Does finance benefit society? A language embedding approach

Manish Jha

Georgia State University

Hongyi Liu

Washington University in St. Louis

Asaf Manela

Washington University in St. Louis and IDC

First draft: June 2020. This draft: November 2021.

Abstract

We measure popular sentiment toward finance using a computational linguistics approach applied to millions of books published in eight countries over hundreds of years. We document persistent differences in finance sentiment across countries despite ample time-series variation. Books written in the languages of more capitalist countries discuss finance in a more positive context. Finance sentiment declines one year before rather than after financial crises. Positive shocks to finance sentiment lead to greater GDP and credit growth.

Keywords: sentiment, text analysis, word embedding, BERT, transfer learning, financial crises

---

We thank Yves-Paul Auffray, Joseph Kloecker, David Lindequist, Jose Martinez, Alan Moreira, Rodrigo Moser, Andrea Paloschi, Rita Podledneva, Tatiana Vdovina, Juan Ignacio Vizcaino, and seminar participants at BigNOMICS, CUHK, NHH, Richmond Fed, Temple University, University of Connecticut, Virtual Finance Seminar, and Washington University in St. Louis for helpful comments. E-mail: mjha@wustl.edu, hongyi.liu@wustl.edu, or amanela@wustl.edu. Computations used the Washington University Center for High Performance Computing, which is partially funded by NIH grant S10 OD018091.

> As finance academics, we should care deeply about the way the financial industry is perceived by society. Not so much because this affects our own reputation, but because there might be some truth in all these criticisms, truths we cannot see because we are too embedded in our own world. And even if we thought there were no truth, we should care about the effects that this reputation has in shaping regulation and government intervention in the financial industry. Last but not least, we should care because the positive role that finance can play in society depends on the public's perception of our industry.
>
> (Zingales, 2015, AFA Presidential Address)

1 Introduction

Positive popular sentiment toward finance can spread its benefits widely, while suspicion toward financial services can restrict credit, risk-sharing, and competition (Zingales, 2012, 2015). Survey evidence reveals that trust in bankers fell following the 2007–2008 financial crisis (Sapienza and Zingales, 2012), that such public perceptions often diverge from those of economists (Sapienza and Zingales, 2013), and that low trust can hinder insurance market efficiency (Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer, forthcoming). The relatively short time series of survey data restricts our understanding of how finance sentiment changes over time and differs across countries. While we cannot survey people who lived through the 20th century, books allow us to travel through time and across borders, and to study public perceptions about the benefits of finance to society.

We measure popular sentiment toward finance in an annual panel covering eight large economies from 1870 to 2009 using a computational linguistics approach applied to the text of millions of books. Our finance sentiment index relies on a recently developed language model (BERT, Devlin, Chang, Lee, and Toutanova, 2018) to measure whether references to finance are, on average, semantically closer to positive versus negative sentences. BERT and its offsprings have shattered records on multiple natural language processing tasks, surpassing human ability on many. We use BERT to embed sentences into relatively low dimensional numerical vectors. Following Kozlowski, Taddy, and Evans (2019), we measure the angle between the embedding of sentences mentioning "finance" and the "positive" minus "negative" dimension. This approach goes beyond the dictionary or bag-of-words approach to sentiment analysis (Zhou, 2018) by capturing not only whether a book excerpt is positive or negative, but also the degree to which the *context* of the word "finance" is positive. By aggregating this positivity angle for all finance-mentioning sentences in each language and in each year, we construct a novel finance sentiment panel.

We find highly persistent differences in finance sentiment across languages. Russian finance sentiment is lowest by far throughout our long sample, followed by German, Italian, Chinese, French, and Spanish, with British and American English at the top. Despite considerable within-country variation, this ordering persists throughout our long sample, with the exception of British English finance sentiment, which is slightly higher than American English sentiment until 1912, and slightly lower thereafter. This ordering is somewhat special to finance sentiment, and differs from sentiment toward other industries or toward relatively neutral words, suggesting it is not about general language-specific sentiment.

Generally, books written in languages of more capitalist countries tend to discuss finance in a more positive context. We find that countries with more right-leaning government, with citizens who prefer a more private to public ownership of business, and where registering a business is easier, all exhibit higher finance sentiment. These facts provide validation that text-based finance sentiment works as expected, but at the same time these measures of capitalism explain only a modest share of the variation in finance sentiment, suggesting there is more to it than just attitudes toward capitalism.

Interestingly, Chinese finance sentiment is about as positive as the French one, though more volatile, temporarily plummeting in 1971 when the People's Republic of China (PRC) is admitted into the United Nations, then rising by a similar amount the following year when US President Nixon visits the PRC, and the Shanghai Communiqué is issued in 1972. Other significant changes in sentiment coincide with major historical events, like wars and revolutions.

Having described and validated our measure of finance sentiment, we relate it to financial crises and to macroeconomic fundamentals, and document two new empirical facts. The first is that finance sentiment drops one year before periods of banking distress as defined by Baron, Verner, and Xiong (2021), but remains relatively flat in their aftermath. Consistent with the idea that a deterioration of sentiment toward finance can transform a mild recession into a severe financial crisis (Sapienza and Zingales, 2012), we find that declines in finance sentiment predict a higher probability of banking distress in the following year, even after controlling for lags of credit growth, which has been shown to rise two years before such crises (Schularick and Taylor, 2012).

The second fact is that shocks to finance sentiment lead to higher future output and credit growth. We estimate impulse responses to finance sentiment shocks on GDP growth using local projections (Jordà, 2005). We study GDP growth, an imperfect measure of economic well-being, simply because it is available for all countries in our panel. We find that a one percent improvement in finance sentiment leads to a gradual and persistent increase in GDP growth of about 20 basis points in each of the ten years following the shock. For a subset of countries that excludes China and Russia, we can include credit growth in the local projections. We find that some but not all of the positive effect of finance sentiment on output growth can be attributed to its positive effect on credit growth.

To the extent that these impulse responses identify a causal effect of finance sentiment, they suggest that positive public perceptions of the financial sector indeed benefit society as Zingales (2015) postulates. But macroeconomic turning points could also affect finance sentiment or be jointly determined by other socioeconomic changes. In subsequent work, we analyze how finance sentiment responds to plausibly exogenous natural disasters (Jha, Liu, and Manela, 2021). We find that finance sentiment declines after epidemics and earthquakes, but rises following severe droughts, floods, and landslides. These heterogeneous effects of natural disasters suggest finance sentiment responds differently to the realization of insured versus uninsured risks.

Our paper relates to recent work on the measurement of public attitude toward the financial sector. Stulz and Williamson (2003) find that a country's language and religion predict its creditor rights. Guiso, Sapienza, and Zingales (2008) find that a general lack of trust reduces stock market participation. Giannetti and Wang (2016) document that after the revelation of corporate fraud in a state, household participation, and trust in the stock market decreases. Gurun, Stoffman, and Yonker (2018) find that communities indirectly exposed to a Ponzi scheme withdraw assets from investment advisers. D'Acunto, Prokopczuk, and Weber (2019) find that present-day demand for finance is lower in German counties where historical antisemitism (and therefore distrust in finance) was higher. Levine, Lin, and Xie (2020) link the African slave trade to household demand and trust of financial services. We contribute to this work by providing a novel measure of sentiment toward finance that spans over a century and several large economies, and documenting how finance sentiment relates to financial crises and to economic and credit growth.

A broader related literature considers the measurement of culture and its effects on economic outcomes (Guiso, Sapienza, and Zingales, 2006). Cultural differences can persist for generations (Spolaore and Wacziarg, 2013). Changes in culture, ideas, and in particular, language, have been tied to the dramatic enrichment the world experienced starting in the 19th century (Mokyr, 2016; McCloskey, 2016). Our finding that finance sentiment drops before financial distress supports the idea that changes in language and in sentiment toward productive parts of society have important effects on economic growth.

A recent increase in the availability of textual data has prompted great interest in its use for analysis of culture in particular (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al., 2011), and in economics and finance more broadly (Gentzkow, Kelly, and Taddy, 2019; Loughran and McDonald, 2020). While this literature has yet to study sentiment toward finance or any particular sector, textual analysis has been used to analyze partisanship (Gentzkow and Shapiro, 2010; Luo, Manconi, and Massa, 2020; Goldman, Gupta, and Israelsen, 2020; Engelberg, Henriksson, Manela, and Williams, 2019), product markets (Hoberg and Phillips, 2016; Chen and Sarkar, 2020), central bank communication (Hansen, McMahon, and Prat, 2018; Cieslak and Vissing-Jorgensen, forthcoming), corporate culture (Grennan, 2019), asset market sentiment (Antweiler and Frank, 2004; Tetlock, 2007; García, 2013; Soo, 2018; Ke, Kelly, and Xiu, 2019), employee expectations (Sheng, 2019), financial constraints (Bodnaruk, Loughran, and McDonald, 2015), subjective wellbeing (Hills, Proto, Sgroi, and Seresinhe, 2019), uncertainty (Baker, Bloom, and Davis, 2016; Manela and Moreira, 2017; Goetzman, Kim, and Shiller, 2017; Hassan, Hollander, van Lent, and Tahoun, 2017; Boudoukh, Feldman, Kogan, and Richardson, 2018), emerging risks (Hanley and Hoberg, 2019), emerging technologies (Chava, Du, and Paradkar, 2019), and the link between business news and business cycles (Bybee, Kelly, Manela, and Xiu, 2019).

While early work relied on simple word counts (the bag-of-words approach), recent work starting with Mikolov, Chen, Corrado, and Dean (2013) shows that using neural networks to embed words in vector spaces improves learning algorithms' performance in natural language processing tasks. Kozlowski, Taddy, and Evans (2019) demonstrate that such word embeddings produce richer insights into cultural associations and categories than prior methods. Our work builds and improves on their methodology by using a pre-trained language model designed to capture context (BERT), both to embed sentences mentioning our object of interest (finance) and to define the dimension on which we project these embeddings (positive − negative). This "transfer learning" approach lowers both estimation error and computation costs.

We proceed as follows. Section 2 describes our text-based finance sentiment measure. Section 3 shows how sentiment evolves over time and across countries. Section 4 considers alternative approaches to sentiment measurement. Section 5 analyzes how finance sentiment relates to financial crises and economic growth. Section 6 concludes. Additional details and results are provided in an online appendix.

2 Text-based sentiment toward finance

In this section, we describe our text data and how we measure a language's sentiment toward finance across time. For each language and year, we start with a sample of finance-mentioning sentences published in the language and year. Next, we measure the degree to which each sentence places finance in a positive context. We then aggregate these scores to an average finance sentiment that reflects the mean sentiment toward finance of books written in the language in that year.

We assume throughout that the choice of words used by book authors, magazine publishers, and journalists whose work is archived in libraries, reflects the sentiment of the average denizen of that language during the time, or at least that of an influential literary elite. For example, in our dataset, the sentence "correcting corruption or financial malpractice" appears first in 1951 and then appears every year after 1959. The sentence was part of the 1959 US labor-management reform legislation hearings, when correcting corruption or financial malpractice became an allowable purpose for establishing a trusteeship by labor unions. Hence, "financial malpractice" is more frequently used in subsequent legal documents and books. The context for the word "financial" here is clearly negative. In this particular case, we assume that the labor unions in particular, and the US English-speaking public in general, are more likely to associate finance with malpractice around that time.

2.1 Data

Our text data includes five-word sentences (5-grams) containing the word "finance" across eight languages, between 1870 and 2009, extracted from the 2012 edition of the Google Books Ngram Corpus (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al., 2011; Lin, Michel, Aiden Lieberman, Orwant, Brockman, and Petrov, 2012). The corpus consists of words and phrases and their annual usage frequency from 1500 to 2009. The data originates from Google scanning over 8 million books or 6% of all books ever published in American English, British English, Simplified Chinese, French, German, Italian, Russian, and Spanish.[1]

Although the original data provides lower complexity n-grams counts as well, we focus on 5-grams because for sentiment analysis, especially with BERT, a word's context is essential. We start our study in 1870 (Google corpora is available from 1500) because from that year, we have more confidence in the accuracy of our macro data. Moreover, the number of sentences becomes sparser as we go back in time, and there are fewer mentions of finance before 1870, which increases the measurement error of our sentiment index.

