UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS

GUILHERME SCHULTZ LINDENMEYER

BOOSTING AND PREDICTABILITY OF MACROECONOMIC VARIABLES: EVIDENCE FROM BRAZIL

Porto Alegre
2022

GUILHERME SCHULTZ LINDENMEYER

BOOSTING AND PREDICTABILITY OF MACROECONOMIC VARIABLES: EVIDENCE FROM BRAZIL

Work submitted to Faculty of Economics at UFRGS as partial requirement for obtaining the Bachelor's Degree in Economics.

Supervisor: Prof. Dr. Hudson da Silva Torrent

Porto Alegre
2022

CIP - Catalogação na Publicação

Schultz Lindenmeyer, Guilherme
Boosting and Predictability of Macroeconomic Variables: Evidence from Brazil / Guilherme Schultz Lindenmeyer. -- 2022.
55 f.
Orientador: Hudson da Silva Torrent.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Curso de Ciências Econômicas, Porto Alegre, BR-RS, 2022.

1. Boosting. 2. Econometria. 3. Não-linear. 4. Previsão. 5. Séries Temporais Macroeconômicas. I. da Silva Torrent, Hudson, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

GUILHERME SCHULTZ LINDENMEYER

BOOSTING AND PREDICTABILITY OF MACROECONOMIC VARIABLES: EVIDENCE FROM BRAZIL

Work submitted to Faculty of Economics at UFRGS as partial requirement for obtaining the Bachelor's Degree in Economics.

Approved in: Porto Alegre, 11th of May, 2022.

THESIS DEFENSE COMMITTEE:

Prof. Dr. Hudson da Silva Torrent – Advisor
UFRGS

Prof. Dr. Cleiton Guollo Taufemback
UFRGS

Prof. Dr. Nelson Seixas dos Santos
UFRGS

ACKNOWLEDGEMENTS

First, I would like to thank my family, in particular my mother and father for their support in helping me make time for my studies. Without them, I would have no time for such studies or food on the table.

I also thank UFRGS and the Department of Economics for the opportunity to be their student as well as for the opportunity to do an exchange program. Also, for the experience in Equilíbrio, the junior enterprise of Economics of which I was glad to be part.

And I thank all the friendships that UFRGS gave me and/or strengthened. My experience at the university would be not nearly as much enjoyable without you. Special thanks to Angelo and Pedro Pablo. Super special thanks to the best partner I could possibly have thanks to UFRGS: Gabrielle. All the best for us.

Finally, my immense thanks to my professor and advisor Dr. Hudson Torrent, not only for being an excellent professor and guide, but also as a great person. I owe all my research experience gained to him.

*"A beginning is the time for taking the most delicate care
that the balances are correct."*
— Frank Herbert, Dune

ABSTRACT

This paper aims to elaborate a treated data set and apply the boosting methodology to monthly Brazilian macroeconomic variables to check its predictability. The forecasting performed here consists in using linear and nonlinear base-learners, as well as a third type of model that has both linear and nonlinear components in the estimation of the variables using the history itself with lag up to 12 periods. The results obtained here through different evaluation approaches point out that, on average, the performance of boosting models using P-Splines as base-learner are the ones that have the best results, especially the methodology with two components: two-stage boosting. Finally, we perform alternative methods to check the robustness of the results.

Keywords: Boosting. Econometrics. Forecasting. Macroeconomic Time Series. Nonlinear.

RESUMO

Este trabalho visa elaborar um conjunto de dados tratados e aplicar a metodologia L_2Boosting às variáveis macroeconômicas mensais brasileiras selecionadas para verificar sua previsibilidade. A previsão aqui realizada consiste no uso de estimações de base lineares e não lineares, assim como um terceiro tipo de modelo que tem componentes lineares e não lineares na estimativa das variáveis usando o próprio histórico com defasagem de até 12 períodos. Os resultados obtidos aqui através de diferentes abordagens de avaliação apontam que, em média, o desempenho dos modelos de boosting usando P-Splines como *base-learner* são os que têm os melhores resultados, especialmente a metodologia com dois componentes: o boosting em dois estágios. Finalmente, realizamos métodos alternativos para verificar a robustez dos resultados.

Palavras-chave: Boosting. Econometria. Não-linear. Previsão. Séries Temporais Macroeconômicas.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

In the book A History of Econometrics by R.J. Epstein, the author writes about the trajectory and directions of econometrics as a research field and science (EPSTEIN, 1987). The book is considered one of the first in its field to write something about the history of econometrics, as until then there were few works dealing with its history. At this time, they used little statistical analysis, and the major emphases in the works were historical narratives or theoretical mathematical models. It is only in the 20th century that more in-depth research and data analysis begins[1]. Nevertheless, almost the entire book's discussion is about structural models and their importance in the field, leaving very little space for time series analysis. The author comments that this was because, until then, little had been developed in such areas and mentions, "I suspect that current developments in the theory of time series ultimately will have a profound influence on our conception of econometrics but this lies outside my 'sample period'." (EPSTEIN, 1987). Time series analysis, according to Epstein, began to gain momentum only after the 1970s.

As we move to the 21st century, much of the methodology narrated by Epstein in his book is still widely used by many economists in empirical work and analysis. The trend that has been taking place is that, in parallel with the advances in economic and econometric theories, the sophistication of statistics has occurred, mainly thanks to advances in processing power and the development of computers to perform calculations. Several areas of science are increasingly making use of machine learning and Data Science models. Economics is no different.

Advances in technology and databases, especially improvements after the 1990s, provide more resources for economists, analysts and researchers to use. In the past it was extremely difficult to use large data sets for empirical work, but with the popularization of computers, free software, and open access, there are huge amounts of data at our disposal without us having to leave the computer to curate it. This allows us to work with highly dimensional data sets, i.e. many variables and many observations. In the paper entitled "Big Data: New Tricks for Econometrics" by Varian (2014), the author mentions "Economists have historically dealt with data that fits in a spreadsheet, but that is changing as new more detailed data become available.". Therefore, it has already been observed that the way of dealing with econometrics is changing[2].

Within the universe of machine learning, this work proposes further exploration and execution of applications with the component-wise boosting method. Instead of aiming to obtain enormous predictive power, as is the case of neural network models or random forests, these component-wise models allow us to maintain some level of interpretability without leaving aside

---

[1] The starting point author is Henry Moore, considered by Epstein as the founder of econometrics with his 1911's work, in which he sought to statistically verify the marginal productivity of wages to support public policies (MOORE, 1911).

[2] The article by Medeiros et al. (2019) is an example of how econometrics is changing through an extensive study on the benefits of using many machine learning models and presenting, in an applied way, their advantages for the field. To make the comparisons between the models, regardless of their origin, they perform the exercise of forecasting the US monthly general inflation using an extensive data set created by them.

the importance of prediction. This is done, for the case of L_2Boosting, which is the targeted model, for instance, through the use of linear learners in its parameters. It is a machine learning model that is increasingly gaining space in the economic literature. The work of Buchen and Wohlrabe (2011) is one of the first papers to consider the boosting methodology as an alternative model to be used in econometric work, in this case to forecast US industrial production growth. Later, the performance of the method in a high-dimensional macroeconomic context was verified by Wohlrabe and Buchen (2014). Also, more empirical work was done with the model for the case of Germany (ROBINZONOV; TUTZ; HOTHORN, 2012; LEHMANN, R.; WOHLRABE, K., 2016), GDP of Japan (YOON, 2021) and case of economic variables from the United States (MEDEIROS et al., 2019; ZENG, 2017; KAUPPI; VIRTANEN, 2021). Finally, the authors Lindenmeyer, Skorin, and Torrent (2021) were one of the first to apply the methodology to the Brazilian case, verifying the model's performance in forecasting monthly electricity consumption in a Brazilian state. We aim to improve the application of the method in the Brazilian scenario.

With the existence of several empirical papers using the boosting method for forecasting, whether in economics or outside the field, this paper seeks to take a step back and enrich the knowledge before the forecasting itself. Therefore, the objective of this paper, however, is not the adoption of this algorithm and its application for macroeconomic forecasting. It is, on the other hand, the application of the boosting methodology in its state of the art to Brazilian macroeconomic variables typically used for economic forecasting in order to analyze, in fact, how predictable it is within its own history. By doing this, we can get a more fundamental sense of how the Brazilian variables actually behave. In this way, it is possible to assign a better prediction model instead of using the same model without having any indication of whether it would be the best. This investigation follows the logic coming from the development of boosting methodology when it comes specifically to economic variables and time series. Therefore, we follow the research from the article Kauppi and Virtanen (2021) where they specifically study the nonlinearities of macroeconomic time series from the United States. With this paper, our aim is to bring this analysis to the Brazilian case, as well as expand the view to apply the linear and component-wise L_2Boosting algorithm and the version of the algorithm obtained from both splines and boosted trees as gradient descent functions. We wish to investigate both the predictability of Brazilian macroeconomic variables and verify whether they are in general linear or nonlinear and to extend the literature on applications of the algorithm in economics.

In order to achieve this goal, we must elaborate a data set of the Brazilian macroeconomic scenario. To do this, it is vital to use variables that are up-to-date and important for economic research. For this reason, we have created a data set based on Brazilian empirical research and in a way that follows the selection logic of other papers that do macroeconomic forecasts. With the data in hand, it is of major importance to transform each of its variables into stationary series, using appropriate methods. With this, a code is developed to adapt the methodology and carry out the data treatment, the modeling and the analysis. To this end, we use the R programming language within the RStudio free software development environment (R CORE TEAM, 2019;

RSTUDIO, 2020). Furthermore, part of the boosting methodology has been adapted to R thanks to the “mboost” library (HOTHORN et al., 2011). The code developed in this paper to perform the forecasts departs from the code created by the authors Kauppi and Virtanen (2021). All our code can be found in our digital repository on GitHub[3].

Given the database created for the application of the methodology, we then did the estimation for a pseudo out-of-sample forecasting with forecast horizon h from 1 to 12, using performance indicators such as the root mean square error and the empirical coefficient of determination. We applied several models, among them: the traditional linear model, linear boosting model and nonlinear boosting (splines and regression trees as base-learner). We used an intermediate model, which has both a linear and a nonlinear component. This model, which, on average for our entire database, obtained the best results according to most of the evaluation indicators used. Finally, we also performed robustness checks, comparing our models with alternative scenarios and comparing two methods of selecting the stopping criterion of the boosting algorithm.

The paper is organized as follows: Chapter 2 we review the literature, Chapter 3 we present all the methodology used here in this paper; Chapter 4 we show details of the data collection, as well as its treatment, and explain how we performed the prediction; Chapter 5 we discuss our results; and finally, Chapter 6 we conclude this paper with a general discussion of the results obtained.

[3] https://github.com/guislinden/Research_UFRGS.

2 LITERATURE REVIEW

To understand how the L_2Boosting method originated, we must first illustrate how the more general methodology, called boosting, arose[4]. The method was first conceived and introduced by Schapire (1990), Freund (1995) and Freund and Schapire (1996) as an Adaptive Boosting algorithm (AdaBoost), and these papers perform the first fundamental step towards the use of feasible boosting algorithms. The AdaBoost methodology introduced by Freund and Schapiro serves as a machine learning procedure for improving the prediction of binary outcomes by averaging predictions from reweighted data. Since then, the algorithm has attracted a lot of attention in the scientific community, mainly because of its good empirical performance on a variety of data sets. Later, in the work of Breiman (1998, 1999), the author, based on the AdaBoost approach, observed that the algorithm can be used as a gradient descent[5] in function space, and through this observation, allowed the use of boosting beyond classification contexts. The algorithm had advances for many different application contexts and areas, so we will continue the discussion with emphasis on the ones that suit our purposes best. Author Friedman (2001) adapted the existing boosting methodology to apply to regression analysis. In this new methodology, the algorithm is used to optimize a squared error loss function, and this creates the methodology we will use in this paper, called L_2Boosting.

For large scale applications, where the number of variables exceeds the number of observations, classical statistical models such as Ordinary Least Squares (OLS) lose their ability to estimate consistent parameters. Bühlmann and Yu (2003) adapt the existing L_2 Boosting methodology until then so that a preselection of variables can be performed to enable correct estimation of a boosted linear regression model. Later, Bühlmann (2006) mathematically proves the consistency of L_2Boosting in applications with large data sets. After all these scientific contributions, we have the model as pointed out by Schmid and Hothorn (2008),

> [. . .] when the number of covariates p in a data set is large (and when selecting a small number of relevant covariates is desirable), boosting is usually superior to standard estimation techniques for regression models (such as backward step-wise linear regression, which, e.g., cannot be applied if p is larger than the number of observations n). (pg.2)

In that paper by Schmid and Hothorn, the authors start from Bühlmann and Yu (2003) and perform the application of L_2Boosting mathematically in two ways: with the base-learner used by Bühlmann and Yu, called smoothing spline, and compare it with the base leaner P-Spline approach conceived by Paul H.C. Eilers and Marx (2010). Schmid and Hothorn conclude that by using P-Spline as base-learners instead of smoothing spline, "[...] the computational effort of component-wise L_2Boosting can be greatly reduced, while there is only a minor effect on

[4] The recent paper by Chu et al. (2020) provides details and models of the different boosting approaches applied also in economic scenarios that we do not address here.

[5] The gradient method, or gradient descent, is a numerical approach used in optimization to find a minimum through an iterative function where, for each step, the negative direction of the gradient is taken.

the predictive performance of the boosting algorithm.". The methodology when dealing with macroeconomic data was improved in Bai and Ng (2009), where the authors perform the first formal adaptation and application of L_2Boosting as a variable selection model with time series. The authors perform tests using regressor variables obtained from Principal Component Analysis (PCA) and also point out: "[...] boosting has the advantage that it does not require *a priori* ordering of the predictors or their lags as conventional model selection procedures do". With this, we arrive at temporally L_2Boosting with its methodology adapted for large data sets, for time series and with both a linear and nonlinear gradient functions. Finally, these authors who developed and proved that the L_2 Boosting model can be used empirically and also they have developed a code library for the R programming and statistical language, which we will use in this work, as already mentioned in Chapter 1.

The use of digital databases and the easy access to them through the Internet allows the elaboration of high-dimensional data sets. The topic of how to perform forecasting exercises with many regressors has been gaining attention in the scientific literature since the 2000s. Stock and Watson (2006) in their paper entitled "Forecasting with Many Predictors" point out that, "Historically, time series forecasts of economic variables have used only a handful of predictor variables, while forecasts based on a large number of predictors have been the province of judgmental forecasts and large structural econometric models." In their paper, the authors prepare a data set with 130 regressors to perform the forecasting exercise of the US industrial production index. For this, they select several models, starting from adaptations to deal with large data sets from classical econometric models such as OLS, and more sophisticated statistical models, such as PCA, dynamic factor models and Bayesian models. The results of the paper emphasize alternative models over OLS, since this model was the only one whose results were worse relative to the autoregressive benchmark.

The discussion regarding forecasting with many variables began to relate to the research line of the boosting methodology. Buchen and Wohlrabe (2011) released the paper "Forecasting with Many Predictors: Is boosting a viable alternative?", that is, a direct response to the problem faced in 2006 by Stock and Watson. In this paper, the authors use the same database and performance evaluation criteria to present the feasibility of the boosting model. The result of the paper is clear, the new methodology is, according to the authors, "[...] a serious contender for forecasting US industrial production." Not only this, but the paper also serves to show that the use of cross-validation rather than the Akaike Information Criteria (AIC) as the stopping criterion for iteration of the boosting algorithm brings significant improvements to the results of the model.

Robinzonov, Tutz, and Hothorn (2012) discuss macroeconomic variables and their non-linearity, as well as the difficulty of selecting lags for forecasting. They offer as a solution the use of boosting. In the paper, the method is used in two ways, linear base-learner, and base-learner with penalized B-Splines. The empirical exercise performed was to forecast monthly German industrial production from 1992 to 2006 with the boosting versus benchmark (AR) methodology.

It is concluded that the boosting method has advantages because of its flexibility, performance and the way variables and lags are selected. The study by Wohlrabe and Buchen (2014) highlights how the use of boosting has been gaining ground in the macroeconomic forecasting literature. In their exercise, the authors develop a database to forecast several macroeconomic indicators for the United States, the Eurozone, and Germany. Overall, the conclusion of the paper emphasizes that the model outperforms its benchmark in almost all scenarios and handles macroeconomic forecasting very well. Furthermore, two other conclusions proposed are related to the maximum number of boosting iterations: the proposal made by Hastie (2007) to use the number of different selected variables instead of the trace when considering the use of AIC brings improvements to forecasting, and the K-fold cross-validation criterion, in general, is dominant among the results.

The role of regional forecasts for policy makers is becoming increasingly important. Of course, such forecasts demand intensive use of local variables, but these variables can suffer from the problems of few observations and a variety of parameters. On the other hand, there is more abundance of international and national variables. The question is, are they considered important, i.e. do they contribute to improved performance, to deal with regional forecasts? In the paper by Kopoin, Moran, and Paré (2013) factor models are used to analyze the impact of national and international variables on regional forecasts. The study is conducted from GDP forecasts of provinces in Canada with quarterly data from 1983Q1 to 2011Q1, plus regional and international variables. The work consists of applying factor model, which are the aggregation of several variables or information into a few factors, and these are used to perform the forecast of the variable of interest, GDP. To verify the impact of different levels of variables, the authors performed several exercises, where for each one they used different data sets, e.g. only regional data, regional and national data, among others. With this, the authors conclude that the use of a database with international and national variables increases the forecast performance, if the horizon is shorter than a year $(h < 12)$.

In Robert Lehmann and Klaus Wohlrabe (2017), the aim of the paper is to use the component-wise boosting model to perform regional GDP forecasting from a database previously assembled by the authors in another paper (LEHMANN, Robert; WOHLRABE, Klaus, 2015) that has data for three German regions: Saxony, Baden-Württemberg and West Germany. The database consists of macroeconomic indicators, price indicators, consumer survey results, international data, and regional data. With this, L_2Boosting is applied, and then its results are verified by comparing it to a benchmark and analyzing the variables selected by the model. Finally, the authors end the research by highlighting the competitiveness of boosting and its applicability. Another macroeconomic application of the methodology took place in Zeng (2017). The work consisted of continuing a line of research that compares forecasting methods for macroeconomic variables and their comparison between country-specific models and models with international predictors, as well as verifying whether forecasts with aggregate data are superior to those with disaggregated data (MARCELLINO; STOCK; WATSON, 2003). In Zeng's paper, the author starts from the conclusion that country-specific forecasts provide better results and tests the feasibility of boosting in the face of the forecasting scenario for macroeconomic variables from 1970 to 2011. Empirical results indicate that using disaggregated data with factor analysis or by selection through boosting results in higher forecasting performances, and the research also considers boosting as a competitive model compared to factor analysis when faced with high-dimensional data.

And as the research has progressed, another recent line of study within boosting has emerged, which is the one we will extend here with this present paper: applications of the methodology in the verification of nonlinear predictability in macroeconomic time series. In Kauppi and Virtanen (2021) there is a study of these nonlinearities in macroeconomic variables in the United States. The importance of the subject is highlighted by the authors,

> While it is often argued that nonlinearity is an inherent feature of macroeconomic time series, linear forecasts have mostly been found to perform better than forecasts based on various nonlinear models. There are cases where nonlinear models have yielded more accurate forecasts than linear models, but it generally remains unclear to what extent and when nonlinear forecasts are likely to be useful in macroeconomic forecasting. (pg.1)

The work consisted in using different boosting approaches to check the predictability of 128 macroeconomic indicators from 1959 to 2016 in the US. With this, the authors show that for a good portion of the selected database, the method can improve the accuracy of the forecast over a linear forecasting approach. They also identified a category of variables where nonlinear modeling is more likely to produce the best results.

An important issue that is first brought up in more detail in Kauppi and Virtanen (2021) on boosting forecasting is multi-step forecasting. The definition of multi-step forecasting is simply forecasting h steps ahead. In the typical context of forecasting monthly economic variables, it is common to deal with forecasting from 1 step ahead up to 12 steps ahead, which consequently results in a year ahead. However, the strategy of how to perform the forecast is not trivial, as there are typically two main ways of forecasting: (i) in a recursive manner, performing the prediction one step at a time up to h times, where h is the number of steps forward that one wishes to obtain the prediction, or (ii) the construction and elaboration of different types of models, one for each type of step forward. That is, in a work consisting of making predictions $h = 1, ..., 12$ steps ahead, we will have in total 12 different models. Each model is trained to perform its respective way of prediction, where usually we use the same data to obtain each model. In this paper, we have based ourselves on the scientific findings of some studies (FINDLEY, 1985; BHANSALI, 1997; ING, 2003) and performed direct forecasting, because in this way the forecast error for all horizons does not depend drastically on the first forecast.

At last, there are papers that address macroeconomic forecasts with a case study of Brazil (CEPNI; GUNEY; SWANSON, 2020; MCNEIL; LETSCHERT, 2005). On the other hand, the scientific literature on the application of variable selection methodologies, such as the use of L_2Boosting, to perform forecasting exercises is still very recent in the Brazilian scenario. In view of this, Lindenmeyer, Skorin, and Torrent (2021) perform the application of the methodology, from a database with 822 regressors from 2002 to 2017, to predict the electricity consumption in the state Rio Grande do Sul at the time of the Brazilian energy crisis. They compare the L_2Boosting algorithm with an autoregressive benchmark and perform the prediction up to 3 horizons ahead. And they conclude by considering the boosting methodology as valid and competitive in the face of short-term forecast scenarios (1 month ahead), since the results were strong compared to the benchmark.

