Journal of Forecasting

WILEY

RESEARCH ARTICLE OPEN ACCESS

The Role of Coincident Information in Real-Time Business Cycle Forecasting

Visa Kuntze[1,2]

[1]Department of Accounting and Finance, Turku School of Economics, University of Turku, Turku, Finland | [2]Department of Mathematics and Statistics, University of Turku, Turku, Finland

Correspondence: Visa Kuntze (visa.kuntze@utu.fi)

Received: 24 January 2024 | **Revised:** 7 April 2026 | **Accepted:** 16 April 2026

Keywords: publication lag | real-time forecasting | recession prediction

ABSTRACT

Official NBER recession dates are announced with substantial delay. Therefore, real-time forecasters cannot condition on the most recent business cycle states even though recessions and expansions are highly persistent. I study whether real-time coincident releases can substitute for this missing information. At each monthly forecast origin, I construct a recession nowcast, using four coincident indicators and several supervised classifiers, and add this nowcast probability to standard probit forecasting models. In an out-of-sample evaluation for US monthly data from 1986 to 2021, nowcast augmentation improves forecast accuracy at short horizons and at the 1-year horizon relative to a term spread benchmark, while including the raw coincident indicators directly is less effective. The gains are incremental once strong leading indicators are included, and model rankings are sensitive to resampling variation.

1 | Introduction

Forecasting recessions is difficult not mainly because leading indicators are uninformative but because official recession dates are available only with substantial delay. In real time, the economy's current and recent past business cycle states are uncertain because the NBER Business Cycle Dating Committee (BCDC) assigns recession and expansion dates only after considerable time has passed. Near turning points, this "publication lag" prevents forecasters from conditioning on the most recent business cycle states, even though these states are highly persistent. As a result, standard recession forecasting models that rely only on leading indicators can perform poorly precisely when timely forecasts are most valuable.

This paper proposes a simple way to incorporate real-time coincident information into multi-horizon recession forecasts. At the forecast origin, I first estimate a coincident recession probability for the most recent month whose state is not yet officially known, using real-time releases of four widely used monthly coincident indicators (industrial production, payroll employment, real personal income excluding transfers, and real manufacturing and trade sales). I consider five classifiers for this nowcasting step: a probit model, k-nearest neighbors, learning vector quantization, random forest, and gradient boosting. This allows the forecasting results to be compared across methods and avoids dependence on any single classifier. I then use this estimated probability as an additional predictor in conventional recession forecasting models (probit models based on the term spread and other leading indicators). This replaces unavailable information on recent recession states with a real-time estimate of the recession probability constructed from coincident releases available at the forecast origin.

My approach connects two strands of literature that are usually treated separately. A substantial nowcasting literature shows that various classifiers can identify business-cycle turning points before official announcements (see, e.g., Chauvet and Piger 2008; Camacho et al. 2018; Giusto and Piger 2017; Piger 2020). In parallel, a large literature on forecasting recessions uses probit

models with financial and macroeconomic predictors, including the term spread (the difference between long- and short-term interest rates), to forecast recession probabilities (see, e.g., Nyberg 2010; Liu and Moench 2016; Hwang 2019). Dynamic probit specifications (e.g., Chauvet and Potter 2005; Kauppi and Saikkonen 2008) can, in principle, exploit persistence by including lagged business cycle states, but publication lags make the most recent lags unavailable in real time forecasting and therefore reduce their usefulness. Using nowcast probabilities as predictors provides a practical alternative, allowing persistence-related information to enter the forecast without imposing ad hoc rules to impute unknown recent states.

Empirically, I conduct a real-time out-of-sample forecasting exercise for the United States at horizons up to 12 months. I evaluate forecast performance using the area under the precision-recall curve (AUPRC) alongside more common measures such as AUROC (e.g., Lahiri and Yang 2023). I study whether adding coincident nowcasts improves forecast accuracy relative to benchmarks that only use leading indicators and whether nowcasts substitute for, or complement, additional leading indicators beyond the term spread across forecast horizons.

The main findings are as follows. Relative to the Liu and Moench (2016) spread benchmark, nowcast augmentation raises AUPRC sharply at short horizons. The improvements are not confined to the near term: At the 1-year horizon, several nowcast variants still deliver a clear AUPRC increase, and bootstrap inference confirms that these gains are unlikely to be just random noise. Including the raw coincident indicators directly, by contrast, is less effective and performs noticeably worse than both the benchmark and the nowcast-augmented models at longer horizons. When the predictor set is broadened to include additional leading indicators from Liu and Moench (2016), the nowcast probability continues to add value as a complement to standard leading indicators rather than a substitute. The incremental gains are visible across horizons, and notably the nowcast probability retains useful information even at the 1-year horizon, on top of already strong financial predictors.

The remainder of the paper is organized as follows. In Section 2, I introduce the static and dynamic probit models as well as a nowcast-augmented model specification. I introduce the nowcasting methodology used in this paper in Section 3. In Section 4, I describe the data and empirical strategy, while Section 5 presents the results of the out-of-sample forecast evaluation. Finally, Section 6 concludes.

2 | Forecasting and Nowcasting Setup

I denote the binary state of the business cycle in month t by y_t, where $y_t = 0$ indicates an expansion and $y_t = 1$ a recession. Throughout this paper, y_t refers to the NBER-referenced business cycle indicator, and the terms "recession" and "expansion" are used as shorthand for NBER-classified recession and expansion months. The supervised methods in this paper are thus trained to reproduce NBER classifications, not recessions defined by other criteria. Forecasts are formed at a fixed forecast origin in the middle of month t. Because official NBER business cycle dates are announced with delay, the current state y_t and several recent values $y_{t-1}, y_{t-2}, \ldots$ are generally unknown. I therefore distinguish between leading indicators used to forecast future states and coincident indicators used to infer the most recent states.

Let $\boldsymbol{x}_t$ denote the vector of leading indicators observed at the mid-month forecast origin in month t. In my empirical implementation, leading indicators are constructed as monthly averages of daily data; hence, at origin t, the latest fully observed leading indicator values correspond to month $t-1$. Accordingly, $\boldsymbol{x}_t$ contains information through month $t-1$. Let $\boldsymbol{x}_{t,N}$ denote the coincident data vintage at origin t, that is, the vector collecting the latest available releases of the coincident indicators at that origin. Because the indicators are released with different publication lags, $\boldsymbol{x}_{t,N}$ has a ragged edge structure: Its components pertain to different calendar months (see Table 1 and Section 4). For example, under the release timing in Table 1, $\boldsymbol{x}_{t,N}$ contains $(E_{t-1}, I_{t-1}, P_{t-2}, M_{t-3})$ at origin t.

At each forecast origin t, I construct recession probability backcasts for the most recent months using $\boldsymbol{x}_{t,N}$ and forecast recessions h months ahead using $\boldsymbol{x}_t$ together with these backcast probabilities. Throughout the paper, I focus on supervised classification methods for forecasting and backcasting. Models are estimated using NBER recession dates as reference labels, while respecting the real-time information available at each forecast origin in out-of-sample forecasting (see Section 4).

2.1 | Predicting the Business Cycle With Probit Models

The conditional probability of a recession h months ahead given the predictors available at the forecast origin t can be expressed as

$$P(y_{t+h} = 1|\boldsymbol{x}_t) = G(\boldsymbol{x}_t), \tag{1}$$

where $G(\cdot)$ denotes the forecasting model or algorithm used. A common choice in the literature is the probit model

$$G(\boldsymbol{x}_t) = \Phi(\pi_{t+h}), \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and π_{t+h} is specified as

$$\pi_{t+h} = \boldsymbol{x}_t'\boldsymbol{\beta}, \tag{3}$$

with the parameter vector $\boldsymbol{\beta}$ which also includes an intercept.

TABLE 1 | Release dates of the values of coincident indicators for month t.

