Empirical Economics (2026) 70:52
https://doi.org/10.1007/s00181-026-02893-7

Similarity-based path forecasting of US recession periods

Visa Kuntze[1,2] · **Henri Nyberg**[1] · **Samuel Rauhala**[1]

Received: 26 November 2024 / Accepted: 25 January 2026

Abstract

We develop a nonparametric similarity-based approach for binary time series that exploits recurring historical patterns to construct probability forecasts for all feasible multi-period outcome sequences. In contrast to conventional horizon-specific parametric models, our path forecasts are obtained simultaneously for all the horizons and remain internally consistent across them. Simulation experiments demonstrate that our method delivers accurate and robust performance in realistic sample sizes. In an empirical application to US business cycle data, our approach successfully anticipates the onset of the past three recessions about one year in advance and provides informative predictions of their expected duration.

Keywords Nonparametrics · Yield curve · Term spread · Interest rates

JEL Classification C14 · C25 · C53 · E32 · E43

We thank the Editor and three anonymous reviewers, Matthijs Lof, and participants of the Nordic Econometric Meeting (Oslo, 2024), Computational and Financial Econometrics (CFE, London, 2024), and GSF Summer Workshop in Finance (Turku, 2024) for useful discussions and comments. We gratefully acknowledge the financial support from the OP Group Research Foundation (grant 20230116), The Foundation for the Advancement of Finnish Securities Markets (grant 20230026), the Foundation for Economic Education (Liikesivistysrahasto, grant 220246), and (Kuntze) the Doctoral Programme in Exact Sciences (EXACTUS, University of Turku) and (Rauhala) The Finnish Doctoral Program Network in Artificial Intelligence, AI-DOC (decision number VN/3137/2024-OKM-6).

✉ Visa Kuntze
visa.kuntze@utu.fi

Henri Nyberg
henri.nyberg@utu.fi

Samuel Rauhala
samuel.j.rauhala@utu.fi

[1] Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland

[2] Department of Finance and Accounting, University of Turku, 20014 Turku, Finland

Published online: 20 February 2026

1 Introduction

In this paper, we propose a nonparametric similarity-based path forecasting approach for binary time series, illustrated by forecasting the state of the business cycle. Path forecasts answer questions such as whether a recession is immediate or more likely to occur later in the year, and how long it might last. By weighting past episodes according to their similarity to current conditions, the method produces internally consistent multi-horizon forecasts for entire sequences of future outcomes, rather than separate horizon-specific probabilities as in conventional binary response models.

Similarity-based forecasting has only recently gained attention in macroeconomic applications for real-valued dependent variables (see Dendramis et al. 2020; Guerròn-Quintana and Zhong 2023). In our approach, similarity is exploited by assigning greater importance to past observations that resemble current conditions and less weight to dissimilar cases. The resulting path forecasts are specific to the prevailing environment at each forecast origin and therefore more flexible than forecasts obtained from horizon-specific parametric models.

When applied to recession forecasting, an additional consideration arises. Our method yields probability forecasts for each viable expansion–recession sequence several quarters ahead and, by building forecasts directly from observed sequences, automatically excludes patterns (e.g., one-quarter recessions or expansions) that violate duration restrictions. Harding and Pagan (2011, 2016) highlight the importance of such duration restrictions and note that conventional parametric models cannot adequately account for them. As in their work, we rely on a nonparametric framework, but our focus is on obtaining multi-period forecasts through path forecasting (see also Jordá, ò., and M. Marcellino, 2010) while incorporating duration restrictions in the binary-valued state of the business cycle.

We assess the finite-sample performance of the method in simulation experiments and find that forecast accuracy improves with sample size while remaining high even in samples comparable in length to quarterly US business cycle data. This suggests that the approach is well suited to forecast real-world business cycles. As an empirical illustration, we forecast US business cycle sequences around the three most recent recession episodes using the term spread (the difference between long-term and short-term interest rates) as the sole leading indicator, consistent with much of the earlier literature. The results show that path forecasts often correctly identify the timing and duration of recessions, particularly their onset.

The remainder of the paper is organized as follows: Sect. 2 reviews the methodological framework, including parametric benchmarks and the proposed similarity-based estimator. Section 3 presents the forecasting framework, simulation studies, and an empirical illustration for the US recessions. Finally, Sect. 4 concludes.

2 Methodology

2.1 Parametric models and recession periods

Our interest lies in predicting the future state of the business cycle multiple periods ahead. Let $y_t \in \{0, 1\}$ denote the state of the business cycle, with $y_t = 1$ indicating a recession, and let $\boldsymbol{x}_t$ be a vector of predictors observed at time t. Following Estrella et al. (2003), we distinguish between two types of recession observations for a forecast horizon h: *marginal* observations and *cumulative* observations,

$$\begin{aligned} y_{t+h}^{(m)} &= \mathbb{1}(y_{t+h} = 1), \\ y_{t+h}^{(c)} &= \mathbb{1}\left(\max_{j=1,\ldots,h} y_{t+j} = 1\right), \end{aligned} \tag{1}$$

where $\mathbb{1}(\cdot)$ denotes an indicator function. The first records whether the economy will be in a recession exactly h periods ahead, while the second records whether a recession will occur at any time within the next h periods.[1] Throughout, we use $\kappa \in \{m, c\}$ to index these two definitions so that $y_{t+h}^{(\kappa)}$ represents either case in a unified notation.

