Can Machine Learning improve inflation forecasting?

Evaluating forecast performance for an emerging economy and a developed economy using machine learning models

Emil Hansson

Master Thesis I – NEKN01

May 2023

Lund University – Department of Economics

Supervisor: Joakim Westerlund

Abstract

This paper aims to compare and evaluate the performance of inflation forecasting performance for benchmark time series models and machine learning models. The process is performed for both a developed economy, the US, and an emerging economy, Mexico. The study examines how forecast performance compares between benchmark time series models and machine learning models, as well as how forecast performance overall compares between an emerging economy and a developed economy. To conduct the study, two separate datasets for the US and Mexico are gathered and to perform the forecasts, a rolling window with a fixed window size is used across a total of 12 horizons.

The study finds that the benchmark time series models outperform the machine learning models in forecasting inflation for the US dataset, but not for the Mexican dataset. Additionally, the study finds that the inflation forecast performance shows smaller forecast errors across all horizons for the developed economy compared to the emerging economy. These results are mainly attributed to the characteristics of each specific dataset. The US dataset with its higher dimension and more potential predictors is concluded to better suit machine learning models and perform better inflation forecasts. In conclusion, the study provides evidence that machine learning can be a useful tool for macroeconomic policymakers when sufficient data is available.

Keywords: Inflation Forecast, Machine Learning, Rolling Window, Time Series

Contents

1. Introduction

One of the more vital elements for economic policymakers is to control inflation, and to achieve this, accurate inflation forecasts are of great interest for policymakers. Groen et al. (2013) argues that other agents in the economy take great interest in accurate inflation forecasts, either to be able to assess how policymakers will make decisions in the future, or to form their inflation expectations when making decisions related to wage negotiations, consuming behavior, and investments. Many central banks have in recent years conducted inflation-forecast targeting, typically with a quantitative inflation target such as 2 percent per year (Woodford, 2007). Inflation targeting has some direct advantages, namely focusing monetary policymaking to achieving low and stable inflation, and inflation targeting provides a measure of credibility of monetary policy. However, implementing and monitoring inflation has some limitations as well and Svensson (1997) concludes that the central bank's inflation forecasts serve best as an intermediate target for efficient and accurate monetary policymaking. Overall, performing accurate and reliable inflation forecasts is of high significance for economic policymakers and other agents in the economy.

The main theoretical framework used in early inflation forecasting studies was the Phillips Curve, presented by Phillips (1958), which was used in one of the original papers on forecasting inflation by Stock and Watson (1999). After that, many other methodological frameworks have been implemented when forecasting inflation. Heeren (2021) states that univariate time series models such as the autoregressive model or autoregressive integrated moving average model have been widely used to forecast inflation. However, according to Hall (2018), forecasting inflation with traditional time series methods presents two main limitations; first, there are many potential predictors, and second, there is a limited number of available time series. These limitations can lead to the "curse of dimensionality" which can lead to a problem of overfitting the models. One way to deal with the dimensionality problem is to incorporate factor models, and benefit from a large potential of predictor variables in inflation forecasting by extracting common factors from the potential predictors. Another set of models that can be beneficial are machine learning models, which are better suited to handle many predictors efficiently according to Zhoua et al. (2017).

Varian (2014) concludes with increased data availability, so called "big data", the implementation of machine learning methods has seen an increased use within the field of economics. Charpentier et al. (2018) states that both econometrics and machine learning have a common goal, which is to build a predictive model for a specific variable of interest by using feature variables. However, both fields have evolved in parallel to each other, where econometrics is set to capture probabilistic models built to describe economic events, whereas machine learning uses algorithms capable of learning from their mistakes. Charpentier et al (2018) continues by stating that machine learning methods in economics are particularly useful when the numbers of potential predictors are large and can improve the forecast performance compared to traditional econometrics methods.

Traditional econometric models often assume that the relationship between the explanatory variable and the predictors are linear, even though this assumption is not always appropriate. Many machine learning methods do not make this assumption and can handle the non-linearity of the models, and therefore improve upon traditional econometric models, Moshiri and Cameron (2000). Forecasting inflation using potential predictors is a complex analysis where the assumption between the predictors and the target variable inflation is not necessarily appropriate to assume linear. Hence, incorporating machine learning as methodology to forecast inflation using many predictors can be a valuable instrument for economic policymakers to include into the analysis.

Athey and Imbens, (2019) states that machine learning can be divided into two categories: unsupervised and supervised machine learning. Supervised machine learning focus primarily on prediction problems, with an outcome variable, y, and a set of predictors, $x = (x_1, \dots, x_n)'$, where x is a vector of potential predictors. The main application of supervised machine learning is to estimate a model on a subset of the data from the values of the predictors, called the training set, and then use the remaining data, the test set, for predictions. Unsupervised machine learning on the other hand does not use a specific prediction task with an explanatory variable and predictors, instead the unsupervised machine learning models use methods that recognize patterns in the data, such as clustering images into groups. Econometrics is intersected mainly by the supervised machine learning field, however, according to Athey and Imbens (2019), one of the main differences between econometrics and supervised machine learning is inference. Supervised machine learning methods rely on data-driven model selection, such as cross-validation, and the goal is prediction performance without regards to implications of inference. Econometrics on the other hand relies on predictive performance on dependent and explanatory variables where the goal is to draw inferences about potential outcomes.

The purpose of this essay is to examine the forecast performance of inflation using a variety of models. Classic econometric time series models, factor models, and supervised machine learning models are evaluated and compared for an emerging economy, Mexico, and a developed economy, the United States (US). The inflation forecasts for both countries are compared and evaluated on two different datasets consisting of monthly macroeconomic variables, and attempts to answer the following questions:

- Do machine learning models outperform benchmark time series models?
- How does the performance of benchmark time series models and machine learning models compare for an emerging economy and a developed economy?

The main contribution of this essay to the existing literature is the use of more potential predictors when forecasting inflation for an emerging economy. Emerging economies are underrepresented in forecast studies with many predictors due to the limited amount of data available for these countries compared to developed countries. This study aims to collect more potential predictors than are usually included in the existing literature when forecasting inflation for an emerging economy, hence the choice of Mexico as the emerging economy to examine since Mexico has sufficient monthly data easily available to collect compared to most emerging economies. Mexico is also interesting to study because it is one of the largest economies in the world globally and is therefore of high significance in many macroeconomic aspects. This study also examines the comparison of forecast performance for an emerging and a developed economy to gain further knowledge of the comparison performance between benchmark models and machine learning models under different dataset structures. The US have been widely used in inflation forecasting, and will mostly serve as a reference point regarding a contribution to the literature and be compared to the Mexican forecasts.

This essay is divided into the following sections: Section 2 presents some of the previous literature on inflation forecasting and some of the methods that have been used in inflation forecasting studies. Section 3 gives an overview of the datasets chosen for the study. Section 4 presents the theoretical framework in which the study is performed under. Section 5 presents the main results of the study. Section 6 discusses the study and its main results, and concludes where it places among the existing literature on the subject.

2. Literature review

There have been several studies with different methodological frameworks for forecasting inflation in both emerging and developed economies, although most of the research has been focused on developed economies and mainly on the US because of the superior amount of available macroeconomic data. The following section presents some notable studies done with some different methodological frameworks.

The Phillips curve was the main method used in early papers. Stock and Watson (1999) forecasts US inflation at the 12-month horizon for data covering the time period 1970-1996 using the Phillips curve. They find that the Phillips curve produced the most accurate short run forecasts of US inflation, and outperformed the univariate forecasting models. Conversely, Atkeson and Ohanian (2001) use a similar methodology to forecast inflation for the US using models based on non-accelerating inflation rate of unemployment (the NAIRU) for the period 1984-1996. They conclude in their study that the Phillips Curve does not even outperform a naïve model that presumes inflation over the next four months will be equal to inflation over the past four months.

Pichler (2008) implements a Dynamic Stochastic General Equilibrium (DSGE) model on US data between 1979 and 2007. They incorporate both a linear and nonlinear approximation to the data, where they find that the nonlinear specification of the data slightly outperforms the linear specification in predictive power. Iversen et al. (2016) evaluated and compared inflation forecast performances for both a DSGE model and a Bayesian VAR model for Sweden from 2007 to 2013. They conclude that the Bayesian VAR model outperformed the DSGE model to forecast inflation in Sweden.