Coincident indicator	Release date
Payroll employment for month t	Beginning of month $t+1$
Industrial production for month t	Middle of month $t+1$
Personal income less transfer payments for month t	End of month $t+1$
Real manufacturing and trade sales for month t	End of month $t+2$

Dynamic extensions incorporate persistence by including lagged business cycle state indicators as predictors:

$$\pi_{t+h} = \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{k=0}^{q} \delta_k y_{t-k}. \tag{4}$$

In the present forecasting environment, however, the most recent values of y_{t-k} are typically unavailable at the forecast origin because NBER dating is announced with delay. This makes dynamic probit models in their standard form difficult to use for real-time forecasting.

Both the static and dynamic probit formulations can be estimated by maximum likelihood (see, e.g., Kauppi and Saikkonen 2008). In the empirical analysis, I consider direct multi-horizon forecasts by estimating horizon-specific models for each h.

2.2 | Forecasting Environment

In real time, only business cycle states up to the most recently announced turning point are known with certainty. Consequently, a forecaster cannot generally use data up to month t when estimating models that include y_t or its recent lags. Applied work has therefore typically imposed simplifying assumptions about the effective publication lag, such as assuming a fixed lag of J months (e.g., Kauppi and Saikkonen 2008; Nyberg 2010) or assuming that the state remains unchanged until the next NBER announcement (Berge 2015). These assumptions can lengthen the effective estimation sample during long expansions but may be misleading around turning points. In my empirical strategy, I instead follow the real-time label availability procedure described in Section 4.

The publication lag is particularly restrictive for dynamic probit specifications. To illustrate, suppose a fixed publication lag of J months is assumed. Then only lags of y_t at least J months old are treated as available at origin t. For direct multi-horizon forecasting, this yields the specification (cf. (4))

$$\pi_{t+h} = \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{k=J}^{q} \delta_k y_{t-k}, \qquad q \geq J. \tag{5}$$

Because y_{t-k} at large k is far removed from the forecast origin, the resulting model can only weakly exploit the persistence in y_t. In practice, the most informative persistence-related information is contained in the most recent lags of y_t, and these are precisely the ones unavailable in real time. Incorporating coincident information to infer these recent states is therefore a natural alternative.

2.3 | Dynamic Probit Model With Constructed Nowcasts

Although the most recent values of y_t are not observed at the forecast origin, coincident macroeconomic indicators related to the business cycle are released in real time. Numerous studies show that these indicators can be used to infer the current business cycle phase before official dating announcements (e.g., Chauvet and Piger 2008; Giusto and Piger 2017; Camacho et al. 2018). At each forecast origin t, I construct backcast recession probabilities for the K most recent months whose official states are not yet known. I denote these by $\widehat{N}_{t-1}^{(t)}, \widehat{N}_{t-2}^{(t)}, \ldots, \widehat{N}_{t-K}^{(t)}$.[1]

Each backcast is defined as the conditional probability:

$$\widehat{N}_{t-l}^{(t)} = P(y_{t-l} = 1 | \boldsymbol{x}_{t,N}^{t-l}), \tag{6}$$

where $l = 1, \ldots, K$. Here, $\boldsymbol{x}_{t,N}^{t-l}$ denotes the feature vector for month $t-l$ constructed from the coincident data vintage available at origin t, that is, using only releases available at that origin. The classifiers used to compute these backcast probabilities are described in Section 3.

With the constructed backcasts and leading indicators $\boldsymbol{x}_t$ available at the forecast origin, direct h-period-ahead forecasts of y_{t+h} can be formulated by augmenting the probit model as

$$\pi_{t+h} = \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{l=1}^{q} \delta_l \widehat{N}_{t-l}^{(t)}, \tag{7}$$

where $q \leq K$ denotes how many lagged backcast probabilities are used as predictors alongside the leading indicators $\boldsymbol{x}_t$. For simplicity, in Section 4, I use only the most recent backcast probability (i.e., $q = 1$).

Instead of using backcast probabilities $\widehat{N}_{t-l}^{(t)}$ directly as predictors, one could convert them into a binary recession/expansion series using a probability threshold and additional turning point rules, and then use these discrete states as regressors in (4). I do not adopt this approach. Thresholding implies that different levels of recession risk (e.g., 0.55 versus 0.95) collapse into the same indicator and therefore discards variation that may help predict future states. Moreover, constructing a discrete real-time dating rule requires specifying a probability threshold and auxiliary criteria (such as minimum duration or consecutive-month requirements), and results can be sensitive to these choices. Using probabilities directly avoids these choices and gives a single specification that applies at all forecast origins.

3 | Nowcasting the State of the Business Cycle

I construct the backcast recession probabilities in Section 2 using supervised classification methods. For each data vintage available at forecast origin t, I estimate a probability series $\{\widehat{N}_{\tau}^{(t)}\}$ over calendar months τ, where $\widehat{N}_{\tau}^{(t)}$ denotes the probability that month τ is a recession based on information available at origin t. In forecasting, I use the most recent backcast probability $\widehat{N}_{t-1}^{(t)}$ as an additional predictor in the recession forecasting model.

I consider learning vector quantization (LVQ), k-nearest neighbors (kNN), random forest (RF), and gradient boosting (GB), and I also include a probit model as a parametric benchmark, described last for ease of comparison with the other classifiers. LVQ is used prominently for real-time business cycle classification by Giusto and Piger (2017), while kNN and tree-based

methods are compared in Piger (2020) and applied to recession prediction in Vrontos et al. (2021).

3.1 | Common Setup: Predictors, Standardization, and Missing Values

For each vintage t and each target month τ, I construct a feature vector $\boldsymbol{x}_{t,N}^{\tau}$ from four monthly coincident indicators: industrial production growth, payroll employment growth, real personal income less transfer payments growth, and real manufacturing and trade sales growth. The feature vector includes the value for month τ and one monthly lag for each indicator, yielding eight predictors in total. The same set of predictors is used for all methods.

Within each vintage, predictors are standardized by z-scoring using the mean and standard deviation computed from the training observations available in that vintage, and the same scaling is applied to all months τ in that vintage. Because the coincident indicators are released with different publication lags, the features in $\boldsymbol{x}_{t,N}^{\tau}$ can be missing in real time for recent target months. The imputation procedure used to handle these missing values is described in Section 4.

Tuning parameters for the machine learning methods are selected within the training sample using the cross-validation procedure described in Section 4. The probit benchmark involves no tuning parameters.

3.2 | k-Nearest Neighbors

The k-nearest neighbors classifier assigns recession probabilities based on the labels of nearby observations in predictor space. For a target month τ in vintage t, let $R_k(\boldsymbol{x}_{t,N}^{\tau})$ denote the set of k training months whose feature vectors are closest to $\boldsymbol{x}_{t,N}^{\tau}$ under Euclidean distance computed on standardized predictors. The kNN recession probability is then

$$\widehat{N}_{\tau}^{(t)} = \frac{1}{k} \sum_{s \in R_k(\boldsymbol{x}_{t,N}^{\tau})} y_s, \tag{8}$$

that is, the share of recession observations among the k nearest neighbors. The tuning parameter is the neighborhood size k.

3.3 | Learning Vector Quantization

Learning vector quantization is a prototype-based classifier that represents each class by a set of codebook vectors in predictor space. During training, the codebook vectors are iteratively adjusted to improve classification performance. Following Giusto and Piger (2017), I estimate LVQ repeatedly under different random initializations of the codebook vectors. I then define $\widehat{N}_{\tau}^{(t)}$ as the fraction of runs in which target month τ is classified as a recession. The main tuning choices are the number of codebook vectors, the number of codebook vectors used for classification, and the number of random initializations.

3.4 | Tree-Based Methods: Random Forest and Gradient Boosting

Both random forest and gradient boosting construct recession probabilities using ensembles of classification trees. A classification tree partitions the predictor space into a set of terminal regions (leaves). For a given leaf, the recession probability is estimated as the fraction of recession observations among training months assigned to that leaf. Predictions for a new observation inherit the estimated probability of the leaf to which it is assigned.

