AMERICAN JOURNAL OF APPLIED STATISTICS AND ECONOMICS (AJASE)

ISSN: 2992-927X (ONLINE)

VOLUME 3 ISSUE 1 (2024)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Applied Statistics and Economics (AJASE)

Volume 3 Issue 1, Year 2024
ISSN: 2992-927X (Online)
DOI: https://doi.org/10.54536/ajase.v3i1.2385
https://journals.e-palli.com/home/index.php/ajase

Evaluating the Efficacy of Supervised Machine Learning Models in Inflation Forecasting in Sri Lanka

W. M. S. Bandara[1*], W. A. R. De Mel[1]

Article Information

Received: January 01, 2024

Accepted: February 09, 2024

Published: February 12, 2024

Keywords

Cross-Validation, Macro Economic, Hyper-Parameter, Inflation Forecasting, Machine Learning

ABSTRACT

This study aims to forecast the inflation rate using supervised machine learning models (SMLM). While SMLMs are widely used in various fields, they have not been widely applied in forecasting inflation rates. Therefore, the main objective of this study is to identify the best model for forecasting inflation among four different SMLMs: LASSO regression (LR), Bayesian Ridge Regression (BRR), Support Vector Machine Regression (SVR), and Random Forest Regression (RFR) models. To achieve this objective, two different types of cross-validation techniques were employed: The k-fold cross-validation method (CV_K) and walk forward validation (WFV) methods. These techniques were used to estimate the parameters and hyper-parameters for each machine learning model with root mean square error. The mean absolute percentage error (MAPE) was used to compare the performance of the different SMLMs. Empirical evidence from Sri Lanka between 1988 and 2021 was used to test the performance of the SMLMs in forecasting inflation rates. The results show that the SVR model with walk-forward validation is the best method for forecasting the future inflation rate of Sri Lanka based on the MAPE value. Overall, this study showcases the effectiveness of Supervised Machine Learning Models (SMLMs) in forecasting inflation rates, emphasizing the critical role of precise cross-validation techniques. These findings are invaluable for policymakers and investors, offering advanced tools for more informed economic decision-making and highlighting the potential of machine learning in enhancing macroeconomic stability and forecasting accuracy.

INTRODUCTION

Inflation, a critical economic indicator, measures the rise in the general price level of goods and services over time. It impacts individuals, businesses, and the overall economy of a country. High inflation can lead to a decrease in the purchasing power of the currency, potentially causing economic instability, social unrest, and political turmoil (Maldeni, 2021) (Malladi, 2023) (Jayasooriya, 2015). Therefore, accurate forecasting of the inflation rate is essential to take preemptive measures to mitigate its adverse effects (Bandara & De Mel, 2021; Jaehyuk Choi, 2023).

Forecasting inflation rates is a challenging task due to several factors. Unpredictable events such as natural disasters, political and social conflicts, and global economic crises can impact the economy unexpectedly. economic variables' complex and dynamic relationships (Jaehyuk Choi, 2023). Hence, there is a need for advanced forecasting models that can handle the complexity of economic data and provide accurate predictions.

This paper focuses on the application of Machine Learning (ML) approaches for forecasting inflation, with empirical evidence from Sri Lanka (Maldeni, 2021). ML, a subset of artificial intelligence, has shown promising results in various fields, including economics. It can analyze large volumes of data, learn from it, and make predictions or decisions without being explicitly programmed. In the context of inflation forecasting, ML models can capture non-linear relationships between variables, adapt to changes, and improve their performance over time with more data.

Economic forecasting is crucial for policy-making and strategic planning (Anagaw, 2023). Accurate inflation forecasts can help the government and central banks implement appropriate monetary policies to maintain price stability. Businesses can also benefit from accurate inflation forecasts for budgeting, pricing, and investment decisions.

ML can contribute significantly to economic forecasting (Rahman *et al.*, 2021). It can handle large datasets, including economic indicators, market data, and social media sentiment, which traditional econometric models may find challenging. ML models can also adapt to new data, making them suitable for dynamic economic environments. Inflation forecasting is a critical aspect of economic stability, particularly for emerging economies like Sri Lanka. The country has faced periods of high inflation, significantly impacting its economy. Therefore, accurate and timely inflation forecasts are vital for maintaining economic growth and stability.

This paper aims to enhance the existing literature by applying Machine Learning (ML) approaches to forecast inflation in Sri Lanka. It will assess various ML models' performance and juxtapose them with traditional econometric models. The study's findings could offer valuable insights for policymakers, economists, and businesses in Sri Lanka and other emerging economies.

The current research intends to develop and evaluate machine learning models' efficacy in forecasting inflation rates in Sri Lanka. By addressing the limitations of traditional forecasting methods and exploring machine

[1] Deportment of Mathematics , University of Ruhuna, Sri Lanka
[*] Corresponding author's e-mail: bandarasudarshana009@gmail.com

learning models' potential, this study contributes to the existing body of knowledge. The findings could have practical implications for policymakers, central banks, and investors, enabling them to make informed decisions based on more precise inflation forecasts.

Therefore, this study holds significant importance, and the application of machine learning models in forecasting inflation rates is necessary to address traditional forecasting methods' limitations. The study's results could lead to more accurate inflation forecasts, contributing to economic stability and informed decision-making processes.

Best predicting performance was achieved with the blocked cross-validation method with respect to the RMSE statistic.

LITERATURE REVIEW

The current study aims to develop and evaluate the performance of machine learning models in forecasting inflation rates in Sri Lanka. The study contributes to the existing literature by addressing the limitations of traditional forecasting methods and exploring the potential of machine learning models in improving inflation forecasts. The findings of this study could have practical implications for policymakers, central banks, and investors in making informed decisions based on more accurate inflation forecasts. Therefore, this study is significant importance, and the application of machine learning models in forecasting inflation rates is necessary to address the limitations of traditional forecasting methods.

