Comparative Analysis of Time Series Forecasting Models: A Case Study on the British Columbia Consumer Price Index

Vu Hoang Anh (Jennie) Nguyen

Bob Gaglardi School of Business and Economics, Thompson Rivers University

ECON 4330: Forecasting in Business & Economics

Dr. Mohammad Mahbobi

April 10, 2024

Table of Contents

Abstract

This paper presents a comparative study of various forecasting models applied to the British Columbia (BC) Consumer Price Index (CPI). The models under consideration include Holt Winters, Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet, Neural Prophet, and Long Short-Term Memory (LSTM). Each model's performance is evaluated d using several statistical metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) where applicable. The results provide insights into the strengths and weaknesses of each model, offering guidance for researchers and practitioners in selecting the most appropriate model for forecasting the BC CPI. Further research directions are suggested to enhance the forecasting accuracy and efficiency of these models.

Keywords: Forecasting models, CPI, model performance evaluation, forecasting accuracy

Introduction

The accuracy of forecasting economic indicators, such as the Consumer Price Index (CPI), is crucial in the field of economic planning and policy making. The domain of economic forecasting is in a state of constant progression, with the emergence of new models and techniques aimed at enhancing accuracy and efficiency. This paper investigates the application and performance assessment of several sophisticated forecasting models, specifically Holt Winters, Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet, Neural Prophet, and Long Short-Term Memory (LSTM), in the context of predicting the Consumer Price Index (CPI) for British Columbia (BC).

These models have found widespread use across diverse forecasting scenarios. For example, Holt Winters and SARIMA have proven their effectiveness in forecasting monthly streamflow (Brito et al., 2021). In addition, machine learning methods like LSTM have demonstrated promising outcomes in time-series forecasting (Kramar & Alchakov, 2023). The Prophet and Neural Prophet models, developed by Facebook, have also earned recognition for their robustness in managing a variety of time series data (Felton, 2021).

In this study, I undertake the evaluation of these models using a variety of statistical metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) where applicable. These metrics offer a comprehensive measure of each model's forecasting performance. The primary objective of this study is to identify the most fitting model for forecasting the BC CPI. The findings from this study will not only enrich the field of economic forecasting but also bear significant implications for economic policy and planning in BC. Moreover, this study sets the stage for future investigations aimed at enhancing the forecasting performance of these models.

Research Objectives

The primary aim of this study is to delve into the realm of economic forecasting, with a particular emphasis on the Consumer Price Index for British Columbia. The CPI is a vital economic indicator that reflects the average change over time in the prices paid by consumers for a market basket of consumer goods and services (Bryan & Cecchetti, 1993). Accurate forecasting of the CPI is essential for effective economic planning and policy making. This research seeks to contribute to this field by achieving the following detailed objectives:

1. Application of Forecasting Models: The first objective is to apply a range of advanced forecasting models to the task of predicting the BC CPI. These models include Holt Winters, SARIMA, Prophet, Neural Prophet, and LSTM. Each of these models has been extensively used in various forecasting scenarios and has proven effective in different contexts.
2. Evaluation of Model Performance: The second objective is to conduct an evaluation of the performance of these models. This will be done using a variety of statistical metrics such as RMSE, MAPE, AIC, and BIC where applicable. These metrics provide a comprehensive

measure of each model's forecasting performance and will allow for a thorough comparison of the models (Brito et al., 2021).

3. Comparison of Models: The third objective is to compare the forecasting performance of the models in the context of BC CPI prediction. This will involve analyzing the strengths and weaknesses of each model and identifying the most suitable model for this specific forecasting task.
4. Future Research Directions: The final objective is to identify potential areas for future research. This will involve suggesting ways to enhance the forecasting performance of the models and proposing new research questions or hypotheses based on the findings of the study.

Literature Review on Methods for Forecasting CPI

Forecasting CPI is crucial for many public policy areas and private business planning. Several traditional methods have been proposed and examined for this purpose.