Another common way of methodology is the use of factor models to forecast inflation. Stock and Watson (2002) show in their paper how the usage of factor models can improve forecasting performance with a principal component analysis for a large set of predictors. They conclude using a principal component analysis for 215 monthly US macroeconomic time series from 1970-1998 that a principal component analysis outperforms univariate regressions, vector autoregressions, and leading indicator models. Forni et al. (2003) utilized the factor model approach on the Euro area when they studied how financial variables can forecast inflation. For a monthly dataset of 447 predictors between 1987-2001, they find that the multivariate factor models perform better than the univariate autoregressive models. Artis et al. (2001) uses a dynamic factor modeling on a large monthly macroeconomic dataset for the UK between 1970-1998 and finds that the factor-based forecast approach is an improvement upon standard benchmark models. This holds true for variables such as prices, real aggregates, and financial variables.

In recent years, with the availability of large datasets, the use of machine learning techniques in macroeconomic forecasting studies have increased. Medeiros et al. (2021) forecasts inflation for the US between 1960 and 2015 with benchmark models, factor models, shrinkage models, ensemble models, random forests, and hybrid linear-random forest models. They conclude that the machine learning models such as the LASSO and Random Forest produce more accurate forecasts than the standard benchmark models and highlight the need for machine learning methods and big data for macroeconomic forecasting. Garcia et al. (2017) uses a similar approach to forecast inflation in an emerging economy, Brazil, and they concluded that the machine learning techniques outperformed the AR model and Random Walk model, although they suggest that the predictive power of the machine learning methods are different depending on the forecast horizons. Özgur and Akkoc (2021) forecasts inflation on an emerging economy, Turkey, using machine learning algorithms combined with benchmark models. They find that the shrinkage methods perform better than conventional econometric methods in the case of Turkish inflation.

To summarize the literature review, factor models and machine learning methods have been seen to improve the performance of forecasts compared to benchmark time series models.

3. Data

The study will consist of two datasets, one for an emerging economy, Mexico, and one for a developed economy, the US. The following sections provide some details about respective datasets.

3.1. The US dataset

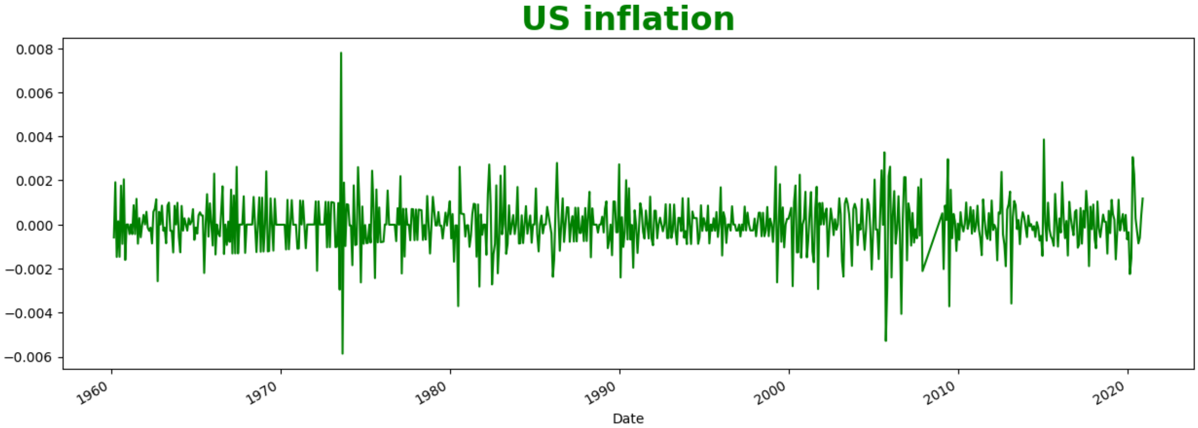
The first dataset of the study contains 122 monthly macroeconomic indicators, including the inflation variable, for the US between 1960-2020, and common for all the variables included are that there are no missing values for the specific sample period. The variables are gathered from the Federal Reserve database, and the dataset is the FRED-MD dataset, which is a high-dimensional dataset of monthly macroeconomic variables designed to be a convenient starting point for macroeconomic analysis using "big data". For more details about the dataset and the specific transformations used for all the variables to achieve stationarity, see McCracken and Ng (2015).

The forecast evaluation for the models is done using a rolling window approach with a fixed window size where the test sample is divided into a training set and a test set, and then for the first date from the sample period, all previous data is used to train the model, and then evaluated with the test set. After the forecast is made for one horizon, then for every next forecast horizon, the window is shifted one month ahead in time. A total of 12 months ahead will be forecasted in this study and evaluated between the different models. In this study, the training sample consists of data from January 1960 to December 2005, and the test data is from January 2006 to December 2020. Hence, there will be 45 years of training data for each model and 15 years of test data for each model in the US dataset. That gives a total of 540 observations for the training set, and of these, the 120 latest observations will be used as the window size, which then will be updated for every month ahead by including the previous forecast in the window sample and excluding the last observation in the window sample, therefore keeping the rolling window fixed and updated with the most relevant observations to use for forecasting.

The motivation for splitting the data into training and test into the given proportions are due to common practice in previous literature. According to Joseph (2022), there is no optimal way of splitting the data, and the optimal training and test split ratio depends on the specific data characteristics. Joseph continues by stating that ratio with 20-30% of the data as test set are commonly used in practice, although this is mostly a split done by rule of thumb without any specific underlying theory. Hence, in this study, the split of 75% training data, and 25% test data is chosen. The size of the window of 10 years is chosen, motivated by an attempt to include the most relevant observations for each forecast, to make the forecast performance better, but to also include as many observations as possible to forecast with. The choice of the window size is also optimal without any specific theoretical way of choosing the window size, but is of high significance to optimize forecast performance (Inoue et al. 2016).

The variable of interest in this study is inflation, and to measure inflation, the Consumer Price Index (CPI) is used. The inflation can be computed for a certain month by taking the following formula: $\pi_t = \log(P_t) - \log(P_{t-1})$ where P_t is the value for the CPI at month *t,* and π_t is the inflation. Figure 1 shows the movement of the CPI variable for the US data across the sample size.

Figure 1: Monthly US CPI movements from 1960-2020



3.2. The Mexican dataset

The Mexican dataset consists of monthly macroeconomic variables that are collected from the Banco de Mexico database, and from the Federal Reserve database, totaling up to 38 variables, including the inflation variable. The monthly Mexican variables span the period January 1997 to December 2017, which is the chosen period for this study, since many monthly macroeconomic variables relevant to include are either not tracked before 1997, or suffer from inconsistent tracking with missing variables in the sample. The goal with choosing variables for the Mexican dataset is to include as many potential macroeconomic variables as possible to forecast inflation, using variables from as many macroeconomic sectors as possible. All variables chosen does not have any missing variables for the sample period, and the variables are also chosen to resemble the US dataset as much as possible for as much consistency as possible. A complete list of variables used in the study is presented in the Appendix.

The main difference between the Mexican and the US dataset is the number of potential predictors and the sample period for both datasets. However, performing the models on two distinct datasets can also be seen as a way to contribute to the existing literature by comparing machine learning models on a dataset with high-dimension, and a dataset with lower dimension. The same forecast approach is made on the Mexican dataset, where a rolling window will be applied, and the forecasts will be made for 12 different horizons with a fixed window size. The dataset will be divided into a training sample, which will be from January 1997 to December 2011, and data from January 2012 to December 2017 will consist of the test sample. The number of training observations total up to 168 observations, where the 50 latest observations will be used as the rolling window, and then for every additional forecast step be updated by removing the last observation in the rolling window, and adding the previous forecast value, therefore keeping the window size fixed and updated for more accurate and reliable forecasts. The motivation for the split ratio between training and test set, and the size of the window, is done by using the same principles as for the US dataset. The split is done to closely resemble the US dataset in proportions for consistency.