In seminal studies, various models such as univariate models and Phillips curve models have been utilized in forecasting inflation rates. (Jesmy, 2010) used Box–Jenkins's method to forecast the monthly mean inflation rate of Sri Lanka by using the historical inflation data (1952-2009). In this study, the univariate ARIMA(1, 1, 2) was selected as the best model using adjusted R-squared statistics. (Bandara, 2011)used the VAR models to forecast the inflation rate of Sri Lanka using the monthly mean historical data (1985-2005). (Jere, 2016) used the univariate time series models to forecast the inflation rate of Zambia by using Holt's Exponential Smoothing. However, due to differences between global and domestic political, social, environmental, country-specific conditions, and sample periods it becomes difficult to compare different models. Standard Phillips curve models (PCM), which rely on economic activity, have acted as a basis to the typical forecasting models of inflation. (Stock, 1999) also argue that these PCM based models outperform the traditional inflation forecasting models. (Atkeson, 2001), however, criticize this claim by showing that Phillips curve forecast of U.S. inflation over a 15-year period are no better than those obtained from a random walk model. Nevertheless, this instability of forecasting relationships is not limited to traditional Phillips curve-based models but extends to other theoretical or ad hoc empirical models used in the literature as well [see, e.g., models that include asset prices, for example, (Marcellino M. S., 2000) , (Goodhart, 2000) (Marcellino M. , 2002),]. Although forecasting specifications built adding one indicator of real activity at the time work poorly and tend to be unstable, some improvements have been documented by (Cristadoro, 2005) (Wright, 2003), (Granger, 2004), and (Inoue & Kilian, 2006), using methods that combine information obtained from many predictors.

In literature, various types of cross-validation methods with traditional forecasting methods were used to forecast inflation. For example, (Bergmeir & Benítez, 2012)used the cross-validation techniques with time series models where the stranded 5- fold cross-validation, blocked cross-validation, last block cross-validation, second block cross-validation, and second cross-validation methods were used. The best predicting performance was achieved with the blocked cross-validation method with respect to the RMSE statistic.

Machine learning models are rarely used in forecasting inflation data. (Volkan *et al.*, 2018) forecasted the core and non-core versions of inflation in the USA by using univariate Auto Regressive Distributed Lag (ARDL), Multivariate time series (VAR), SVR, k-nearest neighbour, and artificial neural network models. According to their results, ARDL provided the highest prediction accuracy for forecasting core-CPI inflation, while SVR outperformed the other models in forecasting core-inflation. All these machine learning models work better with more volatile and irregular series.

In this study, we conducted a simulation to evaluate the performance of four different supervised machine learning models, LR, BRR, SVR, and RFR, for inflation forecasting. The simulation involved training and testing each model using two types of cross-validation methods, Walk Forward Validation (WFV) and K-fold Cross-Validation (CV_K), to estimate the models' hyperparameters.

To compare the models and their performances, we used the mean absolute percentage error (MAPE) statistic. We also evaluated the stability and consistency of each model's performance across different time splits using the mean root mean square error (RMSE) metric.

Overall, our simulation aimed to identify the best machine learning model for forecasting the monthly mean inflation rate in Sri Lanka, considering different cross-validation methods and performance metrics. Our results provide useful insights into the effectiveness of different machine learning models for time-series data and can guide practitioners in selecting the most suitable model for their specific application.

The layout of this article is as follows. Section 2 provides a brief overview of machine learning models, cross-validation techniques and error calculated statistics, section 3 presents the inflation data set and the four covariates that are used in simulation study, the results of a simulation study and algorithms are presented in Section 4, and section 5 is devoted to the conclusion and future work.

MATERIALS AND METHODS

Models

In this subsection, we briefly explain supervised machine learning models, which are used to forecast inflation data, and two cross validation techniques.

LASSO Regression (LR)

According to (Ogutu *et al.*, 2012), "LASSO", "Least Absolute Shrinkage and Selection Operator", is an L_1 regularization technique that uses shrinkage, and because it automatically performs feature selection, LASSO can use a greater number of variables. The LR parameter estimate can be defined as follows;

$$\beta'(\text{lasso}) = \text{argmin}_{\beta}||Y - \beta X||_2^2 + \lambda||\beta||_1 \quad (1)$$

Where $||\beta||_1 = \sum^n_1 |\beta_i|$ is the L_1- norm penalty on β, which induces sparsity in the solution and $\lambda \geq 0$.

Bayesian Ridge Regression (BRR)

In Bayesian ridge regression (Hoerl, 1970), the estimate β is obtained by using the L_2 norm, and it is given by;

$$\beta'(\text{BRR}) = \text{argmin}_{\beta}||Y - \beta X||_2^2 + \lambda||\beta||_2^2 \quad (2)$$

Where $||\beta||_2 = \sum^n_1 \beta^2_i$ is the L_2- norm penalty on β and $\lambda \geq 0$. In this case, we obtain the posterior distribution to estimate β with normal likelihood and normal prior distribution.

Support Vector Regression (SVR)

SVR (Smola & Schölkopf, 2003) gives us the flexibility to define how much error is acceptable in our model. It will compute the parameter estimates by utilizing the following minimization problem.

Minimize

$$\min 1/2 \, ||\beta||^2 \quad (3)$$

Under constant

$$|y_i - \beta_i \, x_i \leq \epsilon| \quad (4)$$

Where we set the absolute error less than or equal to a specified margin, called the maximum error, ϵ. We can tune ϵ to gain the desired accuracy of our model.

Random Forests Regression (RFR)

RFR is a tree-based algorithm with each tree depending on a set of random variables (Cutler *et al.*, 2012). Let X = (X1,X2,...,Xp)' be a p-dimensional random input vector and Y be the response variable. Moreover, we assume that P_{XY} (X,Y) is the unknown joint distribution of X and Y . The objective of the RFR is to find a function f(X) to predict the response variable Y by minimizing the risk function.

$$E_{XY} (L(Y, f(X))) \quad (5)$$

Where $L(Y, f(X)) = (Y - f(X))^2$ is the squared error loss function. Here, one can define f(X) as f(x) = E(Y |X= x), and in regression setting, f(x) can be written as

$$f(x) = 1/J \sum^J_{(j=1)} h_j (x) \quad (6)$$

with respect to a collection of basis functions h_1 (x), h_2 (x),...,h_J (x).

