BRICS EXCHANGE RATE FORECASTING: A STATISTICAL LEARNING APPROACH*

PROJEÇÃO DAS TAXAS DE CÂMBIO DOS BRICS: UMA ABORDAGEM DE APRENDIZADO ESTATÍSTICO

Bernardo Cainelli Gomes da Silva†
Hudson da Silva Torrent‡

ABSTRACT

This paper uses novel tools from the statistical learning literature to the exercise of forecasting the exchange rate of the five currencies of the BRICS countries. Combining these tools with Taylor Rule, purchase power parity, uncover interest rate parity, and monetary models for short and long horizons, the results confirm the improved efficiency over the classical linear regression method. However, the performance varied between methods for different currencies. Also, rolling and expanding window schemes were employed, with no winner[1].

RESUMO

Este artigo utiliza novas ferramentas da literatura de aprendizagem estatística para o exercício de previsão da taxa de câmbio das cinco moedas dos países BRICS. Combinando essas ferramentas com a Regra de Taylor, a paridade do poder de compra, a paridade descoberta da taxa de juros e modelos monetários para horizontes curto e longo, os resultados confirmam melhor eficiência em relação ao método clássico de regressão linear. No entanto, obteve-se desempenho díspar entre diferentes métodos para diferentes moedas. Além disso, foram utilizados esquemas de janela rolante e de janela expansiva, sem qualquer vencedor.

Keywords: Exchange Rate. BRICS. Statistical Learning. Forecasting.
J.E.L. Classifications: C53. F31. F37.

---

*Final paper submitted, in the first semester of 2023, to the Department of Economics and International Relations of the Faculty of Economic Sciences of the Federal University of Rio Grande do Sul (UFRGS) as a partial requirement for obtaining the Bachelor's Degree in Economics.

†Economics undergraduate in Federal University of Rio Grande do Sul. (bernardocgdasilva@gmail.com).
‡Supervisor. PhD in economics from Federal University of Rio Grande do Sul and Oregon State University. Master's degree in economics from Federal University of Rio Grande do Sul. Professor in the Department of Economics and International Relations at UFRGS. (hudson.torrent@ufrgs.br).

[1] The code of this paper can be accessed at github.com/BernardoCGdaSilva/forecasts-thesis.

1 INTRODUCTION

Forecasting exchange rate under the free float regime is consensually difficult since the work of Meese and Rogoff (1983), who first proposed that "the random walk model performed no worse" than other models based on economic fundamentals when the criteria is out-of-sample fit. Their conclusion came to be known as the "Meese-Rogoff puzzle", which the later corpus of work tried to solve with mixed results. The aim of this article is to contribute to this literature on predictive exchange rate models by applying novel methods from the statistical learning field to forecast the relative value between the US Dollar (USD) and BRICS's currencies: Brazilian Real (BRL), Russian Ruble (RUB), Indian Rupee (INR), Chinese Renminbi (CNY) and the South African Rand (ZAR).

The first strategy followed by the literature was to find new macroeconomic variables to model the exchange rate. Cheung, Chinn and Pascual (2005) and later Cheung, Chinn, Pascual and Zhang (2019) updated the investigation by conducting two rounds of tests of various popular macroeconomic models of the 1990s and the 2000s and found no strong evidence that any model could consistently beat the random walk. But some models were noted to perform better on longer horizons, rather than shorter ones. Molodtsova and Papell (2009) branch the investigation towards a new set of models based on the Taylor Rule, with good results even for short horizons. The results, however, were not confirmed in the subsequent works of Wang and Wu (2012) and Gaglianone and Marins (2017).

The bulk of these studies estimated their models using the classical ordinary least squares linear regressor. There is, however, "evidence that the adjustment of exchange rates to equilibrium is non-linear [...], which might diminish the out-of-sample performance of macroeconomic models" (PETROPOULOS et al., 2022). A new strategy has emerged in recent years of applying methods from the statistical learning literature to estimate the exchange rate models.

Statistical learning is a field that merges statistics and computation in order to "learn from data" (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). It studies and develops a diverse set of methods to estimate functions based on data and use these functions to predict some outcome. And it is acquiring popularity due to the high flexibility and improved accuracy that some of its methods possess. Other advantage is that most methods are ready off-the-shelf to be used in a variety of settings, requiring only the adjustment of some parameters[2].

Following the work of Cheung, Chinn, Pascual and Zhang (2019), Colombo and Pelegatti (2020) estimate similar models using non-linear splines, random forests and support vector machines with good results on the one year ahead horizon. A similar investigation was conducted by Ribeiro (2017), where models were estimated using the penalized approaches of the lasso, elastic net, bumping and bagging.

This paper is heavily inspired by the work of Colombo and Pelegatti (2020) who conducted a similar research employing most of the same statistical learning methods on the same horizons of forecast. The difference is in the choice of countries: the authors above focus on developed nations and here the focus is on the developing BRICS. Also, this paper employs more models, methods and estimation window schemes. Another similar work was conducted by Salisu, Gupta and Kim (2022), where the exchange rate of the BRICS currencies was forecasted using 9 models. However, no statistical learning methods were employed.

Section 2 discuss the choice of models. That is, of economic fundamentals as

[2]Hyperparameter tuning is discussed in Section 5.

Table 1 – Models

z	Alias	$X_{z,t}$
1	Taylor	$X_{1,t} = [\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*} \quad q_t]$
2	Taylor PPP	$X_{2,t} = [\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*}]$
3	Taylor PPP Smoothing	$X_{3,t} = [\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*} \quad i_{t-1} - i_{t-1}^*]$
4	Taylor Smoothing	$X_{4,t} = [\pi_t - \pi_t^* \quad y_t^{gap} - y_t^{gap*} \quad q_t \quad i_{t-1} - i_{t-1}^*]$
5	PPP	$X_{5,t} = [q_t]$
6	Foward Premium (UIRP)	$X_{6,t} = [i_t - i_t^*]$
7	Monetary Flexible Prices	$X_{7,t} = [m_t - m_t^* \quad y_t - y_t^*]$
8	Monetary Sticky Prices	$X_{8,t} = [m_t - m_t^* \quad y_t - y_t^* \quad i_t - i_t^* \quad \pi_t - \pi_t^*]$
9	Monetary Sticky Prices and Risk	$X_{9,t} = [m_t - m_t^* \quad y_t - y_t^* \quad i_t - i_t^* \quad \pi_t - \pi_t^* \quad VIX]$

This table shows the $z = \{1, ..., 9\}$ models estimated in this study where π is the inflation rate, y_t^{gap} is the output gap, i is the interest rate, m is the log of the money supply, y is the log of the output, VIX is the 30-day expected volatility of the S&P500 and q is the real exchange rate $q_t = s_t + p_t^* - p_t$, where p is the log of the consumer price index. The superscript * denotes that the variable is from the foreign country. Source: created by the author.

variables to be applied in the training and evaluation of the forecasts. In total, this paper trains nine models: a Purchase Power Parity (PPP), an Uncovered Interest Rate Parity (UIRP), three variations of the monetary model and four variations of the Taylor Rule model. Section 3 discuss the estimation methods employed to train the models. Besides the classical linear regression, two tree based methods are used — a decision tree and a random forest — two Support Vector Machines (SVM) — with linear and with radial kernel — and a regularized regression spline. Section 4 informs the data sources and pre-processing. Section 5 discuss the estimation process and evaluation: the performance measure Root Mean Squared Error (RMSE), the significance test Diebold-Mariano and the benchmark model. Section 6 discuss the results and section 7 concludes.

