Lund University

Department of Economics

Inflation forecasting with Random Forest

A Machine Learning approach to macroeconomic forecasting

By:

Malte Meuller

Supervisor:

Fredrik N G Andersson

Bachelor's thesis

2022-01-24

Acknowledgements

Foremost, I would like to thank my thesis supervisor Fredrik N G Andersson for his continuous support and guidance. Without his constructive critic and enormous knowledge, the writing of this thesis would not be possible.

Secondly, I want to show my gratitude to Isaiah Hull at the Swedish Riksbank for inspiring and helping me to pursue a computational approach to macroeconomic forecasting.

Abstract

The accuracy of inflation forecasts is, and has been, important for economic agents such as governments, central banks, companies, and the general public. Historically it has mainly been conducted with traditional statistical models that limits the usage of bigger datasets. This thesis will examine the performance of the machine learning model called Random Forest by forecasting Swedish inflation between January 2016 and January 2020. The forecasting horizons will be 1, 3, 6 and 12 months and the data used will consist of 250 variables, including lags and growth variables. For all forecasting horizons, except 1 month, Random Forest was able to provide more accurate predictions compared to the benchmark tests. The model also proved to effectively select predictive variables from a extensive data set and could therefore be useful in further quantitative research.

Keywords: *Inflation*, *Forecasting*, *Random Forest*, *Machine Learning*, *Macroeconomic Forecasting*, *Sweden*

Contents

Abbreviations **1**

1 Introduction **2**

2 Previous research **5**

2.1 Forecasting with traditional methods 5

2.2 Forecasting with Random Forest 6

3 Method **7**

3.1 Random Forest 7

3.1.1 Random Forest: Regression Tree 8

3.1.2 Random Forest: Bagging 11

3.1.3 Tuning 13

3.1.4 Random Forest with time series 14

3.1.5 Variable importance 16

4 Benchmark model and evaluation metrics **16**

4.1 Random Walk 16

4.2 ARIMA 17

4.3 Evaluation 18

5 Data and time frame **19**

5.1 Sample period 19

5.2 Inflation 20

5.3 Predictor variables 20

5.4 Data transformation 22

5.5 Software 23

6 Results and analysis **23**

6.1 Results from Auto.Arima test 23

6.2 Forecast results and analysis 24

6.3 Discussion 31

6.4 Limitations and future research 34

Conclusions **35**

Appendix **36**

Sources **39**

Abbreviations

ML: Machine Learning

RF: Random Forest

RW: Random Walk

ARIMA: Autoregressive (AR) Integrated (I) Moving Average (MA)

CART: Classification and Regression Tree

MAE: Mean absolute error

MSE: Mean squared error

CPIF: Consumer price index with fixed interest rate

1 Introduction

Forecasting the state of the economy has always been important for economic agents such as governments, central banks, companies, and the general public. Inflation, which is measured by annual percentage change in CPIF (Riksbank, 2021), is one of the more well-known and crucial indicators for the economy. To be able to achieve more accurate forecasts would benefit the society as whole.

For example, central banks direct their monetary policy, with guidance from future expected inflation (Riksbank, 2021). By adjusting the interest rate the central bank could steer the inflation to their targeted goal. However, the effects of the monetary policy have shown to have a delayed reaction on inflation (Batina & Nelson, 2001) and therefore the implementation of policies often relies on forecasts. Also, wages and prices are determined with regards to future inflation (NIER, 2013). If the inflation turns out much higher than the forecasts, the real wages of workers will decrease. With more accurate forecasts and greater understanding of the forces driving the change, the policies and wage settings could be better implemented and reduce the risk of over- or under-shooting the targets (Riksbank, 2021).

As price stability is one of the main objectives for the Swedish central bank (Riksbank, 2021) it is crucial that economic agents have trust in the forecasts (Svensson, 1999). When economic agents have less faith in the inflation forecasts there is greater room for own conclusions which could foster emotional consumer behaviour, i.e animal spirits (Keynes, 1936). To be able to provide accurate forecasts could reduce animal spirits and contributes to a better and more stable economy for workers, consumers, and companies.

The previous mentioned example regarding monetary policy, wages and price stability usually uses longer forecast horizons ranging from 6 month to several years as guidance (Hull, et al. 2017). Hull et al. mentions that shorter forecasts, like 1 or 3 months, also are important as longer forecast horizons are dependent on accurate short-term forecasts to separate long term effects from temporary deviations. These shorter forecasts are also helpful to financial institutions as stock prices tend to react negatively to increasing inflation Eldomiaty et al. (2019)

Up until now, inflation has typically been forecasted with four types of methods according to Ang et al. (2007). 1.) Univariate inflation time-series models, 2.) Regression models with economic indicators, 3.) Assets price models with embedded information about future inflation, and 4.) Business and consumer survey-data models. All methods have their benefits and differences and there is no consensus which models that are the most accurate. What these models have in common is that they rely on assumptions of a pre-determined relationship between the independent and dependent variable, i.e a stochastic process (Araujo & Gaglianone, 2019). Another way of thinking of these traditional models is that they are easily interpreted and focuses on causal explanations. One potential drawback is that the model needs to be restricted using fewer variables which could weaken the accuracy (Athey & Imbens, 2019).

Machine learning (ML) however, puts little focus on the causality and interpretability and more focus on the accuracy of predictions (Athey & Imbens, 2019). The ML models have little or no underlying assumption about the data and the data process (Araujo & Gaglianone, 2019) which creates possibilities for usage of extensive data sets. The ML method Random Forest (RF) has on several occasions proven to beat traditional forecasting methods in macroeconomics (Araujo & Gaglianone, 2019).

This thesis will use the RF method which was first introduced by Breiman (2001) and is considered one of the most influential ML algorithms. The algorithm uses historical data, called training data, to create a regression tree from which the predictions are being made. To increase accuracy and minimize bias from an individual tree the data is multiplied with a method called bagging and used to create hundreds of trees. The RF algorithm works by continuous splitting the observations into subsamples by choosing a threshold value that minimize the mean square error, MSE. This creates small groups of observations with similar economic conditions from which the forecasts will be based on. Breiman (2001) first applied the algorithm to classification problems. For example, classification of fruit by knowing the colour, size, and taste. But he also states that it could be used for numerical estimations called regression problems.

This thesis will examine the performance of inflation forecasting with RF. Promising results could be found for other countries (Araujo & Gaglianone, 2019; Baybuza, 2018) but the Swedish case is, to the best of my knowledge, still unexplored. As this is the first time RF has been used to forecast inflation in Sweden the aim is to gain further knowledge about inflation forecasting in general, and inflation forecasting for Sweden in particular. This will be done by answering the following research questions: 1.) How accurate does the RF-method predict future inflation? 2.) Could the RF algorithm determine which variables that are predictive?

To conduct the study the RF method for regression problems will be used to forecast inflation. For comparison a Random Walk (RW) and an Autoregressive Integrated Moving Average (ARIMA) model will be used as benchmark tests. RW forecasts predicts that the future value will be equal to the last observed value (Nau, 2014). The ARIMA model focuses on lagged values and previous forecast errors (Hyndman & Athanasopoulos, 2018). If the RF model performs better than the two benchmark models, it indicates that inflation could be more precisely forecasted by including exogenous variables and/or the usage of RF. To understand which variables that are predictive the measure of variable importance, given by the RF-algorithm, will be used.

The sample period stretches from 2000-01-01 until 2020-01-01 when Covid-19 was declared a global health emergency (WHO, 2020). Four different time horizons, 1, 3, 6 and 12 months will be forecasted. The data will consist of monthly observations of inflation and 50 predictor variables representing four subgroups of the economy will be used. These groups are: Consumers, Monetary, Global and Industry. The analysis will use both hard variables like *Fiscal Expenses* and soft variables such as *Consumer Confidence*. Also, lagged variables and derived growth variables will be used summing up to a total of 250 variables.

The following reading will be divided into five sections. Section 2. will give a walkthrough of previous research made in the field. Section 3. will describe the Random Forest method and section 4. the benchmark tests used for comparison. Section 5. will introduce the data used and section 6. will present the results and conclusions.

2 Previous research

The forecasting literature is vast and covers many different dimensional such as choice of methods and data. First some articles upon more traditional inflation forecasting will be covered (Stock & Watson, 2008; Faust and Wright, 2013). Thereafter machine learning literature with focus, but not exclusively, on inflation forecasting, will be discussed (Breiman, 2001; Biau & D´elia, 2010; Chen et al. 2019; Araujo & Gaglianone, 2019, Woloszko, 2020).

2.1 Forecasting with traditional methods

Forecasting inflation has mainly used four groups of methods (Ang et al. 2007). 1.) Inflation time-series models, 2.) Regression models with economic indicators, 3.) Assets price models with embedded inflation expectations, and 4.) Business and consumer survey-data models. Group one includes typical univariate time series models like RW and ARIMA. Group two includes Phillips Curve forecasts that include few economic indicator variables like unemployment or GDP. Group three uses direct or indirect measures of inflation expectation. This could be from asset prices embedding information about future inflation or from measures such as the Fisher equation: $real\,interest\,rate \approx nominal\,interest\,rate - inflation\,rate$ (Fisher, 1907). Group four consists of survey-based methods like business surveys or expert surveys.

A recurring conclusion found in articles that compare these groups of models is that no model is consistently better than the others (Stock & Watson, 2008; Faust & Wright, 2013). All models have their benefits and drawbacks. Univariate model tends to perform best for shorter forecast horizons but gradually decay when looking at a longer horizon. Also, the univariate models tend to perform well in calm times, and multivariate perform better in volatile times. Marcellino (2008) argues that the usage of these univariate models often are justified even if there are numerous advanced models and algorithms to compete with.

On average there seem to be no gain to be made from including many variables compared to a univariate model (Faust & Wright, 2013). But when it comes to variable selection economic indicators and survey data are more predictive than embedded information measures (Ang et al. 2007). Another common finding is that different forms of averaging improves the predictive power. Stock & Watson (2008) finds that averaging over different model tend to erase bias and improve prediction while Faust & Wright (2013) and Ang et al. (2007) states that averaging over expert surveys seem to improve the accuracy.

2.2 Forecasting with Random Forest

Forecasting inflation with RF is an emerging field with limited number of available papers. However, some articles that predict other macroeconomic variables, like GDP growth, use similar methods and input variables. The standard practice in these articles, when comparing the performance of RF, is to use univariate models like RW and/or ARIMA as benchmarks (Biau & D´elia, 2010; Chen et al. 2019; Araujo & Gaglianone, 2019; Woloszko, 2020).

