Bank of Finland Research Discussion Papers
3 • 2026

Kim Ristolainen

Quantifying Minsky Cycles

Bank of Finland
Research

Bank of Finland Research Discussion Papers
Editor-in-Chief Esa Jokivuolle

Bank of Finland Research Discussion Papers 3/2026
13 April 2026

Kim Ristolainen
Quantifying Minsky Cycles
ISSN 1456-6184, online

Bank of Finland
Research Unit

PO Box 160
FIN-00101 Helsinki

Phone: +358 9 1831

Email: research@bof.fi
Website: www.suomenpankki.fi/en/research/research-unit/

Quantifying Minsky Cycles

Kim Ristolainen *

April 2026

Abstract

We develop a novel sentiment measure from survey forecasts that captures the component of beliefs arising from the systematic misaggregation of public information relative to a machine benchmark based on the same information set. 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. We further find that this sentiment component is shaped by memory-related dynamics, as the time elapsed since major crises and the share of young-to-old people in the population predict surges in optimism even when recent economic developments are controlled for. Taken together, the findings provide new historical evidence consistent with the Minsky–Kindleberger view on financial crises.

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

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

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

*Acknowledgements. I thank the Yrjö Jahnsson Foundation and the Emil Aaltonen Foundation for financial support. I am grateful to Gene Ambrocio and Esa Jokivuolle for helpful comments, and especially to Alp Simsek for early discussions and insightful comments on preliminary ideas. I also thank the Research Unit of the Bank of Finland for its hospitality during a research visit undertaken while preparing this paper.
Contact: Department of Accounting and Finance, University of Turku. E-mail: kkrist@utu.fi.

has been limited. A central obstacle has been the lack of a consistent belief-based measure that can be constructed from observed expectations rather than inferred from financial prices, 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 to measure the component of beliefs that arises from the systematic misaggregation of public information from existing survey data on individual beliefs and extends its time and country coverage through state-of-the-art natural language processing methods and historical news texts. Using an extensive historical panel covering 78 countries and more than a century of economic history, we provide evidence that increases in the news-implied component of this sentiment measure predict 'bad' credit growth historically associated with financial crises. We also provide evidence that memory-related forces shape this sentiment component over time, linking belief formation, credit booms, and financial instability at a historical scale not previously analyzed.

The first contribution of this paper is the development of a new belief-based sentiment measure constructed from survey forecasts. Using monthly survey data from 18 countries over 1991–2020, we decompose professional forecasters' beliefs into a component that reflects the systematic misaggregation of public information relative to a machine benchmark formed from the same public information set.[1] We first estimate, for each forecaster, a machine-learning model that predicts the observed individual forecast using only the public data available at the time. The fitted value from this model captures the part of the individual belief[2] that is explained by public information. We then estimate a separate machine-learning model that uses the same public information to predict the future real GDP growth outcome that forecasters are trying to forecast. Finally, we estimate a further machine-learning model that uses the machine benchmark forecast to predict the public-information-explained component of individual beliefs, and interpret the residual from this prediction as the component arising from misaggregation of public information. We refer to this object as *Public-Information Misaggregation Sentiment* (PIMS).

Our approach allows us to measure both the level of PIMS and its dispersion across forecasters. Because the survey data come from professional forecasters, PIMS should be interpreted as a measure of sentiment in a professional forecasting environment rather than as a direct measure of beliefs held by households or firms in that specific country. We find that a large share of both the time-series and cross-sectional variation in professional forecasts is explained by public

[1]Broadly, the public information set includes lagged macroeconomic indicators, financial-market variables, and lagged survey beliefs available at the time the forecast is formed.

[2]Because *Consensus Economics* reports institutional forecasters, the underlying respondent may not always be the same individual over time. In practice, however, the submitted forecasts likely reflect a fairly stable institution-level forecasting process.

information, although this explanatory power declines sharply during recessions. Professional forecasters' beliefs systematically deviate from machine benchmarks that outperform all individual forecasters, suggesting a role for misaggregation of public information. The empirical relevance of PIMS is further assessed through its predictive relationships with financial markets and survey expectations.

The second contribution of this paper is the historical extension of this framework using textual data, which allows us to study the relationship between sentiment, credit growth, and financial crises over much longer horizons and across a broader set of countries. Specifically, we train machine-learning models to predict both the median level and dispersion of survey-based PIMS across countries. This procedure recovers the historical news-implied component of our measure, which allows us to study the long-run dynamics of PIMS and its dispersion among forecasters over more than a century and across 78 countries for periods without survey data. This historical dataset partially addresses the critique of Brunnermeier et al. (2021) regarding the limited time span and narrow coverage of existing belief data. Panel local projection results document that a rise in news-implied PIMS predicts credit booms in the non-tradable corporate sector that recent research has associated with a higher probability of future financial crises (Müller and Verner, 2024) and weaker subsequent output growth. This shows that periods of elevated sentiment related to the misaggregation of public information predict 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 documents 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 empirical evidence that the time since a major economic crisis strongly predicts increases in the component of beliefs arising from the systematic misaggregation of public information, including in specifications that control for recent output growth and financial market movements. We further show that the share of young relative to old people in the population strongly predicts increases in this component of beliefs in specifications controlling for recent output growth. This is consistent with behavioral mechanisms in which elevated sentiment related to the misaggregation of public information re-emerges as the memory of past instability fades.

We conduct a broad range of robustness checks to assess the credibility of the sentiment measure and the sensitivity of our findings on the relationship between sentiment and credit growth. The historical extension naturally raises concerns about language drift, extrapolation across countries, and the possibility that the machine-learning procedure captures spurious structure rather than a stable mapping from text to sentiment. To assess these concerns, we examine both lexical and semantic overlap between the historical and survey-period corpora, the stability of the text-to-sentiment mapping across alternative random seeds, and the extent to which predictive fit is concentrated in the largest countries. We also show that placebo models based on permuting the sentiment target perform dramatically worse than the baseline specification. Finally, we verify that the historical local-projection results are not driven by any single country, the earliest decades of the sample, weakly fit countries, or extreme sentiment realizations.

Related literature. 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 finds 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 historical evidence that optimism arising from systematic misaggregation of public information predicts credit expansions concentrated in sectors historically linked to crises, thereby linking belief dynamics to credit reallocation that the existing literature has identified as closely associated with 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, documenting 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 papers (Patton and Timmermann, 2010; Angeletos et al., 2021; Bianchi et al., 2022; Juodis and Kučinskas, 2023; de Silva and Thesmar, 2023; Bhandari et al., 2024; Maenhout et al., 2025; Bybee, 2024) develop 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 isolates the component of beliefs arising from the systematic misaggregation of public information, relative to a machine benchmark constructed from the same information set. We then find that this component of belief distortion predicts credit dynamics associated in prior work with future crises.

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 historical evidence consistent with a role for memory recall influencing belief formation, showing that fading memories of past crises and demographic shifts are likely to explain peaks in a central component of sentiment, revealing how memory-based belief formation is connected to recurring cycles in sentiment and financial fragility.

The remainder of the paper proceeds as follows. Section 2 introduces the survey-based measure of the misaggregation-driven component of beliefs and explains its construction and empirical interpretation. Section 3 describes the backcasting methodology and presents the extended historical dataset. Section 4 examines the predictive power of this sentiment component for bad credit booms and explores its behavioral foundations, 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 misaggregation of public information can be measured from survey data with machine learning algorithms and how the resulting measure predicts movements in the financial markets.

2.1. Aggregation mistakes with public and private information

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

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

Without loss of generality, suppose the signals are orthogonal (we can always orthogonalize). Forecasters also observe a private signal z, unavailable to the econometrician, with

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

For now, suppose the private information is orthogonal to the public signals. The full-information Bayesian forecast is therefore

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

Suppose the survey forecasts are arbitrary linear functions of available information,

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

where deviations from the Bayesian forecast $b_i^f - b_i$ reflect aggregation mistakes by forecaster f. Since private information is unobserved, b_z^f cannot be identified. However, if the private signal is orthogonal to the public signals, aggregation mistakes in public information can still be partially identified. The Bayesian forecast that conditions only on partial (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 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 public 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 mistakes that forecasters make in aggregating public data.

2.2. A Kalman-filter example

We illustrate how aggregation mistakes arise in a dynamic environment in which agents and a machine update beliefs about a latent economic state following different Kalman filters.

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 component (underlying state of the economy). 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 (mistake or private information) that is assumed to be 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 that is defined by the variances of the shocks and signals. 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 uncorrelated with public information, allowing Eq. (3) to be estimated.

2.2.2. A machine benchmark

When agents have heterogeneous belief-updating models, only one agent can have the *correct* model. Furthermore, the consensus forecast need not coincide with the model that correctly aggregates public information. Following Bianchi et al. (2022), we construct a machine learning model that serves as a benchmark for aggregation of public signals while excluding private information. Suppose the latent state satisfies

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

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 the fact that $\mathbf{g}_t^m$ is an unbiased predictor of $\mathbf{g}_t$, we can write $\mathbf{g}_t = \mathbf{g}_t^m + \tilde{v}_t$, where $\tilde{v}_t$ is a noise term. Combining this expression with $g_t = \mathbf{g}_t + v_t$ 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). As past machine beliefs are not directly observed, we set it to zero for the initial periods. 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 a machine 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 machine forecasts that use the same public information[3] that was available to human forecasters at the time they made their predictions. A sentiment measure is then formed as the component of survey beliefs that can be explained by public data yet remains unexplained by the

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

machine benchmark. This section details the machine-learning specification and the empirical interpretation of the resulting sentiment measure. Because the underlying survey respondents are professional forecasters, the empirical object recovered here should be interpreted as a distortion component in professional expectations.

2.3.1. Predictability of beliefs and the machine benchmark

Eqs. (3) and (5) motivate the use of lagged public data and lagged beliefs as predictors. Empirically, we replace the linear forecasting rules in those equations with flexible random-forest estimators, allowing for nonlinearities and interactions in the mapping from public information to beliefs and outcomes. 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:

$$g^j_{c,t} = f^j_t(\mathcal{X}_{t-1}) + \varepsilon^j_{c,t}, \tag{6}$$

$$\hat{g}^j_{c,t} \equiv \mathbb{E}[g^j_{c,t} \mid \mathcal{X}_{t-1}] = \hat{f}^j_t(\mathcal{X}_{t-1}), \tag{7}$$

where $g^j_{c,t}$ denotes the belief of forecaster j about real GDP growth $g_{c,t+h}$ in country c formed at time t, with forecast horizon $h = 12$ months. Let $\mathcal{X}_{t-1}$ denote the empirical public-information set available up to time $t-1$, including macroeconomic and financial variables and lagged survey (consensus and individual) beliefs. The function $\hat{f}^j_t(\cdot)$ denotes the random-forest (Breiman, 2001) estimate of the conditional expectation of the belief of forecaster j given the public information set. For each forecaster, we estimate this function recursively, using only observations available up to time t to generate an out-of-sample forecast for period $t+1$ for each forecast date.

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[4] is chosen for its flexibility in modeling 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).

[4]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 helping to limit overfitting (Breiman, 2001).

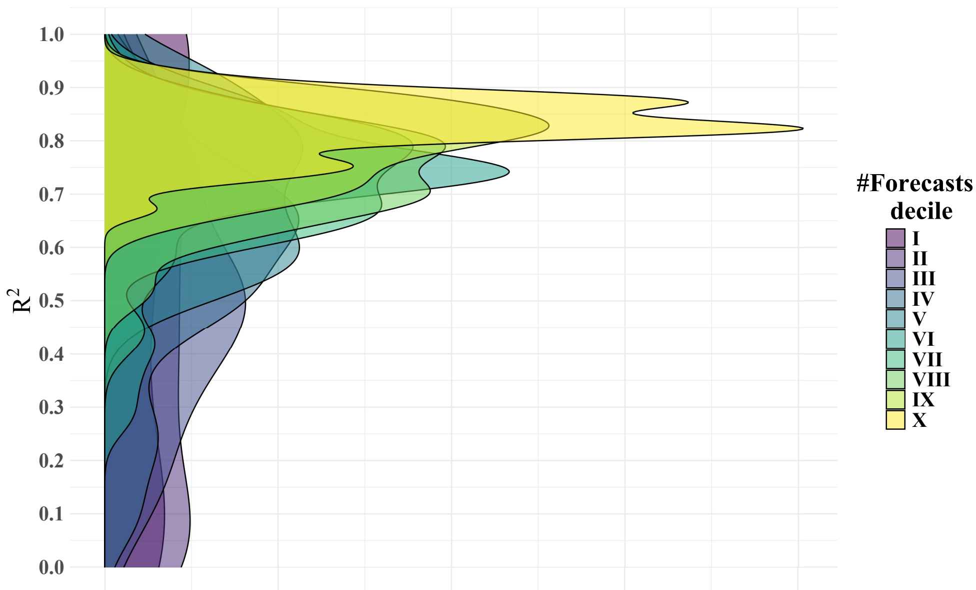


Figure 1: Explainability of beliefs: time-series R^2 distribution across forecasters.

