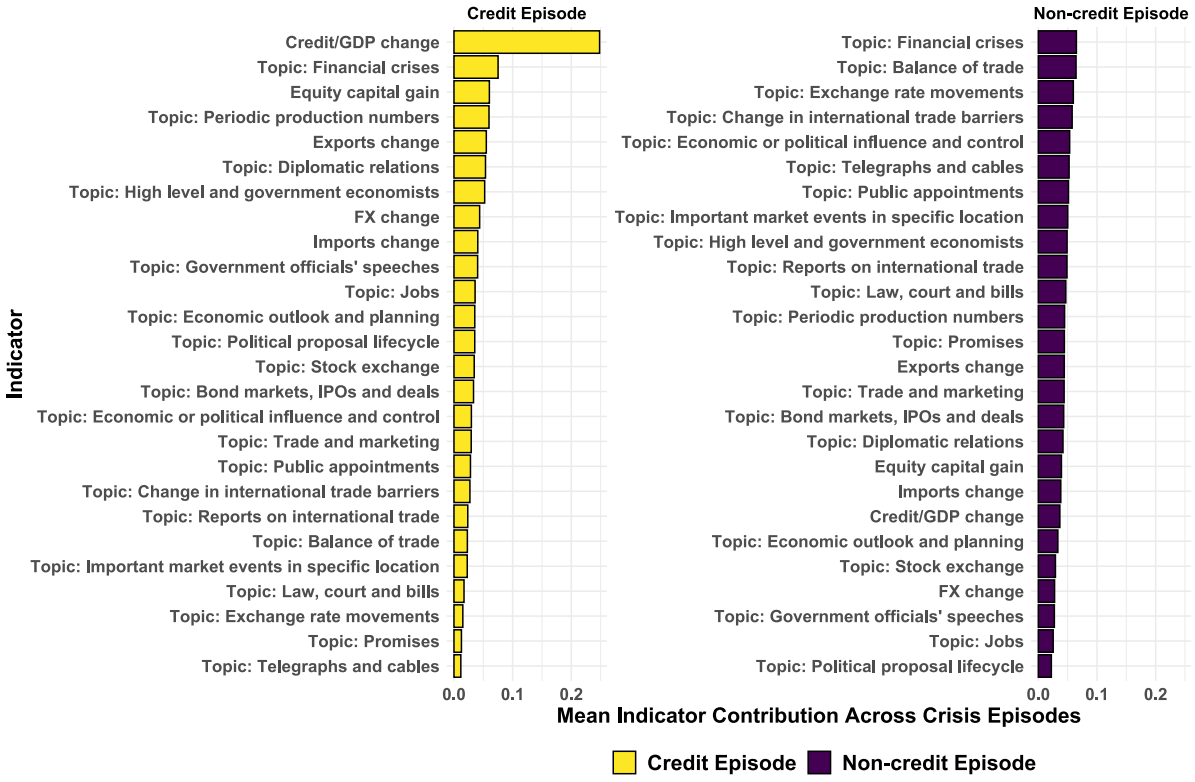

Fig. 10. Contribution of top indicators for non-credit and credit related crises. The contributions are calculated from a random forest crisis prediction model that is estimated with a pre-crisis horizon of two years and uses newsarticle topic frequencies, macroeconomic- and financial variables.

Table 4
Topic correlation with change in credit to GDP and topic contribution to crisis probability for the top contributing topics of non-credit related crisis episodes before and after WWII. The contributions are calculated from a random forest crisis prediction model that is estimated with a pre-crisis horizon of two years and uses newsarticle topic frequencies, macroeconomic- and financial variables.

Topic	Prior WWII			Post WWII			Post/Prior	
	Corr.	*p*-value	Contr.	Corr.	*p*-value	Contr.	Corr.	Contr.
Credit/GDP change			0.139			0.258		1.860
Balance of trade	−0.04	0.328	0.018	−0.15	0.000	0.070	3.75	3.888
Periodic production numbers	0.01	0.787	0.030	0.14	0.000	0.110	14.00	3.703
Exchange rate movements	−0.17	0.000	0.017	−0.11	0.000	0.048	0.66	2.763
High level and government economists	−0.13	0.003	0.040	0.10	0.001	0.100	−0.77	2.495
Financial crises	0.02	0.723	0.109	0.07	0.018	0.195	3.50	1.788
Important market events in specific location	−0.03	0.545	0.045	−0.02	0.437	0.076	0.67	1.699
Trade and marketing	−0.16	0.000	0.029	−0.03	0.383	0.042	0.19	1.439
Economic or political influence and control	−0.03	0.437	0.032	0.04	0.223	0.044	−1.33	1.367
Reports on international trade	−0.04	0.371	0.043	−0.04	0.170	0.054	1.00	1.265
Law, court and bills	0.06	0.145	0.036	−0.01	0.809	0.038	−0.17	1.048
Change in international trade barriers	0.00	0.909	0.036	−0.05	0.132	0.037		1.024
Promises	−0.11	0.014	0.025	0.00	0.906	0.024		0.974
Telegraphs and cables	0.00	0.992	0.034	0.05	0.135	0.030		0.883
Public appointments	−0.03	0.489	0.066	0.00	0.968	0.054		0.820

Notes: *Corr.*: The correlation between values of the crisis indicator and credit to GDP change. *p-value*: The asymptotic p-values for correlations. *Contr.*: Indicator's average contribution to predicted crisis probability. *Corr. in Column (8)*: The ratio of correlations in post and prior WWII. The ratio is not reported, if the numerator or the denominator is zero. *Contr. in Column (9)*: The ratio of contributions in post and prior WWII.

5. Conclusions

We utilize a topic model to identify different economy-related topics found in a group of global newspapers in 17 developed countries for the past 150 years. The topics of the final model are coherent and understandable, and their time series frequencies capture historical events that can be connected to specific economic topics. When this text-based information is included with the usual financial and macroeconomic crisis indicators in a state-of-the-art machine learning financial crisis prediction model, close to perfect prediction ability is achieved for long prediction horizons. The topic information alone achieves nearly similar prediction accuracy, which implies that the attention to economic news captures information relevant to the buildup rather than the onset of a crisis. These results hold in a large number of robustness

checks, including a pseudo out-of-sample forecasting exercise from 1900 onward.

We examine the economic interpretation of the text-augmented model's improved prediction performance by analyzing each indicator's contribution to individual predictions. A visual analysis of the contributions together with crisis case studies on episodes in which the model did not rely on credit growth at all suggests that the *credit booms gone bust* explanation for crises still holds. However, not all credit booms are equal, and the topic series provide important, and more detailed information on the nature of the credit boom and economic developments that possibly increase the risk of a crisis. Although the text-augmented model outperforms the model with the usual macroeconomic and financial indicators across the entire time sample of 150 years and different country and crisis subsamples, certain topic indicators became more relevant after the World War II.

