Revista Colombiana de Estadística - Applied Statistics
July 2025, volume 48, issue 2, pp. 201–226
http://doi.org/10.15446/rce.v48n2.116414

Real-time Forecasting of Colombian Consumer Price Inflation Using High-Dimensional Methods

Pronósticos en tiempo real de la inflación de los precios al consumidor en Colombia con métodos de alta dimensión

Héctor M. Zárate-Solano[a], Miguel A. Manrique-Rodriguez[b]

Econometric Unit, Economic Studies, Banco de la República, Bogotá, Colombia

Abstract

In this paper, we examine the effectiveness of high-dimensional methods for real-time forecasting of inflation in Colombia. We utilize statistical dimension reduction techniques, such as sparse principal components and dynamic factor analysis, alongside machine learning algorithms that incorporate shrinkage methods. Our evaluation of out-of-sample forecasts, using a dataset of 102 macroeconomic and financial indicators, indicates that ensembles of multiple underlying models can enhance forecast accuracy for horizons of 11 and 12 months ahead. Additionally, stepwise models are suitable for horizons between 4 and 10 months ahead, and spectral component models are effective for short horizons.

***Key words*:** Ensembles; Forecasting inflation; Lasso; Machine learning; Penalized regression; Shrinkage.

Resumen

En este documento, se examina la efectividad de los métodos de alta dimensión para pronosticar en tiempo real la inflación en Colombia. Se utilizan técnicas estadísticas de reducción de dimensión, como componentes principales dispersos y análisis factorial dinámico, junto con algoritmos de aprendizaje automático. La evaluación de pronósticos fuera de muestra, utilizando 102 indicadores macroeconómicos y financieros, indica que la síntesis de múltiples modelos subyacentes mejoran la precisión de las predicciones para horizontes de 11 y 12 meses. Además, los modelos secuenciales son adecuados para horizontes de entre 4 y 10 meses, y los modelos de componentes espectrales son efectivos para horizontes cortos.

***Palabras clave*:** Aprendizaje de máquinas; Lasso; Métodos de ensamble; Pronósticos de la inflación; Regresión penalizada.

[a]Ph.D. E-mail: hzaratso@banrep.gov.co
[b]Internship Student. E-mail: usr_practicantegt20@banrep.gov.co

1. Introduction

The inflation forecast targeting framework for monetary policy-making relies on a long-term inflation rate goal. The optimal intermediate tool requires the central bank to compute forecasts to guide its policy by considering the trade-off between deviations of inflation from the target and the output gap. The central bank sets annual inflation targets, and inflation forecasts play a pivotal role in signaling the monetary policy stance, anchoring inflation expectations, and influencing investment decisions, pricing strategies of firms, and wage contracts.

The demand for reliable forecasts has led to the development of various methods, ranging from classical econometrics to economic-based models. These include time series methods such as VAR and ARIMA stochastic processes, structural Phillips curve models, asset price models based on the term structure of interest rates, and survey-based applications. However, despite the significance of accurate inflation forecasts, simple benchmark forecasting models remain among the most competitive options. For more details, refer to the work of Stock & Watson (2012).

Private and public institutions have invested considerable effort into collecting, organizing, and publishing a comprehensive set of economic variables on a regular schedule. However, in situations where datasets contain more variables than observations, classical methods, as noted by Wainwright (2019*a*), can struggle in high-dimensional settings and may not yield reliable predictions. In these cases, suitable statistical models and machine learning algorithms, along with their ensembles, can serve as effective alternatives for forecasting.

The evolution of inflation forecasting could be categorized into four strands: first, reduced inflation variability for most of the sample period. Second, the success of expert and firm surveys, which exhibit high predictive power. Third, the use of forecast combinations. Fourth, incorporation of extensive macroeconomic and financial indicators due to an increasing computational power. Regarding the latter, Kim & Swanson (2018) provided evidence supporting the dominance and advantages of the diffusion index method for inflation forecasting based on high-dimensional data. These results suggest improved forecast accuracy compared to classical models. Additionally, Inoue & Killian (2008) adapted the Bagging method, employing dynamic multiple regression for inflation forecasting, which resulted in significant reductions in prediction mean square error. Garcia et al. (2017) confirmed the strong performance of random subset regression in a real-time forecasting exercise for Brazilian inflation. Other recent forecasting studies using large-scale datasets to forecast inflation include Araujo & Gaglianone (2020), which applied it to Brazilian consumer price inflation at multiple horizons, and Fulton & Hubrisch (2020), which conducted a real-time forecasting exercise for US inflation. They investigated how additional macroeconomic variables, expert judgment, or forecast combinations could improve forecast accuracy and robustness.

From a Bayesian perspective, Gianone et al. (2018) compares sparse and dense representations of predictive models in macroeconomics, particularly when dealing with a large number of potential predictors. The study specifies a prior that accommodates both variable selection and shrinkage. As a result, the posterior distribution tends to focus on a broad range of models, many of which include numerous predictors.

Chernozhukov et al. (2017) classifies the forecasting methodologies with high-dimensional data into two categories: the *sparse modeling* that selects a small set of predictors, and the *dense modeling* that recognizes that all possible variables might be important for prediction. However, an *hybrid* approach that encompasses both sparse and dense approaches favors the low-dimensionality of the models. For more details of this method, see the discussion by Kim & Swanson (2018).

As a summary of the findings in the empirical applications, including large datasets to forecast inflation could improve their accuracy compared to traditional models used in this task. Moreover, most of the methods rely on a two stages strategy. In the first stage, the information content of the big data is extracted through linear or non-linear dimension reduction techniques found in the statistics and machine learning literature. Next, in the second stage, the resulting latent factors are linked to inflation through the linear model's architecture.

Our main contribution lies in advancing macroeconomic forecasting through the implementation of a real-time forecasting exercise using high-dimensional econometric and machine learning models within a data-rich environment characterized by a large number of predictors. We explored various forecasting methods to determine their effectiveness in constructing inflation forecasts. We found that employing hybrid dimensional reduction methods and machine learning offers a superior strategy for achieving more accurate inflation forecasts.

In this paper, we employ statistical high-dimensional reduction methods and machine learning techniques to forecast inflation in real-time.[1] The forecasts are generated using only the information available at the time the forecasts are made. Our approach is based on a modified two-stage procedure. First, we extract the information content from the big data through linear or non-linear dimension reduction techniques found in the statistics and machine learning literature. Then, in the second stage, the resulting latent factors are linked to inflation using the linear model's architecture under penalization. We assess the performance over time by implementing an out-of-sample forecasting exercise using a monthly real-time macroeconomics dataset. Our results provide evidence that dimension reduction is a competitive strategy.

The flow diagram in Figure 1 illustrates the process of forecasting inflation. The process begins with the collection of macroeconomic and financial indicators using web-scraping algorithms on the websites of Banco de la República, the Ministry of Finance, and DANE. Next, we analyze the statistical properties of the time series to ensure stationarity, applying appropriate statistical tests. Afterward, we perform dimensionality reduction using both statistical and machine learning methods. Finally, we generate forecasts with various models and evaluate their performance through out-of-sample exercises.

[1]The machine learning literature (*ML*) is a data-driven approach that addresses non-linear patterns in the data and large volumes of *big data*. Moreover, *ML* comprises two elements: a learning model and an algorithm, enabling it to automate various modeling choices.

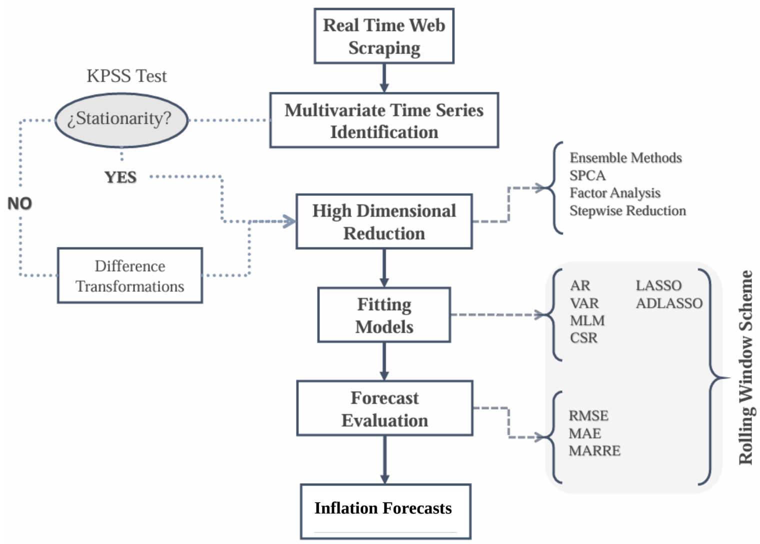


Figure 1: Schematic view to inflation forecasting in a high-dimensional environment.

This paper is organized into five sections, in addition to this introduction. Section 2 provides a descriptive analysis of the high-dimensional macroeconomic indicators dataset related to the Colombian economy. Section 3 offers a concise explanation of the direct forecasting modeling. Section 4 reviews various forecasting models. Section 5 presents a pseudo-out-of-sample forecasting evaluation along with the key results. Finally, Section 6 concludes the paper and suggests directions for future research. Technical results are included in the Appendix.

