Latin American Journal of Central Banking 1 (2020) 100012

Contents lists available at ScienceDirect

Latin American Journal of Central Banking

journal homepage: www.elsevier.com/locate/latcb

Forecasting Costa Rican inflation with machine learning methods

Adolfo Rodríguez-Vargas

Central Bank of Costa Rica, Costa Rica

ARTICLE INFO

JEL codes:
E31
C45
C49
C53

Keywords:
Inflation
Forecasting
Machine learning
Forecast evaluation

ABSTRACT

We present a first assessment of the predictive ability of machine learning methods for inflation forecasting in Costa Rica. We compute forecasts using two variants of k-nearest neighbors, random forests, extreme gradient boosting and a long short-term memory (LSTM) network. We evaluate their properties according to criteria from the optimal forecast literature, and we compare their performance with that of an average of univariate inflation forecasts currently used by the Central Bank of Costa Rica. We find that the best-performing forecasts are those of LSTM, univariate KNN and, to a lesser extent, random forests. Furthermore, a combination performs better than the individual forecasts included in it and the average of the univariate forecasts. This combination not biased; its forecast errors show appropriate properties, and it improves the forecast accuracy at all horizons, both for the level of inflation and for the direction of its changes.

1. Introduction

Article 2 of the Organic Law of the Central Bank of Costa Rica (BCCR) states that its fundamental goal is to ensure internal stability of the currency, which has been understood as maintaining low and stable inflation. As part of an ongoing effort to fulfill that goal, in 2005, the BCCR began a process to migrate to an inflation-targeting scheme that finished in 2018 with the formal adoption of such a regime.

One of the main features of this policy scheme is its prospective nature: since monetary policy operates with a lag, policy decisions take into account the expected trajectory of inflation and other relevant economic variables during the policy horizon. Hence, for an inflation-targeting central bank, it is crucial to have good forecasts for the main economic variables to make adequate policy decisions, and among them, inflation forecasts are of particular importance. Currently, the BCCR forecasts inflation with an ensemble of univariate models (Fuentes and Rodríguez, 2016), an ensemble of Bayesian models (Chavarría Mejía and Chaverri, 2015) and a semi-structural econometric model (Muñoz and Tenorio, 2008).

All these forecasts come from relatively traditional econometric methods. However, machine learning methods have become increasingly popular as a forecasting tool, due to the growing availability of big databases and computing power, and to greater access to specialized software. Their use is widespread in classification problems where the variable of interest is discrete, like prediction of delinquency in loans or consumer purchasing decisions, where they have often outperformed more traditional methods. As such, most applications use cross-sectional data in classification problems. However, machine learning methods can also be adapted for prediction of continuous time-series data, like inflation or GDP growth.

Precisely, the goal of this study is to carry out a first evaluation of the performance of machine learning methods in forecasting Costa Rican inflation. We hope the results of this evaluation will be useful as a guide for future research efforts at the BCCR. The evaluation follows the traditional forecast evaluation literature (see West, 2006, for a review); it aims to verify whether forecasts from these methods comply with properties of optimal forecasts under quadratic loss. Furthermore, we assess whether their performance is superior to that of their univariate counterparts currently in use at the BCCR. Hence, the exercise presented departs from the

E-mail address: rodriguezva@bccr.fi.cr

https://doi.org/10.1016/j.latcb.2020.100012
Received 27 February 2020; Received in revised form 1 October 2020; Accepted 30 October 2020
2666-1438/

traditional approach to machine learning forecasting, which typically excludes any consideration of the properties of the forecasts, other than minimization of loss for out-of-sample prediction.

We evaluate forecasts performed using random forests, extreme gradient boosting, a long short-term memory (LSTM) network, and two variants of k-nearest neighbors. We found that the forecasts with the best performance are those from the LSTM model, univariate KNN, and, to a lesser extent, random forests and extreme gradient boosting. A combination of forecasts improves performance, in comparison with individual forecasts at all horizons, and outperforms the forecasts from univariate methods.

The rest of this document includes the following: Section 2 is a brief review of relevant applied literature; Section 3 provides the methodological details on data, implementation of the methods and evaluation criteria; Section 4 is the discussion of the main results, followed by a succinct concluding section.

2. Overview of related literature

We are not aware of any published study that used KNN to forecast economic variables in Costa Rica, so ours contributes a formal evaluation of the forecasting properties of a relatively simple and inexpensive method. Elsewhere, Diebold and Nason (1990), Mizrach (1992), Lisi and Medio (1997), Lisi and Schiavo (1999), Meade (2002), and Fernandez-Rodriguez et al. (1999) forecast exchange rates using KNN; Barkoulas et al. (1996) and Nowman and Saltoglu (2003) forecast interest rates; and Agnon et al. (1999) applied this type of method to the projection of commodity prices. More recently, Nikolopoulos et al. (2016) used KNN to forecast sporadic demand in a supply chain setting. However, these studies did not attempt a systematic evaluation of the properties of the forecasts.

The random forests (RF) algorithm was proposed by Breiman (2001a). It is a popular method in medicine and other biological sciences, but less so in economics. A notable example is Biau and D'Elia (2011), who applied the method to select variables that feed into a GDP forecasting model for the Euro area from a dataset containing 172 indicators; they found that it compares favorably with autoregressive forecasts and with those of the *Eurozone Economic Outlook*. Thus, their approach is very similar to the goal of our study, albeit with only one ML method being used, as an intermediate step and not to produce the forecast compared with the autoregressive benchmark. David (2017) also explored the use of random forests in variable selection for forecasting models of economic phenomena. In contrast, we do not use random forests as a tool to guide the choice of explanatory variables, but to perform the forecast itself. We note that the variable importance measures computed from the application of random forests present an opportunity to improve other forecasting methods at the BCCR, but we believe that such an endeavor would be best served by a separate study. Furthermore, Bajari et al. (2015) included random forests in their ML toolkit for the estimation of demand for groceries.

To our knowledge, ours is the first study to formally assess the performance of boosting to forecast inflation in Costa Rica. Boosting was proposed by Freund and Schapire (1995, 1996), and introduced to regression problems by Friedman et al. (2000) and Friedman (2001). Variants of the method have proved useful to forecast with large datasets in a computationally efficient way. Some examples are as follow: Wohlrabe and Buchen (2014), who evaluated its performance in forecasting economic variables for the Euro zone and the USA; Lehmann and Wohlrabe (2016), who used German data to assess the type of indicators usually selected by the method; and Zeng (2017), who successfully used the method to select disaggregate variables to forecast aggregate variables. As with random forests, we do not use boosting as a tool for variable selection, but rather to perform the forecast itself.

The widespread use of neural networks to forecast macroeconomic variables started in the 1990s, although the first developments on the topic date from the 1940s (McCullogh and Pitts, 1943; Hebb, 1949). Among the first examples of the application of neural networks for this type of problems are as follow: Swanson and White (1995, 1997), in finance; Tkacz and Hu (1999), to forecast Canada's GDP; Stock and Watson (1998), who found that neural networks perform poorly in comparison to other univariate methods; Refenes and White (1998) and Fernández-Rodríguez et al. (2000), also in finance; and Moshiri and Cameron (2000), who forecast inflation. Nakamura (2006) showed that applying early stopping close to local minima improves forecasting ability, which had already been suggested by González (2000). Cook and Hall (2017) forecast employment indicators using several deep neural network architectures, among them LSTM, and they showed that these improve accuracy with respect to simpler configurations, but they did not mention how these forecasts compare to ML methods other than neural networks. For Costa Rica, Solera (2005) and Esquivel Monge (2007) applied simple neural network models with several specifications for inflation forecasting, and compared their performance with that of forecasting methods used at the time at the BCCR. However, they did not formally evaluate the properties of the forecasts beyond tests of difference in predictive accuracy.

3. Methods, forecasting scheme and evaluation criteria

In this section, we focus on discussing the data and the methodological details of the forecasting and evaluation exercises. We do not present a thorough description of the methods applied. Athey and Imbens (2019) offer a useful introductory review of machine learning methods for economists. Several well-known textbooks present in-depth details of the methods cited below. See for example Hastie et al. (2009), Shalev-Shwartz and Ben-David (2014), Bishop (2006), and Mitchell (1997).

3.1. Data

The variable to forecast is the interannual variation rate of the Consumer Price Index of Costa Rica (with base June-2015). We used a dataset comprising monthly data for the variables detailed in Table 1. We included real sector variables, monetary and internal and

Table 1
Variables used in the forecasts.

Name	Description	Source
IPC	Interannual variation rate of the Consumer Price Index (IPC), base June 2015	National Institute of Statistics and Censuses (INEC)
EXPINF12	12-month inflation expectations	Central Bank of Costa Rica (BCCR)
TCN	Interannual variation rate of the average monthly exchange rate in the MONEX market	BCCR
TCR_M	Interannual variation rate of the multilateral Index of Real Exchange Rate	BCCR
IMAETC	Interannual variation rate of the Monthy Index of Economic Activity (IMAE), trend-cycle	BCCR
ICFNIV	Financial Conditions Index	BCCR
CREDPRIVSF	Interannual variation rate of total credit of the national financial system to the private sector, local currency.	BCCR
BASEM	Interannual variation rate of the monetary base	BCCR
M1	Interannual variation rate of M1	BCCR
TPM	Monetary policy rate	BCCR
TBP	Deposit basic rate (Tasa Básica Pasiva)	BCCR
PRIMERATE	Prime Rate	U.S. Federal Reserve
PETRO	Interannual variation rate of the crude oil barrel price, average ($/bbl)	Pink Sheet, World Bank
GRANOS	Interannual variation rate of the Grains index	Pink Sheet, World Bank
INFSOC	Inflation of trade partners (interannual variation)	BCCR
ISMNNIV	Interannual variation rate of the Index of Minimum Nominal Wages	BCCR
ISMRNIV	Interannual variation rate of the Index of Minimum Real Wages	BCCR
RESPIB	Financial result of the Central Government as a share of GDP	BCCR, with data from the Ministry of Finance
DEUDAPIB	Total internal debt as a share of GDP	BCCR, with data from the Ministry of Finance

external price variables, financial and exchange rate data, and labor market data. Furthermore, we included 12 lags of each of these variables, as well as a set of seasonal dummies, totaling 258 possible explanatory variables. Most of them correspond to interannual variation rates, except for interest rates and the Monetary Conditions Index. Data cover the period January-2003 to February-2019. See Table 16 in the annex for links to sources.

