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

S R C L Gunawardhana

Statistics Department

Central Bank of Sri Lanka

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

13 December 2024

Outline

- Introduction
- Literature Review
- Data
- Methodology
- Empirical Results and Discussion
- Conclusion

Introduction

- Unemployment Rate is a key economic indicator that measures the percentage of the labor force actively seeking work.
- It influences economic stability, output, and social health.
- The Department of Census and Statistics releases quarterly unemployment data, with a 12 to14 weeks time lag, hindering real-time decision-making.
- Focus on forecasting Unemployment Rate in Sri Lanka using different approaches such as Machine Learning and Deep Learning in addition to Econometric Models.

Introduction Contd...

- Traditional Econometric Models such as VAR and VECM are commonly used but may struggle with complex, nonlinear data patterns.
- Advancements in techniques like Machine learning (ML) and deep learning (DL) are revolutionizing forecasting by offering enhanced predictive capabilities.
- Research Gap of limited use of ML/DL techniques for unemployment forecasting in Sri Lanka would be addressed.

Introduction Contd...

- **Study Objectives**:
  - Examine relationships between unemployment and macroeconomic variables in Sri Lanka.
  - Develop and compare forecasting models including econometric, ML, and DL approaches to improve the accuracy of unemployment rate forecasts.
- Contribution to Policy making by providing insights for better short-term decision-making and long-term economic planning.

Literature Review

- The literature review consists of traditional econometric models used in forecasting unemployment and emphasizes the potential of advanced data science techniques to improve forecasting accuracy.
- Previous studies were referred under five key categories,
  - Research Related to Unemployment Rate and Other Macroeconomic Variables in Other Countries
  - Research Related to Unemployment Rate and Other Macroeconomic Variables in Sri Lanka
  - Research on Data Science Approaches to Forecast Unemployment in Other Countries
  - Research on Data Science Approaches to Forecast Unemployment in Sri Lanka
  - Research on Data Science Approaches to Forecast Macroeconomic Variables

Literature Review Contd...

Research Related to Unemployment Rate and Other Macroeconomic Variables in Other Countries

- Unemployment is influenced by macroeconomic variables like GDP, inflation, exchange rates, and exports.
- Studies:
  - Dogan (2012): VAR model used for Turkey; found that GDP growth and exports reduce unemployment, while inflation and exchange rate fluctuations increase it. Supports Okun's Law and the Phillips Curve.
  - Nyahokwe & Ncwadi (2013): South Africa; VAR, GARCH, and VEC models; highlighted the role of exchange rate volatility, especially in export-reliant sectors.
  - Gaston & Rajaguru (2011): Australia; linked rising export prices to reduced unemployment.
  - Asif (2013): Comparative study on Pakistan, India, and China; GDP growth reduced unemployment in India and China but increased it in Pakistan.

Literature Review Contd...

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

- Previous studies have identified inflation, interest rates, exports, and exchange rates as main drivers of unemployment in Sri Lanka.
- Studies:
  - Ariyadasa & Gunaratne (2014): Applied the Phillips Curve to show that inflation raises unemployment by increasing production costs.
  - Jayathilaka & Mahendra (2016): VECM model showed that inflation reduces purchasing power and impacts employment.
  - Fernando & Karunaratne (2018): ARDL model highlighted interest rates as a significant factor; high rates discourage job creation.
  - Senanayake et al. (2019): Granger Causality test found that export growth, especially in textiles, reduces unemployment by creating labor-intensive jobs.
  - Perera & Wijesinghe (2020): Applied Okun's Law and the Phillips Curve in Sri Lanka, showing inverse relationships between GDP growth and unemployment, and a positive relationship between inflation and unemployment.

Literature Review Contd...

Research on Data Science Approaches to Forecast Unemployment in Other Countries

- Data science approaches like ML and DL improve forecasting accuracy by capturing nonlinear relationships in economic data.
- Studies:
  - Guo et al. (2018): U.S.; applied Gradient Boosting Machine (GBM) to labor market data (2005–2017); found it outperformed traditional models like ARIMA, capturing seasonal patterns and complex relationships.
  - Anderson & Broadbent (2019): Europe; used Random Forest on labor data (2010–2018) and showed it outperformed linear regression models by managing multicollinearity and complex variable interactions.

Literature Review Contd...

Research on Data Science Approaches to Forecast Unemployment in Sri Lanka

- ML techniques, including Gradient Boosting, Decision Trees, and Support Vector Regression, are being explored to improve forecasting accuracy in Sri Lanka.
- Studies:
  - Perera & Fernando (2022): Applied GBM to manage multicollinearity between inflation and exchange rates, achieving higher forecast accuracy than linear models.
  - Ranasinghe & Wijeratne (2020): Used Decision Trees to improve forecast accuracy by 15%, effectively capturing interactions among GDP growth, exports, and inflation.
  - Jayaratne & Senanayake (2019): Used SVR to forecast unemployment during volatile periods, finding it outperformed ARIMA models in terms of Mean Absolute Error (MAE) and Mean Squared Error (MSE).