Note: The figure visualizes 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.

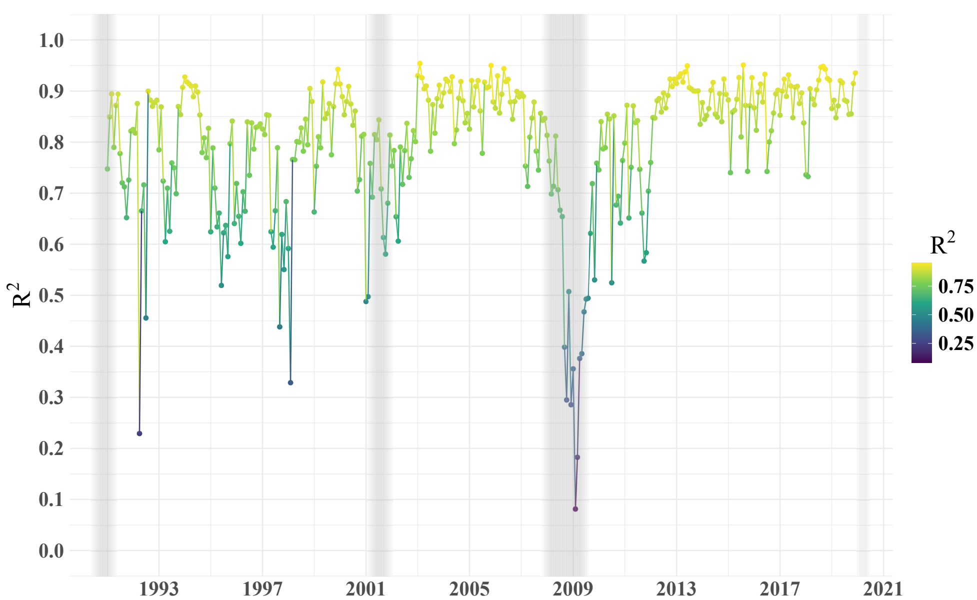


Figure 2: Explainability of beliefs: cross-sectional R^2 in time.

Note: The figure visualizes 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 real GDP growth $g_{c,t+h}$ instead of individual forecasts:

$$g_{c,t+h} = f_t^m(\mathcal{X}_{t-1}) + \varepsilon^m_{c,t+h}. \tag{8}$$

$$\hat{g}^m_{c,t} \equiv \mathbb{E}[g_{c,t+h} \mid \mathcal{X}_{t-1}] = \hat{f}_t^m(\mathcal{X}_{t-1}). \tag{9}$$

Here, $f_t^m(\cdot)$ represents the random-forest model estimated with public information available up to period $t-1$ to produce the machine prediction of future real GDP growth from the same public information set that was available to individual forecasters when they formed their corresponding expectations.

Figures 1 and 2 visualize how much of individual beliefs can be explained with our approach and hence past public data. Figure 1 documents that, on average, about 75% of the time-series variation in individual beliefs is explained by past public information. This share is even higher for forecasters with longer individual time series, exceeding 80% among those with the largest number of forecasts. Figure 2 also shows that 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.

Our analysis, based on a large panel covering 18 countries over 29 years and comprising hundreds of forecasters and nearly 34,000 monthly survey observations, provides evidence that heterogeneity in beliefs is more consistent with differences in forecasting models than with differences in information sets. 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 a large portion of 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 reports two notable patterns. Firstly, the machine learning model outperforms nearly all of the individual forecasters with only past public information. This result suggests that forecasters do not aggregate public information as effectively as the machine benchmark when they are forming beliefs about the future. In addition, the size of the individual survey errors relative to the ones of the machine beliefs are over 2.5 times larger on average in MSE terms.

The systematic difference between individual and machine beliefs is 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 in 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

Table 1: Survey and machine beliefs of one-year-ahead real GDP 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

Note: The table summarizes the survey sample, the explainability of individual beliefs using past public information, and the relative performance of the machine benchmark. The three columns restrict the sample to forecasters with more than 9, 49, and 99 monthly forecasts, respectively. *Cross-sectional* R^2 reports the average monthly share of cross-sectional variation in individual forecasts explained by past public information, while *Time-series* R^2 reports the average forecaster-level time-series fit. The lower panel compares machine and survey beliefs: $MSE_{Machine}$ and $MSE_{Forecaster}$ denote the mean squared errors for the forecasts of the machine and survey beliefs, the MSE ratio reports the average ratio of individual survey mean squared forecast error to machine mean squared forecast error, and the final row reports the share of cases in which the machine benchmark has a lower forecast error than the individual forecaster. All machine predictions are recursive out-of-sample forecasts constructed from the same public information set available to survey respondents.

global financial crisis of 2008. As individual errors to both directions (positive and negative) are larger than machine's belief errors, it suggests that individual forecasts tend to overreact in both directions relative to the machine benchmark.

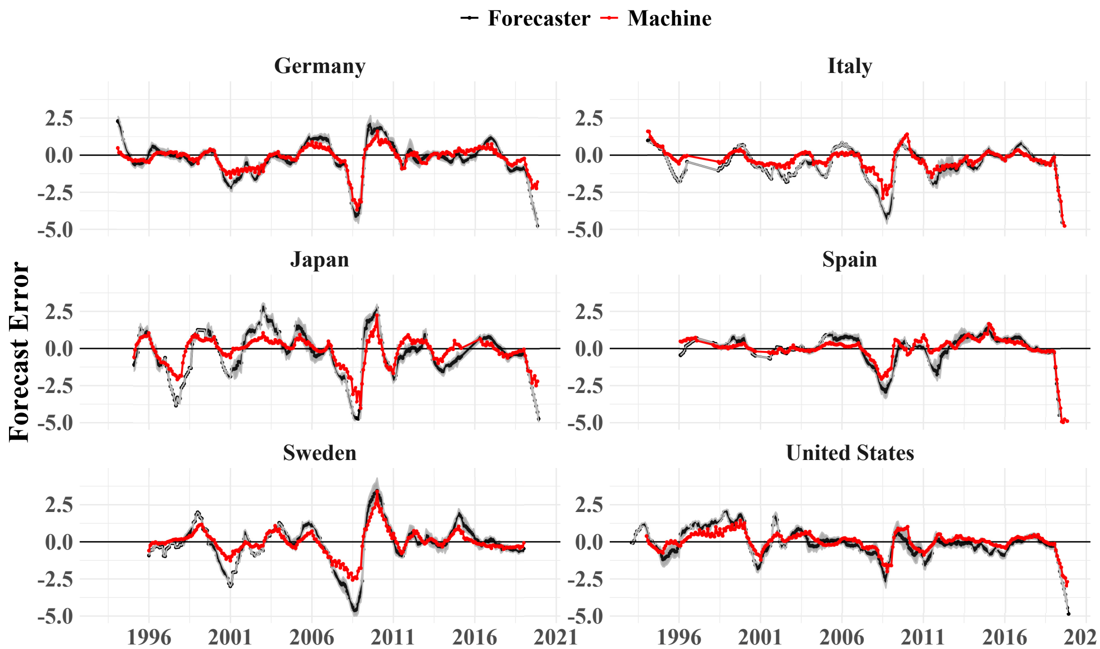


Figure 3: Survey forecast error distribution and machine forecast error in time.

Note: The figure visualizes in red the out-of-sample forecast error (defined as $g_{c,t+h} - \hat{g}^{m}_{c,t}$) of a random forest model predicting one year-ahead ($h = 12$) real GDP growth with only past public data available up to the period when prediction is made. The black line visualizes the median prediction error (defined as $g_{c,t+h} - g^{j}_{c,t}$) 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

Sentiment can be thought of broadly as belief distortion relative to a rational benchmark. Our empirical measure does not identify total sentiment, but rather its identifiable public-information component, which arises from the misaggregation of public information. We refer to this object as *Public-Information Misaggregation Sentiment* (PIMS). For expositional simplicity, we refer to PIMS as sentiment in the remainder of the paper unless stated otherwise. Throughout, however, this shorthand should be understood as referring to the public-information-related distortion component identified from professional forecasts, not to a comprehensive measure of aggregate sentiment in the economy.

At the individual forecaster level, PIMS is denoted by $\mathcal{M}^{j}_{c,t}$ and it captures the component of individual beliefs $\hat{g}^{j}_{c,t}$ that can be explained by past public data, but remains unexplained by the machine belief based on the same information, $\hat{g}^{m}_{c,t}$. Formally,

$$\mathcal{M}^{j}_{c,t} = \hat{g}^{j}_{c,t} - \hat{q}^{j}(\hat{g}^{m}_{c,t}), \tag{10}$$

where $\hat{q}^{j}(\cdot)$ denotes the random forest prediction of $\hat{g}^{j}_{c,t}$ given the machine forecast $\hat{g}^{m}_{c,t}$ of real GDP growth from past public data. $\mathcal{M}^{j}_{c,t}$ is then the residual from a flexible nonlinear projection of fitted forecaster beliefs on the fitted machine benchmark. Under correct aggregation of public information, the fitted individual belief should be fully explained by the fitted machine benchmark, so that $\mathcal{M}^{j}_{c,t} = 0$. We estimate the random forest model separately for each forecaster j in the sample and derive $\mathcal{M}^{j}_{c,t}$ for all available forecasts across countries and time periods.

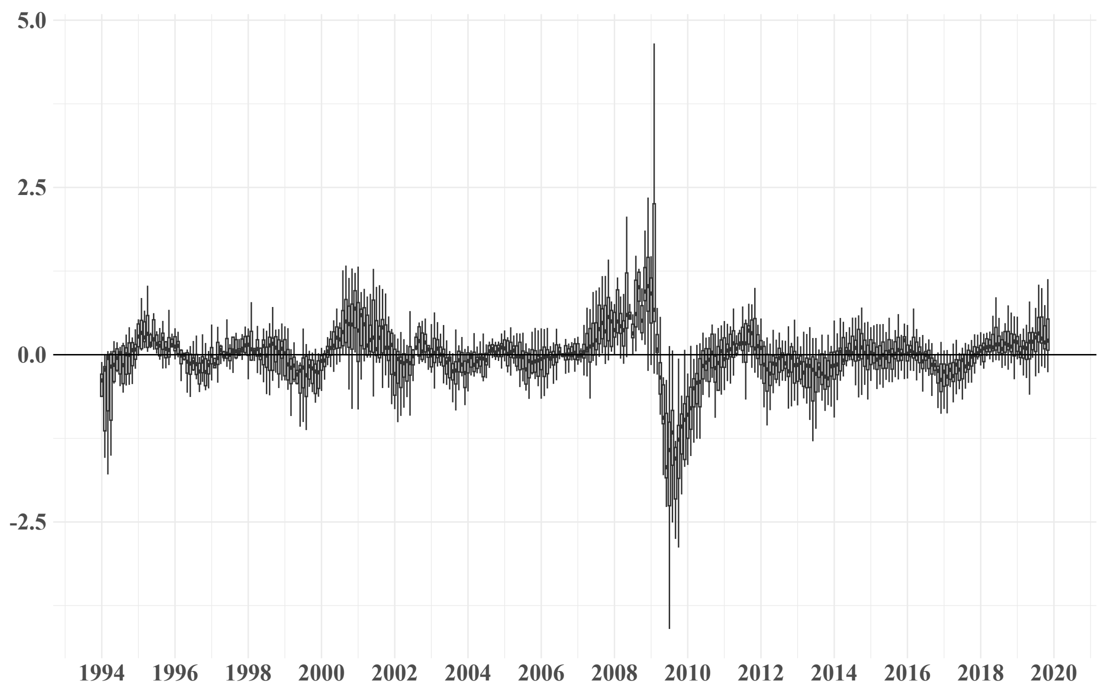


Figure 4: Distribution of median sentiment $\mathcal{M}^{50th}_{c,t}$ across 18 developed countries over time.

Note: Median sentiment $\mathcal{M}^{50th}_{c,t}$ is constructed from the individual forecaster $\mathcal{M}^{f}_{c,t}$ distribution for forecasts of country c in period t. The sentiment $\mathcal{M}^{j}_{c,t}$ for forecaster j is defined as the residual from projecting the fitted forecaster belief $\hat{g}^{j}_{c,t}$ on the fitted machine benchmark $\hat{g}^{m}_{c,t}$, both constructed using the same public information set.

Figure 4 visualizes the monthly distribution of the median sentiment $\mathcal{M}^{50th}_{c,t}$ across countries in time and suggests three patterns. First, there seems to be a global cyclical component in sentiment. From 1994 to 2020 one can see around 12 global sentiment cycles in the data. Some cycles are stronger than others with 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 three years.