We preprocess the Book Corpus by stripping case, symbols, double spaces, part of speech tags, and positional tags. Next, we extract all sentences mentioning the stem of the word for finance. The finance stem word is different across languages, as listed in

[1]We reluctantly omit Hebrew because its word for finance (Mimun) without niqqud is also the name of Maimonides—a famous Jewish philosopher (Rabbi Moshe ben Maimon).

Table 1: Finance mentions across languages

Language	Finance word stem	Unique sentences	Total sentences
American English	financ	220k	79m
British English	financ	48k	15m
Simplified Chinese	金融, 金 融, 金_融	196k	305m
French	financ	100k	43m
German	finanz	28k	7m
Italian	finanz	23k	9m
Russian	финан	187k	250m
Spanish	finan	89k	33m

Note: We report the number of mentions of the word finance, translated and stemmed, in a five-word sequence (5-gram) for each language in the Google Book Ngram Corpus. Our dataset covers the period 1870–2009.

Table 1. We use the word stem "financ" for English to include sentences that contain either "finance" or "financial." Similarly, for other languages, we use a word stem common to the different verb and noun forms of "finance." For example, for (Simplified) Chinese we use "金融" (financial) but also include base words where there is space and underscore between 金 (gold) and 融 (melt). The filtering yields a set of unique sentences mentioning finance for each language. In our data set, American English has the highest number of unique sentences that mention "financ", followed by Simplified Chinese and Russian. Although Simplified Chinese began being promoted only in the 1950s, the Google ngram data for Chinese has been translated to the simplified version throughout the data set.

2.2 Methods

We measure finance sentiment across languages at an annual frequency using a three-step process. First, we embed each sentence in the corpus into a 768-dimensional vector space. Second, we measure the cosine similarity of this sentence embedding with respect to a language specific positive minus negative embedding. Third, we average the cosine similarity of all finance mentioning sentences in each year, weighted by their frequency. We next describe how we calculate a sentence embedding, the positive minus negative embedding, and their cosine similarity.

2.2.1 BERT

Recent work in natural language processing (NLP) has been increasingly successful in capturing the complexity of language by considering words in sequence rather than in isolation. One of the ways this is accomplished is by representing words as embeddings. Word embeddings are high-dimensional vector-space models of text in which each unique word in a corpus is represented as a vector in a shared vector space (Mikolov, Chen, Corrado, and Dean, 2013). The vector for each word is based on the context the word shares with other words in the sentence. The classic flavors of word embeddings, such as Word2Vec (Mikolov, Chen, Corrado, and Dean, 2015), GloVe (Pennington, Socher, and Manning, 2014), and FastText (Bojanowski, Grave, Joulin, and Mikolov, 2016; Joulin, Grave, Bojanowski, and Mikolov, 2016) rely on the Distributional Hypothesis (Harris, 1954) to capture relationships in the embedding space. The hypothesis states that words that occur in the same contexts tend to have similar meanings, with the underlying idea that "a word is characterized by the company it keeps" popularized by Firth (1957). However, there are certain downsides with these flavors. First, the traditional methods assign embeddings from the ground up; this is an issue for our data set in earlier years, when the number of words in the corpus is less than a million (Altszyler, Sigman, Ribeiro, and Slezak, 2017). Second, while these embedding methods work well for word-level embeddings, they are poor sentence encoders in that they often get the context wrong (Perone, Silveira, and Paula, 2018). Thus, we move away from the traditional shallow neural network methods.

We employ a deep neural network-based natural language processing method, Bidirectional Encoder Representations from Transformers (BERT) developed by Devlin, Chang, Lee, and Toutanova (2018). BERT produces meaningful results even with smaller training data and can provide context for words in sentences. The key advantage of this method over classic word vector models is transfer learning – where a model developed for a task is reused as the starting point for a model on a second task. BERT's neural network is pre-trained on 800 million BooksCorpus and 2,500 million Wikipedia words. Thus, the model knows which words have a similar meaning, based on pre-training. Google applies it to both rankings and featured snippets in search.

While a full treatment of BERT is beyond our scope, we wish to provide an intuitive understanding of this method and the structure that it implicitly imposes on the data. BERT uses Transformers (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017), a mechanism that learns contextual relations between words in a text. The model processes each word in relation to all other words in a sentence, rather than one-by-one in order. BERT is also bidirectional, which allows the model to learn the context of a word based on all its surroundings, as opposed to a unidirectional model, which reads the text sequentially. To train the model from unlabeled text from BooksCorpus and Wikipedia text, BERT employs two strategies: (i) Masked Language Modeling –

where 15% of the input words are masked out and then predicted (ii) Next Sentence Prediction – predict if Sentence B is the actual sentence that proceeds Sentence A. Solving the above two problems using its large corpora, BERT is able to place words in the embedding space. Google shares two versions of the pre-trained model: Base (12-layer, 768-hidden features) and Large (24-layer, 1024-hidden features). Both models are available in a cased and uncased variant. We use the base uncased model for English, and Chinese since the extra efficiency we get from the large and cased model is not significant enough to spend more time and resources on them. For French, German, Italian, Russian, and Spanish, we use the cased multilingual model as recommended by Google Research. Thus, we use BERT Base for American and British English, BERT Base Chinese for Simplified Chinese, and BERT Base Multilingual Cased for French, German, Italian, Russian and Spanish.[2]

The following features of BERT make it especially useful for our purposes: First, BERT comes pre-trained, so it works well out of the box. A pre-trained model is important to us, especially in earlier years of our sample, where the Google Books corpus is considerably smaller. Second, it offers contextualized embedding. For example, the word "bank" has a different meaning in the following two sentences "In a crisis, we could bank on financing from the government", and "Government's financing for the bank is in crisis." The context changes what the author conveys. BERT can distinguish the connotation difference between the two sentences, resulting in different embeddings. By contrast, in classical word vector models each word has a unique embedding. Third, to reduce the number of unique words that feature in the model, BERT breaks each word into smaller subwords or tokens. For example, "wonderful" is tokenized to "won #der ##ful," where # denotes subwords. The dimension reduction is especially important in the multilingual model. Finally, BERT is designed to encode entire sentences, up to 512 subwords. The tokenization process adds [CLS], which stands for "classification" at the beginning of each sentence. The embedding for [CLS] is used as the embedding for the entire sentence that follows it.

2.2.2 Cosine Similarity

A major advantage of word embeddings is that they allow language features (such as words, sentences, etc.) to be treated like vector spaces with intuitive mathematical properties. A common example from Mikolov, Yih, and Zweig (2013) is king – man + woman ~ queen. That is, subtracting the male gender vector and adding the female gender vector to the king vector corresponds to a vector that is close to the queen vector. Thus the word queen could be seen as starting at the word king and then moving in the feminine gender

[2]The pre-trained models are available at https://github.com/google-research/bert

direction. Similarly, we could think of dictator + positive - negative ~ king; here, positive minus negative represent a displacement in the positive direction. Thus, if we start from the dictator vector and move a step in the positive direction, we get the king vector. Other word pairs also correspond to the positive dimension, such as (benefit − damage), (good − bad), (good − corrupt), and (help − hurt).

To define our positive minus negative dimension, we average the sentence embedding differences across sentences containing "finance" or "financial" together with the above words, similar to Kozlowski, Taddy, and Evans (2019). The list of sentences for English (both American and British) are shown in Table 2. The corresponding sentence pairs for other languages are included in Appendix A.1. We focus on the broader notion of "finance", as opposed to more specific financial activities or players (e.g. "bank", "lender", etc.), because this sentiment measure speaks directly to our question of interest, attitude toward finance. More specific related words would be close in vector space to "finance" because they are frequently mentioned together, so we expect them to generate similar sentiment estimates, but each brings along its own identification issues. For example, "bank turmoil" can often refer to the financial institution but also to the contested West Bank territory.

Table 2: Positive − negative defining sentences for English

Positive sentences	Negative sentences
financial services benefit society	financial services damage society
finance is good for society	finance is bad for society
finance professionals are mostly good people	finance professionals are mostly corrupt people
finance positively impacts our world	finance negatively impacts our world
financial system helps the economy	financial system hurts the economy

Note: To define the positive minus negative dimension, we average the embeddings of positive sentences less that of their negative counterparts.

To measure sentiment toward finance, for each finance-mentioning sentence j in language i with embedding s_{ji}, we calculate the orthogonal projection of the sentence vector onto the language-specific positivity embedding p_i using cosine similarity:

$$a_{ji} = \frac{s_{ji} \cdot p_i}{|s_{ji}||p_i|} = \frac{\sum_d s_{jid} p_{id}}{\sqrt{\sum_d s_{jid}^2}\sqrt{\sum_d p_{id}^2}}, \tag{1}$$

where d enumerates the elements of s_{ji} and p_i, both 768-dimension vectors. By construction, the cosine similarity in Equation 1 of two positive vectors is bounded between -1 and +1, with zero indicating a neutral sentence. A more negative cosine similarity indicates that the sentence has a more negative sentiment, while a more positive cosine similarity indicates a more positive sentence.

Figure 1 illustrates this method in a two-dimensional space. The five positive (and negative sentences), from Table 2 bunch together in the embedding space, as similar sentences keep similar companies (Firth, 1957). We take the vector difference between positive and negative sentence embeddings to define our positivity dimension. Next, we project finance sentences onto the positive minus negative dimension. Sentences tend to be close to the dimension, which is closer to their connotation. For example, a sentence such as "financial sector supports economic development" lies closer to the positive sentences, at a smaller angle with the positive dimension.

Cosine similarity measures the position between -1 and 1 where the shadow of a given sentence vector falls. If the sentence has a positive connotation, such as the one in our example, we will have a smaller angle between the sentence vector and the positive dimension. A smaller angle is associated with a higher cosine similarity. On the other hand, for a negative sentence such as "financial malpractices stunted our growth" would be closer to the negative dimension, or $\theta_{ij} > 90°$. Thus the cosine similarity for a sentence with negative connotation is negative. Similarly, a neutral sentence such as "finance lessons from the pandemic" would be roughly equidistant from both positive and negative dimensions ($\theta_{ij} \approx 90°$), and thus have a cosine similarity close to zero.

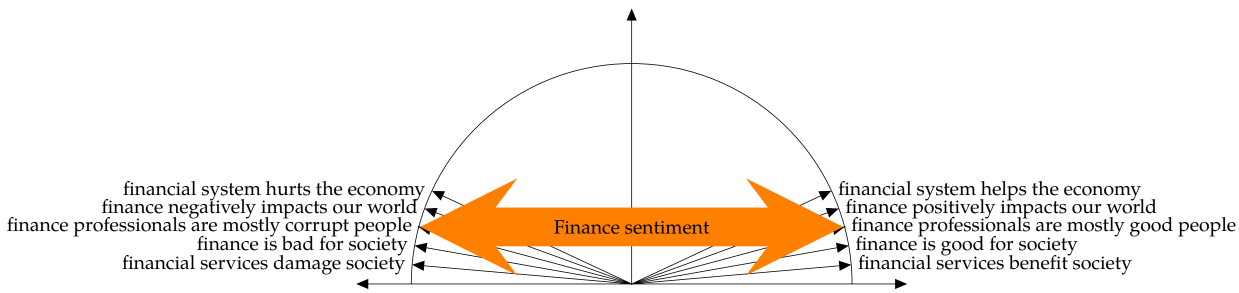
Table 3: Sentences assigned the most positive and negative finance sentiment for American English

Positive sentiment sentences	Negative sentiment sentences
the goal of financial management	turmoil in the financial markets
finance in the graduate school	finances become disordered the
financial support of the center	financial panic swept the country
financial management of the organization	turmoil in financial markets
business and financial experience	financial panic swept the nation
financial support of the graduate	instability in the financial markets
financial support of the science	financial panic in the country
financial support of the course	severe financial setbacks
financial support of the field	a major financial panic
knowledge of the financial structure	world wide financial panic

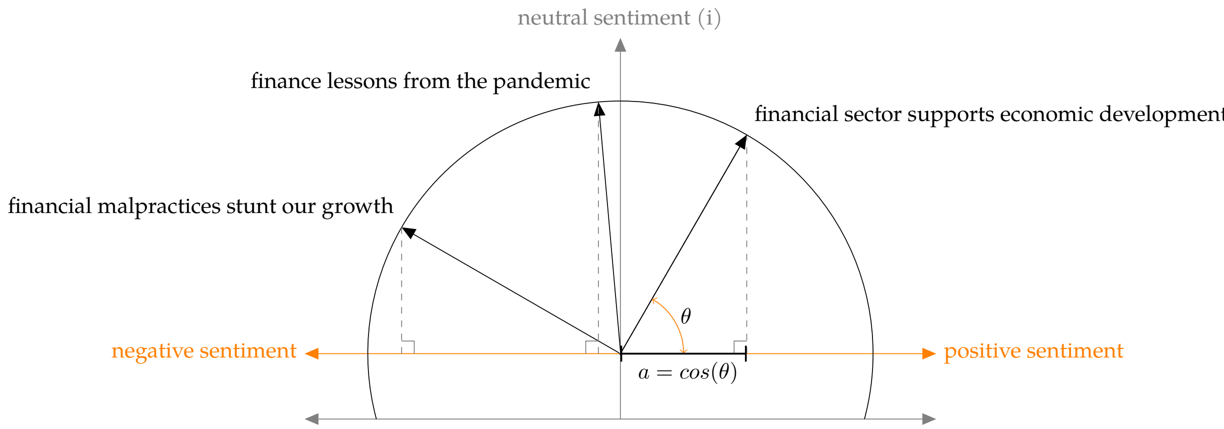
Note: A sentence is assigned positive or negative finance sentiment, based on its projection onto the finance positivity dimension (cosine similarity). Sentences at the top are the most positive or negative in their respective column, and the absolute value of finance sentiment decreases down each list.

