PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Tito Guedes Bruni

Predictors for inflation under different transformations

Monografia de Final de Curso

Orientador: Prof. Gilberto Boaretto

Declaro que o presente trabalho é de minha autoria e que não recorri, para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor.

Rio de Janeiro, Junho de 2023

Agradecimentos

Ao meu orientador, Gilberto Boaretto, por partilhar comigo seu tempo tão escasso durante a fase final de seu doutorado. Além da sugestão do tema, seus comentários e seu conhecimento me permitiram alcançar meu objetivo com esta monografia.

Nestes quatro anos, tive ótimos professores, sendo alguns particularmente marcantes na minha trajetória. Agradeço ao professor Rogério Werneck, por me ensinar que mesmo após uma vida de estudos, ainda se pode aprender algo com aqueles que ainda estão engatinhando. Ao professor Marcelo Abreu, pela paciência na espera de que eu devolvesse os livros que pegava emprestados. Por fim, ao professor Carlos Tomei, por tudo que me ensinou e, em particular, por provar que é possível amar o que se faz mesmo depois de tantos anos.

Aos meus pais, Rejan e Sérgio, por me incentivarem a seguir minhas paixões, mesmo que isso os condenasse a ouvir longos monólogos sobre temas muitas vezes desinteressantes para eles. À minha irmã Flora, pela coragem de interromper os monólogos, nem sempre delicadamente, me lembrando de que há outras coisas na vida.

Aos meus amigos Fredie, Igor, Manu e Pietro, pelas risadas e por me levarem a caminhos que eu não exploraria sem sua influência.

Finalmente, à equipe Data Zoom e, em especial, ao professor Gustavo Gonzaga pela oportunidade de me juntar ao projeto. Meus estudos foram financiados pela PUC-Rio e pela CAPES, e, por isso, serei eternamente grato às duas instituições.

Abstract

We use machine learning (ML) models, namely Random Forest (RF), Complete Subset Regression (CSR), LASSO, adaLASSO, Elatic Net and Ridge to forecast Brazilian yearly CPI inflation. In particular, our goal is to determine whether accumulating predictor variables enhances the accuracy of inflation forecasts. We compare these models with the random walk (RW) and, mainly, with a survey of expectations from the Central Bank of Brazil called Focus. We show that the Random Forest beats the Focus consensus in all possible datasets with gains up to 58% in terms of RMSE. On the other hand, the performance of the shrinkage methods exhibits significant heterogeneity across different datasets. We show that machine learning models consistently outperform the benchmarks when predictor variables are not accumulated or accumulated in 12 months. Finally, we show that the models (especially RF) consistently outperform the benchmarks during periods when inflation is more volatile.

Keywords

Machine Learning; Inflation; Macroeconomics; Time-Series; Forecast.

CONTENTS

1 Introduction

Accurately forecasting inflation is important for several reasons. First of all, modern central banks calibrate their economic policies based on expected inflation (Iversen et al, 2016). Therefore, poor forecasts result in ineffective policies with high social costs. Secondly, many long-term contracts are set in nominal terms and thus bad inflation forecasts would generate undesired uncertainty. Finally, expectations about future prices are a key factor for households' consumption and investment decisions.

In this monograph we use machine learning (ML) methods to forecast 12-month cumulative Brazilian inflation. Namely, the models used are: random forest (Breiman, 2001), CSR (Elliotti; Gargano; Timmermann, 2013), LASSO (Tibshirani, 1996), elastic net (Zou; Hastie, 2005), adaLASSO (Zou, 2006) and ridge (Hoerl; Kennard, 1970). We also compute random walk (RW) forecasts but unlike other papers, we do not use this model as the main benchmark.

Medeiros et al. (2021) showed that ML models outperformed classical time series univariate models specially when US inflation was more volatile. In particular, they showed that Random Forest (RF) improves the root mean squared error (RMSE) by 25% when compared to the RW for 12-month forecasts. A previous work conducted by Garcia et al. (2017) showed similar results for Brazilian inflation: "high-dimensional models, such as shrinkage and complete subset regression, perform very well in the real-time forecasting of inflation in data-rich environments".

The aim of this study is to determine whether accumulating explanatory variables in $h = 1, ..., 12$ months enhances the 12-month inflation's forecasts. Unlike traditional forecasting papers, we are not evaluating multiple horizons but multiple datasets (while fixing a single horizon of inflation accumulated over 12 months). Recently, Coulombe et al. (2021) show that transformations in macroeconomic data can enhance forecasts' accuracy. However, the transformations used by the authors are different from the one we use. Namely, the transformations they use are moving average factors (MAF) and moving average rotation (MARX).

We use three different transformations to compute the percentage changes of the variables during h months. For instance, if one variable is an *index* and another variable is a *monthly percent change*, the ways of computing the percentage change of these variables during h months are different due to the difference between the variables.

Unlike other papers, our main benchmark is not an univariate model, but rather the median (consensus) of the Focus, a Brazilian survey of expectations that includes expectations for inflation. This report projects many economic variables and it takes into consideration the forecasts of more than a hundred professional forecasters.

We compare forecasting performance employing three error measures: root mean squared error (RMSE), mean absolute error (MAE), and median absolute deviation from the median (MAD). We also compute the rolling RMSEs to investigate the performance of each model through time.

The period of analysis goes from January 2006 to January 2023 and we compare rolling and expanding windows forecasts for yearly inflation from January 2014 to January 2023 with the values of yearly inflation. Brazil faced inflationary pressures from 2014 to 2016 due to government intervention in energy prices and from 2020 onward because of the Covid pandemic.

We show that the random forest beats the Focus consensus with all the twelve datasets, showing gains up to 58% RMSE. On the other hand, the performance of the shrinkage methods exhibits significant heterogeneity across the datasets. For instance, considering all the combinations of models and datasets, the smallest and the highest RMSEs came from the adaLASSO (when data is accumulated in 10 and 6 months, respectively).

Our study shows that all the models beat the Focus consensus when data is not accumulated, that is, when we employ the monthly predictors without transformations. In addition, the only other dataset for which almost every model beats the main benchmark is the one where data is accumulated in 12 months. The second result seems intuitive given that the target variable is also accumulated in 12 months.

After computing the rolling RMSEs, we show that RF, LASSO and adaLASSO achieve the best results in terms of *percentage of times* outperforming FOCUS consensus when data is not accumulated, while CSR, elastic net and ridge achieve the best results when data is accumulated in 12 months. These results provide a dynamic perspective for why forecasters should focus on datasets with non-accumulated variables and datasets with 12-month accumulated variables to forecasts yearly inflation. Among all the models, the random forest with non-accumulated data shows the best results by beating the Focus 80% of the time.

