Modelling Open Market Retail Price of Red Onions in Colombo using ARIMA-GARCH Mixed Model

MGA Saumyamala[1#], WPMCN Weerasinghe [2], JLSM Kumara [3], SAL Sachithra[4] and NV Chandrasekara[5]

[1]*Department of Statistics & Computer Science,*
University of Kelaniya,
Sri Lanka

[#] MGA Saumyamala; <avanthismm321@gmail.com>

Abstract— *Red onion is an important commodity in Sri Lankan culture which subject to high frequent fluctuations in retail price due to government policies, trade agreements and weather conditions like heavy rainfall. Objective of this study is to find more accurate time series model to forecast future prices of red onions. This study considers weekly average retail prices (WARP) of red onions in Colombo main markets from January 2014 to April 2019. Several models were fitted and based on model selection criterions, ARIMA(1, 1, 1) was identified as the best model. As the residuals of the model were heteroscedastic, ARCH(9) and GARCH(9,1) models were fitted. According to the literature, WARP of red onion price shows drastic increase in 2017 as a result of production fall. Thus, using change point analysis, series was divided into 3 windows and ARIMA(0,1,0) model was suggested as the best model for each window. Finally using all four models; ARIMA(1,1,1),ARIMA(1,1,1)+ARCH(9), ARIMA(1,1,1)+GARCH(9,1)and ARIMA(0,1,0) price was forecasted for the year 2019 using two methods; static and dynamic forecasting. Forecasting accuracy of the models measured using root mean squared error (RMSE). As a conclusion ARIMA(1,1,1)–GARCH(9,1) model was chosen as the most suitable model with 6.26 RMSE to forecast WARP of red onions. Though there are several studies carried out on behaviour of red onion price in Sri Lanka, no specific model was suggested so far. Therefore this model can be used by the cultivators, intermediaries and government in decision making on production quantity, pricing and import/export regulations.*

Keywords— **Red Onion Price, ARIMA, GARCH, Heteroscedasticity, Change point**

I. INTRODUCTION

Onion is a vital condiment included in most of the Sri Lankan cuisine with a high value. There are two types of onions cultivated in Sri Lanka; red onions and big onions which are close substitutes. There are two varieties of red onions; Sinnan and Vedalan. Red onions are not only a condiment, it is also used as a vegetable and medicine. Therefore, it is considered as a crop with high consumption. National requirement of the red onion is about 100,000 Mt/ year. 90% of the requirement of red onions is supplied from the local market and only 10% is imported. Red onions are mainly cultivated in few agro-ecological regions in the dry zone due to specific requirements in environmental conditions. Major growing areas are Jaffna, Puttalam and Trincomalee. Red onion is considered as a seasonal crop and usual cropping season is the period between May to August. But in Puttalam and Kalpitiya regions red onion farming continues throughout the year(Department of Agriculture., no date).

Currently the red onion farmers are well protected from trade policies and promote the local cultivation under continuous surveillance. But still there exist uncontrollable risks. Red onions are very sensitive to climate, water and other agricultural factors and their changes affect the cultivation. There is a risk of subjecting stocked harvest for floods, plant diseases, and fungusand become non consumable. In such cases supply falls below the demand resulting considerable fluctuations in red onion prices.

In Sri Lanka, detail analysis of retail price of major agricultural commodities was done by Agrarian Research and Training Institute. It shows that, due to germination problem farmers tends to sell their entire harvest soon after harvesting. Due to this reason red onion price fluctuate between very high index in June to very low index by October (Bogahawatta, 1987).Apart from this regular fluctuation,when observing the red onion price over the years it can be clearly observed that the price in 2017 was increased up to twice as other years.

Table 1. Red onion cultivation progress in 2017

Season	Targeted Extent (ha)	Cultivation Progress		Expected Production (Mt)
		Extent (ha)	% of the Target	
Maha	4914	1610	33	17321
Yala	4,475	721	16	8,629

Source: Socio-economic & Planning Centre/DOA MFPAD/HARTI

According to Table 1, in Maha season only 33% of the targeted extent was cultivated and in Yala season it further reduced to 16% (Institute, 2018a).As a result, red onion supply decreased in the markets below the demand followed by increase in retail price of vedalan by about Rs.40.00/kg. Further stocks of sinnan were not available at the retail markets. As per the Table 2, compared to the same period of lastyear, retail price of vedalan increased by about 99%.(Institute, 2018b)

Table 2. Retail prices of red onions in May 2017

Crop	Average (Rs/Kg)			Change compared to (%)	
	May 2017	April 2017	May 2016	April 2017	May 2016
Red Onion(Vedalan)	331.70	291.3	166.64	13.87	99.05

Source: Marketing, Food Policy and Agribusiness Division/HARTI

This high volatile nature of red onion emphasized the necessity of identifying a statistical model that is capable in price forecasting for future with acceptable accuracy.

