# Forecasting Unemployment Rate in Sri Lanka using Selected Macroeconomic Variables: A Comparative Study of Machine Learning/Deep Learning and Econometric Models*

Gunawardhana SRCL†

**Abstract**

This study investigates the impact of key macroeconomic variables on Sri Lanka's unemployment rate and compares various forecasting models, namely, econometric, Machine Learning (ML), and Deep Learning (DL) to improve unemployment rate prediction accuracy. Using quarterly data from 1998 to 2024, including GDP growth, inflation, interest rates, exports, exchange rates, and tourist earnings, the research evaluates relationships between these indicators and unemployment rate. Traditional econometric analysis, specifically the Vector Error Correction Model (VECM), is employed to capture long-term relationships, while ML models (Random Forest, Support Vector Regression, and Extreme Gradient Boosting) and DL models (Feedforward Neural Network) address non-linear and complex patterns in data. Forecast evaluation shows that Random Forest provides the highest accuracy, with a Mean Absolute Error (MAE) of 0.13, outperforming other models. The VECM, while effective in capturing long-term trends, has limitations in short-term forecasting due to linear assumptions.

This study underscores the potential of ML and DL models in economic forecasting, offering robust supportive techniques to improve traditional econometric methods. These insights can support policymakers in proactive labor market interventions. Accordingly, the findings from this study highlight the need for a hybrid approach that combines economic theory, and the adaptability of non-conventional modeling techniques blended with conventional modeling techniques for economic analysis.

**Key Words:** Forecasting, Machine Learning, Macroeconomic Variables, Sri Lanka, Unemployment Rate

**JEL Classification:** C32, C53, E24, E27, E37, E66

Conference Proceedings, 13th International Research Conference (December 2024), Central Bank of Sri Lanka, Colombo.

---

* The views expressed in this paper are the author's own and do not necessarily reflect those of the Central Bank of Sri Lanka.

† Senior Assistant Director, Statistics Department, Central Bank of Sri Lanka (chamili@cbsl.lk)

## 1. Introduction

The unemployment rate, a key economic indicator, measures the percentage of individuals within the labor force who are unemployed and actively seeking work. Its fluctuations influence economic output, social stability, and overall economic health. High unemployment not only reflects an underutilization of labor resources but can also lead to prolonged social and economic issues, while low unemployment may drive inflationary pressures due to economic overheating. In Sri Lanka, the Department of Census and Statistics (DCS) officially reports quarterly unemployment rates. However, the DCS's 12 to 14-week time lag in publishing these figures creates a critical gap for real-time decision-making. This delay underscores the need for effective forecasting methods that provide timely unemployment estimates, allowing stakeholders to proactively respond to changes and enabling more immediate economic assessments and economic analysis, such as GDP and inflation forecasting.

Research across various countries has demonstrated how macroeconomic factors including GDP growth, inflation, interest rates, and external events like pandemics and natural disasters impact unemployment rates. For instance, studies in Turkey, South Africa, and other nations detailed in Literature Review reveal complex interactions between these economic indicators and unemployment. Specific studies on Sri Lanka have shown notable correlations, particularly between unemployment and indicators like exports, inflation, and interest rates. Additionally, the rapid advancements in machine learning (ML) and deep learning (DL) are transforming forecasting methods, offering enhanced capabilities for capturing nonlinear relationships in economic data. Techniques such as Random Forests, Decision Trees, and neural networks have demonstrated significant potential to identify underlying patterns that traditional models may not capture. In global contexts, these methods have proven valuable in economic forecasting, showing promising results in unemployment rate prediction and policy analysis.

While traditional econometric models such as Vector Error Correction Models (VECM) and Auto Regressive Integrated Moving Average (ARIMA) have been widely applied in econometric forecasting, their reliance on linear assumptions may limit their effectiveness in capturing complex, real-world data dynamics. In Sri Lanka, few studies detailed in Literature Review have utilized advanced ML and DL techniques, leaving unexplored opportunities to improve prediction accuracy and robustness in unemployment forecasting. Thus, a critical gap exists in identifying and applying models that effectively balance short-term predictive accuracy with long-term insights. This study aims to address this gap by comparing conventional econometric models with advanced ML and DL techniques to identify the optimal approach for forecasting the unemployment rate in Sri Lanka.

Objectives of this study are sets out to:

1. Examine the relationships between unemployment rates and key macroeconomic variables in Sri Lanka.
2. Develop and compare forecasting models including econometric, ML, and DL approaches to improve the accuracy of unemployment rate forecasts.

   By pursuing these objectives, this research aims to contribute valuable insights for policymakers, demonstrating the practical utility of advanced forecasting techniques in enhancing both short-term responsiveness and long-term economic planning.

## 2. Literature Review

Forecasting unemployment is a multifaceted research area that spans traditional econometric models and more recent advancements in data science approaches. Globally, studies have employed models such as Vector AutoRegression (VAR) and Vector Error Correction Models (VECM) to explore the influence of GDP, inflation, interest rates, and exports on unemployment. In recent years, Machine Learning (ML) and Deep Learning (DL) models such as Gradient Boosting and neural networks have emerged to capture complex, nonlinear relationships in economic data, often achieving higher accuracy. Although research carried out in Sri Lanka has started exploring ML techniques, a gap remains in utilizing advanced models for local unemployment forecasting. This review aims to synthesize insights from existing research to evaluate effective forecasting methods for Sri Lanka.

### *2.1. Research Related to Unemployment Rate and Other Macroeconomic Variables in Other Countries*

International research highlights the intricate relationships between unemployment and macroeconomic variables such as GDP, inflation, exchange rates, and exports. Econometric models are commonly applied to study these interactions. For example, Dogan (2012) used a VAR model to analyze Turkey's unemployment rate, incorporating variables such as GDP growth, exports, inflation, and exchange rates. This study demonstrated that GDP growth and exports tend to reduce unemployment, while inflation and exchange rate fluctuations appear to increase it. These findings align with Okun's Law, which indicates an inverse relationship between economic growth and unemployment, and the Phillips Curve, suggesting a positive correlation between inflation and unemployment.

Similarly, Nyahokwe and Ncwadi (2013) investigated South Africa's unemployment dynamics using VAR, GARCH, and VEC models, highlighting exchange rate volatility as a significant predictor of unemployment, particularly in export-reliant sectors. Their findings underscore the impact of global economic shifts on domestic labor markets, especially in trade-dependent economies.

Other studies, such as Gaston and Rajaguru (2011) in Australia and Asif (2013) in a comparative analysis of Pakistan, India, and China, further demonstrate how macroeconomic factors influence unemployment across different contexts. Gaston and Rajaguru linked rising export prices to reduced unemployment in Australia, while Asif found that GDP growth decreased unemployment in India and China but increased it in Pakistan, highlighting how macroeconomic policies and contexts affect labor market outcomes.

*2.2. Research Related to Unemployment Rate and Other Macroeconomic Variables in Sri Lanka*

In Sri Lanka, studies identify inflation, interest rates, exports, and exchange rates as key determinants of unemployment. Typically, inflation and interest rates are associated with higher unemployment, whereas exports are linked to lower rates. Ariyadasa and Gunaratne (2014) applied the Phillips Curve to demonstrate that inflation raises unemployment by increasing production costs and reducing labor demand. Further, Jayathilaka and Mahendra (2016) confirmed these findings using a VECM, suggesting that inflation lowers purchasing power and suppresses demand, which negatively impacts employment.

Research by Fernando and Karunaratne (2018), employing an ARDL model, highlighted the role of interest rates, showing that high rates discourage job creation in capital-intensive sectors. Senanayake et al. (2019), using a Granger Causality test, found that export growth in sectors like textiles reduces unemployment by creating labor-intensive jobs. Supporting these studies, Perera and Wijesinghe (2020) applied Okun's Law and the Phillips Curve in Sri Lanka, demonstrating that GDP growth is inversely related to unemployment while inflation exerts upward pressure on it. These studies collectively provide an empirical foundation for understanding unemployment in the Sri Lankan economy.

### *2.3. Research on Data Science Approaches to Forecast Unemployment in Other Countries*

Data science advancements have expanded the toolkit for forecasting economic indicators, including unemployment rate. These methods excel at capturing nonlinear relationships, a common limitation in traditional models. For instance, Guo et al. (2018) applied a Gradient Boosting Machine (GBM) to U.S. labor market data (2005–2017), demonstrating superior performance over traditional models like ARIMA by effectively capturing seasonal patterns and complex variable interactions. Similarly, Anderson and Broadbent (2019) employed a Random Forest model on European labor data (2010–2018), focusing on GDP, inflation, and trade indices. Their findings showed that Random Forest was highly effective in managing multicollinearity and variable interactions, resulting in a significant reduction in forecast error compared to linear regression models. These studies underscore the potential of ML models to improve forecasting accuracy by accommodating data complexities.

### *2.4. Research on Data Science Approaches to Forecast Unemployment in Sri Lanka*

In Sri Lanka, the application of ML in unemployment rate forecasting is emerging, with studies focusing on models like Gradient Boosting, Decision Trees, and Support Vector Regression (SVR). Perera and Fernando (2022) applied GBM to manage multicollinearity between inflation and exchange rates, achieving higher forecast accuracy compared to linear models. Their findings underscore GBM’s ability to handle prediction errors iteratively, enhancing model precision.

Similarly, Ranasinghe and Wijeratne (2020) used Decision Trees, which improved forecast accuracy by 15% over traditional methods by effectively capturing interactions among GDP growth, exports, and inflation. Jayaratne and Senanayake (2019) employed SVR to forecast unemployment during volatile periods, finding that it outperformed ARIMA models in Mean Absolute Error (MAE) and Mean Squared Error (MSE), demonstrating adaptability to noisy data.

### *2.5. Research on Data Science Approaches to Forecast Macroeconomic Variables*

Beyond unemployment, ML approaches have been applied in forecasting macroeconomic variables such as GDP and inflation. Gonzalez (2000) demonstrated that neural networks outperform traditional models in GDP forecasting by capturing linear and nonlinear relationships. Fischer and Krauss (2018) used an LSTM model on daily financial data (1992–2015) to predict inflation trends, highlighting the model's strength in sequential data dependencies. Nelson et al. (2017) applied a Recurrent Neural Network (RNN) model to forecast GDP and interest rates, emphasizing RNNs' capability in time-series forecasting. Additionally, studies by Adebiyi et al. (2014), Chen and Guestrin (2016), and Smola and Schölkopf (2004) further illustrate the success of ML models like neural networks, XGBoost, and SVR in predicting volatile, seasonally influenced economic data.

