FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Do Vibes Predict Recessions? Evidence from a Big-Data Forecasting Framework

Nicolas Petrosky-Nadeau, Yeji Sung, and Daniel J. Wilson
Federal Reserve Bank of San Francisco

July 2026

Working Paper 2026-14

https://doi.org/10.24148/wp2026-14

Suggested citation:

Petrosky-Nadeau, Nicolas, Yeji Sung, and Daniel J. Wilson. 2026. "Do Vibes Predict Recessions? Evidence from a Big-Data Forecasting Framework." Federal Reserve Bank of San Francisco Working Paper 2026-14. https://doi.org/10.24148/wp2026-14

Do Vibes Predict Recessions? Evidence from a Big-Data Forecasting Framework*

Nicolas Petrosky-Nadeau† Yeji Sung‡ Daniel J. Wilson§

July 8, 2026

Abstract

Measures of beliefs, sentiment, and narratives often send recession signals that differ from those in hard data, defined as conventional economic and financial indicators. Using a real-time forecasting framework, we compare how soft and hard data predict recessions from one to twelve months ahead. Forecasts based on soft data are more responsive to rising recession risk: they identify more downturns, but also produce more false alarms. Even with far fewer inputs, soft-data forecasts remain competitive with hard-data forecasts out of sample, especially at shorter horizons. Combining hard and soft data often improves forecast performance, suggesting that the two types of information are useful complements.

JEL classification: C53, C55, E32

Keywords: Recession Forecasting, Big Data, Machine Learning, Soft Data

*The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System. We thank Shane Boyle, Taerin Kim, and Rami Najjar for excellent research assistance.

†Federal Reserve Bank of San Francisco, Nicolas.Petrosky-Nadeau@sf.frb.org

‡Federal Reserve Bank of San Francisco, Yeji.Sung@sf.frb.org

§Federal Reserve Bank of San Francisco, Daniel.Wilson@sf.frb.org

1 Introduction

The term "vibecession" entered economic commentary in mid-2022, when consumer sentiment fell to historically low levels and measures of uncertainty spiked, even as the labor market remained strong and GDP continued to grow. The term has since become shorthand for a broader gap between conventional economic data and "soft" data such as sentiment and uncertainty. That gap raises a difficult question for recession forecasting: when soft data deteriorate sharply while conventional indicators remain resilient, should we view soft data as noise, or as an early warning?

This question is especially difficult because recessions are hard to identify and foresee in real time. Policymakers and forecasters often cannot determine whether a recession is approaching, whether one has already begun, or how long it will last. Since 1979, when the NBER began publicly announcing business-cycle peaks and troughs, the committee has taken about seven months on average to announce the start of a recession and about fifteen months to announce its end. These delays matter because forecasters must assess recession risk before official recession dates are announced and before many hard indicators are fully available. Soft data, such as surveys, business reports, and news coverage, may therefore reveal deteriorating expectations or perceived economic stress before it appears clearly in hard data.

In this paper, we systematically evaluate the forecasting content of soft data for recession prediction. Using real-time data from August 1999 through May 2026, we compare out-of-sample recession-risk forecasts based on two types of predictors: hard indicators that measure conventional macroeconomic and financial conditions, and soft indicators that measure beliefs, sentiment, and narratives. We ask whether a relatively small set of soft indicators predicts recessions in real time, both on its own and beyond a much larger panel of hard macroeconomic and financial indicators. We also study how a model using both types of data interprets periods when soft and hard indicators send different signals.

We find that hard and soft data have often pointed to different levels of recession risk, long before the term "vibecession" became popular. Soft-data forecasts are more sensitive to downturn risk: they identify more months followed by a recession within the forecast horizon than do hard-data forecasts, but they also generate more false alarms during expansions. Even so, forecasts based on soft data alone match or improve on forecasts based only on hard data, especially at horizons up to six months ahead. Combining hard and soft data can further improve performance, and Shapley decompositions show why: hard-data categories reinforce soft-data warnings before some recessions but dampen them during expansions when economic conditions do not support elevated recession risk.

We begin by building a real-time recession-forecasting dataset that separates hard indicators, which capture realized economic outcomes and financial prices, from soft indicators, which capture beliefs, sentiment, uncertainty, and narratives about economic conditions. The hard-data panel draws on a much larger set of conventional macroeconomic and financial indicators from FRED-MD real-time vintages, while the soft-data panel contains a smaller set of survey and text-based measures. We group the hard indicators into five categories: real activity, housing, the yield curve, other financial indicators, and inflation. This structure lets us compare recession-risk signals from traditional macroeconomic and financial data with signals from beliefs and narratives.

Recession forecasting is vulnerable to overfitting because the predictor set is large and recessions are rare. We address this problem first by reducing the dimension of the predictor set before estimation. For each hard-data category, we construct factors that summarize common movements among related indicators. We construct factors for the soft-data category in the same way. This approach preserves broad information from each group while preventing the model from relying on many individual series.

We then estimate recession logit models using the elastic-net regularized regression estimator. Elastic net regularization shrinks small coefficients and helps select among correlated predictors, reducing the risk that the model fits patterns from past recessions that may not repeat. The resulting specification preserves information from a broad set of hard and soft indicators while keeping the forecasting model parsimonious.

We generate forecasts in a fully real-time environment. At each forecast date, we re-run the full forecasting procedure, including data construction, factor extraction, hyperparameter selection, and estimation, using only information available at that time. This design matters because many hard-data indicators are released with delays and later revised. Forecasts based on the latest available vintage can therefore use information that real-time forecasters did not have. By restricting each step of the exercise to information available at the forecast date, our framework provides a more realistic assessment of how hard and soft data perform in practice.

We first document that hard and soft data often give different assessments of recession risk. We compare recession probabilities from two elastic-net models estimated using the same real-time forecasting framework: one based only on hard data and the other based only on soft data. The soft-data model is more sensitive to downturn risk. The soft-data model is more sensitive to downturn risk: it raises recession warnings more often ahead of downturns, including around the 2001 recession and in late 2007, but it also generates more false alarms during expansions. In other words, soft data cast a wider net, while hard-data recession signals are more selective.

We then examine how the combined model uses hard and soft data when the two sources send different signals. The combined model does not simply average the hard-data and soft-data forecasts. Instead, it tends to preserve elevated soft-data recession risk when hard-data categories also point to economic weakness, and to mute those warnings when hard data provide little corroboration. This pattern suggests that hard data help distinguish soft-data warnings that coincide with broader economic deterioration from those that remain more isolated.

We next ask whether this balancing improves forecast performance. Because forecast quality can be measured in several ways, we evaluate the models using metrics that capture different aspects of performance: how well recession warnings balance missed recessions against false alarms, how well the model ranks high-risk months, and how well it separates recession-risk periods from expansions more broadly. Although the soft-data model uses far fewer predictors than the hard-data model, it remains competitive out of sample, especially at shorter horizons. Adding soft data to hard data also improves recession-warning performance. These results suggest that soft data provide information about recession risk that is not fully captured by conventional macroeconomic and financial indicators.

These results raise the question of why soft data improve recession forecasts. One possibility is that soft data help because hard data are difficult to use in real time: many hard indicators are released with lags, and initial releases are noisy and later revised. Another possibility is that soft data contain recession-relevant information that hard data do not fully capture, even after delayed observations and revisions are incorporated.

To separate these channels, we compare hard-data models constructed from three vintages: the real-time vintage, a next-month vintage that includes hard-data observations released after the forecast date, and the latest vintage that incorporates subsequent revisions. Hard-data forecasts improve when these later vintages are used, confirming that publication lags and initial-release noise hamper real-time hard-data forecasts. Yet soft data continue to add forecasting power even when combined with these more informative hard-data benchmarks. Soft data therefore both proxy for information that later appears in hard-data releases and contain distinct information about recession risk.

Finally, we use historical episodes to study how the combined model weighs recession signals from hard and soft data when they point in different directions. To do so, we separate the hard-data contribution into its underlying categories rather than treating it as a single group. Using Shapley decompositions, we attribute the model's predicted recession risk to soft data and to each hard-data category. This shows whether elevated recession risk reflects broad corroboration across hard indicators, or instead comes mainly from soft-data weakness that remains isolated.

Hard data play an important role in this distinction, but not all categories contribute in the same way. The decompositions show that the hard-data categories that reinforce or offset elevated recession risk from the soft-data model vary across episodes. Thus, the value of combining hard and soft data is not tied to a single hard-data source. It comes from using the broader hard-data panel to interpret when soft-data weakness coincides with broader economic deterioration.

The rest of the paper proceeds as follows. Section 2 reviews related literature, and Section 3 describes the data and forecasting framework. Section 4 compares recession risks implied by hard and soft data. Section 5 examines how the combined model balances hard- and soft-data recession signals and evaluates forecast performance. Section 6 examines why soft data improve forecasts. Section 7 uses Shapley decompositions to trace the sources of its forecasts across historical episodes. Section 8 reports robustness exercises, and Section 9 concludes.

2 Related Literature

A long recession-forecasting literature has identified useful predictors of downturns, with the yield curve serving as the canonical example since Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998).[1] Related work has examined credit, housing, labor-market, financial, and sentiment indicators, often in settings that focus on individual predictors or small predictor sets.[2] We build on this literature but ask a different question: in a high-dimensional forecasting environment, how much recession information do soft data add beyond a rich set of hard-data indicators? We answer this question by comparing forecasts constructed from different information sets, asking how predictions change when soft indicators are added to hard indicators.

A closely related strand of recent literature treats recession assessment as a real-time probability problem. Furno and Giannone (2024), in particular, develop a framework for nowcasting recession risk using a parsimonious combination of macroeconomic and financial conditions. Their work is close to ours in emphasizing that recession risk is best assessed by combining multiple timely signals, and that any single signal can be misleading in isolation. Our paper differs in both the forecasting object and the information set. Whereas their framework focuses on nowcasting current recession risk with a small set of

[1] More recent analyses of the yield curve's recession forecasting power include Estrella 2005, Wright 2006, Chauvet and Potter 2005, Ang et al. 2006, Rudebusch and Williams 2009, and Liu and Moench 2016.

[2] See also Sahm (2019) and Michaillat (2025) as examples of studies constructing single labor market indicators for real-time recession nowcasting.

indicators, we study recession forecasting across various horizons using a broader high-dimensional information set. This allows us to focus on a different question: whether broad classes of soft data contain recession information beyond what is already captured by hard data.

Our methodology most directly builds on the data-rich forecasting framework of Bianchi et al. (2022), who combine dimension reduction with regularized regression in a real-time forecasting environment.[3] We adapt this approach to recession forecasting, where the outcome is binary and infrequent, and where the information set can be divided into economically distinct hard and soft categories.

Our paper also relates to other work applying machine-learning methods to recession prediction, including Davig and Hall (2019) and Vrontos et al. (2021). In addition, Bluwstein et al. (2023) study a related rare-event early-warning problem, using machine-learning methods to predict financial crises and Shapley-value decompositions to connect predicted crisis risk to economic drivers. These papers motivate our use of a data-rich forecasting framework and our emphasis on interpretation. In our setting, Shapley-value decompositions allow us to assess how hard and soft indicators contribute to predicted recession risk within a high-dimensional model.

The focus on soft data also connects our paper to work on whether sentiment, expectations, and other subjective indicators predict real economic activity. Studies of consumer sentiment ask whether confidence measures help forecast household spending, although their incremental value beyond other indicators is often limited (Ludvigson, 2004; Lahiri et al., 2016). The nowcasting literature gives soft data a different role: because official activity measures are released with delay, timely surveys and expectations can improve assessments of current and near-term macroeconomic conditions (Giannone et al., 2008; Bańbura and Rünstler, 2011; Schnatz and D'Agostino, 2012). This evidence motivates our use of soft indicators to study whether soft information improves recession forecasts beyond a rich set of hard indicators observed in real time.

3 Real-Time Data and Forecasting Framework

This section describes the data and methodology used to construct real-time out-of-sample recession probability forecasts. We first discuss the source data used in the forecasting exercise. We then describe the forecasting approach, including how we use real-time information to estimate recession probabilities. This framework provides the basis for the

[3]Earlier studies on macroeconomic forecasting in big-data environments include Stock and Watson (2014), Medeiros et al. (2021), and Goulet Coulombe et al. (2022).

model comparisons in the following sections.

3.1 Real-Time Big Data

We distinguish between two types of predictors: hard data and soft data. We refer to data as "hard" when they report realized economic quantities or prices, such as output, employment, inflation, interest rates, and asset prices. We refer to data as "soft" when they summarize beliefs, perceptions, and narratives about the economy, such as survey expectations and text-based sentiment indicators.

A central objective of the data construction is to mimic the information set available to forecasters in real time. We therefore use vintage-specific data and avoid using later data revisions or future observations when constructing forecasts. This is especially important for hard macroeconomic indicators, which are often released with publication lags and revised over time. Using revised data would overstate the information available to a forecaster at the time the prediction was made.

Our hard-data predictors come from the FRED-MD monthly real-time vintage database (McCracken and Ng, 2016), which contains over 130 U.S. macroeconomic and financial series, including series originally observed at daily, weekly, and monthly frequencies. The real-time vintages begin in August 1999.[4] Each vintage corresponds to the information available at the end of a given month, which allows us to construct forecasts using only data that would have been available at the forecast date.

Working with real-time vintages also requires addressing the so-called "ragged edge" problem, which arises because variables differ in frequency and publication lag (Wallis, 1986). Daily financial variables are typically available through the end of the month. Weekly indicators may be available for only part of the month, while many monthly macroeconomic indicators are not yet released for the current reference month. We treat these two cases differently. For weekly series, we fill in missing within-month observations through the end of the month using univariate ARMA forecasts within each real-time vintage, as detailed in Online Appendix A.2. For monthly series that have not yet been released because of publication lags, we do not fill in the current-month value.

Our primary soft-data predictors draw on four sources. Three are text-based measures that quantify the tone of economic narratives and the level of economic uncertainty. The Economic Policy Uncertainty (EPU) index tracks policy-related uncertainty reflected in newspaper coverage (Baker et al., 2016). The Daily News Sentiment Index (DNSI) from the Federal Reserve Bank of San Francisco measures the tone of economic news coverage

[4] See Online Appendix Table A1 for the full list of variables.

(Shapiro et al., 2022). A Beige Book sentiment index converts qualitative assessments from Federal Reserve District reports into an indicator of economic conditions (Filippou et al., 2024).

The fourth source is the University of Michigan Survey of Consumers, which provides a real-time read on households' views about current and expected economic conditions. We include five subcomponents of the Consumer Sentiment Index, capturing responses to questions about current and expected personal finances, expected business conditions one and five years ahead, and current durable goods buying conditions. We also include the median household expected change in the unemployment rate over the next year (Carroll, 2003). Together, these soft data series cover perspectives on the economy from journalists, businesses, and households.

The real-time availability of the soft-data variables is source-specific, so the soft-data panel requires additional timing assumptions beyond those used for FRED-MD. We describe these source-specific availability rules in Online Appendix A.1 and assess the importance of the text-based indicators in Section 8.1.

3.2 Forecasting Recession Probabilities

Our goal is to forecast the probability that the economy will be in a recession, as defined by the NBER, at any point over a horizon of h months, using information available during month t. We date recessions as beginning in the month after the NBER business-cycle peak and ending in the NBER trough month.

For each horizon h, we define the forecast target $y_{t,h}$, a binary outcome that equals one if the economy is in recession in any of the following h months, and equals zero otherwise. For example, for $h = 6$, $y_{t,h} = 1$ if the economy is in recession in any month from $t + 1$ through $t + 6$.

Let $p_{t,h} = \Pr(y_{t,h} = 1 \mid X_{t,h})$, where $X_{t,h}$ is a horizon-specific predictor vector, constructed as described in the following subsection, that summarizes the high-dimensional real-time hard and soft data. We model the forecast target as

$$y_{t,h} \mid X_{t,h} \sim \text{Bernoulli}(p_{t,h}).$$

Thus, the conditional probability mass function is

$$\Pr(y_{t,h} \mid X_{t,h}) = p_{t,h}^{y_{t,h}} \left(1 - p_{t,h}\right)^{1-y_{t,h}}.$$

We model $p_{t,h}$ using a logit specification in which the log-odds of recession is linear in

predictors. Specifically, we assume

$$\log\left(\frac{p_{t,h}}{1-p_{t,h}}\right) = X'_{t,h}\,\beta_h, \tag{1}$$

where β_h is horizon-specific. We estimate the model by maximizing the Bernoulli log-likelihood, treating $\{y_{t,h}\}_t$ as conditionally independent given the predictors $X_{t,h}$.

Because the predictor vector is high-dimensional, we use a regularized binary logit model. For each horizon h, $\hat{\beta}_h$ maximizes the sample log-likelihood net of an elastic net penalty. The elastic net combines an ℓ_1 penalty, as in LASSO, with an ℓ_2 penalty, as in ridge regression. The ℓ_1 penalty performs variable selection by setting some coefficients exactly to zero, while the ℓ_2 penalty shrinks coefficients continuously toward zero and stabilizes estimates when predictors are correlated. The elastic net thus allows both sparsity and shrinkage. Formally, the estimator solves

$$\hat{\beta}_h = \arg\max_{\beta_h}\left\{\sum_t\Big[y_{t,h}\log p_{t,h} + (1-y_{t,h})\log(1-p_{t,h})\Big] - \lambda_h\left(\alpha_h\|\beta_h\|_1 + \frac{1-\alpha_h}{2}\|\beta_h\|_2^2\right)\right\},$$

The elastic net has two horizon-specific hyperparameters: $\lambda_h \geq 0$, which controls the overall strength of regularization, and $\alpha_h \in [0,1]$, which controls the mix between ridge ($\alpha = 0$) and LASSO ($\alpha = 1$). We select both hyperparameters using the real-time validation procedure described in Section 3.4.

3.3 Constructing Real-Time Predictors

We next describe how we transform the real-time data into a horizon-specific predictor vector, $X_{t,h}$. The predictor set is high-dimensional, so using the individual series directly would introduce substantial risk of overfitting. At the same time, we want the forecasting model to remain interpretable rather than operate as a black box. We therefore summarize the data within broad categories that have natural economic interpretations.

We construct these summaries separately within six data groups. Five groups summarize hard-data indicators from FRED-MD: *real activity, housing, the yield curve, other financial indicators,* and *inflation*. Online Appendix Table A1 reports the assignment of FRED-MD variables to these groups and shows how each variable is transformed before entering the model. The sixth group summarizes the soft-data measures described above. For each soft-data index, we include both its level and its change relative to a year earlier. This grouping allows the model to use a large information set while still distinguishing recession signals that come from different parts of the economy.

Within each group, we use diffusion-index factors to summarize the information in the underlying series.[5] Following Bianchi et al. (2022), we estimate these factors using the scaled PCA approach of Huang et al. (2022). The purpose of scaled PCA is to combine dimension reduction with predictive relevance. Before extracting principal components, each predictor is scaled by its univariate predictive slope for $y_{t,h}$. At each forecast month, we estimate these slopes using only predictor-outcome pairs for which the recession outcome is already observed. Predictors with stronger univariate forecasting power therefore receive greater weight than they would under standard PCA.

Let $x_s^G = \left(x_{1,s}^G, \ldots, x_{n_G,s}^G\right)'$ denote the $n_G \times 1$ vector of real-time series in group G for month s. For each forecast month t, we apply scaled PCA to the vintage-specific history $\{x_s^G : s \leq t\}$ available at that date. This procedure yields horizon-specific diffusion-index factors for each group, with implementation details provided in Online Appendix A.2.