3.2. Validation of models

Two considerations must be addressed when performing cross validation with time series data: first, the ordering of the observations must be preserved, and second, variables are usually autocorrelated. This means that standard cross-validation techniques, like leave-one-out or k-folds cannot be applied, because they require a random partition of the sample. Random partition is inappropriate because the training sample might end up containing observations that occur later than the validation sample (data leakage), and because validation samples might end up highly autocorrelated, which violates a basic principle of the exercise. Thus, we apply a cross-validation scheme designed specifically to deal with time series variables, discussed by Tashman (2000), who called it rolling-origin evaluation, and Bergmeir and Benítez (2012), who called it rolling-origin-recalibration evaluation. In this procedure, a series of single-observation test sets is constructed, with corresponding training tests containing only information prior to the test set. Hyndman and Athanasopoulos (2018) described the method as follows:

- Assume k is the minimum number of observations for a training set, h is the forecast horizon and T is the total number of observations.
- Observation $t = k + i$ is selected as the test set, observations $1, 2,\dots, k + i - h$ are used to estimate the model, and the forecast error is computed for $t = k + i$.
- This process is repeated for $i = 0, 1,\dots, T - k$.
- An accuracy measure is computed over all errors.

3.3. Forecasting procedure

The general forecasting scheme is as follows. We used a first estimation sample from January 2003 to December 2016 to produce a first set of 12 out-of-sample forecasts, for January 2017 to December 2017. After this, we increased the estimation sample by one month at a time, and then forecast the next 12 months. We repeated this process until the last month forecast was February 2019. From these sets of 12 multi-horizon forecasts, we obtained series of forecasts with fixed horizons of $h = 1$, 3, 6 and 12 months. These were the series included in the evaluation. Properties of the forecasts were assesses using the actual values of inflation for the period January 2017 to February 2019. Details of estimation and forecasting for each method are presented in the next sections. Table 14 of the annex presents software details.

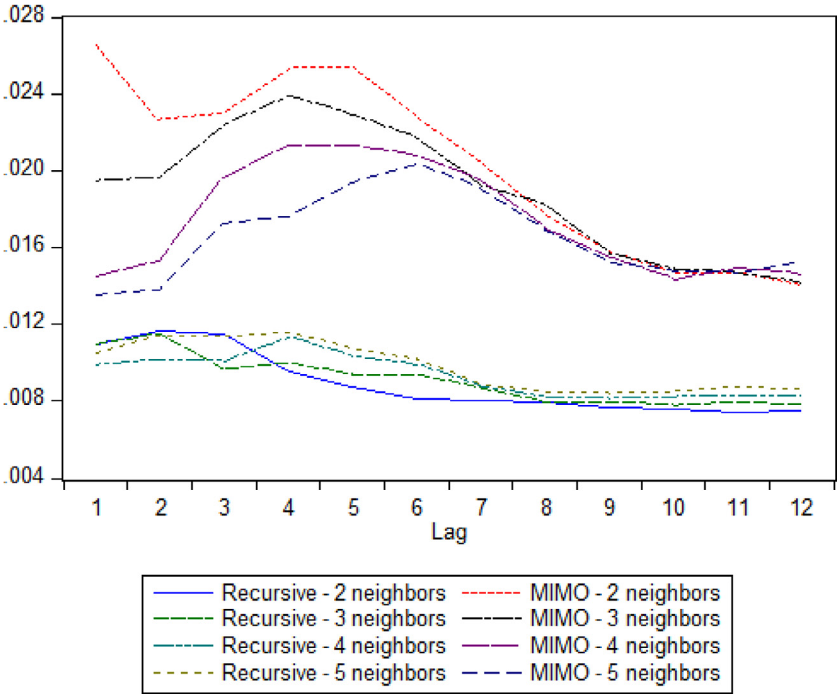

Fig. 1. RMSEs for univariate KNN models
Source: Own elaboration.

3.4. Methods applied

3.4.1. Univariate KNN

KNN regression originated in Mack (1981), and its use for forecasting time series was popularized by Yakowitz and Karlsson (1987), and Yakowitz (1987). Martínez et al. (2017) discussed a methodology for applying the *k-nearest neighbors* algorithm (KNN) to time series forecasting. In this case the algorithm searches a set of periods up to the n most recent lags of the target variable, finds the k most similar periods (i.e., the nearest neighbors) according to a distance metric, and makes the forecast based on the subsequent evolution of the variable of interest. The prediction is done by aggregating the values after the periods where the k nearest neighbors are located, typically using an average (simple or weighting by distance).

We performed the forecasts in R, with the *tsfknn* package (see Martínez et al., 2019). The number of nearest neighbors to search (k), the number of lags of the target variable (n), and the method to perform multi-horizon forecasts were chosen to minimize the average SSR in the forecast sample. Forecasts were computed considering from 2 to 5 nearest neighbors, from 1 to 12 lags of the target variable, and both a multiple input multiple output (MIMO) forecasting strategy and a recursive forecasting strategy. With the MIMO strategy, for each nearest neighbor a vector of target values is defined, with size equal to that of the number of periods to forecast. The forecast is then performed by searching the values of inflation most similar to the last 12 values of the time series, and by aggregating the target vectors that follow each nearest neighbor (see Martínez et al., 2017, for a discussion). The recursive strategy is the usual procedure in autoregressive models, in which past predictions of the target variable are used when there are no more historical data for it. In all cases, Euclidean distance was used, and the aggregation function to generate the forecast was the arithmetic mean.

Fig. 1 shows the average across forecast horizons of the RMSEs for each estimated model. It is clear that the recursive forecasts are more accurate in all cases, with a more homogeneous performance between models with different numbers of neighbors and different numbers of lags. The forecasts with the lowest RMSE are those computed with a recursive strategy, $k = 2$ neighbors and $n = 11$ lags of the target variable. These are the parameters used to compute the forecasts included in the evaluation.

3.4.2. KNN with explanatory variables

The principle of the multivariate KNN algorithm is the same as above, but it considers the lags of the other variables in addition to those of Y. We decided to include 11 lags of inflation, as in the univariate case. It would have been impractical to consider all 258 exogenous variables in the search for nearest neighbors, so a selection process was implemented by computing the correlation of inflation with up to 12 lags of several variables and selecting the lag that resulted in the highest correlation with inflation. The variables considered were those of the inflation equations in the quarterly semi-structural model used for policy analysis at the BCCR (output gap, 12-month inflation expectations, and the interannual variation rates of the nominal exchange rate, the price of grains and the price of oil) and the interannual changes in the monetary base and the monetary policy rate.[1] Finally, we decided to include as explanatory variables the contemporaneous value of 12 months ahead inflation expectations and of the interannual variation rate of nominal exchange rate, the fourth lag of the interannual variation rate of the monetary base, the ninth lag of the policy rate, the eight lag of interannual variation rate of the oil price, and the sixth lag of the interannual variation rate of the price of grains. The number of neighbors was determined as in the univariate case: by computing the average of the RMSEs of each set of forecasts for 2–5 neighbors. The lowest RMSE was reached with 5 neighbors. We used Euclidean distance and aggregation by inverse-distance weighted average.[2]

3.4.3. Extreme gradient boosting

The version of boosting applied in this study is extreme gradient boosting, developed by Chen and Guestrin (2016) as a regularized adaptation of gradient boosting, the end of which is to control overfitting. The differences among other implementations of the algorithm are technical: it is an efficient and *scalable* application of the gradient boosting approach, which is optimized to receive sparse data.

In machine learning methods that involve bootstrapping, like boosting or random forests, the resampling scheme must take into account the time dependence in the data. To deal with the non-i.i.d nature of the time-series variables, we used time-delay embedding to prepare the input data. Thus, we created new variables corresponding to 12 lags of each variable, and included them as features in the input data along with the original variables. In this way, during bootstrapping, each value of the target variable in time t was sampled along with the values in t of the other variables, its own 12 lags and the 12 lags of every other variable.[3]

We used the R package *caret* for validation, training, and prediction. As described in Section 3.2, the validation method is rolling-origin evaluation, implemented here with the function *trainControl* (option *timeSlice*). For this study, we used all variables described in Section 3.1, with the RMSE as the evaluation metric and cyclical selection of variables (deterministic selection, cyclically considering one variable at a time). The parameters calibrated were the maximum number of iterations needed for the underlying gradient descent algorithm to converge (tuned at $nrounds = 50$), the step size shrinkage parameter (tuned at $eta = 0.3$), and the linear booster regularization parameters that limit overfitting (tuned at $lambda = 0.1$ and $alpha = 0.1$).