Table 3 lists the sentences with the most positive and most negative finance sentiment for American English. We can see that finance is placed in the most positive context when it

Figure 1: Conceptual diagram of finance sentiment measurement



(a) Defining the positive minus negative finance sentiment dimension



(b) Projection of sentences onto positive minus negative sentiment dimension

Note: Panel (a) shows a conceptual diagram of how similar sentences aggregate to a two-dimensional embedding space. We take the vector difference between positive and negative embeddings to define the finance sentiment dimension. Panel (b) illustrates the classification of three example sentences by projecting them onto this dimension. For one of the sentences, we illustrate cosine similarity, defined as the cosine of the angle between two vectors. Sentences that are close in terms of meaning have a smaller angle between them in this vector space, thus higher cosine similarity. Positive finance sentences have a smaller angle to the positive dimension and a larger positive projection on the finance sentiment dimension.

is associated with funding for projects and businesses. At the polar opposite, descriptions of financial distress and panics carry the most negative connotation.[3]

Finally, we calculate an annual finance sentiment for each language i based on the cosine similarity of all finance-mentioning sentences that occurs in that language in each year t, weighted by the number of times the sentence occurred that year,

$$f_{it} = \sum_j a_{ji} \times \frac{c_{jit}}{\sum_k c_{kit}}. \tag{2}$$

[3]The Online Appendix A.2 provides similar lists for all languages.

The frequency weighted f_{it} varies over time only because of changes in sentence occurrence c_{jit}, while the sentiment of particular sentences a_{ji} in each language i stays constant. This is an important distinction from the approach of Kozlowski, Taddy, and Evans (2019), who train a language model for each language in each year, and then measure the orthogonal projections based on these year-specific models. While their approach may be more robust when very large amounts of data are available throughout the sample, our approach is more efficient and avoids issues with measurement error in small samples, which are particularly acute early in the Google Books corpus. Computationally, our approach is considerably cheaper, because training neural networks like the one behind BERT is still fairly expensive.

The cost of this reduction in measurement error and computation cost is that we implicitly assume that the language model is constant over time and that only the frequency of language use varies over time. This is obviously not exactly right. Languages evolve. We lack the data to measure the extent to which such changes in the meaning of language matter for our conclusions. Encouragingly, evidence from US newspapers over a similar period suggests that changes to American English do not affect much the ability to predict with text (Manela and Moreira, 2017).

We calculate finance sentiment for every year from 1870 to 2009 for American English, British English, French, German, Italian, Russian, and Spanish. We have a shorter 95-year sample for Chinese because prior to 1922, its corpus is highly sparse and most years feature no mentions of finance. These languages can all be traced to a major geographical area, centered around a distinct country, throughout most of our sample. For example, the concentration of Russian speakers is highest in Russia. Therefore, in what follows, we refer to the finance sentiment of these languages and countries interchangeably, but note that this requires a modest leap of faith. We expect it to introduce more error into our measure toward the end of our sample, when Spanish books, for example, may be published in Latin American countries whose economic condition is no longer highly correlated with that of Spain.

Another caveat to our finance sentiment index is that it is based on published books, which may not represent the average citizen, especially early in our sample, when large parts of the world were illiterate. As a result, we may miss marginalized and underrepresented groups of the population. Nonetheless, this "literary elite" has historically commanded a disproportionately large share of wealth and power, and exerted considerable influence on the opinions of the rest of society. Its sentiment toward finance can therefore be even more economically important.

3 Finance sentiment over time and across countries

Table 4 describes finance sentiment over our sample. Finance sentiment becomes more positive over time, growing 0.2 percent a year over all languages. Within-language, the highest sentiment improvement is in Italy at 0.4 percent. German and British English show no discernible trend, and no language has a negative trend.

We see that sentiment toward finance in languages spoken in more capitalist countries tends to be above that of communist countries. In our sample, American English, on average, is at the very top followed closely by British English. The next set of languages that follow are Spanish, French, Chinese, and Italian. German and Russian are the two languages with an overall negative connotation for finance. Its volatility is highest for Russian, followed by Spanish and French.

Figure 2 plots our finance sentiment panel and reveals several salient features. American English has the most positive sentiment towards finance after 1912; before that, it was slightly below British English. We see a 5% drop in US finance sentiment in 1874, a year after the Panic of 1873, which triggered economic depression in Europe and North America. A similar decline in finance sentiment is in 1896, after the country's gold reserves had dwindled and saved by JP Morgan's, and the Rothschild's gold loan. We see an increase in sentiment in 1885 and 1887, after labor union strikes, which eventually led to the eight-hour workday. The trend across languages is of an improving finance sentiment across time, with a slight dip at the very end in 2007–2008. A possible reason for that could be the great recession, whose impact could be felt across the globe. [4]

Languages do not seem to cross each other, apart from Chinese, which exhibits significant changes in finance sentiment over time. This volatility is in line with historical events. Chinese finance sentiment plunges in 1971 by 31%, and there is an uptick of 28% in 1972, one year after the United Nations recognized the People's Republic of China as "the only legitimate representative of China", followed by a visit from US President Nixon. We also see an 18% increase in 1976, a year after the constitution of the People's Republic of China was formalized.

The three Romance languages in our sample, French, Italian, and Spanish, have similar attitudes towards finance. Spanish has the most favorable view, followed closely by French. We see higher volatility and an uptick in Spanish finance sentiments at the start of the 1874 Bourbon Restoration, which restored the monarchy. French finance sentiment is more volatile during World War II, with the most significant drop of 3% in 1943 when the French surrender to Germany. The highest surge in French finance sentiment is in

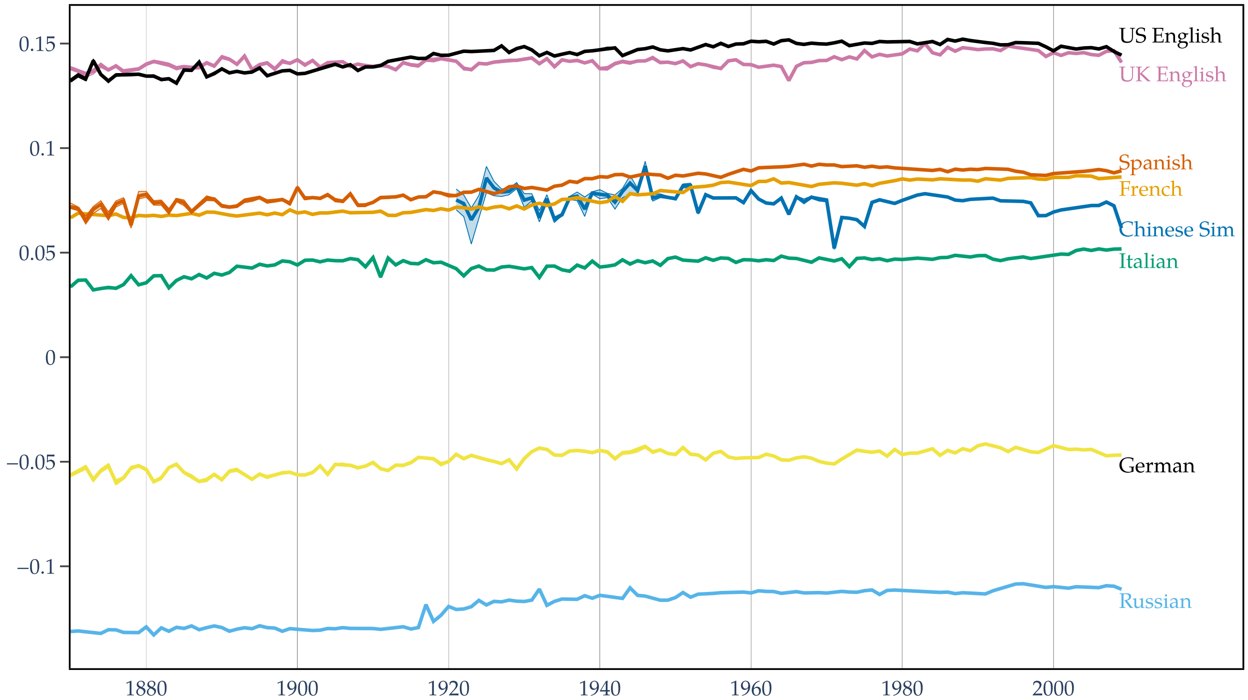
[4] See Wikipedia for historical events mentioned in this section.

Table 4: Finance sentiment and other summary statistics

Country (language)	Variable, %	Mean	Std. Dev.	Obs.
China	Finance sentiment	7.5	0.5	95
(Chinese)	Finance sentiment growth	0.1	7.9	88
	GDP growth	3.2	7.1	119
France	Finance sentiment	7.6	0.7	140
(French)	Finance sentiment growth	0.2	1.3	139
	GDP growth	1.9	6.4	139
	Credit growth	4.9	12.9	101
Germany	Finance sentiment	-4.9	0.5	140
(German)	Finance sentiment growth	0	4.4	139
	GDP growth	2.1	8.1	139
	Credit growth	8.9	17.8	129
Italy	Finance sentiment	4.4	0.4	140
(Italian)	Finance sentiment growth	0.4	5.2	139
	GDP growth	2.0	4.7	139
	Credit growth	6.1	13.9	139
Russia	Finance sentiment	-11.9	0.8	140
(Russian)	Finance sentiment growth	0.1	1.5	139
	GDP growth	2.0	8.5	139
Spain	Finance sentiment	8.3	0.7	140
(Spanish)	Finance sentiment growth	0.2	3.4	139
	GDP growth	2.1	5.0	139
	Credit growth	7.4	11.1	98
UK	Finance sentiment	14.2	0.3	140
(British English)	Finance sentiment growth	0	1.5	139
	GDP growth	1.5	2.9	139
	Credit growth	4.0	8.2	129
US	Finance sentiment	14.5	0.6	140
(American English)	Finance sentiment growth	0.1	1.3	139
	GDP growth	2.1	5.0	139
	Credit growth	4.5	6.7	129
Total	Finance sentiment	4.8	8.7	1075
	Finance sentiment growth	0.2	3.7	1061
	GDP growth	2.1	6.2	1092
	Credit growth	5.9	12.5	725

Note: The sample spans from 1870 to 2009 for 8 country-language pairs. The corpus of sentences for each language is gathered from the Google Book Ngram Corpus. The connotation for each finance-mentioning sentence is measured based on its cosine similarity with respect to the positive minus negative vector. Finance sentiment is the average cosine similarity of finance-mentioning sentences in each language and year. GDP and credit data are from Jordà, Schularick, and Taylor (2017) and Barro and Ursua (2010) when available.

Figure 2: Sentiment toward finance



Note: Finance sentiment is based on the annual average projection of finance-mentioning sentences' embeddings onto the positive minus negative finance sentiment dimension. Sentences are from the Google Books Ngram corpus and embedded using BERT. Bands represent 95 percent confidence intervals produced by subsampling.

1944, the year Paris was liberated. Sentiment dips and recovers for Italian in 1911–1912 at the start of the Italo-Turkish war, then rises in 1933 by 14%, when Fascist membership becomes compulsory for University teachers, prompting more favorable and nationalistic literature.