Random forest reduces the variance of a single tree by averaging predictions across a large number of trees grown on bootstrap samples and with randomization in the selection of candidate predictors at each split. Gradient boosting builds trees sequentially, where each new tree is fit to improve performance on observations that were poorly predicted by the previous trees. In my implementation, RF probabilities are computed as the average of tree-level recession probabilities, equivalently the fraction of trees voting for recession. GB probabilities are the predicted probabilities from the final additive model evaluated through the logistic link function. Tuning parameters include the number of trees and restrictions on tree complexity; the specific choices are described in Section 4.

3.5 | Probit Benchmark

I also estimate a probit model on the coincident indicator feature vectors, using the same predictors as in the machine learning classifiers above and using the NBER recession indicator as the dependent variable. For each vintage t, this yields recession probabilities $\widehat{N}_{\tau}^{(t)} = \Phi\left((\boldsymbol{x}_{t,N}^{\tau})'\boldsymbol{\beta}\right)$ for target months τ. Including this benchmark allows a direct assessment of whether the more sophisticated classifiers add value over a simpler parametric alternative with no tuning parameters.

4 | Empirical Strategy

4.1 | Data

For backcasting, I use real-time vintages of four monthly coincident indicators of the US business cycle: growth rates of industrial production (I), non-farm payroll employment (E), real personal income less transfer payments (P), and real manufacturing and trade sales (M). I use the dataset of Camacho et al. (2018) and extend it to December 2022. The full calendar month sample runs from January 1967 to December 2022, while the first real-time vintages begin in October 1976.

Forecasts are formed at a fixed origin in the middle of month t. The release timing of the coincident indicators implies a ragged-edge information set (Table 1). At the mid-month origin in month t, the latest available releases typically correspond to $E_{t-1}, I_{t-1}, P_{t-2}$, and M_{t-3}. These releases form the coincident-data vintage $\boldsymbol{x}_{t,N}$, from which I construct month-specific feature vectors $\boldsymbol{x}_{t,N}^{\tau}$ and backcast recession probabilities $\widehat{N}_{\tau}^{(t)}$ as described in Section 3. I do not construct pseudo vintages prior to October

1976, and therefore, the out-of-sample evaluation is restricted to forecast origins for which real-time vintages exist and the required inputs are available.

In specifying the forecasting models (3) and (7), I use a set of leading indicators $\boldsymbol{x}_t$ observed at the same mid-month forecast origin. Leading indicators are constructed as monthly averages of daily data. Therefore, at origin t the latest fully observed leading-indicator values correspond to month $t-1$. Accordingly, $\boldsymbol{x}_t$ should be interpreted as information through month $t-1$.

My baseline predictor is the term spread, defined as the difference between 10-year and 3-month interest rates, which has been found to perform well as a leading indicator in a broad literature (see, e.g., Haubrich 2021). I also consider additional candidate predictors based on Liu and Moench (2016), including stock market returns, interest rates and spreads, and selected macro variables. I exclude predictors that are not available from 1967 onward or whose real-time availability is limited. The final set of leading indicators is summarized in Table 2.

4.2 | Real-Time Availability of the Business Cycle Indicator

The dependent variable y_t is based on the NBER Business Cycle Dating Committee (BCDC) chronology. Because turning points are announced with delay, the most recent values of y_t are not observed in real time. In out-of-sample forecasting, I therefore construct a real-time proxy for which values of y_t are treated as known at each forecast origin, following the three-step procedure in Giusto and Piger (2017): (i) a turning point date becomes known once it is announced by the BCDC; (ii) during expansions, I assume that a new peak is announced within 12 months of its occurrence; and (iii) once a turning point is announced, I impose a minimum duration requirement for the subsequent phase. In the baseline implementation, the minimum duration is 2 months, which accommodates short phases such as the 2020 recession. As a robustness check, Appendix S1 reports results with a 6-month minimum duration, as in Giusto and Piger (2017).

This procedure determines, at each forecast origin t, the last month whose value of y_t is treated as known and can be included in model estimation. The forecasting models are then estimated using an expanding window that uses only the labels treated as available at that origin, as described below.

TABLE 2 | Summary of leading indicators $\boldsymbol{x}_t$.

10y–3m spread	S&P 500, 3y change (%)	6m–FF spread
10y rate	Federal funds rate	1y–FF spread
3m rate	Aaa bond yield	2y–FF spread
S&P 500, 1y change (%)	Aaa–FF spread	5y–FF spread
Yen–dollar exchange rate	Baa–FF spread	10y–FF spread

Note: The variables are based on those used by Liu and Moench (2016). Throughout this paper, I use monthly averages.

4.3 | Nowcasting With Ragged-Edge Data: k-Nearest Neighbor Imputation

The ragged-edge structure of the coincident indicators implies that the feature vector $\boldsymbol{x}_{t,N}^{\tau}$ can be partially missing for recent target months τ in a given vintage t. Following Piger (2020), I impute missing predictors for target months using a k-nearest neighbor imputation rule based on the standardized training observations in the same vintage.

Let $\tilde{\boldsymbol{x}}_{t,N}^{\tau}$ denote the observed components of $\boldsymbol{x}_{t,N}^{\tau}$ and let $(\boldsymbol{x}_{t,N}^{\tau})^{*}$ denote the missing components. Let $R_k(\tilde{\boldsymbol{x}}_{t,N}^{\tau})$ denote the set of k training observations that are closest to $\tilde{\boldsymbol{x}}_{t,N}^{\tau}$ under Euclidean distance computed on the observed components. The missing components are imputed as the corresponding neighbor means:

$$(\boldsymbol{x}_{t,N}^{\tau})^{*imp} = \frac{1}{k} \sum_{s \in R_k(\tilde{\boldsymbol{x}}_{t,N}^{\tau})} (\boldsymbol{x}_{t,N}^{s})^{*}. \tag{9}$$

The neighborhood size k used for imputation is fixed at $k = 5$ and is not, for simplicity, tuned alongside the classifier hyperparameters.

4.4 | Optimizing Tuning Parameters

The kNN, RF, and GB classifiers involve tuning parameters that affect predictive performance. Tuning parameters are selected separately within each real-time vintage using only the labels available at that forecast origin. Within a given vintage, I use a rolling-origin cross-validation scheme: Each fold trains the classifier on an expanding window of early observations and evaluates it on the next fixed-length block of months, so that the validation period is always strictly after the training period. For each candidate hyperparameter value, I compute the out-of-fold log loss, average the loss across folds, and select the value that performs best. The classifier is then re-estimated on the full labeled portion of that vintage using the selected hyperparameters, and the resulting model is used to produce the nowcast probabilities for that vintage. This procedure ensures that tuning uses only information available at each forecast origin, and does not rely on a single global hyperparameter choice carried across vintages.

4.5 | Out-of-Sample Forecasting Design

I conduct a pseudo out-of-sample evaluation using an expanding window for updating model parameters. The evaluation sample begins in January 1986, which ensures that real-time coincident-data vintages are available and that the nowcasting classifiers are trained on an initial sample spanning approximately 19 years and including several complete business cycles. Forecasts are produced at monthly origins t from 1986:01 through 2021:12. For each horizon h, the target is month $t+h$, which implies that the last evaluated outcome is 2022:12 for $h = 12$.

At each forecast origin t, I estimate the forecasting models using all available observations up to the last month whose business cycle state is treated as known under the real-time availability

procedure above, and then form direct forecasts of y_{t+h} for horizons $h \in \{1,2,3,4,6,8,10,12\}$. The predicted recession probabilities are compared to the realized values of y_{t+h} for each horizon. Using the terminology of Estrella et al. (2003), these forecasts are marginal: I predict whether a given month at $t+h$ is a recession, as opposed to cumulative forecasts that predict whether a recession occurs at any point between t and $t+h$.