Box-Jenkins Auto-Regressive Models

Traditionally, time series forecasting is performed using well-established statistical tools such as the Box-Jenkins Auto-Regressive models. These models have been widely used in forecasting CPI. The Box-Jenkins methodology involves the identification, estimation, and checking of ARIMA models. These models are fitted to time series data either to better understand the data or to predict future points in the series (Rosado et al., 2021).

Holt-Winters Exponential Smoothing

Another traditional method used in time series forecasting is the Holt-Winters approach to exponential smoothing. This method captures the level, trend, and seasonality in the data, making it suitable for CPI forecasting. The Holt-Winters method forecasts future values using weighted averages of past observations, with the weights decaying exponentially as the observations get older (Rosado et al., 2021).

Bayesian Model Averaging (BMA)

Bravo and Mekkaoui (2022) proposed a flexible Bayesian model averaging approach for CPI inflation models to mitigate conceptual uncertainty and improve short-term out-of-sample forecasting accuracy. The model space includes traditional univariate seasonal time series methods.

Bayesian Vector Autoregression (BVAR)

An extension of the Bayesian vector autoregression methodology has been proposed for forecasting macroeconomic variables, especially CPI inflation. The BVAR model incorporates prior information about the time series data to improve the forecast accuracy. This method has been widely used in macroeconomic forecasting due to its ability to handle multiple time series data and capture the dynamic relationships among different economic variables (Higgins et al., 2016).

Hierarchical Recurrent Neural Networks (HRNN)

Barkan et al. (2023) presented a novel model based on recurrent neural networks for forecasting disaggregated CPI inflation components. While the majority of existing research is focused on predicting headline inflation, many economic and financial institutions are interested in its partial disaggregated components.

Filtered Ensemble Wavelet Neural Network (FEWNet)

A novel approach proposed for forecasting CPI inflation is the filtered ensemble wavelet neural network (FEWNet) that can produce reliable long-term forecasts. This method combines the strengths of ensemble learning and wavelet transformation to enhance the predictive performance of neural networks. The FEWNet model has been shown to be particularly effective in handling non-linear and non-stationary time series data, which are common characteristics of CPI data (Rosado et al., 2021).

Methodology

Forecasting Methods

This research utilizes a spectrum of sophisticated forecasting models to forecast the Consumer Price Index for British Columbia. The models incorporated in this study encompass Holt Winters, SARIMA, Prophet, Neural Prophet, and Long Short-Term Memory.

Holt Winters

Holt Winters is a conventional time series forecasting model that has found extensive application across diverse forecasting scenarios. It is an exponential smoothing method that accounts for both trend and seasonality by applying a smoothing constant to the difference between the actual and forecasted values, allowing the model to adapt to changes over time (Felton, 2021).

SARIMA

SARIMA, on the other hand, is an extension of the Autoregressive Moving Average (ARMA) model. It is equipped with the capability to model seasonal components, making it particularly useful for data with clear periodic patterns. The model incorporates autoregressive (AR), differencing (I), and moving average (MA) parameters, each of which can be seasonally adjusted (Brito et al., 2021).

Prophet and Neural Prophet

Prophet and Neural Prophet are models developed by Facebook’s Core Data Science team. They have gained traction due to their resilience in dealing with a wide array of time series data. These models are engineered to handle common features of business time series, such as seasonality, holidays, and trend changes. They use a decomposable time series model with three main model components: trend, seasonality, and holidays. They are robust to missing data and shifts in the trend, and typically handle outliers well (Garlapati et al., 2021).

Long Short-Term Memory (LSTM)

LSTM is a type of recurrent neural network (RNN) that has the ability to learn and remember over long sequences. This makes it particularly effective for time series forecasting, where temporal dependencies are often crucial. Unlike traditional RNNs, LSTM has a unique design that helps it avoid the vanishing gradient problem, which is a common issue in training traditional neural networks (Bravo and Mekkaoui, 2022).

Model Implementation

The methodology of this study is outlined in the following steps:

Data Preprocessing

The initial step in forecasting the CPI involves preprocessing the data. The raw data is converted into a vertical format with two columns: Date and CPI. The Date column signifies each month of the year, and the CPI column holds the corresponding CPI values. To ensure the data's recency, only the data from the most recent 10 years is utilized for the analysis.

