2025 SICASH MULTICONFERENCE

International Conference on Actuarial Sciences (ICActS)
Available on-line at: www.sicash.sliit.lk

Predictive Model for Monthly Made Tea Production in Sri Lanka

Charu Subasinghe[1*], Nipul Wattegedara[1], Tharushi Silva[1], Sanduni Balasooriya[1], Kavindi Dassanayake[1], M. L. Guruge[1]

[1]*Department of Mathematics and Statistics, Faculty of Humanities and Sciences, Sri Lanka Institute of Information Technology (SLIIT), Malabe, Sri Lanka*

Corresponding author*: HS23275638@my.sliit.lk

Abstract

This study forecasts monthly tea production in Sri Lanka by developing a suitable time series model to identify future trends in the national tea industry. The analysis is based on monthly made tea production data from January 2000 to June 2025, obtained from the Central Bank of Sri Lanka and the Sri Lanka Tea Board. After confirming the non-stationarity of the original series through the Augmented Dickey-Fuller test, both first-order and seasonal differencing were applied to achieve stationarity. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were used to identify potential model structures. Based on model selection criteria such as Akaike Information Criterion (AIC), Schwarz Criterion (SC), Hannan-Quinn Criterion (HQC), and log-likelihood, the $SARIMA(2,1,1)(1,1,2)_{12}$ model was determined as the best-fitted model. Diagnostic tests confirmed the model's residuals to be white noise, with no serial correlation, normality, or heteroskedasticity. The model was validated using data from early 2025, achieving a Mean Absolute Percentage Error (MAPE) below 10%. Forecasted production values from May to August 2025 provide useful insights for policymakers and stakeholders in agricultural planning and export strategy.

Keywords: SARIMA, Tea production, Time series forecasting, Sri Lanka

Introduction

The industry spans diverse agro-ecological zones—high-grown, mid-grown, and low-grown regions—each influencing the volume and quality of tea produced throughout the year. To gain a deeper understanding of this dynamic sector, the Central Bank of Sri Lanka (CBSL), in collaboration with the Sri Lanka Tea Board, has meticulously compiled a comprehensive dataset (Central Bank of Sri Lanka, 2025). This dataset provides monthly observations of tea production, spanning from January 2000 to June 2025, offering a rich and continuous time series for empirical analysis. This extensive period of data captures significant transitions, including climatic anomalies such as prolonged droughts, economic fluctuations, and various policy reforms, all of which have left their mark on production volumes. The data, officially compiled by CBSL, is considered highly reliable, though minor revisions may occur in recent months. It is updated two months after the reference month.

Tea production in Sri Lanka exhibits distinct seasonal patterns, typically peaking in the first quarter of each year and experiencing a decline during the third quarter. These fluctuations are influenced by a combination of natural factors like rainfall and temperature, and institutional factors such as labor availability, input costs, and government regulations. Despite these varied influences, tea remains a

https://doi.org/10.54389/AJFR6117

critical commodity for both domestic economic planning and international competitiveness (Export Development Board (EDB), Sri Lanka, 2022).

This study endeavors to develop a forecasting model based exclusively on historical tea production data, without incorporating external explanatory variables like weather patterns, global demand, or policy changes. By applying robust time series techniques to this extensive and reliable dataset, the objective is to unravel the intricate dynamics of production and generate accurate forecasts. These forecasts are designed to inform strategic decision-making in key areas, including supply chain management, agricultural planning, and the development of effective economic policies.
Thus, the rresearch question is to what extent can historical time series data, in isolation, accurately capture the seasonal and long-term dynamics of Sri Lanka's tea production and provide reliable forecasts for future planning and policymaking.

