ERASMUS UNIVERSITY OF ROTTERDAM
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

Forecasting Inflation using Machine Learning for an Emerging Economy

Name student: Sven Heeren
Student ID number: 499741sh
Supervisor: Eoghan O'Neill
Second Assessor: dr. A.A. Naghi

Date final version: July 4, 2021

Abstract

In this paper, the performance of machine learning methods in forecasting inflation is evaluated. This is done for a developed economy, which is the United States, as in Medeiros, Vasconcelos, Veiga, and Eduardo (2021) and for an emerging economy, which is Mexico. In addition to the models used in Medeiros et al. (2021), two more machine learning models are added, namely the Bayesian additive vector autoregressive tree (BAVART) model and the long short-term memory network (LSTM) model. The results show that for both types of economies, the machine learning methods increase the forecast accuracy significantly. The best model found in this paper is the random forests model. The BAVART model does not increase the accuracy of the forecasts relative to the random forests model but does outperform a number of other models used. The LSTM model shows very poor performance for forecasting inflation in both types of economies, even worse than the random walk model.

Contents

1 Introduction

In recent decades, several central banks have adopted a monetary policy strategy called inflation-targeting monetary policy. Among others, the Bank of England, Sweden's Riksbank and the Reserve Bank of New Zealand have now embraced this strategy (Woodford, 2007), mostly with an inflation rate of 2% as the target. However, according to Svensson (1997), implementing and monitoring this monetary policy have serious complications. Svensson (1997) states that the solution to these problems is to have inflation forecasts as intermediate targets of the monetary policy. It is thus of major importance that the inflation forecasts are as accurate as possible, as better inflation forecasts lead to better outcomes. On top of this, it takes some time before the results of monetary policy are seen in the real inflation, so some decisions are based upon inflation forecasts, according to Haan, Hoeberichts, Maas, and Teppa (2016). Both these things lead to the conclusion that forecasting inflation accurately is of significant importance.

To this end, multiple methods have been used to forecast inflation. The most traditional method is based on the Phillips curve as introduced by Phillips (1958) and extended in the inflation forecasting environment by Stock and Watson (1999). After that, lots of other methods have been developed based on the univariate time series theory, for example, the autoregressive model, which assumes that the future inflation depends on the lags, or the more sophisticated autoregressive integrated moving average models. In recent years, a new set of models have been used more and more, namely machine learning methods. These methods can handle large amounts of data very efficiently, according to Zhoua, Pana, Wanga, and Vasilakos (2017). This is of great significance as the amount of data available has exploded in the recent decade, which is sometimes even called Big Data. However, the traditional methods are not appropriate to handle a large number of variables and data and in this aspect, machine learning methods could thus improve upon traditional methods. On top of this, according to Moshiri and Cameron (2000), the traditional models often assume linearity even if that assumption is not appropriate. There are machine learning methods that do not make such an assumption and in this aspect, these methods may improve upon the traditional methods. Therefore, in this report, an extensive set of machine learning methods is compared with the more traditional methods for forecasting inflation. These machine learning methods consist of shrinkage methods, factor models, ensemble methods, random forests, regression trees and neural networks. In contrast to the paper of Medeiros et al. (2021), two additional machine learning methods are used, namely the long short-term memory network (LSTM) model and the Bayesian additive vector autoregressive tree (BAVART) model. The benchmark methods used in this paper are the autoregressive model, the random walk model and the autoregressive integrated moving average model. These methods are then compared using different test statistics regarding the forecast errors.

Quite some papers have already been produced which test a part of these methods for the United States inflation, such as Medeiros et al. (2021) and Ülke, Sahin, and Subasi (2018). These papers show that machine learning methods do generally increase the accuracy of the inflation forecasts. However, the question which arises is whether this conclusion is valid in other economies as well, such as emerging economies instead of the developed US economy. Quite some research has already been done regarding inflation targeting in emerging economies (e.g. Mandalinci, 2017 and Önder, 2004). However, this research does not take into account the machine learning methods or the research that has been done on this only uses a very restrictive set of methods, for example, Özgür and Akkoç (2021) for Turkish inflation and Garcia, Medeiros, and Vasconcelos (2017) for Brazilian inflation. Hence, in this report, a more extensive list of machine learning methods are used and compared with the purpose of forecasting inflation in an emerging economy, specifically the Mexican economy. Therefore, the following question is answered: "Do machine learning methods improve upon benchmark methods in forecasting inflation for an emerging economy?". This question is answered with the help of some other questions, namely "What is the difference in properties between US and Mexican inflation?", "Are the conclusions of the ranking of models the same in an emerging economy (Mexico) when compared to the US market data?", "Do the new machine learning methods, which are LSTM and BAVART, perform better than the best machine learning method up to this point, namely the random forests model?".

This research finds that the biggest difference between the US and Mexican inflation is the fact that Mexican inflation is much more volatile than US inflation. However, for both data sets the random forests model performs superior to all other models in most cases. However, the Bayesian vector autoregression model shows a very strong performance on the US inflation data, which is not the case for the Mexican data. Overall, most machine learning methods increase the accuracy of inflation forecasts relative to the methods which are not machine learning methods. Even though the random forests model is superior, the ranking of the other models differs between data with high and low volatility. On top of that, the newly introduced BAVART model performs worse than the random forests model, but it outperforms some of the machine learning methods. However, the LSTM model performs very badly and even worse than the random walk model for both data sets.

The report is organized as follows. In Section 2 the already existing literature on this topic is explored. Section 3 describes the data and the way it is obtained and transformed. Section 4 presents the forecasting models and evaluation criteria and Section 5 presents the results. Finally, Section 6 concludes and discusses the report and gives recommendations for further research.

2 Literature

Finding an accurate way to estimate forecasts of inflation has been a research topic that has been studied extensively over the last couple of decades.

The oldest model which has been used extensively is the Phillips curve, as in Stock and Watson (1999). The Phillips curve shows the relationship between inflation and unemployment and this is then used to predict inflation. However, empirical research on 15 years of US inflation data by Atkeson and Ohanian (2001) has shown that the forecasts based on the Phillips curve are not even more accurate than the forecasts from a random walk model. Therefore this model is not used in this research.

Other models which have been around for quite some time are the univariate time series models, such as the autoregressive (AR) or the autoregressive integrated moving average (ARIMA) model. However, as Stock and Watson (2002) and Medeiros et al. (2021) show, the AR model does not perform optimally as both papers indicate that there are certain models which perform significantly better. However, as this model is one of the widely used models in the forecasting literature it is still used in my paper. The paper of Stock and Watson (2002) shows that the usage of factor models improve the performance of the models. Factor models are models which try to extract common components (factors) of the data to describe the data. These factor models are thoroughly explained in Bai (2003). The advantage of these models is that they reduce the number of predictors because usually, the number of factors is much smaller than the number of predictors.

The methods proposed by Medeiros et al. (2021) which beat the AR model are the class of machine learning methods when analysing US inflation. They show that especially the random forests model performs very well, but also shrinkage methods, such as LASSO and ridge regression perform very well. The same conclusion concerning the superior performance of machine learning methods is drawn by Ülke et al. (2018). However, the paper by Ülke et al. (2018) uses other machine learning methods, namely neural networks and the support vector machine method. Therefore, in my paper, a neural network method is added to the methods of Medeiros et al. (2021), namely the long short-term memory network (LTSM) model, as introduced by Hochreiter and Schmidhuber (1997).