CRediT authorship contribution statement

Kim Ristolainen: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Tomi Roukka:** Conceptualization, Data curation, Formal analysis, Investigation, Writing – review & editing. **Henri Nyberg:** Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing.

Data availability

The authors do not have permission to share data.

Appendix A. Additional analysis

A.1. Robustness analysis: out-of-sample crisis predictions

In addition to the robustness analysis in Section 3.3, we analyze possible overfitting issues by performing a pseudo-out-of-sample prediction exercise in which the models are re-estimated every year from 1900 onward, with only observations that were available until that specific year, with the given constructed text-based series. Then, each model is used to predict the state of the economy in the following year since 1900. This provides an additional perspective on the predictive power of the text topics, similar to a forecasting situation in which the full sample is not used to estimate parameters.

Fig. A.1 and Table A.1 report the corresponding out-of-sample prediction results. As expected, and as typically seen in different applications, the AUC values are not as high in this case as they are for the in-sample or even the out-of-bag sample predictions. In any case, the topic-based models seemingly outperform the benchmark model by 4.6 to 7.2 percentage points. This strengthens the previous OOB results regarding the importance of the text-based topic series as predictors substantially.

A.2. Case studies: non-credit episodes?

Next, we examine a subset of the non-credit-predicted crisis episodes in more detail. We present four case studies, in which we explain in more detail the events surrounding the crisis and how those events might be explained by our topic-based indicators that contributed to the most to the predicted crisis probability that improved the crisis prediction ability relative to the model without text topic indicators.

Table A.1
Out-of-sample AUC values for financial crisis prediction models with different pre-crisis horizons and predictor sets re-estimated every year from 1900 onwards.

Horizon (h)	Model	AUC	Difference	p-value
2 years	Benchmark	0.705		
2 years	Benchmark + Text	0.751	0.046	0.055
2 years	Text only	0.734	0.030	0.304
3 years	Benchmark	0.769		
3 years	Benchmark + Text	0.820	0.051	0.005
3 years	Text only	0.798	0.029	0.199
4 years	Benchmark	0.765		
4 years	Benchmark + Text	0.837	0.072	0.000
4 years	Text only	0.820	0.055	0.003

Notes: In this table, in addition to the AUC statistics based on out-of-sample (OOS) data in the third column, the statistical difference in the AUC of the benchmark model (macroeconomic and financial predictors only) is tested, with the p-values using the (DeLong et al., 1988) test. The dependent variable is equal to one either the 2, 3, or 4 years prior crisis.

A.2.1. Portugal, 1890

The 1890 financial crisis in Portugal was mainly due to a massive accumulation of public debt. This led to a fragile state of the economy that resulted in a crisis in 1890, when the main Portuguese export (wine) decreased. In addition, a financial crisis in Brazil contributed to these events, as it decreased the remittances of Portuguese emigrants. These two factors were the main sources of public income. In addition, the financial crisis shook other Latin American countries and led to problems in international financial markets, especially affecting the Baring Brothers, the main financial intermediary of the state of Portugal. However, the roots of the fragile state of Portugal's public sector and the subsequent crisis were not just in the crises in Latin American countries. The ultimate reasons for the crisis and the accumulated public debt were in the policies driven by the financial minister, Mariano de Carvalho, and the government of *Partido Regenerado*. De Carvalho and *Partido Regenerado* revolutionized the way public financing was conducted in Portugal by, for example, nationalizing the tobacco trade. However, at that time, optimism prevailed in the markets, which meant good creditworthiness and the pricing of Portuguese assets (Lains, 2021).

The topics that increased the crisis probability were *high-level and government economists*, *public appointments*, and *trade and marketing*. The first two topics most likely captured the public discussion about the financial policies driven by the *Partido Regenerado*: The appointments and public statements of some high-level and government economists increased before the crisis, and possibly the party or politicians had also made many public promises to enhance their political agenda, which ultimately led to the crisis. Together with the revolution in the Republic of Brazil, the drastic decline in wine production and exports finally triggered the crisis. The *trade and marketing* topic most likely captured relevant information related to these triggers. The 1890 crisis was preceded by a credit boom in the late 1880s, as the government had borrowed extensively from abroad to finance the modernization of the country. This resulted in the expansion of credit and financial speculation, which finally came to an end due to the events described earlier. Although the text-augmented prediction model does not utilize the credit growth variable when making the prediction, this crisis episode also fits the narrative of a credit boom gone bust. However, the path to crisis was not that simple, as there were multiple factors that contributed to the rise in financial instability and finally, to the start of the crisis. Again, the topics captured the credit boom and the reasons why it would probably collapse.

A.2.2. Norway, 1898

Norway experienced rapid economic growth from 1894 to 1897 due to a general shift in the international business cycle, an increase in exports, and the Norges Bank lowering discount rates in previous