Literature Review Contd...

Research on Data Science Approaches to Forecast Macroeconomic Variables

- ML models are also successfully used to forecast other macroeconomic variables like GDP and inflation, demonstrating the broader applicability of data science methods.
- Studies:
  - Gonzalez (2000): Used neural networks to forecast GDP, showing that they outperform traditional models in capturing both linear and nonlinear relationships.
  - Fischer & Krauss (2018): Used LSTM (Long Short-Term Memory) models on daily financial data (1992–2015) to predict inflation trends, showcasing the model’s strength in handling sequential data.
  - Nelson et al. (2017): Applied Recurrent Neural Networks (RNNs) to forecast GDP and interest rates, highlighting their effectiveness in time-series forecasting.
  - Adebiyi et al. (2014), Chen & Guestrin (2016), Smola & Schölkopf (2004): Demonstrated success of ML models (neural networks, XGBoost, SVR) in forecasting volatile, seasonally influenced economic data.

Literature Review Contd…

- Econometric Models like Vector AutoRegression (VAR) and Vector Error Correction Models (VECM) are widely used to study macroeconomic influences on unemployment.
- Despite the global success of ML and DL models, their application in Sri Lanka's unemployment forecasting remains underexplored since there are limited use of advanced data science techniques in Sri Lanka.
- This study aims to address the gap by evaluating ML and DL models alongside traditional methods for more accurate and timely unemployment rate forecasting in Sri Lanka.
- While traditional models provide foundational insights, ML and DL approaches offer enhanced forecasting precision.
- This can improve Sri Lanka’s economic forecasting and policymaking capabilities.

Data

- Seven macroeconomic variables and two dummy variables for the time span of 1998 to 2024 are considered in the study.
- The 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.

Data Contd…

Variable	Description
Unemployment Rate	Proportion of unemployed population to the total labor force above 15 years of age.
GDP Growth	Increase in national income/output, measured by the growth of real GDP. Quarterly growth values are calculated from real GDP (base year 2015).
Inflation	Percentage change in the Colombo consumer price index (CPI), based on a basket of goods. Quarterly values derived from monthly CCPI averages.
Interest Rate	Percentage charged by lenders to borrowers annually. Quarterly average of daily 91-day Treasury Bill rates.
Exports	Goods and services produced and traded internationally. Quarterly data in USD billion terms are used.
Exchange Rate	Units of domestic currency per unit of foreign currency (e.g., LKR per USD). Monthly values converted to quarterly averages.
Tourist Earnings	Earnings from international tourism-related services in USD million terms. Quarterly data are obtained by aggregating the relevant three months pertaining to the quarter.
Terrorist Attacks	Dummy variable indicating periods with significant terrorist attacks.
COVID-19	Dummy variable indicating periods with high economic impact due to COVID-19.

Methodology

- Consists of theories/tests and methods for
  - Econometric Time Series Analysis
  - Data Science Techniques
  - Forecast Evaluation Methods

Econometric Time Series Analysis

Initial Tests:

- Correlation Coefficient: Evaluate the relationships among the variables, helping to identify the strength and direction of linear associations.
- Kruskal-Wallis Test: Assess seasonality in the data, testing if there are significant differences in the distribution of variables across different seasons.
- Augmented Dickey-Fuller (ADF) Test: Check the stationarity of the variables. Non-stationary data is transformed to become stationary before further analysis.
- Johansen Test for Cointegration: Examine whether a long-term equilibrium relationship exists between the non-stationary variables, indicating that they move together over time.
- Granger Causality Test: Investigate the cause-and-effect relationship between variables, helping to understand if one variable can predict another.

Econometric Time Series Analysis

Forecasting Model

- Vector Error Correction Model (VECM): A multivariate model used to capture both short-term dynamics and long-term equilibrium relationships between cointegrated variables.
- Differenced data (stationary) and error correction terms are included to model the adjustment of variables toward equilibrium over time.
- VECM is useful for analyzing interconnected variables and how they adjust to shocks while maintaining long-term stability.

Econometric Time Series Analysis

Model Adequacy Tests:

- Shapiro-Wilk Test: Test for the normality of residuals.
- Ljung-Box Test: Assess autocorrelation in residuals to check for any remaining correlation.
- Breusch-Pagan Test: Test for heteroskedasticity, ensuring that the variance of residuals is constant over time.

Data Science Techniques

- Deep Learning (DL) and Machine Learning (ML) models which are both subsets of artificial intelligence (AI), are used to further strengthen the study.

