Performance evaluation of forecasting models based on time series and machine learning techniques: an application to light fuel consumption in Brazil

Received 1 March 2021
Revised 28 June 2021
Accepted 23 July 2021

Lucas Rodrigues
Luiz de Queiroz College of Agriculture – ESALQ/USP, Piracicaba, Brazil
Luciano Rodrigues
Master's Program in Agribusiness – FGV/EESP, São Paulo, Brazil, and
Mirian Rumenos Piedade Bacchi
Luiz de Queiroz College of Agriculture – ESALQ/USP, Piracicaba, Brazil

Abstract

Purpose – Fuel demand forecast is a fundamental tool to guide private planning actions and public policies aim to guarantee energy supply. This paper aims to evaluate different forecasting methods to project the consumption of light fuels in Brazil (fuel used by vehicles with internal combustion engine).

Design/methodology/approach – Eight different methods were implemented, besides of ensemble learning technics that combine the different models. The evaluation was carried out based on the forecast error for a forecast horizon of 3, 6 and 12 months.

Findings – The statistical tests performed indicated the superiority of the evaluated models compared to a naive forecasting method. As the forecast horizon increase, the heterogeneity between the accuracy of the models becomes evident and the classification by performance becomes easier. Furthermore, for 12 months forecast, it was found methods that outperform, with statistical significance, the SARIMA method, that is widely used. Even with an unprecedented event, such as the COVID-19 crisis, the results proved to be robust.

Practical implications – Some regulation instruments in Brazilian fuel market requires the forecast of light fuel consumption to better deal with supply and environment issues. In that context, the level of accuracy reached allows the use of these models as tools to assist public and private agents that operate in this market.

Originality/value – The study seeks to fill a gap in the literature on the Brazilian light fuel market. In addition, the methodological strategy adopted assesses projection models from different areas of knowledge using a robust evaluation procedure.

Keywords Co-integration, Forecasting, Time series analysis, Biofuels, Econometric, Demand forecasting, Autoregressive, Neural networks, Fuzzy-logic model, Demand-side management, Gasoline, Liquid fuels, Fuel demand, Forecasting methods, Time series, Machine learning, Forecast evaluation

Paper type Research paper

1. Introduction

The management of energy demand has become a mandatory issue for public and private agents working in this area, aiming to plan the need for future resources, promote environmental conservation measures, optimize the use of energy sources, assist on

International Journal of Energy Sector Management

1750-6220
DOI 10.1108/IJESM-02-2021-0009

decision-making and ensure regularity of supply (Suganthi and Samuel, 2012). In this context, the use of quantitative models has become recurrent to improve the forecasts associated with this market and is being consolidated as a fundamental tool to meet the aforementioned objectives.

Among the various sources of energy, liquid fuels used in spark-ignited vehicles (light fuels or Otto cycle [1] fuels) are highlighted, which are mostly used to meet the needs of private transportation around the globe. In Brazil, as it will be detailed in the following topics, the liquid fuels sector has undergone technological, economic and regulatory changes that have significantly altered its functioning. These changes have made the Brazilian light fuel market a unique case in the world and have substantially increased its complexity. Several factors that are difficult to predict started to influence the formation of fuel prices, the dynamics of domestic production, the behavior of consumption and, consequently, the guarantee of internal supply. In addition, in 2020, the health crisis caused by the COVID-19 pandemic generated consequences and uncertainties at a global level that affected several markets, with no previous event that would allow the evaluation, in advance, of the effects caused by the viral pandemic.

Therefore, the present study seeks to evaluate different methods for making predictions about the consumption of light fuels in Brazil. The analysis recognizes the need to anticipate the behavior of demand, considered as one of the main elements associated with greater predictability to the fuel market. The study seeks to answer questions such as: what is the best method to forecast the consumption of light fuels in the country? Is the performance of the models satisfactory for use as a decision-making support tool for public and private agents? Even when an unpredictable event like COVID-19 crisis becomes a fact, is there quantitative models that outperform the accuracy of a naïve method or another alternative method? To answer these questions, the performance of the most common models in the literature on the subject will be evaluated: SARIMA (SAR) and ARIMA with binary variables for seasonality (AR.Bin), vector error correction model (VECM), univariate and multivariate artificial neural networks (ANN and MANN), support vector machine (SVM), hybrid neural fuzzy inference system (HyFIS), random forest (RF) and, finally, combined models (ensemble learning).

As in other nations, the proposed assessment is justified by offering grounds to expand the guarantee of domestic supply. Specifically for the Brazilian market, the results also provide useful information for achieving various public policy standards and measures associated with the future behavior of demand for fuels. Finally, the proposed study seeks to fill a gap in the literature on the Brazilian market. This is because great part of the papers available on the theme did not address consumption forecast. In fact, these studies aimed to understand the effect of prices, income, and other economic variables on the behavior of demand for hydrous ethanol, gasoline or natural gas (NGV).

2. Brazilian light fuel market

The Brazilian light fuel market has peculiarities derived from technological, productive, and institutional complexity that affect, on the one hand, consumption decisions and, on the other, the dynamics of production and supply. Regarding technology, the introduction of flex-fuel vehicles in early 2003 gave the Brazilian consumer the opportunity to decide between ethanol and gasoline or any mixture of both products at each filling. This change affected the dynamics of price formation and, consequently, of the country's consumption of fuels by making the demand for individual fuels more elastic (Freitas and Kaneko, 2011). In 2020, the light fleet with flex-fuel engines represented 80.1% of automobiles and 41.1% of motorcycles in circulation in the country (UNICA-Brazilian Sugarcane Industry Association, 2020). Unlike other countries where gasoline is the main fuel, in the Brazilian market the

configuration established by the flex-fuel fleet requires that the demand for light fuels incorporate the aggregate consumption of gasoline, hydrous ethanol and natural gas or NGV (Rodrigues and Bacchi, 2017).

In the productive sphere, since sugarcane is the raw material for Brazilian ethanol, fuel production is subject to climatic conditions and to the fundamentals observed in the sugar market. International cycles and fluctuations in the sweetener's value affect domestic production decisions, since most sugar plants may, albeit in a restricted way, choose to use this material shared for one or another product (Drabik *et al.*, 2015). In the internal supply of gasoline, Petrobrás [2] has a near-monopoly of the activity, being responsible for more than 90% of the domestic production of the product. This condition allows the oil company to define the domestic prices regardless of market conditions. In the institutional sphere, from 2008 to 2016 there was constant government intervention in the oil sector aiming at inflation control. The main policies were the freezing of domestic gasoline prices and the exemption from federal taxes levied on the oil products (Rodrigues and Bacchi, 2016). After being immersed in an fiscal crisis, as of 2015 the federal government began to tax fuels again to contain the public deficit. At the same time, Petrobras's pricing policy was changed, and the price of gasoline began to be adjusted periodically after the company underwent major financial losses (Rodrigues and Rodrigues, 2018).

In this context, the dimensioning of domestic consumption of fuels becomes essential for the management of resources both by the public sector (need for inspection, infrastructure for the flow of production, etc.), and by the private sector, to define the increase in production capacity, decision of the plants' production strategy, dimensioning of investments in logistics by distributors, among others. In fact, this complexity of the national market has stimulated countless studies on the subject. Specifically in the case of demand, most studies conducted aimed at assessing gasoline, hydrous ethanol, or NGV consumption behavior individually in the face of changes in economic variables. This is the case of those elaborated by Cardoso *et al.* (2019), Santos *et al.* (2018), Isabella *et al.* (2017), Gomez and Legey (2015), Barros *et al.* (2014), Du and Carriquiry (2013), Salvo and Huse (2013), Santos (2013), Freitas and Kaneko (2011), Pacini and Silveira (2011), Alves and Bueno (2003), Burnquist and Bacchi (2002), among others. Among the authors who assessed the total demand for light fuels in Brazil (gasoline, hydrous ethanol and NGV together) are those developed by Rodrigues and Bacchi (2017), Costa *et al.* (2017), Figueira *et al.* (2014) and Rodrigues and Bacchi (2016).

All the aforementioned papers focused on the estimation of demand equations, identifying the price and income elasticities for fuel consumption. There are just a few works whose main objective was use quantitative methods to forecast fuel consumption in Brazil (Castro, 2012; Santiago *et al.*, 2011; Zanini *et al.*, 2001; Figueira *et al.*, 2010).

In general, the studies carried out for the Brazilian market are limited to the following aspects: they are essentially concentrated on the use of traditional time series methods, like ARIMA and Vector Autoregression models; they did not evaluate models to forecast the consumption of light fuels - most studies aim to verify the fit of the model in the training data and the significance of the elasticities; they did not dynamically test several methods in the same database and with the same error metric.

In fact, despite the small number of studies focused on the fuel consumption forecast in the national market, there are numerous studies proposing the forecast of the demand for energy in the international literature, especially for electricity. That said, when reviewing these works, the following techniques were most frequently identified: Fuzzy Logic models (Chang *et al.*, 2011; Kucukali and Baris, 2010; Padmakumari *et al.*, 1999; Pai, 2006), SVM (Ahmad *et al.*, 2014; Che and Wang, 2014; Chen *et al.*, 2017; Debnath and Mourshed, 2018; Fan *et al.*, 2008; García Nieto *et al.*, 2018; Hong, 2009) and RF (Dudek, 2015; Ibarra-Berastegi *et al.*, 2015; Lahouar and

Ben Hadj Slama, 2015; Kane *et al.*, 2014). Although little explored to forecast fuel consumption, the ANN models (Ebrahimpour *et al.*, 2011; Kavaklioglu *et al.*, 2009; Kandananond, 2011; Khashei and Bijari, 2010; Ringwood *et al.*, 2001; Sözen *et al.*, 2005) have been recurrently adopted to forecast the demand for different energy sources and other applications. Other methods applied to forecasting the demand and production of different energy sources can be consulted in Deb *et al.* (2017), Debnath and Mourshed (2018) and Suganthi and Samuel (2012).