Compared to the models used in traditional forecasting the ML models allow for bigger number of observations and varaibles. This possibility creates different strategies for feature selection, i.e the selection of varaibles. The two most common methods are called *cherry picking* and *kitchen sink* and refers to the number of variables and the caution behind the selection (Chen et al, 2019). Woloszko (2020) and Biau & D´elia (2010) adapts a cherry-picking strategy where Woloszko uses 5-8 variables economic indicators and Biau & D´elia only uses soft variables such as consumer and business confidence. In contrast to this, Chen et al. (2019), and Araujo & Gaglianone (2019) use the kitchen sink strategy where they include data from many sources and arrange them in categories reflecting different aspect of the economy such as Finance, Trade and Money. The cherry-picking strategy is better suited when the focus is on comparability between models as non-ML models functions with fewer variables (Woloszko, 2020). When the focus is on predictive power a kitchen-sink strategy result in better forecasts as it utilizes the full capacity of RF (Araujo & Gaglianone, 2019). Baybuza (2018) states that the frequency of the observation reflect the forecasting horizon. For example, monthly data is not suitable for longer forecasts according to Baybuza.

The results from these studies show that RF on average has more accurate forecasts compared to traditional models. In addition to this, Chen et al. (2019) and Araujo &

Gaglianone (2019) concludes that RF performs better than the other ML techniques used. Chen et al. (2019) also states that using more data does not necessarily improve the forecasting results and could instead worsen the interpretability. Even if the performance of RF seems promising, results indicating that RF performs bad during volatile times has been found (Biau & D´elia, 2010). Results from Biau & D´elia shows that a linear regression model based on variables from the RF importance algorithm outperformed RF during the 2008 crisis. According to Biau & D´elia this is because the prediction from RF comes from averaging over several estimator and therefore puts little focus on drastic changes. Woloszko (2020) successfully deals with this issue by using an adaptive version of RF constructed to put more emphasis on structural changes, non-linearity, and combination of variables. For example, rising house prices could mean that the economy is strong, until the prices go up too much, then it could be a house bubble. So, to put emphasis on these interactions and nonlinearities he includes growth variables, moving averages and deviation from mean in the algorithm.

3 Method

3.1 Random Forest

Random Forest was first proposed by Breiman (2001) and originally used as a binary classification tool, e.g, does a person have cancer? Is the person creditworthy for a loan? Progress was mostly made in fields like data science, finance, and medicine but the field of economics has been slow of adapting these new methods (Biau & D´elia, 2010). It may be due to lack of interpretability in the so-called *black box*-methods or the disregard of standard econometric practices (Athey & Imbens, 2019). When the RF method later began showing promising results for numerical estimations, i.e regression problems, the interest became bigger. The method has been proved to handle large numbers of predicting variables without overfitting and still be able to produce good estimation and forecasts (Biau & D´elia, 2010).

The major difference between modern ML methods and traditional statistical methods is that the later assumes an underlying relationship between the predictors and response variable, i.e a stochastic process (Araujo & Gaglianone, 2019). The ML approach on the other hand does not assume any underlying relationships about the predictor and response variable which creates possibilities of bigger datasets.

Due to the non-linearity of economics the RF is a suitable model as it takes nonlinear relationships into account in a way that is more intuitive than in traditional models (Woloszko, 2020). These nonlinearities could be detected by the RF algorithm by combining several threshold values, or nested if-else statements as it is called in code language (Biau & D´elia, 2010). The RF method could be divided into two parts: Classification and Regression Trees (CART) and bootstrapping. The following explanation and mathematical expressions are acquired from Chen et al. (2019), Athey & Imbens (2019), Chakraborty & Joseph (2017).

3.1.1 Random Forest: Regression Tree

To understand the RF method, it is crucial to first understand the Classification and Regression Tree (Breiman et al. 1984), which the algorithm is based on. But before this, an intuitive understanding of how they are built is needed before jumping on to the mathematical aspects. As seen in Figure 1 there are three types of nodes in the tree: Root node, internal node, and leaf node. A node that is being split is called parent node and the two following nodes are child nodes.

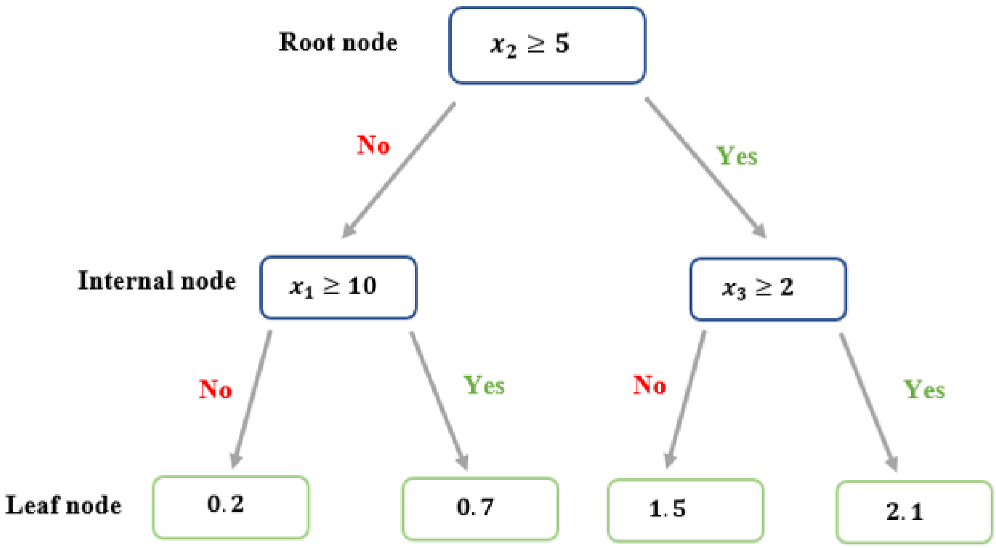

Figure 1: The structure of CART, describing the different parts and how the observations are split up by using threshold values. The numbers at the leaf nodes correspond to the average values of inflation in the observations fulfilling the given thresholds. Source: Own illustration.

In the root node all observations are gathered and then separated according to a threshold value for a variable, in our case if $x_2 \geq 5$. If the observation satisfies the condition, then it will pass on to the right, else to the left. Here it faces the internal node which have the same function as the root node, another if-else statement. By doing these kinds of splits the observations is always divided into smaller groups. At the last split the observation are passed onto the leaf nodes where observations with similar values for the given variables are. In the leaf nodes the focus turns to the y-values instead of the x-values.The average of these known y-values in the same leaf node will become the node value. When passing new observations with unknown y-values the prediction will be the averages of all known y-values in the leaf node that it reaches. To put it in context: the prediction of inflation coming period will be based on values for inflation for months with similar characteristics in the x-variables

The splitting decision, for example $x_2 \geq 5$ or $x_1 \geq 10$, are based on a mathematical equation that minimizes the mean squared error (MSE). The algorithm starts with a given sample with several predictors for the response variable.

$$(x_{i1}, ...x_{ij}, y_i) for\ i = 1, ...n$$

At every step of the tree a new split is made depending on a single variable and a given threshold θ. The mean square error, MSE, before the split is

$$Q = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{1}$$

Where Q is the MSE and $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

After the split is made on variable x_j following the condition: $x_j \geq \theta$, two new nodes are being formed accordingly:

$$L = \{i \colon x_j > \theta\}$$

$$R = \{i \colon x_j < \theta\}$$

The MSE of these two nodes combined are:

$$Q = Q_{Left} + Q_{Right} = \frac{1}{n_L}\sum_{i=1}^{n_L}(y_i - \overline{y})^2 + \frac{1}{n_R}\sum_{i=1}^{n_R}(y_i - \overline{y})^2 \tag{2}$$

Where L and R denotes left and right.

The splitting decision, i.e the value of θ, will be the optimal value that minimizes the MSE of the two child nodes. A full-grown tree minimizes the MSE of all nodes in the tree.

$$Q_{Total} = \frac{1}{n}\sum_{c=1}^{c}\sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{3}$$

Where the inside sum represents one node, and the outside sum represents all nodes.

The predictions from the CART could be expressed as conditional expectations $f(x) = E(Y|X = x)$ and is calculated by using several base learners $h_j(x)$. A base learner is a y-value at the leaf node and by averaging these the prediction will be made for unknown y-values.

$$f(x) = \frac{1}{J}\sum_{j=1}^{J} h_j(x) \tag{4}$$

Where J is the number of observations inside the leaf node.

As seen from equation 4, if there is only a single base learner, the leaf node will contain one observation and thus cause overfitting which is an abnormal accurate fit for the training data. When later introducing test data, i.e unknown y-values, the prediction will take the average of that leaf node which in this case only has one value. The prediction will be highly biased and thus cause low accuracy. To avoid this overfitting problem which often happens with CART a method called bagging is introduced.

3.1.2 Random Forest: Bagging

By combining CART (Breiman et al. 1984) with the bootstrapping aggregating method called bagging (Breiman, 1996), the Random Forest is made. Bagging is a method for creating subsets derived from the original dataset. It works by randomly selecting observations until the subset contains the same number of observations as the original dataset. By introducing *randomness* in the selection of the observation the individual bagged subsets will be unique as some observations are selected multiple times and some observation will be left out. By creating a regression tree from each one of the bagged subsets and combining all trees the Random Forest is now made. The RF will include around 500 unique trees that will decrease the overall bias as an individual tree will not have enough influence.

When the forest is built up containing several hundred trees the algorithm introduces the second part of randomness by randomly selecting a limited number of variables for each node to make the splitting decision on. The RF will therefore end up with several hundred of threes with different observations and unique sets of variables used.

Finally, when the algorithm is used for forecasting the new observation with unknown y-values will go through all trees. The forecast will be the aggregated average of all trees:

$$\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B} f(x) \tag{5}$$

Where B is the number of trees and $f(x)$ is the prediction from a tree. Combining this with equation 4, will give:

$$\hat{y}_i = \frac{1}{JB}\sum_{b=1}^{B}\sum_{j=1}^{J} h_j(x) \tag{6}$$

Where B is the number of trees, J is the number of observation in a leaf node and $h_j(x)$ is the base learner inside a leaf node.

When put togheter the forest will be build up from individual trees seen in figure 1. As all trees have unique sets of observations and variables the new observation will react differently in the bagged trees.

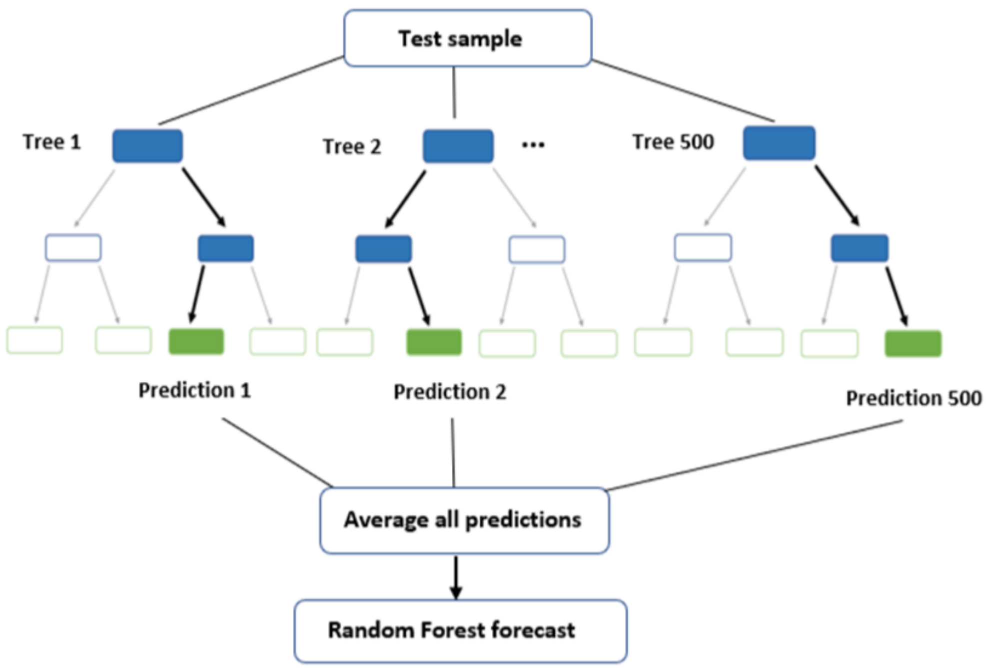

Figure 2: The structure of the Random Forest describing how predictions are being made by being passed through all trees. As all trees are individual they will highlight different aspect. The average of all estimations will be the forecast. Source: Own illustration.