Aspect	**Machine Learning (ML)**	**Deep Learning (DL)**
Definition	A subset of AI using algorithms to learn from data and make decisions without explicit programming.	A subset of ML involving neural networks with many layers to automatically learn hierarchical features.
Models	Decision Trees, SVM, Random Forest, etc.	Neural Networks (e.g., CNNs, RNNs, FFNN)
Data Requirements	Works well with smaller datasets.	Requires large datasets to perform effectively.
Computational Resources	Less computationally intensive.	Requires significant computational power
Complexity	Generally simpler with fewer parameters and layers.	More complex with deeper networks and more parameters.
Performance	Suitable for tasks with clear relationships in data.	Best for complex tasks with unstructured data

Data Science Techniques

- **Deep Learning**
  - Feedforward Neural Network (FFNN)
    - A supervised deep learning model with interconnected layers of neurons, which map complex relationships between inputs and outputs.
    - Effective at capturing non-linear relationships in time series data, especially when interactions among macroeconomic factors are complex.
- **Machine Learning**
  - Random Forest (RF)
    - An ensemble learning method that constructs multiple decision trees for prediction.
    - It’s effective for handling non-linear data and ranking feature importance in economic forecasting.

Data Science Techniques

- **Machine Learning**
  - Support Vector Regression (SVR)
    - A regression technique that finds an optimal hyperplane to separate data points, maximizing the margin for prediction.
    - SVR is robust to noise, handles outliers well, and captures non-linear patterns effectively, making it suitable for economic forecasting with noisy data.
  - Extreme Gradient Boosting (XGBoost)
    - A gradient-boosting ensemble method where decision trees are sequentially built, with each tree aiming to correct the errors of the previous one.
    - Balances model complexity and performance, handling both linear and non-linear relationships, which is crucial for accurate economic forecasting.

Forecast Evaluation

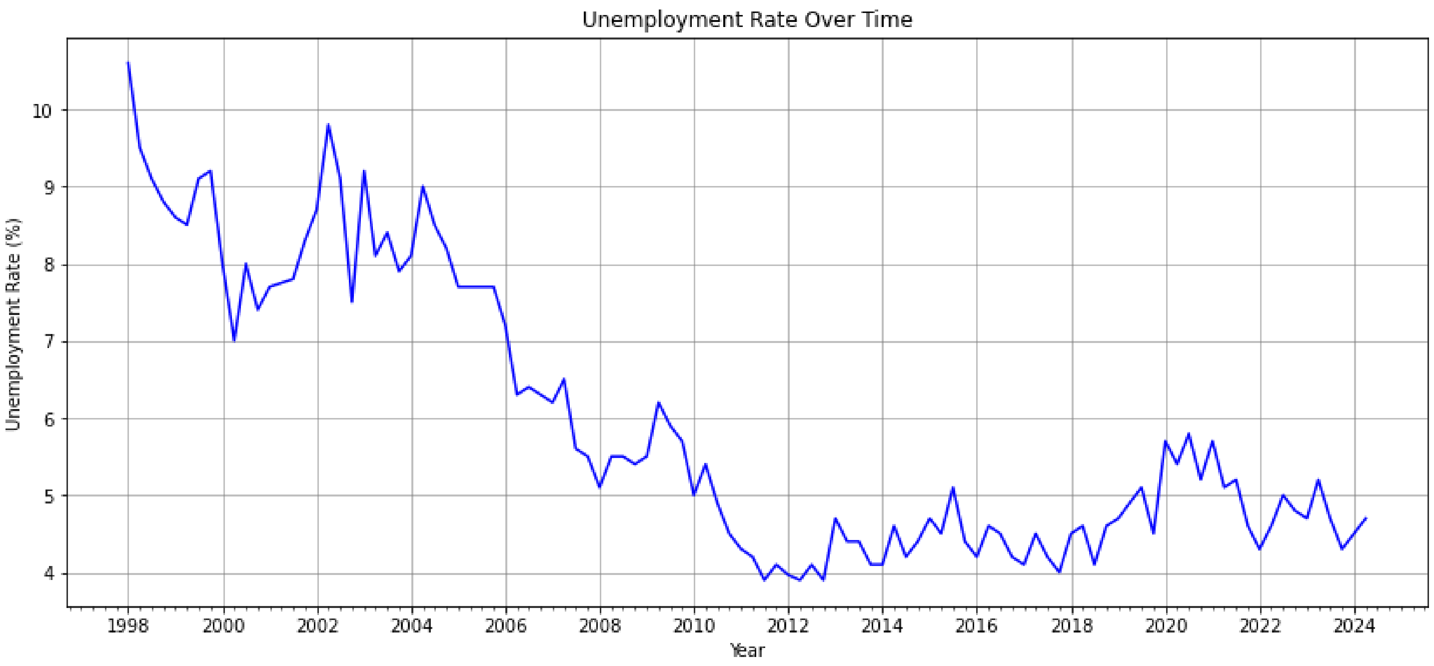
- Forecast accuracy is assessed using four key performance metrics to evaluate the effectiveness of the models in predicting macroeconomic variables.

	Description	Interpretation
Mean Absolute Error (MAE)	Measures the average of the absolute differences between actual and predicted values.	Lower MAE indicates better model performance. MAE reflects the average magnitude of forecast errors.
Mean Absolute Percentage Error (MAPE)	Expresses the average absolute error as a percentage of the actual values.	Lower MAPE indicates better performance, useful for comparing models with different scales of data.
Mean Squared Error (MSE)	Measures the average of the squared differences between actual and predicted values.	Lower MSE values indicate better model performance, as it is sensitive to large errors, emphasizing consistency.
Root Mean Squared Error (RMSE)	Calculates the square root of the average squared differences between actual and predicted values.	RMSE penalizes larger errors more heavily and is useful for assessing how well the model forecasts extreme values or outliers.