With the development and the growing importance of the boosting methodology, the adaptation of this methodology to deal with time series as well as in analyzing the predictability of macroeconomic variables, and along with the little attention given to Brazil in the literature, we have as objective the elaboration of this work. That is, we will extend one more application example of the methodology as well as study and obtain evidence on the Brazilian case.

3 METHODOLOGY

This chapter presents the models chosen, as well as explains the methodology behind each of them. Boosting models require the selection of some macro parameters in order to perform the forecast. Here we also explain these parameters and the reasoning behind them. Later, in Section 4, we comment in detail on how our forecasting strategy and the application of the models explained here was done.

3.1 LINEAR MODEL

The linear modeling is done in a straightforward manner and its main purpose here is to serve as a benchmark for the other models. As our goal is to check the predictability of each time series from its own history, the linear model is autoregressive. Since our data is monthly, we assume a period of up to 12 lags to be considered in the model, so the autoregressive model AR(p) has $p = 0, 1, ..., 12$. Let y_t be the time series we would like to model, then the linear model AR(p) has the following format for a chosen p:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t \tag{1}$$

where β_0 is a constant, β_i are the coefficients for each lag and ε_t is the error term. From Eq. 1 we can see that we obtain 13 models. Since we want to keep only one as a benchmark, we select the best AR(p) model from commonly used model selection strategies: Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) (AKAIKE, 1973; SCHWARZ, 1978).

3.2 BOOSTING

In this section we will comment in a general way on how the Boosting algorithm works, and in the following subsections we will talk about the specifics of each case. For simplicity, we leave aside the variable h for forecast horizons, as their role in the algorithm will be explained later. The algorithm consists in building a model, be it linear or nonlinear, iteratively, and additively (component-wise). Let y_t be some time series in question, x_t a vector of regressor variables (in our case, $x_t(p)$ are the lags from y_t to p, where $p = 1, ..., 12$) and M the stopping criterion of the algorithm, the $\hat{y}_t$ boosting estimation is the result of a sum of M distinct parts plus a constant, having the following form:

$$\hat{y}_t = \hat{f}(x_t) = \hat{f}^{(0)} + v \sum_{m=1}^{M} \hat{g}^{(m)}(x_t; \hat{\beta}_m), \tag{2}$$

where v is a learning rate parameter (usually $0 < v \leq 1$ according to the literature (FRIEDMAN, 2001)) and $\hat{g}^{(m)}(x_t; \hat{\beta}_m)$ is the learner, where $\hat{\beta}_m$ is the set of coefficients obtained from a fitting procedure and a loss function. For a given m,

$$\hat{g}^{(m)}(x_t; \hat{\beta}_m) = \underset{\hat{h}(\cdot)}{\text{argmin}} L(y_t, \hat{h}(x_t)), \tag{3}$$

where $L(\cdot)$ is the chosen loss function, e.g. mean square error, and the function $\hat{h}$ is the fitting procedure which can vary depending on the approach. The stopping criterion M can be chosen by the researcher or also can be retrieved through AIC, BIC or cross-validation. For our research, we fixed an upper bound $M = 300$ and used cross-validation to choose the optimal M, $M^* \leq 300$, in each estimation. Following this, we can describe generically the algorithm to forecast once as follows:
Step 1. We start with $m = 0$ and define $\hat{f}^{(0)} = \bar{y}_t$, where $\bar{y}_t$ is the average.
Step 2. For $m = 1$ to M:

1. Compute the residuals, defined as $\varepsilon_t = y_t - \hat{f}^{(m-1)}$.

2. Do regression on the residuals ε_t on each predictor $x_{(p)}$, with $p = 1, 2, ..., 12$, and compute the sum of squared residuals (SSR).

3. Select the predictor $x_{(p^*)}$ which has the smallest SSR.

4. Define $\hat{g}^{(m)} = \hat{\beta}_{(p^*)} x_{(p^*)}$.

5. Lastly, update the estimation $\hat{f}^{(m)} = \hat{f}^{(m-1)} + v\hat{g}^{(m)}$.

As stated in Park, Lee, and Ha (2009), "L_2 boosting is simply repeated least-squares fitting of residuals.". In our study, we are testing the predictability of macroeconomic variables using only their own history. Therefore, in Step 2 where the regression occurs, we regress only to the $p = 12$ lags of the objective variable. However, the model is easily expandable to use external regressors. The papers in Robert Lehmann and Klaus Wohlrabe (2017) and Lindenmeyer, Skorin, and Torrent (2021) show examples of using the boosting methodology using dozens of external regressors and their lags. To understand the magnitude of the prediction performed here in this paper, we applied the algorithm specified above to each of all variables in our data set, multiplied by the number of observations used in the test set and multiplied by the number of models used.

3.2.1 Linear Boosting

The linear boosting is the use of the algorithm specified in the previous section (see Section 3.2) with the fit of a linear regression in part 2.2 of the algorithm instructions. Let y_t be the time series to be modeled and x_t be its regressors, the model equation for each estimate can be written in the same way as Eq. 1. The difference between the linear model and linear boosting is that here we are selecting the best p for the estimation for each iteration from $m = 1$ to M.

3.2.2 Boosting with Splines

In order to understand how boosting with splines works, we must first understand the concept of splines. We can define splines as a continuous piece-wise curve. Subsequently, the application of splines was extended to be used as a "smoothing spline", that is, an estimation of a function $\hat{f}(x_i)$ given a data set z_i and a number of knots to estimate, which together will form an additive estimate (GREEN; SILVERMAN, 1993).

The use of smoothing splines in boosting methodology started with the paper by Bühlmann and Yu (2003), where they were used as the learner of the algorithm. However, according to Schmid and Hothorn (2008), for the boosting methodology, "smoothing splines are clearly less efficient [computationally] than other smooth base-learners." In that paper, the authors investigate the likelihood of using a modified splines method - P-splines, formulated by Paul H. C. Eilers and Marx (1996). In this method, there is also the use of a penalty, but a discrete one. The authors Schmid and Hothorn (2008) tested the use of P-splines as a replacement for smoothing splines and concluded that

> P-splines have been used successfully in regression as an approximation of smoothing splines. We have shown that this approximation is also successful in a boosting context: By using P-spline base-learners instead of smoothing spline base-learners, the computational effort of component-wise L2Boosting can be greatly reduced, while there is only a minor effect on the predictive performance of the boosting algorithm. (pg.18)

Therefore, this concludes that the smoothing splines method can be replaced by P-splines without loss of quality and with gains in efficiency in the case of boosting. For our scenario, instead of applying a linear regression in step 2.2 of the algorithm, a curve approximation of the data made by the P-spline methodology will be used, which was implemented in the R programming language by the authors of the package "mboost" (HOTHORN et al., 2011).

3.2.2.1 Two-Stage Boosting

There is discussion as to whether macroeconomic and financial variables are predominantly linear or nonlinear, and, regardless, whether the most reasonable model to fit is linear or nonlinear (STOCK; WATSON, 1998). Recapping the discussion made in Chapter 2 on time series nonlinearity, the authors Kauppi and Virtanen (2021) adapt the model from Taieb and Hyndman (2014) and propose it as a direct forecast procedure, and hybrid model between nonlinear and linear: two-stage boosting.

Given a time series y_t, the model starts with the convectional estimation of the linear methodology explained in Section 3.1 for each h. After that, a regression is estimated again, but this time nonlinear boosting (with splines) on the residuals of the estimated series and original series. The result is a two-component estimate that retains information from both linear and nonlinear estimations. Its advantage is that it is expected, on average, to have smaller forecast errors, because if the time series in question has predominantly linear behavior, the nonlinear component will be small. The opposite is also expected to be valid, i.e. a series with nonlinear behavior will have a smaller linear component in the total weight of the estimate. We have the following equation:

$$\widehat{g}_h^{TSBoost}(x_t) = \widehat{g}_h^{Linear}(x_t) + \widehat{r}_h^{BSpline}(x_t) \tag{4}$$

where $\widehat{g}_h^{TSBoost}$ is the two-stage boosting, $\widehat{g}_h^{Linear}$ is the linear estimation and $\widehat{r}_h^{BSpline}$ is the estimation on the residuals of the linear fit.

4 DATA AND FORECASTING APPROACH

In this chapter we explain how we acquired our data set with 140 variables and 288 observations ranging from January 1996 to December 2019. Furthermore, here we discuss important decision points regarding the chosen forecasting strategies, as well as define our strategy used. Finally, we also explain the indicators and evaluation methods chosen to analyze the performance of the models.

4.1 DATA

We gathered a base of 140 Brazilian macroeconomic time series. All variables were limited to start from January 1996, since from then on we already have a certain relative stability because it is post Real Plan and the population is relatively more used to a stable currency. Most of our variables are being still updated until end of 2021 or beginning of 2022, since we aim to use current data that can be used as regressors in other researches. But to standardize the database and leave aside the effect of the pandemic, we consider the data until December 2019. This sums up to mostly 288 monthly observations to each variable. This decision was also made in order to have a more complete set, since at the time of collection there were time series that have not yet been updated for 2020. We divided the 288 observations into a training set, where the model will be developed, and a test set, where we check the effectiveness of the models. As common in many empirical works, we split 75% train set and 25% test set for almost all of our selected variables[6].

To analyze the predictability of Brazilian macroeconomic variables is the construction of a credible data set of utmost importance. For this, the base is grounded from another data set, built by the authors Barbosa, Ferreira, and Silva (2020), where they forecast unemployment, industrial production index, IPCA and IPC (Brazilian CPI indexes). For this, they gathered 117 variables and used factorial models to predict them. Since our focus is on Brazilian variables, we discarded the variables that are not national. Also, since our objective is to analyze the predictability of national variables of interest, we also follow the logic for selecting the variables from Kauppi and Virtanen (2021) and McCracken and Ng (2016).

To do so, we collected the historical time series following a package that integrates the Ipeadata API with the R programming language (GOMES, 2022). The result of this data collection made us take variables from Ipeadata with several different sources that will be shown in the Appendix to this work (see Appendix A). But in order to have an interesting representation of the variables selected, we present in Table 1 the division by subject of the data collected and in Table 2 the source of each series.

[6] There are unemployment variables that had not been updated for the full year 2019 at the time of database development, so the cutoff in those cases between test and training set was 80% and 20% respectively out of 276 observations.

Table 1 – Themes of the Data Set

Theme	Number of Series
Balance of payments	4
Capital stock	1
Consumption and sales	16
Currency and credit	2
Employment	8
Exchange	14
Financial	7
Foreign trade	12
National Accounts	15
Perception and expectation	2
Prices	35
Production	4
Public Finance	15
Salary and income	5
Total	**140**

Table 2 – Sources for the Data Set

Source of the Data	Number of Series
Agência Nacional do Petróleo	8
ANBIMA	2
Banco Central do Brasil	28
Confederação Nacional da Indústria	3
Eletrobras	6
Federação das Indústrias do Estado de Minas Gerais	1
Federação das Indústrias do Estado do Rio de Janeiro	3
Federação do Comércio do Estado de São Paulo	3
Federação e Centro das Indústrias do Estado de São Paulo	3
Fundação Centro de Estudos do Comércio Exterior	10
Fundação Getulio Vargas	8
Fundação Instituto de Pesquisas Econômicas	1
Fundação Sistema Estadual de Análise de Dados	5
Instituto Brasileiro de Geografia e Estatística	19
Instituto de Pesquisa Econômica Aplicada	30
Ministério da Economia	8
Ministério do Desenvolvimento, Indústria e Comércio Exterior	2
Total	**140**

4.1.1 Data Treatment

In order to use the macroeconomic variables obtained in econometric models, we must first treat them. That is, we must perform transformations on each series that leave it with a constant mean and variance over time, i.e. stationarized series.

For this, we apply two tests: the first is the Augmented Dickey Fuller (ADF) test, which was developed by Dickey and Fuller (1979) and developed in R by Pfaff, Zivot, and Stigler (2016), and the second is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test developed by Kwiatkowski et al. (1992), while the coding in R was performed by Trapletti and Hornik (2021). The time series is only considered treated when we interpret the information from both tests as being stationary. For the ADF test, let y_t be the time series in question and is presented as follows,

$$y_t = \phi y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + e_t \tag{5}$$

$$\Delta y_t = \underbrace{(\phi - 1)}_{\delta} y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + e_t \tag{6}$$

where α, ϕ and ϕ_i are coefficients, e_t is the error term, $\delta = \phi - 1$ and p is the number of lags used in the test to bring more robustness. As the test is part of the Unit Root class of tests, the null hypothesis assumes the presence of a unit root ($\alpha = 1$ for Eq. 5 or $\delta = 0$ for Eq. 6) and therefore would be stationary if accepted. The test statistic is given by $\frac{\hat{\delta}}{SE(\hat{\delta})}$ and its critical value is taken from Dickey-Fuller t-distribution (FULLER, 2009). We reject the null hypothesis if the test statistic is smaller than the critical value for 5% significance.

The absence of a unit root does not necessarily mean that the time series in question is stationary. The KPSS test tests the hypothesis that the series is stationary in trend as the null hypothesis or has a unit root as the alternative hypothesis. We decided to bring in this additional test in order to get more robustness to the analysis and treatment of our data. To do so, consider

$$y_t = \beta_t + (r_t + \alpha) + e_t, \tag{7}$$

where y_t is the time series, α is the drift parameter (our case $\alpha = 0$), $r_t = r_{t-1} + u_t$ is a random walk process and both e_t and u_t are independent and identically distributed (i.i.d) error terms with mean equal to zero and constant variance. The KPSS test statistic is given by $\mathrm{KPSS} = \frac{\sum_{t=1}^{N} S_t^2}{N^2 \lambda^2}$, where S_t is the sum of squared accumulated errors, λ is the standard deviation and N is the number of observations. The critical values can be found in the original article by the authors Kwiatkowski et al. (1992). Again we are using 5% significance. If after applying both tests, the series is still interpreted as nonstationary by at least one of them, then we apply (1) difference of series, (2) difference of log of series, or (3) direct growth ($\frac{y_{t+1} - y_t}{y_t}$). If we applied the difference once and then the series is stationary, we consider the series as integration order $I(1)$, if only for the second difference, we consider it as $I(2)$. For none of the series was it necessary to perform more than two differences in order to make it stationary.

4.2 FORECASTING APPROACH

For single-step forecasting for time series, we usually use all the available data up to period t and perform direct estimation for $t+1$. But when we are dealing with multi-step forecasting, i.e. forecasting 2 or more steps ahead, there is a discussion about which method to use: direct forecasting or recursive forecasting. Direct prediction means training a model that always estimates h steps ahead. Therefore, for every h that has to be done in the research, a new model has to be estimated, which can be computationally heavy, and we might also fall into the risk of not using all the available observations at the time of prediction. Recursive forecasting, on the other hand, can be obtained from h different forecasts from one-step ahead models. The advantage of this method is the use of all available information at the time of prediction, but with the downside that we make predictions on top of the predictions, which can lead to carrying errors from $h = 1$ to larger h. In this work, as in the work of Kauppi and Virtanen (2021), we decided to do it by direct prediction. The justification is due to the goal of making forecasts with $h = 1, ..., 12$, and avoiding increasing errors by estimation with low performance in one-step ahead models. Moreover, the forecasting literature on the two strategy differences slightly favors direct estimation (JI et al., 2005; MARCELLINO; STOCK; WATSON, 2006; HAMZAÇEBI; AKAY; KUTAY, 2009). According to the forecast of simulated time series using the boosting methodology carried out by Kauppi and Virtanen (2021), “In the simulations, we find no significant difference between the direct and indirect procedures, while the direct method is on average more accurate than the indirect approach in terms of empirical comparisons”.

All series in our data set are stationary (see Section 4.1), and usually our series are in logarithmic value. Also, the order of integration of the series can be $I(0)$, $I(1)$ and $I(2)$. The estimation here is done as follows, consider y_t to be a series we want to model using the vector of regressors $x_t(p)$, which can be defined as $x_t(p) = y_t, ..., y_{t-p}$, $x_t(p) = \Delta y_t, ..., \Delta y_{t-p}$ or $x_t(p) = \Delta^2 y_t, ..., \Delta^2 y_{t-p}$, for $I(0)$, $I(1)$ and $I(2)$, respectively. In all our prediction exercise we consider $p = 12$. When making the prediction for the observation z_t, we train a model that predicts h steps ahead and that depends on the order of integration of the series. Given h, we can define that variable to be predicted, z_{t+h}, as follows:

$$\begin{aligned} I(0) &: z_{t+h} = y_{t+h} \\ I(1) &: z_{t+h} = y_{t+h} - y_t \\ I(2) &: z_{t+h} = y_{t+h} - y_t - h\Delta y_t \end{aligned} \tag{8}$$

Hence, in this way, the estimation of z_{t+h} will be equal to

$$z_{t+h} = f_h(x_t(p)) + \varepsilon_{t+h} \tag{9}$$

Where $f_h(x_t(p))$ is the estimation given h and using $x_t(p)$ as regressors (e.g. boosting). Also, ε_{t+h} is the prediction error. Thus, using all the information until t and the direct forecast approach for a given h, the boosting method shown in Eq. 2 to perform the prediction exercise on a series y_t of, for example, $I(0)$, is:

$$y_{t+h} = f_h(x_t(p)) + \varepsilon_{t+h} = f_h^{(0)} + v \sum_{m=1}^{M} g^{(m)}(x_t(p); \beta_m) + \varepsilon_{t+h}, \tag{10}$$

$$f_h^{(0)} = \frac{\sum_{i=1}^{t} y_i}{N} = \bar{y}_t. \tag{11}$$

For the parameters, we set $v = 0.1$ as is generally suggested in several papers (FRIEDMAN, 2001; BÜHLMANN; HOTHORN, 2007). Also, our chosen loss function is the minimization of the traditional Mean Squared Error (MSE), defined as the squared difference of the forecast errors. Now the question of interest is how to acquire the optimal stopping criterion, M^*. We could fix a value for all iterations, but nothing guarantees that it would be the best value considered for the data. Therefore, the literature commonly assigns a maximum value, in our case $M_{max} = 300$ and allows some model selection method to define which is the best model with $1 \leq M \leq 300$. To perform a pseudo out-of-sample forecasting, we have to select the best model without knowing the forecast errors. For this, the most commonly used methods are AIC and k-fold cross-validation. Given recent research highlighting the use of k-fold cross-validation over AIC for the boosting methodology, such as (WOHLRABE; BUCHEN, 2014), we opted for cross-validation. The k-fold method separates the training set into k different parts, trains the model on $k - 1$ parts, and tests the performance on the remaining part (STONE, 1974).

One decision point about the forecasting procedure is whether to use an expanding window or a rolling window. Both expanding and rolling window estimates increase the last index of the train set in one for each new estimate to be performed in the iteration. The difference is that the expanding window keeps the first index of the train set fixed, while rolling window also increases it in one for each new estimate to be performed, in order to always use a fixed number of observations for all estimations. As we are using the boosting methodology that performs variable selection, we follow the line of other articles such as Robert Lehmann and Klaus Wohlrabe (2017) and we perform an expanding window on all our forecasts. For this, consider T to be the total number of observations collected (most cases $T = 288$), T_1 to be the t index for the first observation in the test set, and T_2 to be the t index for the last observation also in the test set. For most of our series, T_1 represents January of 2014 and T_2 December of 2019. The advantage of this method is that we always use all the information available until the t-th period of the estimation. It will expand by unity at each i-th iteration of the estimation, in order as to always use all information available at time t, i.e periods $1, ..., (T_1 - 1) + i$.

Finally, to avoid high computational processing time, we chose to perform direct prediction by rounds, where we update the model parameters only once every 12 months. We performed the prediction in 6 rounds, since we have 72 observations in our test set. With this, we start estimating at iteration $i = 1$ using all available information from $t = 1$ to $t = T_1 - 1$ to fit the model. With the estimated parameters, we perform the prediction for steps $h = 1, ..., 12$. After the 12 predictions, we expand the training set by 12 observations and re-estimate a new model, with information from $t = 1$ (since we are using expanding window) to $t = T_1 + 11$, and perform the predictions for the next periods $t = T_1 + 13, T_1 + 14, ..., T_1 + 24$. We keep doing this until we cover the entire test set, thus until T_2.

4.3 EVALUATION

We aim to check the performance of boosting methods, especially the nonlinear and two-stage boosting methods. Since our interest is to see how well these models perform, we selected the linear estimation presented in Section 3.1 as a benchmark. This means that, when possible, we seek to compare the errors and predictive performance of the other proposed models with the linear model. To do this, we need consistent indicators that we can use to compare different models. Among the range of indicators, we have chosen one that is commonly accepted, the root mean squared error (RMSE), because it shows the average deviation between the predicted values and the actual values. From Section 4.2, we know that $T_1 < T_2$, then we can define RMSE as follows:

$$\mathrm{RMSE}_h^{\mathrm{model}} = \sqrt{(T_2 - T_1 + 1)^{-1} \sum_{i=T_1}^{T_2} \left(\mathrm{FE}_{t+h,i}^{\mathrm{model}}\right)^2}, \tag{12}$$

where $\mathrm{FE}_{t+h,k}^{\mathrm{model}} = y_{t+h,i} - \hat{y}_{t+h,i}^{model}$ represents the multi-step forecast error for a given i observation. The smaller the RMSE, the lower the forecast error and the better the model. For the sake of simplifying the indicator and the comparison with the benchmark, consider the following relative indicator (rRMSE):

$$\mathrm{rRMSE}_h = \frac{\mathrm{RMSE}_h^{\mathrm{others}}}{\mathrm{RMSE}_h^{\mathrm{linear}}}. \tag{13}$$

In this modified indicator, when, for a given model other than the linear one, we have $\mathrm{rRMSE} < 1$, it means that the $\mathrm{RMSE}_h^{\mathrm{model}} < \mathrm{RMSE}_h^{\mathrm{linear}}$, so the RMSE indicator of the model to be compared performs better than the linear model. The opposite is true, hence $\mathrm{rRMSE} > 1$ means that the linear model performs better than the selected model.