To work with the Mexican dataset optimally, the monthly macroeconomic variables need to achieve stationarity. To evaluate whether the data is stationary, an Augmented Dickey Fuller (ADF) test, introduced by Dickey and Fuller (1997), is performed on each variable in the dataset, and variables that are non-stationary are transformed in one of the following ways:

(1) Taking the first difference of the variable, which means that the transformation for a variable x at a given month t, is computed as: Δx_t

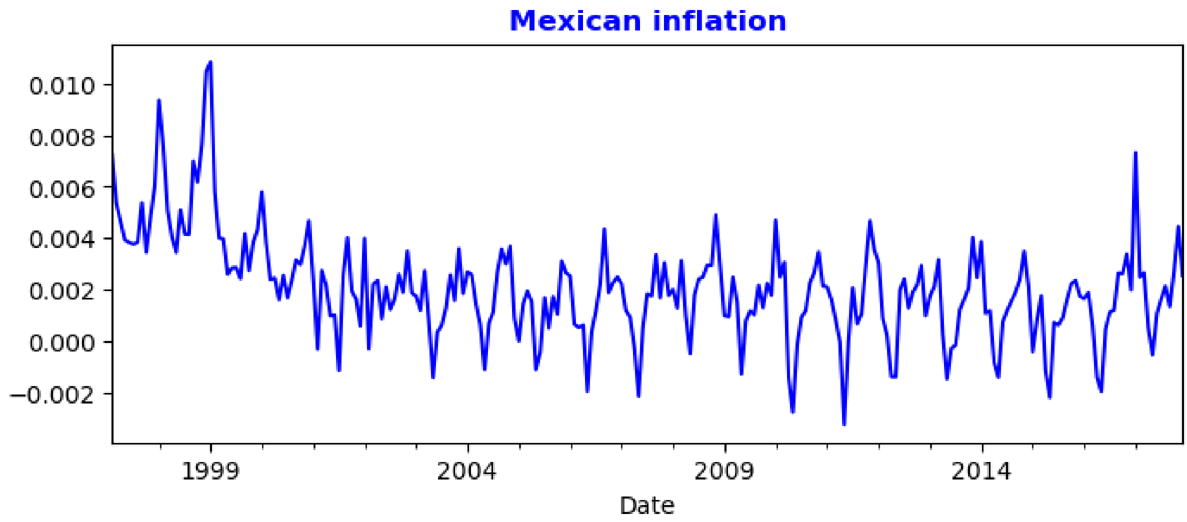
(2) Taking the first log difference of the non-stationary variable, which means the variable x at month t is computed as: $\Delta \log\,(x_t)$

(3) Taking the second log difference of the non-stationary variable. The non-stationary variable x at a month t is computed as: $\Delta^2 \log\,(x_t)$

The transformations for the variables are done in a way to resemble the FRED-MD dataset as much as possible, and after performing the transformations on the non-stationary variables, the p-values obtained from the ADF test are all less than 0.05, which is a good indicator that the data is stationary. The complete list of transformations used on each variable is described in the Appendix.

Just as with the US dataset, the forecasted variable is inflation, and the measurement used is the Consumer Price Index (CPI), which is obtained from the Federal Reserve Database. Inflation is computed by using the same formula as for the US:
$\pi_t = \log(P_t) - \log\,(P_{t-1})$ where P_t is the value for the CPI at month t. Figure 2 shows the Mexican CPI movements for the sample period.

Figure 2: Monthly Mexican CPI movements from 1996-2017



4. Methodology

In this study, the forecasts are based on a rolling window approach of a fixed length. A rolling window approach means that the forecast is revised each period. By using the rolling window framework, it updates the forecasts for $T - 1$ of the periods in the horizon and computes a forecast for the new period T. In this study, the notation of the models will closely follow the structure of Medeiros et al. (2021) which means that the general specification of a model can be written as:

$$\pi_{t+h} = G_h(x_t) + u_{t+h,} \tag{4.1}$$

where $h = 1, \dots, H$ is the forecast horizon and $t = 1, \dots, T$. The dependent variable, π_{t+h} is inflation at month $t + h$ and $x_t = (x_{1t}, \dots, x_{nt})'$ is a vector consisting of the n dependent variables and predictors used in a specific model. $G_h(x_t)$ is a general target function depending on the specific model used to forecast where $G_h(\cdot)$ is the mapping between the predictors and future inflation, where the mapping depends on the specific forecast horizon. Additionally, the direct forecast equation can be written as:

$$\hat{\pi}_{t+h|t} = \hat{G}_{h,t-R_h+1:t}(x_t) \tag{4.2}$$

where $\hat{G}_{h,t-R_h+1:t}$ is the estimated target function based on data from time $t - R_h + 1$ up to t, and where R_h is the window size. $\hat{\pi}_{t+h|t}$ is the forecasted inflation at time $t + h$ given the information available up to time *t*.

4.1. Benchmark Models

The accuracy of the machine learning models is evaluated and compared to some benchmark models. The most basic benchmark model that is included is the Random Walk (RW) model, which will be the reference model when comparing the results. The RW is a naïve forecast model and can be modeled as:

$$\hat{\pi}_{t+h|t} = \pi_t \text{ for } h = 1, \dots, 12$$

4. 3

which means that the forecasted value of inflation at $t + 1$ is the value of inflation at time t for every horizon. The second benchmark model included in the study is the Autoregressive (AR) model of order *p*. The AR(p) model can be modeled as:

$$\hat{\pi}_{t+h|t} = \hat{\phi}_{0,h} + \hat{\phi}_{1,h}\pi_t + \cdots + \hat{\phi}_{p,h}\pi_{t-p+1}$$

4. 4

where the model is different depending on the specific forecast horizon. The hyperparameter *p* determines how many lags are used in the forecast for inflation. $\hat{\phi}_{p,h}$ represents the estimated AR parameters, where each $\hat{\phi}_{p,h}$ corresponds to a weight assigned to the lagged value of inflation, π_{t-p+1} in the forecasted value.

The third benchmark model used in the study is the Autoregressive Moving Average model (ARMA), which is a leverage of a general autoregression AR(p) and a general Moving Average model MA(q). The ARMA model has two components denoted as *p* and *q*, where *p* is the lag order, and *q* is the order of the moving average. The ARMA model of order (p, q) can be formulated as:

$$\hat{\pi}_{t+h|t} = \sum_{h=1}^{p} \hat{\phi}_{p,h}\pi_{t-p+1} + \sum_{h=1}^{q} \hat{\theta}_{p,h}\, \varepsilon_{t-p+1}$$

4. 5

Where $\sum_{h=1}^{q} \hat{\theta}_{p,h}\, \varepsilon_{t-p+1}$ is the MA(q) part of the model and $\sum_{h=1}^{p} \hat{\phi}_{p,h}\pi_{t-p+1}$ is the AR(p) part of the model. $\hat{\theta}_{p,h}$ are the estimated coefficients of the MA(q) model and ε_{t-p+1}

represents the error term, which is the difference between the observed and predicted values in the model at time $t - p + 1$.

The fourth benchmark model is the Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX). This model considers the seasonal components of the time series forecast as well as external factors. This model captures the seasonal components of the analysis and is of order $(p, q, d) \times (P, Q, D)$, where *p* stands for the lag order, *d* stands for the degree of differencing but since the data is already stationary, we have *d=0*, and *q* denotes the order of the moving average. Additionally, the SARIMAX requires another set of components denoted as (P, Q, D) that account for the seasonality aspect, and also includes a parameter s, which is the periodicity of the seasonal cycle of the time series data set, Lazzeri (2020). *P* stands for the seasonal autoregressive lag order, and *Q* stands for the seasonal moving average model. *D* stands for the seasonal differencing order, but since the data is already stationary, we have *D=0* and does not need to be accounted for. This model adds to the forecasting analysis by capturing the seasonal patterns of the data.

When incorporating exogenous variables into the model under the current framework with inflation forecasting, the SARIMAX model can be formulated by modifying the previous equations as the following:

$$\hat{\pi}_{t+h|t} = c + \sum_{h=1}^{p} \hat{\phi}_{p,h}\pi_{t-p+1} + \sum_{h=1}^{q} \hat{\theta}_{q,h}\, \varepsilon_{t-p+1} + \sum_{h=1}^{P} \hat{\varphi}_{P,h}\pi_{t-sp+1} + \sum_{h=1}^{Q} \hat{\vartheta}_{Q,h}\, \varepsilon_{t-sp+1} + \sum_{h=1}^{r} \beta_h\, x_t \tag{4.6}$$

Where $\hat{\pi}_{t+h|t}$ is the forecasted inflation for a specific horizon, and $\sum_{h=1}^{r} \beta_h\, x_t$ is the matrix of predictors multiplied by a vector with the regression coefficients corresponding to the predictors, with a regression model intercept, *c*. The forecasted value of inflation for a specific horizon now depends on both the ARMA components; $\sum_{h=1}^{p} \hat{\phi}_{p,h}\pi_{t-p+1} + \sum_{h=1}^{q} \hat{\theta}_{q,h}\, \varepsilon_{t-p+1}$, the seasonal components of the ARMA model; $\sum_{h=1}^{P} \hat{\varphi}_{P,h}\pi_{t-sp+1} + \sum_{h=1}^{Q} \hat{\vartheta}_{Q,h}\, \varepsilon_{t-sp+1}$, and the exogenous variables plus a constant; $c + \sum_{h=1}^{r} \beta_h\, x_t$. In the SARIMAX equation, $\sum_{h=1}^{P} \hat{\varphi}_{P,h}\pi_{t-sp+1}$ is the seasonal autoregressive components of the model with *P* as the seasonal lag order across all horizons, and $\sum_{h=1}^{Q} \hat{\vartheta}_{Q,h}\, \varepsilon_{t-sp+1}$ is the seasonal moving average components of the model with *Q* as the seasonal moving average order across all horizons.

