Estimating Retail Price Movements of Upcountry Vegetables Using Time Series Analysis

I.J.A. Ruhunuge
Data Management Division
Hector Kobbekaduwa Agrarian Research and Training Institute
Colombo, Sri Lanka
isuriruhunuge999@gmail.com

A.W. Wijeratne
Department of Agribusiness Management
Faculty of Agricultural Sciences
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka

M. Esham
Department of Agribusiness Management
Faculty of Agricultural Sciences
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka

S.P. Fernando
Data Management Division
Hector Kobbekaduwa Agrarian Research and Training Institute
Colombo, Sri Lanka

I.V. Kuruppu
Department of Agribusiness Management
Faculty of Agriculture and Plantation Management
Wayamba University of Sri Lanka
Kuliyapitiya, Sri Lanka

R.M.V.Y. Rathnayake
Data Management Division
Hector Kobbekaduwa Agrarian Research and Training Institute
Colombo, Sri Lanka

***Abstract*—It is essential to build a statistically equipped market information system and a comparative study to identify price fluctuations in upcountry vegetables. That systematic information on prices was important for producers, consumers, suppliers, the government, and international entities. This study aims to analyze retail price movements of upcountry vegetables and forecast prices using ARIMA time series techniques. Hence, this study used five main upcountry vegetables, including carrots, green beans, leeks, tomatoes, and beetroots, to evaluate retail price behavior. For this study, 13 years (2010–2023) of monthly wholesale prices were gathered. The best SARIMA models were ARIMA (1,1,1) (0,0,1) for carrots, ARIMA (2,1,1) (0,0,1) for beans, ARIMA (2,1,0) (0,0,1) for leeks, (1,1,2) (0,0,1) for tomatoes and ARIMA (1,1,1) (0,0,1) for beetroot.**

Keywords—*ARIMA, Up-country vegetables, Price fluctuation, SARIMA*

I. INTRODUCTION

Vegetables are perishables that have a price variation according to supply and demand. Vegetable prices directly impact residents' quality of life and farmers' income, which in turn influences the growth of Sri Lanka's vegetable sector and the country's overall economic balance. It is essential to forecast vegetable prices during both the harvesting season and the off-season for growers to make wise production decisions [1]. Forecasts of food commodity prices are essential for economic policy formulation, as agricultural price stability measures are crucial for breaking the vicious cycle of poverty and food insecurity in developing countries [2]. Inadequate availability of agricultural commodities results in price fluctuations and places a burden on consumers, whereas an excess of agricultural products leads to a decline in vegetable prices and causes financial setbacks for farming households [3]. Hence it is challenging for the government to create policies that adequately address the competing interests of farmers and consumers due to the imbalance in the supply and demand of agricultural products. Moreover, selecting a forecasting method to predict future prices will help policymakers and farmers to make the correct decision. Hence this study aims to analyze retail price movements of upcountry vegetables and forecasts of prices using ARIMA time series techniques.

The presence of price data on agricultural commodities aids private and public organizations in expanding their market. Commodity price analysis has made use of many forms of the autoregressive integrated moving average (ARIMA) model. The pricing behavior of nonperishable goods in South Asia has been studied using both simple and seasonal ARIMA models (SARIMA) [4]. This study used seasonal ARIMA models to predict up-country vegetable prices. The success of this decision-making affects the variation in the supply of food commodities and it ultimately results in price fluctuations.

Therefore, selecting a forecasting method to predict future prices will help farmers to make the correct decision. Understanding pricing patterns will help entrepreneurs make more informed decisions about their businesses and investments. The whole public as consumers benefit from the price information as well. Because agricultural products are essential for daily use, they constitute a significant portion of the market. The prices of these products have a substantial impact on both consumer spending and the income of agricultural households. [5]. The accuracy of the estimated price movements must depend on the quality and reliability of the data used and the appropriateness of the chosen forecasting model.

II. MATERIALS AND METHODS

The study made use of two-time series techniques such as Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). This study involved the development of mathematical models to analyze the monthly prices of up-country vegetables from 2010 to 2023. Price data were collected from the Agrarian Research and Training Institute of Sri Lanka. of upcountry vegetables including carrots, green beans, leeks, tomatoes, and beetroots in the Colombo city region The main data analyzed for the development of the models were the monthly retail prices markets. In the ARIMA modeling approach, lag and the first difference were drawn to utilize the raw data to transform data to stationary, if not the ARIMA models will not be suited with raw data.