The usage of bagging reduces the variance and bias compared to an individual tree. As the trees in the forest comes from the same underlying dataset the trees correlate with each other. Following Bernard et al. (2010) and assuming correlation between two trees of $\rho < |1|$. The variance of the RF could be described as:

$$var_{RF} = \rho\sigma^2 + \frac{1-\rho}{F}\sigma^2 \tag{7}$$

Where F is the number of trees, σ^2 is the variance of a single tree, and ρ is the correlation between the trees.

Hastie et al. (2009) states that the forest has a declining variance when allowing for more trees. As more trees is used the second term will disappear. Random Forest often stabilizes at around 200-500 trees (Hastie et al. (2009).

$$\lim_{F \to \infty} \rho\sigma^2 + \frac{1-\rho}{F}\sigma^2 = \rho\sigma^2 \tag{8}$$

A more intuitive way of understanding the RF method is that it is a mathematical form of *Vox Populi*, more known as Wisdom of the crowd, first stated by psychologist Galton (1907). The logic behind is that a prediction based on an average of many predictions will be better than a single guess as the individual errors from each tree will cancel out each other.

3.1.3 Tuning

When using Random Forest there are some parameters that is needed to be tuned to gain the best predictions. These are:

- Train/test size
- Maximum number of internal nodes, i.e tree size
- Number of tested predictor variables at each node
- Number of variables to include in the RF
- Number of trees in the RF

Random forest algorithm needs training data to teach the algorithm and test data to test how well it performs. The normal split is to divide the data into 80 % training data and 20% test data (Hastie et al. 2009). This thesis will follow a similar split but divide the data set according to time and not random observations. Therefore, the training set will include observations from January 2000 until December 2015, and the test set will include observation from January 2016 until January 2020.

When choosing tree size, the number of internal nodes needs to be adjusted to avoid overfitting or underfitting by changing the depth of the tree (Woloszko, 2020). The optimum depth is chosen by setting a minimum number of observations in each leaf node. The standard value is to allow a minimum leaf node size of 5 (Breiman, 2002). Since this analysis uses time series data, the 5 most similar observation will more likely be dependent on time. To focus the algorithm on activity measures the RF will set the minimum node size to 10. By doing this the forecasts will also tend to predict less extreme values as the effect from single observations will diminish by allowing for more values.

Even if RF is resistant to overfitting (Biau & D´elia, 2010) these problems could still occur when using high dimensional, i.e more variables then observations, time series data (Tyralis & Papacharalampous, 2017). To avoid this problem, it is common to use the variable importance function in R (see 3.1.5 Variable Importance) to select the most predictive variables from the training data to include in the revised model used for predictions (Kursa & Rudnicki, 2011). Tyralis & Papacharalampous (2017) shows that time series forecasting benefits from using fewer predictor variables and to pick them based on the variable importance. This analysis will use the default value from the authors of the RF package *Scikit-learn* (Pedregas et al. 2011) and select the top 10 most important variables.

When it comes to number of variables tested at each point the normal value for regression problems is $m = \frac{p}{3}$ where p is number of predictor variables (Breiman, 2002). As stated, the top 10 variables will be chosen from the variable importance list and therefore the default value would be $m \approx 3$. By allowing for more variables tested at each node the probability for choosing the best variables will increase (Probst et al. 2019). The analysis will try 8 variables at every point to still induce randomness.

The only shortcoming when choosing the number of trees is to have too few. When allowing for more trees the marginal gain will converge to 0 when reaching 300-500 trees (Breiman, 2001). The analysis will use 500 trees in the forest which is the standard value for the algorithm (Breiman, 2002).

3.1.4 Random Forest with time series

As mentioned before, RF was not made for time series data. RF in the normal state wants tubular data, which could be thought of as a normal spreadsheet data structure where the observations are independent (Brownlee, 2020). To use time series data, it is needed to first be translated into a supervised learning problem which is the foundation for predictive modelling using machine learning (Brownlee, 2020). To translate to a supervised learning problem the dataset is copied and then one column is shifted so that y_t will be predicted by y_{t-1} and x_{t-1}. For the 1, 3, 6, 12 months forecasts y_t will be shifted so that:

1 Month: y_t will be predicted by y_{t-1} and $x_{i,t-1}$

3 Months: y_t will be predicted by y_{t-3} and $x_{i,t-3}$

6 Month: y_t will be predicted by y_{t-6} and $x_{i,t-6}$

3 Months: y_t will be predicted by y_{t-12} and $x_{i,t-12}$

The model needs to be tested by using a special technique called walk-forward-validation instead of the normal K-fold cross validation where the test and training data randomly gets appointed observations (Brownlee, 2020). Doing this with time series data would mean that the time structure would be lost and that future values would predict historical data.

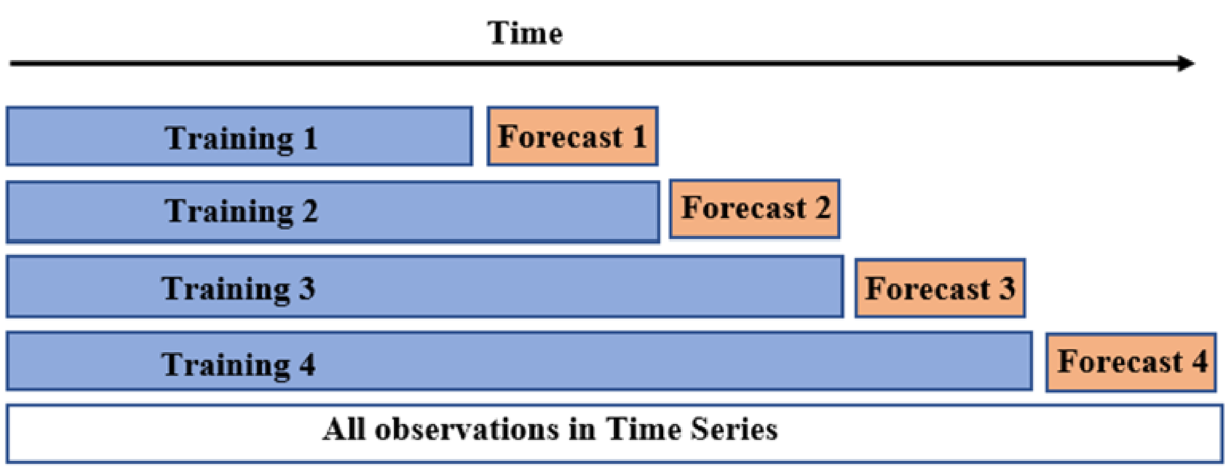

Figure 3: Illustration of how the expanding window works by introducing the passed test data into training data to refit the model. Source: Own illustration.

The walk-forward-validation that will be used is the expanding window method as it continually includes more values to the training data. For example: If the goal is to make 1 month forecasts all values up to the month before forecasting time will be used. After the prediction the expanding window will move on including the x-values and the true y-value of the predicted month to predict next month. Between these predictions the model will refit to train with the latest information available (Brownlee, 2020). Depending on the forecast horizon both one step forecast, and multi-step forecast will be used.

3.1.5 Variable importance

The RF algorithm could also be used to see which variables that have the most explaining power (Biau & D´elia, 2010). This is done by measuring how much the MSE would increase when randomly assigning the variable new values. This could be viewed as simply removing the variable. Changing the values of a predictive variable would damage the forecasts more than a unpredictive variable. The technique is therefore based on a similar equation as the MSE minimizing equations (equation 1, 2 and 3) and will present the variables with most predictive power.

4 Benchmark model and evaluation metrics

When comparing the performance of models it is common to use benchmark models (Biau & D´elia, 2010; Chen et al. 2019; Araujo & Gaglianone, 2019, Woloszko, 2020). These models tend to be linear and univariate but works good as predictive models. Marcellino (2008) argues that these models often are justified even if there are numerous highly advances models and algorithms to compete with. Often when dealing with time series data an ARIMA model or Random Walk is used. In this analysis both will be used.

4.1 Random Walk

Random walk is a one of the more important models in time series forecasting due to its replicability and common usage (Nau, 2014). The model assumes that the predicted value will be the previous value plus an error term. The error term is independent and identical distributed (i.i.d) with a mean of zero.

$$y_t = y_{t-1} + \varepsilon_t$$

When the error term is i.i.d the probability that the future value will increase is as likely a decrease and therefore the best prediction is the previous value (Nau, 2014). In this case the drift component was left out as the inflation pattern showed no signs of drift.

The prediction will therefore be:

$$\hat{y}_t = y_{t-1} \tag{9}$$

Beating the Random Walk indicates that inflation could be more accurate forecasted by using more sophisticated models and/or allowing for exogenous variables.

4.2 ARIMA

ARIMA stands for Autoregressive Integrated Moving Average where the AR, I and MA parameters needs to be specified to get as good results as possible. The AR(p) parameter refers to the number of lags to be used as predictor variables. The I(d) parameter tells how many times the data is needed to be differentiated before it becomes stationary, and the MA(q) refers to the number of lagged forecast errors that should be included (Hyndman & Athanasopoulos, 2018). The analysis will also include seasonal parameters for AR, I and MA as there could be seasonal patterns in inflation data (SCB, 2021). The seasonal parameters, P, D, Q, work in similar ways as the normal ARIMA parameters but instead of lagging/differencing on previous value it does it on previous seasonal value, for example same month 1 year apart.

A common complain about ARIMA is that the selection of parameters is prone to subjectivity (Shaffer et al. 2021). To tackle this problem the automatic ARIMA algorithm, called auto.arima, will be used. The algorithm could be reached from the *forecast* package by Hyndman (2021), through the software R (see 5.5 Software).

The algorithm combines different measures and test on the given data to obtain the best possible results for the future forecasts. The test will also include seasonal parameters of AR, I and MA. This thesis will give a short explanation of how the algorithm works, for more information see Hyndman & Khandakar (2008)

1. The number of differences (d, D) is tested by using the KPSS-test

2. Four basic models with given values for p, P and q, Q are chosen by minimizing the AICc.

3. The best of these models are tested with different values for p, P and q, Q until the model that minimizes AICc is found.