3.4.4. Random forests

We used the R packages *randomForest* and *dyn* for training and prediction, and time-delay embedding to prepare the input data, as explained in Section 3.4.3. We used all of the variables in Table 1, as well as 12 of their lags. Implementation of random forests requires setting the number of trees to generate, the minimum number of terminal observations admissible in each node and the number of variables to include in each split. It is generally accepted that the implementation of this method requires relatively little calibration of these parameters to obtain acceptable predictions (Segal, 2004; Boelaert and Ollion, 2018; Athey and Imbens, 2019). The parameter that is most frequently optimally calibrated is the number of candidate variables entering each split. Regarding the number of trees to generate, Breiman (2001a) proved convergence of the mean squared generalization error in random forest regression, which has been taken as an argument in favor of using a high number of trees. Recently, research by Probst and Boulesteix (2018) on this issue showed that a high, but computationally feasible, number of trees is recommended, as long as classic mean loss error measures are used.

Taking into account these considerations, for the forecasting exercise, the parameters were set in the following way:

- *Number of trees*: It was set at 100 for all forecasting exercises.
- *Size of terminal nodes*: 5 observations (usual value in most applications).
- *Number of variables in each tree:* this parameter was calibrated for each forecasting exercise (see Section 3.3), using the *tuneRF* function of the *randomForest* package, which finds the number of variables that minimizes *out-of-bag* error.

3.4.5. LSTM model

Hochreiter and Schmidhuber (1997) introduced long short-term memory (LTSM) networks, which have been shown to improve accuracy relative to traditional neural network models. These models are a type of recurrent neural network, different from traditional neural networks in that they include a feedback loop between past decisions and the current outcome. Thus, their functional architecture allows us to solve the vanishing gradient problem in the updating rule, which enables us to handle longer-run dependencies.

For estimation and forecasting, we used the *Keras* package, an application programming interface for R. This package runs on *Tensorflow* in Python, which serves as the "backend" engine (data access). We estimated a simple model with one layer, using as features all of the variables included in Table 1, as well as 12 of their lags. Before training the model, data were differenced and scaled to the range of values of the activation function, which is [-1,1] with the sigmoid or hyperbolic tangent. After training and forecasting, the differencing and scaling were reverted to obtain the forecasts in levels.

When training the model in *Keras*, the metric for evaluation was accuracy as measured by the MSE. We used the Adam optimizer by Kingma and Ba (2017), with a default learning rate of 0.02, default learning rate decay (1e-6), and hyperbolic tangent as the

[1] A description of the model will feature in BCCR (forthcoming). An earlier version of the model is Muñoz and Tenorio (2008).

[2] With this method, the forecast is given by $\hat{P} = \sum_{i=1}^{K} w_i(x) P_i / \sum_{i=1}^{K} w_i(x)$, where $w_i(x)$ is the inverse of the distance of each neighbor i.

[3] Tuarob et al. (2017) and Ty et al. (2019) are examples of machine learning exercises where the time-series input data is prepared through time delay embedding.

activation function. We set the batch size to 1, so that the algorithm implemented was stochastic gradient descent, and, to retain temporal dependencies in the training sample, we set the option *shuffle* to false.

3.4.6. Univariate methods

We compared the performance of the methods discussed above with that of the suite of univariate forecasts currently in use at the BCCR. Fuentes and Rodríguez (2016) estimated univariate models for forecasting short-term inflation in Costa Rica, and analyzed the properties of 14 specifications that consider several assumptions about the functional form and the statistical properties of the data-generating process. The models selected are three ARMA specifications (with homoscedastic variance, with GARCH effects, or with stochastic volatility), and three variants of unobserved components models: a simple version, a second one with stochastic volatility, and a third with stochastic volatility and MA errors. In the evaluation exercise, we included the simple average of these methods.

3.5. Evaluation criteria

In this section, we present a summary of the evaluation criteria applied in this study. These are derived from a long literature that includes Mincer and Zarnowitz (1969), Granger and Newbold (1973), Stekler (1991), and Diebold and López (1996), who presented four properties that must be met by optimal forecasts under quadratic loss. West (2006) presented a useful review of the relevant literature. Patton and Timmermann (2007) showed that, under asymmetric loss functions, these standard properties of optimal forecasts can be invalid. In this exercise, we implicitly assumed a symmetric loss function for the Central Bank on deviations from the inflation forecasts. Presentation of the criteria considered in this study closely follows Fuentes and Rodríguez (2016).

a- *Unbiasedness*

If a forecast does not systematically under- or overestimate the true value of the target variable, forecast errors should have zero mean. Mincer and Zarnowitz (1969) proposed an unbiasedness test based on the regression in levels, given by:

$$y_{t+s} = \alpha + \beta \hat{y}_{t+s} + \varepsilon_{t+s}$$

where $\hat{y}_{t+s}$ is the forecast with horizon s for the target variable, and y_{t+s} is its real value. The null hypothesis of unbiasedness is $\alpha = 0$ and $\beta = 1$. In this study, we used Wald tests.

b- *Error correlation*

Diebold and López (1996) showed that forecast errors e_{t+h} for an optimal forecast $\hat{y}_{t+h}$are white noise for horizon $h = 1$ and at most, they follow an MA(h-1) process for $h > 1$. They recommended the modified Wilcoxon signed-rank test proposed by Dufour (1981) for autocorrelation in the case of $h = 1$, and for $h > 1$, the Cumby and Huizinga (1992) test, the null hypothesis of which is that errors follow an MA(q) process, with $0 \leq q \leq h$-1, and the alternative hypothesis of which is that $q > h$.

c- *Forecast accuracy*

We present RMSE and Theil inequality coefficients, and the statistical significance of differences in forecasting ability under quadratic loss for pairs of models were tested using the modified Diebold-Mariano test by Harvey et al. (1997).[4] Additionally, we computed the percentage of changes in direction of the inflation that were correctly predicted.

d- *Forecast error variances*

Diebold and López (1996) stated that optimal forecasts under quadratic loss should have errors with variance that is non-decreasing as the forecast horizon increases. This basically reflects that the uncertainty of the forecast should decrease, and not increase, when more information is available (as is the case for shorter horizons). To verify this property, we conducted F tests for the difference of variances.

e- *Forecast encompassing*

Tests of forecast encompassing are used to determine whether a forecast contains all of the information used in one or several alternative forecasts. This is useful, for example, to decide whether there is an information gain in combining several of them. We used the Chong and Hendry (1986) test for encompassing of point forecasts under quadratic loss function. From the regression:

$$y_{t+s} = \beta_0 + \beta_1 \hat{y}_{t+s,1} + \beta_2 \hat{y}_{t+s,2} + \varepsilon_{t+s}$$

we concluded that Forecast 1 encompasses Forecast 2 if the joint null$\begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$ is not rejected, whereas Forecast 2 encompasses Forecast 1 if $\begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ is not rejected. Otherwise, both forecasts contain useful information.

4. Evaluation results and discussion

Means and standard deviations of the forecasts are presented in the following table. Fig. 4 in the annex presents the forecasts for each horizon, along with observed values. The results of the evaluation are presented in the following sections.

[4] The statistics are $DM = \frac{\bar{d}}{s_d}$ y $HLN = \left(\frac{T+1-2k+k(k-1)/T}{T}\right)^{1/2} DM$, where $\bar{d}$ is the mean of the squared difference of forecast errors, s_d is a consistent estimate of its standard deviation, k is the forecast horizon, and T is the number of observations.

Table 2
Descriptive statistics for forecasts.

	Mean				Standard deviation			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Univariate KNN	0.0185	0.0181	0.0181	0.0161	0.0057	0.0058	0.0063	0.0064
KNN with exogenous variables	0.0434	0.0433	0.0417	0.0314	0.0130	0.0158	0.0163	0.0213
Random forests	0.0236	0.0278	0.0346	0.0407	0.0052	0.0046	0.0040	0.0033
Extreme gradient boosting	0.0185	0.0181	0.0181	0.0161	0.0058	0.0068	0.0066	0.0110
LSTM model	0.0184	0.0192	0.0193	0.0212	0.0055	0.0045	0.0043	0.0019
Average of univariate methods	0.0190	0.0197	0.0201	0.0216	0.0057	0.0060	0.0066	0.0076

Source: Own elaboration.
Note: The mean and standard deviation of observed inflation values for the forecast sample were 0.0190 and 0.0051.

4.1. Unbiasedness

The LSTM model is clearly the best performer, as the hypothesis of unbiasedness is rejected only at the $h = 1$ forecast horizon. Unbiasedness is rejected in all other forecasts, except for the average of the univariate forecasts at $h = 1$.

Table 3
Mincer and Zarnowitz (1969) unbiasedness test
P-value for the joint Wald test.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Univariate KNN	0.0299	0.0000	0.0000	0.0000
KNN with exogenous variables	0.0000	0.0000	0.0000	0.0000
Random forests	0.0000	0.0000	0.0000	0.0000
Extreme gradient boosting	0.0097	0.0000	0.0000	0.0000
LSTM model	0.0454	**0.0554**	**0.0503**	**0.7245**
Average of univariate methods	**0.0631**	0.0000	0.0000	0.0000

Source: Own elaboration.
Note: *P*-values in bold if null hypothesis that the forecast is unbiased is not rejected at 5% significance.

4.2. Forecast error correlations

In most cases, forecast errors complied with the desirable properties established by Diebold and López (1996). For horizons $h >$ 1, Cumby and Huizinga tests showed that in all cases the null that the errors follow, at most, an MA(h-1) process is not rejected. Additionally, for the errors of the univariate KNN, extreme, gradient boosting, the LSTM model and the average of the univariate forecasts there is no statistical evidence of autocorrelation for the errors at the 1-month horizon.

Table 4
Autocorrelation tests for forecast errors
P values.