2. The Colombian Macroeconomic Data

In this section, we provide a comprehensive graphical analysis of the macroeconomic time series' evolution using various statistical methods, including quantile plots and scatter plots. This analysis is based on a dataset of 102 variables obtained from institutional repositories, covering economic activity, labor, and financial indicators. The empirical forecasting study uses monthly data spanning from April 2008 to April 2022 (174 time-periods) with no missing values. Detailed descriptions of the variables and their publication dates are available in Table 7 of the Appendix. The primary focus of the forecasting exercise is on Colombian inflation, as measured by the annual percentage change in the Consumer Price Index (CPI).

Figure 2 illustrates the evolution of inflation and various categories over the sample period. The graphs indicate significant fluctuations in inflation rates. Notably, the highest inflation rate was recorded in July 2016, and there has been a noticeable increase in inflation rates starting from March 2021, indicating a trend change of inflation for the total inflation and its classifications.

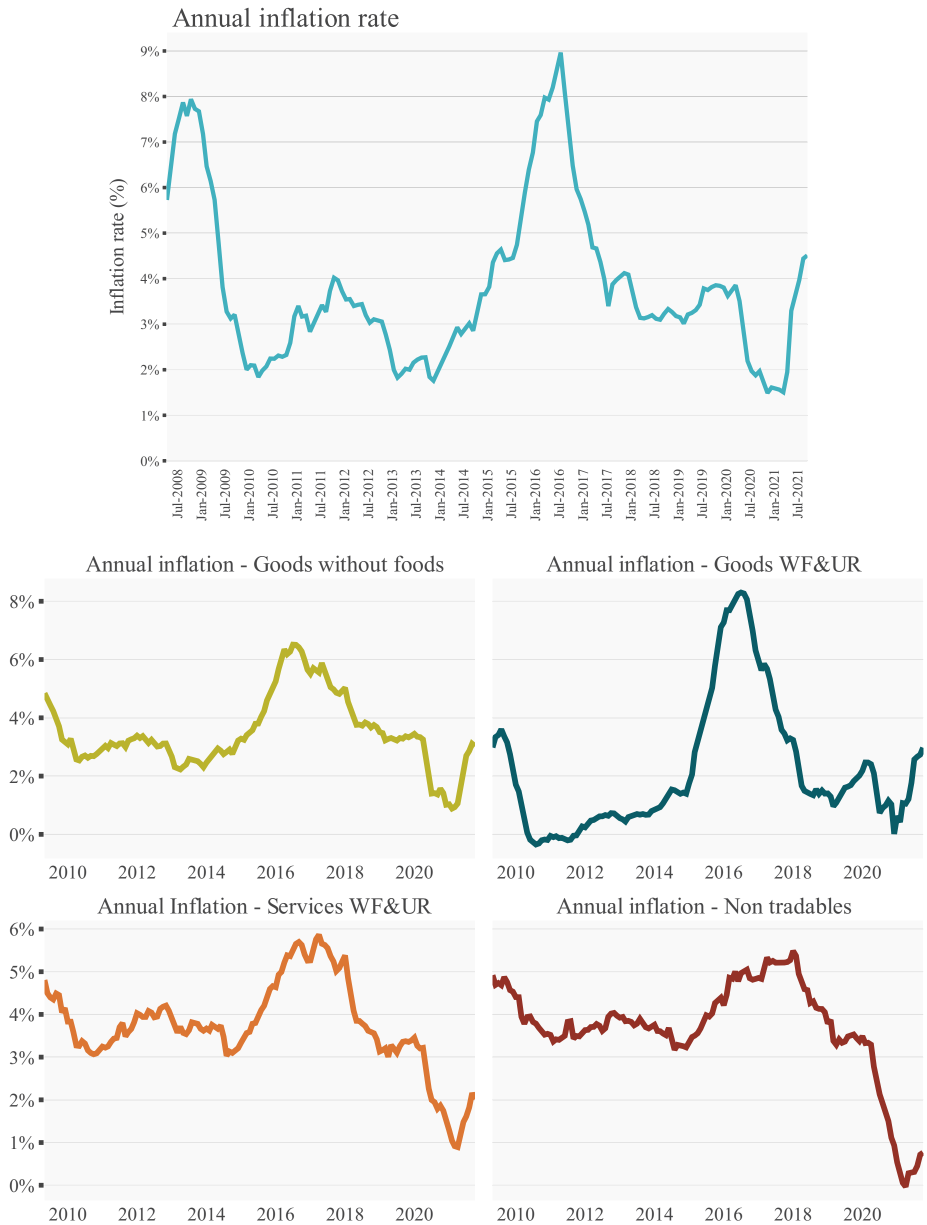


Figure 2: Total inflation and some selected classifications. Banco de la República database. Author's calculations.

On the other hand, in Figure 3 the set of the 102 original macroeconomic time series and their stationary transformations are depicted. The graph exhibited different correlation patterns among groups of indicators, and some outlier observations are identified.

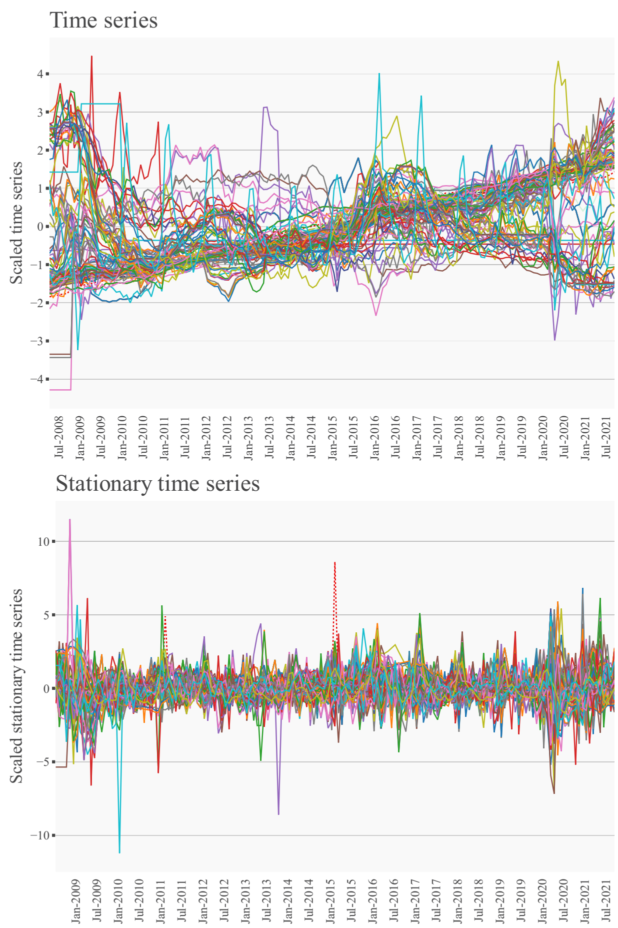


Figure 3: Original and stationary time series standardized with zero mean and unit variance, respectively. Banco de la República database.
Author's calculations.

Quantiles Plots The evolution over time of the macroeconomic indicators is summarized through some selected quantiles of the *dynamic quantile plot* proposed by Peña et al. (2019). This plot has the advantage of preserving the autocorrelation structure of the time series. Thus, the empirical dynamic quantile EDQ solves the following optimization problem:

$$q_q^p = \min \left[\sum_{t=1}^{T} \left(\sum_{z_{it} > y_t} p|z_{it} - y_t| + \sum_{z_{it} < y_t} (1-p)|z_{it} - y_t| \right) \right] \qquad (1)$$

where the series $\{q_q^p\}$ is the $p-th$ empirical dynamic quantile and the observed set of time series is $\mathcal{C}_k = \{z_{it} | 1 < i < 102, 2008.04 < t < 2022.04\}$

Figure 4 shows the EDQ with probabilities 0.25, 0.5, and 0.75, respectively for the original and standardized series. For the original series, the EDQ shows that the series is nonstationary because the three quantiles display an upward trend. Moreover, the interquartile range appear to be larger after the financial crisis of 2008, stable from 2010 to 2016 and larger again after 2016.

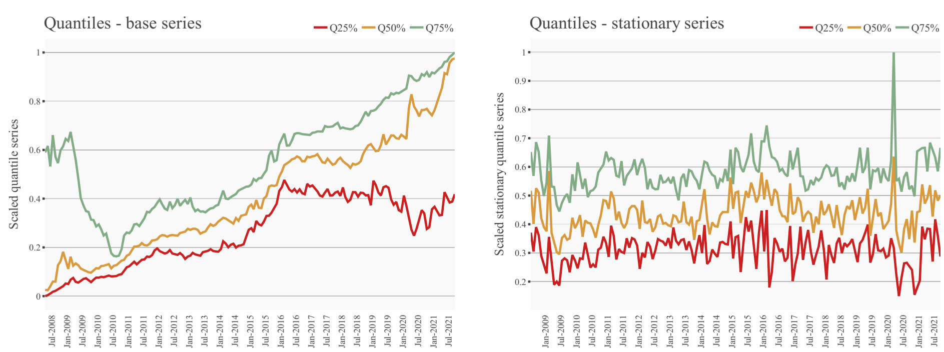


Figure 4: Dynamic quantiles plots for original and standardized series. Banco de la República database. Author's calculations.