Finally, the models (especially RF) consistently outperform the bechmarks during periods when inflation is more volatile (2014-2016 and 2020 onward). This finding is in

line with previous studies that show that machine learning methods perform well during periods of volatile inflation.

Bellow, we briefly summarize the sections of this monograph.

Data and method. In Chapter 2, we explain the features of the selected predictor variables, and how we compute the cumulative percentage changes of the variables. In addition, we show how the models compute the forecasts, and finally which error measures are used to evaluate the performance of the models.

Results. In chapter 3, we provide tables and plots to compare the performance of the models using different error measures. Initially, we analyse the error measures across the entire window of predictions. Then, we analyse 12-month rolling RMSE.

Conclusion. In chapter 4, we enumerate the main results and we conclude that forecasters should evaluate forecasts when predictors are accumulated in 12 months as a robustness exercise after computing the forecasts without accumulating the predictors.

2 Data and Method

2.1 Data

Our data consists of 84 variables, 82 of them extracted from the Time Series Management System. This plataform gathers macroeconomic data from various sources such as the the Brazilian Institute of Geography and Statistics (IBGE) and the Central Bank of Brazil (BCB). We selected variables related to prices, commodities, economic activity, employment, electricity, confidence, finance, credit, government, and international trade. We use data from January 2006 to January 2023 to compute forecasts of inflation from January 2014 to January 2023.

2.2 Transformations

We aim to analyse how accumulating explanatory variables in $h = 1, 2, ..., 12$ months affects the quality of the machine learning models' forecasts of the 12-month ahead Brazilian yearly inflation rate. We apply three types of transformations depending on the features of each variable:

- Transformation 1: $X_t^h = \left(\dfrac{X_t}{X_{t-h+1}} - 1 \right) 100$

- Transformation 2: $X_t^h = \left(\Pi_{t=1}^h \left(1 + \dfrac{X_t}{100} \right) - 1 \right) 100$

- Transformation 3: $X_t^h = X_t - X_{t-h+1}$

Transformation 1 computes the percentage change of *monthly indexes*, such as the Commodities Index (ICBR), during h months. Transformation 2 computes percentage changes of variables which are *monthly percent change*, such as the inflation rate, during h months. Transformation 3 simply applies *differences* (in h months). It is applied to variables such as the unemployment rate (%) and treasury term (months). Appendix A has a complete table describing the transformations applied to each variable.

2.3 Estimation

We compute 109 forecasts based on rolling and expanding window schemes using five linear models (CSR, LASSO, ridge, elastic net, adaLASSO) and one non-linear model (Random Forest). These models present interesting properties such as the shrinkage of coefficients, selection of variables and non-linearity. The initial specification of the linear models is:

$$\pi_{t+12}^{12} = c + \phi_0 \pi_t^{12} + \phi_1 \pi_{t-1}^{12} + \phi_2 \pi_{t-2}^{12} + \phi_3 \pi_{t-3}^{12} + \eta FOCUS_{t+12|t}^{12} + \mathbf{X_t}\Theta + u_{t+12}$$

where π_{t+12}^{12} is the 12-month ahead inflation accumulated in 12 months and $FOCUS_{t+12|t}^{12}$ is the median of the Focus survey of inflation expectations at time t for time $t+12$ and

$$\Theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}, \mathbf{X_t} = \begin{pmatrix} X_{1,t}^h \\ \vdots \\ X_{k,t}^h \end{pmatrix}$$

where Θ contains the coefficients of $X_1^h, ..., X_k^h$ observed at time t. Since we use 84 explanatory variables with the Focus consensus being one of the variables, we have that $k = 83$. From now on, to simplify notation, the coefficients $c, \phi_0, \phi_1, \phi_2, \phi_3, \eta, \theta_1, ..., \theta_k$ will be referred to as $\beta_0, \beta_1, ..., \beta_p$ with $p = k + 5$.

2.3.1 Complete Subset Regression

The complete subset regression (Elliotti; Gargano; Timmermann, 2013) consists of estimating a large number of linear regressions with a fixed number of explanatory variables and computing the mean of the predictions generated by all the models. Our datasets have 88 predictors and even if we estimated models with only four variables, there would be more than 2 million possible combinations of models to be computed. In addition, we would have to compute all these possibilities for each prediction given that we estimate the models using rolling and expanding windows.

To deal with the complexity of the models, we use the R-package, `HDeconometrics` which has a function that previously selects 20 predictors (based on the t-statistic of the coefficients of the predictors obtained by estimating regressions) and computes all possible models with four variables.

2.3.2 Shrinkage Methods

Shrinkage methods are known for their ability to reduce the out-of-sample mean square error (MSE) and mitigating problems associated with overfitting. These methods are able shrink the OLS coefficients by imposing a penalty. All the models within this class follow this general specification:

$$\hat{\boldsymbol{\beta}}(\lambda) = \arg \min_{\boldsymbol{\beta} \in R^p} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \underbrace{\sum_{j=1}^{p} p(\beta_j; \lambda, \boldsymbol{\alpha})}_{\text{penalty}}$$

The first part of this optimization problem corresponds to the OLS problem. The second part corresponds to the penalty composed by a non-negative penalty function $p(.)$, whose arguments are the coefficients $\boldsymbol{\beta}$, the regularization parameter λ and hyper-parameters $\boldsymbol{\alpha}$ (such as weights $\boldsymbol{\omega}$ associated to the variables). For all the shrinkage methods used in this monograph, the value of λ is determined using the BIC criterion (Schwarz, 1978) as other papers have previously done (Medeiros et al, 2021).