The Autoregressive Integrated Moving Average (ARIMA) model is a statistical approach used for analyzing and predicting time series data. It consists of three components: p, d, and q. The parameter p represents the auto-regression, which indicates the lag order. The parameter d represents the degree of differencing, and the parameter q represents the order of the moving average.

ARIMA models are widely used in various fields, such as finance, economics, and forecasting. They utilize past values of a time series to predict future values and can help understand the underlying patterns and trends in the data

The ARIMA model can be extended to the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, denoted as ARIMA (p, d, q) × (P, D, Q)[S]. In SARIMA, the additional parameters P, Q, and D represent the seasonal autoregressive order, seasonal moving average order, and seasonal differencing order, respectively. It is important to thoroughly analyze the time series data before fitting a SARIMA model . Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyper parameters to specify the auto-regression (AR), differencing (I), and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of seasonality.

Then two correlograms of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were used to manually specify an ARIMA model as it gives a clue to find the p, d, and q values. The ACF plot expresses how far the present value is related to its later values while the PACF plot describes the correlation between the time series variable and its lags.

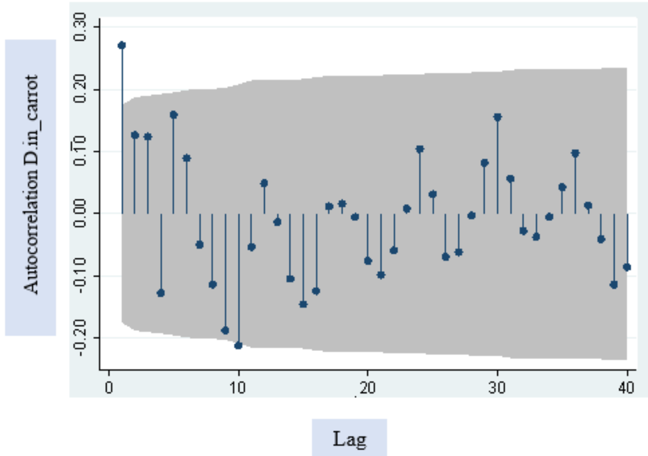
The Ljung and Box "Q" statistics are used to assess whether the autocorrelations of the errors in a time series are significantly different from zero. This test helps determine if there is any remaining autocorrelation in the residuals of a model. Augmented Dickey-Fuller (ADF) test is commonly employed to examine the stationarity property of a time series. It helps determine if a series is stationary or if it requires differencing to achieve stationarity [6]. All the analyses in this study were carried out using the Stata 5.0 package. The Ljung Box test was used to check the accuracy of the proposed model. The likelihood, Akaike information criterion (AIC), and Bayesian Information Criteria (BIC) were used to select the best-fitted model from the proposed models.

III. Results and Discussion

According to the time series plot, there was a slight upward trend, non-constant variance, and a clear seasonal pattern in monthly retail prices of carrots, beans, tomatoes, beetroot, and leeks were observed. As the raw data was non-stationary, transformed them into a stationary time series through one-degree differencing. After completing the appropriate procedures (Fig. 1), a more appropriate mathematical model for estimating the retail price of each commodity was derived.

As mentioned in the below flow chart (Fig. 2), firstly Augmented Dickey-Fuller test was performed and it showed a unit root in the time series which denoted the data set was non-stationary. The first differentiation was taken to make the data stationary. Then plotted Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were employed to identify the order of auto-regression (AR) and moving average (MA).