Additionally, the joint evolution of the time series over time is also depicted with a bivariate plot of the mean and variance.

The scatterplot in Figure 5 shows dates with high variance, and the range of the means of the monthly standardized variables is substantial.

3. Empirical Forecasting Methodology

In this section, we describe the empirical strategy for real-time forecasting inflation. Thus, forecasts are computed exclusively on the data that were available at the time when the forecasts were made. This strategy avoids the trouble with both data revision issues for some variables and inclusion of the uncertainty that forecasters and policymakers face.

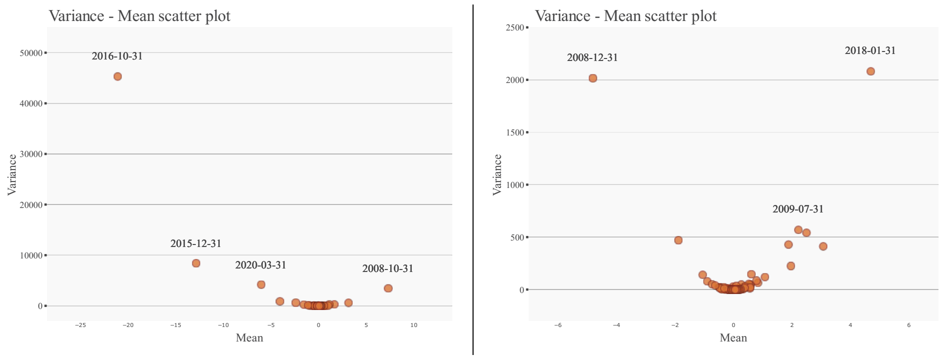


FIGURE 5: Scatter plot of cross-sectional variances and means for the 102 standardized macroeconomic variables.

The forecasting exercise divides the sample period into two sub-periods. The first, known as the *training set* covers the period $(t = 1, \dots T_1)$, from which the parameters of the inflation forecasting models are approximated to reduce the high data dimensionality. Next, the second sub-period defined as the *testing set* considers the period $(T_1 + 1 \cdots T)$. The $T - T_1$ observations are used to measure the accuracy of the forecasting models where the observed inflation rate π_t is compared with the forecasts obtained from the models.

3.1. Direct Forecasting

In this paper, we focus on the direct forecasting procedure, *DMS*,[2] which means that for every forecasting horizon, h, the framework avoids the estimation of a model for the evolution of the predictors. For example, under the linearity assumption, inflation h periods ahead π_{t+h} at period $t+h$ using the information set at period t is given by:

$$\pi_{t+h} = T(\mathbf{x}_t') + \epsilon_{t+h} \tag{2}$$

where $T(\mathbf{x}_t')$ is a linear or non-linear function that combines predictors that comes from a large number of macroeconomic and financial. The $\mathbf{x}_t'$ correspond to a row vector of the set of factors available at time t, and ϵ_{t+h} is the *forecasting error*.

$$\pi_{t+h} = \beta_h' \mathbf{x_t} + \epsilon_{t+h} \tag{3}$$

For each horizon h we estimate a different vector of unknown parameters β_h. by using the observations in the training set, the inflation forecasts are given by

$$\hat{\pi}_{t+h} = \hat{\beta_{\mathbf{h}}}' \mathbf{x_t} \quad h = 1, \dots 12 \tag{4}$$

where $\hat{\beta_h}$ is the estimated vector of parameters based on the training set and h is the forecasting horizon.

[2] The forecasts are obtained by using a horizon-specific model which is known as the multi-period model with a loss function to the forecast horizon. This strategy contains variables that are observed and available in time t. The method is robust to model specification because the tradeoff between bias and variance is reduced in this case. Marcellino et al. (2006)

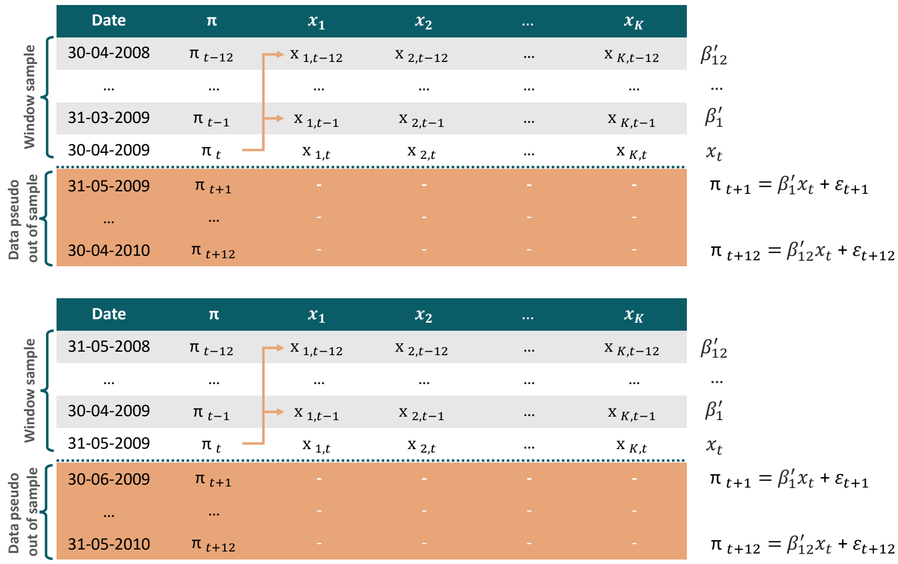


FIGURE 6: Schematic representation of the direct forecasting approach.The rolling window scheme set aside the observations for the last four years to conform the testing set to compute the evaluation statistics at different time horizons $h = 1, \dots, 12$. The competing model fitting is computed in the training period. The process is repeated recursively.

3.2. Augmented Difussion Index Method

In this section, we describe the framework to build the inflation forecasts. First, we rely on the diffusion index method, which is based on a two-step framework. The details of the empirical performance are found in Kim & Swanson (2014) and Stock & Watson (2012)[3]. After that, we augment diffusion index methods with machine learning techniques to enhance the forecasting performance.[4] Movement of the large number of macroeconomic indicators can be driven by a small number of unknown factors that can be used to forecast inflation.

The first stage considers dimension reduction of the data. Let X_{it} the observed data form the unit of i^{th} cross section at time t, for $i = 1 \dots N$ and $t = 1 \dots T$. Consider the following model.

$$X_{it} = \lambda_i' \mathbf{F_t} + \epsilon_{it} \tag{5}$$

where $\mathbf{F_t}$ is the not observable vector of factors λ_i is a vector of non observable coefficients or factor loadings associated with $\mathbf{F_t}$ and ϵ_{it} is the idiosyncratic component of X_{it}. On the other hand, the machine learning techniques are substitutes

[3]The diffusion index method in forecasting utilizes latent factors computed from high-dimensional data as inputs in the specification of subsequent forecasting models.

[4]The idea is to smooth the extracted factors including coefficients penalization related to the factors.

to reduce the dimension by implementing ensembles algorithms. The second stage, consider the forecasting equation for inflation.

$$\pi_{t+h} = \alpha'\mathbf{F_t} + \hat{\beta}'\mathbf{W_t} + e_t \quad h = 1, \dots 12 \tag{6}$$

The variables $\mathbf{W_t}$ are observable and could be added depending the statistician information set.

In summary, we first estimate $\mathbf{F_t}$ and get $\hat{\mathbf{F}}_\mathbf{t}$, next, we can regress π_{t+1} on $\hat{\mathbf{F}}_\mathbf{t-h}$ and $\mathbf{W_{t-h}}$ to obtain the coefficients $\hat{\alpha}$ and $\hat{\beta}$, from which a forecast is formed by

$$\pi_{t+h/T} = \hat{\alpha}'\hat{\mathbf{F}}_\mathbf{t} + \hat{\beta}'\mathbf{W_T} \tag{7}$$

The dimension of F could also be large given their structure of correlation and cross-correlation. We use regularization techniques to penalize the effects of the factors.

4. Forecasting Models

In this section, we briefly outline the forecasting methods employed, emphasizing the importance of selecting relevant variables to effectively analyze high-dimensional data sets. High-dimensional time series present challenges for analyzing and forecasting large dependent data due to the curse of dimensionality and the risk of overfitting. We discuss several reduction methods grounded in statistical and machine learning principles, which aim to ensure optimal prediction accuracy and stability. These methods include traditional economic models like the Phillips curve, subspace methods that reduce the number of predictors, shrinkage or regularization techniques such as LASSO, and machine learning algorithms, particularly ensemble methods.