We denote the estimated factors for group G in month s and horizon h by $f_{s,h}^G \in \mathbb{R}^{r_{G,h}}$, where $r_{G,h}$ is the number of factors extracted for that group and horizon. For each group and horizon, we choose $r_{G,h}$ as the smallest number of factors that explains at least 50 percent of the variation in that group's scaled data matrix.[6] For each forecast month t, the resulting factor series $\{f_{s,h}^G : s \leq t\}$ provides a low-dimensional real-time summary of the high-dimensional history of variables in group G for horizon h. Repeating this factor-construction step for each vintage and forecast horizon ensures that the predictors used at date t do not incorporate later data releases, revisions, or future observations.

To capture both current conditions and recent dynamics while keeping the predictor set parsimonious, we include the current value $f_{t,h}^G$ and its one- to three-month lags, $f_{t-1,h}^G, \cdots, f_{t-3,h}^G$. We include the current factor only when the latest month contains sufficient observed information for that group. Otherwise, we start the predictor block with lagged factors. We also include averages covering lags four to six months back and seven to twelve months back: $\bar{f}_{t-4|t-6,h}^G = \frac{1}{3}\sum_{j=4}^{6} f_{t-j,h}^G$ and $\bar{f}_{t-7|t-12,h}^G = \frac{1}{6}\sum_{j=7}^{12} f_{t-j,h}^G$. Finally, we stack all factor-based terms across groups to form the horizon-specific predictor vector $X_{t,h}$. Although the scaled PCA step substantially reduces dimension of the original predictor set, the inclusion of multiple factors and their lags leaves $X_{t,h}$ high-dimensional.[7]

[5] The use of diffusion-index factors for macroeconomic forecasting follows Stock and Watson (2002). More broadly, our approach builds on the large-dimensional approximate factor-model literature, including Stock and Watson (2002) and Bai and Ng (2002).

[6] Section 8.2 varies this cutoff to 30% and 80%.

[7] In Section 8.2, we also allow simple nonlinearities in two ways. We augment these predictors with (i) factors extracted from the squared series $\{(x_{it}^G)^2\}$ and (ii) element-wise squares of the original horizon-specific factors, $f_{t,h}^G$.

3.4 Real-Time Out-of-Sample Forecasting Procedure

In each forecast month t, we re-estimate the full forecasting procedure using only information available at that month. This includes the construction of predictors, the selection of hyperparameters, and the estimation of the final logit coefficients. Thus, for each horizon h, the forecast $\hat{p}_{t,h}$ is a real-time forecast.

The main timing restriction comes from the forecast target. For horizon h, the target $y_{s,h}$ is observed by month t only if the full forecast window associated with predictor month s has been realized. Let $m_{t,h}$ denote the latest predictor month with a realized horizon-h target.[8] Longer forecast horizons therefore have fewer labeled observations at any forecast date. The labeled sample available at month t is $\{s : s \leq m_{t,h}\}$. Observations after $m_{t,h}$ are not used for estimation or validation, although the model can still produce forecasts for those months.

We tune the elastic-net hyperparameters using recent realized validation periods. For each horizon h and forecast month t, we define validation periods as the most recent non-overlapping 36-month blocks that end at $m_{t,h}$, using up to four blocks subject to a minimum 8-year estimation window before each block:

$$B_{t,h}^{(1)} = \{m_{t,h} - 35, \ldots, m_{t,h}\}, \quad B_{t,h}^{(2)} = \{m_{t,h} - 71, \ldots, m_{t,h} - 36\}, \quad \ldots$$

For early vintages with too few observations to form four validation blocks while preserving the minimum estimation window, we use as many blocks as are feasible. This design helps tune the regularized logit in a rare-event setting by validating candidate hyperparameters on more realized recession outcomes while maintaining a nontrivial estimation history.

For each candidate pair (λ_h, α_h), we evaluate performance recursively across these validation blocks. Let $\tau_{t,h}^{(k)}$ denote the first month in the block $B_{t,h}^{(k)}$. We estimate the penalized logit model using the *training* sample before that block, $\{s : s \leq \tau_{t,h}^{(k)} - 1\}$, generate predicted probabilities $\hat{p}_{s,h}(\lambda_h, \alpha_h)$ for all $s \in B_{t,h}^{(k)}$, and compute block-level average log loss:

$$\ell_{t,h}^{(k)}(\lambda_h, \alpha_h) = \frac{1}{|B_{t,h}^{(k)}|} \sum_{s \in B_{t,h}^{(k)}} \Big[- y_{s,h} \log \hat{p}_{s,h}(\lambda_h, \alpha_h) - (1 - y_{s,h}) \log \big(1 - \hat{p}_{s,h}(\lambda_h, \alpha_h)\big) \Big].$$

We average $\ell_{t,h}^{(k)}(\lambda_h, \alpha_h)$ across all feasible validation blocks and choose the pair $(\hat{\lambda}_{t,h}^*, \hat{\alpha}_{t,h}^*)$ that minimizes this average validation loss.

Holding $(\hat{\lambda}_{t,h}^*, \hat{\alpha}_{t,h}^*)$ fixed, we then refit the elastic-net logit model using all labeled ob-

[8] For example, with data available through May 2026, six-month outcomes can be evaluated only through November 2025: the November 2025 forecast asks whether a recession occurs from December 2025 through May 2026, while the December 2025 forecast depends on outcomes through June 2026.

servations at month t, $\{s : s \leq m_{t,h}\}$, and obtain $\hat{\beta}_{t,h}$. The time-t forecast is then

$$\hat{p}_{t,h} = \Lambda(X'_{t,h}\,\hat{\beta}_{t,h}), \quad \Lambda(z) = (1 + e^{-z})^{-1}$$

where $\Lambda(\cdot)$ maps log-odds into probabilities. Repeating this procedure each month yields a real-time sequence of out-of-sample recession probability forecasts.

The out-of-sample exercise begins in August 1999, the first available FRED-MD real-time vintage. For each vintage, we construct the predictor dataset from January 1980 onward. The estimation sample therefore expands over time, while the validation windows retain a fixed length. We choose the January 1980 start date so that, even at the initial August 1999 vintage and for the 12-month forecast horizon, the available sample contains enough history to estimate the model using at least 96 monthly observations and evaluate candidate hyperparameters on at least three 36-month validation blocks.

4 Recession Signals in Hard and Soft Data

Do soft and hard data send different signals about recession risk? In this section, we study how these signals differ by estimating the elastic-net logit model separately on hard and soft information sets and comparing the resulting real-time out-of-sample forecasts. We compare both the forecast paths and threshold-based diagnostics, such as precision and recall, to characterize the types of recession warnings each model produces.

4.1 Hard-Data and Soft-Data Forecasts

We estimate the model separately using hard and soft data. The hard-data model uses predictors constructed from the hard-data groups, while the soft-data model uses predictors constructed from the soft-data groups. Both use the predictor construction described in the previous section, including the lag terms, so the models differ only in the data source used to construct the predictors.

Figure 1 shows the estimated out-of-sample probability of recession within the next six months as of each forecast date from August 1999 through May 2026. We focus on the six-month horizon for this exercise and present results for other horizons in Online Appendix B (Figures B1–B3).[9] The solid purple line shows probabilities from the soft-data model,

[9]We focus on this horizon because it lies between the very short-term horizons examined in the nowcasting literature and the longer-term horizons at which the yield curve has traditionally been studied, making it well suited for examining the value added of soft data.

while the dashed gray line shows the probabilities from the hard-data model. The figure also includes recession shading and the realized six-month-ahead forecast target.

The figure makes clear that the hard-data and soft-data models often assign very different levels of recession risk. These divergences are not unique to the recent vibecession period. Other episodes with large disagreement include the 2001 recession, the 2008–2009 recession, early 2003, and late 2011. The disagreements take two forms: in some cases, soft data provide a stronger recession signal than hard data before and during actual downturns. In others, soft data send elevated recession-risk signals outside of recession episodes.

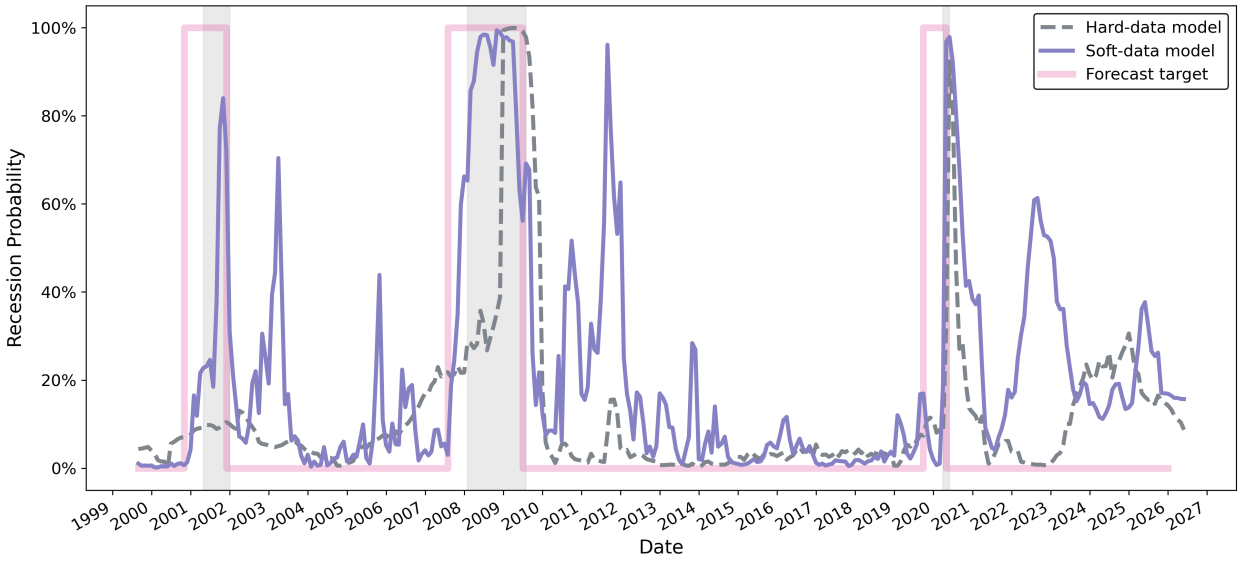
The 2001 recession illustrates the first pattern. The soft-data model's probability rises in the lead-up to the recession, remains elevated during the recession, and then falls near the end of the recession. By contrast, the hard-data model assigns low recession probabilities even during the recession itself and rises somewhat during the recovery. The episode is one in which soft data captured recession risk more clearly than the hard-data indicators. This hard-data miss is consistent with the 2001 recession being mild in many standard activity indicators, which made it challenging for a hard-data model to detect.

The Great Recession shows a different pattern. In mid-2007, the hard-data model already assigns moderate recession risk, around 20 percent, while the soft-data model remains low. The soft-data probability then rises quickly after the six-month-ahead recession window begins, reaching about 25 percent by September 2007 and about 66 percent by December 2007 (the NBER peak). The hard-data probability increases more gradually, staying near 20 percent through late 2007 before rising sharply only later in the recession. Thus, by late 2007, the soft-data model gives a much clearer pre-recession signal than the hard-data model.

The figure also shows episodes in which the soft-data model signals elevated recession risk during expansions. Early 2003 and late 2011 are two historical examples: in both cases, the soft-data model assigns elevated recession probabilities even though no recession follows within six months, while the hard-data model's forecasts remain much lower. These episodes coincide with periods of heightened uncertainty and weak sentiment: early 2003 followed the dot-com recession and took place around the Iraq War, while late 2011 coincided with the European sovereign debt crisis, the U.S. debt-ceiling standoff, and the downgrade of U.S. Treasury debt. The recent vibecession period shows a more sustained version of this divergence. In 2022, the soft-data model rises sharply while the hard-data model remains near zero, consistent with the large deterioration in sentiment and uncertainty despite continued strength in many hard indicators.

Neither model signals the 2020 recession before it begins, which is unsurprising given

Figure 1: Predicted Recession Risk over the Next Six Months: Hard Data versus Soft Data



Notes: The figure shows out-of-sample forecasts of recession risk over the next six months from the soft-data model (purple line) and the hard-data model (dashed grey line) as of each real-time data vintage from August 1999 to May 2026. The shaded areas indicate NBER recessions. The blue line is the realized six-month-ahead forecast target.

that the COVID-19 recession was triggered by an abrupt pandemic shock rather than by the gradual deterioration in macroeconomic conditions that these models are designed to detect.

4.2 Soft data cast a wider net

To further compare how to the models signal recessions, we convert the forecast probabilities into binary recession warning signals. Doing so requires choosing a probability threshold. Because recessions are rare, a forecast can convey meaningful recession risk even when recession remains less likely than not. We therefore use 25 percent as our baseline warning threshold: a model signals recession when its out-of-sample probability is at least 25 percent. This cutoff is high enough to focus on clearly elevated recession risk, but low enough to capture early-warning signals in a rare-event setting. We report analogous results using a more stringent 50 percent threshold in Online Appendix Table B1.

Table 1 compares these binary signals with the realized six-month-ahead forecast target, forming a confusion matrix. A signal is correct if a recession occurs at any point in the six months after the forecast date. True positives (TP) are cases in which the model signals recession and a recession occurs within six months, while false negatives (FN) are cases in which a recession occurs within six months but the model does not signal one.

Table 1: Confusion Matrix for Recession Forecasts

Realized outcome	Soft data only		Hard data only	
	Signaled recession	Did not signal	Signaled recession	Did not signal
Recession occurred	23	20	17	26
No recession occurred	64	209	17	256

Notes: The table classifies recession warning signals from the hard-data and soft-data models. A warning signal is defined as a predicted recession probability of at least 25 percent. The sample covers forecast dates from August 1999 through November 2025.

False positives (FP) are cases in which the model signals recession but no recession occurs within six months, while true negatives (TN) are cases with no signal and no recession within six months.

Two patterns stand out. First, the soft-data model identifies more of the months that are followed by a recession within six months. It correctly identifies 23 such months, compared with 17 for the hard-data model, and has fewer false negatives: 20 versus 26. Second, the soft-data model produces more false alarms, incorrectly signaling recession in 64 months not followed by a recession within six months, compared with 17 for the hard-data model. Thus, the soft-data model casts a wider net: it identifies more months before recessions, but it also signals recession risk more often when no recession occurs.

The confusion matrix also defines two useful summary statistics: precision and recall. Precision measures how reliable a recession signal is once issued: it is the share of predicted recession signals that are followed by a recession within the forecast horizon. Recall measures coverage of the forecast target: among months followed by a recession within the forecast horizon, it is the share that the model correctly signals.

Table 2 reports precision and recall across all forecast horizons for the hard-data and soft-data models. At the six-month horizon, the hard-data model has precision of $17/34 = 0.50$, while the soft-data model has precision of $23/87 \approx 0.26$. The top panel of Table 2 shows that the hard-data model has higher precision than the soft-data model at every horizon. Thus, when the hard-data model signals recession risk, that signal is more likely to correspond to an actual recession.

By contrast, the soft-data model generally has higher recall, capturing a larger share of realized recession episodes. At the six-month horizon, the hard-data model has recall of $17/43 \approx 0.40$, while the soft-data model has recall of $23/43 \approx 0.53$. The bottom panel of Table 2 shows that the soft-data model has higher recall at the 1-, 3-, and 6-month horizons, while the two models have the same recall at the 12-month horizon.

Both precision and recall depend on the probability threshold used to convert forecast probabilities into recession signals. At the baseline 25 percent threshold, the soft-data model captures more recession-target months and therefore tends to have higher recall than the hard-data model, but it also generates more false alarms and therefore has lower precision. Online Appendix Table B1 repeats the comparison using a more stringent 50 percent threshold. At that cutoff, the soft-data model has higher recall at all horizons and also has higher precision at the six- and twelve-month horizons. Its precision is higher because, despite producing more false alarms, the soft-data model also identifies enough additional true positives that a larger share of its warnings are correct.

Taken together, these results show that the two data sources produce different types of recession signals. The soft-data model identifies more recession episodes, while the hard-data model is more selective. We next ask how these differences translate into overall forecast performance, both separately and in combination.

Table 2: Precision & Recall across Models and Forecast Horizons

(a) Precision

Forecast horizon	1 month	3 months	6 months	12 months
Hard data only	0.600	0.609	0.500	0.358
Soft data only	0.355	0.361	0.264	0.279

(b) Recall

Forecast horizon	1 month	3 months	6 months	12 months
Hard data only	0.429	0.412	0.395	0.557
Soft data only	0.786	0.765	0.535	0.557

Notes: The table reports precision and recall for recession warning signals from the hard-data and soft-data models. A warning signal is defined as a predicted recession probability of at least 25 percent. Precision measures the share of warning signals that are followed by a recession within the forecast horizon. Recall measures the share of months followed by a recession within the forecast horizon that the model correctly predicts. The sample covers forecast dates from August 1999 through November 2025.

5 Combining Recession Signals from Hard and Soft Data

In this section, we consider the forecasting performance of the hard-data and soft-data models as well as a combined model that uses both sources of data. We first illustrate how the combined model uses hard and soft data, showing that it does not simply average their forecasts. Instead, it preserves soft-data warnings when hard data corroborate them

and discounts those warnings when hard data do not. We then evaluate the out-of-sample performance of the resulting forecasts.

5.1 The Combined Hard-and-Soft Forecast

The previous section showed that hard and soft data produce different types of recession signals. We now ask what happens when the two information sets are combined. A natural concern is that adding soft data could carry its false alarms into the combined forecast. Comparing the combined model with the two single-source models shows that this is not what happens.

Figure 2 compares the recession probabilities from the combined model with those from the hard-data and soft-data models at the six-month horizon. The top panel compares the combined model with the soft-data model, while the bottom panel compares the combined model with the hard-data model. The panels help answer whether the combined model simply splits the difference between the two single-source forecasts, or whether it treats soft-data signals differently depending on the surrounding hard-data evidence.

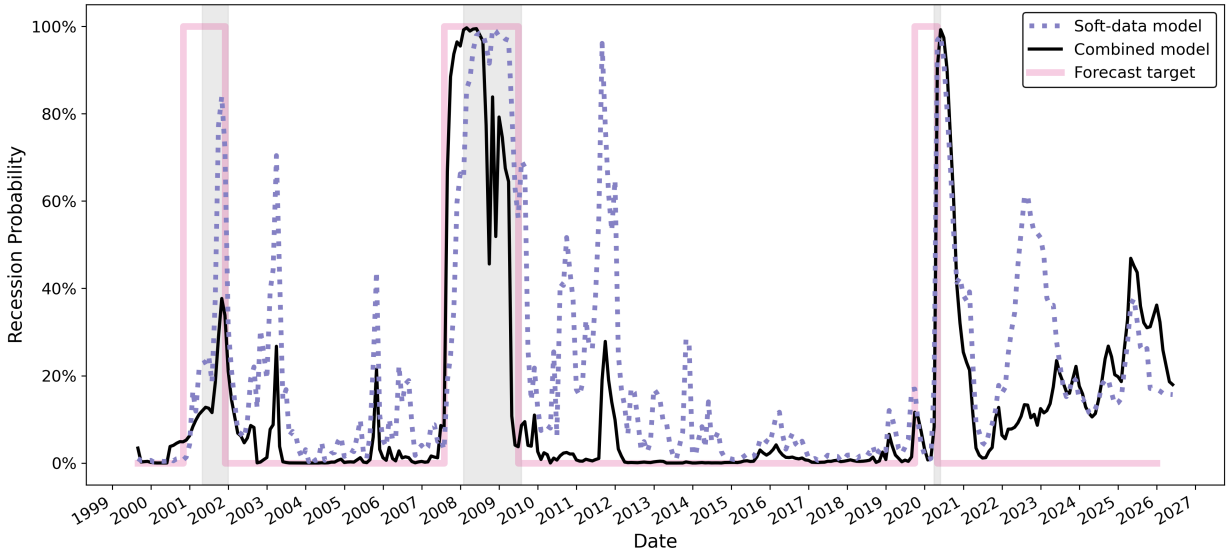
The comparison shows that the combined model is not a mechanical average of the hard-data and soft-data forecasts. Around recessions, the combined model can preserve or amplify soft-data warnings when hard data also point toward recession. In September 2007, for example, the combined model assigns an 88 percent recession probability, compared with 21 percent from the hard-data model and 25 percent from the soft-data model. By December 2007, the combined probability rises to 96 percent, compared with 22 percent for hard data and 66 percent for soft data. The 2001 recession shows a more muted version of the same pattern: the combined model picks up part of the soft-data signal, but keeps the forecast below the soft-data model when hard-data support is weaker.