The two languages in which we find a negative finance sentiment are German and Russian. Similar to Italian, we see finance sentiment becoming more positive as the Nazi party gains power in Germany. Finance sentiment increases by 9% in 1930, the year the Nazi party gained its first minister. For Russia, we see a permanent increase in finance sentiment in 1917, coinciding with the Russian revolution. We also see a permanent increase at the beginning of 1990s after the collapse of the USSR, as Russian-speaking countries adopt a more capitalist system. The largest drop for Russian finance sentiment is in 1933, a year after the Soviet famine of 1932–1933.

3.1 Capitalism and finance sentiment

How does finance sentiment compare with attitudes toward capitalism? In Table 5 we regress finance sentiment on three measures of attitudes toward capitalism that were previously used by Di Tella and Macculloch (2009). We gradually consider more saturated fixed effect specifications.

We find that comparing across countries, higher finance sentiment is associated with more positive attitudes toward capitalism according to all three measures of capitalism. Countries with more right-leaning governments are more likely to exhibit positive sentiment toward finance. So do countries whose citizens say they prefer more private than public ownership of business. By contrast, countries where registering a business is difficult, exhibit lower finance sentiment.

While these results reassure us that our measure of finance sentiment works as intended, the low R-squares in the specifications without fixed effects at the same time imply that finance sentiment is quite different and far from being spanned by capitalism attitudes. Given the persistence of finance sentiment, the inclusion of country fixed effects captures essentially all its variation as evident from the R-squares being close to 1.

Table 5: Finance sentiment and attitudes toward capitalism

	Finance sentiment score											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ideological leaning of government	0.024** (0.010)	0.0005** (0.0002)	0.025** (0.011)	0.0003 (0.0003)								
Preference for private ownership of business					0.032** (0.012)	-0.0008 (0.001)	0.032** (0.013)	-0.0017* (0.0007)				
Difficulty of registering a business									-0.015*** (0.004)	0.00004 (0.00011)	-0.015*** (0.004)	-0.0002 (0.0002)
Country FE		Yes		Yes		Yes		Yes		Yes		Yes
Year FE			Yes	Yes			Yes	Yes			Yes	Yes
Observations	263	263	263	263	18	16	18	16	35	35	35	35
R^2	0.021	0.999	0.047	0.999	0.224	1.000	0.274	1.000	0.221	1.000	0.229	1.000

Note: Ideological leaning is based on the ideology of the largest party in government, according to the classification scheme of the Database of Political Institutions, made available via the World Bank. Scale - Left: 0, Center: 0.5, Right:1. Preference for ownership of businesses is based on World Values Surveys. Scale - State: 1, Private: 10. Difficulty of registering a business measures the total number of procedures required for a startup to obtain legal status, obtained from World Development Indicators published by the World Bank. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. . Standard errors are in parentheses.

3.2 Finance sentiment growth

Motivated by the positive trend and the persistent ordering across countries documented above, we also report finance sentiment growth $\Delta f_{i,t}$, which characterizes the relative change of finance sentiment towards either the positive or negative direction, given the absolute value of the previous year's sentiment for country i and year t:

$$\Delta f_{it} = \frac{f_{it} - f_{it-1}}{|f_{it-1}|} \times 100. \tag{3}$$

Note that while theoretically this measure could be ill-behaved when sentiment is close to zero, as is evident from Figure 2, finance sentiment is far from and does not cross zero for any of the languages we study.

As Table 4 shows, China exhibits the greatest volatility (7.9%), followed by Italy (5.2%) and Germany (4.4%). We can also see that sharp changes in finance sentiment growth tend to partially reverse within a year. We formally investigate this pattern in Section 5.3.

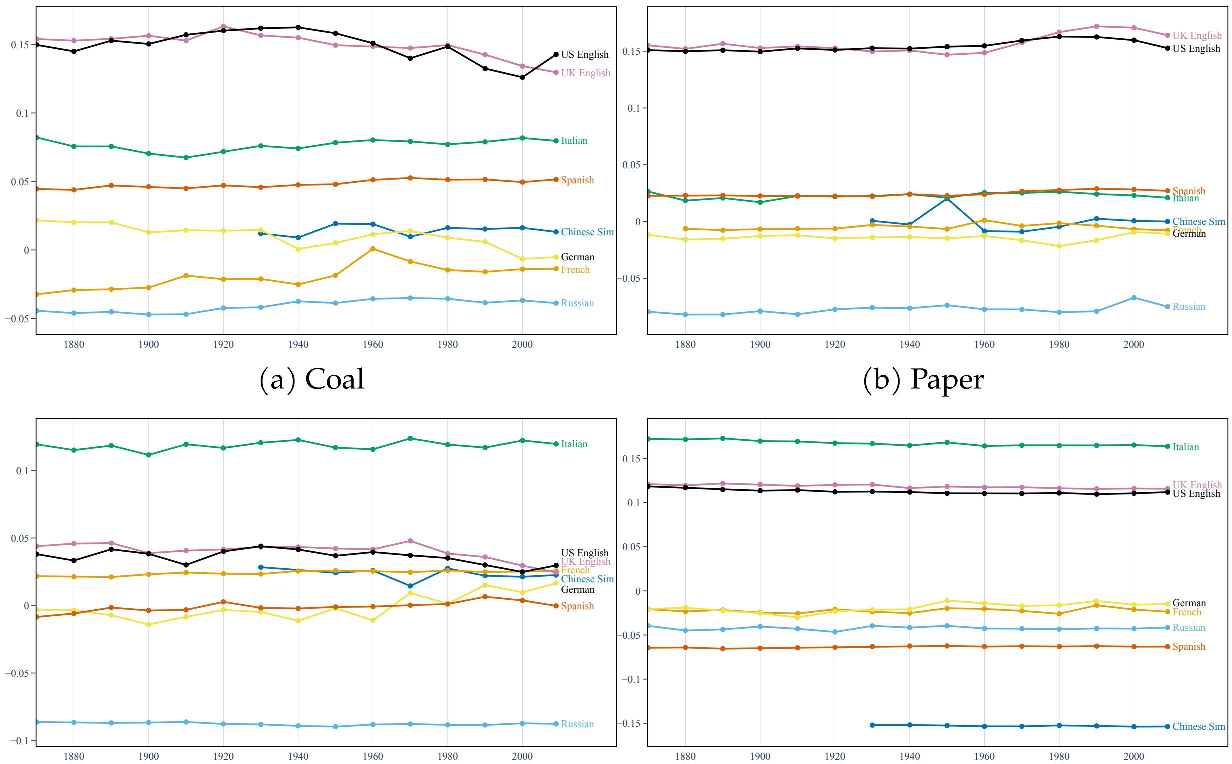
4 Alternative text-based approaches

We next consider several alternative text-based approaches to sentiment measurement.

4.1 Sentiment towards other industries

One potential concern may be that it is not just finance-specific sentiment that is changing over time and across countries, but rather sentiment more generally. To explore this possibility, we construct similar text-based measures of sentiment for three other industries that existed throughout our sample period: Coal, Paper, and Tobacco. Because this calculation is computationally expensive, we produce these sentiment indices only once per decade. This lower frequency analysis suffices for our purposes because we are interested mostly in comparing the ordering and variation across languages to those of finance sentiment.

Figure 3: General sentiment



(a) Coal (b) Paper

(c) Tobacco (d) January

Note: We report sentiment toward the Coal, Paper, and Tobacco industries as well as toward the fairly generic word "January", by redefining the positivity dimension accordingly and focusing on sentences mentioning these alternatives instead of "finance". For example, sentiment for the coal industry is based on the annual average projection of coal-mentioning sentence embeddings onto the positive minus negative coal sentiment dimension. To define the positivity dimension, we average the difference in embedding for the following tuples (and their translations to each language): [("coal is good for society", "coal is bad for society"),("coal miners are mostly good people", "coal miners are mostly corrupt people "),("coal positively impacts our world", "coal negatively impacts our world"),("coal industry helps the economy", "coal industry hurts the economy"),("coal industry benefits society", "coal industry damages society")]. Between similar ngrams such as "coal miners", "coal producers", we select the ngram with a higher Google search volume.

Figure 3 shows that sentiment more generally looks quite different from the one toward finance. Sentiment toward other industries shows no clear upward trend, rises and falls at different times, and the ordering across languages is different. The cross-language variation is also more modest, suggesting these alternative industries carry a more neutral sentiment. We do see, however, that Russian language books tend to be the most negative about all the industries we study, albeit not quite as negative as its finance sentiment.

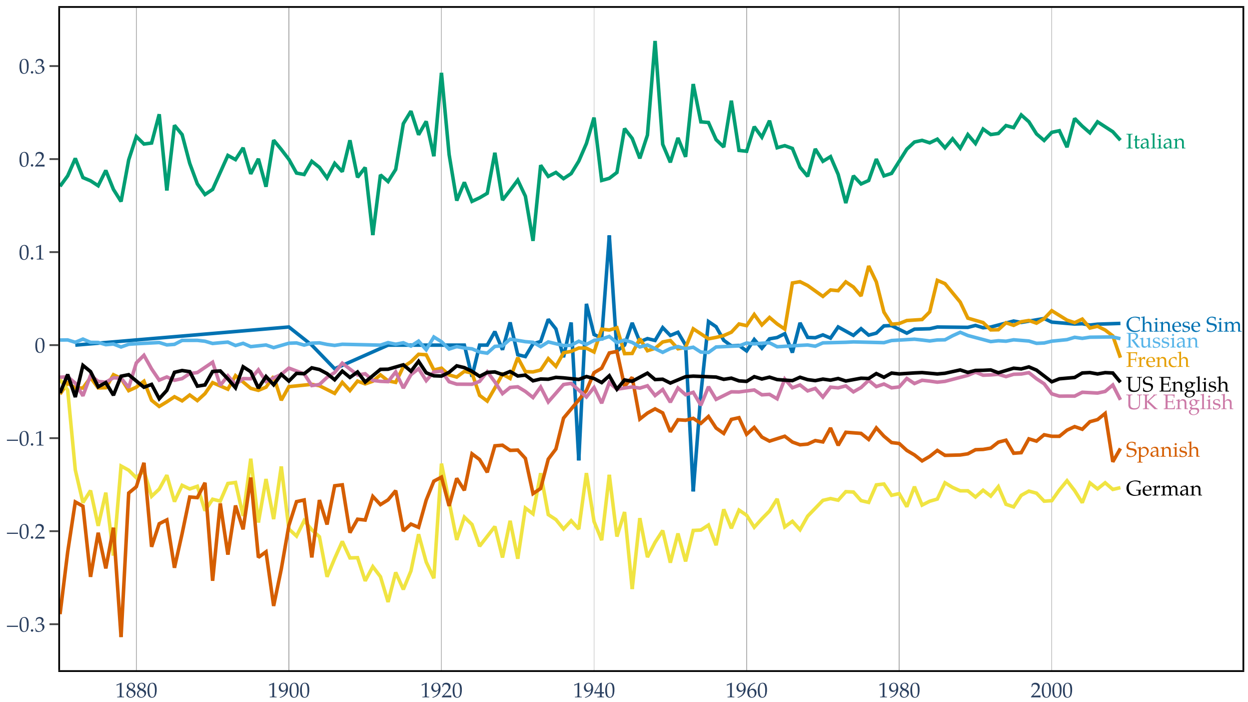
In the last panel of Figure 3 we report another measure of general sentiment as the sentiment associated with the fairly generic word "January" across time, following Gentzkow, Glaeser, and Goldin (2004), who use it to deflate for changes in newspaper reporting volume over time. It is not a perfectly neutral word because its cultural association is quite different across languages and religions. But within language variation over time could still inform us about trends in sentiment more generally. We can see that January sentiment is quite flat across years and does not trend in any language. Its ordering across languages is quite different from that of finance sentiment. January is placed in the most positive context by Italian books and in the most negative context by Chinese ones. Overall, we find little evidence that variation in sentiment more generally is responsible for the variation we see in finance sentiment.

4.2 Dictionary-based approach

A popular and considerably simpler method than ours counts positive versus negative words to measure sentiment (Zhou, 2018). One limitation of this approach is that it often misses the context and subtleties of language, which humans would quickly discern from reading words in sequence. In fact, a major engineering feat of BERT is that its underlying neural network pays attention to longer sequences of words (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017). However, the dictionary-based approach may be a reasonable alternative due to its simplicity.