This model beats the ARIMA model in terms of forecasting time series, according to Siami-Namini, Tavakoli, and Siami Namin (2018). They even show that the rate of reduction of the errors when going from the ARIMA to the LSTM model is between 84 and 87 per cent.

Another method commonly used by central banks is the Bayesian vector autoregressive (BVAR) model. This model is a dynamic multivariate model to forecast inflation, but it has received substantial criticism in recent decades. The reason for this criticism is mainly the linearity of the model. However, macroeconomic and financial variables often have non-linear dependency between them, according to Granger and Terasvirta (1993). Therefore, something to capture these nonlinearities has to be incorporated into the VAR model to make it more accurate. That is where the Bayesian additive regression tree (BART) models come into play, as these show a strong empirical performance to forecasting, according to Prüser (2019). The model used in this report is a model which combines the previous two models, namely the Bayesian additive vector autoregressive tree (BAVART) model of Huber and Rossini (2021). They state that the model is highly flexible to different types of dependencies between variables.

My paper also especially focuses on machine learning methods to forecast inflation in an emerging economy, namely Mexico. Quite some research has already been done on forecasting inflation in emerging markets, for example, Mandalinci (2017) and Önder (2004). However, these papers did not take into account the machine learning methods. There has not been a lot of research on the performance of machine learning methods to forecast inflation in emerging economies. Some studies which have already tried to analyse this are Özgür and Akkoç (2021) and Garcia et al. (2017), which found that machine learning methods also perform very well to make forecasts of inflation in emerging markets. However, these papers did not consider a lot of machine learning methods and in my paper, a more extensive list is used.

3 Data

The first data set used in the report contains 122 macroeconomic variables from the United States. These are variables that do not contain missing values for the sample period from January 1960 to December 2015 (672 observations). A data set provided by one of the authors of Medeiros et al. (2021)[1], containing the same variables as the FRED-MD database, is used. However, in contrast to the variables in the FRED-MD database, the variables have already been transformed such that the

[1]https://github.com/gabrielrvsc/ForecastingInflation

variables are stationary. This is important as most models which are used for forecasting implicitly assume that the variables are stationary. The performance of the models is evaluated over two sample periods, namely from January 1990 to December 2000 (132 observations) and from January 2001 to December 2015 (180 observations). Results are eventually obtained for both sample periods separately. To get the results for the complete sample period from January 1990 to December 2015, the average is taken over the results of the two periods. The forecasts are made based on a rolling window, where it holds that for the first date from the sample period all previous data is used to train the model. Then for every next date this training sample is shifted one month. This means that for the forecasts for the sample period from January 1990-December 2000, a training period of 30 years is used, and for the period from January 2001 to December 2015, a training set of 40 years of data is used. On top of the 122 variables, 4 lags of each variable and 4 autoregressive terms are used as potential predictors which eventually leads to a total of 508 potential predictors.
The variable which is forecasted in this paper is the United States inflation. To get the values of the US inflation in the data set, the Consumer Price Index (CPI) is used as a price index. In contrast with the paper of Medeiros et al. (2021), the Personal Consumption Expenditures (PCE) price index and the core CPI are not used to construct inflation. After obtaining the US CPI, the inflation in a certain month t is computed as $\pi_t = log(P_t) - log(P_{t-1})$ where P_t is the value for the CPI in month t.

The second data set that is used in this paper is a data set that contains Mexican macroeconomic variables. The reason that Mexico is chosen as an appropriate data set is that there is sufficient data available for Mexico, which is also easily accessible. Another reason to use this data set is to test the inflation forecasting methods in the case of a country with an emerging economy (according to Russell (2020)) instead of the United States, whose economy is highly developed. The Mexican data is downloaded from the Banco de Mexico database[2]. The biggest difference between the Mexican and the US data set is the length of the sample period and the number of variables. The sample period for the Mexican data goes from January 2001 to January 2021 (253 observations), as most of the variables are only available from 2001 onwards. The sample period on which the performance is tested contains about half of the observations from February 2011 to January 2021 (120 observations), so that there are enough forecasts to have a reliable say about the relative performance of the models. As with the US data, the forecasts are made based on a rolling window, where the training data consist of 10 years of data. This is thus much less than the 30 and 40 years

[2]https://www.banxico.org.mx/SieInternet/defaultEnglish.do

of training data in the US case, so the performance of models in a case with less training data is also evaluated.

The variable which is forecasted in this paper is Mexican inflation. To get this variable, the Mexican Consumer Price Index (INPC) is used which is also provided by Banco de Mexico. Just as in the US case, the inflation at month t is then computed as $\pi_t = log(P_t) - log(P_{t-1})$ where P_t is the value for the INPC in month t.

The Mexican data consists of 32 variables in which the inflation variable is included as well. This number of variables is significantly less than the number of variables used in the US data set. The reason that this number of variables has been chosen has to do with the fact that these were the variables that had no missing values from January 2001 to January 2021. This also adds to this research to see how the models perform when fewer variables are used. The variables are also chosen which resemble one or some variables from the US data set so that there is some similarity between the two. A complete list of the variables is displayed in Table A1 in the Appendix.

Some characteristics of the inflation variables for the sample periods for which the performance is tested are displayed in Table 1. This is done such that the difference between the different sets becomes apparent and to see if the ranking of the models stays the same between the different periods. It can be observed that the Mexican data is more volatile than both US periods and than the two US periods combined. It is thus of interest to see whether the machine learning methods outperform the other methods for the periods with high and low volatility. As the US and Mexico are neighbouring countries, it could be that the inflation movements are really close to each other. However, this turns out not to be the case. A graph of both inflation time series is displayed in Figure A1.

Period	Standard Deviation	Mean	Max	Min	Range
US (Jan 1990-Dec 2015)	0.2684	0.2031	1.3675	-1.7864	3.1539
US (Jan 1990-Dec 2000)	0.2453	0.1675	0.9456	-0.0585	1.0041
US (Jan 2001-Dec 2015)	0.3200	0.1722	1.3675	-1.7864	3.1539
Mex (Feb 2011-Jan 2021)	0.3544	0.3561	1.6859	-1.0189	2.7048

Table 1: Characteristics of sample periods for which the performance is evaluated for forecasting the US and Mexican inflation data. All numbers are give in percentage (%).

The Mexican data also has to be stationary, just as the US data. To test whether the variables are stationary or not, the augmented Dicky-Fuller (ADF) test is used as first introduced by Dickey and Fuller (1979). This tests whether a unit root is present in the time series of a certain variable. In this test, the null hypothesis is that the data is non-stationary, which is tested against the alternative hypothesis of stationarity. Thus what is needed in this case is an ADF test statistic with a p-value smaller than 0.01 so that the null hypothesis could be rejected in favour of the alternative hypothesis of stationarity. Performing this test on the original data shows that only 4 variables are stationary without any transformations. This means that transformations have to be made on the other variables. For the data set, it is sufficient to use one of the following two transformations:

- Take the first difference of the variable, so for a variable x at month t, the transformed variable z at month t is computed as $z_t = x_t - x_{t-1}$.
- Take the first difference of the logarithm of the variable, so so for a variable x at month t, the transformed variable z at month t is computed as $z_t = log(x_t) - log(x_{t-1})$.