Empirical Results and Discussion

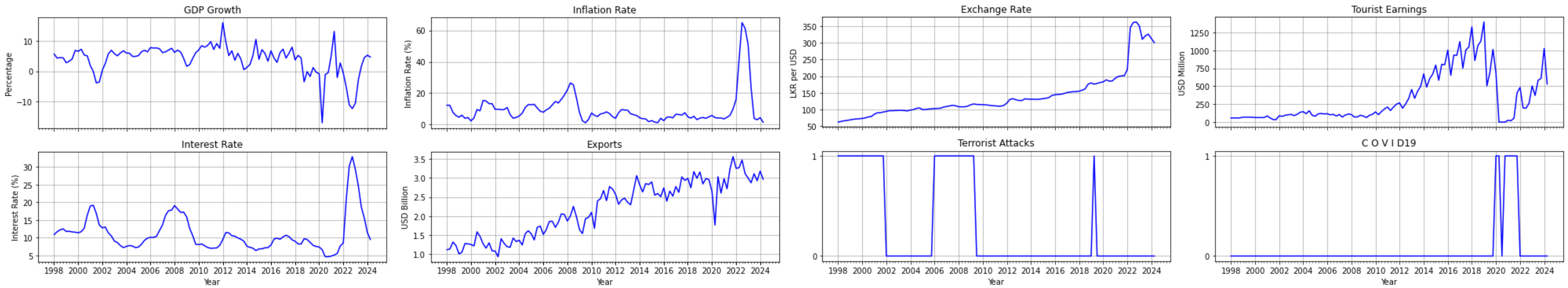
- Temporal Dynamics: Analysis of macroeconomic variables to understand their trends and variations over time.
- Model Estimation:
  - Time Series Econometric Model - Vector Error Correction Model
  - Deep Learning Model - Feed Forward Neural Network Model
  - Machine Learning Models - Random Forest, Support Vector Regression, and Extreme Gradient Boosting.
- Forecasting Performance: MAE, MAPE, RMSE, and MSE to compare model accuracy.
- A comparative analysis is carried out to identify the most accurate model for unemployment rate forecasting in Sri Lanka.

Temporal Variation of Time Series Variables



- **1998 to 2012**: Unemployment rate declined over time.
- **Post-2012 to Mid-2019**: Unemployment remained stable with minor fluctuations.
- **2019 to 2021:** Unemployment rate increased due to Easter Attack and COVID-19 Outbreak
- **From 2022 Onwards**: Unemployment fluctuates with relatively low volatility.

Temporal Variation of Variables



Relationship among Variables

- Pearson's Correlation Coefficient reveals the strength and direction of relationships between unemployment and other macroeconomic variables.

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 a strong negative correlation between unemployment and exports (-0.868) and tourist earnings (-0.585), indicating that growth in these sectors contributes to employment creation.
- There is a moderate negative relationship with the exchange rate (-0.519), while correlations with GDP growth (0.032), inflation (-0.007), and interest rates (0.056) are weak, suggesting limited impact on unemployment.
- Despite weak traditional macroeconomic relationships, all these variables are included in the study due to their interconnectedness in the economy.

Test for Seasonality of Variables

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

- Identification of recurring patterns or fluctuations in time series data influenced by economic cycles or seasonal factors, is important for improving forecasts.
- Kruskal-Wallis Test: Used to assess the statistical significance of seasonality, offering a more reliable measure than visual inspection.
- Test Results: All variables have p-values greater than 0.05, indicating no significant seasonal effects, so seasonal decomposition is not necessary for further analysis.

Test for Stationary of Variables

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	0.02	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)

- Stationary Time Series: Has constant mean, variance, and autocorrelation, making it predictable and suitable for forecasting.
- Non-stationary data can lead to unreliable results, so establishing stationarity is crucial for econometric modeling.
- ADF Test Results:
  - Exports are stationary at the original level.
  - Unemployment rate, GDP growth, inflation rate, and interest rate are stationary after first differencing.
  - Exchange rate and tourist earnings are stationary after second differencing.

Cointegration among Variables

Johansen Test Results

Hypotheses (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

- Cointegration is used to identify long-term relationships between non-stationary variables that move together despite individual trends.
- Cointegration enables the use of models like VECM, which capture both short-term dynamics and long-term equilibrium, providing insights into the interconnectedness of economic variables.
- Johansen Test Results: Identified two cointegrating relationships, indicating that certain variables, including the unemployment rate, are bound together in a long-term equilibrium.

Cointegration among Variables

Normalized Eigenvectors for Cointegration Relations

Variable	Unemployment Rate	GDP Growth	Inflation Rate	Interest Rate	Exports	Exchange Rate	Tourist Earnings
Coefficient	1	0.973	-0.825	0.538	0.927	0.062	-0.005

With the presence of cointegration among variables, Vector Error Correction Model (VECM) is ideal for forecasting unemployment, as it captures both the short-term dynamics and long-term equilibrium relationships between cointegrated variables.