Finally, from 2002 to 2007, global sentiment seems to stay relatively neutral. This is interesting as the narrative about the time preceding the GFC stated that there existed excess optimism for many years until the year the crisis started. However, Figure 4 shows the distribution of median sentiment across the country sample and therefore cannot rule out the possibility that the excess positive sentiment was specific to the United States, to a subset of countries, or to the housing sector rather than to the economy as a whole.

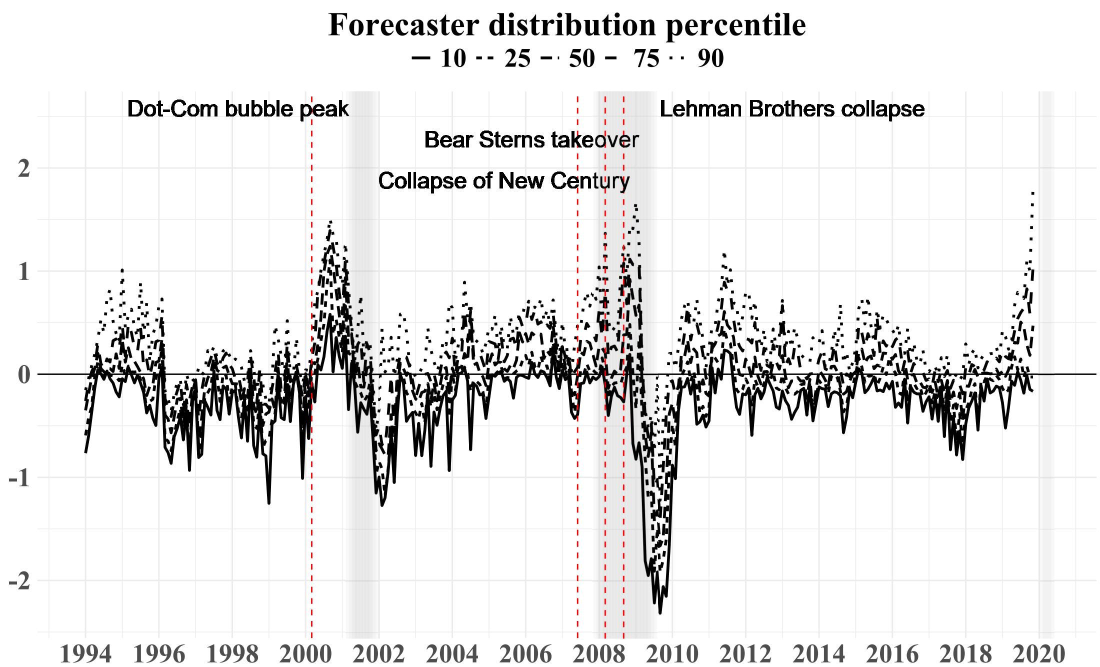


Figure 5: Monthly percentiles of the distribution of individual forecaster sentiment $\mathcal{M}^{j}_{US,t}$ in the United States.

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 three to four 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 machine belief) or became somewhat pessimistic whereas 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 the 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 machine belief was and then it dropped to 2 percentage points too low what the machine belief implied.

2.3.3. Empirical relevance and interpretation

Next, we analyze how financial-market variables and survey expectations react to a shock in the level of median sentiment $\mathcal{M}_{c,t}^{50th}$ by estimating panel local projection regressions (Jorda, 2005), where the response variable $\Delta_h Y_{c,t+h}$ is the $h = 0, \ldots, 20$ month-ahead change in the median survey forecast, median forecast error, policy rate, average corporate bond spread, average dividend-to-price ratio, or cumulative stock return for country c. More formally,

$$\Delta_h Y_{c,t+h} = Y_{c,t+h} - Y_{c,t} = \alpha_c^h + \beta^h \mathcal{M}_{c,t}^{50th} + \Gamma^{h\prime} X_{c,t-1} + \epsilon_{c,t}. \tag{11}$$

In all specifications, the control vector $X_{c,t-1}$ contains two lags of median sentiment and two lags of the response variable. We report both 68% and 95% confidence intervals calculated using Driscoll and Kraay (1998) standard errors with a lag length of six months. The top row of Figure 6 reports the baseline specification with country fixed effects only, while the bottom row adds month fixed effects and therefore absorbs common global month-specific shocks.

Figure 6 reports monthly local projection responses of financial-market and belief variables to shocks in the level of sentiment. In the baseline specification shown in the top row, which includes country fixed effects only, stock prices fall immediately and remain below baseline for an extended period, while dividend–price ratios rise and corporate bond spreads widen. Median one-year-ahead forecasts also decline persistently after the shock. Forecast errors initially move little, but begin to increase after several months, implying that realized outcomes subsequently exceed the lower forecasts formed after the initial adjustment. Policy rates also decline gradually over time. The bottom row adds month fixed effects and therefore absorbs common global month-specific shocks. Under this stricter specification, the qualitative responses of stock prices, dividend–price ratios, bond spreads, median forecasts, and forecast errors remain broadly similar, although the response of policy rates is no longer statistically significant.

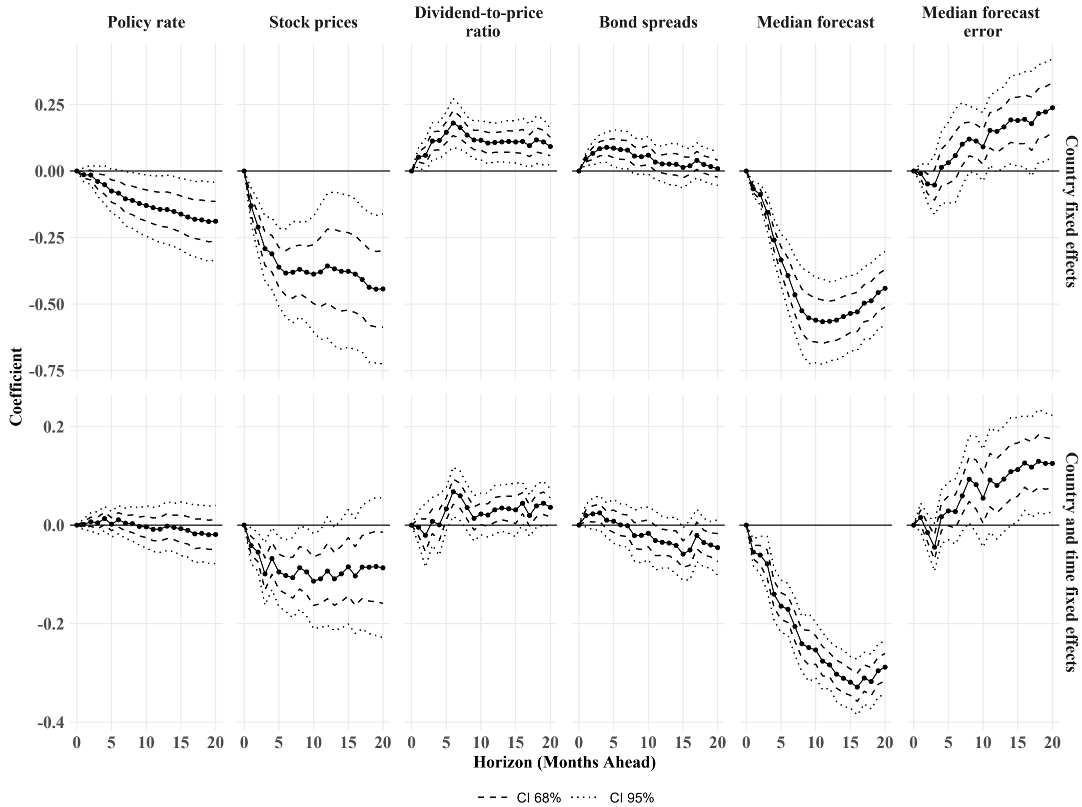


Figure 6: Monthly local projections of financial-market and belief variables following shocks to the level of sentiment.

Note: This figure presents local projection impulse responses of policy rates, stock prices, dividend-to-price ratios, bond spreads, median forecasts, and median forecast errors following innovations in the level of median sentiment $\mathcal{M}^{50th}_{c,t}$. The impulse responses are based on estimation of Eq. 11 with two lags of the sentiment measure and two lags of the response variable. The top row reports specifications with country fixed effects only, while the bottom row additionally includes month fixed effects and therefore absorbs common global month-specific shocks. The responses represent the change in the response variable 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 panel comprises 18 countries observed over 285 months, and the number of monthly observations ranges across specifications from 3475 to 3493.

These results suggest that elevated sentiment is followed by financial-market repricing and downward revision in survey expectations. The baseline specification captures both the country-specific and the common global component of sentiment toward a country, which is useful given that international financial markets are likely to respond to both. Once month fixed effects are included, and the common global component is removed, the responses of stock prices, valuation ratios, bond spreads, forecasts, and forecast errors remain broadly intact. By contrast, the disappearance of a statistically significant policy-rate response indicates that monetary policy reactions are more closely tied to the common global component of sentiment. Overall, the evidence suggests that both the global and country-specific components of sentiment are relevant for subsequent financial-market and belief dynamics, while only the global component appears to be systematically related to policy-rate movements.

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 early 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 news article 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.

A natural starting point is that news content may contain information correlated with sentiment 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 documented 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 lost when a corpus of documents is turned into a so-called document-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 news-title data 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,

$$\mathbf{h}^{(s)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{x}_i^{(s)}, \tag{12}$$

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:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}^{(1)} \\ \mathbf{h}^{(2)} \\ \vdots \\ \mathbf{h}^{(S)} \end{bmatrix} \in \mathbb{R}^{S \times d}, \tag{13}$$

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[5] 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 50 most significant principal components (PCs), ensuring computational efficiency while retaining the most relevant features. The PCA transformation can be expressed as:

$$\mathbf{Z} = \mathbf{H}\mathbf{W},$$

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.

[5]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.

To predict sentiment at a monthly level, we aggregate the PCA-reduced news title embeddings $\mathbf{Z}$ 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, $S_{c,t}$ be the number of news titles in month t that mention a specific country c, $\mathbf{Z}_t \in \mathbb{R}^{S_t \times k}$ be the matrix of PCA-reduced embeddings for all news titles in month t and $\mathbf{Z}_{c,t} \in \mathbb{R}^{S_{c,t} \times k}$ be the matrix of PCA-reduced embeddings for news 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:

$$\bar{\mathbf{Z}}_t = \frac{1}{S_t} \sum_{s=1}^{S_t} \mathbf{z}_t^s, \tag{14}$$

$$\bar{\mathbf{Z}}_{c,t} = \frac{1}{S_{c,t}} \sum_{s=1}^{S_{c,t}} \mathbf{z}_{c,t}^s, \tag{15}$$

where $\mathbf{z}_t^s \in \mathbb{R}^k$ is the PCA-reduced embedding of news title s in month t. $\mathbf{z}_{c,t}^s \in \mathbb{R}^k$ is the PCA-reduced embedding of news title s in month t mentioning country c. $\bar{\mathbf{Z}}_t \in \mathbb{R}^k$ represents the *general monthly news embedding* and $\bar{\mathbf{Z}}_{c,t} \in \mathbb{R}^k$ 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. We consider two sentiment variables that we backcast over time: the median sentiment level $\mathcal{M}_{c,t}^{50\text{th}}$ in country c at month t, and the cross-sectional dispersion of sentiment,

$$\mathcal{M}_{c,t}^{\text{dispersion}} = \max_j \mathcal{M}_{c,t}^j - \min_j \mathcal{M}_{c,t}^j,$$

which measures the dispersion of sentiment across forecasters and thereby captures disagreement in the sentiment distribution. For each country c and month t, our goal is to predict both $\mathcal{M}_{c,t}^{50\text{th}}$ and $\mathcal{M}_{c,t}^{\text{dispersion}}$ with the current embeddings, one month lags of the embeddings, lagged 3-month and 6-month moving averages of the embeddings. Thus, the complete news feature vector for country c in month t is:

$$\mathcal{X}_{c,t}^{News} = \left[\bar{\mathbf{Z}}_t, \bar{\mathbf{Z}}_{c,t}, \bar{\mathbf{Z}}_{t-1}, \bar{\mathbf{Z}}_{c,t-1}, \bar{\mathbf{Z}}_{t-1}^{3m}, \bar{\mathbf{Z}}_{t-1}^{6m}, \bar{\mathbf{Z}}_{c,t-1}^{3m}, \bar{\mathbf{Z}}_{c,t-1}^{6m}\right] \in \mathbb{R}^{8k}, \tag{16}$$

where $\bar{\mathbf{Z}}_t \in \mathbb{R}^k$: general news embedding for month t, $\bar{\mathbf{Z}}_{c,t} \in \mathbb{R}^k$: country-specific news embedding for month t, country c, $\bar{\mathbf{Z}}_{t-1}$, $\bar{\mathbf{Z}}_{c,t-1}$: one-month lagged embeddings, $\bar{\mathbf{Z}}_{t-1}^{3m} = \frac{1}{3}\sum_{\tau=t-3}^{t-1} \bar{\mathbf{Z}}_\tau$, the 3-month moving average of general news embeddings (lagged by one month), $\bar{\mathbf{Z}}_{t-1}^{6m} = \frac{1}{6}\sum_{\tau=t-6}^{t-1} \bar{\mathbf{Z}}_\tau$, the 6-month moving average of general news embeddings (lagged by one month), $\bar{\mathbf{Z}}_{c,t-1}^{3m} = \frac{1}{3}\sum_{\tau=t-3}^{t-1} \bar{\mathbf{Z}}_{\tau,c}$, the 3-month moving average of country-specific news embeddings (lagged by one month), $\bar{\mathbf{Z}}_{c,t-1}^{6m} = \frac{1}{6}\sum_{\tau=t-6}^{t-1} \bar{\mathbf{Z}}_{\tau,c}$ 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.