2 MODELS: ECONOMIC FUNDAMENTALS

This paper estimates $z = \{1, ..., 9\}$ models with the general form:

$$s_{t+h} - s_t = f(X_{z,t}) + \epsilon_{z,t+h}, \quad t = 1, 2, ..., T \tag{1}$$

where $s_{t+h} - s_t$ is the h-period ahead difference of the log of the nominal exchange rate[3], $X_{z,t}$ is a matrix of economic fundamentals, $\epsilon_{z,t+h}$ is the error term and T is the sample size. The $f()$ is an unknown function to be estimated using the methods discussed in the next section.

The assortment of models is expressed in Table 1, where π is the inflation rate, y_t^{gap} is the output gap, i is the interest rate, m is the log of the money supply, y is the log of the output, VIX is the 30-day expected volatility of the S&P500, measured by the Chicago Board Options Exchange and q is the real exchange rate $q_t = s_t + p_t^* - p_t$, where p is the log of the consumer price index. This study focuses on the exchange of the BRICS currency against the US Dollar, therefore the latter is the foreign (*) country, and the former is the home country.

Models 1 to 6 follow the work of Molodtsova and Papell (2009), Wang and Wu (2012) and Gaglianone and Marins (2017). Model 1 is the Taylor Rule model. According to

[3]For small variations, $log(u_2) - log(u_1) \equiv \frac{u_2 - u_1}{u_1}$ (GUJARATI, 2003, p. 142). Therefore, for interpretation's sake, $s_{t+h} - s_t$ can be read as the exchange rate's percentile variation between t and $t + h$.

Taylor (1993), central banks should adjust the short-term nominal interest rate following this relation

$$i_t^* = \pi_t + \phi(\pi_t - \pi^*) + \gamma y_t^{gap} + r^* \tag{2}$$

where i_t^* is the target short-term nominal interest rate, π_t is the inflation rate, π^* is the target level of inflation, y_t^{gap} is the output gap and r^* is the equilibrium level of the real interest rate. Combining π^* and r^* into a constant $\mu = r^* - \phi\pi^*$ and adding the real exchange rate $q_t = s_t + p_t^* - p_t$ (CLARIDA; GALÍ; GERTLER, 1998), equation 2 can be rewritten as

$$i_t^* = \mu + \lambda\pi_t + \gamma y_t^{gap} + \delta q \tag{3}$$

where $\lambda = 1 + \phi$. The difference between two countries results in Model 1. The contribution of the real exchange rate in the model is that central banks can set a target level of the exchange rate to make purchasing power parity (PPP) hold and adjust the nominal interest rate if the exchange rate deviates from the PPP level. If assumed that PPP always holds, real exchange rate can be removed from equation 3 which gives Model 2. Model 3 adds to equation 3 the first lag of the nominal interest rate assuming that it takes some time to adjust to target level. Model 4 is a mixture of Model 2 and 3, with the first lag of the interest rate, but without the real exchange rate.

Model 5 is the PPP model. Model 6 is the forward premium where, under the Uncover Interest Rate Parity, changes in the log of the exchange rate equal the difference in nominal interest rates.

Models 7 and 8 are monetary models and were taken from the work of Rossi (2013). Model 7 is the monetary model with flexible prices where PPP holds and model 8 is the monetary model with sticky prices where prices adjust slowly. Model 9 follow the work of Cheung, Chinn, Pascual and Zhang (2019) which is the Sticky Price Monetary Model Augmented by Risk[4].

3 METHODS: STATISTICAL LEARNING

The statistical learning methods are the way the models are estimated into functions from which forecasts can be made using the data (JAMES et al., 2013, p. 1). Six methods are employed in this study: the linear regression estimated with ordinary least squares, two support vector machines, one with linear kernel and one with radial kernel, multivariate adaptive regression splines, regression tree and random forest. This selection of methods respects the criteria of non-linearity, since the non-linear relationship between the exchange rate and other macroeconomics variables is one hypothesis for the Meese-Rogoff Puzzle. The linear ordinary least squares method is also chosen due to its ubiquitous presence in econometric research and serves as comparison to the other methods.

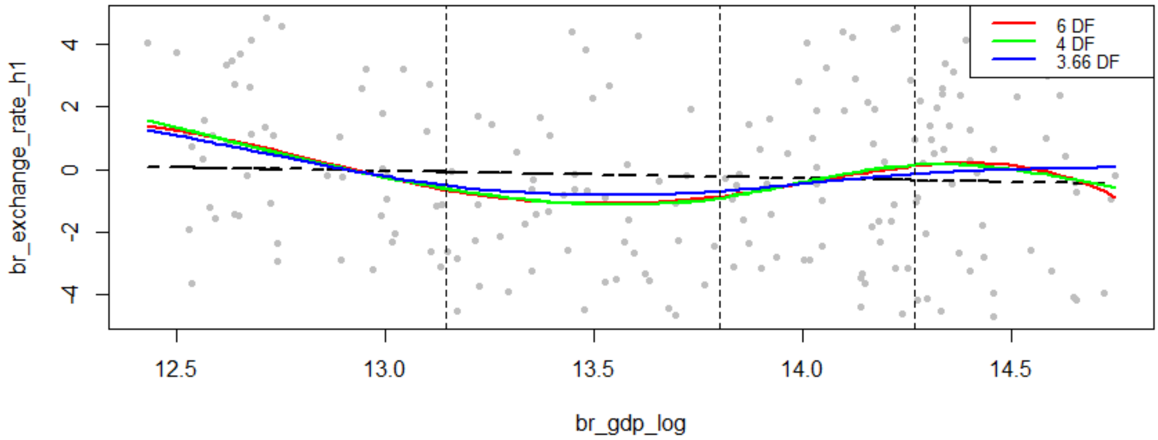
3.1 REGULARIZED REGRESSION SPLINES

The regression spline is a method that separates the regressor variables X into regions and then fits a different cubic function

$$f(X_j) = \beta_0 + \beta_1 X_j + \beta_2 X_j^2 + \beta_3 X_j^3 \tag{4}$$

[4]In the work of Cheung, Chinn, Pascual and Zhang (2019) this model is also augmented by liquidity with the TED Spread, but at the end of 2021 one of its components, the LIBOR, has been discontinued and a consensus on a replacement has not been achieved yet. One often-cited alternative is the SOFR, but as it is a short series, beginning in 2018, it has been decided to drop the liquidity variable altogether in this study.