We use four of these models to compute our inflation's forecasts: LASSO, adaLASSO, elastic net, and ridge. Figure 1 shows the format of the penalties. The value of the coefficients is determined by the point of tangency between the contour lines from $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ and the penalty.
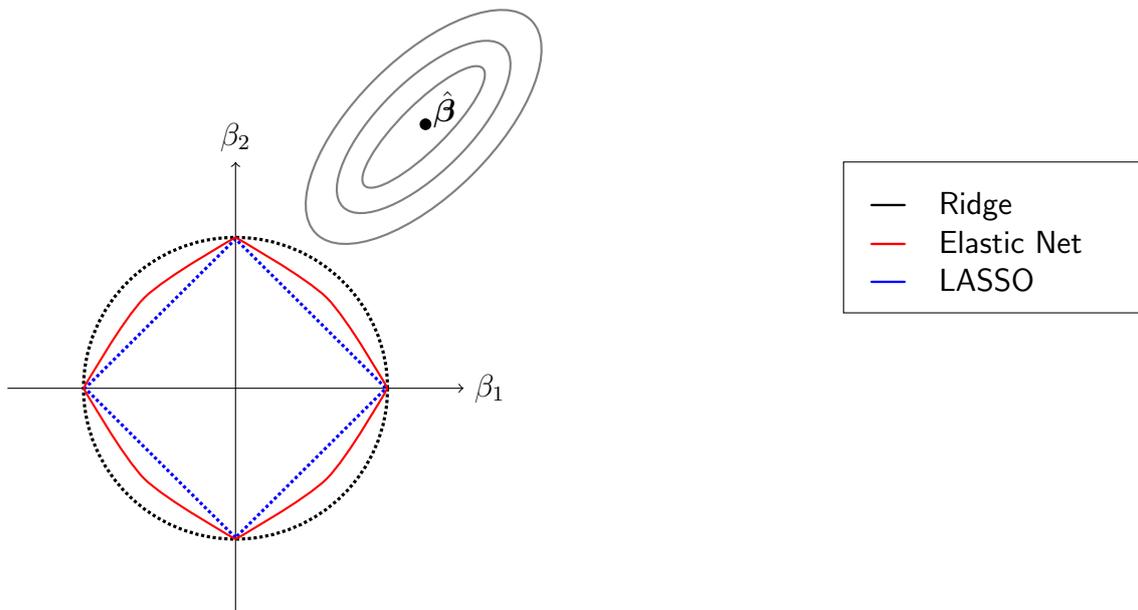

Figure 1 – Regularization Methods for Linear Regression

The penalties from Figure 1 are particular cases of L^p-norms. For $p \geq 1$, the L^p-

norm of $\boldsymbol{x} = (x_1, ..., x_n)$ is:

$$||\boldsymbol{x}||_p = (|x_1|^p + |x_2|^p + ... + |x_n|^p)^{\frac{1}{p}}$$

I) *LASSO*

The Least Absolute Shrinkage and Selection Operator (Tibshirani, 1996) is able to shrink coefficients up to zero and for this reason it is commonly used to select variables. Its penalty assumes the format of the L^1-norm.

$$\hat{\boldsymbol{\beta}}_{\mathsf{LASSO}}(\lambda) = \arg\min_{\boldsymbol{\beta} \in R^p} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \lambda ||\boldsymbol{\beta}||_1$$

II) *Ridge*

The ridge regression (Hoerl; Kennard, 1970) is another method that implements shrinkage of the coefficients' values. However, unlike the LASSO, the ridge regression is not able to implement variable selection. This happens because its *shrinkage penalty* assumes the format of a L^2-norm.

$$\hat{\boldsymbol{\beta}}_{\mathsf{ridge}}(\lambda) = \arg\min_{\boldsymbol{\beta} \in R^p} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \lambda ||\boldsymbol{\beta}||_2$$

III) *Elastic Net*

The elastic net (Zou; Hastie, 2005) rests in between the LASSO and ridge. It tends to shrink the values of the coefficients more than the ridge but it cannot make them go to zero like the LASSO can. Its penalty is a mix of L^1 and L^2 norms:

$$\hat{\boldsymbol{\beta}}_{\mathsf{elnet}}(\lambda_1, \lambda_2) = \arg\min_{\boldsymbol{\beta} \in R^p} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \lambda_1 ||\boldsymbol{\beta}||_1 + \lambda_2 ||\boldsymbol{\beta}||_2$$

IV) *adaLASSO*

The adaptive LASSO (Zou, 2006) is an extension of the LASSO that presents theoretical advantages when compared to LASSO. In particular, the conditions for consistency of its variable selection are weaker than the conditions for the LASSO selection. The adaLASSO incorporates individual weights for each parameter, unlike LASSO which applies the same weighting to all parameters in the penalty. Irrelevant regressors should receive larger weights. To determine the values of the weights, we first compute $\hat{\boldsymbol{\beta}}_{\mathsf{LASSO}}$ and then we define the weights:

$$w_j = \frac{1}{|\hat{\beta}_{\mathsf{LASSO}_j}| + \frac{1}{\sqrt{n}}}, \; j = 1, ..., p$$

$$\hat{\boldsymbol{\beta}}_{\text{adaLASSO}}(\lambda) = \arg \min_{\boldsymbol{\beta} \in R^p} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|$$

2.3.3 Random Forest

The random forest (Breiman, 2001) is a non-parametric method which applies bootstrap to regression trees, that allows to deal with non-linearity. Therefore, to explain the random forest we need to explain what a regression tree is.

Suppose we have a dataset with N observations and $K+1$ variables $x^1, ..., x^K, \pi$, with π being the target variable and $\boldsymbol{x}$ being the predictors. Fix $N^* < N$. A decision tree analyses the first N^* observations of the variables, providing a tree $T^*(.)$ which is a decision rule to forecast the target variable. For instance, the predicted value of π_{N^*+1} is:

$$\hat{\pi}_{N^*+1} = T^*(x_{N^*+1}^1, \ldots, x_{N^*+1}^K)$$

Figure 2 illustrates a decision tree where π_i, $i = 1, ..., N^*$ represents the inflation at line i, and $x^j = x_{N^*+1}^j$, and c_i is a threshold for the decision rule. In this particular example, we presume that $\pi_i \neq \pi_j, \forall i \neq j$. Finally, notice that in our example the number of thresholds is equal to the number of observations (N^*), however in the majority of the cases the dependent variable can be explained with much fewer thresholds.
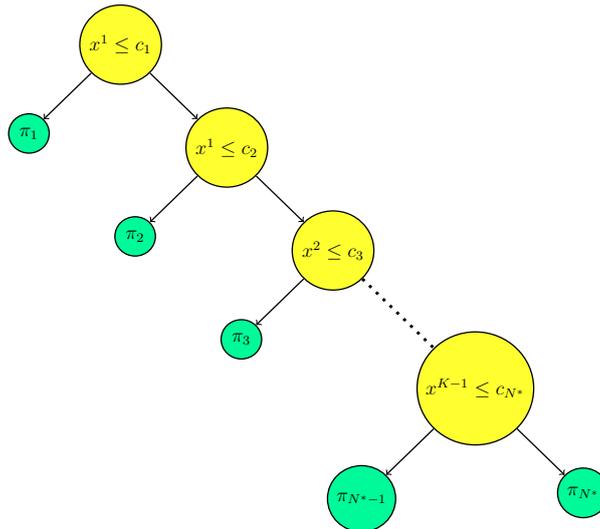
Figure 2 – Illustration of a Regression Tree

The random forest applies bootstrap to regression trees. It generates many regression trees by randomly selecting some variables and some observations. The goal is to

generate regression trees which are considerably different from each other. Finally, we take the average of the output of the regression trees. Let B be the total number of trees. The random forest prediction is given by:

$$\frac{1}{B} \sum_{b=1}^{B} T_b^*(\mathbf{X})$$

where T_1^* is the first bootstrap regression tree, T_2^* is the second, and so on and so forth.