4.1. Traditional Statistical and Economic Reference Models

4.1.1. VAR Models

Vector autoregressions and their extensions has been adopted for macroeconomic forecasting offering competitive forecasts despite their simple formulation. In addition, VAR models are used as the benchmark for comparing the forecast performance of the new methods.

$$\mathbf{z_t} = \phi_0 + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + \mathbf{e}_t \tag{8}$$

Where $t = p+1, \dots T$ and Σ_e is the covariance matrix of the random error. There are $T-p$ observations for the estimation. The matrix representation is given by:

$$\mathbf{Z} = \mathbf{X}\beta + A$$

where

$$\mathbf{Z_t} = (Z_{1t}, \dots Z_{mt})'$$

Therefore,

$$\mathbf{vec}(\mathbf{Z}) = (I_k \otimes X)vec(\beta) + vec(A)$$

Where the covariance matrix of $vec(A)$ is given by $\Sigma_e \times I_{t-p}$. A $VAR(p)$ model could be estimated by using: least squares, maximum likelihood or Bayesian methods. In this forecasting exercise, we follow this VAR methodology to obtain forecasts after the reduction dimension.

4.1.2. Phillips Curve

From a marginal perspective, the Phillips curve has been a key contender for predicting inflation. In this paper, we estimate a backward-looking version of this curve based on the framework established by Silva & Gaglianone (2020). This specification includes three determinants: inflation inertia, approximated by lagged inflation; the pass-through channel, measured by imported inflation; and the monetary policy channel, which is related to aggregate demand and approximated by the output gap. The model specification is as follows:

$$\pi_{t+h} = \alpha_0 + \alpha_1 \pi_t + \alpha_2 \pi_t^{imp} + \alpha_3 g_t + \epsilon_{t+h} \tag{9}$$

where π_{t+h} represents the h-step-ahead inflation, π_t denotes the current inflation rate, and g_t is the output gap, which is approximated using a filter applied to the monthly economic activity index.

4.2. Dimensionality Reduction

Dimensionality reduction encompasses techniques designed to decrease the number of input predictors or features in the training set, where the selection or shrinkage of relevant variables is crucial for managing high-dimensional data in inflation forecasting. For instance, Kim & Swanson (2018) presents evidence supporting factor-based dimension reduction for macro-econometric forecasting. The empirical literature has implemented two primary approaches:

- Statistical Methods: These involve eigenvalue analysis of the covariance matrix, with methods such as sparse principal components analysis and dynamic factor models emerging from this strand.
- Machine Learning Algorithms: These algorithms focus on constructing ensembles that combine various methodologies, thereby eliminating irrelevant factors and reducing the complexity of the original problem.

4.2.1. Statistical Reduction Methods

With the availability of massive datasets for economic analysis, classical theory may struggle to provide accurate predictions when the number of variables is equal to or exceeds the length of the sample period. We will briefly outline several dimension-reduction techniques that could help address this challenge.

4.2.2. Sparse Principal Components

Classical principal component analysis,PCA is a standard technique used for exploratory analysis, estimation in mixture models, and dimension reduction. The objective is to find a low-dimensional subspace that captures the majority of the variance. This subspace is spanned by the top r eigenvectors of the population covariance matrix, $\mathbf{\Sigma}$. However, in practice, the population covariance matrix is unknown, and the practitioners rely on the sample covariance matrix $\hat{\mathbf{\Sigma}}$, which is formed with the sample data. Thus, The purpose of principal component analysis (PCA) is to find a low dimensional approximation by which most of the variation is retained generating minimum information loss.

Classical PCA analysis requires the sample size, n be larger than the dimension d to perform well. Moreover, in a high-dimensional setting where the sample size is lower than the potential predictors, $n < d$, there is no method to obtain both consistent and accurate estimators of the population eigenvectors. However, empirical work is adequate to impose some structure on eigenvectors to develop effective estimators. The simple structure is that of sparsity in the eigenvectors. see Wainwright (2019*b*) for a complete discussion of this topic.

The *sparse principal components analysis* relies on the basis that many regressors are similar to one another and are highly correlated. $SPCA$ aids in the interpretation of principal components by placing *zero* restrictions on various factor loading coefficients. Zou et al. (2006) develop a regression optimization framework. Namely, they consider $\mathbf{X}$ as dependent variables, F as explanatory variables, and loadings as coefficients. They then use the lasso (and elastic net) to derive a sparse loading matrix. Therefore, Sparse Principal components ($SPCA$) decipher the relationship between principal components and the observed data by introducing sparsity in the loadings factors (Kim & Swanson, 2018).

Using the connection between PCA and regression, and using the lasso approach to produce sparse loadings ("regression coefficients"). The following optimization problem is solved

$$(\boldsymbol{A}, \boldsymbol{B}) = \min_{\boldsymbol{A}, B} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} ||X_i - \boldsymbol{A}\boldsymbol{B}'||^2 + \lambda \sum_{j=1}^{k} ||\beta_j||^2 + \sum_{j=1}^{k} \lambda_{1,j} ||\beta_j|| \tag{10}$$

$$\textit{subject to } \boldsymbol{A}'\boldsymbol{A} = I_{k\times k}$$

4.2.3. Factor Analysis

They can represent the dynamic evolution with small numbers of parameters. Thus, the dynamics are driven by factors whereas the noises could be assumed white noise. The dynamic factor model can be expressed by:

$$\boldsymbol{z}_t = \boldsymbol{P}\, F_t + \boldsymbol{\eta_t} \tag{11}$$

where $\boldsymbol{z}_t = (z_{1t}, \ldots z_{kt})$ stationary time series each time series centered. $\boldsymbol{P}$ is the loading matrix, F_t are the common unobserved factors, and $\boldsymbol{\eta_t}$ is the noise or idiosyncratic component. In high-dimensional settings, the assumptions underlying model identification rely on the idiosyncratic component, which may exhibit weak serial and cross-correlations. Additionally, the idiosyncratic component is assumed to be uncorrelated with the factors. Thus, $E[F_t, \eta_t] = 0$. The available information to estimate a *DFM* is contained in the covariance matrix of the series. See (Peña et al., 2019) for details.

4.2.4. Machine Learning Methods

Machine learning algorithms are classified as data-driven methods that avoid assumptions and hypotheses about the underlying statistical relationship in the data. To construct inflation forecasts, we rely on *supervised learning* methods based on a *learning method* that determine the best fit for the predictors and an *algorithm* which models the relationship between the predictors and inflation. Thus, inflation forecasts are built based on model selection techniques such as complete subset and stepwise regression . Additionally, *overfitting* is avoided by using shrinking or regularized techniques. These introduce penalization through the lasso estimator and its variants. We also rely on constructiing *ensemble* methods that combine a myriad of these algorithms. Inoue & Killian (2008) provide an application using *Bagging*.

4.2.5. Complete Subset Regression

The complete subset regression, *CSR*, is a model selection method to estimate a low-dimensional approximation of the original model, in which the forecasts are built from all possible linear regression models with a fixed number of predictors ν. Both & Nibbering (2020) review the *CSR* methodology applied to forecasting.

However, forecasting the inflation rate h periods ahead π_{t+h} by testing all possible combinations of the 102 potential predictors could be an infeasible procedure given its computational intractability in high dimensional settings. Elliot et al. (2015) introduced the *CSR*, which allows selecting an optimal number ν of predictors $\nu < K$ based on the covariance matrix. Thus, the number of possible subset regressions is reduced to

$$\binom{K}{\nu} = \frac{K!}{\nu(K-\nu)!}$$

The *CSR* set-up computes the average of the coefficients over least squares estimates across all ν-dimensional subsets regressions to provide an estimator $\hat{\beta}$ for forecasting.

4.2.6. Stepwise Regression

In this approach, a subset of the predictors is retained and the rest are eliminated from the model. The strategy to choose the subset of the method starts with the intercept, and then sequentially adds into the model the predictor that most improve the fit. The improvement in fit is based on the F statistic. Therefore, the strategy is to add in the predictor producing the largest value of F, stopping when no predictor produces an F-ratio greater than the 90th percentile of the expected F distribution. Thus, the forecasting model is selected from the sequence that minimizes the forecasting error. Hastie et al. (2001) described precisely this method.

4.2.7. Lasso Estimator and Extensions

This sort of estimator is established on the penalized regression framework, which is an estimation method that minimizes a combination of the usual least-squares objective function and the penalty function $\boldsymbol{p}$.

$$\hat{\boldsymbol{\beta}} = minimize \ \ ||\boldsymbol{y} - \boldsymbol{X\beta})||^2 + \lambda \, \mathbf{p}(\boldsymbol{\beta}, \alpha, Z) \tag{12}$$

where $\lambda > 0$ is a tuning parameter and $\boldsymbol{p}$ is the penalty function that penalizes the entries of β for being $\neq 0$ and α is a dimensional parameter. Therefore, given a collection of N macroeconomic variables-inflation pairs, the lasso finds the solution to the optimization problem known as the Least Absolute Shrinkage Selection Operator (Lasso), which considers the penalty function as $\mathbf{p}(\boldsymbol{\beta}, \alpha, Z) = \sum_{i=1}^{k} |\beta_i|$ that only depends on β. The estimation and variable selection is done in one step.

$$minimize \ \ \left[\frac{1}{2N} \sum_{i=1}^{N} (\pi_i - \boldsymbol{X\beta})^2 + \lambda ||\beta|| \right]$$

Moreover, adaptive Lasso is a variant based on the following the penalty function: $\boldsymbol{p}(\boldsymbol{\beta}, \alpha, Z) = \sum_{i=1}^{k} \frac{1}{|\beta_{OLS,i}|} |\beta_i$ Details can be found in Kock & Vasconcelos (2020).

