A new approach to forecasting the probability of recessions after the COVID-19 pandemic *

Maximo Camacho†
Universidad de Murcia

Salvador Ramallo
Universidad de Murcia

Manuel Ruiz
Universidad Politécnica de Cartagena

June 1, 2023

Abstract

The traditional parametric techniques used to forecast recession probabilities from economic indicators have become unsettled due to the few but highly influential observations recorded by most industrialized economies during the COVID-19 pandemic. This paper proposes a new nonparametric approach to computing predictive probabilities of future recessions that is robust to influential observations and other data irregularities, such as structural breaks and heteroskedasticity. The method involves simulating forecasts using past histories of the time series which are embedded into a symbolic space. Then, the forecasts are converted into probability statements, which are weighted by the forecast probabilities of their respective symbols. Using GDP data from the G7 countries, we show that our nonparametric proposal outperforms other linear and nonlinear parametric approaches as a classifier of future national business cycle phases, especially when data from 2020 are included in the sample.

Keywords: Business cycles, Symbolic dynamics, Nonparametric models.
JEL Classification: E32, C22, E27.

*This study is the result of activity carried out in the program Groups of Excellence of the region of Murcia, the Fundación Seneca, Science and Technology Agency of the region of Murcia project 19884/GERM/15. The authors are grateful for the support of grants PID2019-107192GB-I00 (MCIN/AEI/10.13039/501100011033), PID2022-136547NB-I00 and PID2019-107800GB-I00 (MCIN/AEI/10.13039/501100011033). Errors are our responsibility. Data and codes that replicating our results are available from the authors' websites.

†Corresponding author: Universidad de Murcia, Departamento de Metodos Cuantitativos para la Economia y la Empresa, 30100, Murcia, Spain. E-mail: mcamacho@um.es

1 Introduction

Early detection of changes in business cycle phases is crucial for consumption, investment, savings, and production decisions made by economic agents, as well as for making monetary and fiscal policies. Since phase changes are officially recognized long after they start, developing early warning mechanisms to forecast recessions has been a long-standing quest for academics, market practitioners, and policymakers.

To provide timely early warning of an approaching recession, academics usually rely on nonlinear parametric methods that produce probabilistic statements of future phase changes from business cycle indicators, such as the growth rates of quarterly real Gross Domestic Product (GDP). To name only a few, Estrella and Mishkin (1998) used a probit model, Hamilton (1989) developed a Markov-switching autoregressive (MSAR) specification, and Teräsvirta and Anderson (1993) proposed a Smooth Transition Autoregressive (STAR) model.

In addition, Wecker (1979) offered a heuristic solution to compute inferences of future recessions from linear autoregressive (AR) models based on Monte-Carlo simulations of forecast paths. Considering a technical recession (two consecutive quarters of decline in the GDP) as a recession event, the method consists of computing the relative frequency of technical recessions across the simulated forecasts. Despite its simplicity, Hamilton and Perez-Quiros (1996) and Camacho and Perez-Quiros (2002) showed the considerable empirical reliability of this approach to provide forecasts of US recession probabilities.

One way to assess the goodness of fit of recession forecast models is to examine their ability to identify official recessions as determined by national Dating Committees, such as the committee of the National Bureau of Economic Research (NBER) for the US economy. Regardless of how successful the parametric methods have been in the past, in 2020, the parametric dating methods were exposed to unprecedented atypical data that challenged their ability to provide reliable probability forecasts of future recessions. In 2020, due to the economic collapse caused by the COVID-19 pandemic and the rapid countercyclical measures implemented by policymakers, most industrialized countries recorded the sharpest fall and the largest rebound in quarterly GDP since records began. In this paper, we show that these leverage points have dramatically altered the state of affairs in performing business cycle inferences.[1]

To overcome this drawback, the empirical approaches used to compute forecasts of recession probabilities with parametric models rely on the shortcut of manipulating the sample to estimate the model parameters. One example is the Econbrowser GDP-based recession indicator index, which estimates recession probabilities by applying a methodology based on the Markov-switching model developed by Chauvet and Hamilton (2006). To keep the index working after the dramatic drop in the second quarter of 2020, the parameters of the Markov-switching model were not estimated but fixed with the values of the estimated parameters using data only up to the first quarter of 2020. In the same vein, McGrane (2022) obtained the post-COVID recession probabilities from a

[1] Although not in the context of forecasting recessions, Lenza and Primiceri, 2020, Leiva-Leon, Perez-Quiros and Rots, 2020, Antolin-Diaz, Drechsel and Petrella, 2021, Carreiro at al., 2021, and Ng, 2021, have recently investigated some ways to handle the unique features of the COVID-19 recession in time-series forecasting with parametric approaches.

Markov-switching model estimated with data only up to 2019. On a monthly basis, the recession probability index maintained by the St. Louis Fed and documented by Chauvet and Piger (2008), is computed from a Markov-switching model that allows a change in model parameters associated with the period from March 2020 to July 2020.

The aim of this paper is to introduce a new approach to compute h-step ahead predictive probabilities of future recessions that does not require manipulating the sample of the database because the method is robust to influential points and other data irregularities, such as structural breaks, heteroskedasticity and Autoregressive Conditional Heteroskedasticity (ARCH). Specifically, we propose a nonparametric extension of Wecker's method that generates h-period ahead forecast paths at time t by adding to the time series under consideration at time t the past increments occurred in the set of $(h+1)$-dimensional blocks that can be extracted from the past of the time series. Then, the forecast paths can be used to compute the relative frequency of technical recessions.

As we will show below, this strategy implies assuming equal weigths for all the past increments, which seems unreliable in practice. For example, past blocks of the time series referring to recovery periods exhibit upward trends that would hardly ever occur when time t refers to a downturn. To overcome this drawback, we embed all the forecast paths into a symbolic space and derive the expressions required to compute the probability of occurrence of each of the symbols at the time of the forecast. Thus, we compute the relative frequency of a technical recession across the forecast paths by weighting each path differently according to the probability of the corresponding symbol occurring at time t.

The advantage of forecasting recession probabilities in this way compared to linear and nonlinear parametric approaches is twofold. Firstly, the method is nonparametric so there is no need to make any assumption in a specific dynamic model for the given population. In addition, performing of recession probabilities to provide accurate business inferences does not depend on the estimates of the model parameters, which tend to be unstable under structural breaks and large outliers. Secondly, the impact of extreme values appearing in some past blocks of the time series, like those observed in the pandemic period, are expected to average out when computing the weighted relative frequency of the technical recessions and their impact on forecasting performance becomes negligible.

By conducting several Monte Carlo experiments designed to capture standard data problems that characterize economic data sets, we evaluate the finite-sample performance of the proposed algorithm to predict recession probabilities. To assess its forecasting performance, we analyze Receiver Operating Characteristic (ROC) curves, Brier Scores, and Cohen's Kappa coefficients. In absence of data problems, we use these statistics in an out-of-sample forecasting scenario to show that the nonparametric approach developed in this paper behaves similarly to Markov-switching and Wecker's approaches in one-period forecasting but its relative improvement over parametric approaches consistently increases with the forecasting horizon. In the presence of influential observations and structural breaks, and when the errors are heteroskedastic or present ARCH dynamics, the nonparametric approach outperforms the parametric models.

Finally, we evaluate the ability of our nonparametric method to compute accurate in-sample

forecasts of recession probabilities of future recessions (captured by NBER and ECRI recession dates) from national GDP growth rates in the G7 countries. Using pre-pandemic data, the Markov switching approach slightly outperforms the other approaches at one-period forecasting, although there are no sizeable differences with the nonparametric proposal as the forecasting horizon increases. Undoubtedly, the best-performing model is the nonparametric approach when the extreme values of GDP growth rates observed in 2020 are included in the sample because its forecasts of recession probabilities are barely affected by these influential observations.

The paper is organized as follows. Section 2 introduces the nonparametric approach to compute forecasts of recession probabilities. Section 3 shows the results of the Monte Carlo simulations. Section 4 applies the models to forecast recession probabilities in the G7 countries from quarterly GDP growth rate data. This section highlights the estimation problems faced by parametric approaches when the extreme figures observed in 2020 are included in the data set. Section 5 concludes and outlines some further research lines.

2 Robust probabilistic recession statements

Based on a novel extension of Wecker's (1979) proposal, this section describes a new procedure to compute h-period forecasts of recession probabilities. For clarity, we first describe the linear approach and then present our nonparametric extension.

2.1 The linear approach

The approach proposed by Wecker (1979) offers a heuristic solution to compute business cycle inferences with linear autoregressive models based on Monte-Carlo simulations. Let $\{y_1, \ldots, y_t\}$ be the observed values of a stationary and ergodic time series, y_t, and let $\{\hat{y}_{t+1}, \ldots, \hat{y}_{t+h}\}$ be the predictions of its uncertain future values $\{y_{t+1}, \ldots, y_{t+h}\}$.

To adapt the method to our context, let y_t be the growth rates of the seasonally adjusted real GDP series for a given country. We rely on the popular definition of a technical recession to state the occurrence of a recession, which requires the GDP to fall for at least two consecutive quarters.[2] To obtain a probabilistic statement about the event of a recession, we define z_t as the sequence of indicator variables that indicate a recession at t, whose outcomes rely on the time series y_t according to the rule

$$z_t(y_{t-1}, y_t) = \begin{cases} 1 & \text{if } y_{t-1} < 0 \text{ and } y_t < 0 \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

Thus, the rule identifies a recession after two (or more) successive declines.

To estimate the h-step-ahead forecast of the probability of a recession, the vector of present and future values of the time series $Y_h(t+1) = \{y_t, y_{t+1}, \ldots, y_{t+h-1}, y_{t+h}\}$, for $h \geq 1$, must be estimated. Assuming that the data-generating process is a univariate autoregressive Gaussian model, sample paths of the future values of the time series can be repeatedly generated. Concretely, one can draw

[2] Among others, this definition has been used by Hamilton and Perez-Quiros (1996) and by Camacho and Perez-Quiros (2002) to compute business cycle inferences from linear autoregressive models.

a number M of vectors of forecasts $\{(\hat{y}^m_{t+1}, \ldots, \hat{y}^m_{t+h})\}^M_{m=1}$ from $N(\mu_t, Q)$, where explicit forms of the mean and covariance matrix for different forecasting horizons are derived in Appendix A. This leads to $(h+1)$-dimensional forecast paths

$$\hat{Y}^m_{h+1}(t) = (y_t, \hat{y}^m_{t+1}, \ldots, \hat{y}^m_{t+h}), \tag{2}$$

where $m = 1, \ldots, M$ and we set $\hat{y}^m_t = y_t$.