2.4 Error Measures

To measure the performance of our forecasts, we compare the models from three statistics: root mean squared error (RMSE), mean absolute error (MAE), and median absolute deviation from the median (MAD). The forecast error from model m at time t is given by:

$$\hat{e}_{t,m} = \pi_t^{12} - \hat{\pi}_{t|t-12,m}^{12}$$

Then, the error measures are:

$$RMSE_m = \sqrt{\frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \hat{e}_{t,m}^2}$$

$$MAE_m = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} |\hat{e}_{t,m}|$$

$$MAD_m = \mathsf{median}\big|\hat{e}_{t,m} - \mathsf{median}(\hat{e}_{t,m})\big|$$

where T is the time of the last forecast and T_0 is the time of the first forecast.

Given we make forecasts from 2014 until 2023, the value of the error measures of each model in this period can only explain how well each model behaved throughout this long window. Therefore, it does not evaluate how each model behaved during specific periods within the window. For this reason, we also compute the rolling RMSE of each model where each window has 12 months.

3 Results

We forecast Brazilian yearly inflation from January 2014 until January 2023, which means a total of 109 predictions. To compare the performance of each model with respect to the main benchmark, the median of the Focus survey, we normalize the values of their error measures by the values of the error measures from Focus. Initially, we analyse the values of the normalized error measures throughout the entire period of predictions. In a second moment, we compute RMSEs from 12-month rolling windows to have a dynamic view of the performance of models. We repeat this procedure for rolling and expanding windows forecasts. In this chapter, we are going to expose the results from the rolling window forecasts. The expanding window results can be found in Appendix C.

Table 1 shows the values of the error measures from the rolling window forecasts normalized by the Focus RMSE (2.9) and MAE (2.1). When the normalized values are smaller than 1, the model outperform the Focus consensus. Finally, normalized RMSE and MAE from the random walk (RW) are 1.13 and 1.23, respectively, showing that the RW is outperformed by the main benchmark.

Table 1 – RMSE, MAE and MAD for rolling window forecasts

Datasets: X^h, $h = 1, ..., 12$												
RMSE/(MAE)/ {MAD}	1	2	3	4	5	6	7	8	9	10	11	12
RF	0.42	0.62	0.60	0.59	0.60	0.61	0.58	0.56	0.56	0.52	0.52	0.49
	(0.49)	(0.70)	(0.68)	(0.68)	(0.68)	(0.68)	(0.66)	(0.63)	(0.63)	(0.59)	(0.58)	(0.53)
	{0.56}	{0.85}	{0.77}	{0.78}	{0.85}	{0.84}	{0.82}	{0.76}	{0.71}	{0.70}	{0.68}	{0.61}
CSR	0.65	0.74	0.72	0.70	1.51	1.49	1.26	0.63	0.61	0.60	0.59	0.72
	(0.72)	(0.84)	(0.80)	(0.78)	(0.91)	(0.90)	(0.86)	(0.71)	(0.70)	(0.69)	(0.68)	(0.71)
	{0.78}	{1.01}	{0.94}	{0.86}	{0.88}	{0.90}	{0.93}	{0.93}	{0.89}	{0.80}	{0.74}	{0.76}
LASSO	0.44	0.77	2.39	4.15	4.60	7	1.22	5.66	1.99	0.43	2.22	1.67
	(0.46)	(0.85)	(1.18)	(1.42)	(1.28)	(1.46)	(0.65)	(1.31)	(0.72)	(0.41)	(0.70)	(0.67)
	{0.51}	{0.96}	{0.84}	{0.74}	{0.50}	{0.48}	{0.47}	{0.50}	{0.47}	{0.41}	{0.44}	{0.40}
adaLASSO	0.45	0.60	1.63	5.60	4.37	7.24	1.14	4.94	0.51	0.35	1.64	0.46
	(0.49)	(0.68)	(1.06)	(1.57)	(1.22)	(1.48)	(0.62)	(1.15)	(0.45)	(0.38)	(0.61)	(0.46)
	{0.57}	{0.66}	{0.88}	{0.67}	{0.49}	{0.50}	{0.47}	{0.47}	{0.39}	{0.38}	{0.45}	{0.43}
ElNet	0.45	3.23	2.35	4.13	4.50	7.12	1.22	5.46	2.39	0.51	2.68	0.44
	(0.50)	(1.25)	(1.20)	(1.40)	(1.28)	(1.46)	(0.67)	(1.29)	(0.80)	(0.45)	(0.77)	(0.45)
	{0.48}	{0.96}	{0.92}	{0.78}	{0.56}	{0.51}	{0.45}	{0.50}	{0.47}	{0.48}	{0.43}	{0.39}
Ridge	0.52	1.04	1.88	4.05	4.81	5.32	1.90	3.82	3.69	1.40	1.76	0.51
	(0.57)	(1.24)	(1.32)	(1.43)	(1.27)	(1.27)	(0.89)	(1.14)	(0.98)	(0.57)	(0.67)	(0.47)
	{0.58}	{1.66}	{1.15}	{0.92}	{0.69}	{0.54}	{0.51}	{0.48}	{0.48}	{0.43}	{0.45}	{0.35}

The columns represent the datasets. The coloured values correspond to the best result of the models for each error measure.

The random forest consistently beats the benchmark across all datasets and the gains

can be as large as 58% in terms of RMSE. The CSR also beats the benchmark with multiple datasets but unlike the random forest, sometimes it is beaten by the Focus. The shrinkage methods exhibit significant heterogeneity across datasets. For instance, when data is accumulated from 3 to 8 months, the shrinkage methods consistently fail to beat the benchmark. On the other hand, they show significant gains in comparison to the Focus consensus when data is not accumulated or accumulated in 12 months.