### *2.6. Research Gap*

Despite substantial research linking unemployment and macroeconomic variables, a significant gap exists in the application of advanced ML and DL techniques for unemployment forecasting in Sri Lanka. While traditional models like VECM and ARIMA are commonly used, their linear assumptions limit their capacity to capture complex patterns. Moreover, while global studies highlight the effectiveness of ML techniques and DL techniques in forecasting, these methods remain underutilized in Sri Lanka. Addressing this gap could improve the timeliness and accuracy of unemployment rate forecasting, especially considering the delayed availability of Sri Lankan unemployment data. This study aims to bridge this gap by evaluating ML and DL models alongside traditional econometric approaches to identify the most reliable forecasting method.

### *2.7. Summary*

This literature review highlights the complex relationships between macroeconomic variables and unemployment rate. While traditional econometric models provide foundational insights, ML and DL approaches hold potential for enhancing forecast precision, especially in dynamic economic environments. Addressing identified research gaps will contribute to Sri Lanka's economic forecasting capabilities, supporting more responsive and effective policymaking.

## 3. Data and Methodology

### *3.1. Data Collection and Preparation*

#### 3.1.1. Variables, their definitions and sources

The variables used in this study are listed in the below table with their definitions and sources. The Data Library maintained by the Central Bank of Sri Lanka is the main secondary data source for the study. In addition, the Annual Reports and other publications of the Central Bank of Sri Lanka and the Quarterly Labour Force Survey Reports of the Department of Census and Statistics are also used as secondary data sources to collect data and to obtain background information of the data.

**Table 1. Details of Variables**

| No. | Variable | Definition | Source |
|---|---|---|---|
| 1. | Unemployment rate | The proportion of unemployed population to the total labour force above 15 years of age. | Department of Census and Statistics |
| 2. | GDP growth | Economic growth is the increment in the national income and output of an economy. | Department of Census and Statistics |
| 3. | Inflation | The percentage change in consumer price index which is compiled based on the expenditure for a basket of consumer goods. | Department of Census and Statistics |
| 4. | Interest rate | Interest rate is referred to as the amount charged, expressed as a percentage of principal, by a lender to a borrower for the use of assets on annual basis. | Central Bank of Sri Lanka |
| 5. | Exports | The goods and services produced in one country and traded to other countries. | Central Bank of Sri Lanka |
| 6. | Exchange rate | The units of domestic currency per one unit of the foreign currency. | Central Bank of Sri Lanka |
| 7. | Tourist Earnings | The earnings received to the country through from services related to tourism | Central Bank of Sri Lanka |
| 8. | Terrorist Attacks | A dummy variable to identify periods with severe terrorist attacks in the country. | Publicly available information |
| 9. | COVID 19 | A dummy variable to identify periods with comparatively high impact from COVID 19 in the country. | Publicly available information |

### 3.1.2. Data Transformation

The macroeconomic variables that are not available on a quarterly basis but available in other frequencies such as daily or monthly, are converted into quarterly series considering suitable conversion techniques to each of those variables. Further actions with regard to data preparation will be carried out in the analysis. Further details of the variables used are given below.

**Unemployment Rate**

Unemployment rate is a crucial factor of the economy which indicates the conditions of labour market. Unemployment rate is the percentage of the total labour force that is unemployed but actively seeking employment and willing to work. Higher unemployment rates generally reflect worse economic conditions. As mentioned in the Labour Force Survey, Annual Report, published by the DCS, the unemployment rate is calculated by the DCS on quarterly basis after conducting the quarterly Labour Force Survey since the 1$^{st}$ quarter of 1990. Persons available and/or looking for work, and who did not work and taken steps to find a job during last four weeks and ready to accept a job given a work opportunity within next two weeks are said to be unemployed as defined by the DCS in the above publication. Accordingly, in Sri Lanka, the unemployment rate is the proportion of unemployed population to the total labour force above 15 years of age.

Unemployment rate data was extracted by the Data Library of the Central Bank of Sri Lanka and few data gaps were filled by referring to the quarterly Labour Force Survey reports of DCS and by imputing the missing values using the averages of adjoining periods. The Labour Force Survey was not conducted for the 2$^{nd}$ quarter of 2001 by the DCS and thus the unemployment rate for that period was imputed by getting the average of unemployment rates of 1$^{st}$ quarter and 3$^{rd}$ quarter of 2001. Further, since the DCS was involved in post Tsunami related analysis, the quarterly Labour Force Survey was not conducted during the year 2005 and one-off survey was conducted was conducted in August 2005. That value was used for all the four quarters of the year. The quarterly Labour Force Survey was not conducted during the 4$^{th}$ quarter of 2011 and the 1$^{st}$ quarter of 2012 as the DCS was involved in the Census conducted during 2012. Thus, the unemployment rate values for those two quarters were imputed by getting the average of unemployment rates of the other three quarters of the relevant year.

**GDP Growth**

Economic growth is the growth in gross domestic product (GDP) which is the main measure of the nation's economic activities. The GDP is the total value of goods and services produced in a country in a stipulated period of time according to the System of National Accounts 2008 publication which is used globally as the handbook in compiling GDP. Therefore, economic growth is the increment in the national income and output of an economy. A positive economic growth always indicates good economic conditions and potential development of the economy. The GDP is calculated in both real and nominal terms by the Department of Census and Statistics on a quarterly basis.

Currently, the real GDP is calculated considering 2015 as the base year according to the National Accounts Estimates of Sri Lanka - News Release of DCS. The real GDP compiled before the year 2015, of which the base years were 1996, 2002 and 2010 were converted to 2015 base year considering the growth values. In this study the GDP growth is calculated in quarterly basis considering the quarter on quarter growth of real GDP.

$$\text{GDP growth rate} = \frac{(\text{real GDP}_{t} - \text{real GDP}_{t-4}) \times 100}{\text{real GDP}_{t-4}} \qquad (1)$$

real $\text{GDP}_t$ : real GDP in the $t^{th}$ quarter considered in the study

real $\text{GDP}_{t-4}$ : real GDP one year ago to the $t^{th}$ quarter considered in the study

**Inflation Rate**

Inflation rate is also a primary concern of an economy which indicates the acceleration of price levels. As defined in the Consumer Price Index Manual publication which is referred globally in compiling consumer price indices and calculating inflation, briefly, inflation rate is the percentage change in consumer price index which is compiled based on the expenditure for a basket of consumer goods. When the inflation rate is high, the real value of money erodes. In times of high inflation, if the income of people would not increase, the economy would struggle.

The inflation is calculated in Sri Lanka, based on the Colombo Consumers' Price Index (CCPI) and the National Consumers' Price Index (NCPI) by the DCS on a monthly basis. In this study, quarterly figures for CCPI were calculated by taking the average of monthly values of the relevant quarter and the inflation rate is calculated considering the quarter-on-quarter growth of CCPI which is compiled considering the base year as 2021. As per the annual reports of CBSL, the base years of CCPI are 1952, 2002, 2006,2013 and 2021. The CCPI compiled based on previous base years were converted to 2013 base year considering the growth values.

$$\text{Inflation rate} = \frac{(\text{CCPI}_t - \text{CCPI}_{t-4}) \times 100}{\text{CCPI}_{t-4}} \qquad (2)$$

$\text{CCPI}_t$ : CCPI in the $t^{th}$ quarter considered in the study

$\text{CCPI}_{t-4}$ : CCPI one year ago to the $t^{th}$ quarter considered in the study

**Interest Rate**

Interest rate is generally referred to as the amount charged, expressed as a percentage of principal, by a lender to a borrower for the use of assets on annual basis. It is a measure of return and an indicator of risk as well. Higher interest rates discourage investments and lead to bad economic conditions. Interest rate is generally quoted in nominal terms and the inflation adjusted interest rate is referred to as real interest rate which compensates for the time value reduction of money. The Central Bank of Sri Lanka determines the policy interest rate/s in order to control the price and availability of money in the money market.

In this study, 91 days Treasury Bill rate determined by the Central Bank of Sri Lanka on daily basis is considered and the quarterly values were derived by calculating the average of daily values.

**Exports**

Exports are referred to as the goods and services produced in one country and traded to other countries. As the production and sale of such goods and services are directly connected with the economic activities of a country and with labour force activities, exports are a crucial component of an economy. Growth in exports would indicate the improvements in the performance of international trade. In this study, quarterly export values were gathered in USD billion terms.

**Exchange Rate**

Exchange rate is the price of a nation's currency in terms of another currency. It can be used as a measure of appreciation or depreciation of currency with compared to other currencies. The volatility in exchange rates would generate bad economic conditions since a depreciation of currency causes higher cash outflows and an appreciation would cause lower cash inflows. The exchange rate can be determined in direct quote or indirect quote.

As mentioned in the Pamphlet Series No.03 on Exchange Rate published by the CBSL in 2006, in Sri Lanka the exchange rate is calculated in direct quote which indicates the units of domestic currency per one unit of the foreign currency. A positive growth in exchange rate indicates depreciation in Sri Lankan rupee and a negative growth in exchange rate indicates appreciation in Sri Lankan rupee. Sri Lankan Rupee (LKR) per USD, monthly average exchange rate is converted to quarterly terms by getting the average of relevant months to be used in this study.

**Tourism Earnings**

Tourism earnings can be defined as expenditure of international inbound visitors including their payments to national carriers for international transport, and any other payments or payments afterwards made for the services received in the destination country. This variable includes the earnings received from tourism-related services in USD millions. The quarterly data series for tourism earnings was available only from 2001. Therefore, quarterly values for 1998 to 2000 were inputted by equally dividing the annual series. The data were extracted from the Balance of Payments data available at the CBSL.

**Terrorist Attacks**

Sri Lankan economy was affected by terrorist attacks time to time over the period considered in this study. Therefore, a dummy variable is added to represent the periods with severe terrorist attacks since they could be attributable to the economic activities of the country and hence on unemployment.

**COVID 19**

The COVID 19 pandemic prevailed since 2020 affected highly on unemployment. The lock downs, employee reductions and close downs of businesses resulted in job losses. Therefore, a variable to indicate periods with comparatively high impact from COVID 19 in the country was used as a dummy variable in the analysis.