A common parametric specification for the conditional probability of a recession h periods ahead is

$$P\left(y_{t+h}^{(\kappa)} = 1 \middle| \boldsymbol{x}_t\right) = E\left(y_{t+h}^{(\kappa)} \middle| \boldsymbol{x}_t\right) = G(\boldsymbol{x}_t'\boldsymbol{\beta}), \quad t = 1, \ldots, T - H, \tag{2}$$

where $\boldsymbol{\beta}$ denotes the parameter vector and H is the maximum forecast horizon of interest. Using the cumulative distribution function of the logistic or the standard normal distribution for the function $G(\cdot)$ leads to the logit and probit models, respectively, where the unknown parameters $\boldsymbol{\beta}$ are typically estimated by the method of maximum likelihood. The vector $\boldsymbol{x}_t$ contains the predictive variables, such as the term spread, and may also include, for example, lags of $y_t^{(\kappa)}$ or other dynamic structures (see, e.g., Chauvet and Potter 2005; Kauppi and Saikkonen 2008). In practice, horizon-specific models are estimated separately for each h, and the resulting forecasts need not be internally consistent across horizons. For instance, in the cumulative observations case, the probability of a recession within four quarters should logically be at least as large as that within three quarters, yet separate horizon-specific estimates do not guarantee this monotonicity.

Our contribution is motivated by an important but often underappreciated observation highlighted by Harding and Pagan (2011): The marginal business cycle state $y_t^{(m)}$ does not evolve freely between recession and expansion observations. The NBER Business Cycle Dating Committee defines recessions as lasting "more than a few months," and expansions, being the normal state of the economy, are at least as long.

[1] For a comparison of these two approaches, see Lahiri and Yang (2023). Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) consider the construction of related conditional hitting and continued expansion probabilities.

Table 1 Four-quarter US business cycle sequences and their relative sample frequencies, 1968:Q1–2020:Q4

Marginal obs $\left(y_{t+1}^{(m)}, y_{t+2}^{(m)}, y_{t+3}^{(m)}, y_{t+4}^{(m)}\right)$	Cumulative obs $y_{t+4}^{(c)}$	U.S. (Rel. freq.)
(0,0,0,0)	0	0.74
(0,0,0,1)	1	0.04
(0,0,1,1)	1	0.04
(0,1,1,0)	1	0.01
(0,1,1,1)	1	0.02
(1,0,0,1)	1	0.00
(1,0,0,0)	1	0.03
(1,1,0,0)	1	0.04
(1,1,1,0)	1	0.02
(1,1,1,1)	1	0.04

Columns 1–2 present the 10 feasible four-quarter ($H = 4$) sequences of marginal expansion and recession observations that satisfy the business cycle duration restrictions (i.e. both phases last more than one quarter between horizons $h + 1, \ldots, h + 4$) and the corresponding cumulative outcomes (see (1)). Paths such as "(0,0,0,1)" or "(1,1,1,0)" are feasible because $y_t^{(m)}$ is unknown in real time and $y_{t+5}^{(m)}$ is not considered. Column 3 shows the relative frequencies of different paths during 1968:Q1–2020:Q4

Consequently, in quarterly data, both phases last at least two quarters.[2] This duration restriction implies that a recession starting at time t (i.e. $y_{t-1}^{(m)} = 0$ and $y_t^{(m)} = 1$) must be followed by another recessionary observation $y_{t+1}^{(m)} = 1$. Hence, when predicting the values of $y_t^{(m)}$ up to H periods ahead, there are in principle 2^H possible sequences $(y_{t+1}^{(m)}, \ldots, y_{t+H}^{(m)})$, but the number of feasible paths is much smaller once the duration restrictions of business cycles are taken into account. Harding and Pagan (2011, 2016) demonstrate that these restrictions impose complex constraints on the data-generating process of business cycle states that conventional parametric models such as (2) cannot accommodate.

Table 1 illustrates this for a four-quarter horizon ($H = 4$) using quarterly US data from 1968:Q1–2020:Q4. Only 10 out of the $2^4 = 16$ possible paths are feasible once recessions and expansions shorter than one quarter are excluded. The sequence "(1,0,0,1)" is technically feasible under the restriction but does not occur in the data.

To summarize, two distinct challenges arise in modeling and forecasting binary-valued business cycle regimes. First, the duration restrictions imply that the state of the business cycle $y_t^{(m)}$ cannot freely transition between regimes, yet conventional parametric models do not incorporate these constraints. Second, separate estimation of horizon-specific models typically results in forecasts that are not internally consistent across horizons, especially for the cumulative case where the probability of a recession within h periods should weakly increase with h. Addressing both issues requires a

[2] This was also true for the short pandemic recession of 2020. More information is available on the NBER website (https://www.nber.org/research/business-cycle-dating/business-cycle-dating-procedure-frequently-asked-questions)

framework that can produce logically coherent, duration-consistent forecasts without relying on parametric distributional assumptions. In the next section, we introduce such an approach by developing a nonparametric similarity-based estimator for path forecasting.

2.2 Similarity-based path forecasts

We now turn to our nonparametric path forecasting approach, which relies on identifying and weighting past (training sample) periods whose predictor values resemble those prevailing at the forecast origin t. Let $\boldsymbol{x}_t$ denote the current predictor vector and $\boldsymbol{x}_s$, for $s = 1, \ldots, t-H$, the corresponding vectors from the historical training sample. As in the parametric framework (2), forecasts are conditioned on $\boldsymbol{x}_t$, but instead of estimating parameters $\boldsymbol{\beta}$, we infer probabilities directly from the observed data by giving greater weight to past observations characterized by $\boldsymbol{x}_s$ that are more similar to $\boldsymbol{x}_t$.