During expansions, however, the combined model often mutes soft-data spikes. In March 2003, the soft-data model assigns a 70 percent recession probability, while the combined model assigns 27 percent and the hard-data model assigns 5 percent. A similar pattern appears in late 2011: the soft-data model remains above 60 percent in September and December, while the combined model falls from 28 percent to 10 percent. In July 2022, the soft-data model again rises sharply, to 61 percent, while the combined model remains at 13 percent and the hard-data model near zero. These comparisons suggest that hard data help separate soft-data warnings that coincide with broader economic deterioration from soft-data warnings that remain largely uncorroborated.

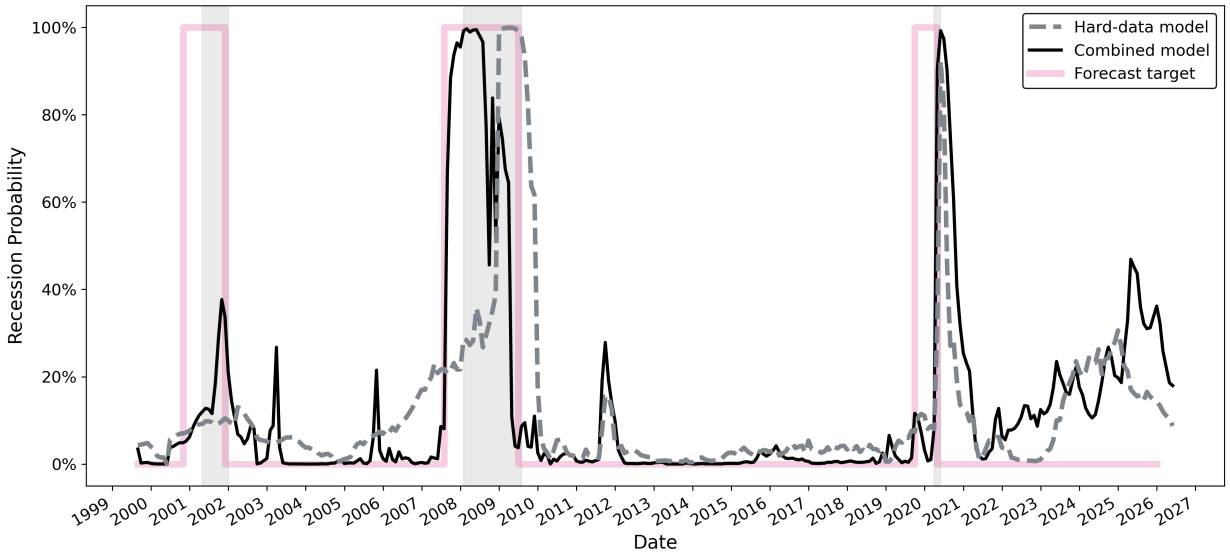
Overall, the combined model appears to use hard data for corroboration of soft-data signals: it preserves or amplifies soft-data warnings when they coincide with broader eco-

Figure 2: Predicted Recession Risk over the Next Six Months: Combining Hard and Soft Data

(a) Soft-data model versus Combined model



(b) Hard-data model versus Combined model



Notes: All plotted probabilities are out-of-sample forecasts of recession risk over the next six months as of each monthly real-time data vintage from August 1999 to May 2026. The solid line in both panels shows the combined model, which uses both hard and soft data. The purple dotted line shows the soft-data model, and the grey dashed line shows the hard-data model. Shaded areas indicate NBER recessions. The blue line is the realized six-month-ahead forecast target.

nomic deterioration, but mutes them when hard-data support is limited. These comparisons show when the combined model moves toward the soft-data signal and when it stays closer to hard data.

5.2 Out-of-Sample Forecast Performance

Figure 2 shows how the combined model uses hard and soft data in specific episodes. We now formalize this comparison by evaluating forecast performance across horizons and using several metrics that capture different aspects of performance. Because the soft-data model uses a much smaller information set than the hard-data model, similar performance would already indicate that soft data contain meaningful recession-predictive information. The combined model then lets us ask whether soft data add information beyond hard data alone.

We begin with the F1 score because it directly summarizes the precision-recall tradeoff documented in Section 4.2. The F1 score combines these two dimensions into a single statistic,

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

Higher values indicate a better balance between reliable warning signals and recession coverage.

The first two panels of Table 3 report the F1 score for each specification and horizon at the 25 and 50 percent thresholds. Comparing the combined model with the hard-data and soft-data models shows whether the two information sets are complementary. At the one-, three-, and six-month horizons, the combined model has a higher F1 score than either model alone, implying that using hard and soft data together improves the balance between precision and recall.

At the twelve-month horizon, however, the hard-data model has the highest F1 score for the 25% threshold while the soft-data model has the highest F1 score for the 50% threshold. This result is a useful caution: adding predictors does not automatically improve forecast performance, and regularization cannot fully prevent weak additional information from worsening forecasts.

The F1 score is intuitive in this setting, but it depends on a subjective threshold choice used to convert probabilities into warning signals. We therefore also report the precision-recall area under the curve (PR-AUC) to evaluate whether the same precision-recall tradeoff holds more broadly. PR-AUC traces precision and recall as the probability cutoff varies, making it a threshold-free counterpart to the F1 score. Because PR-AUC depends on the ranking of predicted probabilities, however, it does not penalize false alarms in the same way as the F1 score.[10]

[10]Months with elevated predicted recession risk but no recession within the forecast horizon reduce PR-AUC mainly when they are ranked above months that are followed by a recession. If months followed by a recession are generally ranked higher, the model can still have a strong PR-AUC even if it would issue many false alarms at a given cutoff.

Table 3: Out-of-Sample Performance by Specification and Forecast Horizons

(a) F1 Score at 25 Percent Cutoff

Forecast horizon	1 month	3 months	6 months	12 months
Hard data only	0.500	0.491	0.442	**0.436**
Soft data only	0.489	0.491	0.354	0.372
Hard and soft data	**0.656**	**0.592**	**0.500**	0.296

(b) F1 Score at 50 Percent Cutoff

Forecast horizon	1 month	3 months	6 months	12 months
Hard data only	0.450	0.298	0.211	0.233
Soft data only	0.645	0.545	0.472	**0.375**
Hard and soft data	**0.717**	**0.655**	**0.559**	0.231

(c) Precision-Recall Area Under the Curve (PR-AUC)

Forecast horizon	1 month	3 months	6 months	12 months
Hard data only	0.610	0.531	0.478	0.315
Soft data only	0.760	**0.682**	0.517	0.334
Hard and soft data	**0.768**	0.653	**0.559**	**0.382**

(d) Area Under the Curve (AUC)

Forecast horizon	1 month	3 months	6 months	12 months
Hard data only	0.887	0.863	0.850	**0.751**
Soft data only	0.922	0.876	0.747	0.599
Hard and soft data	**0.923**	**0.879**	**0.855**	0.732

Notes: Panels (a) and (b) report the F1 score from binary recession signals obtained by thresholding predicted recession probabilities at 25 percent and 50 percent, respectively. Panels (c) and (d) report the PR-AUC and AUC, respectively. Each panel reports out-of-sample recession forecast performance for the three specifications indicated in the rows and for each of the horizons indicated in the columns.

Panel (c) of Table 3 reports PR-AUC for each specification and forecast horizon. The soft-data model has a higher PR-AUC than the hard-data model at every horizon, indicating that soft-data forecasts rank recession-target months more effectively than hard-data forecasts. The combined model generally improves on both, yielding a higher PR-AUC at the one, six, and twelve months and only slightly lower PR-AUC at the three-month horizon. This indicates that soft data add recession-relevant information and that this information is useful alongside hard data.

Panel (d) reports AUC, another threshold-free metric widely used in classification problems (Bradley, 1997; Berge and Jordà, 2011). AUC summarizes the tradeoff between recall

and the false positive rate, rather than the tradeoff between recall and precision.[11] The AUC results also point to an important role for soft data: The combined model yields the highest AUC at the one-, three-, and six-month horizons, while yielding a slightly lower AUC than the hard data model at the twelve month horizon. This reinforces the view that soft data remain useful, especially when combined with hard data.

Overall, the results point to two conclusions. First, the soft-data model performs comparably to, and often better than, the hard-data model despite using a much smaller information set. Second, adding soft data to hard data improves performance across a range of metrics for short- to medium-term horizons. The combined model has the highest score on all four metrics at horizons from one to six months ahead, except that its three-month PR-AUC is slightly below that of the soft-data model. At the twelve-month horizon, the ranking is more mixed: no model dominates across all metrics.

6 Why do soft data help?

The results in the previous section show that soft data improve recession forecasts, especially at short- to medium-term horizons. We next ask why. Soft and hard data differ not only in what they measure, but also in when and how they become available. These differences suggest two possible reasons why soft data may add forecasting value. First, soft data may partly substitute for hard-data observations that are unavailable or noisy in real time. Second, soft data may contain distinct recession signals that are not well captured by hard data, even after revisions.

To distinguish between these channels, we compare forecasts constructed from three hard-data vintages. The real-time vintage uses only the hard data available at the forecast date. The next-month vintage adds the hard-data observations that become available one month later, including many current-month observations that were missing in real time because of publication lags. The latest vintage uses the final available hard-data vintage, incorporating both delayed releases and subsequent data revisions.

This comparison separates the two ways in which soft data may help. If soft data mainly compensate for publication lags or noisy initial releases in hard data, then moving from the real-time vintage to the next-month or latest vintage should improve hard-data forecasts and reduce the incremental value of adding soft data. If soft data contain distinct

[11] This distinction matters for recession forecasting because recessions are rare. A model can generate many false alarms but still have a low false positive rate, since those false alarms are divided by the large number of months not followed by a recession. As a result, AUC can look favorable even when a model would issue too many recession warnings at a given threshold. We therefore treat AUC as a useful comparison with the literature, but place more weight on F1 and PR-AUC for evaluating recession warning signals.

information, then soft data should remain useful even in counterfactual settings where hard data are made more timely or replaced with revised values.

We first examine whether publication lags and ex post revisions limit the forecast performance of the hard-data-only model. The top panel of Table 4 reports the PR-AUC of models trained only on hard data.[12] Hard data in our sample include both financial variables and nonfinancial macroeconomic indicators. Financial variables are observed daily and available at the forecast date, so they do not suffer from publication lags or undergo the revision process that affects macroeconomic releases. The exercises below therefore affect only the nonfinancial hard-data indicators.

The first row uses the real-time vintage available at the forecast date. Forecasts from this specification therefore reflect the actual information set available to the forecaster, including publication lags and potentially noisy initial releases. The second row uses the next-month vintage up to the forecast month. Because many current-month hard-data indicators become available only in the following month, this specification effectively removes publication lags by treating unreleased observations as if they were available at the forecast date. The final row uses current vintage data for all forecast months, incorporating all subsequent revisions. This specification gives the model the most hindsight: it removes both publication lags and the noise associated with initial releases.

The results show that moving from the real-time vintage to the next-month vintage improves the PR-AUC of the hard-data-only model at every horizon. Moving from the next-month vintage to the latest vintage adds little up to six-month horizons, but improves performance at the twelve-month horizon. Overall, the latest-vintage hard-data model outperforms the real-time hard-data model at every horizon, indicating that hard-data observations not yet released by the forecast date, and in some cases subsequent revisions, contain recession-predictive information.

The bottom panel of Table 4 shows analogous results for a model that combines soft and hard data, varying the vintage of the hard data across the three rows. The main result is that adding soft data generally improves PR-AUC relative to the corresponding hard-data-only model, regardless of which hard-data vintage is used. The exception is the twelve-month horizon, where later-vintage hard data alone perform slightly better. This pattern suggests that soft data contain recession-relevant information beyond what hard indicators capture, even when those indicators are observed with a delay or after revision.

The size of the gain depends on the hard-data vintage. Adding soft data improves PR-

[12] We use PR-AUC rather than the F1 score for this exercise because the vintage comparison asks whether later hard-data vintages improve the ranking of recession risk across months. By contrast, the F1 score depends on the chosen warning threshold and can change when revised hard data move forecasts across that cutoff.

AUC the most when hard data are restricted to the real-time vintage, which is the vintage most affected by missing current-month observations and noisy initial releases. The larger gain in this case suggests that soft data partly capture information about current economic conditions that hard data reveal only later.

The vintage exercises suggest that soft data improve real-time recession forecasts through *both* channels. Soft data partly fill gaps left by delayed and noisy hard-data releases, providing a contemporaneous summary of conditions that hard data reveal only later. But timeliness is not the whole story. At horizons up to six months, soft data contain recession-predictive information beyond what is captured by hard data, even after later releases and revisions. These channels help explain why soft data improve forecast performance in real time, as documented in Section 4.2, and why the combined model performs well in the baseline real-time comparison.

Table 4: Precision-Recall AUC Across Hard-Data Vintages

(a) Hard data only

Vintage	1 month	3 months	6 months	12 months
Real-time vintage	0.610	0.531	0.478	0.315
Next-month vintage	0.663	0.562	0.499	0.429
Latest vintage	0.670	0.562	0.503	0.538

(b) Hard and soft data

Vintage	1 month	3 months	6 months	12 months
Real-time vintage	0.768	0.653	0.559	0.382
Next-month vintage	0.692	0.601	0.502	0.389
Latest vintage	0.700	0.630	0.509	0.388

Notes: The table reports precision-recall AUC for models estimated using different vintages of the hard-data variables. The real-time vintage uses the data available at the forecast date. The next-month vintage uses the data available one month after the forecast date, allowing for the release of current-month hard indicators. The latest vintage uses the final available data vintage, incorporating subsequent data revisions. The top panel reports results for models trained only on hard data. The bottom panel reports results for models that combine hard and soft data.

7 Interpreting Recession Signals in Historical Episodes

The comparison of six-month recession probabilities in Section 5 showed that the combined model does not simply average the hard-data and soft-data forecasts: it preserves soft-data signals near recessions while muting them during expansions. That comparison, however, treated hard data as a single block, leaving open which hard-data categories corroborate or offset soft-data signals. Here we use Shapley decompositions to attribute the combined model's fitted six-month recession risk to soft data and to five separate hard-data categories, and we trace how these contributions evolve across four historical episodes.

7.1 Shapley Value Decomposition

The previous sections compare the forecasting performance of the hard-data, soft-data, and combined models, but they do not examine which hard-data indicators drive the combined model's forecasts. We now unpack this mechanism by asking when hard data corroborate soft-data recession signals and when they offset them.

We group hard-data predictors into the five categories introduced in Section 3: real activity, housing, the yield curve, other financial indicators, and inflation. These categories need not move together at a given forecast date. A soft-data warning may be reinforced by weakness in housing or the yield curve, offset by strength in real activity, or left largely unconfirmed by the broader hard-data evidence.

To interpret these fitted forecasts, we use Shapley decompositions (Lundberg and Lee, 2017; Buckmann et al., 2023). Because the forecasts come from a logit model, we decompose fitted log odds rather than predicted probabilities. For each forecast date, the baseline is the fitted log odds when all predictors are set to their sample means. The decomposition then attributes the difference between the fitted log odds and this baseline to soft data and to each hard-data category.

The decomposition accounts for the fact that a predictor's marginal contribution can depend on which other predictors have already been included, especially when predictors are correlated. It does so by averaging each predictor's marginal contribution over possible orderings of the predictors.[13] We implement this using a correlation-dependent linear Shapley decomposition, which accounts for the empirical dependence among predictors.[14]

[13] Formally, let $f(x)$ denote the fitted log odds and let P be the full predictor set. For a subset S, define $f_S(x_S) = E[f(x_S, X_{-S}) \mid X_S = x_S]$, with the conditional expectation evaluated using the empirical mean and covariance matrix of the predictors. The Shapley value for predictor j is the weighted average of $f_{S\cup\{j\}}(x_{S\cup\{j\}}) - f_S(x_S)$ over subsets $S \subseteq P \setminus \{j\}$, with weights determined by the number of possible predictor orderings.

[14] Because the decomposition accounts for predictor dependence, attribution can be shared across corre-

The resulting predictor-level contributions sum to the fitted log odds relative to the baseline. We aggregate these contributions into six groups: soft data and the five hard-data categories. Positive contributions raise fitted recession risk relative to the baseline, while negative contributions lower it.

Because the decomposition is additive in log-odds units, we report the Shapley contributions in log odds. For interpretation, we also translate selected category contributions into implied recession probabilities. If the baseline log odds at date t are $\ell_{0,t}$ and data-category c contributes $\phi_{c,t}$ log-odds points, we report

$$\Lambda\left(\ell_{0,t} + \phi_{c,t}\right),$$

where $\Lambda(x) \equiv \exp(x)/(1+\exp(x))$ is the logistic function.[15] This calculation asks what recession risk the model would assign from category c's date-t contribution alone, holding all other category contributions at zero. These category-implied probabilities put log-odds contributions in more familiar units, but they are not additive across categories.

7.2 Recession Signals in Four Historical Episodes

We focus on four episodes from the combined model's probabilities of recession within six months in Figure 2. The first two episodes, 2001 and the Great Recession, are recession episodes. The remaining two episodes, 2011–2012 and 2022–2025, are soft-data false alarms, as shown in Figure 1: the soft-data model pointed to elevated recession risk, while the hard-data model did not, and no recession followed. We use these cases to study when soft-data signals are reinforced by hard data and when they are offset or left unconfirmed.

Figure 3 decomposes the combined model's probabilities of recession within six months during the 2001 recession and the Great Recession. The top panel shows the 2001 episode. Based on the NBER business-cycle peak of March 2001, the first recession month is April 2001, so an accurate six-month forecast should start signaling recession risk by the end of October 2000. The combined model's recession probability rises modestly over this window and peaks at 38% in October 2001.

The decomposition shows that soft data provide the main positive contribution at the peak. Their contribution is 1.64 log-odds points, which moves implied recession risk from the baseline probability of 13% to 43% when other categories are held at baseline. Real

lated predictors. A predictor whose coefficient is set to zero by the elastic-net logit model can therefore receive a nonzero contribution if it helps explain movements in correlated predictors that enter the fitted forecast.

[15] Online Appendix C shows how log odds are converted into probabilities and applies this conversion to the category-level Shapley contributions.

activity and the yield curve provide only modest support, while other financial conditions, housing, and inflation offset the signal. The combined model therefore raises recession risk, but keeps the signal contained because the broader hard-data evidence does not line up strongly behind the soft-data warning.

The bottom panel shows a different pattern during the Great Recession. Based on the NBER business-cycle peak of December 2007, the first recession month is January 2008, so a six-month forecast should signal recession risk starting at the end of July 2007. The combined model's recession probability rises from 8% at the end of July 2007 to 88% by the end of September 2007.

The early signal is mainly hard-data-driven. In August 2007, the housing contribution raises implied recession risk by about 15 percentage points, while the yield-curve contribution raises it by about 11 percentage points. Soft data contribute less at this point, raising implied risk by about 2 percentage points. This early phase shows why it is useful to unpack the hard-data block: the signal comes not from hard data uniformly, but from housing and the yield curve in particular.