### *3.2. Methodology for Time Series Analysis*

Initially, the relationships among variables are evaluated using Pearson's Correlation Coefficient, followed by the Kruskal-Wallis Test to assess seasonality. The stationary conditions of the variables are then checked using the Augmented Dickey-Fuller (ADF) Test. To examine interconnectedness, the Johansen Test for cointegration and the Granger Causality Test are conducted. A Vector Error Correction Model (VECM) is estimated for unemployment rate forecasting based on other key macroeconomic variables. Model adequacy is subsequently tested through residual diagnostics: the Shapiro-Wilk Test for normality, the Ljung-Box Test for autocorrelation, and the Breusch-Pagan Test for heteroskedasticity.

#### 3.2.1. Vector Error Correction Model (VECM)

The Vector Error Correction Model (VECM) is a multivariate time series model designed to capture both short-term dynamics and long-term equilibrium relationships between non-stationary, cointegrated variables. In VECM, the differenced data (converted to stationary form) is modeled alongside error correction terms, which represent deviations from long-term equilibrium. This allows the model to adjust for discrepancies and gradually revert to the equilibrium state over time. The VECM structure is particularly useful for analyzing interconnected macroeconomic variables like GDP, inflation, and unemployment, as it addresses both immediate shocks and stable, long-term relationships within the dataset.

### *3.3. Methodology for Deep Learning and Machine Learning Analysis*

The study employs a Feedforward Neural Network (FFNN) as the Deep Learning model, using layered neurons to uncover complex patterns across multiple variables. The Machine Learning models include Random Forest (RF), an ensemble method that constructs multiple decision trees to capture non-linear interactions in complex datasets; Support Vector Regression (SVR), which effectively manages noise by optimally separating data points; and Extreme Gradient Boosting (XGBoost), a powerful gradient-boosting technique that iteratively enhances forecasting accuracy by minimizing prediction errors.

### 3.3.1. Feedforward Neural Network Model

The Feedforward Neural Network (FFNN) model is a supervised deep learning model with interconnected layers of neurons that map complex relationships between inputs and outputs. It consists of an input layer, one or more hidden layers, and an output layer, where each layer is connected through weighted neurons. The model minimizes prediction errors by adjusting neuron weights through backpropagation and optimization algorithms. FFNNs are effective in capturing non-linear relationships, making them suitable for time series forecasting where interactions among macroeconomic factors are often complex and multi-layered.

### 3.3.2. Random Forest Model

Random Forest is an ensemble learning method based on constructing multiple decision trees for prediction. In this model, each tree is trained on a random subset of data and variables, allowing it to capture diverse patterns within the dataset. Predictions from individual trees are averaged to produce a final output, enhancing model stability and reducing overfitting. Random Forest's ability to handle non-linear data and rank feature importance makes it highly effective in economic forecasting, where multiple, interrelated indicators influence the outcome.

### 3.3.3. Support Vector Regression Model

Support Vector Regression (SVR) is a regression technique that aims to find an optimal hyperplane that maximizes the margin within which most data points fall. Unlike traditional regression, SVR introduces a margin of tolerance for error, known as the epsilon-insensitive zone, where points within the margin do not contribute to error calculations. This approach makes SVR robust to noise and effective in capturing non-linear patterns when combined with kernel functions. SVR is particularly useful in economic forecasting due to its ability to handle outliers and maintain predictive accuracy.

### 3.3.4. Extreme Gradient Boosting Model

Extreme Gradient Boosting (XGBoost) is a powerful ensemble method based on gradient-boosting principles. In XGBoost, multiple decision trees are sequentially constructed, each aiming to correct the errors of the previous tree. This iterative improvement process allows XGBoost to reduce prediction errors and capture complex patterns in data. With regularization to prevent overfitting, XGBoost effectively balances model complexity and performance. Its application in economic forecasting leverages its ability to handle both linear and non-linear relationships, providing accurate and resilient predictions.

### *3.4. Methodology for Forecast Evaluation*

Forecast accuracy in this study is assessed using four performance metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE). These metrics provide insights into the magnitude, percentage, and distribution of forecast errors, ensuring a comprehensive evaluation of model effectiveness.

#### 3.4.1. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) calculates the average absolute differences between actual and predicted values, providing an easily interpretable measure of forecast accuracy. MAE is a non-negative metric, where lower values indicate better model performance, as it reflects the average magnitude of prediction errors.

#### 3.4.2. Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) expresses the average absolute error as a percentage of actual values, allowing for intuitive comparisons across models. MAPE is particularly useful when comparing models on datasets with varying scales, as it contextualizes the error relative to the observed values.

#### 3.4.3. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) measures the square root of the average squared differences between actual and predicted values. By penalizing larger errors, RMSE emphasizes the model's ability to accurately forecast extreme values or outliers, making it suitable for assessing overall predictive accuracy.

#### 3.4.4. Mean Squared Error (MSE)

Mean Squared Error (MSE) calculates the average of squared differences between actual and predicted values. MSE penalizes large errors more severely than smaller ones, providing a sensitive indicator of forecast accuracy. Lower MSE values reflect better model performance in terms of consistency and reliability.

## 4. Empirical Analysis and Discussion

This chapter presents the empirical findings from the analysis of selected time series variables, focusing on their temporal variations, model estimation, and forecasting performance. Initially, the temporal dynamics of each variable are considered, setting the stage for developing and evaluating several forecasting models. Accordingly, a time series econometric model, a deep learning model and three machine learning models are estimated to forecast unemployment rate of Sri Lanka. Further, a comparative analysis of forecast performance is carried out to identify the most effective model for accurate unemployment rate forecasting.

### *4.1. Data Preparation*

This study uses seven key macro-economic variables, namely, unemployment rate, GDP growth, inflation rate, interest rate, exports, exchange rate and tourist earnings. In addition, two dummy variables, namely, terrorist attacks and COVID19 are also considered. The quarterly data from the year 1998 to the second quarter of 2024 are collected and transformed into quarterly time series as appropriate.

Accordingly, the unemployment rate data collected from the quarterly labour force survey conducted by the Department of Census and Statistics are used for the analysis. The real GDP compiled considering base periods of 1996, 2002,2010 and 2015 are collected, and the 1996,2002 and 2010 base series are converted to combine with 2015 series considering the growth values. The analysis uses quarterly growth values of GDP, calculated based on the quarter-on-quarter percentage increase of the real GDP series. The third variable used in the study, “inflation rate” is calculated based on the monthly CCPI series. The CCPI is also compiled considering different base years, 1952, 2002, 2006,2013 and 2021, and thus, converted into 2021 series using growth values. Using this series, quarterly figures for CCPI are calculated by taking the average of monthly values of the relevant quarter, and the inflation rate is calculated considering the quarter-on-quarter growth of CCPI quarterly series. The variable, “interest rate” is represented in this study by the 91 days Treasury Bill rate determined by the Central Bank of Sri Lanka on daily basis. A quarterly series of interest rates is derived by averaging the daily values relevant to the quarter. In terms of the variable, “exports”, the quarterly export values in USD billion terms published by the Central Bank of Sri Lanka are used for the analysis. The sixth variable, “exchange rate” is calculated based on the monthly average exchange rate values (LKR per USD) available at the Central Bank of Sri Lanka. The quarterly values are obtained by averaging these monthly values. In terms of the variable, “tourist earnings”, the monthly values in USD million terms, collected based on the survey conducted by the Sri Lanka Tourism Development Authority and published by the Central Bank of Sri Lanka are used for the analysis after converting into a quarterly series by aggregating the relevant three months pertaining to the quarter.

*4.2. Temporal Variation of Time Series Variables Selected for the Study and their Association*

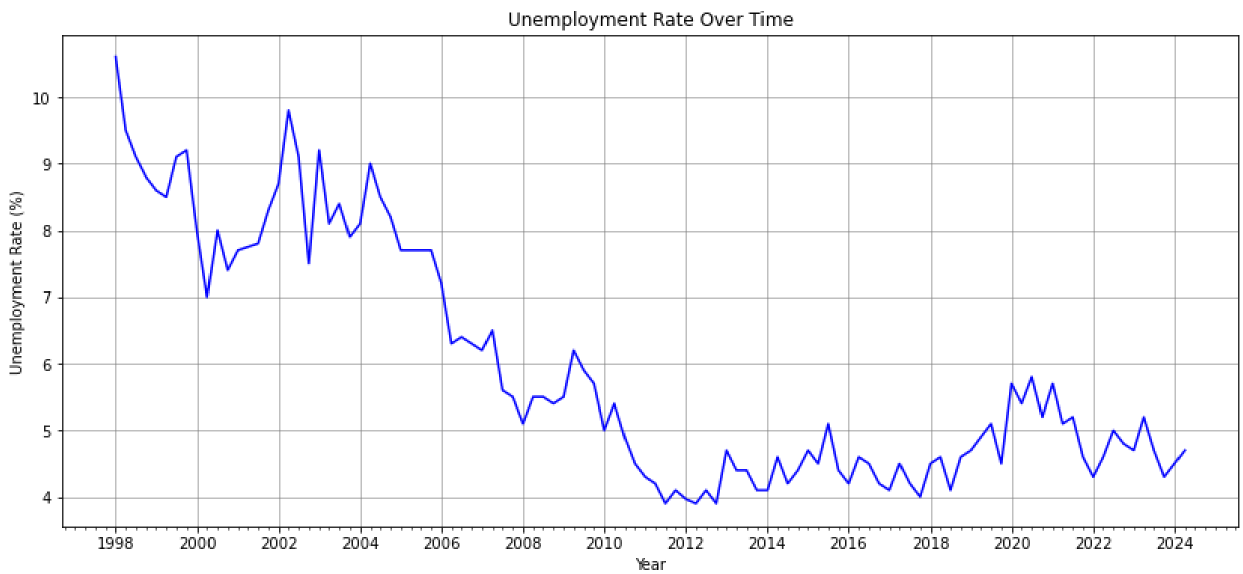


**Figure 1. Unemployment Rate from 1998 to 2024**

Figure 1 illustrates the fluctuations of unemployment rate of Sri Lanka from 1998 to the second quarter of 2024 reflecting broader economic conditions impacting labor markets. The unemployment rate shows a declining trend until the latter part of 2012 and remained relatively stable with minor fluctuations until mid-2019. Thereafter, it increased predominantly due to the effects of Easter Attack followed by COVID19 outbreak. However, from 2022 onwards it fluctuates with relatively low volatility. Analyzing these trends offers valuable insights into the persistence of unemployment in Sri Lanka over time, supporting in the estimation of appropriate models for forecasting.