Formally, the conditional probability of a recession h periods ahead is estimated as:

$$P(y_{t+h}^{(\kappa)} = 1 \mid \boldsymbol{x}_t) = \frac{\sum_{s=1}^{t-H} w_s \, y_{s+h}^{(\kappa)}}{\sum_{s=1}^{t-H} w_s} \equiv \sum_{s=1}^{t-H} w_s^* \, y_{s+h}^{(\kappa)}, \quad (\kappa \in \{m, c\}), \tag{3}$$

where $w_s^* = w_s / \sum_{j=1}^{t-H} w_j$ are normalized similarity weights. The summation extends to $t-H$ because forecasts for all horizons $h \leq H$ are produced simultaneously using the same information set, ensuring internal consistency across horizons. To measure the similarity between the standardized predictor vectors $\widetilde{\boldsymbol{x}}_t$ and $\widetilde{\boldsymbol{x}}_s$, we use the Euclidean distance

$$d_s = \|\widetilde{\boldsymbol{x}}_t - \widetilde{\boldsymbol{x}}_s\|, \quad w_s = \exp(-d_s^a), \tag{4}$$

where the exponential transformation converts distances into positive similarity weights.[3] The exponent $a > 0$ controls how sharply weights decline with distance: a small value of a yields smoother weighting, whereas large a emphasizes the most similar observations. We will return to the determination of a in Sect. 3.

Building on this idea, forecasts for entire future sequences can be obtained simultaneously for all horizons $h = 1, \ldots, H$. Let $(y_{t+1}^{(m)}, \ldots, y_{t+H}^{(m)})$ denote a candidate sequence of marginal business cycle outcomes over the next H periods. The conditional probability of observing a particular sequence given $\boldsymbol{x}_t$ is estimated as:

$$P\Big(y_{t+1}^{(m)} = y_1, \ldots, y_{t+H}^{(m)} = y_H \,\Big|\, \boldsymbol{x}_t\Big) = \sum_{s=1}^{t-H} w_s^* \, \mathbb{1}\Big(y_{s+1}^{(m)} = y_1, \ldots, y_{s+H}^{(m)} = y_H\Big), \tag{5}$$

where the indicator function $\mathbb{1}(\cdot)$ equals one whenever the same sequence occurs in the training data. The weights w_s^* are the same across all horizons h and for both $\kappa = m$ and $\kappa = c$, which guarantees that forecasts remain mutually consistent. Equation (3) is recovered as a special case of (5) for a single horizon h.

[3] Standardization and the removal of possible trends ensure that different variable scales do not affect how distances are measured.

A key advantage of (5) is that it automatically respects the duration restrictions discussed in Sect. 2.1. That is, because the estimator is constructed from realized training sample sequences, only empirically feasible paths receive nonzero probabilities. For instance, with $H = 4$, forecasts are based on the ten feasible US business cycle sequences listed in Table 1. As an example, the conditional probability of the sequence "(0,0,1,1)", that is

$$P(y_{t+1}^{(m)} = 0, y_{t+2}^{(m)} = 0, y_{t+3}^{(m)} = 1, y_{t+4}^{(m)} = 1 \mid \boldsymbol{x}_t),$$

is obtained in (5) as

$$\sum_{s=1}^{t-4} w_s^* \, \mathbb{1}\left(y_{s+1}^{(m)} = 0, y_{s+2}^{(m)} = 0, y_{s+3}^{(m)} = 1, y_{s+4}^{(m)} = 1\right),$$

and analogous expressions apply for the other feasible paths.

The common weighting structure across horizons also ensures internal consistency between path and single-horizon forecasts. For example, in the $H = 4$ case, the probability of being in recession four quarters ahead, $P(y_{t+4}^{(m)} = 1 \mid \boldsymbol{x}_t)$, equals the sum of the conditional probabilities for all feasible paths ending in a recession, "(0,0,0,1)", "(0,0,1,1)", "(0,1,1,1)", "(1,0,0,1)" and "(1,1,1,1)". This property stands in contrast to horizon-specific parametric models, which provide no such cross-horizon coherence.[4]

Conceptually, estimating sequence probabilities as in (5) could be viewed as a multinomial classification problem with one category per feasible path. However, a parametric multinomial model would be difficult to estimate in practice, as some feasible sequences (such as "(1,0,0,1)") are entirely absent from the data. In contrast, the nonparametric estimator simply assigns a zero probability to such cases and guarantees the internal coherence between path forecasts and marginal probabilities.

Finally, the estimator in (5) is closely related to kernel and nearest-neighbor methods (see Li and Racine 2007, for a review; cf. Harding and Pagan 2011). A k-nearest-neighbor version of (5) could be obtained by setting weights $w_s = 1/k$ for the training sample observations with the k smallest distances d_s in (4), and zero otherwise, constructing the forecast as the relative frequency of each path among those nearest neighbors. In practice, determining an appropriate k is challenging for time-series data, especially for recession episodes, where duration restrictions and the limited number of observations make cross-validation unreliable. Our exponential similarity-weighting scheme in (4) avoids these issues by using a continuous weighting parameter a and retaining the duration restrictions by construction. As shown in Sect. 3, our approach performs at least as well as its kNN counterpart in simulations and in the US business cycle data.