3. Methodology

3.1 Data source and pre-treatment

The database used to represent the total demand for light fuels is composed by the sum of the consumption of gasoline, hydrous ethanol and NGV, converted into liters of equivalent gasoline. According to the model proposed by Rodrigues and Bacchi (2017), this conversion is necessary to consider the distinct efficiency of each fuel in the demand composition. Therefore, the consumption of light fuels (C_t) will be given by equation (1):

$$C_t = \sum_{i=1}^{3} C_{it} \times Re_i \tag{1}$$

where C_{it} representing the amount consumed of each fuel i = {hydrous ethanol,gasoline,NGV} in month t; Re_i, is the coefficient of energy equivalence for the conversion into equivalent liters of gasoline (0.70 in case of hydrous ethanol, 1.23 for NGV and 1.00 for gasoline).

The monthly data on gasoline and hydrous ethanol consumption were obtained from information by the Brazilian Agency of Petroleum, Natural Gas, and Biofuels (ANP-National Agency of Petroleum, Natural Gas and Biofuels, 2021a). Additionally, the amount of natural gas commercialized in the country was compiled by the Brazilian Association of Piped Gas Distributing Companies (ABEGÁS-Brazilian Association of Pipeline Gas Distribution Companies, 2021). The historical consumption data, which are available between January 2002 and April 2021, can be viewed in Figure 1. The sharp drop observed between March 2020 and May 2020 is due the beginning of COVID-19 pandemic and the implementation of mobility restriction measures by the Brazilian government. When these months were used

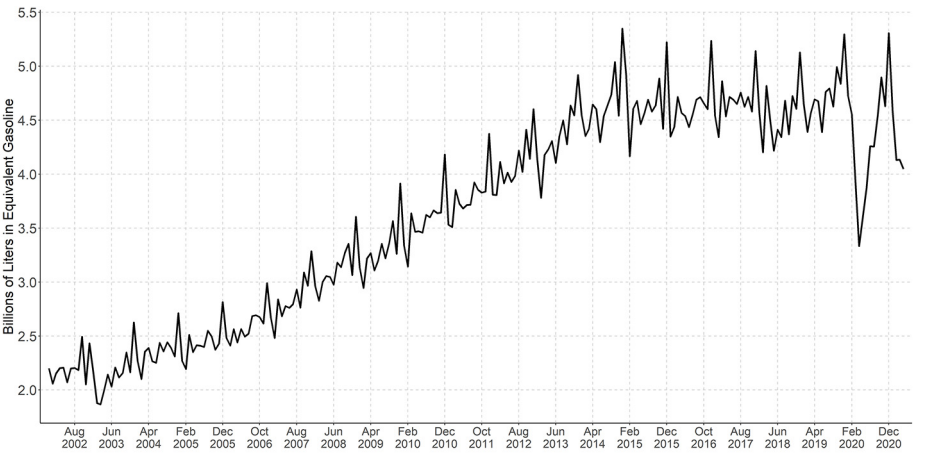

Figure 1.
Consumption of hydrous ethanol, gasoline and NGV in liters of equivalent gasoline monthly from January 2002 to April 2021

as part of the training set for the models, the authors used binary variables as outlier control method, to avoid a distortion in the estimation process.

For multivariate forecast models, two other series were forecast together with the Otto cycle consumption. The first consists of the GDP per capita estimated from the monthly GDP released by the Central Bank of Brazil (BACEN-Central Bank of Brazil, 2021) and the resident population by the Brazilian Institute of Geography and Statistics (IBGE-Brazilian Institute of Geography and Statistics, 2021). The second, in its turn, refers to the average price of light fuels to the final consumer (ANP-National Agency of Petroleum, Natural Gas and Biofuels, 2021a).

3.2 Forecast models evaluated

3.2.1 Seasonal autoregressive integrated moving average model. SARIMA model (p, d, q)(P, D, Q) (SAR) aims at explaining a stationary time series through the past values of the series itself and the model errors (Box *et al.*, 2008).

The estimated model can be represented according to equation (2):

$$\Delta^D(\Delta^{\mathrm{d}}N_t) = \phi_0 + \sum_{i=1}^{p}\phi_i\,\Delta^D(\,\Delta^{\mathrm{d}}N_{t-i}) + \sum_{i=1}^{q}\theta_i e_{t-i} + \sum_{j=1}^{P}\Phi_j\,\Delta^D\Big(\,\Delta^{\mathrm{d}}N_{t-f\times j}\Big) + \sum_{j=1}^{Q}\Theta_j e_{t-f\times j} + e_t,\quad t=1,\ldots,n \tag{2}$$

where Δ^{d} and Δ^D are the difference operator of order d and D, respectively, ϕ_0 is the intercept of the stationary series and ϕ_i, θ_i, Φ_j and Θ_j are coefficients that multiply the autoregressive terms and the moving averages, f is the series frequency that characterizes its seasonality (for a monthly series $f = 12$), and e_t is a normally distributed error with zero mean and variance σ_e^2.

Along with the traditional approach, which uses the series and errors lags to incorporate seasonality effects, this study proposes an alternative approach using binary regressors to control seasonality (AR.Bin). Thus, the model starts being represented by equation (3):

$$\Delta^D(\Delta^{\mathrm{d}}N_t) = \phi_0 + \sum_{i=1}^{p}\phi_i\,\Delta^D(\,\Delta^{\mathrm{d}}N_{t-i}) + \sum_{i=1}^{q}\theta_i e_{t-i} + \sum_{m=1}^{f-1}\beta_m \mathrm{b}_m + e_t,\quad t=1,\ldots,n \tag{3}$$

where b_m is a binary variable representing one month m and β_m is the coefficient representing the differential effect of this month related to the basal month. The seasonality effect on the reference month is given by ϕ_0, while for any $m = 1, \ldots, f-1$ this effect will be $\phi_0 + \beta_{\mathrm{m}}$.

To determine the number of autoregressive components and moving averages, the autocorrelation and partial autocorrelation functions were evaluated. The order of integration of the series was determined from the Kwiatkowski–Phillips–Schmidt–Shin (Kwiatkowski *et al.*, 1992) and Dickey-Fuller with Generalized Least Squares (Elliott *et al.*, 1996) tests for a unit root. The Dickey-Pantula test (Dickey and Pantula, 1987) was implemented to check for the presence of two roots and the presence of seasonal unit root

was evaluated according to the procedure described by Osbom *et al.* (1988). The final selection was based on the Akaike Information Criterion (AIC). The Ljung-Box test for autocorrelation, the Lagrange Multiplier test for autoregressive conditional heteroscedastic (ARCH-LM) test for heteroscedasticity, and the Jarque–Bera (Bueno, 2018) and Shapiro–Wilk (Shapiro and Wilk, 1965) tests for normality were implemented in the verification of the models.

3.2.2 Autoregressive vector with error correction mechanism. The autoregressive VECM consists of an alternative approach to forecast non-stationary time series. Let $\mathbf{Y}_t$ be a vector of m endogenous variables, when some of these variables are non-stationary, but have a common long-term trajectory, it is possible to use a VECM model.

Once the presence of long-term equilibrium is identified, the variables contained in $\mathbf{Y}_t$ are said to be cointegrated in order (d, b) or $C(d, b)$. Thus, it is possible to estimate up to $m - 1$ cointegration vectors. In this case, $\mathbf{Y}_t$ can be expressed by a VECM model according to equation (4):

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha} \mathbf{B}^{\mathrm{T}} \mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Pi}_j \, \Delta \mathbf{Y}_{t-j} + \mathbf{G} \mathbf{Z}_t + \mathbf{e}_t \tag{4}$$

Where $\mathbf{B} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_{\mathrm{r}}]$ and $\boldsymbol{\alpha}$ are matrices $m \times r$ such that $r = rank(\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}})$, $r < m$ is the number of non-zero cointegration vectors and, consequently, error correction vectors. The latter are expressed by $\mathbf{B}^{\mathrm{T}} \mathbf{Y}_{t-1} = \mathbf{u}_{t-1} = \left[u_{t-1}^1, u_{t-1}^2, \ldots, u_{t-1}^r\right]$. $\mathbf{e}_t$ is a white noise, $\mathbf{Z}_t$ is a vector of g exogenous variables, and $\mathbf{G}$ is the matrix of exogenous variable coefficients with dimension $m \times g$. $\boldsymbol{\Pi}_j = -\sum_{i=1+j}^{p-1} \boldsymbol{\Phi}_i$, $j = 1, \ldots, p-1$ are vectors of coefficients that multiply $\Delta \mathbf{Y}_{t-j}$.

The cointegration test proposed by Johansen (1988, 1991) was carried out. The selection of the number of autoregressive terms was based on the AIC. In the verification process, the Portmanteau (Lütkepohl, 2007) and Edgerton and Shukur (1999) autocorrelation tests, the Jarque–Bera (Lütkepohl, 2007) and Shapiro–Wilk (Shapiro and Wilk, 1965) multivariate normality test and the residual multivariate heteroscedasticity test ARCH-LM (Engle, 1982) have been implemented.