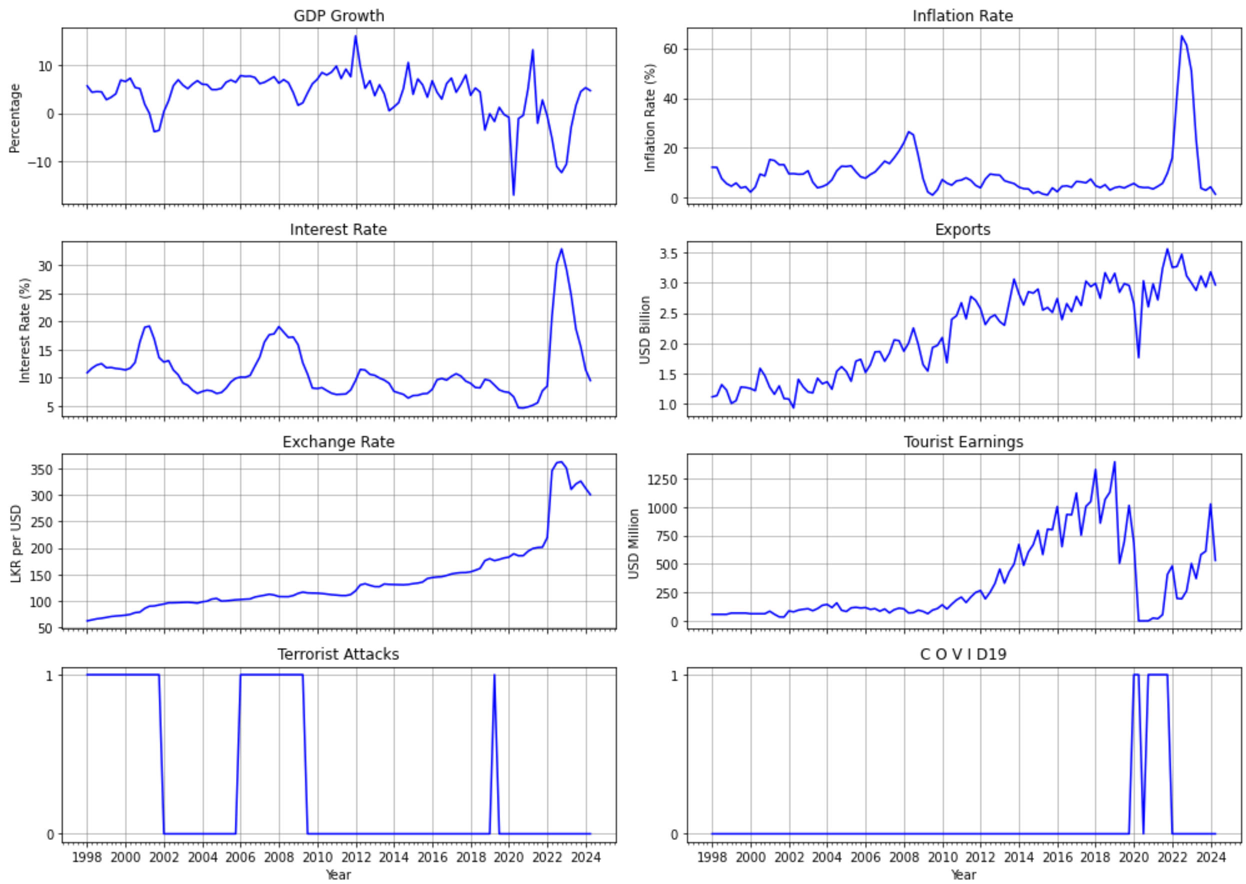


**Figure 2. Time Series Variables Selected for the Study**

Figure 2 presents the temporal patterns of the time series variables selected for this study, including the two dummy variables, namely, terrorist attacks and COVID19. Each variable is plotted over time, providing a comparative view of trends, seasonal effects, and potential structural breaks. These visualizations help highlight how factors such as GDP growth, inflation, interest rate, exports, exchange rate and tourist earnings interact with the unemployment rate.

4.2.1. Relationship of Unemployment Rate with Time Series Variables Selected for the Study

Pearson's correlation coefficient which quantifies the strength and direction of the linear relationship between two continuous variables are used to understand how strongly variables are related, guiding model selection and interpretation.

**Table 2. Pearson's Correlation Coefficient**

| Variable | GDP Growth | Inflation Rate | Interest Rate | Exports | Exchange Rate | Tourist Earnings |
|---|---|---|---|---|---|---|
| Unemployment Rate | 0.032 | -0.007 | 0.056 | -0.868 | -0.519 | -0.585 |

The analysis reveals that unemployment has a strong inverse relationship with exports (-0.868) and tourist earnings (-0.585), indicating that growth in these sectors significantly contributes to employment creation. Additionally, there is a moderate negative correlation between unemployment and the exchange rate (-0.519), suggesting that higher unemployment may coincide with an appreciating local currency. However, correlations with GDP growth (0.032), inflation rate (-0.007), and interest rate (0.056) are weak, implying that traditional macroeconomic relationships, like those proposed by the Phillips Curve and Okun's law, are not strongly evident in the Sri Lankan context under the considered time span. However, all these variables are used in the study considering their interconnectedness in the economy.

4.2.2. Test for Seasonality of Time Series Variables

Seasonality captures recurring patterns or periodic fluctuations within time series data, influenced by economic cycles or seasonal factors. Identifying seasonality is essential for improving forecasting models. The Kruskal-Wallis test is a statistical method superior to visual inspection alone, providing a quantifiable measure of seasonality. This test assesses whether observed seasonal differences in variables are statistically significant, which helps ensure more reliable model adjustments and forecasts.

**Table 3. Kruskal-Wallis Test Results Summary**

| Variable | Kruskal-Wallis Chi-squared | Degrees of Freedom (df) | p-value |
|---|---|---|---|
| UnemploymentRate | 0.75883 | 3 | 0.8593 |
| GDPGrowth | 0.54566 | 3 | 0.9088 |
| InflationRate | 0.11831 | 3 | 0.9896 |
| InterestRate | 0.48288 | 3 | 0.9226 |

| | | | |
|---|---|---|---|
| Exports | 3.2301 | 3 | 0.3575 |
| ExchangeRate | 0.015539 | 3 | 0.9995 |
| TouristEarnings | 1.988 | 3 | 0.5749 |

In all tested variables, the p-values are significantly higher than 0.05, indicating that there are no significant differences in the distributions of these macroeconomic variables across different quarters. Thus, it can be concluded that seasonal effects do not significantly influence these variables over the period under consideration. Therefore, seasonal decomposition is not necessary in conducting further time series analysis.

#### 4.2.3. Test for Stationary of Variables

A stationary time series has a constant mean, variance, and autocorrelation over time, making it more predictable and suitable for forecasting. Stationarity is a critical property in time series analysis, as non-stationary data can lead to unreliable and spurious results in econometric modeling. In this study, the Augmented Dickey-Fuller (ADF) test was employed to assess stationarity. Establishing stationarity is essential before modeling by obtaining differenced series, particularly for techniques like Vector Error Correction Model (VECM), which require non-stationary series that are cointegrated for meaningful long-term relationship analysis.

**Table 4. Augmented Dickey-Fuller (ADF) Test Results Summary**

| Variable | Dickey-Fuller Statistic | Lag Order | p-value | Stationarity Status |
|---|---|---|---|---|
| Diff_UnemploymentRate | -4.2182 | 4 | 0.01 | Stationary (p-value < 0.05) |
| Diff_GDPGrowth | -6.3401 | 4 | 0.01 | Stationary (p-value < 0.05) |
| Diff_InflationRate | -4.625 | 4 | 0.01 | Stationary (p-value < 0.05) |
| Diff_InterestRate | -3.4641 | 4 | 0.04883 | Stationary (p-value < 0.05) |
| Exports | -3.7854 | 4 | 0.02239 | Stationary (p-value < 0.05) |
| Diff2_ExchangeRate | -5.9357 | 4 | 0.01 | Stationary (p-value < 0.05) |
| Diff2_TouristEarnings | -7.0732 | 4 | 0.01 | Stationary (p-value < 0.05) |

The Augmented Dickey-Fuller (ADF) test confirms that all variables in the study have achieved stationarity after appropriate differencing, essential for effective time series modeling. Only exports is stationary at its original level, while unemployment rate, GDP growth, inflation rate, and interest rate are stationary after first differencing. Additionally, Exchange Rate and Tourist Earnings reached stationarity after second differencing, indicated by p-values below 0.05 across all differenced forms. This ensures stable fluctuations in the data, supporting robust forecasting using models like the VECM.

### 4.2.4. Cointegration of Unemployment Rate with other variables selected in the study

Cointegration is an essential measure in time series analysis, particularly when examining multiple variables that are individually non-stationary but may exhibit a stable, long-term relationship. Cointegration occurs when a linear combination of these non-stationary variables is stationary, indicating that despite individual trends, the variables move together in the long run. Detecting cointegration is crucial as it allows for the application of models like the VECM, which accounts for both short-term dynamics and long-term equilibrium among variables. This approach provides valuable insights into the interconnectedness of economic indicators, financial metrics, or other time-dependent variables, helping to capture both transient and persistent effects in complex systems. The Johansen test which is a statistical method used to identify the number of cointegrating relationships in a system of variables is used as follows.

**Johansen Test Results**

**Table 5. Number of Cointegrating Relationships**

| Hypothesis (r) | Test Statistic | 10% Critical Value | 5% Critical Value | 1% Critical Value | Conclusion |
|---|---|---|---|---|---|
| r = 0 | 187.39 | 118.99 | 124.25 | 136.06 | Cointegration exists |
| r ≤ 1 | 108.84 | 85.18 | 90.39 | 104.2 | At least 2 cointegrating relationships |
| r ≤ 2 | 61.22 | 66.49 | 70.6 | 78.87 | No further cointegrating relationships |

As per the results, there are two cointegrating relationships among the variables, suggesting that certain variables, including the unemployment rate, are bound together in a long-term equilibrium.

**Table 6. Eigenvalues (Strength of Cointegration)**

| Rank (r) | Eigen Value |
|---|---|
| 1 | 0.5441 |
| 2 | 0.3788 |
| 3 | 0.2535 |

The Johansen test's eigenvalues help gauge the strength of cointegrating relationships. Here, the highest eigenvalue (0.5441) suggests a strong primary cointegrating relationship, indicating a stable long-term equilibrium among the variables. The subsequent eigenvalues (0.3788 and 0.2535) reflect progressively weaker relationships, showing that while additional equilibrium connections exist, they have less impact on the system's overall stability.