By doing this Hyndman & Khandakar (2008) claims to provide the best possible parameters for forecasting with the given data. A general form of a stationary Seasonal ARIMA (SARIMA) model is written as following:

$$y_t = c + (\phi_1 y_{t-1} + \cdots \phi_p y_{t-p}) + (\theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q}) + \tag{10}$$

$$\ldots (\phi_m y_{t-m} + \cdots \phi_P + (\theta_m \varepsilon_{t-m} + \cdots \theta_Q \varepsilon_{t-Qm}) + \varepsilon_t$$

The two first parentheses are the non-seasonal AR and MA parameters which states how many lags and errors that are included. The two last parentheses show the seasonal AR and MA parameters where m is the length of the season (Hyndman & Athanasopoulos, 2018). If the time series needs to be differentiated first, then y'_t will denote the first difference $y_t - y_{t-1}$

4.3 Evaluation

By comparing the predicted and the observed values, the mean absolute error (MAE) could be calculated. MAE was chosen since it is a scale dependant measure making the errors in the same unit as inflation (Hyndman, 2006).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Where n is the number of observations, y_i is the observed value and $\hat{y}_i$ is the predicted value.

5 Data and time frame

A famous expression regarding forecasting states that a model is only as good as its inputs. The overall mission is therefore to choose the inputs and following characteristics to maximize predictive power.

5.1 Sample period

It is important that the sample period should reflect similar economic conditions as today. If the economy has gone through structural changes the variables could correlate in a different way with the response variable (Hansen, 2001). The Covid-19 pandemic could be viewed as the beginning of structural changes due to the disruptions causing systematic changes (Anayi et al. 2020). When the algorithm trains on historical data containing different economic conditions, the prediction will miss important aspects of how the economy works during the forecasted period.

Lindholm et al. (2018) states that inflation targeting was first introduced in Sweden 1993 but that it took several years until it was an integrated part of the economy. Also, the Industrial Agreement that is of major importance for wage determination and therefore impacts inflation was first introduced 1997. Lindholm et al. argues that the historical data, for predicting Swedish inflation, should be from earliest 1997 to avoid irrelevant data. The Swedish Riksbank (2019) uses data from 1995 in their inflation forecasts and therefore this thesis will follow a similar practice.

In addition to this, many of the datasets used in this analysis start their monthly observations around these years. National Institute of Economic Research, NIER, started their monthly surveys for many important variables in January 2000. Statistics Sweden, SCB, also has a lot of time series starting around 1997. As the RF algorithm cannot handle missing values (Ronaghan, 2018) the data needs a common first and last observation. When taking structural breaks, previous research, and data availability into account, the usage of data between January 2000 - January 2020 is suitable.

5.2 Inflation

Inflation could be measured in several ways. The most known is annual change in consumer price index, *CPI*. In this thesis the inflation measure based on the consumer price index but with fixed interest rate, *CPIF*, will be used

$$Inflation_t = \frac{CPIF_t - CPIF_{t-12}}{CPIF_{t-12}} * 100$$

Where t denotes the month.

Since 2017 the central Bank of Sweden (Riksbank) has used CPIF instead of the previously used CPI as target variable for inflation (Riksbank, 2021). This means that the goal of the monetary policy is to hover the annual change in CPIF around 2 percent from year to year (Riksbank, 2021). The reason that Riksbanken replaced CPI was because it was more volatile as changes in the interest rate affected mortages which was included in the CPI. CPIF is a more suitable measure for underlying inflation as it holds interest rates constant and therefore gives better guidance when conducting monetary policy (Riksbank, 2021). As CPIF-inflation now is their target variable and has de facto been for several years (SCB, 2021) it is reasonable to use this measurement when forecasting. The outcome of the analysis will then have more economic relevance.

The response variable, inflation, will be taken from Statistics Sweden (SCB, 2021). SCB releases the official measurements of inflation with a monthly frequency. The data is released between 10 and 20 days after the month has ended but usually after 14 days (Stenberg, 2021)

5.3 Predictor variables

As Chen et al. (2019) mentions there are two alternatives when selecting data for machine learning algorithms. First off there is the method called *kitchen sink* where you use many variables without thoroughly motivating the choices. The other strategy is called *cherry picking* from which you choose a few variables that has already been proven to correlate with inflation. This thesis will conduct a strategy that lies in between the two and use 50 variables that relate to inflation. Most of these variables are commonly used as economic indicators.

Following Araujo & Gaglianone (2019) and Riksbank (2018) the variables will be chosen to represent several fields of the economy. Riksbank (2018) states that their forecasts are based on economic activity, inflation abroad, financial markets but also soft variable as behavior and confidence measures. This thesis will use similar practice and include variables from four subgroups: Consumers, Monetary, Global and Industry. For more information see Appendix.

Table 1. Variable list

Consumer	**Monetary**	**Global**	**Industry**
Unemployment	Fiscal Expenditure	Balance of Trade	Total Industry Indicator
Consumer Confidence	Interest Rate	KIX Index	Manufacturing Indicator
Households Lending	Inflation	Trade Indicator	Oil Price
Retail Sale	Inflation Expectations	Retail Trade Indicator	Consumer Goods Indicator
Economic Tendency Indicator	M1 Supply	Import Prices	Producer Price Index
Job Vacancies	M3 Supply	Export Prices	Construction Index
Private Sector Index	Central Bank Balance Sheet	Imports	Industry Production
Wages	Government Debt	Inflation Expectations US	Order Books
Bankruptcies	Private Sector Lending	Inflation US	Investments Indicator
Household Consumption	Foreign Exchange Reserves	Inflation Expectations UK	Business Confidence
Food Prices	Budget Value	Inflation UK	Exports

Many of the hard variable are from Riksbank or SCB. The soft variables used come from NIERs monthly economic tendency survey and is the Swedish part of the data set used by Biau & D´elia (2010). International variables are taken from Citigroup, Office for national statistics, Federal Reserve and U.S Bureau of Labor Statistics.

The methods of Tyralis & Papacharalampous (2017) will be used by adding three lags of both inflation and the predictor variables. Following Woloszko (2020), a growth calculated version of all variables is also used to highlight drastic changes.

In addition to these variables, *Date* and *Order* is used as variables which focuses of the time dependency of inflation. The algorithm will connect the time of the observation to the observed inflation instead of connecting inflation to an activity measure like *Fiscal Expenditures* or *M1 Supply*

5.4 Data transformation

One drawback when working with time series in Random Forest is that the algorithm assumes that observations are independent from each other, and therefor does not incorporate the non-stationarity of the predictor variables (Goehry et al. 2021). Non-stationarity implies that the mean or variance change over time (Hyndman & Athanasopoulos, 2018).

When the RF algorithm decides on the splitting thresholds the decisions will thereforbe based on time as the trend is a effect from time. To use the time series with trends as variables a transformation is needed. The goal for the algorithm is that it should be able to spot that a high values of a certain variable, for example fiscal expenses, indicates higher inflation. When the trend is removed by differencing the algorithm will find these patterns.

When calculating the growth variables mentioned in 5.3 Predictor Variables, some values turned to NaN and some to INF. NaN means that it is not a valid number and INF is an infinite value. All calculations done with these values will eventually destroy the forecast as the algorithm cannot handle non numerical values(Ronaghan, 2018). All NaN values came from the error by having a fraction with only zeros as inputs. These observations were for the period where the repo rate was left untouched at 0. As a change from 0 to 0 is a 0% change these values were changed to 0. INF values were obtained if the denominator was equal to 0. For these values I followed the k-nearest neighbor strategy (Augustsson et al. 2019) and assigned the missing value with the mean of the two nearest observations.

5.5 Software

In this thesis the statistical open-source program R, version 4.1.1 (2021-08-10), has been used togheter with additional packages. Standard packages like tidyverse, readxl, tseries, readr were used for normal operation. For RF the package *randomForest*, derived by Liaw & Wiener (2018) from the original code in Fortran by Breiman & Cutler, was used. For the ARIMA and RW forecasts the package *forecast* (Hyndman, 2021) was used.

6 Results and analysis

When comparing the forecasts this thesis will follow the methodology of Biau & D´elia (2010) and Woloszko (2020) and examine the graphs, MAE, and measures of variable importance. This is done to get a better understanding of how and when the RF model could be used. The graphs are forecasted using the expanding window method (see 3.1.4 Random Forest with Time Series) and therefore all predicted values in a graph use all the available observation up to 1, 3, 6 or 12 months before the prediction. This also means that the variables are by default lagged with the forecasting horizon, e.g 1, 3, 6 or 12 months. For example: M1 Supply[t-1] translates to the M1 supply observed 1 month before the predicted value and M1 Supply [t-13] thereby 13 months. The red line indicates the forecasts, and the black line is the monthly values for inflation, i.e predicted vs observed. In the graphs y-axis will be inflation and x-axis the date.

6.1 Results from Auto.Arima test

By doing the tests mentioned in 4.2 ARIMA, the algorithm decided the best specifications. Due to the usage of expanding window the training data is updated and therefore the test resulted in 3 different types of models:

1: SARIMA (3,1,0)(1,0,1) [12]

2: SARIMA (2,1,1)(1,0,1) [12]

3: SARIMA (1,1,0)(0,0,1) [12]

Where 12 is the seasonal length, i.e 1 year.

These specifications translate into these equations:

1: $y'_t = \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \phi_3 y'_{t-3} + \phi_4 y'_{t-12} + \theta_1 \varepsilon'_{t-12} + \varepsilon_t$

2: $y'_t = \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + +\theta_1 \varepsilon'_{t-1} + \phi_3 y'_{t-12} + \theta_2 \varepsilon'_{t-12} + \varepsilon_t$

3: $y'_t = \phi_1 y'_{t-1} + \theta_1 \varepsilon'_{t-12} + \varepsilon_t$

The first equation was suggested from 2016 Jan- April 2017. From 2017 April – 2018 December the second equation was proposed and from January 2019 – January 2022 the last one was suggested.