We use a list of positive and negative words for each language. For English we use the Loughran and McDonald (2011) dictionary. For all other languages, we rely on Chen and Skiena (2014). The sentiment for each sentence is the number of net positive words in a sentence normalized by the total number of positive and negative words in the sentence. We aggregate the sentence sentiments, weighted by their frequency in a year, to get sentiment for that year. Based on the dictionary approach, we get a more volatile score, illustrated in Figure 4. The dictionary approach yields a substantially different ordering

Figure 4: Sentiment toward finance using an alternative dictionary-based approach



Note: Dictionary-based finance sentiment is based on the annual average sentiment of finance-mentioning sentences. Sentiment for a sentence is net positive words in a sentence normalized by total positive and negative words in the sentence. Sentences are from the Google Books Ngram corpus and the positive and negative words are from Loughran and McDonald (2011) and Chen and Skiena (2014).

across languages, which on the face of it seems less likely to represent finance sentiment differences and more likely to be the result of noise and misinterpretation of context.

4.3 Alternative language embedding-based approaches

As mentioned, our language embedding approach builds on Kozlowski, Taddy, and Evans (2019), but differs in an important way. They fit a word embedding model (e.g. word2vec, glove) from scratch to each decade of sentences, and then measure the cosine similarity once for each phrase of interest. The variation in their measures of culture come from variation in term frequencies but also from variation in these fitted language models, whose parameters are measured with error. By contrast, we use a pretrained language model (BERT), measure cosine similarity once for each phrase of interest, and then average these cosine similarities for each year (and language). Variation in our measure is due only to term frequencies, because language model error is held fixed over time.

Table 6 reports the mean and volatility for each of these alternative measures of finance sentiment by language. From the means, we can see that the ordering across languages differs across measures. But the most prominent difference is that the volatility (in parentheses) is on average much higher if we follow the Kozlowski, Taddy, and Evans approach and train a new language model each year. Specifically, the last line, shows that the average within-language volatility is 0.01 using our approach, 0.04 using the dictionary-based approach, and at least 0.07 using the Kozlowski, Taddy, and Evans approach. With arbitrarily large text data this sampling uncertainty may vanish. But it appears that even with the Google Books corpus, whose 5-gram annual counts occupies terabytes of disk space, there is not enough data to reliably train these language embedding models from scratch. Especially in the earlier years of the corpus, when fewer books are available, our BERT-based transfer learning approach appears to be more efficient.

5 Finance sentiment and the macroeconomy

We next study how the finance sentiment relates to financial crises and to macroeconomic growth. We describe the macroeconomic data and model specification before turning to the empirical results.

5.1 Macroeconomic data

The economic and credit data that we use are from the macrohistory dataset compiled by Jordà, Schularick, and Taylor (2017). The macrohistory dataset covers annual data for 17 advanced countries from 1870 to 2016. To merge consistently with our text-based finance sentiment index, we only utilize 6 of them: France, Germany, Italy, Spain, UK, and US, spanning from 1870 to 2009. Together, these 6 countries make up more than 40% of the world economy throughout our sample period. This dataset lacks the GDP and population

Table 6: Alternative approaches

Languages	Dictionary	fastText	GloVe	Word2Vec	BERT
Chinese	-0.06	0.02	0.03	-0.03	0.03
	(0.10)	(0.26)	(0.14)	(0.09)	(0.06)
French	-0.01	-0.25	-0.14	-0.02	0.08
	(0.04)	(0.04)	(0.09)	(0.05)	(0.01)
German	-0.18	-0.10	-0.02	-0.02	-0.05
	(0.04)	(0.07)	(0.09)	(0.07)	(0.00)
Italian	0.20	-0.09	-0.10	0.10	0.04
	(0.03)	(0.06)	(0.15)	(0.07)	(0.00)
Russian	0.00	-0.05	0.04	0.03	-0.12
	(0.00)	(0.10)	(0.11)	(0.07)	(0.01)
Spanish	-0.13	-0.16	0.05	0.06	0.08
	(0.06)	(0.06)	(0.11)	(0.09)	(0.01)
UK English	-0.04	0.16	0.03	0.08	0.14
	(0.01)	(0.09)	(0.12)	(0.05)	(0.00)
US English	-0.03	0.17	0.01	0.14	0.14
	(0.01)	(0.06)	(0.07)	(0.08)	(0.01)
Total	-0.02	-0.03	-0.02	0.04	0.05
	(0.11)	(0.18)	(0.13)	(0.10)	(0.09)
Average	-0.03	-0.04	-0.01	0.04	0.04
	(0.04)	(0.09)	(0.11)	(0.07)	(0.01)

Note: We report the average sentiment and corresponding volatility for each language using different approaches. Total is the pooled mean and standard deviation including all languages, while average is the average across languages of the within-language statistics. Standard deviations are in parentheses.

of China and Russia, which we supplement from the Barro-Ursua Macroeconomic Data (Barro and Ursua, 2010). We incorporate credit growth as one of the key control variables in our model because credit plays an important role in the macroeconomy and in financial development. Following Schularick and Taylor (2012), we use total loans to non-financial private sector as credit proxy.

Table 4 summarizes these macroeconomic variables. It shows that GDP growth is highest for China at 3.2 percent a year, while other economies hover around two percent. Credit growth is six percent on average, with Germany and Spain exhibiting the highest average credit growth.

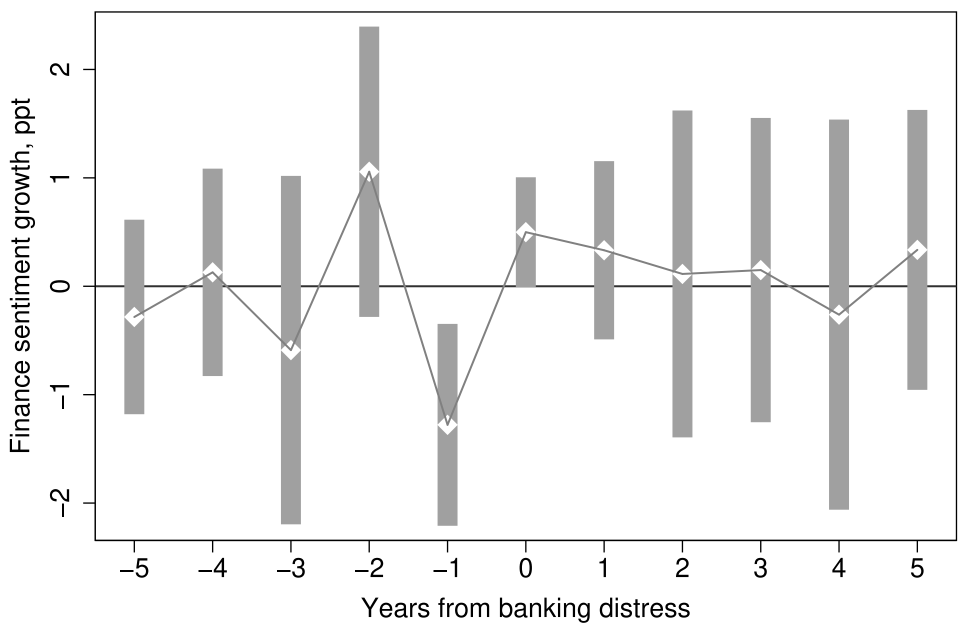
5.2 Financial crises

We next examine the relationship between finance sentiment growth and historical banking crises. One may expect finance sentiment to decline following a financial crisis. But because the financial system requires a large degree of trust, a decline in finance sentiment can itself transform a mild recession into a full-blown financial crisis (Sapienza and Zingales, 2012). To do so, we leverage the data set on banking distress and panics gathered by Baron, Verner, and Xiong (2021), who have shown that bank equity declines precede contractions in credit and output, even in the absence of panics.

Figure 5 plots the mean finance sentiment growth around such bank equity declines. To control for unobserved heterogeneity, we absorb country and year fixed effects. We find that finance sentiment declines by about one percentage point in the year preceding bank equity declines, but stays relatively flat in the five years following these events. Interestingly, these declines in finance sentiment follow a finance sentiment boom two years before the distress period. This result is consistent with the finding of Schularick and Taylor (2012) that financial crises tend to follow credit booms, about two years later.

In Table 7 we investigate whether finance sentiment holds incremental predictive power

Figure 5: Finance sentiment around financial crises



Note: The figure depicts the impact of the crisis of bank equity 30% decline on the finance sentiment growth. It contains a 10-year window, spanning from 5 years before crisis and 5 years after. The gray histograms represent 90% confidence intervals, which are adjusted for country-level clustering. The regression is given as follows: $\Delta f_{it} = c + \beta_1 crisis_{it}^{-5} + \cdots + \beta_{10} crisis_{it}^{+5} + country_i + yr_t + \epsilon_{i,t}$.

Table 7: Finance sentiment declines before bank equity declines

Model	Bank Equity 30% Decline$_{t+1}$ Logistic Regression			Linear Probability Model	
	(1)	(2)	(3)	(4)	(5)
Finance sentiment growth$_t$	-7.70***	-6.08	-9.59**	-0.69**	-0.66**
	(2.97)	(3.72)	(4.06)	(0.22)	(0.23)
Finance sentiment growth$_{t-1}$		5.92	5.56	0.25	0.15
		(5.20)	(6.42)	(0.34)	(0.31)
Credit growth$_t$			-1.12	-0.08	-0.12
			(0.86)	(0.05)	(0.07)
Credit growth$_{t-1}$			3.17***	0.26**	0.21
			(1.03)	(0.07)	(0.12)
Country FE	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	Yes
Obs	1053	1045	709	709	709

Note: This table reports the impact of finance sentiment growth and/or credit growth on predicting banking crises using logistic regression with country fixed effects and linear probability model. The banking crises defined in Baron, Verner, and Xiong (2021) represent that banking equity declines 30% for each country in a year and credit growth is denoted as loan growth as defined in Schularick and Taylor (2012). Model specifications with credit growth exclude China and Russia as credit data of these two countries are unavailable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses for logistic regression model and linear probability model uses clustered standard errors by country.

for financial distress, over and above credit growth. The first three columns report logistic regressions of bank equity declines on lags of finance sentiment growth and credit growth. All three show that finance sentiment declines indicate an increase in the probability of banking distress in the following year. The second lag of finance sentiment growth is statistically insignificant, though positive as in Figure 5. Consistent with the results of Schularick and Taylor (2012), credit growth is higher two years prior to banking distress. But even when we control for credit growth, the first lag of finance sentiment remains a significant predictor of bank equity declines.

Specification (4) shows that this result also manifests in a linear probability model. Because banking panics often involve more than one country, in the last specification we include year fixed effects. We find that finance sentiment growth can also predict idiosyncratic banking distress episodes.

Given the compelling evidence of Baron, Verner, and Xiong (2021) that panics are the

Table 8: Alternative definitions of banking crises

Logistic Regression	Panic_{t+1}	$\text{Bank Failures}_{t+1}$	$\text{Bank Equity Crisis}_{t+1}$	BVX Crisis_{t+1}	JST Crisis_{t+1}
	(1)	(2)	(3)	(4)	(5)
$\text{Finance sentiment growth}_{t}$	-9.36***	-10.22*	-12.57	-8.84**	2.75
	(2.83)	(5.92)	(8.87)	(3.73)	(8.19)
$\text{Finance sentiment growth}_{t-1}$	12.17	14.06	13.60	11.80	12.04
	(10.98)	(11.93)	(13.15)	(10.96)	(8.15)
Credit growth_{t}	-1.48	-2.56*	-2.34	-1.57	-1.37
	(1.86)	(1.51)	(2.23)	(2.13)	(1.76)
$\text{Credit growth}_{t-1}$	4.49***	5.59***	5.51***	4.53***	4.35**
	(1.12)	(1.79)	(1.68)	(1.62)	(1.90)
Country FE	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	No
Obs	709	709	709	709	709

Note: This table shows the robustness result of how finance sentiment growth affects other banking crises using logistic regression with country fixed effects. All other banking crises are defined in Baron, Verner, and Xiong (2021). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses.

result of earlier bank equity declines, our evidence suggests that declines in finance sentiment declines are the cause, rather than the consequence, of financial crises. While timing evidence cannot rule out reverse causality, we find it unlikely that finance sentiment declines in anticipation of a subsequent financial crisis.