4.6 | Measuring Predictive Accuracy

The forecasting models yield predicted recession probabilities $p_{t+h} = P(y_{t+h} = 1|\boldsymbol{x}_t)$ for each horizon h. Forecast accuracy is evaluated using threshold-free measures based on the receiver operating characteristic (ROC) curve and the precision-recall (PR) curve. The area under the ROC curve (AUROC) summarizes the trade-off between the true positive rate and the false positive rate across all thresholds. The area under the PR curve (AUPRC) summarizes the trade-off between precision and recall across all thresholds.

AUROC has been widely used in recession forecasting (see, e.g., Berge and Jordà 2011; Liu and Moench 2016; Choi et al. 2023). However, when forecasting rare events such as recessions, AUROC can give an overly optimistic impression of predictive performance (Saito and Rehmsmeier 2015; Lahiri and Yang 2023). I therefore emphasize AUPRC, which measures how well the model identifies recession months and has a baseline value equal to the unconditional recession frequency. I also report AUROC to facilitate comparison with earlier literature, where this measure is more common. The AUROC-based results are reported in Appendix A.

5 | Results

In this section, I report two sets of results. First, I summarize forecast accuracy in the pseudo out-of-sample evaluation, where models are re-estimated at each monthly forecast origin using only the values of y_t treated as known at that date. Second, I quantify sampling uncertainty and model selection instability using a block bootstrap over forecast origins.

5.1 | Expanding Window Forecasting Results

I take as the benchmark the probit specification of Liu and Moench (2016) that uses the term spread and its 6-month lag as predictors. For brevity, I refer to this two-regressor benchmark as the "spread-only" model throughout. I then examine whether adding real-time coincident information improves recession forecast accuracy. Coincident information enters either through the nowcast probability $\widehat{N}^{(t)}_{t-1,m}$ constructed from monthly coincident indicators (method m), or through a specification that includes the latest available values of the coincident indicators at mid-month t directly ("Coinc."). I refer to nowcast-augmented specifications as NC throughout.

Table 3 reports AUPRC levels for the spread-only benchmark and for versions augmented with the alternative nowcast measures and the coincident indicator benchmark, together with differences relative to spread-only. The most visible gains occur at short horizons, where adding a single real-time nowcast probability raises AUPRC sharply relative to spread-only. The improvements are not confined to the near term: At $h = 12$, several nowcast variants still deliver a clear increase in AUPRC (roughly 0.02–0.07), whereas the coincident-indicator specification performs noticeably worse than both the spread-only and the nowcast-augmented models. Collapsing the ragged-edge coincident releases into a single recession probability therefore works better than including the raw series directly.

I next broaden the predictor set beyond the term spread. Tables 4 and 5 report, for each horizon, the best-performing model among leading-indicators-only specifications and specifications augmented with a single nowcast probability $\widehat{N}^{(t)}_{t-1,m}$, along with Δ, the AUPRC difference relative to the best leading-indicators-only model at the same horizon.

Two patterns stand out. First, at very short horizons the nowcast probability does not improve the already best-performing leading indicator specifications. At $h = 1$–2, the best models are equity-return based, and the best nowcast-augmented variants are essentially tied or slightly worse in AUPRC. Second, the incremental value of nowcast augmentation becomes more visible once the leading-indicator signal weakens and the forecast task is more demanding. In the one-predictor case, the best nowcast-augmented model improves on the best non-NC specification at $h = 8$ and $h = 10$ by about 0.03 units, and at $h = 12$ by about 0.02. In the two-predictor case, the largest gain appears at $h = 6$, where adding $\widehat{N}^{(t)}_{t-1,m}$ raises AUPRC by roughly 0.06, while $h = 8$ and $h = 12$ again show gains on the order of 0.02–0.03. These magnitudes are modest, but real-time coincident information adds incremental value on top of standard leading indicators across a range of horizons.

The tables also show that the best nowcast method is horizon-dependent. At shorter horizons, probit and kNN variants often appear among the top nowcast-augmented specifications, while LVQ and GB are more common at medium and longer horizons. This is consistent with the nowcast acting as a compact summary of current conditions rather than a single dominant transformation across all horizons.

Equity returns and interest rate spreads feature prominently among the best models across horizons, which is in line with Liu and Moench (2016). Two differences are worth noting: I prioritize AUPRC rather than AUROC, so improvements can register even when AUROC changes little, and I evaluate performance under a real-time expanding-window scheme rather than a fixed train/test split. Both differences can shift which specific predictor is most useful at a given horizon.

The accuracy gains from nowcasts are generally more modest when AUROC is used instead of AUPRC, especially at longer horizons, suggesting that nowcasts primarily improve recession classification by raising precision and recall for the rare recession months rather than changing the overall ranking of recession risk across all months. Full AUROC results are reported in Appendix A.

TABLE 3 | AUPRC by horizon: Levels and differences versus the spread-only model.

	Spread	+Probit	+kNN	+LVQ	+RF	+GB	+Coinc.
Panel A: AUPRC (level)							
$h = 1$	0.30	0.48	0.40	0.34	0.38	0.38	0.42
$h = 2$	0.22	0.37	0.29	0.29	0.28	0.28	0.28
$h = 3$	0.22	0.33	0.25	0.26	0.24	0.24	0.24
$h = 4$	0.23	0.30	0.23	0.24	0.23	0.23	0.23
$h = 6$	0.26	0.26	0.26	0.26	0.26	0.26	0.27
$h = 8$	0.30	0.29	0.29	0.30	0.31	0.31	0.28
$h = 10$	0.37	0.39	0.39	0.40	0.39	0.40	0.27
$h = 12$	0.35	0.42	0.40	0.37	0.40	0.40	0.26
Panel B:Δ AUPRC vs. spread-only (level difference)							
$h = 1$		0.18	0.10	0.04	0.08	0.08	0.12
$h = 2$		0.15	0.07	0.07	0.06	0.06	0.06
$h = 3$		0.11	0.03	0.04	0.02	0.02	0.02
$h = 4$		0.07	0.00	0.01	0.00	0.00	0.00
$h = 6$		0.00	0.00	0.00	0.00	0.00	0.01
$h = 8$		−0.01	−0.01	0.00	0.01	0.01	−0.02
$h = 10$		0.02	0.02	0.03	0.02	0.03	−0.10
$h = 12$		0.07	0.05	0.02	0.05	0.05	−0.09

Note: Panel B reports point differences computed from Panel A: Δ = AUPRC(model) − AUPRC(spread-only). "Coinc." augments the spread-only model with the latest available coincident releases at the mid-month origin t, namely, E_{t-1}, I_{t-1}, P_{t-2}, and M_{t-3}.

TABLE 4 | Best one-predictor forecasting models with and without nowcast augmentation.

h	Leading indicators only	AUPRC	Best NC-augmented model	AUPRC	Δ	Second-best NC-augmented model	AUPRC	Δ
1	S&P 500, 1y change (%)	0.80	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.77	−0.03	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.77	−0.03
2	S&P 500, 1y change (%)	0.73	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.72	−0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.72	−0.01
3	S&P 500, 1y change (%)	0.60	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.60	+0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.59	−0.01
4	S&P 500, 1y change (%)	0.50	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.50	−0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.49	−0.01
6	S&P 500, 1y change (%)	0.42	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.42	+0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.42	−0.00
8	1y–FF spread	0.41	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.43	+0.03	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.41	+0.01
10	5y–FF spread	0.44	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.47	+0.03	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.45	+0.01
12	S&P 500, 1y change (%)	0.43	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.44	+0.02	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.44	+0.02

Note: AUPRC is computed over the out-of-sample forecasting evaluation sample. "NC" denotes inclusion of the most recent backcast probability $\widehat{N}^{(t)}_{t-1,m}$ constructed with method $m \in \{\text{probit}, \text{kNN}, \text{LVQ}, \text{RF}, \text{GB}\}$. Δ reports the AUPRC difference relative to the best non-NC model in the same row.