In summary, the proposed estimator (4)–(5) provides a simple yet powerful way to produce coherent multistep forecasts for binary time series without relying on

[4] Our approach also differs from the iterative forecasting schemes of Dueker (2005) and Kauppi and Saikkonen (2008), which do not impose duration restrictions and may assign nonzero probabilities to infeasible sequences.

parametric assumptions. By basing forecasts directly on observed historical sequences, it respects the empirical duration restrictions of business cycle phases and ensures internal consistency across forecast horizons. In the following section, we assess the finite-sample performance of this approach in simulations and illustrate its empirical usefulness for predicting US business cycle sequences.

3 Empirical results

In this section, we evaluate the performance of the proposed similarity-based path forecasting method in both simulated and empirical settings. We first outline the general forecasting framework and evaluation design relevant for business cycle data that is then used consistently in subsequent simulation experiments and for empirical results on the US business cycle regimes.

3.1 Forecasting framework and evaluation criteria

Our forecasting design is common to both the simulation and empirical analyses. We use quarterly US business cycle data for the sample period 1968:Q1–2020:Q4 and the term spread (10-year yield minus 3-month rate) as the sole predictor, consistent with prior work emphasizing methodological contributions (e.g., Estrella et al. 2003; Chauvet and Potter 2005; Kauppi and Saikkonen 2008; Haubrich 2021). Forecasts are produced with an expanding window: at each origin, the similarity-based estimator generates probabilities for all feasible four-quarter sequences ($H = 4$), the training sample is then updated with the new observation, and a new forecast is generated for the subsequent four quarters. To mimic real-time data availability, the business cycle state is assumed to be published with a three-quarter delay, so at time t only observations up to $t - 3$ are used for estimation.

As an evaluation criterion for path forecasts obtained with our method, we adopt a modified version of the pseudo-R^2 measure proposed by Estrella (1998):

$$\text{pseudo}-R^2 = 1 - \left(\frac{l_u}{l_c}\right)^{-\frac{2}{T^*-H} l_c}, \tag{6}$$

where

$$l_u = \sum_{t^*=1}^{T^*-H} \log\left\{ P\left(y^{(m)}_{t^*+1} = y1, \ldots, y^{(m)}_{t^*+H} = yH \middle| \boldsymbol{x}_{t^*}\right)\right\}. \tag{7}$$

The predicted probabilities in (7) for the test sample paths $\left(y^{(m)}_{t^*+1} = y1, \ldots, y^{(m)}_{t^*+H} = yH\right)$, $t^* = 1, \ldots, T^* - H$, are obtained as described in (5). Here, l_u summarizes the average log-scores from our nonparametric forecasts, while l_c is computed analogously from a naive benchmark that assigns probabilities equal to the sample frequencies of different paths in the training data. In rare cases where a feasible path is absent from

the training sample, a minor adjustment is required for the log-score calculation.[5] It is important to note that, unlike in the conventional use of pseudo-R^2 in parametric binary or multinomial models, the present statistic does not rely on a parametric likelihood derived from an assumed distribution. Instead, it should be viewed as a log-score analogue that quantifies an interpretable and comparable summary of proportional improvement in predictive performance over the benchmark. Higher values indicate more accurate forecasts, whereas negative values imply inferior performance to the naive benchmark.

Having established the forecasting framework and evaluation criterion, we now assess the performance of the similarity-based forecasting method in simulation studies before turning to the out-of-sample forecasting results in Sect. 3.3. The simulations replicate the empirical forecasting setup, including the expanding window scheme, the three-quarter publication lag, the four-quarter horizon ($H = 4$), and the pseudo-R^2 evaluation criterion, to ensure full comparability between the simulated and real-data results. We evaluate the small-sample performance of our forecasting method, as well as its robustness to different selections of the similarity-weighting parameter a in (4) and compare its performance to kNN-based weighting.

3.2 Simulation results

The simulation experiments in this section assess how forecast accuracy evolves with sample size and how sensitive it is to the choice of the similarity-weighting parameter a. To mirror the empirical analysis as closely as possible, artificial simulated data are generated from a process that reproduces the joint dynamics of the term spread and the business cycle in the US data. This setup allows us to study the finite-sample predictive accuracy of similarity-based path forecasts under realistic relationships between the variables.

For each Monte Carlo replication $i = 1, \ldots, N_{\text{sim}}$, we generate a sample of paired observations $(x_t^{(i)}, y_{t+4}^{(i)})$ for $t = 1, \ldots, T-4$, where $x_t^{(i)}$ is a continuous predictor and $y_{t+4}^{(i)}$ is a binary outcome that is predicted at time t. The data-generating process is designed to mimic the empirical relationship between the term spread and the marginal recession observations. Specifically, the predictor follows an autoregressive (AR) process, and the binary variable is drawn conditionally on $x_t^{(i)}$ from a parametric probit model

$$y_{t+4}^{(i)} \sim \text{Bernoulli}(p_{t+4}^{(i)}), \quad p_{t+4}^{(i)} = P(y_{t+4}^{(i)} = 1 \mid x_t^{(i)}) = \Phi(\beta_0 + \beta_1 x_t^{(i)}), \tag{8}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cf. $G(\cdot)$ in (2)). The parameter values (β_0, β_1) are obtained by estimating a probit model on the empir-

[5] If a feasible path appears in the test sample but not in the training data, the forecast assigns it probability zero, leading to an undefined log term in (7). For the evaluation metric, we replace such zeros by a small positive constant (e.g., 10^{-12}), while leaving the forecasts themselves unchanged. An alternative is to shrink the forecast toward equally weighted predictions over the feasible paths by replacing $\hat{p}$ entering (7) by $\tilde{p} = \gamma\hat{p} + (1-\gamma)u$, where u assigns equal probability to each feasible path and $1-\gamma$ is a small positive number. In practice, for reasonable choices of γ, this yields numerical results that are nearly identical to the simple constant-based adjustment used in the paper.

