Quantifying Minsky Cycles

Kim Ristolainen *

February 2026

Abstract

We develop a novel sentiment measure derived from survey data to empirically validate the Minsky–Kindleberger view on financial crises. Using survey data from multiple countries, we decompose beliefs into components explained by public information that are orthogonal to optimal machine beliefs, constructing a framework that isolates sentiment and its dispersion among individuals. We show that deviations from machine-optimized benchmarks arise from systematic misaggregation of public information. The sentiment measure is validated through its predictive relationships with financial markets and belief dynamics consistent with heterogeneous-beliefs asset pricing theory. We extend this sentiment measure historically for a panel of 78 countries using machine learning models trained on BERT embeddings of historical news articles (1903–2020). The backcasted sentiment shows that shocks in median sentiment predict credit booms in the non-tradable corporate sector, which prior research has linked to financial crises, providing the first historically large-scale empirical validation of the Minsky cycle. We further show that sentiment, which is a misaggregation of public information, is influenced by memory-related dynamics, as the time elapsed since major crises and the share of young-to-old people in the population strongly predict surges in optimism even when recent economic developments are controlled for.

JEL Codes: E44, E51, G01, D84, G41, E32

Keywords: Survey data, Sentiment, Memory, Machine Learning, Text Data, Credit growth, Financial Crisis

*Acknowledgements. I thank the Yrjö Jahnsson Foundation and the Emil Aaltonen Foundation for financial support, and I thank Alp Simsek for early discussions and comments on preliminary ideas.
Contact: Department of Accounting and Finance, University of Turku. E-mail: kkrist@utu.fi.

1. Introduction

The role of sentiment in driving macroeconomic fluctuations and financial crises has long intrigued economists. From Keynes's notion of "animal spirits" to Minsky's hypothesis of cyclical instability (Minsky, 1977, 1986), sentiment is often invoked as a central force behind economic dynamics. Yet, empirical research that demonstrates its historical role for financial instability has been limited. The main obstacle has been the lack of a consistent sentiment measure that is derived from actual beliefs rather than financial market data, as well as the absence of data covering long historical periods and broad country samples, given that survey-based indicators exist only for recent decades. This paper develops a novel framework that quantifies sentiment from existing survey data on individual beliefs and extends its time and country coverage through state-of-the-art natural language processing of historical news text. We validate the so-called Minsky cycle by utilizing an unusually extensive historical panel dataset on sentiment covering 78 countries and more than a century of economic history, showing that increases in median sentiment predict "bad" credit growth historically associated with financial crises. We further provide historical evidence on the role of memory-based belief formation mechanisms behind rising sentiment, thereby linking belief formation, bad credit booms, and financial instability in historical data at a scale not previously analyzed.

The first contribution of this paper is the development of a new sentiment measure based on survey beliefs. Using monthly survey data from 18 countries for the period from 1991 to 2020, we decompose individual beliefs into components explained by past public information that is orthogonal to optimal machine beliefs with public information. This approach enables us to measure both the level of sentiment and its dispersion across forecasters. We find that the majority of both the time-series and cross-sectional variation in beliefs can be explained by public information, though this explanatory power declines sharply during recessions. Survey forecasters' beliefs systematically deviate from machine-optimized benchmarks that outperform all individual forecasters, primarily due to misaggregation of public information. The sentiment measure is further validated through its predictive relationships with financial markets and expectations, all of which react in ways consistent with the heterogeneous-beliefs asset pricing framework of Thesmar and Verner (2025) where sophisticated investors require higher expected returns in response to the rising sentiment of naive investors.

To credibly study the relationship between sentiment, credit growth, and financial crises, one needs belief data that covers much longer time periods and a far broader set of countries than existing survey datasets allow. We address this limitation by extending our sentiment measure historically and across a panel of 78 countries, using textual data to reconstruct sentiment for country-months without survey responses. Specifically, we construct BERT embeddings of

Wall Street Journal article titles spanning 1903–2020 separately for general and country specific news, reduce their dimensionality using principal component analysis, and train random forest models to predict both the median sentiment and its dispersion among forecasters. This approach enables the reconstruction of sentiment measures for periods without survey data, yielding a long-run historical dataset that addresses the critique of Brunnermeier et al. (2021) regarding the limited time span and narrow coverage of existing belief data, and makes it possible to analyze sentiment dynamics over more than a century with cross sectional variation across countries and individuals. Panel local projection results show that rising sentiment predicts credit booms in the non-tradable corporate sector that past research has associated with a higher probability of future financial crises (Müller and Verner, 2024) and weaker subsequent output growth. This provides the first empirical validation of the Minsky–Kindleberger view using a new long-run sentiment measure that has not been available before, demonstrating that optimism rooted in misaggregation of public information predicts credit booms in sectors most closely associated with subsequent crises.

Sufi and Taylor (2022) emphasize that understanding financial crises requires studying the causes of the "bad" credit booms that precede them. The third contribution of this paper is to provide evidence on these origins. Recent theoretical and survey experimental work by Bordalo et al. (2025a); Graeber et al. (2024) and Jiang et al. (2025) rooted in psychological research on memory recall shows that individuals' expectations are shaped by the experiences they can recall relative to the question at hand — experiences that may be relevant or irrelevant to the belief they are forming. Consistent with these findings, we present descriptive evidence that the time since a major economic crisis and the share of young relative to old people in the population strongly predict peaks in sentiment, even when recent output growth and financial market movements are controlled for. This highlights the behavioral foundation of Minsky cycles: optimism, driven by systematic misaggregation of the same public information, tends to re-emerge as the memory of past instability fades.

This work relates to several strands of literature. First, it connects to the extensive research on the endogenous nature of financial crises, from the classic accounts of Minsky (1977) and Kindleberger (1978) to more recent syntheses such as Sufi and Taylor (2022) and Frydman and Xu (2023). A central insight of this literature is that financial instability emerges endogenously through credit booms that build up during periods of optimism and apparent stability. A large body of empirical work shows that crises are preceded by rapid credit expansions, especially to non-tradable sectors (Reinhart and Rogoff, 2009; Schularick and Taylor, 2012; Mian et al., 2017; Baron and Xiong, 2017; Baron et al., 2021; Krishnamurthy and Muir, 2025; Müller and Verner, 2024). Our paper contributes to this literature by providing the first empirical validation of the Minsky–Kindleberger view based on a previously unavailable long-run sentiment measure. We show that optimism arising from systematic misaggregation of public information predicts credit expansions concentrated in sectors historically linked to crises, thereby empirically connecting belief dynamics to the mechanism at the heart of financial instability.

A second related strand is the growing literature on belief formation and belief distortions. Seminal work by Gennaioli and Shleifer (2018) links investor psychology to financial fragility, showing how expectations shaped by memory and salient experiences lead to underestimated risks and overextended credit, and how their abrupt revision can trigger crises. A large empirical literature further documents that individuals systematically deviate from rational updating, tending to over- or underreact to news (Coibion and Gorodnichenko, 2015; Bordalo et al., 2020, 2024).

A number of research (Patton and Timmermann, 2010; Angeletos et al., 2021; Bianchi et al., 2022a; Juodis and Kučinskas, 2023; de Silva and Thesmar, 2023; Bhandari et al., 2024; Maenhout et al., 2025; Bybee, 2024) develops empirical measures of beliefs and sentiment with varying procedures to quantify expectation distortions and their relevance for the economy. This paper contributes to that literature by introducing a new measurement framework that identifies sentiment as the component of individual beliefs explained by past public information but orthogonal to an optimal benchmark formed with the same information set. We then show how these belief distortions propagate into credit dynamics and crisis risk.

The paper also relates to the rapidly expanding literature using text data in economics to recover economic signals (Baker et al., 2016; Manela and Moreira, 2017; Gentzkow et al., 2019; Hassan et al., 2019; Bybee et al., 2023; Ash and Hansen, 2023; Bybee et al., 2024). By reconstructing sentiment from historical news and extending belief measures over more than a century and across many countries, our approach addresses the limitations of existing survey-based measures and enables new empirical analyses of belief dynamics over long horizons.

Finally, our work relates to the literature on how individual experience and memory shape economic behavior and expectations (Malmendier and Nagel, 2011, 2016; Nagel and Xu, 2022; Malmendier and Wachter, 2024), as well as to recent theoretical and experimental research on belief formation grounded in psychological evidence on memory recall and similarity (Kahana and Kahana, 2012; Bordalo et al., 2025b, 2023, 2025a; Jiang et al., 2025; Wachter and Kahana, 2024; Graeber et al., 2024). We contribute by providing large-scale historical evidence consistent with memory recall influencing belief formation, showing that fading memories of past crises and demographic shifts help explain peaks in sentiment, revealing how memory-based belief formation is connected to the Minsky cycle.

The remainder of the paper proceeds as follows. Section 2 introduces the survey-based sentiment measure, explaining its construction and validating its relevance. Section 3 describes the backcasting methodology and presents the extended historical dataset. Section 4 examines the predictive power of sentiment for bad credit booms, linking the findings to the Minsky cycle and explores the behavioral foundations of sentiment, focusing on the role of memory and its implications for belief dynamics. Section 5 concludes with a discussion of broader implications.

2. Measuring sentiment

In this section, we describe how sentiment can be measured from survey data with machine learning algorithms and how the resulting measures correlate with financial markets.

2.1. Aggregation mistakes with public and private information

Consider predictions of a future outcome y (e.g. GDP growth) based on a set of publicly observable signals $x = (x_1, x_2)$. Suppose the objective signal structure is given by

$$\begin{aligned} x_1 &= y + \varepsilon_1, \quad \varepsilon_1 \sim N(0, \tau_1^{-1}), \\ x_2 &= y + \varepsilon_2, \quad \varepsilon_2 \sim N(0, \tau_2^{-1}), \end{aligned}$$

where the signal errors ε_1 and ε_2 are orthogonal to each other and to y. Forecasters also observe a private signal z, unavailable to the econometrician, with

$$z = y + \varepsilon_z, \quad \varepsilon_z \sim N(0, \tau_z^{-1}),$$

whose innovation ε_z is orthogonal to the public-information errors $\varepsilon_1, \varepsilon_2$. The full-information Bayesian forecast is therefore

$$\begin{aligned} E[y|x, z] &= b_1 x_1 + b_2 x_2 + b_z z, \\ b_i &= \frac{\tau_i}{\tau_1 + \tau_2 + \tau_z}, \quad i \in \{1, 2, z\}. \end{aligned}$$

We allow survey forecasts to be arbitrary linear functions of available information,

$$y^f = b_1^f x_1 + b_2^f x_2 + b_z^f z,$$

where deviations $b_i^f - b_i$ reflect aggregation mistakes. Since private information is unobserved, b_z^f cannot be identified. However, under orthogonality of the private-information innovation, aggregation mistakes in public information can still be partially recovered. The Bayesian forecast that conditions only on public information is

$$E^m[y|x] = b_1^m x_1 + b_2^m x_2, \quad b_i^m = \frac{\tau_i}{\tau_1 + \tau_2}, \ i \in \{1, 2\}.$$

Importantly, public signals receive the same *relative* weights under partial- and full-information Bayesian forecasts: $b_1^m / b_2^m = b_1 / b_2$.

Private information affects only the overall weight placed on public signals. This property allows us to identify misaggregation of public information. Projecting outcomes and survey forecasts onto public signals yields

$$\begin{aligned} \hat{y} &= P_{y|x} y = \hat{b}_1^m x_1 + \hat{b}_2^m x_2, \\ \hat{y}^f &= P_{y^f|x} y^f = \hat{b}_1^f x_1 + \hat{b}_2^f x_2, \end{aligned}$$

where the second line follows because the private-information innovation is orthogonal to x, so the projection recovers b_1^f and b_2^f. Proportional differences between $(\hat{b}_1^f, \hat{b}_2^f)$ and $(\hat{b}_1^m, \hat{b}_2^m)$ are consistent with rational use of private information, while deviations in relative weights indicate misaggregation. Operationally, we extract these deviations by projecting $\hat{y}^f$ onto $\hat{y}$ and retaining the residual,

$$\tilde{y}^f = M_{\hat{y}^f|\hat{y}} \hat{y}^f.$$

Under correct aggregation of public information, this residual is zero. A nonzero residual captures systematic mistakes in aggregating public data. In summary, our procedure consists of two steps: (i) project outcomes and forecasts onto public information, and (ii) isolate the component of predicted forecasts that is orthogonal to predicted outcomes. Under mild assumptions, this component measures sentiment as misaggregation of public signals.

2.2. A Kalman-filter example

We illustrate how aggregation mistakes arise in a dynamic environment in which agents update beliefs about a latent economic state using Kalman filters. Differences in belief-updating rules generate systematic misaggregation of public information, which our sentiment measure is designed to capture.

2.2.1. Heterogeneous belief formation

Denote agents by the superscript $j \in J \cup \{m\}$, where $j \neq m$ indexes survey respondents and $j = m$ denotes the machine benchmark. Agents seek to forecast a variable with a persistent and a transitory component,

$$
\begin{aligned}
g_t &= \mathbf{g}_t + v_t, \\
\mathbf{g}_t &= \rho \mathbf{g}_{t-1} + \eta_t, \qquad v_t \sim N(0, \sigma_v^2),\ \eta_t \sim N(0, \sigma_\eta^2),
\end{aligned} \tag{1}
$$

where $\mathbf{g}_t$ is a latent persistent state. The long-run mean is normalized to $g^* = 0$. Beliefs about future outcomes are fully characterized by beliefs about the current state $E_t^j[\mathbf{g}_t]$. Let $\mathbf{x}_t$ denote a vector of publicly observable variables that are informative about $\mathbf{g}_t$, including the current realization g_t. Agent j believes that

$$
\mathbf{g}_t = \mathbf{b}^j \mathbf{x}_t + \mu_t^j + e_t^j, \qquad e_t^j \sim N\left(0, (\sigma_e^j)^2\right), \tag{2}
$$

where $\mathbf{b}^j$ captures how the agent aggregates public information and μ_t^j represents an idiosyncratic interpretation or private signal, whose innovations are orthogonal to public data. The residual variance $(\sigma_e^j)^2$ is allowed to differ across agents, implying heterogeneous Kalman gains.

Let $\mathbf{g}_t^j$ denote agent j's belief about the persistent state. Under a learning steady state, beliefs follow the Kalman filter

$$
\mathbf{g}_t^j = (1-\kappa^j)\rho \mathbf{g}_{t-1}^j + \kappa^j \left(\mathbf{b}^j \mathbf{x}_t + \mu_t^j\right),
$$

where κ^j is the agent-specific Kalman gain. Iterating this expression for L periods yields the regression representation

$$
\begin{aligned}
\mathbf{g}_t^j &= \left((1-\kappa^j)\rho\right)^L \mathbf{g}_{t-L}^j + \sum_{l=0}^{L} \left((1-\kappa^j)\rho\right)^l \kappa^j \mathbf{b}^j \mathbf{x}_{t-l} + \sum_{l=0}^{L} \left((1-\kappa^j)\rho\right)^l \kappa^j \mu_{t-l}^j \\
&= \underbrace{\text{constant} + \beta_g^j \mathbf{g}_{t-L}^j + \sum_{l=0}^{L} \beta_{\mathbf{x},l}^j \mathbf{x}_{t-l}}_{\text{predictable}} + \underbrace{\varepsilon_t^j}_{\text{residual}},
\end{aligned} \tag{3}
$$

which decomposes beliefs into a component predictable from public data and past beliefs, and a residual capturing idiosyncratic interpretations. We maintain the assumption that ε_t^j is orthogonal to public information, allowing Eq. (3) to be estimated.