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.[6] 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 by comparing each observation to the average prediction from the subset of trees that did not include that observation in their bootstrap samples.

To improve model performance, hyperparameter tuning is conducted over the minimum number of observations required in a terminal node and the number of features randomly considered at each split. A smaller minimum node size allows deeper trees and can capture finer patterns in the data, although it may also increase the risk of overfitting at the level of individual trees. A smaller number of candidate features at each split increases diversity across

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

trees, whereas a larger number can improve the quality of individual splits but also increase correlation among trees. The final model is selected as the combination of minimum node size and number of features considered at each split that yields the lowest out-of-bag mean squared error, with the aim of achieving a good bias-variance trade-off.

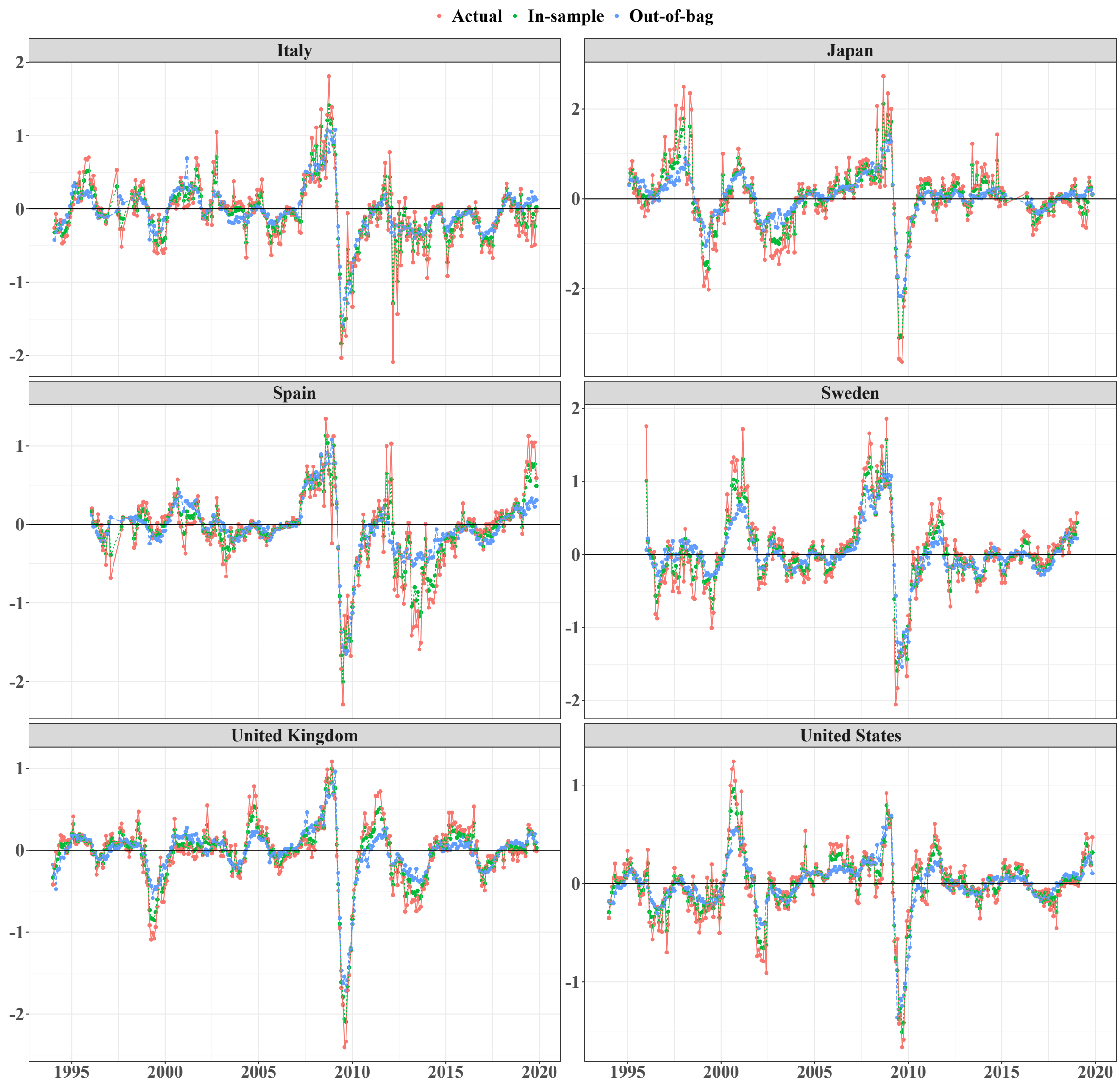


Figure 7: Random forest predictions of survey derived median sentiment with newsarticle BERT embedding principal components.

For each sentiment measure, we train separate random forest models:

$$\widehat{\mathcal{M}}_{c,t}^{50th,\text{news}} = \hat{f}^{\text{50th}}(\mathcal{X}_{c,t}^{news}), \tag{17}$$

$$\widehat{\mathcal{M}}_{c,t}^{\text{dispersion,news}} = \hat{f}^{\text{dispersion}}(\mathcal{X}_{c,t}^{news}), \tag{18}$$

where f^{50th} is the random forest model predicting median sentiment $\mathcal{M}_{c,t}^{50th,}$, $f^{\text{dispersion}}$ is the random forest model predicting sentiment dispersion $\mathcal{M}_{c,t}^{dispersion,}$. We define the predictions $\widehat{\mathcal{M}}_{c,t}^{50th,\text{news}}$ and $\widehat{\mathcal{M}}_{c,t}^{dispersion,\text{news}}$ as the news implied median sentiment and dispersion of sentiment for country c in month t. The model is trained with initial 18 countries[7] 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.

3.2.2. Model performance and backcasting of sentiment

Figure 7 visualizes the performance of the random forest model in predicting the median sentiment $\mathcal{M}_{c,t}^{50th}$ for a selected subset of six countries. Importantly, while the backcasting procedure aims to approximate the historical sentiment measures as closely as possible, it is not realistic to expect news text alone to explain their full variation. Instead, the procedure recovers the component of sentiment that is systematically linked to the news information set, which accounts for a substantial share of observed sentiment variation.

The OOB predictions (blue) show that the model is able to generalize the predictive relationship between news information and median sentiment reasonably well. The OOB R^2 values are 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 news-article title information. The close alignment of the true values of sentiment, the in-sample and OOB-sample prediction suggests that the fitted relationship generalizes reasonably well and has not overfitted to the training data.

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. Figure 8 plots an example time series of the news-implied median sentiment for the United States. The shaded bands report out-of-bag 90% conformal prediction intervals for the backcasted portion of the series. They are constructed from the empirical distribution of out-of-bag prediction errors in the observed sample and should therefore be interpreted as

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

approximate uncertainty bands around the historical backcasts. 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[8] was chosen to cover the countries for which Müller and Verner (2024) have collected detailed corporate sector credit data.

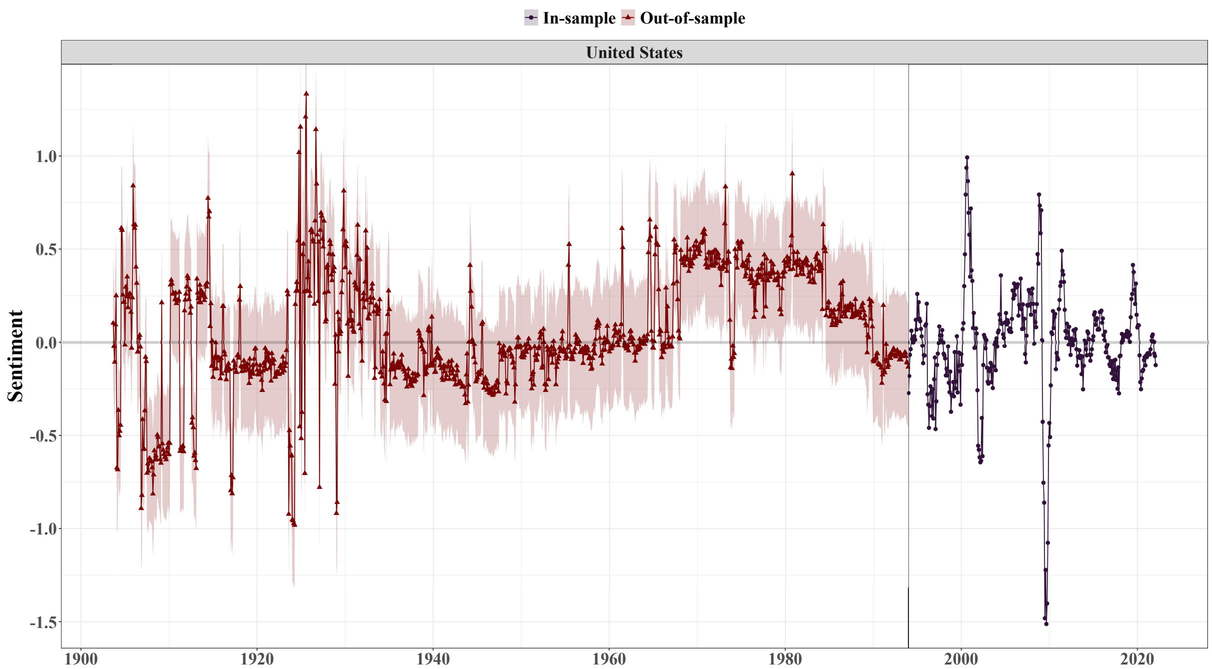


Figure 8: Backcasted sentiment $\widehat{\mathcal{M}}^{50th,\text{news}}$ for the US.

Note: Shaded bands show 90% out-of-bag conformal prediction intervals for the backcasted portion of the series.

A natural concern in the backcasting exercise is that the language of news may evolve over time, potentially weakening the connection between the historical text corpus and the survey-period mapping used for estimation. Appendix B examines this issue using both lexical and semantic diagnostics. First, we measure how much of the historical corpus remains covered by the survey-period vocabulary and by a more restrictive stable vocabulary consisting of sufficiently frequent survey-period words. This addresses the concern that the historical backcast may rely

[8]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.

heavily on words or expressions that are absent from the period in which the text-to-sentiment mapping is trained. Second, we construct BERT-based semantic similarity measures that compare each historical news month to its nearest survey-period month in embedding space. This provides a complementary check on whether the historical corpus remains close to the support of the survey-period text even when wording changes over time.