Model Training

The preprocessed data is subsequently divided into training and testing datasets. The standard practice is to allocate 80% of the data for training and 20% for testing. The training data is employed to train various forecasting models. Each model is trained separately, and the model parameters are tuned to achieve the best performance on the training data.

The parameters for both the SARIMA and Holt-Winter models were not arbitrarily chosen. They were determined through a systematic process of parameter tuning. This involved running a loop over various combinations of parameters and selecting the combination that resulted in the best performance according to the evaluation metrics (MAPE, RMSE, AIC, BIC). This approach ensures that the models are well-suited to the specific characteristics of the BC CPI data.

Model Evaluation

Once the models are trained, they are evaluated on the testing dataset, which was not used during the training process. This provides an unbiased assessment of the model's performance. The evaluation metrics could be MAPE, RMSE, AIC, and BIC.

- *MAPE* is a measure of prediction accuracy of a forecasting method in statistics. It expresses the accuracy as a ratio, where the numerator is the absolute difference between the actual and the forecast value, and the denominator is the actual value. The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points. MAPE is commonly used because it's easy to interpret (Chicco et al., 2021).
- *RMSE* is a metric used to evaluate the performance of a regression model. It is defined as the square root of the mean squared error (MSE), which is the average squared difference between the predicted values of the model and the true values of the data. The lower the RMSE, the better a given model is able to "fit" a dataset (Chicco et al., 2021).
- *AIC* and *BIC* are statistical measures used for model selection, taking into account both the goodness of fit and the complexity of the model (Susko & Roger, 2019). These criteria are typically applied in the context of models such as linear regression or time series models like SARIMA, where the likelihood can be easily computed due to specific assumptions about the distribution of the residuals (errors). However, models like Prophet and Neural Networks, which do not make specific distributional assumptions and use more complex, non-linear functions to fit the data, do not have a straightforward way to compute a likelihood value. This makes it challenging to compute AIC and BIC values for these models (Berk, 2022).

Model Comparison

The results from the different models are compared to identify the most suitable model for forecasting the BC CPI. The comparison is based on the evaluation metrics calculated in the previous step. The model with the best performance (i.e., lowest error metrics) is typically chosen as the final model.

Data Description

The dataset under consideration contains Consumer Price Index (CPI) data sourced from Statistics Canada (StatCan). The comprehensive dataset spans a period of 44 years, from 1979 to 2023. However, for the purpose of the analysis, I have chosen to focus on the most recent decade, encompassing the years 2014 to 2023. This selection has been made to enable the examination of the most current and relevant trends in the CPI data.