In general, the six models consistently outperform the benchmark when data is not accumulated or accumulated in 12 months. However, the overall performance of the models varies across the other datasets. For this reason, our analysis emphasizes the results obtained with these two datasets.

When regressors are not accumulated, all the models outperform the benchmark showing gains around 50% in terms of RMSE. In particular, the random forest achieves the best result.

When regressors are accumulated in 12 months, all the models (except for LASSO) show significant gains in terms of RMSE (up to 56%) when compared to the Focus. This result seems reasonable given that we are forecasting a variable that is also accumulated in 12 months. The operation of accumulating the regressors in 12 months should make them smoother and this could be an interesting property when forecasting a yearly variable.

Table 2 shows descriptive statistics about the performance of models across all datasets. RF and CSR are the only models with an average RMSE smaller than the benchmark. This result is expected given that in Table 1 we see these two models consistently beating the benchmark across multiple datasets. Finally, we find that adaLASSO obtains at the same time the lowest and the highest RMSE and MAE across our datasets. Similarly, ridge shows at the same time the lowest and the highest MAD. These two results are in line with our previous finding that shrinkage methods performance varies markedly across different datasets.

Medeiros et al. (2021) showed that ML models (especially the RF) outperformed traditional times-series univariate models in forecasting US inflation, mainly during volatile periods. We expect Brazilian inflation to be more volatile in comparison to the US given that "emerging markets usually exhibit higher and more volatile inflation" (Garcia; Medeiros; Vasconcelos, 2017). For this reason, Brazilian data should be an interesting resource to test machine learning methods (high) performance when inflation is more volatile.

Table 2 – Descriptive statistics: rolling window forecasts errors

	$\overline{RMSE}$	$\overline{MAE}$	$\overline{MAD}$	max_{RMSE}	max_{MAE}	max_{MAD}	min_{RMSE}	min_{MAE}	min_{MAD}
RF	0.56	0.63	0.74	0.62	0.70	0.85	0.42	0.49	0.56
CSR	0.85	0.77	0.87	1.51	0.91	1.01	0.59	0.68	0.74
LASSO	2.71	0.92	0.56	7.00	1.46	0.96	0.43	0.41	0.40
adaLASSO	2.41	0.85	0.53	7.24	1.57	0.88	0.35	0.38	0.38
ElNet	2.87	0.96	0.58	7.12	1.46	0.96	0.44	0.45	0.39
Ridge	2.56	0.98	0.69	5.32	1.43	1.66	0.51	0.47	0.35
RW	1.13	1.23	1.04	1.13	1.23	1.04	1.13	1.23	1.04

Our 12-month rolling RMSE computations start in January 2015 and end in January 2023. In Figure 3, we can see that from 2015 until the beginning of 2017 and from 2021 onward the majority of the ML models beat the Focus. This is expected given that these are two periods when Brazilian inflation grew more than it normally does. In the first period, the growth of inflation is related to the government's intervention in energy prices and, more recently, is related to the Covid. Notice that even though the pandemic starts in March 2020, it only completely impacts the 12-month accumulated inflation from February 2021 onward.
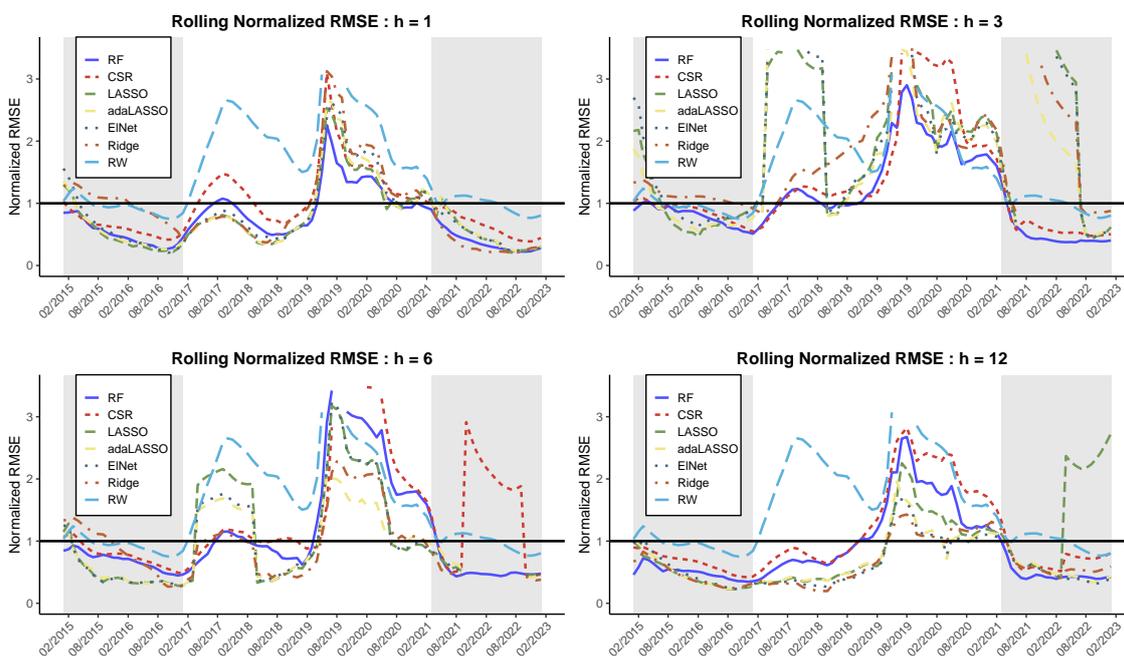

Figure 3 – Rolling RMSE

These plots illustrate a similar result to that shown in Table 1: in general, the models are more accurate when we do not accumulate the data or accumulate it in 12 months. With these two datasets, the only period of time when the models are generally outperformed by the Focus is from February 2019 to February 2021. However, when we

use the other datasets, the group of models presents more heterogeneous results through time. In addition, we get bigger and more volatile errors through time when we use these other datasets. The complete version of the plots can be found in the Appendix B.

A desired property for a statistical model is the ability of beating a benchmark as many times as possible. Table 3 shows the percentage of the times when the rolling RMSE of each model-dataset pair is smaller than the rolling RMSE of the Focus. The total number of 12-month rolling RMSEs computed from January 2015 to January 2023 is 97. We show that for all the models, the best results are either obtained when data is not accumulated or accumulated in 12 months. In particular, the RF with non-accumulated regressors beats the FOCUS more than 80% of the time.