TABLE 5 | Best two-predictor forecasting models with and without nowcast augmentation.

h	Leading indicators only	AUPRC	Best NC-augmented model	AUPRC	Δ	Second-best NC-augmented model	AUPRC	Δ
1	S&P 500, 1y change (%) + S&P 500, 3y change (%)	0.81	S&P 500, 1y change (%) + S&P 500, 3y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.80	−0.01	S&P 500, 1y change (%) + S&P 500, 3y change (%) + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.79	−0.02
2	S&P 500, 1y change (%) + 3m–FF spread	0.75	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.74	−0.01	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.73	−0.01
3	S&P 500, 1y change (%) + 10y–FF spread	0.65	S&P 500, 1y change (%) + 5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.67	+0.02	S&P 500, 1y change (%) + 10y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.66	+0.02
4	S&P 500, 1y change (%) + 5y–FF spread	0.60	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{RF}}$	0.61	+0.01	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.61	+0.01
6	S&P 500, 1y change (%) + 5y–FF spread	0.46	S&P 500, 1y change (%) + Aaa bond yield + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.52	+0.06	S&P 500, 1y change (%) + 6m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.49	+0.03
8	S&P 500, 1y change (%) + 5y–FF spread	0.41	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.43	+0.03	S&P 500, 1y change (%) + 6m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.43	+0.03
10	Baa–Aaa spread + 5y–FF spread	0.44	Baa–Aaa spread + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.45	+0.01	Baa–Aaa spread + 5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.45	+0.01
12	S&P 500, 1y change (%) + 6m–FF spread	0.46	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.48	+0.02	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.48	+0.02

Note: See notes to Table 4.

5.2 | Regime Separation

In Figure 1, I plot the median predicted recession probability separately for recession months and expansion months in the evaluation sample, for horizons $h \in \{1,4,8,12\}$ and three reference models: the spread-only benchmark, the best leading-indicators-only model, and the best nowcast-augmented model. A useful forecasting model should assign higher probabilities in recession months and lower probabilities in expansion months, so the gap between the two regime medians provides a transparent diagnostic of economic relevance that complements the aggregate AUPRC rankings.

The figure shows that the main gains from moving beyond the spread-only benchmark come from raising recession-month probabilities while keeping expansion-month probabilities near zero. At $h = 1$, the nowcast-augmented model achieves the highest recession-month median, consistent with the short-horizon AUPRC improvements. At intermediate horizons, differences between the nowcast-augmented and non-nowcast specifications are smaller and not uniformly in favor of nowcast augmentation. At $h = 12$, the nowcast-augmented model again delivers slightly higher recession-month probabilities, suggesting that coincident information retains some value even at the 1-year horizon. The visual gaps are modest throughout, and the figure is consistent with the view that nowcasts provide incremental rather than transformative improvements.

5.3 | Turning Point Detection

A natural concern is that the nowcast gains may reflect mainly persistence: If the economy is already in a recession, forecasting a recession a few months ahead is mechanically easier. To check this, I split the evaluation targets into months within a ± 2 month window around NBER peaks and troughs and all other months, comparing AUPRC improvements relative to spread-only within each subset. The ± 2 month window is narrow enough to focus on the immediate neighborhood of a turning point while avoiding dilution by long stretches of ordinary recession or expansion months.

Table 6 shows that away from turning points, both leading-indicator models and their nowcast-augmented counterparts improve sharply on the spread-only benchmark, consistent with much of the predictive accuracy coming from tracking the prevailing regime. Near turning points, the evidence is more mixed at short horizons: at $h = 1$–2, the leading-indicators-only models outperform the nowcast-augmented alternatives, though both still beat the spread benchmark by a wide margin. Nowcast augmentation shows clearer gains in the turning point window especially at $h = 12$, with some improvement also visible at $h = 6$. Nowcasts therefore do not resolve the difficulty of forecasting near turning points at short horizons, but they do add useful information relative to the spread benchmark even in the turning point window at longer horizons.

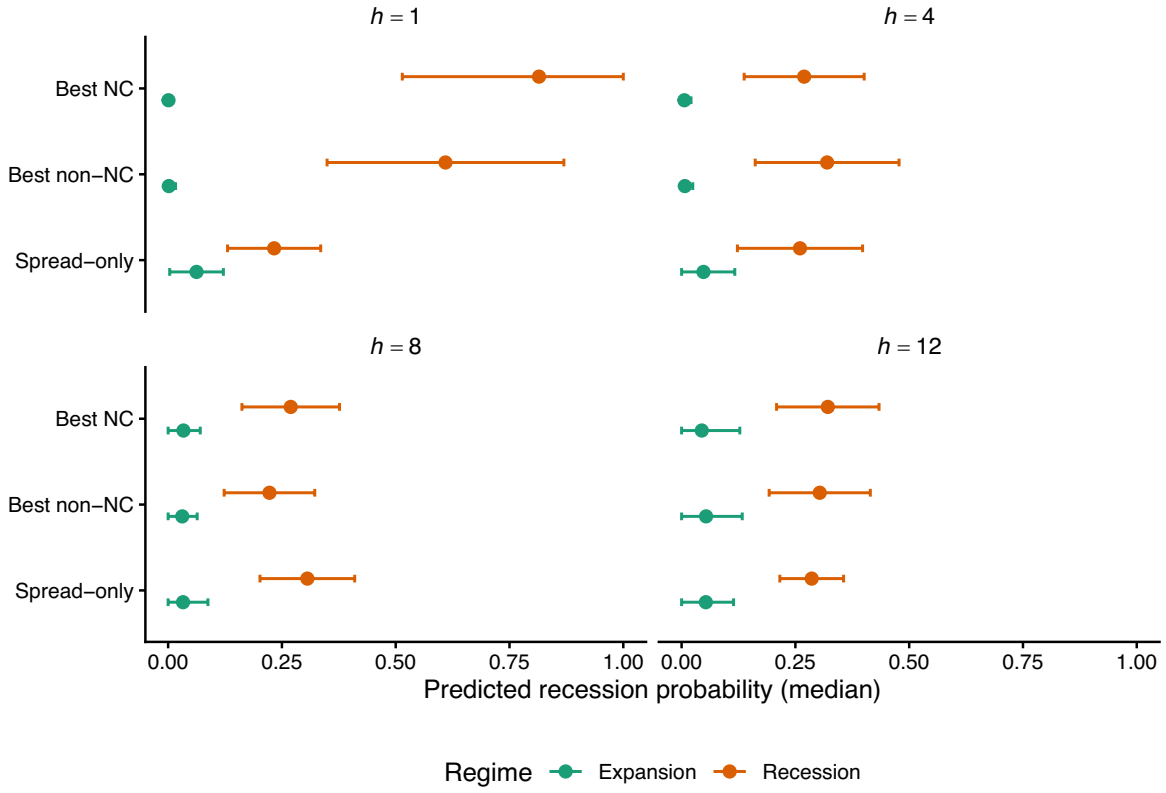

FIGURE 1 | Recession and expansion probability separation (selected horizons). For horizons $h \in \{1,4,8,12\}$, the figure plots the median predicted recession probability in NBER recession months and in expansion months for three specifications: the spread-only benchmark, the best leading-indicators-only model, and the best nowcast-augmented model. Markers indicate regime-specific medians and horizontal bars show the interquartile range within each regime. Larger gaps between the recession and expansion distributions indicate stronger regime separation.

TABLE 6 | Turning point versus non-turning months: AUPRC improvements relative to the spread-only model.

Panel A: Near turning points (± 2 months)			Panel B: Away from turning points		
h	Δ Best non-NC	Δ Best NC	h	Δ Best non-NC	Δ Best NC
1	0.25	0.14	1	0.65	0.72
2	0.17	0.13	2	0.79	0.81
3	0.04	0.03	3	0.75	0.79
4	0.03	0.02	4	0.68	0.73
6	0.05	0.11	6	0.26	0.14
8	0.04	0.05	8	0.16	0.23
10	0.03	0.00	10	0.07	0.13
12	0.16	0.19	12	0.12	0.13

Note: Entries report ΔAUPRC = AUPRC(model) − AUPRC(spread-only) computed within each subset. "Best non-NC" is the best-performing leading-indicators-only model in the subset; "Best NC" is the best-performing NC-augmented model in the subset. Turning point months are defined as ± 2 months around NBER peaks and troughs.