3.2.3 Artificial neural networks. ANN can be represented by a network of oriented links that are interconnected at certain points called nodes or neurons (Haykin, 2001). Let the parameters vector of the model given by $\boldsymbol{\omega} = \left\{ \left\{ \omega_{k'}^l \right\}_{k' \in l-1} \right\}_{l=2}^{L}$, $\{\mathbf{Y}_t, \mathbf{x}_t\}_{t=1}^n$, a set of known data where $\mathbf{Y}_t$ is vector m-varied of the optimal response to a vector p–varied of input $\mathbf{x}_t$. The estimated value of $\boldsymbol{\omega}$ is given by equation (5):

$$\boldsymbol{\omega} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{j=1}^{m} \sum_{t=1}^{n} \left(Y_{t,j} - \hat{f}_j^L(\boldsymbol{\omega}, \mathbf{x}_t) \right)^2 \right\} \tag{5}$$

where $\hat{f}^L(\boldsymbol{\omega}, \mathbf{x}_t) = \left\{ \hat{f}_j^L(\boldsymbol{\omega}, \mathbf{x}_t) \right\}_{j=1}^m$ are nonlinear mappings of $\mathbf{x}_t \rightarrow \mathbf{Y}_t$ resulting in layer L, called the ANN output layer. In the present study $L = 3$ and the number of neurons in the output layer is equal to the number of endogenous variables (m) in the model. To optimize

the function, the Broyden–Fletcher–Goldfarb–Shanno algorithm was used because it presents better performance in smaller dimension problems (Haykin, 2001).

Two ANN structures were estimated. The first is a univariate model ($m = 1$) with Otto cycle fuel consumption as the output variable and p autoregressive terms as input variables. The second is a multivariate model ($m = 3$) that jointly forecasts the Otto cycle consumption, the average fuel price and GDP per capita through p autoregressive terms for each of these variables as an input vector.

3.2.4 Support vector machines. SVM is a non-parametric model capable of estimating nonlinear relations from an input vector to an output variable, using the so-called kernel trick (Hastie *et al.*, 2009). It consists of applying a nonlinear transformation $\left\{\left\{\varphi_j(\mathbf{x}_t)\right\}_{j=1}^{m_1}\right\}_{t=1}^{n}$ on the input variables $\left\{\left\{x_{i,t}\right\}_{i=1}^{m_0}\right\}_{t=1}^{n}$ so as the dimension increases from m_0 to m_1. The objective of the method is to transform a nonlinear relation of $\mathbf{x}_t \rightarrow \mathbf{Y}_t$ in a linear relation of $\boldsymbol{\varphi}(\mathbf{x}_t) \rightarrow Y_t$. It is necessary to estimate the vector of weights $\boldsymbol{\omega}$ that allows the construction of a hyperplane $L\left(Y_t, \hat{f}(\boldsymbol{\omega}, \mathbf{x}_t)\right) = 0$, so as $\hat{f}(\boldsymbol{\omega}, \mathbf{x}_t) = \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_t) = \sum_{j=0}^{m_1} w_j \varphi_j(\mathbf{x}_t)$. The estimate of coefficients vector takes place indirectly, from the resolution of the problem of optimization given by equation (6):

$$\begin{aligned}
&\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}'} Q(\boldsymbol{\alpha},\boldsymbol{\alpha}') = \sum_{t=1}^{n} Y_t\left(\alpha_t - \alpha_t'\right) - \epsilon \sum_{t=1}^{n}\left(\alpha_t - \alpha_t'\right) - \frac{1}{2}\sum_{t=1}^{n}\sum_{j=1}^{n}\left(\alpha_t - \alpha_t'\right)\left(\alpha_j - \alpha_j'\right)K(\mathbf{x}_t, \mathbf{x}_j) \\
&s.t \\
&\sum_{t=1}^{n}\left(\alpha_t - \alpha_t'\right) = 0 \\
&0 \le \alpha_t \le C \ , \ i = 1, 2, \ldots, n \\
&0 \le \alpha_t' \le C \ , \ i = 1, 2, \ldots, n
\end{aligned} \tag{6}$$

In equation (6) $K(\mathbf{x}_t, \mathbf{x}_j) = \boldsymbol{\varphi}^{\mathbf{T}}(\mathbf{x_t})\boldsymbol{\varphi}(\mathbf{x_j}) = \sum_{j=0}^{m_1} \varphi_j(\mathbf{x}_t)\varphi_j(\mathbf{x}_j)$ is called kernel function. In practice, to estimate an SVM it is not necessary to know the function $\boldsymbol{\varphi}(\mathbf{x})$, only the kernel function and that it is in accordance with Mercer's theorem (Haykin, 2001). All input vectors $\mathbf{x}_t$ whose $\left(\alpha_t, \alpha_t'\right) \neq 0$ are called support vectors. In fact, these observations are the ones that define the estimated hyperplane since the optimal weight vector will be given by $\boldsymbol{\omega} = \sum_{t=1}^{n}\left(\alpha_t - \alpha_t'\right)\boldsymbol{\varphi}(\mathbf{x}_t)$ and the forecast value for an input $\mathbf{x}_h$ will be $Y_h = \sum_{t=1}^{n}\left(\alpha_t - \alpha_t'\right)K(\mathbf{x}_h, \mathbf{x}_t)$.

3.2.5 Hybrid neural fuzzy inference system. The fuzzy controls are ways to establish relations among linguistic variables from modeling of rules connecting a condition (If) and a consequence (Then). The systems mixing the fuzzy systems and the neural networks paradigms can be called "neural *fuzzy*" (Jang *et al.*, 1997). This study implemented the method – HyFIS (Kim and Kasabov, 1999). The step of rules learning was performed using the procedure proposed by Wang and Mendel (1992).

3.2.6 Random forest. Random Forest model uses decision trees to make forecasts of y_t in function of an input vector $\mathbf{x}_t$. In this case, several regression trees are estimated, and the

final forecast of the algorithm is given by the average of trees forecasts (Breiman, 2001; Hastie *et al.*, 2009). If a regression tree set $\{T_b\}_{b=1}^{B}$ will be estimated to map the relation $\mathbf{x} \rightarrow Y$, then the algorithm performs the following steps:

- Obtain a random subsample of size n_b from the training sample.
- Generating a regression tree T_b with the simulated sample and, using the stop criterion that each terminal node reaches the minimum size n_{min}^{b}, or the variation of the training sample residual sum of squares is below a tolerance limit.
- The remaining observations on the terminal nodes are used in the forecast of an individually tree given by $T(\mathbf{x},\Theta_b)$. Where Θ_b represents the variables selected, the optimal division points and the observations in each terminal node.
- The forecast value for an input vector $\mathbf{x}$ is given by $\hat{f} = \frac{1}{B}\sum_{b=1}^{B} T(\mathbf{x},\ \Theta_b)$.

Knowing that there is a set of S possible divisions for the nodes of a tree T_b, the selection of the optimal node break point was made according to the algorithm CART for decision trees (Breiman *et al.*, 1984).

3.2.7 Ensemble learning models. The ensemble learning models aim at combining the results of different algorithms to reduce forecast error variance and to increase its accuracy. The most usual form is based on the linear combination of forecasts according to equations (7) and (8):

$$\hat{\mathbf{y}} = \hat{\mathbf{Y}}\mathbf{W} \tag{7}$$

$$\sum_{j}^{J} w_j = 1 \tag{8}$$

where $\hat{\mathbf{y}}$ is a combined forecast vector with dimension $T \times 1$, $\hat{\mathbf{Y}}$ is a matrix $T \times J$ containing the forecasts of the J different algorithms for T periods and $\boldsymbol{W}$ is the dimension vector $J \times 1$ which contains the weights w_j for the linear combination of forecasts.

The study results proposing a combination of forecasts have been unanimous, favoring the improvement of forecast accuracy (Clemen, 1989; Hyndman and Athanasopoulos, 2018). Two forms of combination of predictions were evaluated from an empirical point of view. The first estimates the weights by means of restricted least squares (RLS) imposing that the sum of the weights is equal to one. In the second method, a simple average (AVG) of the predictions of the different models is taken as the final projection of the data.

3.3 Determination of hyperparameters and evaluation metrics

3.3.1 Determination of hyperparameters of the models. The determination of the set of hyperparameters was obtained by minimizing the estimate of the generalization error or forecast error (Hastie *et al.*, 2009). The procedures described below, which determine the hyperparameters of the models, did not use data after February 2020, when the measures to restrict mobility caused by the COVID-19 health crisis began. The noise contained in the data, caused by this event, should not affect the choice of hyperparameters, which represent a structural component of the models and are decisive for their ability of generalization.

For linear models, a simple way to assess the generalization capacity is to compare the AIC (Hastie *et al.*, 2009; Hyndman and Athanasopoulos, 2018) after estimating the models with all available data. These are the cases of the SARIMA and VECM methods. For non-linear models, the cross-validation process was used. Cross-validation involving time series

is different from problems involving panel data due to the series autoregressive dynamics (Bergmeir *et al.*, 2018). In the present study the procedure was structured as follows:

- A set of parameters Θ_i is initially selected for the model.
- An initial cut is performed in the series of consumption of the Otto cycle, with only the data from January 2002 to December 2011 remaining, that is, 120 observations of a total of 218. The selection of this cut point is explained by the change of trend in the series that was observed from 2013 onwards.
- This subseries is used to estimate the models and obtained the h step ahead forecasts ($\hat{Y}$), and the forecast error vector is calculated and stored.
- A new observation is added to the training base and step 3 is repeated.
- Steps 3 and 4 are repeated until the h-step ahead forecast has as its last forecast period the last observation known of the series. The root mean square error (RMSE) is calculated for the model with parameters Θ_i according to equation (9):

$$RMSE_i = \sqrt{\frac{1}{(81-h)h}\sum_{j=1}^{(81-h)h}\left(Y_j - \hat{Y}_j\right)^2} \tag{9}$$

Obtaining the forecast errors of the subseries, it was possible to estimate RMSE as metrics of the model generalization error, and to select the set of hyperparameter with better performance.