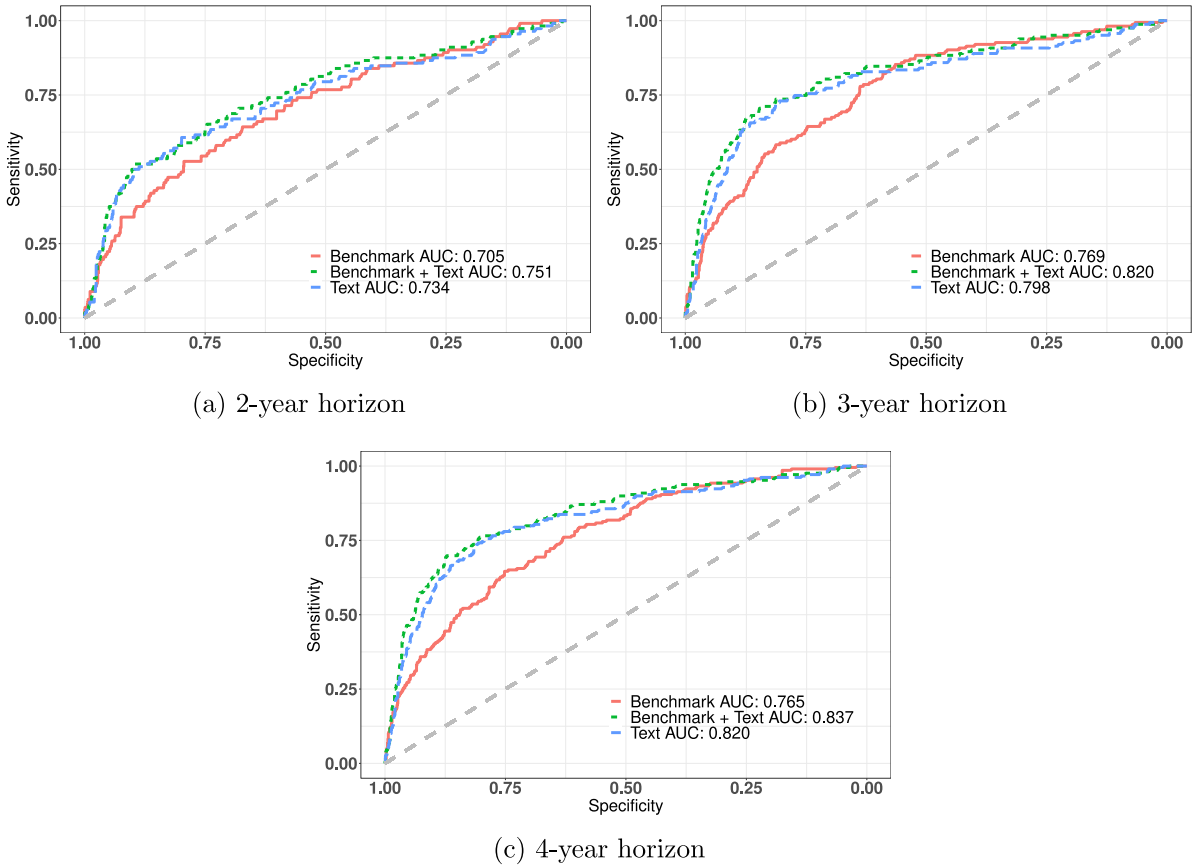

Fig. A.1. Out-of-sample ROC curves for financial crisis prediction models with different pre-crisis horizons and predictor sets re-estimated every year from 1900 onwards.

years. This led to a real estate boom, which was later amplified by the parliament's decision to continue constructing railroads from Oslo to other cities. In 1898, exports decreased dramatically, and the Norges Bank increased the discount rate 1.5 percentage points during that year. The increase in the discount rates and the tightness of the money markets created uncertainty, which with increased bank risk-taking in previous years resulted in a fall in the real estate market and the failure of all relatively new banks in Oslo (Gerdrup, 2003).

Our text-augmented prediction model relies heavily on the topics *promises*, *trade and marketing*, and *economic or political influence and control*. These topics likely captured the government's actions that supported the real estate market, the evolution of exports, and the general rise in positive sentiment in the years preceding the crisis.

A.2.3. Japan, 1927

After World War I, Japan experienced a recession and two crises in 1920 and 1922. In September 1923, the Great Kanto earthquake occurred and collapsed the buildings of many banks and businesses in the greater Tokyo area. Many businesses lost their revenue streams due to the loss of their business premises. Because of this, they could not pay back their loans, and the banks suffered heavily from these so-called earthquake bills. The loan losses were one of the major causes of the 1927 financial crisis. The crisis was triggered by a speech by the minister of Finance for the Diet (Japan's national legislature), in which he falsely stated that one of the major banks—the Tokyo Watanabe bank—had failed that day. This resulted in a banking panic, in which Tokyo Watanabe and one of its affiliated banks collapsed. Later, the panic spread, as people were worried about which banks were also holding earthquake bills. In addition, a new panic occurred when the banks that had earthquake bills were disclosed to the public. In a two-month period, around 30 banks collapsed due to these panics (Yabushita and Inoue, 1993; Hunter, 2014).

The text-augmented prediction model relies most on the contribution of the topics *bond markets, IPOs and deals*, *Financial market movements*, *reports on international trade*, and *specific industry*. The first two most likely captured the financial risk spanning from the earthquake bills, whereas the latter two could be related to the macroeconomic movements in Japan after World War I and the two earlier Japanese financial crises in the 1920s. This does not fit the credit booms gone bust narrative, as the crisis was mainly due to a number of subsequent panics because of fears of bad loans related to a natural disaster.

A.2.4. Great Britain, 1973

Great Britain experienced a financial crisis in 1973 due to the government's previous expansionary economic policies to sustain economic growth. However, this resulted in increasing inflation pressures, rising interest rates, and ultimately, a decline in economic growth. The balance of trade was deteriorating in the late 1960s due to increased government spending, inflation, and a decrease in productivity. In addition, the markets and economists were raising doubts about the government's ability to sustain these policies. The expansionary economic policies also led to boom in the housing market. Banks had been giving loans with housing as collateral, with the assumption that house prices were going to increase. House prices collapsed when interest rates were raised, which resulted in many banks holding loans in which the value of the collateral was smaller than the loan. Many loans were not paid back, and multiple banks were bailed out (Morrison, 1974; Bank of England, 1978).

According to the text-augmented prediction model, the topics *balance of trade* and *economists' expectations* contributed the most to the predicted crisis probability in this case. Again, these topics fit to the

narrative. Although the historical explanation partly includes credit growth among other factors as causes of the crisis, the model emphasizes the importance of the topic indicators, giving more detailed information on the buildup of the crisis. It might be that other indicators (e.g., topics) capture the credit growth, and these most important indicators reveal the heightened risk of it being unsustainable.

Appendix B. Text collection and preprocessing

B.1. Text collection procedure

The news article texts used in the paper were collected from Proquest Historical newspapers and Proquest Newstream databases. The articles were downloaded from the year 1870 to the most recent year that was available. The four newspapers from the historical newspapers database had articles from 1870 to 2010 (*Times of India*) or from 1870 to 2016 (*The Washington Post, The New York Times, and The Globe and Mail*), and the remaining two newspapers had articles from 1982 to 2017 (*The Wall Street Journal*) or from 1989 to 2017 (*The Guardian*). The purpose was to use search keywords that would maximize the likelihood of selecting articles related to the economy. The list of keywords is as follows: *economy, economic, economics, economist, finance, financial, financed, financing, business, production, product, producing, productive, productivity, producer, trade, trading, trader, traded, export, exporting, exported, exporter, import, importing, imported, importer, employment, employed, employee, employer, unemployment, unemployed, jobless, job, industry, industrial, output, depression, recession, deficit, forecast, forecasted, forecasting, forecaster, market, GNP, gross-national-product, GDP, gross-domestic-product, investment, investing, investor, inflation, consumer, consumption, asset, stock, CPI, consumer-price-index, bank, banking, banker, interest-rate, currency, projection, manufacturing, manufacture, manufactured, manufacturer, exchange, boom, boost, devaluation, outlook, prospect, price-level, company, corporation, enterprise, retail, retailer, factory, labor, work, works, worked, working, worker, fiscal, supplier, commerce, commercial, sale, income, profit, revenue, earning, expenditure, expense, spending, resource, dividend*, and *purchase*. Articles were searched so that the results had to include at least one of the mentioned "economy" keywords and a keyword related to a country name. The keywords for the 17 countries were *Australia, Australian, Belgium, Belgian, Canada, Canadian, Denmark, Danish, Finland, Finnish, France, French, Germany, German, Italy, Italian, Japan, Japanese, Netherlands, Dutch, Norway, Norwegian, Portugal OR Portuguese, Spain, Spanish, Sweden, Swedish, United Kingdom, UK, U.K., Great Britain, England, British, United States, USA, U.S., Switzerland, and Swiss*. In addition to these two keyword groups, the searched articles had to be from a specific year from 1870 to 2016. The metadata of the 200 most relevant articles with these keywords for each year and country combination were downloaded to get a panel of articles (200 economy-related texts for every newspaper-year-country combinations). The metadata includes the title, abstract, publication date, start page, and document type of each text article. We narrow the texts to front page articles, articles, and editorials. For the two newspapers with articles with shorter time spans, we also include cover stories, news, features, and commentaries. The searches were conducted between February 2020 and May 2020.[19] There were a few newspaper-country-year combinations that did not have any results, but the majority of the combinations had the full 200 results.