5.4 | Bootstrap Inference and Selection Instability

Many incremental gains from nowcast augmentation are modest in magnitude. I therefore complement the point estimates with bootstrap inference to guard against over-interpreting small differences that may be sensitive to the particular sequence of recession episodes in the evaluation period.

The bootstrap is applied to the evaluation sample of forecast origins from 1986:01 to 2021:12. For a given horizon h and metric (AUPRC or AUROC), I treat the realized outcomes $\{y_{t+h}\}$ and the corresponding forecast probability sequences $\{\widehat{p}_{t+h}(m)\}$, produced in the out-of-sample evaluation, as fixed objects. I then resample the sequence of forecast origins using the stationary bootstrap of Politis and Romano (1994). Following Liu and Moench (2016), I implement the stationary bootstrap with an average block length of 96 months (8 years) and verify robustness to a shorter 48-month choice in Appendix S1. This procedure draws random-length blocks of consecutive months and concatenates them to form a bootstrap index set of the same length as the original evaluation period. Denoting the bootstrap index set by $\mathcal{I}_b$, I compute, for each model m,

$$\mathrm{AUPRC}_b(m) = \mathrm{AUPRC}\big(\{y_{t+h}: t \in \mathcal{I}_b\}, \{\widehat{p}_{t+h}(m): t \in \mathcal{I}_b\}\big),$$

and analogously for AUROC. The object of interest is the benchmark-relative difference computed on the same resampled index set,

$$\Delta_b(m) = \mathrm{AUPRC}_b(m) - \mathrm{AUPRC}_b(\text{spread-only}).$$

This procedure yields an empirical distribution for $\Delta_b(m)$, from which I report percentile confidence intervals and $f(m)$, the fraction of bootstrap draws where $\Delta_b(m) \le 0$.

This design differs from Liu and Moench (2016) and Göbel (2023), who also use the stationary block bootstrap to resample dependent and independent variables and re-estimate models under a fixed train and test split. That approach is natural when model comparisons are tied to a single hold-out evaluation period generated from one estimation window. Here, forecasts are generated at many monthly origins under an expanding-window scheme, and the object of interest is the distribution of relative performance across those origins. Resampling blocks of forecast origins targets that object directly while preserving the dependence structure in forecast performance over time.

Table 7 shows positive median AUPRC improvements at short horizons, with the probit-based nowcast standing out at $h = 3$ and $h = 4$, where the confidence intervals are largely above zero and the fraction of negative bootstrap draws is small. At $h = 12$, several nowcast variants including probit, GB, and kNN deliver positive median improvements with similarly small fractions of negative draws, even though intermediate horizons show effects closer to zero. The bootstrap evidence thus indicates that nowcasts are helpful not only at short horizons but also for forecasting recessions 1 year ahead.

When the candidate set is expanded beyond the spread-only versus NC comparison, the identity of the top ranked model can vary across plausible resamples of the evaluation period. I therefore take the model that is best on the original evaluation sample as the reference, denoted $m^\star$, and report both its benchmark-relative improvement across bootstrap draws, $\Delta_b(m^\star) = \mathrm{AUPRC}_b(m^\star) - \mathrm{AUPRC}_b(\text{spread-only})$, and its selection share, defined as the fraction of draws in which the same $m^\star$ also attains the highest AUPRC among all candidates. This separates two questions: whether the original-sample winner tends to beat the benchmark in resamples, and whether its identity as the best model is stable under resampling. The results are presented in Table 8.

The mean improvement $\Delta_b(m^\star)$ is largest at short horizons, consistent with coincident information being most useful when forecasting only a few months ahead. At $h = 12$, $m^\star$ also delivers a positive mean and median improvement, larger than in the restricted spread-only versus NC comparison, albeit with wider confidence intervals. At intermediate horizons such as $h = 8$, the confidence intervals are wide and span zero, reflecting genuine uncertainty about the magnitude of the gain, while the fraction of negative draws remains small because the bootstrap distribution, though wide, is centered above zero. The selection shares make clear that these gains should not be read as evidence for a uniquely best specification: Even when $\Delta_b(m^\star)$ is typically positive, the identity of the top ranked model can vary across bootstrap resamples.

Table 9 makes the model uncertainty visible by showing the top three specifications for each horizon by winner share. Even at a fixed horizon, the bootstrap rarely concentrates on a single specification. At $h = 3$ and $h = 4$, the most common winner has a share around 0.26 and 0.18, and the residual "Other" mass is large (0.46 and 0.53), meaning a long tail of alternative models wins often. At $h = 6$, $h = 10$, and $h = 12$, the top model captures close to half the draws, but the remaining mass is still substantial. The ranking is therefore fragile when evaluated over

TABLE 7 | Bootstrap results for Δ AUPRC versus the spread-only model.

	+Probit	+kNN	+LVQ	+RF	+GB	+Coinc.
$h = 1$	0.20 [− 0. 19, 0.42]	0.11 [− 0. 31, 0.37]	0.04 [− 0. 38, 0.28]	0.09 [− 0. 31, 0.33]	0.10 [− 0. 28, 0.32]	0.15 [− 0. 14, 0.31]
$h = 2$	0.15* [− 0. 08, 0.31]	0.07 [− 0. 18, 0.19]	0.06 [− 0. 29, 0.20]	0.06 [− 0. 25, 0.19]	0.05 [− 0. 23, 0.18]	0.06 [− 0. 10, 0.13]
$h = 3$	0.11** [− 0. 01, 0.27]	0.04* [− 0. 02, 0.13]	0.05 [− 0. 20, 0.14]	0.02 [− 0. 12, 0.11]	0.02 [− 0. 11, 0.08]	0.02 [− 0. 04, 0.06]
$h = 4$	0.08*** [0. 02, 0.22]	0.01 [− 0. 01, 0.08]	0.02 [− 0. 16, 0.09]	0.01 [− 0. 09, 0.07]	0.01 [− 0. 02, 0.04]	0.01 [− 0. 03, 0.04]
$h = 6$	0.00 [− 0. 02, 0.04]	0.00 [− 0. 03, 0.02]	−0.01 [− 0. 05, 0.01]	0.00 [− 0. 02, 0.02]	0.00 [− 0. 02, 0.02]	0.01* [0. 00, 0.03]
$h = 8$	−0.01 [− 0. 06, 0.01]	−0.01 [− 0. 07, 0.00]	−0.01 [− 0. 04, 0.06]	0.00 [− 0. 02, 0.02]	0.00 [− 0. 02, 0.02]	−0.01 [− 0. 09, 0.02]
$h = 10$	0.01 [− 0. 08, 0.06]	0.00 [− 0. 06, 0.03]	0.03 [− 0. 05, 0.10]	0.01 [− 0. 09, 0.06]	0.02 [− 0. 02, 0.06]	−0.09 [− 0. 21, 0.00]
$h = 12$	0.06** [0. 00, 0.11]	0.04*** [0. 00, 0.07]	0.01 [− 0. 09, 0.06]	0.05 [− 0. 04, 0.09]	0.04*** [− 0. 01, 0.07]	−0.08 [− 0.13, − 0.01]

Note: Entries report the bootstrap median Δ and percentile 95% confidence interval for Δ = AUPRC(model) − AUPRC(spread-only). Stars indicate the fraction of bootstrap draws with $\Delta_b(m) \le 0$: $^{*}f < 0.10$, $^{**}f < 0.05$, $^{***}f < 0.01$.

