Comparative Study of Models for Predicting Sri Lanka Consumer Price Index (CCPI) and Developing of Macroeconomic Condition Index

PROJECT PROPOSAL

Department of Mathematics
University of Ruhuna

by

SC/2021/12480- M.F. Abdullah

Supervisor: Senior Prof. Leslie Jayasekara

2025

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background of the Study

Among the most critical factors are price stability and control of inflation aspects of a nation's economic management. In Sri Lanka, the Consumer Price Index is the main indicator for inflation and cost of living, providing the main indications of household purchasing power and economic stability. Accurately predicting the CCPI is, therefore, crucial for fiscal policy formulation, investment decisions, and business planning.

In the Sri Lankan context, inflation is highly sensitive to supply side dynamics and external sector shocks. Key indicators such as domestic industrial production, which captures supply-chain constraints and domestic demand, and tourism arrivals, a critical source of foreign exchange and a driver of service sector demand, are therefore crucial for understanding and forecasting price movements.

Past studies conducted in Sri Lanka have mostly employed traditional econometric models including Vector Autoregression (VAR), and Autoregressive Distributed Lag (ARDL), and Ordinary Least Squares (OLS) regression in order to analyze the determinants of inflation. However, such methods assume linear relationships among variables and may fail to capture complex nonlinear interactions that exist in real-world data.

With increasing access to data analytics tools and computational Powerful machine learning techniques have recently gained much more popularity for forecasting and predictive modeling. It aims to combine both regression and machine learning approaches for predicting the CCPI in Sri Lanka and to develop a composite Macroeconomic Condition Index (MCI) using Principal Component Analysis (PCA).

Therefore, this study will focus on key monthly indicators including the Industrial Production Index (IPI), Purchasing Managers' Index (PMI), Broad Money Supply (M_2), Average Weighted Prime Lending Rate (AWPR), Exchange Rate, Diesel Sales Volume, and Tourism Arrivals as primary determinants for CCPI forecasting.

1.2 Research Problem

Although various studies have used statistical methods to analyze inflation and CCPI determinants, most of these methods implicitly assume linearity, and thus are poorly suitable to model the dynamic relationships between various macroeconomic factors. Moreover, only a few studies have applied computerized techniques(such as machine learning) to forecast the inflation in Sri Lanka.
There is also a lack of an integrated composite index that reflects the overall state of macroeconomic conditions. Therefore, This study tries to fill these gaps by comparing various statistical(regression) and computerized(machine learning) methods to obtain the best predictive model and construct a PCA based composite Macroeconomic Condition Index (MCI) in Sri Lanka.

1.3 Research Gap

The majority of Sri Lanka's current inflation research uses economic models rather than modern data driven techniques such as machine learning. In addition, these studies usually involve a limited number of variables and pass over high frequency(monthly) data.
There is limited research focusing on the specific roles of Industrial Production Index and tourism arrivals as key determinants of inflation in the Sri Lankan context, despite their theoretical importance in capturing supply side constraints and external sector dynamics.

Existing studies rely on analyzing numerous separate indicators like exchange rates, interest rates, and money supply which can be complex and difficult to interpret together. There is a clear need for a single, comprehensive index that can summarize this complex economic information into an easy to understand index.

1.4 Significance of the Study

This is important research both from the academic and practical point of view. The research contributes academically to the empirical literature by combining regression, machine learning, and dimensionality reduction in the context of Sri Lankan inflation analysis.

This research will provide practical tools to help government planners, businesses, and financial analysts better predict future changes in CCPI. To

make the findings easy to use, the project will also create an interactive, visual dashboard (using Power BI). This dashboard will clearly show the trends and forecasts, helping everyone make smarter, faster decisions about economic policy, business strategy, and investments.

1.5 Aims and Objectives

This project aims to build a smarter system for predicting inflation in Sri Lanka. Currently, economists rely on traditional statistical formulas, which can be limited and hard to interpret. I plan to improve this by testing modern methods against these older techniques to see which gives more accurate forecasts. Additionally, I will create a simple "Macro Economic Condition Index" that combines many complex economic indicators into one easy to understand number. The goal is to produce forecasts that are not only more reliable, but also clearer and more useful for policymakers, businesses, and everyday citizens who need to understand where prices are headed.