	$h = 1$ Wilcoxon/Dufour	$h = 3$ Cumby and Huizinga (1992)	$h = 6$	$h = 12$	
Univariate KNN	**0.1618**	**0.0740**	**0.3841**	**0.9331**	**0.3173**
KNN with exogenous variables	0.0000	0.0412	**0.4250**	**0.7268**	**0.3173**
Random forests	0.0000	**0.0977**	**0.1797**	**0.4401**	**0.4167**
Extreme gradient boosting	**0.3746**	**0.1433**	**0.4072**	**0.1934**	**0.3173**
LSTM model	**0.3603**	**0.2642**	**0.1524**	**0.5528**	**0.3173**
Average of univariate methods	**1.0000**	**0.4717**	**0.1933**	**0.5658**	**0.3173**

Note: *P*-values for Cumby and Huizinga (1992) statistics for the null of errors following a MA(h-1) versus alternative of MA(h). For $h = 1$, the Wilcoxon test proposed by Dufour (1981) was used (null is process is white noise). *P*-values in bold if null is not rejected.

4.3. Forecast accuracy

Table 5 shows the RSME and the Theil coefficient for the forecasts. At all horizons, the LSTM model produces the most accurate predictions, followed by the average of the univariate methods, the forecasts of the univariate KNN, and those of extreme gradient boosting and random forests. The forecasts computed by KNN with exogenous variables perform poorly.

Table 5
RMSE and Theil inequality coefficient for forecasts.

	RMSE				Theil			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Univariate KNN	0.0039	0.0062	0.0067	0.0100	0.1006	0.1587	0.1676	0.25743
KNN with exogenous variables	0.0267	0.0277	0.0256	0.0244	0.4111	0.4174	0.3910	0.41104
Random forests	0.0057	0.0089	0.0145	0.0193	0.1295	0.1834	0.2612	0.30792
Extreme gradient boosting	0.0044	0.0072	0.0072	0.0144	0.1130	0.1851	0.1819	0.38501
LSTM model	0.0039	0.0036	0.0042	0.0024	0.1003	0.0892	0.1027	0.05648
Average of univariate methods	0.0032	0.0061	0.0065	0.0094	0.0805	0.1487	0.1552	0.21178

Source: Own elaboration.
Note: Theil values closer to zero indicate more accuracy.

Table 6
Prediction of the direction of changes in inflation
In bold if greater than 50%.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Univariate KNN	**0.5600**	**0.6087**	**0.5500**	0.4286
KNN with exogenous variables	**0.6400**	**0.6522**	0.3500	0.3571
Random forests	**0.6800**	**0.6522**	**0.6000**	**0.7143**
Extreme gradient boosting	0.4800	**0.6522**	**0.5500**	0.4286
LSTM model	**0.6000**	**0.5652**	**0.5500**	0.5000
Average of univariate methods	0.4800	**0.6087**	0.4500	0.2857

Source: Own elaboration.

Are these differences in forecasting ability statistically significant? Table 10 of the annex shows the p-values for the Harvey, Leybourne, and Newbold tests for each pair of forecasts at each horizon. At 5% significance, the tests show that the forecasts of LSTM and univariate KNN have the best performance, as those forecasts have the largest number of significant differences in comparison with the rest of methods, and few significant differences between them. In particular, for horizons of 1 to 6 months, we could not conclude that these two forecasts are significantly different.

For the shortest horizon, univariate forecasts are significantly more accurate than all methods, except univariate KNN and extreme gradient boosting. Next are LSTM model forecasts, which outperform 2 methods. Furthermore, at this horizon, the forecast obtained by random forests, extreme gradient boosting, and univariate KNN are not significantly different. At longer horizons, the performance of the LSTM model forecasts is very good: for horizons of 3 and 6 months, they outperform all methods except univariate KNN and extreme gradient boosting; and for $h = 12$, they are significantly more accurate than all the others. Extreme gradient boosting forecasts deteriorate sharply at $h = 12$. Overall, the LSTM model shows the best performance, followed by univariate KNN and extreme gradient boosting, which are very similar to one another, and random forests. KNN with exogenous variables is the worst performer, as it is significantly less accurate than all methods at most horizons.

Accuracy in the prediction of the direction of changes in inflation is summarized in Table 6. The good performance of the random forests forecasts is notable, as at all horizons, the direction of the change in inflation is correctly predicted in at least 60% of the cases. This can be assessed from Fig. 4 in the annex: The series of random forests forecasts closely follow the movements of inflation, although generally at different levels. Univariate KNN and the LSTM model also show good performance, but for horizons of 1, 3 and 6 months, whereas extreme gradient boosting has a modest performance at the 3-and 6-month horizons.

4.4. Encompassing tests

Tables 11 and 12 of the annex present the results of Chong and Hendry encompassing tests for each pair of forecasts at each horizon. At the shortest horizon, univariate-methods forecasts encompass those of KNN and boosting. At higher horizons, the LSTM model forecasts again show a better performance, as they are the only ones to encompass other forecasts, particularly for $h = 3$ and $h = 12$, when they encompass 4 and 5 of the 5 forecasts in comparison, respectively.

4.5. Forecast error variance

Most of the forecasts evaluated have variances that increase with the horizon, with the exceptions of the LSTM model and KNN with exogenous variables. However, as seen in Table 13 of the annex, in most cases, it cannot be concluded that these variances at different forecast horizons are significantly different from each other.

Table 7
Variances of forecast errors.

	$h=1$	$h=3$	$h=6$	$h=12$
Univariate KNN	0.00002	0.00004	0.00004	0.00007
KNN with exogenous variables	0.00012	0.00023	0.00021	0.00054
Random forests	0.00001	0.00002	0.00001	0.00001
Extreme gradient boosting	0.00002	0.00005	0.00005	0.00011
LSTM model	0.00002	0.00001	0.00002	0.00001
Average of univariate methods	0.00001	0.00004	0.00004	0.00010

Source: Own elaboration.

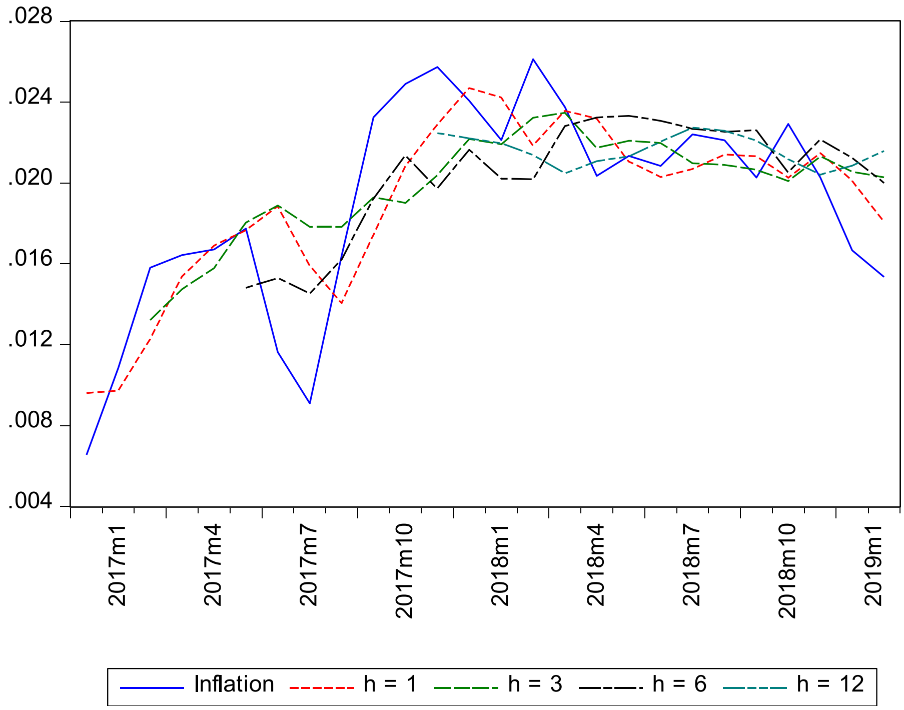

Fig. 2. Combination of forecast at several horizons
Source: Own elaboration.

4.6. Combination of forecasts

In this section, we evaluate a combination of the forecasts with the best individual performance. This exercise follows the literature on traditional forecast combination, (see the reviews by Granger (1989), Timmermann (2006), and Wallis (2011), among others). Ideally, the forecasts resulting from combination should exhibit superior, or at least similar, properties to the individual forecasts. In particular, it must be noted that, of the methods assessed, only the LSTM model produces forecasts that are not biased, so it would be desirable to obtain a combination that shows this property, while keeping the other desirable properties of the forecasts in its calculation.

From the analysis in Sections 4.1–4.5, summarized in the rankings of Table 15 in the annex, it can be concluded that the methods that consistently show desirable properties are LSTM and univariate KNN, although their ability to predict the direction of changes in inflation, while adequate, is lower than that of random forests. We decided to combine the forecasts of these three methods with the procedure suggested by Capistrán and Timmermann (2009). With this method, the combination is obtained as the fitted values of a regression of the true values of the target variable on a constant and the simple average of the individual forecasts. The results of the combination are shown in Fig. 2 and those of its evaluation in Tables 8 and 9.

In general, the results of the combination are very satisfactory. The null of unbiasedness is not rejected at all horizons, their forecast errors are not autocorrelated at the $h = 1$ horizon and at most, they follow a MA(h-1) process for $h > 1$, and it has an ability to predict the direction of changes in inflation equal to or higher than 50% at all horizons, but particularly those that are longer. Furthermore, at horizons of 3, 6, and 12 months, the combination is significantly more accurate than the individual forecasts, with the

Table 8
Evaluation of the forecast combination
Several tests.