B.2. Text preprocessing

We describe the necessary text preprocessing procedures conducted on the full set of articles originally gathered from the Proquest database before we fit the topic model or preform any further analyses.

[19] The specifics can be provided on request from the authors if needed.

Manual audition of the texts

As the first step before the articles were passed to an extensive text cleaning phase, we glanced at some randomly picked articles. We immediately noticed that the corpus included articles that did not exactly describe the content that was intended by containing the specific keyword or country. This was mostly because the articles might include *some* form of the specific keyword or country name, e.g., "france" - "franceskos", "company" - "accompany". As a result, we decided to change some of the keywords and country-specific indicators to more *precisely* defined, that is, no other forms of a certain word were allowed, but the exact form of the word itself. In addition, we noticed that words related to the base word "work" were frequently not related to the economy, as those articles included a large share of news about some artistic and cultural piece of work or performance; thus, these keywords were removed from the title keyword search.

Consequently, we decided to do a systematic manual audit of the articles included in the original corpus to see if some of the keywords still resulted in articles that are frequently related to something other than the economy. In many cases, these "bad" articles included news about culture and sports. For the audit, we took all the articles from a certain decade and country and manually checked them one by one to see if they were economy related to at least some level and what the keyword was if the article was considered "bad". We tried to include as many decades and countries as possible to avoid timing or location bias, as some topics (e.g., cultural) might be more relevant during some times and in some places than others. The audit sample of articles that we checked manually consisted of a total 5727 articles. We kept track of if the article was "bad" according to our criterion and what keyword was found in that article. As a result, the share of articles that we labeled "bad" was 6.1%, and we did not exclude any further keywords.

Furthermore, during the audition, we noticed that the length of the abstracts varied considerably, especially across years. Thus, we decided to use the titles of the news articles instead of abstracts. This was also because that usually the title includes the main information of the news, and the actual text might bring up more random noise in the data. We decided to form our corpus by using the keyword and word list and that those were have to appear in the title of the article; that is, for the inclusion of an article, it had to mention at least one of the keywords and countries in the title. With this definition, we wanted to increase the probability that the included titles "make sense".

Text cleaning

Before we can actually use the historical newspaper data for the analysis purposes, the text data must be "cleaned" thoroughly. This is done before forming the document-term-matrix (DTM), which is used in the fitting of the topic model. It is particularly important to remove all the unnecessary signs and words as we want to identify interesting and distinct patterns across the titles apart from random noise. We did the text preparation and cleaning with the following steps:

(1) We use so-called uni-grams, one-word terms, to capture the essential information regarding the economic news.

(2) We remove any special characters, numbers, URLs and a list of common English stop-words (e.g., "about", "each", "have", "very").

(3) We remove a custom list of proper nouns, such as names of people, places, companies, etc.

(4) We remove text written numbers, months, weekdays, and currencies.

(5) We remove all terms of three signs or less.

(6) We include only words that occur in all three of the 50-year periods (1870–1919,1920–1969, 1970–2016) and at least in 8 out of the 17 sample countries, thus, implying vocabulary that is independent across different time periods and countries.

(7) We remove all articles that contain no words that are included in the final corpus (technical issue as the zero rows in the DTM do not add any information to the topic model).

Appendix C. Additional tables and figures

See Tables C.1–C.3 and Figs. C.1–C.3.

Table C.1
Labels and most common words of topics 1–55.