After performing one of these transformations on the variables, all the p-values of the ADF test statistics are lower than 0.01, which indicates stationarity. Which of the two transformations is used for which variable can be found in Table A1.

4 Methodology

In this report, the same notation is used as in Medeiros et al. (2021). That means that the general specification for a model is

$$\pi_{t+h} = G_h(x_t) + u_{t+h}, \tag{1}$$

where $h = 1, \ldots, H$ and $t = 1, \ldots, T$. The π_{t+h} is the inflation in month $t+h$ and $x_t = (x_{1t}, \ldots, x_{nt})$ is a vector with the n covariates and potential predictors used in the model. $G_h(x_t)$ is some target function depending on the model and u_{t+h} is some zero-mean random error. Furthermore, there is a direct forecast equation which is

$$\hat{\pi}_{t+h|t} = \hat{G}_{h,t-R_h+1:t}(x_t), \tag{2}$$

with $\hat{G}_{h,t-R_h+1:t}(x_t)$ uses data from time $t - R_h + 1$ up to t with R_h the window size. The covariates are not forecasted, except for the BVAR, the BAVART and the ARIMA model, which are discussed later. In the AR model, the other covariates are not used as this is a univariate model.

4.1 Benchmark Models

Some benchmark models are used against which the machine learning models are tested. The most basic model is the random walk (RW) model. The forecasts in this model are $\hat{\pi}_{t+h|t} = \pi_t$ for $h = 1, \ldots, 12$, so this means that the forecast is just the value of the inflation at time t which is the same as in Medeiros et al. (2021). On top of this, the autoregressive (AR) model of order p is used, as also done by Medeiros et al. (2021). The p is first chosen based on the Bayesian information criterion. Then OLS is used to compute the parameters over the in-sample period and then forecasts are made based on $\hat{\pi}_{t+h|t} = \hat{\rho}_{o,h} + \hat{\rho}_{1,h}\pi_t + \ldots + \hat{\rho}_{p,h}\pi_{t-p-1}$. A model for which all the covariates are forecasted is the Bayesian vector autoregressive (BVAR) model, as in the paper of Bańbura, Giannone, and Reichlin (2010). On top of these three models, one more benchmark model is added to this paper, namely the autoregressive integrated moving average (ARIMA) model, as in Hyndman and Khandakar (2008).

4.2 Machine learning models

Machine learning models are used which are from the paper by Medeiros et al. (2021) and two more models are added, namely the BAVART and the LSTM model. All the models are explained in this section, where for the models used in Medeiros et al. (2021), the values of fixed constants are taken as in that paper, for example, the number of bootstrapping samples in a certain method.

4.2.1 Shrinkage Models

For the shrinkage models, as in Medeiros et al. (2021), the model is $G_h(x_t) = \beta_h' x_t$ where

$$\hat{\beta}_h = \arg\min_{\beta_h} [\sum_{t=1}^{T-h} (y_{t+h} - \beta_h' x_t)^2 + \sum_{i=1}^{n} p(\beta_{h,i}; \lambda, \omega_i)], \tag{3}$$

where λ and ω are the parameters of the penalty function $p(\beta_{h,i}; \lambda, \omega_i)$. The shrinkage methods differ in their choice of the penalty function. Consider the general formulation of the penalty function

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

the methods differ when it comes to choosing the parameters. The shrinkage methods are:

- Ridge regression (RR) puts $\alpha = 1$.
- Least absolute shrinkage and selection operator (LASSO) puts $\alpha = 0$ and all weights equal one, so $\omega_i = 1$ for $i = 1, \ldots, n$.

- Adaptive least absolute shrinkage and selection operator (adaLASSO) puts $\alpha = 0$, but treats ω_i as a parameter to be estimated.
- Elastic net (ElNet) puts $\alpha = 0.5$ and all weights equal one, so $\omega_i = 1$ for $i = 1, \dots, n$.
- Adaptive elastic net (adaElNet) puts $\alpha = 0.5$, but treats ω_i as a parameter to be estimated.

4.2.2 Ensemble Models

All of the ensemble models have in common that these are weighted averages over a set of forecasts. All of these models are used as in Medeiros et al. (2021).

First of all, the bagging model is used as a bagging model of OLS after variable selection. It makes forecasts for $B = 100$ bootstrap samples and then takes the average of the B forecasts as the value of the forecast computed by bagging. Every bootstrap iteration, the methods starts by performing an OLS regression with all variables (could be separated in groups when the number of variables is larger than number of observations) and then picks the variables with a t-statistic above a certain value c. Then a new OLS regression is estimated using only these variables after which that model is used to produce a forecast for that one bootstrap sample.

The second ensemble method is the complete subset regressions (CSR) method. This method first performs a linear regression of the inflation variable on every covariate individually (with lags) and then picks the $n = 20$ covariates with the highest t-statistic values in absolute terms. Then on these n variables forecasts are made using all possible different subsets with size $q = 4$ of the n variables using a linear regression. After this the average over the forecasts is taken to get the final forecast for this model.

The final ensemble method which is used is the jackknife model averaging (JMA) model. The difference between this model and the other two ensemble models is that JMA does not use the simple average over the forecasts. Instead, it computes optimal weights of the model forecasts using a cross-validation procedure as described in Hansen and Racine (2012). So this model first computes forecasts for different models, after which the weights are optimized. Thereafter, the final forecast from the JMA model is computed using the forecasts and the weights.

4.2.3 Factor Models

Factor models are models which are used to reduce the dimension of the model by trying to get common components from the predictors. So the goal is to extract common factors from x_t to forecast the inflation. The exact way that factors could be extracted can be found in Bai (2003). Three models are used which are using factors.

The first model, named 'Factors', computes the principle components of the covariates. After that, it uses the first four principle components and the inflation variable to estimate an OLS model with which forecasts for the inflation are made.

The second model which is estimated is the target factors (T. Factors) model. In this model, only the variables that are important for forecasting inflation are used to compute factors, as otherwise there is a possibility of the factors becoming very noisy, according to Bai and Ng (2008). First, a regression of the inflation on w_t, which is a set of controls, and on the candidate variables with importance for forecasting inflation is performed. Then all the variables which have a significant coefficient using a significance level of 0.05 are selected. After this, the principle components of these variables are computed and then the inflation is regressed on the controls and a number of factors, where the number of factors is determined by the Bayesian information criterion.

The final model using factors is the boosting factor (B. Factor) model. The boosting algorithm is used here to select the optimal number of lags and the factors in the model, as in Bai and Ng (2008). First of all, all of the n factors are computed from the n variables, together with the four lags of the factors. After that, an algorithm is used to find out which factors and lags to use. The exact algorithm can be found in Medeiros et al. (2021).