	P-value for null hypothesis of unbiasedness	Prediction of direction of changes in inflation (% success)	*P*-values for autocorrelation tets [1/]	Forecast error variances
***h* = 1**	1.00	0.6400	0.2940	0.000010
***h* = 3**	1.00	0.6957	0.2832	0.000013
***h* = 6**	1.00	0.5000	0.5932	0.000012
***h* = 12**	1.00	0.5714	0.3173	0.000008

Source: Own elaboration.
[1/] $h = 1$: Wilcoxon, others: Cumby and Huizinga.

Table 9
P-values for the Harvey et al., (1997) test Combination vs. individual methods
In bold if the forecast error of the column is significantly lower than that of the row.

	KNN univariado Combination $h = 1$	Bosques aleatorios $h = 3$	LSTM $h = 6$	$h = 12$
Univariate KNN	0.0847	**0.0490**	**0.0234**	**0.0457**
KNN with exogenous variables	**0.0000**	**0.0000**	**0.0003**	**0.0013**
Random forests	**0.0091**	**0.0015**	**0.0000**	**0.0000**
Extreme gradient boosting	0.0999	**0.0037**	**0.0226**	0.0526
LSTM model	**0.0112**	0.4333	0.1966	0.6729
Average of univariate methods	0.4132	**0.0327**	**0.0029**	**0.0288**

Source: Own elaboration.

notable exception of the LSTM model. For the shortest horizon, however, the combination is not more accurate than the univariate KNN forecast, extreme gradient boosting, nor the average of the univariate methods.[5]

5. Conclusions

The goal of this study was to perform a first evaluation of the capacity of machine learning methods to forecast inflation in Costa Rica. Forecasts were computed with 5 methods: two variants of KNN, random forests, extreme gradient boosting and a *long short-term memory* (LSTM) model. Their properties were evaluated following the literature on optimal forecasts, and their performance was compared with that of the average of univariate methods currently in use at the Central Bank of Costa Rica.

We found that the forecasts with the best performance are those from the LSTM model, univariate KNN, and to a lesser extent, random forests and extreme gradient boosting. In particular, LSTM model forecasts are not biased for horizons longer than 1 month, show more accuracy than the rest of the forecasts, and encompass most other forecasts.

A combination of forecasts improves the performance in comparison with individual forecasts at all horizons, and crucially, also outperforms the forecasts from univariate methods. The combination is not biased, their errors do not show undesirable correlation patterns, and it improves forecasting ability at all horizons, both for the level of inflation and for the direction of its changes.

Given these results, we consider that the implementation of machine learning methods for forecasting at the BCCR is a promising endeavor. A first line of work could be the improvement in the application of methods that underperformed in the study, as well as the potential extension of the work to include additional ML methods.

Declarations of Competing Interest

None

Disclaimer

The ideas expressed in this paper are those of the author and not necessarily represent the view of the Central Bank of Costa Rica.

Annex

Figs. 3 and 4. Tables 10-16.

[5] The mean of these three individual forecasts was also evaluated, but their properties were not satisfactory. In particular, it is not unbiased at all horizons, it does not improve on the accuracy of individual forecast, and is considerably less accurate than the CT combination presented. Several CT combinations, including extreme gradient boosting, were also considered, resulting in undesirable properties (bias at longer horizons, poor prediction of direction of changes).

Fig. 3. Variables used
Source: Own elaboration.