Cross Validation Methods

Cross validation is a method used to increase the effectiveness of a machine learning model. It is based on re-sampling training data to train and test groups and evaluating model performance under over-fitting and under-fitting conditions. In this study, we use two types of cross validation methods, namely, K-fold cross-validation and walk forward validation.

K-Folds Cross Validation (CV_K)

In regression and classification settings, the K-fold cross-validation (Trevor Hastie, 2009) (CV_K) technique is commonly used due to its simplicity, fairness, and high effectiveness. The dataset is divided into K intervals, and one subinterval is used as test data while the remaining K-1 intervals are used as training data. By fitting the model K times and selecting the optimal K value that minimizes the Root Mean Square Error (RMSE), we can evaluate the model's performance.

Walk Forward Validation (WFV)

Walk forward validation (WFV), also known as time-series validation, is a technique commonly used for evaluating time series data. In this method, the entire dataset is divided into K intervals. The model is trained on the first interval and tested on the second. Then, the first two intervals are combined to train the model, which is tested on the third. This process is repeated until the first K - 1 intervals are used for training, and the remaining interval is used for testing. At each step, the model is fit, and the Root Mean Square Error (RMSE) is computed. The optimal K value is selected based on the minimum RMSE.

Hyper-Parameter Tuning

These hyper-parameters control various aspects of the model, such as the regularization strength, learning rate, and number of hidden layers in neural networks. Selecting optimal hyper-parameters is crucial for achieving high model performance, and grid search or random search techniques are often used to explore the hyper-parameter space and find the optimal values.

Error Calculation Methods

Let n,Y_t, and $\hat{Y}_t$ be the number of fitted points, the actual value of the response variable Y at time t, and the predicted value of Y_t, respectively.

Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage (Armstrong, 1992) Error (MAPE) can be calculated by using the following formula.

$MAPE = 1/n \sum^n_{(i=1)} (|Y_t - \hat{Y}_t|)/Y_t$.

MAPE works best in the absence of extreme values in the data set.

Root Mean Square Error (RMSE)

The Root Mean Square Error (Hyndman, 2006) (RMSE) is given by;

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{Y}_t)^2}.$$

The RMSE and MAPE are commonly used measures of the forecasting error.

Data Set

In this study, we consider the monthly inflation rate data in Sri Lanka from January 1988 to August 2021 (Tradingview, 2023) Figure 1 depicts this data.

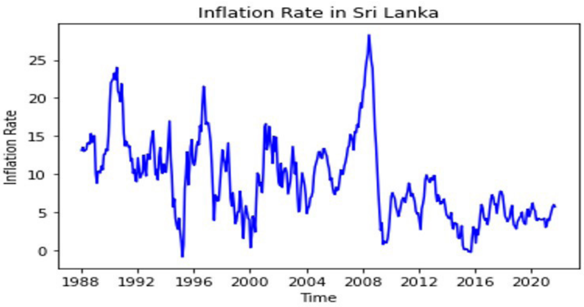

Figure 1: Monthly Mean Inflation Rate of Sri Lanka (1988-2021)

The time series plot in Figure 1 shows a stochastic behaviour of the inflation data, and one can notice that there are a few unusual data points, especially one at June in 2008. The reasoning for this may be that during this time period, the war in Sri Lanka was at a critical stage.

In order to fit the above four machining learning models, we have to convert the inflation rate data into a machine learning data set by introducing a new set of variables. The response variable, Y (t) is the inflation rate at time t. Here, we use four predictor variables: the first, second, third, and fourth differences of Y (t) and denote them as Y (t - 1),Y (t - 2),Y (t - 3) and Y (t - 4), respectively.

Simulation Study

The dataset used for the simulation study includes monthly data starting from January 1988. The dataset contains a total of 405 data points. To evaluate the performance of the model, the dataset was divided into three subsets: a training dataset consisting of 368 data points from January 1988 to February 2018, a test dataset consisting of 24 data points from March 2018 to March 2020, and a validation dataset consisting of 12 data points from April 2020 to August 2021. The simulation studies were performed using Python 3.8.5.

In this simulation, we extend our previous study by exploring four different supervised machine learning models: Lasso, Bayesian Ridge Regression (BRR), Support Vector Regression (SVR), and Random Forest Regression (RFR) to forecast the monthly mean inflation rate in Sri Lanka.

In the simulation, we randomly select n rows from the whole data set for different sample sizes ranging from 50 to 405. We repeat each sample size 100 times and calculate the mean of the RMSE for each machine learning model. The results for each machine learning model are plotted against the sample size, and the mean RMSE is used as a measure of the model's performance. The plots show the model performance for different sample sizes and highlight the optimal sample size required for each model. Overall, this simulation aims to evaluate the performance of four different machine learning models and identify the best model for forecasting the monthly mean inflation rate in Sri Lanka using CV_K.

Figure 2 compares the performance of the LR, BRR, SVR, and RF models using the CV_K approach at different sample sizes ranging from 50 to 405. The figure includes four subplots, each representing a different machine learning model. The x-axis represents the sample size, while the y-axis represents the mean RMSE value obtained for each sample size. The results demonstrate that the performance of each model varies with the sample size, with some models performing better than others at certain sample sizes. The LR and BRR models consistently exhibit the lowest mean RMSE values across all sample sizes. On the other hand, the SVR and RF models perform comparatively worse, particularly at smaller sample sizes. These findings suggest that the LR and BRR models may be more suitable for predicting the monthly mean inflation rate in Sri Lanka, especially when dealing with smaller sample sizes with CV_K.