Table 2 Average pseudo-R^2s by sample size

Sample size (T)	Mean pseudo-R^2	SD of pseudo-R^2
200	0.37	0.15
300	0.41	0.11
400	0.42	0.09
500	0.43	0.08
600	0.44	0.07
750	0.44	0.06

Each entry is based on $N_{\text{sim}} = 1{,}000$ simulated replications of (8). Means and standard deviations (SDs) are calculated over these 1,000 replications and T is the length of the simulated time series

ical US data with the term spread as predictor, while the AR parameters of x_t are calibrated from an autoregressive model fit to the observed term spread.[6] We set the weighting parameter to $a = 2$ in the baseline analysis and later examine its robustness to alternative values.

Each simulated replication uses the first half of observations as the initial training sample, after which the forecasts are produced recursively in an expanding window scheme that mirrors the real-time setup in the empirical application, including a publication lag of three quarters and a fixed four-quarter forecast horizon ($h = 4$). To reflect the empirical duration restrictions of the business cycle, infeasible short sequences—such as isolated recessions or expansions lasting only one quarter (patterns "010" or "101")—are removed by merging any run shorter than two quarters with the adjacent state, yielding a binary series $y_t^{(i)}$ consistent with the observed minimum duration properties.

Table 2 reports the average pseudo-R^2 predictive statistics across different sample sizes. Each reported entry is based on 1,000 Monte Carlo replications. Forecast accuracy improves systematically with sample size: the mean pseudo-R^2 increases from approximately 0.37 at $T = 200$ to around 0.43 at $T = 500$, while its dispersion declines steadily, indicating more stable forecast performance in larger samples. Beyond $T = 600$, improvements are minor and the pseudo-R^2 stabilizes near 0.44, suggesting that the method reaches its effective prediction precision limit well within the range of empirically relevant sample sizes. Given that the US postwar data (up to 2020s) contain roughly 300 quarterly observations, this result implies that the method performs robustly under realistic sample sizes and that additional data are expected to yield only modest further gains in predictive accuracy.

To assess the robustness of the results to the specific choice of similarity weighting, we examine alternative values of the exponent parameter a in the weighting scheme (4), which governs how sharply the weights decline with distance. As mentioned, the results in Table 2 are based on the selection $a = 2$. Smaller values of a assign relatively uniform weights across observations, producing smoother but less localized forecasts, whereas larger values emphasize only the most similar historical episodes. We consider

[6] Specifically, we estimate an AR(2) model with a constant term using the US data. The AR model order is chosen by AIC.

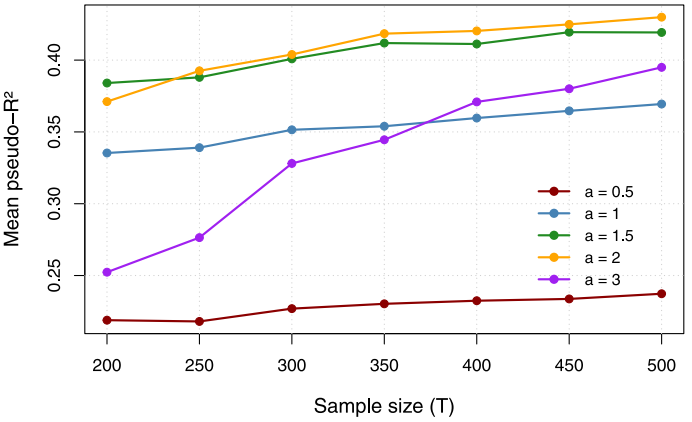

Fig. 1 Average pseudo-R^2 across $a \in \{0.5, 1, 1.5, 2, 3\}$ and sample sizes T. Each line represents results across 1,000 replications

$a \in \{0.5, 1, 1.5, 2, 3\}$ across sample sizes $T \in \{200, 250, 300, 350, 400, 450, 500\}$, and simulation results are again based on 1,000 simulated replications.

Figure 1 shows a consistent pattern across all time-series lengths T. Forecast accuracy improves markedly when moving from very small exponents ($a = 0.5$) to moderate values ($a = 1.5$–2.0), but declines again when the weights in (5) become too concentrated ($a = 3$). Overly diffuse weighting leads to underfitting, while overly sharp weighting produces unstable, noise-sensitive forecasts. Across settings, the pseudo-R^2 stabilizes around 0.40–0.43 for $a \approx 1.5$–2.0, with $a = 2$ yielding slightly superior accuracy. Consequently, we adopt $a = 2$ in the empirical analysis as a robust compromise between flexibility and localization.

As a final simulation experiment, we compare our similarity-based approach with a discrete local-averaging alternative based on the k-nearest-neighbor (kNN) estimator described in Sect. 2.2. In contrast to the exponentially decaying similarity weights (4) over all the past training sample observations, the kNN version forms the forecast using only the k historical sequences of y_t associated with the most similar values of the predictor at each forecast origin. To evaluate kNN performance, we simulate data from the same data-generating process as before and compute pseudo-R^2 values across 1,000 replications for nearest-neighborhood sizes $k \in \{10, 20, \ldots, 140\}$, using $T = 300$ as a representative sample length relevant for the quarterly US data (Table 3).