Topic	Label	Title
1	Positive market movements[a]	Gain-total-record-purchases-larger-farm-jump
2	Seasonal reporting/bookkeepping	Review-weekly-spring-reviews-books-features-failures
3	Industrial production	Output-production-industrial-speed-newsprint-aluminum-copper
4	Economic growth	Economy-growth-spur-signs-grows-slowing-pace
5	Central bank	Central-bank-governor-institution-raises-eyes-credit
6	Fiscal policy[a]	Cuts-defense-plans-spending-worried-tourists-program
7	Surveys	Finds-survey-study-profitable-manufacturer-edge-poll
8	Economic atmosphere	Confidence-sentiment-game-farmer-mood-caution-spree
9	Commercial banking	Banks-savings-accounts-cash-lending-institutions-fail
10	Specific Industry	Industry-textile-competition-silk-film-iron-electrical
11	Political proposal lifecycle	Proposals-representatives-resistance-pledges-restored-freedom-cease
12	Investors' economic outlook	Investors-briefs-look-wary-feel-yields-interest
13	Culture[a]	Asset-dilemma-times-management-cultural-weighing-consequences
14	Commercial treaties	Commercial-relations-treaty-reciprocity-treaties-interests-arrangement
15	Colonial issues	Colony-floor-cutting-pearson-assembly-resented-armed
16	Investigations and charges	Inquiry-trader-involves-investigating-charges-probe-bribery
17	Competition in international trade[a]	Revival-dispute-losing-scope-returns-pushing-consul
18	Important market events in specific location	Topics-markets-incident-wall-comment-elsewhere-local
19	Crime or abuse revealed	Banker-wife-book-admits-leave-charged-arrested
20	Telegraphs and cables	Latest-telegraph-transatlantic-prince-cables-cable-burned
21	Plans and ideas[a]	Idea-whose-forward-importing-stress-revived-accept
22	Suppliers[a]	Major-appears-supplier-becomes-bought-pork-steadily
23	Money markets[a]	Money-current-call-events-tight-developments-speculation
24	Unemployment rate	Rate-jobless-unemployment-highest-lowest-discount-drops
25	Economic policy agreements and cooperation	Economic-ties-problems-impact-policies-stability-summit
26	Prospects	Prospects-mail-bright-letter-poor-brighter-good
27	Economists' expectations	Expect-economists-predict-believe-optimistic-forecast-weighed
28	Banking crises	Banking-system-giant-scandal-beats-sector-crises
29	Bond markets, IPOs and deals	Enter-watch-offering-bacon-enters-expanding-bond
30	International trade restrictions - Imports	Curbs-import-restrictions-barriers-retaliation-quota-eased
31	International trade restrictions - Exports	Drive-export-prepares-uranium-doubtful-world-wide-approved
32	Strifes and conflictions[a]	Satisfaction-source-supports-indications-seeds-renews-dictator
33	High officials and representatives	Prime-minister-dinner-resigns-foster-quits-ambassador
34	Taxation	Taxes-income-fixed-taxation-affects-levies-billions
35	Counseling and lecturing[a]	Extent-bound-lakes-seizures-mountains-engage-enterprises
36	Subsidiaries and business actions[a]	Foundation-popular-disposal-initiative-loyal-completes-opposed
37	Labor, unions and strikes	Labor-party-unions-leader-strikes-reds-vote
38	Inflation	Consumer-inflation-check-leaving-wage-restraint-perils
39	Company acquisitions	Company-maker-phone-telephone-electric-venture-trust
40	Analysis and outlook[a]	Improved-avoided-alarm-tumble-hurting-term-response
41	Movements in trade	Expects-hopeful-drop-benefit-wars-district-agency
42	Battles, hostilities and war	Fight-style-winning-battle-supremacy-hostile-breaking
43	Foreign investments	Investments-lift-foreigners-changing-shift-safe-welcomes
44	Firm level sales and purchases	Importer-shut-firm-option-buys-sold-owner
45	High level and government economists	Economics-politics-sights-lesson-lessons-search-nobel
46	Financial market movements	Trading-issues-advances-quiet-list-rails-mixed
47	Alcohol	Wine-wines-liquor-grievances-spirits-beer-whiskey
48	Government bills, resolutions and statements	Colonies-bill-senate-measure-islands-senator-passed
49	Government level conflicts and wars	Story-hear-false-rumored-prisoners-disasters-disappointed
50	Reports on international trade	Exports-imports-exceed-curbed-exceeded-beef-rice
51	Countries' economic relations	Large-finance-postwar-retains-responsibility-thus-autonomy
52	Balance of trade	Deficit-billion-widens-payments-budget-surplus-widened
53	Trade policy	Policy-fiscal-industries-defends-adopt-technical-fair
54	Rearrangement of firm operations and mergers	Firms-companies-concerns-insurance-invest-chain-mergers
55	Noble metals	Gold-metal-standard-silver-reserves-reserve-shipment

Notes:
[a] Indicates that the labeling a specific topic was not as clear as for the majority of topics.

Table C.2
Labels and most common words of topics 56–110.

Topic	Label	Title
56	Exchange rate movements	Moves-currency-falls-devaluation-single-traders-closes
57	Transactions	Horses-class-purchased-water-bread-kind-chiefly
58	National supplies shortage	Needs-food-supply-requirements-adequate-assure-material
59	Agriculture and nature	Live-disease-timber-sweeping-breeders-poultry-diamonds
60	Pacts and treaties	Pact-accord-agreement-sign-chamber-signed-pacts
61	Freezing and overtaking of assets	Assets-sell-blocked-claim-freeze-property-blocks
62	Recession[a]	Felt-recession-unlikely-predicted-toll-worry-mild
63	Rumors, stories and propaganda	Spirit-stories-caused-spots-streets-sick-ended
64	Highly specialized products and producers	Enterprise-route-developing-road-conversion-success-things
65	Positive quarterly corporate reports	Profits-earnings-corporate-dividend-reports-dividends-corporation
66	Jobs	Jobs-lose-create-women-creates-employment-eliminate
67	Periodic production numbers	International-annual-journal-cent-manufacturing-awaits-climb
68	Loans and aid[a]	Dependent-less-gasoline-novel-resort-owing-mainly
69	Sovereign seeking economic aid or cooperation	Seeks-halt-bigger-group-access-avert-branch
70	Industrial manufacturing	Manufacture-drugs-process-project-cars-engines-fine
71	Stamps	Placed-stamp-stamps-penalties-adopts-freely-stations
72	Reports on business conditions[a]	Looking-ways-throughout-world-talking-means-shaken
73	Law, court and bills	Supreme-court-case-judge-decisions-decision-settles
74	Antique and design	Built-design-furniture-unexpected-submarines-produces-castle
75	Public appointments	Department-town-division-navy-vessel-officers-corps
76	Economic outlook and planning	Post-war-model-later-tension-proves-carry-believes
77	Promises	Promise-given-failing-breaks-authorized-drawn-advice
78	Free trade difficulties and barriers	Freer-empire-free-merchants-discuss-extend-dominions
79	Wheat, crops and harvesting	Effects-estimates-drought-widespread-corn-acreage-wheat
80	Common market	Back-entry-common-commonwealth-door-bloc-beat
81	Economic or political influence and control	Mine-critics-fate-learned-influence-retreat-summer
82	Diplomatic relations	Reply-grants-granted-powers-issued-encouragement-recognize
83	Change in international trade barriers	Duty-materials-items-cloth-bans-autos-wool
84	Firm financing	Million-spend-link-distribution-mass-spent-least
85	Firm profits	Report-profit-trio-revise-posts-perks-shows
86	Boost of something	Boost-gives-gets-steps-hopes-green-defenses
87	Clash	Holders-strategy-aims-clash-aggressive-pool-reveal
88	Foreign investment	Like-potential-climate-investment-foreseen-shape-welcome
89	Trade and marketing	Marketing-guarantee-differ-widely-advertising-differences-vary
90	Gains and ahead	Ahead-without-puts-tough-keeping-prove-exporting
91	Planes	Bars-planes-remove-blow-coffee-moral-virtually
92	Commodities	Product-domestic-gross-butter-enlarged-dairy-type
93	Economists	Economist-predicts-noted-leading-professor-criticizes-diplomat
94	Global business	Business-around-wire-exporter-area-forum-third
95	Individual solutions and resolutions	Solution-worked-agents-truth-rule-arrive-newspapers
96	Factories and resources[a]	Resources-factories-wealth-rich-natural-vast-development
97	Relief and help	Relief-busy-unemployed-fund-cross-idle-glad
98	Businessmen visits	Businessmen-visit-mean-ability-visiting-businessman-lacking
99	Crises and dangers	Coal-fuel-acute-nears-miners-middle-east
100	Royals and diplomats	Unique-happy-appear-queen-indication-couple-fault
101	Matches and fights[a]	Prospect-hill-sports-invasion-cheered-complications-dark
102	Trade control and regulation	Laws-regulations-tighten-shippers-neutral-stiff-fought
103	Stock exchange	Closing-tone-prices-boerse-quotations-trend-upward
104	Commodity producers	Producers-pleased-longer-tactics-importers-regret-dealers
105	Communities and institutions	Notes-school-city-church-society-college-association
106	Marine ships and other big purchases[a]	Ships-instead-running-size-potash-pictures-voted
107	Government officials' speeches	Praises-speaks-official-snag-practice-tribute-employes
108	Financial crises	Position-difficulties-services-matters-panic-circles-condition
109	Investing[a]	Economies-ending-test-revised-trials-pessimistic-fights
110	Miscellaneous[a]	Board-effect-talk-still-realism-foreign-dominion