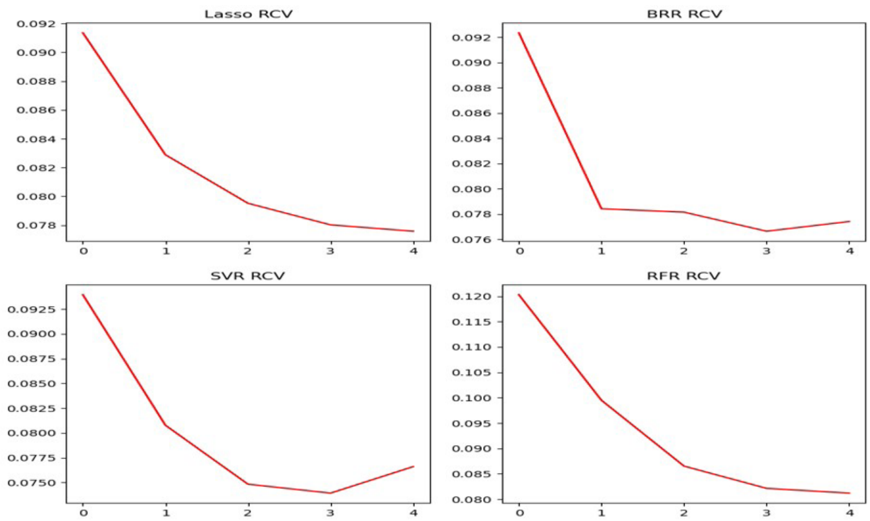

Figure 2: Comparison of Model Performance using K-fold Cross-Validation

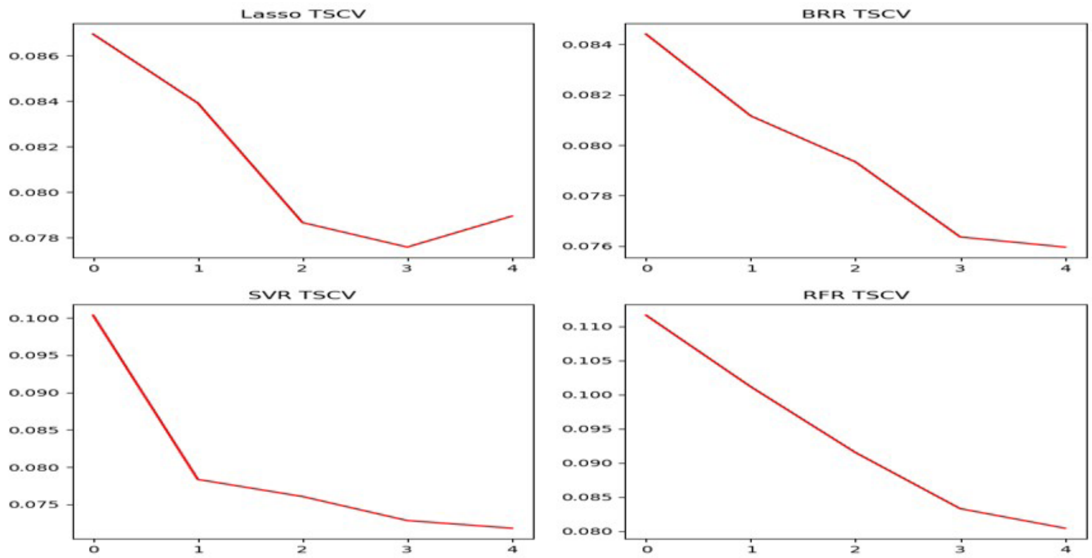

Figure 3: Comparison of Model Performance using WFV

Figure 3 is presented to compare the performance of LR, BRR, SVR, and RFR models using the WFV technique. Each subplot in the figure represents a different time split, and the mean RMSE is used to measure the model's performance. The plots demonstrate the stability and consistency of the models' performance across different time splits, indicating that WFV is a valuable technique for evaluating the performance of time-series data models. The simulation's objective is to assess the effectiveness of the four machine learning models and identify the most suitable model for predicting the monthly mean inflation rate in Sri Lanka using WFV.

Table 1: Comparison of SMLM 's RMSE with CV_K and WFV Techniques at Different Sample Sizes

RMSE					
Sample Size	**50**	**100**	**200**	**350**	**405**
LR_CV_K	0.091351	0.082883	0.079515	0.078016	0.077575
LR_ WFV	0.08694	0.083908	0.078658	0.07758	0.078942
BRR_CV_K	0.092358	0.078415	0.078144	0.076632	0.077401
BRR_WFV	0.084405	0.08116	0.079336	0.076359	0.075956
SVR_CV_K	0.093935	0.080772	0.074827	0.073928	0.076598
SVR_ WFV	0.100356	0.078344	0.076048	0.07283	0.071805
RF_CV_K	0.120294	0.099488	0.086512	0.082122	0.081185
RF_ WFV	0.111612	0.101143	0.091514	0.083301	0.080442

The Table 1 presents the performance of four different machine learning models, LR, BRR, SVR, and RF, using two different techniques: CV_K and WFV. The models were tested on five different sample sizes ranging from 50 to 405. The performance of each model was measured using mean RMSE.

The results indicate that, in general, the models performed better with larger sample sizes. Additionally, some models showed better performance with a particular technique. For example, LR, BRR, and RF models achieved lower RMSE values using WFV technique, while SVR showed better performance using CV_K technique.

Overall, the table highlights the importance of selecting an appropriate technique and sample size when building machine learning models for time-series data. It also provides useful insights into the performance of different models under different conditions, which can be helpful in selecting the most suitable model for specific applications.

Algorithm 1: Pseudo code for LR

Input: Response (**Y**) dependent on predictor(**X**) error tolerance ϵ

: LASSO solution for (1)

- Begin
- Data Preposing: Normalization X and Y
- Initialization $U = 0, \hat{y} = y - U$
- $c = X^T Y$
- $C = \text{Max}_j \{|c_j|\}$
- $\hat{j} = j\{|C_j|\}, A = \hat{j}$
- Calculate weight by R_w =PART_PAC(X) or $R_w = IW(X)$ or $R_w\ CRITIC(X)$
- Centralization R_w
- When $\| \hat{y} \|_L$ 1 and $|A| < m$, do β cycle
- End cycle
- Return (β)

Algorithm 1 is a pseudo code for the LR algorithm, which is used to predict the response (Y) dependent on predictor (X) with an error tolerance ϵ. The algorithm starts with data preprocessing and initialization, followed by weight calculation based on the chosen method. The LR algorithm iteratively cycles through β until the desired result is achieved. Finally, the algorithm returns the calculated β values.