A proxy to estimate the predictability of the models performed is the coefficient of determination. It is through this indicator that we can have an assessment of the performance of the models used in the prediction of each of the variables collected using only their history. The authors Kauppi and Virtanen (2021) define the empirical coefficient of determination that we use for given p in this paper as:

$$\widehat{R}^2(\widehat{g}_h^{\mathrm{model}}(x_t)) = 1 - \frac{\widehat{\mathrm{MSE}}(\widehat{g}_h^{\mathrm{model}}(x_t))}{\widehat{\mathrm{var}}(y_{t+h})} \tag{14}$$

where R^2 is the coefficient of determination and $\widehat{\mathrm{var}}(y_{t+h})$ is the variance of the sample from the actual values outside the sample. Also, t varies from T_1 to T_2, where T_1 is the index for the beginning of the test set and T_2 is the last index of the test set in each variable. If the estimated coefficient $\widehat{R}^2(\widehat{g}_h^{\mathrm{model}}(x_t))$ is less than 0, we replace the value by 0.

There are currently no econometric tests yet to compare in a generic way the predictive ability of one model compared to another regardless of the model specifications. The closest to this was the Giacomini-White test, which measures the statistical significance of the difference in predictive ability of two models (GIACOMINI; WHITE, 2006). According to the authors,

> We implement this different focus by conducting inference about conditional, rather than unconditional, moments of forecasts and forecast errors. Recognizing that even a good model may produce bad forecasts due to estimation uncertainty or model instability, we make the object of evaluation the entire forecasting method (including the model, the estimation procedure and the size of the estimation window), whereas the existing literature concentrates solely on the model. In so doing, we are also able to handle more general data assumptions (heterogeneity rather than stationarity) and estimation methods, as well as providing a unified framework for comparing forecasts based on nested or non-nested models, which was not previously available. (GIACOMINI; WHITE, 2006, pg.23)

We implement this statistical test from the “afmtools” package in R (CONTRERAS-REYES; GEORG M.; PALMA, 2013). The null hypothesis of the Giacomini-White test states that the predictive ability of two models are equal, while the alternative hypothesis, when using the one-sided statistic, considers that one model has greater predictive power than the other. A point worth noting regarding the use of this test in our work is the same one raised by the authors Kauppi and Virtanen (2021),

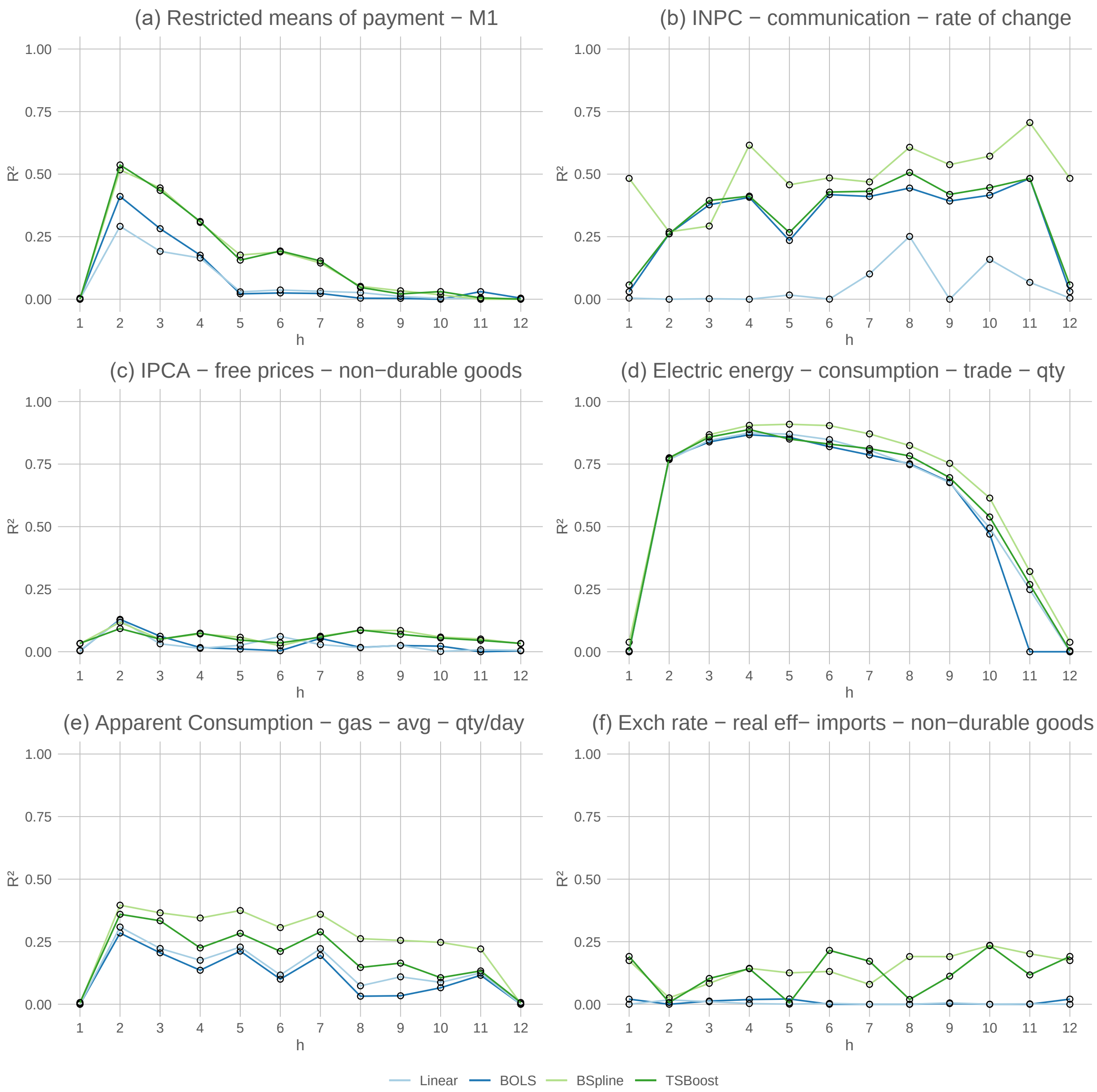
> One problem with the GW test is that its validity rests on the assumption of “nonvanishing estimation errors” and it is thus designed for situations where the underlying simulated out-of-sample prediction errors are obtained by using a fixed (or a rolling) window rather than an expanding window estimation scheme applied here.” (pg.13)

Even if the scenario where we apply the Giacomini-White test is not ideal, we believe that it can be used as a reference because we calculate other indicators specified here in this section, which also serve to ascertain the predictive power. Through them, we can get an idea about the performance of each of the models in the 140 verified series, and we can have a way to compare the models among themselves.

5 EMPIRICAL FINDINGS

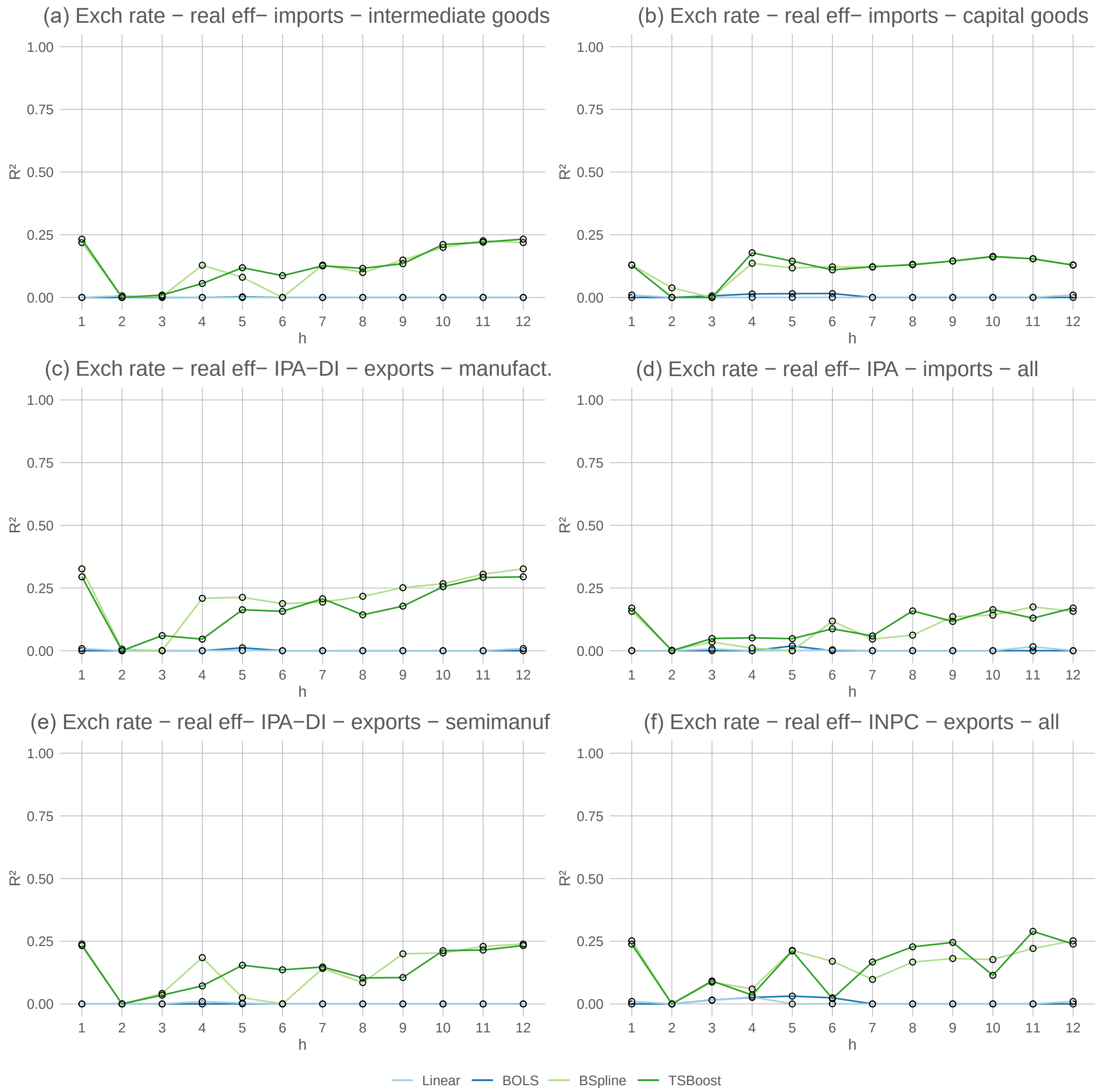
Here we analyze the performance of each of the models applied in forecasting the 140 variables collected for the test period, which is from January 2014 to December 2019. The models used are explained in the methodology section (see Chapter 3), i.e. simple linear estimation (Linear), linear boosting (BOLS), boosting with splines as base-learner (BSpline) and two-stage boosting (TSBoost). Tables 3, 4 and 5 contain other models: BSpline* and TSBoost*, which refer, respectively, to boosting with splines without extrapolation and two-stage boosting without extrapolation, and finally Tree, which is boosting with regression trees as base-learner (more details in Sections 5.1 and 5.2).

Figure 1 – Selected Plots 1 - Period 2014-2019 - R^2



Source: Author.

Figure 2 – Selected Plots 2 - Period 2014-2019 - R^2



Source: Author.

A result that better evokes the predictive performance of each model in its own history is the empirical coefficient of determination R^2 calculated from the prediction error and variance of the test set sample, as defined in Eq. 14. According to the R^2 methodology, it is assumed that the higher the variance of a variable, the more difficult it is to predict. Therefore, for highly volatile series, there is a certain compensation at the moment of coefficient calculation (see details in Section 4.3). Since we did the estimation for all variables in our database, there are 140 different plots. To allow us to select some of them to present here, we made a special condition for choosing them[7]. Figures 1 and 2 contain some plots where the y-axis is the estimated coefficient R^2 and the x-axis is the forecast horizon $h = 1, ..., 12$.

[7] The selected graphs are obtained for the variables where the logical condition below is realized for at least 9 or

From the figures, we can see that there seems to be a combined movement between the linear models (Linear and BOLS) and the nonlinear models (BSpline and TSBoost). Furthermore, since our goal is to test the validity and applicability of boosting methodology, especially with nonlinear base-learners (P-Spline) to forecast macroeconomic variables, we are focusing the analysis on the cases where the performance of these models is better. With this, we can visually see if there are differences in the curves estimated by the different models for the selected variables. Both Figures 1 and 2 show graphs where there is a clear difference between the estimation with the linear base-learner and with splines as base-learner.

In particular, in Figure 1, we can see that, for example, the curves in graphs (a), (d), and (e) present the greatest difference between the estimation methods. The nature of the variables is about the monetary base, apparent gas consumption and electricity consumption, hence, from different types of economic data. Nevertheless, these are scenarios where estimation with splines has a clear advantage. Curve (d) presents the best determination coefficient, reaching values above 0.80 depending on the forecast horizon. Another point of emphasis is the low value of R^2 when $h = 1$. Mispecifications when $h = 1$ show one of the reasons why we chose to perform direct as opposed to recursive forecasting. For subsequent h's, we have, for most cases, a higher value for R^2.

In Figure 2, we have a selection of variables where all are specifications of the real effective exchange rate. According to the methodology of the time series presented in Nonnenberg (2015), we can define the real effective exchange rate as the weighted average of Brazil's bilateral real exchange rates against each of its major trading partners. The weighting of each country is given by the share of each country in Brazilian exports or imports. The Brazilian inflation index used is the INPC/IBGE. This indicator brings information about the level of prices, inflation, and Brazil's main trading partners into one index. According to the figure, we can see that there is a considerable advantage for nonlinear methods over linear methods in relation to the R^2 value in almost all horizons. Moreover, an interesting point is that they are indicators with low predictive power if using linear methods, since we can see that, for example, curves (a) and (e) have $R^2 = 0$ for almost all h's. In contrast, boosting with splines and two-stage boosting have relatively high value for the coefficient, especially at $h = 1$ and $h = 12$ in all scenarios.

more different h's. Let h be in $\{1, ..., 12\}$ and P be the models with P-Spline as base-learner, then

$$(R^2_{P|h} > R^2_{\text{Linear}\,|h}) \cap (R^2_{P|h} > R^2_{\text{BOLS}\,|h}).$$

Table 3 – R^2 from the Data Set†

R^2 min	Method	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
		(a) All cases											
0	**Linear**	0.33	0.35	0.35	0.34	0.32	0.29	0.30	0.30	0.28	0.26	0.24	0.11
	BOLS	0.30	0.33	0.32	0.33	0.31	0.27	0.29	0.29	0.27	0.26	0.23	0.11
	BSpline	0.30	0.31	0.31	0.32	0.30	0.27	0.29	0.29	0.28	0.28	0.25	0.13
	TSBoost	0.33	0.34	0.34	0.34	0.32	0.29	0.30	0.30	0.29	0.28	0.26	0.13
	BSpline*	0.30	0.32	0.32	0.33	0.31	0.28	0.30	0.30	0.29	0.28	0.26	0.13
	TSBoost*	0.33	0.34	0.34	0.34	0.32	0.29	0.31	0.30	0.29	0.28	0.26	0.13
	Tree	0.25	0.26	0.26	0.26	0.24	0.21	0.23	0.24	0.22	0.21	0.20	0.09
	N	140	140	140	140	140	140	140	140	140	140	140	140
0.1	**Linear**	0.43	0.46	0.49	0.52	0.52	0.52	0.53	0.54	0.48	0.49	0.47	0.50
	BOLS	0.40	0.45	0.47	0.48	0.50	0.50	0.51	0.50	0.47	0.49	0.46	0.50
	BSpline	0.41	0.44	0.45	0.45	0.47	0.42	0.44	0.46	0.44	0.42	0.40	0.38
	TSBoost	0.44	0.48	0.51	0.52	0.50	0.51	0.49	0.48	0.45	0.45	0.41	0.41
	BSpline*	0.41	0.44	0.46	0.45	0.47	0.42	0.45	0.47	0.43	0.42	0.40	0.36
	TSBoost*	0.44	0.48	0.50	0.51	0.50	0.51	0.50	0.48	0.46	0.46	0.41	0.40
	Tree	0.38	0.42	0.42	0.42	0.40	0.39	0.40	0.39	0.40	0.37	0.34	0.32
	N	101	98	92	89	88	79	85	85	88	85	85	43
		(b) Splines more accurate on average											
0	**Linear**	0.35	0.33	0.31	0.32	0.31	0.26	0.29	0.28	0.26	0.25	0.24	0.14
	BOLS	0.33	0.32	0.30	0.31	0.31	0.25	0.29	0.28	0.26	0.26	0.25	0.14
	BSpline	0.35	0.34	0.33	0.35	0.33	0.28	0.31	0.32	0.30	0.32	0.32	0.19
	TSBoost	0.36	0.35	0.33	0.34	0.33	0.29	0.31	0.31	0.29	0.31	0.31	0.19
	BSpline*	0.36	0.34	0.33	0.35	0.33	0.28	0.32	0.32	0.30	0.32	0.31	0.20
	TSBoost*	0.36	0.35	0.33	0.34	0.33	0.29	0.31	0.31	0.29	0.30	0.31	0.19
	Tree	0.30	0.30	0.27	0.28	0.26	0.20	0.25	0.26	0.24	0.24	0.24	0.14
	N	75	68	67	78	84	80	76	88	82	82	74	71
0.1	**Linear**	0.46	0.54	0.50	0.52	0.59	0.52	0.54	0.55	0.47	0.51	0.46	0.49
	BOLS	0.43	0.52	0.49	0.51	0.57	0.50	0.53	0.56	0.47	0.52	0.47	0.50
	BSpline	0.46	0.53	0.51	0.52	0.57	0.48	0.52	0.55	0.49	0.53	0.47	0.47
	TSBoost	0.47	0.56	0.52	0.53	0.59	0.53	0.54	0.56	0.49	0.53	0.48	0.52
	BSpline*	0.46	0.53	0.51	0.52	0.57	0.49	0.53	0.56	0.49	0.53	0.47	0.48
	TSBoost*	0.47	0.56	0.51	0.53	0.59	0.54	0.54	0.56	0.49	0.53	0.48	0.51
	Tree	0.40	0.47	0.44	0.45	0.48	0.37	0.43	0.46	0.40	0.42	0.40	0.39
	N	55	40	39	45	40	36	38	42	41	37	36	19
		(c) Linear more accurate on average											
0	**Linear**	0.30	0.36	0.38	0.37	0.34	0.34	0.33	0.33	0.32	0.28	0.23	0.09
	BOLS	0.26	0.33	0.35	0.35	0.31	0.31	0.30	0.29	0.30	0.27	0.21	0.08
	BSpline	0.23	0.29	0.29	0.28	0.26	0.26	0.25	0.25	0.25	0.22	0.18	0.06
	TSBoost	0.28	0.34	0.35	0.34	0.31	0.30	0.29	0.29	0.29	0.25	0.21	0.07
	BSpline*	0.24	0.30	0.30	0.30	0.28	0.28	0.27	0.26	0.26	0.23	0.19	0.06
	TSBoost*	0.28	0.34	0.36	0.34	0.31	0.30	0.30	0.30	0.29	0.26	0.21	0.08
	Tree	0.18	0.23	0.25	0.23	0.22	0.22	0.21	0.21	0.20	0.17	0.15	0.04
	N	65	72	73	62	56	60	64	52	58	58	66	69
0.1	**Linear**	0.45	0.50	0.55	0.58	0.60	0.70	0.62	0.64	0.70	0.58	0.51	0.63
	BOLS	0.40	0.47	0.51	0.54	0.56	0.64	0.58	0.59	0.66	0.56	0.49	0.56
	BSpline	0.38	0.42	0.44	0.46	0.48	0.57	0.51	0.54	0.58	0.48	0.43	0.47
	TSBoost	0.43	0.48	0.53	0.54	0.58	0.66	0.58	0.61	0.67	0.55	0.48	0.57
	BSpline*	0.38	0.43	0.46	0.47	0.50	0.60	0.53	0.54	0.60	0.49	0.44	0.47
	TSBoost*	0.43	0.48	0.53	0.54	0.58	0.66	0.58	0.60	0.67	0.55	0.48	0.57
	Tree	0.30	0.34	0.36	0.37	0.39	0.48	0.42	0.43	0.45	0.38	0.35	0.28
	N	39	47	48	37	29	26	31	23	24	26	27	8

† Automatic colors, where greener is relatively better and redder worse.
* Splines prediction without extrapolation.

Source: Author.