1.5.1 Specific Objectives

1. To build and prepare a comprehensive monthly dataset (January 2014 - December 2023) of Sri Lanka's key economic indicators including: Purchasing Managers' Index (PMI), Industrial Production Index (IPI), Broad Money Supply (M), Average Weighted Prime Lending Rate (AWPR), Exchange Rate (Rs./USD), Diesel Sales Volume, and Tourism Arrivals, ensuring the data is clean, consistent, and ready for analysis.

2. To develop and compare three different prediction models: a traditional statistical model (Multiple Linear Regression) and two advanced machine learning models (Random Forest and XGBoost), to determine which most accurately forecasts the Consumer Price Index (CCPI).

3. To create a single, easy to understand "Macroeconomic Composite Index" using a statistical technique (PCA) that combines the essential information from multiple economic indicators, simplifying the complex economic landscape.

4. To thoroughly evaluate the best model and create an interactive Power BI dashboard that clearly visualizes the forecasts, explains the economic drivers behind them, and serves as a practical tool for decision makers.

1.6 Research Questions

1. What are the most significant macroeconomic indicators that influence Consumer Price Index (CCPI) fluctuations in Sri Lanka?

2. Which modeling approach traditional Multiple Linear Regression or advanced machine learning algorithms (Random Forest and XGBoost) provides superior predictive accuracy for CCPI forecasting in Sri Lanka?

3. Can a significant proportion of the variance in multiple macroeconomic variables be captured by a single PCA derived Macroeconomic Condition Index (MCI)?

2 Literature Review

Inflation forecasting is central to macroeconomic policy and is grounded in several foundational theories. The Quantity Theory of Money Fisher [1911] establishes a direct link between money supply and the general price level. The Phillips Curve Phillips [1958] describes the inverse relationship between unemployment and wage inflation, later extended to price inflation. For small, open economies like Sri Lanka, the Exchange Rate Pass-Through Theory is particularly relevant, explaining how domestic prices respond to currency depreciation due to higher import costsGoldberg and Knetter [1997]. These theoretical frameworks provide the basis for empirical modeling but often rely on linear assumptions that may not capture real-world complexities.

Early empirical studies on Sri Lankan inflation predominantly used traditional econometric methods. Ratnasiri [2009] applied Vector Autoregression (VAR) models, identifying money supply growth and exchange rate fluctuations as key inflation drivers. Later, Jayawickrema and Lokupitiya [2018] employed Autoregressive Distributed Lag (ARDL) cointegration analysis, confirming long-run relationships between money supply, exchange rates, and inflation. While these studies provide valuable insights, they assume linear and static interactions among variables—a significant limitation given Sri Lanka's exposure to nonlinear shocks such as global crises and supply disruptions Stock and Watson [2002].

Beyond monetary factors, sectoral and high-frequency indicators offer deeper insights into inflation dynamics. Tourism arrivals significantly influence foreign exchange reserves and aggregate demand, yet their direct impact on inflation remains underexplored in Sri Lanka World Bank [2021]. Forward-looking indicators such as the Purchasing Managers' Index (PMI) provide early signals of economic activity and supply-chain pressuresKoenig [2002]. Diesel sales volume reflects economic activity and vulnerability to global fuel price shocksBernanke et al. [1997], while the Industrial Production Index (IPI) captures broader supply-side conditions. Food price volatility, driven by both domestic and global factors, has been shown to explain substantial variation in the Consumer Price Index Wickramasinghe [2019]. Despite their relevance, these indicators are seldom integrated into a unified forecasting framework for Sri Lanka.

Machine learning (ML) has revolutionized macroeconomic forecasting by capturing complex nonlinear patterns. Algorithms such as Random Forest

and XGBoost consistently outperform traditional models in environments with high-dimensional and interactive data Chen and Guestrin [2016]. In developed economies, ML is increasingly adopted by central banks and financial institutions for inflation and GDP forecasting Coulombe et al. [2022], Medeiros et al. [2021]. However, in Sri Lanka, the application of ML to inflation forecasting remains limited. Preliminary studies highlight its potential but often lack rigorous validation, model interpretability, and the use of high-frequency data, reducing their practical utility for policymakers.