6.2 Forecast results and analysis

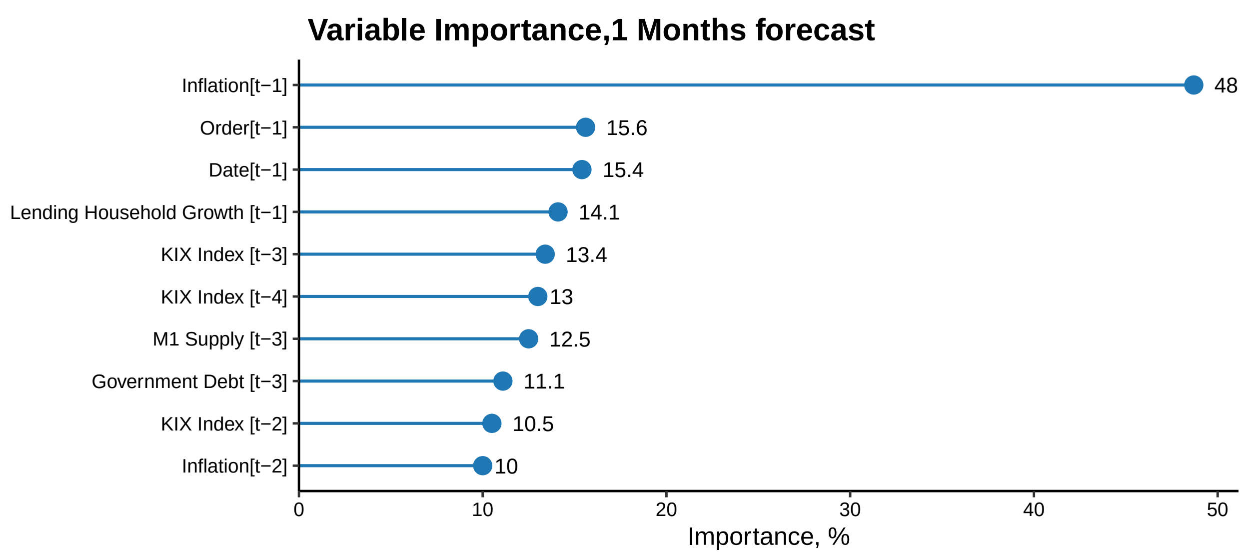

Figure 5: The 10 most important variables for RF 1 months forecast. As seen the forecasts depends highly on past values

By analysing the 1-month forecasts the results of basing the prediction on previous inflation were evident. This could be seen by having a close resemblance to the RW model which

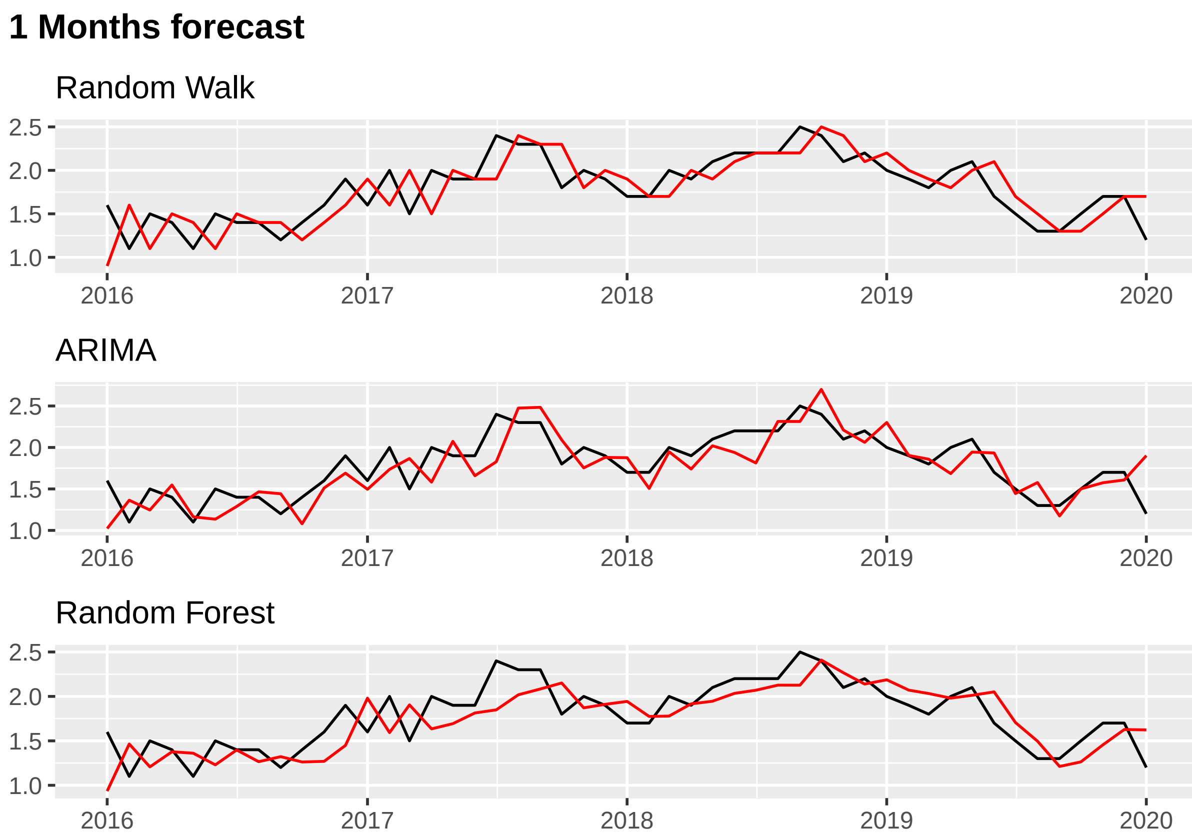

Figure 4: 1 Month forecast using RW, ARIMA and RF. All models have roughly the same MAE and follow the structure of a RW

predicts that the future value will be the current value. All models have roughly the same MAE, 0.21 compared to 0.22 (See table 2.), even if ARIMA performed slightly better. The importance of lagged values should come to no surprise for the RW and ARIMA forecasts as the equations, from 6.1 Results from Auto.Arima test, involves lags.

This pattern could be seen in the RF plot and especially prevalent from Oktober 2016 to Mars 2017 and after May 2019. The variable importance plot suggests that *Inflation*, *Order* and *Date* was the most important variables. Inflation has an importance of 48.7% which translates to a 48.7% increase in overall MSE if it was removed. These three variables all highlight the time dependency of inflation. Order and Date fill out the same purpose as the lagged inflation values as the algorithm will predict inflation on time instead of measurements of activity. If all removed the MSE would increase with roughly 80% according to the algorithm. If they where removed other variables which have similar characteristics could replace them and therefore it is hard to conclude that the MSE would increase with 80%. But it would certainly increase.

Other variables that were important for the forecast was *Lending Household Growth*, *KIX-Index*, *M1 Supply*, and *Government Debt.* KIX Index, which is a weight calculated exchange rate for the Swedish Krona, showed importance for values lagged 2, 3 and 4 months. This indicates that the strength of the Swedish Krona correlates with inflation.

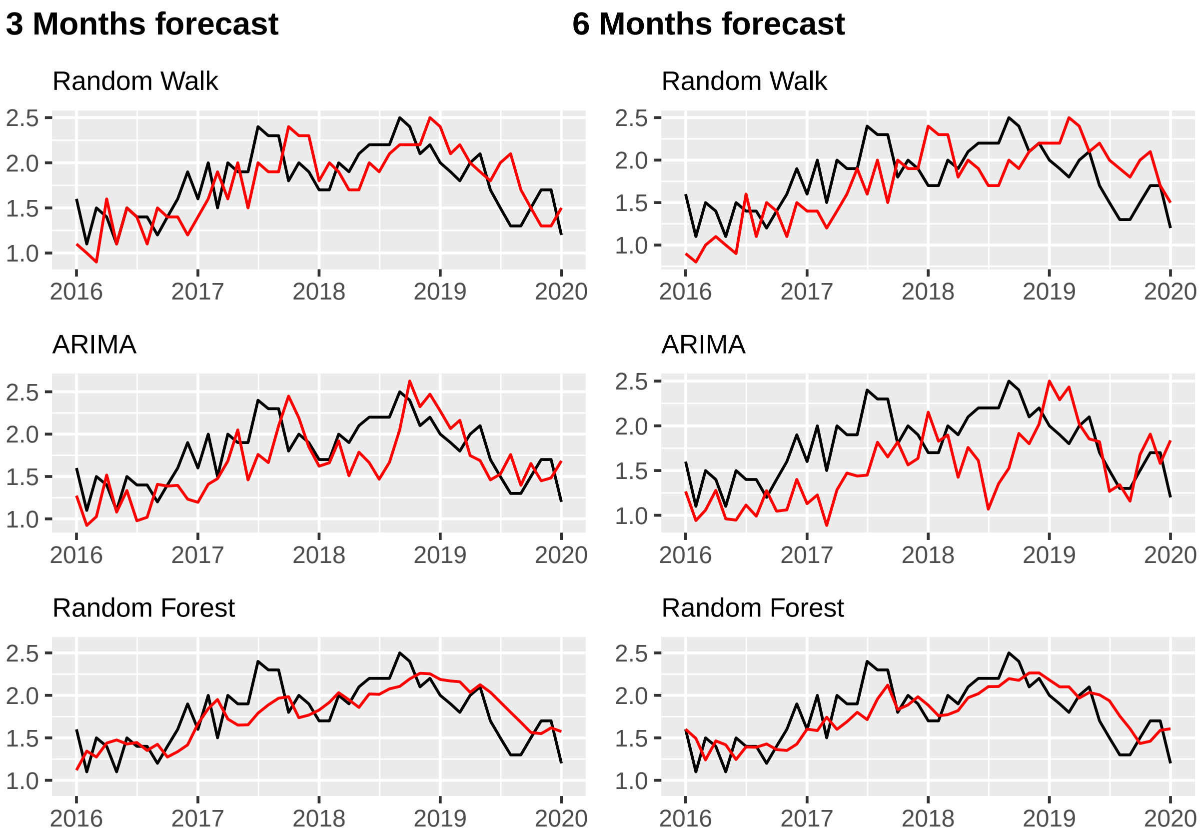

Figure 6: 3 and 6 months forecasts. From comparing these two the accuracy of RF more visible

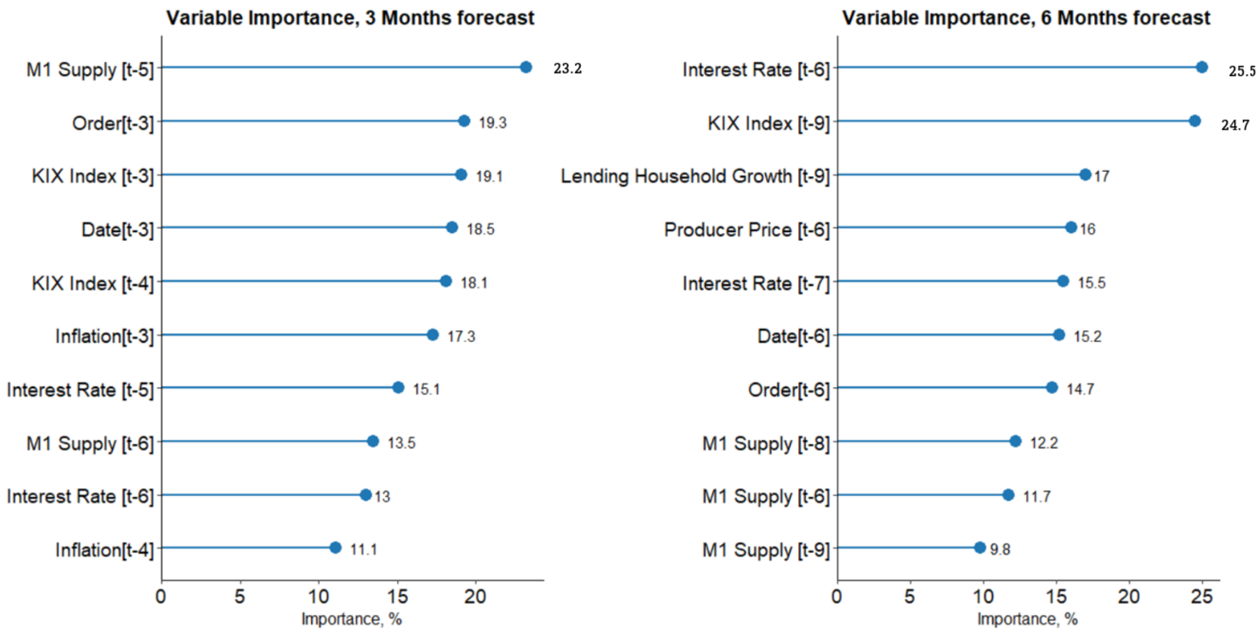

Figure 7: Variable importance for 3 and 6 months. When having a longer forecasts other variables are being used for the predictions

For the 3 and 6 months forecast the performance of RF is more prevalent. For the 3 months forecast the MAE of RF was 0.228 compared to 0.301 (ARIMA) and 0.286 (RW). For the 6 months forecast the MAE of RF was 0.206 compared to 0.401 and 0.355. This corresponds to a reduction in MAE of 20.3% and 42% compared to RW and 24.5% and 48.6% compared to the ARIMA (see table 3.) When increasing the forecast horizon in this study, RF adapt better compared to the benchmark models.