- Positive Relationship with GDP Growth: Higher economic growth is associated with higher unemployment, suggesting a possible mismatch between economic expansion and job creation.
- Inverse Relationship with Inflation: Aligns with the Phillips curve, indicating that inflation could potentially reduce unemployment under certain conditions.
- Positive Relationship with Interest Rates: Higher borrowing costs could restrict employment creation by hindering investment and economic activity.
- Positive Relationship with Exports: Higher exports is associated with higher unemployment.
- Minor Positive Relationships with Exchange Rate: Suggest limited direct impact on unemployment.
- Slight Inverse Relationship with Tourist Earnings: Implies marginal employment creation within tourism-related sectors, suggesting that growth in tourism may slightly reduce unemployment.

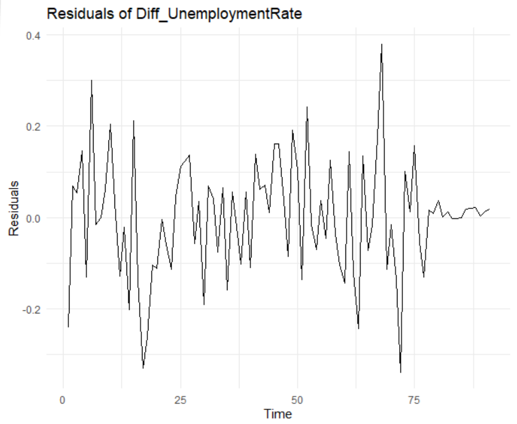
Estimation of Vector Error Correction Model

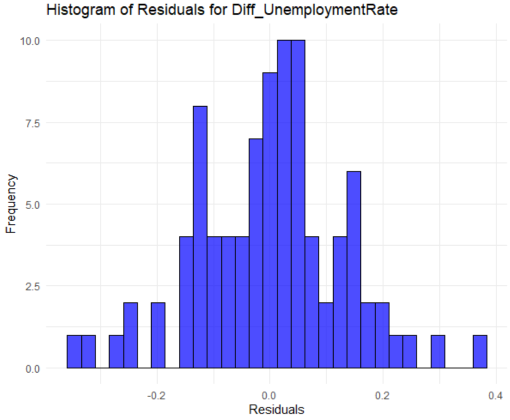
- Cointegrating Relationships:
  - The Johansen test confirms two cointegrating relationships between variables, which suggests a stable, long-term equilibrium.
  - Therefore, the number of cointegrating relationships (r) for the VECM model is set to r = 2.
- Lag Length Selection: Various criteria (AIC, HQ, SC, FPE) suggest an optimal lag length around 10.
  - However, due to the limited data, lag length 8 is selected to maintain a balance between model fit and simplicity.
- A sample of 100 data points were used for model fitting and another 4 data points were kept aside for forecast evaluation.

Residual Tests for VECM Model

- Residuals are the differences between actual values and model predictions, crucial for assessing model performance and assumptions.
- Key Assumptions:
  - Normality
    - Shapiro-Wilk Test: Assesses normality of residuals.
  - No Autocorrelation
    - Ljung-Box Test: Ensures no significant autocorrelation.
  - Homoscedasticity
    - Breusch-Pagan Test: Confirms homoscedasticity, ensuring constant variance across residuals.
- These tests ensure the VECM model's reliability for forecasting the unemployment rate and indicate whether adjustments are needed to improve accuracy.

Normality of Residuals





Shapiro-Wilk Test

- Test Statistic (W): 0.98779

  (close to 1, indicating normal distribution)

- p-value: 0.5618

  (greater than 0.05, suggesting normality)

- Residuals are approximately normally distributed, supporting the reliability of the VECM model.

Autocorrelation of Residuals

Ljung-Box Test

- Test Statistic (X-squared) = 9.2151
- p-value = 0.5118

  (greater than 0.05, suggesting no autocorrelation)
- No significant autocorrelation in residuals, indicating the VECM model accounts for time dependencies in the unemployment rate.

Heteroskedasticity of Residuals

Breusch-Pagan Test for Heteroskedasticity

- Test Statistic (BP) = 4.4872
- Degrees of Freedom (df) = 9
- p-value = 0.8765 (greater than 0.05, suggesting no heteroskedasticity)
- No heteroskedasticity in residuals, indicating consistent variance across different levels of independent variables.

Model Robustness

- The favorable results from the Shapiro-Wilk (normality), Ljung-Box (autocorrelation), and Breusch-Pagan (homoscedasticity) tests confirm that the VECM model is statistically robust.
- This strong diagnostic performance ensures the model's reliability for:
  - Understanding long-term dynamics.
  - Making accurate short-term forecasts of the unemployment rate.