2.2.2. A machine benchmark

When agents employ heterogeneous belief-updating rules, the consensus forecast need not coincide with the model that correctly aggregates public information. Following Bianchi et al. (2022a), we construct a machine-learning benchmark that enforces optimal aggregation of public signals while excluding private information. Suppose the latent state satisfies

$$\mathbf{g}_t - g^* = \mathbf{b}^m \mathbf{x}_t + e_t^m, \qquad e_t^m \sim N\Big(0, (\sigma_e^m)^2\Big). \tag{4}$$

Unlike agents, the machine has no idiosyncratic interpretation ($\mu_t^m = 0$). In a learning steady state, machine beliefs evolve according to

$$\begin{aligned}
\mathbf{g}_t^m &= (1-\kappa^m)\rho \mathbf{g}_{t-1}^m + \kappa^m \mathbf{b}^m \mathbf{x}_t \\
&= ((1-\kappa^m)\rho)^L \, \mathbf{g}_{t-L}^m + \sum_{l=0}^{L} ((1-\kappa^m)\rho)^l \, \kappa^m \mathbf{b}^m \mathbf{x}_{t-l}.
\end{aligned}$$

Because $\mathbf{g}_t^m$ is unobserved, we recover it recursively by regressing the observed outcome g_t on public information and lagged machine beliefs. Using $g_t = \mathbf{g}_t + v_t$ yields

$$\begin{aligned}
g_t &= ((1-\kappa^m)\rho)^L \, \mathbf{g}_{t-L}^m + \sum_{l=0}^{L} ((1-\kappa^m)\rho)^l \, \kappa^m \mathbf{b}^m \mathbf{x}_{t-l} + \tilde{v}_t + v_t \\
&= \underbrace{\text{constant} + \beta_g^m \mathbf{g}_{t-L}^m + \sum_{l=0}^{L} \beta_{\mathbf{x},l}^m \mathbf{x}_{t-l}}_{\mathbf{g}_t^m} + \tilde{v}_t + v_t,
\end{aligned} \tag{5}$$

which provides an empirical analogue to Eq. (3). Initial machine beliefs are set to zero.

Our analysis compares estimated versions of Eqs. (3) and (5) to identify systematic differences in how agents and the machine aggregate public information. These differences form the basis of our empirical sentiment measure, which we apply to survey forecasts of GDP growth in the next section.

2.3. From theory to measurement with survey data

The preceding subsections illustrated, in stylized form, how forecasters can misaggregate public information relative to an optimal benchmark. We now implement this framework empirically using *Consensus Economics* monthly survey expectations of real GDP growth for 18 countries from 1991 to 2020. Because the survey asks for calendar-year forecasts, the horizon for a forecast differs across months. We therefore linearly interpolate between the current-year and next-year forecasts to obtain a balanced monthly one-year-ahead forecast series (see Appendix A.1). Next, we construct optimal machine forecasts that use the same public information[1] that was available to human forecasters at the time they made their predictions. Sentiment is then measured as

[1]The survey data, as well as the macroeconomic and financial variables, are described in Sections A.1 and A.2.

the component of survey beliefs that can be explained by public data yet remains orthogonal to the optimal machine benchmark. This section details the machine-learning specification and validation of the resulting sentiment measure.

2.3.1. Predictability of beliefs and the machine benchmark

To obtain the component of individual beliefs that can be predicted from public information, we estimate for each forecaster j and period t the following specification:

$$\hat{g}^{j}_{c,t} \equiv \mathbb{E}[g^{j}_{c,t} \mid \mathbf{x}_{t-1}] = f^{j}_{t}(\mathbf{x}_{t-1}) + \varepsilon^{j}_{c,t}, \tag{6}$$

where $g^{j}_{c,t}$ denotes the belief (forecast) of individual j about real GDP growth $g_{c,t+h}$ in country c formed at time t, with forecast horizon $h = 12$ months. The vector $\mathbf{x}_{t-1}$ is the empirical analogue of the public information set G_{t-1} from Section 2.2, including publicly available macroeconomic and financial variables as well as past consensus and individual beliefs observed up to period $t-1$. The function $f^{j}_{t}(\cdot)$ denotes the random-forest (Breiman, 2001) predictor estimated separately for each forecaster j and time period t given public information available up to that period.

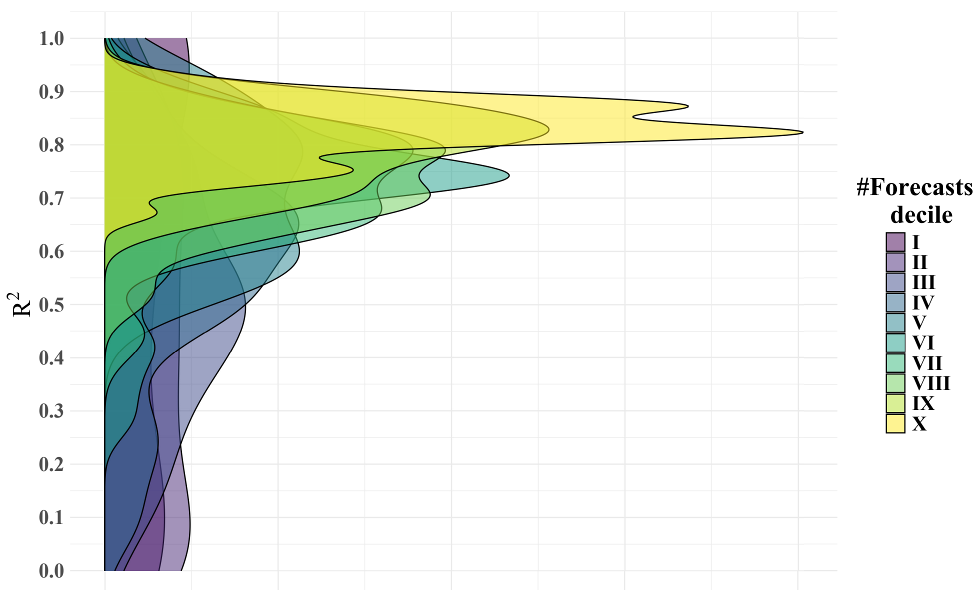


Figure 1: Explainability of beliefs: time-series R^2 distribution across forecasters.

Note: The figure shows the distribution of time-series R^2 values measuring how well individual beliefs are explained by past public information, separately for forecaster groups differing in the number of forecasts made.

In this framework, differences in $f^{j}_{t}(\cdot)$ across individuals and over time reflect heterogeneity in the mental models agents use to process the same public data. The estimation procedure yields out-of-sample predictions $\hat{g}^j_{c,t}$ that capture the portion of each forecaster's belief explainable by past public information. The random forest algorithm[2] is chosen for its flexibility in modeling complex, nonlinear relationships, its robustness to overfitting, and its consistently strong predictive performance across a wide range of forecasting problems relative to other machine learning methods (Fernández-Delgado et al., 2014).

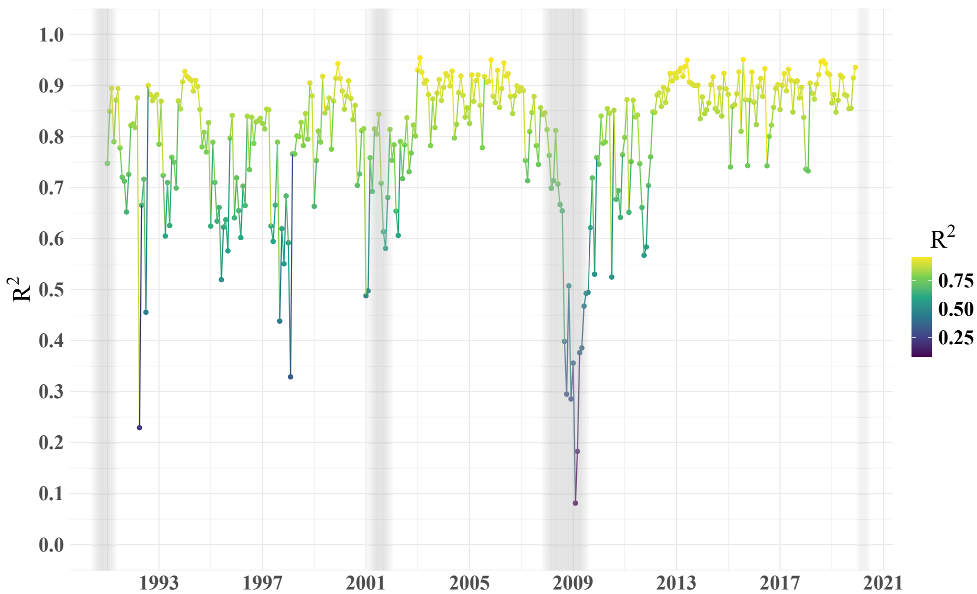


Figure 2: Explainability of beliefs: cross-sectional R^2 in time.

Note: The figure shows the evolution of the monthly cross-sectional R^2 over time, measuring how well individual beliefs are explained by past public information.

Analogously, we estimate the machine (benchmark) belief, denoted by $\hat{g}^m_{c,t}$, which corresponds to the theoretical machine belief g^m_t in Section 2.2. The same procedure is applied, except that the model is trained to predict the actual GDP growth $g_{c,t+h}$ instead of individual forecasts:

$$\hat{g}^m_{c,t} \equiv \mathbb{E}[g_{c,t+h} \mid \mathbf{x}_{t-1}] = f^m_t(\mathbf{x}_{t-1}) + \varepsilon^m_{c,t}. \tag{7}$$

Here, $f^m_t(\cdot)$ represents the random-forest model estimated with public information available up to period t to produce the optimal machine prediction of future GDP growth from the same public information set that was available to individual forecasters when they formed their corresponding expectations.

[2]A random forest is an ensemble learning method that combines the predictions of multiple decision trees. Each tree is trained on a random subset of the data and features, and their predictions are aggregated (e.g., via averaging) to produce the final output. This approach reduces variance and improves prediction accuracy while avoiding overfitting (Breiman, 2001).

Figures 1 and 2 visualize how much of individual beliefs can be explained with our approach and hence past public data. Figure 1 shows that on average around 75% of the time-series variation of an individual beliefs can be explained with past public information. This share is likely to be even larger as the estimate is over 80% for individuals with the largest number of forecasts made. Figure 2 shows that also a majority of the cross-sectional variation in individual beliefs can be explained with public information and different mental models. This share varies in time as most of the time, clearly over 75% of this variation can be explained, but during recessions or periods of uncertainty this share drops dramatically to 10-60% range.

Acknowledging the fact that these results are derived with a large panel dataset with 18 countries, 29 years of data, hundreds of forecasters and close to 34.000 monthly survey observations, our analysis provides strong evidence on belief variations spanning from individuals having different mental (aggregation) models with the same data rather having varying information when forming beliefs. In particular, a significant fraction of cross-sectional belief heterogeneity can be explained from public aggregate data and agents' past individual beliefs, which are themselves public (available to forecasters). This suggests that heterogeneous information is unlikely to explain observed belief heterogeneity. Rather agents seem to have different belief-updating *models*: they look at the same data and come to different conclusions.

Table 1 provides quite astonishing results. Firstly, the machine learning model outperforms basically all of the individual forecasters with only past public information. This result already implies that people do not optimally aggregate information when they are forming beliefs about the future. In addition, the size of the individual survey errors relative to the ones of the optimal machine beliefs are over 2.5 times larger on average in MSE terms. The systematic difference between individual and machine beliefs is clearly illustrated in Figure 3, which shows the machine forecast error alongside the distribution of individual forecasters' errors across countries and over time. It seems that both the individuals and the machine are usually making mistakes to the same direction, but the errors of the individuals are larger. These larger errors for individuals occur during the dot-com bubble, 9/11 turmoil in 2001 and the global financial crisis of 2008. This implies that individuals underestimate tail-risks or their impact to the economy. As individual errors to both directions (positive and negative) are larger than machine's belief errors, it seems that individuals overreact to both directions.

Table 1: Survey and machine beliefs of one year ahead growth.

	Survey forecaster sample		
	Forecasts > 9	Forecasts > 49	Forecasts > 99
Survey information			
Countries (N)	18	18	18
Periods (N)	348	348	348
Period	1/1991–12/2019	1/1991–12/2019	1/1991–12/2019
Forecasters (N)	319	175	109
Monthly forecasts (N)	33,906	29,822	24,915
Explainability of beliefs			
Cross-sectional R^2 (%)	80.2	81.1	81.7
Time series R^2 (%)	67.8	76.2	79.0
Machine beliefs vs. survey beliefs			
$MSE_{Machine}$	0.773	0.795	0.772
$MSE_{Forecaster}$	2.21	2.22	2.14
Average $MSE_{Machine}/MSE_{Forecaster}$	0.402	0.370	0.379
$MSE_{Machine} < MSE_{Forecaster}$ (%)	98.1	100	100

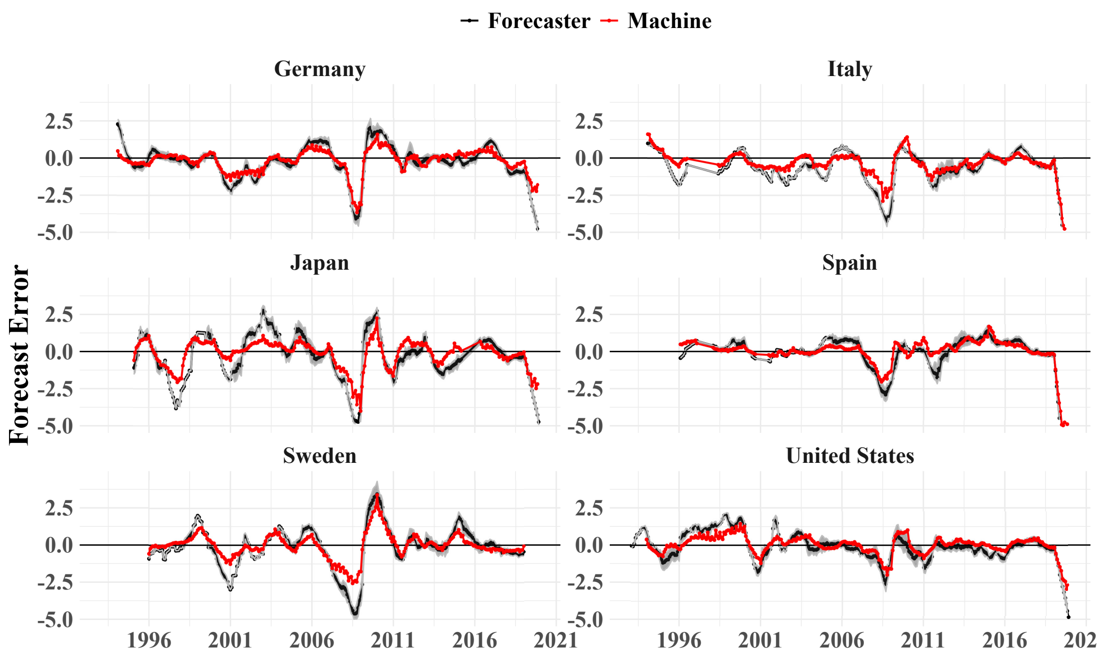


Figure 3: Survey forecast error distribution and machine forecast error in time.