In contrast to the 1 month forecast, the 3- and 6-months forecast substitutes the lagged inflation values for measures of economic activity. This could be seen from the smoothness of the RF curve instead of the spiky features seen in the RW and ARIMA graphs. The variable importance plot also points to the significance of other varaibles. The time dependent variables place 2, 4 and 6 in the three months forecast but 6 and 7 for the six-month forecast. Lagged Inflation which was the most important variable in 1 month's forecast is not in the top 10 for the 6 months. These results indicates that the importance of these time dependent variables become less when having a longer forecast horizon.

From the variable importance plot of 3 months the 5- and 6-months lags of *M1 supply* and *Interest* Rate showed important for predicting inflation. The algorithm finds these longer lags more predictive then values closer to the predicted month. In the 6 months forecast these two variables with 6- and 7-months lag are also among the most predictive. In addition to this, several lagged variations of the mentioned variables are placed in the top 10, indicating that the predictive power could increase by allowing the algorithm to combine newer data with older data.

In both the 3 and 6 months forecast *KIX Index* was among the most predictive variables. Other variables include was *Producer Price*, and *Lending Households Growth*, but the later only for the 6 months forecast.

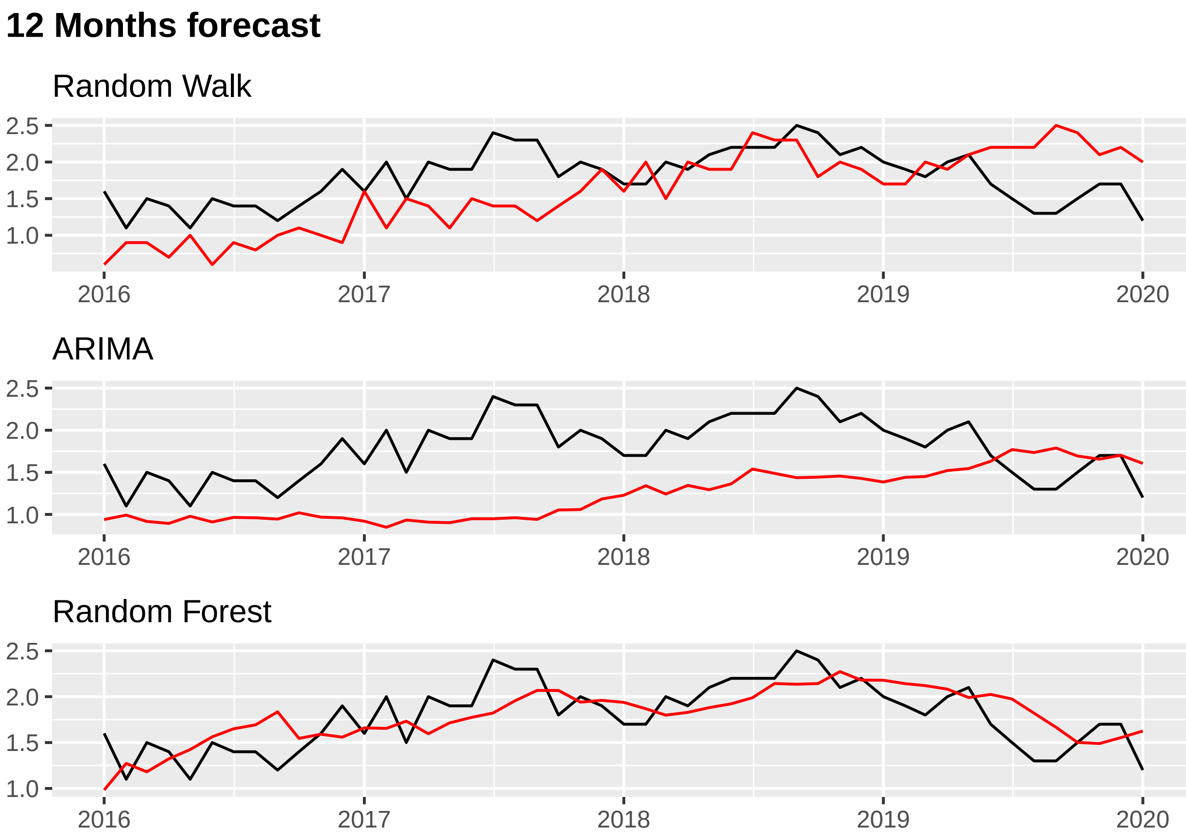

Figure 8: 12 Months forecast. The RW and ARIMA has worsened but RF gained a more smooth forecast

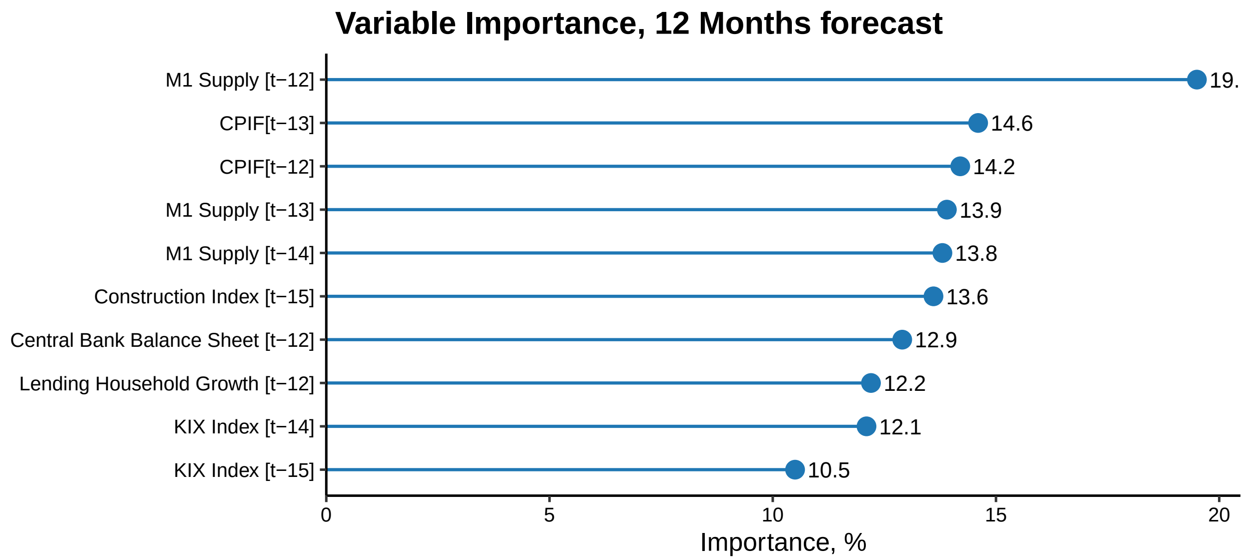

Figure 9: Variable importance for 12 months forecast

For the 12-month forecast the RF outperformed the other methods with an MAE of 0.241 compared to 0.625 (ARIMA) and 0.467 (RW). The corresponding MAE reduction was 48.4% and 61.4% which is a noticeable improvement. RW performed better then ARIMA which could show tendency of seasonal patterns as the inflation 12 months in the future were predicted by this month's inflation, but is could also be due to randomness. If it were a more volatile sample period, the RW would probably not predict that well. The ARIMA model for the 12 months forecast performs the worst. When using ARIMA for longer horizon the forecasts tend to converge to the mean of the training data (Hyndman & Athanasopoulos, 2018) which could be seen in the plot. The RF forecasts manages to follow the general path of inflation except for January 2016 - August 2016 where it predicted a rise when it actually turned out to float around an inflation of 1.3. Compared to the shorter inflation forecasts with RF the 12 months forecast is the smoothest but also the worst in terms of MAE.

From looking at the variable importance plot for the 12-month horizon *M1 supply*, *CPIF* and *KIX Index* gave best predictive power with several lag included. The reason why CPIF with 12 months lag is included could be understood by looking at equation 8 describing the inflation formula. CPIF from 12 months ago are included in the formula and therefore has a direct effect on inflation. However, why the algorithm decided to include CPIF with 13 months lag is mysterious. It could be due to similar value as CPIF with 12 lags, but additional test is needed to understand. *Central Bank Balance Sheet*, *Lending Household Growth and Construction Index*, was also included in the top 10.

6.3 Discussion

Table 2. Performance, in MAE.

Model/Horizon	1 Month	3 Month	6 Month	12 Month
Random Walk	0.218	0.286	0.355	0.467
ARIMA	0.212	0.301	0.401	0.625
Random Forest	0.219	0.228	0.206	0.241

Table 3. Random Forest MAE reduction compared to other models.

Model/Horizon	1 Month	3 Month	6 Month	12 Month
Random Walk	−0.5%	20.3%	42%	48.4%
ARIMA	−3.3%	24.5%	48.6%	68.4%

When comparing MAE for the models and forecasting horizons the RF method beats both ARIMA and RW for all horizons except 1 month. For 1-month the ARIMA performed best. The conclusion drawn by Stock & Watson (2008) that no model consistently performs better seems valid for this study also. But as seen in this analysis, when forecasting for longer time periods RF outperforms the others. Table 3. describes the reduction in MAE compared to the benchmarks test. Compared to the RW, RF reduced the MAE with 48.4% and compared to the ARIMA 61.4% for the 12 months forecasts. As stated by Marcellino (2008) these univariate models are justified for time series analysis due to their predictive power. RF outperforms these models which indicates that RF has could be used for forecasting inflation. The predictive power of RF corresponds to the findings of Biau & D´elia (2010) Chen et al. (2019) and Araujo & Gaglianone (2019) in their articles about macroeconomic forecasting with RF. Also, Baybuza (2018) and Araujo & Gaglianone (2019), shows that inflation forecasting with RF generally predicts better compared to univariate benchmark models, except for the 1-month horizon.