Forecast accuracy depends strongly on k. For small neighborhoods ($k \leq 70$), the kNN estimator overfits individual observations, producing extreme probabilities and negative pseudo-R^2. As k increases, forecasts smooth and stabilize, peaking near $k = 120$ (mean pseudo-$R^2 = 0.26$), still below the similarity-weighted estimator despite $k = 120$ covering about 40% of the sample. Overall, the results indicate that our similarity weighting in (4) produces more stable and accurate forecasts than the discrete kNN alternative, which is sensitive to the choice of the neighborhood size.

In summary, the simulation results demonstrate that our similarity-based approach provides accurate and stable forecasts in realistic data environments. Predictive accuracy improves steadily with sample size, remains robust to moderate changes in the

Table 3 Average pseudo-R^2 by kNN neighborhood size k

k	Mean pseudo-R^2	SD of pseudo-R^2
≤ 70	Negative	
80	0.11	0.41
90	0.20	0.33
100	0.23	0.29
110	0.25	0.30
120	0.26	0.20
130	0.22	0.20
140	0.21	0.17

The table reports mean and standard deviation (SD) of pseudo-R^2 statistics across 1,000 simulated replications for each neighborhood size k in kNN similarity weighting. The sample size is fixed at $T = 300$

similarity-weight parameter a, and clearly outperforms the k-nearest-neighbor alternative. These findings indicate that the estimator is well behaved in finite samples and achieves a favorable balance between flexibility and smoothness. Next, we turn to the empirical application to assess how the findings of this section on different tuning selections translate to actual US business cycle path forecasting performance.

3.3 Out-of-sample forecasting

Having established the finite-sample properties of our similarity-based method in simulations, we now apply it to US business cycle data to evaluate its practical forecasting performance. The results below follow the same recursive expanding window framework and forecast horizon ($H = 4$) as in the simulation studies, using the term spread as the sole predictor. At each forecast origin, our method produces a full set of probability forecasts for all feasible four-quarter sequences of recession and expansion outcomes, allowing us to examine how accurate our predictions are around historical turning points.

To enable practical decision-making, the forecaster must translate the probability forecasts into a single predicted path. Because the expansionary sequence "(0,0,0,0)" is almost always the most probable due to the rarity of recessions, a simple argmax choice would be uninformative. We therefore employ a decision rule that classifies a sequence as expansionary if the probability of "(0,0,0,0)" exceeds a pre-specified threshold ξ, and as recessionary otherwise, selecting in that case the most likely recessionary path. This rule preserves interpretability under severe class imbalance and enables a meaningful assessment of how well the forecasts capture the onset and duration of recessions.

In practice, the decision rule is implemented in three steps:

1. Compute the probability forecasts for all feasible sequences. For $H = 4$, there are ten possible business cycle paths (see Table 1).
2. If the probability of a continuing expansion, "(0,0,0,0)", exceeds a pre-specified threshold ξ, classify the forecast as expansionary.

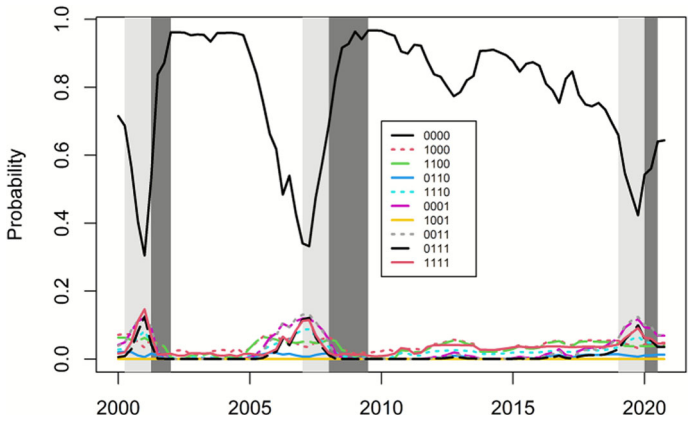

Fig. 2 Out-of-sample forecasts for the 10 different feasible business cycle sequences 4 quarters ahead. Observations up to 1999:Q4 are used for the initial training of our path forecasts obtained with (3), while an expanding window strategy is subsequently used to recursively update the training sample. The actual recession episodes and the four preceding quarters, for which the four-period path forecasts contain the upcoming recession, are highlighted with dark and light gray, respectively

3. Otherwise, select the most likely recessionary sequence as the predicted path.

To illustrate how the decision rule operates in practice, Fig. 2 shows the resulting out-of-sample forecasts for all ten feasible four-quarter business cycle sequences. We use $\xi = 0.72$, corresponding to the historical frequency of the expansionary sequence "(0,0,0,0)" in the training sample (1967:Q1–1999:Q4). At each forecast origin, the similarity-based estimator assigns the highest probability to the continuing-expansion path, but this probability declines markedly in the quarters preceding recessions. Whenever it falls below the threshold ξ, the rule switches to reporting the most likely recessionary sequence, effectively signaling a downturn. The figure highlights these episodes: light gray areas mark periods when the four-quarter forecast window already contains an upcoming recession, and dark gray areas show the actual recessions.

Table 4 reports the predicted probabilities for all feasible four-quarter business cycle sequences around major turning points in the 2000s. Around the 2001 recession, our approach signals a downturn about one year in advance (2000:Q2), with the most likely recessionary path matching the realized "(0,0,0,1)". Subsequent origins place highest weight on paths where the recession begins later in the horizon, consistent with the moderate leading time of the term spread. By 2001:Q3, the continuing-expansion probability exceeds the threshold, correctly indicating recovery albeit one quarter too early.