Then, using the rule of two consecutive periods of declining stated in (1), we can compute M realizations of the indicator variable $\mathcal{Z}^M_{t+h} = \{z_{t+h}(\hat{y}^m_{t+h-1}, \hat{y}^m_{t+h})\}^M_{m=1}$. Notice that, since the distribution of future values of the time series conditioned to its past values, $g_y(y_{t+1}, \ldots, y_{t+h}|y_1, \ldots, y_t)$, can be approximated by the empirical distribution of the generated forecast paths $\hat{Y}^m_{h+1}(t)$, the distribution of z_{t+h} can also be approximated by the empirical distribution of $\mathcal{Z}^M_{t+h}$, as stated in Wecker (1979). Therefore, the sample mean of this empirical distribution is taken to be a forecast of the probability that the economy will be in recession at date $t+h$

$$P_L(z_{t+h} = 1) = \frac{1}{M}\sum_{m=1}^{M} z_{t+h}(\hat{y}^m_{t+h-1}, \hat{y}^m_{t+h}), \tag{3}$$

for any look-ahead horizon h.

From this expression, it becomes clear that the ability of P_L to detect future recessions depends crucially on the performance of the forecasting model used to compute reliable forecasts of $(y_{t+1}, \ldots, y_{t+h})$. In this context, it is evident that model misspecification, structural breaks, or extreme values inducing instability in the autoregressive parameters will negatively impact the performance of P_L.

2.2 Nonparametric forecasts

As in the linear approach, our proposal to infer whether an economy will be in recession with forecast horizon h, requires estimating $Y_{h+1}(t) = \{y_t, y_{t+1}, \ldots, y_{t+h-1}, y_{t+h}\}$. However, instead of generating forecast paths from parametric autoregressive models, we rely on simulating nonparametric forecasts by embedding the time series $\{y_t\}^T_{t=1}$ in a $(h+1)$-dimensional space by computing the histories

$$Y_{h+1}(\tau) = (y_\tau, y_{\tau+1}, \ldots, y_{\tau+h-1}, y_{\tau+h}), \tag{4}$$

where $\tau = 1, \ldots, T-h$. Each of these vectors summarizes the behavior of the time series in the neighborhood of τ, accounting for the value of the stationary time series at τ and the subsequent steps $\tau+1, \ldots, \tau+h$.

For each τ, we use $Y_{h+1}(\tau)$ to generate realizations of the forecast of $Y_{h+1}(t)$ as follows:

$$\hat{Y}^\tau_{h+1}(t) = (y_t, \hat{y}^\tau_{t+1}, \ldots, \hat{y}^\tau_{t+h}), \tag{5}$$

where $\hat{y}^\tau_{t+k} = y_t + (y_{\tau+k} - y_\tau)$ is the τ-th generation of the forecast y_{t+k}, for $k = 1, 2, \ldots, h$ and $\tau = 1, 2, \ldots, t-h$. In this proposal, the τ-th forecast of y_{t+k}, given by $\hat{y}^\tau_{t+k}$, is the value of y_t plus the increment produced in the time series in the following k step starting at a given period τ.

Apart from it being stationary and ergodic, we do not require further assumptions about the data-generating process of y_t, its population probability distribution. In this case, it is straightforward to show that $E(\hat{y}^{\tau}_{t+k}) = E(y_t)$. In addition, the increments in the time series, $y_{t+k} - y_t$ as well as $y_{\tau+k} - y_{\tau}$, are stationary because the time series is also stationary, which implies that they are realizations of the same distribution distribution that depends on k but not on t. This supports our approach as a natural way to perform the different forecasts.

Using the $t-h$ forecast paths $\hat{Y}^{\tau}_{h+1}(t)$, we can generate a sequence of indicators of a technical recession $z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h})$ for $\tau = 1, \ldots, t-h$, whose empirical distribution approximates the distribution of z_{t+h}. Thus, as a natural extension of the linear approach, the probability of a recession at $t+h$ can be estimated by

$$P(z_{t+h} = 1) = \frac{1}{t-h} \sum_{\tau=1}^{t-h} z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h}), \tag{6}$$

which is the frequency of a technical recession across the nonparametric simulations of the forecast path.

It is worth emphasizing that this extension of the linear approach to producing probabilistic statements for the occurrence of recessions could lead to invalid inferences. Notice that in the linear autoregressive case, the forecasts of the variable of interest $y_{t+1}, ..., y_{t+h}$, account for the inertia of the time series since they are computed using linear combinations of past values of the variable, with the recent past having the highest weights.

By contrast, our nonparametric projections do not consider the typical business-cycle inertia of GDP growth rates when developing the short-term forecasts at t. In fact, expression (6) computes the probability of a recession at $t+h$ by averaging the $t-h$ estimates of the indicator variable z, which assigns equal weights to all the indicators of recession $z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h})$, regardless of the neighborhoods of the time series at t and τ. Thus, when the economy is in the middle of an expansion at t, the method would assign the same weight to $\hat{Y}^{\tau}_{h+1}(t)$ regardless of whether τ refers to an expansionary or to a recessionary period.

The following section proposes a modification of (6) that overcomes this drawback using a weighting algorithm based on symbolic dynamics.

2.3 Symbolic dynamics based weights

Symbolic dynamics involves the simple process of labeling each of the $(h+1)$-history $Y_{h+1}(t) = (y_t, y_{t+1}, \ldots, y_{t+h})$, for $t = 1, \ldots, T-h$, with a symbol. Thus, instead of following the trajectory of the time series point by point, one only keeps recording the alternation of the symbols. According to Collet and Eckmann (2009), the evolution of the symbols can capture the complete description of the dynamic system[3].

The proposed symbolization is as follows. Let S_{h+1} be the symmetric group of order $(h+1)!$, that is the group formed by all the permutations of length $h+1$ of the elements in the set $\{0, 1, 2, \ldots, h\}$. An element of this group, namely $\pi = (i_0, i_1, ..., i_h) \in S_{h+1}$, is called a symbol. The

[3]Some other applications of symbolic dynamics in economics are Tino et al. (2000), Matilla, Ruiz, and Dore (2014), Hou et al. (2017) and Camacho, Romeu, and Ruiz (2021).

symbolization procedure consists on assigning a unique permutation that sorts out its entries from the smallest to the largest to any $(h+1)$-tuple $Y_{h+1}(t)$. Formally, this procedure maps $Y_{h+1}(t)$ to the unique permutation $\pi = (i_0, i_1, ..., i_h) \in S_{h+1}$, satisfying the following two conditions:

$$y_{t+i_0} \leq y_{t+i_1} \leq ... \leq y_{t+i_h}, \tag{7}$$

$$i_{s-1} < i_s \; if \; y_{t+i_{s-1}} = y_{t+i_s}. \tag{8}$$

The first condition imposes an ordinal pattern and the second is a technical condition that guarantees the uniqueness of the symbol in the case of equal values, which theoretically has zero probability of occurring in the case of continuous distributions. Thus, symbolic dynamics converts the sequences of $(h+1)$-histories $\{Y_{h+1}(t)\}_{t=1}^{T-h}$ into sequences of ordinal patterns labeled with symbols $\{\pi(t)\}_{t=1}^{T-h}$.

As a quick example, consider the time series $\{y_t\} = \{5, 3, 2, 1, 8, 9, 3, 4, 5, 2\}$ of length $T = 10$. For a forecasting horizon $h = 2$, we can obtain eight 3-histories

$$\{Y_3(t)\}_{t=1}^{8} = \{(5,3,2), (3,2,1), (2,1,8), (1,8,9), (8,9,3), (9,3,4), (3,4,5), (4,5,2)\}. \tag{9}$$

Using the integers $\{0, 1, 2\}$ to construct the symbols, the set of potential symbols is

$$S_3 = \{(0,1,2), (0,2,1), (1,0,2), (1,2,0), (2,0,1), (2,1,0)\}. \tag{10}$$

Now, symbolic dynamics yields the symbolized series as

$$\{\pi(t)\}_{t=1}^{8} = \{(2,1,0), (2,1,0), (1,0,2), (0,1,2), (1,2,0), (2,0,1), (0,1,2), (1,2,0)\}. \tag{11}$$

It is worth emphasizing that symbol (0,1,2) will be associated with increasing patterns (expansions) in the time series, while symbol (2,1,0) will typically refer to decreasing dynamics (recessions).

In addition, the probability at t of symbol $\pi \in S_{h+1}$, which we call P_π^t, can be computed analytically. Based on an extension of Abd Alla (2004), Appendix B shows the expressions of these forecasts for $h = 1, 2, 3$. These forecasts of symbol probabilities can be used to improve the accuracy of the recession probability forecasts stated in (6).

To this end, let τ be any time period smaller or equal to $t-h$, and denote with $\pi(\tau)$ the symbol associated with $(h+1)$-history $Y_{h+1}(\tau)$. Let $P_{\pi(\tau)}^t$ be the probability that symbol $\pi(\tau)$ appears at t.[4] We propose the nonparametric forecast at t of a recession occurring at $t+h$ as

$$P_{NP}(z_{t+h} = 1) = \frac{\sum_{\tau=1}^{t-h} z_{t+h}(\hat{y}_{t+h-1}^{\tau}, \hat{y}_{t+h}^{\tau}) P_{\pi(\tau)}^t}{\sum_{\tau=1}^{t-h} P_{\pi(\tau)}^t}. \tag{12}$$

This expression implies that the contribution to $P_{NP}(z_{t+h} = 1)$ of the τ-th recession indicator is weighted by the probability that the ordinal pattern of $(h+1)$-history $Y_{h+1}(\tau) = (y_\tau, y_{\tau+1}, \ldots, y_{\tau+h})$

[4]It implies that $P_{\pi(\tau)}^t = P_\pi^t$ if $\pi(\tau) = \pi$, for all $\pi \in S_{h+1}$.

(i. e. symbol $\pi(\tau)$) would occur at t, where the weights are normalized to add up to one.[5]

It is easy to check that forecasting probabilities of recession with the weighted average of the recession indicators as in (12) overcomes the drawback of (6). Let us denote with π_E those symbols associated with expansions and with π_R those associated with recessions. If we assume that the economy is in the middle of an expansion at t, then probability $P^t_{\pi_E}$ should be much higher than probability $P^t_{\pi_R}$. Therefore, if period τ is in the middle of a recession, the corresponding $(h+1)$-history, $Y_{h+1}(\tau)$, will be associated with $\pi(\tau) = \pi_R$ and the weight of the recessionary indicator $z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h})$ appearing in the numerator of (12) will be as low as $P^t_{\pi_R}$.