Figure 1

Decomposition of the data

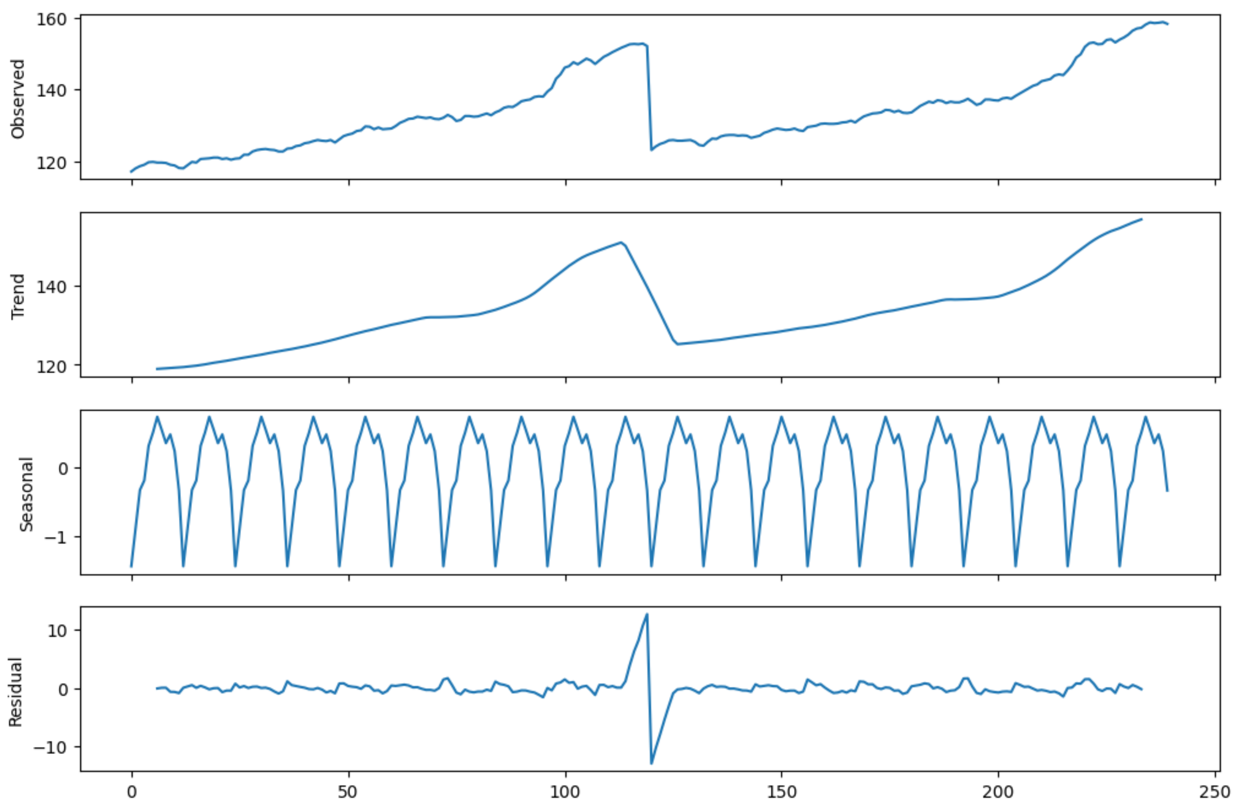

Figure 1 shows the decomposition of the data, which help understand the different components that contribute to the changes in CPI over time. The observed data is the actual data, the trend shows the overall direction of the data, the seasonal component shows the regular pattern of fluctuations, and the residual is what remains after the trend and seasonal components have been removed from the observed data. This analysis can help in understanding the underlying patterns in the CPI data and can be useful in forecasting future values.

- *Observed*: The observed (actual) data shows an overall increase over time but with a significant dip in the middle. It starts from below 120 and increases gradually until it reaches close to 160.
- *Trend*: The trend graph smoothens out the fluctuations seen in the observed data. It highlights an overall upward trajectory with a noticeable valley around the midpoint. This indicates that despite short-term fluctuations, there is a general increase in CPI over time.
- *Seasonal*: This graph exhibits regular oscillations indicating seasonal variations. The pattern is consistent, showing that there are regular increases and decreases in CPI at specific intervals throughout each period.
- *Residual*: The residual plot shows the noise left after extracting the trend and seasonal components from the observed data. There's a spike corresponding to where we saw a dip in both observed and trend graphs. Apart from this anomaly, residuals fluctuate mildly around zero.

Empirical Results

This section presents the empirical results obtained from applying several forecasting models to the BC CPI data. The performance of each model is evaluated using several metrics, including MAPE, RMSE, AIC, and BIC.

SARIMA Model

The SARIMA model was configured with an order of (1, 0, 1) and a seasonal order of (1, 0, 1, 12).

- The order (p, d, q) represents the autoregressive, differencing, and moving average terms of the model respectively. In this case, p=1 indicates that the model includes one lagged observation in the prediction, d=0 means that no differencing is used, and q=1 means that a moving average term is included in the model.
- The seasonal order (P, D, Q, s) is similar to the order (p, d, q), but it applies to the seasonal component of the model. In this case, P=1, D=0, and Q=1 indicate that the model includes one lagged observation, no differencing, and a moving average term in the seasonal component of the model. The s=12 indicates that the seasonality is annual (12 months).

The model achieved a MAPE of 2.39%, indicating a high level of accuracy. The RMSE was 4.72, suggesting a good fit of the model to the data. The AIC and BIC values were 132.97 and 145.79 respectively, suggesting a relatively good balance between model complexity and goodness of fit.