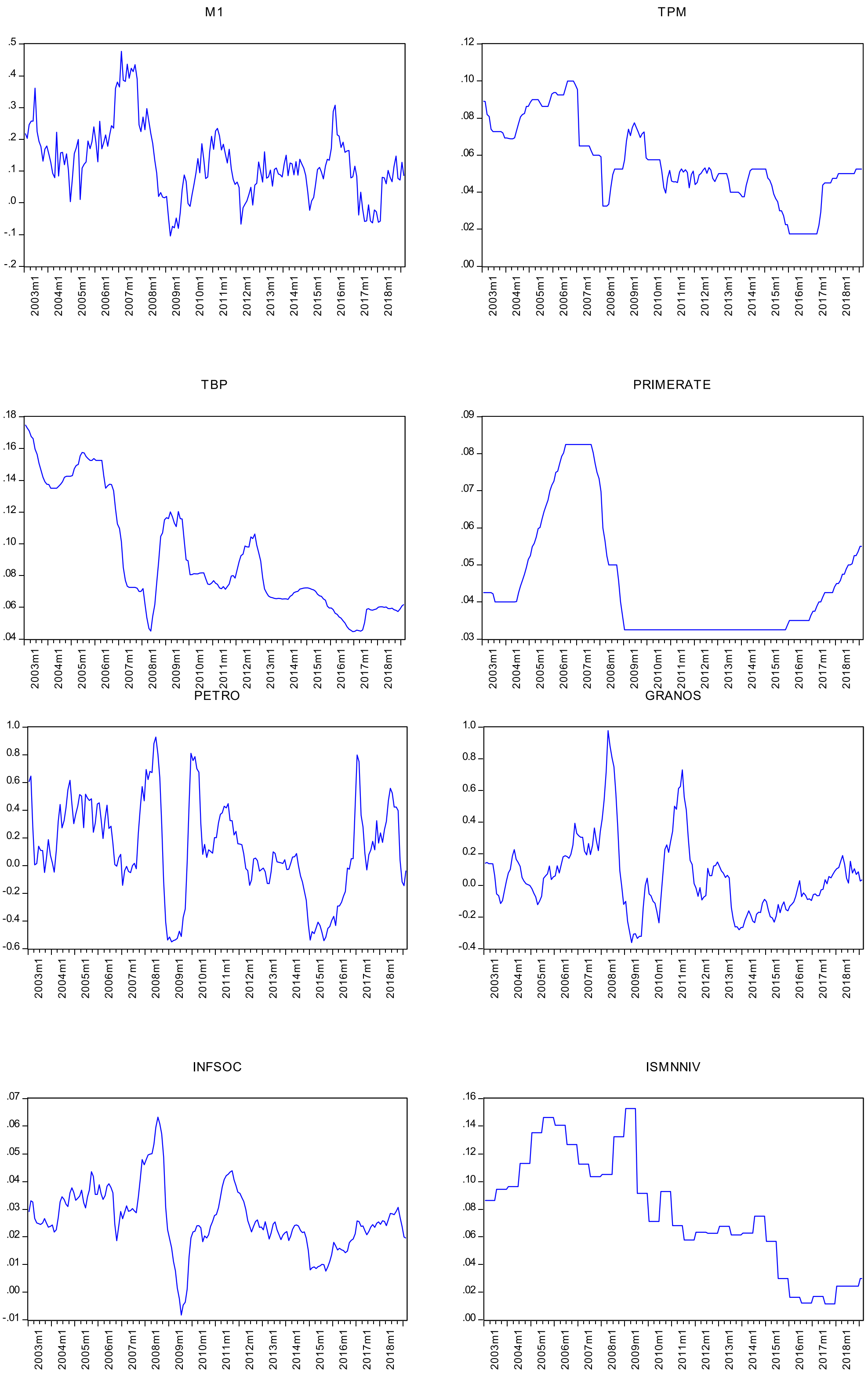

Fig. 3. Continued

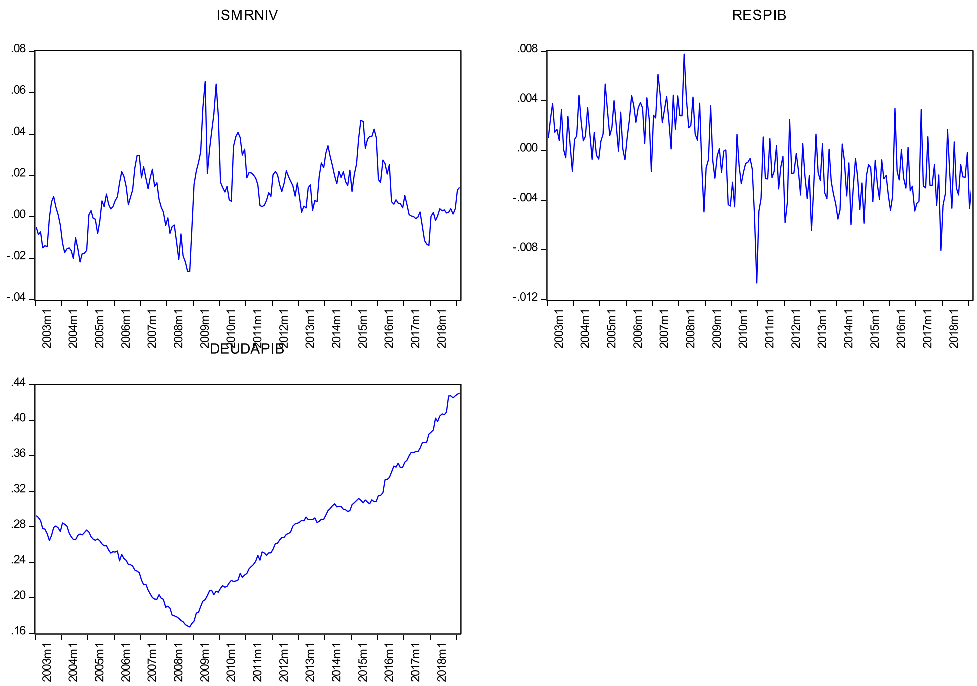

Fig. 3. Continued

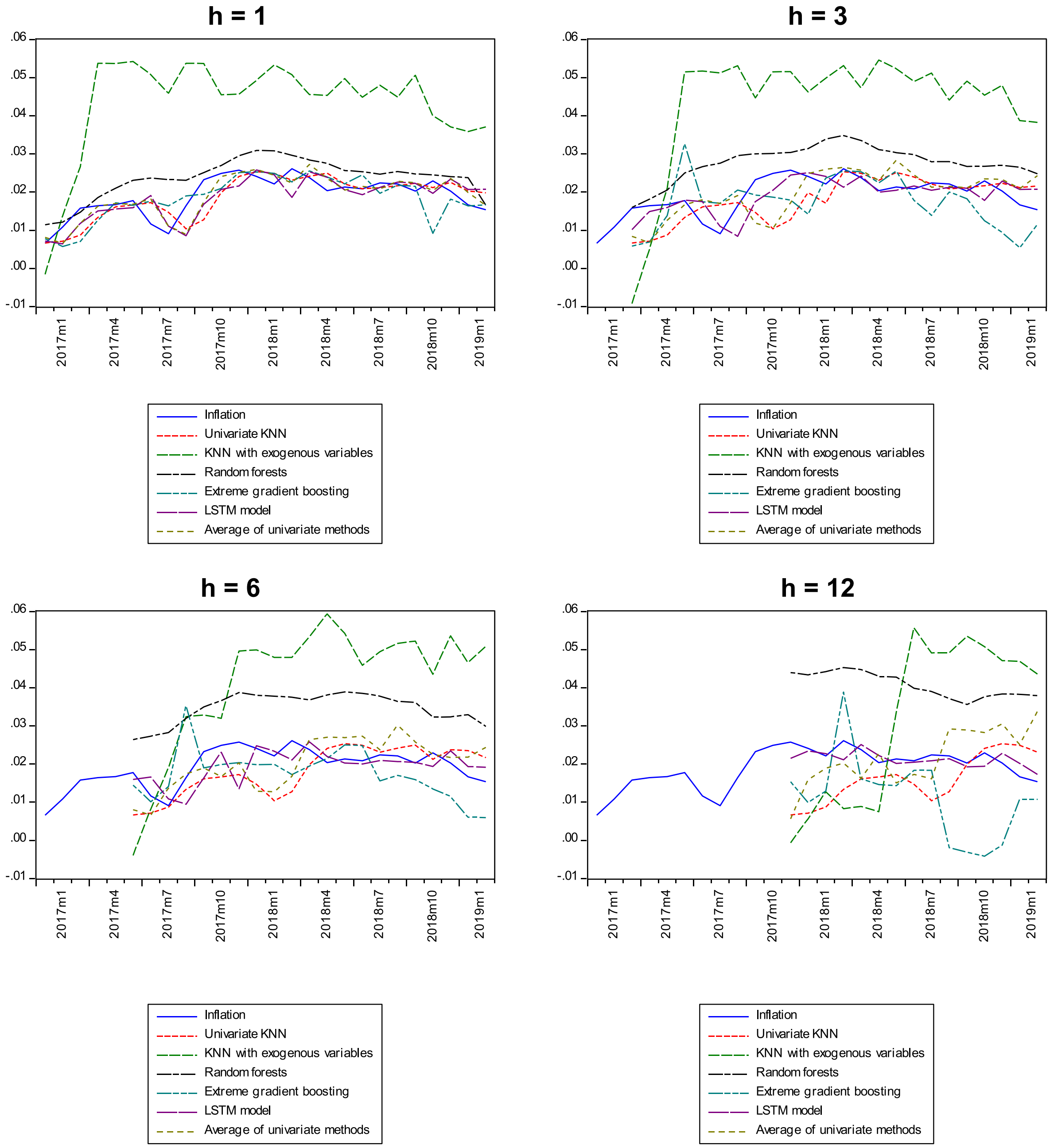

Fig. 4. Forecasts at several horizons
Source: Own elaboration.

Table 10
P-values for Leybourne y Newbold (1997) test
In bold if the forecast error of the column is significantly lower than that of the row.

$h = 1$	Univariate KNN	KNN with exogenous variables	Random forests	Extreme gradient boosting	LSTM model	Average of univariate methods
Univariate KNN		1.0000	0.9252	0.6721	0.4751	0.0743
KNN with exogenous variables	**0.0000**		**0.0000**	**0.0000**	**0.0000**	**0.0000**
Random forests	0.0748	1.0000		0.1709	**0.0484**	**0.0172**
Extreme gradient boosting	0.3279	1.0000	0.8291		0.3043	0.0998
LSTM model	0.5249	1.0000	0.9516	0.6957		**0.0057**
Average of univariate methods	0.9257	1.0000	0.9828	0.9002	0.9943	
Cases where column error is lower than row error:	1	0	1	1	2	3
$h = 3$						
	Univariate KNN	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		1.0000	0.8898	0.7423	0.0579	0.3927
KNN with exogenous variables	**0.0000**		**0.0000**	**0.0000**	**0.0000**	**0.0000**
Random forests	0.1102	1.0000		0.1781	**0.0031**	0.0762
Extreme gradient boosting	0.2577	1.0000	0.8219		**0.0055**	0.2259
LSTM model	0.9421	1.0000	0.9969	0.9945		0.9645
Average of univariate methods	0.6073	1.0000	0.9238	0.7741	**0.0355**	
Cases where column error is lower than row error:	1	0	1	1	4	1
$h = 6$						
	Univariate KNN	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.9995	1.0000	0.5882	0.0535	0.3980
KNN with exogenous variables	**0.0005**		**0.0041**	**0.0007**	**0.0004**	**0.0004**
Random forests	**0.0000**	0.9959		**0.0000**	**0.0000**	**0.0000**
Extreme gradient boosting	0.4118	0.9993	1.0000		0.0506	0.3657
LSTM model	0.9465	0.9996	1.0000	0.9494		0.9721
Average of univariate methods	0.6020	0.9996	1.0000	0.6343	**0.0279**	
Cases where column error is lower than row error:	2	0	1	2	3	2
$h = 12$						
	Univariate KNN	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.9903	0.9995	0.7712	**0.0420**	0.4078
KNN with exogenous variables	**0.0097**		0.1374	**0.0046**	**0.0013**	**0.0031**
Random forests	**0.0005**	0.8626		0.1543	**0.0000**	**0.0001**
Extreme gradient boosting	0.2288	0.9954	0.8457		**0.0495**	0.1800
LSTM model	0.9580	0.9987	1.0000	0.9505		0.9690
Average of univariate methods	0.5923	0.9969	0.9999	0.8200	**0.0310**	
Cases where column error is lower than row error:	2	0	0	1	5	2

Table 11
Results of Chong and Hendry (1986) encompassing tests
For each pair of forecasts the table shows which forecast encompasses the other.

$h = 1$	Univariate KNN	KNN with exogenous variables	Random forests	Extreme gradient boosting	LSTM model	Average of univariate methods
Univariate KNN						Column
KNN with exogenous variables						Column
Random forests						
Extreme gradient boosting						Column
LSTM model						
Average of univariate methods	Row	Row		Row		
Cases where column forecast encompasses row forecast:	0	0	0	0	0	3
$h = 3$						
	Univariate KNN	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN					Column	
KNN with exogenous variables					Column	
Random forests						
Extreme gradient boosting					Column	
LSTM model	Row	Row		Row		Row
Average of univariate methods					Column	
Cases where column forecast encompasses row forecast:	0	0	0	0	4	0
$h = 6$						
	Univariate KNN	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN					Column	
KNN with exogenous variables						
Random forests						
Extreme gradient boosting						
LSTM model	Row					Row
Average of univariate methods					Column	
Cases where column forecast encompasses row forecast:	0	0	0	0	2	0
$h = 12$						
	Univariate KNN	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN					Column	
KNN with exogenous variables					Column	
Random forests					Column	
Extreme gradient boosting					Column	
LSTM model	Row	Row	Row	Row		Row
Average of univariate methods					Column	
Cases where column forecast encompasses row forecast:	0	0	0	0	5	0

Table 12
P-values of Chong and Hendry (1986) encompassing tests.