Secondary Data

The national level monthly made tea production statistics were acquired from the Central Bank of Sri Lanka and the Sri Lanka Tea Board under the Real Sector classification, encompassing the period January 2000 to June 2025. The data, sourced from formal financial data repositories, incorporates revisions publicized two months following the conclusion of each month. The dataset reflects on the manufacturing patterns across shifting financial situations, such as post-pandemic recovery stages and the impacts of the catastrophic 2022 economic crisis (Sri Lanka Tea Board, 2022). Additionally, insights regarding seasonal fluctuations in output were gleaned from discussions with regional agricultural experts, providing textured context to the statistical series. Government announcements and high-level strategic plans were also surveyed to round out the macroeconomic perspective.

Methodology

The Seasonal Auto Regressive Integrated Moving Average (SARIMA) was used to model and forecast the monthly tea production in Sri Lanka. SARIMA is an extension of the ARIMA model that allows for seasonal autoregressive (SAR), seasonal differencing (SD), and seasonal moving average (SMA) terms. It is utilized in the case that data exhibits regular seasonal variation. The notation for a SARIMA model is SARIMA(p, d, q)(P, D, Q)s, where p, d, and q are the non-seasonal ARIMA parameters; P, D, and Q are the seasonal parameters; and s is the number of periods in a season (for example, s = 12 for monthly data with yearly seasonal periodicity). The general form of the SARIMA model is:

$$\left(\Phi_P(B^S)\phi_p(B)\right)(1-B)^d(1-B^S)^D Y_t = (\Theta_Q(B^S)\theta_Q(B))\epsilon_t \text{ ---------------- } (1)$$

Where Y_t is the observed tea production at time t, $\phi_p(B)$ and $\theta_Q(B)$ are the nonseasonal AR and MA polynomials, $\Phi_P(B^S)$ and $\Theta_Q(B^S)$ [D5.1] are the seasonal AR and MA polynomials, and ε_t is a white noise error term with zero mean and constant variance.

The modelling procedure began by dividing the data into training and testing subsets. The period from January 2000 to October 2024 was used as the training set, while the period from November 2024 to April 2025 was reserved for model validation. The procedure began with testing the series for stationarity using the Augmented Dickey-Fuller (ADF) Test. The series was found to be non-stationary, then non-seasonal differencing and seasonal differencing were performed. For initial identification of the model orders, we used the number of seasonal and non-seasonal lags for which are significant in the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). Several SARIMA models were evaluated, and the best-fitting model was selected based on AIC, SC, HQC, and log-likelihood values.

Out-of-sample forecast accuracy was assessed using RMSE and MAPE, ensuring that predictions were both statistically sound and practically reliable for planning purposes. The analysis was conducted in **E-Views 12**, with the methodology designed to evaluate whether historical tea production data alone can provide accurate forecasts for strategic decision-making in supply chain management, agricultural planning, and policy formulation.