3.3.2 Evaluation metrics and statistical tests for comparing algorithms. After defining the hyperparameters, one can proceed to compare the different forecasting methods. An interactive process was conducted to generate a robust estimate of the accuracy of the models. The following steps were performed:

- An initial cut is made in the Otto cycle consumption series leaving only the data from January 2002 to December 2011.
- This subseries is used to estimate the models and carry out the h-step ahead forecast. The RMSE is calculated, according to equation (12), and stored as a measure of accuracy in the period. In this case, each cutoff period has an RMSE value for the h-step ahead forecast.
- A new observation is added to the training base and step 3 is repeated.

Step 3 is repeated until the h-step ahead forecast has the last known observation of the data as the last predicted period. Thus, a series with the RMSE is obtained for each point in time and for each model. In this study, two nonparametric tests were implemented (Shingala, 2016). Friedman's test was performed to check for a difference between the average ranking of an algorithm i, given by $\bar{r}_i$, and the average performance of all models $\bar{r}$. The test hypotheses were shown on equation (10):

$$\text{Friedman}\begin{matrix}\text{H0}: \bar{r}_i = \bar{r}, \text{there is no difference in the ranking of methods in general} \\ \text{HA}: \bar{r}_i \neq \bar{r}, \text{there is a difference in the ranking of methods in general}\end{matrix} \tag{10}$$

If Friedman's test rejects *H0*, Nemenyi test can be performed to compare the pairs of methods. In this case, these are compared pair-to-pair through its average ranking. equation (11) shows the null hypotheses and test alternatives"

$$\textit{Nemenyi}\begin{cases}\text{H0}: \bar{r}_i = \bar{r}_j \ \forall\ i \neq j, \text{there is no difference in the ranking of methods } i \text{ and } y.\\ \text{HA}: \bar{r}_i \neq \bar{r}_j \ \forall\ i \neq j, \text{there is a difference in the ranking of methods } i \text{ and } y.\end{cases} \tag{11}$$

The execution of multiple tests for pairwise comparison of multiple methods leads to the inflation of type I error. To avoid this issue, the Bonferroni control method (Shingala, 2016; Demsar, 2006) was performed. A naive forecasting method was introduced between the models to verify the superiority of the methods in relation to a model with low cost of implementation. In this case, the months of the last year will be used as a forecast for a subsequent year.

The RMSE was used as a criterion for ranking the models. However, the Mean Absolute Percentage Error (MAPE) will be displayed for each point in time (that is, for the h-step ahead forecast from that date), calculated according to equation (13), to facilitate the interpretation of results by excluding the dimensional issue in errors. Additionally, the maximum Percentage Error (PE) considering the h step ahead forecast horizon is also presented (equation (14)):

$$RMSE_t = \sqrt{\frac{1}{h}\sum_{j=1}^{h}\left(Y_j - \hat{Y}_j\right)^2} \tag{12}$$

$$MAPE_t = \frac{100\%}{h}\sum_{j=1}^{h}\frac{|Y_j - \hat{Y}_j|}{Y_j} \tag{13}$$

$$\max PE_t = 100\%\max\left\{\frac{|Y_j - \hat{Y}_j|}{Y_j}\right\}_{j=1}^{h} \tag{14}$$

4. Results and discussion

4.1 Selected hyperparameters

Table 1 shows the results of the selection of hyperparameters. Models whose parameters are chosen according to the forecast horizon show a break for 3, 6 and 12 months.

The use of non-seasonally adjusted series in the ANN, multi ANN, SVM, HyFIS and RF models comprises the alternative with the smaller forecast error. The additional diagnostic tests performed for the SARIMA, ARIMA with binary variables and VECM models indicate good adherence of the models to the data, except for the normality tests of the residues. Despite the results of the normality tests, it is worth noting that they are highly sensitive to the presence of outliers and, in such cases, tend to reject the hypothesis of normally distributed data with ease (Coin, 2008). Thus, these approaches were also considered in the model comparison procedures.

4.2 Accuracy evaluation and model comparison

At first, the comparison with the combined models was disregarded to choose the algorithms with superior individual performances. The Friedman test rejects the null hypothesis. Therefore, it indicates that there is difference in the ranking of methods in

Table 1.
Hyperparameters selected from cross-validation procedures and information criteria

Model	Treatment of seasonality	Data pre-processing	Optimal hyperparameters	Hyperparameters
Sarima	Seasonal autoregressive terms	Logarithm	$p = 2, d = 1, q = 3$ $P = 1, D = 0, Q = 1$	Number of autoregressive terms (p), moving average (q), integration order (d), number of seasonal autoregressive terms (P), seasonal moving average (Q) and seasonal integration order (D)
Arima w/ Binarys	Binary variables	Logarithm	$p = 2, d = 1, q = 3$	Number of autoregressive terms (p), moving average (q), integration order (d)
VECM	Binary variables	Logarithm	$p = 5, r = 1$	Number of autoregressive terms (p) and cointegration vectors (r)
ANN	Moving Averages Filter	Standardization by mean and standard deviation	3 months $p = 4, k_1 = 1, k_2 = 2$ 6 months $p = 3, k_1 = 1, k_2 = 1$ 12 months	Number of autoregressive terms (p), number of neurons in the first and second hidden layer (k_1, k_2)
Multi ANN	3 meses Moving Averages 6 meses Filter 12 meses Binary variables	Standardization by mean and standard deviation	3 months $p = 4, k_1 = 10, k_2 = 10$ 6 months $p = 3, k_1 = 5, k_2 = 0$ 12 months $p = 3, k_1 = 10, k_2 = 9$	Number of autoregressive terms (p) of the endogenous variables, i.e, Otto cycle consumption, average price of light fuels and GDP per capita, number of neurons in the first and second hidden layer (k_1, k_2)
SVM	Moving Averages Filter	Standardization by mean and standard deviation	3 months $p = 3, \epsilon = 0{,}1, C = 6,$ kernel=gaussian 6 months $p = 3, \epsilon = 0{,}1, C = 10,$ 12 months kernel=gaussian	Number of autoregressive terms (p), error tolerance threshold (ϵ), constant C and the kernel function for input space transformation
HyFIS	Moving Averages Filter	Standardization by mean and standard deviation	3 months $p = 3, L = 22$ 6 months 12 months $p = 3, L = 16$	Number of autoregressive terms (p), number of labels that define the input and output speech universe (L)
Random Forest	Moving Averages Filter	No transformation	3 months $p = 6, B = 100,$ m = 5, $n^B_{min} = 4$ 6 months $p = 6, B = 1000,$ 12 months $m = 6,$ $n^B_{min} = 3$	Number of autoregressive terms (p), number of trees (B), number of explanatory variables used by a tree (m) and the minimum size of a terminal node (n^B_{min})

general. The results of the Nemenyi test and the average ranking for each method, in its turn, are presented in Figure 2. The best algorithm identified by this procedure was SVM, followed by RF and AR.Bin for the 12 and 6months forecast. Only in the case of the 3-month horizon did the latter exceed SVM's performance. However, it is not possible to affirm that the first 5 best ranked models have a statistically lower performance than SVM in the 12-month forecast. The same is valid for the forecast of 6 and 3 months ahead. The confidence interval of the test is represented in Figure 2 by the gray area.

The performances of the best-ranked methods were statistically superior to the naive model. This fact suggests greater robustness of these procedures, in contrast to the higher computational cost necessary to implement them. Although all data were considered in statistical tests, the error metrics are presented discriminating the period before and after the first wave of COVID-19 contamination. The metrics, for both periods, are presented in Table A1 in the Appendix and it show the mean and standard deviation of the MAPE of the forecast, the MAPE average of the training sample, the maximum and minimum MAPE value and the maximum percentage error of each model considering all periods used in the validation.

Table A1 end up reflecting the results obtained in the Nemenyi test. SVM and AR.Bin lead with the best accuracy. It is possible to notice that when considering all periods of validation or, only the period before COVID-19 first wave, the longer forecast horizon implies an increase in the average MAPE of the models, in general, which have greater accuracy in the 3-month forecast, followed by the 6- and then 12-month forecast. In addition, this increase in the forecast horizon does not lead to generalized degradation of the forecasting capacity of the methods and, on the other hand, allows to discriminate them more precisely. Consequently, there is a greater heterogeneity of rankings and MAPE in the 12-month forecast. The SAR and VECM models are no longer part of the significant region of Nemenyi test. Therefore, there is a strong indication that the SVM showed to be superior to these traditional time series methods for the data used. The performance of the VEC

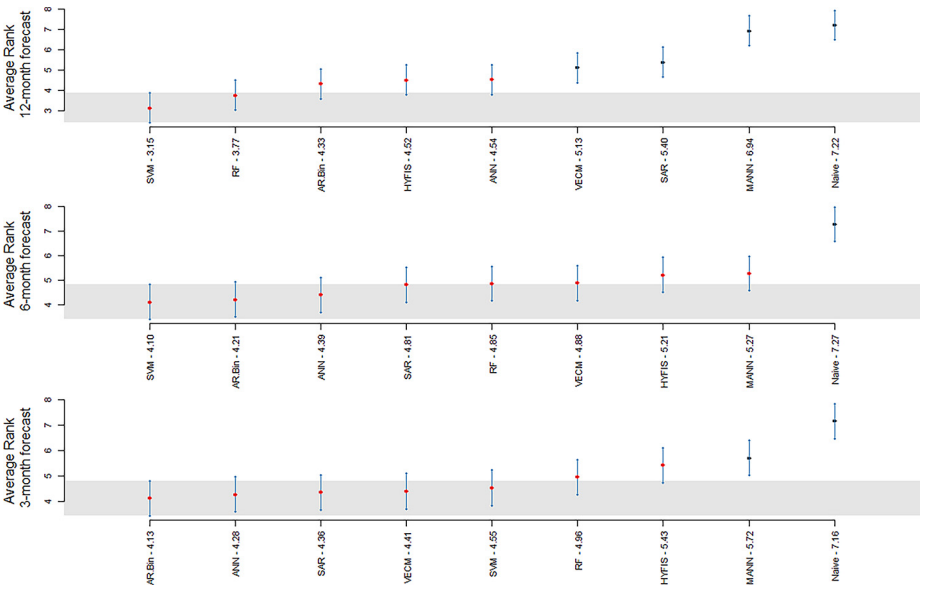

Figure 2.
Result and confidence interval of the Nemenyi test for forecast 12,6 and 3 months ahead

depreciates considerably after the start of the COVID-19 pandemic. The forecast error of the explanatory variables (price and income), which is carried over to fuel consumption forecast, is an innate consequence of multivariate models, which reduce their accuracy.