As Table 8 shows, this result is fairly robust to varying the definition of banking distress. Finance sentiment declines one year before banking panics, bank failures, bank equity crises, or the union of these indicators, as defined by Baron, Verner, and Xiong (2021). The one exception is that it does not forecast Jordà, Schularick, and Taylor (2017) crises, which are based on narrative accounts of bank runs, bank failures, and government interventions. This could be because such narrative accounts miss important but smaller episodes or because the episodes they classify as crises are fundamentally different. As Krishnamurthy and Muir (2017) explain, these chronologies sometimes date the same episodes in different years.

Finally, we note that if the shift in public sentiment toward finance appears in the Google Books corpus with a lag, this would strengthen our result that finance sentiment changes lead banking distress periods. The evidence on publication lags of severe natural disasters we report in Jha, Liu, and Manela (2021) suggests modest lags of less than a year before such disasters are first mentioned in the corpus. This may be surprising if we consider the time it now takes to publish a bestseller, but the Google Books corpus includes many serial publications like government reports that are often more timely.

5.3 Economic growth

We next analyze how shocks to finance sentiment affect macroeconomic activity through the lens of local projection-estimated impulse responses, which separate these shocks from other macroeconomic shocks to output and credit. Specifically, we estimate cumulative impulse response functions via local projections (Jordà, 2005):

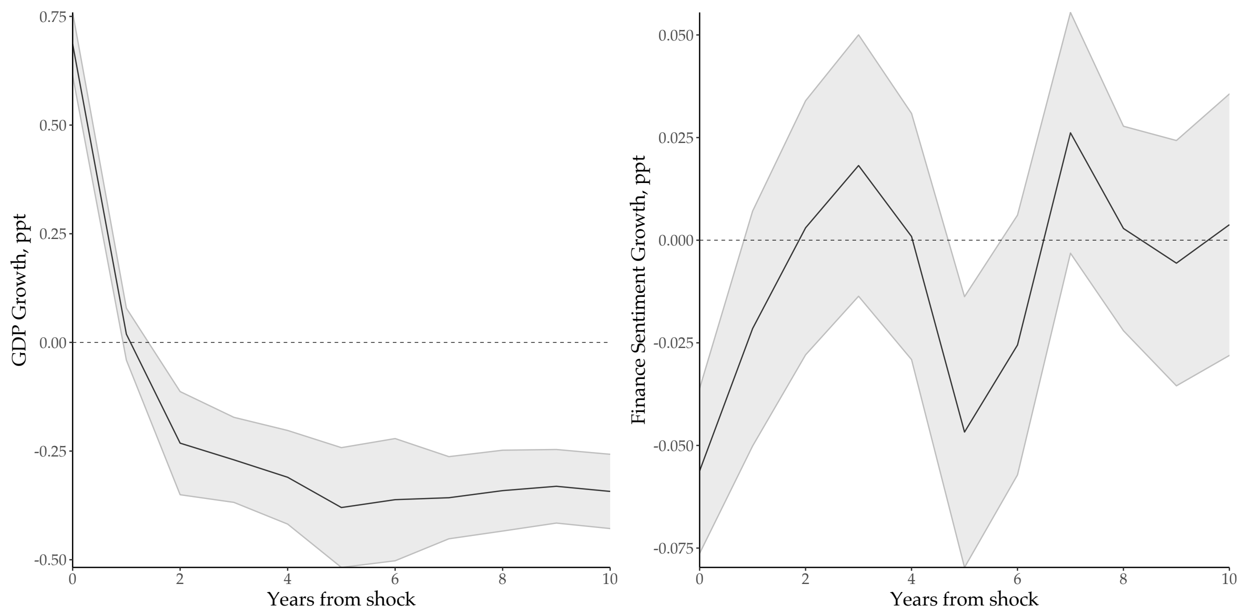
$$\Delta_h y_{i,t+h} = \alpha_i^h + \sum_{k=1}^{3} \beta_k^h \Delta f_{i,t-k} + \sum_{k=0}^{3} \gamma_k^h X_{i,t-k} + \epsilon_{i,t+h}, \quad h = 0, ..., H, \tag{4}$$

where i represents the country and t represents the year. $\Delta_h y_{i,t+h} = y_{i,t+h} - y_{i,t-1}$ indicate the h-year cumulative growth of interest, e.g. GDP growth rate and credit growth. α_i^h are country fixed effects. $\Delta f_{i,t}$ is finance sentiment growth, $X_{i,t}$ is a vector of control variables, and $\epsilon_{i,t+h}$ are disturbance terms.

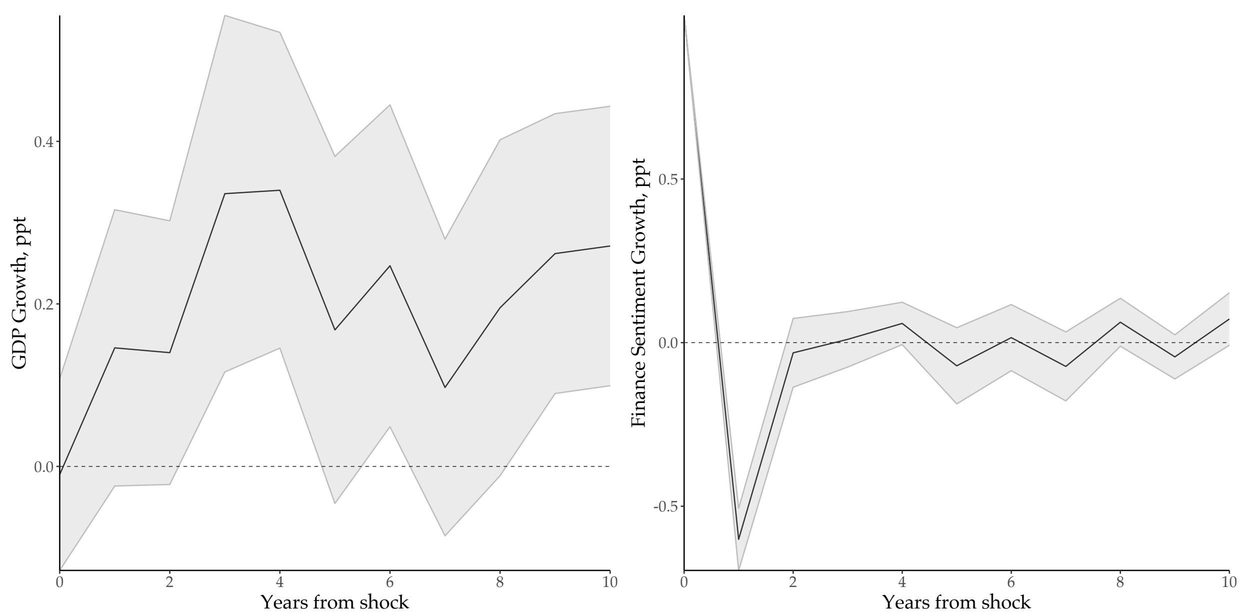
This model estimates the response of $\Delta_h y_{i,t+h}$ from a shock to $\Delta f_{i,t}$. To capture the direct link of such shocks to economic growth, we control for the first 3 lags of credit growth and finance sentiment growth. Similarly, the first 3 lags of economic growth and finance sentiment growth are control variables when credit growth is our target variable.

The results in Figure 6 include all countries in the sample and therefore focus on GDP and finance sentiment alone because no credit data is available for China or Russia. The bottom left panel shows that a one percentage point increase in finance sentiment growth increases GDP growth by about 0.3 percentage points four years out, though it has no contemporaneous effect. This effect is quite large compared with the mean annual GDP growth of 2.1 percent.

Figure 6: Impulse response of GDP growth to a finance sentiment growth shock



(a) GDP growth shock



(b) Finance sentiment growth shock

Note: Impulse responses estimated via local projections indicate the change of the cumulative response to a one percentage point shock. Bands are 90% confidence intervals based on Driscoll and Kraay nonparametric robust standard errors.

The top right panel shows increases in GDP tend to coincide with declines in finance sentiment growth. While this latter effect is statistically different from zero, its economic magnitude is quite modest.

Interestingly, the bottom right panel reveals that finance sentiment growth tends to oscillate after shocks. This is somewhat surprising, as we expected it to gradually mean-revert like GDP growth does on the top left panel. These oscillations could be the result of book writers and publishers attempting to continuously innovate with contrarian books.

It is likely, however, that finance sentiment affects economic growth not directly, but indirectly, by changing the demand for financial services. It may also affect the supply for financial services by changing how the sector is regulated. Both mechanisms should manifest as changes in the quantity of credit. To investigate this channel, we focus next on the subsample of advanced economies for which we have credit data.

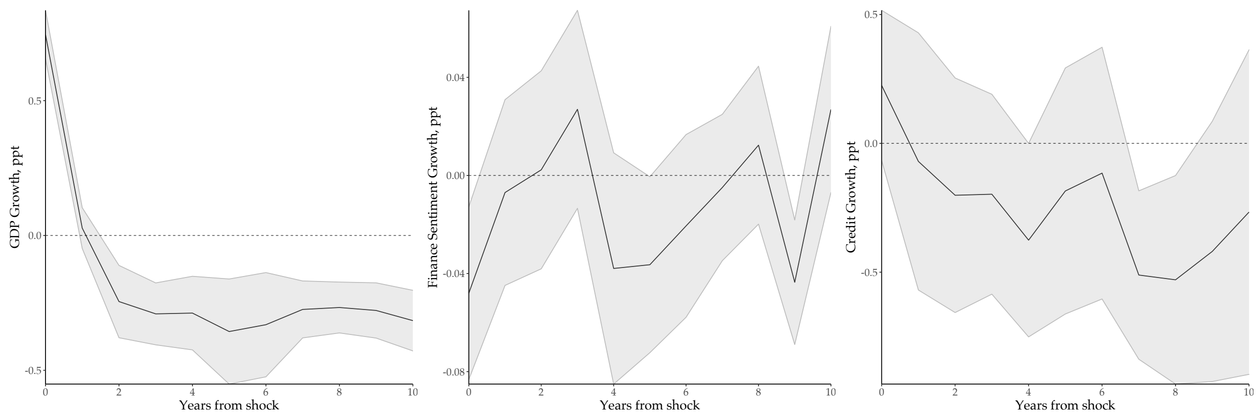
Figure 7 depicts the response of economic growth, finance sentiment growth, and credit growth to shocks by the same three variables. Regardless of the oscillating impact of finance sentiment growth, the cumulative response to the shock in finance sentiment growth is positive for both economic growth and credit growth after a year. A one percentage point increase in finance sentiment growth is associated with a 0.4 percentage point increase in credit growth. The addition of credit growth also reduces somewhat the impulse response of GDP to finance sentiment. It seems, therefore, that some but not all of the effect of finance sentiment on GDP is through credit growth.

6 Conclusion

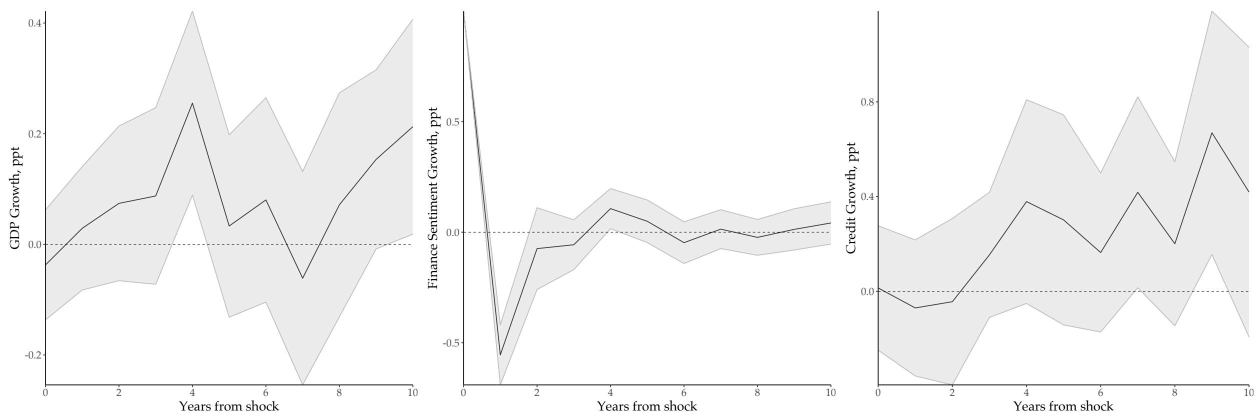
We measure popular sentiment toward finance using a computational linguistics approach applied to millions of books published in eight countries over hundreds of years, and document several new facts.