3 Monte Carlo simulation

In this section, we set up several Monte Carlo experiments to assess the finite-sample performance of our nonparametric proposal to compute h-step ahead predictions of recession probabilities, $P_{NP}(z_{t+h} = 1)$, for $h = 1, 2, 3$. In addition, we use the simulations to evaluate how data problems, such as influential points, structural breaks, heteroskedasticity, and ARCH effects, might affect forecast performance. To facilitate comparisons with Wecker's (1979) linear approach, we also include the forecasts of recession probabilities using an autoregressive model, $P_L(z_{t+h} = 1)$.

For the sake of comparison, we have included the forecasts of a Markov-switching autoregressive model of order q, MSAR(q), which is one of the most popular approaches used to compute recession probabilities. Following Hamilton (1989), we assume that the dynamics of y_t are governed by an unobservable regime-switching state variable, s_t. The model can be stated as

$$y_t = \mu_{s_t} + a_1(y_{t-1} - \mu_{s_{t-1}}) + \ldots + a_q(y_{t-q} - \mu_{s_{t-q}}) + \epsilon_t, \tag{13}$$

where $\epsilon_t \sim iidN(0, \sigma^2)$.[6]

Within this framework and assuming that $\mu_0 > \mu_1$, one can label $s_t = 0$ and $s_t = 1$ as the expansion and recession states at time t, respectively. In addition, it is commonly supposed that the state variable evolves following an irreducible 2-state Markov chain whose transition probabilities are defined by

$$p(s_t = j | s_{t-1} = i, s_{t-2} = h, ..., I_{t-1}) = p(s_t = j | s_{t-1} = i) = p_{ij}, \tag{14}$$

where $i, j = 0, 1$ and $I_t = \{y_1, \ldots, y_t\}$ is the information set up to period t.

Hamilton (1989) described a forward filter to store the filtered probabilities of recession $P(s_t = 1|\theta, I_t)$, where $\theta = (\mu_0, \mu_1, a_1, \ldots, a_q, p_{00}, p_{11})$ and to provide a maximum likelihood estimation of model parameters $\hat{\theta}$. Using the parameter estimates, it is easy to compute inferences about the

[5] It is worth pointing out that, since $P_{NP}(z_{t+h} = 1) = E(z_{t+h})$, expression (3) is an unbiased estimator of the probability of recession.

[6] The dynamics can be adapted to account for regime shifts in the autoregressive parameters and in the variance. In addition, the nonlinearities can be imposed in the mean or in the drift of the time series.

business cycle regime at $t+1$ as

$$P(s_{t+1}=1|\hat{\theta},I_t)=\sum_{i=0}^{1}P(s_t=i|\hat{\theta},I_t)p_{i1}. \tag{15}$$

This expression can be used recursively to obtain h-period ahead forecasts $P(s_{t+h}=1|\hat{\theta},I_t)$.

We quantify the ability of the Wecker (1979) method, the Markov-switching model, and our nonparametric approach to forecasting the h-step ahead state of the business cycle with the help of three different metrics. The first metric is the Brier score, BS, which is the mean square error of recession probability. A Brier score of 0 means perfect accuracy, and a Brier score of 1 means perfect inaccuracy. For more detailed information this metric is also computed only for recessions (BSR) and expansions (BSE).

The second metric is the area under the receiver operating characteristic curve, namely $AUROC$ (Berje and Jorda, 2011), which is a measure of the overall performance of a binary classifier, it considers the trade-off between the true positive rate and the false positive rate. The $AUROC$ takes values between 0.5 for a random classifier and 1 for a perfect classifier. Regarding this metric, we define the True Posititive rate (TPR) as the probability of predicting recessions that actually become recessions and the True Negative rate (TNR) as the probability of predicting expansions that actually become expansions.

The third metric is Cohen's Kappa coefficient (Cohen, 1960), which is a chance-corrected measure of agreement between the classification of the forecasting techniques and actual recessions. This metric, denoted by κ, reaches 1 for complete agreement. A more detailed explanation of these metrics can be found in Appendix C.

3.1 Nonparametric model performance

We start the simulations by generating $r=1,\ldots,R=500$ business cycle sequences s_t^r of expansions ($s_t^r=0$) and recessions ($s_t^r=1$) of length $T=500$ that follow 2-state Markov chains. To ensure that these dummies share the standard business cycle dynamics, we use the NBER dates to compute the percentage of quarters classified as expansions followed by expansions and the percentage of quarters classified as recessions followed by recessions in the period 1955.2-2022.3. According to this analysis, we set $p_{00}=0.9$ and $p_{11}=0.6$.

The data-generating process is an MSAR(1) as outlined in (13) with noisy terms $\epsilon_t^2\sim iidN(0,0.5)$. Using s_t^r, the dynamics of time series y_t^r are generated by setting the differences of the within-state means $\mu_0-\mu_1=0.5$ and a state-independent autoregressive parameter $a_1=0.2$. With this data-generating process, we examine forecasting performance in an out-of-sample scenario.

The results of this exercise for the nonparametric approach are displayed in Table 1. The baseline simulations, whose results are presented in Panel A, show that the accuracy of the probability forecasts deteriorates very little as the forecasting horizon increases, as documented by the Brier score. Similarly, the $AUROC$ metric is around 0.8 regardless of the forecasting horizon, indicating that our nonparametric method presents a good discriminating ability to distinguish the state of the generated business cycles. Finally, Cohen's kappa suggests a reasonable level of agreement

between the one-period forecasts of the nonparametric method and the generated business cycles. Nevertheless, κ tends to diminish slightly as the forecasting horizon increases.

Panel B of Table 1 evaluates the effect of the persistence of the generated time series, measured by the autoregressive parameter a_1, on the nonparametric probability forecasts. The results indicate that the forecasting performance is somehow worsened by higher persistence, mainly through BSR. However, the deterioration of the forecasting performance is not dramatic since the persistence does not change this metric substantially when the autoregressive parameter rises from $a_1 = 0.2$ to $a_1 = 0.5$ or $a_1 = 0.8$.

To examine the effect of the sample size in the nonparametric approach, Panel C displays the estimates of the forecasting performance metrics for data-generating processes of sample sizes $T = 250$ and $T = 1000$. In these two cases, the figures are similar to those obtained in the baseline scenario, indicating that the model's performance is invariant to time series of reasonable sample sizes.

We also study the effects of uncertainty on forecasting performance by generating noisy terms with variances of $\sigma^2 = 1$ and $\sigma^2 = 1.5$, shown in Panel D. As expected, the ability of the model to compute inferences on the generated business cycle deteriorates substantially when variance σ^2 increases. Thus, we find that noisy scenarios deteriorate the ability of the nonparametric procedure to classify the periods into recessions and expansions.

In addition, we evaluate the effects of business cycle phase persistence on forecasting performance by performing simulations with combinations (p_{00}, p_{11}) of $(0.6, 0.6)$ and $(0.9, 0.9)$. The results are shown in Panel E. When a business cycle phase becomes more persistent, the Brier score falls, whereas the overall performance of the binary classifier ($AUROC$) and the overall agreement (κ_h) of the forecasts and the generated cycles increase.

Finally, to examine how the business cycle signal affects forecasting performance, we also set the within-state difference $\mu_0 - \mu_1$ to 1 and 2. Panel F shows that larger differences of within-state means substantially improve the performance of the nonparametric model as the signal-to-noise of the data-generating process increases. Thus, large differences between within-state means facilitate the classification of the time periods into recessions and expansions.

3.2 Comparison of the methods' performance under data problems

Despite the good performance of the model in providing statistical inferences of the generated business cycles in the baseline scenario, the data problems that characterize the economic dynamics in empirical applications could lead to potential performance deterioration. To evaluate these potential adverse effects, we conduct an out-of-sample forecasting exercise with outliers, structural breaks, heteroskedasticity, and ARCH dynamics.

In addition, we are interested in evaluating the differences in performance deterioration that the data problems may cause in Markov-switching, linear and nonparametric specifications to forecast recession probabilities. For this purpose, Panels A, B, and C of Table 2 display the Brier scores, $AUROC$, and Kappa coefficients achieved by these three alternative forecasting proposals. To facilitate comparisons, the first row in each panel reports the results of the baseline scenario. The table shows that the three models perform well forecasting business cycle phases. For $h = 1$,

the three models display similarly low BS, and high $AUROC$ and κ. Notably, the forecasts that deteriorate the least with the prediction horizon are those of the nonparametric model. Regardless of the statistic, the nonparametric model has the greatest classification ability for $h = 2$ and $h = 3$.

To assess the performance deterioration caused by extremely large observations, we generate additive outliers in the simulated time series that are consistent with the magnitude of the large GDP growth rates observed in 2020. Specifically, we add an additive outlier of -15 standard deviations from the mean of the simulated time series at $t = 100$, followed by a 10 standard deviations outlier in $t+1$ to the baseline data-generating process. In line with the results reported in Table 2, of the three models, the one that fails the most in performing business cycle inferences due to the large number of observations is the Markov-switching model. Intuition indicates that the estimated mean in the low-growth regime is dominated by the first outlier and the sample after this date is classified as expansion regardless of the value of the time series. For this reason, BSR tends to 1 while TPR and $AUROC$ are close to 0 and 0.5, respectively.

For one-period forecasting horizons, the performance deterioration of the probability forecasts computed with the linear model is not as significant as that of the Markov-switching model. However, the deterioration is much greater for forecasting horizons larger than 1, as demonstrated by the low values of $AUROC$, which fall to about 0.5, indicating no better classification ability than the toss of a coin strategy. Performance deterioration in forecasting recessions is also evident in the large value of BSR, which is 0.72, and the low value of κ, which tends to 0.

Remarkably, the forecasts computed with the nonparametric algorithm do not show any significant deterioration in the metrics used to examine the performance of the h-step ahead forecasts of the business cycle, regardless of the forecasting horizon considered. As Panel C of Table 2 shows, the metrics are almost invariant when the outlier is introduced into the data-generating process and BS, $AUROC$, and κ are roughly similar to the baseline scenario.

The second robustness check focuses on structural breaks. To examine this data problem, we generate the last three-fourths of the sample as in the baseline scenario ($\mu_0 = 0.25$) while we set $\mu_0 = 1$ for the first fourth of the sample, with the rest of the parameters unaltered. The third row of each panel of Table 2 shows that, as in the case of the outlier, the structural break produces substantial deterioration in the performance of the Markov-switching and linear specifications. However, the changes in the performance of the nonparametric proposal are, again, negligible.