4.2.4 Random Forests and its combinations

The best model according to Medeiros et al. (2021) in forecasting inflation is the random forests (RF) model, which was introduced by Breiman (2001). This method uses regression trees, which are nonparametric models, in which the covariate space is recursively partitioned. If the amount of terminal nodes (K) is determined, the splits are chosen in such a way that for the regression

$$\pi_{t+h} = \sum_{k=1}^{K} c_k I_k(x_t; \theta_k), \tag{5}$$

the sum of squared errors in minimized with θ_k being the set of parameters. For $I_k(x_t; \theta_k)$ the formula is

$$I_k(x_t; \theta_k) = \begin{cases} 1 & \text{if} \quad x_t \in R_k(\theta_k), \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $R_k(\theta_k)$ is the kth region. The randomly constructed regression trees are aggregated based on bootstrapping, where $B = 500$ bootstrap samples are used, for which a subset of original regressors is chosen at random. For every bootstrap sample eventually a tree with K_b regions will exist. Finally, the forecasts are then given by

$$\hat{\pi}_{t+h} = \frac{1}{B}\sum_{b=1}^{B}[\sum_{k=1}^{K_b}\hat{c}_{k,b}I_{k,b}(x_t;\hat{\theta}_{k,b})]. \tag{7}$$

On top of this model, two combinations of models with RF are used, namely the RF/OLS and the adaLASSO/RF model. The RF/OLS model starts with splitting the variables using a regression tree, after which the N split variables with the highest importance are used to perform OLS and to make a forecast. This is done for $B = 500$ bootstrap samples. After all the B forecasts are made, the average is taken to get the final forecast. For the adaLASSO/RF model, the adaLASSO method is used to select the variables, after which RF is used to make forecasts only using the selected variables. The reason that these two models are used is to evaluate the importance of variable selection and nonlinearities in the performance of RF, according to Medeiros et al. (2021). If RF/OLS performs similar to RF, nonlinearities are not very important, but when it performs significantly less than RF, nonlinearities are likely to be an issue for linear models. The adaLASSO/RF model is used for the exact opposite reason, so if that model performs close to RF, variable selection is not very important and nonlinearity is.

4.2.5 Long Short-Term Memory Network Model

The first model which is new in this paper is the long short-term memory network (LSTM) model, as introduced by Hochreiter and Schmidhuber (1997). This model belongs to a set of models called the recurrent neural network models. These models are neural network models which can "remember" previously handled input and output and use this in later steps of the process, instead of throwing this information away after the step is done. The reason that specifically the LSTM model is used in this paper is that this model can also capture long-term dependencies between the variables, which are often present in macroeconomic settings.

The specification of the model is based on the papers by Paranhos (2021) and Rodríguez-Vargas (2020) using the LSTM functions in the Keras interface in R from Tensorflow. In this model, the function G from equation 1 has a neural network structure and is defined as $G(x_t, \hat{\Theta}_h) = g(f_{t|L})$ with

$$\Theta_h = \arg\min_{\Theta_h}(\frac{1}{T-h}\sum_{t=1}^{T-h}(y_{t+h} - G(x_t, \Theta_h))^2), \tag{8}$$

where Θ_h is the set of parameters used in the model and where x_t both contains the dependent and independent variables and a number of its lags. The model uses internal memory to make point forecasts where the internal memory is denoted as $f_{t|L}$, where this means that only information up to lag L is used. The $f_{t|L}$ is computed with a couple of functions using the operators *tanh* and *sigmoid* and the g function is given in Paranhos (2021). The model contains four LSTM layers, because this was computed to be optimal in Paranhos (2021). Also 200 epochs are used and this amount is chosen, as more epochs would be too computationally expensive and 200 is the minimum choice in the paper of Paranhos (2021). One epoch is finished when the learning model has worked through the entire training data set once. The loss function which is used in the model is the mean squared error and the optimizer is the Adam optimizer from Kingma and Ba (2017).

4.2.6 Bayesian Additive Vector Autoregressive Tree

The second model which is new in this paper is a combination between the vector autoregressive (VAR) model and the Bayesian additive regression tree (BART) model, which is the Bayesian additive vector autoregressive tree (BAVART) model, as described by Huber and Rossini (2021). This model is especially useful as it can capture nonlinearities between variables and it is likely that there are nonlinearities when a lot of variables are used, which is the case here.

The BAVART model is

$$Y = F(X) + \varepsilon = F(X) + \epsilon A_0' \tag{9}$$

where $Y = (y_1, \ldots, y_T)'$ is a $T \times M$ matrix with the M endogenous variables for T time periods and $X = (X_1, \ldots, X_T)$ with $X_t = (y_{t-1}', \ldots, y_{t-P}')$ for $t = 1, \ldots, T$. The $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$ are the shocks with $\varepsilon_t \sim N(0_M, \Sigma)$ in the first formula and the $\epsilon = (\epsilon_1, \ldots, \epsilon_T)'$ are the shocks with $\epsilon_t \sim N(0_M, H)$ in the second formula for the BAVART model. It thus holds that $\Sigma = A_0 H A_0'$. This second formula for the BAVART model (the part which contains $\epsilon A_0'$) uses the structural form of the VAR which leads to significant computational gains, according to Huber and Rossini (2021). This is because this form makes the shocks independent and as such, the equations can be estimated seperately. It also holds that $F(X) = (f_1(X), \ldots, f_M(X))'$ where the $f_j(X)$ are approximated just as in the standard BART. The standard BART model does this by summing N regression trees:

$$f_j(X) = \sum_{k=1}^{N} g_{jk}(X|\mathcal{T}_{jk}, m_{jk}), \tag{10}$$

where g_{jk} is an equation-specific step function and $\mathcal{T}_{jk}$ is a tree structure with the vector m_{jk} being the terminal node structure associated with that tree.

The second formulation of the BAVART model makes that all the equations except the first one can be written as

$$y_{\bullet j} = \sum_{k=1}^{N} g_{jk}(X|\mathcal{T}_{jk}, m_{jk}) + \sum_{l=1}^{j-1} a_{jl}\varepsilon_{\bullet l} + \epsilon_{\bullet j}, \tag{11}$$

where j and l refer to the jth and lth column of Y, ϵ and ε. The first part of the summation is the non-parametric part and the second part of the summation is the regression part. The idea is to determine linear relations between contemporary ε_t's (the covariances between variables) while also determining the non-linear relationships between the variables and its lags.

On top of this, a stochastic volatility model is included to account for heteroskedasticity in the errors, as in Clark and Ravazzolo (2015), and this model has been shown to greatly increase forecasting accuracy in macroeconomic time series. This assumes that H is $H_t = \text{diag}(e^{h_{1t}}, \ldots, e^{h_{Mt}})$, where h_{jt} is determined by

$$\begin{cases} h_{jt} = c_j + \rho_j(h_{jt-1} - c_j) + \sigma_{jh}v_{jt},\ v_{jt} \sim N(0,1) \\ h_{j0} \sim N(c_j, \frac{\sigma_{jh}^2}{1-\rho_j^2}). \end{cases} \tag{12}$$

In these formulas, c_j is the unconditional mean, ρ_j is a persistence parameter and σ_{jh}^2 is the variance of the error terms. The exact way to generate the trees using a three step stochastic process and to compute the forecasts is described in Huber and Rossini (2021). This process also uses a Markov chain Monte Carlo (MCMC) algorithm with a certain number of burn-ins and draws.