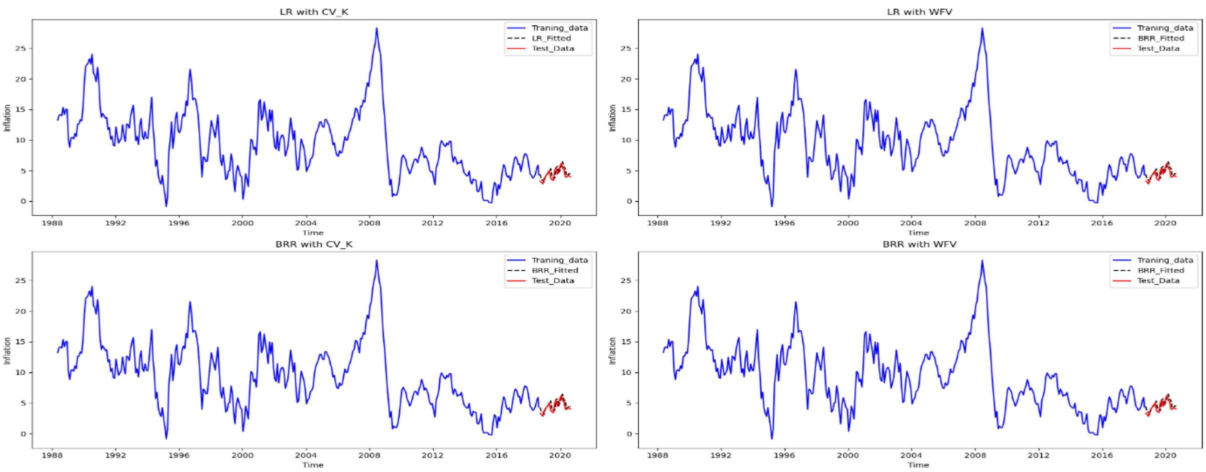

Figure 4: Fitted Test Inflation Rates Data for LR and BBR Models with CV_K and WFV Techniques

Figure 4 shows the fitted test inflation rates data for the LR and BRR models using two different techniques: CV_K and WFV. The LR model with CV_K technique predicts the inflation rates data using lasso regression while minimizing the prediction error with K-fold cross-validation. On the other hand, the BRR model with WFV technique uses Bayesian Ridge Regression to predict the inflation rates data while weighing the features by their variance.

From the graph, both LR and BRR models with CV_K technique provide similar fitted test inflation rates data for the entire period from October 2018 to September 2020. However, the LR model with CV_K technique predicts slightly higher inflation rates compared to the BRR model with CV_K technique. On the other hand, the BRR model with WFV technique predicts lower inflation rates than the LR model with CV_K technique, especially from February 2019 to September 2020. Overall, the LR model with CV_K technique and the BRR model with WFV technique provide different predictions for the inflation rates data, indicating the importance of choosing the appropriate model and technique for inflation rate prediction.

Algorithm 2: Pseudo code for BRR
Input: Response Y dependent on predictor X, error tolerance ϵ
Output: Bayesian Ridge Regression solution for (2)

- Data Preposing: Normalization X and Y
- Begin Initialization $U = 0, \hat{y} = y - U$
- $c = (Y - X\beta)^T (Y - X\beta) + \lambda\beta^T \beta$
- $C = \mathrm{Max}_j \{|c_j|\}$
- $\hat{j} = \arg\max j \{|C_j|\}, A = \hat{j}$
- Calculate weight by R_w =PART_PAC(X) or $R_w = IW(X)$ or $R_w\ CRITIC(X)$
- Centralization R_w
- When $\| \hat{y} \|_{L2} < 1$ and $|A| < m$, do β cycle
- End cycle
- Return (β)

The presented pseudo (Algorithm 2) code outlines the implementation of BRR algorithm, which is a popular technique for linear regression analysis. The algorithm takes input of response Y dependent on predictor X and error tolerance ϵ, and outputs the BRR solution.

The algorithm involves preprocessing of data by normalizing X and Y, followed by initialization of U, ŷ, and other variables. It then calculates the weight by either PART_PAC, IW, or CRITIC method, centralizes R_w, and performs the β cycle until specific conditions are met.

Overall, the BRR algorithm provides a robust and efficient solution for linear regression analysis, and its implementation can be tailored to different research needs based on the choice of weight calculation method.

Algorithm 3: Pseudo code for SVR
Input: S, λ, T, k
Initialize Choose: w_i s.t $\| w \| \leq \frac{1}{\lambda}$

For $t = 1, 2 \ldots T$
Choose $(A_i \not\subset S)$ Where $\| A \| = k$
Set $A_i^{\dagger} = \{(\mathrm{x,y}) \in \mathrm{A}_t : (\mathrm{W,x}) < 1\}$
Set $\eta_t = \frac{1}{\lambda t}$

Set $N_{t+1} = min\left\{1, \frac{1/\sqrt{\lambda}}{\|w_{t+\frac{1}{2}}\|}\right\} w_{t+\frac{1}{2}}$

Output: $w_{T+\frac{1}{2}}$

Algorithm 3 is the pseudo code for SVR, a popular regression algorithm used in machine learning. SVR involves choosing a set of support vectors from the input data and constructing a linear model to minimize the error between the predicted values and the actual values. The algorithm involves several iterations where support vectors are chosen, and the model is updated with the chosen support vectors until convergence is reached.

Figure 5 displays the fitted test inflation rates data for the SVR and RFR models using two different techniques: CV_K and WFV. The SVR model with CV_K technique predicts the inflation rates data using support vector regression while minimizing the prediction error with CV_K. On the other hand, the RFR model with WFV technique uses random forest regression to predict the inflation rates data while weighing the features by their variance.

From the graph, both SVR and RFR models with CV_K technique provide similar fitted test inflation rates data for the entire period from October 2018 to September 2020. However, the SVR model with CV_K technique predicts slightly higher inflation rates compared to the RFR model with WFV technique. On the other hand, the RFR model with WFV technique predicts lower inflation rates than the SVR model with CV_K technique, especially from February 2019 to September 2020.