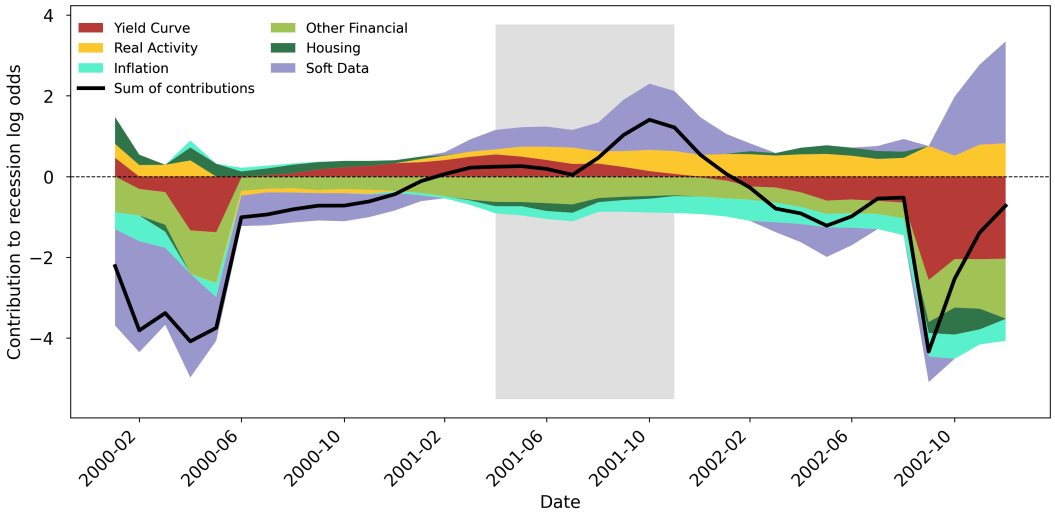
As the downturn becomes visible, soft data become central to the fitted forecast. By June 2008, the combined model's recession probability reaches 98%. The soft-data contribution alone raises implied recession risk by about 86 percentage points. By late 2008, the source of risk shifts again: other financial conditions become the dominant recessionary hard-data signal, while the yield-curve contribution becomes less positive. The episode therefore shows a rotation in the source of recession risk: housing and the yield curve provide the first six-month signal, soft data dominate as the downturn becomes visible, and financial conditions become important during the crisis phase.

Figure 4 decomposes the combined model's predicted probabilities of recession within six months during the two periods in which the soft-data model signaled elevated recession risk that did not materialize. The top panel shows the 2011–2012 mixed-signal episode. No recession occurred, but soft data deteriorated sharply in late 2011, consistent with weak sentiment around the European sovereign debt crisis, the U.S. debt-ceiling standoff, and the downgrade of U.S. Treasury debt. The combined model's recession probability peaks at 28% in September 2011.

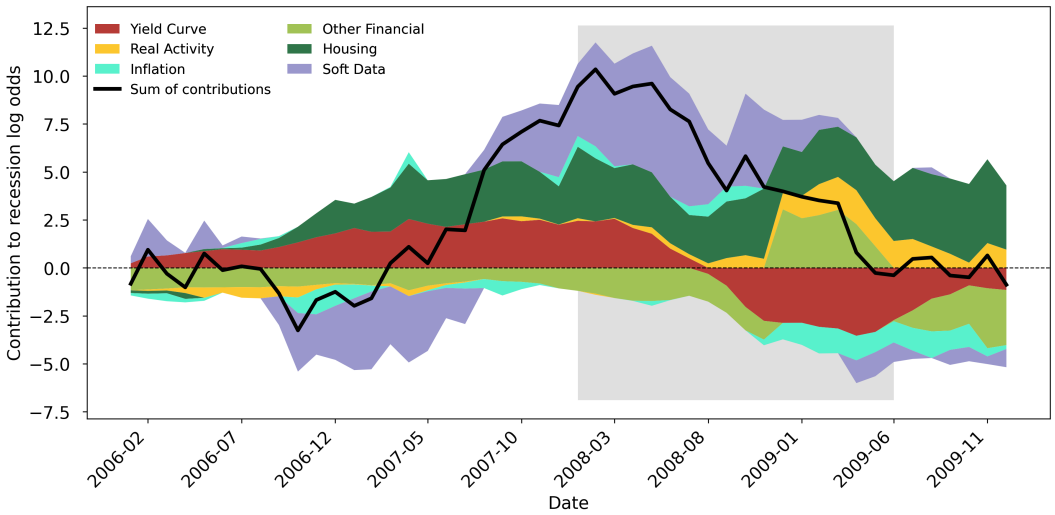
The decomposition shows that the soft-data signal is much larger than the total forecast would suggest. Soft data contribute 2.58 log-odds points, moving implied recession risk from the baseline probability of 5% to 42% when other categories are held at baseline. Other financial conditions also move in a recessionary direction later in 2011, but the offset comes from the rest of the hard-data block. By December 2011, real activity, housing, and the yield curve all point away from recession, and the combined model's predicted

Figure 3: Sources of Predicted Recession Risk: Recession Episodes

(a) Jan. 2000–Dec. 2002



(b) Jan. 2006–Dec. 2009



Notes: The solid black line is the sum of the colored category contributions and equals the combined model's fitted log odds of a recession within the next six months minus the baseline log odds. Each colored area shows the contribution of the indicated category to that difference. Each category's contribution is the sum of the Shapley values from all individual variables (including lags) within the category. The grey shaded bars indicate NBER recessions.

probability is only 10%. The episode therefore shows how hard-data categories can mute a soft-data false alarm rather than allowing the combined forecast to follow the soft-data signal one-for-one.

The bottom panel shows the 2022–2025 period, during which the gap between weak soft indicators and resilient realized activity remained unusually pronounced. The episode is useful because it shows both sides of the combined model's filtering mechanism. In the first phase, the model restrains the soft-data signal because hard-data corroboration is limited. In the second phase, it amplifies the signal as the yield curve begins to point in the same direction.

From 2022 through mid-2023, soft data deteriorate sharply while the broader hard-data block does not align behind a recession signal. Predicted risk rises from 6% in January 2022 to a first peak of 23% in May 2023. In July 2022, soft data contribute 2.32 log-odds points, moving implied recession risk from the baseline probability of 9% to 50% when other categories are held at baseline. Inflation also points in a recessionary direction, consistent with the inflation surge and cost-of-living concerns that weighed on sentiment, but the yield curve, housing, and other financial conditions imply risks below the baseline. The combined model therefore treats this period as elevated but contained risk rather than translating soft-data weakness into a broad recession signal.

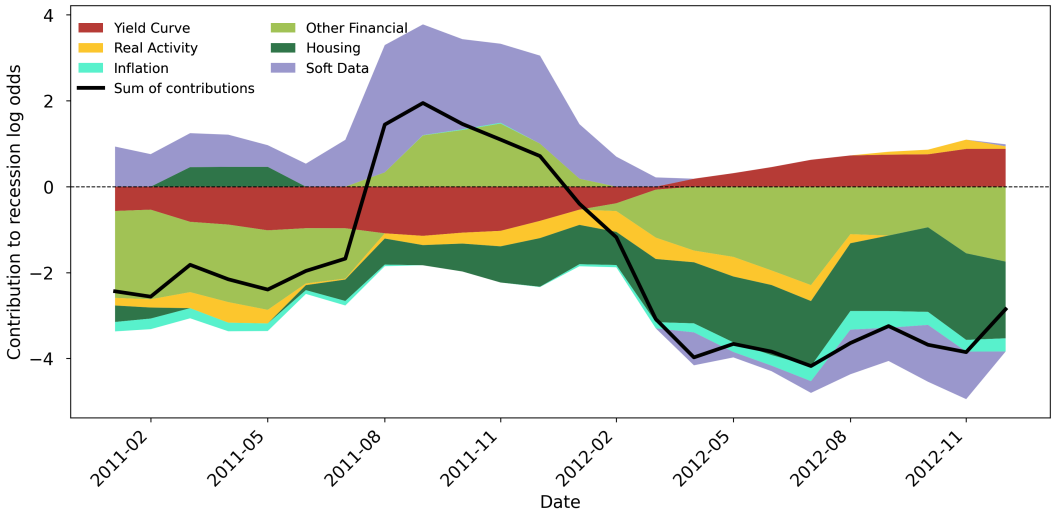
In the second phase, soft data remain weak, and the yield curve also contributes to higher recession risk.[16] By December 2024, the yield-curve category contributes 1.17 log-odds points, implying a recession probability of 21%. Predicted recession risk peaks at 47% in April 2025 and remains elevated at 45% in May 2025, when soft data contribute 2.32 log-odds points, implying a recession probability of 46%. This renewed soft-data weakness is consistent with heightened policy uncertainty, especially around trade policy and its expected effects on prices and labor markets. The 2022–2025 episode therefore shows both sides of the filtering mechanism: hard data can mute soft-data false alarms when corroboration is limited, but they can also amplify soft-data signals when hard-data evidence begins to line up behind them.

Lessons from the episode decompositions. Across the four episodes, the decomposition shows why it is useful to unpack the hard-data block. Hard data do not enter the combined model as a single recession signal: different categories matter across episodes and at different phases of the same episode. In the Great Recession, housing and the yield

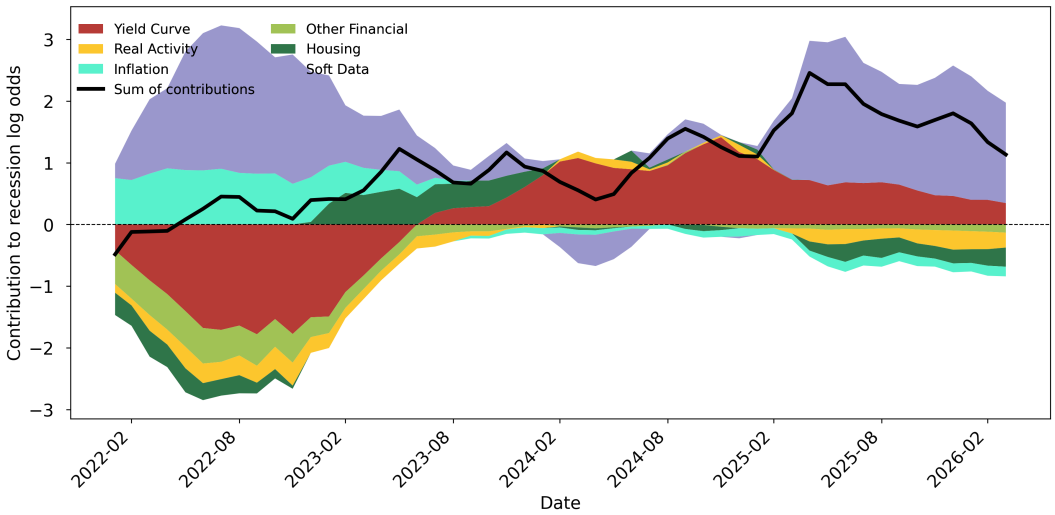
[16] In the monthly data used here, the 10-year minus 2-year Treasury spread turns negative in July 2022, and the 10-year minus 3-month spread turns negative in November 2022. The 10-year minus 3-month spread remains inverted through most of 2024.

Figure 4: Sources of Predicted Recession Risk: Expansion Episodes

(a) Jan. 2011–Dec. 2012



(b) Jan. 2022–Mar. 2026



Notes: The solid black line is the sum of the colored category contributions and equals the combined model's fitted log odds of a recession within the next six months minus the baseline log odds. Each colored area shows the contribution of the indicated category to that difference. Each category's contribution is the sum of the Shapley values from all individual variables (including lags) within the category. The grey shaded bars indicate NBER recessions.

curve provide the first six-month signal, while other financial conditions become more important during the crisis phase. In 2001, by contrast, hard-data support is weaker and more mixed.

The combined model raises recession risk most strongly when soft-data weakness is corroborated by hard-data categories. This helps explain why the Great Recession receives a much stronger signal than the 2001 recession, even though soft data contribute in both episodes. The same mechanism works in the other direction during expansions. In 2011–2012 and much of 2022–2024, soft data point toward recession risk, but hard-data categories are mixed or offsetting, so the combined model keeps risk more contained. By 2025, the yield curve begins to contribute positively alongside soft data, and the combined model raises recession risk as well. These patterns explain why adding hard data to soft data can reduce false alarms while preserving the early-warning content of soft data.

8 Robustness of the Soft-Data Contribution

The preceding sections show that soft data improve recession forecasts when they are added to hard data, especially at short- to medium-term horizons. This section asks how robust that contribution is. We consider two related checks. We first vary the soft-data information set. This asks which sources drive the main result and whether the contribution survives narrower or broader definitions of soft data. We then ask whether richer forecasting specifications, especially richer representations of data, absorb the apparent contribution of soft data. As in Section 6, we report PR-AUC to keep the robustness comparison focused on threshold-free ranking performance.

8.1 Soft-Data Composition and Boundaries

We first ask whether the soft-data contribution depends on which variables are included in the soft-data set. In the main specification, the soft-data category includes six Michigan survey variables and three text-based indicators: the Economic Policy Uncertainty index, the Daily News Sentiment Index, and Beige Book sentiment. We consider two kinds of changes. First, we remove one source at a time from this main soft-data set to see what drives the gains. Second, we compare the main set with narrower and broader alternatives: excluding the text-based indicators and including professional forecasts.

Table 5 keeps the hard-data model as the benchmark and removes one soft-data source at a time from the combined model. This exercise asks whether the soft-data contribution is concentrated in one source or spread across the category.

We find that Michigan survey variables account for much of the near-term contribution. The leave-one-out models that retain the Michigan variables but exclude one text-based source remain above the hard-data benchmark at every horizon, and their PR-AUCs stay close to the baseline hard-and-soft-data model. By contrast, excluding the Michigan variables produces a substantial deterioration relative to the baseline combined model at the one-, three-, and six-month horizons, and the resulting model falls below the hard-data benchmark at one and six months. The twelve-month horizon is different: excluding Michigan improves PR-AUC relative to the baseline combined model, suggesting that household sentiment may be a noisier signal of longer-term recession risk. Overall, the near-term gains appear to come primarily from Michigan survey variables, while the baseline combined model's performance does not appear to depend heavily on any single text-based indicator.

Table 5: PR-AUC When Removing One Soft-Data Source at a Time

Model	1 month	3 months	6 months	12 months
Hard data only	0.610	0.531	0.478	0.315
Hard data with baseline soft data	0.768	0.653	0.559	0.382
Hard data with leave-one-out soft-data sets				
Excluding Michigan	0.600	0.549	0.461	0.411
Excluding EPU	0.760	0.654	0.548	0.364
Excluding DNSI	0.784	0.688	0.569	0.352
Excluding Beige Book	0.777	0.614	0.544	0.373

Notes: The table reports out-of-sample PR-AUC. The first row reports the hard-data model, and the second row reports the model that combines hard data with the baseline soft-data set. The leave-one-out rows keep the hard-data predictors and remove the indicated source from the baseline soft-data set. Forecast horizons are measured in months.

We then vary the boundary of the soft-data set in two directions. First, we narrow the set to Michigan survey variables only. This restriction removes the text-based indicators: the underlying newspaper articles and Beige Book reports were available in real time, but the specific measures we use were constructed later. The exercise therefore asks whether the result survives under a stricter real-time definition of soft data. Second, we broaden the set by adding Blue Chip professional forecasts of real GDP growth and unemployment. Professional forecasts are not sentiment or narrative measures, but they are forward-looking assessments of macroeconomic conditions. Including them asks whether the baseline soft-data set omits an important expectations-based source that could materially change the estimated contribution of soft information. Table 6 compares each variant with the hard-data benchmark and the baseline hard-and-soft-data model.

The table shows that changing this boundary affects the size of the soft-data contribution, but the contribution does not disappear. When the soft-data set is narrowed to Michigan survey variables, the combined model still outperforms the hard-data model at every horizon, although the gain is smaller at six and twelve months. This indicates that the incremental contribution of soft data does not depend entirely on the text-based indicators, even though those indicators add useful medium- and longer-horizon information. When the boundary is expanded to include Blue Chip professional forecasts, short-horizon performance changes little, but twelve-month PR-AUC rises to 0.440, compared with 0.382 in the baseline hard-and-soft-data model. Professional forecasts therefore appear most relevant for longer-horizon recession risk, rather than for the near-term gains emphasized above. The performance gain at the longer horizon from adding professional forecasts is interesting contrast to the performance loss from including forward-looking household variables at the same horizon, as shown in Table 5. Households appear to provide better signals of near-term recession risk than professionals, while the opposite is true for longer-term risk.

Table 6: PR-AUC with Alternative Soft-Data Sets

Model	1 month	3 months	6 months	12 months
Hard data only	0.610	0.531	0.478	0.315
Hard data with baseline soft data	0.768	0.653	0.559	0.382
Hard data with alternative soft-data boundaries:				
Excluding text-based indicators	0.746	0.651	0.543	0.329
Including professional forecasts	0.756	0.658	0.524	0.440

Notes: The table reports out-of-sample PR-AUC. The first row reports the hard-data model, and the second row reports the model that combines hard data with the baseline soft-data set. The boundary rows keep the hard-data predictors and change the soft-data set as indicated. The row excluding text-based indicators keeps only the Michigan survey variables, while the professional-forecasts row adds Blue Chip forecasts of unemployment and real GDP growth. Forecast horizons are measured in months.

More generally, these exercises show that the soft-data contribution is not an artifact of a single source or of one particular boundary for the soft-data set. The exact ranking changes by horizon, especially at twelve months, but the main conclusion is stable: across these soft-data-set variants, the combined model generally outperforms the hard-data benchmark.

8.2 Forecasting Specification Choices

We next ask whether the soft-data contribution depends on how we construct the predictors used in the forecasting model. One possibility is that the baseline specification

summarizes the underlying data too restrictively. In that case, soft data could appear useful because they capture information that a richer hard-data specification would otherwise include. We examine this possibility by considering alternative specifications that allow the forecasting model to use the underlying data in richer ways. First, we allow for simple nonlinearities to capture the possibility that recession risk responds disproportionately when indicators move far from normal conditions. Second, we change the variance-explained threshold used to choose how many factors to retain from each of the five hard-data categories and the soft-data category. Table 7 reports, for each specification, PR-AUC for the hard-data model and for the model that adds soft data to that same specification.

The first two alternatives allow simple nonlinearities in two ways: one constructs squared baseline factors, while the other extracts additional factors from squared underlying predictors.[17] Both nonlinear specifications improve the hard-data model relative to the baseline hard-data specification at every horizon, so these exercises strengthen the hard-data benchmark. Even so, the combined model still improves on the strengthened hard-data benchmark in nearly all nonlinear specifications. The only exception is the six-month horizon when factors are extracted from squared variables. The improvement is especially large at the twelve-month horizon when squared factors are included.

The second set of alternatives changes the variance-explained threshold used to choose the number of factors within each predictor group. In the baseline specification, we retain enough factors to explain at least 50 percent of the variation in each group. The 30 percent threshold usually retains fewer factors, while the 80 percent threshold retains more. Lowering the threshold generally weakens the hard-data model, while raising it improves hard-data PR-AUC at the one-, three-, and twelve-month horizons. Adding soft data still improves PR-AUC under both threshold choices in nearly all cases. The only exception is the twelve-month horizon under the 80 percent factor threshold, in which case the combined model's PR-AUC is slightly below that of the hard data model.

[17]The squared-factor specification uses squared first principal components from each data group. The squared-variable specification instead extracts one additional factor from squared predictors within each group. In both cases, we use the contemporaneous extra factor when available and otherwise use the first available lag up to two months.