Looking only at the period prior to the start of COVID-19 contamination, all forecast horizons have model options with an average MAPE below 3%. Except for MANN for 12 months forecast, apparently no method has an overfitting problem, since the training and validation errors do not differ significantly. In order to assess the stability of the forecast error, one can analyze the MAPE standard deviation and the maximum MAPE and percentage error within a point in time. Before COVID-19 first wave, the SVM, whose performance ranking was the best among all methods in the 12 and 6 months forecast, also obtained a low standard deviation of MAPE considering all horizons. Considering the 3 and 6 months forecast horizons, in which the AR.Bin model obtained the smallest average MAPE, the maximum percentage error of this model was 12.4% and the maximum MAPE was 6.1% and 5.7%, respectively. On the other hand, the SVM model, whose average MAPE is not as low as AR.Bin, has a more stable performance when looked by those metrics.

Observing the period after the beginning of the first wave of COVID-19 contamination in Brazil, it is possible to notice a general worsening in the performance of the models for the forecasts that incorporate into their horizon the months after March 2020. The stability of the models also had a negative impact particularly in the early months of the pandemic, as the training data set did not have any information about this scenario. However, while increasing the forecast horizon impaired prediction accuracy, increasing the number of forecasted months reduces the weight of the initial months in the calculation of the MAPE, which showed a greater detachment of forecasts in relation to the value actually observed. For this reason, the mean and standard deviation of the MAPE are lower for a 12-month forecast. SVM and AR.Bin continue to present the lowest MAPE among the estimated individual models. SVM also presents the most stable performance in all forecast horizons.

After analyzing the individual methods, the joint learning models were implemented considering the top two methods with the lowest MAPE in each forecast horizon, these being SVM and RF for 12 months and AR.Bin and SVM for 3- and 6-months.

4.3 Results of the ensemble learning models

The Nemenyi test were reevaluated including the combined methods. Figure 3 presents the result of the Nemenyi test.

Accuracy metrics previously mentioned, in the individual models, can also be checked in the Table A1 in the Appendix for the combined models. The combination using the mean of the forecasts generates a MAPE as good as the combination based on the inference of the weights using the least squares method. Although the combined models have lower forecasting errors and ranking for the 3 and 6 months forecast, the performed tests do not reject the null hypothesis that their accuracy is as good as that generated by the individual methods. For 12-month forecast horizon the combined models outperformed the MAPE of SVM only in periods before COVID-19.

Finally, considering only the period after the beginning of the COVID-19 crisis, the Nemenyi test does not allow us to state that there is a difference between the models and the naive method for forecasts of 3 and 6 months. However, if considering the 12-month horizon in which the accuracy heterogeneity is more evident, it is possible to conclude that only the SVM and the RLS combined model are statistically superior to the naive method when performing predictions for this period (Figure 4). Still, the traditionally used SAR and VECM models now occupy the last positions in the performance ranking in this period.

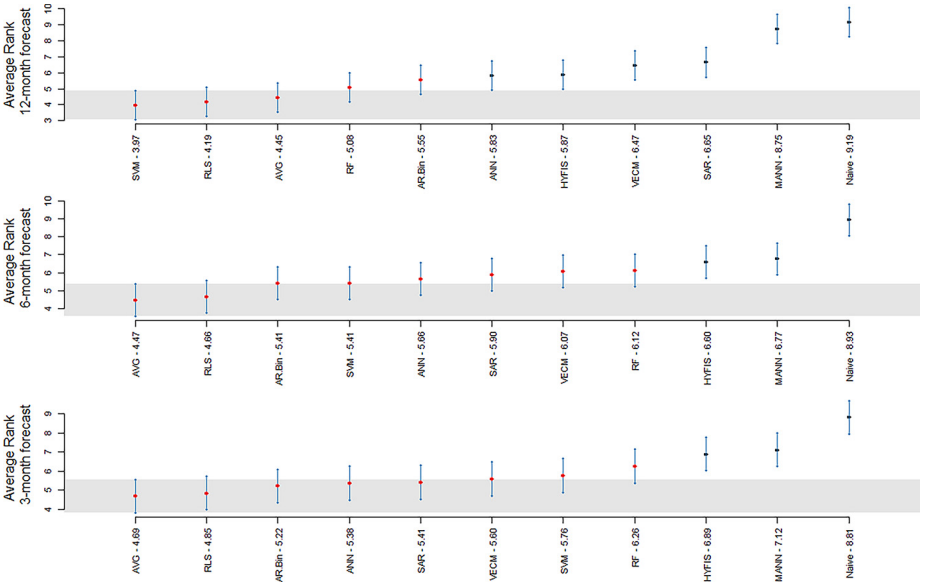

Figure 3.
Result and confidence interval of the Nemenyi test for forecast 12 ,6 and 3 months ahead including the ensemble learning models

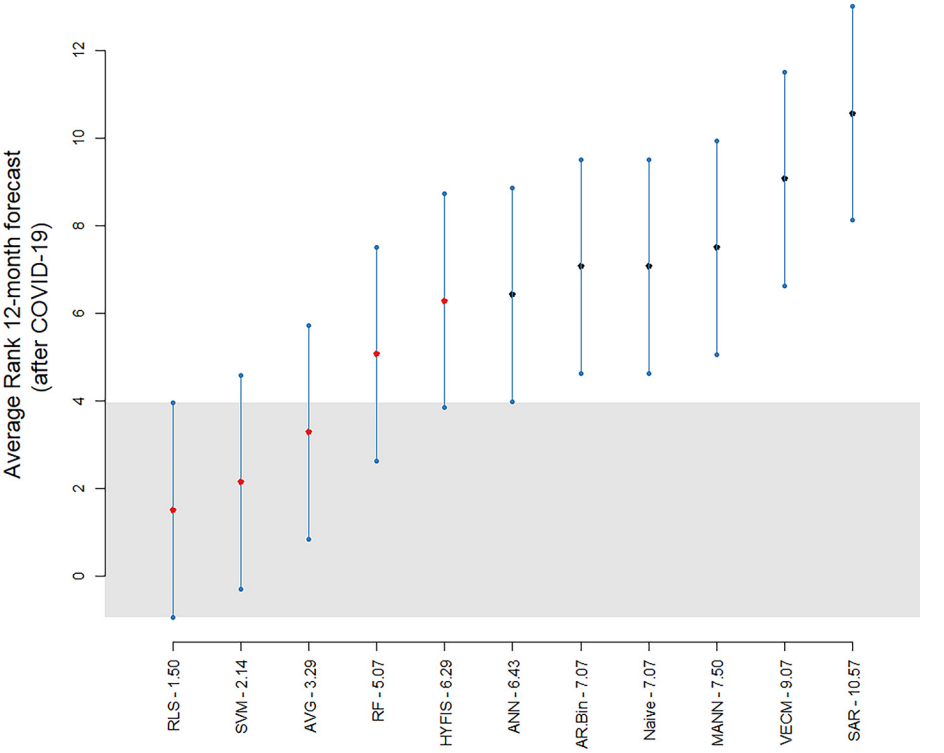

Figure 4.
Result and confidence interval of the Nemenyi test for forecast 12 months ahead whose predicted horizon contains months from March 2020

4.4 General evaluation of results and considerations on the application of the models

In general terms, in the period prior to the beginning of the COVID-19 pandemic, model forecasts for three months of consumption showed an average MAPE below 2.3%, with maximum variations (Max MAPE) below 6%. These values indicate that the structure developed can be useful for commercial and logistical planning carried out routinely by agents of the private sector, in the dimensioning of their operations in the short term. In the case of fuel supply contracts, for example, the negotiations cover permitted variations of up to 5% in the withdrawal of the volume initially contracted.

Short-term projections can be used by the public sector in the process of monitoring domestic light fuel supply. This activity is carried out periodically by the Supply Monitoring Committee, coordinated by the Ministry of Mines and Energy. This committee was created by Resolution No. 14/2017 of the National Energy Policy Council (CNPE-National Energy Policy Council, 2017) and, among other aspects, aims to monitor the conditions of supply and demand for light fuels, in addition to discussing strategies to ensure adequate supply of the light fuels. In the case of annual forecast, the results obtained show options for models with an average MAPE below 3%. These parameters also reveal the possibility of using the evaluated models in the coordination and planning of agents operating in this sector.

To cite an example, the projection of light fuel consumption for 12 months is one of the parameters to be evaluated by agents in the process of contracting anhydrous ethanol. This is because ANP Resolution No. 67/2011 (ANP-National Agency of Petroleum, Natural Gas and Biofuels, 2011) established incentives for the advance contracting of the supply of this biofuel between producers and distributors. Currently, the minimum volume to be contracted is defined based on the consumption of gasoline observed in the previous year. Still in the regulatory sphere, the approval of Law No. 13,576 of December 2017 and the institution of the National Biofuels Policy in Brazil (RenovaBio [3]) started to require regular forecasts for the definition of the ten-year goals of reducing the carbon intensity of the Brazilian fuel matrix (CRBIO-RenovaBio Commitment, 2020; CNPE-National Energy Policy Council, 2020). This definition is annually approved after a wide public consultation process. Among the parameters used in that definition, there is the consumption of light fuels in the country. Thus, the tools developed here can also be useful for the calibration of the decarbonization goals mentioned above.