4.3 Evaluation

The evaluation of the forecasts in the different models is done using three different statistics, as done in Medeiros et al. (2021). These are the root mean squared error (RMSE), the mean absolute error (MAE) and the median absolute deviation from the median (MAD). The formulas for these are

$$RMSE_{m,h} = \sqrt{\frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \hat{e}_{t,m,h}^2}, \tag{13}$$

$$MAE_{m.h} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} |\hat{e}_{t,m,h}^2|, \tag{14}$$

$$MAD_{m,h} = \text{median}[|\hat{e}_{t,m,h}^2 - \text{median}(\hat{e}_{t,m,h}^2)|], \tag{15}$$

with $\hat{e}_{t,m,h}$ is the forecast error at time t with model m with information up to time $t-h$ and where T_0 is the initial time in the sample period.

For all these models it holds that the performance of a model is better relative to another model based on a certain statistic when the value of the statistic is smaller.

5 Results

5.1 United States Data Analysis

First of all, the results for the US data are discussed. The average values of the evaluation statistics over all the 12 forecast horizons are shown in Table 2. For this data set, the BAVART model could not be estimated using the complete set of variables and this holds for all of the sample periods. The reason for this is that when more than 85 variables are used, the number of regressors used highly exceeds the number of observations in the training sample. As BAVART uses the inverse of the cross product of the data matrix to estimate the hyperprior of the model, the method does not work if that inverse is non-existent. Therefore, two models using only 85 variables have been added, namely a random forests model (RF-85) and a BAVART model (BAVART-85), to still be able to compare these models. The reason that the random forests model is used here, is because that model performs better than all the other models except the BVAR model in terms of RMSE, MAE and MAD for most sample periods and because it is a machine learning method.

However, the BAVART-85 model uses only 100 burn-ins and 100 draws for the MCMC algorithm, which could be insufficient for the convergence of the model with the number of variables used. To see if this weakens the performance a lot, another model (BAVART-moreN) is used which uses only 10 variables with 200 burn-ins and 2000 draws. This model turns out to enhance the performance of the BAVART model a lot when looking at the RMSE and the MAE. The BAVART-moreN model performs better than or similar to most other machine learning models, but it is still outperformed by the random forests model and its combinations. However, this could be because fewer variables are used which is done because of computational efficiency.

Surprisingly, the BVAR model is the superior model when it comes to forecasting US inflation both for the high and low volatility periods. After that, the RF model is the best model when it comes to forecasting inflation for all sample periods, thus both for the high and the low volatility periods. The conclusion about the RF model is in line with the findings of Medeiros et al. (2021), who also find that RF performs superior to the other models. Another finding is that for the low volatility period, the factor models perform very poor when compared with the ensemble and shrinkage methods and even worse than the RW model for the MAE. However, for the high volatility period, the factors models perform relatively better than the shrinkage methods and approximately equal to the ensemble methods.

Another finding is that the BAVART-moreN model performs worse than the RF model, which indicates that the RF model is performing better for forecasting inflation than the BAVART model. However, the RF model uses 122 variables instead of only 10, which could be the reason that the RF model performs better.

The other model which is added to this report, namely the LSTM model, performs very poorly for forecasting US inflation. It even performs worse than the random walk model in most cases. For example, the average RMSE of the LSTM model is 2.36 times the average RMSE of the random walk model for the sample period from January 1990 to December 2000.

Table 2: The averages of the different statistics (RMSE, MAE, MAD) relative to the random walk model over all 12 forecast horizons for the different models of the United States inflation based on the CPI. The blue coloured cells are the statistics for the best-performing model.

	Jan 1990-Dec 2000			Jan 2001-Dec 2015			Complete sample period		
Model	RMSE	MAE	MAD	RMSE	MAE	MAD	RMSE	MAE	MAD
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.84	0.89	0.69	0.80	0.78	0.78	0.82	0.83	0.74
ARIMA	0.76	0.75	0.63	0.74	0.71	0.68	0.75	0.73	0.66
BVAR	0.72	0.80	0.66	0.57	0.58	0.53	0.63	0.64	0.59
LASSO	0.85	0.91	0.72	0.89	0.94	0.74	0.87	0.93	0.73
adaLASSO	0.82	0.84	0.75	0.89	0.93	0.73	0.86	0.89	0.74
ElNet	0.86	0.93	0.71	0.87	0.91	0.75	0.87	0.92	0.73
adaElNet	0.82	0.85	0.74	0.87	0.91	0.74	0.85	0.88	0.74
RR	0.81	0.85	0.78	0.81	0.82	0.67	0.81	0.83	0.72
Bagging	0.90	0.93	0.91	0.79	0.80	0.81	0.84	0.86	0.85
CSR	0.82	0.87	0.79	0.80	0.77	0.71	0.81	0.81	0.74
JMA	1.06	1.09	1.09	1.04	1.05	1.06	1.05	1.07	1.07
Factor	0.92	1.00	0.86	0.78	0.76	0.72	0.84	0.86	0.78
T. Factor	0.93	1.01	0.85	0.79	0.76	0.72	0.85	0.87	0.78
B. Factor	0.95	1.04	1.07	0.78	0.78	0.76	0.85	0.89	0.89
RF	0.79	0.81	0.66	0.73	0.70	0.70	0.76	0.75	0.68
RF/OLS	0.80	0.84	0.81	0.74	0.73	0.69	0.77	0.78	0.74
adaLASSO/RF	0.80	0.83	0.69	0.75	0.72	0.69	0.77	0.77	0.69
RF-85	0.79	0.81	0.62	0.75	0.71	0.72	0.77	0.75	0.68
BAVART-85	0.99	1.08	0.75	0.82	0.82	0.72	0.89	0.93	0.73
BAVART-moreN	0.83	0.86	0.77	0.80	0.80	0.78	0.81	0.83	0.76
LSTM	2.36	1.84	1.09	1.72	1.30	0.95	1.99	1.53	1.01

As described in Section 4.2.4, the comparisons between the RF model and the RF/OLS and adaLASSO/RF models shine some light on the importance of nonlinearities and variable selection in the superior performance of the RF model. As both combination models perform less than the RF model on its own, this is an indication that both nonlinearities and variable selection are important factors for the RF model to perform as well as it does. On top of that, as the BAVART-moreN model outperforms linear models even with a significantly lower number of variables, the nonlinearities of the variables seem to play a significant role in the performance of the forecasting models.

5.2 Mexican Data Analysis

Secondly, the forecasts for Mexican inflation will be evaluated. These forecasts for the different evaluation statistics and forecast horizons are displayed in Tables 3 to 5. The RF model is the best model when looking at the RMSE and MAE for most of the 12 forecasting horizons. However, for the average MAD, the RF/OLS model is the best with the RF model in second place. It can also be seen that the shrinkage methods perform better than the factor and ensemble methods for all forecast horizons and are pretty close to the RF model for all statistics. Some shrinkage methods even perform better than or equal to the RF model for some forecast horizons when looking at the MAE and MAD. However, on average RF stays the superior model.