Table 7: Alternative Forecasting Specifications

(a) Hard data only

Specification	1 month	3 months	6 months	12 months
Baseline	0.610	0.531	0.478	0.315
Squared factors	0.680	0.551	0.524	0.370
Factors from squared variables	0.693	0.557	0.522	0.345
30% factor threshold	0.569	0.450	0.375	0.380
80% factor threshold	0.680	0.597	0.451	0.335

(b) Hard and soft data

Specification	1 month	3 months	6 months	12 months
Baseline	0.768	0.653	0.559	0.382
Squared factors	0.716	0.664	0.550	0.500
Factors from squared variables	0.787	0.676	0.499	0.415
30% factor threshold	0.735	0.594	0.467	0.408
80% factor threshold	0.802	0.678	0.487	0.313

Notes: The table reports out-of-sample PR-AUC for models estimated using hard data only and using both hard and soft data under the same forecasting specification. The squared-factors specification adds element-wise squares of the main factors to the logit. The factors-from-squared-variables specification extracts additional factors from squared raw predictors. The factor-threshold rows change the variance-explained threshold used to select the number of factors in each category. Forecast horizons are measured in months.

Overall, these robustness checks support the main conclusion that soft data add recession-predictive information beyond the baseline hard-data model. The gains vary across horizons and specifications, but the soft-data contribution remains present across most alternatives, including several richer hard-data benchmarks. The results also suggest that the baseline finding is not an artifact of one particular modeling choice or soft-data source: the contribution does not depend entirely on any one soft-data source, on the inclusion of text-based indicators, or on an obviously weak hard-data benchmark. Online Appendix B.3 shows similar results when we estimate the logit with LASSO or ridge penalties instead of the elastic-net baseline.

9 Conclusion

This paper shows that a relatively small set of soft indicators contains meaningful real-time information about recession risk. Despite using far fewer predictors than the hard-data model, the soft-data model performs competitively on its own, especially at short and

medium horizons.

We furthermore show that combining hard and soft indicators tend to improve recession-warning performance. It also helps distinguish cases in which weak sentiment, uncertainty, and narratives line up with broader economic deterioration from cases in which the hard data provide little support. Viewed this way, vibecessions are not simply episodes in which sentiment diverges from the hard data. They are cases in which forecasters need to assess whether soft-data weakness contains independent warning information or instead reflects an uncorroborated false alarm. Our results suggest that soft data should be monitored as a complement to traditional hard indicators.

These results leave two questions for future work. First, measurement matters. The soft indicators used here are only a narrow subset of the information contained in surveys, text, and other real-time accounts of economic conditions. Building more systematic measures of these signals could improve recession assessment and help identify which dimensions of soft information are most useful. Second, our results raise the question of what distinct information soft data capture. They suggest that soft indicators do more than anticipate later hard-data releases. That additional content may reflect expectations, uncertainty, narratives, or other forces that shape spending, hiring, and investment. Understanding these channels would help interpret soft-data signals and connect them to broader macroeconomic outcomes.

References

Ang, A., M. Piazzesi, and M. Wei (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics 131*(1–2), 359–403.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

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

Bańbura, M. and G. Rünstler (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting 27*(2), 333–346.

Berge, T. J. and Ò. Jordà (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics 3*(2), 246–277.

Bianchi, F., S. C. Ludvigson, and S. Ma (2022). Belief distortions and macroeconomic fluctuations. *American Economic Review 112*(7), 2269–2315.

Bluwstein, K., M. Buckmann, A. Joseph, S. Kapadia, and Ö. Şimşek (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics 145*, 103773.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition 30*(7), 1145–1159.

Buckmann, M., A. Joseph, and H. Robertson (2023). An interpretable machine learning workflow with an application to economic forecasting. *International Journal of Central Banking 19*(4), 449–552.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics 118*(1), 269–298.

Chauvet, M. and S. Potter (2005). Forecasting recessions using the yield curve. *Journal of Forecasting 24*(2), 77–103.

Davig, T. and A. S. Hall (2019). Recession forecasting using Bayesian classification. *International Journal of Forecasting 35*(3), 848–867.

Estrella, A. (2005). The yield curve and recessions. *The International Economy 19*(3), 8.

Estrella, A. and G. A. Hardouvelis (1991). The term structure as a predictor of real economic activity. *The Journal of Finance 46*(2), 555–576.

Estrella, A. and F. S. Mishkin (1998). Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics 80*(1), 45–61.

Filippou, I., C. Garciga, J. Mitchell, and M. T. Nguyen (2024, April). Regional economic sentiment: Constructing quantitative estimates from the Beige Book and testing their ability to forecast recessions. *Economic Commentary 2024*(08), 1–8.

Furno, F. and D. Giannone (2024). Nowcasting recession risk. In *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, pp. 156–186. Edward Elgar Publishing.

Giannone, D., L. Reichlin, and D. Small (2008, May). Nowcasting: The Real-Time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Goulet Coulombe, P., M. Leroux, D. Stevanovic, and S. Surprenant (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics 37*(5), 920–964.

Huang, D., F. Jiang, K. Li, G. Tong, and G. Zhou (2022). Scaled PCA: A new approach to dimension reduction. *Management Science 68*(3), 1678–1695.

Lahiri, K., G. Monokroussos, and Y. Zhao (2016). Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics 31*(7), 1254–1275.

Liu, W. and E. Moench (2016). What predicts US recessions? *International Journal of Forecasting 32*(4), 1138–1150.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic Perspectives 18*(2), 29–50.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, Volume 30. Curran Associates, Inc.

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

Medeiros, M. C., G. F. Vasconcelos, Á. Veiga, and E. Zilberman (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics 39*(1), 98–119.

Michaillat, P. (2025). Early and accurate recession detection using classifiers on the anticipation-precision frontier. Preprint 2506.09664, arXiv.

Rudebusch, G. D. and J. C. Williams (2009). Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics 27*(4), 492–503.

Sahm, C. (2019). Direct stimulus payments to individuals. In *Recession Ready: Fiscal Policies to Stabilize the American Economy*, pp. 67–92. The Hamilton Project.

Schnatz, B. and A. D'Agostino (2012, August). Survey-based nowcasting of US growth: A Real-Time forecast comparison over more than 40 years. Working Paper Series 1455, European Central Bank.

Shapiro, A. H., M. Sudhof, and D. J. Wilson (2022). Measuring news sentiment. *Journal of Econometrics 228*(2), 221–243.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2014). Estimating turning points using large data sets. *Journal of Econometrics 178*, 368–381.

Vrontos, S. D., J. Galakis, and I. D. Vrontos (2021). Modeling and predicting US recessions using machine learning techniques. *International Journal of Forecasting 37*(2), 647–671.

Wallis, K. F. (1986). Forecasting with an econometric model: The "ragged edge" problem. *Journal of Forecasting 5*(1), 1–13.

Wright, J. H. (2006, February). The yield curve and predicting recessions. Finance and Economics Discussion Series 2006-07, Board of Governors of the Federal Reserve System.

For Online Publication
APPENDIX

Petrosky-Nadeau, Sung, and Wilson,
"Do Vibes Predict Recessions? Evidence from a Big-Data Forecasting Framework"

A Real-Time Out-of-Sample Forecasts

In this section, we describe how we construct real-time out-of-sample recession forecasts. We first describe the data sources and their real-time availability. We then explain how we transform those data into low-dimensional predictors. Finally, we describe forecast estimation and hyperparameter tuning.

A.1 Data Sources and Availability

FRED-MD. The FRED-MD monthly vintage dataset (McCracken and Ng, 2016) provides real-time monthly vintages for more than 130 macroeconomic series. The public real-time FRED-MD vintages begin in August 1999, so our out-of-sample forecast evaluation begins with that vintage. For our analysis, we reorganize the series into five groups: real activity, housing, the yield curve, other financial indicators, and inflation.[18] Table A1 lists our grouping and the variables in each group. The source column reports the source used in the constructed panel: most variables come from FRED-MD, while FRED and ALFRED identify high-frequency source pulls used to align or supplement selected series.

We modify the FRED-MD datasets in three ways. First, we exclude the University of Michigan Consumer Sentiment Index because the soft-data block separately includes the five Michigan survey variables that make up the index, along with expected unemployment. Second, for a limited set of daily or weekly series available from FRED or ALFRED, we keep direct source pulls as a backup to the monthly FRED-MD vintage values. In the month-end historical forecasting exercise studied in this paper, daily series are generally complete by the forecast date and are aggregated to monthly values. The direct high-frequency pulls are therefore most useful for aligning selected weekly series and for supporting current-vintage or middle-of-the-month forecast extensions, which are not the focus of this paper. The S&P 500 itself is included in FRED-MD; the separate S&P industrials series, when included, comes from the optional Haver source. Third, at each vintage date, we drop variables with fewer than ten years of available observations in that vintage sample. In our sample, this rule affects a small set of variables, including CBOE volatility indexes, ISM manufacturing survey indexes, money and credit aggregates, and a few price-index and equity-market series.

[18] The original release groups these into eight categories: output and income, labor market, housing, consumption, orders, and inventories, money and credit, interest and exchange rates, prices, and stock market.

Soft data. We introduce soft-data variables that capture expectations and perceptions across different groups in the economy. The baseline soft-data panel contains six Michigan survey variables and three text-based measures. Five of the Michigan variables are the components of the Index of Consumer Sentiment: current personal finances, expected personal finances, expected business conditions over one year, expected business conditions over five years, and current buying conditions. The sixth Michigan variable is expected unemployment over the next year. The text-based measures are the Economic Policy Uncertainty Index, the Daily News Sentiment Index, and the Beige Book sentiment index. Blue Chip professional forecasts are optional proprietary inputs and enter only in the robustness exercise that expands the soft-data boundary. Tables A2 and A3 list the baseline variables, source and timing assumptions, and the optional Blue Chip inputs used in the professional-forecasts robustness exercise.

The soft-data panel combines survey and text-based sources with different publication schedules. For the University of Michigan Survey of Consumers, the data files available from the Michigan website provide the long historical series used to construct the historical panel, but these files lag the latest monthly release by four weeks. To better align the real-time availability of Michigan data, we complement these files with saved final-results PDFs from the main Michigan release page. These final-results releases provide timely latest-month values for the five Index of Consumer Sentiment components listed above. We parse saved PDFs matching `umich_final_results_*.pdf` to fill these latest-month component values.

The Michigan unemployment-expectations variable is treated differently. It is not part of the final-results component table behind the Index of Consumer Sentiment. We therefore treat the expected change in unemployment over the next year as lagged by about one month. In constructing real-time predictors, vintage assembly masks its same-month value in the concurrent soft-data panel so the latest vintage does not assume that the unemployment-expectations data were available by month-end.

The text-based sources require a separate timing caveat. The underlying newspaper articles, daily news stories, and Beige Book reports are contemporaneous information, but the public indexes are later data products. The EPU index is a monthly newspaper-based index introduced by Baker et al. (2016); the Daily News Sentiment Index is a daily news-based index introduced by Shapiro et al. (2022); and the Beige Book sentiment index is constructed from Beige Book text and introduced by Filippou et al. (2024). We therefore interpret the text-based variables as measurements of contemporaneous narratives rather than as fully real-time historical data products, and we assess this timing concern by excluding text-based indicators in the robustness exercises.

Transformations and variable lists. We transform each variable to achieve stationarity using one of the following transformation codes. For log transformations, we compute the log only for strictly positive source values and treat nonpositive values as missing before differencing. For the year-over-year percent-change transformation in the final row, we require both x_{it} and $x_{i,t-12}$ to be strictly positive. If either value is nonpositive, the transformed observation is treated as missing.

Symbol	Transformation
lv	x_{it}
Δlv	$x_{it} - x_{it-1}$
$\bar{\Delta}$lv	$x_{it} - x_{it-12}$
ln	$\ln(x_{it})$
$\Delta \ln$	$\ln(x_{it}) - \ln(x_{it-1})$
$\bar{\Delta} \ln$	$\ln(x_{it}) - \ln(x_{it-12})$
Δlv/lv	$(x_{it} - x_{it-1})/x_{it-1}$
$\bar{\Delta}$lv/lv	$(x_{it} - x_{it-12})/x_{i,t-12}$

Table A1: **FRED-MD Data Variables**

No.	Variable	Description	Trans.	Freq.	Data Start	Source
Real Activity						
1	RPI	Real personal income	$\bar{\Delta}\ln$	M	1959M1	FREDMD
2	W875RX1	Real personal income excl. transfers	$\bar{\Delta}\ln$	M	1959M1	FREDMD
3	INDPRO	Industrial production index	$\bar{\Delta}\ln$	M	1959M1	FREDMD
4	IPFPNSS	IP: Final products and nonindustrial supplies	$\bar{\Delta}\ln$	M	1959M1	FREDMD
5	IPFINAL	IP: Final products	$\bar{\Delta}\ln$	M	1959M1	FREDMD
6	IPCONGD	IP: Consumer goods	$\bar{\Delta}\ln$	M	1959M1	FREDMD
7	IPDCONGD	IP: Durable consumer goods	$\bar{\Delta}\ln$	M	1959M1	FREDMD
8	IPNCONGD	IP: Nondurable consumer goods	$\bar{\Delta}\ln$	M	1959M1	FREDMD
9	IPBUSEQ	IP: Business equipment	$\bar{\Delta}\ln$	M	1959M1	FREDMD
10	IPMAT	IP: Materials	$\bar{\Delta}\ln$	M	1959M1	FREDMD
11	IPDMAT	IP: Durable materials	$\bar{\Delta}\ln$	M	1959M1	FREDMD
12	IPNMAT	IP: Nondurable materials	$\bar{\Delta}\ln$	M	1959M1	FREDMD
13	IPMANSICS	IP: Manufacturing (SIC)	$\bar{\Delta}\ln$	M	1959M1	FREDMD
14	IPB51222S	IP: Residential utilities	$\bar{\Delta}\ln$	M	1972M1	FREDMD
15	IPFUELS	IP: Fuels	$\bar{\Delta}\ln$	M	1959M1	FREDMD
16	NAPMPI	ISM Manufacturing: Production Index	lv	M	1959M1	FREDMD
17	CUMFNS	Capacity utilization: Manufacturing	lv	M	1959M1	FREDMD
18	HWI	Help-wanted index	lv	M	1959M1	FREDMD
19	HWIURATIO	Help wanted/unemployed ratio	lv	M	1959M1	FREDMD
20	CLF16OV	Civilian labor force	$\bar{\Delta}\ln$	M	1959M1	FREDMD
21	CE16OV	Civilian employment	$\bar{\Delta}\ln$	M	1959M1	FREDMD
22	UNRATE	Unemployment rate	lv	M	1959M1	FREDMD
23	UEMPMEAN	Average duration of unemployment (weeks)	lv	M	1959M1	FREDMD
24	UEMPLT5	Unemployed less than 5 weeks	lv	M	1959M1	FREDMD
25	UEMP5TO14	Unemployed 5–14 weeks	lv	M	1959M1	FREDMD
26	UEMP15OV	Unemployed 15 weeks and over	lv	M	1959M1	FREDMD
27	UEMP15T26	Unemployed 15–26 weeks	lv	M	1959M1	FREDMD
28	UEMP27OV	Unemployed 27 weeks and over	lv	M	1959M1	FREDMD
29	initial_claims	Initial unemployment claims	lv	W	1959M1	ALFRED
30	PAYEMS	Total nonfarm payroll employment	$\bar{\Delta}\ln$	M	1959M1	ALFRED
31	USGOOD	Employees: Goods-producing	$\bar{\Delta}\ln$	M	1959M1	FREDMD
32	CES1021000001	Employees: Mining and logging	$\bar{\Delta}\ln$	M	1959M1	FREDMD
33	USCONS	Employees: Construction	$\bar{\Delta}\ln$	M	1959M1	FREDMD
34	MANEMP	Employees: Manufacturing	$\bar{\Delta}\ln$	M	1959M1	FREDMD
35	DMANEMP	Employees: Durable goods	$\bar{\Delta}\ln$	M	1959M1	FREDMD
36	NDMANEMP	Employees: Nondurable goods	$\bar{\Delta}\ln$	M	1959M1	FREDMD
37	SRVPRD	Employees: Service-providing	$\bar{\Delta}\ln$	M	1939M1	FREDMD
38	USTPU	Employees: Trade, transportation, utilities	$\bar{\Delta}\ln$	M	1959M1	FREDMD
39	USWTRADE	Employees: Wholesale trade	$\bar{\Delta}\ln$	M	1959M1	FREDMD
40	USTRADE	Employees: Retail trade	$\bar{\Delta}\ln$	M	1959M1	FREDMD
41	USFIRE	Employees: Financial activities	$\bar{\Delta}\ln$	M	1959M1	FREDMD
42	USGOVT	Employees: Government	$\bar{\Delta}\ln$	M	1959M1	FREDMD
43	CES0600000007	Avg weekly hours: Goods production	lv	M	1959M1	FREDMD
44	AWOTMAN	Avg overtime hours: Manufacturing	lv	M	1959M1	FREDMD
45	AWHMAN	Avg weekly hours: Manufacturing	lv	M	1959M1	FREDMD
46	NAPMEI	ISM Manufacturing: Employment Index	lv	M	1959M1	FREDMD
47	CES0600000008	Avg hourly earnings: Goods production	$\bar{\Delta}\ln$	M	1959M1	FREDMD
48	CES2000000008	Avg hourly earnings: Construction	$\bar{\Delta}\ln$	M	1959M1	FREDMD
49	CES3000000008	Avg hourly earnings: Manufacturing	$\bar{\Delta}\ln$	M	1959M1	FREDMD