Granger Causality Tests

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	0.334	No Granger causality; Diff_InflationRate does not significantly predict other variables.
Diff_InterestRate	1.142	0.249	No Granger causality; Diff_InterestRate does not significantly predict other variables.
Exports	1.1993	0.179	No Granger causality; Exports does not significantly predict other variables.
Diff2_ExchangeRate	1.7171	0.003	Granger causality present; Diff2_ExchangeRate significantly predicts other variables.
Diff2_TouristEarnings	1.4867	0.023	Granger causality present; Diff2_TouristEarnings significantly predicts other variables.
TerroristAttacks	1.1412	0.250	No Granger causality; TerroristAttacks does not significantly predict other variables.
COVID19	5.7941	< 2.2 e-16	Strong Granger causality; COVID19 has a significant predictive effect on other variables.

- Granger causality tests assess whether past values of one variable can significantly predict future values of another.
- Test results show that exchange rate, tourism earnings, and COVID-19 significantly predict changes in unemployment.
- GDP growth, inflation, interest rate, exports, and terrorist attacks do not significantly predict unemployment changes in this model.

Forecast using VECM Model

Quater	Unemployment Rate	Forecasted Diff_Unemployment Rate	Forecasted Unemployment Rate	Difference
Q3 2023	4.7	2.108915	7.308915	-2.60892
Q4 2023	4.3	-3.74166	3.567255	0.732745
Q1 2024	4.5	0.095111	3.662366	0.837634
Q2 2024	4.7	5.674544	9.33691	-4.63691

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

- The VECM model forecasts the unemployment rate for four quarters, showing substantial deviations from actual unemployment values, particularly in the first and last forecast periods.
- The model's accuracy is assessed using error metrics, including MAE, MAPE, MSE and RMSE, all indicating significant forecasting errors, especially in terms of larger deviations from actual unemployment rates.
- Despite capturing long-term equilibrium relationships, the VECM model struggles with accuracy in short-term predictions, with high error metrics suggesting limited reliability for precise unemployment forecasting.
- The analysis suggests exploring alternative forecasting techniques, like machine learning and deep learning models, to improve prediction accuracy, handle non-linear relationships, and adapt to the volatile economic environment.

Deep Learning - FFNN Model

	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 Feed Forward Neural Network (FFNN) model was used to predict unemployment rates, leveraging its ability to capture non-linear patterns in economic data.
- The model showed moderate accuracy, with performance improving during the forecast period compared to cross-validation.
- The R^2 value increased significantly, indicating that the model explained the data better over time.
- The FFNN model was reasonably accurate in terms of forecast precision.

Machine Learning - Random Forest Model

	Value (Cross-Validation)	Value (Forecast Period)
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 was used to predict unemployment rates, effectively capturing complex, non-linear relationships between economic indicators.
- The model showed good accuracy, with performance improving significantly from cross-validation to the forecast phase.
- With low error metrics, including a low Mean Absolute Percentage Error (MAPE), the model proved to be a reliable tool for forecasting unemployment rates with minimal deviations.

Machine Learning - SVR Model

	Value (Cross-Validation)	Value (Forecast Period)
Mean Absolute Error (MAE)	0.5084	0.2342
Mean Squared Error (MSE)	0.4812	0.0797
Root Mean Squared Error (RMSE)	0.6782	0.2824
R^2 (Coefficient of Determination)	0.7456	-1.8993
Mean Absolute Percentage Error (MAPE)	8.4052	5.2836

- The SVR model captures non-linear relationships in unemployment rate data and offers the advantage of effectively handling complex patterns, making it reliable for prediction tasks.
- In cross-validation, the model showed moderate accuracy, with some improvement in performance during the forecast period.
- The model's performance improved during the forecast phase, with better alignment to actual values and lower error metrics, indicating stronger accuracy.
- The results demonstrate that the SVR model is effective in predicting unemployment rates, particularly for shorter forecast periods, with satisfactory accuracy in both phases.