Table 3 – Percentage of the times when the models beat Focus

	1	2	3	4	5	6	7	8	9	10	11	12
RF	80.41	53.61	58.76	59.79	67.01	68.04	70.10	75.26	76.29	75.26	76.29	70.10
CSR	62.89	45.36	48.45	48.45	44.33	43.30	45.36	54.64	55.67	58.76	60.82	69.07
LASSO	73.20	54.64	32.99	41.24	49.48	58.76	65.98	43.30	55.67	73.20	73.20	64.95
adaLASSO	75.26	60.82	36.08	43.30	45.36	53.61	67.01	56.70	76.29	75.26	70.10	75.26
ElNet	74.23	40.21	31.96	39.18	46.39	58.76	55.67	43.30	53.61	75.26	74.23	77.32
Ridge	60.82	28.87	16.49	31.96	45.36	47.42	40.21	35.05	58.76	68.04	67.01	76.29

Percentage of times that 12-month rolling RMSE of a model-dataset pair is smaller than the 12-month rolling RMSE of the Focus consensus. RW beats the benchmark only about 30% of the time.

Finally, the results from the expanding window forecasts are similar to the results from the rolling window forecasts. One remarkable difference is that when using expanding window, all the models (except for the RF) obtain the best MAD when data is accumulated in 12 months.

4 Conclusion

We investigate the performance of six machine learning methods for forecasting 12-month-ahead yearly inflation, considering the Brazilian case. In particular, we study whether there are gains from accumulating the regressors in 1 to 12 months. Our benchmark is the median of the Focus survey of inflation expectations.

Our main findings are:

1. Machine learning models consistently outperform the Focus consensus when data is not accumulated or it is accumulated in 12 months.

2. Machine learning models (especially RF) consistently outperform the Focus during periods of more volatile inflation.

3. The random forest (RF) is the only model that outperforms the benchmark in every dataset. In particular, the best model is the RF without accumulating regressors.

4. The performance of shrinkage methods varies considerably across different datasets.

Following our first finding, we recommend evaluating forecasts when regressors are accumulated in 12 months as a robustness exercise after computing the forecasts without applying transformations to the regressors.

Bibliography

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001.

COULOMBE, P. G. et al. Macroeconomic data transformations matter. *International Journal of Forecasting*, Elsevier, v. 37, n. 4, p. 1338–1354, 2021.

ELLIOTT, G.; GARGANO, A.; TIMMERMANN, A. Complete subset regressions. *Journal of Econometrics*, Elsevier, v. 177, n. 2, p. 357–373, 2013.

GARCIA, M. G.; MEDEIROS, M. C.; VASCONCELOS, G. F. Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting*, v. 33, n. 3, p. 679–693, 2017. ISSN 0169-2070. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S0169207017300262⟩.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Taylor & Francis, v. 12, n. 1, p. 55–67, 1970.

IVERSEN, J. et al. Real-time forecasting for monetary policy analysis: The case of sveriges riksbank. *Riksbank Research Paper Series*, n. 142, 2016.

MEDEIROS, M. C. et al. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, Taylor Francis, v. 39, n. 1, p. 98–119, 2021. Disponível em: ⟨https://doi.org/10.1080/07350015.2019.1637745⟩.

SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, JSTOR, p. 461–464, 1978.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996.

ZOU, H. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, Taylor & Francis, v. 101, n. 476, p. 1418–1429, 2006.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, Wiley Online Library, v. 67, n. 2, p. 301–320, 2005.

A Table of transformations

The table provides the classification, the code (from the Time Series Management System), the frequency, the unity of measure, the lags, and finally the transformation applied to each predictor. When Transformation=0, the variable is not transformed. Transformations 1, 2, and 3 are explained here.

	Group	Variable	Code	Frequency	Unity	Transformation	Lag
1	PRICES	ipca	433	M	month_%_var	2	1
2	PRICES	ipca_ali	1635	M	month_%_var	2	1
3	PRICES	ipca_hab	1636	M	month_%_var	2	1
4	PRICES	ipca_resid	1637	M	month_%_var	2	1
5	PRICES	ipca_vest	1638	M	month_%_var	2	1
6	PRICES	ipca_transp	1639	M	month_%_var	2	1
7	PRICES	ipca_comunic	1640	M	month_%_var	2	1
8	PRICES	ipca_saude	1641	M	month_%_var	2	1
9	PRICES	ipca_desp	1642	M	month_%_var	2	1
10	PRICES	ipca_educ	1643	M	month_%_var	2	1
11	PRICES	ipc_BR	191	M	month_%_var	2	1
12	PRICES	igp_M	189	M	month_%_var	2	1
13	PRICES	igp_DI	190	M	month_%_var	2	1
14	PRICES	igp10	7447	M	month_%_var	2	1
15	PRICES	ipca15	7478	M	month_%_var	2	1
16	PRICES	bm_broad	1788	M	cmu_thousand	1	2
17	PRICES	bm	1785	M	cmu_thousand	1	2
18	PRICES	m1	1783	M	cmu_thousand	1	2
19	PRICES	m2	1786	M	cmu_thousand	1	2
20	PRICES	m3	27813	M	cmu_thousand	1	2
21	PRICES	m4	27815	M	cmu_thousand	1	2
22	COMMODITIES	icbr	27574	M	index	1	1
23	COMMODITIES	icbr_agr	27575	M	index	1	1
24	COMMODITIES	icbr_metal	27576	M	index	1	1
25	COMMODITIES	icbr_energy	27577	M	index	1	1
26	ACTIVITY	ibcbr	24363	M	index	1	3
27	ACTIVITY	pimpf	21859	M	index	1	2
28	ACTIVITY	pimpf_extract	21861	M	index	1	2
29	ACTIVITY	pimpf_manufac	21862	M	index	1	2
30	ACTIVITY	retail_total	1455	M	index	1	2
31	ACTIVITY	retail_fuel	1483	M	index	1	2
32	ACTIVITY	retail_supermarket	1496	M	index	1	2
33	ACTIVITY	retail_clothing	1509	M	index	1	2
34	ACTIVITY	retail_house	1522	M	index	1	2
35	ACTIVITY	retail_drugstore	20099	M	index	1	2
36	ACTIVITY	retail_paper	20101	M	index	1	2