Results and Discussion

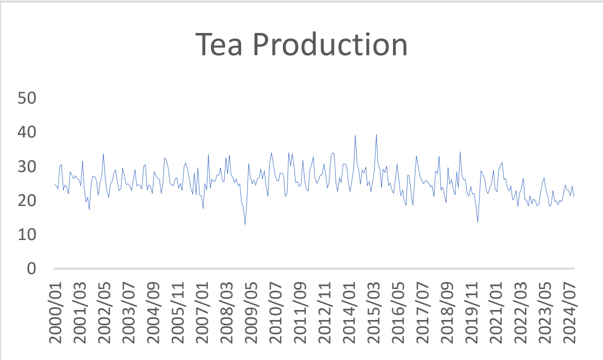


Figure 3: *Temporal Variability of Original Series*

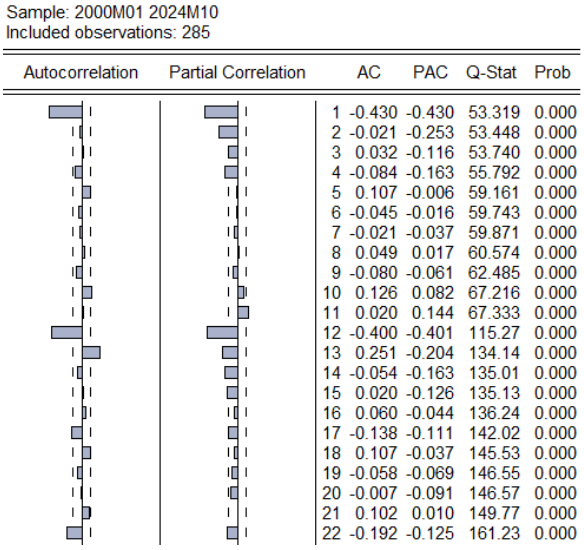
As seen in Figure 1, the temporal variability of the national-level monthly tea production data covers the period up to June 2025. Tea production exhibits considerable variation over time, with frequent fluctuations, indicating high short-term volatility. There is no clear upward or downward trend, indicating that production levels remain relatively constant around an average.

The data varies from a minimum of 12.9 metric tons to a maximum of 39.2 metric tons, with a mean of 25.41 metric tons and a standard deviation of 3.976. The **skewness** of 0.213 suggests a slight rightward skew, while the kurtosis value of 3.509 indicates a distribution slightly more peaked than normal. Since the p-value of the Jarque-Bera test is 0.065 ($p > 0.05$), it indicates that monthly tea production is not significantly deviated from the y normally distribution. The series was tested for the stationarity condition using ADF test. As the corresponding test statistic is not significant it can be concluded with 95% confidence that the original series is non-stationary.

Therefore, the difference series was taken to make the series stationary, and it was found that the p-value (0.000) of the ADF test is less than 0.05, indicating that the series is stationary at 5% level of significance. ACF and PACF plots of the first difference series were taken to confirm the stationarity of the series. However, the presence of statistically significant spikes at specific lags, notably at lag 1 and at seasonal lags like 12, suggests that the differenced series is not completely stationary. This indicates that a more advanced time series model, possibly including seasonal components, will be necessary to capture and forecast the remaining patterns in tea production fully. Therefore, the short-term 1^{st} difference and the long-term 12^{th} difference of the original series are obtained, and the ADF test is carried out. Its correlogram reveals that the short-term 1^{st} difference and the long-term 12^{th} difference of the original series is stationary as the test statistic is significant with a corresponding p value (0.000) less than 0.05.

Possible models

As seen in Figure 2, the comparison between the observed and theoretical ACF and PACF plots led to the identification of five potential SARIMA models. The significance of the 1^{st} and 12^{th} autocorrelations, along with the 1^{st}, 2^{nd}, and 12^{th} partial autocorrelations, suggests the presence of both non-seasonal and seasonal dependencies. Accordingly, the following SARIMA models were considered most suitable: SARIMA (1,1,1) $(1,1,1)_{12}$ SARIMA (2,1,2) $(1,1,1)_{12}$ and ARIMA (2,1,1) $(1,1,2)_{12}$.



Sample: 2000M01 2024M10
Included observations: 285

	AC	PAC	Q-Stat	Prob
1	-0.430	-0.430	53.319	0.000
2	-0.021	-0.253	53.448	0.000
3	0.032	-0.116	53.740	0.000
4	-0.084	-0.163	55.792	0.000
5	0.107	-0.006	59.161	0.000
6	-0.045	-0.016	59.743	0.000
7	-0.021	-0.037	59.871	0.000
8	0.049	0.017	60.574	0.000
9	-0.080	-0.061	62.485	0.000
10	0.126	0.082	67.216	0.000
11	0.020	0.144	67.333	0.000
12	-0.400	-0.401	115.27	0.000
13	0.251	-0.204	134.14	0.000
14	-0.054	-0.163	135.01	0.000
15	0.020	-0.126	135.13	0.000
16	0.060	-0.044	136.24	0.000
17	-0.138	-0.111	142.02	0.000
18	0.107	-0.037	145.53	0.000
19	-0.058	-0.069	146.55	0.000
20	-0.007	-0.091	146.57	0.000
21	0.102	0.010	149.77	0.000
22	-0.192	-0.125	161.23	0.000