A similar pattern emerges around the 2007–2009 Great Recession. Beginning in 2007:Q1, the probability of a recession within the next year rises sharply, with the path forecasts correctly identifying recessionary paths as most likely. The highest-probability sequences, "(0,0,1,1)" and "(0,0,0,1)", indicate that a downturn was expected to start within two to four quarters, which is close to what materialized. In 2009:Q1, one quarter before the trough, the probability of the expansionary path "(0,0,0,0)" again exceeds the threshold, signaling recovery, while the correct short

Table 4 Four-quarter ($H = 4$) path forecasts for the US business cycle around different recession episodes and forecast origins

Marginal obs $\left(y_{t+1}^{(m)}, y_{t+2}^{(m)}, y_{t+3}^{(m)}, y_{t+4}^{(m)}\right)$	Cum. obs $y_{t+4}^{(c)}$	Forecast origin ($\boldsymbol{x}_t$)							
		2000:Q2	2000:Q3	2000:Q4	2001:Q3	2007:Q1	2007:Q2	2007:Q3	2009:Q1
(0,0,0,0)	0	0.580	0.435	0.343	0.872	0.363	0.495	0.590	0.941
(1,0,0,0)	1	0.072	0.049	0.034	0.035	0.035	0.050	0.059	0.020
(1,1,0,0)	1	0.056	0.054	0.062	0.038	0.051	0.046	0.049	0.016
(0,1,1,0)	1	0.021	0.017	0.022	0.008	0.016	0.015	0.016	0.001
(1,1,1,0)	1	0.039	0.058	0.067	0.017	0.078	0.065	0.051	0.005
(0,0,0,1)	1	0.082	0.112	0.121	0.010	0.119	0.105	0.087	0.001
(1,0,0,1)	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0,0,1,1)	1	0.079	0.108	0.119	0.007	0.129	0.104	0.079	0.001
(0,1,1,1)	1	0.021	0.055	0.086	0.000	0.092	0.049	0.027	0.000
(1,1,1,1)	1	0.050	0.112	0.146	0.014	0.115	0.071	0.042	0.015

The probability forecasts corresponding to the sequences that subsequently took place are highlighted by boxes in the table

recession-ending sequence "(1,0,0,0)" remains the most likely among recessionary paths.

Table 5 summarizes the results for the COVID-19 recession. During 2019:Q1—2019:Q2, when the term spread was near zero, the risk of recession increases and the correct paths "(0,0,0,1)" and "(0,0,1,1)" receive top or second-highest probability. In 2019:Q3, the model points to an impending downturn, but not the unusually brief COVID-19 recession, reflecting the scarcity of such short episodes in the training data. The late 2019 and early 2020 forecasts miss the abrupt end of the recession, a limitation consistent with the exceptional nature of the event.

In addition to the threshold-based evaluation above, we also report pseudo-R^2 values for the empirical out-of-sample forecasts to maintain comparability with the simulation evidence in Sect. 3.2. Table 6 presents the out-of-sample pseudo-R^2 statistics for different values of the parameter a in the similarity-weighting scheme (4) over the evaluation period 2000:Q1–2020:Q4. The values are positive and of broadly similar magnitude to those obtained in the simulation study, confirming that the method delivers comparable predictive performance on real data. Forecast accuracy remains stable across a wide range of a, with the pseudo-R^2 peaking near $a = 1.5$–2.0 and declining slightly thereafter. Although $a = 1.5$ yields the marginally highest pseudo-R^2, the differences are small, and the baseline choice $a = 2$ provides a balanced and robust specification. Additional robustness checks using the k-nearest-neighbor version of the estimator, along with a discussion of its practical implementation challenges and comparison to the similarity-weighted approach, are provided in Appendix A.

Overall, the path forecasts made two to four quarters before the onset of each recession align closely with the realized sequences, while the end of recessions is harder to predict. This asymmetry reflects the well-known fact that the term spread is most informative about the onset of downturns, while offering little guidance on when recessions end.[7] Consistent with this, forecast accuracy is highest three to four quarters before recessions and declines when the forecast horizon shortens toward the turning point.

4 Conclusions

This paper introduces a novel nonparametric similarity-based approach for producing path forecasts for binary time series. The method provides probability forecasts for all feasible multi-period sequences of future outcomes, offering a coherent alternative to standard direct or iterative forecasting approaches. By generating internally consistent probability forecasts across horizons, the framework accommodates duration restrictions on feasible future paths, which is particularly relevant in recession forecasting.

Our simulation results show that the method performs accurately and robustly in realistic sample sizes, with forecast precision improving steadily as the amount of data increases. In the empirical application to US business cycles, the method successfully

[7] In line with this, Paap et al. (2009) find that the Composite Index of Leading Indicators (CLI), which includes the term spread, exhibits much longer lead times at business cycle peaks than at troughs.

Table 5 Four-quarter ($H = 4$) path forecasts for the US business cycle around the COVID-19 pandemic recession

Marginal obs $\left(y_{t+1}^{(m)}, y_{t+2}^{(m)}, y_{t+3}^{(m)}, y_{t+4}^{(m)}\right)$	Cum. obs $y_{t+4}^{(c)}$	Forecast origin ($\boldsymbol{x}_t$) 2019:Q1	2019:Q2	2019:Q3	2019:Q4	2020:Q1
(0,0,0,0)	0	0.560	0.503	0.446	0.557	0.574
(1,0,0,0)	1	0.045	0.039	0.032	0.043	0.044
(1,1,0,0)	1	0.039	0.038	0.039	0.038	0.038
(0,1,1,0)	1	0.013	0.012	0.012	0.012	0.012
(1,1,1,0)	1	0.048	0.055	0.061	0.049	0.046
(0,0,0,1)	1	0.093	0.106	0.117	0.094	0.090
(1,0,0,1)	1	0.000	0.000	0.000	0.000	0.000
(0,0,1,1)	1	0.093	0.109	0.121	0.095	0.090
(0,1,1,1)	1	0.048	0.063	0.079	0.050	0.046
(1,1,1,1)	1	0.061	0.076	0.091	0.063	0.060