Finally, we evaluate the models' performance to static and serially correlated heteroskedasticity. To simulate data with static heteroskedasticity, we generate the last three-fourths of the sample as in the baseline scenario ($\sigma^2 = 0.5$) while we set $\sigma^2 = 2.5$ for the first fourth of the sample. Serially correlated heteroskedasticity is achieved by generating disturbances $\epsilon_t = \sigma_t u_t$, where u_t is a normalized independent Gaussian process and $\sigma_t = 0.2 + 0.8\epsilon_{t-1}$. Again, the performance deterioration of Markov-switching and linear autoregressive forecasts is substantial, although a bit less severe than in the cases of outliers and structural breaks. As in the previous two scenarios, the deterioration in the performance of the linear and Markov-switching approaches is much greater than in our nonparametric proposal.

4 Empirical example

In this section, we assess the empirical reliability of Markov-switching specifications, linear autoregressive models, and our nonparametric approach to provide accurate in-sample forecasts in the run-up to business cycle recessions in the G7 countries. In addition, we examine the impact of the COVID-19 recession, which led to record falls and recoveries, on forecasting performance. For this purpose, we develop the analysis with a sample that ends in 2019 and a complete sample that includes the COVID-19 pandemic data ending in the third quarter of 2022.

The COVID-19 recession is the latest in our data sets. Thus, we are precluded from examining its effect in out-of-sample exercises as in the section devoted to Monte Carlo simulations. To assess the effect that the observations recorded in 2020 will have on the future performance of the forecasting approaches, we focus on the impact of these influential observations on historical business cycle dating. Notice that we do not pursue ad-hoc shortcuts as temporary solutions to this problem, such as using additional regimes or shortening the sample used to estimate model parameters.

We use data sets of the growth rates of seasonally adjusted real GDP for the Group of Seven (G7) countries (USA, UK, Germany, France, Italy, Canada, and Japan). The data come from the OECD Main Economic Indicators. The sample starts in 1955.2 for the UK and US, 1960.2 for Italy, Canada, and Japan, 1964.1 for Germany, and 1969.1 for France. It ends in 2022.3 for all the countries. To evaluate the empirical performance of the forecasting approaches, we use the reference cycle dates provided by the dating committees of the NBER in the case of the US and the Economic Cycle Research Institute (ECRI) for the rest of the countries.

4.1 Pre-COVID-19 data

In the first approximation to the analysis of forecasting performance, we focus on the historical ability of the three forecasting approaches to classify the dates into expansion and recession in constrained samples that end in the last quarter of 2019. Remarkably, the results reported in Panel C of Table 3 show that the probability forecast of our nonparametric procedure is in close agreement with the reference cycles for all the countries. Regardless of the forecasting horizon, our approach results in a low Brier score, an $AUROC$ much higher than 0.5, and a large kappa coefficient, all of which are comparable to those reached by the Markov-switching model (Panel A) and the linear specification (Panel B).

In terms of ROC curves, the Markov-switching model achieves the best business cycle performance across all the possible classification thresholds for all the countries but France, Germany, and Italy, in part because their higher volatilities tend to diminish the ability of the Markov-switching model to separate the states.[7] Although this approach shows high values for the US, UK, and Canada, the magnitudes reported for the linear and the nonparametric approaches are also significantly greater than 0.5, indicating their considerable discriminating ability. The apparently better performance of the Markov-switching approach relies on its ability to classify expansions because this advantage vanishes when examining the ability to correctly identify recessions with

[7] This result agrees with those of Carstensen et al.(2020).

TPR statistics.

Measured by the Brier score, the differences in forecasting performance across the three competing approaches are minor in all the cases. The nonparametric model slightly outperforms in one and two-quarter forecasts, and the linear approach outperforms in three-quarter forecasts. Regarding the kappa index, the data reveal uniformly closer agreement between the nonparametric probabilities of recession the official recessions than with the other two forecasting approaches. The numbers reported in Table 3 show that the kappa coefficients of the nonparametric approach are substantially larger than those of the other two competitors, regardless of the forecasting horizon and the country of the sample.

To illustrate the good performance of the three dating processes with samples that do not include the pandemic period, Figure 1 displays the growth rates of quarterly real GDP for the US from 1955.2 to 2019.4 (Panel A) and the 2-quarter ahead predictions of the probabilities of recession obtained from a Markov-switching model (Panel B), a linear specification (Panel C) and our nonparametric approach (Panel D), which are obtained using full-sample parameter estimates. To facilitate comparisons, the panels include the dates of economic recessions as determined by the NBER, which are shaded.

The Markov-switching model estimates within-expansion and within-recession means of $\hat{\mu}_0 = 0.91$ and $\hat{\mu}_1 = -0.56$, respectively. Panel B of Figure 1 plots the two-horizon forecast recession probabilities $P(s_{t+2} = 1|\hat{\theta}, I_t)$, with $\hat{\theta}$ estimated using data up to 2019.4. These probabilities produce satisfactory data classification into expansions and recessions, reproducing the NBER chronology very closely. During periods that the NBER classifies as expansions, the probabilities of recession are usually close to zero. At the NBER peaks, the probabilities rise above 30% and remain at these levels until the NBER troughs.

Moving to Wecker's approach, we fit an AR(2) model to the GDP growth rates, whose parameters are also estimated using data up to 2019.4. This specification produces estimates of the autoregressive parameters of $\hat{a}_1 = 0.26$ and $\hat{a}_2 = 0.13$, respectively, which are statistically significant. At each period, we use (3) to compute the 2-period forecasts of recession probability, $P_L(z_{t+2} = 1)$, which are plotted in Panel C of Figure 1. The predicted probabilities of recession also align well with the official NBER business cycle, although to a lesser degree than the Markov-switching specification. Generally, when the probability of recession exceeds 40%, a recession follows, while recession probabilities fall below 40% in recession troughs.

Finally, Panel D of Figure 1 plots the 2-quarter forecasts of recession probabilities computed as expression (12). Despite the simplicity of computing the forecasts, the figure shows that the new, nonparametric assessment of recession probabilities is proficient at capturing the NBER-referenced business cycle chronology. Specifically, the forecasts of recession probabilities always jump to almost one at the point of NBER recessions (shown in shaded areas) and remain much lower at NBER expansions.

4.2 Complete data set

Despite the good performance of the three alternative methods when the samples end in 2019, the COVID-19 pandemic in 2020 hit the national economies worldwide with unprecedented force

causing record fluctuations in GDP growth that have substantially altered the findings obtained with pre-pandemic data. Particularly, the G7 countries faced one of their sharpest declines during the health restrictions established in the first quarter of 2020, followed by exceptionally rapid rebounds when the restrictions were relaxed and stimulus measures came into effect.

In Table 4, we evaluate the impact of the extraordinaryly high growth rates reported in 2020 on forecasting performance by extending the sample to the third quarter of 2022. Regardless of the country, the table shows that the Markov-switching specification has lost its business cycle classification abilities. For each country, the model identifies only one recession in 2020 while classifying the rest of the sample dates as expansions. This implies that BSR tends to be one, and BSE tends to be zero. For this model, the $AUROC$ metrics are close to 0.5, suggesting no better performance than a random classifier. In addition, the kappa coefficients are always close to zero.

The deterioration suffered in the linear autoregressive forecasts is a bit less severe than in the Markov-switching approach. However, we find that BSR increases substantially and the TPR falls dramatically, as does κ_h, indicating that this model is not very useful for anticipating future recessions. In addition, Table 4 reveals that the losses in forecasting accuracy when pandemic data is included increase as the forecasting horizon expands.

In contrast to the Markov-switching model and the linear autoregressive specification, Table 4 shows that our nonparametric approach is scarcely affected by the 2020 data. Regardless of the country, when the extreme data are included, the model still agrees closely with the reference cycles, has good overall performance as a binary classifier, and is consistent with the turning points determined by the national dating committees.

To illustrate this graphically, we enlarge the sample in Figure 1 with US GDP growth rates observed up to 2022.3. Panel A of Figure 2 shows that the US suffered from the deepest recession and the most significant recovery on record in two consecutive quarters (-8.9% and 7.5% in 2020.2 and 2020.3, respectively). These two extreme observations blur the patterns of the time series dynamics observed in the pre-pandemic data. A straightforward explanation is that the two influential observations drastically skewed the empirical distribution of the GDP.

In the case of the Markov-switching specification, the two influential observations in 2020 affect the estimate of the within-recession mean dramatically. It falls to $\hat{\mu}_1 = -9.07$, leaving the within-recession mean almost unaltered at $\hat{\mu}_0 = 0.77$. The resulting 2-period ahead forecasts of recession probability, displayed in Panel B of Figure 2 show that the cyclical interpretation of high and low growth states fails spectacularly. The substantial drop in GDP growth documented in 2020.2, which is identified as a low-growth state, is so influential that the model relegates all the previous recessions to a high-growth state. This invalidates the standard Markov-switching approach as a tool for anticipating future recessions unless we use ad-hoc shortcuts such as allowing changes in model parameters in 2020 or estimating model parameters using data only up to 2019.

Regarding Wecker's technique, it is well-known that extreme observations greatly affect the forecasting performance of linear autoregressive models and tend to bias the full-sample parameter estimates. Fitting an AR(2) to the enlarged sample of growth rates produces autoregressive parameters that fall dramatically to $\hat{a}_1 = 0.03$ and $\hat{a}_2 = 0.08$ and become statistically non-significant. We use these estimates to simulate the forecast paths and the recessionary indicator required to

compute the two-period forecasts of recession probabilities. Panel C of Figure 2, which displays these probabilities, shows that Wecker's technique also fails to predict the NBER-referenced recessions when the two extreme observations released in 2020 are used to compute the model-based forecasts.

Notably, Panel D of Figure 2 shows that enlarging the sample with data up to 2022.3 does not deteriorate the performance of the new nonparametric approach. In this case, estimating the model parameters from samples that use the extreme observations reported in 2020 leads to recession probabilities that continue to be remarkably similar to NBER business cycle dating. This shows that, unlike the cases of autoregressive models and Markov-switching specifications, our proposal minimizes the importance of extreme observations in the time series used to perform the forecasts of recession probabilities.

To sum up, the two influential observations of GDP growth rates in 2020 have (and will have in the future) devastating effects on business cycle identification from standard parametric models and calls into question whether they will be useful for dating business cycles from the COVID-19 period on. We show that our nonparametric proposal is robust to this and other data-generating problems, such as structural breaks and when errors are heteroscedastic or present ARCH dynamics.