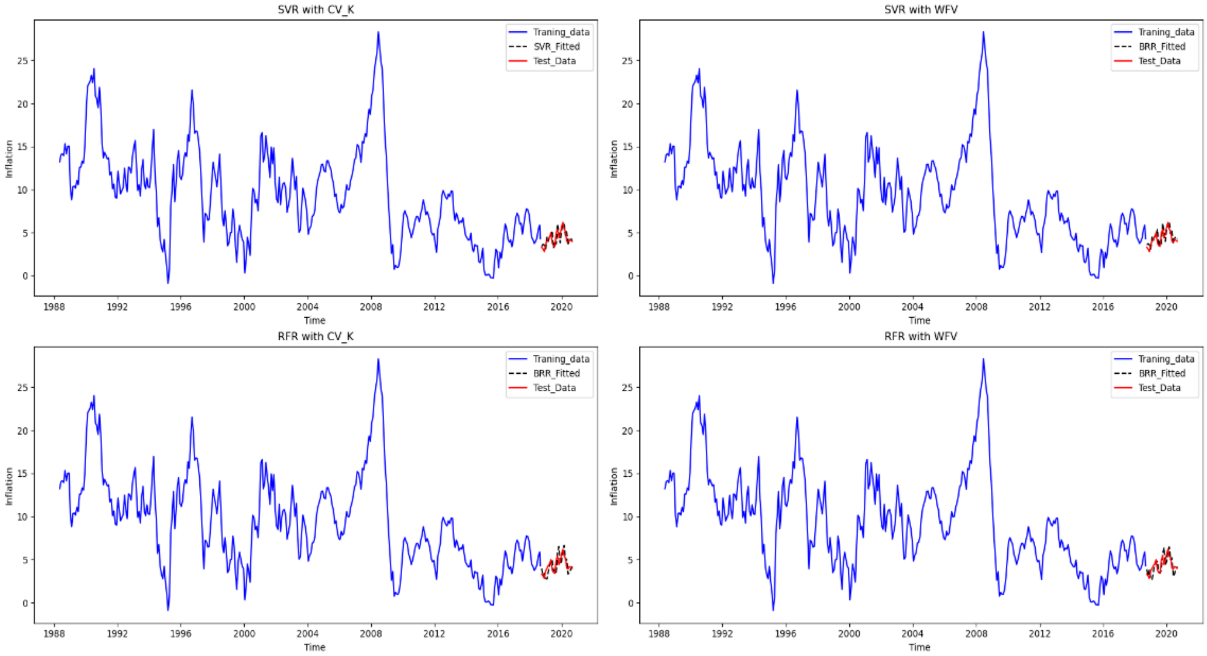

Figure 5: Fitted Test Inflation Rates Data for SVR and RFR Models with CV_K and WFV Techniques

Overall, the SVR model with CV_K technique and the RFR model with WFV technique provide different predictions for the inflation rates data, indicating the importance of choosing the appropriate model and technique for inflation rate prediction. The choice between these models and techniques may depend on the specific requirements of the application and the underlying data characteristics. Table 2 presents the predicted values of the test data for four different machine learning models: RFR, SVR, LASSO, and BRR. The table includes the actual values and predicted values for each model with CV_K. and WFV. The table spans from October 2018 to September 2020, with monthly predictions for each model. Overall, the table provides a comparison of the performance of the different models in predicting the test data over the two-year period.

Table 2: Predicted values of the test data for SMLM

Date		**RFR**		**SVR**		**LASSO**		**BRR**	
	Real	**CV_K**	**WFV**	**CV_K**	**WFV**	**CV_K**	**WFV**	**CV_K**	**WFV**
2018 Oct	3.3	3.98	3.86	3.42	3.61	4.38	4.38	4.23	4.23
2018 Nov	3.1	3.11	3.10	3.66	3.72	3.52	3.52	3.50	3.50
2018 Dec	2.8	3.42	3.79	3.55	3.57	3.44	3.44	3.47	3.47
2019 Jan	3.7	2.82	2.81	3.12	3.11	3.14	3.14	3.12	3.12
2019 Feb	4	2.74	2.71	4.48	4.49	4.15	4.14	4.20	4.20
2019 Mar	4.3	3.54	3.55	4.00	4.11	4.35	4.35	4.31	4.31
2019 Apr	4.5	4.27	4.47	4.54	4.66	4.63	4.63	4.60	4.60
2019 May	5	4.68	4.59	4.60	4.74	4.81	4.81	4.78	4.78
2019 Jun	3.8	4.94	4.81	5.21	5.38	5.32	5.32	5.32	5.32
2019 Jul	3.5	3.67	3.74	3.27	3.42	3.96	3.97	3.84	3.84
2019 Aug	3.4	3.81	3.42	4.00	4.03	3.80	3.80	3.80	3.80
2019 Sep	5	3.51	3.80	3.66	3.71	3.73	3.73	3.74	3.74
2019 Oct	5.4	5.77	5.68	5.97	6.08	5.46	5.46	5.55	5.55
2019 Nov	4.4	6.51	6.40	5.00	5.26	5.68	5.68	5.62	5.62
2019 Dec	4.8	4.49	4.41	3.83	4.00	4.55	4.56	4.43	4.43
2020 Jan	5.7	4.78	4.76	5.25	5.38	5.11	5.11	5.16	5.16
2020 Feb	6.2	6.08	6.04	5.90	6.09	6.03	6.03	6.07	6.07
2020 Mar	5.4	6.66	6.48	6.02	6.28	6.44	6.44	6.42	6.42
2020 Apr	5.2	4.97	5.01	4.66	4.89	5.52	5.52	5.42	5.42

2020 May	4	5.06	5.21	5.12	5.29	5.41	5.41	5.41	5.41
2020 Jun	3.9	3.33	3.04	3.56	3.72	4.15	4.15	4.08	4.08
2020 Jul	4.2	3.53	3.36	4.29	4.36	4.20	4.20	4.23	4.23
2020 Aug	4.1	3.85	3.97	4.39	4.49	4.54	4.54	4.56	4.56
2020 Sep	4	4.24	4.21	4.09	4.21	4.39	4.39	4.35	4.35

Algorithm 4: Pseudo code for RFR
Precondition: A training set $S: = (x_1, y_1), \ldots, (x_n, y_n)$, features F, and number of trees in forest B

- Start
- Function Random Forest (F, S)
- $H \leftarrow \emptyset$
- for $i \in 1 \cdot B$ do
- $S^i \leftarrow$ A bootstrap sample
- $h_i \leftarrow$ Randomized Tree Learn (Si , F)
- $H \leftarrow H \cup \{(hi)\}$
- End for, return H
- function: Randomized Tree Learn(S, F)
- At each node: f ← very small subset of F
- Split on best feature in f
- Return - learned tree

Algorithm 4 presents the pseudo code for RFR. RFR is a popular ensemble learning method used for both classification and regression problems. The algorithm builds a specified number of decision trees using a bootstrapped sample of the training data and selects a random subset of features at each node to split on. The final prediction is the average of the predictions from all the trees in the forest. The RFR algorithm is known for its ability to handle high dimensional data and avoid overfitting.