Note: The figure visualizes in red the out-of-sample forecast error (defined as $g_{c,t} - \hat{g}^m_{c,t}$) of a random forest model predicting one year-ahead real GDP growth with only public past data available up to the period when prediction is made. The black line visualizes the median prediction error (defined as $g_{c,t} - g^j_{c,t}$) across individual forecasters. The shaded grey are represents the 10th and 90th percentile of the individual forecast error distribution. The forecasters who have made at least 50 forecasts are included in the graph.

2.3.2. Constructing sentiment and dispersion measures

Next, we estimate our main measure of sentiment $B^j_{c,t}$, that captures the component of individual beliefs $\hat{g}^j_{c,t}$ that can be explained by past public data, but is orthogonal to the optimal machine belief based on the same information, $\hat{g}^m_{c,t}$. Formally,

$$B^j_{c,t} = \hat{g}^j_{c,t} - f^j(\hat{g}^m_{c,t}), \tag{8}$$

where $f^j(\cdot)$ denotes the random forest prediction of $\hat{g}^j_{c,t}$ given the optimal machine forecast $\hat{g}^m_{c,t}$ of real GDP growth from past public data. In addition, we construct an alternative sentiment measure for belief updates in an analogous way:

$$B^j_{c,t} = \Delta\hat{g}^j_{c,t} - f^j(\Delta\hat{g}^m_{c,t}), \tag{9}$$

where sentiment is defined as the component of belief updates that is orthogonal to optimal machine updates based on public data. We estimate the random forest model separately for each forecaster j in the sample and derive sentiment $B^j_{c,t}$ for all available forecasts across countries c and time periods t.

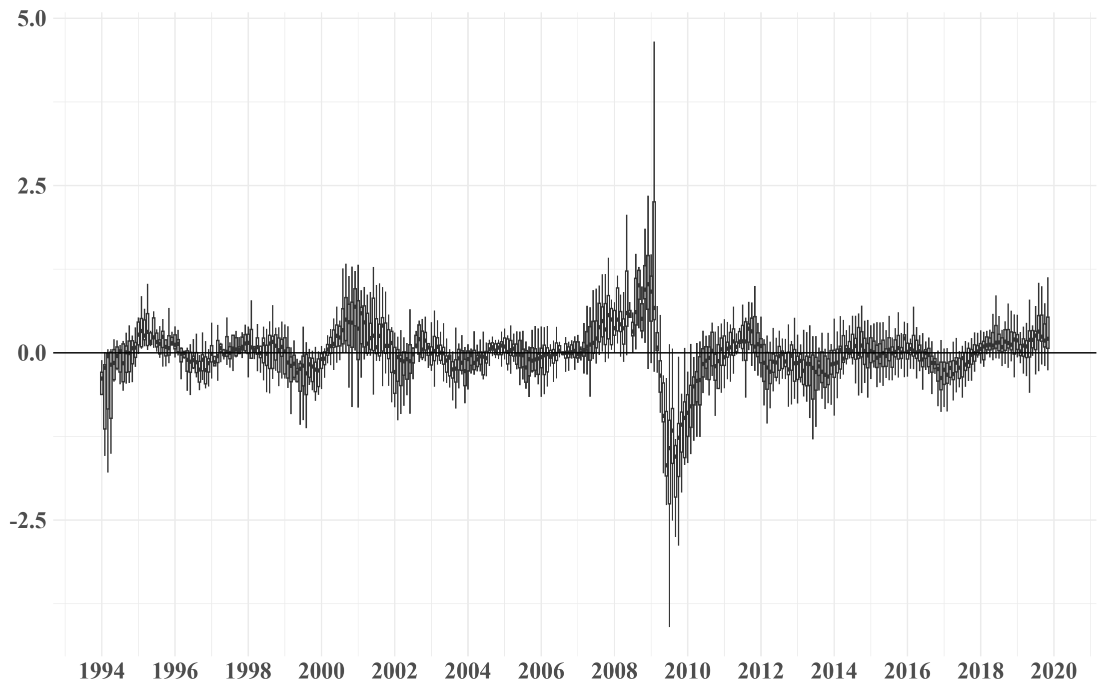


Figure 4: Population median sentiment distribution across 18 developed countries in time.

Note: Forecaster sentiment is defined as $B^j_{c,t} = \hat{g}^j_{c,t} - f^j(\hat{g}^m_{c,t})$, where $\hat{g}^j_{c,t}$ is the objective past public information prediction of the forecaster j's forecast and $\hat{g}^m_{c,t}$ is the objective past public information prediction of one year ahead real GDP growth. Sentiment $B^j_{c,t}$ is the part of forecaster's subjective forecast that can be explained with past public information, but that is also orthogonal to the objective forecast of the forecasted outcome with past public information.

Figure 4 visualizes the distribution of the median sentiment across countries for a given month in our time sample. There are a couple of interesting takeaways from this visualization. First, there seems to be a global cyclical component in sentiment. From 1994 to 2020 one can see around twelve global sentiment cycles in the data. Some cycles are stronger that others e.g. greater distance from trough to peak sentiment.

The cycles seem to last only around two years with the exception of the cycles prior and post the global financial crisis of 2008 where the cycles seem to last around 3 years. Finally, from 2002 to 2007 global sentiment seem to stay surprisingly neutral. This is interesting as the narrative about the time preceding the GFC was that there was excess optimism for many years until the year the crisis started. However, this figure visualizes the distribution of the median sentiment across the country sample and the sentiment might have been specific to the US or a group of countries and specifically to the housing sector and not the economy as a whole.

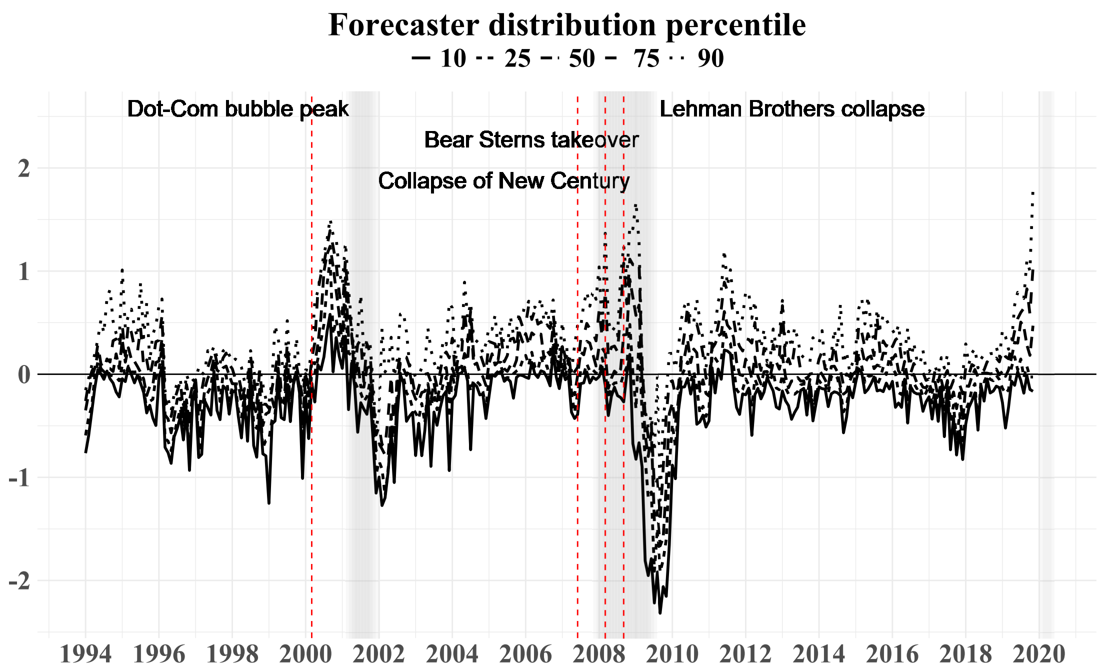


Figure 5: Individual forecaster monthly sentiment distribution percentiles for the United States.

Figure 5 visualizes the distribution of individual forecasters' sentiment only for the US. In addition, to the location of the sentiment distribution, here we can also see the dispersion or concentration of sentiment among forecasters in a given time period. Interestingly the dispersion of sentiment varies in time. As an example the 3-4 years prior to the GFC, the most pessimistic individuals were quite neutral for a prolonged period, but the most optimistic individuals' sentiment increased steadily from mid 2004 all the way up to the end 2007. A closer look at the major events prior to the GFC shows that after the collapse of New Century in April 2007, the takeover of Bear Sterns in March 2008 and the bankruptcy of Lehman Brothers in September 2008 the most pessimistic forecasters stayed neutral (close to optimal machine belief) or became somewhat pessimistic where as the more optimistic forecasters jumped back to optimism. Excess pessimism took hold of all forecasters after the Lehman Brothers bankruptcy, but for a couple of months after this event the disagreement of forecasters increased to highest levels seen in the data as the pessimistic forecasters became increasingly more pessimistic at the same time when the most optimistic forecasters became more and more optimistic. After the disagreement peak of the end of 2008, even the most optimistic forecasters changed their sentiment from extremely optimistic to extremely pessimistic. To interpret the size of this sentiment change, the real GDP growth was first expected to be almost 2 percentage points higher relative to what the optimal machine belief was and then it dropped to 2 percentage points too low what the optimal machine belief implied.

2.3.3. Validation and interpretation

Next, we analyze how different financial market variables react to a shock in the median sentiment by estimating panel local projection regressions (Jorda, 2005), were the response variable $\Delta Y_{c,t+h}$ is the $h = 0, \ldots, 20$ month ahead change of the median survey forecast, median forecast error, median sentiment, policy rate, average corporate bond spread, average dividend-to-price ratio, or the cumulative stock return for country c. More formally,

$$\Delta Y_{c,t+h} = Y_{c,t+h} - Y_{c,t} = \alpha_c^h + \beta^h S_{c,t} + \gamma^h \text{controls}_{c,t-1} + \epsilon_{i,t+h}. \tag{10}$$

We include country fixed effects α_c^h and also control for six lags of policy rates, stock prices, median sentiment and the response variable. We report both 68% and 95% confidence intervals calculated according to Driscoll and Kraay (1998) standard errors with a lag length of six months. We run the regression separately for countries with restrictions in monetary policy that are in the European Exchange Rate Mechanism (ERM) and for countries that are not (non-ERM) so that we can rule out the effect that monetary policy has on the response variables as in Caballero and Simsek (2020).

Figure 6 shows the local projection estimates for shocks to sentiment in belief updates. In both ERM and non-ERM countries, median one-year-ahead forecasts rise after a positive sentiment shock and only slowly return to their earlier levels after around 10 months. Forecast errors widen at the same time, as expectations become systematically too optimistic relative to realized outcomes, and these errors only unwind roughly twelve months later. Asset prices adjust much faster as equity prices fall on impact, while dividend–price ratios and corporate bond spreads rise, consistent with an increase in required returns.

These patterns match the framework and evidence of Thesmar and Verner (2025) who show that most movements in valuation ratios reflect time-varying expected returns, not revisions in expected cash flows, and that sophisticated investors' required returns move negatively with the swings in naive expectations. In that light, our results can be read as follows: a positive sentiment shock captures a rise in naive optimism, while rational investors respond by demanding higher returns, producing an immediate fall in stock prices and higher dividend yields and spreads.

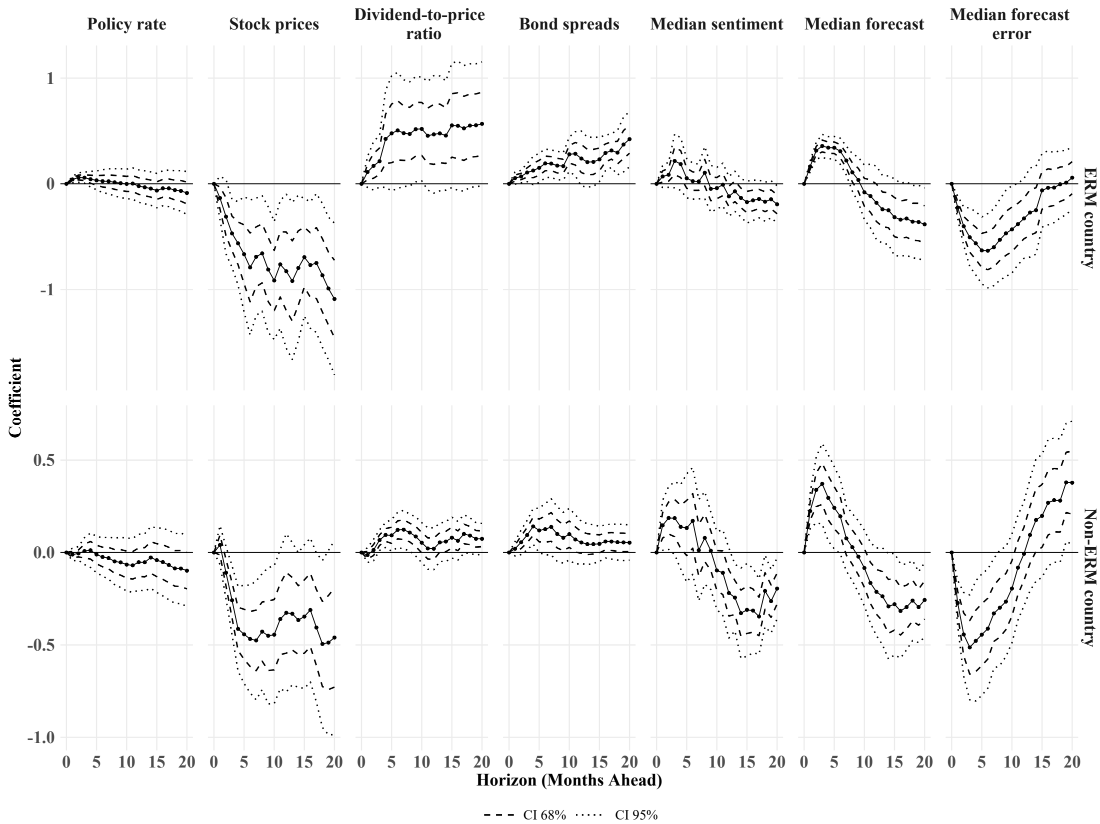


Figure 6: Local projection regression responses to a shock in belief update sentiment.