Figure 4: *ACF and PACF for the seasonality differenced series*

Selecting the best-fitted model

Table 2: Comparison of the possible models

Model	Log-Likelihood	AIC	SC	HQC
ARIMA $(2,1,1)(1,1,2)_{12}$	-642.15	4.790	4.883	4.828
SARIMA$(2,1,2)(1,1,1)_{12}$	-642.68	4.794	4.887	4.832
SARIMA$(1,1,1)(1,1,1)_{12}$	-651.07	4.824	4.890	4.850

Table 2 indicates that, based on the significance of the parameters, the lowest Akaike info criterion (AIC), the lowest Schwarz criterion (SC), the lowest Hannan-Quinn Criterion (HQC), and the highest log likelihood, among the three possible models, the SARIMA $(2,1,1)(1,1,2)_{12}$ was found to be the best-fitted model. Thus, the best-fitted model can be derived as,

$$(1-0.2695B-0.1610B^2+0.2243B^{12})\ Y_t= -0.0056+(1+0.9995B+0.9989B^{12}+0.3669B^{24})\ \varepsilon_t \quad \text{-----------} \ (2)$$

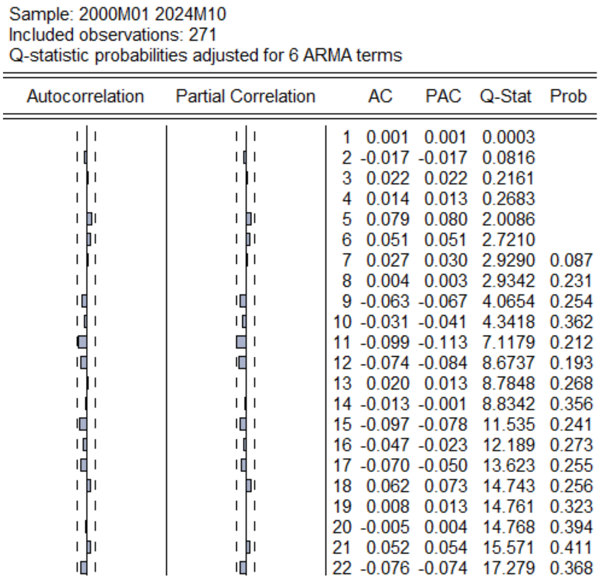
Model Diagnostics

To confirm model stability, the inverse roots of the AR and MA polynomials were examined. It was found that the inverse roots of the AR and MA polynomials lie within the unit circle, confirming that the SARIMA model is both stationary and invertible.

Residual Diagnostics

The p-value of the Jarque-Bera test was found to be 0.065142, which is greater than the 0.05 significance level, indicating that the residuals are approximately normally distributed at the 5% level of significance.
According to Figure 3, the Q statistics of all residual probabilities are not significant at the 5% level of significance, as their p-values are greater than 0.05. Thus, it can be concluded with 95% confidence that

the residuals are random, independent, and identically distributed. Table 3 indicates the results of diagnostic tests carried out to check for serial correlation and heteroskedasticity in the model residuals. The Breusch-Godfrey Serial Correlation LM Test yields a p-value of 0.297, indicating no significant autocorrelation. Similarly, the Heteroskedasticity Test shows a p-value of 0.359, suggesting that the variance of the residuals remains constant. These results confirm that the model satisfies the assumptions of independence and homoskedasticity of residuals.


Sample: 2000M01 2024M10
Included observations: 271
Q-statistic probabilities adjusted for 6 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.001	0.001	0.0003	
		2	-0.017	-0.017	0.0816	
		3	0.022	0.022	0.2161	
		4	0.014	0.013	0.2683	
		5	0.079	0.080	2.0086	
		6	0.051	0.051	2.7210	
		7	0.027	0.030	2.9290	0.087
		8	0.004	0.003	2.9342	0.231
		9	-0.063	-0.067	4.0654	0.254
		10	-0.031	-0.041	4.3418	0.362
		11	-0.099	-0.113	7.1179	0.212
		12	-0.074	-0.084	8.6737	0.193
		13	0.020	0.013	8.7848	0.268
		14	-0.013	-0.001	8.8342	0.356
		15	-0.097	-0.078	11.535	0.241
		16	-0.047	-0.023	12.189	0.273
		17	-0.070	-0.050	13.623	0.255
		18	0.062	0.073	14.743	0.256
		19	0.008	0.013	14.761	0.323
		20	-0.005	0.004	14.768	0.394
		21	0.052	0.054	15.571	0.411
		22	-0.076	-0.074	17.279	0.368