(A)



(B)

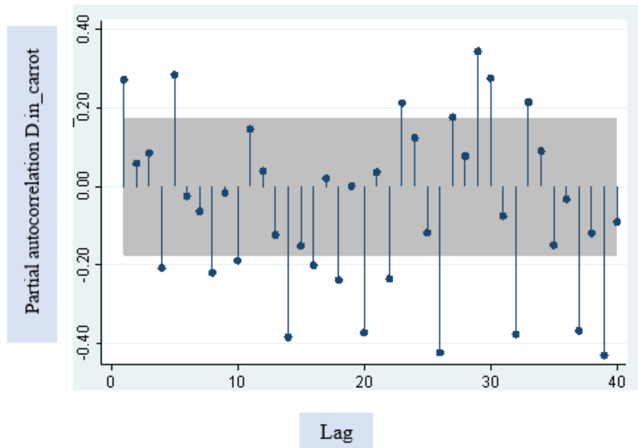


Fig 1. shows (A) the Autocorrelation graph of the carrot (B) the Partial autocorrelation graph of the carrot

The best models were ARIMA (1,1,1) (0,0,1) for carrots, ARIMA (2,1,1) (0,0,1) for beans, ARIMA (2,1,0) (0,0,1) for leeks, (1,1,2) (0,0,1) for tomatoes and ARIMA (1,1,1) (0,0,1) for beetroot.

As mentioned in Tab. 1, the results of the ARIMA model predictions indicated that the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for the models of five vegetables were all close to zero. Additionally, the Mean Absolute Percentage Error (MAPE) was found to be less than 30%. When comparing the ARIMA models for the prices of the five vegetables, it was observed that the carrot and tomato price models provided the most accurate predictions. The absolute percentage error for this model was within 5%, and it demonstrated a high level of accuracy in simulating the true values.

In conclusion, a variety of factors interact to affect vegetable pricing along the whole supply chain, and changes in these aspects in the future are uncertain. The price forecast trend line that is closest to the actual value, namely the ARIMA forecast curve, should be chosen as the reference when vegetable producers make production and planting plans as a result, helping producers decide whether to increase, decrease, or keep the output unchanged and lowering the production risk of vegetables with perishable characteristics.

IV. CONCLUSION

The best SARIMA models were ARIMA (1,1,1) (0,0,1) for carrots, ARIMA (2,1,1) (0,0,1) for beans, ARIMA (2,1,0) (0,0,1) for leeks, (1,1,2) (0,0,1) for tomatoes and ARIMA (1,1,1) (0,0,1) for beetroot.

When vegetable producers are making production and planting plans, it is important for them to consider the price forecast trend line that closely aligns with the actual value. In this case, the ARIMA forecast curve serves as a reliable reference. It is evident that all five main upcountry vegetables experience significant fluctuations throughout the year.

Price forecasts should be communicated among vegetable growers using all media platforms.

REFERENCES

[1] Yoo, D. Developing vegetable price forecasting model with climate factors. Korean J. Agric. Econ, 275, 2831–2844, 2016.

[2] Illankoon, I. M. G. L., & Kumara, B. T. G. S. Analyzing the Influence of Various Factors for Vegetable Price using Data Mining. 13th International Research Conference. General Sir John Kotelawala Defence University, 2020.

[3] Xiong, T.; Li, C.; Bao, Y. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China Neuro computing.275, 2831–2844, 2018.

[4] Reddy, A. A. Price forecasting of tomatoes. International Journal of Vegetable Science, 25(2), 176–184, 2019.

[5] FAO. 2018 FAOSTAT OlineDatabase. Available online: http:// www.fao.org / faostat/en/#data (accessed on 7December 2020).

[6] Xu, X.,Corn cash price forecasting. American Journal of Agricultural Economics, 102(4), 1297–1320. doi:10.1002/ajae.12041,2020.