5 Conclusions

Providing economic agents with early warning systems that give advanced notice of future business cycle developments has become an issue of great interest in economics. For this purpose, some parametric approaches, such as the Markov-switching model advocated by Hamilton (1989) or the autoregressive forecasts used by Wecker (1979), have become very popular methods to provide recession probabilities. However, the unprecedented magnitude of the recession caused by the recent global pandemic has shown that these tools fail to produce robust dating of business cycle turning points.

To contribute to this literature, this paper develops a nonparametric extension of the autoregressive forecast method, combined with symbolic dynamics, to compute robust inferences of reference cycle turning points. Our simulations suggest that the method performs well in recession predictions and that, unlike the parametric alternatives, it is robust against outliers, structural breaks, heteroscedasticity, and ARCH effects. This is desirable in early warning systems because, in practice, these data problems are the norm rather than the exception.

Using pre-COVID-19 data, the empirical evidence shows that the historical ability of the nonparametric approach to forecasting the business cycle phases of the G7 countries is similar to that of its parametric competitors. However, when the sample is enlarged with COVID-data, the nonparametric approach substantially outperforms the parametric forecasts, whose performance is hardly better than a random classifier.

We look forward to carrying out future work addressing the following issues. First, we see a natural extension of our approach to developing early warning systems for determined events in many other situations by defining the event under consideration differently. For example, further application areas are forecasting critical transitions in temperature regimes, detecting product de-

fects in manufacturing, or providing statistical classifications of diseases. Second, our empirical application relies on an in-sample approach because the post-COVID-19 data have only been available for roughly two years. As new vintages become available, real-time reassessments will provide new insights into the model's forecasting performance. Third, we see the possibility of adjusting the methodology to a multivariate approach, to take several series into account when predicting recessions.

References

[1] Abd Alla, F. 2004. Distributions of order patterns in time series. PhD dissertation, Greifswald.

[2] Anderson H., Teräsvirta T., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregression models. *Journal of Applied Econometrics* 7: S199-136.

[3] Antolin-Diaz, J., Drechsel, T., and Petrella, I. 2021. Advances in nowcasting economic activity: Secular trends, large shocks and new data. CEPR Discussion Papers 15926.

[4] Berge, T.J. and Jordà, O. 2011. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics* Vol 3, N0. 2, 246-277.

[5] Camacho, M., and Perez Quirós, G. 2002. This is what the leading indicator lead. *Journal of Applied Econometrics* 17: 61-80.

[6] Camacho, M., Romeu, A., and Ruiz, M. 2021. Symbolic transfer entropy test for causality in longitudinal data. *Economic Modelling* 94: 646-661.

[7] Carriero, A., Clark, T., Marcellino, M., and Mertens, E. 2021. Measuring uncertainty and its effects in the COVID-19 era. CEPR Discussion Papers 15965.

[8] Carstensen,K., Heinrich, M., Reif, M., and Wolters, M.H. 2020. Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model. *International Journal of Forecasting* 37 (3), 1328-1328.

[9] Chauvet, M., and Hamilton, J. 2006. Dating Business Cycle Turning Points. In Milas, C., Rothman, P., van Dijk, D., and Wildasin, E. (eds) *Nonlinear Time Series Analysis of Business Cycles*. Emerald Group Publishing Limited: Bingley.

[10] Chauvet, M., and Piger J. 2008. A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics* 26: 42-49.

[11] Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46.

[12] Collet, P., and Eckmann, J. 2009. Iterated maps on the interval as dynamical systems. Birkhauser, Basel.

[13] Dore, M., Matilla, M. and Ruiz, M. 2014. A permutation entropy based test for causality: The volume-stock price relation. *PHYSICA A* 398: 280-288.

[14] Estrella, A. and Mishkin, F. 1998. Predicting US recessions: financial variables as leading indicator. *Review of Economic and Statistics* 80: 45-61.

[15] Hamilton, J., and Perez-Quiros, G. 1996. What do the leading indicators lead? *Journal of Business* 69: 27-49.

[16] Hamilton, J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle *Econometrica* 57: 357-384.

[17] Hou, Y., Liu, F., Gao, J., Cheng, Ch., and Song, Ch. 2017. Characterizing complexity changes in Chinese stock markets by permutation entropy. *Entropy* 19: 514.

[18] Leiva-Leon, D., Perez-Quiros, G., and Rots, E. 2020. Real-time weakness of the global economy. European Central Bank Working Paper No 2381.

[19] McGrane, M. 2022. A Markov-switching model of the unemployment rate. Congressional Budget Office No. 2022-05.

[20] Ng, S. 2021. Modeling macroeconomic variations after COVID-19. NBER Working Papers No. 29060.

[21] Lenza, M., and Primiceri, G. 2020. How to estimate a VAR after March 2020. NBER Working Papers No. 27771.

[22] Tino P., Schittenkopf, Ch., Dorffner, G., and Dockner, E. 2000. A symbolic dynamics approach to volatility prediction. In Abu-Mostafa, Y., LeBaron, B., Lo, A. and Weigend, A. (eds) *Computational Finance*. MIT Press: Cambridge.

[23] Wecker, W. 1979. Predicting the turning points of a time series. *Journal of Business* 52: 35-50.

Appendix A

Let y_t be a time series that follows the following autoregressive process of order 2

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t, \tag{A.1}$$

where ϵ_t is a white noise Gaussian disturbance term with a mean of 0 and variance σ^2. For a one-step forecasting horizon $h = 1$, simulations of y_{t+1} can be developed from the distribution

$$y_{t+1} \sim N(c + a_1 y_t + a_2 y_{t-1}, \sigma_2). \tag{A.2}$$

When the forecasting horizon is $h = 2$, simulations of $(y_{t+2}, y_{t+1})'$ can be performed by

$$\begin{pmatrix} y_{t+2} \\ y_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} (1+a_1)c + (a_1^2 + a_2^2)y_t + a_1 a_2 y_{t-1} \\ c + a_1 y_t + a_2 y_{t-1} \end{pmatrix}, \sigma^2 \begin{pmatrix} (1+a_1^2) & a_1 \\ a_1 & 1 \end{pmatrix} \right). \tag{A.3}$$

Finally, when $h = 3$, simulations of $(y_{t+3}, y_{t+2}, y_{t+1})'$ can be obtained from

$$\begin{pmatrix} y_{t+3} \\ y_{t+2} \\ y_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} (1+a_1+a_2+a_1^2)c + (a_1^3 + 2a_1 a_2)y_t + (a_1^2 a_2 + a_2^2)y_{t-1} \\ (1+a_1)c + (a_1^2 + a_2^2)y_t + a_1 a_2 y_{t-1} \\ c + a_1 y_t + a_2 y_{t-1} \end{pmatrix}, Q \right), \tag{A.4}$$

where

$$Q = \sigma^2 \begin{pmatrix} (a_1^2 + a_2)^2 + a_1^2 + 1 & a_1(a_1^2 + a_2) + a_1 & a_1^2 + a_2 \\ a_1(a_1^2 + a_2) + a_1 & (1 + a_1^2) & a_1 \\ a_1^2 + a_2 & a_1 & 1 \end{pmatrix}. \tag{A.5}$$

Appendix B

Let $\{y_t\}_{t=1}^{T}$ be a stationary Gaussian process and ρ_i^t its autorrelation of order i computed with information up to t. In addition, let $h = 1, 2, 3$ be the forecasting horizon, let $Y_{h+1}(\tau)$ be the $(h+1)$-dimensional history of this time series starting at τ, and let $\pi \in S_{h+1}$ be its corresponding symbol. Finally, let P_π^t be the forecast probability of that symbol. For a forecasting horizon $h = 1$, the probability of the symbols that belong to S_2 are

$$P_{(0,1)}^t = P_{(1,0)}^t = \frac{1}{2}. \tag{B.1}$$

For a forecasting horizon $h = 2$, the probability of the symbols that belong to S_3 become

$$P_{(0,1,2)}^t = P_{(2,1,0)}^t = \frac{1}{\pi} arcsin(\frac{1}{2}\sqrt{\frac{1-\rho_2^t}{1-\rho_1^t}}) \tag{B.2}$$

and

$$P_{(0,2,1)}^t = P_{(1,2,0)}^t = P_{(2,0,1)}^t = P_{(1,0,2)}^t = \frac{1}{4}(1 - \frac{2}{\pi} arcsin(\frac{1}{2}\sqrt{\frac{1-\rho_2^t}{1-\rho_1^t}})) \tag{B.3}$$

Finally, for a forecasting horizon $h = 3$, the probability of the symbols that belong to S_4 are

$$P_{(0,1,2,3)}^t = P_{(3,2,1,0)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(arcsin(\alpha_1^t) + 2arcsin(\alpha_2^t))), \tag{B.4}$$

$$P_{(2,0,3,1)}^t = P_{(1,3,0,2)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(2arcsin(\alpha_3^t) + arcsin(\alpha_4^t))), \tag{B.5}$$

$$P_{(3,1,2,0)}^t = P_{(0,2,1,3)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(arcsin(\alpha_4^t) - 2arcsin(\alpha_5^t))). \tag{B.6}$$

$$P_{(1,0,3,2)}^t = P_{(2,3,0,1)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(2arcsin(\alpha_6^t) + arcsin(\alpha_1^t))), \tag{B.7}$$

$$P_{(0,1,3,2)}^t = P_{(1,0,2,3)}^t = P_{(2,3,1,0)}^t = P_{(3,2,0,1)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(arcsin(\alpha_7^t) - arcsin(\alpha_1^t) - arcsin(\alpha_5^t))), \tag{B.8}$$

$$P_{(2,0,1,3)}^t = P_{(0,2,3,1)}^t = P_{(3,1,0,2)}^t = P_{(1,3,2,0)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(arcsin(\alpha_7^t) - arcsin(\alpha_4^t) - arcsin(\alpha_5^t))), \tag{B.9}$$

$$P_{(0,3,1,2)}^t = P_{(3,0,2,1)}^t = P_{(2,1,3,0)}^t = P_{(1,2,0,3)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(arcsin(\alpha_3^t) + arcsin(\alpha_8^t) - arcsin(\alpha_5^t))), \tag{B.10}$$

and

$$P_{(0,3,2,1)}^t = P_{(3,0,1,2)}^t = P_{(1,2,3,0)}^t = P_{(2,1,0,3)}^t = \frac{1}{8}(1 + \frac{2}{\pi}(arcsin(\alpha_6^t) - arcsin(\alpha_8^t) + arcsin(\alpha_2^t))), \tag{B.11}$$

where

$$\alpha_1^t = \frac{2\rho_2^t - \rho_1^t - \rho_3^t}{2(1-\rho_1^t)}, \qquad \alpha_2^t = \frac{2\rho_1^t - \rho_2^t - 1}{2(1-\rho_1^t)}, \qquad \alpha_3^t = \frac{\rho_2^t - \rho_1^t + \rho_3^t - 1}{2\sqrt{(1-\rho_2^t)(1-\rho_3^t)}},$$

$$\alpha_4^t = \frac{\rho_1^t - \rho_3^t}{2(1-\rho_2^t)}, \qquad \alpha_5^t = \frac{1}{2}\sqrt{\frac{1-\rho_2^t}{1-\rho_1^t}}, \qquad \alpha_6^t = \frac{\rho_1^t + \rho_3^t - \rho_2^t - 1}{2\sqrt{(1-\rho_1^t)(1-\rho_2^t)}},$$

$$\alpha_7^t = \frac{\rho_1^t + \rho_2^t - \rho_3^t - 1}{2\sqrt{(1-\rho_1^t)(1-\rho_2^t)}}, \qquad \alpha_8^t = \frac{\rho_1^t - \rho_2^t}{\sqrt{(1-\rho_1^t)(1-\rho_3^t)}}.$$