Note: This figure presents local projection impulse responses of policy rates, stock returns, dividend-to-price ratios, bond spreads, median sentiment, median forecasts and median forecast error following innovations in the median (belief update) sentiment separately for ERM and non-ERM countries. The impulse responses are based on estimation of Eq. 10 with six lags of the sentiment measure, six lags of the response variable, six lags of policy rates and stock prices as controls. Country fixed effects are included and the responses represent the change in credit to GDP from period 0 to period h. Dashed (dotted) lines represent 68% (95%) confidence intervals computed using Driscoll and Kraay (1998) standard errors with a lag length of six. The data ranges across specifications from 8 (non-ERM) to 11 (ERM) countries, 260 to 280 months and from 1438 to 1739 monthly observations.

The persistence of forecast optimism and forecast errors are in line with the slow mean reversion of extrapolative beliefs that Thesmar and Verner (2025) emphasize, whereas the swift response of prices illustrates the repricing channel through which expected returns dominate short-run asset price movements. The absence of monetary policy reactions further suggests that these shocks are perceived as temporary or too short-lived to elicit a policy response, reinforcing the interpretation that the dynamics we observe originate in the interaction of heterogeneous beliefs and discount-rate movements rather than in central bank behavior.

3. Extending the sentiment measure historically

Our novel sentiment measure relies on the existence of survey data on individual beliefs, which limits the time coverage only back to the mid 1990s and the country coverage to around 18 countries. To test credibly the possible relationship of sentiment and its concentration to credit growth and financial stability, we need the sentiment measure for a much longer time span so that it covers multiple credit cycles and financial crises for a sufficiently large country sample. To mitigate this shortcoming, we utilize historical newsarticle titles from the *Wall Street Journal* together with machine learning and natural language processing techniques to estimate a news text approximation of our survey derived sentiment measure all the way to the beginning of the 20th century.

It is a realistic assumption that news and individual sentiment (subjective belief mistake) have a relationship as people collect the majority of their information about the economy and related issues from this information source. Hence, if one has a sufficient amount of data on news and beliefs for the same time and country coverage, one can estimate the relationship between them and utilize the longer time coverage of newspaper data to predict sentiment out-of-sample for periods that we do not have survey data available. We are basically *backcasting* sentiment in time as the out-of-sample predictions are not for future time periods after the estimation sample, but rather for time periods prior to the estimation sample. Manela and Moreira (2017) performed a similar exercise to approximate the VIX index with text data further back in time for periods that the original VIX index was not available for. The authors used a support vector regression to predict the VIX with monthly frequencies of n-grams in the WSJ.

3.1. Data and representation learning

We use historical news text to approximate the informational environment faced by forecasters in earlier decades. This subsection introduces the text corpus, describes how article titles are embedded using transformer-based models, and explains the dimensionality reduction that produces structured representations of news content at the monthly and country level.

3.1.1. Transforming news information into numerical format

News data and text data in general may hold a lot more information than usual macroeconomic and financial variables. Many studies in economics and finance have already shown great gains in utilizing text data in empirical studies of new and old problems in the field. A crucial part in the empirical use of text data is the way that it is transformed to a numerical format that is often further on used as an input in regression and machine learning models. A common way of using a bag-of-words representation of word frequencies in documents can often result into a significant loss of information that might be crucial for the secondary analysis. This is because information about the word order, context and semantic relationships are lossed when a corpus of documents is turned into a so called data feature matrix (DFM) that holds the counts of all unique words (columns) found in the corpus for each document (rows) in the corpus. This means that terms like bank, banker and banking crisis are treated as completely separate words and that the word *market* is treated as the same in sentences like *The stock market closed at an all-time high today*, *There is a growing market for electric vehicles* and *We went to the market to buy groceries*.

To extract as much information as possible from the news texts, we transform each news title into word embeddings, where each word is represented as a numerical vector in a continuous k-dimensional space. These vectors are estimated by training on large text corpora, optimizing an objective function that captures word co-occurrence patterns, such as predicting a word given its surrounding context (as in Word2Vec by Mikolov et al. (2013)) or factorizing word co-occurrence matrices (as in GloVe by Pennington et al. (2014)). The resulting embeddings encode semantic similarity, positioning words with similar meanings closer together in the vector space. Traditional word embedding models, however, produce static representations, meaning that the vector for a given word remains the same regardless of its surrounding context.

We utilize the BERT (Bidirectional Encoder Representations from Transformers) model originally developed by Devlin et al. (2019) that takes into account the context of a word in a sentence or a document when forming the vector representation of words. Unlike simpler models, BERT leverages a transformer architecture and self-attention mechanisms to understand both the context of words within a sentence and the broader relationships across the text. This property makes it particularly well-suited for extracting meaningful information from text, even when dealing with fragmented or headline-style content, as is the case with WSJ titles. BERT embeddings provide a dense and context-aware representation of each news title, similar to the underlying mechanisms of more advanced systems like ChatGPT, which also build upon transformer-based architectures. For efficiency reasons, we do not use the original base-BERT model, but instead a lighter and faster pre-trained version called DistilBERT (Sanh et al., 2019). This model has close to the same performance (97%) as the original, but is around 60% faster in processing a given input sequence of text data.

3.1.2. BERT embeddings and aggregation to monthly news information

Let S be the total number of news titles in the corpus, N_s be the number of tokens in a news title s and d be the embedding dimension (e.g., 768 for BERT). The output of the DistilBERT model is a BERT embedding matrix $\mathbf{X}^{(s)} \in \mathbb{R}^{N_s \times d}$ for title s, where each row represents a token embedding. Hence the DistilBERT embedding for the i-th word in news title s can be represented as $\mathbf{x}_i^{(s)} \in \mathbb{R}^d$. For later analysis, we need to match the newstitle information with the monthly survey derived sentiment data, hence we use mean-pooling to obtain a fixed-length vector representation for a single news title s. More formally,

$$\mathbf{h}^{(s)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{x}_i^{(s)}, \tag{11}$$

where $\mathbf{h}^{(s)} \in \mathbb{R}^d$ is the final mean-pooled embedding for title s. Once we obtain a mean-pooled embedding for each news title, we can represent the entire corpus as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}^{(1)} \\ \mathbf{h}^{(2)} \\ \vdots \\ \mathbf{h}^{(S)} \end{bmatrix} \in \mathbb{R}^{S \times d}, \tag{12}$$

where each row $\mathbf{h}^{(s)}$ represents the mean-pooled embedding of the s-th news title. Our final corpus includes 1.066.317 distinct news titles[3] and these documents are represented in 768 dimensions. Given the high dimensionality of the BERT embeddings, we apply principal component analysis (PCA) to reduce these embeddings to the fifty most significant principal components (PCs), ensuring computational efficiency while retaining the most relevant features. The PCA transformation can be expressed as:

$$\mathbf{Z} = \mathbf{XW},$$

where $\mathbf{W} \in \mathbb{R}^{d \times k}$ is the matrix of the top $k = 50$ principal components, and $\mathbf{Z} \in \mathbb{R}^{S \times k}$ is the resulting lower-dimensional representation.

To predict sentiment at a monthly level, we aggregate the news title embeddings into monthly representations. Specifically, we compute a *general news embedding* that represents the overall monthly news coverage and a *country-specific news embedding* that represents the subset of news titles mentioning a particular country. Let T be the total number of months in the dataset, S_t be the number of news titles in month t, S_t^c be the number of news titles in month t that mention a specific country c, $\mathbf{H}^{(t)} \in \mathbb{R}^{S_t \times k}$ be the matrix of PCA-reduced embeddings for all news titles in month t and $\mathbf{H}^{(t,c)} \in \mathbb{R}^{S_t^c \times k}$ be the matrix of PCA-reduced embeddings for news

[3]As BERT takes into account the context of each word, it is not necessary to perform text cleaning procedures like removing stopwords etc. To get enough context and exclude short news titles that might not hold that much information, we include only the titles with at least 10 words.

titles in month t mentioning country c. To obtain fixed-length monthly embeddings, we compute the mean-pooling over all news titles for each category:

$$\mathbf{h}^{(t)} = \frac{1}{S_t} \sum_{s=1}^{S_t} \mathbf{z}_s^{(t)}, \tag{13}$$

$$\mathbf{h}^{(t,c)} = \frac{1}{S_t^c} \sum_{s=1}^{S_t^c} \mathbf{z}_s^{(t,c)}, \tag{14}$$

where $\mathbf{z}_s^{(t)} \in \mathbb{R}^k$ is the PCA-reduced embedding of news title s in month t. $\mathbf{z}_s^{(t,c)} \in \mathbb{R}^k$ is the PCA-reduced embedding of news title s in month t mentioning country c. $\mathbf{h}^{(t)} \in \mathbb{R}^k$ represents the *general monthly news embedding* and $\mathbf{h}^{(t,c)} \in \mathbb{R}^k$ is the *country-specific monthly news embedding*.

3.2. Predicting sentiment from news text

Using the text representations introduced above, we train supervised machine-learning models to map news content to the survey-based sentiment and dispersion measures developed in Section 2.3. This approach allows us to infer sentiment historically by applying the trained models to periods and countries without survey data, thereby extending the belief-based framework over a much longer horizon.

3.2.1. Random forest prediction model

Our objective is to predict monthly sentiment for each country based on aggregated news embeddings, incorporating both *contemporaneous and lagged embeddings* as predictors. The sentiment measure is available at the country-month level, denoted as $S^{(t,c)}$. We consider two sentiment variables that we want to backcast in time: $B_{\text{med}}^{(t,c)}$, representing the median sentiment score for country c in month t and $B_{\text{conc}}^{(t,c)}$, measuring how concentrated sentiment is around specific values, capturing polarization in sentiment distribution. Both of these measure are related to the level of beliefs rather than to belief updates. We chose this as the sentiment derived from the level of beliefs is likely to hold more information important for credit growth and crisis than its counter part derived from belief updates (e.g. changes).

For each country c and month t, our goal is to predict both $B_{\text{med}}^{(t,c)}$ and $B_{\text{conc}}^{(t,c)}$ with the current embeddings, one month lags of the embeddings, lagged 3-month and 6-month moving averages of the embeddings. Thus, the complete feature vector for country c in month t is:

$$\mathbf{x}^{(t,c)} = \left[\mathbf{h}^{(t)}, \mathbf{h}^{(t,c)}, \mathbf{h}^{(t-1)}, \mathbf{h}^{(t-1,c)}, \bar{\mathbf{h}}^{(t,3)}, \bar{\mathbf{h}}^{(t,6)}, \bar{\mathbf{h}}^{(t,c,3)}, \bar{\mathbf{h}}^{(t,c,6)}\right] \in \mathbb{R}^{8k}, \tag{15}$$

where $\mathbf{h}^{(t)} \in \mathbb{R}^k$: general news embedding for month t, $\mathbf{h}^{(t,c)} \in \mathbb{R}^k$: country-specific news embedding for month t, country c, $\mathbf{h}^{(t-1)}$, $\mathbf{h}^{(t-1,c)}$: one-month lagged embeddings, $\bar{\mathbf{h}}^{(t-1,3)} = \frac{1}{3}\sum_{\tau=t-3}^{t-1} \mathbf{h}^{(\tau)}$, the 3-month moving average of general news embeddings (lagged by one month), $\bar{\mathbf{h}}^{(t-1,6)} = \frac{1}{6}\sum_{\tau=t-6}^{t-1} \mathbf{h}^{(\tau)}$, the 6-month moving average of general news embeddings (lagged by one month), $\bar{\mathbf{h}}^{(t-1,c,3)} = \frac{1}{3}\sum_{\tau=t-3}^{t-1} \mathbf{h}^{(\tau,c)}$, the 3-month moving average of country-specific news embeddings (lagged by one month), $\bar{\mathbf{h}}^{(t-1,c,6)} = \frac{1}{6}\sum_{\tau=t-6}^{t-1} \mathbf{h}^{(\tau,c)}$ the 6-month moving average of country-specific news embeddings (lagged by one month). Hence we utilize only the past and current news to predict current sentiment.

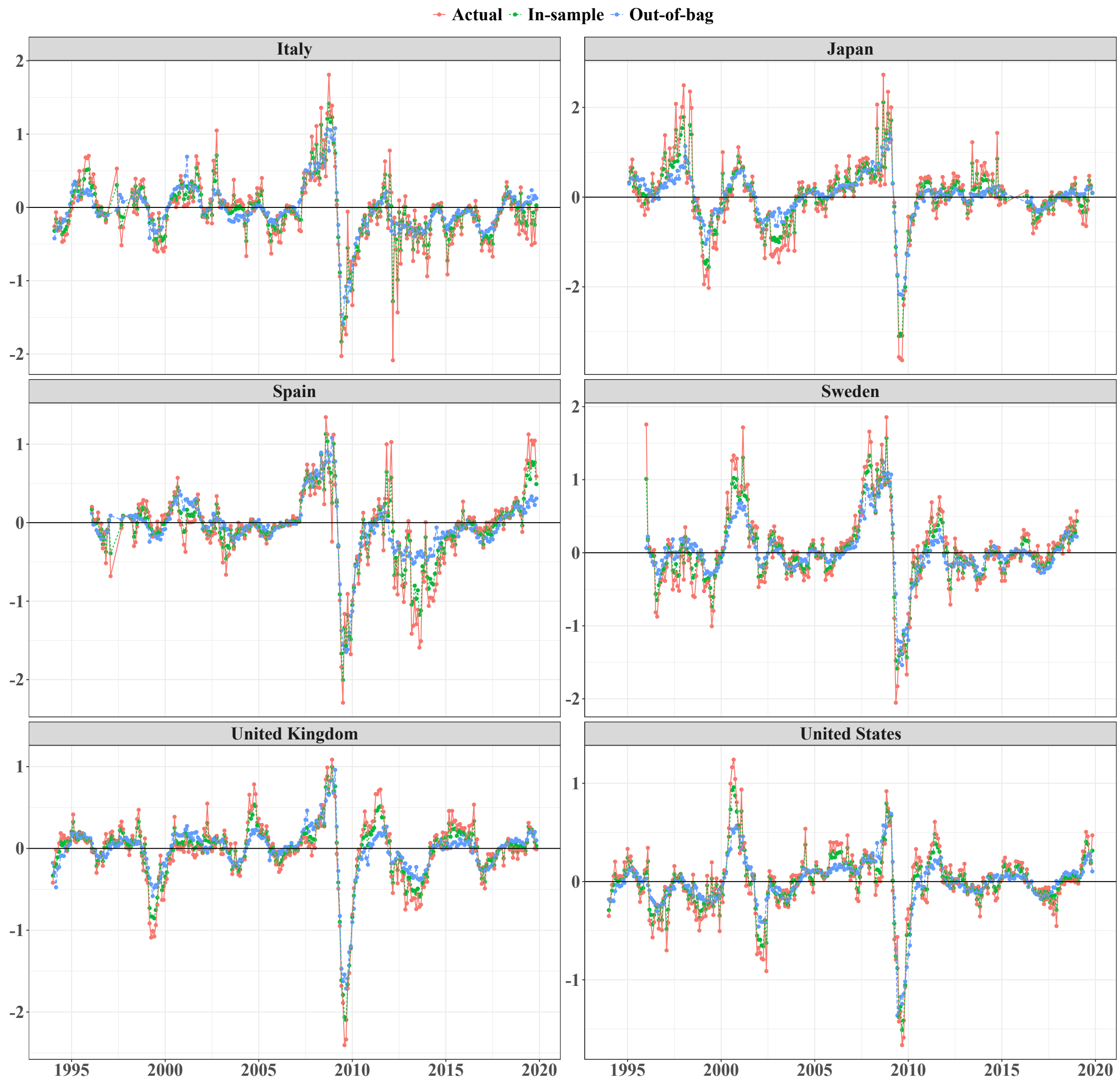


Figure 7: Random forest predictions of survey derived median sentiment with newsarticle BERT embedding principal components.

