Conference Paper No: PF-05

An application of time series techniques to forecast the Open market weekly average retail price of lime in Sri Lanka

R. A. S. Wickramarathne*, M. P. A. T. Wickramanayaka, K. R. T. S. Mahanama and N. V. Chandrasekara

Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka
wickrama_ps17209@stu.kln.ac.lk*

Abstract

Limes are known for their acidic and tangy flavour and are commonly used in cooking, as a garnish, or to add flavour to drinks. The lime market in Sri Lanka is highly volatile, with prices fluctuating significantly on a weekly basis. In this research study, the main objective is to forecast the weekly lime price in Sri Lanka. Even though some research has been conducted on forecasting fruit prices in Sri Lanka, there is currently a lack of research on forecasting lime prices. The weekly price of lime from 1st week of January 2010 to 3rd week of February 2023 was considered for this study (632 observations). The first 600 observations were used as the training set and reserved data were used as the testing set. The time series plot of the weekly lime price of Sri Lanka indicates a slight upward trend and a non-constant variance with a seasonal pattern. The presence of a seasonal pattern motivated the development of a Seasonal Autoregressive Integrated Moving Average (SARIMA) model. When comparing Akaike's Information Criterion (AIC), ARIMA(1,1,2)(0,1,1)$_{[24]}$ generated the minimum AIC value (-1.125469). Assumptions of autocorrelation and heteroscedasticity were not violated and the normality was violated. Although, the performance measures of ARIMA(1,1,2)(0,1,1)$_{[24]}$ were very low, ARIMA(1,1,2)(0,1,1)$_{[24]}$ was identified as the better model with mean absolute error of 40.799, mean absolute percentage error of 7.543, and root mean squared error of 49.793. The results obtained from this analysis would be helpful to mitigate price risks and uncertainties in the lime industry.

Keywords

Forecasting, Lime, Seasonal Autoregressive Integrated Moving Average (SARIMA)

Introduction

The lime industry in Sri Lanka holds historical significance, tracing its roots back to ancient times when Arab traders introduced the fruit to the island. Over the centuries, lime cultivation expanded and became integral to local cuisine, beverages, and traditional Ayurvedic medicine. Beyond culinary and medicinal use, lime's essential oils have gained popularity in cosmetic and fragrance production, contributing to Sri Lanka's exports. Fluctuations in supply and demand impact its price, affecting farmers, traders, and consumers. Forecasting can optimize supply chains, minimize waste, and boost profits. However, despite its significance, no predictive models exist for lime prices in Sri Lanka. This study aims to bridge this gap by identifying a suitable time series model to forecast the weekly average retail price of lime per kilogram.

The importance of forecasting the price of tomatoes due to their high nature in perishability and seasonality has identified by Mathenge Mutwiri, 2019, Adanacioglu & Yercan, 2012 and Reddy, 2019. All three studies have employed the seasonal autoregressive integrated moving average (SARIMA) model. In 2019, Perera et al. have imputed missing values using mean imputation, linear interpolation, log-linear interpolation, and exponential smoothing in their study on seasonal ARIMA model to forecast monthly potato yield in Sri Lanka. The better technique to explore missing values has been identified as the technique that generates the least root mean squared error (RMSE) and mean absolute percentage error (MAPE) by them. Another study (Mathenge Mutwiri, 2019) has identified a better model using the minimum Akaike Information Criterion (AIC). Performances of the models have been compared using root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). Aryani et al., 2018 modelled ARIMA(1,1,1) with variable exchange rate to predict the effect of the predictor variable on Islamic bank profitability even though the assumption for normality was violated. Their paper claims that the unfulfilled assumption for normality has proved the high volatility in data.

Methodology/materials and methods

Data were collected from the Department of Census and Statistics. It consisted of 632 total observations from 1st week of January 2010 to 3rd week of February 2023 with six missing values. The first 600 observations and the remaining 32 observations were utilized to build the model and evaluate the performance respectively.

Missing value imputation techniques

Missing value imputation techniques involve predicting unknown data points using existing data in a series. Some common methods include linear and log-linear imputation, more advanced techniques like Catmull-Rom and Cardinal Spline, and cubic spline interpolation.

Stationarity series

Stationarity in a time series implies that its mean, variance, and covariance remain constant over time, with the covariance between time periods depending solely on the lag. Tests like the Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) are utilized to assess stationarity.

Transformations

Transformations such as differencing, seasonal differencing, and log transformation are employed to convert non-stationary time series into stationary ones by mitigating trends and seasonality.

Seasonal ARIMA

Seasonal ARIMA, which we denote as ARIMA(p,d,q)(P,D,Q)$_{[S]}$, is the product of the two polynomials generated by the (p,d,q) ARIMA model and the (P,D,Q)$_{[S]}$ ARIMA model (Equation 1).