Appendix C

The Brier score is computed as the average over the R simulations of the squared deviations of the h-period recession probability forecasts, $\hat{P}_h^{tr}$, from a binary value, s_{t+h}^r, which takes the value of one in actual recessions:

$$BS_h = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}(\hat{P}_h^{tr} - s_{t+h}^r)^2, \tag{B.1}$$

where $\hat{P}_h^{tr}$ is $P_{NP}(z_{t+h}=1)$ in the case of the nonparametric forecasts proposed in this paper. It is $P_L(z_{t+h}=1)$ in the case of the autoregressive forecasts, and it is $P(s_{t+h}=1|\hat{\theta}^r, I_t)$ in the case of the Markov-switching autoregressive forecasts for $h=1,2,3$, and $r=1,\ldots,R$. A Brier score of 0 means perfect accuracy, and a Brier score of 1 means perfect inaccuracy. When the measure focuses only on those time periods where a recession occurs ($s_t^r=1$), we call it the Brier Score of Recessions (BSR_h), whereas when it focuses on expansions ($s_t^r=0$), we call it the Brier Score of Expansions (BSE_h).

The second metric used to measure the models' performance follows the lines suggested by Berje and Jorda (2011). In particular, we measure the recession/expansion classification ability of the three forecasting methods using the Receiver Operating Characteristic (ROC) framework. In particular, given a threshold c, a recession is called when $\hat{P}_h^{tr} > c$, whereas an expansion is called when $\hat{P}_h^{tr} \leq c$. For each simulation, we can define the True Positive rate, $TP_h^r(c)$, and the False Positive rate, $FP_h^r(c)$, as

$$TP_h^r(c) = P(\hat{P}_h^{tr} > c | s_{t+h}^r = 1), \tag{B.2}$$

$$FP_h^r(c) = P(\hat{P}_h^{tr} > c | s_{t+h}^r = 0). \tag{B.3}$$

This implies that $TP_h^r(c)$ is the probability of calling recessions when there are actually recessions, and $FP_h^r(c)$ is the probability of calling recessions when there are actually expansions. We refer to the True Positive Rate, TPR_h, as the average of $TP_h^r(c)$ for all $r=1,2,\ldots,R$, and all c from 0 to 1 at steps of 0.001,

$$TPR_h = \frac{1}{1000R}\sum_{r=1}^{R}\sum_{c=0.001}^{1} TP_h^r(c). \tag{B.4}$$

Similarly, we define the True Negative rate, TNR_h as

$$TNR_h = \frac{1}{1000R}\sum_{r=1}^{R}\sum_{c=0.001}^{1} (1 - FP_h^r(c)). \tag{B.5}$$

The ROC curve represents the trade-off between $TP_h^r(c)$ and $FP_h^r(c)$ for different thresholds c. Concretely the ROC curve is represented plotting the points $(FP_h^r(c), TP_h^r(c))$ on a $[0,1]\times[0,1]$ plane for all possible thresholds c. When $\hat{P}_h^{tr}$ is an uninformative classifier with respect to the phase cycle, the ROC curve coincides with the main diagonal line, and, when it is a perfect classifier the ROC curve is on the upper left part of the unit quadrant. A standard measure of

overall classification ability is the area under the *ROC* curve, denoted by $AUROC_h^r$. This quantity takes values between 0.5 for a random classifier and 1 for a perfect classifier (see Berge and Jorda, 2011). Then, we take

$$AUROC_h = \frac{1}{R}\sum_{r=1}^{R} AUROC_h^r, \tag{B.6}$$

which measures the business cycle classification ability of the h-step forecasting probabilities.

Finally, we use Cohen's kappa coefficient, originally developed by Cohen (1960). It is a chance-corrected measure of agreement between forecasting technique classifications and real recessions. For a given threshold c, the kappa coefficient is calculated as

$$\kappa_h^r(c) = \frac{P_a^r(c) - P_r^r(c)}{1 - P_u^r(c)}, \tag{B.7}$$

where P_a^r denotes the probability of overall agreement and P_u^r is the probability of hypothetical probability of chance agreement for the r-th simulation. Specifically, these probabilities are calculated as:

$$P_a^r(c) = \frac{1}{T-h}\sum_{t=1}^{T-h}\left[I(\hat{P}_h^{tr} > c)s_{t+h}^r + I(\hat{P}_h^{tr} < c)(1 - s_{t+h}^r)\right], \tag{B.8}$$

and

$$P_u^r(c) = \frac{1}{T-h}\sum_{t=1}^{T-h} I(\hat{P}_h^{tr} > c)\frac{1}{T-h}\sum_{t=1}^{T-h} s_{t+h}^r + \frac{1}{T-h}\sum_{t=1}^{T-h} I(\hat{P}_h^{tr} < c)\frac{1}{T-h}\sum_{t=1}^{T-h}(1 - s_{t+h}^r) \tag{B.9}$$

where $I(\cdot)$ is an indicator function taking the value 1 for a true statement.

Using this notation, the calculation of Cohen's metric, κ_h^r, for a particular forecasting horizon h, is given by the average of all kappa coefficients:

$$\kappa_h = \frac{1}{1000R}\sum_{r=1}^{R}\sum_{c=0.001}^{1}\kappa_h^r(c) \tag{B.10}$$

The interpretation of a given magnitude of κ_h is somehow problematic.[8] Nonetheless, it is straightforward to check that Cohen's κ is equal to 1 when there is complete agreement between the two classifiers while it is equal to 0 when there is no agreement between the classifiers other than what would be expected by chance. The metric can be negative when the agreement between the two classifiers is worse than random.

[8] Cohen's κ tends to increase when recessions and expansions are equiprobable and when expansions and recessions are distributed asymmetrically by the two classifiers.

Table 1. Nonparametric model's performance

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Baseline model: $a_1 = 0.2$; $T = 500$; $\sigma^2 = 0.5$; $p_{00} = 0.9$; $p_{11} = 0.6$; $\mu_0 - \mu_1 = 0.5$																					
	0.16	0.36	0.11	0.75	0.52	0.83	0.31	0.13	0.33	0.09	0.80	0.48	0.78	0.23	0.14	0.30	0.10	0.80	0.50	0.74	0.21
Panel B. Changes in time series' inertia																					
$a_1 = 0.5$	0.15	0.42	0.08	0.75	0.49	0.89	0.37	0.12	0.39	0.06	0.81	0.45	0.85	0.29	0.13	0.32	0.08	0.81	0.49	0.77	0.24
$a_1 = 0.8$	0.18	0.65	0.06	0.64	0.31	0.93	0.27	0.16	0.63	0.04	0.74	0.28	0.92	0.23	0.14	0.51	0.05	0.74	0.35	0.87	0.21
Panel C. Changes in sample size																					
$T = 250$	0.17	0.40	0.11	0.73	0.49	0.83	0.28	0.14	0.34	0.09	0.78	0.47	0.78	0.21	0.14	0.32	0.10	0.78	0.48	0.74	0.19
$T = 1000$	0.16	0.36	0.11	0.76	0.52	0.84	0.33	0.13	0.33	0.08	0.80	0.48	0.78	0.24	0.14	0.30	0.10	0.80	0.50	0.75	0.22
Panel D. Changes in variance																					
$\sigma^2 = 1$	0.23	0.44	0.18	0.64	0.45	0.75	0.15	0.18	0.39	0.14	0.67	0.43	0.71	0.11	0.19	0.36	0.15	0.67	0.45	0.68	0.11
$\sigma^2 = 1.5$	0.26	0.47	0.21	0.60	0.42	0.71	0.10	0.20	0.41	0.15	0.62	0.41	0.69	0.08	0.21	0.38	0.16	0.62	0.43	0.66	0.07
Panel E. Changes in states' inertia																					
p_{00}=p_{11}=0.6	0.25	0.36	0.14	0.73	0.52	0.79	0.31	0.22	0.33	0.10	0.76	0.48	0.74	0.23	0.21	0.30	0.12	0.76	0.50	0.71	0.21
p_{00}=p_{11}=0.9	0.21	0.31	0.12	0.78	0.58	0.83	0.40	0.19	0.29	0.09	0.83	0.52	0.78	0.30	0.18	0.26	0.10	0.83	0.54	0.74	0.28
Panel F. Changes in differences of within-state means																					
μ_0-μ_1=1	0.08	0.23	0.04	0.89	0.65	0.94	0.58	0.07	0.24	0.04	0.94	0.57	0.87	0.43	0.09	0.21	0.05	0.94	0.59	0.82	0.37
μ_0-μ_1=2	0.02	0.10	0.01	0.96	0.79	0.99	0.81	0.03	0.15	0.01	0.99	0.68	0.94	0.64	0.04	0.13	0.02	0.99	0.67	0.87	0.54

Notes. For each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ), for the nonparametric approach. Panel A reports the results of the baseline specification and Panels B to F refer to the extent to which departures from the baseline model affect the classification measures.