Figure 1 – Splines



Spline example estimated with filtered Brazilian data: log GDP to predict exchange rate variation one month ahead. The black dashed line is a linear model fit to the data. The red line is a cubic spline with 6 degrees of freedom with knots at the dashed vertical lines (13.15, 13.80 and 14.27). Inside every region, a different cubic function is fit. Three restriction guarantee smooth continuity: the functions must be continuous at zero, first and second derivative. The green line is a natural spline where the functions at the boundaries are linear, reducing two degrees of freedom, and knots at the same positions as the red line. The blue line is a smooth spline where every observation is a knot, but the degrees of freedom are reduced via a regularizing parameter. Source: created by the author.

to each region, where j denotes the predictors. The point where regions meet is called knot (JAMES et al., 2013, p. 297). In Figure 1, the red line depicts an example of cubic spline with three knots (vertical dashed lines at 13.15, 13.80 and 14.27). In each region, a new cubic function is fit. To ensure continuity between regions, additional three constraints are imposed: the functions must be continuous to avoid breaks, and the first and the second derivatives must be continuous for smoothness of transition between regions. The supposed reason cubic splines are preferred is because the human eye has difficulty to detect the discontinuity of the knots.

Polynomial splines have the characteristic to exhibit high variance at the borders of X. To control this behaviour, two extra constraints can be applied: the functions at the boundary regions have to be linear. This extra constrained spline is called a natural spline. In Figure 1, the green line is a natural spline, therefore below 13.15 and above 14.27 it is a linear function. The more constraints added to a method, the fewer degrees of freedom it contains, which can be understood as ”a quantity that summarizes the flexibility of a curve” (JAMES et al., 2013, p. 31). A cubic spline with K knots has $4 + K$ degrees of freedom. A natural spline has two extra restrictions, therefore has $2 + K$ degrees of freedom.

The number and position of knots can be achieved by two means: it can be manually chosen or the degree of freedom can be chosen and the knots be placed at uniform quantiles of the data. This second approach has the benefit that the degrees of freedom can be chosen using cross-validation. This technique consists in holding out a random portion of the data, fitting one alternative function to the remaining data and measuring the residual squared error with the hold-out. The process is repeated until all alternatives are fit and the one with the smallest residual squared error is chosen.

An alternative method of fitting a natural spline is finding a function $f(X)$ that minimizes

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int f''(t)^2 dt \tag{5}$$

where $\sum_{i=1}^{n}(y_i - f(x_i))^2$ is the residual sum of squares, a measure of quality of fit, and the second derivative $f''(t)^2dt$ can be seen as the measure of curvature. A second derivative at zero is a straight line and as it increases, the more wiggly it becomes at t. The integral captures the measure of curvature for the entire function. And λ is a positive shrinking hyperparameter.

This approach is called a smooth spline and has the same characteristics of a natural spline: cubic in the middle and linear at the boundaries. It, however, can have n degrees of freedom, as every observation is treated as a knot. To appease this, the λ hyperparameter acts as a restrainer, reducing the degrees of freedom. The blue line in Figure 1 is a smooth spline with 3.66 degrees of freedom.

This study employs smooth splines as one of its methods via the framework of Generalized Additive Models

$$f(X) = \beta_0 + \sum_{j=1}^{p} f_j(X_j) + \epsilon \tag{6}$$

where for each p predictor, a different smooth spline f_j is estimated and added together. The λ hyperparamenter is chosen via hyperparameter tuning, discussed in Section 5.

3.2 REGRESSION TREE AND RANDOM FOREST

The regression tree method consists of (1) splitting the regressor space into J non-overlapping regions R and (2) fitting a simple model into each one (HASTIE; TIBSHIRANI; FRIEDMAN, 2009, p. 305). The most common model is the mean of the response values in R_J. To perform the splitting procedure, all predictors $X_1, ..., X_p$ and all cut-points s are evaluated and the pair predictor-cut-point that most reduces the residual sum of squares (RSS)

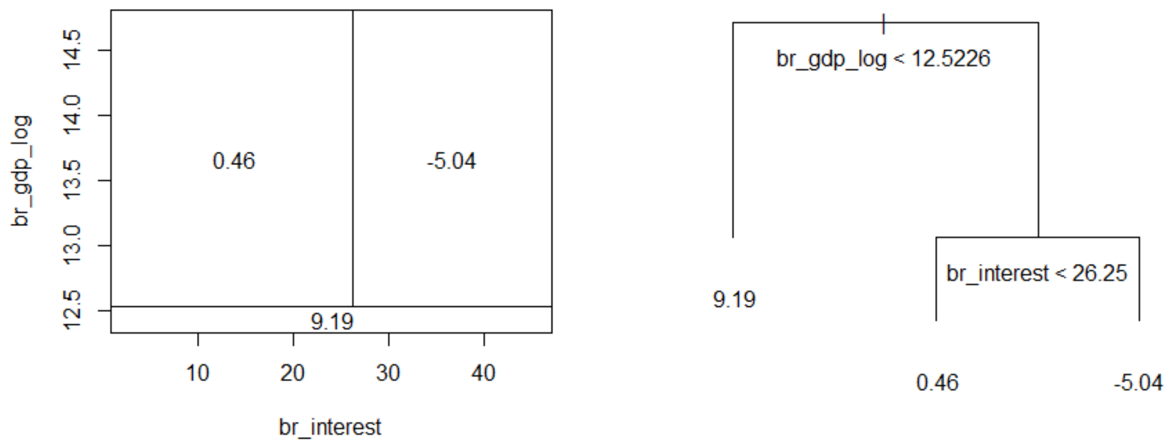
$$\sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{7}$$

is selected (JAMES et al., 2013, p. 331). R_1 and R_2 are the two split regions from the predictor space j and $\hat{y}_{R_1}$ and $\hat{y}_{R_2}$ are the mean response for the cutpoint s. Next, the process is repeated in each region until a stop criteria is reached.