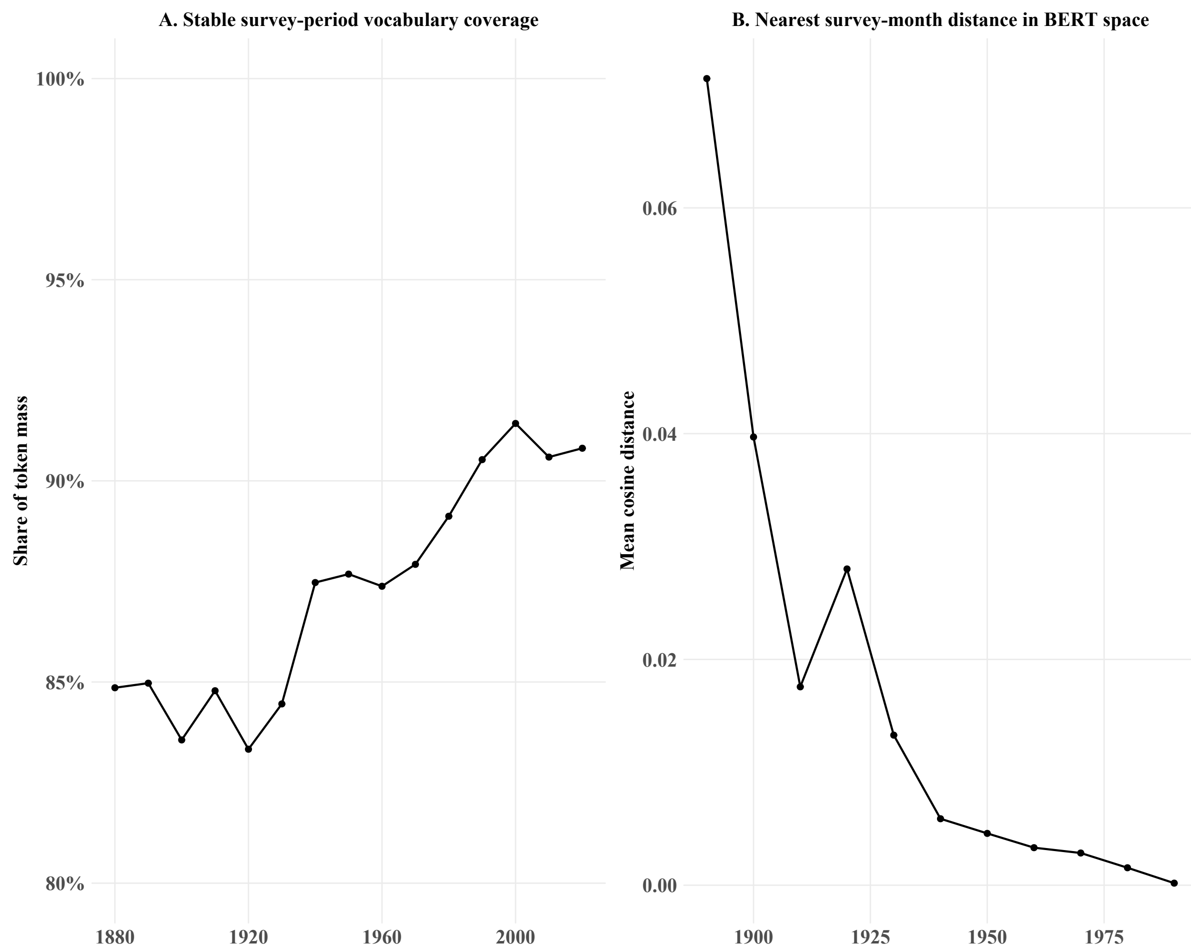


Figure 9: Language-drift diagnostics for the backcasting exercise.

Note: The left panel reports the share of token mass retained when the historical news corpus is restricted to a stable vocabulary defined from the survey period. The right panel reports the mean cosine distance between each historical month and its nearest survey-period month in BERT embedding space. Together, the panels indicate substantial lexical continuity and meaningful semantic support between the historical and survey-period corpora, although similarity declines gradually further back in time.

Figure 9 summarizes these diagnostics. The left panel shows that a large share of historical token mass remains within the survey-period vocabulary and, more stringently, within the stable survey-period vocabulary throughout the historical sample. The right panel shows that historical news months remain semantically close to at least some survey-period months in BERT embedding space, although this similarity declines gradually further back in time. Taken together, these diagnostics suggest that language drift is present, but does not imply a collapse in overlap between the historical and survey-period news environments.

Appendix B further evaluates whether the estimated text-to-sentiment mapping is stable and informative within the survey period before it is used for historical backcasting. First, we re-estimate the random-forest models across many alternative random seeds while holding the baseline hyperparameters fixed. This isolates the role of stochastic variation in forest construction and shows that out-of-bag fit is essentially unchanged across replications, indicating that the first-stage mapping is not sensitive to arbitrary algorithmic randomness.

Second, we examine predictive fit separately across countries. Because the historical text source is the *Wall Street Journal*, a natural concern is that the mapping may work well only for the largest and most internationally salient countries. The appendix shows that although predictive performance is somewhat stronger for major countries, it is not confined to them, and a substantial part of the cross-country heterogeneity in fit is related to differences in survey sample length rather than country salience alone.

Third, we conduct a falsification exercise in which the sentiment target is randomly permuted across observations while the news-based predictors and baseline hyperparameters are held fixed. Under this placebo design, out-of-bag fit deteriorates sharply, with the placebo models performing dramatically worse than the baseline models. Taken together, these checks suggest that the backcast does not arise from spurious machine-learning fit or a fragile algorithmic specification, but instead reflects a reasonably stable and economically meaningful mapping from news text to the sentiment measure.

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.

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) find 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) documents 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 the news-implied component of sentiment about the economy, anchored to our survey-based PIMS measure. In this way, we can examine whether a sudden collective misaggregation of public information among investors initiates the Minsky Cycle—that is, a bad credit boom ending in a bust and a crisis.

To analyze the relationship between sentiment and credit growth, we utilize 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 $\widehat{\mathcal{M}}_{c,t}^{50th,\text{news}}$ and sentiment concentration $\widehat{\mathcal{M}}_{c,t}^{dispersion,\text{news}}$. In this section, the relevant historical object is the news-implied component of sentiment recovered from WSJ coverage. For non-U.S. countries, it should therefore be interpreted as internationally mediated sentiment about the country rather than as a direct measure of domestic beliefs.

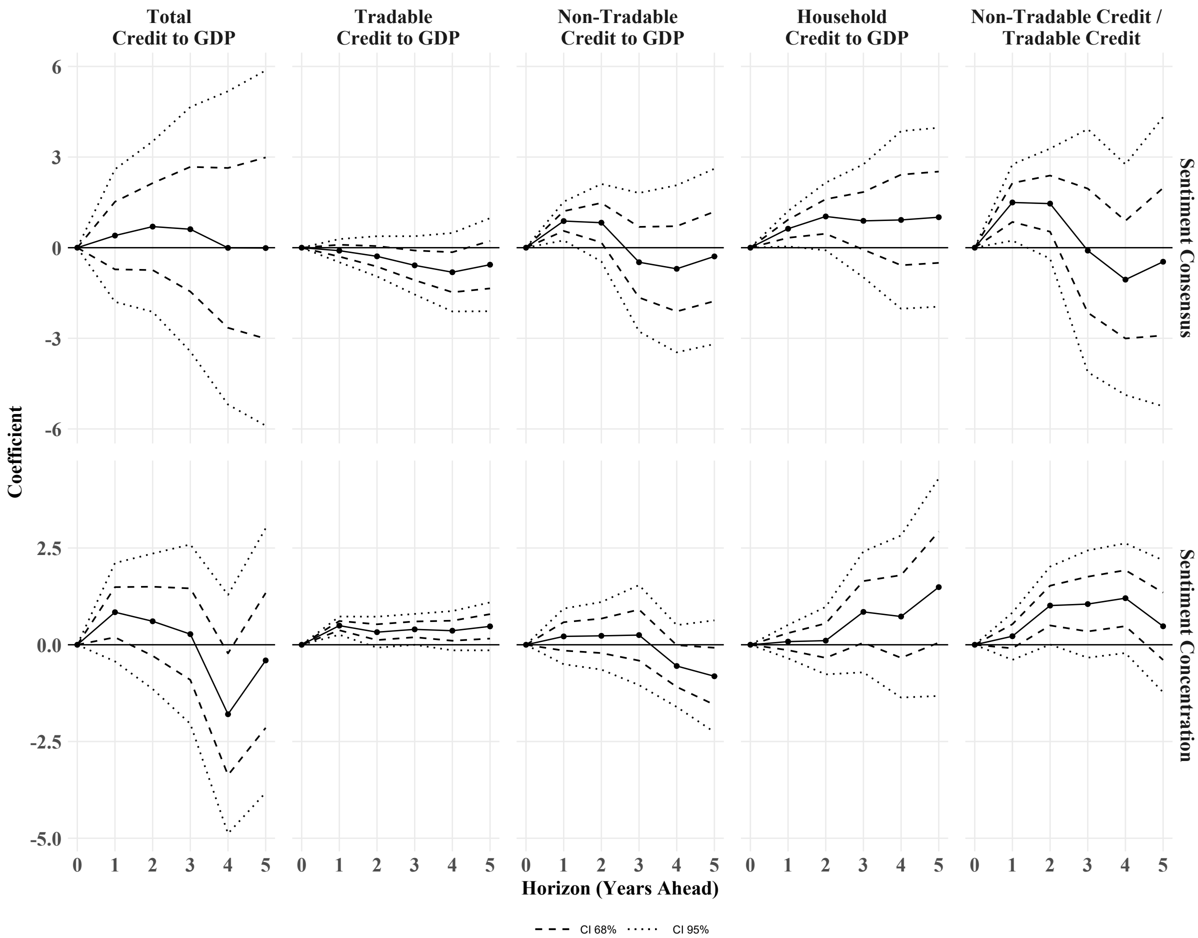


Figure 10: Historical local projections of aggregate and broad sectoral credit following shocks to news-implied sentiment.

Note: This figure reports panel local-projection impulse responses of total credit-to-GDP, tradable-sector credit-to-GDP, non-tradable-sector credit-to-GDP, household credit-to-GDP, and the non-tradable-to-tradable credit ratio to shocks in news-implied median sentiment and sentiment concentration. The sentiment measures are the historical text-based counterparts of the survey-period PIMS measures. The impulse responses are based on estimation of equation 19 with two lags of the sentiment measure and one lag of the corresponding credit variable as controls. Country and year fixed effects are included, and the responses represent the change in the credit variable from period 0 to period h. Dashed (dotted) lines denote 68% (95%) confidence intervals computed using Driscoll and Kraay (1998) standard errors with a lag length of two. Depending on the specification, the sample ranges from 74 to 78 countries, 62 to 103 years, and 1704 to 4698 annual observations.

We estimate panel local projection regressions where the response variable $\Delta_h Y_{c,t+h}$ is the change of a measure of credit-to-GDP from t to $t+h$, for $h = 1, \ldots, 5$ and the shock variable $\hat{\mathcal{M}}^{\text{news}}_{c,t}$ is a backcasted measure of sentiment for country c at year t. More formally,

$$\Delta_h Y_{c,t+h} = Y_{c,t+h} - Y_{c,t} = \alpha_c^h + \lambda_t^h + \beta^h \hat{\mathcal{M}}_{c,t}^{\text{news}} + \Gamma^{h\prime} X_{c,t-1} + \epsilon_{c,t+h}, \tag{19}$$

where $h = 1, \ldots, 5$, α_c^h denote country fixed effects, and λ_t^h denote year fixed effects that control for common global shocks. We include two lags of the sentiment measure and one lag of the credit variable as controls denoted by $X_{c,t-1}$. We report both 68% and 95% confidence intervals calculated according to Driscoll and Kraay (1998) standard errors with a lag length of two years.

Figure 10 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 10 provides evidence 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 find that the reallocation of credit between sectors predicts subsequent economic slowdowns and financial crises.

The local projection results displayed in Figure 11 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 12 presents the local projection estimates from Figure 10 separately for periods with very low sentiment concentration and for all other periods. 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 not very low. This pattern is consistent with the idea that harmful surges in sentiment are collective in nature, in the spirit of Keynes (1936)'s animal spirits.

Appendix B further examines the robustness of these historical local-projection results through a series of targeted exercises designed to address the main vulnerabilities of the back-casted sentiment measure. First, we re-estimate the local projections while excluding one country at a time, which allows us to assess whether the baseline coefficients are driven disproportionately by any single country in the 78-country panel. Second, we restrict the sample to the post-1950 period in order to reduce reliance on the earliest historical decades, where language drift and extrapolation concerns are likely to be strongest. Third, we exclude the countries for which the survey-period text-to-sentiment mapping has the weakest predictive fit, thereby focusing on the part of the historical panel where the first-stage sentiment measure is likely to be most reliable. We also combine these two restrictions and, separately, winsorize the top and bottom 1% of the sentiment distribution to verify that the results are not driven by a small number of extreme historical realizations.