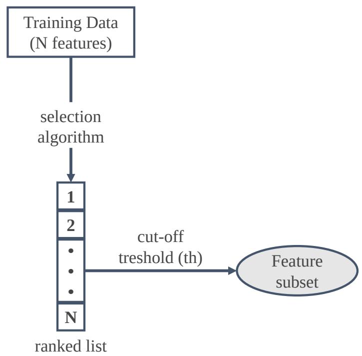
4.3. Ensemble Methods

Ensembles are structured as a model that represents a combination of multiple underlying models. Thus, the output of the ensemble is the combination of inflation forecasts from much more simple models. As a result, the *ensemble* is related to the combination of different variables that seek to eliminate irrelevant factors and to reduce the dimensionality and complexity of the original problem. Under this setup, there are different techniques to detect the most relevant characteristics for forecasting inflation and to choose the appropriate method.[5]

[5]For example, although heterogeneous methods apply different selection algorithms to the same dataset, homogeneous methods use the same algorithm for different disturbing versions of the data. this is issue is related to the stability requirement. Severeal functions appear to control the ensambling.

For example: *Filters* minimizes redundancy given the mutual correlation among predictors. *Wrappers* optimize the execution of a given classifier. The learning algorithm evaluates the merit of each subset of variables.

The selection process is based on the ability of the algorithms to assign weights of the variables. In this paper, we rely on regularization techniques, which are based on penalizing the sum of squares of the model with a roughness penalty on the following form: $SCRP(f, \lambda) = SCR(f) + \lambda J(f)$, where λ is the penalty measure. If there is no penalty term, $\lambda = 0$ is related to *interpolation*, and if $\lambda = \infty$ it is referred to *smoothing*. Two schematic representations of ensemble methods are presented in Figure 7.



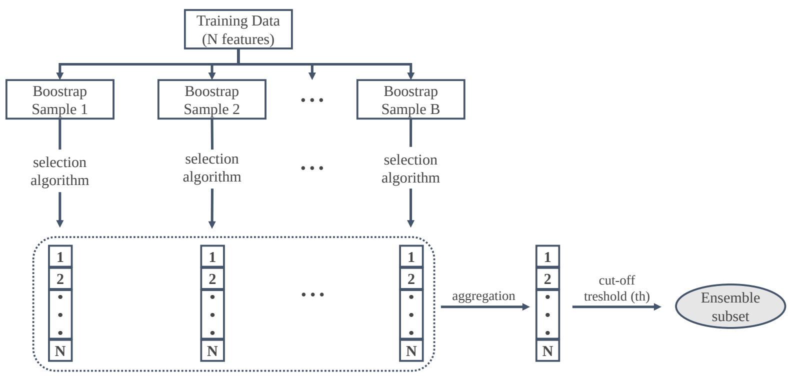


FIGURE 7: Diagrams to build the forecasting ensemble environment. Author's elaborations.

The Figure A1 in the Appendix shows the similarity matrix for one step lagged selected variables for annual inflation rate forecast. Table 1 present the main results of the ensemble algorithms.

Table 1: How High Dimensionality Reduction Methods (HDRM) did select features.

HDRM	Selected Variables / Components	Method / Model	Criterion
ENSEMBLE	V1, V28, V45, V58, V79, V83, V87, V96	OWNLAGS	–
CCF	–	–	–
GLASSO	–	–	–
MI	–	–	–
RF	–	Majority	–
DFMPC	V1, F1, F2, F3, F4, F5	Caro, Peña (2020)	–
SPCA	V1, PC1, PC2, PC3, PC4, PC5, PC6, PC7	PENALTY	Explained Variance (First 7)
STEPWISE	V1, V79, V83, V87, V45, V52, V49, V43	–	AIC (First 15)

Note: Vi refers to the variable list, with names shown in the annex. The term Fi represents factors, and PCi stands for principal components. CCF: Cross Correlation Function, GLASSO: Graphical Lasso, MI: Mutual Information, RF: Random Forest, AIC: Akaike Information Criterion.

5. Forecast Evaluation and Results

In this section, we describe an out-of-sample forecasting exercise to evaluate the performance of the competing inflation forecasting models described above. The criteria for assessing the forecast accuracy relies on forecast error statistics computed on the *testing* period. Thus, we used a scheme that selects a period sample, from which we implement each model to produce point h-step ahead forecasts. Next, we repeat the processes to conform multiple out-of-sample forecasts according to every forecast horizon.

Error measurements based on absolute or squared errors, such as the Mean Absolute Error (MAE) and the Root Mean Squared Error ($RMSE$), evaluate the distance of the observations from the forecasts. However, these statistics are scale-dependent, meaning the forecasting error metrics are influenced by the units of the underlying data. To address this, we incorporate a scale-independent metric based on percentage error metrics. This allows for the comparison of forecast performance across different inflation categories using an alternative measure known as the Mean Absolute Relative Risk Error ($MARRE$).

The findings are summarized as follows: Table 2 presents the Root Mean Squared Error (RMSE). The rows of Table 2 detail the high-dimensional forecast models used during the sample period from April 2008 to April 2022, including ensemble, stepwise, dynamic factor, and spectral principal components models. The columns indicate the horizon at which each model is evaluated. Looking at the forecast performance for each model, we conclude that according to the ordering based on the $RMSE$, the high dimension reduction based on dynamic factors combined with the LASSO penalization tends to outperform the other models in the horizons $h = 1$ to $h = 3$. Forecasting horizons between four and ten months ahead prefer the stepwise-lasso method. The ensemble methods are better for forecasting eleven and twelve horizons.

Tables 3 and 4 in the appendix present the main results of the evaluation, using two forecasting error statistics: Mean Absolute Error (MAE) and Mean Absolute Relative Error (MARRE). In addition, Tables 5 and 6 contain the Giacomini-White test, which compares the predictive accuracy of pairs of forecasting models across different time horizons.

TABLE 2: RMSE evaluation for 12 steps ahead forecasts.

HDR METHOD	MODEL	METRICS	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12
ENSEMBLE	ADLASSO	RMSE	0.007	0.011	0.011	0.011	0.009	0.009	0.009	0.010	0.010	0.009	**0.010**	**0.010**
ENSEMBLE	AR	RMSE	0.020	0.019	0.018	0.016	0.016	0.015	0.015	0.014	0.014	0.012	0.012	0.012
ENSEMBLE	CSR	RMSE	0.007	0.011	0.012	0.012	0.010	0.010	0.010	0.011	0.010	0.011	0.011	0.011
ENSEMBLE	LASSO	RMSE	0.007	0.011	0.011	0.011	0.009	0.009	0.009	0.009	0.010	0.009	**0.010**	**0.010**
STEPWISE	ADLASSO	RMSE	0.007	0.011	0.011	**0.010**	**0.008**	**0.008**	**0.008**	**0.008**	**0.008**	**0.008**	0.009	0.009
STEPWISE	AR	RMSE	0.020	0.019	0.018	0.016	0.016	0.015	0.015	0.014	0.014	0.012	0.012	0.012
STEPWISE	CSR	RMSE	0.009	0.011	0.012	0.012	0.011	0.010	0.011	0.011	0.011	0.011	0.011	0.011
STEPWISE	LASSO	RMSE	0.007	0.011	0.011	**0.010**	**0.008**	**0.008**	**0.008**	**0.008**	0.009	**0.008**	0.009	0.009
DFMPC	ADLASSO	RMSE	0.007	**0.010**	0.011	0.011	0.010	0.009	0.010	0.010	0.010	0.009	**0.010**	**0.010**
DFMPC	AR	RMSE	0.020	0.019	0.018	0.016	0.016	0.015	0.015	0.014	0.014	0.012	0.012	0.012
DFMPC	CSR	RMSE	0.008	0.011	0.011	**0.010**	0.009	0.009	0.010	0.010	0.010	0.011	0.011	0.011
DFMPC	LASSO	RMSE	**0.006**	**0.010**	**0.010**	0.011	0.010	0.010	0.009	0.010	0.010	0.010	**0.010**	**0.010**
SPCA	ADLASSO	RMSE	**0.006**	**0.010**	0.011	0.011	0.010	0.009	0.009	0.009	0.009	0.009	0.009	0.009
SPCA	AR	RMSE	0.020	0.019	0.018	0.016	0.016	0.015	0.015	0.014	0.014	0.012	0.012	0.012
SPCA	CSR	RMSE	0.011	0.011	**0.010**	**0.010**	0.009	0.010	0.010	0.011	0.011	0.011	0.011	0.011
SPCA	LASSO	RMSE	**0.006**	**0.010**	0.011	0.011	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
NA	PHC	RMSE	0.029	0.029	0.033	0.035	0.034	0.039	0.042	0.036	0.032	0.040	0.033	0.038

Authors computations based on the best orecasting models

Finally, Figure 8 presents the forecasts plot from some alternative models. Thus, according to *LASSO*, the predicted annual inflation will increase steeply during the remaining months of 2022, the year in which the inflation rate will be 6.5 percent in December.