As stated by Stock & Watson (2008) the univariate model tends to decay when forecasting for longer horizons. One reason why RF does not gradually worsen in the way that RW and ARIMA does could be due to the large number of possible variables to choose from. For example, when comparing the shorter and longer forecasts conducted in this thesis the importance of variables highlighting the time dependency such as *Inflation*, *Order* and *Date* diminish. RW and ARIMA has no variables to substitute from, but since the RF was provided 250 variables covering different aspects of the economy it could use the ones suitable for the forecasting horizon. Also, the increasing importance of other variables like *Producer Price Index* and *Construction Index*, could also demonstrate the flexibility of RF. Producer Price Index which measures cost of production showed predictive power for the 6-month forecast. As production comes before consumption, higher productions costs will put an upwards pressure on prices, but with a delayed effect. Additional test is needed to confirm any causal effect.

The previous mentioned argument could also indicate why the 6 months forecast gained better results, in terms of MAE, then the 1- and 3-months forecasts. The shocks or changes may need time before influencing the economy and inflation. If the central bank should raise the M1 supply, or if every household would take a loan today, then the effects would not be seen the following day as it takes time for people to spend money or make investments. Similiar results are also found in Batina & Nelson (2001) stating that shocks to variables like *M1 supply* and *Interest Rate* have a lagged effect on inflation which could provide explanations for the findings in this thesis. The lagged effect could explain why some variables will give better predictive power when allowing for more time to pass between observations and forecasts. The results of this thesis oppose the findings of Baybuza (2018) stating that predictions based on monthly observation perform bad for longer time periods.

The smoothing function seen in the longer RF forecasts is one of the reasons why the RF predictions are more accurate. By smoothing out the predictions the general trends are captured, and the forecasts minimizes big errors. Even if Faust & Wright (2013) and Stock & Watson (2008) did not use RF in their papers, their results could be applied to these findings. They concluded that averaging over multiple estimation, either by averaging over models or surveys, would increase the predictive power. As seen in 3.1.2 Random Forest: Bagging, RF builds the predictions by averaging over a large amount of estimation and thus creating less risk of biased estimations. Also, as stated in 3.1.3 Tuning, by allowing for bigger node sizes the predictions will take less extreme values which contributes to the smoothing pattern.

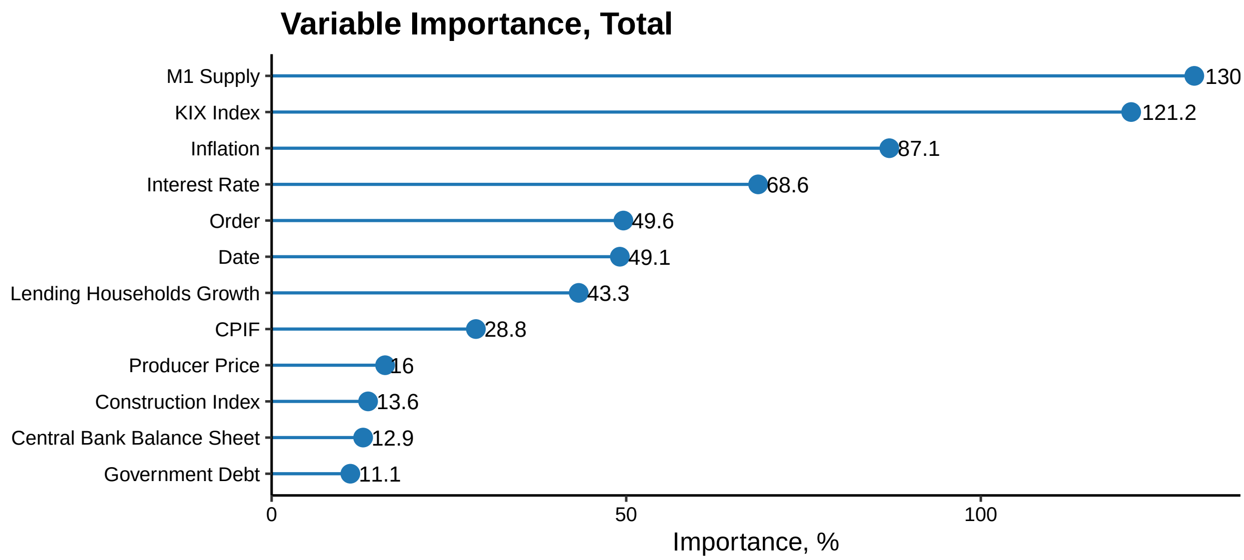

Figure 10: Variable importance for all forecast. M1 Supply, KIX Index and Inflation are the best predictors

The variable importance plot for all forecasted horizons indicates some variables that are repeatably showed to have predictive power. These are: *M1 Supply, KIX-index, Inflation, Interest Rate, Order, Date*, and *Lending Household Growth*. Date, Order, and Inflation all demonstrate the time dependency of inflation. The purpose of these variables is to mimic the function of lagged inflation values and therefor it could be treated as one. The significance of lagged values when forecasting inflation is shown in several articles, for example Stock & Watson (2008) and Baybuza (2018). *KIX-index, M1 Supply, Interest rate*, and *Lending Households Growth* also shows robust results throughout this study. The results from Ang et al. (2007) states that the most predictive variables usually are economic indicators and survey-data. In this thesis the economic indicators, such as the previous mentioned, had greater importance compared to the survey data used. These economic indicator variables are commonly used in inflation forecasting, for example by Riksbank (2021) for making their forecasts. Araujo & Gaglianone (2020) also find that these varaibles, but for Brazil, is predictive. To be completely sure of the variable importance additional tests and comparison would be needed to isolate the effects of the model and variable. But since these varaibles are known for their predictive power and commonly used in inflation forecasts (Ang et al, 2007) the conclusion that RF successfully determined variables that are predictive could be made

6.4 Limitations and future research

Despite showing promising results in this case study of Swedish inflation from 2016-2020 it is hard to draw any firm conclusions about the capacity of the RF model. As mentioned in 5.1 Sample Period, the chosen period was picked for several reasons. It had good data, showed little structural changes, and involved no big crises. When forecasting inflation these things are not possible to choose from to the same extent. Biau & D´elia (2010) mentions that the RF does not predict well in times of crisis as the model works by averaging values and therefore bad at predicting outliers. To overcome this problem in future research the adaptive version of the RF, proposes by Woloszko (2020) could be used as it puts more emphasis on drastic changes to the economy. Also, the minimum node size, discussed in 3.1.3 Tuning could be lowered to allow the predictions to take more extreme values.

A problem arising from the lack of interpretability is when trying to understand specific time frames. By looking at the RF plot in figure 8, the forecast for 2016 august and September shows a clear spike and then a drop. But since the RF model works as a black box, it is hard to detect what caused these flawed predictions. In future research the use of multivariate non-ML methods could be used to overcome these problems. For example, a linear regression model based of the RF variable importance selection, as done in Biau & D´elia (2010), could be made. By doing this it is possible to understand which variables that influenced the monthly prediction from comparong with the regression analysis. It would also be easier to conclude, by comparing MAE, if the gain in forecast accuracy is due to usage of multivariate data, the usage of RF or a combination of them.

It would also be interesting to compare the results from this study with forecasts from economic institutions such as OECD or the Riksbank. Due to different methodology and disposition in the forecasts such comparison where not possible in this thesis. The previously mentioned suggestions for future research would increase the relevance for policy makers as more rigid conclusion and comparison could be made.

Conclusions

This thesis has examined the performance of the machine learning method called Random Forest by forecasting Swedish monthly inflation between January 2016 and January 2020. It has also examined to what extent the RF algorithm could determine which variables that show predictive power.

The results from this thesis could be used to gain further knowledge about inflation and forecasting as it contributes to the fast-growing field of ML methods in economics. As this is the first time Swedish inflation has been forecasted with RF it provides evidence that the method could be used for Sweden as well. For all horizons, except 1 month, it was able to provide more accurate predictions then the benchmark tests. The largest gain, compared to the benchmark tests, was for longer forecasting horizons which indicates a flexibility of the RF method when choosing variables to include. This insight could be useful for policy makers and researchers when conducting analysis containing large number of variables. The study has also provided further arguments for the use lagged inflation values, and economic indicator variable, *KIX-index, M1 Supply, Interest rate, Lending Households Growth* in particular, when forecasting inflation.

To draw firmer conclusions about the findings of this thesis, additional tests are needed to fully understand the effect of the variables and the usage of RF. This could be done by conducting a multivariate analysis of the variables from the variable importance measures given by the RF algorithm.

Appendix

Table 4.1 Data Sources.

Variable Name	**Source**
Unemployment	Statistics Sweden
Consumer Confidence	National Institute of Economic Research
Households Lending Growth	Statistics Sweden
Retail Sale	Statistics Sweden
Economic Tendency Indicator	National Institute of Economic Research
Job Vacancies	Swedish Public Employment Agency
Private Sector Index	National Institute of Economic Research
Wages	Swedish National Mediation Office
Bankruptcies	Statistics Sweden
Households Consumption	Statistics Sweden
Food Prices	Statistics Sweden
Fiscal Expenditure	Swedish National Management Authority
Interest Rate	Central Bank of Sweden (Riksbank)
Inflation	Statistics Sweden
Inflation Expectations	National Institute of Economic Research
M1 Supply	Central Bank of Sweden (Riksbank)
M3 Supply	Central Bank of Sweden (Riksbank)
Central Bank Balance Sheet	Central Bank of Sweden (Riksbank)
Government Debt	Swedish National Debt Office
Private Sector Lending	Central Bank of Sweden (Riksbank)
Foreign Exchange Reserves	Central Bank of Sweden (Riksbank)

Table 4.2. Data Sources.

Variable Name	**Source**
Retail Trade Indicator	National Institute of Economic Research
Import Prices	Statistics Sweden
Export Prices	Statistics Sweden
Imports	Statistics Sweden
Inflation Expectations US	Federal Reserve Bank of New York
Inflation US	U.S Bureau of Labor Statistics
Inflation ExpectationK UK	YouGov/Citigroup
Inflation UK	Office for National Statistics
Total Industry Indicator	National Institute of Economic Research
Manufacturing Indicator	National Institute of Economic Research
Oil Princes	Statistics Sweden
Consumer Goods Indicator	National Institute of Economic Research
Producer Price Index	National Institute of Economic Research
Construction Index	National Institute of Economic Research
Industry Production	Statistics Sweden
Order Books	Statistics Sweden
Investments Indicator	National Institute of Economic Research
Business Confidence	National Institute of Economic Research

Table 4.3. Data Sources.