Note: The figure OOB R^2 55.9% median sentiment and 37.3% sentiment dispersion.

A random forest (Breiman, 2001) is an ensemble learning method that constructs multiple *decision trees* and aggregates their predictions to improve accuracy and reduce overfitting. Since the sentiment variables in this study are continuous, the model is used for regression rather than classification. In a regression random forest, each tree outputs a numerical prediction, and the final prediction is obtained by averaging the outputs of all trees. The model is estimated using bootstrap aggregation (bagging), where each tree is trained on a random subset of the training data drawn with replacement[4]. Additionally, at each split within a tree, only a random subset of features is considered. This reduces correlation among trees and improves generalization. The model is evaluated using the *out-of-bag* (OOB) mean squared error (MSE). Since each tree is trained on a bootstrapped subset of the data, about one-third of the observations are left out (*OOB samples*) for each tree. These OOB samples serve as an internal validation set, allowing for performance estimation without requiring a separate validation set. The OOB MSE is computed as:

$$\text{OOB MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_{\text{OOB},i} \right)^2 \tag{16}$$

where y_i is the actual sentiment value, and $\hat{y}_{\text{OOB},i}$ is the predicted sentiment for observation i, obtained by averaging predictions from only the trees that did not include i in their bootstrap sample.

To improve model performance, hyperparameter tuning is performed by adjusting the minimum number of observations required in a terminal node and the number of features randomly selected at each split. Smaller minimum number for observations required in a terminal node allow deeper trees and capture finer details in the data, but may lead to overfitting. A smaller number of features in each tree increases diversity among trees, while a larger number can reduce variance, but increase correlation among trees. The final model is selected based on the combination of the minimum node size and the number of features considered at each split that results in the lowest out-of-bag mean squared error, ensuring an optimal balance between bias and variance. For each sentiment measure $S^{(t,c)}$, we train separate random forest models:

$$\hat{B}^{(t,c)}_{\text{news,med}} = f_{\text{RF, med}}(\mathbf{x}^{(t,c)}), \tag{17}$$

$$\hat{B}^{(t,c)}_{\text{news,conc}} = f_{\text{RF, conc}}(\mathbf{x}^{(t,c)}), \tag{18}$$

where $f_{\text{RF, med}}$ is the random forest model predicting median sentiment $B^{(t,c)}_{\text{med}}$, $f_{\text{RF, conc}}$ is the random forest model predicting sentiment concentration $B^{(t,c)}_{\text{conc}}$. We define the predictions $\hat{B}^{(t,c)}_{\text{news,med}}$ and $\hat{B}^{(t,c)}_{\text{news,conc}}$ as the news implied median sentiment and concentration of sentiment for country c in month t. The model is trained with initial 18 countries[5] for which we have sentiment observations derived from *Consensus Economics* surveys for the time interval between January 1994 and November 2019 with a total number of 4155 country-month observations.

[4] The number of trees is fixed at 500.

[5] Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

3.2.2. Model performance and backcasting of sentiment

Figure 7 visualizes the performance of the random forest model in predicting the median sentiment $B_{\text{med}}^{(t,c)}$ for a selected subset of six countries. The OOB predictions (blue) show that the model is able to generalize the predictive relationship between news information and median sentiment impressively well. The OOB R^2 values is 56% for median sentiment and 37% for sentiment dispersion, implying that a large share of the variation in the level of median sentiment and its dispersion among forecasters can be explained with only newsarticle title information. The close alignment of the true values of sentiment, the in-sample and OOB-sample prediction indicates the random forest has learned a generalizable relationship and has not overfitted to the training data.

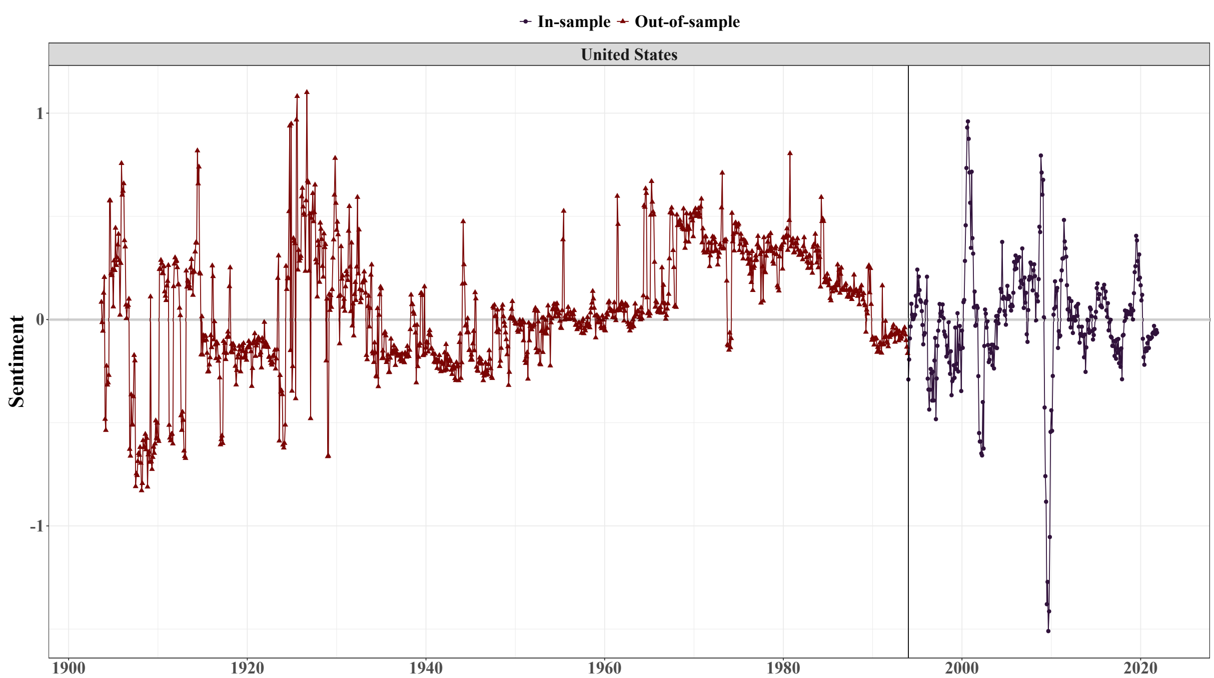


Figure 8: Backcasted sentiment $\hat{B}_{\text{news,med}}$.

Next, we backcast the data back in time for the 18 countries all the way to September 1903 and we also extend the cross-sectional dimension of the survey dataset by estimating the news-implied sentiment for additional 60 countries bringing the total number of country-month observations to 78.387. An example time-series of the news-implied median sentiment is visualized in Figure 8 for the United States. The cross-sectional extension is made possible by using the general word embeddings and also the country specific ones that mention these additional 60 countries. The final country set[6] was chosen to cover the countries for which Müller and Verner (2024) have collected detailed corporate sector credit data.

4. Empirical evidence on Minsky cycles

4.1. Sentiment and bad credit booms

The consensus in the economic literature on financial crises has increasingly converged on the *credit boom gone bust* explanation. This interpretation was originally put forward by Minsky (1977) and Kindleberger (1978), and later brought back into the mainstream discussion in the aftermath of the global financial crisis by Reinhart and Rogoff (2009), who presented empirical evidence based on an unprecedentedly large historical dataset on crises. This re-emergence of financial crisis research also attracted attention from the general public. Since then, the empirical literature has expanded substantially, increasingly relying on newly compiled long-run historical datasets rather than the shorter panel datasets commonly used in the so-called early warning system literature. Earlier studies in that tradition often emphasized prediction accuracy and the use of increasingly sophisticated forecasting models. In contrast, more recent research utilizing long-run historical data has tended to focus more on substantive interpretation, emphasizing the underlying causes of crises while employing simpler statistical modeling frameworks.

[6] Albania, Argentina, Armenia, Australia, Austria, Belgium, Botswana, Bulgaria, Canada, Chile, Czech Republic, Denmark, Dominican Republic, Estonia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Haiti, Honduras, Hong Kong, Hungary, India, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kyrgyz Republic, Latvia, Lithuania, Macedonia, Malawi, Malaysia, Mauritius, Mexico, Mongolia, Morocco, Nepal, Netherlands, New Zealand, Nigeria, Norway, Oman, Pakistan, Panama, Peru, Philippines, Portugal, Russia, Saudi Arabia, Sierra Leone, Singapore, Slovak Republic, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Tanzania, Thailand, Trinidad & Tobago, Tunisia, Turkey, Uganda, Ukraine, United Arab Emirates, United Kingdom, United States, and Venezuela.

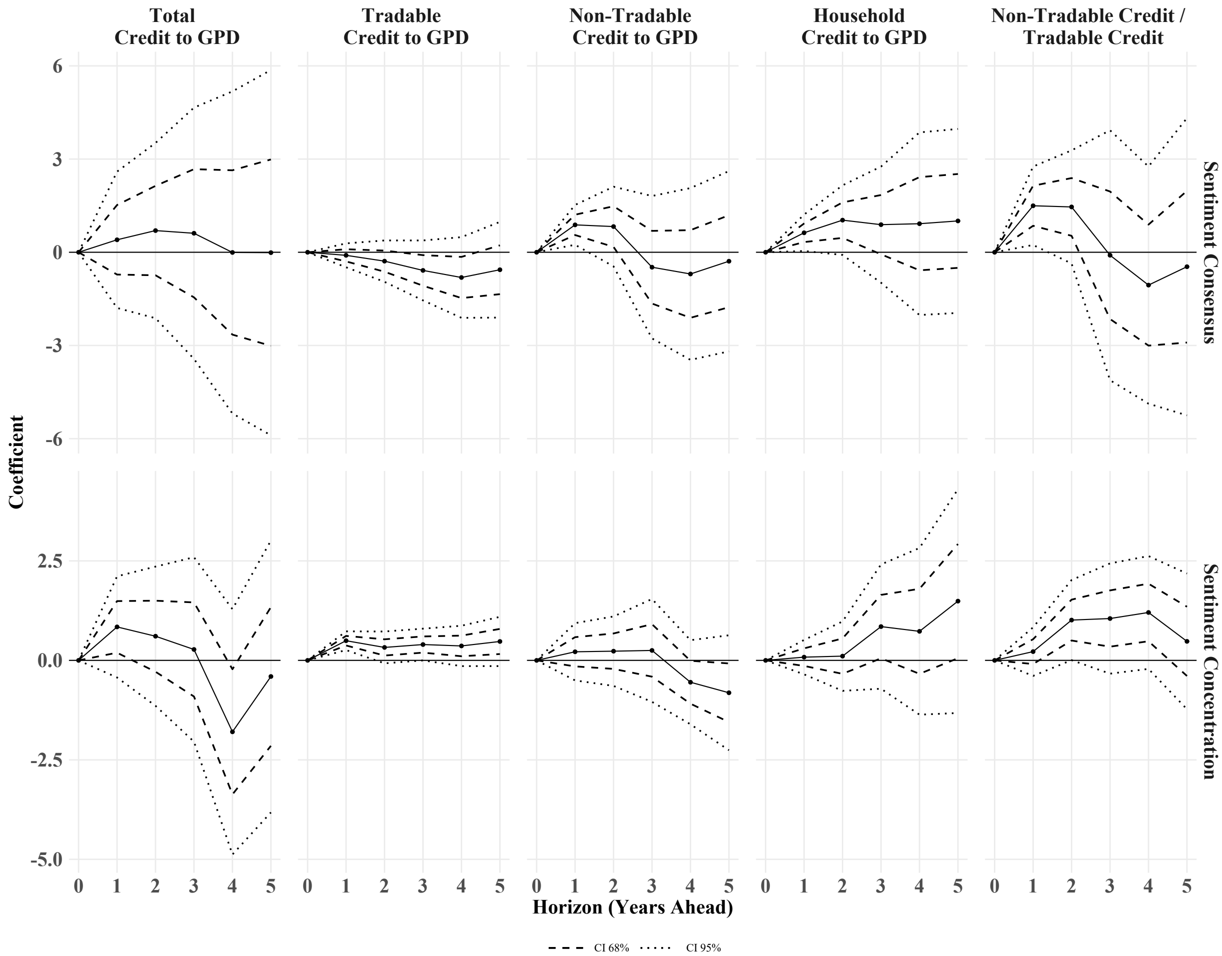


Figure 9: Local Projections: Sentiment shocks on credit growth.

Note: This figure presents local projection impulse responses of credit (total credit, tradable sector credit, non-tradable sector credit, household credit) to GDP following innovations in the median sentiment (consensus) and sentiment concentration (negative of belief dispersion). The impulse responses are based on estimation of equation 19 with two lags of the sentiment measure and one lag of the credit variable as controls. Country and year fixed effects are included and the responses represent the change in credit to GDP from period 0 to period h. Dashed (dotted) lines represent 68% (95%) confidence intervals computed using Driscoll and Kraay (1998) standard errors with a lag length of two. The data ranges across specifications from 74 to 78 countries, 62 to 103 years and from 1704 to 4698 annual observations.

A substantial number of studies have linked credit expansions to subsequent financial crises. Schularick and Taylor (2012) provided the first empirical evidence from historical data covering a long time span of 140 years and 14 developed countries—rather than only emerging economies—that financial crises are strongly predicted by credit growth in preceding years. Relatedly, Mian et al. (2017) presented evidence from 30 countries between 1960 and 2012 showing that household debt-to-GDP ratios predict lower subsequent GDP growth and higher unemployment. Baron and Xiong (2017) provided evidence that expansions in bank credit predict bank equity crashes, and Greenwood et al. (2022) emphasized the role of simultaneous growth in credit and asset prices. Furthermore, Krishnamurthy and Muir (2025) showed that the interaction of high credit growth and low interest rates is important in predicting crises in a historical setting.