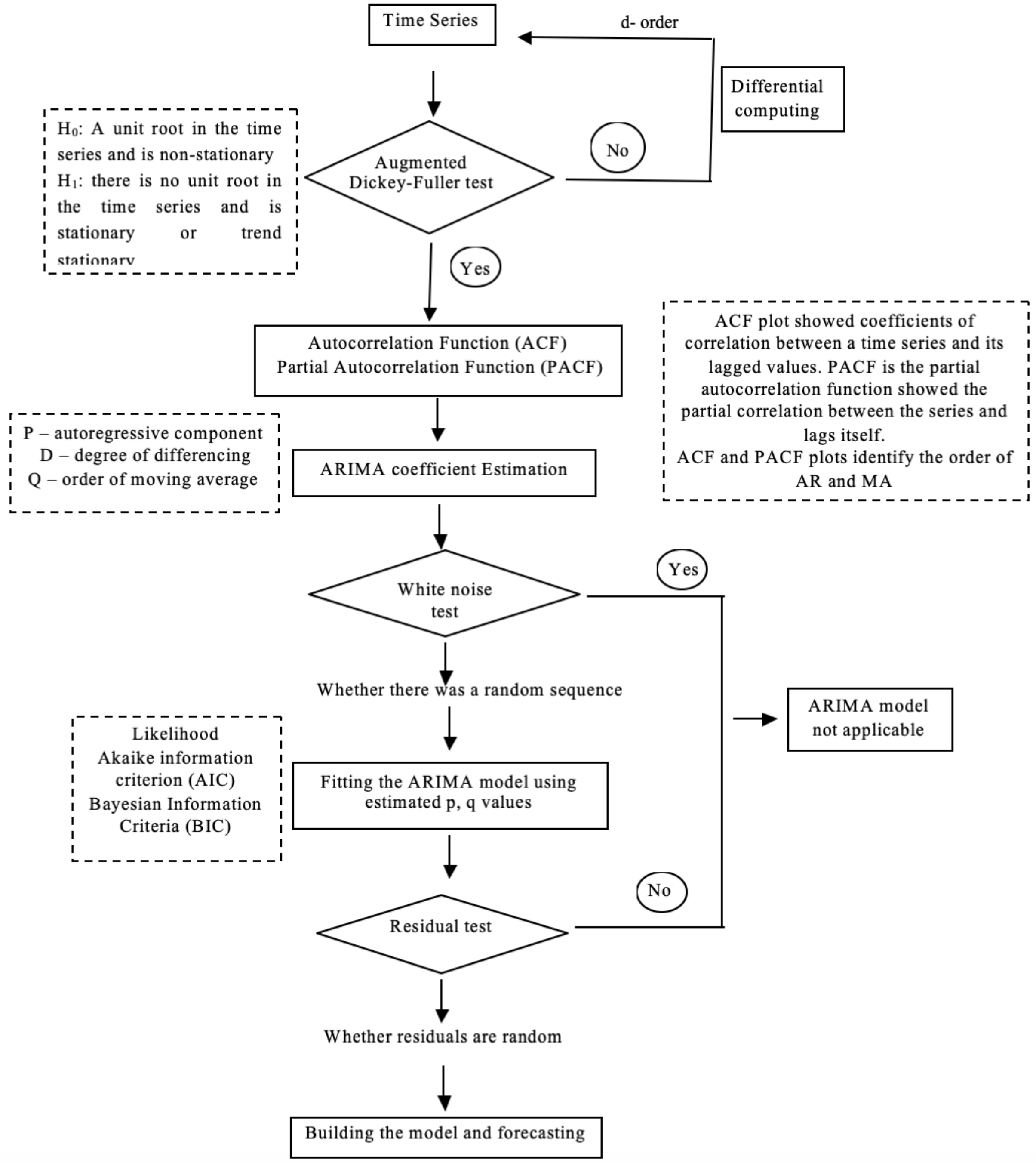


Fig. 2. ARIMA model flowchart for price prediction

TABLE 1. EFFECTIVENESS OF AVERAGE VEGETABLE PRICE FORECASTS (*SOURCE: AUTHORS*)

		MSE	*RMSE*	*MAE*	*MAPE%*
Carrot	ARIMA (1,1,1) (0,0,1)	0.23	0.14	0.12	4.8
Beans	ARIMA (2,1,1) (0,0,1)	0.33	0.58	0.61	29.5
Tomato	ARIMA (1,1,2) (0,0,1)	0.12	0.11	0.14	4.1
Leeks	ARIMA (2,1,0) (0,0,1)	0.27	0.46	0.48	13.2
Beetroot	ARIMA ((1,1,1) (0,0,1)	0.48	0.68	0.71	18.6

* Average error values in percentage terms. Values at least 0.15 equate to 15% error was a better fit. Smaller values indicate better fit.

* Squared errors are based on the square of the differences between the fitted values and the observed values. It's similar to a standard deviation value. Smaller values indicate a better fit.