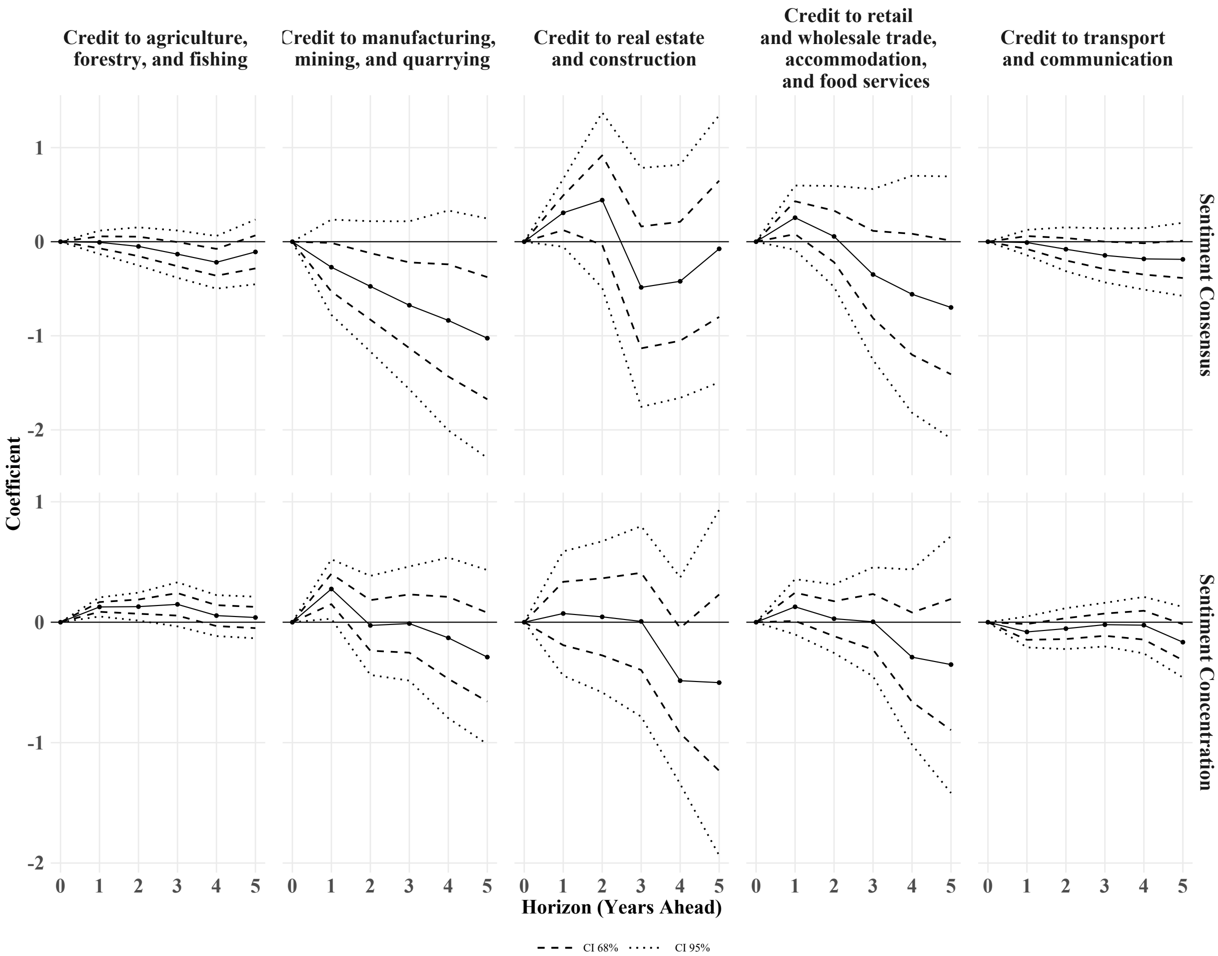


Figure 11: Historical local projections of credit growth across detailed industries following shocks to news-implied sentiment.

Note: This figure presents local projection impulse responses of credit (for different industrial 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.

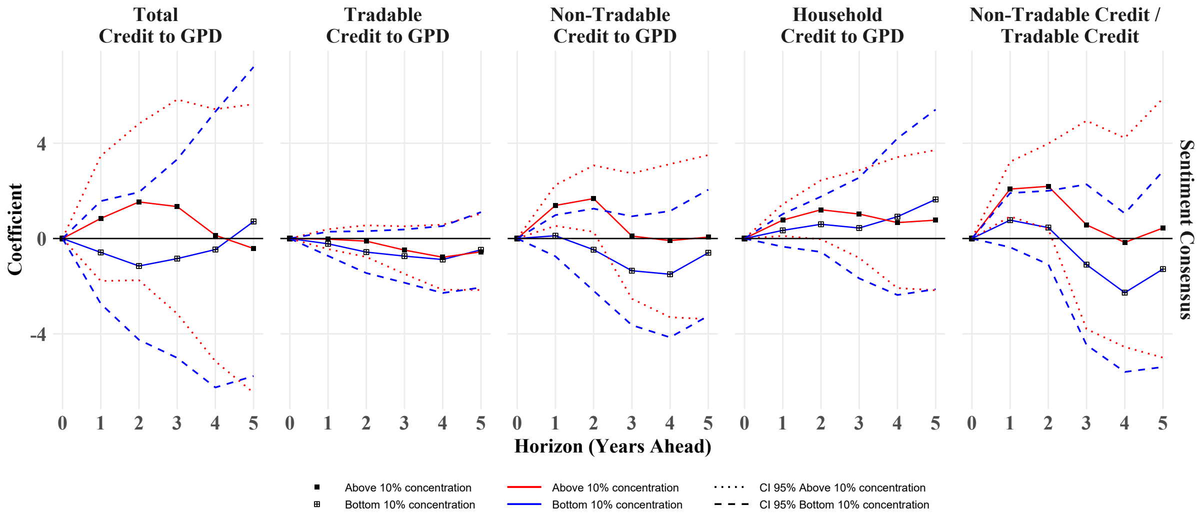


Figure 12: Historical local projections separately for periods of very low and higher sentiment concentration.

Note: This figure presents local projection impulse responses of credit to GDP following innovations in the median sentiment (consensus) separately for very low and higher sentiment concentration (negative of belief dispersion) periods. Very low (higher) sentiment concentration periods are the ones with sentiment concentration below (above) the 10th 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 78 countries, 62 to 103 years and from 1704 to 4672 annual observations.

Taken together, these exercises show that the main qualitative response patterns remain broadly intact. The leave-one-country-out estimates indicate that the results are generally not driven by any single country, while the restricted-sample and winsorization checks show that the findings are not solely a consequence of the earliest historical observations, the weakest-fit countries, or a handful of extreme sentiment shocks. Across specifications, the strongest evidence remains concentrated in household credit and in the short-run response of non-tradable to tradable credit ratio. These findings provide historical evidence that elevated misaggregation-driven sentiment predicts credit growth in sectors historically associated with financial fragility.

4.2. Memory, experience, and the formation of sentiment

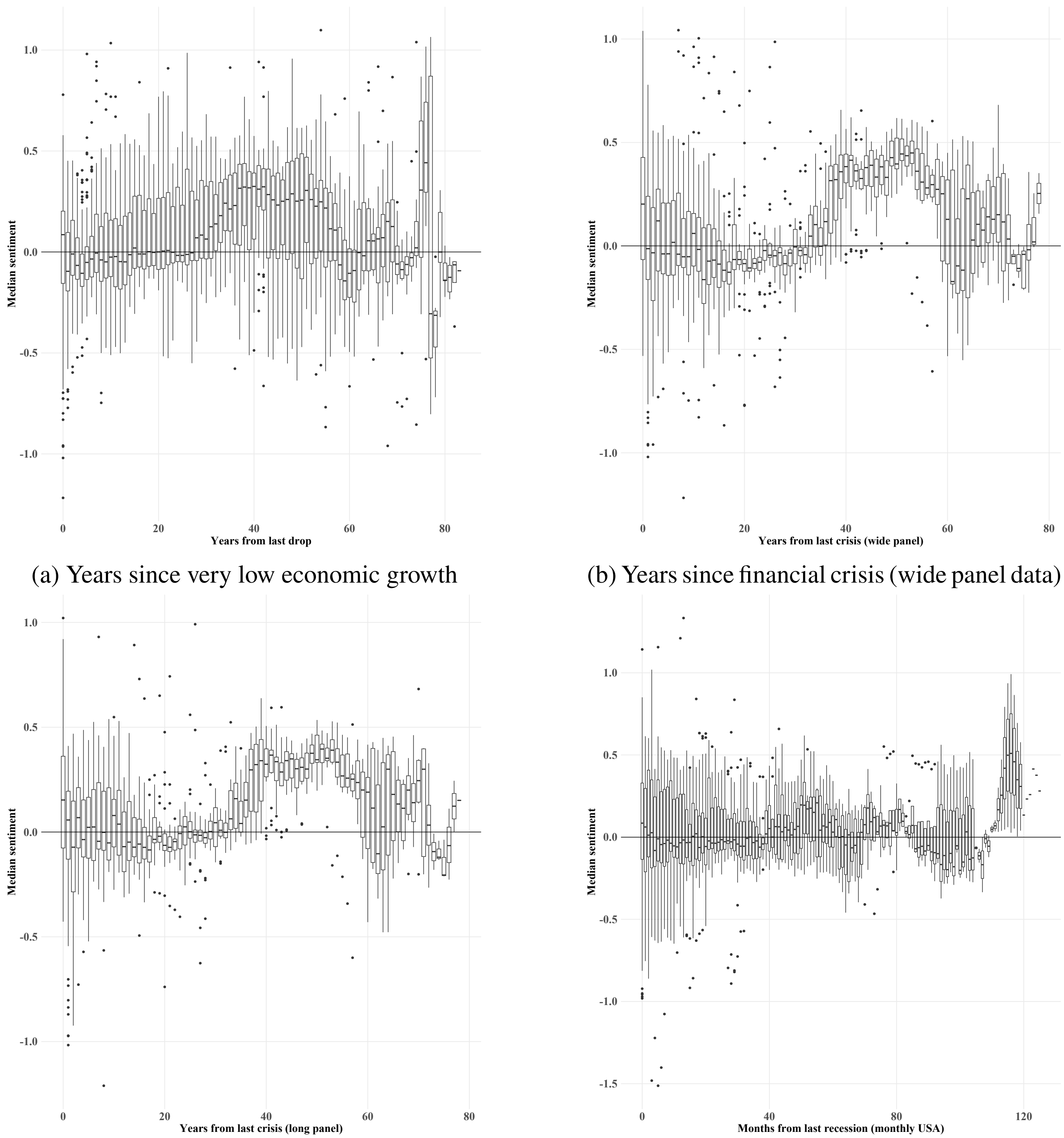
The results of the previous section show that increases in the news-implied component of sentiment predict bad credit booms linked in prior work to financial crises. In the historical sample, this object should be interpreted as the component of country-level sentiment captured by news coverage and anchored to the survey-based PIMS measure estimated from professional forecasts. This still does not explain where this increase in misaggregation comes from. Neither Minsky (1977, 1986) nor 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 documented 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 pessimistic about future returns. Similarly in Malmendier and Nagel (2016), the authors find 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 is 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



(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 13: Sentiment and time since negative economic events.

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 affect on belief formation relative to statistical information. They test the theoretical predictions with survey experiment evidence and find 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 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 13 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 13a-13c 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 13d, 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 empirical 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 panel regressions where the median sentiment and the concentration

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

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 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 financial crisis, 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, there is no strong evidence that time since a crisis or severe economic downturn increases sentiment concentration, although the coefficient is positive and marginally significant in specifications that control for GDP growth and stock returns.

To summarize the empirical results, economic sentiment tends to increase with the time elapsed since a crisis or severe negative economic event, although at a diminishing rate. This pattern suggests that, as the crisis recedes into the past, individuals gradually become less cautious as the memory of the rare event fades and their beliefs diverge more from the objective benchmark implied by machine-based beliefs constructed from public information. At the same time, the estimated quadratic relationship implies that this increase does not continue indefinitely: sentiment eventually reaches a peak and may decline thereafter.

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. A robust empirical finding from historical macro-finance research is that credit growth predicts crises, and more recent evidence shows that credit booms concentrated in the non-tradable corporate sector are especially informative about future financial distress (Müller and Verner, 2024). Even so, much less is known about the forces that generate these "bad" credit booms in the first place.

This paper develops a new way to measure a sentiment component from survey beliefs, defined as the part of professional forecasts that reflects the systematic misaggregation of public information relative to a machine benchmark constructed from the same information set. Because the object is identified from professional forecasts, it should be interpreted as a professional-expectations measure rather than as a direct measure of household or firm beliefs. We show that a large share of variation in beliefs, both across forecasters and over time, is predictable from public information, yet professional forecasts still deviate systematically from machine benchmarks that outperform individual forecasters.

We then extend this measure historically using machine-learning models trained on textual representations of historical news coverage, allowing us to construct a news-implied counterpart of the survey-period sentiment measure for a panel of 78 countries over more than a century. The main historical result is that increases in median sentiment predict credit booms in the non-tradable corporate sector, which prior research has linked to subsequent financial crises. In addition, we find that the time elapsed since the last major crisis and the ratio of young to old people in the population predict increases in this sentiment measure, consistent with memory-related forces playing an important role in belief formation over time.

Taken together, the results provide historical evidence consistent with the Minsky–Kindleberger view that periods of elevated sentiment are linked to crisis-prone credit booms. More broadly, they suggest that memory-related forces help shape the dynamics of misaggregation-driven beliefs and, through them, the build-up of financial fragility.

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 (2022): "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*.

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.