Finance sentiment differences across countries mostly persist throughout our long sam-

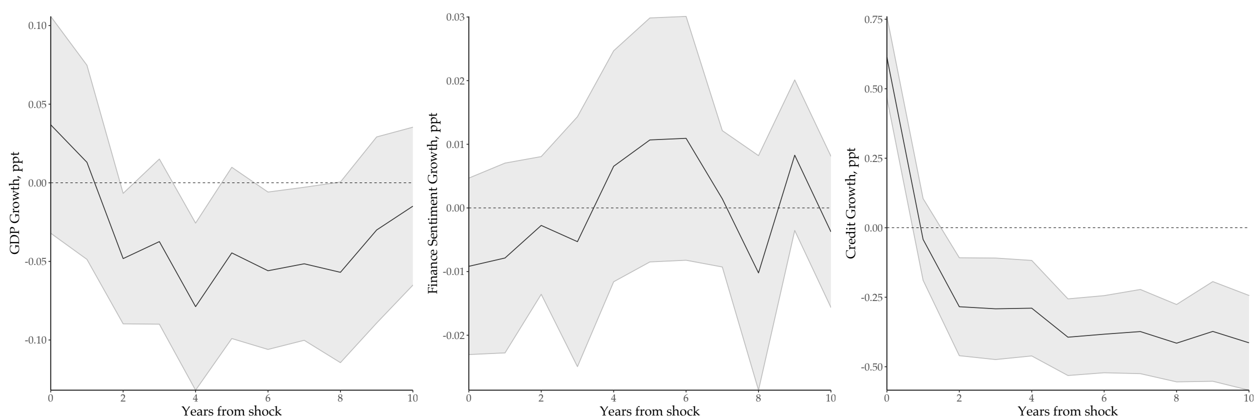
Figure 7: Impulse response of GDP growth and credit growth to finance sentiment growth shocks (without China and Russia)



(a) GDP growth shock



(b) Finance sentiment growth shock



(c) Credit growth shock

Note: Impulse responses estimated via local projections indicate the change of the cumulative response to a one percentage point shock. Bands are 90% confidence intervals based on Driscoll and Kraay nonparametric robust standard errors.

ple, with the exception of China, which exhibits greater volatility and a level of finance sentiment about as positive as that of Italy and France. Generally, books written in languages of more capitalist countries tend to discuss finance in a more positive context. Finance sentiment declines one year before periods of banking distress. Using local projections, we find that shocks to finance sentiment positively affect long term economic and credit growth.

In follow-up work, we investigate the origins of finance sentiment fluctuations, and in particular, how finance sentiment responds to natural disasters (Jha, Liu, and Manela, 2021). The long panel of finance sentiment that we provide could also be used by future work to understand other consequences of finance sentiment variation. For example, combining our methods and data with the fraud-based identification approach of Giannetti and Wang (2016) and Gurun, Stoffman, and Yonker (2018) would be interesting. Further analysis of sentiment toward other industries, like the ones we explore in Figure 3, could be another promising avenue for future research. More generally, the language embedding-based approach we develop could prove useful for other text-based measures that are of interest to economists.

References

Altszyler, Edgar, Mariano Sigman, Sidarta Ribeiro, and Diego Fernández Slezak, 2017, Comparative study of LSA vs Word2vec embeddings in small corpora: A case study in dreams database, *Consciousness and Cognition* 56, 178–187.

Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? the information content of Internet stock message boards, *Journal of Finance* 59, 1259–1293.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, 2016, Measuring economic policy uncertainty, *Quarterly Journal of Economics* 131, 1593–1636.

Baron, Matthew, Emil Verner, and Wei Xiong, 2021, Banking Crises Without Panics*, *The Quarterly Journal of Economics* 136, 51–113.

Barro, Robert J, and Jose F Ursua, 2010, Barro-ursua macroeconomic data, .

Bodnaruk, Andriy, Tim Loughran, and Bill McDonald, 2015, Using 10-K Text to Gauge Financial Constraints, *Journal of Financial and Quantitative Analysis* 50, 623–646.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov, 2016, Enriching Word Vectors with Subword Information, .

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, and Matthew Richardson, 2018, Information, Trading, and Volatility: Evidence from Firm-Specific News, *Review of Financial Studies*.

Bybee, Leland, Bryan T. Kelly, Asaf Manela, and Dacheng Xiu, 2019, Business News and Business Cycles, Working Paper.

Chava, Sudheer, Wendi Du, and Nikhil Paradkar, 2019, Buzzwords?, SSRN Scholarly Paper ID 3862645 Social Science Research Network Rochester, NY.

Chen, Jiafeng, and Suproteem K Sarkar, 2020, A Semantic Approach to Financial Fundamentals, p. 5.

Chen, Yanqing, and Steven Skiena, 2014, Building Sentiment Lexicons for All Major Languages, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* pp. 383–389 Baltimore, Maryland. Association for Computational Linguistics.

Cieslak, Anna, and Annette Vissing-Jorgensen, forthcoming, The Economics of the Fed Put, *Review of Financial Studies*.

D'Acunto, Francesco, Marcel Prokopczuk, and Michael Weber, 2019, Historical Antisemitism, Ethnic Specialization, and Financial Development, *Review of Economic Studies* 86, 1170–1206.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv:1810.04805 [cs]*.

Di Tella, Rafael, and Robert Macculloch, 2009, Why Doesn't Capitalism Flow to Poor Countries?, *Brookings Papers on Economic Activity* 2009, 285–321.

Engelberg, Joseph, Matthew Henriksson, Asaf Manela, and Jared Williams, 2019, The Partisanship of Financial Regulators, Working Paper.

Firth, JR, 1957, Applications of General Linguistics - Firth - 1957 - Transactions of the Philological Society - Wiley Online Library, .

García, Diego, 2013, Sentiment during recessions, *Journal of Finance* 68, 1267–1300.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, forthcoming, Trust and Insurance Contracts, *Review of Financial Studies*.

Gentzkow, Matthew, Edward L Glaeser, and Claudia Goldin, 2004, The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered, Working Paper 10791 National Bureau of Economic Research.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as Data, *Journal of Economic Literature* 57, 535–574.

Gentzkow, Matthew, and Jesse M. Shapiro, 2010, What Drives Media Slant? evidence From U.S. Daily Newspapers, *Econometrica* 78, 35–71.

Giannetti, Mariassunta, and Tracy Yue Wang, 2016, Corporate Scandals and Household Stock Market Participation, *Journal of Finance* 71, 2591–2636.

Goetzman, W., D. Kim, and R. J. Shiller, 2017, Affect, Media and Earthquakes: Determinants of Crash Beliefs from Investor Surveys, Discussion paper Yale University Working Paper.

Goldman, Eitan, Nandini Gupta, and Ryan D. Israelsen, 2020, Political Polarization in Financial News, SSRN Scholarly Paper ID 3537841 Social Science Research Network Rochester, NY.

Grennan, Jillian, 2019, A Corporate Culture Channel: How Increased Shareholder Governance Reduces Firm Value, Working Paper ID 2345384 Rochester, NY.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2006, Does Culture Affect Economic Outcomes?, *Journal of Economic Perspectives* 20, 23–48.

——— , 2008, Trusting the stock market, *Journal of Finance* 63, 2557–2600.

Gurun, Umit G., Noah Stoffman, and Scott E. Yonker, 2018, Trust Busting: The Effect of Fraud on Investor Behavior, *Review of Financial Studies* 31, 1341–1376.

Hanley, Kathleen Weiss, and Gerard Hoberg, 2019, Dynamic Interpretation of Emerging Risks in the Financial Sector, *Review of Financial Studies*.

Hansen, Stephen, Michael McMahon, and Andrea Prat, 2018, Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach, *Quarterly Journal of Economics* 133, 801–870.

Harris, Zellig S., 1954, Distributional Structure, *WORD* 10, 146–162.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun, 2017, Firm-Level Political Risk: Measurement and Effects, SSRN Scholarly Paper ID 2838644 Social Science Research Network Rochester, NY.

Hills, Thomas T., Eugenio Proto, Daniel Sgroi, and Chanuki Illushka Seresinhe, 2019, Historical analysis of national subjective wellbeing using millions of digitized books, *Nature Human Behaviour* 3, 1271–1275.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-Based Network Industries and Endogenous Product Differentiation, *Journal of Political Economy* 124, 1423–1465.

Jha, Manish, Hongyi Liu, and Asaf Manela, 2021, Natural disaster effects on popular sentiment toward finance, Working Paper.

Jordà, Òscar, 2005, Estimation and Inference of Impulse Responses by Local Projections, *American Economic Review* 95, 161–182.

———, Moritz Schularick, and Alan M. Taylor, 2017, Macrofinancial History and the New Business Cycle Facts, *NBER Macroeconomics Annual* 31, 213–263.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov, 2016, Bag of Tricks for Efficient Text Classification, *arXiv e-prints* 1607, arXiv:1607.01759.

Ke, Zheng Tracy, Bryan T. Kelly, and Dacheng Xiu, 2019, Predicting Returns with Text Data, Working Paper.

Kozlowski, Austin C., Matt Taddy, and James A. Evans, 2019, The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings:, *American Sociological Review*.

Krishnamurthy, Arvind, and Tyler Muir, 2017, How Credit Cycles across a Financial Crisis, Discussion Paper w23850 National Bureau of Economic Research Cambridge, MA.

Levine, Ross, Chen Lin, and Wensi Xie, 2020, The African Slave Trade and Modern Household Finance, *Economic Journal* 130, 1817–1841.

Lin, Yuri, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, and Slav Petrov, 2012, Syntactic Annotations for the Google Books NGram Corpus, in *Proceedings of the ACL 2012 System Demonstrations* pp. 169–174 Jeju Island, Korea. Association for Computational Linguistics.

Loughran, T., and B. McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

Loughran, Tim, and Bill McDonald, 2020, Textual Analysis in Finance, SSRN Scholarly Paper ID 3470272 Social Science Research Network Rochester, NY.

Luo, Mancy, Alberto Manconi, and Massimo Massa, 2020, Blinded by Perception? the Stock Market's Reaction to Politically Aligned Media, SSRN Scholarly Paper ID 2879939 Social Science Research Network Rochester, NY.

Manela, Asaf, and Alan Moreira, 2017, News implied volatility and disaster concerns, *Journal of Financial Economics* 123, 137–162.

McCloskey, Deirdre Nansen, 2016, *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World* (University of Chicago Press).

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, et al., 2011, Quantitative analysis of culture using millions of digitized books, *Science* 331, 176–182.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, 2013, Efficient Estimation of Word Representations in Vector Space, *arXiv:1301.3781 [cs]*.

Mikolov, Tomas, Kai Chen, Gregory S. Corrado, and Jeffrey A. Dean, 2015, Computing numeric representations of words in a high-dimensional space, .

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig, 2013, Linguistic Regularities in Continuous Space Word Representations, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* pp. 746–751 Atlanta, Georgia. Association for Computational Linguistics.

Mokyr, Joel, 2016, *A Culture of Growth: The Origins of the Modern Economy* (Princeton University Press).

Pennington, Jeffrey, Richard Socher, and Christopher Manning, 2014, GloVe: Global Vectors for Word Representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* pp. 1532–1543 Doha, Qatar. Association for Computational Linguistics.

Perone, Christian S., Roberto Silveira, and Thomas S. Paula, 2018, Evaluation of sentence embeddings in downstream and linguistic probing tasks, *arXiv:1806.06259 [cs]*.

Sapienza, Paola, and Luigi Zingales, 2012, A Trust Crisis, *International Review of Finance* 12, 123–131.

——— , 2013, Economic Experts versus Average Americans, *American Economic Review* 103, 636–642.

Schularick, Moritz, and Alan M Taylor, 2012, Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008, *American Economic Review* 102, 1029–1061.

Sheng, Jinfei, 2019, Asset Pricing in the Information Age: Employee Expectations and Stock Returns, SSRN Scholarly Paper ID 3321275 Social Science Research Network Rochester, NY.

Soo, Cindy K., 2018, Quantifying Sentiment with News Media across Local Housing Markets, *Review of Financial Studies* 31, 3689–3719.