Figure 2 depicts an example of the variation of the exchange rate one month ahead predicted by two variables of the Brazilian data. The first split is in $j = br\_gdp\_log$ at $s = 12.5226$. The predictor region $br\_gdp\_log > 12.5226$ was further split with $j = br\_interest$ at $s = 26.25$. In this example, every new observation with the predictors $br\_gdp\_log > 12.5226$ and $br\_interest > 26.25$ predict the variation one month ahead of the exchange rate to be -5.04, the mean response of the training data.

For this study, the stop criteria selected is the depth of any node, where a node any s, the point where two regions are split. In Figure 2, the node $br\_gdp\_log < 12.5226$ has depth zero and the node $br\_interest < 26.25$ has depth 1. The optimal depth is defined via hyperparameter tuning.

Figure 2 – Regression Tree



Regression tree example estimated with Brazilian data: log GDP and interest rate to predict exchange rate variation one month ahead with mean response values. Left: regressor space split into three regions. The mean exchange rate variation when log GDP is smaller than 12.5226 is predicted to be 9.19. When it is bigger, if the interest rate is smaller than 26.25, then 0.46, otherwise, -5.04. Right: tree visualization of the same partition. Source: created by the author.

Random forests are an improvement over regression trees, which are prone to overfitting on very large trees and are not very robust to data updates; new information can change drastically the fitted tree. The ensemble method of bagging has some capacity to reduce these weaknesses. To perform bagging, B random samples are taken from the data and used to train B trees. Then, the predictions are averaged. This procedure results in bagged trees. To achieve a random forest, at every branch of every tree of a bagged tree, only a random subset of the predictors is available for training. This decorrelates the trees before bagging, improving predictive performance. This can be thought as a solution to the case where one predictor is stronger than the rest. Then, most of the bagged trees would choose this predictor as first node and would be very similar. By restraining the predictor available in training, some trees will not have this predictor available for the first node, forcing a different choice.

There is also a hyperparameter in the random forest method, which is the number of predictors at each sample. The optimal number is also chosen via hyperparameter tuning.

3.3 SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a statistical learning method developed for the classification setting and later adapted to regression — it was first aimed at sorting observations between qualitative outputs and later adapted to project observations into a quantitative range (HASTIE; TIBSHIRANI; FRIEDMAN, 2009, p. 434). For the SVM classifier, the first step is to construct a hyperplane that separates different classes of observations by maximizing their distance to it. This approach linearly divides the data and has the characteristic that only the closest observations to the hyperplane are determinant in its positioning. These data points are the supporting vectors and the distance between them and the hyperplane without any observations is the maximum margin classifier. New training observations positioned inside the margin can restrict it and move the hyperplane. New observations far from the margin do not affect it.

However, not all data can be linearly divided by a hyperplane. In these cases, two solutions are added. The first is a cost hyperparameter that allows some observations to be inside the maximum margin classifier and at the wrong side of the hyperplane (JAMES et al., 2013, p. 375). The second seeks to account for a non-linear boundary between classes by projecting the feature space to a higher-dimensional space where the boundary is linear using a kernel function. Which can be thought as "a function that quantifies the similarity of two observations" (JAMES et al., 2013, p. 381). The SVM classifier can be written as

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i K(x_i, x_{i'}) \tag{8}$$

where n is the number of data points, α_i is a parameter that is non-zero only for supporting vectors and $K(x_i, x_i')$ is the kernel function.

The SVM regressor is built over the linear regression method, but incorporates the ideas from the SVM classifier. The process is to estimate the linear model

$$f(x) = \beta_0 + X^T\beta \tag{9}$$

by minimizing the cost function

$$H(\beta, \beta_0) = C\sum_{i=i}^{n} V(y_i - \hat{y}_i) + \frac{\lambda}{2}||\beta||^2 \tag{10}$$

where V is the ϵ-insensitive error measure

$$V_\epsilon(r) = \begin{cases} 0, & \text{if } |r| < \epsilon \\ |r| - \epsilon, & \text{otherwise} \end{cases} \tag{11}$$

and C is the cost hyperparameter.

Solving this minimization problem enables again the use of kernel functions and results in Equation 8 but for regression (HASTIE; TIBSHIRANI; FRIEDMAN, 2009, p. 436). This choice of error measure has the capacity to filter between observations. Predictor with errors below the threshold ϵ does not contribute to the regression fit. Only observations with bigger errors than ϵ influence the fit. This is similar, in concept, to selecting support vectors from the SVM classifier, but differs where in that case the closest observations to the boundary were chosen to be the supporting vectors while the outliers were disregarded. In the SVM regressor, the outliers are the top contributors (KUHN; JOHNSON, 2013, p. 153).

Two kernel functions are employed into the SVM regressor in this study: the linear kernel

$$K(x_i, x_{i'}) = \sum_{j=i}^{p} x_{ij}x_{i'j} \tag{12}$$

and the radial kernel

$$K(x_i, x_{i'}) = exp(-\gamma\sum_{j=i}^{p}(x_{ij} - x_{i'j})^2) \tag{13}$$

where γ is a positive constant defined via hyperparameter tuning, as well as C. As the objective of the kernel function is to project the data to a higher-dimension, the alignment of the kernel will tell how this projection is conducted. If the data is linearly related, then the linear kernel will capture this characteristic, if is non-linear, then the radial kernel will show better results at separating the observations.

Table 2 – Sample size

Country	T	Period
Brazil	275	12/1998 - 07/2022
Russia	182	12/2005 - 07/2021
India	179	03/2007 - 07/2022
China	240	12/2001 - 07/2022
South Africa	310	12/1995 - 07/2022

Source: created by the author.

4 DATA

The choice made was to investigate the nominal exchange rate of the BRICS currencies — Brazilian Real (BRL), Russian Ruble (RUB), Indian Rupee (INR), Chinese Renminbi (CNY) and the South African Rand (ZAR) — in relation to the US Dollar (USD), therefore all the US variables are treated as the foreign country variables. Output is the nominal GDP, money supply is M1 and interest is the overnight rate.

All series are monthly, but two. Nominal GDP is available on a quarterly basis and was disaggregated to the higher monthly frequency using the Chow-Lin Max Log algorithm with the industrial production index for each country as a proxy variable (SAX; STEINER, 2013). And VIX is daily, filtered on the last day of each month. Output gap was estimated on the monthly-adjusted GDP series using the Hamilton Filter (HAMILTON, 2017). Table 2 shows sample size and period for every country.