The complexity of macroeconomic analysis often involves numerous correlated variables, leading to challenges in interpretation and modeling. Principal Component Analysis (PCA) is a widely used dimensionality reduction technique that creates composite indices summarizing the shared variance among multiple indicators Stock and Watson [2002]. Such indices, such as a Macroeconomic Condition Index (MCI), simplify the economic landscape into a single, interpretable measure. While PCA has been applied in Sri Lanka for financial and regional indices, no study has yet constructed a comprehensive MCI specifically designed for inflation forecasting—a gap that limits the ability to communicate economic conditions succinctly to stakeholders.

The existing literature reveals three critical gaps in Sri Lankan inflation forecasting research: (1) an overreliance on linear econometric models that fail to capture nonlinear dynamics, (2) the exclusion or underutilization of high-frequency, sector-specific indicators such as tourism arrivals and PMI, and (3) the absence of a composite index that synthesizes multiple economic signals into an interpretable measure for forecasting. This study addresses these gaps by integrating seven key monthly indicators, comparing multiple linear regression with advanced ML models (Random Forest and XGBoost), and constructing a novel PCA-based Macroeconomic Condition Index. The inclusion of SHAP analysis ensures model interpretability, bridging the gap between technical accuracy and policy relevance.

3 Methodology

3.1 Study Design

This research adopts a quantitative, comparative, and data driven research design that integrates both statistical and machine learning techniques. The purpose of the study is to model and forecast the Consumer Price Index (CCPI) in Sri Lanka using macroeconomic indicators and to develop a PCA based Macroeconomic Condition Index (MCI) that captures the combined impact of multiple variables.

The study consists of three major analytical phases:

- Development of a regression model to establish linear relationships.
- Development of machine learning models to capture complex non-linear interactions.
- Construction of a composite index (MCI) using Principal Component Analysis (PCA).

The findings of the regression and ML models are compared based on performance accuracy and interpretability.

3.2 Study Area

The study focuses on Sri Lanka, a developing South Asian nation with a diverse economic structure that has experienced significant inflationary fluctuations over the past decade.The analysis concentrates on the national level macroeconomic environment, with emphasis on the relationships between key macroeconomic indicators, specifically the Purchasing Managers' Index (PMI), Industrial Production Index (IPI), Broad Money Supply (M), Average Weighted Prime Lending Rate (AWPR), Exchange Rate, Diesel Sales Volume, and Tourism Arrivals, and the Consumer Price Index (CCPI). The CCPI, published monthly by the Department of Census and Statistics (DCS), serves as the primary measure of inflation in Sri Lanka. The study period covers January 2014 to December 2023, encompassing major policy shifts, global crises (COVID-19 pandemic, 2022 economic crisis), and post-crisis recovery trends, which together provide a comprehensive view of inflation dynamics.

3.3 Data Sources

All data used in this study will be obtained from reliable and publicly accessible official sources to ensure accuracy and authenticity. The main data sources include:

- **Central Bank of Sri Lanka (CBSL):** Monthly Economic Indicators (providing PMI, IIP, Broad Money Supply M, AWPR, Exchange Rate, and Diesel Sales Volume data).
- **Department of Census and Statistics (DCS):** Consumer Price Index (CCPI) data and inflation statistics
- **Sri Lanka Tourism Development Authority (SLTDA):** Monthly tourism arrivals data
- **World Bank and International Monetary Fund (IMF) databases:** Supplementary data for cross verification of macroeconomic indicators when required

All data represent monthly time series observations, recorded in consistent units and adjusted for base year comparability (base year 2021 = 100).