The most recent evidence by Müller and Verner (2024) shows that not all credit growth is necessarily bad in the sense of predicting future financial crises. Using a newly collected historical dataset that separates credit by corporate sector, they demonstrate that crises are predicted specifically by credit booms in non-tradable sectors such as real estate and construction. Hence, the current consensus from a vast literature utilizing long historical country panels is that credit growth in corporate sectors where demand and supply are primarily driven by domestic factors is the most important and robust predictor of financial crises. One important question remains: what predicts these bad credit booms? In the spirit of Minsky (1977) and Kindleberger (1978), we seek to empirically test whether harmful credit expansions arise following surges in aggregate sentiment about the economy, which we define as a misaggregation of public information relative to the optimal machine aggregation of that same information. In this way, we can examine whether a sudden collective misinterpretation of public information among investors initiates the Minsky Cycle—that is, a bad credit boom ending in a bust and a crisis.

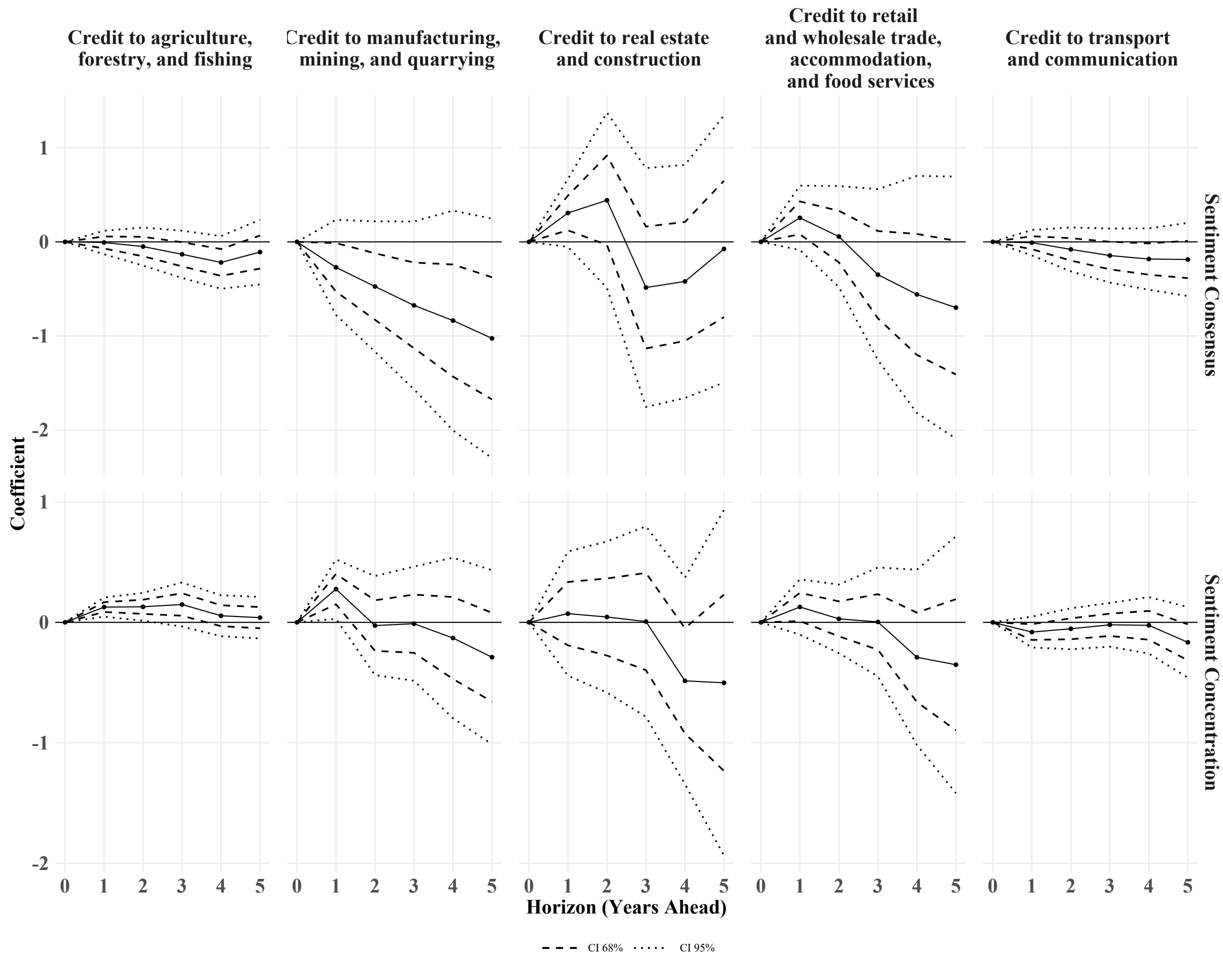


Figure 10: Local Projections: Sentiment shocks on credit growth of different sectors.

Note: This figure presents local projection impulse responses of credit (for different sectors) to GDP following innovations in the median sentiment (consensus) and sentiment concentration (negative of belief dispersion). The impulse responses are based on estimation of equation 19 with two lags of the sentiment measure and one lag of the credit variable as controls. Country and year fixed effects are included and the responses represent the change in credit to GDP from period 0 to period h. Dashed (dotted) lines represent 68% (95%) confidence intervals computed using Driscoll and Kraay (1998) standard errors with a lag length of two. The data ranges across specifications from 74 to 77 countries, 66 to 77 years and from 1873 to 2743 annual observations.

To analyze the relationship between sentiment and credit growth, we estimate panel local projection regressions (Jorda, 2005) with a dataset constructed from the vast annual historical credit dataset by Müller and Verner (2024) merged with our backcasted measure of median sentiment and sentiment concentration. We estimate a panel local projection regression were the response variable $\Delta Y_{c,t+h}$ is the $h = 0, \ldots, 5$ year ahead change of a measure of credit-to-GDP and the shock variable $\hat{B}^{\text{news}}_{c,t}$ is a backcasted measure of sentiment for country c at year t. More formally,

$$\Delta Y_{c,t+h} = Y_{c,t+h} - Y_{c,t} = \alpha_c^h + \beta^h \hat{B}_{c,t}^{\text{news}} + \gamma^h \text{controls}_{c,t-1} + \epsilon_{i,t+h}, \tag{19}$$

where $h = 0, 1, 2, ..., 5$, α_i^h are the country fixed effects, and λ_t are the fixed year effects controlling for common global shocks. We include two lags of the sentiment measure and one lag of the credit variable as controls. We report both 68% and 95% confidence intervals calculated according to Driscoll and Kraay (1998) standard errors with a lag length of six months.

Figure 9 visualizes the response of different credit-to-GDP variables to a shock in either the median sentiment or the concentration of sentiment. Interestingly, total credit-to-GDP does not change after a positive median sentiment shock. However, the allocation of credit changes, as credit to households and credit to non-tradable sectors increase significantly. If anything, credit to the tradable sector declines. The last column of Figure 9 shows how the ratio of non-tradable to tradable sector credit increases during the first two years after a sentiment shock, after which it declines. These results are very much in line with Müller and Verner (2024), who show that the reallocation of credit between sectors predicts subsequent economic slowdowns and financial crises.

The local projection results displayed in Figure 10 utilize more detailed data on credit to different sectors and reveal that during the first two years after a positive sentiment shock, credit increases mainly in the real estate and construction sectors. The opposite occurs in manufacturing, mining and quarrying, and in transport and communication, as a sentiment shock is followed by a longer-term decline in credit to these tradable sectors. This highlights how mistaken growth expectations disproportionately benefit industries tied to domestic demand in the short-term.

Figure 11 presents the local projection regressions from Figure 9 separately for periods with high and low concentration of sentiment. It appears that for credit booms, disagreement in beliefs does not play a major role, as positive sentiment shocks precede bad credit booms only when agreement (concentration) about sentiment is high. This result emphasizes that these harmful surges in sentiment are collective in nature, in the spirit of Keynes (1936)'s animal spirits.

These findings historically validate the Kindleberger-Minsky view on the endogenous nature of financial crises. By quantifying sentiment as systemic misinterpretation of public information, our results provide a behavioral mechanism that bridges the qualitative insights of Minsky (1977) and Kindleberger (1978) and historical empirical evidence. Specifically, our results highlight how belief mistakes from people's varying mental models used to interpret same information differently amplify credit misallocation to vulnerable sectors, such as non-tradables and real estate, reinforcing financial fragility.

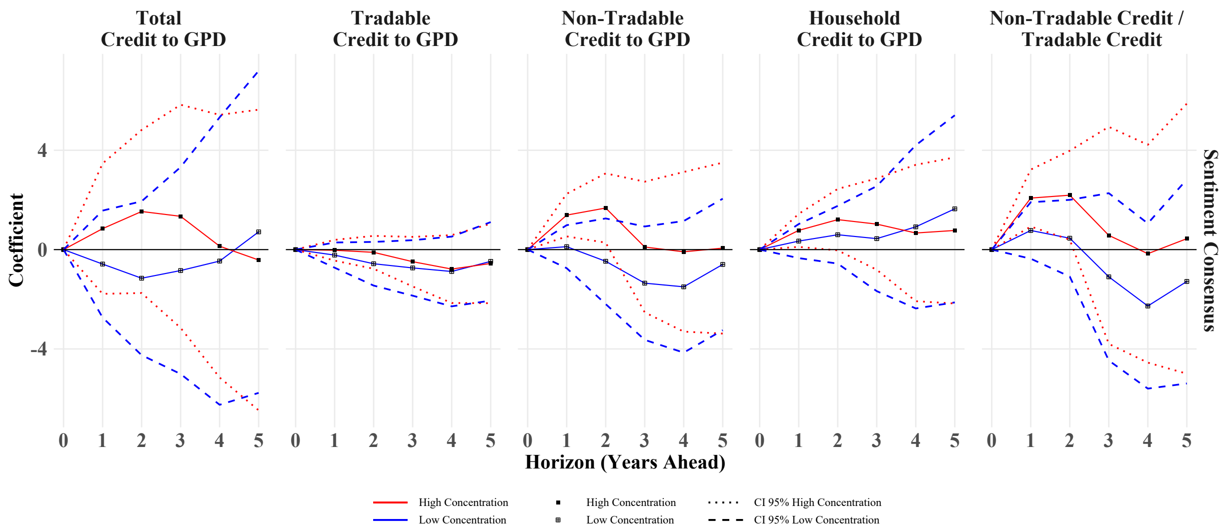


Figure 11: Local Projections: Sentiment shocks on credit growth separately for low and high sentiment concentration periods.

Note: This figure presents local projection impulse responses of credit (for different sectors) to GDP following innovations in the median sentiment (consensus) separately for high and low sentiment concentration (negative of belief dispersion) periods. Low (high) sentiment concentration periods are the ones with sentiment concentration below (above) the lowest 15 percentile. The impulse responses are based on estimation of equation 19 with two lags of the sentiment measure and one lag of the credit variable as controls. Country and year fixed effects are included and the responses represent the change in credit to GDP from period 0 to period h. Dashed and dotted lines represent 95% confidence intervals computed using Driscoll and Kraay (1998) standard errors with a lag length of two. The data ranges across specifications from 74 to 77 countries, 66 to 77 years and from 1873 to 2743 annual observations.

4.2. Memory, experience, and the formation of sentiment

The results of the previous section empirically validate that bad credit booms and financial crisis are predicted with increases in aggregate sentiment about future economic growth. Furthermore, this sentiment is shown to be misinterpretation of public information by economic agents relative to an objective optimal machine belief given the same information set. This still does not explain where this increase in misinterpretation comes from. Neither Minsky (1977, 1986) or Kindleberger (1978) provide any detailed explanation or process why sentiment starts to increase. Rather they refer to a *displacement* from the common path of expectations due to positive news shocks as the reason why sentiment starts to rise.

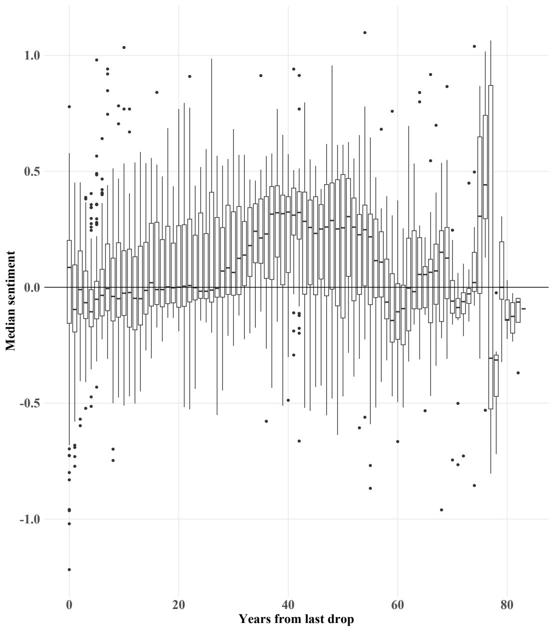
A number of papers have shown how past experiences can affect current economic behavior. For example, Malmendier and Nagel (2011) show how people who have experienced low stock returns earlier in their lives, are less likely to invest into stock markets and are more pessimist about about future returns. Similarly in Malmendier and Nagel (2016), the authors show that current inflation expectations are affected by a person's inflation experiences during her lifetime. The vast evidence of the role of experiences in shaping economic behavior is summarized in Malmendier and Wachter (2024).

In financial crisis context, a famous simple explanation about the underlying behavioral reasons for crisis is the *This time is different - syndrome*. This phrase was made famous by Reinhart and Rogoff (2009) in their book with the same title, where the authors argued that humanity has not yet *graduated* from financial crisis as these events seem to occur time and a time again in history with similar preceding patterns in current thinking by economic agents. This syndrome as a situation where people forget the teachings of the past crisis and justify overly optimistic beliefs about low risks related to financial stability by claiming that the current status is somehow significantly different from the pre-crisis times of the past. In summary, experiences matter for belief formation and anecdotal evidence suggests that people tend to underweight rare past episodes when assessing the likelihood of future events.