4.2. Shrinkage Models

For the shrinkage models used in this study, the general notation can be represented as in Medeiros et al. (2021):

$$G_h(x_t) = \beta_h' x_t,$$

4. 7

where:

$$\hat{\beta}_h = arg \min_{\beta_h} \left[\sum_{t=1}^{T-h} (y_{t+h} - \beta_h' x_t)^2 + \sum_{i=1}^{n} p(\beta_{h,i}; \lambda, \omega_i) \right],$$

4. 8

and $p(\beta_{h,i}; \lambda, \omega_i)$ is the general penalty function that depends on the penalty parameter λ and on a specific weight ω_i, where $\omega_i > 0$. In this paper, three shrinkage models are used to forecast inflation: The Ridge Regression (RR), the Least Absolute Shrinkage and Selection Operator (LASSO), and the Elastic Net regression (ENet).

4.2.1. Ridge Regression

The RR was first introduced by Hoerl and Kennard (1970) and the RR shrinks the coefficients towards zero for less relevant predictors. The parameter λ controls the size of the shrinkage and the penalty function for the RR can be written as:

$$\sum_{i=1}^{n} p\left(\beta_{h,i}; \lambda, \omega_i\right) = \lambda \sum_{i=1}^{n} {\beta_{h,i}}^2$$

4. 9

Medeiros et al. (2021) points out that RR gives the advantage of providing an analytical solution that can easily be computed, and variables that are less relevant shrink towards zero. This kind of regularization is referred to as ι_2 regularization.

4.2.2. Lasso Regression

The LASSO regression was first proposed by Tibshirani (1996), and its penalty function can be defined as:

$$\sum_{i=1}^{n} p\left(\beta_{h,i}; \lambda, \omega_i\right) = \lambda \sum_{i=1}^{n} |\beta_{h,i}|$$

4. 10

The main difference from the RR is that the LASSO shrinks the irrelevant variables to zero. The constraint given by the LASSO makes the solutions nonlinear in the π_{t+h}, and unlike the RR, gives no closed form expression. This kind of regularization is referred to as ι_1 regularization.

4.2.3. Elastic Net Regression

The Elastic Net (ENet) is a generalization method that includes both the RR and LASSO as special cases. The ENet shrinkage method is a convex combination of the ι_1 and ι_2 norms (Zou and Hastie, 2005). The elastic net penalty function under the current framework can be defined as:

$$\sum_{i=1}^{n} p\left(\beta_{h,i}; \lambda, \omega_i\right) = \alpha\lambda \sum_{i=1}^{n} {\beta_{h,i}}^2 + (1-\alpha)\lambda \sum_{i=1}^{n} \left|\beta_{h,i}\right|, \tag{4. 11}$$

Where $\alpha \in [0,1]$ is the ratio between the ι_1 and ι_2 regularization. The ENet estimator selects variables the same way as the LASSO and shrinks together the coefficients of correlated predictors the same way as the RR (Hastie et al. 2001).

4.3. Target Factors

The concept of target factors is a methodology that attempts to reduce the dimensionality of the potential predictors by extracting common components from the predictors, x_t, to forecast inflation. To use target factors, a Principal Component Analysis (PCA) of the predictors will be used to extract common factors, and then it will use the optimal number of principal components as the dependent variable to forecast inflation using an OLS regression. Medeiros et al. (2021) describes that the idea behind target factors is that many variables are irrelevant predictors of the targeted inflation π_{t+h}, and by computing the principal components of the variables with high predictive power, a more accurate forecast can be performed. The formal description behind how factor models work and how the principal component estimator works can be found in Bai and Ng (2003).

4.4. Random Forest Regression

The Random Forest (RF) model was first presented by Breiman (2001) with the main goal to reduce the variance of regression trees that are non-parametric models that can approximate an unknown nonlinear function. It uses bootstrap aggregation (bagging) of the regression trees to achieve a lower variance (Breiman, 1996). The regression trees used in the RF model works by splitting the dependent variables into regions, denoted by R_k such as each region corresponds to a terminal node (K). Given the dependent variable inflation, π_{t+h}, a set of predictors, x_t and a given number of terminal nodes, K, all the splits are chosen to a regression model with the goal to minimize the sum of squared errors in the following way:

$$\pi_{t+h} = \sum_{k=1}^{K} c_k I_k(x_t; \theta_k),$$

4. 12

where $I_k(x_t; \theta_k)$ is an indicator function such as:

$$I_k(x_t; \theta_k) = \begin{cases} 1 \; if \; x_t \; \in \; R_k(\theta_k) \\ 0 \; otherwise \end{cases}$$

4. 13

,where $R_k(\theta_k)$ is the k:th region. The collection of regression trees in the RF is specified in a bootstrap sample of the data, where for each sample $b = 1, \dots, B$ a tree with K_B regions are estimated for a selected subset of the original regressors. The forecast regression applied to the data can be written as:

$$\hat{\pi}_{t+h|t} = \frac{1}{B} \sum_{b=1}^{B} \left[\sum_{k=1}^{K_B} \hat{c}_{k,b} I_{k,b}(x_t; \hat{\theta}_{k,b}) \right]$$

4. 14

4.5. K-Nearest Neighbor Regression

The K-Nearest Neighbor (KNN), first introduced by Fix and Hodges (1951) is a form of nonparametric method that uses dependent variables to predict according to the closest training examples. This method looks at the K points, or neighbors, in the training set that are nearest to the test input, x_t, and counts how many members of each class are in the set and then returns an estimate (Murphy, 2012). The idea behind the KNN can be described by a distance measure such as the Euclidean distance between the predictors and the inflation variable:

$$d(\pi_{t+h}, x_t) = \sqrt{\sum_{t=1}^{n} (\pi_{t+h} - x_t)^2}$$

4. 15

4.6. Support Vector Regression

Support Vector Regression (SVR) is a machine learning algorithm based on the classification paradigm Support Vector Machines that transforms a nonlinear regression into a linear one by using kernel functions to map the original input into a new feature space with higher dimensions and finding a hyperplane that maximizes the margin while also minimizing prediction errors, and then perform a linear regression (Murphy, 2012). To formalize the SVR, the corresponding objective function can be written as:

$$J = C \sum_{t=1}^{N} L_{\epsilon}(\pi_{t+h}, \hat{\pi}_{t+h}) + \frac{1}{2}||w||^2 \tag{4.16}$$

where C is a regularization constant and $\hat{\pi}_{t+h} = f_h(x_t)$. The term $L_{\epsilon}(\pi_{t+h}, \hat{\pi}_{t+h})$ is called the epsilon insensitive loss function which measures the tolerance, ϵ, around the predicted values that are negligible. However, this is a convex and unconstrained objective function, but not differentiable because of the absolute value of the loss term, $\frac{1}{2}||w||^2$. To account for this and make it a constrained optimization problem, a possible approach is to introduce slack variables. The slack variables represent the degree to which each point lies outside the tolerance region which makes the optimization problem of finding the hyperplane more relaxed (Murphy, 2012):

$$\pi_{t+h} \leq f_h(x_t) + \epsilon + \delta_{t+h}^{+} \tag{4.17}$$

$$\pi_{t+h} \leq f_h(x_t) + \epsilon + \delta_{t+h}^{-} \tag{4.18}$$

We can now rewrite our original objective function by using the slack variables to:

$$J = C \sum_{t=1}^{N} (\delta_{t+h}^{+} + \delta_{t+h}^{-}) + \frac{1}{2}||w||^2 \tag{4.19}$$

where δ_{t+h}^{+} and δ_{t+h}^{-} are the slack variables.

4.7. Python Packages and Tuning Parameters

To conduct the inflation forecasts for all the models, the Python software is used in this study using the *statsmodels* library for the traditional time series models, and the *sklearn* library is used for the machine learning models. Then from these libraries, the specific packages relevant for each model are imported and used to forecast inflation using a rolling window approach with a fixed window size. All variables are normalized by importing the *StandardScaler* package from the *sklearn* library for optimal performance across all models.