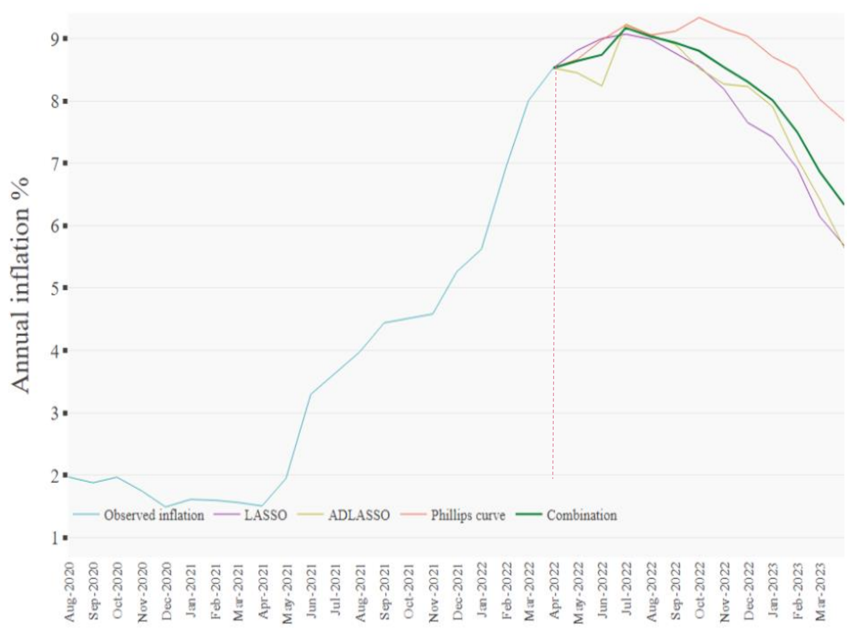


FIGURE 8: Monthly inflation and inflation forecasts for 12 steps ahead. Banco de la República database. Author's calculations.

6. Conclusions

Forecasting accurately the inflation rate at various horizons is one of the most valuable applications for monetary policy and investment decisions. In this paper, we perform a real-time forecasting exercise with high-dimensional econometric models in a data-rich environment. The number of predictors is large. Specifically, We outline a number of methods developed in both machine learning and statistical basis. We examined which methods are useful for constructing inflation forecasts based on principal components analysis, and dynamic factors analysis in conjunction with shrinkage algorithms.

The forecast comparison is based mainly on the smaller root mean forecast error and other hypothesis tests usually found in time series applications. The findings suggest that combining statistical models with learning and shrinkage methods dominates the predictive accuracy of short-term forecasts. indicates that ensembles of multiple underlying models can enhance forecast accuracy for horizons of 11 and 12 months ahead. Additionally, stepwise models are suitable for horizons between 4 and 10 months ahead, and spectral component models are effective for short horizons. up to the one-quarter horizon.

Moreover, the forecasting experiments indicate that the proposed strategy can identify subsets of explanatory variables or latent factors to provide alternative methods to empirically evaluate the predictive accuracy of macroeconomic forecasting of various linear and non-linear models. Recent issues are proposed for future research, such as the Bayesian forecasting framework and structural breaks. Moreover, we should seek to expand more complex models by including versions of economic models such as the Phillips curve.

Acknowledgements

The views expressed in this document are solely those of the authors and do not necessarily reflect the opinions of Banco de la República. We appreciate the contributions of participants at the 42nd International Symposium on Forecasting held in Oxford, England, as well as those from the XIX internal seminar at Banco de la República. We would also like to thank the anonymous referees from the Colombian Journal of Statistics for their feedback, Dylan Torres for his statistical assistance, and José Celis for his help in editing this document.

[Received: August 2024 — Accepted: January 2025]

References

Araujo, G. & Gaglianone, W. (2020), 'Machine learning methods for inflation forecasting in brazil: new contenders versus classical models', *Central Bank of Barzil Working Paper* pp. 1–39.

Both, T. & Nibbering, D. (2020), *Subspace methods*, Springer, chapter 9, pp. 267–326. In: Macroeconomic Forecasting in the Era of Big Data: Theory and Practice.

Chernozhukov, V., Hansen, C. & Liao, Y. (2017), 'A lava attack on the recovery of sums of dense and sparse signals', *Annals of Statistics* **45**, 37–76.

Elliot, G., Gargano, A. & Timmermann, A. (2015), 'Complete subset regression with large dimensional sets of predictors', *Journal of Economics Dynamics and Control* **54**(54), 86–110.

Fulton, C. & Hubrisch, K. (2020), 'Forecasting us inflation in real time', *Finance and Economics Discussion Series 2021-014. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2021.014.* pp. 1–30.

Garcia, M., Medeiros, M. & Vasconcelos, G. (2017), 'Real-time inflation forecasting with high-dimensional models: The case of brazil', *International Journal of Forecasting* **33**(3), 679–693.

Gianone, D., Lenza, W. & Primiceri, G. (2018), 'Economic predictions with big data: The illusion of sparsity', *Federal Reserve Bank of New York Staff Reports, no. 847* pp. 1–26.

Hastie, T., Tishbirani, R. & Friedman, J. (2001), *The elemnts of statistical learning: data mining, inference, and prediction*, 2 edn, Springer.

Inoue, A. & Killian, L. (2008), 'How useful is bagging in forecasting economic time series? a case study of u.s. consumer price inflation', *Journal of the American Statistical Association* **103**(482), 511–522.

Kim, H. & Swanson, N. (2014), 'Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence', *Journal of econometrics* (178), 352–367.

Kim, H. & Swanson, N. (2018), 'Mining big data using parsimonous factor, machine learning, variable selection and shrinkage methods', *International Journal of Forecasting* **34**(2), 339–354.

Kock, A., M. M. & Vasconcelos, G. (2020), *Penalized Time Series Regression*, Springer, chapter 7, pp. 193–277. In: Arango, L. and Hamann, F. ed., (2020). Macroeconomic Forecasting in the Era of Big Data: Theory and Practice.

Marcellino, M., Stock, J. & Watson, M. (2006), 'A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series', *Journal of econometrics* **135**(135), 499–526.

Peña, D., Tsay, R. & Zamar, R. (2019), 'Empirical dynamic quantiles for visualizating of high-dimensional time series', *Technometrics* **61**, 429–444.

Silva, G. & Gaglianone, W. (2020), 'Machine learning methods for inflation forecasting in brazil. new contenders versus classical models', *Journal of Business and Economic Statistics* **30**(30), 481–493.

Stock, J. & Watson, M. (2012), 'Generalized shrinkage methods for forecasting using many predictors', *Journal of Business and Economic Statistics* **30**(30), 481–493.

Wainwright, M. (2019*a*), *Introduction*, Cambridge Series in Statistical and Probabilistic Mathematics, chapter 1, pp. 236–257. In: High-Dimensional Statistics: A Non-Asymptotic Viewpoint.

Wainwright, M. (2019*b*), *Principal component anlysis in high-dimensions*, Cambridge Series in Statistical and Probabilistic Mathematics, chapter 8, pp. 236–257. In: High-Dimensional Statistics: A Non-Asymptotic Viewpoint.

Zou, H., Hastie, T. & Tibshirani, R. (2006), 'Sparse principal component analysis', *Journal of Computational and Graphical Statistics* **15**(2), 265–286.

Appendix. Out-of-Sample Evaluation

Table A1: MARRE evaluation for 12 steps ahead forecasts.