Notes:
[a] Indicates that the labeling a specific topic was not as clear as for the majority of topics.

Table C.3
The most representative texts of selected topics.

Topic	Label	Proportion	Text
1	Positive market movements∗	92%	Canada maintains balance in trade: favorable november figures show $8,852,377 over imports.
3	Industrial production	95%	Newsprint output at record level: u. s. and canadian production in may totals 599,412 tons – consumption also up newsprint output at regord level
4	Economic growth	93%	British fear u.s. hurts recovery: effect of strong dollar, weak economy cited british fear u.s. hurts recovery
5	Central bank	93%	Addresses made at the annual meeting of shareholders: the royal bank of canada address of chairman and president
6	Fiscal policy∗	95%	Denmark budgets defense increase: maps spending of $57,100,000 in 2 years–french pledge set at 10% of revenue denmark budgets defense increase france pledges increase norway to double outlay
8	Economic atmosphere	84%	Canadian athletes feeling the spending pinch
9	Commercial banking	92%	U.s. banks balk at long terms for polish debt: u.s. banks balk at extension of polish debt
10	Specific Industry	93%	Japanese automobile industry now faces u.s. competition: japanese automobile industry now faces u.s. competition
17	Competition in international trade∗	92%	East german sees trade 'obstacles': east german sees trade obstacles
18	Important market event in specific location	96%	Wheat is lower: the western market dull and weaker canadian stocks easier wall street securities inactive and easier failures in canada– local breadstuffs and other markets– features in groceries
23	Money markets∗	92%	Current affairs in england: events of the week mr. gladstone at greenwich money market state of trade
24	Unemployment rate	94%	Unemployment rate drops to 6.1 percent: u.s. jobless level is lowest in 7 1/2 years nation unemployment rate hits 7 1/2-year low
25	Economic policy agreements and cooperation	94%	Economic doubt on uniting korea: the problems of german unity are causing seoul to turn cautious. economic doubts emerge on efforts to unite the 2 koreas
27	Economists' expectations	89%	England sees u.s. boom as just begun: british economists assert they are in second prosperity year.
28	Banking crises	92%	Scandal could alter swiss banking: swiss banking facing changes in wake of scandal
29	Bond markets, IPOs and deals	90%	On italy black market, dollar continues strong: dollar is strong on black market
30	International trade restrictions - Imports	93%	France removes 200 import curbs: lifts quotas on goods from u.s., canada and europe france removes 200 import curbs
31	International trade restrictions - Exports	93%	U.s. lifts controls on export of drilling gear to soviets: u.s. permits export of drilling gear
38	Inflation	94%	U.s. fears 'runaway inflation': treasury hints at controls on wages, prices u.s. hints controls on wages, prices
40	Analysis and outlook∗	84%	Near term outlook for u.s. called favorable
43	Foreign investments	92%	U.s. investments upset canadians: ottawa eager to implement law curbing take-overs legislation is spurred canada upset over u.s. investments
44	Firm level sales and purchases	88%	Data sought on thomson deals with libya, iraq: u.s. probes sale of ltv unit to french firm
45	High level and government economists	93%	Norwegian awarded nobel for economics: trygve haavelmo honored for '40s work in science of econometrics norwegian awarded nobel prize
47	Alcohol	92%	Defeating prohibition: one merhod employed on the maine border swedish settlers introduced skin– the cute yankee had them made hollow, and thes brought in whiskey from canada
50	Report on international trade	93%	Heavy decrease in trade with europe: february imports from europe $51,000,000 below 1920, exports down $142,000,000.german trade larger exports to asia and south americadecrease heavily; also imports from them.
51	Countries' economic relations	87%	Finance at large: canada hears from germany last democratic chancellor how germans forgecf constitution
52	Balance of trade	95%	'85 trade deficit is worst ever: $148.5 billion gap expected to intensify pressures on congress '85 u.s. trade deficit hits $148 billion; protectionist pressures seen intensifying
53	Trade policy	94%	Tariff advocated by britsh liberal: ardent free trade advocate, sir john simon, reverses own stand. tariff advocated by british liberal
54	Rearrangement of firm operations and mergers	92%	Set merger to cut costs: usm acquires last 20% of british subsidiary in $34-million deal companies take merger actions
56	Exchange rate movements	94%	Dollar plagued by peso: dollar dips as the peso falls again mexico crisis hurts u.s. currency value dollar slips as the peso continues sharp fall
60	Pacts and treaties	96%	Italy, america sign 'stop-gap' pact on trade: hull announces temporary treaty, effective wednesday. u.s and italy sign trade pact pending draft of new treaty
62	Recession∗	93%	Pain spreads to the heart of europe: euro zone remains mired in six-quarter slowdown as france slips into recession and germany teeters on the brink of one
66	Jobs	93%	Civil service discloses 400 high-pay jobs: log jam of promotions broken; u.s. posts in $11,200–14,000 range civil service discloses 400 high-pay jobs
67	Periodic production numbers	91%	Eurobonds decline on slumping dollar, u.s. retail sales leap — special to the wall street journal
69	Sovereign seeking economic aid or cooperation	88%	Danish concern is go-between for u.s. clients on china deals: danish concern is go-between for u.s. clients on china trade

(*continued on next page*)

Table C.3 (*continued*).