	Group	Variable	Code	Frequency	Unity	Transformation	Lag
37	ACTIVITY	retail_office	20102	M	index	1	2
38	ACTIVITY	retail_others	20104	M	index	1	2
39	ACTIVITY	retail_building	20105	M	index	1	2
40	ACTIVITY	retail_auto	1548	M	index	1	2
41	ACTIVITY	prod_vehicles	1373	M	units	1	1
42	ACTIVITY	prod_agr_mach	1388	M	units	1	1
43	ACTIVITY	vehicle_sales	7389	M	barrels_day_thousand	1	1
44	ACTIVITY	tcu	24352	M	%	3	1
45	EMPLOYMENT	min_wage	1619	M	cmu	1	2
46	EMPLOYMENT	aggreg_wage	10790	M	R$	1	0
47	EMPLOYMENT	unem	NA	M	%	3	3
48	ELECTRICITY	elec	1406	M	GWh	1	3
49	ELECTRICITY	elec_com	1402	M	GWh	1	3
50	ELECTRICITY	elec_res	1403	M	GWh	1	3
51	ELECTRICITY	elec_ind	1404	M	GWh	1	3
52	CONFIDENCE	cons_confidence	4393	M	index	1	1
53	CONFIDENCE	future_expec	4395	M	index	1	1
54	FINANCE	irf_m	12461	D	index	1	1
55	FINANCE	ima_s	12462	D	index	1	1
56	FINANCE	ima_b	12466	D	index	1	1
57	FINANCE	ima	12469	D	index	1	1
58	FINANCE	saving_deposits	1838	M	cmu_thousand	1	2
59	FINANCE	selic	4390	M	%_pm	0	1
60	FINANCE	cdi	4391	M	%_pm	0	1
61	FINANCE	tjlp	256	M	%_py	0	1
62	FINANCE	ibovespa	NA	D	%_pm	0	1
63	CREDIT	cred_total	28183	M	R$_million	1	2
64	CREDIT	cred_gdp	28215	M	%	3	2
65	CREDIT	indebt_house_19882	19882	M	%	3	4
66	CREDIT	indebt_house_20400	20400	M	%	3	4
67	GOVERNMENT	net_debt_gdp	4513	M	%	3	2
68	GOVERNMENT	net_debt	4478	M	R$_million	3	2
69	GOVERNMENT	net_debt_fedgov_bcb	4468	M	R$_million	3	2
70	GOVERNMENT	net_debt_states	4472	M	R$_million	3	2
71	GOVERNMENT	net_debt_cities	4473	M	R$_million	3	2
72	GOVERNMENT	primary_result	4649	M	R$_million	3	2
73	GOVERNMENT	debt_fedgov_old	4502	M	R$_million	1	2
74	GOVERNMENT	debt_fedgov_new	13761	M	R$_million	1	2
75	GOVERNMENT	treasury_emit	4151	M	cmu_million	1	2
76	GOVERNMENT	treasury_mkt	4154	M	cmu_million	1	2
77	GOVERNMENT	treasury_term	10616	M	months	3	2
78	GOVERNMENT	treasury_dur	10617	M	months	3	2
79	INTERNATIONAL	reer	11752	M	index	1	2
80	INTERNATIONAL	usd_brl_end	3695	M	cmu_US$	1	1
81	INTERNATIONAL	usd_brl_avg	3697	M	cmu_US$	1	1
82	INTERNATIONAL	current_account	22701	M	US$_million	1	2
83	INTERNATIONAL	trade_balance	22707	M	US$_million	1	2
84	INTERNATIONAL	imports	22709	M	US$_million	1	2

B Rolling window forecasts: rolling RMSE
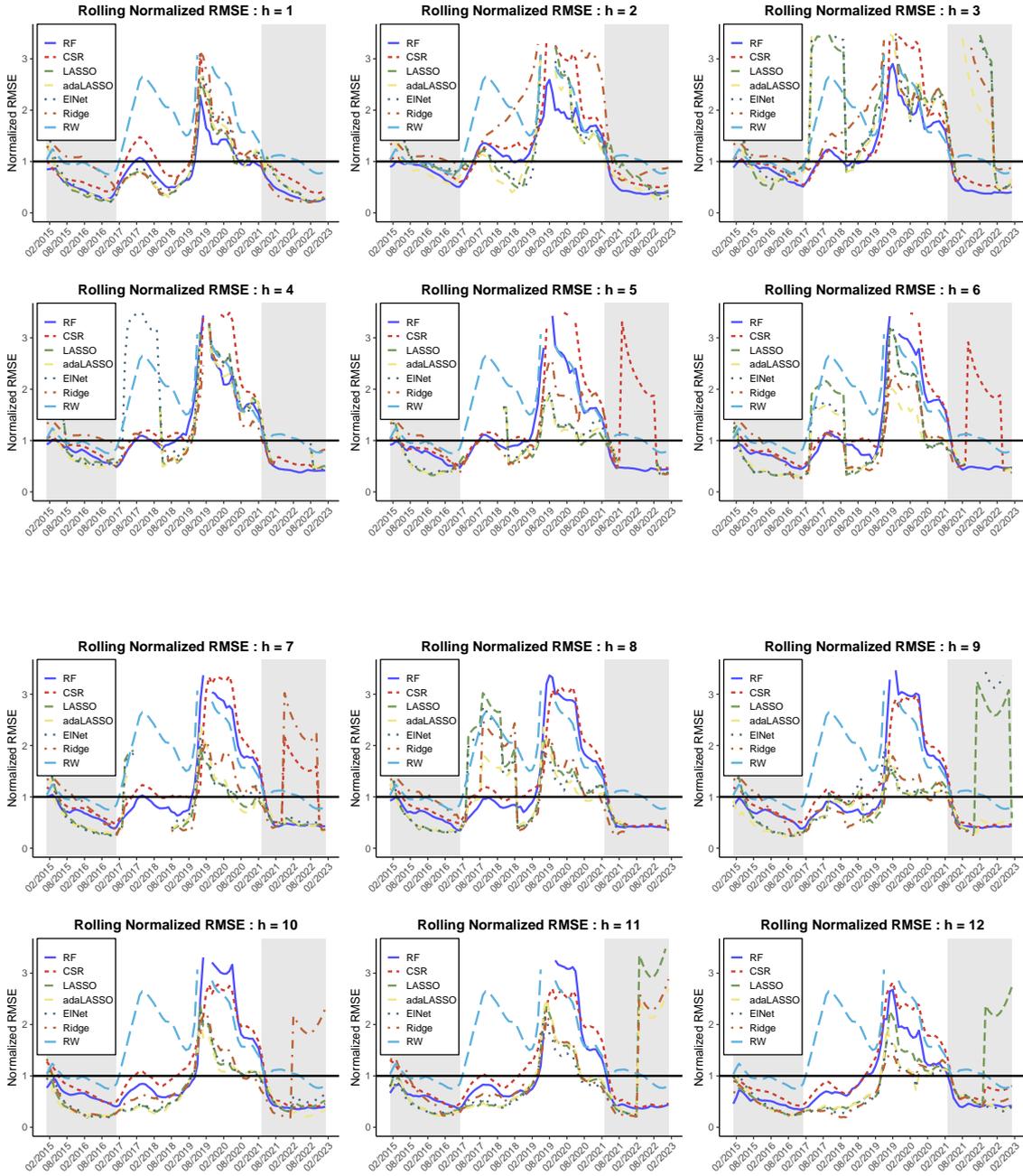
C Expanding Window Results

