ANALYZING THE LOW COUNTRY TEA YIELD IN SOUTHWESTERN REGION OF SRI LANKA, WITH SPECIAL REFERENCE TO WEATHER VARIABLES

Samithige Amila Perera

(189064T)

Degree of Master of Science in Business Statistics

Department of Mathematics

University of Moratuwa

Sri Lanka

September 2023

ANALYZING THE LOW COUNTRY TEA YIELD IN SOUTHWESTERN REGION OF SRI LANKA, WITH SPECIAL REFERENCE TO WEATHER VARIABLES

Samithige Amila Perera

(189064T)

Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Business Statistics

Department of Mathematics

University of Moratuwa

Sri Lanka

September 2023

DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the PhD/MPhil/Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. P.M. Edirisinghe

Signature of the Supervisor: Date:

ACKNOWLEDGEMENT

I would like to express my heartiest gratitude to my supervisor, Dr. P.M. Edirisinghe, Senior Lecturer, Department of Mathematics, University of Moratuwa for his supervision and kind attention throughout the research project to make it a success.

I would also like to thank the coordinator of this postgraduate course, Ms. Ravindi Jayasundara, Department of Mathematics, Faculty of Engineering, University of Moratuwa for her greatest support throughout my research project.

Furthermore, I would like to thank the Manager and the staff of Kahawatta Plantation, Rathnapura for giving opportunity, guidance, and endless help to the success of this research.

I would like to thank the management and the staff of Dilmah Conservation, Endana Plantation, Dilmah Ceylon Tea PLC and Tea Research Institute, Rathnapura for supporting me in many ways to complete this research project successfully.

ABSTRACT

Tea cultivation is a crucial economic activity in southwestern region of Sri Lanka, contributing significantly to the nation's economy. Understanding the impact of the weather on tea production is crucial for food security and economic stability in the region amid climate change. The research uses historical sample of weather records and tea yield data to assess the impact of weather conditions on daily tea crop yield in southwestern region in low country of Sri Lanka to develop a forecasting model which uses historical data for future tea yield predictions.

Preliminary findings suggested that temperature fluctuations, particularly extreme heat conditions have a discernible effect on tea yield. Furthermore, the role of rainfall patterns, including seasonal distribution and intensity, was pivotal in shaping the success of tea cultivation. Additionally, the influence of humidity, wind speed and irradiance were explored on tea production, recognizing their potential as key drivers of yield variation.

In this study, both conventional time series approaches and more recent machine learning methods were used to analyze the datasets for all variables. The success of the tea yield forecasting with the SARIMA model was observed in univariate time series. Additionally, when considering multivariate time series, it was found that other variables could be forecasted for our dependent variable, tea yield. The VAR model produced forecasts with a lag order of 1 in a stationary context. Machine learning forecasting techniques demonstrated higher precision and practical applicability, making their approach more essential for future studies.

This study aims to enhance sustainable agriculture and climate resilience in Sri Lanka's low-country region by identifying critical weather variables and their impacts. It will inform local farmers, policymakers, and stakeholders, optimizing cultivation practices, resource allocation, and risk management strategies, ultimately enhancing tea farming's resilience and contributing to the global tea industry's stability midst climate change.