3.4 Data Collection

Data will be collected primarily through secondary sources. Monthly datasets will be downloaded directly from official Central Bank of Sri Lanka (CBSL) and Department of Census and Statistics (DCS) online publications in PDF or Excel format. The collection process involves the following steps:

1. **Data Extraction:** The seven key macroeconomic variables (PMI, IIP, M, AWPR, Exchange Rate, Diesel Sales Volume) will be systematically extracted from each monthly CBSL MEI report.
2. **Tourism Data Acquisition:** Monthly tourism arrivals data will be obtained from Sri Lanka Tourism Development Authority (SLTDA) reports.
3. **Data Integration:** All variables will be compiled and merged into a single structured dataset using Microsoft Excel.

4. **Data Cleaning:** The dataset will be cleaned and preprocessed by:
    - Checking for missing values
    - Identifying and addressing outliers
    - Verifying consistent time stamps across all series
5. **Data Transformation:** Variables will be transformed when necessary (e.g., logarithmic transformation or differencing) to ensure stationarity and meet modeling assumptions.

After preprocessing, the cleaned dataset will contain approximately 120 monthly observations covering all selected macroeconomic indicators for the period 2014–2023.

3.5 Proposed Methodology

This study adopts a quantitative comparative methodology integrating both statistical and machine learning approaches to model and forecast the CCPI in Sri Lanka. The methodological framework consists of six major phases, as outlined below:

3.5.1 Descriptive Data Analysis

Prior to model development, comprehensive descriptive analysis will be conducted to understand data characteristics and inform modeling decisions. This includes:

- Calculation of descriptive statistics (mean, median, standard deviation, ranges)
- Correlation analysis between CCPI and each of the seven independent variables (PMI, IIP, M, AWPR, Exchange Rate, Diesel Sales Volume, Tourism Arrivals).
- Time series visualization to identify trends and structural breaks
- Distribution analysis using histograms and box plots
- Stationarity testing using Augmented Dickey-Fuller test

3.5.2 Data Preprocessing

The preprocessing phase ensures data quality and prepares variables for analysis:

- **Standardization & Normalization:** Macroeconomic indicators will be standardized using Z-score normalization to ensure comparable scales across variables.
- **Missing Value Treatment:** Missing values will be handled using appropriate imputation techniques (mean/median imputation or time-series specific methods).
- **Multicollinearity Assessment:** Correlation analysis using Pearson correlation coefficients and Variance Inflation Factor (VIF) statistics will identify and address multicollinearity issues.

3.5.3 Principal Component Analysis (PCA)

PCA will be applied to the seven macroeconomic indicators (PMI, IIP, M, AWPR, Exchange Rate, Diesel Sales Volume, Tourism Arrivals) to derive a Macroeconomic Condition Index (MCI):

- PCA will be applied to macroeconomic indicators to derive a Macroeconomic Condition Index (MCI).
- The first principal component (PC1), explaining the maximum variance, will serve as the composite variable representing overall macroeconomic health.
- The MCI will be tested as both an alternative and complementary predictor in forecasting models.

3.5.4 Model Development

Three distinct modeling approaches will be implemented:

- **Statistical Model:** Multiple Linear Regression (MLR) will capture linear relationships between CCPI and predictors, serving as a baseline model.
- **Machine Learning Models:**

  - **Random Forest Regressor:** An ensemble method robust to non-linearity and outliers
  - **XGBoost Regressor:** A gradient boosting algorithm known for high predictive performance

3.5.5 Model Evaluation and Comparison

Model performance will be rigorously assessed using multiple metrics:

- **Root Mean Square Error (RMSE):** Measures average prediction error magnitude
- **Mean Absolute Error (MAE):** Provides robust error measurement less sensitive to outliers
- **Coefficient of Determination (R^2):** Quantifies proportion of variance explained by the model

3.5.6 Visualization and Interpretation

Results will be made accessible through comprehensive visualization:

- **Power BI Dashboard:** An interactive dashboard will visualize CCPI trends, feature importance, and model comparison results.
- **Model Interpretability:** Feature importance analysis and SHAP values will provide economic insights from machine learning models.

3.5.7 Validation

Robust validation techniques will ensure model reliability:

- **Time-series Split Validation:** Sequential training-test splits will maintain temporal ordering and prevent data leakage.
- **Overfitting Prevention:** Cross-validation and regularization techniques will be applied to ensure model generalizability.