HDR METHOD	MODEL	METRICS	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12
ENSEMBLE	ADLASSO	**MARRE**	0.058	0.100	0.125	0.146	0.157	0.191	0.166	0.239	0.214	0.189	0.191	0.218
ENSEMBLE	AR	**MARRE**	17.414	16.625	16.683	15.729	16.699	17.173	16.373	17.269	15.384	14.385	14.611	13.834
ENSEMBLE	CSR	**MARRE**	0.077	0.122	0.163	0.186	0.183	0.232	0.233	0.301	0.279	0.182	**0.161**	0.172
ENSEMBLE	LASSO	**MARRE**	0.054	0.098	0.124	0.141	0.154	0.192	0.164	0.234	0.209	0.187	0.188	0.214
STEPWISE	ADLASSO	**MARRE**	0.059	0.097	0.116	0.131	**0.129**	0.159	**0.126**	**0.191**	0.201	0.155	0.168	0.167
STEPWISE	AR	**MARRE**	17.414	16.625	16.683	15.729	16.699	17.173	16.373	17.269	15.384	14.385	14.611	13.834
STEPWISE	CSR	**MARRE**	0.132	0.126	0.161	0.182	0.196	0.205	0.198	0.265	0.211	0.178	0.164	0.168
STEPWISE	LASSO	**MARRE**	0.056	0.096	0.116	**0.130**	**0.129**	**0.155**	**0.126**	**0.191**	0.198	**0.154**	0.163	**0.164**
DFMPC	ADLASSO	**MARRE**	0.059	0.093	0.115	0.143	0.166	0.212	0.199	0.262	0.212	0.198	0.186	0.198
DFMPC	AR	**MARRE**	17.414	16.625	16.683	15.729	16.699	17.173	16.373	17.269	15.384	14.385	14.611	13.834
DFMPC	CSR	**MARRE**	0.095	0.144	0.139	0.146	0.158	0.176	0.212	0.272	0.290	0.175	0.175	0.176
DFMPC	LASSO	**MARRE**	0.058	0.092	0.112	0.143	0.163	0.235	0.178	0.282	0.222	0.198	0.186	0.198
SPCA	ADLASSO	**MARRE**	0.054	0.091	0.112	0.138	0.158	0.196	0.176	0.273	**0.176**	0.175	0.182	0.187
SPCA	AR	**MARRE**	17.414	16.625	16.683	15.729	16.699	17.173	16.373	17.269	15.384	14.385	14.611	13.834
SPCA	CSR	**MARRE**	0.188	0.136	0.129	0.144	0.146	0.236	0.189	0.277	0.314	0.172	0.166	0.169
SPCA	LASSO	**MARRE**	**0.053**	0.090	**0.111**	0.137	0.156	0.196	0.177	0.275	0.210	0.174	0.183	0.187
NA	PHC	**MARRE**	37.226	37.630	41.736	43.662	42.289	48.966	53.552	44.336	39.073	50.052	39.677	45.476

TABLE A2: MAE evaluation for 12 steps ahead forecasts.

HDR METHOD	MODEL	METRICS	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12
ENSEMBLE	ADLASSO	MAE	**0.003**	0.006	0.007	**0.007**	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.008
ENSEMBLE	AR	MAE	0.016	0.016	0.014	0.013	0.012	0.012	0.012	0.011	0.011	0.010	0.010	0.009
ENSEMBLE	CSR	MAE	0.004	0.006	0.008	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
ENSEMBLE	LASSO	MAE	**0.003**	0.006	0.007	**0.007**	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.008
STEPWISE	ADLASSO	MAE	**0.003**	0.006	0.007	**0.007**	**0.006**	**0.005**	**0.005**	**0.006**	**0.006**	**0.006**	**0.006**	**0.006**
STEPWISE	AR	MAE	0.016	0.016	0.014	0.013	0.012	0.012	0.012	0.011	0.011	0.010	0.010	0.009
STEPWISE	CSR	MAE	0.005	0.007	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
STEPWISE	LASSO	MAE	**0.003**	0.006	0.007	**0.007**	**0.006**	**0.005**	**0.005**	**0.006**	**0.006**	**0.006**	**0.006**	**0.006**
DFMPC	ADLASSO	MAE	**0.003**	**0.005**	0.007	**0.007**	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
DFMPC	AR	MAE	0.016	0.016	0.014	0.013	0.012	0.012	0.012	0.011	0.011	0.010	0.010	0.009
DFMPC	CSR	MAE	0.005	0.007	0.007	**0.007**	0.007	0.007	0.007	0.007	0.007	0.008	0.008	0.008
DFMPC	LASSO	MAE	**0.003**	**0.005**	**0.006**	**0.007**	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
SPCA	ADLASSO	MAE	**0.003**	**0.005**	**0.006**	**0.007**	0.007	0.006	0.006	0.007	0.007	**0.006**	0.007	0.007
SPCA	AR	MAE	0.016	0.016	0.014	0.013	0.012	0.012	0.012	0.011	0.011	0.010	0.010	0.009
SPCA	CSR	MAE	0.006	0.007	0.007	**0.007**	0.007	0.008	0.008	0.008	0.008	0.008	0.008	0.008
SPCA	LASSO	MAE	**0.003**	**0.005**	**0.006**	**0.007**	0.007	0.006	0.006	0.007	0.007	**0.006**	0.007	0.007
NA	PHC	MAE	0.026	0.026	0.029	0.031	0.030	0.035	0.039	0.032	0.029	0.037	0.029	0.034

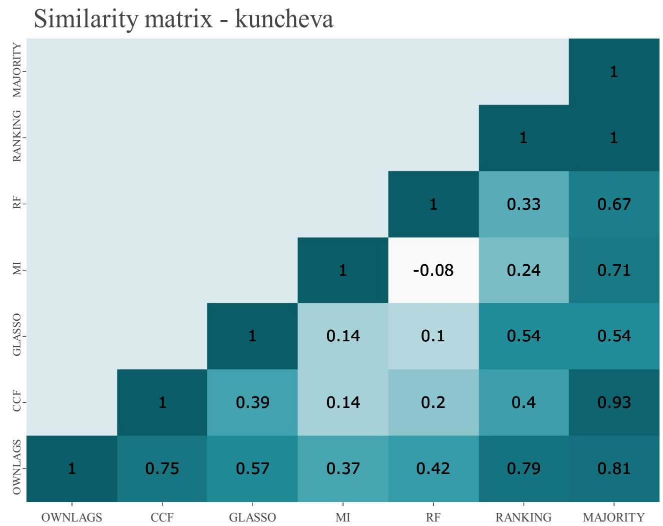


FIGURE A1: The graph shows 1 step lagged selected variables for annual inflation rate forecast, for more details see `fsMTS` package in R. Banco de la República database.
Author's calculations.

TABLE A3: Giacomini - White test by HDRM at t+3.

MLM t+3	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		1	0.9988	0.9999
STEPWISE	0		0.017	0.0089
DFMPC	0.0012	0.983		0.5087
SPCA	1e-04	0.9911	0.4913	

CSR t+3	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		1	0.6726	0.8324
STEPWISE	0		0	0
DFMPC	0.3274	1		0.428
SPCA	0.1676	1	0.572	

LASSO t+3	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		1	0.9925	1
STEPWISE	0		0.0026	0
DFMPC	0.0075	0.9974		0.6684
SPCA	0	1	0.3316	

ADLASOO t+3	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		1	0.9962	1
STEPWISE	0		4e-04	0
DFMPC	0.0038	0.9996		0.7382
SPCA	0	1	0.2618	

Table A4: Giacomini - White test by HDRM at t+12.

MLM t+12	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		0.9982	0.976	0.9994
STEPWISE	0.0018		0.1297	0.1363
DFMPC	0.024	0.8703		0.6019
SPCA	6e-04	0.8637	0.3981	

CSR t+12	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		0.9999	0.0014	0.2916
STEPWISE	1e-04		0	0
DFMPC	0.9986	1		0.999
SPCA	0.7084	1	0.001	

LASSO $t+12$	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		1	0.9258	0.9999
STEPWISE	0		0.0479	0.0097
DFMPC	0.0742	0.9521		0.6764
SPCA	1e-04	0.9903	0.3236	

ADLASOO t+12	ENSEMBLE	STEPWISE	DFMPC	SPCA
ENSEMBLE		1	0.9732	0.9999
STEPWISE	0		0.0366	0.0038
DFMPC	0.0268	0.9634		0.6134
SPCA	1e-04	0.9962	0.3866	

The test compares the conditional predictive accuracy of two forecasts. H_0 : forecast of the models in the rows are better. Thus, rejecting H_0 means that the forecasts of the columns are significantly more accurate than forecasts from the rows.

Table A5: List of macroeconomic and finance variables.

ID	Name	Frequency	LastUpdate	NextUpdate	Source
V1	Total inflation	Monthly	08-11-2021	06-12-2021	BanRep
V2	Commercial portfolio in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V3	Trading portfolio in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V4	Consumer portfolio in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V5	Consumer portfolio in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V6	Microcredit portfolio in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V7	Microcredit portfolio in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V8	Adjusted mortgage portfolio in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V9	Mortgage portfolio in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V10	Total Gross portfolio adjusted for securitization in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V11	Total Gross Portfolio with securitization in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V12	Total Gross Portfolio without adjustment for securitization in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V13	Total Gross Portfolio without adjustment for securitization in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V14	Total Net Portfolio adjusted for securitization in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V15	Total Net Portfolio with securitization adjustment in legal currency	Monthly	12-11-2021	10-12-2021	BanRep