**Table 7. Normalized Eigenvectors for Cointegration Relations**

| Variable | UnemploymentRate | GDPGrowth | Inflation Rate | Interest Rate | Exports | Exchange Rate | TouristEarnings |
|---|---|---|---|---|---|---|---|
| Coefficient | 1 | 0.973 | -0.825 | 0.538 | 0.927 | 0.062 | -0.005 |

The normalized eigenvectors for cointegration relations reveal key long-term associations between unemployment rate and other economic variables. A positive relationship with GDP growth suggests that higher economic growth is associated with higher unemployment. The inverse relationship with inflation aligns with the Phillips curve, indicating that inflation may help reduce unemployment under certain conditions. The positive link with interest rates implies that higher borrowing costs could hinder employment creation. Exports and exchange rate have minor positive relationships, suggesting limited direct impact on unemployment, while tourist earnings show a slight inverse relationship, implying at marginal employment creation within tourism-related sectors.

**Table 8. Adjustment to Equilibrium (Loading Matrix)**

| Variable | UnemploymentRate | GDPGrowth | InflationRate | InterestRate | Exports | ExchangeRate | TouristEarnings |
|---|---|---|---|---|---|---|---|
| Adjustment Coefficient | -0.014 | -1.352 | 0.379 | 0.076 | 0.007 | 0.402 | 0.824 |

The adjustment coefficients reveal that while unemployment rate adjusts slowly to restore equilibrium, GDP growth and tourist earnings respond more rapidly, underscoring their active roles in economic stabilization. This indicates that GDP growth, inflation, and interest rates are primary drivers in the long-term equilibrium with unemployment rate. Policy measures targeting these areas may significantly impact unemployment, though the slow adjustment of unemployment rate itself suggests that structural changes are required for quick shifts. This interconnected framework emphasizes the importance of a balanced policy approach to managing unemployment rate within broader economic conditions.

According to the above Johansen test results, cointegration exists among the variables. Specifically, there are two significant cointegrating relationships, indicating a long-term equilibrium between the unemployment rate and other economic indicators. This suggests that these variables do not drift apart indefinitely but are bound together by underlying economic forces. At this background using VECM to forecast the unemployment rate is well-suited due to its ability to model long-term equilibrium relationships among variables that are cointegrated. The VECM framework allows for incorporating both the short-term dynamics and the long-term relationships between unemployment and key economic indicators considered.

### *4.3. Estimation of Vector Error Correction Model*

The Vector Error Correction Model (VECM) is specifically designed for time series that are non-stationary in levels but exhibit stationarity after differencing, and that are cointegrated, meaning they share a long-term equilibrium relationship. Further, the VECM framework is well-suited to understanding both the transient responses to economic shocks and the persistent relationships among the variables, making it valuable for policy analysis and forecasting in an interconnected economic system. Therefore, a VECM model is estimated to forecast unemployment rate, and it is crucial to decide the parameters, number of cointegrating relationships (r) and lag length in fitting the VECM model.

### 4.3.1. Select number of cointegrating relationships for VECM

The Johansen test results confirm at least two cointegrating relationships, as the trace statistic for $r = 0$ (187.39) exceeds critical values, rejecting the null hypothesis of no cointegration. For $r \leq 2$, the statistic (61.22) falls below the critical values, establishing two cointegrating relationships. Thus, $r = 2$ is chosen for the VECM, capturing two stable, long-term relationships within the system.

### 4.3.2. Select Optimal Lag Length

**Table 9. Lag Length**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Optimal Lag Length | | | | | | | | | | | | |
| AIC (n) | HQ (n) | SC (n) | FPE (n) | | | | | | | | | |
| 10 | 10 | 10 | 10 | | | | | | | | | |
| Lag Length Selection Criteria | | | | | | | | | | | | |
| Lag length | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| AIC(n) | 10.9920 | 10.2294 | 7.7499 | 5.4564 | 4.0581 | 2.2726 | -0.6199 | -3.3000 | -118.3050 | -Inf | -Inf | -Inf |
| HQ(n) | 12.0127 | 12.1688 | 10.6080 | 9.2331 | 8.7535 | 7.8866 | 5.9128 | 4.1514 | -109.9350 | -Inf | -Inf | -Inf |
| SC(n) | 13.5256 | 15.0433 | 14.8441 | 14.8309 | 15.7128 | 16.2076 | 15.5954 | 15.1955 | -97.5292 | -Inf | -Inf | -Inf |
| FPE(n) | 59925.9321 | 29481.8700 | 2852.2466 | 386.3011 | 162.4784 | 68.0062 | 18.2036 | 22.8356 | 0.0000 | 0 | 0 | 0 |

In time series analysis, model selection criteria such as Akaike Information Criterion (AIC), Hannan-Quinn Criterion (HQ), Schwarz Criterion (SC/BIC), Final Prediction Error (FPE) are essential tools for determining the optimal lag length. When selecting the optimal lag length, aiming for a balance between model fit and simplicity can be crucial, especially when the model needs to be interpretable and robust. In this case, while the criteria indicate an optimal lag length around 10, experimenting with slightly shorter lags (around 8 to 10) can help assess if the model maintains adequate fit with fewer parameters. Due to the limited data available for the study, lag length 8 is selected for the study.

#### 4.3.3. Model Estimation

Using two cointegrating relationships and eight lags, a VECM model is fitted focusing on forecasting unemployment rate. The stationary variables, namely, first difference of unemployment rate, first difference of GDP growth, first difference of inflation rate, first difference of interest rate, exports, second difference of exchange rate, second difference of tourist earnings and two dummy variables indicating the presence of terrorist attacks and COVID19 during the quarter were used for the model estimation. A sample of 100 data points were used for model fitting and another 4 data points were kept aside for forecast evaluation.

According to the long-term relationship identified through cointegrating vectors, unemployment is inversely related to inflation, suggesting that higher inflation tend to reduce unemployment rate in the long term. The positive association with interest rates suggests that high borrowing costs might discourage employment. COVID-19, with a high positive coefficient, indicates that the pandemic has had a lasting, negative impact on employment. Further, GDP growth shows a strong relationship with inflation, interest rates, and exports, suggesting that economic growth indirectly affects unemployment by influencing inflation and interest rates. As per the short-term dynamics of lagged variables, past increases of unemployment rate tend to self-correct in subsequent periods. In addition, short-term increases in inflation raise unemployment, while increases in interest rates can have a stabilizing effect by reducing it. Further, export growth lowers unemployment in the short term, while exchange rate depreciations increase it. The Error Correction Terms (ECTs) in VECM, indicate the speed at which each variable adjusts to restore long-term equilibrium following economic disruptions. Accordingly, unemployment rate adjusts slowly to economic shocks, which may be due to structural labor market characteristics.

At this background, VECM analysis highlights the interconnected nature of unemployment with other economic indicators and underscores the roles of GDP growth, inflation, and interest rates in influencing employment trends. This information provides a comprehensive framework for policy interventions that address both short-term fluctuations and long-term stability in unemployment forecasting.

#### 4.3.4. Residual Assumptions

In econometric modeling, examining residuals is essential to validate the reliability and accuracy of the model's predictions. Residuals, which are the differences between the actual observed values and the model’s predicted values, serve as indicators of the model's performance. Analyzing residuals helps determine if the model assumptions hold and if the model is an appropriate fit for the data. Key assumptions include that residuals should ideally be normally distributed, exhibit no significant autocorrelation, and display homoscedasticity (constant variance) across time. Violations of these assumptions could suggest potential model mis-specification or the need for further refinement.

Therefore, to assess the reliability of the VECM in forecasting the unemployment rate, a series of diagnostic tests were performed on the residuals. The Shapiro-Wilk test checks if residuals follow a normal distribution, which can be used for balanced forecasting. Normality can be observed through residual plots as well. The Ljung-Box test ensures no significant autocorrelation, indicating that residuals are independent and free from time-based patterns. Lastly, the Breusch-Pagan test confirms homoscedasticity, implying consistent variance across residuals. Together, these tests help determine if the model is statistically reliable or if adjustments are needed to enhance forecast accuracy.

#### 4.3.4.1. Shapiro-Wilk Test for Normality of *Diff_UnemploymentRate* Residuals

**Test Statistic and Results:**

- W = 0.98779: The Shapiro-Wilk test statistic (W) is close to 1, suggesting that the distribution of the residuals is nearly normal.
- p-value = 0.5618: With a p-value well above the conventional significance levels (e.g., 0.05 or 0.01), we do not reject the null hypothesis of normality.

The Shapiro-Wilk test results indicate that the residuals from the *Diff_UnemploymentRate* equation of VECM are approximately normally distributed. This is a favorable outcome for the VECM model, as normality in residuals supports the model's assumptions and enhances the reliability of parameter estimates. This result suggests that the model is capturing the underlying dynamics of *Diff_UnemploymentRate* reasonably well, with minimal unexplained, non-normal deviations.

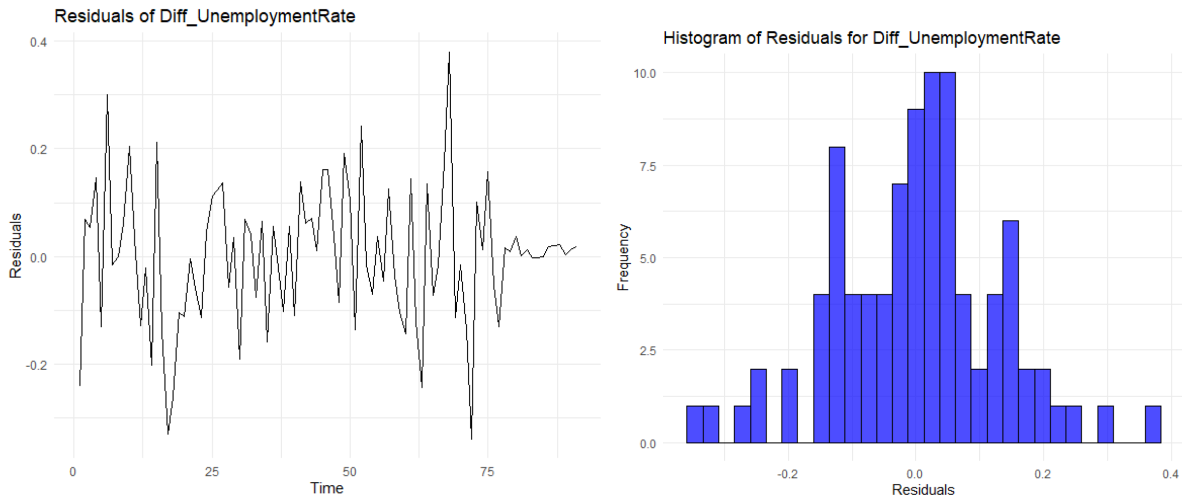


**Figure 3. Residual Plots**

Further, the above residual plots are also indicate normal distributions.