Holt-Winter Model

The Holt-Winter model was configured with parameters (‘add’, ‘add’, 12).

- The first ‘add’ indicates that the model uses additive trend. This means that the trend component of the time series data is modeled as a linearly increasing or decreasing series.
- The second ‘add’ indicates that the model uses additive seasonality. This means that the seasonal component of the time series data is modeled as a series of fixed additive changes across each season.
- The 12 indicates that the length of the seasonal cycle is 12 periods, which typically represents a year in monthly data.

The model achieved a RMSE of 14.3 and a MAPE of 9.21%. The AIC and BIC values were 164.86 and 123.83 respectively.

Prophet Model

The Prophet model achieved a MAPE of 7.53% and an RMSE of 11.72, suggesting a moderate level of accuracy and fit to the data.

Neural Prophet Model

The Neural Network model achieved a MAPE of 9.07% and an RMSE of 14.18, suggesting a moderate level of accuracy and fit to the data.

LSTM Model

The LSTM model, in terms of MAPE and RMSE, achieving values of 0.55% and 0.83 respectively, indicating a high level of accuracy and a strong fit to the data. The AIC and BIC values were 161.04 and 163.39 respectively.

Figure 2.

Actual and Forecasted Values for Different Models

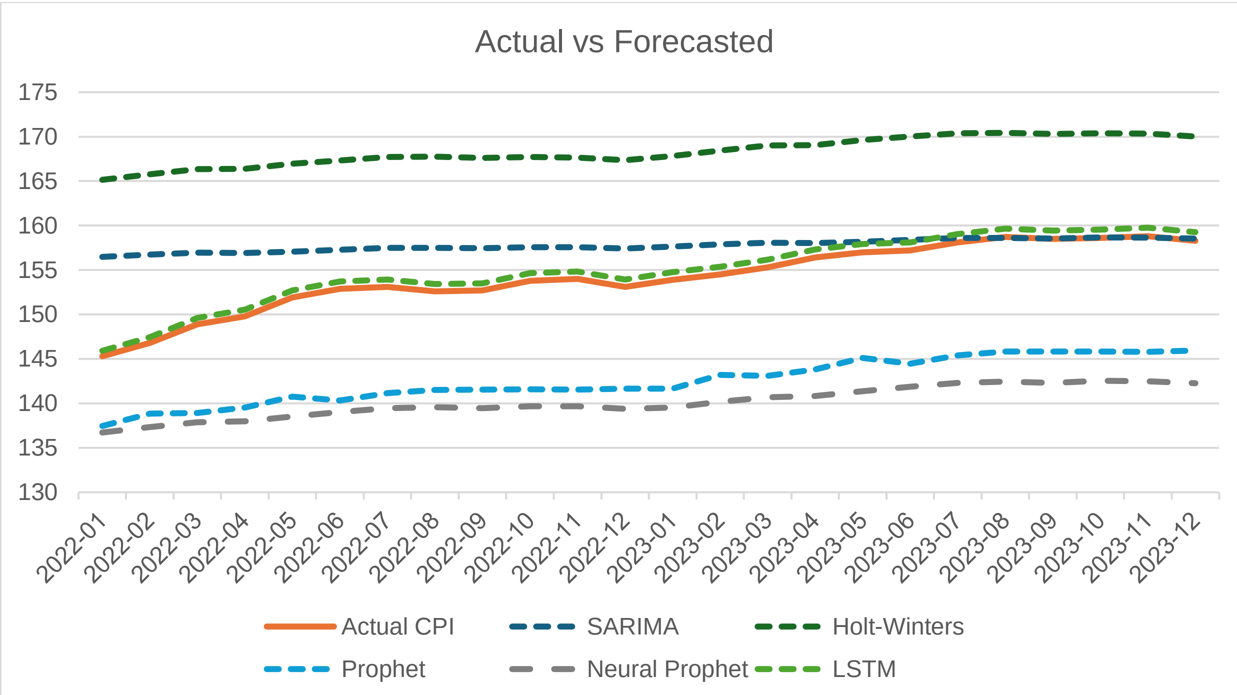

Table 1

Performance Metrics of Different Forecasting Models

Model	MAPE	RMSE	AIC	BIC
LSTM	0.55%	0.83	161.04	163.39
SARIMA	2.39%	4.72	132.97	145.79
Prophet	7.53%	11.72	-	-
Neural Prophet	9.07%	14.8	-	-
Holt-Winters	9.21%	14.3	164.86	123.83