Variable Name	**Source**
Exports	Statistics Sweden
Budget Value	Swedish National Financial Management Authority
Balance of Trade	Statistics Sweden
KIX Index	National Institute of Economic Research
Trade Indicator	National Institute of Economic Research

Sources

Anayi, Lena; Bloom, Nicholas; Bunn, Philip; Mizen, Paul; Thwaites, Gregory; Young, Chris. 2021. Covid-19 and structural change. VoxEU. October 28. https://voxeu.org/article/covid-19-and-structural-change

Ang, Andrew; Bekaert, Geert; Wei, Min. 2007. Do macro variables, asset markets, or surveys forecast inflation better? Journal of Monetary Economics 54(4): 1163-1212. https://doi.org/10.1016/j.jmoneco.2006.04.006

Silva Araujo, Gustavo & Piazza Gaglianone, Wagner. 2020. Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models. https://www.cemla.org/actividades/2020-final/2020-10-xxv-meeting-cbrn/Session%202/3.%20Machine_Learning...%20Wagner%20Piazza.pdf (Retrieved: 2021-11-15)

Athey, Susan & W. Imbens, Guido. 2019. Machine Learning Methods Economists Should Know About. https://arxiv.org/pdf/1903.10075.pdf (Retrieved: 2021-12-10)

Augustsson, Malin; Güner, Levent; Zidic, Tanja. 2019. Comparing approaches for handling missing values in random forests. DOI:10.13140/RG.2.2.25683.43041 (Retrieved: 2021-11-13)

Batini, Nicoletta & Nelson, Edward. 2003. The Lag from Monetary Policy Actions to Inflation: Friedman Revisited. International Finance 4(3): 381-400. https://doi.org/10.1111/1468-2362.00079

Baybuza, Ivan. 2018. Inflation Forecasting Using Machine Learning Methods. Russian Journal of Money and Finance 77(4): 42-59. doi: 10.31477/rjmf.201804.42

Bernard, Simon; Heutte, Laurent; Adam, Sébastien. 2010. A Study of Strength and Correlation in Random Forests. Communications in Computer and Information Science 93:186-191. DOI:10.1007/978-3-642-14831-6_25

Biau, Olivier & D´elia, Angela. 2010. Euro Area GDP Forecast Using Large Survey Dataset - A Random Forest Approach, No 259600029, EcoMod2010, EcoMod, https://EconPapers.repec.org/RePEc:ekd:002596:259600029.

Breiman, Leo. 2001. Random Forests. Machine Learning 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

Breiman, Leo; Friedman, Jerome; J. Stone, Charles; Olshen, R.A. 1984. Classification and Regression Trees. Boca Raton: Taylor & Francis.

Breiman, Leo. 1996. Bagging predictors. Machine Learning 24(2): 123–140. https://doi.org/10.1007/BF00058655

Leo Breiman (2002). Manual for Setting Up, Using, and Understanding Random Forest V3.1. https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf

Brownlee, Jason. 2020. Random Forest for Time Series Forecasting. Machine Learning Mastery. [Blog]. November 2. https://machinelearningmastery.com/random-forest-for-time-series-forecasting/ (Retrieved: 2021-12-22)

Brownlee, Jason. 2016. Time Series Forecasting as Supervised Learning. Machine Learning Mastery. [Blog]. December 5. https://machinelearningmastery.com/time-series-forecasting-supervised-learning/ (Retrieved: 2021-12-22)

Chakraborty, Chiranjit & Joseph, Andreas. 2017. Machine Learning at Central Banks. Bank of England Working Paper No. 674. http://dx.doi.org/10.2139/ssrn.3031796

C. Chen, Jeffrey; Dunn, Abe; Hood, Kyle; Driessen, Alexander; Batch, Andrea. 2019. Off to the Races: A Comparison of Machine Learning and Alternative Data for Predicting Economic Indicators. Abraham, Katharine G.; S. Jarmin, Ron; Moyer, Brian; D. Shapiro, Matthew (Eds.). Big Data for Twenty-First-Century Economic Statistics. Chicago: University of Chicago Press.

Eldomiaty, Tarek; Saeed, Yasmeen; Hammam, Rasha; AboulSoud, Salma. 2020. The associations between stock prices, inflation rates, interest rates are still persistent. Empirical evidence from stock duration model. Journal of Economics, Finance and Administrative Science 25(49): 149-162. http://dx.doi.org/10.1108/jefas-10-2018-0105.

Faust, Jon & Wright, Jonathan. 2013. Forecasting Inflation. Handbook of Economic Forecasting vol 2: 2-56. DOI: 10.1016/B978-0-444-53683-9.00001-3

Fisher, Irving. 1907. The Rate of Interest. New York: The MacMillan Company

Galton, Francis. 1907. Vox Populi. Nature vol. 75: 450–451.
https://doi.org/10.1038/075450a0

Goehry, Benjamin; Yan, Hui; Goude, Yannig; Massart, Pascal; Poggi, Jean-Michel. 2021. Random Forests for Time Series. https://hal.archives-ouvertes.fr/hal-03129751 (Retrieved: 2021-12-28)

E. Hansen, Bruce. 2001. The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity. Journal of Economic Perspectives 15(4): 117–128.
https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.15.4.117

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. 2008. The Elements of Statistical Learning. Second edition. New York: Springer.
https://hastie.su.domains/Papers/ESLII.pdf

Hull, Isaiah; Löf, Mårten; Tibblin, Markus. 2017. Price information collected online and short-term inflation forecasts. Stockholm: Sveriges Riksbank.
https://www.riksbank.se/globalassets/media/rapporter/ekonomiska-kommentarer/engelska/2017/price-information-collected-online-and-short-term-inflation-forecasts.pdf (Retrieved: 2021-01-10)

Hyndman, Rob. 2021. forecast.
https://www.rdocumentation.org/packages/forecast/versions/8.15

Hyndman, Rob. 2006. Another Look at Forecast-Accuracy Metrics For Intermittent Demand. The International Journal of Applied Forecasting. 4:43-46
https://robjhyndman.com/papers/foresight.pdf

Hyndman, Rob J. & Athanasopoulos, George. 2018. Forecasting: Principles and Practice. 2nd edn. https://otexts.org/fpp2/

Hyndman, Rob J. & Khandakar, Yeasmin. 2008. Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software 27(3): 1-22.
https://doi.org/10.18637/jss.v027.i03

Keynes: https://link.springer.com/book/10.1007/978-3-319-70344-2 Keynes, John Maynard. 1936. The General Theory of Employment, Interest, and Money. Cambridge:

Palgrave Macmillan. https://link.springer.com/book/10.1007/978-3-319-70344-2

Kursa, Miron & Rudnicki, Witold. 2011. The All Relevant Feature Selection using Random Forest. University of Warsaw. https://www.researchgate.net/publication/51913500_The_All_Relevant_Feature_Selection_using_Random_Forest

Liaw, Andy. 2018. randomForest: Classification and Regression with Random Forest. https://www.rdocumentation.org/packages/randomForest/versions/4.6-14 (Retrieved: 2022-01-04)

Lindholm, Unn; Mossfeldt, Marcus; Stockhammar, Pär. 2018. Forecasting Inflation in Sweden. Stockholm: The National Institute of Economic Research (NIER). https://www.konj.se/download/18.1d3ad268167441d30e9a419b/1567433866935/Working%20paper%20152%20Forecasting%20inflation%20in%20Sweden.pdf%C2%A8

Marcellino, Massimiliano. 2008. A linear benchmark for forecasting GDP growth and inflation?. Journal of Forecasting 27(4): 305-340. https://doi.org/10.1002/for.1059 Nau, Robert. 2020. Statistical forecasting: notes on regression and time series analysis. https://people.duke.edu/~rnau/411home.htm

National Institute of Economic Research. 2021. User guide to the Economic Tendency Survey.
Stockholm.https://www.konj.se/download/18.1514d07715f05ba61e2eff52/1637913721511/Metodbok-UK.pdf

Pedregosa, Fabian; Varoquaux, Gaël; Gramfort, Alexandre; Michel, Vincent; Thirion, Bertrand; Grisel, Olivier; Blondel, Mathieu; Prettenhofer, Peter; Weiss, Ron; Dubourg, Vincent; Vanderplas, Jake; Passos, Alexandre; Cournapeau, David; Brucher, Matthieu; Perrot, Matthieu; Duchesnay, Édouard. 2011. Scikit-learn: Machine Learning in Python. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html (Retrieved: 2021-12-02)

Petalas, Diamantis Petropoulos; van Schie, Hein; Vettehen, Paul Hendriks. 2017. Forecasted economic change and the self-fulfilling prophecy in economic decision-making. PLoS ONE 12(3): Article e0174353 https://psycnet.apa.org/record/2017-13847-001

Probst, Philipp; Wright, Marvin; Boulesteix, Anne-Laure. 2019. Hyperparameters and Tuning Strategies for Random Forest. Wiri's Data Mining and Knowledge Discovery 9(3).

https://doi.org/10.1002/widm.1301

Sveriges Riksbank. 2018. The inflation target. https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target/ (Retrieved: 2021-12-10)

Sveriges Riksbank. 2018. The Riksbank makes forecasts. https://www.riksbank.se/en-gb/monetary-policy/the-path-to-a-monetary-policy-decision/the-riksbank-makes-forecasts/ (Retrieved: 2021-12-08)

Ronaghan, Stacey. 2018. The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark. https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3 (Retrieved: 2021-12-05)

SCB. 2021. Inflationstakten 3,6 procent i november 2021. https://scb.se/hitta-statistik/statistik-efter-amne/priser-och-konsumtion/konsumentprisindex/konsumentprisindex-kpi/pong/statistiknyhet/konsumentprisindex-kpi-november-2021/ (Retrieved: 2022-01-10)

SCB. 2017. Så mäter SCB inflation – skillnaden mellan KPI, KPIF och HIKP. https://www.scb.se/hitta-statistik/artiklar/2017/Sa-mater-SCB-inflation--skillnaden-mellan-KPI-KPIF-och-HIKP/ (Retrieved:2021-10-22)

L. Schaffer, Andrea; A. Dobbins, Timothy; Pearson, Sallie-Anne. 2021. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. BMC Medical Research Methodology volume 21(58). https://doi.org/10.1186/s12874-021-01235-8

Stenberg, Hannes, Statistikservice Avdelningen för social statistik och analys, SCB, kontakt via mail, 2021-11-22

Svensson, Lars E.O. 1999. Price Stability as a Target For Monetary Policy: Defining and Maintaining Price Stability. Cambridge: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w7276/w7276.pdf (Retrieved: 2021-11-10)

Stock, James H & Watson, Mark W. 2008. Phillips Curve Inflation Forecasts. Cambridge:

National Bureau of Economic Research.
https://www.nber.org/system/files/working_papers/w14322/w14322.pdf (Retrieved: 2021-12-12)

Tyralis, Hristos & Papacharalampous, Georgia A. 2017. Variable Selection in Time Series Forecasting Using Random Forests. Algorithms 10(4): 114-129.
https://doi.org/10.3390/a10040114

World Health Organization. 2020. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)

Woloszko, Nicolas. 2020. Adaptive Trees: a new approach to economic forecasting. OECD Economics Department Working Papers, No. 1593, OECD Publishing, Paris, https://doi.org/10.1787/5569a0aa-en.