Equation 1. *SARIMA model and its components*

ARIMA(p, d, q)x(P, D, Q)$_{[s]}$
p - non-seasonal AR order
d - non-seasonal differencing
q - non-seasonal MA order
P - seasonal AR order
D - seasonal differencing
Q - seasonal MA order

S – a time span of repeating seasonal pattern

<u>Model diagnostic</u>

Model diagnostic involves checking for heteroscedasticity, autocorrelation, and normality to assess a statistical model's assumptions and quality.

<u>Model performance</u>

Adequate models are used to generate predicted values for the reserved data (test set). The performance of those adequate models is measured by comparing root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The model with the least errors is identified as the better-performing model.

Results and Discussion

The dataset was divided into five portions due to containing six missing values. Once imputing the missing values using five imputation techniques the RMSE and were calculated. The below table represents the calculated RMSE values and MAE for each interpolation technique.

Table 2. *Performances of missing value imputation techniques*

	Linear	Log-Linear	Catmull-Room Spline	Cardinal Spline	Cubic Spline
RMSE	31.1641	32.5418	32.1810	31.8065	34.6068
MAE	23.9685	24.6768	25.3731	25.0146	29.0654

Considering the minimum RMSE and MAE values, the Linear interpolation technique was identified as the better imputation technique. Therefore, six missing values in the dataset were imputed using the Linear interpolation technique.

The training set consisted of observations from 1st week of January 2010 to 3rd week of June 2022 (600 data points). The remaining observations up to 3rd week of February 2023 were considered as a test set (32 data points).

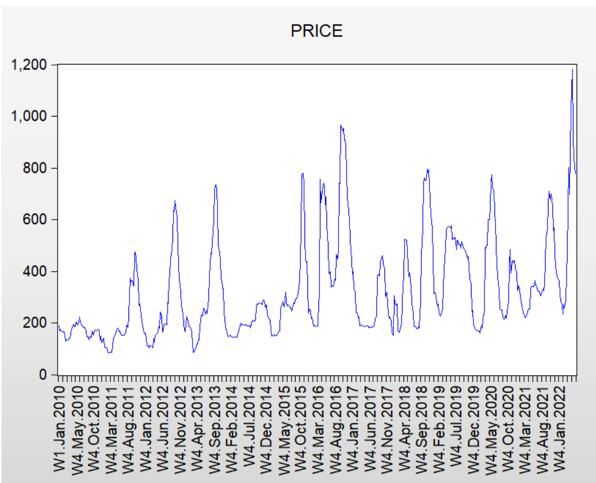


Figure 2. *Time series plot of weekly lime price*

Figure 1 illustrates that, the original lime price series depicts a slight upward trend and seasonal pattern. Three-unit root tests were used to ensure the seasonality of the series.

According to Table 2, ADF and PP tests indicated that the series is stationary while the KPSS test indicated that the series is not stationary at a 5% level of significance. Since the KPSS test is the most robust test for checking stationarity compared to ADF and PP tests (Afriyie et al., 2020), several transformation techniques were applied to make the series stationary.

Table 3. *Checking for stationarity of the original series*

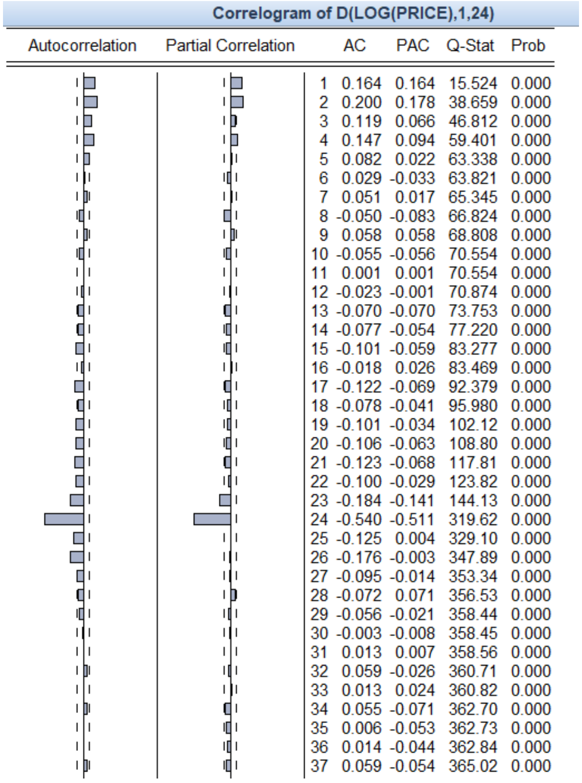
Test	**Criteria**	**Value**
ADF	p-value	0.0000
PP	p-value	0.0002
KPSS	Test statistic	1.2583

Difference transformation was used to eliminate the trend component and log transformation was used to reduce the high variance in the series. Further, the seasonal difference was utilized by identifying the seasonal lag considering the Webel-Ollech (WO) test. According to the WO test, it could be identified that the seasonality of 1^{st} difference log series of lime price is 24. Afterwards, stationarity was investigated again using unit root tests.