$h = 1$						
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 1\ 0)$	Univariate KNN	KNN with exogenous variables	Random forests	Extreme gradient boosting	LSTM model	Average of univariate methods
Univariate KNN		0.0214	0.0014	0.0374	0.0550	0.0001
KNN with exogenous variables	0.0000		0.0000	0.0000	0.0000	0.0000
Random forests	0.0000	0.0000		0.0000	0.0000	0.0000
Extreme gradient boosting	0.0009	0.0178	0.0000		0.0005	0.0000
LSTM model	0.0672	0.0233	0.0006	0.0318		0.0000
Average of univariate methods	0.0910	0.0544	0.0338	0.1365	0.0150	
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 0\ 1)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0009	0.0672	0.0910
KNN with exogenous variables	0.0214		0.0000	0.0178	0.0233	0.0544
Random forests	0.0014	0.0000		0.0000	0.0006	0.0338
Extreme gradient boosting	0.0374	0.0000	0.0000		0.0318	0.1365
LSTM model	0.0550	0.0000	0.0000	0.0005		0.0150
Average of univariate methods	0.0001	0.0000	0.0000	0.0000	0.0000	
$h = 3$						
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 1\ 0)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0000	0.0000	0.0000
KNN with exogenous variables	0.0000		0.0000	0.0000	0.0000	0.0000
Random forests	0.0000	0.0000		0.0000	0.0000	0.0000
Extreme gradient boosting	0.0000	0.0000	0.0000		0.0000	0.0000
LSTM model	0.1324	0.1361	0.0371	0.0726		0.1152
Average of univariate methods	0.0000	0.0000	0.0000	0.0000	0.0000	
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 0\ 1)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0000	0.1324	0.0000
KNN with exogenous variables	0.0000		0.0000	0.0000	0.1361	0.0000
Random forests	0.0000	0.0000		0.0000	0.0371	0.0000
Extreme gradient boosting	0.0000	0.0000	0.0000		0.0726	0.0000
LSTM model	0.0000	0.0000	0.0000	0.0000		0.0000
Average of univariate methods	0.0000	0.0000	0.0000	0.0000	0.1152	
$h = 6$						
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 1\ 0)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0000	0.0000	0.0000
KNN with exogenous variables	0.0000		0.0000	0.0000	0.0000	0.0000
Random forests	0.0000	0.0000		0.0000	0.0000	0.0000
Extreme gradient boosting	0.0000	0.0000	0.0000		0.0000	0.0000
LSTM model	0.1105	0.0358	0.0000	0.0035		0.0868
Average of univariate methods	0.0000	0.0000	0.0000	0.0000	0.0000	
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 0\ 1)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0000	0.1105	0.0000
KNN with exogenous variables	0.0000		0.0000	0.0000	0.0358	0.0000
Random forests	0.0000	0.0000		0.0000	0.0000	0.0000
Extreme gradient boosting	0.0000	0.0000	0.0000		0.0035	0.0000
LSTM model	0.0000	0.0000	0.0000	0.0000		0.0000
Average of univariate methods	0.0000	0.0000	0.0000	0.0000	0.0868	
$h = 12$						
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 1\ 0)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0000	0.0000	0.0000
KNN with exogenous variables	0.0000		0.0000	0.0000	0.0000	0.0000
Random forests	0.0000	0.0000		0.0000	0.0000	0.0000
Extreme gradient boosting	0.0000	0.0000	0.0000		0.0000	0.0000
LSTM model	0.0693	0.4695	0.3764	0.0001		0.1260
Average of univariate methods	0.0000	0.0000	0.0000	0.0000	0.0000	
H$_0$: $(\beta_0\ \beta_1\ \beta_2) = (0\ 0\ 1)$	**Univariate KNN**	**KNN with exogenous variables**	**Random forests**	**Extreme gradient boosting**	**LSTM model**	**Average of univariate methods**
Univariate KNN		0.0000	0.0000	0.0000	0.0693	0.0000
KNN with exogenous variables	0.0000		0.0000	0.0000	0.4695	0.0000
Random forests	0.0000	0.0000		0.0000	0.3764	0.0000
Extreme gradient boosting	0.0000	0.0000	0.0000		0.0001	0.0000
LSTM model	0.0000	0.0000	0.0000	0.0000		0.0000
Average of univariate methods	0.0000	0.0000	0.0000	0.0000	0.1260	

Table 13
Tests of equality of variances
P-values of tests for forecasts at horizons indicated in row/column.

Univariate KNN					KNN with exogenous variables				
	1	**3**	**6**	**12**		**1**	**3**	**6**	**12**
1		**0.0398**	**0.0230**	**0.0007**	**1**		0.1357	0.2013	**0.0013**
3	**0.0398**		0.7835	0.1320	**3**	0.1357		0.8693	0.0629
6	**0.0230**	0.7835		0.2265	**6**	0.2013	0.8693		0.0531
12	**0.0007**	0.1320	0.2265		**12**	**0.0013**	0.0629	0.0531	
Random forests					**Extreme gradient boosting**				
	1	**3**	**6**	**12**		**1**	**3**	**6**	**12**
1		0.3090	0.4251	0.4631	**1**		**0.0360**	0.0520	**0.0002**
3	0.3090		0.0864	0.1250	**3**	**0.0360**		0.9243	0.0774
6	0.4251	0.0864		0.9788	**6**	0.0520	0.9243		0.0744
12	0.4631	0.1250	0.9788		**12**	**0.0002**	0.0774	0.0744	
LSTM model					**Average of univariate methods**				
	1	**3**	**6**	**12**		**1**	**3**	**6**	**12**
1		0.6502	0.7874	0.0759	**1**		**0.0021**	**0.0009**	**0.0000**
3	0.6502		0.4876	0.1593	**3**	**0.0021**		0.7385	0.0513
6	0.7874	0.4876		0.0534	**6**	**0.0009**	0.7385		0.1137
12	0.0759	0.1593	0.0534		**12**	**0.0000**	0.0513	0.1137	

Table 14
Software used.

Method	Software	Details
Univariate KNN	R	*Package tsfknn* by Martínez et al. (2017) for the application of KNN to time series.
KNN with exogenous variables	MATLAB	Function *knnsearch* was used to find the neareast neighbours
Random forests	R	Package *randomForest* by Breiman et al. (2018), for the generation of random forests for regression and classification. Package *dyn* by Grothendieck (2018), that allows the creation of interfaces with several functions that do regression, *randomForests* among them. This package makes possible to use those functions with time series data, including specifications with lags and differences.
Extreme gradient boosting	R	Package *xgboost* by Chen et al. (2015).
LSTM model	R	*Keras* package, requires TensorFlow and Python.
Average of univariate methods	MATLAB, Eviews	

Table 15
Ranking of methods according to forecast evaluation criteria.

Unbiasedness (Table 3)		Forecast accuracy			
		Levels (Table 10)		*Direction of changes* (Table 6)	
Rank	*Model*	*Rank*	*Model*	*Rank*	*Model*
1	LSTM model	1	LSTM model	1	Random forests
2	Average of univariate methods	2	Average of univariate methods	2	LSTM model
3	Univariate KNN	3	Univariate KNN	3	Univariate KNN
3	Extreme gradient boosting	4	Extreme gradient boosting	4	Extreme gradient boosting
3	Random forests	5	Random forests	5	KNN with exogenous variables
3	KNN with exogenous variables	6	KNN with exogenous variables	6	Average of univariate methods
Encompassing (Tables 11 & 12)		**Forecast errors**			
		Correlations (Table 4)		***Variances (Table 7)***	
Rank	*Model*	*Rank*	*Model*	*Rank*	*Model*
1	LSTM model	1	LSTM model	1	Average of univariate methods
2	Average of univariate methods	1	Average of univariate methods	1	Univariate KNN
3	Univariate KNN	1	Univariate KNN	1	Extreme gradient boosting
4	Extreme gradient boosting	1	Extreme gradient boosting	1	Random forests
5	Random forests	2	Random forests	2	LSTM model
6	KNN with exogenous variables	3	KNN with exogenous variables	2	KNN with exogenous variables

Table 16
Links to data sources

Name	Source	Link
IPC	National Institute of Statistics and Censuses (INEC)	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%202732
EXPINF12	Central Bank of Costa Rica (BCCR)	https://gee.bccr.fi.cr/indicadoreseconomicos/Cuadros/frmVerCatCuadro.aspx?idioma=2&CodCuadro=%20759
TCN	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/Cuadros/frmVerCatCuadro.aspx?idioma=2&CodCuadro=%20748
TCR_M	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/Cuadros/frmVerCatCuadro.aspx?idioma=2&CodCuadro=%202501
IMAETC	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%203478
ICFNIV	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%204986
CREDPRIVSF	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%202449
BASEM	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%2058
M1	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%20164
TPM	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%20779
TBP	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%2017
PRIMERATE	U.S. Federal Reserve	https://fred.stlouisfed.org/series/MPRIME
PETRO	Pink Sheet, World Bank	https://www.worldbank.org/en/research/commodity-markets
GRANOS	Pink Sheet, World Bank	https://www.worldbank.org/en/research/commodity-markets
INFSOC	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%202499
ISMNNIV	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%20295
ISMRNIV	BCCR	https://gee.bccr.fi.cr/indicadoreseconomicos/cuadros/frmvercatcuadro.aspx?idioma=2&codcuadro=%20333
RESPIB	BCCR, with data from the Ministry of Finance	https://www.hacienda.go.cr/contenido/139-cifras-mensuales-de-ingresos-gastos-y-financiamiento-del-gobierno-central
DEUDAPIB	BCCR, with data from the Ministry of Finance	https://www.hacienda.go.cr/contenido/12520-gobierno-central-detalle