The source of the exchange rate and GDP series is the International Monetary Fund's (IMF) International Financial Statistics database; of the consumer price index is IMF's Consumer Price Index database. Money supply, industrial production index and interest were obtained from Organization for Economic Co-operation and Development (OECD), "Main Economic Indicators - complete database", retrieved from FRED, Federal Reserve Bank of St. Louis. VIX was also retrieved from FRED.

5 SET UP AND EVALUATION

This study is conducted on monthly time series and forecasts the exchange rate of each BRICS country one month ($h = 1$) and one year ($h = 12$) ahead through the pseudo-out-of-sample approach. The sample is split into two groups: the in-sample group and the out-of-sample group; and the models' functions are estimated using the in-sample data and evaluated using the out-of-sample data.

The size of the in-sample group is also called the window size and two window estimation schemes are employed in this study. The first is the rolling window in which window size is fixed at 120 observations (10 years). For every new observation, the last observation is dropped. The second scheme is the expanding window or recursive window. In it, the first window has the same size as the rolling window, 120 observations, but as the window advances, no observation is dropped and the window size increases. In both cases, the window start at the oldest observation (the beginning of the series) and moves forward in time. Every time the window change, the models are re-estimated in-sample and tested out-of-sample. The trade-off between the two schemes is that "Shorter estimation window sizes allow the parameter to adapt more quickly to structural changes; on the other hand, the parameter is less efficiently estimated in a smaller sample." (ROSSI, 2013).

Table 3 – Relative RMSE — Brazil

h	z	Expanding Window						Roling Window					
		LR	SPLINE	SVM_L	SVM_R	TREE	RF	LR	SPLINE	SVM_L	SVM_R	TREE	RF
1	1	1.00	1.09	1.02	0.99	1.00	1.06	1.02	1.05	1.06	1.02	1.05	1.10
	2	1.01	1.09	1.02	1.02	1.00	1.14	1.01	1.05	1.04	1.02	1.03	1.09
	3	1.02	1.11	1.02	1.01	1.00	1.08	1.03	1.07	1.06	1.03	1.03	1.07
	4	1.03	1.13	1.02	1.01	1.00	1.07	1.02	1.07	1.04	1.01	1.05	1.08
	5	1.00	1.01	1.02	1.05	1.02	1.28	1.01	1.01	1.03	1.07	1.11	1.18
	6	1.00	1.02	1.01	1.04	1.07	1.16	1.02	1.02	1.03	1.04	1.11	1.16
	7	1.04	1.03	1.05	1.02	1.08	1.13	1.02	1.10	1.04	1.02	1.13	1.13
	8	1.06	1.10	1.06	1.04	1.01	1.07	1.05	1.25	1.06	1.03	1.06	1.08
	9	1.05	1.08	1.05	1.04	1.01	1.08	1.04	1.21	1.05	1.02	1.07	1.08
12	1	0.86*	0.88	0.87*	0.92	1.02	0.89	0.92	0.90	0.95	0.80*	0.99	0.89
	2	1.00	1.00	1.05	1.05	1.00	1.05	1.07	1.08	1.11	1.04	1.09	1.07
	3	0.97	0.94	1.01	0.97	0.93	0.79*	1.08	0.92	1.10	0.94	0.92	0.80*
	4	0.84*	0.78*	0.84*	0.80	1.02	0.77*	0.83*	0.69*	0.89*	0.63*	0.92	0.73*
	5	0.88*	0.99	0.91	1.02	1.03	1.09	0.93**	0.98	0.95*	1.01	1.02	1.02
	6	0.92*	0.83*	0.93*	0.85	0.86*	0.88	1.02	0.93	1.05	0.96	0.96	0.92
	7	1.05	0.86	1.09	0.75	0.96	0.73	0.95	1.43	0.91	0.66	0.97	0.73
	8	1.07	0.86	1.12	0.77	0.87	0.65*	0.96	1.47	0.90	0.66*	0.87	0.65*
	9	1.08	0.88	1.13	0.74*	0.88	**0.63***	0.97	1.52	0.91	**0.61***	0.89	0.64*

This table report the forecasting results for Brazil: the relative RMSE is the ratio of the model RMSE to the random walk without drift RMSE. Values smaller than one denote that the model forecasted better than the Random Walk. *, **, and ***, denote significance levels at 10%, 5%, and 1% of the p-values for the Diebold–Mariano test. h is the horizon of prediction: one month or one year ahead. z is the index for the models. The methods are Linear Regression OLS (LR), Regularized Regression Splines (SPLINE), Support Vector Machine with linear kernel SVM_L and with radial kernel SVM_R, Regression Tree (TREE) and Random Forest (RF). Bolded are the best method and model for each quadrant if significant and better than the benchmark. Source: created by the author.

Table 4 – Relative RMSE — Russia

h	z	Expanding Window						Roling Window					
		LR	SPLINE	SVM_L	SVM_R	TREE	RF	LR	SPLINE	SVM_L	SVM_R	TREE	RF
1	1	1.03	1.03	1.01	1.03	1.00	1.11	1.02	1.04	1.00	1.04	1.02	1.13
	2	1.02	1.02	1.00	1.00	1.01	1.11	1.02	1.04	1.00*	1.02	1.00	1.10
	3	1.02	1.06	1.01	1.03	1.01	1.18	1.04	1.11	1.02	1.02	1.01	1.20
	4	1.02	1.11	1.02	1.02	1.03	1.14	1.03	1.24	1.00	1.03	1.03	1.16
	5	1.01	1.00	1.00	1.04	1.04	1.28	1.01	1.02	0.99	1.05	1.10	1.30
	6	1.04	1.08	1.03	1.03	1.09	1.17	1.04	1.10	1.03	1.01	1.12	1.18
	7	1.11	1.20	1.07	1.06	1.45	1.45	1.20	1.61	1.15	1.05	1.40	1.50
	8	1.12	1.25	1.09	1.10	1.23	1.20	1.53	2.30	1.34	1.10	1.32	1.26
	9	1.11	1.19	1.10	1.08	1.24	1.17	1.49	1.90	1.31	1.08	1.33	1.18
12	1	0.95	0.84**	1.06	0.92	0.87*	0.83**	0.91	0.86*	0.96	0.90	0.84**	0.83*
	2	1.08	1.17	1.02	1.09	1.11	1.30	1.07	1.15	0.98	1.06	1.07	1.27
	3	1.09	0.97	1.05	0.96	1.04	1.05	1.18	1.12	1.00	1.04	1.07	1.13
	4	1.01	0.81*	1.15	0.87	0.86**	**0.76****	0.79*	**0.75***	0.97	0.70	0.86**	**0.75****
	5	0.91	0.91**	1.07	0.94	0.91	0.94	0.83	0.90**	0.93	0.85**	0.81**	0.93
	6	1.10	0.96	1.01	1.09	1.03	1.23	1.16	1.20	1.02	1.05	1.22	1.18
	7	1.26	1.26	1.07	1.15	1.04	1.21	1.32	3.43	1.08	1.11	0.98	1.22
	8	1.40	1.16	1.13	0.98	1.04	0.90	1.48	1.51	1.19	1.02	0.98	0.90
	9	1.23	1.16	1.07	0.92	1.04	0.83	1.30	1.98	1.14	0.93	0.98	0.85