Discussion

Based on the empirical results, we can observe that different models have varying levels of accuracy and fit when applied to the BC CPI data.

- The LSTM model stands out as the most well-performed models, achieving an exceptionally low MAPE value of 0.55% and RMSE of 0.83. This suggests that the LSTM model was not only able to predict the BC CPI data with a high degree of accuracy but also had a strong fit to the data.
- The SARIMA model also stands out as one of the most well-performed models based on the MAPE metric with a value of 2.39%. This suggests that the SARIMA model was able to predict the BC CPI data with a high degree of accuracy. Additionally, it achieved a relatively low RMSE of 4.72, indicating a good fit to the data.
- The Holt-Winter model, Prophet model, and Neural Prophet model showed moderate levels of accuracy and fit. The Holt-Winter model had a MAPE of 9.21% and an RMSE of 14.3. The Prophet model achieved a MAPE of 7.53% and an RMSE of 11.72. The Neural Prophet model had a MAPE of 9.07% and an RMSE of 14.18. These models may be less accurate in predicting the BC CPI data compared to the SARIMA and LSTM models.

In terms of model complexity and goodness of fit, as indicated by the AIC and BIC values, all models showed a relatively good balance. However, it’s worth noting that the SARIMA model had the lowest AIC and BIC values, suggesting that it might provide the best balance between model complexity and goodness of fit.

Future Research

The results of this study provide valuable insights into the performance of various forecasting models on the BC CPI data. However, there are several avenues for future research that could further enhance our understanding and prediction accuracy. By pursuing the following avenues, future research can continue to improve upon the forecasting accuracy achieved in this study and provide more robust and interpretable models for predicting BC CPI.

– Parameter Tuning: The LSTM and SARIMA models showed promising results in this study. However, there may be room for improvement through more extensive parameter tuning. Future research could explore a wider range of parameters to optimize these models further. This could involve using grid search or random search methodologies to systematically explore different combinations of parameters.
– Incorporating Additional Variables: This study used a univariate approach for forecasting, considering only a single variable at a time. Future research could explore the impact of incorporating additional variables into the models, such as economic indicators or other relevant factors. This could potentially improve forecast accuracy by capturing more complex relationships in the data.
– Exploring Other Models: This study examined a range of models, including LSTM, SARIMA, Prophet, Neural Prophet, and Holt-Winters. However, there are other forecasting models not covered in this study, such as Vector Autoregressive (VAR) models, ARIMA with exogenous variables (ARIMAX), or other types of neural networks. Future research could explore the performance of these models on the BC CPI data.
– Long-term Forecasting: This study focused on short-term forecasting. Future research could examine the performance of these models in long-term forecasting scenarios. This could provide valuable insights into the stability of the models over longer time horizons and their ability to capture long-term trends in the data.
– Model Interpretability: While LSTM showed high accuracy, it lacks interpretability compared to simpler models like SARIMA. Future research could explore methods to increase the interpretability of complex models like LSTM. This could involve techniques such as LIME

(Local Interpretable Model-Agnostic Explanations) or SHAP (SHapley Additive exPlanations).

Conclusion

This paper presented a comprehensive comparative study of various forecasting models applied to the British Columbia (BC) Consumer Price Index (CPI). The models under consideration, including Holt Winters, Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet, Neural Prophet, and Long Short-Term Memory (LSTM), were evaluated based on several statistical metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC).

The results of this study provide valuable insights into the strengths and weaknesses of each model. The LSTM and SARIMA models stood out as the most accurate models, offering promising avenues for future research and application. However, the choice of model should also consider other factors such as computational efficiency, interpretability, and the specific requirements of the forecasting task.