Figure 5: *Correlogram of the residuals*

Table 3: *Diagnostic Tests for Residual Autocorrelation and Heteroskedasticity*

Test	t-statistic	p -value
Breusch-Godfrey Serial Correlation LM Test	1.218	0.297
Heteroskedasticity Test	0.844	0.359

Forecast values for the Training Dataset

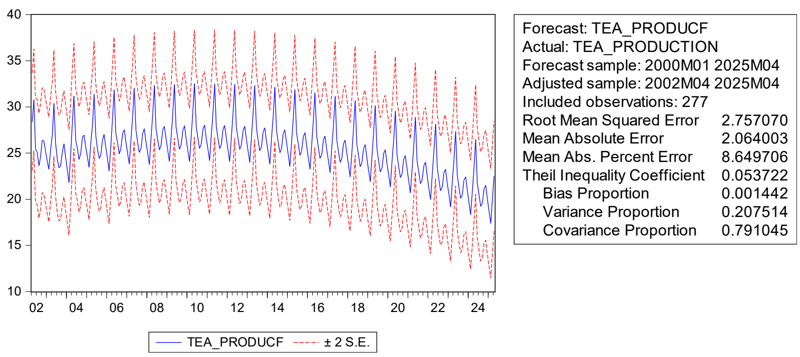


Figure 4: *The Forecast and its 95% Confidence Limits for the Training Dataset*

As seen in Figure 4, the model forecasts tea production with reasonable accuracy for the training data, as indicated by a low Root Mean Square (RMSE) of 2.75 and Mean Absolute Error (MAE) of 2.06, along with an MAPE of 8.64%, suggesting that average forecast errors are within an acceptable range of 10% (Bobbitt, 2021). Also, Theil's U statistic is 0.053 with a minute bias (0.0014). Theil's U being close to zero confirms the model's strong predictive performance (Kolassa, 2021).

Validation for the Testing Dataset

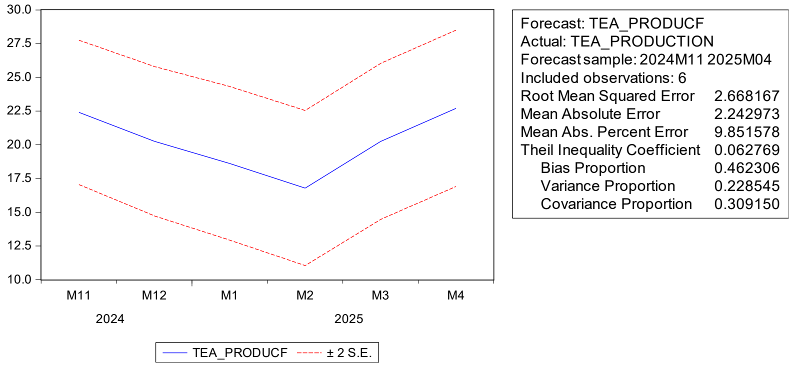


Figure 5: *The Forecast and Its 95% Confidence Limits for the Testing Dataset*

As seen in Figure 5, the model forecasts tea production with reasonable accuracy for the testing data, as indicated by a low RMSE (2.67) and MAE (2.24), along with an MAPE of 9.85%, suggesting that average forecast errors are within an acceptable range of 10% (Bobbitt, 2021). Also, Theil's U statistic is 0.0628 and, being close to zero, confirms the model's strong predictive performance. Although there is some bias present, the overall trend and structure of the data are well captured.