Table 4 – RMSE, MAE and MAD for expanding window forecasts

| Datasets: X^h, $h = 1, ..., 12$ | | | | | | | | | | | | |
RMSE/(MAE)/ {MAD}	1	2	3	4	5	6	7	8	9	10	11	12
RF	0.45	0.65	0.62	0.61	0.60	0.59	0.59	0.58	0.57	0.55	0.53	0.50
	(0.50)	(0.76)	(0.73)	(0.70)	(0.68)	(0.66)	(0.66)	(0.65)	(0.63)	(0.62)	(0.59)	(0.54)
	{0.56}	{0.92}	{0.90}	{0.93}	{0.89}	{0.74}	{0.78}	{0.76}	{0.69}	{0.73}	{0.69}	{0.66}
CSR	0.78	0.76	0.75	0.75	0.74	0.74	0.74	0.73	0.72	0.71	0.69	0.80
	(0.83)	(0.89)	(0.87)	(0.86)	(0.85)	(0.84)	(0.83)	(0.83)	(0.83)	(0.81)	(0.79)	(0.86)
	{0.93}	{1.13}	{1.05}	{1.00}	{0.97}	{1.01}	{1.02}	{1.06}	{1.00}	{1.00}	{0.94}	{0.83}
LASSO	0.45	0.66	0.63	0.62	2.04	0.51	0.52	1.56	1.64	1.65	0.49	2.31
	(0.50)	(0.77)	(0.73)	(0.73)	(0.92)	(0.57)	(0.54)	(0.72)	(0.69)	(0.64)	(0.48)	(0.78)
	{0.51}	{1.04}	{0.89}	{0.76}	{0.77}	{0.65}	{0.61}	{0.48}	{0.49}	{0.40}	{0.50}	{0.39}
adaLASSO	0.46	0.59	0.59	0.59	0.56	0.50	0.47	0.64	1.48	1.64	0.42	1.72
	(0.52)	(0.69)	(0.68)	(0.69)	(0.65)	(0.56)	(0.50)	(0.52)	(0.64)	(0.64)	(0.45)	(0.70)
	0.49	0.90	0.90	0.83	0.75	0.67	0.55	0.51	0.45	0.43	0.52	0.41
ElNet	0.48	0.68	0.64	0.63	0.64	0.55	0.53	1.63	1.33	1.66	0.47	2.28
	(0.54)	(0.79)	(0.74)	(0.73)	(0.69)	(0.60)	(0.56)	(0.73)	(0.64)	(0.64)	(0.48)	(0.78)
	0.53	0.99	0.89	0.82	0.77	0.66	0.64	0.57	0.53	0.45	0.49	0.41
Ridge	0.58	0.98	0.98	1.22	1.86	1.15	0.92	3.16	2.88	1.62	1.01	1.32
	(0.63)	(1.17)	(1.17)	(1.16)	(1.01)	(0.82)	(0.74)	(1.07)	(0.96)	(0.68)	(0.66)	(0.67)
	0.59	1.50	1.50	1.37	0.92	0.64	0.65	0.62	0.56	0.52	0.54	0.42

The columns represent the datasets. The coloured values correspond to the best result of the models for each error measure.

Table 5 – Descriptive statistics: expanding window forecasts errors

	$\overline{\text{RMSE}}$	$\overline{\text{MAE}}$	$\overline{\text{MAD}}$	max_{RMSE}	max_{MAE}	max_{MAD}	min_{RMSE}	min_{MAE}	min_{MAD}
RF	0.57	0.64	0.77	0.65	0.76	0.93	0.45	0.50	0.56
CSR	0.74	0.84	0.99	0.80	0.89	1.13	0.69	0.79	0.83
LASSO	1.09	0.67	0.62	2.31	0.92	1.04	0.45	0.48	0.39
adaLASSO	0.81	0.60	0.62	1.72	0.70	0.90	0.42	0.45	0.41
ElNet	0.96	0.66	0.64	2.28	0.79	0.99	0.47	0.48	0.41
Ridge	1.47	0.89	0.82	3.16	1.17	1.50	0.58	0.63	0.42
RW	1.13	1.23	1.04	1.13	1.23	1.04	1.13	1.23	1.04

Table 6 – Percentage of the times when the models beat Focus

	1	2	3	4	5	6	7	8	9	10	11	12
RF	80.41	53.61	61.86	68.04	70.10	73.20	74.23	73.20	75.26	74.23	76.29	70.10
CSR	55.67	41.24	49.48	50.52	51.55	49.48	49.48	50.52	52.58	55.67	56.70	65.98
LASSO	73.20	65.98	64.95	67.01	57.73	74.23	75.26	50.52	62.89	69.07	71.13	62.89
adaLASSO	74.23	72.16	72.16	70.10	73.20	74.23	78.35	77.32	62.89	69.07	75.26	62.89
ElNet	72.16	59.79	60.82	64.95	71.13	74.23	69.07	62.89	65.98	71.13	73.20	61.86
Ridge	61.86	25.77	25.77	13.40	37.11	43.30	46.39	47.42	51.55	61.86	57.73	61.86

Percentage of times that 12-month rolling RMSE of a model-dataset pair is smaller than the 12-month rolling RMSE of the Focus consensus. RW beats the benchmark only about 30% of the time.

Rolling Normalized RMSE : h = 7

Rolling Normalized RMSE : h = 8

Rolling Normalized RMSE : h = 9

Rolling Normalized RMSE : h = 10

Rolling Normalized RMSE : h = 11

Rolling Normalized RMSE : h = 12