Continued on next page

Table A1 continued from previous page

No.	Variable	Description	Trans.	Freq.	Data Start	Source
50	DPCERA3M086SBEA	Real personal consumption expenditures	$\Delta \ln$	M	1959M1	FREDMD
51	CMRMTSPLx	Real manufacturing and trade sales	$\bar{\Delta} \ln$	M	1959M1	FREDMD
52	RETAILx	Retail and food services sales	$\bar{\Delta} \ln$	M	1959M1	FREDMD
53	NAPM	ISM Manufacturing: PMI Composite Index	lv	M	1959M1	FREDMD
54	NAPMNOI	ISM Manufacturing: New Orders Index	lv	M	1959M1	FREDMD
55	NAPMSDI	ISM Manufacturing: Supplier Deliveries Index	lv	M	1959M1	FREDMD
56	NAPMII	ISM Manufacturing: Inventories Index	lv	M	1959M1	FREDMD
57	ACOGNO	New orders for consumer goods	$\bar{\Delta} \ln$	M	1992M2	FREDMD
58	AMDMNOx	New orders for durable goods	$\bar{\Delta} \ln$	M	1959M1	FREDMD
59	ANDENOx	New orders for nondefense capital goods	$\bar{\Delta} \ln$	M	1968M2	FREDMD
60	AMDMUOx	Unfilled orders for durable goods	$\bar{\Delta} \ln$	M	1959M1	FREDMD
61	BUSINVx	Total business inventories	$\bar{\Delta} \ln$	M	1959M1	FREDMD
62	ISRATIOx	Inventories to sales ratio	lv	M	1959M1	FREDMD
Housing						
63	HOUST	Housing starts: Total	$\bar{\Delta} \ln$	M	1959M1	FREDMD
64	HOUSTNE	Housing starts: Northeast	$\bar{\Delta} \ln$	M	1959M1	FREDMD
65	HOUSTMW	Housing starts: Midwest	$\bar{\Delta} \ln$	M	1959M1	FREDMD
66	HOUSTS	Housing starts: South	$\bar{\Delta} \ln$	M	1959M1	FREDMD
67	HOUSTW	Housing starts: West	$\bar{\Delta} \ln$	M	1959M1	FREDMD
68	PERMIT	New private housing permits	$\bar{\Delta} \ln$	M	1960M1	FREDMD
69	PERMITNE	New permits: Northeast	$\bar{\Delta} \ln$	M	1960M1	FREDMD
70	PERMITMW	New permits: Midwest	$\bar{\Delta} \ln$	M	1960M1	FREDMD
71	PERMITS	New permits: South	$\bar{\Delta} \ln$	M	1960M1	FREDMD
72	PERMITW	New permits: West	$\bar{\Delta} \ln$	M	1960M1	FREDMD
Yield Curve						
73	T10Y3M	10-year - 3-month TBill	lv	D	1982M1	FRED
74	T10Y2Y	10-year - 2-year TBill	lv	D	1976M6	FRED
75	T5Y3M	5-year - 3-month TBill	lv	D	1962M1	FRED
76	COMPAPFFx	3-month commercial paper - FFR	lv	D	1959M1	FRED
77	TB3SMFFM	3-month TBill - FFR	lv	D	1959M1	FRED
78	TB6SMFFM	6-month TBill - FFR	lv	D	1959M1	FRED
79	T1YFFM	1-year TBill - FFR	lv	D	1959M1	FRED
80	T5YFFM	5-year TBill - FFR	lv	D	1959M1	FRED
81	T10YFFM	10-year TBill - FFR	lv	D	1959M1	FRED
82	AAAFFM	AAA corporate bond - FFR	lv	D	1959M1	FRED
83	BAAFFM	BAA corporate bond - FFR	lv	D	1959M1	FRED
Other Financial						
84	FEDFUNDS	Effective Federal Funds Rate	lv	D	1954M7	FRED
85	CP3Mx	3-month commercial paper rate	lv	D	1959M1	FRED
86	TB3MS	3-month TBill	lv	D	1954M1	FRED
87	TB6MS	6-month TBill	lv	D	1958M12	FRED
88	GS1	1-year Treasury constant maturity	lv	D	1959M1	FRED
89	GS5	5-year Treasury constant maturity	lv	D	1959M1	FRED
90	GS10	10-year Treasury constant maturity	lv	D	1959M1	FRED
91	AAA	Moody's seasoned AAA corporate bond	lv	D	1959M1	FRED
92	BAA	Moody's seasoned BAA corporate bond	lv	D	1959M1	FRED
93	TWEXAFEGSMTHx	Trade-weighted US dollar index	$\bar{\Delta} \ln$	D	1973M1	FRED
94	EXSZUSx	Swiss franc/US dollar exchange rate	$\bar{\Delta} \ln$	D	1959M1	FRED

Continued on next page

Table A1 continued from previous page

No.	Variable	Description	Trans.	Freq.	Data Start	Source
95	EXJPUSx	Japanese yen/US dollar exchange rate	$\Delta \ln$	D	1959M1	FRED
96	EXUSUKx	UK pound/US dollar exchange rate	$\bar{\Delta} \ln$	D	1959M1	FRED
97	EXCAUSx	Canadian dollar/US dollar exchange rate	$\bar{\Delta} \ln$	D	1959M1	FRED
98	M1SL	M1 money stock	$\bar{\Delta} \ln$	M	1959M1	FRED
99	M2SL	M2 money stock	$\bar{\Delta} \ln$	M	1959M1	FRED
100	M2REAL	Real M2 money stock	$\bar{\Delta} \ln$	M	1959M1	FRED
101	BOGMBASE	Monetary base	$\bar{\Delta} \ln$	M	1959M1	FRED
102	TOTRESNS	Total reserves of depository institutions	$\bar{\Delta} \ln$	M	1959M1	FRED
103	NONBORRES	Non-borrowed reserves	$\bar{\Delta}$lv/lv	M	1959M1	FRED
104	BUSLOANS	Commercial and industrial loans	$\bar{\Delta} \ln$	M	1959M1	FRED
105	REALLN	Real estate loans	$\bar{\Delta} \ln$	M	1959M1	FRED
106	NONREVSL	Non-revolving consumer credit	$\bar{\Delta} \ln$	M	1959M1	FRED
107	CONSPI	Non-revolving credit to personal income ratio	$\bar{\Delta} \ln$	M	1959M1	FRED
108	MZMSL	MZM money stock	$\bar{\Delta} \ln$	M	1959M1	FRED
109	DTCOLNVHFNM	Consumer loans: Auto	$\bar{\Delta} \ln$	M	1959M1	FRED
110	DTCTHFNM	Consumer loans outstanding	$\bar{\Delta} \ln$	M	1959M1	FRED
111	INVEST	Securities in bank credit	$\bar{\Delta} \ln$	M	1959M1	FRED
112	VIXCLSx	CBOE Volatility Index (VIX)	lv	D	1962M7	FRED
113	S&P 500	S&P 500 stock market index	$\Delta \ln$	M	1959M1	FREDMD
114	S&P: indust	S&P Industrial sector index	$\Delta \ln$	M	1959M1	Haver
Inflation						
115	WPSFD49207	PPI: Finished goods	$\bar{\Delta} \ln$	M	1959M1	FREDMD
116	WPSFD49502	PPI: Finished consumer goods	$\bar{\Delta} \ln$	M	1959M1	FREDMD
117	WPSID61	PPI: Intermediate materials	$\bar{\Delta} \ln$	M	1959M1	FREDMD
118	WPSID62	PPI: Crude materials	$\bar{\Delta} \ln$	M	1959M1	FREDMD
119	PPICMM	PPI: Metals and metal products	$\bar{\Delta} \ln$	M	1959M1	FREDMD
120	NAPMPRI	ISM Manufacturing: Prices Index	$\bar{\Delta} \ln$	M	1959M1	FREDMD
121	CPIAUCSL	Consumer Price Index: All items	$\bar{\Delta} \ln$	M	1959M1	FREDMD
122	CPIAPPSL	CPI: Apparel	$\bar{\Delta} \ln$	M	1959M1	FREDMD
123	CPITRNSL	CPI: Transportation	$\bar{\Delta} \ln$	M	1959M1	FREDMD
124	CPIMEDSL	CPI: Medical care	$\bar{\Delta} \ln$	M	1959M1	FREDMD
125	CUSR0000SAC	CPI: Commodities	$\bar{\Delta} \ln$	M	1959M1	FREDMD
126	CUSR0000SAD	CPI: Durables	$\bar{\Delta} \ln$	M	1959M1	FREDMD
127	CUSR0000SAS	CPI: Services	$\bar{\Delta} \ln$	M	1959M1	FREDMD
128	CPIULFSL	CPI: All items less food and energy	$\bar{\Delta} \ln$	M	1959M1	FREDMD
129	CUSR0000SA0L2	CPI: All items less shelter	$\bar{\Delta} \ln$	M	1959M1	FREDMD
130	CUSR0000SA0L5	CPI: All items less medical care	$\bar{\Delta} \ln$	M	1959M1	FREDMD
131	PCEPI	PCE price index	$\bar{\Delta} \ln$	M	1959M1	FREDMD
132	DDURRG3M086SBEA	PCE: Durable goods	$\bar{\Delta} \ln$	M	1959M1	FREDMD
133	DNDGRG3M086SBEA	PCE: Nondurable goods	$\bar{\Delta} \ln$	M	1959M1	FREDMD
134	DSERRG3M086SBEA	PCE: Services	$\bar{\Delta} \ln$	M	1959M1	FREDMD
135	OILPRICEx	Crude oil price (WTI/Brent composite)	$\bar{\Delta} \ln$	D	1959M1	FRED

Table A2: **Soft Data Variables**

No.	Variable	Description	Trans.	Freq.	Data Start
1	epu	Economic Policy Uncertainty Index	lv	M	1900M1
2	news_sent	Daily News Sentiment Index	lv	D	1980M1
3	bb_sent	Beige Book sentiment	lv	M	1970M5
4	personal_finance_current	Current financial situation	lv	M	1946M2
5	personal_finance_expected	Expected change in financial situation	lv	M	1947M2
6	business_cond_1y	Expected business conditions, next year	lv	M	1947M2
7	business_cond_5y	Expected business conditions, next 5 years	lv	M	1951M11
8	buying_cond	Buying conditions, large household goods	lv	M	1951M2
9	unemp_change_1y	Expected change in unemployment	lv	M	1960M11

Table A3: **Soft-Data Sources and Timing Assumptions**

Source family	Variables	Source and timing assumption
Economic Policy Uncertainty	epu	Public monthly index from policyuncertainty.com. The newspaper coverage is contemporaneous, but the constructed index is a data product introduced by Baker et al. (2016).
Daily News Sentiment	news_sent	Public daily index from the Federal Reserve Bank of San Francisco. Daily observations are aggregated to month end. The index is introduced by Shapiro et al. (2022).
Beige Book sentiment	bb_sent	Public sentiment index from openICPSR. Beige Book text is released on the FOMC cycle, while the constructed index is introduced by Filippou et al. (2024).
Michigan Survey of Consumers	Six Michigan survey variables	Michigan website data files provide the long historical series and lag the latest monthly release by four weeks. Saved final-results releases fill timely latest-month values for the five Consumer Sentiment Index variables. The unemployment-expectations variable is treated as lagged by about one month.
Blue Chip professional forecasts	bc_*	Optional proprietary monthly inputs used only in the robustness exercise that adds professional forecasts. When supplied, they are treated as available by month end because the release occurs on the 10th of each month.

A.2 Predictor Construction

This subsection describes how we turn the real-time data panels into predictors for the recession-forecasting model. The main challenge is that the raw data are not a balanced monthly panel observed cleanly at each forecast date. Series are released with different

lags and frequencies, some current-month values are not yet available, some soft-data measures are observed irregularly, and some observations are missing or treated as unusable because they are extreme. We therefore first align each vintage to the information set available at month end, then handle missing values, and finally extract group-level factors from the transformed data.

Ragged edge problem. Our dataset combines variables with different sampling frequencies and release timing. Most FRED-MD series are monthly and typically become available with about a one-month lag. Other series are daily or weekly and update more frequently. The soft-data measures also have mixed timing: the Michigan survey and EPU are monthly, the Daily News Sentiment Index is daily and aggregated to month end, and the Beige Book sentiment index follows the Federal Open Market Committee (FOMC) release cycle.

We align all predictors to the information set available at the end of month t. For monthly series, we use only values available by month-end t, which typically correspond to reference month $t-1$. For daily and weekly series, we use all observations available as of the vintage date in month t, including within-month releases, to construct month-t predictors. Specifically, we project the series forward to month-end using an ARMA(p, q) model, choosing the lag order by minimizing the Bayesian information criterion over candidate specifications. We estimate the model using the most recent two years of daily observations or the most recent five years of weekly observations, selecting the AR and MA orders (p, q) by grid search over $p \in \{0, \ldots, 5\}$ and $q \in \{0, \ldots, 5\}$.[19]

Handling missing values. Our dataset contains missing values for four reasons. First, many monthly series are released with a lag, so the value for month t is not yet available at time t and is treated as missing in the real-time information set. Second, some series start later than others, so they are missing early in the sample. Third, some soft-data measures are not observed every month; for example, Beige Book sentiment follows the FOMC release cycle. Fourth, we treat a small number of extreme observations as missing. For example, initial unemployment claims at the onset of the COVID-19 pandemic exhibit unusually large spikes that reflect exceptional conditions, and we do not feed those raw values into the forecasting model.

We handle missing variables using an expectation-maximization (EM) algorithm. Intuitively, EM fills gaps by projecting each series onto the low-dimensional factor space implied by the rest of the panel, so missing values reflect the co-movement in the data rather than ad hoc interpolation. We describe the implementation details below.

Real-time estimation of factors. Before describing the factor-estimation algorithm, we define the horizon-specific forecast target used both for scaling predictors and for estimating the recession-forecasting model. For predictor month s and forecast horizon h, let $y_{s,h}$

[19] Because our historical out-of-sample evaluation uses month-end vintages, daily series are typically complete and rarely require ragged-edge adjustments. Weekly series are more likely to be incomplete even at month end, since the final weekly observation for the month may not yet be available. This issue is most pronounced for the most recent vintage, when both weekly series and daily series may have not yet realized all observations for the current month.

equal one if the economy is in recession in any month from $s+1$ through $s+h$, and zero otherwise. At forecast date t, this target is observed only when the full horizon-h window has been realized. Let $m_{t,h}$ denote the latest predictor month with a realized horizon-h target. For example, with data available through May 2026, six-month targets are realized through November 2025: the November 2025 target asks whether a recession occurs from December 2025 through May 2026, while the December 2025 target depends on June 2026 and is not yet realized.

For each out-of-sample forecasting exercise, we estimate factors using the corresponding vintage data, implementing the procedure in Bianchi et al. (2022). We retain the horizon index h in the notation below because the scaled PCA step is repeated separately for each forecast horizon. We repeat the following steps for each forecast date t and horizon h.

1. For each series, we flag outliers as observations that lie more than $10\times$ the interquartile range (IQR) from the distribution of that series. We set flagged observations as missing.

2. We standardize each series by subtracting its sample mean and dividing by its sample standard deviation. We denote the standardized series by $\bar{x}_{i,s}$, where s indexes months inside the vintage panel.

3. We scale each series by its estimated association with the horizon-specific forecast target over labeled months. For each predictor $\bar{x}_{i,s}$, we estimate the univariate OLS regression

$$y_{s,h} = \beta_{0,i,h} + \beta_{i,h}\bar{x}_{i,s} + \nu_{i,s,h}$$

   using only labeled months $s \leq m_{t,h}$, whose horizon-h targets have already been realized. Following Huang et al. (2022), we use the estimated slope $\hat{\beta}_{i,h}$ as a scaling factor and apply it to the full vintage panel. We then form the scaled series $\tilde{x}_{i,s,h} \equiv \hat{\beta}_{i,h}\bar{x}_{i,s}$ and construct the dataset from $\{\tilde{x}_{i,s,h}\}_i$.

4. Let $\tilde{x}^G_{s,h} = \left(\tilde{x}^G_{1,s,h}, \ldots, \tilde{x}^G_{n_G,s,h}\right)'$ denote the $n_G \times 1$ vector of scaled series in group G at month s for horizon h. For each group G, we extract r static factors from $\{\tilde{x}^G_{s,h}\}_s$ using principal component analysis (PCA). We assume the approximate factor structure

$$\tilde{x}^G_{i,s,h} = (\lambda^G_{i,h})' f^G_{s,h} + e^G_{i,s,h},$$

   where $f^G_{s,h}$ is an $r \times 1$ vector of common factors, $\lambda^G_{i,h}$ is a corresponding $r \times 1$ vector of factor loadings, and $e^G_{i,s,h}$ is an idiosyncratic error term. For each group and horizon, we choose the number of principal components as the smallest r such that the cumulative explained variance in the scaled data matrix is at least 50 percent.

   To handle missing values, we implement an EM algorithm on $\tilde{x}^G_{s,h}$: we initialize missing entries at zero (the series mean after standardization), estimate factor loadings and factors by PCA, and then update missing entries using the fitted factor model's implied values. We iterate this estimate-impute procedure until the mean squared change in the imputed values falls below 10^{-6}.

When constructing the forecasting predictors, we also decide whether the current-month factor is available in the latest month of a vintage. This decision is made before EM imputation. For each group, we compute the concurrent observed share as the number of nonmissing transformed group columns in the latest month divided by the number of transformed columns available for that group. We assign the current-month factor only when this observed share is at least 20 percent. If the observed share is below 20 percent, the current-month factor is left missing even though EM can fill the rest of the panel for PCA. This rule prevents a large category from receiving a concurrent factor simply because one variable is observed in the latest month.

The final predictor vector collects current conditions and recent dynamics from each group. For group G and horizon h, we include the current factor $f^G_{t,h}$ when it satisfies the observed-share rule, one- to three-month lags $f^G_{t-1,h}, \ldots, f^G_{t-3,h}$, the four- to six-month lag average $\frac{1}{3}\sum_{j=4}^{6} f^G_{t-j,h}$, and the seven- to twelve-month lag average $\frac{1}{6}\sum_{j=7}^{12} f^G_{t-j,h}$. Stacking these terms across the hard-data groups and the soft-data group gives the baseline horizon-specific predictor vector $X_{t,h}$. In the robustness exercises, we also consider nonlinear extensions that augment $X_{t,h}$ with squared factors or with factors extracted from squared underlying predictors; Section 8.2 describes these specifications.

A.3 Forecast Estimation and Hyperparameter Tuning

For each forecast horizon h, we estimate the regularized logit model using the same labeled-sample cutoff used in the scaled PCA step. At forecast date t, observations through $m_{t,h}$ are available for training and validation, while later observations are excluded from estimation even if their predictors are observed.

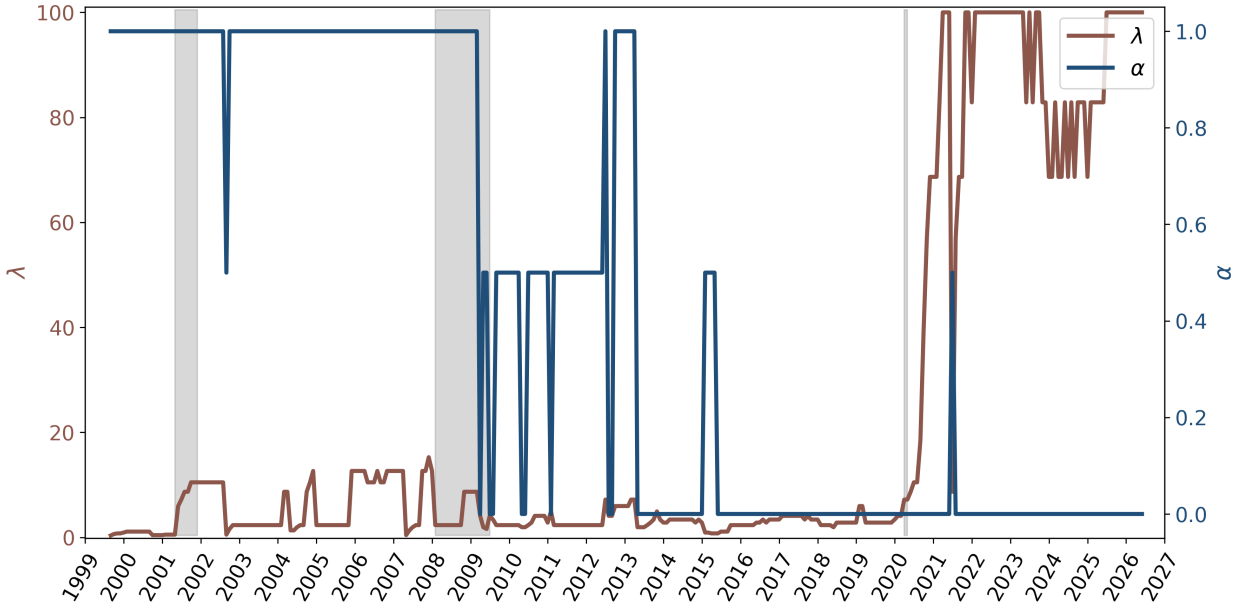
The elastic-net penalty depends on two hyperparameters: λ, which controls the overall strength of regularization, and α, which controls the mix between ridge and LASSO penalties. We evaluate λ on a 50-point geometric grid from 100 to 0.01 and evaluate $\alpha \in \{0, 0.5, 1\}$. The endpoints correspond to ridge and LASSO, respectively, while the midpoint is the elastic-net case.