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.[9]

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.[10]

Formally, let $y_0(t)$ denote the current calendar year and $y_1(t) = y_0(t) + 1$ the next calendar year from the perspective of month t. Let $k \in \{1, \ldots, 12\}$ denote the calendar month of the forecast (January = 1, ..., December = 12). Then the interpolated one-year-ahead forecast of forecaster j and the corresponding interpolated realization are approximated by

$$\begin{aligned}
\mathbb{F}_t^j[Y_{t+12}] &\approx \frac{13-k}{12}\,\mathbb{F}_t^j[Y_{y_0(t)}] + \frac{k-1}{12}\,\mathbb{F}_t^j[Y_{y_1(t)}], \\
Y_{t+12} &\approx \frac{13-k}{12}\,Y_{y_0(t)} + \frac{k-1}{12}\,Y_{y_1(t)}.
\end{aligned}$$

Here, $Y_{y_0(t)}$ and $Y_{y_1(t)}$ denote realized real GDP growth in the current and next calendar year, respectively, while $\mathbb{F}_t^j[Y_{y_0(t)}]$ and $\mathbb{F}_t^j[Y_{y_1(t)}]$ are forecaster j's corresponding annual growth forecasts made at time t.

Under this scheme, the interpolated forecast reflects the current-year forecast only in January.

[9] 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.

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

Table 4: Variable definitions, frequencies, data sources, and use in the paper.

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 real 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
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

Note: The table lists the variables used in the survey-based measurement exercise, the monthly local projections, the historical backcasting procedure, and the annual historical panel analysis. Frequencies refer to the frequency at which the variable enters the estimation, and the final column indicates the section in which the variable is used.

In February, the interpolated forecast is a weighted average of 11/12 for the current year and 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.

A.2 Macroeconomic and financial data

In section 2.3, we use 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, inflation, and the interpolated year a head consensus forecasts of real GDP growth. For 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. For the same reason, we exclude the first quarterly lag and include only lags 2 through 12 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.

B. Robustness of the historical backcasting exercise

In this section we report three sets of robustness checks. We first examine lexical and semantic overlap between the historical and survey-period corpora. We then study the stability of the survey-period text-to-sentiment mapping across alternative seeds, country composition, and placebo specifications. Finally, we assess the sensitivity of the historical local-projection results to excluding individual countries, restricting the time period, and trimming extreme sentiment observations.

B.1 Language drift and support in the historical text corpus

A natural concern in the historical backcasting exercise is that the language of news may evolve over time, weakening the connection between the historical text corpus and the survey-period mapping used to estimate sentiment. To assess this concern, this appendix reports several robustness checks. We examine lexical and semantic overlap between the historical and survey-period corpora.

Table 5: Language-drift diagnostics by decade.

Decade	In-vocab tok.	Stable tok.	Mean BERT	Med. BERT	P90 BERT
1890	96.4	85.0	0.071	0.066	0.112
1900	96.3	83.6	0.040	0.021	0.076
1910	96.4	84.8	0.018	0.015	0.027
1920	95.4	83.3	0.028	0.027	0.041
1930	95.9	84.5	0.013	0.008	0.028
1940	97.0	87.5	0.006	0.006	0.007
1950	97.1	87.7	0.005	0.004	0.006
1960	96.7	87.4	0.003	0.003	0.004
1970	97.0	87.9	0.003	0.003	0.003
1980	98.1	89.1	0.002	0.001	0.003
1990	98.8	90.5	0.000	0.000	0.000

Note: *In-vocab tok.* reports the share of historical token mass covered by the survey-period vocabulary. *Stable tok.* reports the share of token mass retained when the historical corpus is restricted to a stable vocabulary defined by sufficiently frequent survey-period words. *Mean BERT*, *Med. BERT*, and *P90 BERT* report the mean, median, and 90th percentile of the cosine distance between each historical month and its nearest survey-period month in BERT embedding space. Lower values indicate greater semantic similarity. The embedding-based diagnostics begin in 1890.

Table 5 provides two complementary diagnostics of temporal language drift. First, lexical overlap between the historical and survey-period corpora is high throughout the sample: roughly 95–99% of token mass in each decade is covered by the survey-period vocabulary, and about 83–91% remains when the corpus is restricted to a stable vocabulary defined from sufficiently frequent survey-period words. This suggests that the historical corpus continues to rely heavily on a lexical core that is also present in the survey period.

Second, the embedding-based diagnostics indicate that historical news months remain semantically close to at least some survey-period months, although this similarity declines gradually further back in time. The nearest-neighbour BERT distances are largest in the earliest decades and fall steadily toward zero as one approaches the survey period. Taken together, these results suggest that language drift is present, but does not imply a collapse in the overlap between the historical and survey-period news environments.

B.2 Stability of the survey-period mapping

The next set of robustness checks examines whether the survey-period mapping itself is sensitive to alternative estimation choices. Table 6 shows that the out-of-bag fit of the random forest is essentially unchanged across 100 alternative random seeds. Both median sentiment and sentiment dispersion display nearly identical average out-of-bag R^2 and mean squared error across seeds relative to the baseline specification. This indicates that the estimated mapping is highly stable with respect to stochastic variation in forest construction.

Table 6: Random-seed robustness of the survey-period random forest fit.

	OOB R^2			OOB MSE		
Dependent	Baseline	Mean seed	SD seed	Baseline	Mean seed	SD seed
Median sentiment	0.565	0.564	0.002	0.128	0.128	0.001
Sentiment dispersion	0.368	0.368	0.003	0.273	0.273	0.001

Note: The table reports baseline out-of-bag (OOB) fit for the survey-period random forest models and the corresponding mean and standard deviation across 100 alternative random seeds, holding the baseline hyperparameters fixed. The close similarity between the baseline values and the averages across seeds indicates that the first-stage text-to-sentiment mapping is highly stable with respect to stochastic variation in forest construction.

A related concern is that, because the historical text corpus is based on the *Wall Street Journal*, the estimated mapping may fit larger and more internationally salient countries better than smaller countries. Figure 14 examines this issue for median sentiment by plotting country-level out-of-bag root mean squared error against the number of available country-month observations in the survey period. The figure shows that countries with longer survey histories tend to exhibit lower prediction error. The negative correlation between RMSE and sample length is -0.588, while the corresponding correlation between the actual-predicted correlation and sample length is 0.692. Thus, although predictive performance is somewhat stronger for larger core countries, a substantial part of the cross-country heterogeneity in fit appears to reflect differences in sample length rather than only country salience.

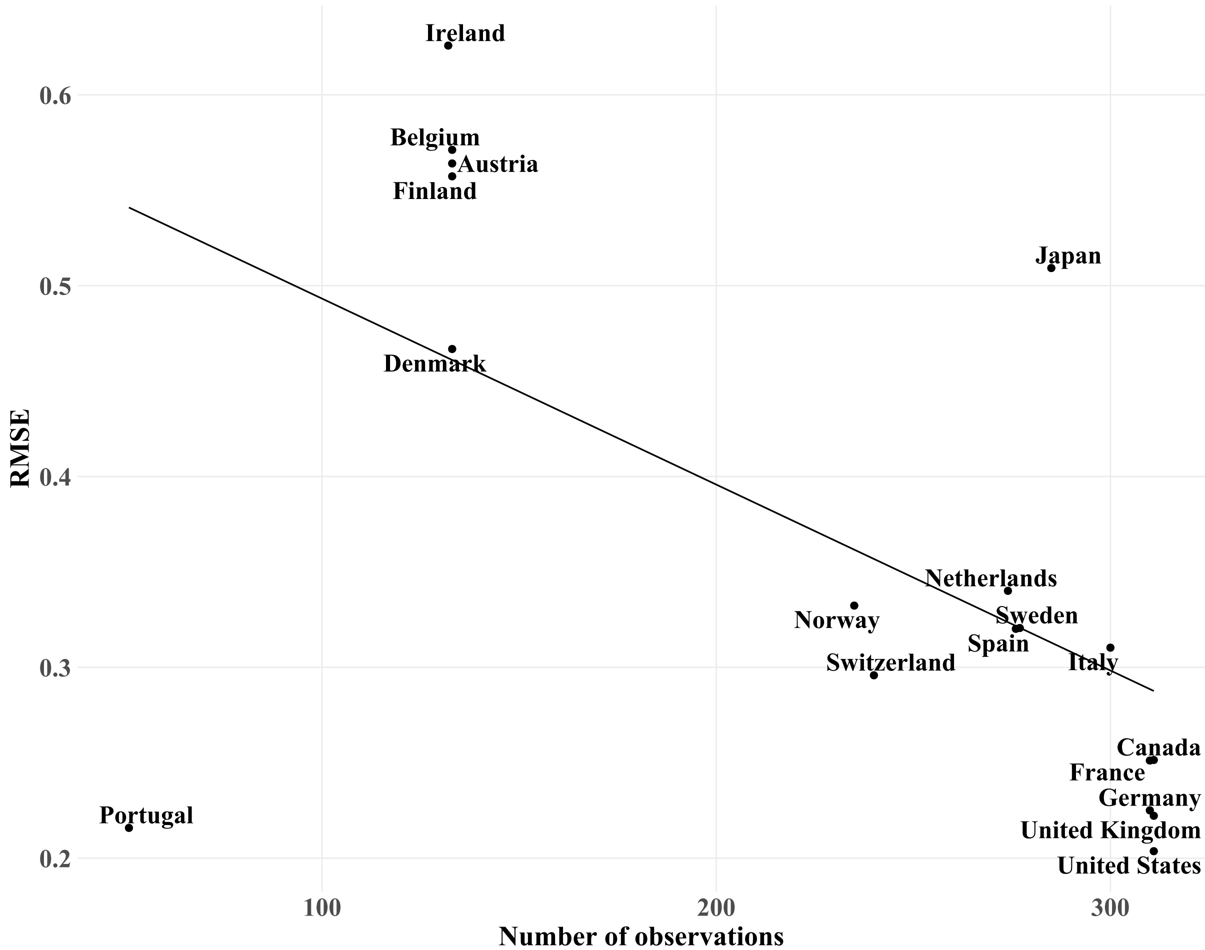


Figure 14: Country-level out-of-bag predictive fit and sample length for median sentiment in the survey period.

Note: The figure plots, for each country, the out-of-bag root mean squared error (RMSE) for median sentiment in the survey period against the number of available country-month observations. The solid line shows the fitted linear relationship. The correlation between RMSE and number of observations is -0.588. Countries with longer survey histories tend to exhibit lower out-of-bag prediction error, suggesting that part of the cross-country heterogeneity in predictive fit is related to sample length.

Finally, we report a falsification exercise designed to test whether the estimated mapping is extracting genuine information from the news-based predictors rather than merely generating spurious fit. In the placebo exercise, the sentiment target is randomly permuted across observations while the predictors and baseline hyperparameters are held fixed. Figure 15 shows that the distribution of placebo out-of-bag R^2 is far below the baseline value for both sentiment measures, and Table 7 shows that placebo performance deteriorates sharply in both R^2 and mean squared error. In particular, the placebo models deliver negative average out-of-bag R^2 for both median sentiment and sentiment dispersion, while the baseline models fit the data substantially better. This provides strong evidence that the estimated mapping is capturing meaningful information in the news-based features.