With machine learning algorithms, an important feature is to optimally choose the hyperparameters to optimize the performance from the models. For the shrinkage models, consisting of the RR, LASSO and ENet, a cross-validation test is performed in Python to optimally choose the penalty parameter λ using the K-Fold cross-validation with K=5. Using a K-Fold cross-validation with K=5 means that the training set is split into 5 distinct subsets called folds, and then trains and make predictions on each model 5 times, picking a different fold for evaluation each time and training on the other 4 folds (Geron, 2017). Using K-Fold cross-validation with K=5 is common in practice because of its advantages from a computational standpoint when the dataset is large (Fushiki, 2009), and will be used in this study. The optimal penalty parameter for each shrinkage model is determined by importing the following packages from the *sklearn* library: *LassoCV*, *RidgeCV* and *ElasticNetCV*. After finding the optimal λ for each model, this value is then maintained across all forecast horizons. Additionally, for the ENet, the optimal ratio α between the ι_1 and ι_2 regularization is also determined using cross-validation with K-Fold in the same hyperparameter selection as the penalty parameter.

In this paper, to extract the optimal numbers of factors for the target factor model, a scree-test with eigenvalues will be performed. According to Yong and Pearce (2013), the optimal number of factors to extract can be performed by plotting a scree-test of the eigenvalues of the factors against the corresponding number of principal components. They continue by stating that the optimal number of factors to extract can be found by applying Kaiser's criterion, which is a criterion that suggests retaining all the factors with an eigenvalue of 1. After importing the *PCA* package and plotting the scree-tests for both datasets, the optimal number of factors to extract based on Kaiser's criterion from each dataset is 4 factors for the US dataset, and 5 factors for the Mexican dataset. The plots of the scree-tests for both datasets can be found in the Appendix.

For the SVR, RF and KNN, the optimal hyperparameters are determined by using the *GridSearchCV* package, which evaluates and test different values of the hyperparameters for each model, and then determines which hyperparameters are the best to use for optimal performance. After the optimal hyperparameters have been determined, these will then be maintained for each forecast horizon.

The optimal lag order for the AR(p) model is determined by plotting the Partial Autocorrelation Function from the *plot_pacf* package, which gives the partial correlation of a stationary time series with its lagged values after removing the effect on any correlations due to the terms of shorter lags, (Lazzeri, 2020). The optimal components for the ARMA model and the SARIMAX model can be determined by importing the *auto_arima* package, which tests out different combinations of the components and automatically selects the most optimal components for both models. The components for the time series models are then kept constant throughout all forecasting windows. The PACF plot for the inflation variable can be found in the Appendix for both datasets.

4.8. Evaluation Metrics

To evaluate and compare the different modeling frameworks, the Root Mean Squared Error (RMSE) is used to measure the average magnitude of the forecast errors. The formula for the RMSE can be written as:

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

4. 20

where $\hat{e}^2_{t,m,h}$ is the error of the forecast at a given time t, with a specific model *m* and with information up to time $t - h$. T_0 is the initial time at the beginning in the sample period. For all models used, it holds that a lower value of the RMSE statistics is better for forecast performance.

Additionally, the Mean Absolute Error (MAE) is reported for both the samples. The MAE can be formulated as:

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

4. 21

Which instead captures the average absolute difference between the actual and predicted values for each model. Just as for the RMSE, it holds that a lower value equates to a better forecast performance for each model.

5. Results

This section provides the results of the study by first providing and discussing the RMSE values obtained for the US dataset. Next, the Mexican RMSE values are presented and discussed. Finally, the test-results are compared and discussed for both datasets.

5.1. US Data Results

First, the US dataset is presented. The RMSE and MAE test results across every forecast horizon are presented in Table 1 and Table 2. We can see that the machine learning methods consistently outperform the benchmark models across all forecasting horizons. All the machine learning algorithms perform similar forecasting errors in relation to the RW model and the errors are consistent throughout all horizons. The best performing model for all horizons is the RF model, which has been seen to have been the case in some previous research using data for the US (see: Medeiros et al. (2021) and Heeren (2021)). The performance of the RF could be due to its benefits in handling nonlinearities and the underlying variables, which holds true for the SVR and KNN as well. For the shrinkage model, the RR performs worse than both the LASSO and ENet, who produces similar results due to similar feature variable selection for optimal results. The ENet and LASSO can handle irrelevant coefficients by shrinking them to zero, unlike the RR who only shrinks irrelevant variables towards zero. The results therefore reflect that there are some variables in the dataset who are irrelevant variables for predicting inflation.

Table 1: The ratio for the RMSE test statistics for different forecast horizons relative to the RW model for the sample 1960-2020. The best performing model for each horizon is highlighted in grey. The average RMSE ratio across all horizons is calculated in the last column (Ave).

	Forecasting horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Ave.
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.51	0.51	0.68	0.46	0.65	0.59	0.39	0.46	0.43	0.46	0.67	0.78	0.55
ARMA	0.50	0.51	0.68	0.48	0.64	0.62	0.43	0.44	0.46	0.47	0.65	0.78	0.55
SARIMAX	0.50	0.54	0.62	0.45	0.69	0.61	0.40	0.44	0.44	0.49	0.65	0.74	0.55
Factor Model	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
LASSO	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
RR	0.43	0.43	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.43	0.44	0.43
ENet	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
RF	0.31	0.30	0.30	0.29	0.29	0.29	0.28	0.27	0.27	0.26	0.27	0.26	0.28
KNN	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
SVR	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33

The benchmark models perform similar results on average in forecasting inflation, but give different forecast errors across the horizons. Overall, the results for the US dataset are expected, since the high-dimension of predictors in the dataset makes the use of machine learning algorithms more beneficial because the machine learning models included can use regularization and variable selection techniques to extract relevant predictors. Therefore, since there are many potential predictors to forecast inflation and the sample period is large consisting of many total observations, it is expected that the machine learning models can produce well-performing forecast results for the US dataset.

Table 2: The ratio for the MAE test statistics for different forecast horizons relative to the RW model for the sample 1960-2020. The best performing model for each horizon is highlighted in grey. The average MAE ratio across all horizons is calculated in the last column (Ave).

	Forecasting horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Ave.
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.70	0.70	0.91	0.63	0.88	0.79	0.54	0.62	0.58	0.62	0.90	1.05	0.74
ARMA	0.68	0.68	0.91	0.64	0.87	0.84	0.58	0.60	0.62	0.64	0.87	1.06	0.75
SARIMAX	0.68	0.73	0.84	0.61	0.93	0.82	0.54	0.60	0.59	0.66	0.88	1.00	0.74
Factor Model	0.34	0.34	0.34	0.34	0.34	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
LASSO	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
RR	0.46	0.47	0.48	0.47	0.47	0.47	0.46	0.47	0.46	0.47	0.46	0.47	0.47
ENet	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
RF	0.30	0.29	0.29	0.29	0.28	0.28	0.27	0.26	0.25	0.24	0.25	0.24	0.27
SVR	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
KNN	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30

5.2. Mexican Data Results

Secondly, the Mexican dataset is presented. The RMSE values relative to the RW are presented in Table 3 and the MAE relative to the RW are presented in Table 4. The forecasts for the Mexican dataset produce higher forecast errors across all models, and the benchmark models outperforms the machine learning algorithms across all horizons. The AR model performs best overall with the lowest average RMSE and MAE across all horizons. The benchmark models give more inconsistent forecast errors across the different horizons where they hold both the highest reported forecast error and the lowest reported forecast error. This may be due to seasonality patterns in the data for the inflation variable. Overall, the forecast errors are inconsistent for most models across the horizons, which could also be an indication that the sample size is too small for the model to be trained on, which can be an indicator that the models do not capture the relationship between the predictors and the inflation variable, producing worse results.

Table 3: The ratio for the RMSE test statistics for different forecast horizons relative to the RW model for the sample 1997-2017 using the Mexican dataset. The best performing model for each horizon is highlighted in grey. The average RMSE ratio across all horizons is calculated in the last column (Ave).