At each forecast date and horizon, we choose hyperparameters by block validation. Validation blocks are the most recent non-overlapping 36-month blocks ending at $m_{t,h}$, using up to four blocks while requiring at least eight years of training data before each block. For a candidate pair (λ, α), the model is estimated on the observations before each validation block and evaluated on the block using average log loss. We aggregate the validation scores across available blocks and select the candidate with the lowest average log loss. After selecting the hyperparameters, we refit the model using all labeled observations through $m_{t,h}$ and then form the real-time forecast for date t.

This tuning procedure is repeated separately for every forecast date, forecast horizon, and information set. Figure A1 reports the selected λ and α over time for the combined model at each forecast horizon.

Figure A1: Selected Elastic-Net Hyperparameters in the Combined Model

(a) One-month horizon



(b) Three-month horizon

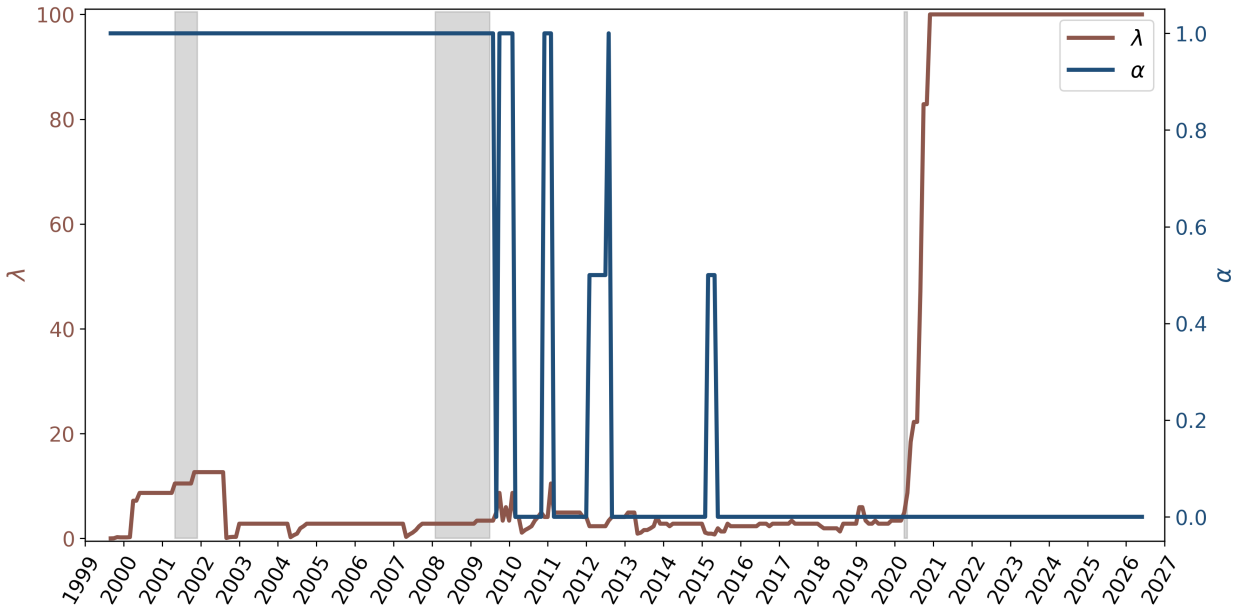
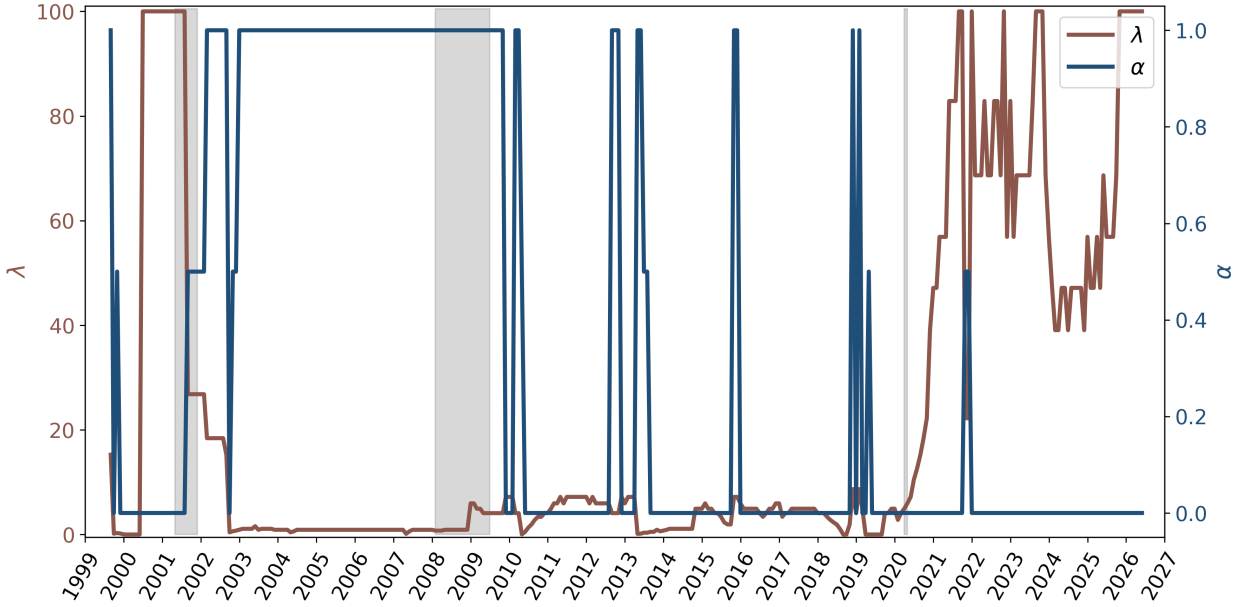
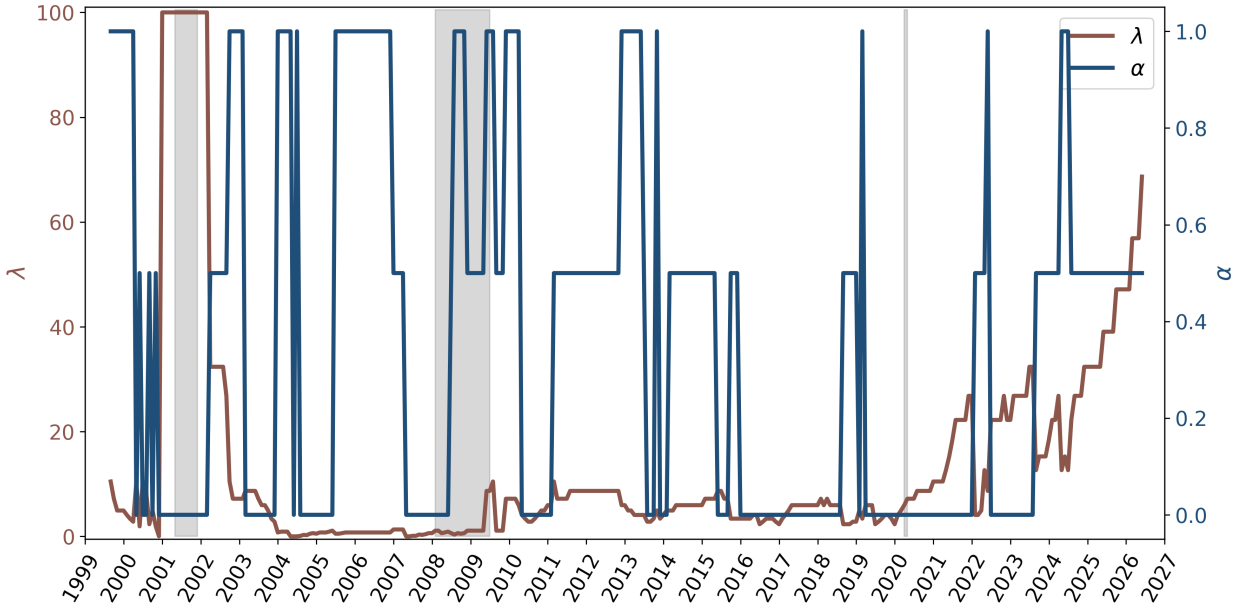


Figure A1: Selected Elastic-Net Hyperparameters in the Combined Model (continued)

(c) Six-month horizon



(d) Twelve-month horizon



Notes: The figure reports the selected elastic-net hyperparameters for the combined model at each real-time forecast date. The brown line shows λ, the overall regularization strength, and the navy line shows α, the mix between ridge and LASSO penalties.

B Supplementary Forecast Results

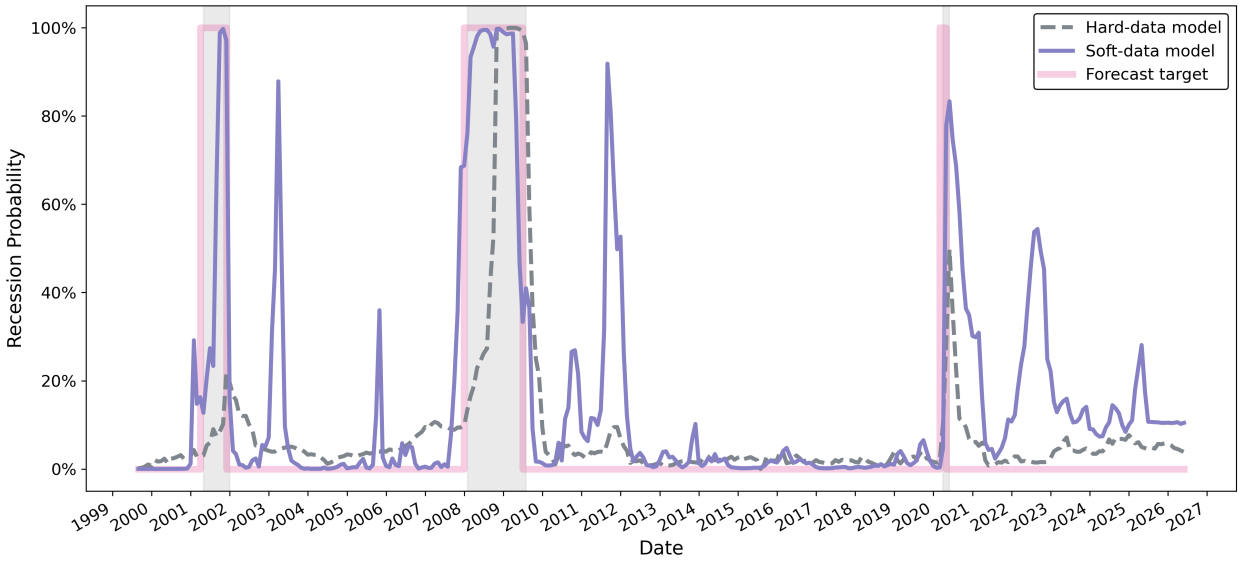
B.1 Forecast Probabilities at Other Horizons

The main text focuses on the six-month horizon. This subsection reports analogous out-of-sample recession probabilities at the one-, three-, and twelve-month horizons. For each horizon, panel (a) compares the hard-data and soft-data models, while panel (b) reports the combined model.

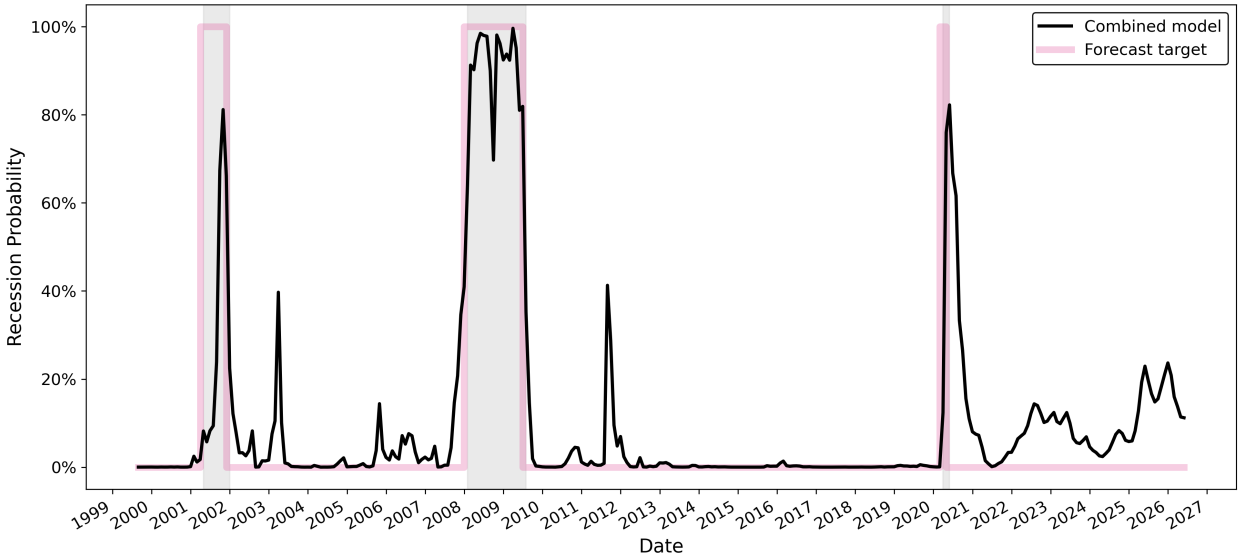
The figures show the same broad contrast emphasized in the main text. Soft-data forecasts tend to move more sharply than hard-data forecasts, especially around the 2001 recession and several expansion episodes. The combined model generally preserves large recession signals near downturns but remains more muted than the soft-data model in many non-recession periods. The twelve-month horizon is less sharp than the shorter horizons, which is consistent with the weaker twelve-month performance gains reported in the main tables.

Figure B1: Predicted Recession Risk over the Next One Month

(a) Hard-data model vs. soft-data model



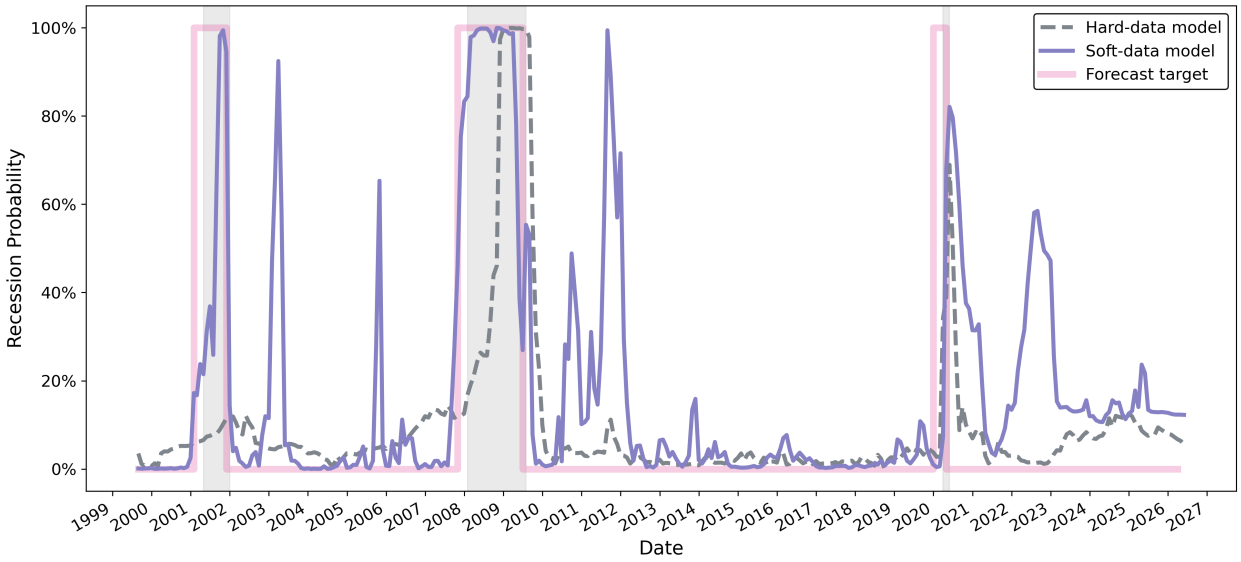
(b) Combined model



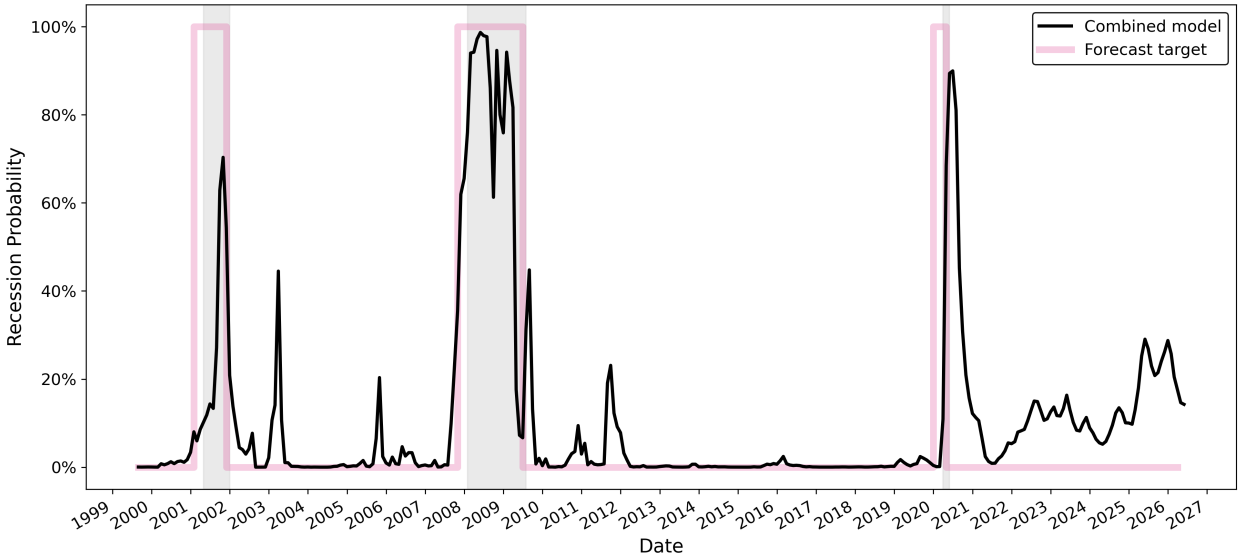
Notes: All plotted probabilities are out-of-sample forecasts of recession risk over the next one month as of each monthly real-time data vintage from August 1999 to May 2026. Panel (a) compares the hard-data model and the soft-data model. Panel (b) shows the combined model. The shaded areas indicate NBER recessions, and the pink line is the realized one-month-ahead forecast target.