Validation Data with Forecasted Values and Percentage Errors

Table 4: *Validation Data with Predicted Values and Errors*

Date	Actual Tea Production(Metric Tons)	Predicted Tea Production(Metric Tons)	Percentage error
2024/11	22.3	21.511	0.035
2024/12	21.7	20.005	0.076
2025/01	21.5	18.719	0.129
2025/02	15.6	17.396	0.115
2025/03	24.4	20.024	0.179
2025/04	26.4	22.515	0.147

Table 4 indicates that a percentage error below 15% is generally acceptable for real-world forecasting, especially in agriculture, where production varies due to external factors (Hyndman & Athanasopoulos, 2018). This model is reliable for short-term planning but may need slight adjustments to improve accuracy in certain months.

Short-Term Forecasting

Table 5: *Predicted Values for May 2025 to August 2025*

Date	Tea production (Metric Tons)
2025/05	27.646
2025/06	21.973
2025/07	20.101
2025/08	18.487

Table 5 indicates the monthly tea production for the next four months forecast using the fitted model.

Conclusions and Recommendations

The initial analysis, which included a correlogram and the Augmented Dickey-Fuller (ADF) test ($p = 0.2582$), indicated that the original tea production series was non-stationary. To address this, both first-order and seasonal differencing were applied, after which the series was confirmed to be stationary (ADF $p = 0.000$). Among the models assessed, the SARIMA $(2,1,1)(1,1,2)_{12}$ model proved to be the best fit based on the significance of its parameters, as well as the lowest AIC, SC, and HQC values, and the highest log-likelihood. The final model met all key diagnostic requirements, showing stable behavior (all roots within the unit circle), no residual autocorrelation, and consistent variance. It also demonstrated strong forecasting ability, with MAPE under 10% and forecast errors generally below 15% for both the training and validation sets. These forecasts can support policymakers and stakeholders by guiding export planning, labor allocation, and resource management according to expected production trends. Monitoring actual production against these predictions can further improve short-term planning and operational decisions. Based on these results, the SARIMA model is recommended for short-term forecasting of monthly tea production from May to October 2025. However, slightly higher errors observed in March and April suggest that continued monitoring and occasional model refinement may help maintain accuracy. For long-term forecasting or atypical variations, incorporating external factors such as weather, labor availability, or policy changes could further improve accuracy (Tea Exporters Association, 2024), especially for long-term planning and understanding unusual variations in output.

Acknowledgment

Special thanks are extended to the academic staff of the Department of Mathematics and Statistics, including the Head of the Department, for their valuable guidance, continuous support, and encouragement throughout the study. The successful completion of this work was made possible by their expertise and insightful contributions.

References

Bobbitt, Z. (2021, May 10). What is a good MAPE? *Statology*. https://www.statology.org/what-is-a-good-mape/

Central Bank of Sri Lanka. (2025, July 17). *Central Bank of Sri Lanka*. https://www.cbsl.lk/eResearch/Modules/RD/SearchPages/Search_Result.aspx?R=3M2Aaw0Iy5k=

Ceylon Tea. (2025). *Pure Ceylon Tea*. https://www.pureceylontea.com/story/

Cryer, J., & Chang, K.-S. (2008). *Time Series Analysis with Application in R.* Springer Science+Business Media, LLC.

Export Development Board (EDB), Sri Lanka. (2022). *Tea sector: Industry capability report*.

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Kolassa, S. (2021, March 31). Interpretation of Theil's U2 statistic: Forecasting methods and applications. *Stack Exchange*. https://stats.stackexchange.com/questions/345178/interpretation-of-theils-u2-statistic-forecasting-methods-and-applications

Sri Lanka Tea Board. (2022). *Annual report 2022*. https://srilankateaboard.lk/wp-content/uploads/2024/01/Annual-Report-2022.pdf

Tea Exporters Association. (2024). *Tea growing regions*. Tea Exporters Association Sri Lanka. https://teasrilanka.org/growing-regions