References

Agnon, Y., Golan, A., Shearer, M., 1999. Nonparametric, nonlinear, short-term forecasting: theory and evidence for nonlinearities in the commodity markets. Econ. Lett. 65 (3), 293–299.
Athey, S., Imbens, G.W., 2019. Machine learning methods economists should know about. Annu. Rev. Econ. 11 (1), 685–725.
Bajari, P., Nekipelov, D., Ryan, S.P., Yang, M, 2015. Machine learning methods for demand estimation. Am. Econ. Rev. 105 (5), 481–485.
Barkoulas, J., Baum, C.F., Chakraborty, A., 1996. Nearest-Neighbor Forecasts of U.S. Interest Rates. Boston College Department of Economics Boston College Working Papers in Economics 313Revised in 2003.
Bergmeir, C., Benítez, J.M., 2012. On the use of cross-validation for time series predictor evaluation. Inf. Sci. Ny 191, 192–213.
Biau, O., D'Elia, A., 2011. Euro Area GDP Forecasting Using Large Survey Datasets. A Random Forest Approach.. EWP Euroindicators working papers.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.
Boelaert, J., Ollion, É, 2018. The great regression: machine learning, econometrics, and the future of quantitative social sciences. Rev. Fr. Sociol. 59 (3), 475–506. doi:10.3917/rfs.593.0475.
Breiman, L., 2001a. Random forests. Mach. Learn. 45, 5–32.
Breiman, L., Cutler, A., Liaw, A., Wiener, M., 2018. R Package "randomForest" for Classification and Regression,, pp. 6–14 version 4.
Capistrán, C., Timmermann, A., 2009. Forecast combination with entry and exit of experts. J. Bus. Econ. Stat. 428–440.
Chavarría Mejía, J.D, Chaverri Morales, C., 2015. Pronósticos de inflación mediante técnicas bayesianas. (Documento de Investigación, DI-05-2015). Department of Economic Research, Central Bank of Costa Rica.
Chen, T, Guestrin, C, 2016. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '16), New York, NY, USA. ACM, pp. 785–794. doi:10.1145/2939672.2939785.
Chong, Y.Y., Hendry, D., 1986. Econometric evaluation of linear macro-economic models. Rev. Econ. Stud. 53, 671–690.
Cook, T, Hall, A, 2017. Macroeconomic Indicator Forecasting with Deep Neural Networks. Federal Reserve Bureau of Kansas City RWP-17-11.
Cumby, R.E., Huizinga, J., 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regression. Econometrica 60, 185–195.
David, B., 2017. Model economic phenomena with CART and Random Forest algorithms. Economix Working Paper 2017-46. Université Paris Nanterre.
Diebold, F., Nason, J., 1990. Nonparametric exchange rate prediction? J. Int. Econ. 28, 315–332.
Diebold, F.X., López, J.A., 1996. Forecast evaluation and combination. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics. Elsevier, pp. 241–268.
Dufour, J.M., 1981. Rank tests for serial dependence. me Ser. Anal. 2 (3), 117–128.
Esquivel Monge, M., 2007. Performance of Artificial Neural Networks in Forecasting Costa Rican Inflation. Department of Economic Research, Central Bank of Costa Rica DEC-DIE-029-2009-DI..
Fernández-Rodríguez, F., Sosvilla-Rivero, S., Andrada-Felix, J., 1999. Exchange-rate forecasts with simultaneous nearest-neighbour methods: evidence from the EMS. Int. J. Forecast. 15, 383–392.
Fernández-Rodríguez, F., González-Martel, C., Sosvilla-Rivero, S., 2000. On the profitability of technical trading rules based on artificial neural networks: evidence from the Madrid stock market. Econ. Lett. 69 (1), 89–94.
Freund, Y., Schapire, R., 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In: Paul, V (Ed.), Computational Learning Theory. Springer, Berlin, pp. 23–37.
Freund, Y., Schapire, R., 1996. Experiments with a new boosting algorithm. In: Proceedings of the Thirteenth International Conference on Machine Learning, Morgan Kaufmann: Burlington, MA, pp. 148–156.
Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.
Friedman, J., Hastie, T, Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. Ann. Stat. 28, 337–407.
Fuentes, M., Rodríguez-Vargas, A., 2016. Pronósticos Univariados de Inflación Para Costa Rica con Volatilidad Estocástica y Efectos GARCH. Department of Economic Research, Central Bank of Costa Rica (Documento de Investigación No. 04-2016).
González, S., 2000. Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models. Finance Canada Working Paper 2000-07.
Granger, C.W.J., 1989. Invited review combining forecasts—twenty years later. J. Forecast. 8 (3), 167–173.
Granger, C.W., Newbold, P., 1973. Some comments on the evaluation of economic forecasts. Appl Econ. 5, 35–47.
Grothendieck, G., 2018. Paquete "Dyn" Para R Para Regresión con Series de Tiempo Versión 0.2-9.6, 18 de marzo de 2018.
Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. Int. J. Forecast. 13, 281–291.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition Springer, New York Springer Series in Statistics.
Hebb, D., 1949. The Organization of Behavior. Wiley, New York.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hyndman, R.J., Athanasopoulos, G, 2018. Forecasting: Principles and Practice, 2nd edition OTexts, Melbourne, Australia.
Kingma, D., Ba, J., 2017. Adam: A Method for Stochastic Optimization Eprint en arXiv:1412.6980 [cs.LG].
Lehmann, R., Wohlrabe, K., 2016. Looking into the black box of boosting: the case of Germany. Appl. Econ. Lett. 23 (17), 1229–1233.
Lisi, F., Medio, A., 1997. Is a random walk the best exchange rate predictor? Int. J. Forecast. 13, 255–267.
Lisi, F., Schiavo, R., 1999. A comparison between neural networks and chaotic models for exchange rate prediction. Comput. Stat. Data Anal. 30, 87–102.
Martínez, F., Frías, M.P., Pérez, M.D., Rivera, A.J., 2017. A methodology for applying k-nearest neighbor to time series forecasting. Artif. Intell. Rev. 1–19. doi:10.1007/s10462-017-9593-z.
Martínez, F., Frías, M.P., Charte, F., Rivera, A.J., 2019. Time series forecasting with KNN in R: the TSFKNN Package. R J 11 (2), 229–242.
Meade, M., 2002. A comparison of the accuracy of short term foreign exchange forecasting methods. Int. J. Forecast. 18, 67–83.
Mack, Y.P., 1981. Local properties of k-NN regression estimates. SIAM J. Algebr. Discret. Methods 2 (3), 311–323.
McCulloch, W., Pitts, W., 1943. A logical calculus of ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133.
Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), *Economic Forecasts and Expectations: Analyses of Forecasting Behavior* (NBER Studies in Business Cycles, Vol. 19). National Bureau of Economic Research, Cambridge, MA.
Mitchell, T., 1997. Machine Learning. McGraw-Hill, New York.
Mizrach, B., 1992. Multivariate nearest-neighbour forecasts of EMS exchange rates. J. Appl. Econ. 7, 151–163.
Moshiri, S., Cameron, N., 2000. Econometrics versus ANN model forecasting inflation. J. Forecast. 19 (3), 201–217.
Muñoz, E., Tenorio, E., 2008. El Modelo Macroeconómico de Proyección Trimestral del Banco Central de Costa Rica en la Transición a la Flexibilidad del Tipo de Cambio. Department of Economic Research, Central Bank of Costa Rica (Documento de Investigación DIE-08-2008-DI).
Nakamura, 2006. Inflation forecasting using a neural network. Econ. Lett. 86 (3), 373–378.
Nowman, B., Saltoglu, B., 2003. Continuous time and nonparametric modelling of U.S. interest rate models. Int. Rev. Financ. Anal. 12, 25–34.
Nikolopoulos, K.I, Babai, Z., Bozos, K., 2016. Forecasting supply chain sporadic demand with nearest neighbor approaches. Int. J. Prod. Econ. 177, 139–148.
Patton, A.J., Timmermann, A., 2007. Properties of optimal forecasts under asymmetric loss and nonlinearity. J. Econ. 140 (2), 884–918.
Probst, P., Boulesteix, A.L., 2018. To tune or not to tune the number of trees in random forest. J. Mach. Learn. Res. 18, 1–18.
Refenes, A.P., White, H., 1998. Neural networks and financial economics. Int. J. Forecast. 6 (17), 347–495.
Segal, M.R., 2004. Machine Learning Benchmarks and Random Forest Regression. UCSF: Center for Bioinformatics and Molecular Biostatistics Retrieved from https://escholarship.org/uc/item/35x3v9t4 .
Shalev-Shwartz, S., Ben-David, S., 2014. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, New York.

Solera, A., 2005. Pronóstico de Inflación en Costa Rica: Una Estimación con Redes Neuronales Artificiales. Department of Social Accounting, Central Bank of Costa Rica Documento de Trabajo. DCS-155-2005.

Stekler, H. O., 1991. Macroeconomic forecast evaluation techniques. Int. J. Forecast. 7, 375–384.

Stock, J.H., Watson, M., 1998. A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. NBER Working Paper 6607.

Swanson, N., White, H., 1995. A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. J. Bus. Econ. Stat. 13, 265–275.

Swanson, N., White, H., 1997. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. Rev. Econ. Stat. 79 (4), 540–555.

Tashman, L.J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. Int. J. Forecast. 16 (4), 437–450. https://ideas.repec.org/a/eee/intfor/v16y2000i4p437-450.html.

Timmermann, A., 2006. Forecast combinations. Handb. Econ. Forecast. 1, 135–196.

Tkacz, G., Hu, S., 1999. Forecasting GDP Growth Using Artificial Neural Networks. Bank of Canada Working Paper 99-3.

Tuarob, S., Tucker, C.S., Kumarab, S., Giles, C.L, Pincus, A.L, Conroy, D, Ram, N., 2017. How are you feeling? A personalized methodology for predicting mental states from temporally observable physical and behavioral information. J. Biomed. Inform. 68, 1–19.

Ty, A.J.A., Fang, Z., Gonzalez, R.A., Rozdeba, P.J., Abarbanel, H.D.I., 2019. Machine learning of time series using time-delay embedding and precision annealing. Neural Comput. 31 (10), 2004–2024.

Yakowitz, S., 1987. Nearest-neighbour methods for time series analysis. J. Time Ser. Anal. 8, 10–26.

Yakowitz, S., Karlsson, M., 1987. Nearest neighbour methods for time series, with application to rainfall/runoff prediction. In: Macneill, J.B., Umphrey, G.J. (Eds.), Stochastic Hydrology. D. Reidel Publishing Co., pp. 149–160.

Wallis, K.F., 2011. Combining forecasts–forty years later. Appl. Financ. Econ. 21 (1–2), 33–41.

West, K. D., 2006. Forecast Evaluation. In: Elliot, G., Granger, C.W., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier.

Wohlrabe, K., Buchen, T., 2014. Assessing the macroeconomic forecasting performance of boosting: evidence for the United States, the Euro area and Germany. J. Forecast. 33, 231–242.

Zeng, J., 2017. Forecasting aggregates with disaggregate variables: does boosting help to select the most relevant predictors? J. Forecast. 36, 74–90.