4.3.4.2. Ljung-Box Test for Autocorrelation in VECM Residuals

The Ljung-Box test assesses whether residuals from each equation in the VECM model exhibit significant autocorrelation, with results as follows:

**Key Test Results for Diff_UnemploymentRate:**

X-squared = 9.2151, p-value = 0.5118

The high p-value (0.5118) indicates no significant autocorrelation in the *Diff_UnemploymentRate* residuals, suggesting that the VECM model sufficiently accounts for time dependencies in this variable.

4.3.4.3. Breusch-Pagan Test for Heteroskedasticity in *Diff_UnemploymentRate* Residuals

Test Results:

- Test Statistic (BP): 4.4872
- Degrees of Freedom (df): 9
- p-value: 0.8765

The p-value of 0.8765 is significantly higher than conventional significance levels (0.05), indicating that the residuals for *Diff_UnemploymentRate* do not exhibit heteroskedasticity, and the variance of residuals is likely consistent across different levels of the independent variables.

At this background, the favorable results across normality, autocorrelation, and homoscedasticity tests indicate that the VECM model is statistically robust for forecasting Diff_UnemploymentRate. This strong diagnostic performance enhances the model's reliability for both understanding long-term dynamics and making short-term forecasts in the unemployment rate.

#### 4.3.5. Granger Causality Tests

The Granger causality tests assess whether past values of one variable can significantly predict future values of another. By examining these relationships, the test helps identify potential predictive linkages among economic indicators, offering insights into which factors may influence changes in dependent variable over time. This analysis aids in understanding dynamic interactions within the dataset.

**Table 10. Granger Causality Statistics**

| Variable | F-Test Value | p-value | Interpretation |
|---|---|---|---|
| Diff_GDPGrowth | 1.1451 | 0.245 | No Granger causality; Diff_GDPGrowth does not significantly predict other variables. |
| Diff_InflationRate | 1.0855 | 0.3341 | No Granger causality; Diff_InflationRate does not significantly predict other variables. |
| Diff_InterestRate | 1.142 | 0.2492 | No Granger causality; Diff_InterestRate does not significantly predict other variables. |
| Exports | 1.1993 | 0.1793 | No Granger causality; Exports does not significantly predict other variables. |
| Diff2_ExchangeRate | 1.7171 | 0.003157 | Granger causality present; Diff2_ExchangeRate significantly predicts other variables. |

| | | | |
|---|---|---|---|
| Diff2_TouristEarnings | 1.4867 | 0.02306 | Granger causality present; Diff2_TouristEarnings significantly predicts other variables. |
| TerroristAttacks | 1.1412 | 0.2503 | No Granger causality; TerroristAttacks does not significantly predict other variables. |
| COVID19 | 5.7941 | < 2.2e-16 | Strong Granger causality; COVID19 has a significant predictive effect on other variables. |

The Granger causality tests reveal that among the variables examined, Diff2_ExchangeRate, Diff2_TouristEarnings, and COVID19 show significant predictive power for changes in the unemployment rate (Diff_UnemploymentRate). The exchange rate and tourism earnings have a predictive effect on unemployment, suggesting that fluctuations in exchange rates influence export competitiveness, which in turn may impact employment in export-oriented sectors, while tourism earnings affect employment in tourism-dependent industries. The COVID-19 variable demonstrates a particularly strong predictive relationship with unemployment, highlighting the substantial and enduring economic disruptions caused by the pandemic, especially in labor-intensive sectors such as hospitality and services. In contrast, variables like GDP growth, inflation rate, interest rate, exports, and terrorist attacks do not significantly predict unemployment changes in this model, indicating that while these may influence unemployment in broader contexts, they do not provide predictive value within this specific time series framework. These results underscore the impact of sectoral and external shocks on unemployment trends, informing targeted economic responses.

#### 4.3.6. Forecast

The fitted VECM model is used to forecast unemployment rate covering four quarters. Since the model consists of the first difference of unemployment rate, the forecasts were initially drawn for that and then converted to unemployment rate.

In evaluating forecasts, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) are key indicators of accuracy, where lower values across these metrics generally signify a better model fit. MAE assesses the average magnitude of errors, with smaller values reflecting fewer absolute deviations from actual values. MAPE represents error as a percentage, allowing for intuitive comparisons across models, where lower percentages denote higher accuracy. RMSE, which squares errors before averaging, is particularly sensitive to larger errors, making it an ideal measure for capturing models' overall accuracy with lower values preferred. MSE, similar to RMSE but without the square root, penalizes larger errors and reflects model consistency.

**Table 11. Forecast Results**

| Date | Unemployment Rate | Forecasted Diff_UnemploymentRate | Forecasted Unemployment Rate | Difference |
|---|---|---|---|---|
| 7/1/2023 | 4.7 | 2.108915 | 7.308915 | -2.60892 |
| 10/1/2023 | 4.3 | -3.74166 | 3.567255 | 0.732745 |
| 1/1/2024 | 4.5 | 0.095111 | 3.662366 | 0.837634 |
| 4/1/2024 | 4.7 | 5.674544 | 9.33691 | -4.63691 |

**Table 12. Forecast Accuracy**

| | Value |
|---|---|
| Mean Absolute Error (MAE) | 2.204 |
| Mean Absolute Percentage Error (MAPE) | 47.46% |
| Root Mean Squared Error (RMSE) | 2.718 |
| Mean Squared Error (MSE) | 7.386 |

The error metrics indicate varying levels of accuracy in the model's unemployment rate forecasts. The Mean Absolute Error (MAE) of 2.204 suggests that, on average, the model's forecasts deviate from the actual unemployment values by about 2.2 percentage points. The Mean Absolute Percentage Error (MAPE) of 47.46% reveals a relatively high percentage error, implying that forecast deviations are quite significant when considered as a proportion of the actual unemployment values, which may reduce reliability in practical contexts. The Root Mean Squared Error (RMSE) of 2.718 highlights the presence of larger errors, as RMSE penalizes larger deviations more heavily, further emphasizing the model's difficulty in achieving precise forecasts. The Mean Squared Error (MSE) of 7.386 reinforces this, as it represents the average squared error, illustrating that overall forecast consistency may be affected. Overall, these metrics suggest that while the model captures general trends, it struggles with precise accuracy, particularly with larger errors, as indicated by the high MAPE.

The analysis above effectively outlines the results and implications of using the Vector Error Correction Model (VECM) to forecast the unemployment rate based on macroeconomic variables. By incorporating econometric tests and residual analyses, it underscores the value of traditional time-series modeling in capturing long-term economic relationships and assessing the impact of key drivers like GDP growth, inflation, and external shocks (e.g., COVID-19, terrorist attacks). However, the model's limitations in handling complex, non-linear relationships and its relatively high error metrics suggest that alternative forecasting approaches could improve prediction accuracy. While VECM provides a strong foundation by modeling equilibrium relationships, incorporating deep learning and machine learning models can significantly enhance forecast precision, adapt to non-linear and complex interactions, and provide a more resilient approach in highly volatile economic environments.

### *4.4. Deep Learning and Machine Learning Approaches for Unemployment Rate Forecasting*

In this study, Deep Learning (DL) and Machine Learning (ML) models are employed to enhance the accuracy of unemployment rate forecasting by leveraging their advanced capabilities in handling complex, non-linear relationships. The Deep Learning models applied here is the Feedforward Neural Network (FFNN), which uses layered neurons to uncover intricate patterns across multiple variables. The Machine Learning models include Random Forest (RF), an ensemble method that constructs multiple decision trees to capture non-linear interactions, making it ideal for complex datasets, Support Vector Regression (SVR), which optimally separates data points for effective handling of noise in predictions and Extreme Gradient Boosting (XGBoost), an efficient and powerful gradient-boosting approach that iteratively improves forecasting accuracy by reducing prediction errors. This combination of DL and ML models enables a more robust analysis, capturing subtle dynamics within the economic indicators influencing unemployment.

#### 4.4.1. Feed Forward Neural Network (FFNN) model

The Feed Forward Neural Network (FFNN) is a deep learning model often applied to forecast complex relationships in datasets, including economic indicators. It uses interconnected layers of neurons to capture non-linear relationships, making it ideal for scenarios where traditional linear models may fall short.

In this analysis, the FFNN model was trained to forecast unemployment rates. The model fitting process began with data preprocessing to ensure a consistent input structure, followed by configuring the FFNN with an appropriate architecture for time series forecasting. K-Fold cross-validation was employed to evaluate the model's performance and assess its generalization capability. Over 100 training epochs, the model's loss gradually reduced, indicating effective learning from the training data.

**Table 13. FFNN Model Performance Summary**

| | Value (Cross-Validation) | Value (Forecast Period) |
|---|---|---|
| Mean Absolute Error (MAE) | 0.55 | 0.51 |
| Mean Squared Error (MSE) | 0.39 | 0.48 |
| Root Mean Squared Error (RMSE) | 0.62 | 0.68 |
| $R^2$ (Coefficient of Determination) | -13.03 | 0.75 |
| Mean Absolute Percentage Error (MAPE) | 12.36% | 8.40% |

The FFNN model's forecast performance reveals a moderate level of accuracy based on various key metrics. The Mean Absolute Error (MAE) of 0.55 suggests that the forecasted unemployment rates deviate by approximately 0.55 percentage points from actual values on average, indicating reasonable accuracy. The Root Mean Squared Error (RMSE) of 0.62 further supports this, showing minor but present deviations in forecasted values. Additionally, a Mean Absolute Percentage Error (MAPE) of 12.36% highlights moderate reliability, with lower MAPE values generally indicating higher model robustness. Together, these metrics indicate that while the FFNN model captures general patterns, its capacity to fully explain variance is limited, suggesting room for improvement in forecast precision. Overall, the FFNN provides moderate forecasting accuracy with the given data.