This table report the forecasting results for Russia: the relative RMSE is the ratio of the model RMSE to the random walk without drift RMSE. Values smaller than one denote that the model forecasted better than the Random Walk. *, **, and ***, denote significance levels at 10%, 5%, and 1% of the p-values for the Diebold–Mariano test. h is the horizon of prediction: one month or one year ahead. z is the index for the models. The methods are Linear Regression OLS (LR), Regularized Regression Splines (SPLINE), Support Vector Machine with linear kernel SVM_L and with radial kernel SVM_R, Regression Tree (TREE) and Random Forest (RF). Bolded are the best method and model for each quadrant if significant and better than the benchmark. Source: created by the author.

All the methods employed, but the linear regression, have hyperparameters which need to be set before the estimation of the model. The selection of the best fit is conducted via hyperparameter tuning: at every window, the method is estimated multiple times with different combinations of hyperparameters. The fit with the smaller RMSE is chosen. The possible combinations are 15 steps inside a tuning grid (a range of values). Every method has a specific tuning grid.

The benchmark is the reference against which performance is measured. For exchange rate forecasting, the consensus is the random walk since the work of Meese and Rogoff (1983). Here, the choice is for the random walk without drift:

$$s_{t+h} - s_t = 0 \tag{14}$$

which implies that "the best predictor of exchange rates tomorrow is the exchange rate today." (ROSSI, 2013). To perform this measurement, the loss function chosen was the Root Mean Squared Error (RMSE) and the significance test was the Diebold-Mariano (DIEBOLD; MARIANO, 1995). Both are widely used in the exchange-rate forecasting literature.

The RMSE is expressed as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{15}$$

where $y = s_{t+h} - s_t$ and $\hat{y}$ is the $s_{t+h} - s_t$ forecasted by the model. n is the number of estimation windows. This means that at each window, the error is squared and the mean of them is root-squared. This loss function has the characteristic of penalizing more rigorously greater errors than smaller ones. The relative RMSE is the ratio of the model RMSE to that of the Random Walk RMSE. As smaller values are best, a relative RMSE smaller than one denotes that the model forecasted better than the benchmark.

The necessity of a significance test is "to know whether a victory 'in-sample' was merely good luck, or truly indicates a difference 'in population'."(DIEBOLD, 2012). The Diebold-Mariano test compares the error term of two models under the null hypothesis that both have the same accuracy and the alternative hypothesis that they have different levels of accuracy. p-values below 10% are considered significant. Therefore, a good fit has both relative RMSE smaller than one and a p-value smaller than 10%.

6 RESULTS

The results are show in Tables 3 to 7, one for each one of the BRICS and each with four quadrants: upper left for one month ahead forecasts with the expanding window scheme, upper right for one month ahead forecast with the rolling window scheme, lower left for one year ahead forecast with the expanding window scheme and lower right for one year ahead forecast with the rolling window scheme. The lines of every quadrant enumerate the models according to Table 1 and the columns represent the methods: Linear Regression (LR), Multivariate Adaptive Regression Splines (SPLINE), Support Vector Machine with linear kernel (SVM_L) and with radial kernel (SVM_R), Regression Tree (TREE) and Random Forest (RF).

In accordance with the literature, all the exchange rate models, in all settings, failed to substantially outperform the random walk without drift benchmark in the short horizon of one month ahead.

Table 5 – Relative RMSE — India

h	z	Expanding Window						Roling Window					
		LR	SPLINE	SVM_L	SVM_R	TREE	RF	LR	SPLINE	SVM_L	SVM_R	TREE	RF
1	1	1.05	1.08	1.03	1.03	1.09	1.11	1.15	1.22	1.07	1.01	1.12	1.23
	2	1.04	1.05	1.02	1.07	1.09	1.20	1.07	1.06	1.02	1.07	1.05	1.16
	3	1.04	1.07	1.02	1.04	1.09	1.15	1.06	1.05	1.03	1.05	1.12	1.16
	4	1.05	1.20	1.03	1.04	1.07	1.11	1.15	1.19	1.09	1.07	1.18	1.18
	5	0.99	0.99	0.99	1.01	1.00	1.32	1.03	1.11	1.02	1.02	1.06	1.46
	6	0.99	1.02	1.00	1.06	1.03	1.21	1.00	1.01	1.00	1.06	1.07	1.14
	7	1.03	1.11	1.01	1.02	1.07	1.09	1.04	1.04	1.03	1.02	1.00	1.08
	8	1.11	1.16	1.05	1.04	1.20	1.12	1.12	1.34	1.07	1.04	1.13	1.12
	9	1.09	1.08	1.04	1.05	1.08	1.10	1.08	1.18	1.05	1.04	1.10	1.09
12	1	0.73	0.71	0.66*	0.65**	0.63**	0.79	0.69*	0.67**	0.73*	0.65**	0.69*	0.71*
	2	0.67**	0.68**	0.67**	0.79**	0.66**	0.87	0.66**	0.67**	0.67**	0.79**	0.68**	0.83**
	3	0.67*	0.96	0.68**	0.83**	0.80*	0.94	0.65**	0.91	0.68**	0.80**	0.85	0.89
	4	0.77	0.96	0.68*	0.64**	0.63**	0.71	0.75*	0.75*	0.79	0.64**	0.71**	0.63**
	5	**0.60****	0.72	0.61**	0.75	0.63**	1.01	0.70*	0.70*	0.73	0.66*	0.68**	0.79
	6	0.67*	0.94	0.66**	0.83	0.79*	1.13	0.63**	0.92	0.66**	0.82	0.75**	0.97
	7	0.74**	1.22	0.73**	0.68**	0.71**	0.73**	0.82*	1.55	0.86*	0.69**	0.70**	0.74**
	8	0.84	1.09	0.94	0.65*	0.72**	**0.60****	0.73**	0.77	0.73**	0.65**	0.76**	**0.61****
	9	0.74**	0.98	0.84**	0.67	0.71**	0.62**	0.78**	0.89	0.89*	0.67*	0.70**	0.62**

This table report the forecasting results for India: the relative RMSE is the ratio of the model RMSE to the random walk without drift RMSE. Values smaller than one denote that the model forecasted better than the Random Walk. *, **, and ***, denote significance levels at 10%, 5%, and 1% of the p-values for the Diebold–Mariano test. h is the horizon of prediction: one month or one year ahead. z is the index for the models. The methods are Linear Regression OLS (LR), Regularized Regression Splines (SPLINE), Support Vector Machine with linear kernel SVM_L and with radial kernel SVM_R, Regression Tree (TREE) and Random Forest (RF). Bolded are the best method and model for each quadrant if significant and better than the benchmark. Source: created by the author.