Topic	Label	Proportion	Text
70	Industrial manufacturing	92%	Remarkable activity in oil engine manufacture: three new british oil engines big oil engine plant for b. b. c. north regional station
72	Reports on business conditions*	88%	The advertising world: good market-testing centres hard to find in canada
77	Promises	92%	Canada as pioneer: credit given us for the preferential tariff british manufacturers should take full advantage of the opening–the london outlook on the situation
80	Common market	93%	Snags still exist in kennedy round: u.s. and common market at odds on '3d countries' u.s. and the common market hit snags at the kennedy round
83	Change in international trade barriers	93%	Higher duties on implements: australia tariff changes affect products supplied by canada want reciprocal trade
85	Firm profits	86%	U.s. corporate profit margins up in quarter
86	Boost of something	88%	Gilbert comes through for u.s.: gives americans lead in davis cup; agassi-becker tied gilbert gives u.s. a boost in davis cup; agassi-becker tied
88	Foreign investment	92%	Foreign investment is vexing australia: australia vexed by foreign funds one-quarter foreign ownership
89	Trade and marketing	92%	Session nears end and tory cohorts prepare for fray: but liberals are planning to enlarge canada trade relationship u. s. tariff reactions
90	Gains and ahead	87%	Manufacturing not the only thing being outsourced by the u.s.: ahead of the curve
92	Commodities	92%	Potato is everywhere: from bakery to feed bin, u.s. product is in demand; even farmers speculate in futures market
93	Economists	89%	Hon head lauds canadian employer: hillman predicts faster clothing trade pick-up in canada british outlook here
94	Global business	86%	Business around the world: breweries merging in norway
102	Trade control and regulation	92%	Quarantine abuses.: still further charges against the quarantine officials. further serious charges–violation of the united states revenue laws–spanish vessels and their ballast.
107	Government officials' speeches	88%	Treasury chief defends outsourcing of u.s. work: snow calls the practice 'part of trade'
108	Financial crises	92%	Swiss are critical of german terms: bankers disappointed military government will not permit financial transfers
109	Investing*	90%	Investing in funds & etfs: a monthly analysis — global investing: swiss pros and cons — is the value still there for the loved country?

Notes: Column 3 shows how much of a specific article titles consists of a specific topic. We have selected here examples of articles were the proportion of an article title is over 80% of a specific topic and hence highly representative of that topic.

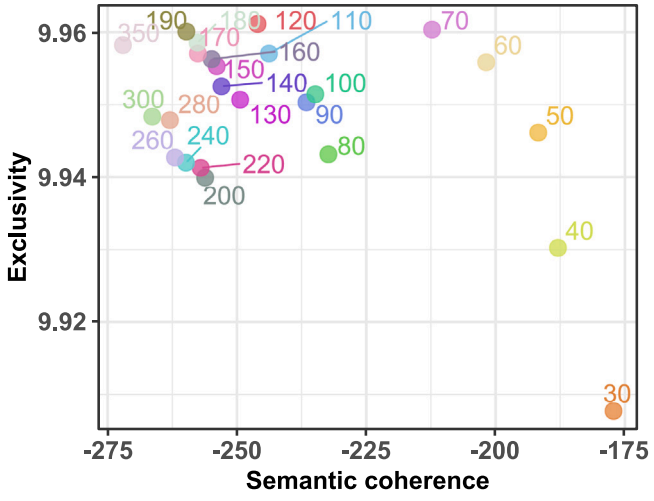

Fig. C.1. The topic model's semantic coherence and exclusivity for different values of K (predetermined number of topics). Semantic coherence measures how often the most frequent words of a given topic co-occur in the news titles; the higher the number of topics, the lower the probability of co-occurring topic top words. Exclusivity measures how unique the most frequent words of a given topic are; the higher the number of topics, the higher the probability of unique topic top words.

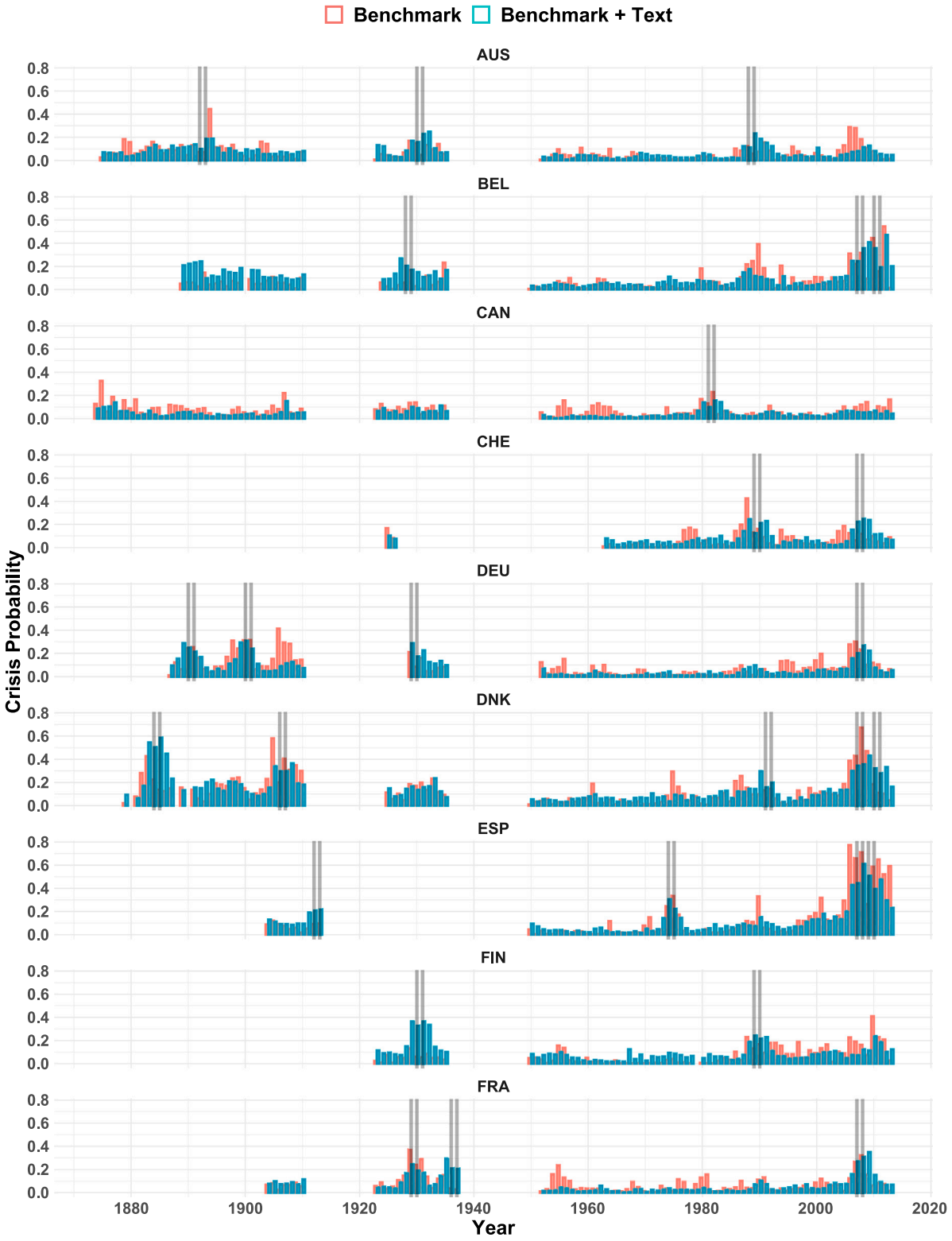

Fig. C.2. Out-of-bag crisis probabilities. The predicted probabilities are from random forest crisis prediction models that are estimated with a pre-crisis horizon of two years.