Table 4. *Checking for stationarity of d(log(price),1,24)*

Test	**Criteria**	**Value**
ADF	p-value	0.0000
PP	p-value	0.0000
KPSS	Test statistic	0.01252

All three-unit root tests indicated that the series which has applied seasonal difference, first difference and log transformation is stationary at a 5% level of significance.



Correlogram of D(LOG(PRICE),1,24)

	AC	PAC	Q-Stat	Prob
1	0.164	0.164	15.524	0.000
2	0.200	0.178	38.659	0.000
3	0.119	0.066	46.812	0.000
4	0.147	0.094	59.401	0.000
5	0.082	0.022	63.338	0.000
6	0.029	-0.033	63.821	0.000
7	0.051	0.017	65.345	0.000
8	-0.050	-0.083	66.824	0.000
9	0.058	0.058	68.808	0.000
10	-0.055	-0.056	70.554	0.000
11	0.001	0.001	70.554	0.000
12	-0.023	-0.001	70.874	0.000
13	-0.070	-0.070	73.753	0.000
14	-0.077	-0.054	77.220	0.000
15	-0.101	-0.059	83.277	0.000
16	-0.018	0.026	83.469	0.000
17	-0.122	-0.069	92.379	0.000
18	-0.078	-0.041	95.980	0.000
19	-0.101	-0.034	102.12	0.000
20	-0.106	-0.063	108.80	0.000
21	-0.123	-0.068	117.81	0.000
22	-0.100	-0.029	123.82	0.000
23	-0.184	-0.141	144.13	0.000
24	-0.540	-0.511	319.62	0.000
25	-0.125	0.004	329.10	0.000
26	-0.176	-0.003	347.89	0.000
27	-0.095	-0.014	353.34	0.000
28	-0.072	0.071	356.53	0.000
29	-0.056	-0.021	358.44	0.000
30	-0.003	-0.008	358.45	0.000
31	0.013	0.007	358.56	0.000
32	0.059	-0.026	360.71	0.000
33	0.013	0.024	360.82	0.000
34	0.055	-0.071	362.70	0.000
35	0.006	-0.053	362.73	0.000
36	0.014	-0.044	362.84	0.000
37	0.059	-0.054	365.02	0.000

Figure 3: Correlogram of d(log(price),1,24)

According to the correlogram (Figure 2), the significant seasonal and non-seasonal lags are identified as below.

Considering ACF cutoff lags, 1 was identified as seasonal lag while 1,2,3 and 4 were identified as non-seasonal lags. Considering PACF cutoff lags, 1,2 and 3 were identified as seasonal lags while 1 and 2 were identified as non-seasonal lags.

Twenty-four seasonal ARIMA models were identified as candidate models. Once fitting those models, the suggested models were reduced to five models due to the non-significance of the coefficients of the parameter terms. Table 4 depicts the reduced seasonal ARIMA models and the AIC values.

***Table 5.** AIC values of significant models*

Model	AIC
ARIMA(1,1,0)(01,1)$_{[24]}$	-1.098958
ARIMA(1,1,1)(0,1,1)$_{[24]}$	-1.119620
ARIMA(1,1,2)(0,1,1)$_{[24]}$	-1.125469
ARIMA(2,1,1)(0,1,1)$_{[24]}$	-1.125249
ARIMA(0,1,0)(0,1,1)$_{[24]}$	-1.092755

According to Table 4, ARIMA (1,1,2)(0,1,1)$_{[24]}$ was selected as the better model since it represented the minimum AIC value. Afterwards, the three assumptions of autocorrelation, heteroscedasticity and normality were checked, and Table 5 summarizes the results.

***Table 6.** Results of the adequacy checking of the fitted model*

Assumption	Test	p-value	Decision
Heteroscedasticity	ARCH test	0.2418	satisfied
Autocorrelation	Ljung-Box Q test	p-values are greater than 0.05	satisfied
Normality	Jarque-Bera	0.0000	Not satisfied

According to the previous literature (Aryani et al., 2018), most of the time series violates the normality assumption. But forecasting can be done considering the fitted model.

Hence, testing set values were forecasted using ARIMA (1,1,2)(0,1,1)[24] to evaluate the performance of the fitted model.