While all models provided valuable insights into the BC CPI data, there is always room for improvement and optimization. Future research directions suggested in this study aim to enhance the forecasting accuracy and efficiency of these models. These include more extensive parameter tuning, incorporating additional variables, exploring other models, focusing on long-term forecasting, and increasing the interpretability of complex models like LSTM.

In conclusion, this study contributes to the ongoing research in the field of time series forecasting, providing guidance for researchers and practitioners in selecting the most appropriate model for forecasting the BC CPI. The findings of this study highlight the importance of model selection in forecasting and pave the way for future research in this area.

Appendices

This study was conducted by building the forecasting models in Python, using the Google Colab platform. Google Colab is a cloud-based Python development environment that provides a platform for executing Python code and includes several machine learning libraries. This environment was chosen for its ease of use, extensive library support, and the ability to share and collaborate on code.

The Python code for the models was developed iteratively, with careful parameter tuning and model evaluation at each step. The code includes data preprocessing, model building, model training, and model evaluation steps. Each model's performance was evaluated using several metrics, including Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The link to the Python notebook on Google Colab is provided below:

Python Notebook on Google Colab

References

Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., & Koenigstein, N. (2023). Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting*, *39*(3), 1145–1162. https://doi.org/10.1016/j.ijforecast.2022.04.009

Berk, M. (2022, January 4). Prophet vs. NeuralProphet - Towards Data Science. *Medium*. https://towardsdatascience.com/prophet-vs-neuralprophet-fc717ab7a9d8

Bravo, J. M., & Mekkaoui, N. E. (2022). Short-Term CPI Inflation Forecasting: Probing with Model Combinations. In *Lecture notes in networks and systems* (pp. 564–578). https://doi.org/10.1007/978-3-031-04826-5_56

Brito, G. R. A., Villaverde, A. R., Quan, A. L., & Pérez, M. E. R. (2021). Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Applied Sciences/SN Applied Sciences*, *3*(6). https://doi.org/10.1007/s42452-021-04667-5

Bryan, M., & Cecchetti, S. G. (1993). *The consumer price index as a measure of inflation*. https://doi.org/10.3386/w4505

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ. Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Felton, I. (2021, December 10). A quick run-through of Holt-Winters, Seasonal ARIMA and FB Prophet. *Medium*. https://towardsdatascience.com/a-quick-run-through-of-holt-winters-sarima-and-fb-prophet-c6370c6f32f1

Garlapati, A., Krishna, D. R., Garlapati, K., Yaswanth, N. M. S., Rahul, U., & Narayanan, G. (2021). Stock Price Prediction Using Facebook Prophet and Arima Models. *2021 6th International Conference for Convergence in Technology (I2CT)*. https://doi.org/10.1109/i2ct51068.2021.9418057

Higgins, P. D., Zha, T., & Zhong, K. (2016). *Forecasting China's economic growth and inflation*. https://doi.org/10.3386/w22402

Kramar, V., & Alchakov, V. (2023). Time-Series forecasting of seasonal data using machine learning methods. *Algorithms*, *16*(5), 248. https://doi.org/10.3390/a16050248

Rosado, R., Abreu, A. J., Arencibia, J. C., González, H., & Hernandez, Y. (2021). Consumer price index forecasting based on univariate time series and a deep neural network. In *Lecture notes in computer science* (pp. 33–42). https://doi.org/10.1007/978-3-030-89691-1_4

Susko, E., & Roger, A. J. (2019). On the Use of Information Criteria for Model Selection in Phylogenetics. *Molecular Biology and Evolution*, *37*(2), 549–562. https://doi.org/10.1093/molbev/msz228

Verbrugge, R. J., & Zaman, S. (2021). Whose inflation expectations best predict inflation? *Economic Commentary (Cleveland)*, 1–7. https://doi.org/10.26509/frbc-ec-202119

Wu, Y., Arora, G., & Mei, X. (2024). Using CPI in loss given default forecasting models for commercial real estate portfolio. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2402.15498