Fig. C.3. Out-of-bag crisis probabilities. The predicted probabilities are from random forest crisis prediction models that are estimated with a pre-crisis horizon of two years.

References

Airoldi, E.M., Bischof, J.M., 2016. Improving and evaluating topic models and other models of text. J. Amer. Statist. Assoc. 111 (516), 1381–1403.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Q. J. Econ. 131 (4), 1593–1636.

Baron, M., Verner, E., Xiong, W., 2020. Banking crises without panics. Q. J. Econ. 136 (1), 51–113.

Berge, T., Jordà, Ò., 2011. Evaluating the classification of economic activity into recessions and expansions. Am. Econ. J.: Macroecon. 3 (2), 246–277.

Blei, D.M., Ng, A.Y., Jordan, M.J., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bybee, L., Kelly, B.T., Manela, A., Xiu, D., 2023a. Business news and business cycles. J. Finance forthcoming.

Bybee, L., Kelly, B., Su, Y., 2023b. Narrative asset pricing: Interpretable systematic risk factors from news text. Rev. Financ. Stud. 36 (12), 4759–4787.

Correa, R., Garud, K., Londono, J.M., Mislang, N., 2021. Sentiment in central banks' financial stability reports. Rev. Finance 25 (1), 85–120.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.

Bank of England, 1978. The Secondary Banking Crisis and the Bank of England's Support Operations. Quarterly Bulletin Q2, Bank of England.

Gerdrup, K.R., 2003. Three Episodes of Financial Fragility in Norway Since the 1890s. BIS Working Papers 142, Bank for International Settlements.

Gorodnichenko, Y., Pham, T., Talavera, O., 2021. Social media, sentiment and public opinions: Evidence from brexit and USelection. Eur. Econ. Rev. 136, 103772.

Greenwood, R., Hanson, S.G., Shleifer, A., Sørensen, J.A., 2022. Predictable financial crises. J. Finance 77 (2), 863–921.

Hansen, S., McMahon, M., 2016. Shocking language: Understanding the macroeconomic effects of central bank communication. J. Int. Econ. 99 (S1), S114–S133.

Hassan, T.A., Hollander, S., van Lent, L., Tahoun, A., 2019. Firm-level political risk: Measurement and effects. Q. J. Econ. 134 (4), 2135–2202.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hunter, J., 2014. Extreme confusion and disorder? The Japanese economy in the great kantō earthquake of 1923. J. Asian Stud. 73 (3), 753–773.

Jordà, Ò., Richter, B., Schularick, M., Taylor, A.M., 2020. Bank capital redux: Solvency, liquidity, and crisis. Rev. Econom. Stud. 88 (1), 260–286.

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

Jordà, Ò., Schularick, M., Taylor, A.M., 2013. When credit bites back. J. Money Credit Bank. 45 (2), 3–28.

Jordà, Ò., Schularick, M., Taylor, A.M., 2015a. Betting the house. J. Int. Econ. 96, S2–S18, 37th Annual NBER International Seminar on Macroeconomics.

Jordà, O., Schularick, M., Taylor, A.M., 2015b. Leveraged bubbles. J. Monetary Econ. 76, S1–S20, Supplement Issue: November 7-8, 2014 Research Conference on "Asset Price Fluctuations and Economic Policy".

Jordà, Ò., Schularick, M., Taylor, A.M., 2016. Sovereigns versus banks: Credit, crises, and consequences. J. Eur. Econom. Assoc. 14 (1), 45–79.

Jordà, Ò., Schularick, M., Taylor, A.M., 2017. Macrofinancial history and the new business cycle facts. NBER Macroecon. Annu. 31 (1), 213–263.

Jordà, Ò., Taylor, A.M., 2012. The carry trade and fundamentals: Nothing to fear but FEER itself. J. Int. Econ. 88 (1), 74–90.

Lains, P., 2021. A History of Public Banking in Portugal in the 19th and 20th Centuries: the Caixa Geral de DepÓsitos.. Routledge.

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

Larsen, V.H., Thorsrud, L.A., 2022. Asset returns, news topics, and media effects. Scand. J. Econ. 124 (3), 838–868.

Larsen, V.H., Thorsrud, L.A., Zhulanova, J., 2021. News-driven inflation expectations and information rigidities. J. Monetary Econ. 117, 507–520.

Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A., 2011. Optimizing semantic coherence in topic models. EMNLP '11, Association for Computational Linguistics, USA, pp. 262–272.

Morrison, R.J., 1974. The british economy: On the edge of the precipice. Curr. Hist. 66 (391), 101–105.

Reinhart, C.M., Rogoff, K., 2009. This Time is Different: Eight Centuries of Financial Folly. Princeton University Press.

Reinhart, C.M., Rogoff, K.S., 2011. From financial crash to debt crisis. Amer. Econ. Rev. 101 (5), 1676–1706.

Richter, B., Schularick, M., Wachtel, P., 2021. When to lean against the wind. J. Money Credit Bank. 53 (1), 5–39.

Rinker, T.W., 2020. qdap: Quantitative Discourse Analysis Package. Buffalo, New York, 2.4.2.

Schularick, M., Taylor, A.M., 2012. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. Amer. Econ. Rev. 102 (2), 1029–1061.

Shiller, R.J., 2017. Narrative economics. Amer. Econ. Rev. 107 (4), 967–1004.

Sufi, A., Taylor, A.M., 2021. Financial crises: A survey. Working Paper 29155, National Bureau of Economic Research.

Svensson, L.E., 2017. Cost-benefit analysis of leaning against the wind. J. Monetary Econ. 90, 193–213.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. J. Finance 62 (3), 1139–1168.

Ward, F., 2017. Spotting the danger zone: Forecasting financial crises with classification tree ensembles and many predictors. J. Appl. Econometrics 32 (2), 359–378.

Yabushita, S., Inoue, A., 1993. The stability of the Japanese banking system: A historical perspective. J. Jpn. Int. Econ. 7 (4), 387–407.