Table 3 shows the calculated R^2 as the average of all selected series given the forecast horizon and model. As done in Kauppi and Virtanen (2021), we separated in minimum R^2, which can be 0 or 0.1. We did this because we can distinguish series where forecasting using the series' own history is really difficult with the models used, and that means the value for R^2 is closer to 0. We apply colors to the tables to make it easier to visualize the results. The closer the coefficient is to 1, the better the prediction given h and the model considered, so the greener the cell in the table. The worse it is, the redder the cell is. In part (a), considering all cases (number of observations $N = 140$), we find that the best models for all series are the TSBoost (mean 0.295) and the linear model (mean 0.290), followed by the BSpline model (mean 0.285). What is worth to note is that the TSBoost average is higher than the Linear's and the BSpline's, which shows that it is a powerful model to use when one does not know the best model to estimate for a specific series, because on average it is the most accurate. Part (b) of Figure 3 shows us the selection for the subset where series with splines (either BSpline or TSBoost) are more accurate on average. We can see that N has dropped from 140 to 77 on average, i.e., 55% of the series used. As expected, we can see that as much as for the N equal to 0 or 0.1, we have that the performance of models with splines are considerably better, having means equal to 0.31 while linear models have means equal to 0.27. In case (c), we select the series whose indicators for the linear models are superior. We then obtained 45% of all the series on N average, but we can see that the performance of the Linear model is superior to all the others, including the BOLS model. We comment on the results for the models without extrapolation and the Tree model in subsequent sections. One thing to note here in part (c) of the table, where we have the selection of series where the Linear model is better, the performance of the TSBoost model is almost as good as the linear model. So, with these results, for selection (b), the TSBoost model is as good as the BSPline, and for part (c), the TSBoost model is almost as good as the linear one, making it a suitable model in general.

Another way to visualize the preditive performance is from the estimated rRMSE of the series. Table 4 presents the indicator values compared to the Linear model, which we consider as a benchmark. The table shows the average rRMSE of N observations considering the model and the forecast step. In general, part (a) shows us that the Linear model is, on average, superior to all models from $h = 1$ to $h = 7$. Thereafter, we have marginal superiority to the TSBoost model. Note that BSpline is not better on average in any case compared to the Linear model. In scenario (b), on the other hand, we are selecting only those series where the estimation with splines is superior on average. For 63% of the total series, the performance for the spline models is better than the linear models for all h's. Scenario (c) brings us the series where the Linear model is better, but we have only 37% of the series. In this case, all other models for all h's have the value of $\mathrm{rRMSE} > 1$, but still the TSBoost model is superior to the BOLS model on average. Here, the same thing that occurs in the previous table is true again: the TSBoost model is as good as BSpline in part (b), and the TSBoost model is almost as good as the Linear model in part (c). Also, we expose the rRMSE indicator in full tables for each of the variables for the forecast horizons $h = 1, 6, 12$ in Appendix B.

Table 4 – rRMSE relative to Linear Model†

Method	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
	(a) All cases											
BOLS	1.029	1.028	1.039	1.064	1.085	1.100	1.082	1.058	1.039	1.029	1.031	1.020
BSpline	1.023	1.038	1.064	1.056	1.056	1.063	1.056	1.050	1.048	1.039	1.035	1.026
TSBoost	1.004	1.009	1.002	1.012	1.010	1.015	1.003	0.993	0.996	0.992	0.989	0.981
BSpline*	1.016	1.027	1.042	1.038	1.043	1.046	1.044	1.040	1.040	1.032	1.029	1.020
TSBoost*	1.002	1.001	0.998	1.003	1.000	1.002	0.996	0.991	0.991	0.988	0.984	0.979
Tree	1.104	1.127	1.149	1.156	1.161	1.179	1.175	1.161	1.167	1.171	1.150	1.070
N	140	140	140	140	140	140	140	140	140	140	140	140
	(b) Splines more accurate on average											
BOLS	1.016	1.013	1.029	1.054	1.042	1.024	1.024	1.005	0.998	1.005	1.028	1.006
BSpline	0.987	0.988	1.003	1.001	0.997	0.986	0.981	0.981	0.972	0.973	0.994	0.986
TSBoost	0.990	0.993	0.980	0.983	0.984	0.978	0.979	0.971	0.969	0.959	0.954	0.947
BSpline*	0.983	0.986	0.998	0.999	0.992	0.982	0.979	0.974	0.968	0.972	0.993	0.981
TSBoost*	0.990	0.980	0.978	0.983	0.984	0.977	0.971	0.969	0.968	0.960	0.952	0.946
Tree	1.037	1.039	1.072	1.088	1.068	1.058	1.061	1.050	1.030	1.058	1.073	1.034
N	80	73	72	85	94	91	89	100	95	93	89	94
	(c) Linear more accurate on average											
BOLS	1.047	1.045	1.050	1.079	1.174	1.242	1.183	1.188	1.124	1.077	1.036	1.048
BSpline	1.072	1.094	1.128	1.143	1.178	1.207	1.187	1.223	1.209	1.169	1.108	1.109
TSBoost	1.022	1.027	1.026	1.057	1.064	1.085	1.044	1.048	1.054	1.057	1.051	1.052
BSpline*	1.061	1.071	1.089	1.100	1.146	1.166	1.159	1.203	1.190	1.152	1.093	1.101
TSBoost*	1.019	1.025	1.020	1.033	1.034	1.048	1.040	1.046	1.041	1.045	1.039	1.045
Tree	1.193	1.222	1.230	1.261	1.350	1.403	1.375	1.441	1.456	1.394	1.285	1.143
N	60	67	68	55	46	49	51	40	45	47	51	46

† Automatic colors, where greener is relatively better and redder worse.
* Splines prediction without extrapolation.

Source: Author.

The remaining strategy to analyze the results we report is the Giacomini-White test. This test measures the statistical difference in prediction between two models. But, as discussed in Section 4.3, we are not sure of the validity of the test, since we are expected to do this with a rolling window of estimation. Since the number of observations in the test set is 36, we have a relatively low n, so we believe the results of the test are not dismissible and we chose to show them here. Table 5 shows the results of applying the test. We iteratively ran and applied the test for each of the 140 series, and compared them to the Linear model. The null hypothesis of the test is that there is statistically no difference in prediction between the models, and the alternative hypothesis used is that the tested model is superior to the Linear model. The results show, given the model and given h, the number of series, out of 140, which are statistically superior in forecasting in comparison to the Linear model. We separate them into two significance levels, namely 10% and 5%. We can easily see that for most series, the test indicts no difference in prediction between the models. But for the series whose superiority is statistically superior, we have, on average, better predictions for the Tree model (we will explain in more detail in Section 5.2) and the TSBoost model. As seen from this and the other tables, we can see that the predictive performance of the two-stage boosting model is always high on average, being an excellent candidate to be used in all cases, especially when the nature of the series is unknown (linear or nonlinear).

Table 5 – Giacomini-White Test - Quantity of Series Statistically Superior to the Linear Model†

p-value	Method	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
0.90	**BOLS**	19	19	22	21	24	26	22	31	28	26	21	36
	BSpline	12	13	15	17	20	22	20	29	26	29	25	37
	TSBoost	16	23	21	19	25	23	23	27	30	34	40	40
	BSpline*	14	17	21	22	24	25	21	29	30	32	30	40
	TSBoost*	20	23	22	21	24	23	21	29	30	37	44	40
	Tree	42	48	47	47	40	36	35	33	32	26	24	16
0.95	**BOLS**	15	12	14	12	13	14	13	19	17	18	16	21
	BSpline	7	7	9	10	13	14	12	20	19	22	21	19
	TSBoost	10	14	12	10	12	14	14	11	23	23	30	22
	BSpline*	9	11	12	14	14	16	12	20	20	22	22	22
	TSBoost*	11	14	15	11	11	14	13	11	23	23	31	24
	Tree	27	34	39	35	29	30	28	27	27	21	23	13

† Automatic colors, where greener is relatively better and redder worse.
* Splines prediction without extrapolation.

Source: Author.

5.1 SPLINES WITHOUT EXTRAPOLATION

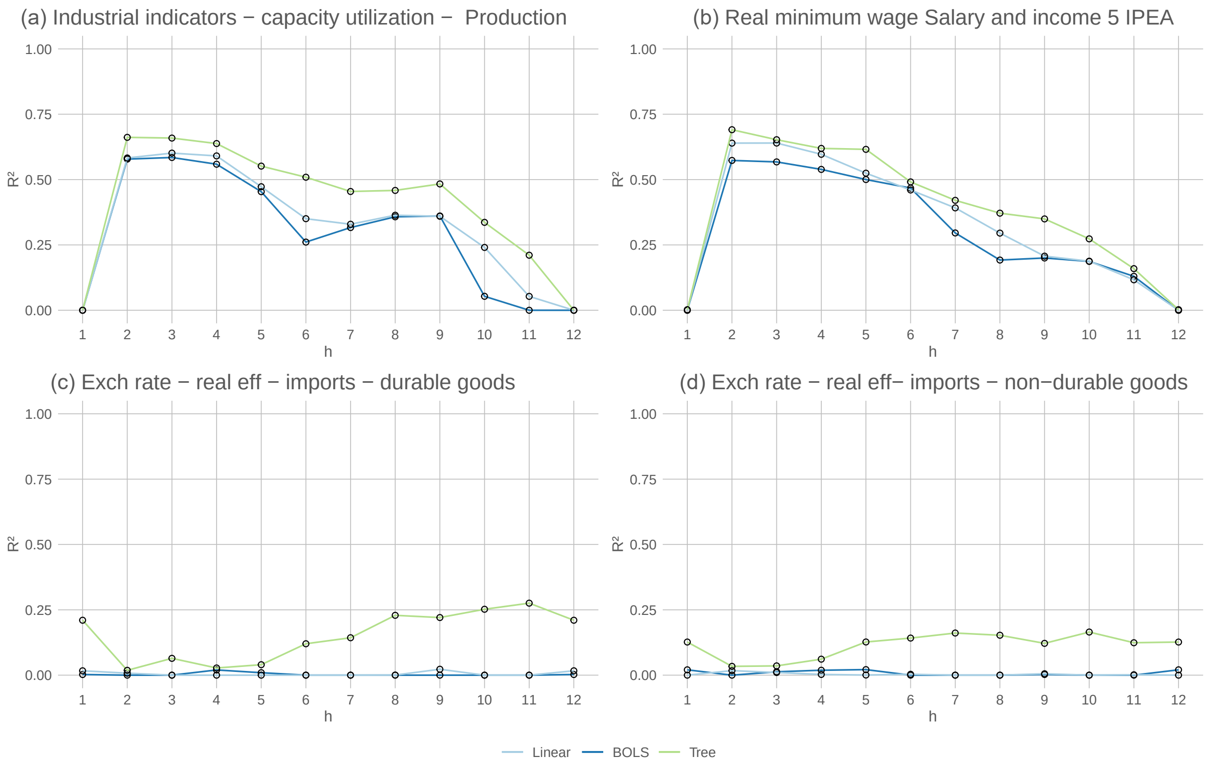
From the methodology explained for Section 3.2.2, whenever we use the boosting algorithm considering the P-Spline as a base-learner, we are smoothing the noisy data. However, one of the problems we may have when doing a smoothing, is the consequence of extrapolation. The extrapolation does not necessarily lead to bad results, but for some cases, mostly when estimation with extrapolation occurs, it generates very different values in comparison to when there is interpolation, causing an increase in the variance of the estimates and consequently reducing the overall performance of the model with splines. One of the corrections, presented by the authors Kauppi and Virtanen (2021) as "Hybrid Model", is whenever extrapolation occurs, to replace the extrapolated estimation by the Linear model estimation. The results of these hybrid models are the models shown in Tables 3, 4 and 5 such as BSpline* and TSBoost*.

From the results we can see in Table 3 that the models without extrapolation, on average, have marginal gains in predictive power when compared to the models with extrapolation. Even in scenario (b), we always have the BSpline* with superior performance compared to the conventional BSpline in both R^2 scenarios. In Table 4, the TSBoost* model is the most superior model on average in (a) all cases and in (b) splines better on average. Finally, both BSpline* and TSBoost* are better than their counterparts with extrapolation in Table 5, showing that the strategy of removing extrapolation can generate marginal gains in predictive power, and also supporting the view that using the two-stage boosting model is superior on average. Visually, we chose not to show them because qualitatively the curves are very similar to the curves of the splines with extrapolation.

5.2 ROBUSTNESS CHECK: BOOSTING REGRESSION TREES

We perform the estimation with an alternative base-learner that is common to apply in the boosting methodology: regression trees. The boosting method was augmented and the "mboost" statistical package was improved from the contributions on regression trees of the paper Hothorn, Hornik, and Zeileis (2006). With this, we were able to apply the boosting algorithm and construct a piece-wise and recursive estimation of the binary partition model. This is a non-parametric model and potentially estimates a nonlinear relationship between the predicted variable and the regressors. However, according to the authors of the Hothorn et al. (2011) statistical package, "The regression fit is a black box prediction machine and thus hardly interpretable." We chose to apply this method as a robustness check due to its wide applicability in the forecasting literature.

Figure 3 – Selected Plots 3 - Period 2014-2019 - R^2



Source: Author.

The performance of the Tree model in all scenarios in Tables 3 and 4 is, for all h and compared to all other models, the worst. Our interpretation of the results is, therefore, the Tree model realizes such a nonlinear relationship between the regressors and the regressed, that when the variable to be estimated is not of nonlinear intrinsic behavior, the Tree model has low predictive power. But on the other hand, Table 5 shows the Tree model as the model that has the most statistically superior predictions to the Linear model, by a relatively large margin. Figure 3 visually shows the performance of the Tree model compared to the linear models for some selected variables, from which we can see that there is a relatively wide visual difference between the models.

5.3 ROBUSTNESS CHECK: COMPARING AIC WITH CROSS-VALIDATION

To make another robustness check of the results, we chose to elaborate and present the table of the empirical coefficient of determination R^2 calculated for the estimations, where the $\mathrm{M}_{\mathrm{stop}} = M^*$ of the boosting algorithm was acquired through the AIC methodology, as opposed to k-fold cross-validation. Using Table 6, we can view the results in a manner analogous to Table 3. We do not present the AIC method for selecting M^* for the model boosting regression trees.

Starting from scenario (a) and comparing the results between the two tables, we can see that, in general, the R^2 results obtained a marginal gain, but the average for TSBoost remained the same. The model that obtained the most gains in scenario (a) was BSpline, but in scenario (a) with $R^2 \min = 0.1$, the results are almost identical. In part (b), again the BSpline model obtained some marginal gains at $R^2 \min = 0$, but at $R^2 \min = 0.1$ most models had a small relative loss. Finally, in part (c), all models on average had a small gain. The use of AIC found approximately 58% of the database to have more accurate splines on average, considering all h's. Overall, both results from using AIC or cross-validation to select the $\mathrm{M}_{\mathrm{stop}} = M^*$ for each estimation are bringing qualitatively similar and numerically very close results.

Table 6 – R^2 from the Data Set - AIC†

R^2 min	Method	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
	(a) All cases												
0	**Linear**	0.33	0.35	0.35	0.34	0.32	0.29	0.30	0.30	0.28	0.26	0.24	0.11
	BOLS	0.31	0.33	0.33	0.34	0.32	0.28	0.30	0.30	0.28	0.27	0.24	0.11
	BSpline	0.31	0.32	0.33	0.34	0.32	0.29	0.30	0.31	0.30	0.29	0.26	0.13
	TSBoost	0.33	0.34	0.34	0.34	0.32	0.29	0.30	0.30	0.29	0.28	0.26	0.13
	BSpline*	0.32	0.33	0.34	0.35	0.33	0.30	0.32	0.32	0.30	0.29	0.26	0.13
	TSBoost*	0.33	0.35	0.35	0.34	0.32	0.29	0.31	0.30	0.29	0.28	0.26	0.13
	N	140	140	140	140	140	140	140	140	140	140	140	140
0.1	**Linear**	0.43	0.46	0.49	0.52	0.52	0.52	0.53	0.54	0.48	0.49	0.47	0.50
	BOLS	0.40	0.44	0.48	0.48	0.48	0.48	0.51	0.49	0.48	0.48	0.47	0.49
	BSpline	0.41	0.45	0.46	0.46	0.45	0.41	0.45	0.45	0.43	0.44	0.40	0.38
	TSBoost	0.44	0.48	0.51	0.52	0.50	0.50	0.49	0.49	0.46	0.45	0.41	0.41
	BSpline*	0.41	0.45	0.47	0.45	0.46	0.42	0.45	0.45	0.43	0.44	0.39	0.37
	TSBoost*	0.44	0.48	0.50	0.52	0.50	0.51	0.50	0.49	0.46	0.45	0.40	0.40
	N	101	98	92	89	88	79	85	85	88	85	85	43
	(b) Splines more accurate on average												
0	**Linear**	0.36	0.36	0.30	0.32	0.32	0.25	0.28	0.29	0.26	0.26	0.25	0.12
	BOLS	0.35	0.36	0.30	0.32	0.32	0.25	0.29	0.30	0.27	0.27	0.26	0.13
	BSpline	0.37	0.38	0.33	0.36	0.35	0.29	0.33	0.34	0.32	0.33	0.32	0.18
	TSBoost	0.37	0.38	0.31	0.33	0.33	0.27	0.31	0.32	0.29	0.30	0.30	0.17
	BSpline*	0.38	0.38	0.34	0.36	0.36	0.29	0.33	0.35	0.32	0.33	0.32	0.18
	TSBoost*	0.37	0.38	0.31	0.33	0.33	0.27	0.31	0.32	0.29	0.30	0.30	0.17
	N	75	68	67	78	84	80	76	88	82	82	74	71
0.1	**Linear**	0.46	0.51	0.49	0.52	0.55	0.50	0.53	0.54	0.45	0.49	0.47	0.51
	BOLS	0.44	0.49	0.48	0.52	0.54	0.49	0.53	0.55	0.46	0.50	0.47	0.52
	BSpline	0.46	0.52	0.51	0.55	0.55	0.49	0.53	0.56	0.48	0.53	0.49	0.49
	TSBoost	0.47	0.52	0.50	0.53	0.55	0.52	0.54	0.55	0.46	0.51	0.48	0.54
	BSpline*	0.47	0.52	0.51	0.55	0.56	0.50	0.54	0.57	0.48	0.52	0.48	0.51
	TSBoost*	0.47	0.52	0.50	0.53	0.55	0.52	0.54	0.55	0.46	0.50	0.48	0.53
	N	55	40	39	45	40	36	38	42	41	37	36	19
	(c) Linear more accurate on average												
0	**Linear**	0.29	0.33	0.40	0.39	0.34	0.36	0.33	0.31	0.32	0.27	0.22	0.10
	BOLS	0.26	0.31	0.37	0.36	0.31	0.33	0.31	0.28	0.30	0.26	0.21	0.09
	BSpline	0.23	0.27	0.32	0.30	0.27	0.29	0.27	0.24	0.26	0.22	0.18	0.07
	TSBoost	0.27	0.31	0.38	0.36	0.31	0.32	0.30	0.27	0.29	0.25	0.20	0.08
	BSpline*	0.24	0.29	0.34	0.32	0.29	0.31	0.29	0.26	0.28	0.23	0.19	0.07
	TSBoost*	0.27	0.31	0.38	0.36	0.31	0.33	0.31	0.28	0.30	0.25	0.20	0.08
	N	65	72	73	62	56	60	64	52	58	58	66	69
0.1	**Linear**	0.45	0.50	0.57	0.58	0.61	0.69	0.62	0.65	0.73	0.61	0.50	0.62
	BOLS	0.41	0.47	0.53	0.54	0.56	0.63	0.58	0.60	0.69	0.59	0.48	0.55
	BSpline	0.38	0.42	0.48	0.48	0.51	0.57	0.52	0.55	0.64	0.54	0.44	0.48
	TSBoost	0.42	0.48	0.55	0.55	0.59	0.64	0.58	0.61	0.70	0.58	0.46	0.56
	BSpline*	0.38	0.44	0.50	0.49	0.53	0.60	0.55	0.56	0.65	0.54	0.45	0.49
	TSBoost*	0.42	0.48	0.55	0.55	0.59	0.64	0.59	0.61	0.71	0.58	0.47	0.56
	N	39	47	48	37	29	26	31	23	24	26	27	8

† Automatic colors, where greener is relatively better and redder worse.
* Splines prediction without extrapolation.

Source: Author.

6 CONCLUSION

This work aimed to elaborate and treat a new and recent Brazilian data set with macroeconomic time series in a way that it follows the molds of other economic and econometric research. Then, the application of boosting methodology to verify the linear and nonlinear predictability of the variables studied. And, finally, the application of the two-stage boosting model, a model initially formalized by Kauppi and Virtanen (2021) and having its performance verified by the empirical coefficient of determination indicator, calculated based on the out-of-sample forecasts. As secondary objectives, we had the application of robustness checking of the selected models from the comparison with other models, namely the hybrid models - splines without extrapolation - and the trees regression as base-learner, as well as the comparison of the applied methodology of selecting the optimal model via cross-validation with the AIC method.

We can say that the application of the models obtained solid results for all the models applied, but, on average, the two stage boosting modeling had the highest performance. Thus, we can say that this model is the best candidate among the models used here to forecast Brazilian macroeconomic variables with their own history. As presented in the results part of this paper, the two-stage boosting model achieved high performance generally when boosting with splines also performed well, but it also retained as strong predictive power together with the linear model in cases where the linear model was superior to purely nonlinear modeling. Also, regarding boosting with splines, it is a strong base-learner method with good results for our case, specially for series where nonlinear forecasting is a good model. Modifying the modeling to remove the extrapolations caused by splines can marginally increase the predictive power, depending on the linear model prediction, or whichever way the prediction is replaced. Finally, between the AIC or cross-validation selection methods in time series, we can conclude that their use does not qualitatively change the results, but only marginally the values, both of which can be used to obtain close results.

It is worth noting that the conclusions obtained here are limited to the applied context, and that further study of the results of the methodology is therefore desired. However, taking into account the research conducted on Kauppi and Virtanen (2021) regarding macroeconomic variables from the United States and our research on Brazilian macroeconomic variables, we are moving in the direction of consensus between the results. We suggest future research for applying the modeling in other contexts, as well as looking for ways that can be done to further improve the two-stage boosting model.