Table 3: MAPE Values in Test Data of each SMLM

CV_K		WFV	
Model	**MAPE**	**Model**	**MAPE**
LR	13.42	LR	13.43
BRR	13.04	BRR	13.40
SVR	12.79	SVR	13.34
RFR	15.82	RFR	15.63

Table 3 shows the MAPE values for four different supervised machine learning models (SMLMs) used to predict the inflation rate in Sri Lanka. The table presents the MAPE values for each model in two columns for two different cross validation methods CV_K and WFV.

Upon examining the MAPE values, it can be concluded that the support vector regression (SVR) model outperformed all the other models for both feature extraction techniques. For WFV, SVR had the lowest MAPE of 13.34%, followed by Bayesian ridge regression (BRR) at 13.40%, lasso regression (LR) at 13.43%, and random forest regression (RFR) at 15.63%. Similarly, for the other feature extraction technique, SVR had the lowest MAPE of 12.79%, followed by BRR at 13.04%, LR at 13.42%, and RFR at 15.82%. Therefore, SVR is the most accurate model for predicting the inflation rate in Sri Lanka based on the given features.

However, it is important to note that the differences in MAPE values among the models were relatively small, with differences of only a few percentage points. Therefore, it may be more appropriate to consider other factors, such as computational complexity and ease of implementation, when choosing a model for practical applications. In conclusion, the results suggest that the SVR model with WFV may be a suitable choice for predicting the inflation rate in Sri Lanka, but further validation and evaluation may be required to ensure the reliability of the results.

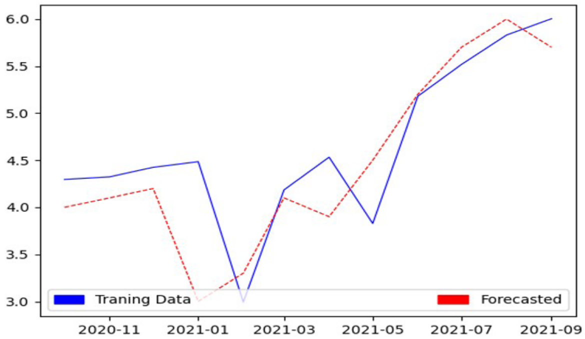

Figure 6: Forecasted values- SVR with WFV Techniques

The graph shows the actual values and forecast values of a certain variable over a period. The variable is denoted by Y_t while the forecasted values are generated using a machine learning model, namely Support Vector Regression (SVR). The graph consists of two lines: one line representing the actual values and another line representing the forecasted values.

The actual values are plotted as points on the graph, while the forecasted values are connected by a line. The graph enables a visual comparison between the actual values and the forecasted values. The closeness of the forecasted values to the actual values can be seen from the graph. Looking at the graph, it can be observed that the SVR model generally performed well in forecasting the variable. However, there are some instances where the forecasted values deviate from the actual values. For example, in February 2021, the actual value was 3.3, but the forecasted value was 2.992228, which is considerably lower. Similarly, in May 2021, the actual value was 4.5, but the forecasted value was 3.827932, which is also lower.

On the other hand, in September 2021, the actual value was 5.7, and the forecasted value was 6.004278, which is slightly higher. The closeness of the forecasted values to the actual values can be seen from the trend line of the forecasted values, which closely follows the trend line of the actual values. Overall, the graph provides a clear visualization of the performance of the SVR model in forecasting the variable. It highlights the instances where the model performed well and the instances where it deviated from the actual values. The insights from the graph can be used to further refine the machine learning model and improve its forecasting accuracy.

Table 4: Foretasted values for SVR model with WFV technique

Date	Yt	Forecast
2020 Oct	4.0	4.296212
2020 Nov	4.1	4.323517
2020 Dec	4.2	4.425066
2021 Jan	3.0	4.485011
2021 Feb	3.3	2.992228
2021 Mar	4.1	4.185231
2021 Apr	3.9	4.532705
2021 May	4.5	3.827932
2021 Jun	5.2	5.180228
2021 Jul	5.7	5.517756
2021 Aug	6.0	5.830174
2021 Sept	5.7	6.004278
MAPE		10.16

Table 4 shows the forecasted values generated by a Support Vector Regression (SVR) model using the wavelet-based feature vector (WFV) technique. The table includes the date, the actual values of the variable being forecasted (Yt), and the forecasted values generated by the model. The MAPE value of 10.16 % indicates that the average error of the forecasted values is approximately 10% of the actual values. The table demonstrates that the model was relatively accurate in predicting the values for the first few months, but the accuracy decreased in later months, with the largest discrepancy occurring in May. The model showed improvement in June and July but still underestimated the actual values in August and September. Overall, the table suggests that the SVR model with the WFV technique may be a suitable choice for forecasting the variable of interest, but further analysis and model refinement may be necessary to improve accuracy.

CONCLUSIONS

In conclusion, the support vector regression (SVR) model with the Walk Forward Validation (WFV) technique demonstrated superior performance in forecasting the inflation rate in Sri Lanka. The model's accuracy was measured using the mean absolute percentage error (MAPE) value, which was found to be 10.16%. The predicted values of the SVR model with WFV technique were compared to the actual inflation rates, and it was observed that the model's forecasts closely matched the actual values. However, the model's sensitivity to unusual data was noted, which could be addressed by using a more robust machine learning model. Overall, the findings suggest that the SVR model with WFV technique can be a valuable tool for forecasting inflation rates in Sri Lanka. As a future direction, it would be beneficial to consider the presence of outliers in the data and explore robust machine learning models that can handle them effectively. Additionally, incorporating other relevant economic and financial indicators into the model can potentially improve the accuracy of inflation rate forecasts.