The probability forecasts corresponding to the sequences that actually took place are highlighted by boxes in the table

Table 6 Out-of-sample pseudo-R^2 values for different tuning parameter a selections

a	pseudo-R^2
0.5	0.143
1.0	0.205
1.5	0.215
2.0	0.207
3.0	0.184

The table reports pseudo-R^2 values for the empirical out-of-sample forecasts over 2000:Q1–2020:Q4. The parameter value $a = 2$ corresponds to the baseline specification used throughout the empirical analysis

signals the onset of the past three recessions about one year in advance and provides reasonably accurate indications of their duration and recovery phases. The predictive accuracy is stable across a broad range of specific selections in similarity weighting, reinforcing the robustness of our results. Taken together, the simulation and empirical results demonstrate that the proposed path forecasting framework offers a practical and interpretable tool for multi-period business cycle forecasts.

Several extensions remain for future research. Throughout our analysis, we have remained agnostic about the contemporaneous state of the business cycle at the forecast origin, which in real time is typically uncertain. Combining the proposed path forecasting approach with nowcast probabilities of the current regime (see, for example, Kuntze 2023 would provide a natural extension. Another promising avenue would be to incorporate a broader set of leading indicators beyond the term spread. Our results indicate that already this simple single-predictor specification captures much of the predictive power that is relevant for forecasting US recessions.

Appendix A: Robustness results with kNN

To complement the simulation evidence, this appendix reports an additional robustness check using the empirical US data from 2000:Q1 to 2020:Q4. Table 7 reports results for the k-nearest-neighbor (kNN) version of our path forecasting method.

Table 7 Out-of-sample pseudo-R^2 values for different neighborhood sizes k in the k-nearest-neighbor (kNN) variant of path forecasts (US data, 2000:Q1–2020:Q4)

k	pseudo-R^2
≤ 80	negative
90	0.210
100	0.187
110	0.157
120	0.117

Values below zero (for $k \leq 80$) are grouped as "negative". See also the notes to Table 6

In contrast to the simulation study, the empirical pseudo-R^2 values for the kNN specification are quite similar to those of the exponentially decaying similarity weights (4), suggesting that both versions of path forecasts capture comparable out-of-sample forecasting information from the term spread in real data. Forecast accuracy is relatively robust for larger neighborhoods ($k \geq 90$) and stabilizes around 0.20, whereas very small values of k yield highly unstable forecasts with negative pseudo-R^2s.

In practice, determining an appropriate neighborhood size k for expanding window recession forecasts is challenging. Standard cross-validation procedures are ill-suited in this context because y_t exhibits strong serial persistence and duration restrictions, with expansions and recessions occurring in blocks of at least two observations. Splitting the sample into independent folds would distort this underlying structure. Moreover, the limited number of distinct business cycle episodes offers little scope for constructing separate validation samples. These features make the similarity-weighted approach in (4) particularly attractive, as it avoids discrete bandwidth selection entirely, adapts smoothly to local data density through the continuous weighting parameter a, and, by construction, preserves the duration structure inherent in the observed business cycle sequences.

Overall, these results indicate that while the exponential similarity-weighting scheme provides smoother forecasts, the kNN alternative performs quite well in practice, but poses greater implementation challenges and uncertainties.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital).

References

Chauvet M, Potter S (2005) Forecasting recessions using the yield curve. J Forecast 24:77–103

Dendramis Y, Kapetanios G, Marcellino M (2020) A similarity-based approach for macroeconomic forecasting. J R Stat Soc Ser A 183:801–827

Dueker M (Dueker) Dynamic forecasts of qualitative variables: a qual VAR model of U.S. recessions. J Bus Econ Stat 23:96–104

Estrella A (1998) A new measure of fit for equations with dichotomous dependent variables. J Bus Econ Stat 16:198–205

Estrella A, Rodrigues A, Schich S (2003) How stable is the predictive power of the yield curve? Evidence from Germany and the United States. Rev Econ Stat 85:629–644

Guerròn-Quintana P, Zhong M (2023) Macroeconomic forecasting in times of crises. J Appl Economet 38(3):295–320

Harding D, Pagan A (2011) An econometric analysis of some models for constructed binary time series. J Bus Econ Stat 29(1):86–95

Harding D, Pagan A (2016) The econometric analysis of recurrent events in macroeconomics and finance. Princeton, New Jersey

Haubrich J (2021) Does the yield curve predict output? Annu Rev Financ Econ 13:341–62
Jordá Ò, Marcellino M (2010) Path forecast evaluation. J Appl Economet 25:635–662
Kauppi H, Saikkonen P (2008) Predicting U.S. recessions with dynamic binary response models. Rev Econ Stat 90:777–791
Kuntze V (2023) The role of coincident information in real-time business cycle forecasting. SSRN Research Papers, id. 4692581
Lahiri K, Yang C (2023) A tale of two recession-derivative indicators. Empiric Econ 65(2):925–947
Li Q, Racine JS (2007) Nonparametric econometrics: theory and practice. Princeton University Press, Princeton
Paap R, Segers R, van Dijk D (2009) Do leading indicators lead peaks more than troughs? J Bus Econ Stat 27(4):528–543