	Forecasting horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Ave.
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.32	0.04	0.27	0.47	0.42	0.66	0.93	0.26	0.45	0.43	0.13	0.16	0.38
ARMA	0.17	0.21	0.70	0.76	1.11	0.89	1.47	0.23	0.40	0.61	0.08	0.04	0.56
SARIMAX	0.09	0.29	0.81	0.08	0.75	1.38	1.56	0.34	0.67	0.46	0.43	0.35	0.60
Factor Model	0.65	0.65	0.65	0.65	0.66	0.65	0.66	0.66	0.67	0.65	0.65	0.65	0.65
LASSO	0.92	0.93	0.97	0.91	0.87	0.84	0.84	0.83	0.83	0.83	0.80	0.78	0.86
RR	0.84	0.85	0.85	0.84	0.84	0.80	0.80	0.80	0.83	0.83	0.82	0.80	0.83
ENet	0.66	0.66	0.66	0.67	0.67	0.67	0.68	0.68	0.67	0.66	0.66	0.65	0.67
SVR	0.89	0.92	0.88	0.82	0.85	0.86	0.82	0.82	0.83	0.82	0.85	0.79	0.85
RF	0.82	0.80	0.80	0.80	0.79	0.85	0.78	0.75	0.80	0.71	0.73	0.72	0.78
KNN	0.68	0.68	0.68	0.69	0.70	0.71	0.73	0.73	0.73	0.72	0.72	0.72	0.71

All machine learning algorithms outperform the naïve RW model across all horizons but do not outperform the benchmark models. The best performing machine learning algorithm for all horizons is the ENet, who can combine the LASSO and RR regularization terms, which can produce a more flexible approach to variable selection in order to select the best variables in forecasting inflation. The ENet gives the best MAE value for a 6-months ahead forecast. The weaker performance of the machine learning algorithms for the Mexican variables can be due to the lower dimension of the dataset, which makes the use of machine learning algorithms less beneficial. The Factor Model does not perform better than the benchmark models, but outperforms the machine learning models. The Factor Model works in a similar way to the machine learning models in the sense that it reduces the dimensionality of the dataset by extracting common factors from the feature variables.

Table 4: The ratio for the MAE test statistics for different forecast horizons relative to the RW model for the sample 1997-2017 using the Mexican dataset. The best performing model for each horizon is highlighted in grey. The average MAE ratio across all horizons is calculated in the last column (Ave).

	Forecasting horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Ave.
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.39	0.04	0.32	0.57	0.51	0.80	1.13	0.32	0.55	0.52	0.15	0.19	0.49
ARMA	0.20	0.26	0.85	0.92	1.34	1.08	1.78	0.28	0.48	0.73	0.10	0.04	0.67
SARIMAX	0.11	0.25	0.98	0.10	0.90	1.67	1.89	0.41	0.81	0.55	0.52	0.43	0.72
Factor Model	0.58	0.59	0.59	0.60	0.61	0.61	0.61	0.61	0.62	0.60	0.59	0.59	0.60
LASSO	0.87	0.89	0.93	0.88	0.84	0.81	0.81	0.79	0.81	0.81	0.76	0.74	0.83
RR	0.81	0.82	0.82	0.81	0.82	0.78	0.79	0.79	0.82	0.82	0.82	0.79	0.77
ENet	0.58	0.58	0.59	0.59	0.60	0.60	0.61	0.61	0.60	0.60	0.59	0.59	0.60
SVR	0.84	0.88	0.86	0.81	0.85	0.86	0.82	0.81	0.82	0.80	0.81	0.75	0.83
RF	0.78	0.74	0.74	0.75	0.74	0.78	0.73	0.71	0.73	0.65	0.65	0.63	0.72
KNN	0.60	0.63	0.60	0.62	0.64	0.64	0.65	0.65	0.65	0.64	0.64	0.64	0.63

5.3. Results comparison

Finally, the results for both the US dataset and the Mexican dataset are compared. For both datasets, all models across all horizons outperform the naïve RW model. However, the extent to that varies across the datasets. The US dataset is performing better and showing less forecast error both in terms of RMSE and MAE values compared to the Mexican dataset. Additionally, all the machine learning models perform much better for the US dataset, where the dimensionality is higher with more potential predictors to forecast inflation. The higher dimensionality of the US dataset makes the benefits of machine learning algorithms more relevant. Another aspect to why the machine learning algorithms is better suited than the benchmark models for forecasting inflation using the US dataset can be attributed to the multicollinearity aspect of the dataset. The US dataset consists of a lot of predictors which are similar to each other, hence highly correlated to each other, which can be mitigated by applying machine learning techniques (Chan et al. 2022). Chan et al. continues by stating that machine learning algorithms tend to perform better with datasets consisting of high degrees of multicollinearity due to the effectiveness of machine learning techniques and its built-in features in handling multicollinearity by variable selection and dimensionality reduction. The machine learning models also seem to benefit from the increased complexity of the dataset, where machine learning models better can identify these patterns, such as nonlinearity between the inflation variable and the predictors to produce more accurate forecasts compared to the benchmark models.

For the Mexican dataset, the machine learning algorithms does not outperform the benchmark models in forecasting inflation. One possible explanation as previously mentioned is the variable selection, where the potential predictors of forecasting inflation are significantly less than the US dataset. The sample period is also significantly smaller, which affects the model in terms of the training and test size aspect of the forecasting. With less historical data to work with, the models may not be able to capture the underlying factors for forecasting inflation as optimally as possible, and the forecast results suffer from it. Additionally, forecasting inflation in an emerging market can be more challenging in general, and Duncan & Garcia (2019) point out several factors that make forecasting inflation harder for emerging markets. They point out that emerging economies have experienced changes in their macroeconomic dependencies attributed to changes in polices and globalization. They also point out that emerging economies are more exposed to external shocks, and changes in global factors such as commodity prices and exchange rates have a larger impact on domestic inflation. These aspects make the variables more volatile over time. These factors can influence inflation forecasting and make it harder overall to produce good forecast results.

To summarize the results for the two datasets, all models outperform the naïve RW model across every horizon. For the US dataset, the machine learning models outperforms the benchmark time series models and produces lower forecast errors on average compared to the Mexican dataset, in which the machine learning models do not outperform the benchmark models on average in forecasting inflation. Additionally, the forecast errors are similar on average for most models on average across all horizons, which can be attributed to the robustness imposed by using a fixed window across all horizons.

6. Discussion

This paper investigates and compare the performance of forecasting inflation for an emerging economy, Mexico, and a developed economy, the US. The inflation forecast was made using a rolling window approach for 1-12 months ahead in the future, using a fixed window size for traditional benchmark models, a target factor model, and machine learning models. The forecast evaluation was made by comparing the RMSE values and the MAE values across all forecast horizons. To conduct optimal results for the forecast, hyperparameter selection is made by applying cross-validation tests for the machine learning models, and a scree-test with Kaiser's criterion for the target factor model. The optimal components of the benchmark models are determined by plotting the PACF, and by using an automized Python package for optimal model component selection.

The methodology described above are applied on two datasets, the FRED-MD dataset for the US, which is a high-dimensional dataset consisting of monthly macroeconomic variables, and a dataset of monthly Mexican macroeconomic variables, gathered from the Federal Reserve database and from the Banco de Mexico database. Both these datasets also contain the inflation variable, measured as the logarithm-difference of the Consumer Price Index (CPI). Additionally, the study was attempted to answer the following two questions:

- Do machine learning methods outperform traditional time series models?
- How does the evaluation of time series models and machine learning methods compare for an emerging economy and a developed economy?

To answer the first question, machine learning methods was seen to outperform the traditional benchmark time series models for the high-dimensional US dataset, but not for the Mexican dataset. The intuitive reason behind these results can be attributed to the difference in dimensionality for the datasets. The US dataset contain both more potential predictors in forecasting inflation, and the dataset contain more total monthly observations for each variable, which suits the machine learning models better, since these models are better suited when applied to a more complex dataset, where variable selection and regularization can be a useful tool in extracting relevant features to optimize forecast performance. For the Mexican dataset, the potential predictors and monthly observations are fewer in total, which creates less complexity of the data and therefore less benefits of using machine learning models, and as a result, the machine learning models does not outperform the benchmark models. To answer the second question, it is seen that the performance of forecasting inflation is better for the developed economy compared to the emerging economy. The reported RMSE and MAE values for the US dataset are lower across all forecast horizons and models. This can also be attributed to the characteristics of the dataset, where a larger sample size and more potential predictors compared to the Mexican dataset makes the forecast errors smaller for the models. Additionally, forecasting inflation is generally harder for emerging economies due to more volatile movements of the predictors across the sample period, which makes it harder to identify patterns in the data for forecasting performance. Overall, the results are in line with the existing literature that machine learning models was seen to outperform benchmark time series models for the US dataset, but machine learning models did not outperform the benchmark models for the Mexican dataset, going against the existing literature.

The best machine learning model overall in this study was seen to be the RF model, mainly due to its ability to identify complex patterns such as nonlinearities in the data, and its ability to reduce importance of less significant variables in forecasting inflation, putting more weight on variables that have more impact on inflation. As for macroeconomic policymakers, based on the results from this study, machine learning models can be useful tool for policymakers in forecasting, due to its abilities to work in a complex environment, in which forecasting inflation using potential predictors are. However, as seen in this study, to optimally use the machine learning models, sufficient macroeconomic data needs to be available, which is more applicable to developed economies.