Table 6 – Relative RMSE — China

h	z	Expanding Window						Roling Window					
		LR	SPLINE	SVM_L	SVM_R	TREE	RF	LR	SPLINE	SVM_L	SVM_R	TREE	RF
1	1	1.00	1.00	1.00	1.01	1.02	1.04	1.00	1.02	1.00	1.01	1.06	1.06
	2	1.01	1.02	1.01	1.01	1.00	1.06	1.01	1.01	1.00	1.01	1.04	1.07
	3	1.01	1.02	1.01	1.01	1.01	1.04	1.00	1.02	0.99**	1.01	1.04	1.05
	4	1.01	1.02	1.02	1.01	1.02	1.04	1.01	1.04	1.01	1.01	1.05	1.05
	5	1.02	1.02	1.01	1.02	1.05	1.14	1.01	1.01	1.01	1.00	1.05	1.12
	6	1.02	1.00	1.01	0.98	**0.95****	0.99*	1.00	0.97*	0.99	**0.93****	0.98*	1.02*
	7	1.06	1.05	1.02	0.99	1.06	1.06	1.06	1.05	1.04	1.00	1.07	1.06
	8	1.07	1.07	1.04	1.02	1.02	0.98	1.13	1.06	1.07	1.02	0.99	1.00
	9	1.07	1.08	1.04	1.02	1.06	1.01	1.13	1.13	1.06	1.01	1.04	1.02
12	1	0.91	0.85**	0.94	0.96	0.97	0.88	0.90**	0.92	0.94	0.91	0.96	0.86
	2	1.11	1.04	1.09	1.07	1.06	1.05	1.10	1.10	1.12	1.05	1.08	1.01
	3	1.03	0.94	1.05	0.88	0.96	0.80	0.97	0.99	1.01	0.84	0.95	0.80
	4	0.96	0.80***	1.01	0.83	0.87	0.75**	0.94	0.85	1.01	**0.68****	0.85	0.72**
	5	1.01	0.94	1.03	0.94	0.90	0.92	0.98	0.92	1.02	0.90	0.89	0.90
	6	1.05	1.04	1.06	1.09	1.04	1.03	1.01	1.04	1.03	1.11	1.06	1.02
	7	1.16	0.95	1.40	**0.74****	0.98	0.79	1.21	2.63	1.25	0.73*	1.01	0.79
	8	1.17	1.04	1.38	0.91	1.02	0.79*	1.34	1.77	1.36	0.94	1.02	0.78
	9	1.16	1.02	1.33	1.03	1.03	0.75*	1.36	1.66	1.36	0.84	0.98	0.76

This table report the forecasting results for China: the relative RMSE is the ratio of the model RMSE to the random walk without drift RMSE. Values smaller than one denote that the model forecasted better than the Random Walk. *, **, and ***, denote significance levels at 10%, 5%, and 1% of the p-values for the Diebold–Mariano test. h is the horizon of prediction: one month or one year ahead. z is the index for the models. The methods are Linear Regression OLS (LR), Regularized Regression Splines (SPLINE), Support Vector Machine with linear kernel SVM_L and with radial kernel SVM_R, Regression Tree (TREE) and Random Forest (RF). Bolded are the best method and model for each quadrant if significant and better than the benchmark. Source: created by the author.

Table 7 – Relative RMSE — South Africa

h	z	Expanding Window						Roling Window					
		LR	SPLINE	SVM_L	SVM_R	TREE	RF	LR	SPLINE	SVM_L	SVM_R	TREE	RF
1	1	1.01	1.04	1.03	1.03	1.05	1.15	1.01	1.04	1.06	1.02	1.14	1.14
	2	1.01	1.02	1.02	1.02	1.01	1.19	1.02	1.03	1.03	1.01	1.06	1.21
	3	1.01	1.03	1.02	1.01	1.01	1.14	1.02	1.07	1.03	1.02	1.11	1.17
	4	1.00	1.04	1.03	1.02	1.04	1.12	1.02	1.07	1.07	1.02	1.15	1.13
	5	**0.99***	**0.99***	**0.99***	1.03	1.01	1.11	1.00	1.01	1.02	1.04	1.09	1.14
	6	1.00*	1.00*	1.00	1.02	1.06	1.10	1.00*	1.01	1.01	1.03	1.06	1.12
	7	1.02	1.05	1.05	1.02	1.05	1.09	1.10	1.12	1.07	1.02	1.09	1.10
	8	1.04	1.10	1.10	1.02	1.03	1.07	1.12	1.66	1.10	1.02	1.12	1.07
	9	1.03	1.05	1.07	1.01	1.01	1.07	1.09	1.42	1.09	1.02	1.10	1.07
12	1	0.94	0.93	0.90	0.78***	0.93	0.81**	0.90	0.90**	0.87	0.78**	0.93	0.77**
	2	0.99	1.06	0.96	0.96	1.06	1.11	1.06	1.07	1.04	0.97	1.05	1.07
	3	0.98	0.86	1.01	0.85	0.89	0.74**	1.03	0.86	1.10	0.85	0.91	0.76**
	4	0.86	0.80	0.91	0.67***	0.87	**0.61*****	0.91	0.79*	0.94	0.69***	0.90	0.64***
	5	0.90*	0.88	0.88**	0.93	0.94	0.99	0.91*	0.98	0.87**	0.95	1.00	1.00
	6	0.90	0.87	0.92	0.86	0.88	0.85	0.92	0.78**	0.95	0.75*	0.77**	0.76*
	7	1.03	0.87	1.00	0.75**	0.88	0.73**	1.31	0.84	1.28	0.76**	0.91	0.74**
	8	1.08	0.78	1.13	0.82	0.87	0.62***	1.21	0.81	1.23	0.76**	0.86	**0.62*****
	9	1.10	0.81	1.16	0.79	0.88	0.62**	1.30	0.86	1.30	0.78*	0.86	0.63**

This table report the forecasting results for South Africa: the relative RMSE is the ratio of the model RMSE to the random walk without drift RMSE. Values smaller than one denote that the model forecasted better than the Random Walk. *, **, and ***, denote significance levels at 10%, 5%, and 1% of the p-values for the Diebold–Mariano test. h is the horizon of prediction: one month or one year ahead. z is the index for the models. The methods are Linear Regression OLS (LR), Regularized Regression Splines (SPLINE), Support Vector Machine with linear kernel SVM_L and with radial kernel SVM_R, Regression Tree (TREE) and Random Forest (RF). Bolded are the best method and model for each quadrant if significant and better than the benchmark. Source: created by the author.