Table 2. Comparison of the three different approaches under data problems

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Markov-switching model																					
Baseline	0.28	0.47	0.28	0.66	0.47	0.64	0.10	0.28	0.42	0.25	0.63	0.44	0.63	0.06	0.29	0.44	0.24	0.61	0.43	0.63	0.05
Outliers	0.23	0.75	0.11	0.55	0.19	0.85	0.05	0.23	0.75	0.10	0.55	0.19	0.84	0.03	0.23	0.75	0.10	0.55	0.18	0.84	0.03
Breaks	0.42	0.09	0.50	0.64	0.79	0.38	0.09	0.39	0.14	0.46	0.61	0.71	0.40	0.07	0.37	0.18	0.42	0.59	0.66	0.42	0.05
Hetero	0.41	0.41	0.41	0.52	0.51	0.51	0.01	0.40	0.41	0.40	0.51	0.50	0.51	0.01	0.40	0.41	0.40	0.51	0.50	0.51	0.01
ARCH	0.30	0.66	0.22	0.52	0.28	0.72	0.00	0.30	0.66	0.22	0.51	0.28	0.71	0.00	0.30	0.66	0.22	0.51	0.27	0.71	-0.01
Panel B. Linear model																					
Baseline	0.14	0.51	0.04	0.76	0.32	0.90	0.20	0.15	0.69	0.02	0.75	0.17	0.88	0.05	0.17	0.78	0.01	0.61	0.12	0.90	0.01
Outliers	0.15	0.55	0.04	0.73	0.28	0.90	0.16	0.17	0.72	0.02	0.54	0.16	0.86	0.01	0.17	0.72	0.02	0.52	0.15	0.85	0.00
Breaks	0.14	0.64	0.01	0.73	0.23	0.97	0.23	0.17	0.81	0.01	0.79	0.10	0.97	0.07	0.18	0.87	0.01	0.69	0.07	0.96	0.03
Hetero	0.17	0.51	0.07	0.68	0.32	0.85	0.14	0.16	0.58	0.06	0.56	0.24	0.77	0.01	0.17	0.62	0.05	0.51	0.22	0.79	0.00
ARCH	0.16	0.53	0.06	0.69	0.32	0.87	0.15	0.15	0.64	0.03	0.65	0.21	0.83	0.03	0.16	0.70	0.03	0.57	0.17	0.84	0.01
Panel C. Nonparametric specification																					
Baseline	0.16	0.36	0.11	0.75	0.52	0.83	0.31	0.13	0.33	0.09	0.80	0.48	0.78	0.23	0.14	0.30	0.10	0.80	0.50	0.74	0.21
Outliers	0.16	0.36	0.12	0.75	0.52	0.83	0.31	0.14	0.31	0.10	0.80	0.50	0.76	0.23	0.14	0.31	0.10	0.79	0.50	0.76	0.22
Breaks	0.14	0.41	0.08	0.76	0.47	0.89	0.35	0.12	0.38	0.06	0.84	0.43	0.84	0.25	0.12	0.31	0.08	0.84	0.48	0.79	0.22
Hetero	0.18	0.43	0.12	0.69	0.43	0.81	0.19	0.16	0.39	0.11	0.72	0.40	0.72	0.10	0.17	0.35	0.12	0.72	0.43	0.68	0.09
ARCH	0.20	0.39	0.15	0.70	0.50	0.79	0.24	0.17	0.36	0.12	0.73	0.46	0.74	0.18	0.17	0.32	0.13	0.73	0.48	0.71	0.16

Notes. For the Markov switching, linear and nonparametric approaches, and each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ).

Table 3. Empirical performance in G7 countries: 1955.2-2019.4

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Markov-switching model																					
US	0.08	0.53	0.01	0.92	0.31	0.93	0.26	0.11	0.70	0.02	0.81	0.17	0.90	0.07	0.11	0.77	0.01	0.69	0.13	0.89	0.02
UK	0.07	0.62	0.01	0.90	0.24	0.96	0.24	0.09	0.74	0.01	0.84	0.15	0.95	0.12	0.10	0.80	0.01	0.76	0.11	0.85	0.07
Canada	0.06	0.51	0.01	0.86	0.36	0.96	0.37	0.07	0.63	0.01	0.81	0.24	0.94	0.21	0.08	0.73	0.01	0.76	0.16	0.93	0.10
France	0.45	0.09	0.54	0.72	0.83	0.35	0.11	0.44	0.10	0.52	0.65	0.80	0.35	0.08	0.43	0.10	0.51	0.58	0.76	0.34	0.06
Germany	0.25	0.97	0.01	0.71	0.02	0.98	-0.01	0.25	0.96	0.01	0.66	0.02	0.98	0.00	0.25	0.96	0.01	0.58	0.02	0.98	0.00
Italy	0.32	0.43	0.29	0.60	0.48	0.62	0.09	0.33	0.47	0.29	0.56	0.46	0.61	0.06	0.34	0.48	0.29	0.52	0.44	0.60	0.04
Japan	0.47	0.01	0.61	0.83	0.95	0.31	0.15	0.46	0.02	0.60	0.76	0.92	0.30	0.12	0.46	0.03	0.59	0.73	0.89	0.29	0.11
Panel B. Linear model																					
US	0.17	0.19	0.17	0.85	0.71	0.74	0.31	0.13	0.29	0.11	0.83	0.50	0.73	0.17	0.13	0.45	0.08	0.68	0.34	0.73	0.05
UK	0.14	0.29	0.12	0.81	0.53	0.78	0.20	0.11	0.42	0.07	0.76	0.37	0.76	0.11	0.11	0.48	0.06	0.73	0.32	0.76	0.07
Canada	0.19	0.21	0.19	0.82	0.70	0.72	0.25	0.12	0.31	0.10	0.76	0.49	0.72	0.16	0.12	0.44	0.08	0.71	0.34	0.73	0.05
France	0.16	0.56	0.07	0.65	0.28	0.83	0.07	0.16	0.69	0.05	0.66	0.17	0.80	-0.02	0.16	0.55	0.07	0.51	0.26	0.74	0.00
Germany	0.18	0.43	0.10	0.69	0.38	0.81	0.16	0.19	0.55	0.07	0.50	0.26	0.74	0.00	0.20	0.59	0.07	0.60	0.24	0.75	-0.01
Italy	0.19	0.32	0.14	0.76	0.55	0.78	0.29	0.18	0.41	0.10	0.69	0.38	0.72	0.10	0.20	0.53	0.09	0.55	0.28	0.72	0.00
Japan	0.19	0.36	0.14	0.71	0.48	0.76	0.20	0.17	0.46	0.09	0.66	0.34	0.74	0.07	0.18	0.53	0.07	0.57	0.28	0.74	0.02
Panel C. Nonparametric specification																					
US	0.07	0.41	0.02	0.79	0.5	0.97	0.53	0.11	0.49	0.05	0.78	0.37	0.87	0.23	0.13	0.54	0.06	0.68	0.31	0.84	0.13
UK	0.09	0.39	0.06	0.78	0.51	0.92	0.40	0.10	0.37	0.07	0.77	0.48	0.82	0.26	0.11	0.43	0.06	0.71	0.44	0.82	0.24
Canada	0.08	0.45	0.04	0.76	0.49	0.94	0.42	0.08	0.42	0.05	0.78	0.44	0.86	0.27	0.10	0.44	0.07	0.74	0.40	0.82	0.18
France	0.17	0.75	0.04	0.60	0.20	0.94	0.17	0.16	0.64	0.06	0.68	0.23	0.86	0.09	0.18	0.75	0.06	0.60	0.15	0.86	0.00
Germany	0.19	0.52	0.09	0.68	0.37	0.88	0.25	0.19	0.51	0.09	0.68	0.31	0.80	0.10	0.22	0.57	0.10	0.60	0.27	0.77	0.03
Italy	0.18	0.50	0.08	0.71	0.41	0.89	0.32	0.17	0.46	0.08	0.72	0.38	0.81	0.18	0.20	0.50	0.11	0.61	0.33	0.76	0.08
Japan	0.18	0.61	0.05	0.66	0.32	0.92	0.27	0.17	0.51	0.07	0.73	0.32	0.83	0.15	0.19	0.53	0.08	0.69	0.30	0.82	0.11

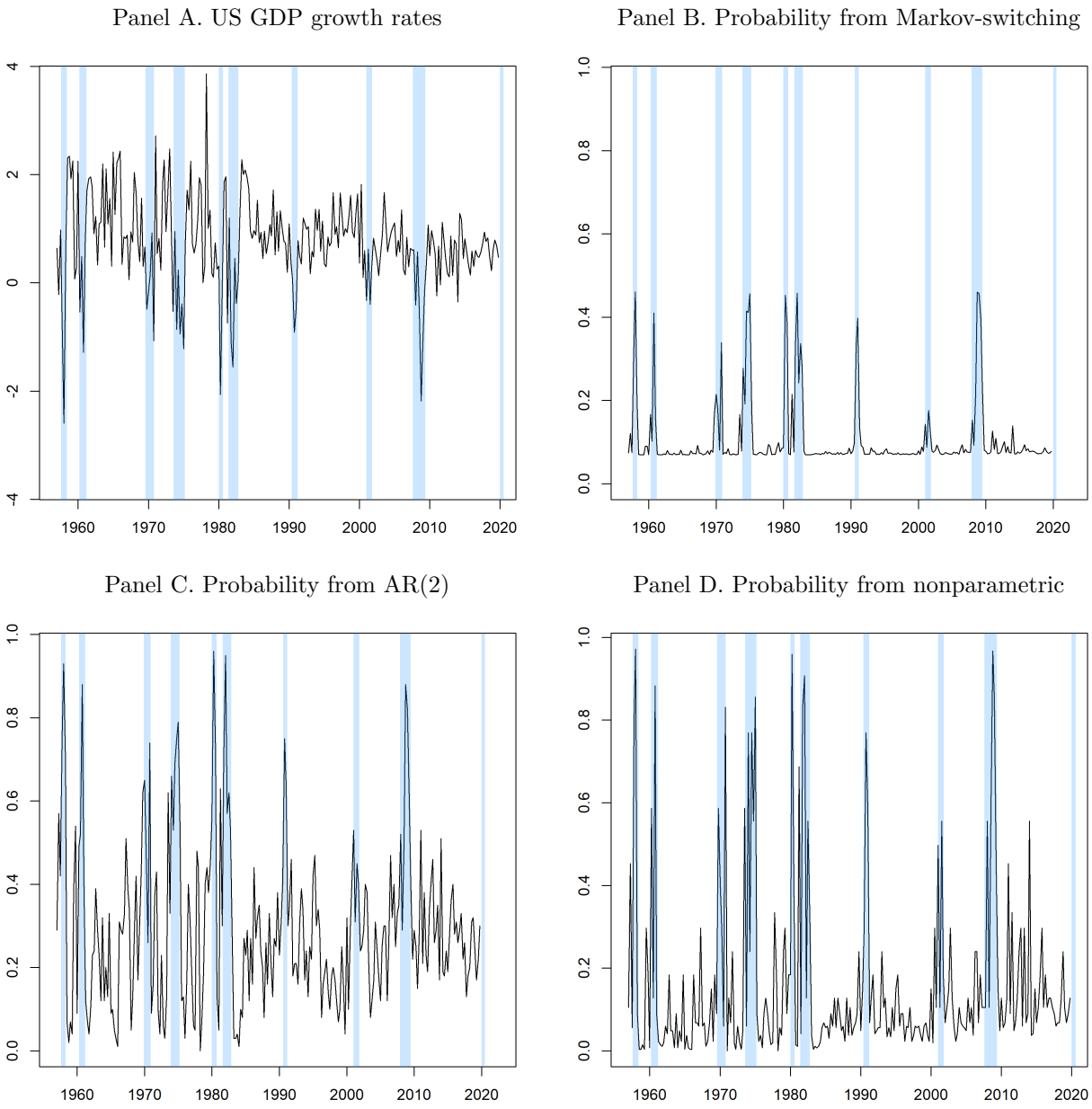
Notes. For the Markov switching, linear and nonparametric approaches, and each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ).