As well as with the US data, the BAVART model for the Mexican data also uses 100 burn-ins and 100 draws, whereas the BAVART-moreN model uses 200 burn-ins and 2000 draws with only 10 variables. For all three evaluation statistics, both BAVART models perform significantly worse than the RF model, which is in line with the findings for the US data. For the Mexican data set, the BAVART model outperforms the factors and target factors model, the ARIMA model for higher forecasting horizons, the Bagging and JMA model and it performs similarly to the BVAR model for higher forecasting horizons. However, it is outperformed by the shrinkage models and by the CSR and boosting factors models. When comparing the BAVART model with the BAVART-moreN model, the BAVART-moreN model has better performance for the one-step ahead forecasts, but thereafter the BAVART model is better again. This suggests that the losses of decreasing the number of variables from 32 to 10 are stronger than the gains in forecasting accuracy of increasing the number of burn-ins and draws. Overall, the BAVART model performs similarly on Mexican data and US data. It does outperform several models, but it is not the superior model.

The other model which is added to this report, namely the LSTM model, performs very poorly for forecasting Mexican inflation. This is the same conclusion as drawn for the US inflation, as it also performs worse than the random walk model for Mexican inflation in most cases. Possible explanations for this could be the small number of epochs used, which is only 200 and could have been higher, or the insufficient depth of the LSTM layers in the model.

Table 3: The values of the RMSE statistic relative to the random walk model for the different forecast horizons and the average of this (in column Ave.) for forecasting Mexican inflation. The blue coloured cells are the RMSE statistics of the best-performing models.

	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	0.87	0.71	0.60	0.56	0.52	0.51	0.60	0.61	0.56	0.65	0.79	0.92	0.68
ARIMA	0.81	0.69	0.72	0.74	0.72	0.72	0.76	0.76	0.74	0.80	0.92	1.02	0.78
BVAR	0.76	0.70	0.67	0.64	0.60	0.58	0.62	0.67	0.70	0.79	0.94	1.07	0.73
LASSO	0.77	0.60	0.52	0.52	0.46	0.44	0.46	0.50	0.56	0.63	0.77	0.87	0.59
adaLASSO	0.77	0.62	0.54	0.52	0.47	0.45	0.47	0.50	0.55	0.63	0.77	0.89	0.60
ElNet	0.79	0.63	0.56	0.53	0.46	0.43	0.47	0.51	0.56	0.65	0.78	0.87	0.60
adaElNet	0.85	0.61	0.55	0.54	0.47	0.46	0.47	0.50	0.55	0.63	0.76	0.88	0.60
RR	0.98	0.76	0.70	0.65	0.60	0.58	0.63	0.67	0.69	0.80	0.96	1.10	0.76
Bagging	1.07	1.23	1.01	1.00	0.88	0.55	0.55	0.66	1.07	1.14	0.96	1.19	0.95
CSR	0.76	0.58	0.53	0.52	0.47	0.46	0.53	0.55	0.56	0.65	0.78	0.90	0.61
JMA	1.50	0.72	0.91	0.60	0.51	0.65	0.82	0.64	0.79	0.75	1.06	1.20	0.85
Factor	0.93	0.69	0.59	0.58	0.52	0.50	0.60	0.65	0.62	0.69	0.84	0.92	0.68
T. Factor	0.92	0.69	0.59	0.58	0.52	0.50	0.60	0.65	0.62	0.69	0.82	0.92	0.68
B. Factor	0.83	0.59	0.51	0.54	0.51	0.48	0.53	0.56	0.60	0.71	0.85	0.99	0.64
RF	0.70	0.54	0.49	0.48	0.43	0.42	0.47	0.49	0.50	0.59	0.72	0.85	0.56
RF/OLS	0.74	0.57	0.51	0.50	0.45	0.43	0.48	0.51	0.56	0.65	0.77	0.90	0.59
adaLASSO/RF	0.70	0.54	0.55	0.49	0.45	0.44	0.45	0.50	0.58	0.61	0.76	0.84	0.57
BAVART	0.91	0.71	0.68	0.70	0.54	0.53	0.53	0.64	0.65	0.77	0.94	1.07	0.72
BAVART-moreN	0.77	0.69	0.72	0.71	0.65	0.60	0.64	0.68	0.70	0.81	0.97	1.11	0.76
LSTM	3.71	0.92	3.32	3.44	1.67	1.98	1.91	1.06	0.91	1.73	1.17	1.17	1.92

For this data set, also the RF/OLS and adaLASSO/RF models have been estimated to consider the importance of nonlinearities and variable selection in the performance of the RF model. When looking at the RMSE and MAE criteria, especially the adaLASSO/RF model performs very similar or even better than the RF model. This indicates that variable selection is not of major importance in the performance of the RF model. On the contrary, the RF/OLS model performs slightly worse than the RF model for most forecasting horizons which indicates that nonlinearities play a more prominent role in the superior performance of the RF model than the variable selection for Mexican data.

Table 4: The values of the MAE statistic relative to the random walk model for the different forecast horizons and the average of this (in column Ave.) for forecasting Mexican inflation. The blue coloured cells are the MAE statistics of the best-performing models.

	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	0.84	0.67	0.56	0.53	0.50	0.50	0.57	0.57	0.53	0.61	0.76	0.95	0.63
ARIMA	0.78	0.66	0.71	0.71	0.72	0.74	0.74	0.73	0.71	0.77	0.88	1.02	0.76
BVAR	0.73	0.65	0.62	0.60	0.56	0.57	0.59	0.63	0.66	0.75	0.90	1.10	0.70
LASSO	0.72	0.55	0.46	0.47	0.40	0.40	0.41	0.45	0.51	0.57	0.70	0.85	0.54
adaLASSO	0.71	0.57	0.49	0.47	0.40	0.41	0.42	0.45	0.51	0.57	0.71	0.86	0.55
ElNet	0.73	0.56	0.49	0.47	0.40	0.40	0.42	0.46	0.52	0.59	0.72	0.85	0.55
adaElNet	0.77	0.56	0.50	0.49	0.41	0.42	0.42	0.46	0.51	0.58	0.70	0.87	0.56
RR	0.95	0.71	0.65	0.61	0.57	0.57	0.59	0.63	0.65	0.75	0.91	1.11	0.73
Bagging	0.85	0.73	0.69	0.68	0.58	0.51	0.49	0.52	0.64	0.78	0.82	1.09	0.70
CSR	0.74	0.57	0.51	0.47	0.41	0.43	0.49	0.52	0.52	0.59	0.72	0.86	0.57
JMA	0.95	0.64	0.62	0.55	0.46	0.48	0.55	0.57	0.61	0.67	0.92	1.07	0.67
Factor	0.85	0.68	0.59	0.57	0.51	0.49	0.58	0.62	0.58	0.65	0.78	0.91	0.65
T. Factor	0.85	0.68	0.59	0.57	0.51	0.49	0.58	0.62	0.58	0.65	0.78	0.91	0.65
B. Factor	0.82	0.56	0.47	0.50	0.49	0.48	0.50	0.53	0.57	0.65	0.81	1.00	0.61
RF	0.68	0.51	0.46	0.42	0.39	0.40	0.41	0.44	0.45	0.54	0.66	0.84	0.52
RF/OLS	0.70	0.55	0.49	0.45	0.40	0.40	0.43	0.46	0.50	0.58	0.69	0.87	0.54
adaLASSO/RF	0.67	0.50	0.51	0.42	0.39	0.41	0.41	0.44	0.53	0.57	0.74	0.84	0.54
BAVART	0.91	0.71	0.68	0.70	0.54	0.53	0.53	0.64	0.65	0.77	0.94	1.07	0.69
BAVART-moreN	0.72	0.63	0.67	0.67	0.61	0.60	0.62	0.64	0.66	0.76	0.92	1.13	0.72
LSTM	3.25	0.90	1.84	1.75	0.95	1.17	1.06	0.88	0.85	1.07	1.07	1.25	1.33

Comparing the results for both data sets, the most important conclusion is that the RF model is the best performing model in most cases. However, for the US data, the BVAR model has a very strong performance as well. The relative ranking of the other models to each other, however, does vary between the data sets with a high and a low volatility. In both data sets, nonlinearities play a prominent role in the performance of the models. Variable selection is also very important for the US data set, but this is less important for the Mexican data set.