### 4.4.2. Random Forest Model

The Random Forest model is a powerful ensemble learning algorithm that combines multiple decision trees to enhance predictive accuracy and robustness. Known for its resilience to overfitting, especially in the presence of noisy data, Random Forest averages the results from individual trees, which helps in capturing complex, non-linear relationships between variables. Additionally, the model provides insights into feature importance, aiding in the understanding of which variables most significantly impact the target outcome. These qualities make Random Forest particularly well-suited for tasks involving complex datasets and a mix of categorical and continuous variables.

In this analysis, a Random Forest model was trained to forecast the unemployment rate, utilizing K-Fold cross-validation to evaluate and refine model performance. Initially, default parameters were used; however, careful parameter tuning, such as adjusting the number of trees, maximum tree depth, and minimum samples per leaf, was identified as a potential improvement. A grid search or randomized search could be employed for optimal parameter selection. The cross-validation process allowed for iterative evaluation, providing robust metrics to ensure the model's generalization. Additionally, the model included relevant economic indicators (e.g., GDP growth, inflation rate) to capture factors influencing unemployment, helping improve its predictive power for the targeted timeframe.

**Table 14. Random Forest Model Performance Summary**

| | Value (Cross-Validation) | Value (Forecasting) |
|---|---|---|
| Mean Absolute Error (MAE) | 0.388 | 0.13 |
| Mean Squared Error (MSE) | 0.248 | 0.044 |
| Root Mean Squared Error (RMSE) | 0.489 | 0.21 |
| $R^2$ (Coefficient of Determination) | 0.888 | -0.605 |
| Mean Absolute Percentage Error (MAPE) | 6.33% | 2.97% |

The Random Forest model demonstrated strong performance in both cross-validation and forecasting phases, showcasing its predictive accuracy for unemployment rates. During cross-validation, the Mean Absolute Error (MAE) of 0.388 indicates that predictions deviate by less than 0.4 percentage points from actual values, underscoring minimal errors. The Mean Squared Error

(MSE) and Root Mean Squared Error (RMSE) were also low at 0.248 and 0.489, respectively, reflecting an accurate fit with minimal significant errors. In the forecasting phase, the model's MAE reduced further to 0.130, and MAPE was notably low at 2.97%, indicating close alignment with actual unemployment values. Overall, the Random Forest model's consistent performance makes it a reliable choice for forecasting unemployment rates with minimal deviations from true values.

### 4.4.3. Support Vector Regression (SVR) Model

Support Vector Regression (SVR) is a robust and flexible regression technique that is effective in modeling both linear and non-linear relationships. By utilizing a kernel function, like the Radial Basis Function (RBF), SVR can capture complex patterns in the data, making it well-suited for economic indicators that exhibit non-linear trends. This method focuses on minimizing error within a specified margin, which enhances its accuracy and resilience to outliers.

In this study, the SVR model was applied with an RBF kernel to capture non-linear relationships in the unemployment rate data. To ensure reliable performance, K-Fold cross-validation was used, which divides the dataset into multiple subsets to validate the model across different data splits. This approach allows the model to generalize better, reducing the risk of overfitting. For parameter selection, grid search techniques could be applied to optimize SVR's hyperparameters for further refinement. This setup helped in creating a model responsive to the dynamic patterns in economic indicators.

**Table 15. SVR Model Performance Summary**

| | Value (Cross-Validation) | Value (Forecast Period) |
|---|---:|---:|
| Average Mean Absolute Error (MAE) | 0.5084 | 0.2342 |
| Average Mean Squared Error (MSE) | 0.4812 | 0.0797 |
| Average Root Mean Squared Error (RMSE) | 0.6782 | 0.2824 |
| Average $R^2$ | 0.7456 | -1.8993 |

| | | |
|---|---|---|
| Average Mean Absolute Percentage Error (MAPE) | 8.4052 | 5.2836 |

The SVR model showed consistent performance across cross-validation and forecasting phases. In cross-validation, the Mean Absolute Error (MAE) was 0.5084, indicating moderate accuracy with an average deviation of about 0.51 percentage points from actual values. The Mean Squared Error (MSE) of 0.4812 and RMSE of 0.6782 also suggest satisfactory predictive precision. During the forecast period, the model improved with an MAE of 0.2342 and a MAPE of 5.28%, indicating closer alignment with actual values. This improvement in forecast performance demonstrates that the SVR model can effectively predict unemployment rates with satisfactory closeness to observed values, particularly in shorter forecast periods.

### 4.4.4. XGBoost model

XGBoost (Extreme Gradient Boosting) is an advanced machine learning algorithm known for its speed and efficiency in handling structured data and producing accurate predictions. It uses gradient boosting, an ensemble learning technique, which iteratively adds decision trees to minimize prediction errors. XGBoost is particularly effective for time series forecasting tasks, like predicting unemployment rates, as it can handle non-linear relationships and capture complex interactions between variables.

For this analysis, the XGBoost model was configured to predict unemployment rates by training on economic indicators. The model was set up with cross-validation to assess generalization across different subsets of the data. Model fitting involved determining key hyperparameters, such as the learning rate, maximum depth of trees, and the number of boosting rounds, which were optimized using a grid search approach. The training phase focused on minimizing error metrics, and the final model was fitted on the entire dataset after cross-validation to generate forecasts.

**Table 16. XGBoost Model Performance Summary**

| | Value (Cross-Validation) | Value (Forecast) |
|---|---|---|
| Mean Absolute Error (MAE) | 0.4286 | 0.276 |
| Mean Squared Error (MSE) | 0.306 | 0.1521 |

| | | |
|---|---|---|
| Root Mean Squared Error (RMSE) | 0.5507 | 0.3899 |
| $R^2$ | 0.8469 | -4.5294 |
| Mean Absolute Percentage Error (MAPE) | 7.03% | 6.22% |

The XGBoost model showed robust performance during cross-validation, with an $R^2$ of 0.8469, indicating it explains around 84.69% of the unemployment rate variance. The Mean Absolute Error (MAE) of 0.4286 and Root Mean Squared Error (RMSE) of 0.5507 suggest a low average error, demonstrating accuracy in general predictions. In the forecast period, XGBoost achieved a reduced MAE of 0.2760 and a MAPE of 6.22%, suggesting satisfactory prediction closeness to actual values. However, the model's performance variations across periods indicate it may benefit from further refinement to enhance consistency in generalization. Overall, the XGBoost model performs well as a predictive tool, with strong potential for accurate unemployment rate forecasts.

### *4.5. Forecast Comparison*

Unemployment rate of Sri Lanka is forecasted for a one-year time span from third quarter of 2023 to second quarter of 2024 employing five different models. The forecasts estimated are plotted along with the actual unemployment rate in Figure 4. However, the forecasts obtained from the fitted VECM model significantly deviate from the actual unemployment rate. Therefore, to get a clear graphical representation of other forecasts, VECM forecasts are excluded from Figure 5.