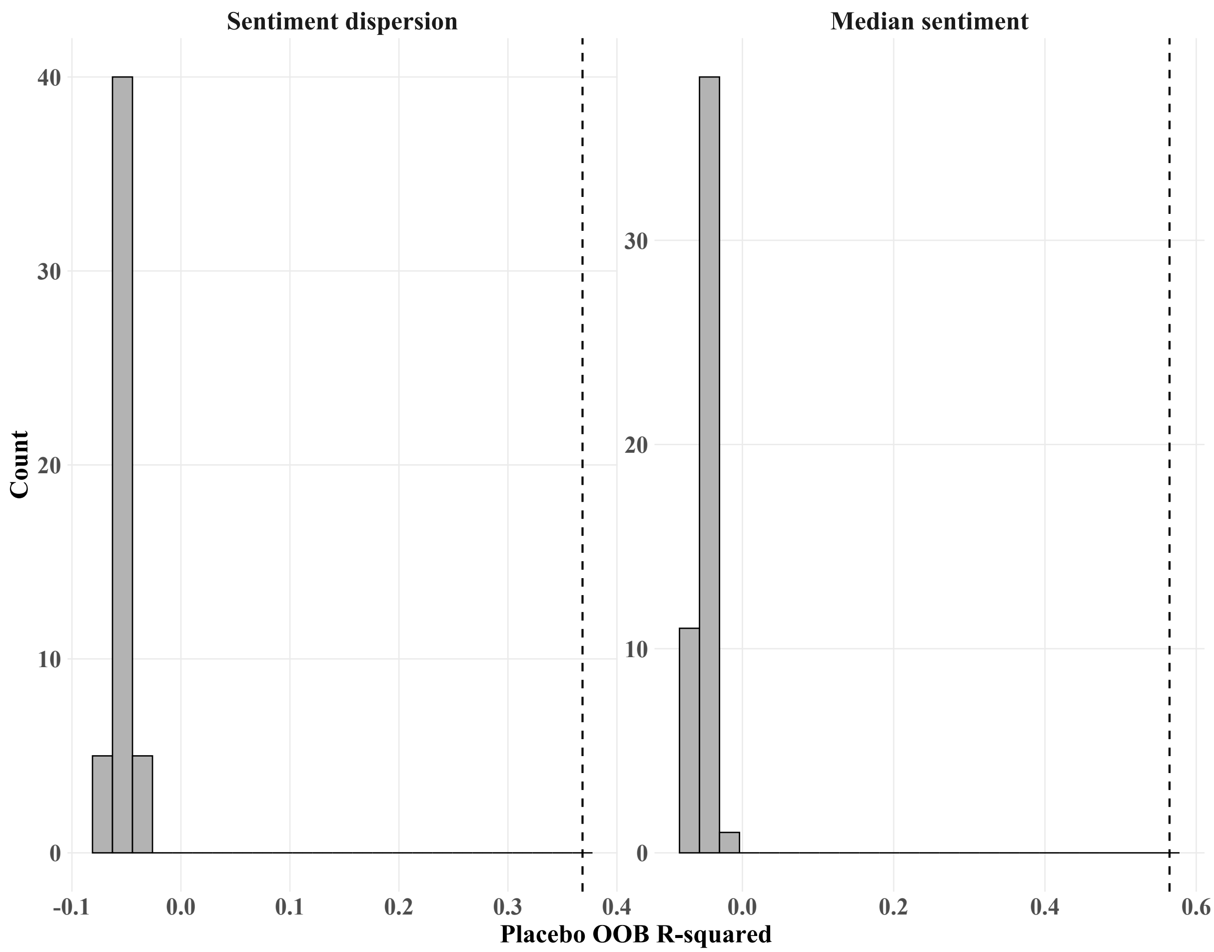


Figure 15: Placebo distribution of out-of-bag R^2 for the survey-period random forest models.

Note: The histograms show out-of-bag R^2 from 50 placebo models estimated after randomly permuting the sentiment target across observations. The predictors and baseline hyperparameters are held fixed. The dashed line marks the out-of-bag R^2 of the corresponding baseline model. Placebo fit is substantially worse than baseline fit for both sentiment measures.

Taken together, these results suggest that the historical backcasting exercise is supported by substantial lexical and semantic continuity between the historical and survey-period corpora, that the survey-period mapping is stable across alternative random seeds, that predictive performance is not confined solely to the largest countries, and that the random forest extracts genuine information from the news-based predictors rather than spurious structure generated by the estimation procedure. The remaining robustness checks therefore focus on whether the historical

Table 7: Placebo falsification test for the survey-period random forest models.

	OOB MSE				OOB R^2			
Dependent	Baseline	Mean placebo	SD placebo	p-value	Baseline	Mean placebo	SD placebo	p-value
Median sentiment	0.13	0.31	0.00	0.00	0.56	-0.05	0.01	0.00
Sentiment dispersion	0.27	0.45	0.00	0.00	0.37	-0.05	0.01	0.00

Note: The 50 placebo models are estimated after randomly permuting the sentiment target across observations while holding the predictors and baseline hyperparameters fixed. The table reports baseline out-of-bag (OOB) fit, the mean and standard deviation of placebo OOB fit, and empirical p-values equal to the fraction of placebo replications that weakly outperform the baseline model. Lower placebo R^2 and higher placebo mean squared error indicate that the baseline mapping captures genuine information in the news-based predictors rather than spurious machine-learning fit.

local-projection results remain similar when the earliest decades or the weakest-fit countries are excluded.

B.3 Robustness of the historical local projections

The leave-one-country-out exercise in Figure 16 shows that the historical local-projection estimates are generally not driven by any single country. For each horizon and response variable, the shaded bands report the range of coefficients obtained when the specification is re-estimated after excluding one country at a time. These bands are narrow for most horizons, indicating that the baseline estimates are highly stable to individual country exclusions. In a small number of cases the bands widen, implying that some countries are more influential for coefficient magnitudes than others, but the overall qualitative response patterns remain intact.

Table 8 further examines the sensitivity of the historical local projections to broader sample restrictions. We consider four alternative perturbations of the baseline specification: excluding all observations prior to 1950, excluding the six countries with the weakest survey-period predictive fit, imposing both restrictions simultaneously, and winsorizing the top and bottom 1% of the sentiment distribution. Across these alternative samples, the strongest evidence continues to concern household credit and the short-run response of non-tradable to tradable credit. In particular, the first-year response of non-tradable to tradable credit remains positive across specifications, while the household-credit response remains positive throughout and is statistically significant at short horizons in several cases.

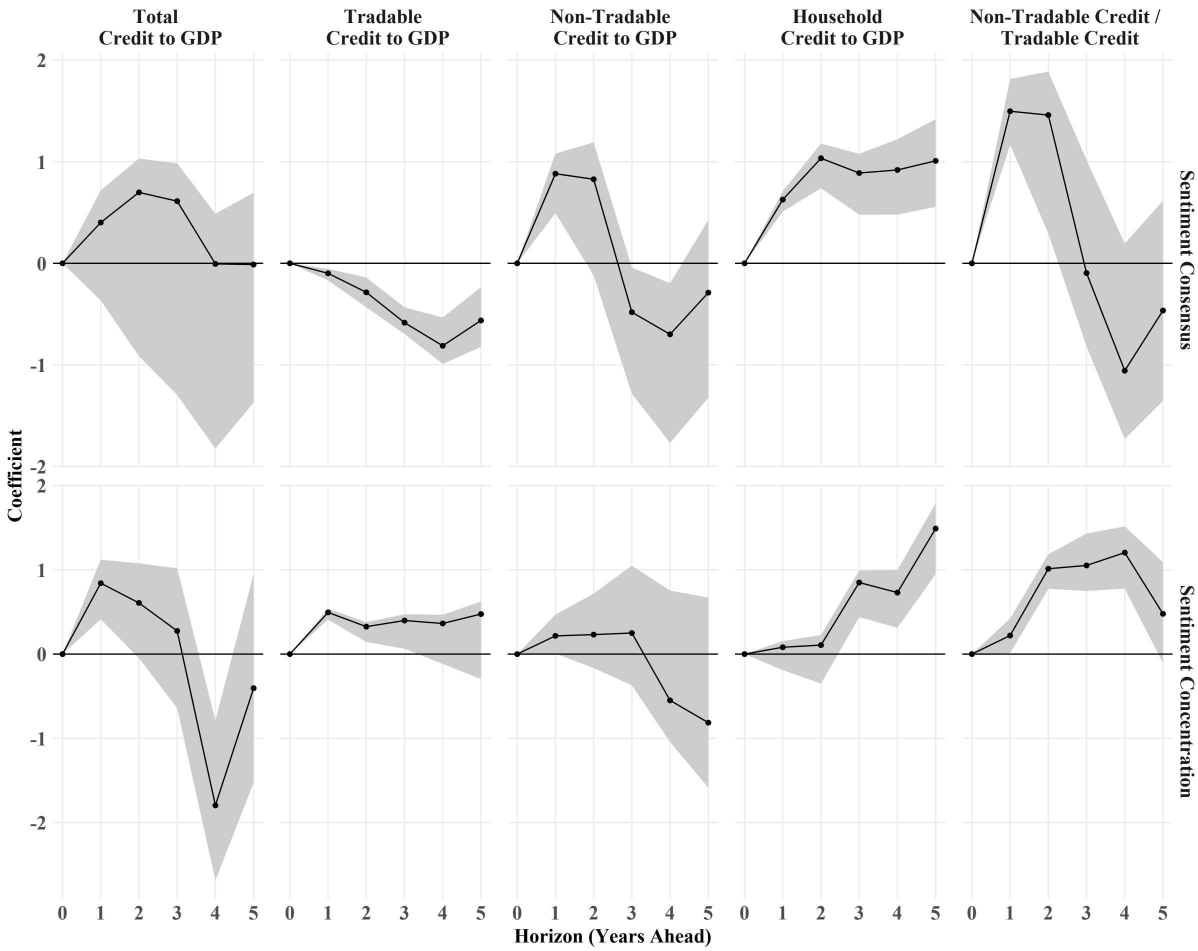


Figure 16: Leave-one-country-out robustness of the historical local projections.

Note: The figure reports the baseline local projection coefficients together with shaded bands constructed from repeated re-estimation of the specification after excluding one country at a time from the sample. The sample contains 78 countries, so the bands are based on 78 leave-one-country-out estimations for each horizon and response variable. The shaded areas span the minimum and maximum coefficient values across these re-estimations. The narrowness of the bands for most horizons indicates that the results are not driven by any single country, while the wider bands at some horizons reflect a limited number of more influential country exclusions.

Taken together, these results suggest that the main historical findings are not solely driven by the earliest decades, by the countries for which the survey-period mapping is weakest, or by a small number of extreme sentiment observations. At the same time, the robustness checks also indicate that the strongest and most stable evidence concerns the sectoral composition of credit rather than aggregate credit alone, and especially the short-run response of non-tradable to tradable credit and household credit.

Table 8: Robustness of historical local-projection coefficients across restricted samples.

outcome	horizon	Baseline	Post-1950	Exclude weak-fit	Post-1950 + exclude weak-fit	Winsorize 1% extremes
Household Credit to GDP	1	0.627**	0.708**	0.836**	0.935***	0.703**
Household Credit to GDP	2	1.034*	1.101*	1.036*	1.094*	1.050
Household Credit to GDP	3	0.889	0.938	0.733	0.753	0.904
Household Credit to GDP	4	0.919	1.032	1.473	1.587	1.008
Household Credit to GDP	5	1.009	0.864	2.243	1.997	0.871
Non-Tradable Credit to GDP	1	0.882***	0.914***	0.425	0.443	1.033***
Non-Tradable Credit to GDP	2	0.827	0.862	-0.573	-0.573	0.989
Non-Tradable Credit to GDP	3	-0.481	-0.464	-1.802*	-1.795*	-0.355
Non-Tradable Credit to GDP	4	-0.699	-0.704	-2.337*	-2.346*	-0.701
Non-Tradable Credit to GDP	5	-0.289	-0.312	-2.121	-2.169*	-0.578
Non-Tradable Credit / Tradable Credit	1	1.497**	1.536**	2.029**	2.076**	2.114***
Non-Tradable Credit / Tradable Credit	2	1.459	1.503	0.894	0.923	1.999**
Non-Tradable Credit / Tradable Credit	3	-0.096	-0.017	-0.428	-0.334	0.876
Non-Tradable Credit / Tradable Credit	4	-1.057	-0.995	-0.928	-0.858	-0.318
Non-Tradable Credit / Tradable Credit	5	-0.465	-0.443	-0.220	-0.201	-0.052
Total Credit to GDP	1	0.402	0.829	-0.496	-0.462	0.068
Total Credit to GDP	2	0.698	0.868	-1.158	-1.391	0.481
Total Credit to GDP	3	0.612	0.017	-2.196	-2.709	0.615
Total Credit to GDP	4	-0.006	-1.279	-2.536	-2.988	-0.182
Total Credit to GDP	5	-0.012	-1.228	-1.307	-2.128	-0.514
Tradable Credit to GDP	1	-0.098	-0.083	-0.109	-0.121	-0.168
Tradable Credit to GDP	2	-0.286	-0.275	-0.428	-0.458	-0.288
Tradable Credit to GDP	3	-0.584	-0.585	-0.794*	-0.846*	-0.700
Tradable Credit to GDP	4	-0.812	-0.823	-1.127*	-1.187*	-0.989
Tradable Credit to GDP	5	-0.563	-0.632	-0.714	-0.846	-0.903

Note: The table reports historical local-projection coefficients for the response of credit variables to shocks in median sentiment across five alternative samples: the baseline sample, a post-1950 sample, a sample excluding the six countries with the weakest survey-period predictive fit, a sample imposing both restrictions simultaneously, and a sample in which the top and bottom 1% of the sentiment distribution are winsorized. Coefficients are shown for horizons 1 through 5. Standard errors are Driscoll–Kraay. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The table is intended to assess whether the main historical credit-boom findings are driven by the earliest decades, weakly fit countries, or extreme sentiment realizations.

Bank of Finland Research Discussion Papers 2026

ISSN 1456-6184, online

1/2026 Fabio Verona
Forecasting inflation: The sum of the cycles outperforms the whole

2/2026 Gene Ambrocio – Dien Giau Bui – Iftekhar Hasan – Chih-Yung Lin
Pyrrhic diversification:Foreign institutional ownership and stock return sensitivity to the global financial cycle

3/2026 Kim Ristolainen
Quantifying Minsky Cycles