Even for the post-COVID period, despite the lower accuracy of the models, the results proved to be relevant to support decision-making by public and private agents. During this period of greater uncertainty and continuous changes, the estimated models maintained a greater performance compared to the naive model and outperformed even traditional models of time series. In 2020, for example, the National Agency for Petroleum, Natural Gas and Biofuels promoted an exceptional reduction of 16% (ANP-National Agency of Petroleum, Natural Gas and Biofuels, 2020) in the requirements for contracting anhydrous ethanol, governed by the aforementioned resolution, due to the expected drop in fuel demand. Also, the National Energy Policy Council, through CNPE Resolution No. 08/2020 (CNPE-National Energy Policy Council, 2020), also promoted a reduction of 50% in the decarbonization targets of RenovaBio, due, among other aspects, to the perspective of reducing consumption resulting from the effects of the pandemic. In both situations, the models estimated here could have been used to support the mentioned decisions, since the initially projected drop that justified these measures proved to be overestimated (ANP-National Agency of Petroleum, Natural Gas and Biofuels, 2021b).

5. Conclusions

The present study compared different methods for predicting the consumption of Otto cycle fuels in Brazil. The cross-validation procedure allowed estimate the generalization error of the models and the average MAPE obtained was less than 3% in all forecast horizons for periods before COVID-19 crisis and, even after this event, the results proved to be robust, and the superior performance of some models still hold. When all validation data were considered, the most part of the methods proved to be significantly superior to the naive forecasting model.

In addition, for the 12-month forecast whose accuracy heterogeneity is more evident, the SVM and the combined RLS method were found to be statistically superior to SARIMA and VECM. This fact is relevant given the wide use of these methods, which makes it an interesting benchmarking for forecasting. Furthermore, the longer the horizon to forecast, the more relevant becomes the execution of a robust comparison process between the models, since the difference in performance between them intensifies.

Thus, it is possible to conclude that, despite the technological, market, and institutional changes observed in the Brazilian light fuel sector in the last 10 years, the forecast models evaluated indicate that it is possible to make forecasts for the consumption of light fuels with a satisfactory level of accuracy. Total consumption forecast for light fuels is useful for private agents in the definition of operational and investment actions related to the production, distribution, transport, and resale of fuels in the country. In the case of the public sector, the models implemented can be used in the dimensioning of resources to adjust domestic production capacity, as well as the logistical infrastructure necessary to supply the demand for transport. In addition, in the Brazilian case, consumption forecast is critical for the more efficient operation of marketing rules and established public policies.

Finally, the analysis conducted here does not cover all the need for further studies related to the topic. Models for carrying out regional forecasts and forecasting the individual demand for each fuel are important for a deeper understanding of the topic. New assessments should also take place when more data is available for the post-COVID-19 period, as the analysis conducted here incorporated only fourteen observations from that period. The limited number of observations does not allow an assertive conclusion about possible changes in consumption patterns promoted by changes in the dynamics of mobility introduced by the global health crisis.

Notes

1. The term “Otto cycle” refers to the thermodynamic cycle associated with the operation of internal combustion engines with spark ignition.
2. The company Petróleo Brasileiro SA (Petrobrás) is a Brazilian mixed capital company whose majority shareholder is the federal government.
3. The program aims to establish emission targets for fuel distributors in view of the level of pollution related to each fuel.

References

ABEGÁS-Brazilian Association of Pipeline Gas Distribution Companies (2021), “Consumption statistics”, available at: www.abegas.org.br/estatisticas-de-consumo (accessed 5 May 2021).

Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H. and Saidur, R. (2014), “A review on applications of ANN and SVM for building electrical energy consumption

forecasting", *Renewable and Sustainable Energy Reviews*, Vol. 33, pp. 102-109, doi: 10.1016/j.rser.2014.01.069.

Alves, D.C.O. and Bueno, R.L.S. (2003), "Short-run, long-run and cross elasticities of gasoline demand in Brazil", *Energy Economics*, Vol. 25 No. 2, pp. 191-199, doi: 10.1016/S0140-9883(02)00108-1.

ANP-National Agency of Petroleum, Natural Gas and Biofuels (2011), "Resolution no. 67, of 9.12.2011", available at: http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2011/dezembro&item=ranp-67-2011 (accessed 31 October 2020).

ANP-National Agency of Petroleum, Natural Gas and Biofuels (2020), "Resolution no. 819, of 6.5.2020", available at: www.in.gov.br/en/web/dou/-/resolucao-n-819-de-5-de-junho-de-2020-260556565 (accessed 20 June 2021).

ANP-National Agency of Petroleum, Natural Gas and Biofuels (2021a), "Fuel sales in 2020 dropped 5.97% compared to 2019 due to the pandemic", available at: www.gov.br/anp/pt-br/canais_atendimento/imprensa/noticias-comunicados/comercializacao-de-combustiveis-em-2020-teve-queda-de-5-97-na-comparacao-com-2019-devido-a-pandemia (accessed 20 May 2021).

ANP-National Agency of Petroleum, Natural Gas and Biofuels (2021b), "Statistical data", available at: www.gov.br/anp/pt-br/centrais-de-conteudo/dados-estatisticos (accessed 5 May 2021).

BACEN-Central Bank of Brazil (2021), "Time series management system", available at: www3.bcb.gov.br/sgspub (accessed 5 May 2021).

Barros, C.P., Gil-Alana, L.A. and Wanke, P. (2014), "Ethanol consumption in Brazil: empirical facts based on persistence, seasonality and breaks", *Biomass and Bioenergy*, Vol. 63, pp. 313-320, doi: 10.1016/j.biombioe.2014.02.012.

Bergmeir, C., Hyndman, R.J. and Koo, B. (2018), "A note on the validity of cross-validation for evaluating autoregressive time series prediction", *Computational Statistics and Data Analysis*, Vol. 120, pp. 70-83, doi: 10.1016/j.csda.2017.11.003.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008), *Time Series Analysis: Forecasting and Control*, 4th ed., John Wiley and Sons, Hoboken, NJ.

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984), *Classification and Regression Trees*, CRC Press, New York, NY.

Bueno, R.L.S. (2018), *Econometria de Séries Temporais*, 2nd ed., Cengage Learning, São Paulo.

Burnquist, H.L. and Bacchi, M.R.P. (2002), "A demanda por gasolina no Brasil: uma análise utilizando técnicas de co-integração", paper presented at XL Congresso Brasileiro de Economia e Sociologia Rural, Passo Fundo, Rio Grande do Sul.

Cardoso, L.C.B., Bittencourt, M.V.L., Litt, W.H. and Irwin, E.G. (2019), "Biofuels policies and fuel demand elasticities in Brazil", *Energy Policy*, Vol. 128, pp. 296-305, doi: 10.1016/j.enpol.2018.12.035.

Castro, G.R. (2012), "Modelo de previsão de demanda por combustíveis automotivos no Brasil", Graduation conclusion work, Universidade Federal do Rio de Janeiro, available at: http://hdl.handle.net/11422/1867

Chang, P.C., Fan, C.Y. and Lin, J.J. (2011), "Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach", *International Journal of Electrical Power and Energy Systems*, Vol. 33 No. 1, pp. 17-27, doi: 10.1016/j.ijepes.2010.08.008.

Che, J. and Wang, J. (2014), "Short-term load forecasting using a kernel-based support vector regression combination model", *Applied Energy*, Vol. 132, pp. 602-609, doi: 10.1016/j.apenergy.2014.07.064.

Chen, Y., Xu, P., Chu, Y., Li, W., Wu, Y., Ni, L., Bao, Y. and Wang, K. (2017), "Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings", *Applied Energy*, Vol. 195, pp. 659-670, doi: 10.1016/j.apenergy.2017.03.034.

Clemen, R.T. (1989), "Combining forecasts: a review and annotated bibliography", *International Journal of Forecasting*, Vol. 5 No. 4, pp. 559-583.

CNPE-National Energy Policy Council (2017), "Resolution no. 17, of 8.6.2017", available at: http://legislacao.anp.gov.br/?path=legislacao-federal/resolucoes/resol-cnpe/2017 (accessed 19 August 2020).

CNPE-National Energy Policy Council (2020), "Resolution no. 8, of 18.8.2020", available at: www.mme.gov.br/documents (accessed 19 August 2020).

Coin, D. (2008), "Testing normality in the presence of outliers", *Statistical Methods and Applications*, Vol. 17 No. 1, pp. 3-12, doi: 10.1007/s10260-007-0046-8.

Costa, C.C., Burnquist, H.L., Souza, M.J.P. and Valdez, C. (2017), "Modeling fuel demand in Brazil: the importance of the flex-fuel vehicle", *Revista Econômica Do Nordeste*, Vol. 48 No. 3, pp. 113-127.

CRBIO-RenovaBio Commitment (2020), "Public consultation no. 94 of June 5, 2020", available at: www.mme.gov.br/web/guest/conselhos-e-comites/crbio (accessed 19 August 2020).

Deb, C., Zhang, F., Yang, J., Lee, S.E. and Shah, K.W. (2017), "A review on time series forecasting techniques for building energy consumption", *Renewable and Sustainable Energy Reviews*, Vol. 74, pp. 902-924, doi: 10.1016/j.rser.2017.02.085.

Debnath, K.B. and Mourshed, M. (2018), "Forecasting methods in energy planning models", *Renewable and Sustainable Energy Reviews*, Vol. 88, pp. 297-325, doi: 10.1016/j.rser.2018.02.002.