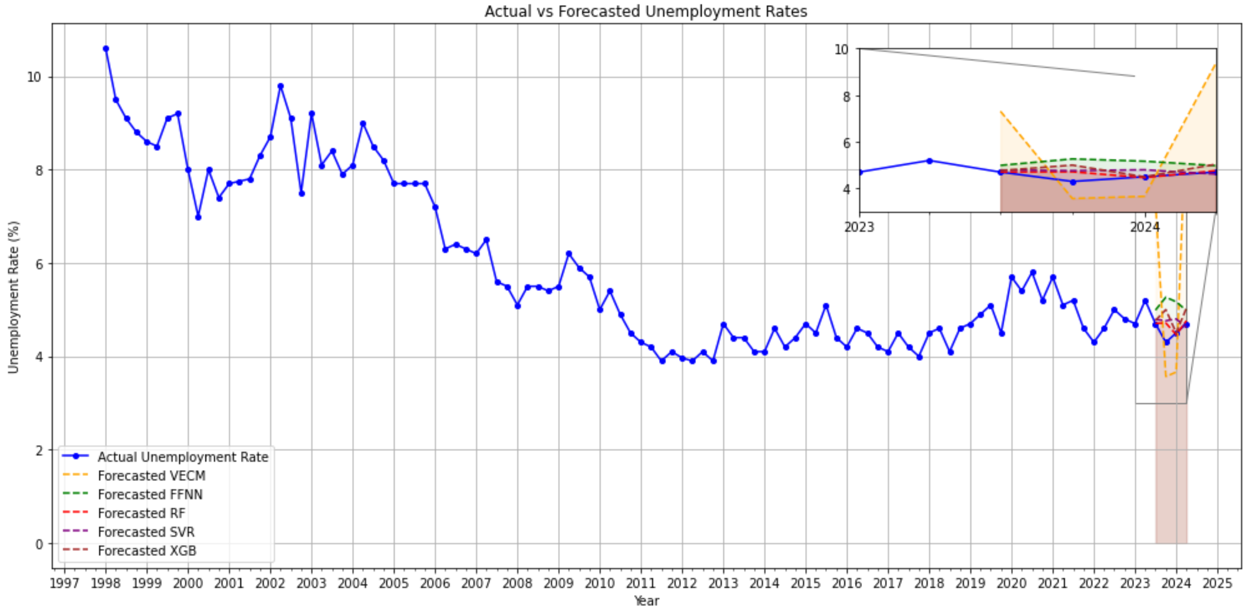


**Figure 4. Actual and Forecasted Unemployment Rate**

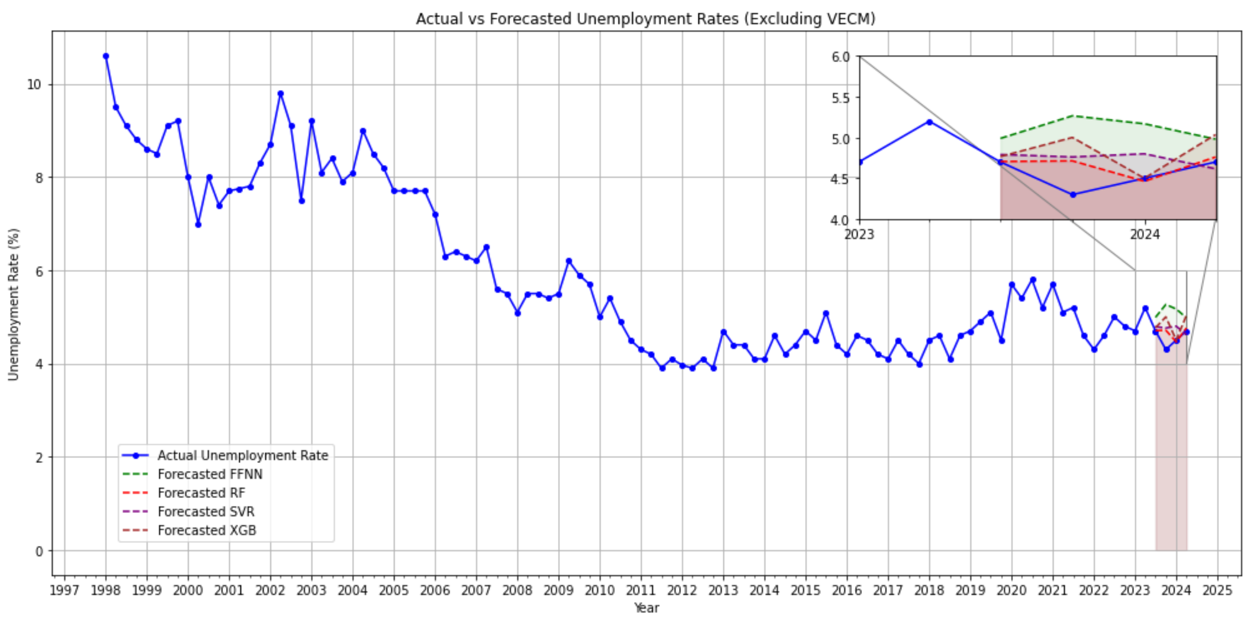


**Figure 5. – Actual and Forecasted Unemployment Rate (Excluding VECM)**

**Table 17. Actual and Forecasted Unemployment Rate**

| Year | Quarter | UnemploymentRate | Forecasted Unemployment Rate_ VECM | Forecasted Unemployment Rate_ FFNN | Forecasted Unemployment Rate_ RF | Forecasted Unemployment Rate_ SVR | Forecasted Unemployment Rate_ XGB |
|---|---|---|---|---|---|---|---|
| 2023 | 3 | 4.70 | 7.31 | 4.99 | 4.71 | 4.79 | 4.77 |
| 2023 | 4 | 4.30 | 3.57 | 5.27 | 4.71 | 4.76 | 5.00 |
| 2024 | 1 | 4.50 | 3.66 | 5.17 | 4.46 | 4.80 | 4.50 |
| 2024 | 2 | 4.70 | 9.34 | 4.98 | 4.76 | 4.62 | 5.04 |

The table illustrates the actual unemployment rates alongside forecasts from five models: VECM, FFNN, RF, SVR, and XGB, over four quarters from Q3 2023 to Q2 2024. Among these models:

- Random Forest (RF) consistently provides the most accurate forecasts, closely aligning with actual rates across all quarters, underscoring its effectiveness for unemployment rate prediction.
- FFNN and SVR offer reasonable forecasts but display slight overestimations in some quarters, indicating moderate predictive capability.
- VECM shows the highest deviations, particularly in Q2 2024, indicating challenges in capturing short-term changes effectively.
- XGB produces forecasts that are fairly close to the actual values, though with minor overestimations in a couple of periods.

The Random Forest model emerges as the most reliable option for forecasting unemployment rates in this dataset, while VECM's performance highlights its limitations in short-term forecasting.

**Table 18. Summary of Forecasting Model Performance**

| Model | MAE | MSE | RMSE | MAPE |
|---|---|---|---|---|
| VECM | 2.2041 | 7.3865 | 2.7178 | 47.46% |
| FFNN | 0.5512 | 0.3858 | 0.6212 | 8.40% |

| | | | | |
|---|---|---|---|---|
| RF | 0.1295 | 0.0441 | 0.2101 | 2.97% |
| SVR | 0.2342 | 0.0797 | 0.2824 | 5.28% |
| XGB | 0.276 | 0.1521 | 0.3899 | 6.22% |

The evaluation of various models for forecasting unemployment rates reveals that the Random Forest (RF) model outperforms the others in both cross-validation and forecasting metrics. RF achieves the lowest errors across Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), demonstrating superior accuracy and minimal deviation from actual values. Its ensemble approach effectively captures both linear and non-linear dependencies in economic data, proving robust approach for complex patterns.

Other models, such as Support Vector Regression (SVR) and XGBoost, perform well but fall short of RF in terms of accuracy. While SVR captures non-linear relationships through kernel functions and XGBoost leverages gradient boosting for effective response to macroeconomic volatility, both models show signs of overfitting in forecasting, suggesting they may require further tuning for long-term stability. The Feed Forward Neural Network (FFNN) also shows potential with moderate accuracy but needs enhanced tuning to improve generalization.

The Vector Error Correction Model (VECM), grounded in economic theory, captures long-term equilibrium relationships but shows the highest error rates due to limitations in handling non-linear shifts and rapid economic changes. Overall, Random Forest stands out as the best model for forecasting unemployment, balancing accuracy, and resilience to economic fluctuations, making it highly suitable for real-time forecasting and policy-making. This analysis highlights the strength of ensemble models, such as RF, in providing reliable forecasts that adapt to economic trends effectively.

Accordingly, it is important to use a combination of ML, DL, and artificial intelligence (AI) techniques with econometric techniques and economic theories to enhance and improve the conventional approach in economic analysis and forecasting. While econometrics relies on established models and assumptions to understand economic relationships, ML and DL can process large volumes of data and capture complex, nonlinear patterns that traditional methods may miss. AI further supports these techniques by improving predictive accuracy, optimizing decision-making, and revealing hidden trends. Integrating these advanced technologies with economic theories results in more dynamic, adaptable models that can respond to changing conditions and provide better forecasts. This approach allows analysts to refine traditional economic methods, offering more precise insights and empowering better-informed decisions, ultimately leading to improved economic outcomes.

**Conclusion**

This study examined the relationships between unemployment rates and key macroeconomic variables in Sri Lanka and evaluated the effectiveness of econometric, machine learning (ML), and deep learning (DL) models for forecasting unemployment rate. By analyzing variables such as GDP growth, inflation, interest rates, and exports, significant relationships were identified that align with established economic theories like Phillips Curve. These findings underscore the importance of macroeconomic stability in managing labor market conditions in Sri Lanka.

The study also demonstrated the potential of ML and DL approaches in improving forecast accuracy, particularly when capturing non-linear patterns that traditional econometric models may overlook. Specifically, Random Forest emerged as the most reliable model for forecasting unemployment rate in Sri Lanka, showing minimal deviation from actual values across multiple error metrics. In contrast, while the Vector Error Correction Model (VECM) effectively captured long-term equilibrium relationships, it exhibited limitations in handling short-term fluctuations and non-linear interactions. This suggests that integrating ML and DL models with traditional econometric approaches could enhance forecast accuracy and enable policymakers to respond to labor market changes more proactively.

In summary, while traditional econometric models offer valuable insights into unemployment dynamics, advanced ML and DL methods add predictive power, especially in volatile economic environments. The findings from this study contribute to the understanding of unemployment rate forecasting in Sri Lanka, highlighting a more comprehensive approach that combines conventional and non-conventional modeling techniques for more responsive policy formulation.

## References


Adebiyi, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Comparison of ARIMA and neural network models for stock price prediction. *Journal of Applied Mathematics*, 2014, 1-7. https://doi.org/10.1155/2014/614342.

Anderson, P., and Broadbent, E. (2019). Application of random forest in labor market predictions. *Journal of Economic and Financial Research*, 4(1), 75-90.

Ariyadasa, K., and Gunaratne, W. (2014). Inflation and unemployment: The Sri Lankan experience. *Sri Lanka Economic Journal*, 1(1), 33-45.

Asif, M. (2013). Determinants of unemployment in Pakistan, India, and China: A comparative study. *International Journal of Economics and Financial Issues*, 3(1), 2-5.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785.

Dogan, E. (2012). Macroeconomic variables and unemployment: The case of Turkey. *International Journal of Economics and Financial Issues*, 2(1), 71-78.

Fernando, M., and Karunaratne, T. (2018). The impact of interest rate on unemployment in Sri Lanka. *Asian Economic and Financial Review*, 8(4), 532-543.

Fischer, T., and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054.

Gaston, N., and Rajaguru, G. (2011). How an export boom affects unemployment. *Economic Modelling*, 28(3), 1229-1240. https://doi.org/10.1016/j.econmod.2010.12.005.

Gonzalez, R. (2000). The use of neural networks in macroeconomic forecasting. *Canadian Journal of Economics*, 33(3), 413-424.

Guo, C., Yu, H., and Liu, J. (2018). Forecasting unemployment with gradient boosting. *Applied Economics Letters*, 25(3), 212-216.

Jayaratne, H., and Senanayake, S. (2019). Forecasting unemployment rates in Sri Lanka using support vector regression. *Asia-Pacific Journal of Economics and Business*, 11(3), 65-76.

Jayathilaka, R., and Mahendra, R. (2016). The relationship between inflation and unemployment in Sri Lanka. *Colombo Business Journal*, 7(2), 55-72.

Nelson, D. M., Pereira, A. C. M., and Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. *International Joint Conference on Neural Networks (IJCNN)*, 1419-1426. https://doi.org/10.1109/IJCNN.2017.7966019.

Nyahokwe, O., and Ncwadi, R. (2013). The impact of exchange rate volatility on unemployment in South Africa. *Mediterranean Journal of Social Sciences*, 4(3), 109-116.

Perera, H., and Fernando, J. (2022). Using gradient boosting machines to predict unemployment in Sri Lanka. *Sri Lanka Journal of Economics*, 3(4), 221-239.

Perera, H., and Wijesinghe, R. (2020). Examining Okun's Law and the Phillips Curve in Sri Lanka. *International Journal of Applied Economics*, 17(1), 1-18.

Ranasinghe, M., and Wijeratne, K. (2020). Decision tree models in forecasting unemployment in Sri Lanka. *Sri Lankan Journal of Economic Research*, 7(1), 33-48.

Senanayake, S., Wijesinghe, A., and Perera, K. (2019). Export growth and employment in Sri Lanka. *South Asian Journal of Business and Management Cases*, 8(2), 35-45.

Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Yu, L., Huang, W., Lai, K. K., Nakamori, Y., and Wang, S. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology & Decision Making (IJITDM)*, 6, 113-140. https://doi.org/10.1142/S021962200700237X.