REFERENCES

AKAIKE, Htrotugu. Maximum likelihood identification of Gaussian autoregressive moving average models. en. **Biometrika**, v. 60, n. 2, p. 255–265, 1973. ISSN 0006-3444, 1464-3510. DOI: 10.1093/biomet/60.2.255. Available from: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/60.2.255. Visited on: 12 Mar. 2022.

BAI, Jushan; NG, Serena. Boosting diffusion indices. **Journal of Applied Econometrics**, v. 24, n. 4, 2009. DOI: 10.1002/jae.1063.

BARBOSA, Rafael Barros; FERREIRA, Roberto Tatiwa; SILVA, Thibério Mota da. Previsão de variáveis macroeconômicas brasileiras usando modelos de séries temporais de alta dimensão. pt. **Estudos Econômicos (São Paulo)**, v. 50, n. 1, p. 67–98, Mar. 2020. Number: 1. ISSN 1980-5357. DOI: 10.1590/0101-41615013rrt. Available from: https://www.revistas.usp.br/ee/article/view/142262. Visited on: 15 Jan. 2022.

BHANSALI, R. J. DIRECT AUTOREGRESSIVE PREDICTORS FOR MULTISTEP PREDICTION: ORDER SELECTION AND PERFORMANCE RELATIVE TO THE PLUG IN PREDICTORS. **Statistica Sinica**, v. 7, n. 2, p. 425–449, 1997. Publisher: Institute of Statistical Science, Academia Sinica. ISSN 1017-0405. Available from: https://www.jstor.org/stable/24306087. Visited on: 24 Jan. 2022.

BREIMAN, Leo. Arcing classifier (with discussion and a rejoinder by the author). **The Annals of Statistics**, v. 26, n. 3, 1998. DOI: 10.1214/aos/1024691079.

BREIMAN, Leo. Prediction Games and Arcing Algorithms. **Neural Computation**, v. 11, n. 7, 1999. DOI: 10.1162/089976699300016106.

BUCHEN, Teresa; WOHLRABE, Klaus. Forecasting with many predictors: Is boosting a viable alternative? **Economics Letters**, v. 113, n. 1, p. 16–18, 2011. DOI: 10.1016/j.econlet.2011.05.040.

BÜHLMANN, Peter. Boosting for high-dimensional linear models. **Annals of Statistics**, v. 34, n. 2, 2006. DOI: 10.1214/009053606000000092.

BÜHLMANN, Peter; HOTHORN, Torsten. Boosting algorithms: Regularization, prediction and model fitting. **Statistical Science**, v. 22, n. 4, p. 477–505, 2007. DOI: 10.1214/07-STS242.

BÜHLMANN, Peter; YU, Bin. Boosting with the L2 loss: Regression and classification. **Journal of the American Statistical Association**, v. 98, n. 462, p. 324–339, 2003. DOI: 10.1198/016214503000125.

CEPNI, Oguzhan; GUNEY, I. Ethem; SWANSON, Norman R. Forecasting and nowcasting emerging market GDP growth rates: The role of latent global economic policy uncertainty and macroeconomic data surprise factors. **Journal of Forecasting**, v. 39, n. 1, 2020. DOI: 10.1002/for.2602.

CHU, Jianghao et al. Boosting. In: ADVANCED Studies in Theoretical and Applied Econometrics. 2020. v. 52. P. 431–463. DOI: 10.1007/978-3-030-31150-6_14.

CONTRERAS-REYES, Javier E; GEORG M., Goerg; PALMA, Wilfredo. **afmtools: Estimation, Diagnostic and Forecasting Functions for ARFIMA models**. Feb. 2013. Available from: http://www2.uaem.mx/r-mirror/web/packages/afmtools/index.html. Visited on: 7 Apr. 2022.

DICKEY, David A.; FULLER, Wayne A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. **Journal of the American Statistical Association**, v. 74, 366a, p. 427–431, June 1979. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.1979.10482531. ISSN 0162-1459. DOI: 10.1080/01621459.1979.10482531. Available from: https://doi.org/10.1080/01621459.1979.10482531. Visited on: 9 Mar. 2022.

EILERS, Paul H. C.; MARX, Brian D. Flexible smoothing with B-splines and penalties. **Statistical Science**, v. 11, n. 2, p. 89–121, May 1996. Publisher: Institute of Mathematical Statistics. ISSN 0883-4237, 2168-8745. DOI: 10.1214/ss/1038425655. Available from: https://projecteuclid.org/journals/statistical-science/volume-11/issue-2/Flexible-smoothing-with-B-splines-and-penalties/10.1214/ss/1038425655.full. Visited on: 5 Apr. 2022.

EILERS, Paul H.C.; MARX, Brian D. Splines, knots, and penalties. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 6, 2010. DOI: 10.1002/wics.125.

EPSTEIN, R. J. **A History of Econometrics**. Elsevier, 1987. Google-Books-ID: k6yjBQAAQBAJ. ISBN 978-1-4832-9422-3.

FINDLEY, D. F. Model Selection for Multi-Step-Ahead Forecasting. en. **IFAC Proceedings Volumes**, v. 18, n. 5, p. 1039–1044, July 1985. ISSN 1474-6670. DOI: 10.1016/S1474-6670(17)60699-2. Available from: https://www.sciencedirect.com/science/article/pii/S1474667017606992. Visited on: 24 Jan. 2022.

FREUND, Yoav. Boosting a weak learning algorithm by majority. **Information and Computation**, v. 121, n. 2, 1995. DOI: 10.1006/inco.1995.1136.

FREUND, Yoav; SCHAPIRE, Robert E. Experiments with a New Boosting Algorithm. **Proceedings of the 13th International Conference on Machine Learning**, 1996.

FRIEDMAN, Jerome H. Greedy function approximation: A gradient boosting machine. **Annals of Statistics**, v. 29, n. 5, 2001. DOI: 10.1214/aos/1013203451.

FULLER, Wayne A. **Introduction to Statistical Time Series**. John Wiley & Sons, Sept. 2009. Google-Books-ID: tI6j47m4tVwC. ISBN 978-0-470-31775-4.

GIACOMINI, Raffaella; WHITE, Halbert. Tests of conditional predictive ability. **Econometrica**, v. 74, n. 6, p. 1545–1578, Nov. 2006. DOI: 10.1111/j.1468-0262.2006.00718.x.

GOMES, Luiz Eduardo. **An R package for Ipeadata API database**. Jan. 2022. original-date: 2019-01-22T13:38:38Z. Available from: https://github.com/gomesleduardo/ipeadatar. Visited on: 8 Mar. 2022.

GREEN, P. J.; SILVERMAN, Bernard W. **Nonparametric Regression and Generalized Linear Models: A roughness penalty approach**. New York: Chapman and Hall/CRC, May 1993. ISBN 978-0-429-16105-6. DOI: 10.1201/b15710.

HAMZAÇEBI, Coşkun; AKAY, Diyar; KUTAY, Fevzi. Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. en. **Expert Systems with Applications**, v. 36, 2, Part 2, p. 3839–3844, Mar. 2009. ISSN 0957-4174. DOI: 10.1016/j.eswa.2008.02.042. Available from: https://www.sciencedirect.com/science/article/pii/S0957417408001383. Visited on: 6 Apr. 2022.

HASTIE, Trevor. Comment: Boosting algorithms: Regularization, prediction and model fitting. **Statistical Science**, v. 22, n. 4, 2007. DOI: 10.1214/07-STS242A.

HOTHORN, Torsten; HORNIK, Kurt; ZEILEIS, Achim. Unbiased Recursive Partitioning: A Conditional Inference Framework. en. **Journal of Computational and Graphical Statistics**, v. 15, n. 3, p. 651–674, Sept. 2006. ISSN 1061-8600, 1537-2715. DOI: 10.1198/106186006X133933. Available from: http://www.tandfonline.com/doi/abs/10.1198/106186006X133933. Visited on: 11 Apr. 2022.

HOTHORN, Torsten et al. mboost: Model-Based Boosting. **R package version 2.0-12**, 2011.

ING, Ching-Kang. MULTISTEP PREDICTION IN AUTOREGRESSIVE PROCESSES. en. **Econometric Theory**, v. 19, n. 2, p. 254–279, Apr. 2003. Publisher: Cambridge University Press. ISSN 1469-4360, 0266-4666. DOI: 10.1017/S0266466603192031. Available from: https://www.cambridge.org/core/journals/econometric-theory/article/abs/multistep-prediction-in-autoregressive-processes/305D7DFC3F4E65372D1433C641949E73. Visited on: 24 Jan. 2022.

JI, Yongnan et al. Direct and Recursive Prediction of Time Series Using Mutual Information Selection. en. In: CABESTANY, Joan; PRIETO, Alberto; SANDOVAL, Francisco (Eds.). **Computational Intelligence and Bioinspired Systems**. Berlin, Heidelberg: Springer, 2005. P. 1010–1017. ISBN 978-3-540-32106-4. DOI: 10.1007/11494669_124.

KAUPPI, Heikki; VIRTANEN, Timo. Boosting nonlinear predictability of macroeconomic time series. **International Journal of Forecasting**, v. 37, n. 1, p. 151–170, 2021. DOI: 10.1016/j.ijforecast.2020.03.008. Visited on: 6 Jan. 2022.

KOPOIN, Alexandre; MORAN, Kevin; PARÉ, Jean Pierre. Forecasting regional GDP with factor models: How useful are national and international data? **Economics Letters**, v. 121, n. 2, p. 267–270, 2013. DOI: 10.1016/j.econlet.2013.08.007.

KWIATKOWSKI, Denis et al. 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? en. **Journal of Econometrics**, v. 54, n. 1, p. 159–178, Oct. 1992. ISSN 0304-4076. DOI:

10.1016/0304-4076(92)90104-Y. Available from: https://www.sciencedirect.com/science/article/pii/030440769290104Y. Visited on: 9 Mar. 2022.

LEHMANN, R.; WOHLRABE, K. Looking into the black box of boosting: the case of Germany. **Applied Economics Letters**, v. 23, n. 17, 2016. DOI: 10.1080/13504851.2016.1148246.

LEHMANN, Robert; WOHLRABE, Klaus. Boosting and regional economic forecasting: the case of Germany. **Letters in Spatial and Resource Sciences**, v. 10, n. 2, 2017. DOI: 10.1007/s12076-016-0179-1.

LEHMANN, Robert; WOHLRABE, Klaus. Forecasting GDP at the regional level with many predictors. **German Economic Review**, v. 16, n. 2, p. 226–254, 2015. DOI: 10.1111/geer.12042.

LINDENMEYER, Guilherme; SKORIN, Pedro Pablo; TORRENT, Hudson da Silva. Using boosting for forecasting electric energy consumption during a recession: a case study for the Brazilian State Rio Grande do Sul. **Letters in Spatial and Resource Sciences**, 2021. DOI: 10.1007/s12076-021-00268-3.

MARCELLINO, Massimiliano; STOCK, James H.; WATSON, Mark W. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. en. **Journal of Econometrics**, v. 135, n. 1, p. 499–526, Nov. 2006. ISSN 0304-4076. DOI: 10.1016/j.jeconom.2005.07.020. Available from: https://www.sciencedirect.com/science/article/pii/S030440760500165X. Visited on: 6 Apr. 2022.

MARCELLINO, Massimiliano; STOCK, James H.; WATSON, Mark W. Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. **European Economic Review**, v. 47, n. 1, p. 1–18, Feb. 2003. Publisher: North-Holland. DOI: 10.1016/S0014-2921(02)00206-4.

MCCRACKEN, Michael W.; NG, Serena. FRED-MD: A Monthly Database for Macroeconomic Research. **Journal of Business & Economic Statistics**, v. 34, n. 4, p. 574–589, 2016. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07350015.2015.1086655. ISSN 0735-0015. DOI: 10.1080/07350015.2015.1086655. Available from: https://doi.org/10.1080/07350015.2015.1086655. Visited on: 25 Jan. 2022.

MCNEIL, Ma; LETSCHERT, Ve. Forecasting electricity demand in developing countries: A study of household income and appliance ownership. **Proceedings from eceee Summer Studies**, Waide 1997, 2005.

MEDEIROS, Marcelo C. et al. Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. **Journal of Business and Economic Statistics**, v. 39, n. 1, p. 98–119, 2019. DOI: 10.1080/07350015.2019.1637745.

MOORE, Henry Ludwell. Laws of Wages: An Essay in Statistical Economics. **The Economic Journal**, v. 22, n. 85, p. 66–66, 1911. DOI: 10.2307/2221628.

NONNENBERG, Marcelo José Braga. NOVOS CÁLCULOS DA TAXA EFETIVA REAL DE CÂMBIO PARA O BRASIL. pt, p. 20, 2015.

PARK, B.U.; LEE, Y.K.; HA, S. L2 boosting in kernel regression. **Bernoulli**, v. 15, n. 3, p. 599–613, 2009. Publisher: [Bernoulli Society for Mathematical Statistics and Probability, International Statistical Institute (ISI)]. ISSN 1350-7265. Available from: https://www.jstor.org/stable/20680169. Visited on: 4 Apr. 2022.

PFAFF, Bernhard; ZIVOT, Eric; STIGLER, Matthieu. **urca: Unit Root and Cointegration Tests for Time Series Data**. Sept. 2016. Available from: https://CRAN.R-project.org/package=urca. Visited on: 9 Mar. 2022.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, 2019.

ROBINZONOV, Nikolay; TUTZ, Gerhard; HOTHORN, Torsten. Boosting techniques for nonlinear time series models. **AStA Advances in Statistical Analysis**, v. 96, n. 1, p. 99–122, 2012. DOI: 10.1007/s10182-011-0163-4.

RSTUDIO, Team. RStudio: Integrated Development for R. **Rstudio Team, PBC, Boston, MA URL http://www.rstudio.com/**, 2020. DOI: 10.1145/3132847.3132886.

SCHAPIRE, Robert E. The strength of weak learnability. **Machine Learning**, v. 5, n. 2, 1990. DOI: 10.1007/bf00116037.

SCHMID, Matthias; HOTHORN, Torsten. Boosting additive models using component-wise P-Splines. **Computational Statistics and Data Analysis**, v. 53, n. 2, p. 298–311, 2008. DOI: 10.1016/j.csda.2008.09.009.

SCHWARZ, Gideon. Estimating the Dimension of a Model. **The Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978. Publisher: Institute of Mathematical Statistics. ISSN 0090-5364. Available from: https://www.jstor.org/stable/2958889. Visited on: 12 Mar. 2022.

STOCK, James H.; WATSON, Mark W. **A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series**. June 1998. Series: Working Paper Series. DOI: 10.3386/w6607. Available from: https://www.nber.org/papers/w6607. Visited on: 6 Apr. 2022.

STOCK, James H.; WATSON, Mark W. Forecasting with Many Predictors. **Handbook of Economic Forecasting**, v. 1, p. 515–554, Jan. 2006. Publisher: Elsevier. ISSN 9780444513953. DOI: 10.1016/S1574-0706(05)01010-4.

STONE, M. Cross-Validatory Choice and Assessment of Statistical Predictions. en. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 36, n. 2, p. 111–133, 1974. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1974.tb00994.x. ISSN 2517-6161. DOI: 10.1111/j.2517-6161.1974.tb00994.x. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1974.tb00994.x. Visited on: 6 Apr. 2022.

TAIEB, Souhaib Ben; HYNDMAN, Rob. Boosting multi-step autoregressive forecasts. en. In: PROCEEDINGS of the 31st International Conference on Machine Learning. PMLR, Jan. 2014.

P. 109–117. ISSN: 1938-7228. Available from: https://proceedings.mlr.press/v32/taieb14.html. Visited on: 24 Jan. 2022.

TRAPLETTI, Adrian; HORNIK, Kurt. **tseries: Time Series Analysis and Computational Finance**. Nov. 2021. Available from: https://CRAN.R-project.org/package=tseries. Visited on: 9 Mar. 2022.

VARIAN, Hal R. Big data: New tricks for econometrics. **Journal of Economic Perspectives**, v. 28, n. 2, p. 3–28, 2014. DOI: 10.1257/jep.28.2.3.

WOHLRABE, Klaus; BUCHEN, Teresa. Assessing the macroeconomic forecasting performance of boosting: Evidence for the United States, the Euro area and Germany. **Journal of Forecasting**, v. 33, n. 4, 2014. DOI: 10.1002/for.2293.

YOON, Jaehyun. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. en. **Computational Economics**, v. 57, n. 1, p. 247–265, Jan. 2021. ISSN 1572-9974. DOI: 10.1007/s10614-020-10054-w. Available from: https://doi.org/10.1007/s10614-020-10054-w. Visited on: 15 Jan. 2022.

ZENG, Jing. Forecasting Aggregates with Disaggregate Variables: Does Boosting Help to Select the Most Relevant Predictors? **Journal of Forecasting**, v. 36, n. 1, p. 74–90, 2017. DOI: 10.1002/for.2415.

APPENDIX A – DATA

Here we present the data set gathered and used to do our research. Let y_t be the time series in question, transformations 1, 2 and 3 mean respectively y_t, Δy_t and $\Delta^2 y_t$. Transformations 4, 5 and 6 mean respectively $log(y_t)$, $\Delta log(y_t)$ and $\Delta^2 log(y_t)$. Transformation 7 means direct growth ($\frac{y_{t+1}-y_t}{y_t}$).

Table 7 – Data Set

Series Name	Theme	Transf.	Source
Stock index - Ibovespa - closing	Financial	1	Anbima
Savings - profitability - deposits from 04.05.2012 (1st day of the month)	Financial	5	Anbima
Apparent consumption - fuel alcohol - average - quantity/day	Consumption and sales	5	ANP
Apparent Consumption - oil products - average - quantity/day	Consumption and sales	5	ANP
Apparent Consumption - gasoline - average - quantity/day	Consumption and sales	5	ANP
Apparent Consumption - LPG - average - quantity/day	Consumption and sales	5	ANP
Apparent Consumption - oil products - other - average - quantity/day	Consumption and sales	5	ANP
Apparent Consumption - fuel oil - average - quantity/day	Consumption and sales	5	ANP
Apparent Consumption - diesel oil - average - quantity/day	Consumption and sales	5	ANP
Production - petroleum - average - quantity/day	Production	5	ANP
Restricted means of payment - M1 - demand deposits - average	Currency and credit	5	Bacen
DLSP - net external debt - states and municipalities	Public Finances	6	Bacen
DLSP - net external debt - federal government and Central Bank	Public Finances	2	Bacen
DLSP - net internal debt - states and municipalities	Public Finances	6	Bacen
DLSP - net internal debt - federal government and Central Bank	Public Finances	6	Bacen
DSLP - total - states and municipalities	Public Finances	6	Bacen
DSLP - total - federal government and Central Bank	Public Finances	6	Bacen
Exchange rate - R$/US$ - commercial - purchase - average	Exchange	5	Bacen
IPCA - free prices - commercial - rate of change	Prices	1	Bacen
IPCA - free prices - non-tradable - rate of change	Prices	2	Bacen
IPCA - free prices - rate of change	Prices	2	Bacen
IPCA - free prices - durable - rate of change	Prices	2	Bacen
IPCA - free prices - non-durable goods - rate of change	Prices	1	Bacen
IPCA - free prices - semi-durable goods - rate of change	Prices	2	Bacen
IPCA - free prices - services - rate of change	Prices	2	Bacen
IPCA - monitored prices - rate of change	Prices	2	Bacen
Restricted monetary base - M0 - average	Currency and credit	5	Bacen
NFSP - states and municipalities - nominal	Public Finances	2	Bacen
GDP	National Accounts	6	Bacen
International Reserves	Balance of payments	6	Bacen
Gold - monthly percentage variation	Financial	1	Bacen
Interest rate - TJLP	Financial	5	Bacen
Interest rate - Over / Selic - accumulated in the month	Financial	5	Bacen
Interest rate - TR - first day of the month	Financial	2	Bacen
Balance of payments - Financial account - balance (borrowings - concessions)	Balance of payments	2	Bacen
Balance of payments - capital account - balance	Balance of payments	2	Bacen
Balance of payments - current transactions - balance	Balance of payments	2	Bacen
Installed capacity utilization - industry - average	Production	5	FGV/Conj. Econ.
Industrial indicators - hours worked - industry - index (average 2006 = 100)	Employment	5	CNI
Industrial indicators - capacity utilization - industry	Production	5	CNI
Industrial Indicators - personnel employed - industry	Employment	6	CNI
Tax on Distribution of Goods (ICMS) - Brazil	Public Finances	5	Min. Economia
Net fixed capital stock (2010 prices)	Capital stock	6	IPEA
Electric energy - consumption - trade - quantity	Consumption and sales	5	Eletrobras
Electric energy - consumption - industry - quantity	Consumption and sales	5	Eletrobras
Electric energy - consumption - residential - quantity	Consumption and sales	5	Eletrobras
Electric energy - consumption - trade - average rate per MWh	Consumption and sales	5	Eletrobras