Machine Learning - XGBoost Model

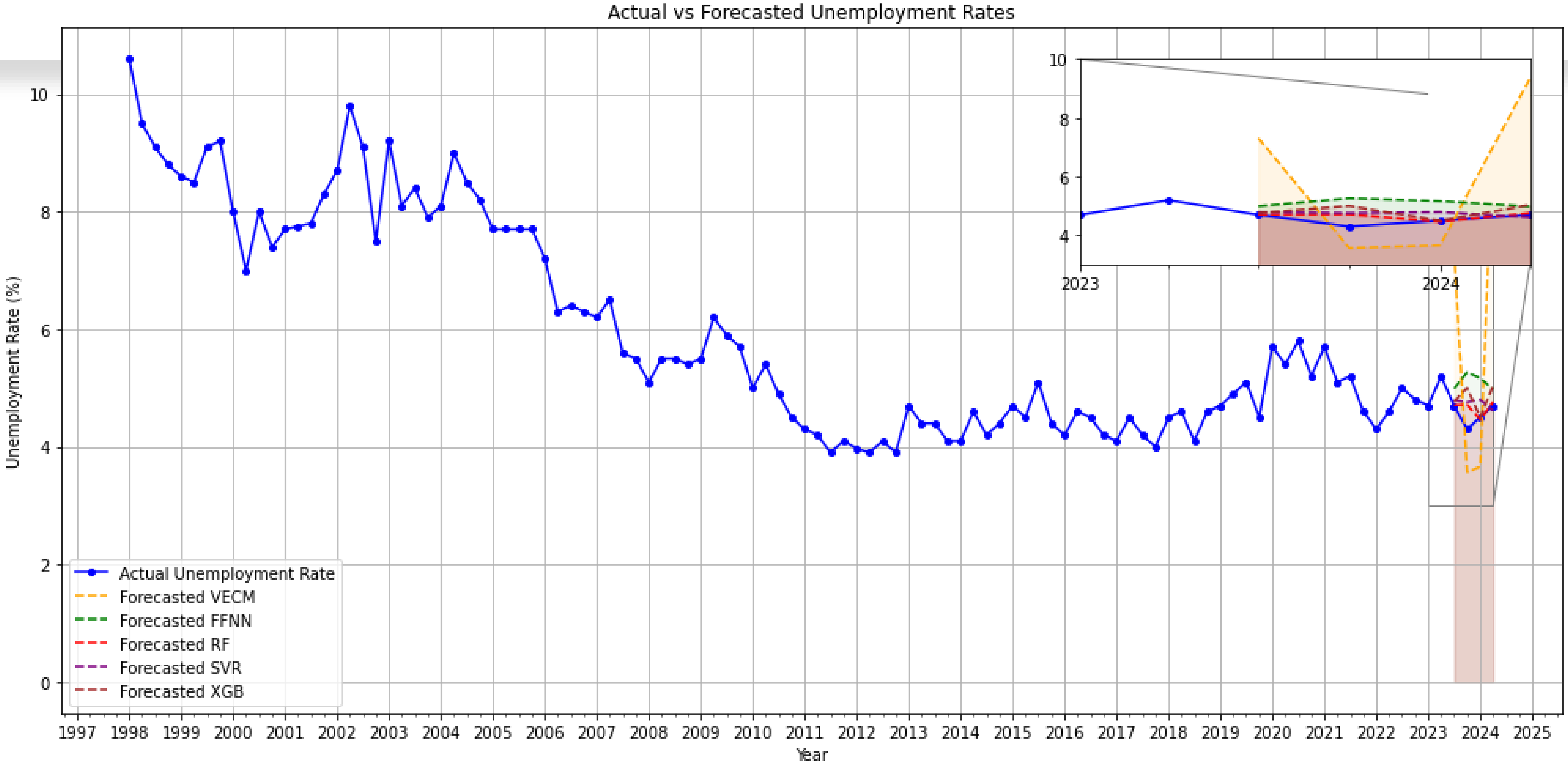
	Value (Cross-Validation)	Value (Forecast Period)
Mean Absolute Error (MAE)	0.4286	0.276
Mean Squared Error (MSE)	0.306	0.1521
Root Mean Squared Error (RMSE)	0.5507	0.3899
R^2 (Coefficient of Determination)	0.8469	-4.5294
Mean Absolute Percentage Error (MAPE)	7.03%	6.22%

- XGBoost, is highly effective for time series forecasting, capturing non-linear relationships and interactions between economic indicators to predict unemployment rates.
- During cross-validation, the model explained a high percentage of the variance in unemployment rates, with good accuracy based on its error metrics.
- XGBoost proved to be a reliable and effective tool for forecasting unemployment rates, capturing non-linear relationships and showing strong performance with improved accuracy in the forecast period..