Spolaore, Enrico, and Romain Wacziarg, 2013, How Deep Are the Roots of Economic Development?, *Journal of Economic Literature* 51, 325–369.

Stulz, René M., and Rohan Williamson, 2003, Culture, openness, and finance, *Journal of Financial Economics* 70, 313–349.

Tetlock, Paul C., 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance* 62, 1139–1168.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, 2017, Attention Is All You Need, *arXiv:1706.03762 [cs]*.

Zhou, Guofu, 2018, Measuring Investor Sentiment, *Annual Review of Financial Economics* 10, 239–259.

Zingales, Luigi, 2012, *A Capitalism for the People: Recapturing the Lost Genius of American Prosperity* (Basic Books).

———, 2015, Presidential Address: Does Finance Benefit Society?, *Journal of Finance* 70, 1327–1363.

A Online Appendix

A.1 Positive and negative sentences used to define the positivity dimension across languages

Table OA.1: Positive and negative sentences

Positive sentences	Negative sentences
financial services benefit society	financial services damage society
finance is good for society	finance is bad for society
finance professionals are mostly good people	finance professionals are mostly corrupt people
finance positively impacts our world	finance negatively impacts our world
financial system helps the economy	financial system hurts the economy

(a) English

金融服务有益社会	金融服务损害社会
金融对社会好	金融对社会不好
财务专业人员大多很好	财务专业人员大多邪恶
金融对世界产生积极影响	金融对世界产生消极影响
金融系统帮助经济	金融系统有害金融

(b) Chinese

les services financiers profitent à la société	les services financiers nuisent à la société
la finance est bonne pour la société	la finance est mauvaise pour la société
les professionnels de la finance sont surtout bons	les professionnels de la finance sont surtout mauvais
la finance a un impact positif sur notre monde	la finance a un impact négatif notre monde
le système financier aide l'économie	le système financier nuit à l'économie

(c) French

Table OA.1: Positive and negative sentences, continued.

Positive sentences	Negative sentences
Finanzdienstleistungen kommen der Gesellschaft zugute	Finanzdienstleistungen schaden der Gesellschaft
Finanzen sind gut für die Gesellschaft	Finanzen sind schlecht für die Gesellschaft
Finanzprofis sind meistens gut	Finanzprofis sind meistens böse
Finanzen wirken sich positiv auf unsere Welt aus	Finanzen wirken sich negativ aus unsere Welt
Finanzsystem hilft der Wirtschaft	Finanzsystem schadet der Wirtschaft

(d) German

i servizi finanziari avvantaggiano la società	i servizi finanziari danneggiano la società
la finanza fa bene alla società	la finanza fa male alla società
i professionisti della finanza sono per lo più buoni	i professionisti della finanza sono principalmente cattivi
la finanza ha un impatto positivo sul nostro mondo	la finanza ha un impatto negativo il nostro mondo
il sistema finanziario aiuta l'economia	il sistema finanziario danneggia l'economia

(e) Italian

общество оказывает финансовую помощ	общество наносит ущерб финансовым услугам
финансы полезны для общества	финансы вредны для общества
профессионалы в области финансов в основном хороши	профессионалы в области финансов в основном злые
финансы положительно влияют на наш мир	финансы негативно влияют наш мир
финансовая система помогает экономике	финансовая система наносит ущерб экономике

(f) Russian

los servicios financieros benefician a la sociedad	los servicios financieros perjudican a la sociedad
los profesionales financieros son en su mayoría buenos	los profesionales financieros son en su mayoría malos
las finanzas impactan positivamente en nuestro mundo	las finanzas impactan negativamente nuestro mundo
el sistema financiero ayuda a la economía	el sistema financiero perjudica a la economía

(g) Spanish

Note: In line with Kozlowski, Taddy, and Evans (2019), we start with five pairs of words for the positive minus negative dimension for English (both American and British). The word pair includes: (positive – negative), (benefit – damage), (good – bad), (good – corrupt), and (help – hurt). We then create positive and negative sentences which discuss finance, using these words. For other languages, we translate these sentences with the help of native speakers.

A.2 Top ten worst to best ngrams sorted by finance sentiment

Table OA.2: Worst to best sentence sorted by finance sentiment

American English	British English
turmoil in the financial markets	turmoil in the financial markets
finances become disordered the	instability in the financial markets
financial panic swept the country	lack of money to finance
turmoil in financial markets	a financial panic
financial panic swept the nation	the financial panic
instability in the financial markets	financial panic in the united
financial panic in the country	international financial instability
severe financial setbacks	lack of funds to finance
a major financial panic	my finances falling short
world wide financial panic	the financial deficit
⋮	⋮
knowledge of the financial structure	finance graduate school of
financial support of the field	finance for small and medium
financial support of the course	understanding of the financial system
financial support of the science	financial support of the work
financial support of the graduate	financial management initiative
business and financial experience	financial support of this project
financial management of the organization	financial management of the business
financial support of the center	financial support of the research
finance in the graduate school	financial management of the school
the goal of financial management	financial support of the science

(a) English

Table OA.2: Worst to best sentence sorted by finance sentiment, continued

Chinese	English Translation
严重 扰乱 了 金融 秩序	Seriously disturbed the financial order
扰乱 了 国家 金融 秩序	Disrupt the national financial order
严重 扰乱 了 金融	Seriously disrupting the financial
扰乱 了 正常 的 金融	Disrupt the normal financial
扰乱 了 金融 秩 序	Disrupt the financial order of rank
扰乱 了 金融 秩序	Disrupt the financial order
扰乱 了 金融 市场	Disrupt the financial markets
干扰 了 金融 秩序	Disturb financial order
既 不 利于 金融	Not only is not conducive to financial
扰乱 了 金融 序	Disrupt the financial order
⋮	⋮
经济 发展 提供 金融	Economic development has provided financial
农村 发展 提供 金融	Rural Development provides financial
金融 推动 发展	Promote the development of financial
金融 服务 促进 农村	Promotion of rural financial services
金融 务 促进	Promote financial affairs
服务 促进 金融	Promoting financial services
金融 立足	Financial foothold
服务 农村 金融	Financial services in rural areas
金融 服务 社会	Financial services community
服务 规范 发展 金融	Regulate the development of financial services

(b) Chinese

Table OA.2: Worst to best sentence sorted by finance sentiment, continued

French	English Translation
ny a pas de finances	ny no finances
ministre des finances rené pleven	Finance Minister Rene pleven
the financial revolution in england	the financial revolution in england
état des finances était déplorable	financial condition was deplorable
bérenger finances et absolutisme	bérenger Finance and absolutism
finances est rejeté	Finance is rejected
mauvais état des finances royales	poor state of the royal finances
état des finances na pas	financial condition didnt
finances étaient en mauvais état	finances were in bad condition
finances na pu être déposé	na been filed Finance
⋮	⋮
encourager et à soutenir financièrement	encourage and support financially
mobiliser les ressources financières et	mobilize financial resources and
la gestion financière en	financial management
assistance financière et technique avec	financial and technical assistance with
à la coopération financière avec	financial cooperation with
réaliser la solidarité financière des	achieve financial solidarity
assurer la gestion financière et	the financial management and
organiser et de financer les	organize and finance
de promouvoir et de financer	promote and finance
à promouvoir et à financer	to promote and finance

(c) French

Table OA.2: Worst to best sentence sorted by finance sentiment, continued

German	English Translation
christian watrin bochum finanzpolitik	christian watrin Bochum financial policy
finanzmarkt kapitalismus	financial market capitalism
renzsch wolfgang finanzverfassung	renzsch wolfgang financial constitution
imperialismus staatsfinanzen rüstung	imperialismus government finances armor
gemeindefinanzgesetz vom dezember	community financial law from december
neoabsolutismus staatsfinanzen und politik	Absolutism government finances and politics
r a finanztheorie	r a financial theory
r a musgrave finanztheorie	r a musgrave finance theory
schmölders finanzpolitik berlin	Schmölders financial policy berlin
mayer geschichte der finanzwirtschaft	mayer history of finance economy
⋮	⋮
ist zuständig für die finanzielle	is responsible for the financial
finanziell und organisatorisch zu unterstützen	financial and organizational support
hilfe bei der finanzierung der	help with the financing of
finanzen die zur durchführung und	finance the implementation and to
hilfe bei der finanzierung von	help with the financing of
finanzieren mit	fund with
finanzierung erfolgt durch beiträge der	financed through contributions of the
unternehmen damit derartige finanzierungen	company so that such financing
sorgt für die finanzierung	provides for the financing
finanzen und mit zustimmung des	Finance and with the approval of

(d) German

Table OA.2: Worst to best sentence sorted by finance sentiment, continued

Italian	English Translation
dimissioni del ministro delle finanze	resignation of Finance Minister
il finanziamento è stato concesso	The loan was granted
le finanze sono condannate dai	finances are condemned by
scioglimento del contratto di finanziamento	termination of the loan agreement
ministro delle finanze è autorizzato	Minister of Finance is authorized
grave crisi finanziaria	serious financial crisis
il ministro delle finanze dichiarava	Finance Minister declared
le finanze saranno emanate	finances will be issued
la finanza sabauda allaprirsi	finance Savoy allaprirsi
l esercizio finanziario ha inizio	the financial year
⋮	⋮
disponibilità di risorse finanziarie che	availability of financial resources
di gestire le risorse finanziarie	to manage the financial resources
a soddisfare le esigenze finanziarie	to meet the financial needs
effettuare la gestione finanziaria di	make the financial management of
gestione delle risorse finanziarie e	management of financial resources and
relazioni economiche e finanziarie con	economic and financial relations with
coordinamento della finanza regionale con	coordination of regional finance with
assistere tecnicamente e finanziariamente i	assist technically and financially
gestione delle risorse finanziarie idonee	management of the financial resources
relazioni commerciali e finanziarie con	commercial and financial relations with

(e) Italian

Table OA.2: Worst to best sentence sorted by finance sentiment, continued

Russian	English Translation
обращение финансы кредит	recourse finance loan
плутократия бароны финансового	plutocracy financial barons
буржуазии финансовый срыв	Financial breakdown of the bourgeoisie
протекционизм господство финансистов	Protectionism domination of financiers
финансов кредита социализме	Finance socialism loan
империализм финансовый капитализм	financial capitalism, imperialism
обращение кредит финансы	recourse loan finance
финансовое банкротство	financial bankruptcy
империализма колониальное финансовое порабощение	colonial imperialism financial enslavement
страшных финансовых грозных	terrible financial formidable
⋮	⋮
оказывает и финансовую поддержку	and providing financial support
оказывает колхозам финансовую помощь	It provides financial assistance to collective farms
финансовая деятельность колхоза осуществляется на	financial activities carried out on a collective farm
обеспечивается финансирование мероприятий	provided funding
оказывает финансовую и помощь	and provides financial assistance
оказывает финансовую и поддержку	It provides financial support and
оказывает финансовую и политическую поддержку	It provides financial and political support
оказывает большую финансовую помощь	providing more financial aid
оказывает значительную финансовую помощь	providing substantial financial assistance
оказывает финансовую и техническую помощь	It provides financial and technical assistance

(f) Russian

Table OA.2: Worst to best sentence sorted by finance sentiment, continued

Spanish	English Translation
la finanza no era	finance was not
la situación financiera no era	the financial situation was not
el capital financiero se sentirá	financial capital will feel
el capital financiero no es	financial capital is not
el mercado financiero no es	the financial market is not
la especulación financiera domine su	financial speculation dominates its
el sistema financiero se vio	the financial system was
una desgraciada situación financiera pudiese	an unfortunate financial situation could
el déficit se financió	The deficit was financed
su situación financiera no era	its financial situation was not
⋮	⋮
actividades financieras y de servicios	financial activities and services
asesoría técnica y apoyo financiero	technical advice and financial support
financiamiento de las diversas actividades	financing various activities
apoyo técnico y financiero internacional	international technical and financial support
apoyo financiero y asistencia técnica	financial support and technical assistance
apoyo financiero a las actividades	financial support to activities
apoyo financiero para las actividades	Financial support for activities
asistencia técnica y recursos financieros	technical assistance and financial resources
financiamiento de las actividades culturales	financing of cultural activities
asistencia técnica y de financiamiento	technical assistance and financing

(g) Spanish

Note: The sentences are sorted from worst to best in terms of their cosine similarity with positive minus negative vector. The English translation is provided using Google Translate.