Key words: Tea cultivation, SARIMA, VAR

Contents

DECLARATION 1
ACKNOWLEDGEMENT 2
ABSTRACT 3
List of Figures 6
List of Tables 7
List of Appendices 8
CHAPTER 1 INTRODUCTION 9
1.1 Background of the study 9
1.2 Asian Region Tea 10
1.3 Types of Tea Produced in Sri Lanka 10
1.4 Tea Varieties in the world 11
1.5 Vegetatively Propagated Tea and Seedling Tea 12
1.6 Problem identification and justification 12
1.7 Objectives the study 13
CHAPTER 2 LITERATURE REVIEW 14
2.1 World Tea Production and History 14
2.2 Tea Plantation in Sri Lanka 15
2.3 Climate - Weather Conditions and Tea Production 16
2.4 Rainfall Effect for the Tea Yield 18
2.5 Wind Effect for the Tea Yield 18
2.6 Humidity Effect for Tea Yield 19
2.7 Irradiance effect for Tea Yield 19
CHAPTER 3. MATERIALS AND METHODOLOGY 21
3.1 Study Area 21
3.2 Data Collection and Preprocessing 22
3.3 Weather Data 23
3.3.1 Wind Speed (km/h) 23
3.3.2 Wind Direction (°) 24
3.3.3 Irradiance (W/m²) 24
3.3.4 Total Energy (kWh) 24
3.3.5 Temperature (°C) 24
3.3.6 Humidity (RH) 24
3.3.7 Rainfall (mm) 25
3.4 Data Preparation 25

3.5 Formulation 25
3.6 Time Series Analysis 25
3.6.1 Stationary Series 26
3.6.2 Decomposition of the Time Series 27
3.6.3 Components of Time Series 28
3.6.4 Auto-Correlation (ACF) 29
3.6.5 ARIMA/ SARIMA Models 31
3.6.6 Time series regression 32
3.6.7 VAR – (Vector Auto Regression) 33
3.6.8 VECM – (Vector Error Correction Model) 34
3.6.9 Time series forecasting 35
CHAPTER 4. RESULTS AND DISCUSSION 38
4.1 Descriptive statistics 38
4.2 Univariate time series 39
4.3 Decomposition Test 40
4.4 SARIMA Forecast 42
4.5 Multivariate Time series 44
4.6 Cointegration test for Multivariate Time series 45
4.7 The Dickey-Fuller test 47
4.8 KPSS statistic 48
4.9 The Granger Causality 49
4.1 VIF (Variance Inflation Factor) Test 51
4.10 Correlation between dependent variable and independent variables 53
4.11 VAR Model 53
4.12 VECM Model 55
4.12.1 The Rank of the Vector Error Correction Model (VECM) 56
4.13 VECM Model 57
4.14 Ensemble for Multivariate Time Series Forecasting 59
4.14.1 XGBoost - lagged features and rolling statistics 60
4.14.2 Lagged Features 60
5. Conclusion and Future Directions 62
5.1 Conclusion 62
5.2 Limitation 63
5.3 Future Direction 63
REFERENCES 64

Appendix I x
Appendix II xiv

List of Figures

Figure 3-1: The Study Site, Endana Tea Estate- Kahawatta 22
Figure 3-2: The Weather Station of Endana Field Research station 23
Figure 4-1: Dailly smoothed variable plots 39
Figure 4-2 Decomposition components plot 40
Figure 4-3: Partial Autocorrelation Function of tea yield 41
Figure 4-4: Plot of SARIMA (0,0,0) (1,0,1)12 model 42
Figure 4-5: Time series plot of SARIMA Forecast of Tea Yield 43
Figure 4-6: Correlation matrix for variables 53
Figure 4-7: Actual Values and Forecasted values of tea yield of VAR 55
Figure 4-8: VECM Coefficients 57
Figure 4-9: Actual Values and Forecasted values of tea yield of VECM 58
Figure 4-10: Ensemble Model Forecast of Tea Yield 59
Figure 4-11: Actual and predicted tea yield of Lagged version of forecasting 61

List of Tables

Table 4-1: Summary Statistics of each weather variables and tea yield 38
Table 4-2: Cointegration test 46
Table 4-3: Dickey-Fuller test 48
Table 4-4: KPSS Test 49
Table 4-5: The Variance Inflation Factor (VIF) values 51
Table 4.6: VAR Model Coefficients 54
Table 4.7: coefficient of Ensemble model 59

List of Appendices

Appendix	Description	Page
Appendix - I	Granger Causality Test	x
Appendix – II	VAR Model Summary	xiv
Appendix – III	VECM Summery	xxi
Appendix – IV	Python code	xxiii