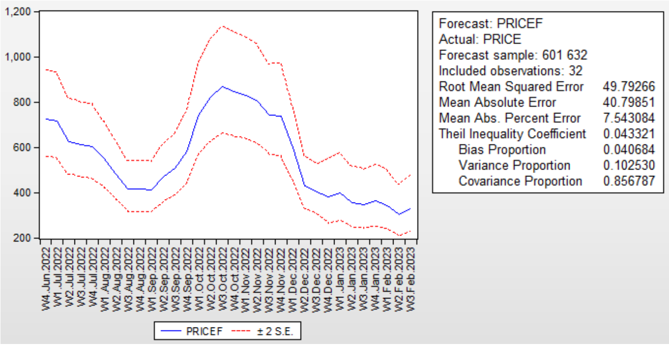


***Figure 3.** Performance measures of the fitted model*

According to Figure 3, MAE is 40.799 and MAPE is 7.54. Therefore, ARIMA (1,1,2)(0,1,1)[24] model can be suggested as a better forecasting model to forecast weekly price of lime in Sri Lanka.

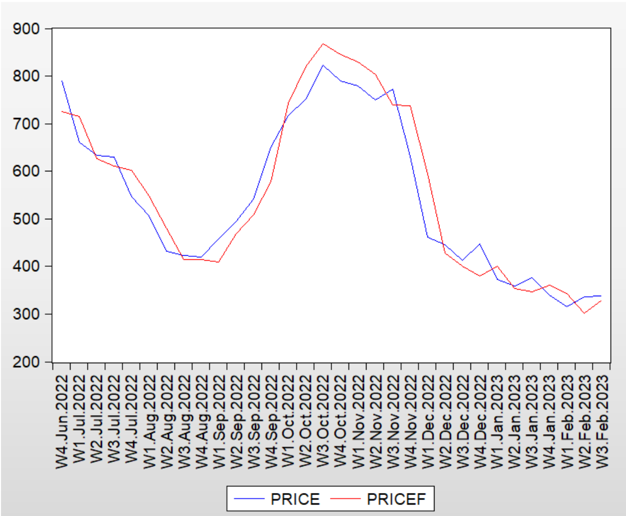


***Figure 4.** Actual vs Fitted graph of ARIMA(1,1,2)(0,1,1)[24] model*

Figure 4 depicts the extracted actual value (Blue colour) vs forecasted value (Red colour) graph of the ARIMA (1,1,2)(0,1,1)[24] model for the test set. It can be observed that the fitted model has well captured the pattern of actual data. Moreover, implication of the violated normality assumption and inclusion of other influential factors will be continued in further analysis.

Conclusion

Lime is a very important citrus fruit for Sri Lanka due to its economic and agricultural significance. The lime market in Sri Lanka is highly volatile, with prices fluctuating significantly on a weekly basis. Therefore, this research identified an univariate time series model, ARIMA (1,1,2)(0,1,1)[24] as the better model for forecasting weekly lime prices in Sri Lanka. The performance of this model is based on Mean Absolute Error (40.799), Mean Absolute Percentage Error (7.54), and Root Mean Squared Error (49.7926). The assumptions of heteroscedasticity and autocorrelation among residuals were absent in the suggested model while the assumption for normality in residuals was not fulfilled. The lack of support for the normality assumption demonstrated considerable volatility in data. For further analysis, some factors such as weather conditions, pest outbreaks, and government policies were identified as factors that affect the weekly lime price in Sri Lanka. Moreover, data mining techniques such as Feed Forward Neural Network (FFNN) and Time Delay Neural Network (TDNN) can be employed to deliver a better model. This model can be beneficial in businesses to better manage their supply chains by allowing them to plan their purchases and deliveries more efficiently.

References

Adanacioglu, H., & Yercan, M. (2012). An analysis of tomato prices at wholesale level in Turkey: An application of SARIMA model. *Custos e Agronegocio*, *8*(4), 52–75.

Aryani, S., Aidi, M. N., & Syafitri, U. D. (2018). Analysis of The Profitability of Islamic Banking Using Arimax Model and Regression with Arima Errors Model. *International Journal of Scientific Research in Science, Engineering and Technology*, *4*(8), 49–53.

Limes: A Citrus Fruit with Powerful Benefits. (n.d.). Healthline. Retrieved February 16, 2023, from https://www.healthline.com/nutrition/limes

Mathenge Mutwiri, R. (2019). Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya: Time Series Analysis Using Sarima Model. *International Journal of Statistical Distributions and Applications*, *5*(3), 46. https://doi.org/10.11648/j.ijsd.20190503.11

OPEN MARKET RETAIL PRICES DASHBOARD. (n.d.). Department of Census and Statistics Sri Lanka. Retrieved February 16, 2023

Afriyie, J. K., Twumasi-Ankrah, S., Gyamfi, K. B., Arthur, D., & Pels, W. A. (2020). Evaluating the performance of unit root tests in single time series processes. *Mathematics and Statistics*, *8*(6), 656–664. https://doi.org/10.13189/ms.2020.080605