Table A5. Continued

ID	Name	Frequency	LastUpdate	NextUpdate	Source
V16	Total Net Portfolio without adjustment for securitization in foreign currency	Monthly	12-11-2021	10-12-2021	BanRep
V17	Total Net Portfolio without securitization adjustment in legal currency	Monthly	12-11-2021	10-12-2021	BanRep
V18	Monetary base	Monthly	12-11-2021	10-12-2021	BanRep
V19	Quasi-Moneys - CDT	Monthly	12-11-2021	10-12-2021	BanRep
V20	Quasi-money - savings accounts	Monthly	12-11-2021	10-12-2021	BanRep
V21	Cash	Monthly	12-11-2021	10-12-2021	BanRep
V22	Monetary Aggregate - M1	Monthly	12-11-2021	10-12-2021	BanRep
V23	Monetary Aggregate - M2	Monthly	12-11-2021	10-12-2021	BanRep
V24	Monetary Aggregate - M3	Monthly	12-11-2021	10-12-2021	BanRep
V25	Bank Reserve	Monthly	12-11-2021	10-12-2021	BanRep
V26	Total Repos,Simultaneous and Temporary Transfer of Securities with the real sector	Monthly	12-11-2021	10-12-2021	BanRep
V27	Export price index, according to foreign trade	Monthly	30-11-2021	12-12-2021	BanRep
V28	Import price index, according to foreign trade	Monthly	30-11-2021	12-12-2021	BanRep
V29	Terms of trade index, according to foreign trade	Monthly	30-11-2021	12-12-2021	BanRep
V30	Aggregate Series of Workers' Remittances	Monthly	26-11-2021	15-12-2021	BanRep
V31	Real Exchange Rate Index with Germany	Monthly	11-11-2021	07-12-2021	BanRep
V32	Real Exchange Rate Index with Canada	Monthly	11-11-2021	07-12-2021	BanRep
V33	Real Exchange Rate Index with China	Monthly	11-11-2021	07-12-2021	BanRep
V34	Real Exchange Rate Index with Spain	Monthly	11-11-2021	07-12-2021	BanRep
V35	Real Exchange Rate Index with the United States	Monthly	11-11-2021	07-12-2021	BanRep
V36	Real Exchange Rate Index with France	Monthly	11-11-2021	07-12-2021	BanRep
V37	Real Exchange Rate Index with Italy	Monthly	11-11-2021	07-12-2021	BanRep
V38	Real Exchange Rate Index with IMF Members	Monthly	11-11-2021	07-12-2021	BanRep
V39	Real exchange rate with CPI deflactor	Monthly	11-11-2021	07-12-2021	BanRep
V40	Real Exchange Rate Index deflacted with PPI	Monthly	11-11-2021	07-12-2021	BanRep
V41	Inflation target	Annual	01-01-2021	01-01-2022	BanRep
V42	Consumer Price Index: Non-food and non-regulated goods	Monthly	08-11-2021	06-12-2021	BanRep
V43	Consumer Price Index: Core 15	Monthly	08-11-2021	06-12-2021	BanRep
V44	Consumer Price Index: Core 20	Monthly	08-11-2021	06-12-2021	BanRep

Table A5. Continued

ID	Name	Frequency	LastUpdate	NextUpdate	Source
V45	Consumer Price Index: Regulated	Monthly	08-11-2021	06-12-2021	BanRep
V46	Consumer Price Index: Non-Food and Non-Regulated Services	Monthly	08-11-2021	06-12-2021	BanRep
V47	Consumer Price Index: No food	Monthly	08-11-2021	06-12-2021	BanRep
V48	Consumer Price Index: No Food BR	Monthly	08-11-2021	06-12-2021	BanRep
V49	Consumer Price Index: No Food or Regulated	Monthly	08-11-2021	06-12-2021	BanRep
V50	Consumer Price Index: No Food or Regulated BR	Monthly	08-11-2021	06-12-2021	BanRep
V51	Consumer Price Index: No primary foodstuffs, utilities and fuels	Monthly	08-11-2021	06-12-2021	BanRep
V52	Consumer Price Index: BR Food	Monthly	08-11-2021	06-12-2021	BanRep
V53	Consumer Price Index: Non-tradable	Monthly	08-11-2021	06-12-2021	BanRep
V54	Consumer Price Index: Tradable	Monthly	08-11-2021	06-12-2021	BanRep
V55	Producer Price Index	Monthly	08-11-2021	06-12-2021	BanRep
V56	Producer Price Index, Domestic Supply - Agriculture, Livestock, Hunting, Forestry and Fisheries	Monthly	08-11-2021	06-12-2021	BanRep
V57	Producer Price Index, Domestic Supply - Mining and Quarrying	Monthly	08-11-2021	06-12-2021	BanRep
V58	Producer Price Index, Domestic Supply - Manufacturing Industry	Monthly	08-11-2021	06-12-2021	BanRep
V59	Producer Price Index, Domestic Supply - Total	Monthly	08-11-2021	06-12-2021	BanRep
V60	Producer Price Index, Domestic Production - Agriculture, Livestock, Hunting, Forestry and Fisheries	Monthly	08-11-2021	06-12-2021	BanRep
V61	Producer Price Index (PPI), Domestic Production - Mining and Quarrying	Monthly	08-11-2021	06-12-2021	BanRep
V62	Producer Price Index, Domestic Production - Manufacturing Industries	Monthly	08-11-2021	06-12-2021	BanRep
V63	Producer Price Index, By Origin - Exported	Monthly	08-11-2021	06-12-2021	BanRep
V64	Producer Price Index, According to Origin - Imported	Monthly	08-11-2021	06-12-2021	BanRep
V65	Producer Price Index, According to Origin - produced and consumed	Monthly	08-11-2021	06-12-2021	BanRep

Table A5. Continued

ID	Name	Frequency	LastUpdate	NextUpdate	Source
V66	Producer Price Index, According to use or economic destination - Final consumption	Monthly	08-11-2021	06-12-2021	BanRep
V67	Producer Price Index, According to Economic Use or Destination - Construction Materials	Monthly	08-11-2021	06-12-2021	BanRep
V68	Producer Price Index, According to Economic Use or Destination - Capital Goods	Monthly	08-11-2021	06-12-2021	BanRep
V69	Producer Price Index, According to economic use or destination - Intermediate consumption	Monthly	08-11-2021	06-12-2021	BanRep
V70	Unemployment rate	Monthly	30-11-2021	30-12-2021	BanRep
V71	Gross International Reserves	Monthly	10-11-2021	06-12-2021	BanRep
V72	Gross International Reserves (FLAR Free)	Monthly	10-11-2021	06-12-2021	BanRep
V73	Net international reserves	Monthly	10-11-2021	06-12-2021	BanRep
V74	Net International Reserves (FLAR Free)	Monthly	10-11-2021	06-12-2021	BanRep
V75	Acquisition of a home other than VIS (placement in COP), Interest rate	Monthly	12-11-2021	10-12-2021	BanRep
V76	Acquisition of VIS housing (placement in COP), Interest rate	Monthly	12-11-2021	10-12-2021	BanRep
V77	Non-VIS housing construction (placement in COP), Interest rate	Monthly	12-11-2021	10-12-2021	BanRep
V78	VIS housing construction (placement in COP), Interest rate	Monthly	12-11-2021	10-12-2021	BanRep
V79	Consumer Credit, Interest Rate	Monthly	12-11-2021	10-12-2021	BanRep
V80	Commercial Loans (Ordinary), Interest Rate	Monthly	12-11-2021	10-12-2021	BanRep
V81	Commercial Loans (Preferential or Corporate), Interest Rate	Monthly	12-11-2021	10-12-2021	BanRep
V82	Commercial Loans (Treasury), Interest Rate	Monthly	12-11-2021	10-12-2021	BanRep
V83	Microcredits (other than leasing), Interest rate	Monthly	12-11-2021	10-12-2021	BanRep
V84	90-day Fixed Term Deposit (DTF) Rate	Monthly	03-11-2021	01-12-2021	BanRep
V85	Banco de la República Placement Interest Rate	Monthly	12-11-2021	10-12-2021	BanRep
V86	Total Placement Interest Rate	Monthly	12-11-2021	10-12-2021	BanRep
V87	Euro - COP/EUR - Average rate	Daily	01-12-2021	02-12-2021	BanRep
V88	Euro - USD/EUR	Daily	01-12-2021	02-12-2021	BanRep
V89	Pound Sterling - COP/GBP - Average Rate	Daily	01-12-2021	02-12-2021	BanRep

Table A5. Continued

ID	Name	Frequency	LastUpdate	NextUpdate	Source
V90	Representative Market Rate (TRM)	Daily	30-11-2021	31-12-2021	BanRep
V91	Overnight Banking Benchmark (IBR), effective	Daily	31-11-2021	32-12-2021	BanRep
V92	Overnight Banking Reference Indicator (IBR), nominal	Daily	32-11-2021	33-12-2021	BanRep
V93	Interest Rate Zero Coupon, Treasury Securities (TES), pesos - 1 year	Daily	33-11-2021	34-12-2021	BanRep
V94	Interest Rate Zero Coupon, Treasury Securities (TES), pesos - 10 years	Daily	34-11-2021	35-12-2021	BanRep
V95	Zero Coupon Interest Rate, Treasury Securities (TES), pesos - 5 years	Daily	35-11-2021	36-12-2021	BanRep
V96	Monetary Policy Rate	Daily	36-11-2021	37-12-2021	BanRep
V97	Interbank Rate (TIB)	Daily	37-11-2021	38-12-2021	BanRep
V98	Economy Monitoring Indicator-ISE	Monthly	31-11-2021	2-12-2021	DANE
V99	Domestic Debt	Monthly	31-09-2001	nan	MinHacienda
V100	External Debt	Monthly	32-09-2001	nan	MinHacienda
V101	Average Monthly Inflation Expectations	Monthly	30-11-2021	31-12-2021	BanRep
V102	Average 12-month inflation expectations	Monthly	30-11-2021	31-12-2021	Banrep