Demsar, J. (2006), "Statistical comparisons of classifiers over multiple data sets", *Journal of Machine Learning Research*, Vol. 7, pp. 1-30.

Dickey, D.A. and Pantula, S.G. (1987), "Determining the order of differencing in autoregressive processes", *Journal of Business and Economic Statistics*, Vol. 5 No. 4, pp. 455-461, doi: 10.1080/07350015.1987.10509614.

Drabik, D., Gorter, H., Just, D.R. and Timilsina, G.R. (2015), "The economics of Brazil's ethanol-sugar markets, mandates, and tax exemptions", *American Journal of Agricultural Economics*, Vol. 97 No. 5, pp. 1433-1450, doi: 10.1093/ajae/aau109.

Du, X. and Carriquiry, M.A. (2013), "Flex-fuel vehicle adoption and dynamics of ethanol prices: lessons from Brazil", *Energy Policy*, Vol. 59, pp. 507-512, doi: 10.1016/j.enpol.2013.04.008.

Dudek, G. (2015), "Short-Term load forecasting using random forests. Intelligent systems' 2014", in Filev, D., Jabłkowski, J., Kacprzyk, J., Krawczak, M., Popchev, I., Rutkowski, L., Sgurev, V., Sotirova, E., Szynkarczyk, P. and Zadrozny, S. (Eds), *Advances in Intelligent Systems and Computing*, Springer, Cham, pp. 821-828, doi: 10.1007/978-3-319-11310-4_71.

Ebrahimpour, R., Nikoo, H., Masoudnia, S., Yousefi, M.R. and Ghaemi, M.S. (2011), "Mixture of mlp-experts for trend forecasting of time series: a case study of the Tehran stock exchange", *International Journal of Forecasting*, Vol. 27 No. 3, pp. 804-816, doi: 10.1016/j.ijforecast.2010.02.015.

Edgerton, D. and Shukur, G. (1999), "Testing autocorrelation in a system perspective testing autocorrelation", *Econometric Reviews*, Vol. 18 No. 4, pp. 343-386, doi: 10.1080/07474939908800351.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996), "Efficient tests for an autoregressive unit root", *Econometrica*, Vol. 64 No. 4, pp. 813-836, doi: 10.2307/2171846.

Engle, R.F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, Vol. 50 No. 4, pp. 987-1007, doi: 10.2307/1912773.

Fan, S., Chen, L. and Lee, W.J. (2008), "Machine learning based switching model for electricity load forecasting", *Energy Conversion and Management*, Vol. 49 No. 6, pp. 1331-1344, doi: 10.1016/j.enconman.2008.01.008.

Figueira, S.R., Burnquist, H.L. and Bacchi, M.R.P. (2010), "Forecasting fuel ethanol consumption in Brazil by time series models: 2006-2012", *Applied Economics*, Vol. 42 No. 7, pp. 865-874, doi: 10.1080/00036840701720978.

Figueira, S.R.F., Santos, D.F.L., Borges, A.C.G., Sanches, A. and Campos, A.L.S. (2014), "Demand for automotive fuel in Brazil: an empirical analysis using cointegration techniques", *Business and Management Review*, Vol. 4 No. 3, pp. 343-356.

Freitas, L.C. and Kaneko, S. (2011), "Ethanol demand under the flex-fuel technology regime in Brazil", *Energy Economics*, Vol. 33 No. 6, pp. 1146-1154, doi: 10.1016/j.eneco.2011.03.011.

García Nieto, P.J., Lasheras, F.S., García-Gonzalo, E. and Cos Juez, F.J. (2018), "PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: a case study", *Science of the Total Environment*, Vol. 621, pp. 753-761, doi: 10.1016/j.scitotenv.2017.11.291.

Gomez, J.M.A. and Legey, L.F.L. (2015), "An analysis of the impact of Flex-Fuel vehicles on fuel consumption in Brazil, applying cointegration and the Kalman filter", *Energy*, Vol. 81, pp. 696-705, doi: 10.1016/j.energy.2015.01.015.

Hastie, T., Tibshirani, R. and Frieman, J. (2009), *The Elements of Statistical Learning*, 2nd ed., Springer, New York, NY.

Haykin, S. (2001), *Redes Neurais: Princípios e Prática*, 2nd ed., Bookman, Porto Alegre.

Hong, W.-C. (2009), "Electric load forecasting by support vector model", *Applied Mathematical Modelling*, Vol. 33 No. 5, pp. 2444-2454, doi: 10.1016/j.apm.2008.07.010.

Hyndman, R.J. and Athanasopoulos, G. (2018), *Forecasting: Principles and Practice*, 2nd ed., Otexts, Melbourne.

Ibarra-Berastegi, G., Saénz, J., Esnaola, G., Ezcurra, A. and Ulazia, A. (2015), "Short-term forecasting of the wave energy flux: analogues, random forests, and physics-based models", *Ocean Engineering*, Vol. 104, pp. 530-539, doi: 10.1016/j.oceaneng.2015.05.038.

IBGE-Brazilian Institute of Geography and Statistics (2021), "Population statistics", available at: www.ibge.gov.br/estatisticas/sociais/populacao (accessed 5 May 2021).

Isabella, G., Yu, A.S.O., Silva, A.M. and Pegetti, A.L. (2017), "Another driver of the Brazilian fuel ethanol supply chain: the consumers' preferences", *Revista de Administração*, Vol. 52 No. 3, pp. 304-316, doi: 10.1016/j.rausp.2017.05.003.

Jang, J.-S.R., Sun, C.-R. and Mizutani, E. (1997), *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, NJ.

Johansen, S. (1988), "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control*, Vol. 12 Nos 2/3, pp. 231-254, doi: 10.1016/0165-1889(88)90041-3.

Johansen, S. (1991), "Estimation and hypothesis testing of cointegration vectors in Gaussian vector", *Econometrica*, Vol. 59 No. 6, pp. 1551-1580, doi: 10.2307/2938278.

Kandananond, K. (2011), "Forecasting electricity demand in Thailand with an artificial neural network approach", *Energies*, Vol. 4 No. 8, pp. 1246-1257, doi: 10.3390/en4081246.

Kane, M.J., Price, N., Scotch, M. and Rabinowitz, P. (2014), "Comparison of ARIMA and random forest time series models for prediction of avian influenza H5N1 outbreaks", *BMC Bioinformatics*, Vol. 15 No. 1, doi: 10.1186/1471-2105-15-276.

Kavaklioglu, K., Ceylan, H., Ozturk, H.K. and Canyurt, O.E. (2009), "Modeling and prediction of turkey's electricity consumption using artificial neural networks", *Energy Conversion and Management*, Vol. 50 No. 11, pp. 2719-2727, doi: 10.1016/j.enconman.2009.06.016.

Khashei, M. and Bijari, M. (2010), "An artificial neural network (p, d, q) model for timeseries forecasting", *Expert Systems with Applications*, Vol. 37 No. 1, pp. 479-489, doi: 10.1016/j.eswa.2009.05.044.

Kim, J. and Kasabov, N. (1999), "HyFIS: adaptive neuro-fuzzy inference systems e their application to nonlinear dynamical systems", *Neural Networks*, Vol. 12 No. 9, pp. 1301-1319, doi: 10.1016/S0893-6080(99)00067-2.

Kucukali, S. and Baris, K. (2010), "Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach", *Energy Policy*, Vol. 38 No. 5, pp. 2438-2445, doi: 10.1016/j.enpol.2009.12.037.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?", *Journal of Econometrics*, Vol. 54 Nos 1/3, pp. 159-178, doi: 10.1016/0304-4076(92)90104-Y.

Lahouar, A. and Ben Hadj Slama, J. (2015), "Day-ahead load forecast using random forest and expert input selection", *Energy Conversion and Management*, Vol. 103, pp. 1040-1051, doi: 10.1016/j.enconman.2015.07.041.

Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis*, 2nd ed., Springer, Berlin.

Osbom, D.R., Chui, A.P.L., Smith, J.P. and Birchenhall, C.R. (1988), "Seasonality and the order of integration for consumption", *Oxford Bulletin of Economics and Statistics*, Vol. 50 No. 4, pp. 361-377, doi: 10.1111/j.1468-0084.1988.mp50004002.x.

Pacini, H. and Silveira, S. (2011), "Consumer choice between ethanol and gasoline: lessons from Brazil and Sweden", *Energy Policy*, Vol. 39 No. 11, pp. 6936-6942, doi: 10.1016/j.enpol.2010.09.024.

Padmakumari, K., Mohandas, K.P. and Thiruvengadam, S. (1999), "Long term distribution demand forecasting using neuro fuzzy computations", *International Journal of Electrical Power and Energy Systems*, Vol. 21 No. 5, pp. 315-322, doi: 10.1016/S0142-0615(98)00056-8.

Pai, P.F. (2006), "Hybrid ellipsoidal fuzzy systems in forecasting regional electricity loads", *Energy Conversion and Management*, Vol. 47 Nos 15/16, pp. 2283-2289, doi: 10.1016/j.enconman.2005.11.017.

Ringwood, J.V., Bofelli, D. and Murray, F.T. (2001), "Forecasting electricity demand on short, medium and long time scales using neural networks", *Journal of Intelligent and Robotic Systems*, Vol. 31 Nos 1/3, pp. 129-147, doi: 10.1023/A:1012046824237.

Rodrigues, L. and Bacchi, M.R.P. (2016), "Light fuel demand and public policies in Brazil, 2003-2013", *Applied Economics*, Vol. 48 No. 54, pp. 5300-5313, doi: 10.1080/00036846.2016.1176115.