3.6 Software Packages

The following software packages will be utilized in the research for data preprocessing, statistical analysis, machine learning model development, and visualization.

Table 1: Software Packages and Applications

Software/Package	**Specific Application**
Python (Pandas, NumPy)	Data manipulation, cleaning, preprocessing, and feature engineering
Python (Scikit-learn)	PCA implementation, ML models (Random Forest), evaluation metrics, cross-validation
Python (XGBoost)	XGBoost algorithm implementation and hyperparameter tuning
Python (Statsmodels)	Statistical tests (ADF, VIF), regression analysis, econometric modeling
Python (Matplotlib, Seaborn)	Data visualization, exploratory data analysis, result plotting
Python (SHAP)	Model interpretation, feature importance analysis, explainable AI
Jupyter Notebook	Interactive development environment for code execution and documentation
Microsoft Excel	Initial data compilation, organization, and basic data cleaning
Power BI	Interactive dashboard development for result visualization and reporting

4 Expected Outcomes and Deliverables

4.1 Academic Contributions

- Comprehensive descriptive statistics and correlation analysis of seven key Sri Lankan macroeconomic indicators (PMI, IIP, M, AWPR, Exchange Rate, Diesel Sales Volume, Tourism Arrivals) in relation to CCPI.
- Novel PCA-based Macroeconomic Condition Index for Sri Lanka
- Framework for interpretable machine learning in economic forecasting

4.2 Practical Applications

- Decision support system for inflation management
- Interactive dashboard for policy visualization
- Validated forecasting tools for financial institutions
- **Baseline descriptive analysis** for future economic research on Sri Lanka

4.3 Project Deliverables

- Comprehensive research thesis
- Reproducible Python code and analysis notebooks
- Operational Power BI dashboard
- Cleaned and documented dataset
- **Descriptive statistics report** summarizing key data characteristics

5 Time Schedule

5.1 Work Plan/ Gantt Chart

Table 2: Detailed Project Timeline (12-Month Schedule)

Activity	**Month 2**	**Month 4**	**Month 6**	**Month 8**	**Month 10**	**Month 12**
Literature Review & Proposal Finalization	**XXXXX**					
Data Collection & Preprocessing		**XXXXX**	**XXX**			
Exploratory Data Analysis & PCA			**XX**	**XXXXX**		
Model Development (MLR, RF, XGBoost)				**XX**	**XXXXX**	
Model Evaluation & SHAP Analysis					**XX**	**XXX**
Dashboard Development & Thesis Writing					**X**	**XXXXX**

References

Ben S Bernanke, Mark Gertler, and Mark Watson. Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity*, 1997(1):91–157, 1997.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5):920–964, 2022.

Irving Fisher. *The purchasing power of money*. Macmillan, 1911.

Pinelopi Koujianou Goldberg and Michael M Knetter. Goods prices and exchange rates: What have we learned? *Journal of Economic Literature*, 35(3):1243–1272, 1997.

V Jayawickrema and R Lokupitiya. Determinants of inflation in sri lanka: An ardl approach. *Sri Lankan Journal of Applied Statistics*, 19(2):45–62, 2018.

Evan F Koenig. Using the purchasing managers' index to assess the economy's strength and the likely direction of monetary policy. *Federal Reserve Bank of Dallas Economic and Financial Policy Review*, 1(6):1–14, 2002.

Marcelo C Medeiros, Gabriel FR Vasconcelos, Alvaro Veiga, and Eduardo Zilberman. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119, 2021.

Alban William Phillips. The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–1957. *Economica*, 25(100):283–299, 1958.

H P G S Ratnasiri. The main determinants of inflation in sri lanka: A var based approach. *Central Bank of Sri Lanka, Staff Studies*, 39(1&2):1–14, 2009.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.

D Wickramasinghe. Food price inflation and volatility in sri lanka: Trends and determinants. *Sri Lanka Economic Journal*, 20(1):23–45, 2019.

World Bank. World development report 2021: Data for better lives, 2021.