Table 5: The values of the MAD statistic relative to the random walk model for the different forecast horizons and the average of this (in column Ave.) for forecasting Mexican inflation. The blue coloured cells are the MAD statistics of the best-performing models.

	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	0.74	0.68	0.52	0.46	0.48	0.55	0.53	0.47	0.46	0.52	0.76	1.01	0.60
ARIMA	0.73	0.65	0.73	0.64	0.68	0.89	0.78	0.65	0.63	0.65	0.86	1.01	0.73
BVAR	0.65	0.56	0.57	0.44	0.47	0.58	0.52	0.49	0.53	0.60	0.83	1.03	0.61
LASSO	0.64	0.51	0.42	0.36	0.33	0.40	0.35	0.35	0.45	0.54	0.69	0.89	0.49
adaLASSO	0.65	0.53	0.41	0.40	0.32	0.45	0.36	0.47	0.47	0.56	0.67	0.81	0.50
ElNet	0.62	0.47	0.45	0.35	0.33	0.39	0.34	0.37	0.48	0.51	0.71	0.90	0.49
adaElNet	0.73	0.50	0.43	0.40	0.35	0.48	0.35	0.32	0.45	0.51	0.72	0.98	0.52
RR	0.85	0.63	0.57	0.48	0.49	0.60	0.55	0.48	0.53	0.60	0.83	1.04	0.60
Bagging	0.69	0.66	0.51	0.46	0.40	0.50	0.43	0.36	0.40	0.51	0.70	0.87	0.62
CSR	0.69	0.51	0.47	0.38	0.36	0.46	0.48	0.39	0.45	0.54	0.76	0.93	0.53
JMA	0.66	0.53	0.45	0.44	0.39	0.39	0.43	0.44	0.46	0.62	0.71	0.94	0.54
Factor	0.74	0.67	0.59	0.47	0.45	0.53	0.57	0.52	0.50	0.62	0.89	0.90	0.62
T. Factor	0.74	0.67	0.59	0.47	0.45	0.53	0.57	0.52	0.50	0.62	0.89	0.90	0.62
B. Factor	0.71	0.54	0.43	0.41	0.41	0.46	0.45	0.44	0.48	0.51	0.74	0.84	0.53
RF	0.64	0.49	0.47	0.37	0.36	0.44	0.39	0.37	0.40	0.43	0.62	0.79	0.48
RF/OLS	0.61	0.49	0.39	0.37	0.33	0.39	0.37	0.36	0.42	0.46	0.60	0.75	0.46
adaLASSO/RF	0.61	0.41	0.47	0.32	0.33	0.45	0.38	0.33	0.48	0.52	0.74	0.89	0.49
BAVART	0.87	0.61	0.61	0.57	0.45	0.50	0.43	0.48	0.52	0.60	0.80	0.96	0.62
BAVART-moreN	0.63	0.54	0.56	0.56	0.56	0.63	0.58	0.51	0.53	0.61	0.86	1.09	0.64
LSTM	2.01	0.71	0.79	0.60	0.54	0.76	0.74	0.54	0.60	0.59	0.83	1.07	0.82

6 Conclusion

In this paper, research was conducted as to whether the machine learning methods would increase the forecast accuracy in an emerging economy. Overall, it is found that these methods indeed perform better than the methods which are not machine learning methods. This is found both in a developed economy and in an emerging economy, namely the United States and Mexico respectively. However, a strong performance of the BVAR model is found for the US inflation forecasts, but not for the Mexican inflation forecasts. The biggest difference between these data sets is that the Mexican inflation is significantly more volatile than the US inflation, which thus leads to the conclusion that the machine learning methods increase forecast accuracy in both low and high volatility periods. On top of this, it has been found that the random forests model is the superior model in most cases. The added BAVART model does perform averagely, outperforming some, but not all models.

Furthermore, it is shown that the number of draws for the MCMC algorithm in the BAVART model plays an important role in the performance of this model. For the US data, increasing the number of draws from 100 to 2000 increases the forecast accuracy significantly even when 10 instead of 85 variables are used. This same conclusion cannot be drawn for the Mexican inflation forecasts. This all leads to the conclusion that the BAVART model could potentially be the superior model when a high number of draws and a lot of variables are used.

The LSTM model shows very bad performance and performs even worse than the random walk model for both data sets. This could have multiple reasons. The number of epochs could be insufficient, which was only 200, or an insufficient amount of layers are used or the depth of the layers is insufficient.

For further research, more different machine learning models could be examined to see whether these models are even better than the ones used in this paper. For example, different neural network models or support-vector machine models could be considered. On top of that, a more extensive set of variables could be used or different variables could be included which may have more value for predicting inflation. On top of this, the BAVART model could be estimated with a higher number of draws and a higher number of variables. The LSTM model could also be more optimized, by adding more layers, using more epochs or even using different functions within the LSTM layers. All in all, more research on this topic is still needed, but machine learning methods are certainly useful for predicting inflation in both developed and emerging economies.

References

Atkeson, A., & Ohanian, L. E. (2001). Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, *25*(1), 2–11.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, *71*(1), 135-171.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304-317.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71-92.

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

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, *30*(4), 551-575.

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.

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*, *33*(3), 679-693.

Granger, C. W. J., & Terasvirta, T. (1993). *Modelling Non-Linear Economic Relationships* (No. 9780198773207). Oxford University Press.

Haan, J. d., Hoeberichts, M., Maas, R., & Teppa, F. (2016). *Inflation in the Euro Area and Why It Matters.*

Hansen, B., & Racine, J. (2012). Jackknife model averaging. *Journal of Econometrics*, *167*(1), 38-46.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., & Schreiner, J. (2020). Nowcasting in a pandemic using non-parametric mixed frequency VARs. *Journal of Econometrics*.

Huber, F., & Rossini, L. (2021). Inference in Bayesian Additive Vector Autoregressive Tree Models. Retrieved from `https://arxiv.org/pdf/2006.16333.pdf`

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3).

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization.*

Mandalinci, Z. (2017). Forecasting inflation in emerging markets: An evaluation of alternative models. *International Journal of Forecasting*, *33*(4), 1082–1104.

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.

Paranhos, L. (2021). Predicting Inflation with Neural Networks. Retrieved from https://arxiv.org/pdf/2104.03757.pdf

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.

Prüser, J. (2019). Forecasting with many predictors using Bayesian additive regression trees. *Journal of Forecasting*, *38*(7), 621-631.

Rodríguez-Vargas, A. (2020). Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking*, *1*(1), 100012.