There are several ways in which further research can be extended for inflation forecasting. First, a different set of machine learning models could be examined. Machine learning algorithms such as Neural Networks, Complete Subset Regressions, and Deep Learning models such as Long-Short-Term-Memory can all be used to forecast inflation and would provide some valuable extensions on the subject. With more time at hand, more variables can be extracted for an emerging economy to increase the dimensionality of the dataset to improve performance. Another interesting extension of the study would be to forecast with another window approach. By changing from a rolling window to an expanding window, the forecast results could be different. The size of the window and the split of the training and test data could also be examined in more detail to improve the forecast performance. Other emerging economies could also be examined and be compared with to the case of Mexico, to see if the forecast results for this study holds for other emerging economies.

7. Reference list

Athey, S. & Imbens G.W. (2019). “Machine Learning Methods that Economists Should Know About.” *Annual Review of Economics*. 11: 685–725.

Artis, M., Banerjee, A. & Marcellino, M. (2001). “Factor forecasts for the UK”. *Bocconi University – IGIER*. Manuscript.

Bai, J. & Ng, S. (2003), “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–171.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45 (1), 5–32.

Chan, J.Y.-L.; Leow, S.M.H.; Bea, K.T.; Cheng, W.K.; Phoong, S.W.; Hong, Z.-W.; Chen, Y.-L. (2022). “Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review”. *Mathematics*, 10, 1283.

Charpentier, A., Flachaire, E., & Ly, A. (2018). “Econometrics and machine learning”. *Economics and Statistics*, 505(1):147–169.

Dickey, D. A., & Fuller, W. A. (1979). “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”. *Journal of the American Statistical Association*, 74 (366), 427-431.

Duncan, R. & Martínez-Garcìa, E. (2019), “New perspectives on forecasting inflation in emerging market economies: an empirical assessment”, *International Journal of Forecasting*, Vol. 35 No. 3, pp. 1008-1031.

Fix, E. and Hodges, J. (1951) “Discriminatory analysis. Nonparametric discrimination: Consistency properties”. *USAF School of Aviation Medicine, Project Number.* 21-49-004, Report Number: 4.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2003), “Do financial variables help forecasting inflation and real activity in the euro area?”, *Journal of Monetary Economics*, Vol. 50 No. 6, pp. 1243-1255

Fushiki, T. (2011). “Estimation of prediction error by using k-fold cross-validation,” *Statistics and Computing*, vol. 21, no. 2, pp. 137–146.

Garcia, M.G., Medeiros, M.C. & Vasconcelos, G.F. (2017), "Real-time inflation forecasting with high-dimensional models: the case of Brazil", *International Journal of Forecasting*, Vol. 33 No. 3, pp. 679-693.

Géron A. (2017). *Hands-on machine learning with Scikit-Learn, Keras. In: and TensorFlow: concepts, tools, and techniques to build intelligent systems*. First edition. O'Reilly Media.

Groen, J., Paap, R., & Ravazzolo, F. (2013), "Real-Time Inflation Forecasting in a Changing World," *Journal of Business and Economic Statistics*, 31, 29– 44.

Hall, A. S. (2018), "Machine Learning Approaches to Macroeconomic Forecasting," *The Federal Reserve Bank of Kansas City Economic Review*, 103, 63.

Hastie, T., Tibshirami, R., & Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Second Edition. New York: Springer.

Heeren, S. (2021). "Forecasting inflation using Machine Learning for an Emerging Economy". *Econometrie*.

Hoerl, A. E. & Kennard, R. W. (1970). "Ridge Regression: Biased Estimation for Nonorthogonal Problems". *Technometrics*, 12(1):55–67.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. Journal of Econometrics, 196(1), 55–67.

Iversen, J., Laséen, S., Lundvall, H. & Soderstrom, U. (2016), "Real-time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank", *Riksbank. Research Paper Series, Stockholm*, Vol. 142.

Joseph, V. R. (2022). "Optimal ratio for data splitting". *Statistical Analysis and Data Mining: The ASA Data Science Journal.*

Lazzeri, F. (2020). *Machine learning for time series forecasting with Python*. First Edition. John Wiley & Sons.

McCracken, M. W., & S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34(4), 574–589.

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, A., & Eduardo, Z. (2021). "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods". *Journal of Business & Economic Statistics*, 39 (1), 98–119.

Moshiri, S., & Cameron, N. (2000). "Neural Network Versus Econometric Models in Forecasting Inflation". *Journal of Forecasting*, 19 (3), 201–217.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. First Edition. MIT Press.

Phillips, A. W. (1958). "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom", 1861-1957. *Economica*, 25 (100), 283–299.

Pichler, P. (2008), "Forecasting with DSGE models: the role of nonlinearities", *The B.E. Journal of Macroeconomics*, Vol. 8 No. 1, pp. 1-33.

Stock, J., and Watson, M. (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

Stock, J. H., & Watson, M. W. (2002). "Macroeconomic Forecasting Using Diffusion Indexes". *Journal of Business & Economic Statistics*, 20 (2), 147-162.

Svensson, L. E. (1997). "Inflation forecast targeting: Implementing and monitoring inflation targets". *European Economic Review*, 41 (6), 1111–1146.

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society: Series B* (Methodological) 58(1), 267–288.

Varian, Hal R. (2014). "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–28.

Woodford, M. (2007). "The Case for Forecast Targeting as a Monetary Policy Strategy". *Journal of Economic Perspectives*, 21 (4), 3–24.

Yong, A. G., & Pearce, S. (2013). "A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis". *Tutorials in Quantitative Methods for Psychology*, 9(2), 79-94.

Zou, H., and Hastie, T. (2005), "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society, Series B*, 67, 301–320.

Özgur, O., & Akkoc, U. (2021). "Inflation forecasting in an emerging economy: Selecting variables with machine learning algorithms". *International Journal of Emerging Markets.*

8. Appendix

Table 5: Variable names and corresponding transformations for the Mexican dataset. The transformation codes represent (1) no transformation, (2) first difference of the variable, (3) first difference of the log of the variable, (4) second difference of the log of the variable.

	Description of variable	Transformation Code	Source	Series ID
1	Monthly inflation		FRED	CPALCY01MXM661N
2	Mexican Stock Market Index, General Index	3	Banco de Mexico	CF57
3	Monetary Base	3	Banco de Mexico	CF2
4	Currency in Circulation	3	Banco de Mexico	SF1
5	Cash in vault	3	Banco de Mexico	SF5
6	CLI: Trend	2	FRED	MEXLOLITOTRSTSAM
7	International Reserves	3	Banco de Mexico	SF7
8	Financial sources (Total)	3	Banco de Mexico	SF321170
9	Assets and Liabilities, Credit	3	Banco de Mexico	SF337
10	Credit to Banks	3	Banco de Mexico	SF343
11	Net International Assets	3	Banco de Mexico	SF29654
12	91-day Cetes	3	Banco de Mexico	SF17801
12	Weighted average government funding rate	3	Banco de Mexico	SF17864

13	Real Exchange Rate Index	2	Banco de Mexico	SR28
14	External Price Index	2	Banco de Mexico	SR464
15	INPC, Core, Merchandise	4	Banco de Mexico	SP74626
16	INPC, Core, Services	4	Banco de Mexico	SP74628
17	INPC, Non-Core	4	Banco de Mexico	SP74630
18	Public Sector Budgetary Expenditures	4	Banco de Mexico	SG1
19	Total Debt of the Public Sector	3	Banco de Mexico	SG193
20	Internal Debt of Public Sector	3	Banco de Mexico	SG194
21	External Debt of Public Sector	3	Banco de Mexico	SG195
22	Total Exports	3	FRED	XTEXVA01MXM664S
23	Total Imports	3	FRED	XTIMVA01MXM664S
24	M1 for Mexico	3	FRED	MANMM101MXM189S
25	Current Account Deposits in Resident Banks	3	Banco de Mexico	SF12721
26	Pension Funds	3	Banco de Mexico	SF10478
27	Financial Savings including Public Sector Savings	3	Banco de Mexico	SF221973
28	Monetary Aggregates including public sector	3	Banco de Mexico	SF13724

29	Federal Government securities	3	Banco de Mexico	SF13764
30	Domestic Private Consumption, Total	3	Banco de Mexico	SR16558
31	Production: Total Energy	2	FRED	MEXPRENTO01IXOBSAM
32	Production: Total Industry excluding construction	2	FRED	MEXPRINTO01GPSAM
33	Production: Total Industry including construction	2	FRED	MEXPRINTO02IXOBSAM
34	Production: Total Manufacturing	1	FRED	MEXPROMANMISMEI
35	Labor Compensation: Earnings	2	FRED	MEXLCEAMN01GPSAM
36	Total Retail Trade	2	FRED	MEXSARTMISMEI
37	Sales: Wholesale Trade	2	FRED	MEXSLWHTO01IXOBSAM
38	Harmonized Unemployment Rate	2	FRED	LRHUTTTTMXM156S

Figure1: The PACF plot of the inflation variable for the Mexican dataset with a maximum of 20 lags.