TABLE 8 | Best-in-class improvement over the spread-only benchmark (AUPRC, fixed $m^{\star}$, BL = 96).

h	Best model on original sample $m^{\star}$	Sel. share	Mean Δ	Median Δ	95% CI	f
1	S&P 500, 1y change (%) + S&P 500, 3y change (%)	0.138	0.47	0.50	[0.11, 0.75]	0.001***
2	S&P 500, 1y change (%) + 3m–FF spread	0.342	0.46	0.52	[0.02, 0.71]	0.019**
3	S&P 500, 1y change (%) + 5y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$	0.255	0.39	0.44	[0.01, 0.63]	0.021**
4	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{RF}}$	0.128	0.34	0.37	[0.01, 0.54]	0.019**
6	S&P 500, 1y change (%) + 6m–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{GB}}$	0.525	0.20	0.22	[0.02, 0.31]	0.018**
8	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{GB}}$	0.333	0.10	0.11	[−0.14, 0.19]	0.042**
10	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$	0.484	0.08	0.09	[−0.10, 0.18]	0.101
12	S&P 500, 1y change (%) + 3m–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{probit}}$	0.474	0.14	0.13	[−0.04, 0.39]	0.063*

Note: Δ denotes the improvement in AUPRC relative to the spread-only benchmark for the full-sample winning model $m^{\star}$, computed on each bootstrap resample of forecast origins with block length 96. Sel. share is the fraction of draws in which $m^{\star}$ also wins. f is the fraction of draws with $\Delta \le 0$. Stars: $^{*}f < 0.10$, $^{**}f < 0.05$, $^{***}f < 0.01$. Total number of bootstrap draws is 1000.

TABLE 9 | Selection instability across bootstrap draws: top-3 winners by horizon (AUPRC, BL = 96).

h	Top 1	Top 2	Top 3	Other
1	S&P 500, 1y (%) + 3m–FF (0.328)	S&P 500, 1y (%) + S&P 500, 3y (%) + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$ (0.176)	S&P 500, 1y (%) + S&P 500, 3y (%) (0.138)	0.358
2	S&P 500, 1y (%) + 3m–FF (0.342)	S&P 500, 1y (%) + S&P 500, 3y (%) (0.220)	S&P 500, 1y (%) + 1y–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$ (0.102)	0.336
3	S&P 500, 1y (%) + 5y–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$ (0.255)	S&P 500, 1y (%) + 3m–FF (0.144)	S&P 500, 1y (%) + 10y–FF (0.137)	0.464
4	S&P 500, 1y (%) + 3m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{kNN}}$ (0.178)	S&P 500, 1y (%) + 3m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{GB}}$ (0.168)	S&P 500, 1y (%) + 3m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{RF}}$ (0.128)	0.526
6	S&P 500, 1y (%) + 6m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{GB}}$ (0.525)	S&P 500, 1y (%) + Aaa yield + $\widehat{N}^{(t)}_{t-1,\mathrm{kNN}}$ (0.175)	S&P 500, 1y (%) + 5y–FF (0.078)	0.222
8	S&P 500, 1y (%) + 6m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{GB}}$ (0.333)	S&P 500, 1y (%) + Aaa yield + $\widehat{N}^{(t)}_{t-1,\mathrm{kNN}}$ (0.177)	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$ (0.162)	0.328
10	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$ (0.484)	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{LVQ}}$ (0.169)	5y–FF spread (0.083)	0.264
12	S&P 500, 1y (%) + 3m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{probit}}$ (0.474)	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\mathrm{probit}}$ (0.097)	S&P 500, 1y (%) + 3m–FF + $\widehat{N}^{(t)}_{t-1,\mathrm{GB}}$ (0.097)	0.332

Note: Winner shares are the fractions of bootstrap draws in which each model attains the highest AUPRC among all candidates. "Other" is the combined share of all models outside the top three. Block length is 96 months, and total number of draws is 1000.

plausible resamples of the evaluation period, which is why I emphasize benchmark-relative improvements and their uncertainty rather than any single winning specification.

A limitation of the bootstrap inference is that it conditions on the realized probability sequences from the nowcasting step and does not resample the first-stage estimation. The reported confidence intervals and fractions of negative draws therefore reflect uncertainty about evaluation-sample composition under serial dependence, not the additional uncertainty that would arise from re-estimating the nowcast model in each bootstrap draw. This focus is deliberate: The dominant practical uncertainty for the model comparison at hand is how benchmark-relative performance changes across plausible blocks of forecast origins and recession episodes in the evaluation sample, and the stationary bootstrap captures that key source of uncertainty directly.

5.5 | Summary and Robustness

The evidence points to a consistent contribution from real-time nowcast probabilities. Several nowcast variants deliver statistically supported AUPRC improvements at both short and longer horizons, and the best-in-class analysis confirms that gains relative to the benchmark remain meaningful even at $h = 12$. Once richer predictor sets are allowed, the incremental gains from adding a nowcast probability are typically small, and nowcasts do not overturn the leading indicators as the primary drivers of forecast accuracy. The nowcast contribution is therefore most visible when the benchmark is parsimonious.

The main results are robust to alternative specifications. Repeating the bootstrap with a block length of 48 months instead of 96, and using a minimum phase duration of 6 months instead of two in the real-time label availability procedure, leaves the main conclusions unchanged; these results are reported in Appendix S1. I also examined two alternative ways of incorporating coincident information: a principal component extracted from the four coincident indicators and an interactions model where the nowcast probability is interacted with the leading indicators. Neither approach improves on the nowcast-augmented specifications reported above, confirming that collapsing coincident releases into a supervised nowcast probability is the most effective way to add real-time coincident information to the forecasting model. These results are also reported in Appendix S1.

6 | Conclusions

This paper studies whether real-time coincident information can improve multi-horizon recession forecasts, using nowcast probabilities constructed from four coincident indicators as additional predictors in standard probit models. The main finding of this paper is that collapsing coincident releases into a nowcast probability is more useful than adding the raw indicators directly. Relative to the term spread benchmark, several nowcast variants raise AUPRC at short horizons, and the gains extend to the 1-year horizon, while the raw coincident indicator specification performs worse than both the benchmark and the nowcast-augmented models.

Once strong leading indicators are included, the nowcast gains are naturally smaller, and at very short horizons the leading indicators alone can be sufficient. The nowcast probability is therefore best seen as a complement to standard leading indicators rather than a substitute, adding useful information especially when the signal of the leading indicators weakens at medium and longer horizons.

Bootstrap inference confirms that several nowcast variants deliver statistically supported AUPRC gains, most notably at short horizons and at the 1-year horizon, indicating that the improvements reflect genuine incremental information in the nowcast probabilities rather than an artifact of the particular recession episodes in the sample.

A simulation-based generalization of these findings would require generating synthetic business cycle data that reproduces the persistence, asymmetry, and rarity of recessions together with realistic release structures, which is a major undertaking in its own right. I leave this to future research, noting that the bootstrap evidence in Section 5 provides at least partial insight into the sensitivity of the results to the particular sequence of recession episodes in the evaluation sample.

Acknowledgments

The author thanks Paolo Fornaro, Henri Nyberg, and participants at Helsinki GSE Econometrics Workshop (2023) for useful comments. I gratefully acknowledge the financial support from the Academy of Finland (grant 321968), the Foundation for Economic Education (Liikesivistysrahasto, grant 220246), and the EXACTUS Doctoral Programme at the University of Turku. The author used OpenAI Codex (model gpt-5.1-codex-max) to assist with code development and optimization, and Claude (Anthropic, Sonnet 4.6) to assist with editing and structuring the manuscript text. All research design, analysis, interpretation of results, and final content remain the sole responsibility of the author. AI-generated content was reviewed and verified throughout. Open access publishing facilitated by Turun yliopisto, as part of the Wiley - FinELib agreement.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] Because values of the coincident indicators for month t are released no earlier than month $t+1$, predictions for month t are not feasible at the forecast origin. Hence, $\hat{N}_{t-l}^{(t)}$ are backcasts of recent past months rather than predictions for the current month.

References

Berge, T. 2015. "Predicting Recessions With Leading Indicators: Model Averaging and Selection Over the Business Cycle." *Journal of Forecasting* 34: 455–471.