Table 4. Empirical performance in G7 countries: 1955.2-2022.3

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Markov-switching model																					
US	0.13	0.98	0.01	0.50	0.01	0.99	0.00	0.13	0.98	0.01	0.50	0.01	0.99	0.00	0.13	0.98	0.01	0.50	0.01	0.99	0.00
UK	0.12	0.99	0.01	0.50	0.01	0.99	-0.01	0.12	0.97	0.01	0.50	0.01	0.98	-0.01	0.12	0.96	0.01	0.50	0.02	0.98	-0.01
Canada	0.11	0.98	0.01	0.50	0.01	0.99	0.00	0.10	0.98	0.01	0.50	0.01	0.99	0.00	0.10	0.98	0.01	0.50	0.01	0.99	0.00
France	0.21	0.97	0.01	0.50	0.02	0.99	0.01	0.21	0.98	0.01	0.49	0.01	0.98	-0.01	0.21	0.97	0.01	0.49	0.01	0.98	-0.01
Germany	0.34	0.13	0.40	0.68	0.64	0.37	0.00	0.35	0.13	0.40	0.60	0.64	0.37	0.00	0.35	0.13	0.40	0.55	0.64	0.36	0.00
Italy	0.24	0.97	0.01	0.50	0.02	0.99	0.01	0.24	0.97	0.01	0.49	0.01	0.98	-0.01	0.24	0.96	0.01	0.49	0.02	0.97	-0.01
Japan	0.23	0.97	0.01	0.51	0.01	0.99	0.00	0.23	0.97	0.01	0.50	0.01	0.99	0.00	0.23	0.97	0.01	0.50	0.01	0.99	0.00
Panel B. Linear model																					
US	0.13	0.31	0.11	0.81	0.49	0.79	0.17	0.13	0.49	0.07	0.75	0.30	0.75	0.03	0.13	0.55	0.07	0.56	0.26	0.75	0.01
UK	0.14	0.55	0.08	0.59	0.27	0.81	0.03	0.13	0.70	0.06	0.69	0.17	0.78	-0.03	0.12	0.53	0.07	0.58	0.27	0.74	0.01
Canada	0.14	0.41	0.10	0.69	0.39	0.79	0.07	0.12	0.57	0.06	0.52	0.25	0.75	0.00	0.12	0.56	0.07	0.48	0.25	0.75	0.00
France	0.17	0.52	0.09	0.64	0.30	0.81	0.07	0.17	0.69	0.06	0.65	0.17	0.79	-0.04	0.16	0.53	0.07	0.54	0.27	0.74	0.00
Germany	0.19	0.54	0.07	0.66	0.29	0.84	0.11	0.20	0.62	0.07	0.62	0.21	0.76	-0.03	0.18	0.54	0.07	0.56	0.27	0.75	0.01
Italy	0.18	0.47	0.09	0.67	0.34	0.82	0.13	0.19	0.60	0.06	0.56	0.23	0.76	-0.01	0.19	0.55	0.07	0.53	0.25	0.75	0.00
Japan	0.18	0.43	0.11	0.70	0.40	0.79	0.16	0.17	0.54	0.06	0.64	0.27	0.77	0.03	0.18	0.55	0.06	0.59	0.26	0.76	0.01
Panel C. Nonparametric specification																					
US	0.08	0.42	0.03	0.78	0.50	0.96	0.49	0.12	0.49	0.06	0.77	0.36	0.86	0.20	0.14	0.56	0.07	0.67	0.30	0.83	0.11
UK	0.10	0.39	0.06	0.78	0.51	0.92	0.39	0.11	0.37	0.08	0.76	0.46	0.80	0.22	0.11	0.44	0.07	0.77	0.42	0.81	0.21
Canada	0.09	0.48	0.05	0.74	0.47	0.94	0.39	0.10	0.44	0.06	0.76	0.42	0.84	0.22	0.11	0.46	0.07	0.72	0.38	0.82	0.16
France	0.17	0.70	0.05	0.62	0.24	0.94	0.05	0.17	0.63	0.07	0.68	0.24	0.85	0.08	0.19	0.72	0.07	0.60	0.17	0.84	-0.01
Germany	0.20	0.52	0.09	0.68	0.37	0.88	0.25	0.20	0.50	0.10	0.68	0.32	0.79	0.09	0.22	0.57	0.11	0.60	0.26	0.77	0.02
Italy	0.18	0.50	0.08	0.71	0.41	0.89	0.32	0.18	0.44	0.09	0.72	0.39	0.80	0.18	0.21	0.51	0.11	0.61	0.32	0.76	0.07
Japan	0.19	0.62	0.06	0.65	0.31	0.91	0.25	0.17	0.49	0.07	0.73	0.34	0.82	0.15	0.19	0.55	0.08	0.69	0.29	0.81	0.09

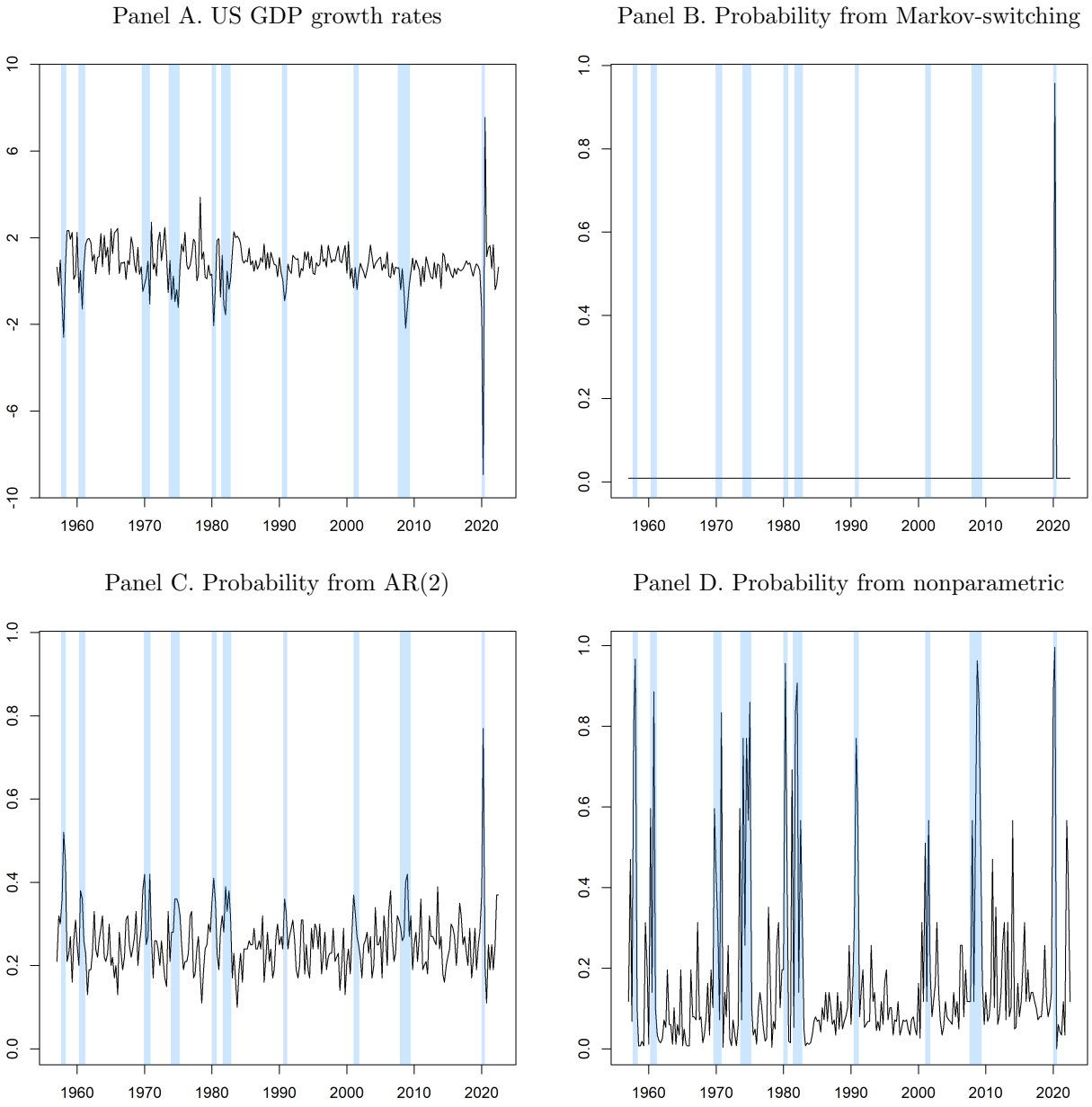
Notes. For the Markov switching, linear and nonparametric approaches, and each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ).

Figure 1: Business cycle inferences: 1955.2-2019.4



Notes. Panel A displays the US quarterly GDP growth rate for the period 1955.2-2019.4. Panels B, C, and D plot the 2-quarter ahead predictions of the probability of recession from AR(2), Markov-switching and nonparametric models, respectively. The shaded areas represent the NBER-referenced recessions.

Figure 2: Business cycle inferences: 1955.2-2022.3



Notes. Panel A displays the US quarterly GDP growth rate for the period 1955.2-2022.3. Panels B, C, and D plot the 2-quarter ahead predictions of the probability of recession from AR(2), Markov-switching and nonparametric models, respectively. The shaded areas represent the NBER-referenced recessions.