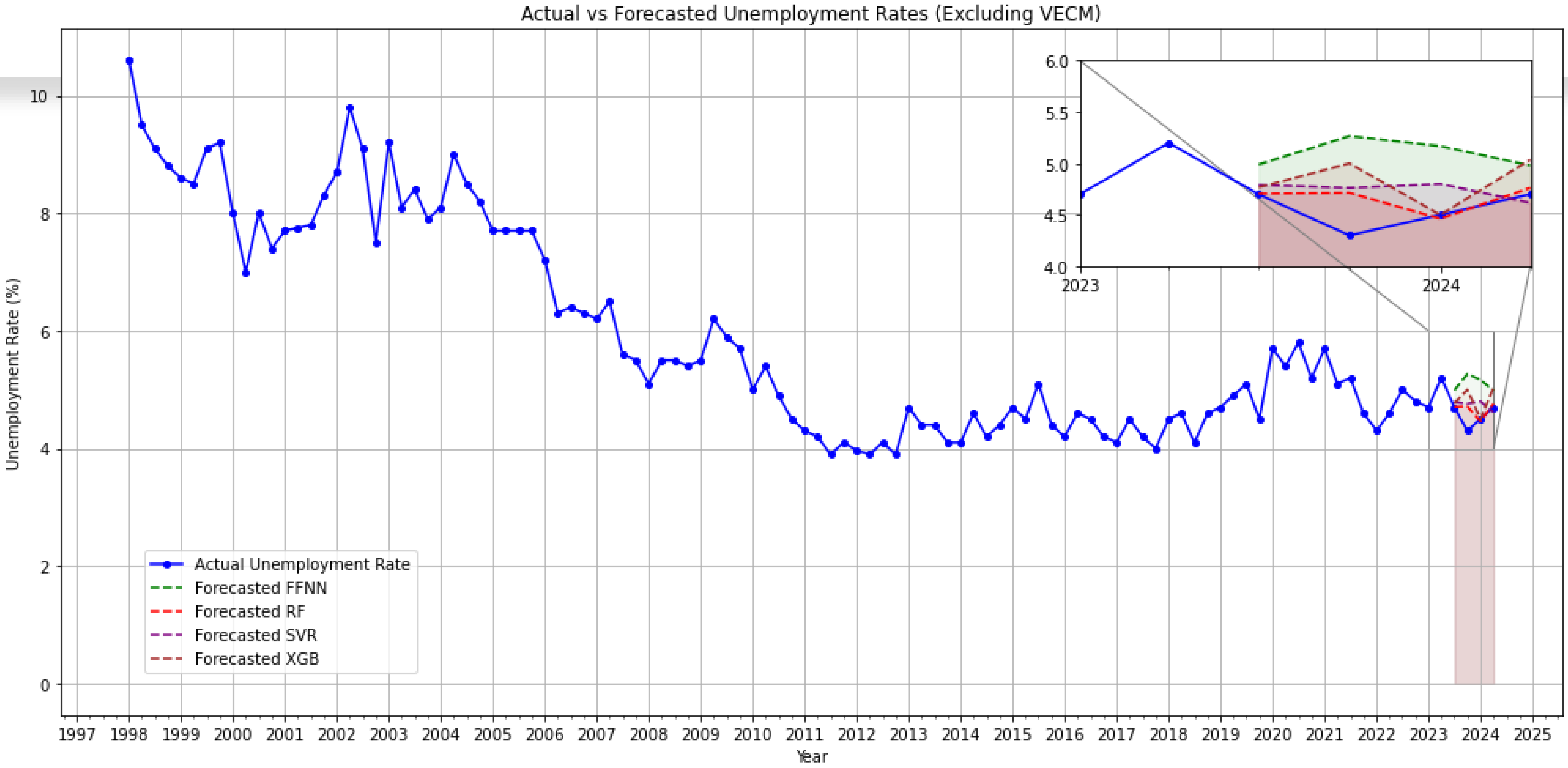
Forecast Comparison

- The unemployment rate of Sri Lanka is forecasted for a one-year period from the third quarter of 2023 to the second quarter of 2024 using five different models.
- The forecasts obtained from these models are compared with the actual unemployment rate, and the results are plotted in the first figure below.
- The forecasts from the fitted VECM model significantly deviate from the actual unemployment rate.
- To provide a clearer visual comparison of the other model forecasts, the VECM forecasts are excluded from the second figure.

Actual and Forecasted Unemployment Rate



Actual and Forecasted Unemployment Rate



Actual and Forecasted Unemployment Rate

- The table presents the actual unemployment rates and forecasts from five models: VECM, FFNN, RF, SVR, and XGB, covering four quarters from Q3 2023 to Q2 2024.

Year	Quarter	Unemployment Rate	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

Actual and Forecasted Unemployment Rate

- VECM exhibits the highest deviations, especially in Q2 2024, highlighting its difficulty in capturing short-term fluctuations.
- Random Forest model consistently provides the most accurate forecasts, aligning closely with the actual rates in most of the times, indicating its strong predictive ability for unemployment.
- FFNN and SVR show reasonable accuracy but slightly overestimate in certain quarters, suggesting moderate forecasting performance.
- XGB offers forecasts close to the actual values, with minor overestimations in a few periods.
- Overall, Random Forest stands out as the most reliable model for forecasting unemployment, while the econometric model, VECM shows limitations in short-term prediction accuracy.

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%

- Achieving the lowest errors in MAE, MSE, RMSE, and MAPE, Random Forest model, demonstrates its superior predictive accuracy.
- Support Vector Regression and XGBoost show reasonable performance but fall short of RF, with some overfitting and minor forecast inconsistencies suggesting the need for further tuning.
- Feed Forward Neural Network demonstrates moderate accuracy but requires improved tuning to enhance its generalization for better forecasting reliability.
- Vector Error Correction Model, shows the highest error rates due to its limitations in handling non-linear shifts and rapid economic changes.

Conclusion

- Random Forest is identified as the most effective model for forecasting unemployment in Sri Lanka, excelling in accuracy and resilience to economic fluctuations, making it ideal for forecasting economic variables.
- The study explored identifying the relationships of key macroeconomic variables like GDP growth, inflation, interest rates, and exports with unemployment rate, highlighting the importance of macroeconomic stability for managing labor market conditions in Sri Lanka.
- While VECM provided valuable long-term insights, its inability to handle short-term fluctuations and non-linear interactions suggests that combining traditional econometric methods with ML and DL can improve forecasting accuracy and allow for more proactive policy responses to labor market dynamics.
- It would help in capturing complex, non-linear patterns that traditional methods may miss and ultimately improving economic decision-making.

Thank You!