Berge, T., and O. Jordà. 2011. "Evaluating the Classification of Economic Activity Into Recessions and Expansions." *American Economic Journal: Macroeconomics* 3, no. 2: 246–277.

Camacho, M., G. Perez-Quiros, and P. Poncela. 2018. "Markov-Switching Dynamic Factor Models in Real Time." *International Journal of Forecasting* 34, no. 4: 598–611.

Chauvet, M., and J. Piger. 2008. "A Comparison of the Real-Time Performance of Business Cycle Dating Methods." *Journal of Business and Economics Statistics* 26, no. 1: 42–49.

Chauvet, M., and S. Potter. 2005. "Forecasting Recessions Using the Yield Curve." *Journal of Forecasting* 24: 77–103.

Choi, J., D. Ge, K. H. Kang, and S. Sohn. 2023. "Yield Spread Selection in Predicting Recession Probabilities." *Journal of Forecasting* 42, no. 7: 1772–1785.

Estrella, A., A. R. Rodrigues, and S. Schich. 2003. "How Stable Is the Predictive Power of the Yield Curve? Evidence from Germany and the United States." *Review of Economics and Statistics* 85, no. 3: 629–644.

Giusto, A., and J. Piger. 2017. "Identifying Business Cycle Turning Points in Real Time With Vector Quantization." *International Journal of Forecasting* 33: 174–184.

Göbel, M. 2023. "Forecasting U.S. Recessions: The Yield Curve—What Else?!" Unpublished manuscript. https://www.maximiliangoebel.com/research.

Haubrich, J. 2021. "Does the Yield Curve Predict Output?"*Annual Review of Financial Economics* 13: 341–362.

Hwang, Y. 2019. "Forecasting Recessions With Time-Varying Models." *Journal of Macroeconomics* 62: 103153.

Kauppi, H., and P. Saikkonen. 2008. "Predicting U.S. Recessions With Dynamic Binary Response Models." *Review of Economics and Statistics* 90: 777–791.

Lahiri, K., and C. Yang. 2023. "ROC and PRC Approaches to Evaluate Recession Forecasts." *Journal of Business Cycle Research* 2023: 1–30.

Liu, W., and E. Moench. 2016. "What Predicts US Recessions?" *International Journal of Forecasting* 32, no. 4: 1138–1150.

Nyberg, H. 2010. "Dynamic Probit Models and Financial Variables in Recession Forecasting." *Journal of Forecasting* 29, no. 1-2: 215–230.

Piger, J. 2020. "Turning Points and Classification." *Chapter 18 in Macroeconomic forecasting in the Era of Big Data—Theory and Practice*. Springer Nature Switzerland AG.

Politis, D., and J. Romano. 1994. "The Stationary Bootstrap." *Journal of the American Statistical Association* 89, no. 428: 1303–1313.

Saito, T., and M. Rehmsmeier. 2015. "The Precision-Recall Plot Is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets." *PLoS ONE* 10, no. 3: e0118432.

Vrontos, S., J. Galakis, and I. Vrontos. 2021. "Modeling and Predicting U.S. Recessions Using Machine Learning Techniques." *International Journal of Forecasting* 37, no. 2: 647–671.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. Supplementary Appendix S1.

Appendix A

AUROC by horizon: Levels and differences versus spread-only.

	Spread	+Probit	+kNN	+LVQ	+RF	+GB	+Coinc.
Panel A: AUROC (level)							
$h = 1$	0.75	0.93	0.90	0.88	0.89	0.89	0.92
$h = 2$	0.75	0.90	0.86	0.87	0.84	0.85	0.87
$h = 3$	0.75	0.87	0.84	0.86	0.82	0.82	0.82
$h = 4$	0.76	0.84	0.81	0.83	0.81	0.81	0.80
$h = 6$	0.78	0.80	0.79	0.80	0.79	0.79	0.80
$h = 8$	0.82	0.82	0.82	0.82	0.82	0.82	0.82
$h = 10$	0.85	0.84	0.85	0.84	0.85	0.85	0.82
$h = 12$	0.86	0.86	0.87	0.84	0.86	0.87	0.83
Panel B:Δ AUROC vs. spread-only (level difference)							
$h = 1$		0.18	0.15	0.13	0.14	0.14	0.17
$h = 2$		0.15	0.11	0.12	0.09	0.10	0.12
$h = 3$		0.12	0.09	0.11	0.08	0.08	0.08
$h = 4$		0.08	0.05	0.08	0.05	0.05	0.05
$h = 6$		0.02	0.01	0.02	0.01	0.01	0.02
$h = 8$		0.00	0.00	−0.00	0.01	0.01	0.00
$h = 10$		−0.00	0.00	−0.01	0.00	0.01	-0.02
$h = 12$		0.00	0.01	−0.02	0.00	0.01	-0.03

Note: Panel B reports differences computed from Panel A: $\Delta = \text{AUROC(model)} - \text{AUROC(spread-only)}$.

Best one-predictor forecasting models with and without nowcast augmentation (AUROC).

h	Leading indicators only	AUROC	Best NC-augmented model	AUROC	Δ	Second-best NC-augmented model	AUROC	Δ
1	S&P 500, 1y change (%)	0.97	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.97	+0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.97	+0.00
2	S&P 500, 1y change (%)	0.96	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.96	−0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.96	−0.00
3	S&P 500, 1y change (%)	0.95	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.94	−0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.94	−0.00
4	S&P 500, 1y change (%)	0.94	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.94	+0.00	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.94	+0.00
6	S&P 500, 1y change (%)	0.89	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.90	+0.01	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{RF}}$	0.90	+0.01
8	S&P 500, 1y change (%)	0.85	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.86	+0.01	S&P 500, 1y change (%) + $\widehat{N}^{(t)}_{t-1,\text{RF}}$	0.86	+0.01
10	5y–FF spread	0.86	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.86	+0.01	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.86	+0.00
12	5y–FF spread	0.87	5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.88	+0.01	1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.87	+0.00

Note: See the notes to Table 4.

Best two-predictor forecasting models with and without nowcast augmentation (AUROC).

h	Leading indicators only	AUROC	Best NC-augmented model	AUROC	Δ	Second-best NC-augmented model	AUROC	Δ
1	S&P 500, 1y change (%) + 1y–FF spread	0.97	S&P 500, 1y change (%) + S&P 500, 3y change (%) + $\widehat{N}^{(t)}_{t-1,\text{probit}}$	0.97	+0.00	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.97	+0.00
2	S&P 500, 1y change (%) + 1y–FF spread	0.96	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.96	+0.00	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.96	+0.00
3	S&P 500, 1y change (%) + 1y–FF spread	0.95	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.95	+0.00	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{kNN}}$	0.95	+0.00
4	S&P 500, 1y change (%) + 6m–FF spread	0.95	S&P 500, 1y change (%) + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{LVQ}}$	0.95	+0.00	S&P 500, 1y change (%) + 6m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.95	+0.00
6	S&P 500, 1y change (%) + 6m–FF spread	0.91	S&P 500, 1y change (%) + 6m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{RF}}$	0.92	+0.01	S&P 500, 1y change (%) + 6m–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.92	+0.01
8	S&P 500, 1y change (%) + 5y–FF spread	0.85	S&P 500, 1y change (%) + Baa–Aaa spread + $\widehat{N}^{(t)}_{t-1,\text{RF}}$	0.88	+0.03	S&P 500, 1y change (%) + Baa–Aaa spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.88	+0.03
10	Baa–Aaa spread + 5y–FF spread	0.86	Baa–Aaa spread + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.87	+0.01	S&P 500, 1y change (%) + 5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{RF}}$	0.87	+0.01
12	6m–FF spread + 5y–FF spread	0.87	Baa–Aaa spread + 5y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.88	+0.01	Baa–Aaa spread + 1y–FF spread + $\widehat{N}^{(t)}_{t-1,\text{GB}}$	0.88	+0.01

Note: See the notes to Table 4.