Regarding the performance of the linearly estimated models, the Taylor Rule models where PPP does not hold, with the real exchange rate q, as well as the PPP model (Models 1, 4 and 5) showed the most improvement in the longer horizon for all countries. In some cases substantially beating the benchmark. Taylor Rule models where PPP always hold (Models 2 and 3) did not improve much and for some currencies worsened for the longer horizon. Model 6, the UIRP model, showed mixed performance, but improved in the long horizon for most countries. The monetary models (Models 7, 8 and 9) mostly resulted in worst predictions one year ahead than one month ahead. The exception was India, where all models and methods improved on the long horizon and Taylor Rule models where PPP always hold outperformed than the ones where it does not. This finding aligns the revision of Rossi (2013), who identifies predictability capacity on monetary models from horizons longer than 3 years, while Taylor Rule models show better performance at shorter horizons.

Non-linear methods outperformed the traditional ordinary least squares in about half the forecasts for every currency. The methods with better performance, however, were all non-linear statistical learning methods (with one exception for the Indian Rupee, where the linear model tied with the random forest). Regarding the best forecast for each quadrant, the random forest scored 7 out of 16 and the support vector machine with radial kernel scored 4. This contrasts with the work of Colombo and Pelegatti (2020) where this score was reversed: the SVM was found to better forecast. Also, this paper has found poorer performance of the statistical learning methods and less consistency across countries. In Colombo and Pelegatti (2020), the SVM_R displayed similar performance across countries.

The evidence that rolling windows can improve forecast by allowing the model to adjust to structural changes was not found. Results vary across currencies and across models where some forecasts performed better under the rolling window scheme and others under the expanding window scheme.

7 CONCLUSIONS

An exhaustive investigation was conducted where 9 forecasting models using 6 methods and 2 window scheme were estimated for two horizons and 5 currencies, resulting in 1080 predictions with the primary aim of validating if the new statistical learning tools would perform better than the classical linear method. It was confirmed that "The superior performance of a particular model/specification/currency combination does not typically carry over from one out-of-sample period to the other, nor from one specification to the other." (CHEUNG; CHINN; PASCUAL; ZHANG, 2019). However, it was also confirmed the improvement of non-linear statistical learning tools over the classical linear method.

REFERENCES

CHEUNG, Yin-Wong; CHINN, Menzie D.; PASCUAL, Antonio Garcia. Empirical exchange rate models of the nineties: Are any fit to survive? **Journal of International Money and Finance**, Netherlands, v. 24, n. 7, p. 1150–1175, Nov. 2005.

CHEUNG, Yin-Wong; CHINN, Menzie D.; PASCUAL, Antonio Garcia; ZHANG, Yi. Exchange rate prediction redux: New models, new data, new currencies. **Journal of International Money and Finance**, Netherlands, v. 95, p. 332–362, July 2019.

CLARIDA, Richard; GALÍ, Jordi; GERTLER, Mark. Monetary policy rules in practice some international evidence. English (US). **European Economic Review**, Elsevier, United States, v. 42, n. 6, p. 1033–1067, 1998.

COLOMBO, Emilio; PELAGATTI, Matteo. Statistical learning and exchange rate forecasting. **International Journal of Forecasting**, Netherlands, v. 36, n. 4, p. 1260–1289, Oct. 2020.

DIEBOLD, Francis. **Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests**. United States, Sept. 2012.

DIEBOLD, Francis X.; MARIANO, Roberto S. Comparing Predictive Accuracy. **Journal of Business and Economic Statistics**, United States, v. 13, n. 3, p. 253–263, July 1995.

GAGLIANONE, Wagner Piazza; MARINS, Jaqueline Terra Moura. Evaluation of exchange rate point and density forecasts: An application to Brazil. **International Journal of Forecasting**, Netherlands, v. 33, n. 3, p. 707–728, July 2017.

GUJARATI, Damodar N. **Basic Econometrics**. 4th. New York: McGraw-Hill, Jan. 2003.

HAMILTON, James. **Why You Should Never Use the Hodrick-Prescott Filter**. United States, May 2017.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning**. United States: Springer New York, 2009.

JAMES, Gareth et al. **An Introduction to Statistical Learning**. United States: Springer New York, 2013.

KUHN, Max; JOHNSON, Kjell. **Applied Predictive Modeling**. United States: Springer New York, 2013.

MEESE, Richard A.; ROGOFF, Kenneth. Empirical exchange rate models of the seventies. **Journal of International Economics**, Netherlands, v. 14, n. 1-2, p. 3–24, Feb. 1983.

MOLODTSOVA, Tanya; PAPELL, David H. Out-of-sample exchange rate predictability with Taylor rule fundamentals. **Journal of International Economics**, Netherlands, v. 77, n. 2, p. 167–180, Apr. 2009.

PETROPOULOS, Fotios et al. Forecasting: theory and practice. **International Journal of Forecasting**, Elsevier BV, v. 38, n. 3, p. 705–871, July 2022.

RIBEIRO, Pinho J. Selecting exchange rate fundamentals by bootstrap. **International Journal of Forecasting**, Elsevier BV, v. 33, n. 4, p. 894–914, Oct. 2017.

ROSSI, Barbara. Exchange Rate Predictability. **Journal of Economic Literature**, United States, v. 51, n. 4, p. 1063–1119, Dec. 2013.

SALISU, Afees A.; GUPTA, Rangan; KIM, Won Joong. Exchange rate predictability with nine alternative models for BRICS countries. **Journal of Macroeconomics**, Elsevier BV, v. 71, p. 103374, Mar. 2022.

SAX, Christoph; STEINER, Peter. Temporal Disaggregation of Time Series. **The R Journal**, Austria, v. 5, n. 2, p. 80, 2013.

TAYLOR, John B. Discretion versus policy rules in practice. **Carnegie-Rochester Conference Series on Public Policy**, Netherlands, v. 39, p. 195–214, 1993.

WANG, JIAN; WU, JASON J. The Taylor Rule and Forecast Intervals for Exchange Rates. **Journal of Money, Credit and Banking**, United States, v. 44, n. 1, p. 103–144, Jan. 2012.