Table 7 continued from previous page

Series Name	Theme	Transf.	Source
Electric energy - consumption - industry - average tariff per MWh	Consumption and sales	5	Eletrobras
Electric energy - consumption - household - average tariff per MWh	Consumption and sales	5	Eletrobras
Consumer Confidence Index (ICC)	Perception and expectation	5	Fecomercio SP
Current economic conditions index (ICEA)	Consumption and sales	5	Fecomercio SP
Consumer expectations index (IEC)	Perception and expectation	5	Fecomercio SP
MG industry - capacity utilization - mineral extraction - average	Production	5	Fiemg
Real salary - average - industry - index (average 2006 = 100) - SP	Salary and income	5	Fiesp
Real salary - industry - index (average 2006 = 100) - SP	Salary and income	5	Fiesp
Real Sales - Industry - Index (average 2006 = 100) - SP	Consumption and sales	5	Fiesp
CPI - general - index (Jun. 1994 = 100) - RMSP	Prices	6	Fipe
Employment - industry - RJ - index (average 2006 = 100)	Employment	6	Firjan
Real income - industry - index - RJ - (average 2006 = 100)	Consumption and sales	5	Firjan
Wage bill - industry - index - RJ - (average 2006 = 100)	Employment	5	Firjan
Imports - prices - index - (average 2018 = 100)	Foreign trade	6	Funcex
Imports - consumer durables - (FOB) (new series)	Foreign trade	5	Funcex
Imports - non-durable consumer goods - (FOB) (new series)	Foreign trade	5	Funcex
Imports - intermediate goods - (FOB) (new series)	Foreign trade	5	Funcex
Imports - capital goods - (FOB) (new series)	Foreign trade	5	Funcex
Exports - prices - index (average 2018 = 100)	Foreign trade	6	Funcex
Exports - consumer durables - (FOB) (new series)	Foreign trade	5	Funcex
Exports - non-durable consumer goods - (FOB) (new series)	Foreign trade	5	Funcex
Exports - intermediate goods - (FOB) (new series)	Foreign trade	5	Funcex
Exports - capital goods - (FOB) (new series)	Foreign trade	5	Funcex
Apparent Consumption - consumer goods - index (2012 average = 100)	National Accounts	5	IPEA
Apparent Consumption - Consumer Durables - Index (2012 average = 100)	National Accounts	5	IPEA
Apparent consumption - consumer durables	National Accounts	5	IPEA
Apparent consumption - consumer goods	National Accounts	5	IPEA
Apparent consumption - semi- and non-durable consumer goods	National Accounts	5	IPEA
Apparent consumption - semi- and non-durable consumer goods	National Accounts	5	IPEA
Apparent consumption - intermediate goods	National Accounts	5	IPEA
Apparent consumption - intermediate goods	National Accounts	5	IPEA
Apparent Consumption - Capital Goods - Index (2012 average = 100)	National Accounts	5	IPEA
Apparent consumption - capital goods - seasonally	National Accounts	5	IPEA
IPEA GFCF indicator - index (average 1995 = 100)	National Accounts	5	IPEA
IPEA GFCF indicator - civil construction - index (average 1995 = 100)	National Accounts	5	IPEA
IPEA GFCF indicator - civil construction	National Accounts	5	IPEA
IPEA GFCF indicator - seasonally adjusted	National Accounts	5	IPEA
Minimum wage - purchasing power parity (PPP)	Salary and income	5	IPEA
Real minimum wage	Salary and income	5	IPEA
Exchange rate - real effective - IPA-EP-DI - imports - durable goods	Exchange	5	IPEA
Exchange rate - real effective - IPA-EP-DI - imports - non-durable goods	Exchange	5	IPEA
Exchange rate - real effective - IPA-EP-DI - imports - intermediate goods	Exchange	5	IPEA
Exchange rate - real effective - IPA-EP-DI - imports - capital goods	Exchange	5	IPEA
Exchange rate - real effective - IPA-EP-DI - imports - fuel	Exchange	5	IPEA
Exchange rate - real effective - IPA-DI - imports - index (average 2010 = 100)	Exchange	5	IPEA
Exchange rate - real effective - INPC - imports - index (average 2010 = 100)	Exchange	5	IPEA
Exchange rate - real effective - Weighted IPA - exports - basic	Exchange	5	IPEA
Exchange rate - real effective - IPA-DI-Origin - exports - manufactured	Exchange	5	IPEA
Exchange rate - real effective - INPC - exports - manufactured	Exchange	5	IPEA
Exchange rate - real effective - IPA-DI - Origin - exports - semimanufactured	Exchange	5	IPEA
Exchange rate - real effective - IPA-DI - exports - index (average 2010 = 100)	Exchange	5	IPEA
Exchange rate - real effective - INPC - exports - index (average 2010 = 100)	Exchange	5	IPEA
IGP-10 - general - index (aug. 1994 = 100)	Prices	5	FGV/Conj. Econ.
IGP-M - general - index (aug. 1994 = 100)	Prices	5	FGV/Conj. Econ.
IGP-OG - general - index (aug. 1994 = 100)	Prices	5	FGV/Conj. Econ.
INCC-DI - general - index (Aug. 1994 = 100)	Prices	6	FGV/Conj. Econ.
IPA-10 - general - index (aug. 1994 = 100)	Prices	5	FGV/Conj. Econ.
IPA-DI - Origin - Agricultural Products - index (aug. 1994 = 100)	Prices	5	FGV/Conj. Econ.

Table 7 continued from previous page

Series Name	Theme	Transf.	Source
IPA-DI - Origin - Industrial Products - Index (aug. 1994 = 100)	Prices	5	FGV/Conj. Econ.
INPC - general - index (dec. 1993 = 100)	Prices	6	IBGE/SNIPC
INPC - food & beverages - rate of change	Prices	1	IBGE/SNIPC
INPC - household goods - rate of change	Prices	1	IBGE/SNIPC
INPC - personal expenses - rate of change	Prices	2	IBGE/SNIPC
INPC - communication - rate of change	Prices	2	IBGE/SNIPC
INPC - education, reading and stationery - rate of change	Prices	2	IBGE/SNIPC
INPC - housing - rate of change	Prices	1	IBGE/SNIPC
INPC - health and personal care - rate of change	Prices	2	IBGE/SNIPC
INPC - transport - rate of change	Prices	2	IBGE/SNIPC
INPC - clothing - rate of change	Prices	2	IBGE/SNIPC
IPCA - food and beverages - rate of change	Prices	1	IBGE/SNIPC
IPCA - household goods - rate of change	Prices	1	IBGE/SNIPC
IPCA - personal expenses - rate of change	Prices	2	IBGE/SNIPC
IPCA - communication - rate of change	Prices	2	IBGE/SNIPC
IPCA - education, reading and stationery - rate of change	Prices	2	IBGE/SNIPC
IPCA - housing - rate of change	Prices	1	IBGE/SNIPC
IPCA - health and personal care - rate of change	Prices	2	IBGE/SNIPC
IPCA - transport - rate of change	Prices	2	IBGE/SNIPC
IPCA - clothing - rate of change	Prices	2	IBGE/SNIPC
Average income - real - wage earners - main work	Salary and income	5	Seade/PED
Unemployment rate - open - RMSP	Employment	5	Seade/PED
Hidden - precarious unemployment rate - RMSP	Employment	6	Seade/PED
Unemployment rate - hidden - RMSP	Employment	6	Seade/PED
Unemployment Rate - RMSP	Employment	5	Seade/PED
Imports - (FOB)	Foreign trade	5	MDIC/SECEX
Exports - (FOB)	Foreign trade	5	MDIC/SECEX
Savings - deposit - SBPE and rural - balance	Financial	6	Bacen Outras/SGS
Tax on imports (II) - total - gross revenue	Public Finances	5	Min. Economia
Tax on Financial Transactions (IOF) - total - gross revenue	Public Finances	5	Min. Economia
Tax on Industrialized Products (IPI) - total - gross revenue	Public Finances	5	Min. Economia
Income tax (IR) - withholding - capital income - gross revenue	Public Finances	5	Min. Economia
Income tax (IR) - withholding - remittance income abroad - gross revenue	Public Finances	5	Min. Economia
Income tax (IR) - withheld - labor income - gross revenue	Public Finances	5	Min. Economia
Tax on rural land property (ITR) - gross revenue	Public Finances	2	Min. Economia

APPENDIX B – RELATIVE ROOT MEAN SQUARE ERROR - ALL SERIES

Table 8 – All Cases - rRMSE Relative to Linear Model - $h = 1$

Series	BOLS	BSpline	TSBoost	BSpline*	TSBoost*	Tree
Stock index - Ibovespa - closing Financial 1 Anbima	1.000	1.000	1.000	1.000	1.000	0.999
Production - petroleum - average - quantity/day Production 5 ANP	1.001	0.990	0.997	0.990	0.997	0.977
IGP-10 - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.991	1.018	1.019	1.020	1.019	1.111
IGP-M - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.005	1.018	1.028	1.012	1.028	1.117
IGP-OG - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.977	0.997	0.986	0.997	0.986	1.038
INCC-DI - general - index (Aug. 1994 = 100) Prices 6 FGV/Conj. Econ.	1.145	1.116	1.008	1.116	1.008	1.065
IPA-10 - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.003	1.027	1.018	1.027	1.018	1.072
IPA-DI - Origin - Agricultural Products - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.002	1.000	0.997	1.000	0.997	1.028
IPA-DI - Origin - Industrial Products - Index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.037	1.013	1.018	1.014	1.018	1.039
INPC - general - index (dec. 1993 = 100) Prices 6 IBGE/SNIPC	1.036	1.050	0.984	1.035	0.983	1.040
INPC - food & beverages - rate of change Prices 1 IBGE/SNIPC	0.961	0.959	0.994	0.959	0.994	0.956
INPC - household goods - rate of change Prices 1 IBGE/SNIPC	0.991	0.994	1.010	0.994	1.010	0.980
Restricted means of payment - M1 - demand deposits - average Currency and credit 5 Bacen	1.126	0.950	0.979	0.948	0.979	1.062
INPC - personal expenses - rate of change Prices 2 IBGE/SNIPC	1.087	1.070	1.053	1.070	1.053	1.102
INPC - communication - rate of change Prices 2 IBGE/SNIPC	1.141	1.097	1.004	1.097	1.004	1.268
INPC - education, reading and stationery - rate of change Prices 2 IBGE/SNIPC	1.279	0.887	1.000	0.970	1.000	1.076
INPC - housing - rate of change Prices 1 IBGE/SNIPC	0.968	0.965	0.961	0.965	0.961	0.958
INPC - health and personal care - rate of change Prices 2 IBGE/SNIPC	1.080	1.116	0.997	0.983	0.994	1.106
INPC - transport - rate of change Prices 2 IBGE/SNIPC	1.042	1.021	1.009	1.009	1.029	1.062
INPC - clothing - rate of change Prices 2 IBGE/SNIPC	1.109	1.084	0.967	1.050	0.967	1.155
IPCA - food and beverages - rate of change Prices 1 IBGE/SNIPC	0.966	0.959	0.992	0.959	0.992	0.976
IPCA - household goods - rate of change Prices 1 IBGE/SNIPC	0.990	1.008	1.020	1.008	1.020	0.987
IPCA - personal expenses - rate of change Prices 2 IBGE/SNIPC	1.111	1.137	1.001	1.137	1.001	1.213
DLSP - net external debt - states and municipalities Public Finances 6 Bacen	1.058	1.130	1.018	1.130	1.018	1.250
IPCA - communication - rate of change Prices 2 IBGE/SNIPC	1.263	1.141	0.976	1.141	0.976	1.384
IPCA - education, reading and stationery - rate of change Prices 2 IBGE/SNIPC	1.433	0.976	0.934	0.976	0.934	1.571
IPCA - housing - rate of change Prices 1 IBGE/SNIPC	0.976	0.973	0.982	0.980	0.982	0.967
IPCA - health and personal care - rate of change Prices 2 IBGE/SNIPC	1.068	1.041	1.014	1.041	1.014	1.079
IPCA - transport - rate of change Prices 2 IBGE/SNIPC	1.061	1.121	1.119	1.121	1.119	1.217
IPCA - clothing - rate of change Prices 2 IBGE/SNIPC	1.139	1.080	0.979	1.041	0.974	1.213
Average income - real - wage earners - main work Salary and income 5 Seade/PED	1.023	1.088	1.082	1.088	1.082	1.081
Unemployment rate - open - RMSP Employment 5 Seade/PED	1.018	1.026	1.025	1.026	1.025	1.125
Hidden - precarious unemployment rate - RMSP Employment 6 Seade/PED	1.129	1.241	1.024	1.201	1.024	1.333
Unemployment rate - hidden - RMSP Employment 6 Seade/PED	1.112	1.202	1.053	1.202	1.053	1.237
DLSP - net external debt - federal government and Central Bank Public Finances 2 Bacen	0.965	0.988	0.985	0.960	0.960	0.960
Unemployment Rate - RMSP Employment 5 Seade/PED	1.000	1.007	1.017	1.110	1.014	0.982
Imports - (FOB) Foreign trade 5 MDIC/SECEX	0.976	0.974	0.981	0.974	0.981	0.994
Exports - (FOB) Foreign trade 5 MDIC/SECEX	1.028	1.032	0.991	0.986	0.984	1.045
Savings - deposit - SBPE and rural - balance Financial 6 Bacen Outras/SGS	1.036	0.971	0.984	0.971	0.984	1.049
Tax on imports (II) - total - gross revenue Public Finances 5 Min. Economia	0.987	0.990	1.000	0.990	1.000	1.017
Tax on Financial Transactions (IOF) - total - gross revenue Public Finances 5 Min. Economia	0.996	0.989	1.000	0.989	1.000	1.054
Tax on Industrialized Products (IPI) - total - gross revenue Public Finances 5 Min. Economia	1.038	0.959	0.978	0.988	0.981	1.019
Income tax (IR) - withholding - capital income - gross revenue Public Finances 5 Min. Economia	1.097	0.988	0.990	0.988	0.990	1.381
Income tax (IR) - withholding - remittance income abroad	1.156	1.026	0.968	1.026	0.968	1.061
Income tax (IR) - withheld - labor income - gross revenue Public Finances 5 Min. Economia	1.188	1.131	1.034	1.131	1.034	1.396
DLSP - net internal debt - states and municipalities Public Finances 6 Bacen	1.042	1.168	1.016	1.179	1.016	1.351
Tax on rural land property (ITR) - gross revenue Public Finances 2 Min. Economi	1.137	1.313	1.129	0.932	1.000	5.792
DLSP - net internal debt - federal government and Central Bank Public Finances 6 Bacen	1.049	1.057	0.902	1.057	0.902	1.309
DSLP - total - states and municipalities Public Finances 6 Bacen	1.052	1.133	1.040	1.134	1.042	1.212
DSLP - total - federal government and Central Bank Public Finances 6 Bacen	0.996	1.091	0.961	1.091	0.961	1.289
Exchange rate - R/U S - commercial - purchase - average Exchange 5 Bacen	1.007	1.012	1.000	1.012	1.000	0.992
IPCA - free prices - commercial - rate of change Prices 1 Bacen	0.964	0.976	1.000	0.991	1.000	1.011
Savings - profitability - deposits from 04.05.2012 (1st day of the month) Financial 5 Anbima	1.046	1.024	0.988	1.024	0.988	1.138
IPCA - free prices - non-tradable - rate of change Prices 2 Bacen	1.104	1.234	1.004	1.234	1.004	1.250
IPCA - free prices - rate of change Prices 2 Bacen	1.035	1.065	0.997	1.051	0.997	1.085
IPCA - free prices - durable - rate of change Prices 2 Bacen	1.089	1.135	0.998	1.135	0.998	1.164
IPCA - free prices - non-durable goods - rate of change Prices 1 Bacen	0.982	0.982	1.025	1.021	1.025	1.004
IPCA - free prices - semi-durable goods - rate of change Prices 2 Bacen	1.136	1.189	0.986	1.122	0.987	1.314
IPCA - free prices - services - rate of change Prices 2 Bacen	1.166	1.202	0.984	1.202	0.984	1.299
IPCA - monitored prices - rate of change Prices 2 Bacen	1.035	1.053	0.982	1.053	0.982	1.101
Restricted monetary base - M0 - average Currency and credit 5 Bacen	1.014	1.017	1.058	1.017	1.058	1.450
NFSP - states and municipalities - nominal Public Finances 2 Bacen	0.942	0.918	0.994	0.918	0.994	0.936
GDP National Accounts 6 Bacen	1.425	1.418	1.015	1.405	1.015	1.509
Apparent consumption - fuel alcohol - average - quantity/day Consumption and sales 5 ANP	1.035	1.026	1.006	1.026	1.006	1.003
International Reserves Balance of payments 6 Bacen	1.023	1.042	1.015	1.042	1.015	1.150
Gold - monthly percentage variation Financial 1 Bacen	1.005	1.004	1.005	1.004	1.005	1.005
Interest rate - TJLP Financial 5 Bacen	1.054	0.957	1.125	0.957	1.125	0.932
Interest rate - Over / Selic - accumulated in the month Financial 5 Bacen	1.055	0.997	0.966	0.997	0.966	1.041