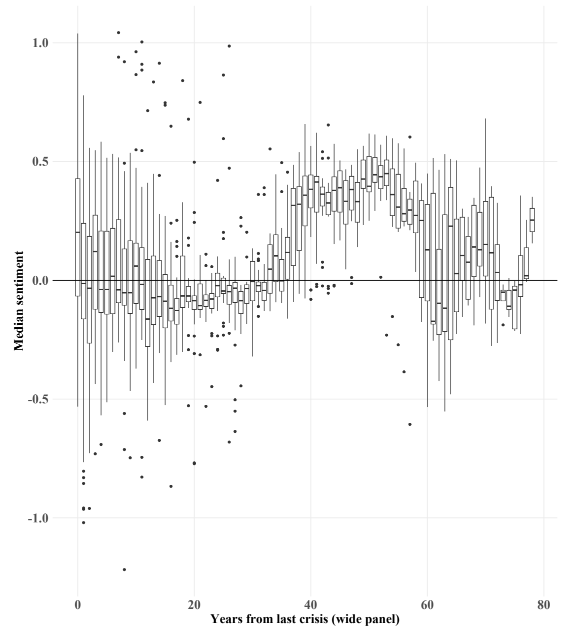
Recent literature on memory-recall and similarity on belief formation provides a theoretical explanation for overoptimism that is rooted in psychology and memory research. The theoretical model by Bordalo et al. (2025a) describes how people form their beliefs on events that they either do not have much experience of or that they have not experienced at all. The model predicts that when people form their beliefs about an event, they simulate the event from their experiences and different memories fight for retrieval. They argue that more similar events are recalled with higher likelihood, which is why non-domain specific–that is non-relevant experiences for assessing the likelihood of a future event– can bias beliefs due to similarity. The similarity of an experience increases the likelihood of recalling it (and hence the assessment of the likelihood of the future event being assessed), but this will also bring interference to the memory sampling process as it reduces the likelihood of recalling other experiences that might be relevant for assessing the likelihood of a future event.

Bordalo et al. (2025a) confirm the predictions of the theoretical model with two survey experiments on COVID and cyber-attacks, where the latter utilized random priming of experiences to be recalled. In addition, Graeber et al. (2024) in a similar manner extend this theoretical framework to study how differently experiences or information that is of narrative/story-like effect on belief formation relative to statistical information. They test the theoretical predictions with survey experiment evidence and show that stories affect belief formation more persistently than statistical information.

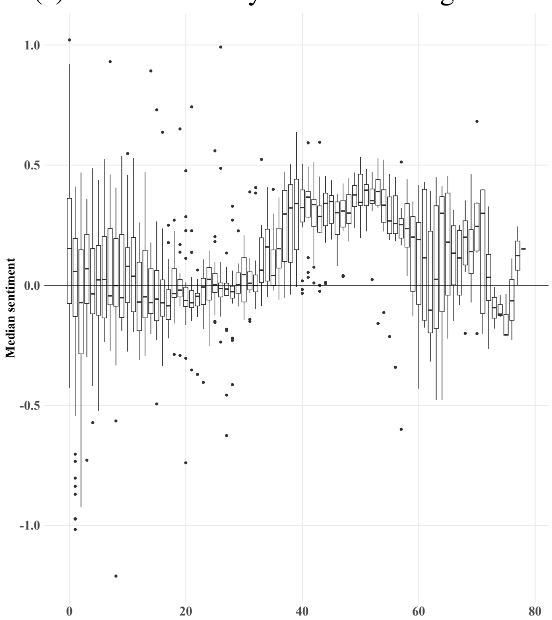
When assessing the likelihood of financial crisis in the future, people sample their memory database of experiences similar to financial crisis that they will use to simulate the likelihood



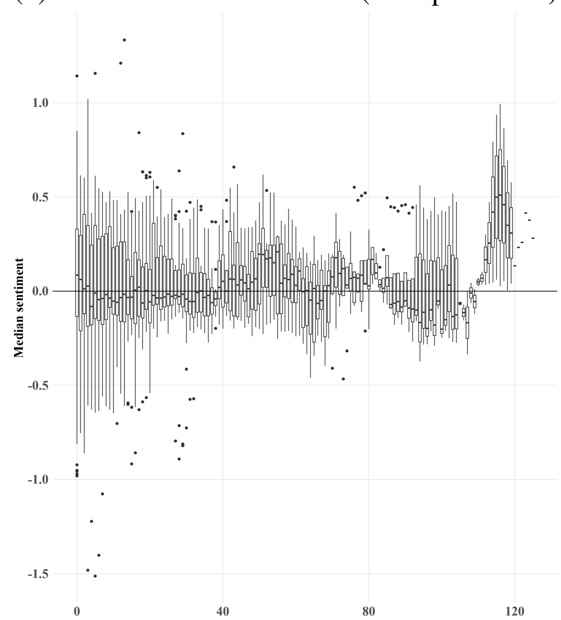
(a) Years since very low economic growth



(b) Years since financial crisis (wide panel data)



(c) Years since financial crisis (long panel data)



(d) Months since last recession (monthly USA)

Figure 12: Sentiment and time since negative economic events.

of this event. Experience relates to this framework directly as if a person has experienced an event it is likely to be in their memory database compared to a situation without lived experience. Financial crisis is conceptually a very abnormal event to which it is hard to find domain (large finance related negative events) or non-domain (crisis type events not related to the macroeconomy e.g. crisis in family or crisis in you personal financial situation) specific experiences in vast numbers.

Given the memory recall process by Bordalo et al. (2025a) we have two testable predictions for forming beliefs on the likelihood of crisis. First, a person who has experienced a financial crisis earlier in their life is more likely to recall that event and use that for simulation when assessing the probability of a similar event in the future relative to a person who has not that experience in their memory database. Secondly, when time goes by from that past rare event, the memory database has been expanded with a large number of new experiences so the relative portion of crisis experiences deteriorates in time implying a smaller change of recalling and simulating their belief in crisis with that experience. As some part of sentiment is seen to relate to the neglecting of tail-risk, experiences and memory on financial crisis should have an effect on sentiment.

Figure 12 visualizes sentiment and the time that has passed since a negative economic event with different definitions of a negative event. It can be seen from figures 12a-12c that when approximately 30 years has passed since a negative event, the median sentiment starts to rise across countries in the sample. The years passed since a crisis seems to have a clearer and stronger effect on sentiment relative to very low economic growth periods. The closer similarity of a financial crisis as an experience to the picture that one has of a tail-event in their mind might be the explanation for this difference. This is supported by Figure 12d, which shows that time since recessions does not have an increasing relationship with median sentiment. However, this visual analysis does not take into account omitted variables like GDP growth.

Next, we will empirically test whether experiences and memory could play a role in the rise of sentiment and whether we can provide descriptive evidence from history that is in line with the recent experiment survey evidence of Bordalo et al. (2025a) and Graeber et al. (2024). We will do so by estimating a panel regression where the median sentiment and the concentration of sentiment are predicted with the time in years that has passed since a very low economic growth period. We define this as a year with GDP growth among the lowest 5 percentile in our sample. This predictor will capture how more distant experiences are more difficult to recall when forming beliefs about similar events and furthermore taking account the possibility of negative events when forming expectations about future economic growth. As the relationship can be nonlinear (e.g. time at first does not count, but when enough time has passed the effect on memory could be strong), we will include the squared term of years since a major

Table 2: Time since bad growth, share of young people, median sentiment and concentration of sentiment distribution.

	Dependent variable:					
	Median sentiment			Sentiment concentration		
	(1)	(2)	(3)	(4)	(5)	(6)
Years since drop_{t-1}	0.004^{**}		0.003^{*}	0.001		0.0005
	(0.002)		(0.002)	(0.001)		(0.001)
Years since drop^2_{t-1}	-0.0001^{**}		-0.0001^{**}	-0.00001		-0.00001
	(0.00002)		(0.00002)	(0.00002)		(0.00002)
Very $\text{young/old}_{5MA,t-1}$		0.016^{**}	0.014^{**}		0.006	0.001
		(0.007)	(0.006)		(0.006)	(0.004)
Very $\text{young/old}^2_{5MA,t-1}$		-0.0003^{**}	-0.0003^{**}		-0.0001	-0.0001
		(0.0002)	(0.0001)		(0.0001)	(0.0001)
Median sentiment_{t-1}	0.225	0.222^{*}	0.216	-0.271^{**}	-0.261^{*}	-0.274^{**}
	(0.139)	(0.125)	(0.137)	(0.134)	(0.133)	(0.133)
Median $\text{sentiment}_{5MA,t-1}$	0.460^{***}	0.456^{**}	0.432^{***}	0.187	0.168	0.181
	(0.157)	(0.180)	(0.163)	(0.165)	(0.165)	(0.164)
Sentiment $\text{concentration}_{t-1}$	0.342^{*}	0.361^{**}	0.340^{**}	0.563^{***}	0.557^{***}	0.562^{***}
	(0.176)	(0.183)	(0.171)	(0.068)	(0.074)	(0.069)
Sentiment $\text{concentration}_{5MA,t-1}$	-0.033	-0.031	-0.027	0.077	0.066	0.077
	(0.167)	(0.170)	(0.165)	(0.058)	(0.057)	(0.058)
log(Population)	-0.098^{*}	-0.095^{*}	-0.098^{*}	0.010	-0.002	-0.001
	(0.057)	(0.052)	(0.055)	(0.020)	(0.023)	(0.019)
GDP $\text{Growth}_{5MA,t-1}$	-0.004	-0.004	-0.005	-0.004^{*}	-0.003	-0.003^{*}
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	2572	3643	2456	2572	3643	2456
Number of Countries	58	78	58	58	78	58
Year Range	65	60	60	65	60	60
Observations	2,572	3,643	2,456	2,572	3,643	2,456
R^2	0.312	0.283	0.307	0.446	0.452	0.450

Note: *p<0.1; **p<0.05; ***p<0.01

economic drop as a predictor as well. In addition, we include the share of very young to old people in the population as a predictor. Older (younger) people are more (less) likely to have experienced multiple negative events and hence recall them easier (harder). We will control for past movements in sentiment, GDP growth, and the total amount of population. We will also include country fixed effects to control for unobservable country-specific factors.

Table 3: Time since bad growth, share of young people, median sentiment and concentration of sentiment distribution. Long panel historical panel of 18 developed countries from Jordà et al. (2017).

	Dependent variable:					
	Median sentiment			Sentiment concentration		
	(1)	(2)	(3)	(4)	(5)	(6)
Years since crisis$_{t-1}$	0.007**	0.009***	0.009***	0.003	0.006*	0.006*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Years since crisis$^2_{t-1}$	−0.0001*	−0.0001***	−0.0001***	−0.00003	−0.0001*	−0.0001*
	(0.00004)	(0.00003)	(0.00003)	(0.00004)	(0.00004)	(0.00004)
Median sentiment$_{5MA,t-1}$	0.576***	0.505***	0.499***	−0.152**	−0.254***	−0.261***
	(0.155)	(0.191)	(0.188)	(0.072)	(0.086)	(0.090)
Sentiment concentration$_{5MA,t-1}$	0.050	0.018	0.018	0.562***	0.529***	0.529***
	(0.076)	(0.084)	(0.084)	(0.138)	(0.145)	(0.147)
GDP Growth$_{5MA,t-1}$			0.015			0.148
			(0.184)			(0.139)
Total Return on risky assets$_{5MA,t-1}$		0.349	0.336		−0.156	−0.228
		(0.266)	(0.303)		(0.143)	(0.181)
Total Return on Safe assets$_{5MA,t-1}$		−0.105	−0.100		1.166*	1.209*
		(0.501)	(0.503)		(0.656)	(0.670)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
# Countries	18	14	14	18	14	14
# Years	98	98	98	98	98	98
Observations	1,422	1,015	1,004	1,422	1,015	1,004
R^2	0.271	0.287	0.283	0.229	0.236	0.237
Note:						*p<0.1; **p<0.05; ***p<0.01

The empirical results in Table 2 show that even when controlling for past sentiment, its dispersion and GDP growth, the time since a severe economic downturn has a significant increasing effect to the median sentiment in a country. In addition, the share of very young to old people in the population increases sentiment. Interestingly, the concentration of sentiment is not explained by these variables at all. As the number of years in the sample for each country is relevant for the strength of the analysis, we repeat the analysis with the long country sample with the 18 countries included in macro-history database of Jordà et al. (2017). In this analysis, we have 98 years of observations for each country and this enables us to use the time since a financial crisis as a predictor. In addition to past GDP growth, we also control for past returns in the financial market. Table 3 shows more significant and stronger effects to median sentiment from time since a negative economic event than the results in Table 2. Controlling for financial returns and even GDP growth does not weaken the results, but the coefficients stay similar and become more significant. Interestingly, the time since crisis or severe economic downturn does not increase the concentration of sentiment.

To summarize the empirical results, we observe that economic sentiment tends to rise after a significant amount of time has passed since a crisis or severe negative economic event. This delayed rise in sentiment suggests that individuals and markets initially remain rationally cautious and their belief formation process does not deviate significantly from the objective one that a machine would make, but as time goes by and the memory of the rare event fades, optimism gradually builds. Younger individuals, who may have no direct experience with the prior crisis or severe economic events, are more likely to form their beliefs based on a more optimistic view of the future as they cannot sample these things from memory due to the lack of experiences of that event or even experiences similar to that event. Older individuals, conversely, may retain memories of the crisis (and other negative rare events) and therefore remain more cautious. As younger generations take on a larger share of the population over time, their more optimistic outlook may drive an overall increase in sentiment.

5. Conclusions

The origins of financial crises remain one of the central unresolved questions in economics and have continued to attract interest, especially after the global financial crisis of 2008. Nevertheless, significant progress has been made in recent decades, primarily through the use of newly assembled historical panel datasets. A robust empirical finding, supported by numerous studies, is that credit growth predicts crises. More recent evidence has refined this result further by showing that credit booms concentrated in the non-tradable corporate sector predict future crises (Müller and Verner, 2024). Even with these convincing results, the story of financial crises is still only partially uncovered, with its beginnings remaining largely obscure. We still do not know what causes the "bad" credit booms that culminate in crises.

Following the work of Minsky (1977) and Kindleberger (1978), a widely accepted hypothesis is that increases in aggregate sentiment trigger credit booms and, ultimately, crises. However, the lack of historical data on survey-based beliefs, and the absence of a method for properly measuring sentiment from these beliefs, has long hindered empirical research on this question. In this paper, we develop a novel way of measuring sentiment from survey beliefs, defined as the component of beliefs that can be explained by past public data but is orthogonal to an optimal machine belief formed with the same information set. We show that the majority of variation in beliefs—both across forecasters and over time—can be explained by public information and by the different mental models that individuals use to misaggregate this information relative to the optimal benchmark.

We extend this monthly panel of median sentiment and its dispersion among forecasters to cover 78 countries and 120 years, using state-of-the-art natural language processing and machine learning methods applied to historical Wall Street Journal text data. Our empirical analysis reveals that increases in median sentiment about the economy predict credit booms in the non-tradable corporate sector, which recent research has identified as predictor of financial crises. In addition, we show that the time elapsed since the last financial crisis and the ratio of young to old people in the population predict increases in sentiment. This finding aligns with research on the role of memory recall and similarity in belief formation: when people have not personally experienced a crisis—or as time passes since such rare events—they are less likely to recall them when forming beliefs about the likelihood of future crises.

The results of this paper provide historical validation for the Minsky–Kindleberger view of financial crises and, more importantly, offer evidence of their behavioral foundations: people forget the lessons of rare economic events and increasingly misinterpret the same public information when forming beliefs about crisis risk and the broader economy.

References

ANGELETOS, G.-M., Z. HUO, AND K. A. SASTRY (2021): "Imperfect Macroeconomic Expectations: Evidence and Theory," *NBER Macroeconomics Annual*, 35, 1–86.