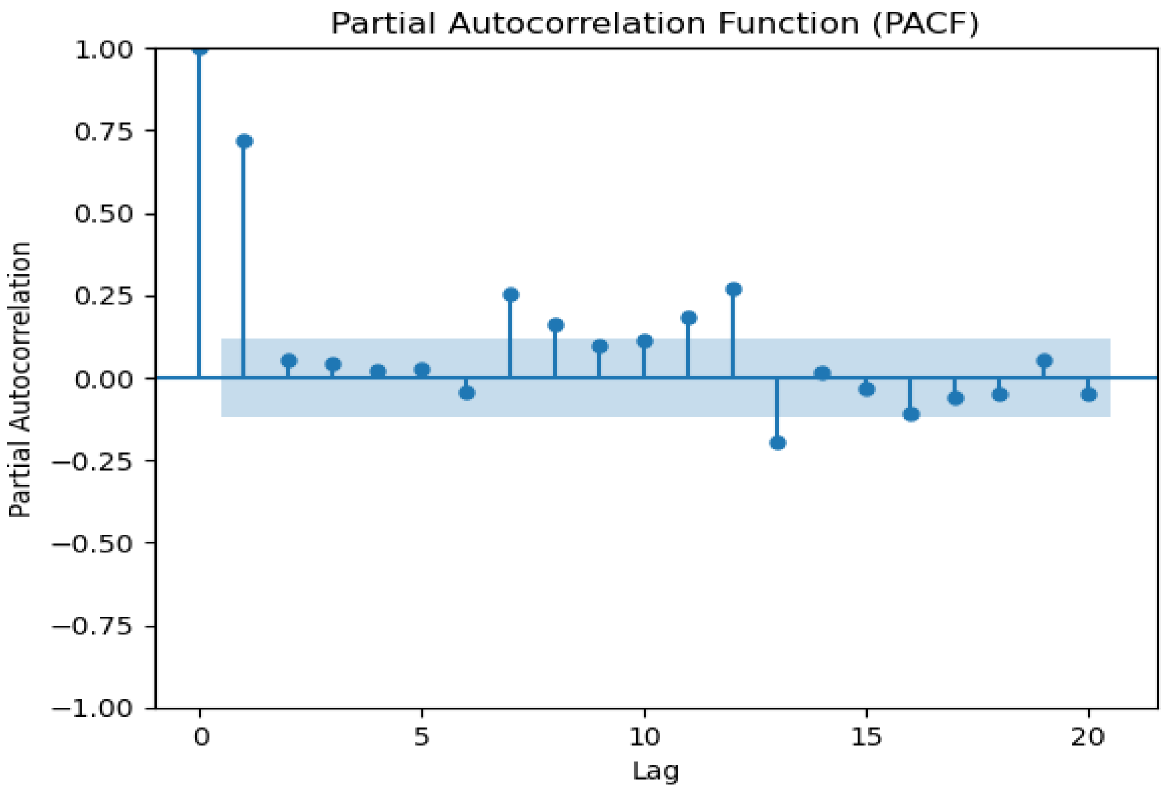


Figure 2: The PACF plot of the inflation variable for the US dataset with a maximum of 20 lags.

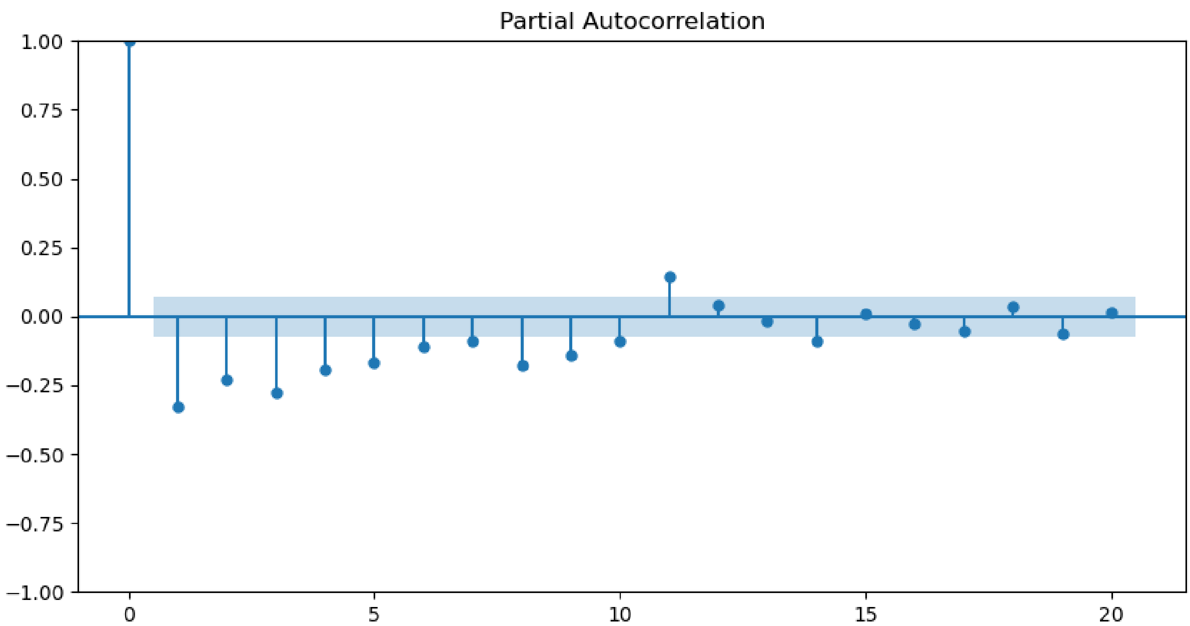


Figure 3: The Scree plot for the Mexican predictors to determine the optimal number of factors to extract for the target factor model.

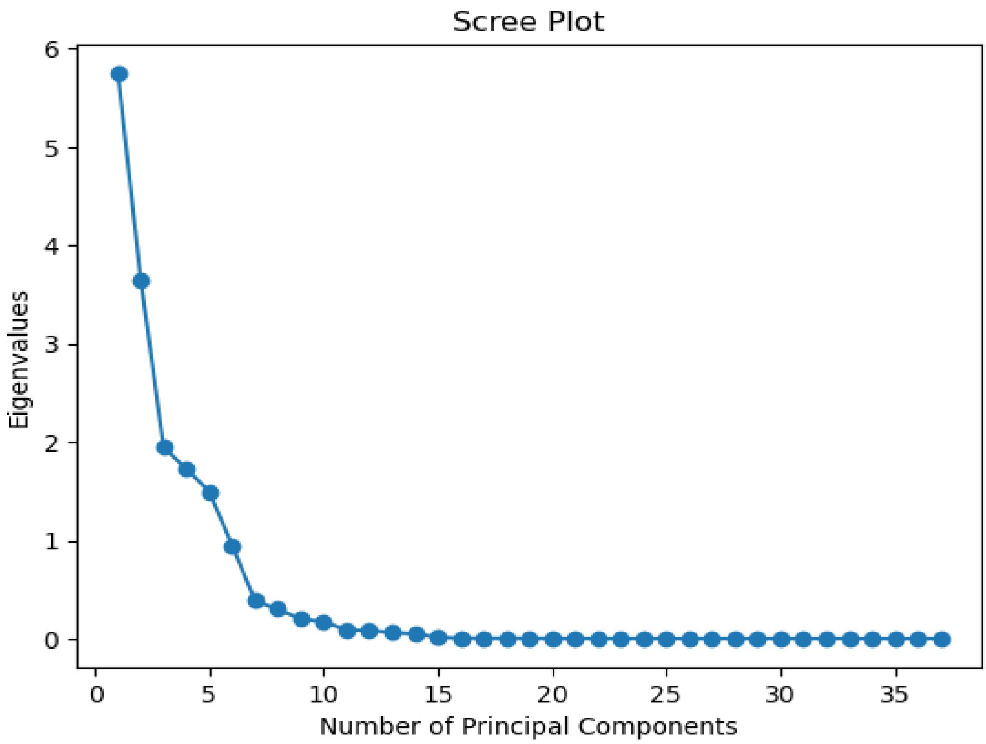


Figure 4: The Scree plot for the US predictors to determine the optimal number of factors to extract for the target factor model.