Table 8 continued from previous page

Series	BOLS	BSpline	TSBoost	BSpline*	TSBoost*	Tree
Interest rate - TR - first day of the month Financial 2 Bacen	1.038	0.944	0.943	0.944	0.943	1.147
Balance of payments - Financial account - balance (borrowings - concessions) BP	0.985	1.036	0.999	1.045	1.001	1.122
Balance of payments - capital account - balance Balance of payments 2 Bacen	1.060	1.292	1.039	1.292	1.039	1.294
Balance of payments - current transactions - balance Balance of payments 2 Bacen	1.005	1.023	1.004	1.001	1.005	1.016
Installed capacity utilization - industry - average Production 5 FGV/Conj. Econ.	0.978	0.983	1.034	0.977	1.034	1.012
Industrial indicators - hours worked - industry - index (average 2006 = 100)	0.983	0.829	0.949	0.829	0.949	0.803
Apparent Consumption - oil products - average - quantity/day	1.027	0.992	0.994	0.990	1.002	0.974
Industrial indicators - capacity utilization - industry Production 5 CNI	1.024	0.886	0.940	0.886	0.940	0.873
Industrial Indicators - personnel employed - industry Employment 6 CNI	1.050	1.026	1.000	1.026	1.000	1.170
Tax on Distribution of Goods (ICMS) - Brazil Public Finances 5 Min. Economia	1.043	1.061	1.053	1.061	1.053	1.025
Net fixed capital stock (2010 prices) Capital stock 6 IPEA	1.021	1.028	1.014	1.023	1.008	1.039
Electric energy - consumption - trade - quantity Consumption and sales 5 Eletrobras	1.012	1.058	0.996	1.058	0.996	1.058
Electric energy - consumption - industry - quantity Consumption and sales 5 Eletrobras	1.028	1.008	0.967	1.008	0.967	1.021
Electric energy - consumption - residential - quantity Consumption and sales 5 Eletrobras	1.067	1.030	1.005	1.030	1.005	1.030
Electric energy - consumption - trade - average rate per MWh	0.980	0.968	0.949	0.959	0.945	0.971
Electric energy - consumption - industry - average tariff per MWh	0.931	0.933	1.007	1.001	1.007	0.836
Apparent Consumption - gasoline - average - quantity/day Consumption and sales 5 ANP	0.997	0.959	0.992	0.959	0.992	1.040
Consumer Confidence Index (ICC) Perception and expectation 5 Fecomercio SP	0.974	0.988	0.986	0.988	0.986	0.998
Current economic conditions index (ICEA) Consumption and sales 5 Fecomercio SP	0.963	0.981	0.989	0.981	0.989	0.996
Consumer expectations index (IEC) Perception and expectation 5 Fecomercio SP	0.994	1.003	0.995	1.003	0.995	0.997
MG industry - capacity utilization - mineral extraction - average Production 5 Fiemg	0.997	1.022	1.023	0.998	0.998	1.001
Real salary - average - industry - index (average 2006 = 100) - SP Salary and income 5 Fiesp	0.992	1.004	1.000	1.004	1.000	1.120
Real salary - industry - index (average 2006 = 100) - SP Salary and income 5 Fiesp	1.040	1.066	0.978	1.066	0.978	1.171
Real Sales - Industry - Index (average 2006 = 100) - SP Consumption and sales 5 Fiesp	0.963	0.953	1.003	0.907	0.993	0.994
CPI - general - index (Jun. 1994 = 100) - RMSP Prices 6 Fipe	1.036	1.043	1.010	1.043	1.010	1.148
Employment - industry - RJ - index (average 2006 = 100) Employment 6 Firjan	1.006	1.049	0.999	1.049	0.999	1.066
Real income - industry - index - RJ - (average 2006 = 100) Consumption and sales 5 Firjan	1.020	1.054	1.034	1.056	1.064	1.110
Apparent Consumption - LPG - average - quantity/day Consumption and sales 5 ANP	1.186	1.180	0.993	1.073	1.001	1.205
Wage bill - industry - index - RJ - (average 2006 = 100) Employment 5 Firjan	1.027	0.960	0.999	0.960	0.999	0.971
Imports - prices - index - (average 2018 = 100) Foreign trade 6 Funcex	1.044	1.095	1.033	1.095	1.033	1.160
Imports - consumer durables - (FOB) (new series) Foreign trade 5 Funcex	0.988	0.982	1.019	0.982	1.019	0.996
Imports - non-durable consumer goods - (FOB) (new series) Foreign trade 5 Funcex	1.006	1.040	1.002	1.040	1.002	1.099
Imports - intermediate goods - (FOB) (new series) Foreign trade 5 Funcex	1.006	0.975	0.983	0.975	0.983	0.982
Imports - capital goods - (FOB) (new series) Foreign trade 5 Funcex	1.020	1.029	1.027	1.060	1.033	1.115
Exports - prices - index (average 2018 = 100) Foreign trade 6 Funcex	1.048	1.057	1.036	1.057	1.036	1.077
Exports - consumer durables - (FOB) (new series) Foreign trade 5 Funcex	0.984	0.975	1.002	0.975	1.002	1.024
Exports - non-durable consumer goods - (FOB) (new series) Foreign trade 5 Funcex	1.017	1.073	1.000	1.057	1.000	1.063
Exports - intermediate goods - (FOB) (new series) Foreign trade 5 Funcex	1.015	1.033	1.002	1.017	1.002	1.114
Apparent Consumption - oil products - other - average - quantity/day Consumption and sales 5 ANP	1.033	0.909	1.007	0.909	1.007	0.992
Exports - capital goods - (FOB) (new series) Foreign trade 5 Funcex	1.042	0.996	0.935	0.996	0.935	1.046
Apparent Consumption - consumer goods - index (2012 average = 100) National Accounts 5 IPEA	1.048	0.954	0.976	0.974	0.979	0.972
Apparent Consumption - Consumer Durables - Index (2012 average = 100) National Accounts 5 IPEA	1.027	1.010	1.000	1.010	1.000	1.064
Apparent consumption - consumer durables National Accounts 5 IPEA	1.001	1.015	1.015	1.015	1.015	1.017
Apparent consumption - consumer goods National Accounts 5 IPEA	1.051	1.043	1.069	1.004	1.004	1.102
Apparent consumption - semi- and non-durable consumer goods National Accounts 5 IPEA	1.142	1.030	0.965	1.028	0.995	1.013
Apparent consumption - semi- and non-durable consumer goods National Accounts 5 IPEA	1.069	1.146	1.011	1.011	1.010	1.113
Apparent consumption - intermediate goods National Accounts 5 IPEA	1.007	0.949	1.016	0.949	1.016	1.013
Apparent consumption - intermediate goods National Accounts 5 IPEA	1.010	1.016	1.000	1.016	1.000	1.018
Apparent Consumption - Capital Goods - Index (2012 average = 100) National Accounts 5 IPEA	1.010	1.015	0.999	1.016	1.000	1.077
Apparent Consumption - fuel oil - average - quantity/day Consumption and sales 5 ANP	1.007	1.056	1.073	1.056	1.066	1.049
Apparent consumption - capital goods - seasonally National Accounts 5 IPEA	0.936	0.944	0.966	0.909	0.939	1.004
IPEA GFCF indicator - index (average 1995 = 100) National Accounts 5 IPEA	0.975	1.021	1.000	1.007	1.000	1.028
IPEA GFCF indicator - civil construction - index (average 1995 = 100) National Accounts 5 IPEA	0.964	0.975	1.003	0.986	1.003	1.068
IPEA GFCF indicator - civil construction National Accounts 5 IPEA	0.957	0.987	1.010	0.987	1.010	0.973
IPEA GFCF indicator - seasonally adjusted National Accounts 5 IPEA	0.965	0.970	0.994	0.985	0.994	0.946
Minimum wage - purchasing power parity (PPP) Salary and income 5 IPEA	0.919	0.833	1.004	0.833	1.004	0.768
Real minimum wage Salary and income 5 IPEA	0.937	0.885	1.014	0.885	1.014	0.906
Exchange rate - real effective - IPA-EP-DI - imports - durable goods Exchange 5 IPEA	1.001	1.001	1.002	1.001	1.002	1.011
Exchange rate - real effective - IPA-EP-DI - imports - non-durable goods Exchange 5 IPEA	0.997	1.009	0.998	1.009	0.998	1.018
Exchange rate - real effective - IPA-EP-DI - imports - intermediate goods Exchange 5 IPEA	1.000	1.013	1.000	1.013	1.000	1.008
Apparent Consumption - diesel oil - average - quantity/day Consumption and sales 5 ANP	1.105	1.144	1.001	1.028	1.001	1.075
Exchange rate - real effective - IPA-EP-DI - imports - capital goods Exchange 5 IPEA	0.996	1.009	0.998	1.009	0.998	1.004
Exchange rate - real effective - IPA-EP-DI - imports - fuel Exchange 5 IPEA	1.005	1.001	0.996	1.001	0.996	1.027
Exchange rate - real effective - IPA-DI - imports - index (average 2010 = 100) Exchange 5 IPEA	1.011	1.029	0.997	1.029	0.997	1.034
Exchange rate - real effective - INPC - imports - index (average 2010 = 100) Exchange 5 IPEA	0.997	1.008	1.000	1.008	1.000	1.009
Exchange rate - real effective - Weighted IPA - exports - basic Exchange 5 IPEA	1.009	1.018	1.003	1.018	1.003	1.048
Exchange rate - real effective - IPA-DI-Origin - exports - manufactured Exchange 5 IPEA	1.012	1.024	1.001	1.024	1.001	1.013
Exchange rate - real effective - INPC - exports - manufactured Exchange 5 IPEA	1.009	1.016	1.000	1.016	1.000	1.016
Exchange rate - real effective - IPA-DI - Origin - exports - semimanufactured Exchange 5 IPEA	1.000	1.012	1.000	1.012	1.000	1.030
Exchange rate - real effective - IPA-DI - exports - index (average 2010 = 100) Exchange 5 IPEA	1.017	1.024	0.998	1.024	0.998	1.044
Exchange rate - real effective - INPC - exports - index (average 2010 = 100) Exchange 5 IPEA	1.009	0.998	1.000	0.998	1.000	1.020

Table 9 – All Cases - rRMSE Relative to Linear Model - $h = 6$

Series	BOLS	BSpline	TSBoost	BSpline*	TSBoost*	Tree
Stock index - Ibovespa - closing Financial 1 Anbima	0.999	1.006	1.009	1.006	1.009	1.004
Production - petroleum - average - quantity/day Production 5 ANP	1.008	0.993	1.000	0.993	1.000	1.022
IGP-10 - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.993	1.003	0.999	1.002	0.999	1.009
IGP-M - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.986	0.991	0.999	0.983	0.999	1.024
IGP-OG - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.987	0.991	1.000	0.989	1.000	1.020
INCC-DI - general - index (Aug. 1994 = 100) Prices 6 FGV/Conj. Econ.	1.592	1.608	0.967	1.608	0.967	1.892
IPA-10 - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.991	0.986	0.995	0.987	0.995	0.998
IPA-DI - Origin - Agricultural Products - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.995	1.009	1.015	1.009	1.015	1.009
IPA-DI - Origin - Industrial Products - Index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.969	0.945	0.980	0.955	0.979	0.893
INPC - general - index (dec. 1993 = 100) Prices 6 IBGE/SNIPC	1.038	1.128	1.002	1.091	1.003	1.366
INPC - food & beverages - rate of change Prices 1 IBGE/SNIPC	1.007	1.015	0.999	1.015	0.999	0.993
INPC - household goods - rate of change Prices 1 IBGE/SNIPC	0.979	1.051	1.040	1.051	1.040	0.942
Restricted means of payment - M1 - demand deposits - average Currency and credit 5 Bacen	0.994	0.912	0.910	0.912	0.910	0.928
INPC - personal expenses - rate of change Prices 2 IBGE/SNIPC	1.011	1.161	0.982	1.161	0.982	1.107
INPC - communication - rate of change Prices 2 IBGE/SNIPC	1.310	0.941	0.991	0.941	0.991	1.601
INPC - education, reading and stationery - rate of change Prices 2 IBGE/SNIPC	2.631	1.002	1.019	1.009	1.027	1.558
INPC - housing - rate of change Prices 1 IBGE/SNIPC	0.984	0.986	0.991	0.970	0.991	0.987
INPC - health and personal care - rate of change Prices 2 IBGE/SNIPC	1.047	1.288	1.024	1.045	0.977	1.346
INPC - transport - rate of change Prices 2 IBGE/SNIPC	1.000	1.053	0.955	1.011	0.924	1.156
INPC - clothing - rate of change Prices 2 IBGE/SNIPC	1.006	1.087	0.946	0.990	0.947	1.165
IPCA - food and beverages - rate of change Prices 1 IBGE/SNIPC	0.994	1.001	0.975	1.001	0.975	0.998
IPCA - household goods - rate of change Prices 1 IBGE/SNIPC	0.997	1.038	1.042	1.038	1.042	0.998
IPCA - personal expenses - rate of change Prices 2 IBGE/SNIPC	1.031	1.216	0.986	1.216	0.986	1.287
DLSP - net external debt - states and municipalities Public Finances 6 Bacen	1.047	1.098	0.961	1.098	0.961	1.378
IPCA - communication - rate of change Prices 2 IBGE/SNIPC	1.237	0.887	0.851	0.887	0.851	1.131
IPCA - education, reading and stationery - rate of change Prices 2 IBGE/SNIPC	3.086	0.983	0.962	0.983	0.962	1.619
IPCA - housing - rate of change Prices 1 IBGE/SNIPC	0.988	0.989	1.005	0.993	1.000	0.991
IPCA - health and personal care - rate of change Prices 2 IBGE/SNIPC	1.056	1.168	0.992	1.168	0.992	1.246
IPCA - transport - rate of change Prices 2 IBGE/SNIPC	1.016	1.296	0.984	1.296	0.984	1.438
IPCA - clothing - rate of change Prices 2 IBGE/SNIPC	1.009	1.088	0.960	0.998	0.958	1.160
Average income - real - wage earners - main work Salary and income 5 Seade/PED	0.975	0.952	0.922	0.952	0.922	0.955
Unemployment rate - open - RMSP Employment 5 Seade/PED	0.986	1.019	1.226	1.019	1.226	1.027
Hidden - precarious unemployment rate - RMSP Employment 6 Seade/PED	1.725	1.890	1.001	1.807	1.001	2.277
Unemployment rate - hidden - RMSP Employment 6 Seade/PED	1.464	1.556	1.012	1.556	1.012	1.944
DLSP - net external debt - federal government and Central Bank Public Finances 2 Bacen	0.972	1.051	1.037	0.983	0.995	1.129
Unemployment Rate - RMSP Employment 5 Seade/PED	0.999	1.064	0.994	1.050	0.998	1.052
Imports - (FOB) Foreign trade 5 MDIC/SECEX	0.964	0.955	1.100	0.955	1.100	0.976
Exports - (FOB) Foreign trade 5 MDIC/SECEX	1.005	0.969	1.157	0.977	1.157	1.076
Savings - deposit - SBPE and rural - balance Financial 6 Bacen Outras/SGS	0.959	0.970	0.920	0.970	0.920	1.021
Tax on imports (II) - total - gross revenue Public Finances 5 Min. Economia	0.953	0.924	1.003	0.924	1.003	0.914
Tax on Financial Transactions (IOF) - total - gross revenue Public Finances 5 Min. Economia	1.065	1.059	1.073	1.059	1.073	1.062
Tax on Industrialized Products (IPI) - total - gross revenue Public Finances 5 Min. Economia	0.984	0.908	1.000	0.988	1.000	0.895
Income tax (IR) - withholding - capital income - gross revenue Public Finances 5 Min. Economia	1.217	1.075	0.948	1.075	0.948	1.214
Income tax (IR) - withholding - remittance income abroad	1.272	1.130	0.957	1.130	0.957	1.069
Income tax (IR) - withheld - labor income - gross revenue Public Finances 5 Min. Economia	1.874	1.323	1.057	1.323	1.057	1.519
DLSP - net internal debt - states and municipalities Public Finances 6 Bacen	1.227	1.720	1.000	1.732	1.000	1.925
Tax on rural land property (ITR) - gross revenue Public Finances 2 Min. Economi	6.130	1.122	1.063	1.044	1.042	7.119
DLSP - net internal debt - federal government and Central Bank Public Finances 6 Bacen	0.906	0.984	1.000	0.984	1.000	1.176
DSLP - total - states and municipalities Public Finances 6 Bacen	1.132	1.412	1.009	1.428	1.008	1.568
DSLP - total - federal government and Central Bank Public Finances 6 Bacen	0.915	0.986	1.031	0.986	1.031	1.622
Exchange rate - R/U S - commercial - purchase - average Exchange 5 Bacen	1.011	0.979	0.953	0.979	0.953	1.024
IPCA - free prices - commercial - rate of change Prices 1 Bacen	0.985	1.017	1.017	1.022	1.021	0.990
Savings - profitability - deposits from 04.05.2012 (1st day of the month) Financial 5 Anbima	0.908	0.872	0.921	0.872	0.921	0.871
IPCA - free prices - non-tradable - rate of change Prices 2 Bacen	1.004	1.221	0.995	1.221	0.995	1.531
IPCA - free prices - rate of change Prices 2 Bacen	0.968	1.052	0.969	1.027	0.969	1.162
IPCA - free prices - durable - rate of change Prices 2 Bacen	1.024	1.103	0.998	1.103	0.998	1.133
IPCA - free prices - non-durable goods - rate of change Prices 1 Bacen	0.971	0.990	0.984	0.986	0.984	0.980
IPCA - free prices - semi-durable goods - rate of change Prices 2 Bacen	1.008	1.045	0.997	1.019	0.997	1.171
IPCA - free prices - services - rate of change Prices 2 Bacen	1.133	1.445	1.047	1.445	1.047	1.609
IPCA - monitored prices - rate of change Prices 2 Bacen	0.993	1.045	0.964	1.045	0.964	1.152
Restricted monetary base - M0 - average Currency and credit 5 Bacen	1.003	0.940	1.064	0.940	1.064	0.924
NFSP - states and municipalities - nominal Public Finances 2 Bacen	0.974	1.072	1.040	1.072	1.040	1.192
GDP National Accounts 6 Bacen	3.311	3.586	1.001	3.483	1.001	4.330
Apparent consumption - fuel alcohol - average - quantity/day Consumption and sales 5 ANP	1.033	1.045	1.023	1.045	1.023	1.055
International Reserves Balance of payments 6 Bacen	1.444	1.592	1.156	1.592	1.156	1.777
Gold - monthly percentage variation Financial 1 Bacen	0.991	0.991	0.996	0.991	0.996	0.991
Interest rate - TJLP Financial 5 Bacen	1.018	1.085	1.111	1.085	1.111	1.084
Interest rate - Over / Selic - accumulated in the month Financial 5 Bacen	0.903	0.880	0.900	0.880	0.900	0.882
Interest rate - TR - first day of the month Financial 2 Bacen	0.859	0.829	0.988	0.829	0.988	0.863
Balance of payments - Financial account - balance (borrowings - concessions) BP	0.979	1.028	1.029	1.026	1.029	1.021
Balance of payments - capital account - balance Balance of payments 2 Bacen	1.180	1.188	1.011	1.188	1.011	1.220
Balance of payments - current transactions - balance Balance of payments 2 Bacen	1.000	1.024	1.051	1.051	1.063	1.046
Installed capacity utilization - industry - average Production 5 FGV/Conj. Econ.	0.947	0.871	0.890	0.894	0.913	0.948

Table 9 continued from previous page

Series	BOLS	BSpline	TSBoost	BSpline*	TSBoost*	Tree
Industrial indicators - hours worked - industry - index (average 2006 = 100)	1.010	0.871	1.000	0.871	1.000	1.115
Apparent Consumption - oil products - average - quantity/day	1.009	1.093	1.070	0.936	1.070	1.119
Industrial indicators - capacity utilization - industry Production 5 CNI	0.937	0.826	1.000	0.826	1.000	0.815
Industrial Indicators - personnel employed - industry Employment 6 CNI	1.015	1.098	1.000	1.098	1.000	1.131
Tax on Distribution of Goods (ICMS) - Brazil Public Finances 5 Min. Economia	1.008	1.121	0.975	1.121	0.975	1.087
Net fixed capital stock (2010 prices) Capital stock 6 IPEA	0.966	0.960	0.984	0.939	0.963	0.945
Electric energy - consumption - trade - quantity Consumption and sales 5 Eletrobras	0.918	0.730	0.971	0.730	0.971	0.955
Electric energy - consumption - industry - quantity Consumption and sales 5 Eletrobras	0.983	0.802	0.997	0.802	0.997	0.866
Electric energy - consumption - residential - quantity Consumption and sales 5 Eletrobras	1.091	0.948	0.904	0.948	0.904	1.323
Electric energy - consumption - trade - average rate per MWh	0.968	0.981	0.978	0.935	0.935	0.971
Electric energy - consumption - industry - average tariff per MWh	0.973	0.919	0.984	0.945	0.984	0.875
Apparent Consumption - gasoline - average - quantity/day Consumption and sales 5 ANP	0.991	0.878	0.936	0.878	0.936	0.933
Consumer Confidence Index (ICC) Perception and expectation 5 Fecomercio SP	1.023	0.975	0.991	0.975	0.991	0.942
Current economic conditions index (ICEA) Consumption and sales 5 Fecomercio SP	0.937	0.923	0.916	0.923	0.916	0.918
Consumer expectations index (IEC) Perception and expectation 5 Fecomercio SP	1.029	1.058	1.030	1.058	1.030	1.022
MG industry - capacity utilization - mineral extraction - average Production 5 Fiemg	1.003	1.152	1.067	1.032	1.006	1.108
Real salary - average - industry - index (average 2006 = 100) - SP Salary and income 5 Fiesp	0.921	1.011	1.025	1.011	1.025	1.031
Real salary - industry - index (average 2006 = 100) - SP Salary and income 5 Fiesp	0.943	1.012	1.011	1.012	1.011	0.980
Real Sales - Industry - Index (average 2006 = 100) - SP Consumption and sales 5 Fiesp	0.947	0.934	1.073	0.874	1.074	0.933
CPI - general - index (Jun. 1994 = 100) - RMSP Prices 6 Fipe	0.992	1.005	0.961	1.005	0.961	1.100
Employment - industry - RJ - index (average 2006 = 100) Employment 6 Firjan	1.102	1.152	1.002	1.152	1.002	1.306
Real income - industry - index - RJ - (average 2006 = 100) Consumption and sales 5 Firjan	0.993	0.974	0.990	0.964	0.976	1.062
Apparent Consumption - LPG - average - quantity/day Consumption and sales 5 ANP	1.025	1.125	1.001	1.012	1.000	1.493
Wage bill - industry - index - RJ - (average 2006 = 100) Employment 5 Firjan	1.028	1.016	1.072	1.016	1.072	1.006
Imports - prices - index - (average 2018 = 100) Foreign trade 6 Funcex	1.165	1.244	1.009	1.244	1.009	1.386
Imports - consumer durables - (FOB) (new series) Foreign trade 5 Funcex	1.004	1.030	1.072	1.030	1.072	1.087
Imports - non-durable consumer goods - (FOB) (new series) Foreign trade 5 Funcex	1.007	1.076	1.000	1.076	1.000	1.108
Imports - intermediate goods - (FOB) (new series) Foreign trade 5 Funcex	0.914	0.874	0.996	0.874	0.996	0.892
Imports - capital goods - (FOB) (new series) Foreign trade 5 Funcex	1.031	1.120	1.103	1.106	1.097	1.189
Exports - prices - index (average 2018 = 100) Foreign trade 6 Funcex	1.018	1.070	0.995	1.070	0.995	1.199
Exports - consumer durables - (FOB) (new series) Foreign trade 5 Funcex	0.956	0.900	1.000	0.900	1.000	0.892
Exports - non-durable consumer goods - (FOB) (new series) Foreign trade 5 Funcex	1.014	1.008	1.075	0.998	1.075	1.169
Exports - intermediate goods - (FOB) (new series) Foreign trade 5 Funcex	1.008	1.021	1.164	1.047	1.164	1.115
Apparent Consumption - oil products - other - average - quantity/day Consumption and sales 5 ANP	0.995	1.001	0.999	1.001	0.999	0.999
Exports - capital goods - (FOB) (new series) Foreign trade 5 Funcex	1.001	0.960	1.001	0.960	1.001	1.083
Apparent Consumption - consumer goods - index (2012 average = 100) National Accounts 5 IPEA	1.021	1.125	1.175	1.077	1.175	1.311
Apparent Consumption - Consumer Durables - Index (2012 average = 100) National Accounts 5 IPEA	0.923	0.915	1.006	0.915	1.006	1.006
Apparent consumption - consumer durables National Accounts 5 IPEA	0.996	0.995	0.996	0.995	0.996	0.996
Apparent consumption - consumer goods National Accounts 5 IPEA	1.012	1.000	1.048	1.028	1.048	1.032
Apparent consumption - semi- and non-durable consumer goods National Accounts 5 IPEA	1.002	1.167	1.095	1.131	1.096	1.701
Apparent consumption - semi- and non-durable consumer goods National Accounts 5 IPEA	1.031	1.019	1.005	0.988	0.994	1.059
Apparent consumption - intermediate goods National Accounts 5 IPEA	0.940	0.929	1.017	0.929	1.017	1.251
Apparent consumption - intermediate goods National Accounts 5 IPEA	0.926	0.834	0.887	0.834	0.887	0.878
Apparent Consumption - Capital Goods - Index (2012 average = 100) National Accounts 5 IPEA	0.958	1.002	1.019	0.991	1.020	1.026
Apparent Consumption - fuel oil - average - quantity/day Consumption and sales 5 ANP	1.041	1.095	0.999	0.956	0.999	1.075
Apparent consumption - capital goods - seasonally National Accounts 5 IPEA	1.014	0.938	0.923	0.968	0.964	1.028
IPEA GFCF indicator - index (average 1995 = 100) National Accounts 5 IPEA	0.930	0.985	1.034	0.944	1.034	0.969
IPEA GFCF indicator - civil construction - index (average 1995 = 100) National Accounts 5 IPEA	0.948	0.965	1.090	1.080	1.090	1.034
IPEA GFCF indicator - civil construction National Accounts 5 IPEA	0.890	0.899	1.050	0.899	1.050	0.847
IPEA GFCF indicator - seasonally adjusted National Accounts 5 IPEA	0.886	0.926	0.931	0.914	0.928	0.881
Minimum wage - purchasing power parity (PPP) Salary and income 5 IPEA	1.004	1.043	1.172	1.043	1.172	1.049
Real minimum wage Salary and income 5 IPEA	1.007	1.032	1.178	1.032	1.178	0.978
Exchange rate - real effective - IPA-EP-DI - imports - durable goods Exchange 5 IPEA	0.993	0.936	0.961	0.936	0.961	0.930
Exchange rate - real effective - IPA-EP-DI - imports - non-durable goods Exchange 5 IPEA	0.996	0.930	0.884	0.930	0.884	0.924
Exchange rate - real effective - IPA-EP-DI - imports - intermediate goods Exchange 5 IPEA	1.000	1.011	0.956	1.011	0.956	1.005
Apparent Consumption - diesel oil - average - quantity/day Consumption and sales 5 ANP	1.027	1.365	1.104	0.993	1.104	1.183
Exchange rate - real effective - IPA-EP-DI - imports - capital goods Exchange 5 IPEA	1.008	0.944	0.951	0.944	0.951	0.998
Exchange rate - real effective - IPA-EP-DI - imports - fuel Exchange 5 IPEA	0.933	0.955	0.965	0.955	0.965	0.954
Exchange rate - real effective - IPA-DI - imports - index (average 2010 = 100) Exchange 5 IPEA	0.996	0.937	0.953	0.937	0.953	0.978
Exchange rate - real effective - INPC - imports - index (average 2010 = 100) Exchange 5 IPEA	1.014	0.925	0.890	0.925	0.890	0.972
Exchange rate - real effective - Weighted IPA - exports - basic Exchange 5 IPEA	1.001	1.000	1.003	1.000	1.003	0.988
Exchange rate - real effective - IPA-DI-Origin - exports - manufactured Exchange 5 IPEA	1.000	0.901	0.918	0.901	0.918	0.960
Exchange rate - real effective - INPC - exports - manufactured Exchange 5 IPEA	1.018	0.881	0.888	0.881	0.888	0.938
Exchange rate - real effective - IPA-DI - Origin - exports - semimanufactured Exchange 5 IPEA	1.001	1.014	0.930	1.014	0.930	1.001
Exchange rate - real effective - IPA-DI - exports - index (average 2010 = 100) Exchange 5 IPEA	0.999	0.983	1.007	0.983	1.007	1.006
Exchange rate - real effective - INPC - exports - index (average 2010 = 100) Exchange 5 IPEA	1.012	0.922	1.002	0.922	1.002	0.974