Figure B2: Predicted Recession Risk over the Next Three Months

(a) Hard-data model vs. soft-data model



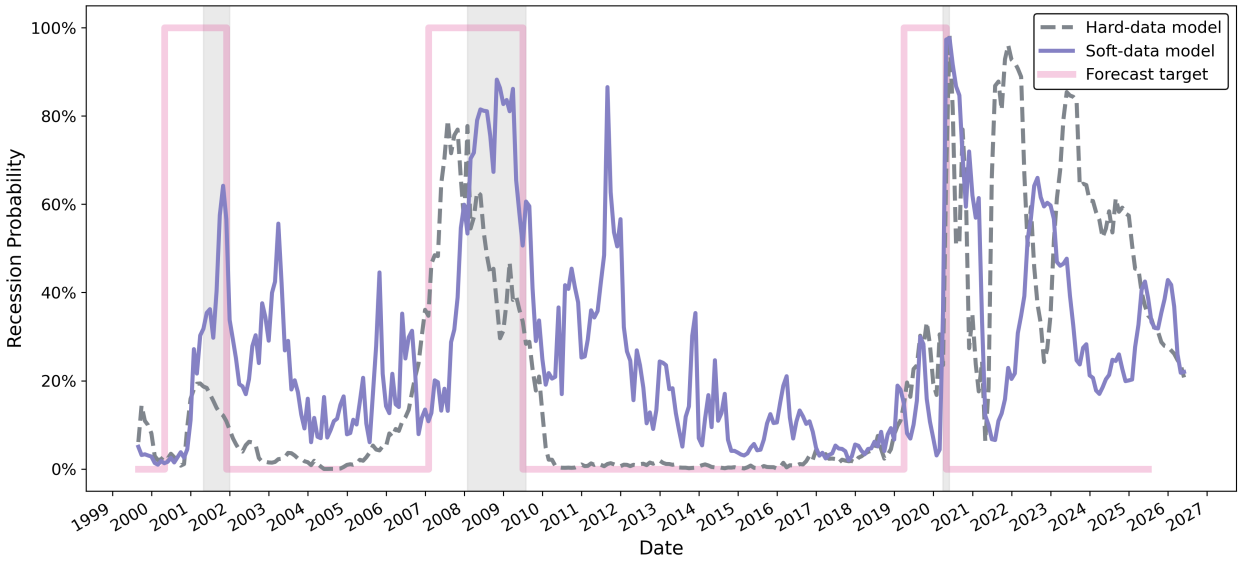
(b) Combined model



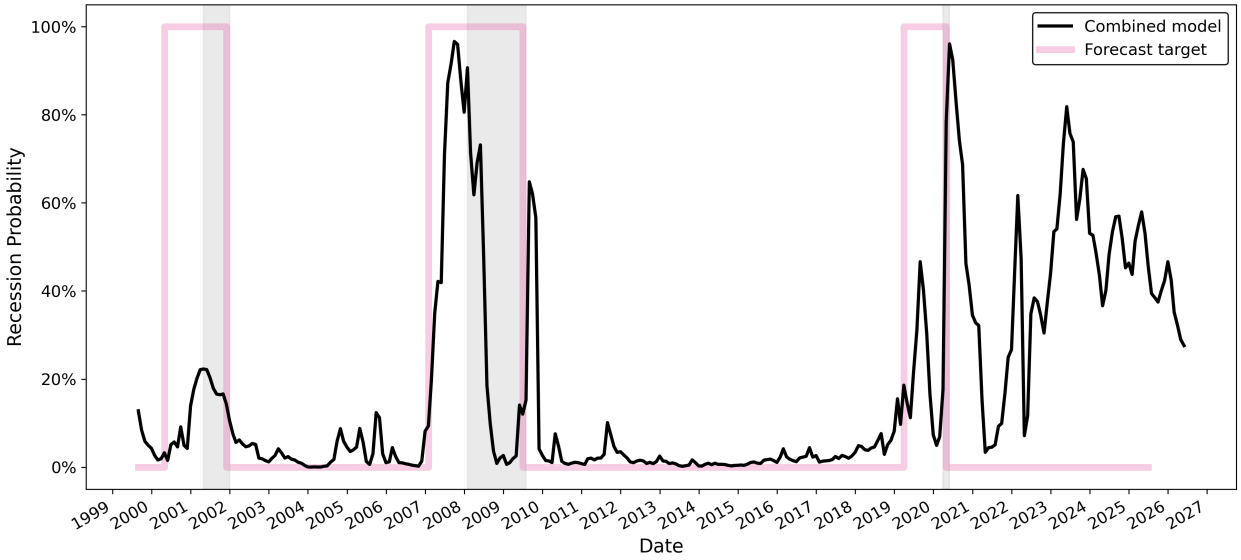
Notes: All plotted probabilities are out-of-sample forecasts of recession risk over the next three months as of each monthly real-time data vintage from August 1999 to May 2026. Panel (a) compares the hard-data model and the soft-data model. Panel (b) shows the combined model. The shaded areas indicate NBER recessions, and the pink line is the realized three-month-ahead forecast target.

Figure B3: Predicted Recession Risk over the Next Twelve Months

(a) Hard-data model vs. soft-data model



(b) Combined model



Notes: All plotted probabilities are out-of-sample forecasts of recession risk over the next twelve months as of each monthly real-time data vintage from August 1999 to May 2026. Panel (a) compares the hard-data model and the soft-data model. Panel (b) shows the combined model. The shaded areas indicate NBER recessions, and the pink line is the realized twelve-month-ahead forecast target.

B.2 Recession Signals at a 50 Percent Threshold

The main text reports binary recession-signal results using a 25 percent warning threshold. This subsection reports analogous results using a more stringent 50 percent threshold. This cutoff asks whether the signal-quality patterns in the main text depend on using a lower early-warning threshold.

Table B1: Recession Warning Signals at a 50 Percent Threshold

(a) Confusion matrix, six-month horizon

Realized outcome	Soft data only		Hard data only	
	Signaled recession	Did not signal	Signaled recession	Did not signal
Recession occurred	21	22	6	37
No recession occurred	25	248	8	265

(b) Precision and recall across horizons

Metric	Model	1 month	3 months	6 months	12 months
Precision	Hard data only	0.750	0.538	0.429	0.237
	Soft data only	0.588	0.488	0.457	0.412
Recall	Hard data only	0.321	0.206	0.140	0.230
	Soft data only	0.714	0.618	0.488	0.344

(c) F1 score across horizons

Model	1 month	3 months	6 months	12 months
Hard data only	0.450	0.298	0.211	0.233
Soft data only	0.645	0.545	0.472	0.375
Hard and soft data	0.717	0.655	0.559	0.231

Notes: Recession warning signals are defined as predicted recession probabilities of at least 50 percent. Panel (a) reports the six-month confusion matrix. Panels (b) and (c) report precision, recall, and F1 scores across forecast horizons. The evaluation sample includes forecast dates with realized horizon-specific targets.

At this stricter threshold, the soft-data model still has higher recall than the hard-data model at the one-, three-, and six-month horizons. The hard-data model has higher precision at the one- and three-month horizons, while precision is similar at six months and lower at twelve months. The combined model has the highest F1 score at the one-, three-, and six-month horizons, but not at the twelve-month horizon.

B.3 Alternative Regularization Methods

The baseline forecasts use an elastic-net penalty, which combines sparse selection from LASSO with continuous shrinkage from ridge regression. Table B2 compares the baseline elastic-net forecasts with pure LASSO and pure ridge forecasts for the hard-data model and the combined model. These are specification checks rather than alternative preferred models: the question is whether the main forecasting patterns depend on the particular mix of sparsity and shrinkage in the elastic net.

Table B2: Forecast Performance with Alternative Regularization Methods

(a) Hard data only

Specification	1 month	3 months	6 months	12 months
Baseline	0.610	0.531	0.478	0.315
LASSO	0.655	0.526	0.472	0.350
Ridge	0.617	0.524	0.504	0.346

(b) Hard and soft data

Specification	1 month	3 months	6 months	12 months
Baseline	0.768	0.653	0.559	0.382
LASSO	0.684	0.605	0.596	0.385
Ridge	0.741	0.673	0.553	0.391

Notes: The table reports out-of-sample PR-AUC across forecast horizons. The baseline specification uses the elastic-net penalty.

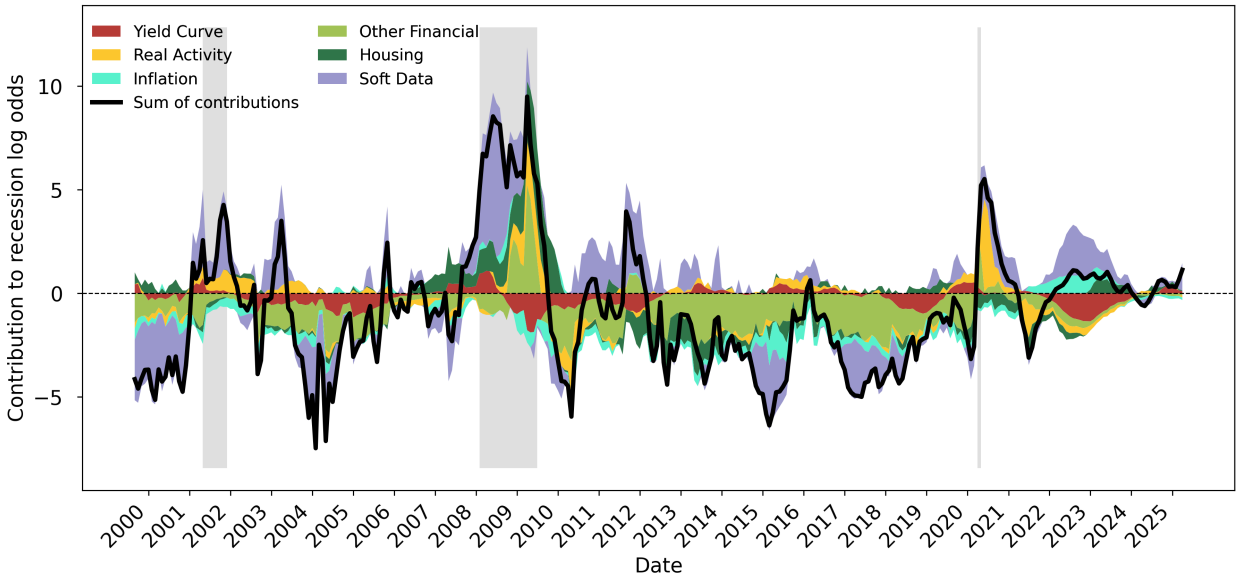
No single penalty dominates across horizons and information sets. LASSO and ridge improve some entries, but the elastic-net baseline remains competitive and gives a stable maintained specification for the main analysis.

B.4 Decomposition Figures at Other Horizons

The main text uses selected six-month episodes to explain how the combined model filters soft-data signals. This subsection reports the full Shapley decompositions for the one-, three-, six-, and twelve-month horizons. The decomposition is additive in fitted log odds, so positive values raise predicted recession log odds relative to the baseline and negative values lower them. Appendix C explains how to translate these log-odds contributions into probability terms.

Figure B4: Shapley Decomposition of Predicted Recession Log Odds by Horizon

(a) One-month horizon



(b) Three-month horizon

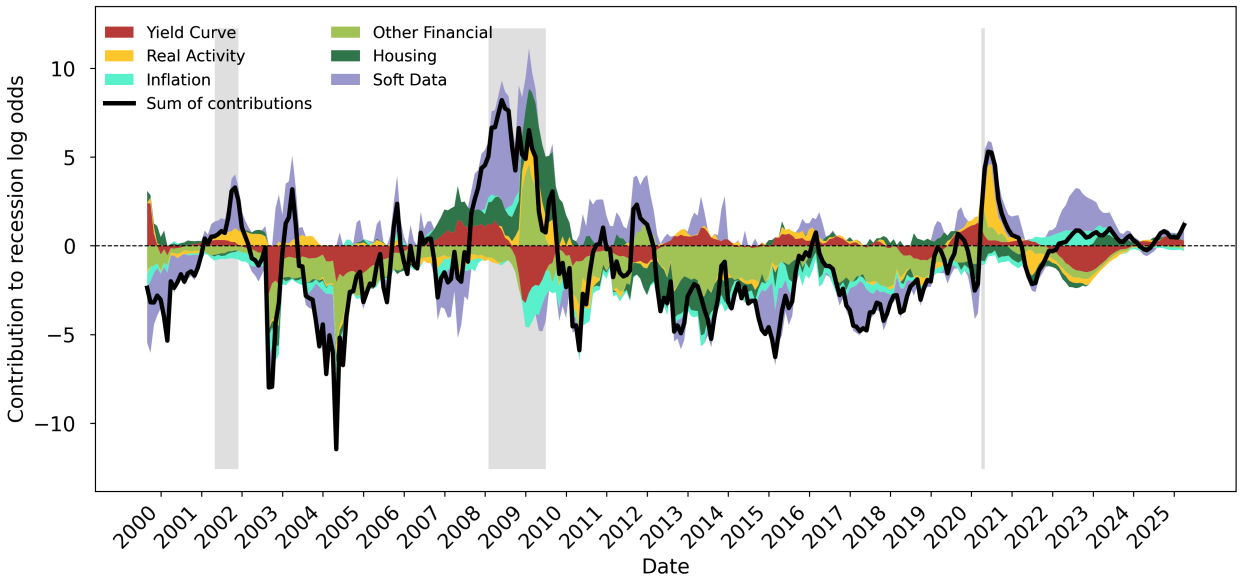
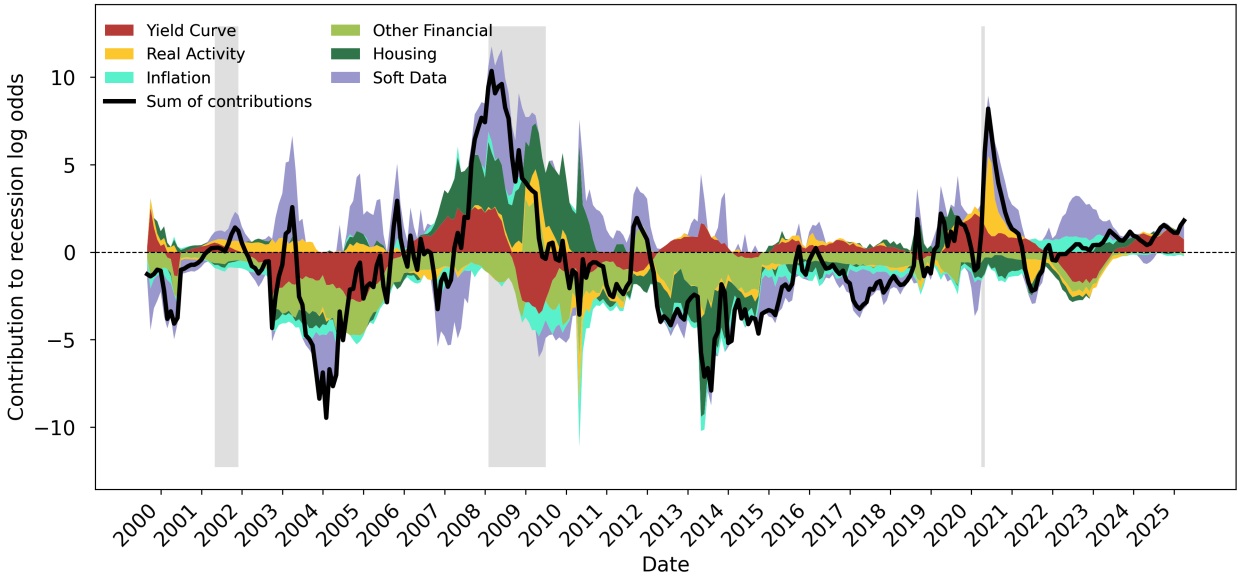
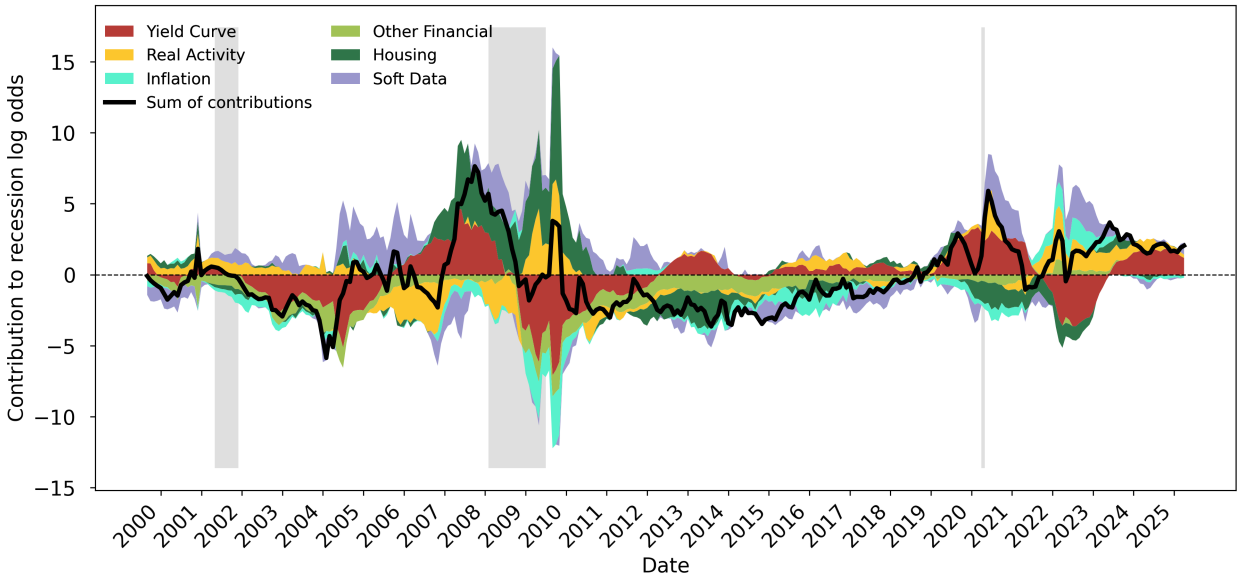


Figure B4: Shapley Decomposition by Horizon (continued)

(c) Six-month horizon



(d) Twelve-month horizon



Notes: The solid black line is the sum of the colored category contributions and equals the combined model's fitted recession log odds for the indicated forecast horizon minus the baseline log odds. Each colored area shows the contribution of the indicated category to that difference. Each category's contribution is the sum of the Shapley values from all individual variables (including lags) within the category. The grey shaded bars indicate NBER recessions.

Across horizons, the decompositions show that the category-level source of recession risk is not fixed. Soft data matter in several episodes, but their effect is amplified or offset by different hard-data categories depending on the horizon and historical period.

C From Log-Odds to Probability Points

This appendix explains how to interpret the category-level contributions shown in the decomposition figures. The decomposition is additive on the log-odds scale rather than on the probability scale. As a result, a given contribution has a direct interpretation in log odds.

Let $p(x)$ denote the model's predicted recession probability for observation x. Define the corresponding log odds as

$$\eta(x) = \log\left(\frac{p(x)}{1-p(x)}\right).$$

The category-level Shapley decomposition writes this log-odds prediction as

$$\eta(x) = \eta_{\text{base}} + \sum_c \phi_c.$$

where η_{base} is the baseline log odds and ϕ_c is the contribution of category c. The baseline log odds correspond to a baseline probability p_0, so

$$\eta_{\text{base}} = \log\left(\frac{p_0}{1-p_0}\right).$$

Subtracting the baseline log odds from the model prediction gives

$$\eta(x) - \eta_{\text{base}} = \sum_c \phi_c.$$

Thus, the category contributions sum to the change in predicted recession log odds relative to the baseline.

The same contribution can also be interpreted on the odds scale. Let

$$o_0 = \frac{p_0}{1-p_0}$$

denote the baseline odds. Since log odds are the logarithm of odds, exponentiating a contribution gives an odds multiplier:

$$m_c = \exp(\phi_c).$$

For example, a contribution of $\phi_c = 0.5$ multiplies recession odds by $\exp(0.5) \approx 1.65$. A contribution of $\phi_c = -0.5$ multiplies recession odds by $\exp(-0.5) \approx 0.61$.

Combining all category contributions, the model-implied odds are

$$o(x) = o_0 \times \exp\left(\sum_j \phi_j\right).$$

Converting these odds back to probability gives

$$p(x) = \frac{o(x)}{1 + o(x)} = \frac{\exp\left(\sum_j \phi_j\right) p_0}{1 - p_0 + \exp\left(\sum_j \phi_j\right) p_0}.$$

This expression shows why the plotted contributions should not be read as probability-point contributions. They are additive in log odds, or equivalently multiplicative in odds. The mapping from log odds to probability is nonlinear.

To express the contribution of a single category c in probability points, hold all other category contributions at their baseline values and apply ϕ_c to the baseline log odds. The implied probability is

$$p_c = \frac{\exp(\phi_c) p_0}{1 - p_0 + \exp(\phi_c) p_0}.$$