Rodrigues, L. and Bacchi, M.R.P. (2017), "Analyzing light fuel demand elasticities in Brazil using cointegration techniques", *Energy Economics*, Vol. 63, pp. 322-331, doi: 10.1016/j.eneco.2017.02.012.

Rodrigues, L. and Rodrigues, L. (2018), "Economic-financial performance of the Brazilian sugarcane energy industry: an empirical evaluation using financial ratio, cluster and discriminant analysis", *Biomass and Bioenergy*, Vol. 108, pp. 289-296, doi: 10.1016/j.biombioe.2017.11.013.

Salvo, A. and Huse, C. (2013), "Build it, but will they come? Evidence from consumer choice between gasoline and sugarcane ethanol", *Journal of Environmental Economics and Management*, Vol. 66 No. 2, pp. 251-279, doi: 10.1016/j.jeem.2013.04.001.

Santiago, F.S., Mattos, R.S. and Perobelli, F.S. (2011), "Um modelo integrado econométrico+insumo-produto Para previsão de longo prazo da demanda de combustíveis no Brasil", *Nova Economia*, Vol. 21 No. 3, pp. 423-455, doi: 10.1590/S0103-63512011000300005.

Santos, G.F. (2013), "Fuel demand in Brazil in a dynamic panel data approach", *Energy Economics*, Vol. 36, pp. 229-240, doi: 10.1016/j.eneco.2012.08.012.

Santos, A.S., Gilio, L., Halmenschlager, V., Diniza, T.B. and Almeida, A.N. (2018), "Flexible-fuel automobiles and CO2 emissions in Brazil: parametric and semiparametric analysis using panel data", *Habitat International*, Vol. 71, pp. 147-155, doi: 10.1016/j.habitatint.2017.11.014.

Shapiro, S.S. and Wilk, M.B. (1965), "An analysis of variance test for normality (complete samples)", *Biometrika*, Vol. 52 Nos 3/4, pp. 591-611, doi: 10.2307/2333709.

Shingala, M.C. (2016), "Comparison of different post HOC tests", Thesis for the degree of Doctor of Philosophy, available at: http://hdl.handle.net/10603/171827

Sözen, A., Arcaklioğlu, E. and Özkaymak, M. (2005), "Turkey's net energy consumption", *Applied Energy*, Vol. 81 No. 2, pp. 209-221, doi: 10.1016/j.apenergy.2004.07.001.

Suganthi, L. and Samuel, A.A. (2012), "Energy models for demand forecasting – a review", *Renewable and Sustainable Energy Reviews*, Vol. 16 No. 2, pp. 1223-1240, doi: 10.1016/j.rser.2011.08.014.

UNICA-Brazilian Sugarcane Industry Association (2020), "Brazilian automobile and light fuel fleet (Otto cycle)", available at: https://observatoriodacana.com.br/listagem.php?idMn=55&idioma=2 (accessed 31 January 2021).

Wang, L.X. and Mendel, J.M. (1992), "Generating fuzzy rules by learning from examples", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22 No. 6, pp. 1414-1427, doi: 10.1109/21.199466.

Zanini, A., Souza, R.C. and Pedreira, C.E. (2001), "Um modelo híbrido para previsão de curto prazo da demanda de gasolina automotiva no Brasil", Paper presented at: *Proceedings of the V Brazilian Conference on Neural Networks*, *April 2-5*, *Rio de Janeiro*, doi: 10.21528/CBRN2001-086.

Corresponding author
Lucas Rodrigues can be contacted at: lucas2.rodrigues@outlook.com

Table A1. Summary of accuracy metrics of methods for forecasting 12, 6 and 3 months ahead

		All period		Before COVID-19 first wave*					After COVID-19 first wave**					
	Model	Average MAPE	Average MAPE***	Average MAPE (Training)	MAPE standard deviation	Max MAPE	Min MAPE	Max EP	Average MAPE***	Average MAPE (Training)	MAPE standard deviation	Max MAPE	Min MAPE	Max EP
12-months forecast	SVM	3.66	2.80 (2)	2.12	1.01	5.67	1.47	11.61	9.02 (2)	2.11	2.01	10.71	3.65	35.97
	RLS	3.68	2.85 (4)	2.10	1.05	5.94	1.46	11.86	8.83 (1)	2.08	1.96	10.49	3.59	35.06
	AVG	3.70	2.78 (1)	2.33	0.96	5.59	1.52	11.20	9.41 (3)	2.09	2.08	11.20	3.78	38.08
	RF	3.79	2.82 (3)	2.62	0.94	6.58	1.38	11.15	9.82 (4)	2.56	2.17	11.50	3.97	40.13
	AR.Bin	4.02	3.00 (5)	2.04	1.21	6.96	1.28	14.37	10.42 (7)	2.09	3.11	14.65	3.51	46.07
	ANN	4.15	3.19 (7)	2.20	1.30	6.96	1.54	15.74	10.12 (5)	2.21	2.25	12.61	4.28	42.26
	HYFIS	4.18	3.19 (8)	2.96	1.35	7.14	1.21	12.81	10.37 (6)	2.85	2.29	14.06	4.51	43.52
	VECM	4.19	3.05 (6)	1.99	1.21	7.25	1.33	16.89	11.26 (10)	2.06	3.56	16.37	3.09	44.53
	SAR	4.59	3.34 (9)	2.16	1.78	8.62	1.22	19.80	12.33 (11)	2.18	3.89	17.20	3.16	48.36
	Naive	5.33	4.46 (10)	–	1.58	7.37	1.98	13.23	10.75 (8)	–	1.89	12.36	5.44	40.78
	MANN	5.48	4.65 (11)	2.34	1.83	11.61	1.71	19.45	10.76 (9)	2.39	2.71	13.21	3.25	46.71
6-months forecast	AVG	3.51	2.38 (1)	1.94	0.92	5.02	0.93	10.16	10.97 (2)	2.00	6.15	23.30	5.13	43.25
	RLS	3.58	2.39 (2)	1.94	0.91	4.98	0.93	10.18	11.52 (3)	2.00	5.88	23.20	5.13	43.04
	AR.Bin	3.68	2.49 (3)	2.05	1.07	5.66	0.79	12.43	11.60 (4)	2.24	6.64	24.71	4.90	46.07
	SVM	3.71	2.50 (4)	2.12	0.82	4.48	1.11	8.94	11.72 (5)	2.17	5.00	19.69	5.27	35.97
	ANN	3.88	2.58 (6)	2.20	1.02	5.65	1.00	8.94	12.56 (7)	2.53	6.46	23.69	4.77	43.16
	RF	3.89	2.58 (7)	2.62	0.81	5.73	1.28	9.97	12.59 (8)	2.65	4.93	21.34	5.66	40.07
	HYFIS	4.00	2.74 (9)	2.62	0.97	5.34	0.90	10.22	12.35 (6)	2.60	6.11	24.00	5.02	42.92
	SAR	4.01	2.56 (5)	2.16	1.36	7.13	0.84	13.86	13.59 (9)	2.36	6.80	26.33	4.95	48.36
	VECM	4.04	2.59 (8)	1.99	1.07	5.17	0.79	10.10	13.66 (10)	2.21	6.21	25.45	5.07	45.04
	MANN	4.40	3.00 (10)	3.26	1.38	7.33	0.88	11.14	13.75 (11)	3.47	6.52	24.44	4.46	42.77
	Naive	5.37	4.54 (11)	–	1.86	8.64	1.24	13.23	10.88 (1)	–	6.34	20.72	3.62	40.78
	RLS	3.17	2.19 (1)	1.95	1.08	5.79	0.86	10.09	9.84 (2)	2.05	7.85	31.72	2.41	43.11
	AVG	3.12	2.19 (2)	1.95	1.07	5.80	0.88	10.20	9.52 (1)	2.05	8.03	31.80	2.43	43.27
3-months forecast	AR.Bin	3.30	2.21 (3)	2.05	1.21	6.05	0.56	12.43	10.73 (4)	2.34	8.46	33.14	2.17	46.07
	SVM	3.32	2.33 (7)	2.13	0.99	5.54	0.68	7.97	10.15 (3)	2.23	6.27	26.40	2.51	36.15
	SAR	3.43	2.23 (4)	2.16	1.26	7.32	0.46	13.22	11.68 (8)	2.47	9.09	34.29	1.48	47.35
	RF	3.48	2.40 (8)	2.60	1.09	6.22	0.66	10.06	10.91 (5)	2.74	7.79	30.08	2.05	40.21
	VECM	3.49	2.31 (6)	1.99	1.23	5.63	0.63	9.19	11.61 (7)	2.32	7.82	33.57	2.83	44.79
	ANN	3.63	2.29 (5)	2.21	1.14	6.08	0.36	8.96	12.81 (10)	2.62	9.84	31.01	1.97	45.37

(*continued*)

	All period		Before COVID-19 first wave*					After COVID-19 first wave**				
Model	Average MAPE	Average MAPE***	Average MAPE (Training)	MAPE standard deviation	Max MAPE	Min MAPE	Max EP	Average MAPE***	Average MAPE (Training)	MAPE standard deviation	Max MAPE	Min MAPE	Max EP
HYFIS	3.71	2.53 (9)	2.62	1.14	5.45	0.70	9.46	11.77 (9)	2.68	8.80	32.59	3.05	42.92
MANN	4.74	2.92 (10)	3.31	1.56	9.20	0.79	20.34	17.15 (11)	3.48	11.89	36.49	1.55	53.48
Naive	5.41	4.55 (11)	–	2.20	9.54	0.33	13.23	11.27 (6)	–	9.13	28.80	2.22	40.78

Notes: *Accuracy metrics were calculated with data between January 2002 and February 2020. **Accuracy measures were calculated using all time points whose predicted horizon contains months from March 2020. ***Number in parentheses indicates the ranking of the model by Average MAPE value

Table A1.