Russell, F. (2020, September 24). *FTSE Equity Country Classification September 2020 Annual Announcement.* https://research.ftserussell.com/products/downloads/FTSE-Country-Classification-Update-2020.pdf?_ga=2.96202076.1569865430.1620979058-735106260.1620716451.

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2018 17th ieee international conference on machine learning and applications (icmla)* (p. 1394-1401). doi: 10.1109/ICMLA.2018.00227

Stock, J. H., & Watson, M. W. (1999). Forecasting Inflation. *Journal of Monetary Economics*, *44*(2), 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.

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

Zhoua, L., Pana, S., Wanga, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350–361.

Önder, A. (2004). Forecasting Inflation in Emerging Markets by Using the Phillips Curve and Alternative Time Series Models. *Emerging Markets Finance and Trade*, *40*(2), 71-82.

Özgür, O., & Akkoç, U. (2021). Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms. *International Journal of Emerging Markets*, *ahead-of-print*. Retrieved from https://doi-org.eur.idm.oclc.org/10.1108/IJOEM-05-2020-0577

Ülke, V., Sahin, A., & Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing and Applications*, *30*, 1519—1527.

7 Appendix

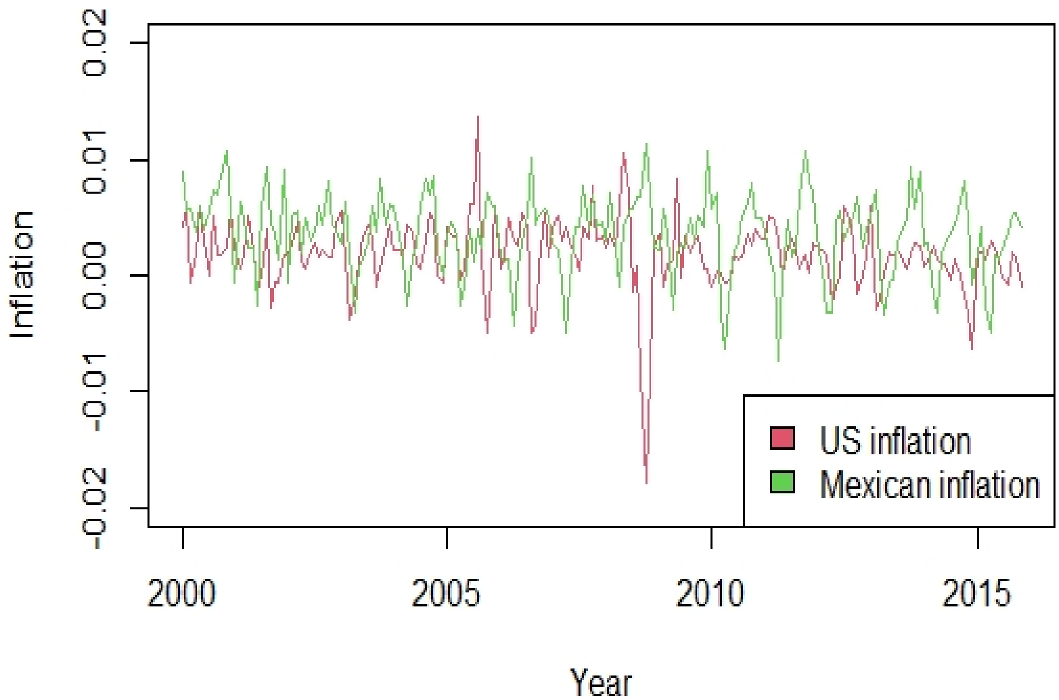

Figure A1: US and Mexican inflation from January 2000 to December 2015.

Table A1: The variables used in the Mexican data set. The id refers to the official id codes of the different variables in the Banco de Mexico database which are needed to obtain these variables. The transformation codes stand for (1) no transformation, (2) first difference of variable and (3) first difference of the logarithm of the variable. The TIIE rate is the interbank equilibrium rate in Mexico.

	id	description	transformation code
1	SP1	Mexican Consumer Price Index (INPC)	
2	SF6	Uses of Monetary Base	3
3	SF332	Banco de Mexico Assets and Liabilities	3
4	SF337	Banco de Mexico Credit	2
5	SF7	International Reserves	3
6	SF1	Currency in Circulation	3
7	SF29655	Net Domestic Credit (Pesos)	2
8	SF17908	Mexico/U.S. Exchange Rate	2
9	SF57923	Mexico/Euro Exchange Rate	2
10	SF57771	Mexico/Canada Exchange Rate	2
11	SF57775	Mexico/China Exchange Rate	2
12	SF3338	91-day Treasury Certificates Rate	2
13	SF3270	182-day Treasury Certificate Rate	2
14	SF40823	Short-term Private Debt Instuments Rate	2
15	SF283	28-day TIIE rate	2
16	SF17801	91-day TIIE rate	2
17	SF17863	Weighted Average Bank Funding Rate	2
18	SF17864	Weighted Average Government Funding Rate	2
19	SR28	Pesos's Real Exchange Rate Index	2
20	SR9	Mexican GDP	1
21	SP66540	INPC: Merchandise, Foods, Beverage, Tobacco	2
22	SP74627	INPC: Non-Food Merchandise	2
23	SP66542	INPC: Housing	2
24	SP56339	INPC: Education (tuition)	2
25	SP56337	INPC: Agriculture	3
26	SP74631	INPC: Energy and Prices Approved by Government	3
27	SG1	Accumulated Public Sector Budgetary Expenditures	1
28	SG8	Accumulated Public Sector Revenues	1
29	SG193	Total Broad Economic Debt of Public Sector	3
30	SG231	Average Maturity of Government Securities	3
31	SE36593	Merchandise Trade Balance of Mexico Total Exports	2
32	SF10478	Total Pension Funds	3

8 Programming Code

For the purpose of obtaining the results multiple programming codes are used, which are all written in RStudio, so using R. The codes are contained in a zip-file named "Codes_Thesis_Sven_Heeren" which is sent to the supervisor. The zip-file contains two folders. The "ForecastingInflation-master" folder contains the slightly adjusted models from the paper by Medeiros et al. (2021). The "Own_code" folder contains the codes written by me and contains the following files:

- **arima.R**: code used to run the ARIMA model function using the **func-arima.R** function.
- **BAVART.R**: code containing the function to run the BAVART model using the **mfbavart_func.R** function for the different data sets (both the US and the Mexican data set).
- **Fast.R**: code used to compute the RMSE, MAE and MAD with respect to the random walk model.
- **func-arima.R**: code containing the function of the ARIMA model.
- **LSTM_code.R**: code containing the LSTM model function.
- **MexDataFinal.RData**: file which contains the final Mexican R data used to perform the models.
- **MexicoDataRetrieval.R**: code used to obtain the data from the Banco de Mexico database.
- **mfbavart_func.R**: code containing the BAVART model, which is an highly adjusted version of the code from the paper by Huber, Koop, Onorante, Pfarrhofer, and Schreiner (2020).
- **RandomWalk.R**: code used to compute the random walk model forecasts and the error statistics.
- **Run-LSTM.R**: code containing the function to run the LSTM model for the different data sets (so for the US and Mexican data set) using the **LSTM_code.R** function.