ASH, E. AND S. HANSEN (2023): "Text algorithms in economics," *Annual Review of Economics*, 15, 659–688.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring Economic Policy Uncertainty*," *The Quarterly Journal of Economics*, 131, 1593–1636.

BARON, M., E. VERNER, AND W. XIONG (2021): "Banking Crises Without Panics*," *The Quarterly Journal of Economics*, 136, 51–113.

BARON, M. AND W. XIONG (2017): "Credit Expansion and Neglected Crash Risk*," *The Quarterly Journal of Economics*, 132, 713–764.

BHANDARI, A., J. BOROVIČKA, AND P. HO (2024): "Survey Data and Subjective Beliefs in Business Cycle Models," *The Review of Economic Studies*, rdae054.

Bianchi, F., S. C. Ludvigson, and S. Ma (2022a): "Belief Distortions and Macroeconomic Fluctuations," *American Economic Review*, 112, 2269–2315.

——— (2022b): "Belief Distortions and Macroeconomic Fluctuations," *American Economic Review*, 112, 2269–2315.

Bordalo, P., G. Burro, K. Coffman, N. Gennaioli, and A. Shleifer (2025a): "Imagining the Future: Memory, Simulation, and Beliefs," *The Review of Economic Studies*, 92, 1532–1563.

Bordalo, P., J. Conlon, N. Gennaioli, S. Kwon, and A. Shleifer (2025b): "How People Use Statistics," *The Review of Economic Studies*, rdaf022.

Bordalo, P., J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer (2023): "Memory and Probability*," *The Quarterly Journal of Economics*, 138, 265–311.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–2782.

Bordalo, P., N. Gennaioli, R. L. Porta, M. OBrien, and A. Shleifer (2024): "Long-Term Expectations and Aggregate Fluctuations," *NBER Macroeconomics Annual*, 38, 311–347, publisher: The University of Chicago Press.

Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

Brunnermeier, M., E. Farhi, R. S. J. Koijen, A. Krishnamurthy, S. C. Ludvigson, H. Lustig, S. Nagel, and M. Piazzesi (2021): "Review Article: Perspectives on the Future of Asset Pricing," *The Review of Financial Studies*, 34, 2126–2160.

Bybee, L. (2024): "The Ghost in the Machine: Generating Beliefs with Large Language Models," Tech. rep., Yale School of Mangement.

Bybee, L., B. Kelly, A. Manela, and D. Xiu (2024): "Business News and Business Cycles," *The Journal of Finance*, 79, 3105–3147.

Bybee, L., B. Kelly, and Y. Su (2023): "Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text," *Review of Financial Studies*.

Caballero, R. J. and A. Simsek (2020): "A Risk-Centric Model of Demand Recessions and Speculation*," *The Quarterly Journal of Economics*, 135, 1493–1566.

Coibion, O. and Y. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *The American Economic Review*, 105, 2644–2678.

de Silva, T. and D. Thesmar (2023): "Noise in Expectations: Evidence from Analyst Forecasts," *The Review of Financial Studies*, 37, 1494–1537.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *North American Chapter of the Association for Computational Linguistics*.

Driscoll, J. and A. Kraay (1998): "Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data," *The Review of Economics and Statistics*, 80, 549–560.

Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim (2014): "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?" *Journal of Machine Learning Research*, 15, 3133–3181.

Frydman, C. and C. Xu (2023): "Banking Crises in Historical Perspective," *Annual Review of Financial Economics*, 15, 265–290, publisher: Annual Reviews.

Gennaioli, N. and A. Shleifer (2018): *A Crisis of Beliefs: Investor Psychology and Financial Fragility*, Princeton University Press.

Gentzkow, M., B. Kelly, and M. Taddy (2019): "Text as Data," *Journal of Economic Literature*, 57, 535–574.

Graeber, T., C. Roth, and F. Zimmermann (2024): "Stories, Statistics, and Memory*," *The Quarterly Journal of Economics*, 139, 2181–2225.

Greenwood, R., S. G. Hanson, A. Shleifer, and J. A. Sorensen (2022): "Predictable Financial Crises," *The Journal of Finance*, 77, 863–921.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun (2019): "Firm-Level Political Risk: Measurement and Effects*," *The Quarterly Journal of Economics*, 134, 2135–2202.

Jiang, Z., H. Liu, C. Peng, and H. Yan (2025): "Investor Memory and Biased Beliefs: Evidence from the Field*," *The Quarterly Journal of Economics*, qjaf035.

Jorda, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

Jordà, O., M. Schularick, and A. M. Taylor (2017): "Macrofinancial History and the New Business Cycle Facts," *NBER Macroeconomics Annual*, 31, 213–263.

Jordà, , K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor (2019): "The Rate of Return on Everything, 1870–2015*," *The Quarterly Journal of Economics*, 134, 1225–1298.

Juodis, A. and S. Kučinskas (2023): "Quantifying noise in survey expectations," *Quantitative Economics*, 14, 609–650.

Kahana, M. J. and M. J. Kahana (2012): *Foundations of Human Memory*, Oxford, New York: Oxford University Press.

Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*, Palgrave Macmillan.

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

Krishnamurthy, A. and T. Muir (2025): "How Credit Cycles across a Financial Crisis," *The Journal of Finance*, 80, 1339–1378, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13431.

Maenhout, P. J., A. Vedolin, and H. Xing (2025): "Robustness and dynamic sentiment," *Journal of Financial Economics*, 163.

Malmendier, U. and S. Nagel (2011): "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*," *The Quarterly Journal of Economics*, 126, 373–416.

——— (2016): "Learning from Inflation Experiences *," *The Quarterly Journal of Economics*, 131, 53–87.

Malmendier, U. and J. A. Wachter (2024): "Memory of Past Experiences and Economic Decisions," in *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*, ed. by M. J. Kahana and A. D. Wagner, Oxford University Press, 0.

Manela, A. and A. Moreira (2017): "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123, 137–162.

Mian, A., A. Sufi, and E. Verner (2017): "Household Debt and Business Cycles Worldwide*," *The Quarterly Journal of Economics*, 132, 1755–1817.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013): "Efficient Estimation of Word Representations in Vector Space," .

Minsky, H. (1986): *Stabilizing an Unstable Economy*, Yale University Press, New Haven, CT.

Minsky, H. P. (1977): "The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to "Standard" Theory," *Nebraska Journal of Economics and Business*, 16, 5–16.

Müller, K. and E. Verner (2024): "Credit Allocation and Macroeconomic Fluctuations," *The Review of Economic Studies*, 91, 3645–3676.

Nagel, S. and Z. Xu (2022): "Asset Pricing with Fading Memory," *The Review of Financial Studies*, 35, 2190–2245.

Patton, A. J. and A. Timmermann (2010): "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion," *Journal of Monetary Economics*, 57, 803–820.

Pennington, J., R. Socher, and C. Manning (2014): "GloVe: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, ed. by A. Moschitti, B. Pang, and W. Daelemans, Doha, Qatar: Association for Computational Linguistics, 1532–1543.

Reinhart, C. M. and K. S. Rogoff (2009): *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf (2019): "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *ArXiv*, abs/1910.01108.

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

Sufi, A. and A. M. Taylor (2022): "Financial crises: a survey," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 6 of *Handbook of International Economics: International Macroeconomics, Volume 6*, 291–340.

Thesmar, D. and E. Verner (2025): "Beliefs and Stock Market Fluctuations: New Evidence from the Past Seven Decades," .

Wachter, J. A. and M. J. Kahana (2024): "A Retrieved-Context Theory of Financial Decisions*," *The Quarterly Journal of Economics*, 139, 1095–1147.

A. Data Sources

A.1 Survey data

Consensus Economics provides monthly forecasts from 635 financial and other professional forecasting institutions and corporations for a wide range of economic indicators, including real GDP growth, unemployment, inflation, the current account, industrial production, wholesale prices, and 10-year government bond yields. Forecasts are available both at the consensus and individual levels for 25 countries.[7]

Most of the forecasted indicators are expressed at an annual frequency, while the forecasts themselves are collected monthly. We focus on the one-year-ahead real GDP growth forecasts, which have the broadest coverage among all variables. This series is available for all 25 countries, though with varying starting dates beginning in October 1989.

The main advantage of the *Consensus Economics* survey—beyond its wide coverage across countries, indicators, and forecasters—is its monthly frequency. However, because we require a fixed forecast horizon, we interpolate between the forecasts for the next year's and the current year's real GDP growth, both at the consensus and individual levels. Without interpolation, the forecast horizon would vary across successive months. For example, roughly 11½ months for forecasts made in January and 10½ months for those made in February.[8]

Formally, the interpolated one-year-ahead forecast made by forecaster j in period t, and the corresponding interpolated actual growth rate, can be represented as:

$$\begin{aligned} \mathbb{F}_t^j[Y_{t+12}] \approx \mathbb{F}_t^j[Y_{y1(t)}] \times \frac{k-1}{12} + \mathbb{F}_t^j[Y_{y0(t)}] \times \frac{13-k}{12}, \\ Y_{t+12} \approx Y_{y1(t)} \times \frac{k-1}{12} + Y_{y0(t)} \times \frac{13-k}{12}, \\ k \in (1, 2, ...12). \end{aligned}$$

Here, $y0(t)$ and $y1(t)$ refer to the real GDP growth forecasts for the current and next calendar years, respectively, from the perspective of month t, and k corresponds to the month of the forecast (e.g., January = 1, February = 2, . . . , December = 12).

Under this scheme, the interpolated forecast reflects the current-year forecast only in January. In February, the interpolated forecast is a weighted average of 11/12 for the current year and

[7]Austria, Belgium, Canada, Denmark, Egypt, Euro area, Finland, France, Germany, Greece, Ireland, Israel, Italy, Japan, Netherlands, Nigeria, Norway, Portugal, Saudi Arabia, South Africa, Spain, Sweden, Switzerland, United Kingdom, and United States.

[8]The effective horizon is around 11½ months because the survey deadline typically falls in the middle of each month.

Table 4: Data description

Variable	Frequency	Source	Further information	Usage
Consensus forecast (%)	Monthly	Consensus Economics	Interpolated year ahead median forecast of real GDP growth. Detailed description A.1.	2.3
Individual forecast (%)	Monthly	Consensus Economics	Interpolated year ahead individual forecast of real GDP growth. Detailed description A.1.	2.3
Real GDP growth annual (%)	Monthly	IMF WEO	Interpolated real GDP growth. Detailed description A.1.	2.3
Dividend-to-price ratio (%)	Monthly	GFD	Dividends as percent of Stock Price.	2.3
Stock return (%)	Monthly	GFD	Percentage change in stock price index from last month.	2.3
Interest rate (short term) (%)	Monthly	GFD/BIS/National Central Banks	3 month Treasury Yield.	2.3
Interest rate (long term) (%)	Monthly	GFD/National Central Banks	10 year Treasury Yield.	2.3
Inflation (%)	Monthly	IMF	Percent change in CPI from previous month.	2.3
Corporate bond Spread (%)	Monthly	GFD/Datastream	Corporate bond index minus treasury yield.	2.3
Real GDP growth quarterly (%)	Quarterly	OECD	Percent change in seasonally adjusted nominal GDP from previous quarter.	2.3
Industrial Production (%)	Quarterly	OECD	Percentage change of index of production output volume from last quarter.	2.3
Unemployment (%)	Quarterly	ILO/IMF/St Louis Fed/Eurostat/WB	Taken from various sources to cover the entire country set.	2.3
House price index change (%)	Quarterly	OECD	Percentage change of nominal house price index from last quarter.	2.3
House price-to-rent ratio	Quarterly	OECD	The price to rent ratio is the nominal house price index divided by the housing rent price index.	2.3
Investment (%)	Quarterly	OECD	Gross fixed capital formation percentage change from the previous quarter.	2.3
Policy rate (%)	Monthly	Eurostat/IMF/OECD/BIS	Taken from various sources to cover the entire country set.	2.3
ERM-status	Annual	CS	Data taken from Caballero and Simsek (2020).	2.3
Credit-to-GDP ratio	Annual	MV	Various credit-to-GDP ratios constructed from the Müller and Verner (2024) dataset.	4.1
Population	Annual	UN	Logarithm of the total number of population.	4.2
Young-to-old ratio	Annual	UN	Number of people age 0-24 divided by people age over 65.	4.2
Return on risky assets (%)	Annual	JST	Data taken from The Jorda-Schularick-Taylor Macrohistory Database Jordà et al. (2019).	4.2
Return on safe assets (%)	Annual	JST	Data taken from The Jorda-Schularick-Taylor Macrohistory Database Jordà et al. (2019).	4.2

1/12 for the next year, and by December, the weights have reversed (1/12 for the current year and 11/12 for the next year).

This interpolation procedure yields monthly series of one-year-ahead consensus forecasts, individual forecasts, and corresponding realized growth rates. The interpolated consensus forecasts from the past 12 months are used as inputs for the machine learning (ML) model. We exclude the current-period (t) consensus forecast from the ML model's input set to avoid giving the ML model any informational advantage over the survey forecasters.

Bianchi et al. (2022b)

A.2 Macroeconomic and financial data

In section 2.3, we use the following the previous 12 lags of the following monthly variables as predictors in the models forecasting either the individual belief of a forecaster or the actual real GDP growth: dividend to price ratio, interest rates (short and long term), stock returns, dividend-to-price ratios, corporate bond spreads, change in consumer price index, and the interpolated year a head consensus forecasts of real GDP growth. For the industrial production and inflation, we do not include the first lag of the variable to make sure that the machine learning model does not have any information that the forecasters might not have due to delayed publication of the latest numbers. We also include the 2 to 12 lags of the following quarterly variables: real GDP growth, unemployment rate, investments growth, industrial production growth, house price growth, and housing price-to-rent ratio. In addition to the macroeconomic and financial variables, we include a country and month dummy in the prediction models. The sources of these variables can be found in Table 4.

A.3 Historical newsarticle text data

The text data used in this paper consists of all news article titles in the *Wall Street Journal* between the period 1890–2022. The text data were gathered from Proquest Historical Newspapers using their text and data mining (TDM) tool. To obtain this data, one needs to obtain a subsription to the Proquest Historical Newspapers database and to the use of the TDM tool. The text data used in this paper was downloaded with the TDM tool in March 2022.

We included the titles of the following text types: article, commentary, correction, correspondence, editorial, feature, front page, interview, letter to editor, news, review, table of contents and undefined. We excluded titles with less than 10 words. With this filtering the total number of documents in the final corpus is 1,066,317. As BERT embeddings utilize the context or surrounding words of each word when transforming them into a numerical vector, it is not necessary to clean irrelevant words to the similar extent as is common in natural language processing model estimation procedures (e.g. topic models) usually. Excluding some words might deteriorate the quality of the embeddings as the surrounding words describe the context of a word. Hence the only text cleaning procedures included removing empty titles, extra white space and turning all letters to lowercase prior to using the pre-trained DistillBERT model.