REFERANCES

Anagaw, T. (2023). Review on: Effect of Inflation on Economic Growth in Ethiopia. *American Journal of Applied Statistics and Economics, 2*(1), 7–10. https://doi.org/10.54536/ajase.v2i1.1658

Armstrong, J. S. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of forecasting, 8*(1), 69-80.

Atkeson, A. O. (2001). Are Phillips curves useful for forecasting inflation. pp. 2–11.

Bandara, R. (2011). The Determinants of Inflation in Sri Lanka: An Application of the Vector Autoregression Model. *South Asia Economic Journal, 12*(2), 271-286. https://doi.org/10.1177/139156141101200204

Bandara,W. M. S & De Mel, W. A. R. (2021). ARIMA-Neural Hybrid Estimates of Inflationary Expectations: Some Evidence from Sri Lanka. *20th Academic Sessions Univercity of Ruhuna, 1*(1), 1. Retrieved from http://ir.lib.ruh.ac.lk/handle/iruor/13401

Bergmeir, C., & Benítez, J. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192-213. https://doi.org/10.1016/j.ins.2011.12.028

Cristadoro, R. M. (2005). A core inflation indicator for euro area. *Journal of Money, Credit and Banking, 37*(3), 539-560. Retrieved from http://www.jstor.org/stable/3839167.

Cutler, A., Cutler, D., & Stevens, J. R. (2012). Ensemble Machine Learning: Methods and Applications. (C. Zhang, & Y. Ma, Eds.) New york: Springer. https://doi.org/10.1007/978-1-4419-9326-7_5

Goodhart, C. a. (2000). Do asset prices help to predict consumer price inflation? (Vol. 5). Manchester School, . Retrieved from https://ssrn.com/abstract=242532

Granger, C. a. (2004). Thick modeling. *Economic Modelling, 21*(2), 323-343. Retrieved from https://EconPapers.repec.org/RePEc:eee:ecmode:v:21:y:2004:i:2:p:323-343

Hoerl, A. E. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics, 12*(1). Retrieved from https://doi.org/10.2307/1267351

Hyndman, R. J. (2006). Another look at measures of forecast accuracy. *International Journal of forecasting, 22*(4), 679-688.

Inoue, A., & Kilian, L. (2006). On the selection of forecasting models. *Journal of Econometrics, 137*(2), 273–306. https://doi.org/10.1016/j.jeconom.2005.03.003

Jaehyuk Choi, D. G. (2023). Yield Spread Selection in Predicting Recession Probabilities: A Machine Learning Approach. *Journal of Forecasting, 42*, 7.

Jayasooriya, D. (2015). MONEY SUPPLY AND INFLATION: EVIDENCE FROM SRI. *Asian Studies International Journal, 1*(1), 28.

Jere, S. ,. (2016). Forecasting inflation rate of zambia using holt's exponential. *Open Journal of Statistics*, 363-372. https://doi.org/ 10.4236/ojs.2016.62031

Jesmy, A. (2010). Estimation of future inflation in Sri Lanka using ARIMA model. ,. Kalam, 21-27.

Maldeni, R. &. (2021). A Machine Learning Approach to CCPI-Based Inflation Prediction. *Proceedings of Sixth*

International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, p. 236. https://doi.org/10.1007/978-981-16-2380-6_50

Malladi, R. K. (2023). enchmark Analysis of Machine Learning Methods to Forecast the U.S. Annual Inflation Rate During a High-Decile Inflation Period. *Computational Economics*. https://doi.org/10.1007/s10614-023-10436-w

Marcellino, M. (2002). Forecast Pooling for Short Time Series of Macroeconomic Variables. IGIER Innocenzo Gasparini Institute for Economic Research. Retrieved from https://ideas.repec.org/p/igi/igierp/212.html

Marcellino, M. S. (2000). A Dynamic Factor and Neural Networks Analysis of the Co-movement of Public Revenues in the EMU. *Ital Econ J 8*, 289–338. https://doi.org/10.1007/s40797-021-00155-2

Ogutu, J., Schulz, S., & Torben, P. (2012). Genomic selection using regularized linear regression models: Ridge regression, lasso, elastic net and their extensions. *BMC proceedings, 6*, S10. https://doi.org/10.1186/1753-6561-6-S2-S10

Rahman, M. A., Kabir, M. A., Haque, M. E., & Hossain, B. M. (2021). A Machine Learning-Based Price Prediction for Cows. *merican Journal of Agricultural Science, Engineering, and Technology, 5*(1), 64–69. https://doi.org/10.54536/ajaset.v5i1.63

Smola, A. J., & Schölkopf, B. (2003). A tutorial on support vector regression. *Statistics and computing, 14*(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Stock, J. H. (1999). Forecasting Inflation new index of aggregate activity. *Journal of Monetary Economics, 44*(2), 293-335. https://doi.org/10.1016/S0304-3932(99)00027-6

Tradingview. (2023). Sri Lanka Inflation Rate YoY. Trading View. Retrieved from https://www.tradingview.com/symbols/economics-lkiryy/

Trevor Hastie, R. T. (2009). *The Elements of Statistical Learning* (2 ed.). New York: Springer. https://doi.org/10.1007/978-0-387-84858-7

Volkan, U., Afsin, S., & Abdulhamit, S. (2018). Comparison of Time Series and Machine Learning Models for Inflation Forecasting: Empirical Evidence from the US. *Neural Computing and Applications, 30*, 1519–1527. https://doi.org/10.1007/s00521-016-2766-x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Melbourne: Springer.

Wright, J. H. (2003, September). Forecasting U.S. Inflation by Bayesian Model Averaging (September 2003). International Finance Discussion Paper, 1-33.http://dx.doi.org/10.2139/ssrn.457360