Table 10 – All Cases - rRMSE Relative to Linear Model - $h = 12$

Series	BOLS	BSpline	TSBoost	BSpline*	TSBoost*	Tree
Stock index - Ibovespa - closing Financial 1 Anbima	0.988	0.991	0.983	0.991	0.983	0.987
Production - petroleum - average - quantity/day Production 5 ANP	1.001	0.977	0.981	0.977	0.981	0.996
IGP-10 - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.002	1.005	1.000	1.003	1.000	1.023
IGP-M - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.000	1.003	1.001	0.997	0.999	1.017
IGP-OG - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.000	1.008	1.000	1.004	1.000	0.979
INCC-DI - general - index (Aug. 1994 = 100) Prices 6 FGV/Conj. Econ.	1.520	1.668	0.946	1.668	0.946	1.954
IPA-10 - general - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.004	0.996	0.998	0.988	0.993	0.986
IPA-DI - Origin - Agricultural Products - index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	1.015	1.015	1.016	1.015	1.016	1.009
IPA-DI - Origin - Industrial Products - Index (aug. 1994 = 100) Prices 5 FGV/Conj. Econ.	0.982	0.930	0.970	0.943	0.985	1.062
INPC - general - index (dec. 1993 = 100) Prices 6 IBGE/SNIPC	0.952	1.042	0.963	1.003	0.963	1.202
INPC - food & beverages - rate of change Prices 1 IBGE/SNIPC	1.000	0.997	0.990	0.997	0.990	0.996
INPC - household goods - rate of change Prices 1 IBGE/SNIPC	0.998	1.018	1.051	1.018	1.051	1.002
Restricted means of payment - M1 - demand deposits - average Currency and credit 5 Bacen	1.002	1.003	1.002	1.003	1.002	1.004
INPC - personal expenses - rate of change Prices 2 IBGE/SNIPC	0.944	0.882	0.945	0.882	0.945	0.863
INPC - communication - rate of change Prices 2 IBGE/SNIPC	1.013	0.730	0.986	0.730	0.986	0.999
INPC - education, reading and stationery - rate of change Prices 2 IBGE/SNIPC	0.970	0.973	0.948	0.997	0.990	0.981
INPC - housing - rate of change Prices 1 IBGE/SNIPC	0.975	1.006	0.997	1.007	0.998	0.974
INPC - health and personal care - rate of change Prices 2 IBGE/SNIPC	0.994	1.057	1.017	0.964	0.960	1.002
INPC - transport - rate of change Prices 2 IBGE/SNIPC	0.994	0.953	0.925	0.950	0.930	0.976
INPC - clothing - rate of change Prices 2 IBGE/SNIPC	1.042	0.983	1.022	1.001	1.025	1.014
IPCA - food and beverages - rate of change Prices 1 IBGE/SNIPC	1.000	0.990	0.990	0.990	0.990	1.002
IPCA - household goods - rate of change Prices 1 IBGE/SNIPC	1.003	1.082	1.013	1.082	1.013	1.021
IPCA - personal expenses - rate of change Prices 2 IBGE/SNIPC	0.949	0.954	0.982	0.954	0.982	1.004
DLSP - net external debt - states and municipalities Public Finances 6 Bacen	1.054	1.087	0.961	1.087	0.961	1.342
IPCA - communication - rate of change Prices 2 IBGE/SNIPC	1.000	0.697	1.000	0.697	1.000	0.925
IPCA - education, reading and stationery - rate of change Prices 2 IBGE/SNIPC	1.005	1.035	1.000	1.035	1.000	0.975
IPCA - housing - rate of change Prices 1 IBGE/SNIPC	0.990	0.998	1.003	1.000	1.003	0.985
IPCA - health and personal care - rate of change Prices 2 IBGE/SNIPC	0.972	0.951	0.992	0.951	0.992	1.005
IPCA - transport - rate of change Prices 2 IBGE/SNIPC	1.030	1.129	1.062	1.129	1.062	1.183
IPCA - clothing - rate of change Prices 2 IBGE/SNIPC	1.053	1.021	1.027	1.031	1.029	1.024
Average income - real - wage earners - main work Salary and income 5 Seade/PED	0.983	0.940	0.932	0.940	0.932	0.920
Unemployment rate - open - RMSP Employment 5 Seade/PED	1.000	1.000	1.000	1.000	1.000	1.000
Hidden - precarious unemployment rate - RMSP Employment 6 Seade/PED	1.633	1.796	1.058	1.725	1.002	2.286
Unemployment rate - hidden - RMSP Employment 6 Seade/PED	1.489	1.666	1.000	1.666	1.000	2.089
DLSP - net external debt - federal government and Central Bank Public Finances 2 Bacen	0.960	1.043	1.079	0.946	0.956	0.968
Unemployment Rate - RMSP Employment 5 Seade/PED	0.988	0.988	0.991	0.988	0.989	0.988
Imports - (FOB) Foreign trade 5 MDIC/SECEX	0.993	0.991	1.001	0.991	1.001	0.990
Exports - (FOB) Foreign trade 5 MDIC/SECEX	0.999	0.999	0.999	0.999	0.999	0.999
Savings - deposit - SBPE and rural - balance Financial 6 Bacen Outras/SGS	0.891	0.884	0.924	0.884	0.924	0.951
Tax on imports (II) - total - gross revenue Public Finances 5 Min. Economia	0.973	0.944	0.959	0.944	0.959	0.935
Tax on Financial Transactions (IOF) - total - gross revenue Public Finances 5 Min. Economia	0.993	1.425	1.424	1.425	1.424	0.970
Tax on Industrialized Products (IPI) - total - gross revenue Public Finances 5 Min. Economia	0.945	0.950	0.950	0.950	0.950	0.944
Income tax (IR) - withholding - capital income - gross revenue Public Finances 5 Min. Economia	0.973	0.967	0.988	0.967	0.988	0.990
Income tax (IR) - withholding - remittance income abroad	1.032	1.040	1.060	1.040	1.060	1.028
Income tax (IR) - withheld - labor income - gross revenue Public Finances 5 Min. Economia	0.995	0.990	0.985	0.990	0.985	0.978
DLSP - net internal debt - states and municipalities Public Finances 6 Bacen	1.304	1.931	1.000	1.880	1.000	2.432
Tax on rural land property (ITR) - gross revenue Public Finances 2 Min. Economi	0.993	1.136	0.986	0.923	1.054	1.001
DLSP - net internal debt - federal government and Central Bank Public Finances 6 Bacen	0.829	0.854	1.002	0.854	1.002	1.056
DSLP - total - states and municipalities Public Finances 6 Bacen	1.139	1.408	0.991	1.427	0.991	1.675
DSLP - total - federal government and Central Bank Public Finances 6 Bacen	1.024	1.070	1.010	1.070	1.010	1.639
Exchange rate - R/U S - commercial - purchase - average Exchange 5 Bacen	1.010	0.857	0.871	0.857	0.871	0.961
IPCA - free prices - commercial - rate of change Prices 1 Bacen	0.997	1.005	0.999	1.003	0.997	1.005
Savings - profitability - deposits from 04.05.2012 (1st day of the month) Financial 5 Anbima	0.951	0.958	0.963	0.958	0.963	0.948
IPCA - free prices - non-tradable - rate of change Prices 2 Bacen	1.023	1.051	1.009	1.051	1.009	1.024
IPCA - free prices - rate of change Prices 2 Bacen	1.010	1.029	1.042	1.034	1.042	1.039
IPCA - free prices - durable - rate of change Prices 2 Bacen	0.998	1.018	1.015	1.018	1.015	1.083
IPCA - free prices - non-durable goods - rate of change Prices 1 Bacen	0.999	0.985	0.985	0.991	0.991	0.996
IPCA - free prices - semi-durable goods - rate of change Prices 2 Bacen	0.996	1.022	1.044	1.026	1.046	1.028
IPCA - free prices - services - rate of change Prices 2 Bacen	1.026	1.060	0.997	1.060	0.997	1.083
IPCA - monitored prices - rate of change Prices 2 Bacen	1.036	1.067	0.997	1.067	0.997	1.099
Restricted monetary base - M0 - average Currency and credit 5 Bacen	1.000	0.995	0.999	0.995	0.999	1.000
NFSP - states and municipalities - nominal Public Finances 2 Bacen	0.983	1.150	1.147	1.150	1.147	1.059
GDP National Accounts 6 Bacen	3.293	3.597	0.999	3.454	0.999	5.117
Apparent consumption - fuel alcohol - average - quantity/day Consumption and sales 5 ANP	1.000	1.000	1.000	1.000	1.000	0.998
International Reserves Balance of payments 6 Bacen	1.563	1.989	1.884	1.989	1.884	2.377
Gold - monthly percentage variation Financial 1 Bacen	1.000	1.001	1.002	1.001	1.002	1.004
Interest rate - TJLP Financial 5 Bacen	1.004	0.978	0.973	0.978	0.973	0.988
Interest rate - Over / Selic - accumulated in the month Financial 5 Bacen	0.917	0.864	0.864	0.864	0.864	0.864
Interest rate - TR - first day of the month Financial 2 Bacen	0.951	0.872	0.928	0.872	0.928	0.858
Balance of payments - Financial account - balance (borrowings - concessions) BP	0.952	1.019	0.978	0.997	0.972	1.013
Balance of payments - capital account - balance Balance of payments 2 Bacen	1.031	1.092	1.037	1.092	1.037	1.133
Balance of payments - current transactions - balance Balance of payments 2 Bacen	0.983	0.975	0.980	0.975	0.980	0.984
Installed capacity utilization - industry - average Production 5 FGV/Conj. Econ.	0.965	0.758	0.797	0.746	0.785	0.791

Table 10 continued from previous page

Series	BOLS	BSpline	TSBoost	BSpline*	TSBoost*	Tree
Industrial indicators - hours worked - industry - index (average 2006 = 100)	1.000	1.000	1.000	1.000	1.000	1.000
Apparent Consumption - oil products - average - quantity/day	1.007	1.003	1.015	1.018	1.014	1.014
Industrial indicators - capacity utilization - industry Production 5 CNI	0.808	0.774	0.774	0.774	0.774	0.774
Industrial Indicators - personnel employed - industry Employment 6 CNI	0.923	0.991	0.988	0.991	0.988	1.047
Tax on Distribution of Goods (ICMS) - Brazil Public Finances 5 Min. Economia	0.850	0.737	0.827	0.737	0.827	0.775
Net fixed capital stock (2010 prices) Capital stock 6 IPEA	0.905	0.936	0.938	0.911	0.913	0.905
Electric energy - consumption - trade - quantity Consumption and sales 5 Eletrobras	0.830	0.779	0.793	0.779	0.793	0.795
Electric energy - consumption - industry - quantity Consumption and sales 5 Eletrobras	0.953	0.826	0.832	0.826	0.832	0.887
Electric energy - consumption - residential - quantity Consumption and sales 5 Eletrobras	0.990	1.111	1.122	1.111	1.122	1.019
Electric energy - consumption - trade - average rate per MWh	0.990	1.039	1.037	0.986	0.989	1.058
Electric energy - consumption - industry - average tariff per MWh	1.004	1.003	1.026	0.995	1.026	1.004
Apparent Consumption - gasoline - average - quantity/day Consumption and sales 5 ANP	0.996	0.995	0.993	0.995	0.993	0.994
Consumer Confidence Index (ICC) Perception and expectation 5 Fecomercio SP	0.973	0.972	0.980	0.972	0.980	0.972
Current economic conditions index (ICEA) Consumption and sales 5 Fecomercio SP	0.974	0.970	0.964	0.970	0.964	0.975
Consumer expectations index (IEC) Perception and expectation 5 Fecomercio SP	0.966	0.969	0.978	0.969	0.978	0.967
MG industry - capacity utilization - mineral extraction - average Production 5 Fiemg	0.995	1.028	1.027	1.028	1.027	1.014
Real salary - average - industry - index (average 2006 = 100) - SP Salary and income 5 Fiesp	0.817	0.819	0.816	0.819	0.816	0.816
Real salary - industry - index (average 2006 = 100) - SP Salary and income 5 Fiesp	1.000	1.000	1.000	1.000	1.000	1.000
Real Sales - Industry - Index (average 2006 = 100) - SP Consumption and sales 5 Fiesp	0.917	0.920	0.917	0.920	0.917	0.917
CPI - general - index (Jun. 1994 = 100) - RMSP Prices 6 Fipe	0.910	0.987	0.975	0.987	0.975	0.944
Employment - industry - RJ - index (average 2006 = 100) Employment 6 Firjan	1.103	1.170	0.989	1.170	0.989	1.377
Real income - industry - index - RJ - (average 2006 = 100) Consumption and sales 5 Firjan	1.010	0.979	0.966	0.988	0.998	0.980
Apparent Consumption - LPG - average - quantity/day Consumption and sales 5 ANP	0.986	0.935	0.975	0.985	0.976	1.017
Wage bill - industry - index - RJ - (average 2006 = 100) Employment 5 Firjan	1.012	1.029	1.047	1.029	1.047	1.032
Imports - prices - index - (average 2018 = 100) Foreign trade 6 Funcex	1.132	1.167	1.000	1.167	1.000	1.330
Imports - consumer durables - (FOB) (new series) Foreign trade 5 Funcex	0.995	0.995	0.995	0.995	0.995	0.996
Imports - non-durable consumer goods - (FOB) (new series) Foreign trade 5 Funcex	1.005	1.006	1.006	1.006	1.006	1.005
Imports - intermediate goods - (FOB) (new series) Foreign trade 5 Funcex	0.995	0.992	0.992	0.992	0.992	0.996
Imports - capital goods - (FOB) (new series) Foreign trade 5 Funcex	0.988	0.984	0.990	0.978	0.990	0.997
Exports - prices - index (average 2018 = 100) Foreign trade 6 Funcex	1.037	1.065	0.997	1.065	0.997	1.180
Exports - consumer durables - (FOB) (new series) Foreign trade 5 Funcex	0.899	0.912	0.913	0.912	0.913	0.914
Exports - non-durable consumer goods - (FOB) (new series) Foreign trade 5 Funcex	1.019	1.021	1.012	1.021	1.012	1.017
Exports - intermediate goods - (FOB) (new series) Foreign trade 5 Funcex	1.002	1.008	1.008	1.008	1.008	1.004
Apparent Consumption - oil products - other - average - quantity/day Consumption and sales 5 ANP	0.983	0.983	0.988	0.983	0.988	0.983
Exports - capital goods - (FOB) (new series) Foreign trade 5 Funcex	1.004	1.009	0.994	1.009	0.994	1.016
Apparent Consumption - consumer goods - index (2012 average = 100) National Accounts 5 IPEA	1.009	1.009	1.009	1.009	1.009	1.008
Apparent Consumption - Consumer Durables - Index (2012 average = 100) National Accounts 5 IPEA	0.999	0.999	0.999	0.999	0.999	1.000
Apparent consumption - consumer durables National Accounts 5 IPEA	1.000	1.000	1.000	1.000	1.000	1.000
Apparent consumption - consumer goods National Accounts 5 IPEA	0.988	0.988	0.988	0.988	0.988	0.987
Apparent consumption - semi- and non-durable consumer goods National Accounts 5 IPEA	1.013	1.020	1.021	1.020	1.021	1.018
Apparent consumption - semi- and non-durable consumer goods National Accounts 5 IPEA	1.000	1.017	0.983	0.985	0.983	1.044
Apparent consumption - intermediate goods National Accounts 5 IPEA	0.778	0.817	0.816	0.817	0.816	0.778
Apparent consumption - intermediate goods National Accounts 5 IPEA	0.959	0.849	0.774	0.849	0.774	0.915
Apparent Consumption - Capital Goods - Index (2012 average = 100) National Accounts 5 IPEA	1.003	1.004	1.006	1.006	1.007	1.004
Apparent Consumption - fuel oil - average - quantity/day Consumption and sales 5 ANP	0.975	0.998	0.980	0.998	0.980	0.976
Apparent consumption - capital goods - seasonally National Accounts 5 IPEA	1.004	0.989	1.027	0.963	1.005	0.997
IPEA GFCF indicator - index (average 1995 = 100) National Accounts 5 IPEA	0.993	0.993	0.993	0.993	0.993	1.003
IPEA GFCF indicator - civil construction - index (average 1995 = 100) National Accounts 5 IPEA	1.000	1.029	1.035	1.010	1.016	1.000
IPEA GFCF indicator - civil construction National Accounts 5 IPEA	1.000	1.000	1.000	1.000	1.000	0.997
IPEA GFCF indicator - seasonally adjusted National Accounts 5 IPEA	0.934	0.925	0.924	0.925	0.924	0.923
Minimum wage - purchasing power parity (PPP) Salary and income 5 IPEA	1.000	1.000	1.000	1.000	1.000	1.000
Real minimum wage Salary and income 5 IPEA	1.000	1.004	1.000	1.004	1.000	0.999
Exchange rate - real effective - IPA-EP-DI - imports - durable goods Exchange 5 IPEA	0.993	0.903	0.898	0.903	0.898	0.890
Exchange rate - real effective - IPA-EP-DI - imports - non-durable goods Exchange 5 IPEA	1.013	0.918	0.909	0.918	0.909	0.944
Exchange rate - real effective - IPA-EP-DI - imports - intermediate goods Exchange 5 IPEA	1.000	0.884	0.876	0.884	0.876	0.922
Apparent Consumption - diesel oil - average - quantity/day Consumption and sales 5 ANP	0.946	0.949	0.930	0.934	0.928	0.956
Exchange rate - real effective - IPA-EP-DI - imports - capital goods Exchange 5 IPEA	0.995	0.933	0.933	0.933	0.933	0.979
Exchange rate - real effective - IPA-EP-DI - imports - fuel Exchange 5 IPEA	0.983	0.941	0.940	0.941	0.940	0.936
Exchange rate - real effective - IPA-DI - imports - index (average 2010 = 100) Exchange 5 IPEA	1.003	0.918	0.911	0.918	0.911	0.918
Exchange rate - real effective - INPC - imports - index (average 2010 = 100) Exchange 5 IPEA	0.994	0.908	0.867	0.908	0.867	0.902
Exchange rate - real effective - Weighted IPA - exports - basic Exchange 5 IPEA	0.974	0.985	0.956	0.985	0.956	0.941
Exchange rate - real effective - IPA-DI-Origin - exports - manufactured Exchange 5 IPEA	0.995	0.820	0.839	0.820	0.839	0.907
Exchange rate - real effective - INPC - exports - manufactured Exchange 5 IPEA	0.999	0.841	0.841	0.841	0.841	0.933
Exchange rate - real effective - IPA-DI - Origin - exports - semimanufactured Exchange 5 IPEA	1.000	0.872	0.876	0.872	0.876	0.927
Exchange rate - real effective - IPA-DI - exports - index (average 2010 = 100) Exchange 5 IPEA	0.997	0.951	0.937	0.951	0.937	0.912
Exchange rate - real effective - INPC - exports - index (average 2010 = 100) Exchange 5 IPEA	0.995	0.865	0.873	0.865	0.873	0.920