Many researchers have used time series approach in modelling red onion prices and other commodities as well. Onion price in Hubali market of Northern Karnataka was modelled using ARIMA model and the model (1,1,1) (2,1,1) was suggested as the best model(Jalikatti,V.N,&Patil, 2015). In Sri Lanka case study was done for red onion prices in Jaffna district to identify the dynamic behavior. This study has estimated trend, volatility and the distributions of red onion pricesand emphasize the significance of modelling of weekly red onion prices.(Sivarajasingham and Mustafa, 2011)

II. METHODOLOGY

Considering the existing background of red onion market, an advanced time series approach was used to model the red onion price. E-views and R statistical software were used throughout the study.

A. *Data Collection*

Secondary data was captured from the official web site of Department of Census and Statistics Sri Lanka under the inflation and prices category. From available 254 records, average retail prices of 240 records from 1st week of year 2014 to 48th week of year 2018 were used to develop the model. The rest, from 1st week of year 2019 to 14th week of year 2019 was set aside to test the accuracy of the model.

B. *Preliminary Analysis*

In the data set there were no missing values and few outliers were found. To maintain continuity of the series original data set was used for further analysis. Most of the statistical models and econometrics models are based on the assumption that data series is stationary. Therefore, it is necessary to identify whether series is stationary or not and if not, should be converted to stationary series.

1) Stationary Time Series: A time series considered to be stationary if sample mean, sample variance and autocorrelation are not different over time period. The stationarity of a time series is tested using Augmented Dickey Fuller (ADF) test and Phillips-Perron (PP) unit root tests where hypothesis is;

H_0: The series possess unit root (series is not stationary)

H_1: The series do not possess a unit root (series is stationary)

Since the series we considered consist of upward trend, the series was not stationary. Thus, ‘Differencing’ technique was used to convert the series into a stationary series.

2) Differencing: Differencing is removing trend by applying the difference operator to the original time series to obtain a new time series, say,

$$X_t = Y_t - Y_{t-1} = \nabla \qquad (1)$$

where ∇ is the (backward shift) difference operator.

C. *Univariate Time Series Analysis*

Univariate time series consist of sequence of measurements of same variable recorded over equal time increments. In this study WARP of red onions modelled using an ARIMA model.

1) Auto Regressive Integrated Moving Average (ARIMA) process: Let Zt be a discrete purely random process with mean zero and varianceσ_z^2.Then,

$\{Xt\}$ is said to be an autoregressive process of order p, AR (p) if,

$$X_t = \delta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots . + \alpha_p X_{t-p} + \qquad (2)$$

$\{Xt\}$ is said to be a moving average process of order q, MA (q) if,

$$X_t = \mu + Z_t - \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \cdots . + \beta_p Z_t \qquad (3)$$

The process formed by combining AR (p) and MA (q) processes is called mixed autoregressive-moving average process ARMA (p, q)

$$X_t = \delta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots . + \alpha_p X_{t-p} + Z_t - \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \cdots . + \beta_p Z_{t-p} \quad (4)$$

When the series is differenced by order of d, the ARMA process becomes Autoregressive Integrated Moving average processARIMA (p, d, q);

$$\varphi(B)(1 - B)dX_t = \delta + \theta(B) \quad (5)$$

Where ϕ(B) is the AR polynomial and θ(B) is the MA polynomial. B is the backshift operator.

D. Heteroscedasticity

One key assumption of residuals is that the variance should be constant across observations i.e. Homoscedasticity. To test the presence of heteroscedasticity in residuals we use ARCH test and the hypothesis is as follows,

H_0: There is no heteroscedasticity between residuals

H_1: There is heteroscedasticity between residuals.

1) Autoregressive Conditionally Heteroscedastic (ARCH) process: ARCH models are used to model a volatile variance in the series.

ARCH (q) model specification

The random variable z_t ($\epsilon_t = \sigma_t z_t$) is a strong white noise process. The series $\sigma t^2$is modelled by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon \quad (6)$$

Where $\alpha_0 > 0$ $and \alpha i \geq 0$, $i > 0$

2)Generalized Autoregressive Conditionally Heteroscedastic (GARCH) process: If an autoregressive moving average (ARMA) model is assumed for the error variance, the model is a generalized autoregressive conditional heteroscedasticity (GARCH) model. The GARCH model is equivalent to an infinite ARCH model. In that case, the GARCH (p, q) model, where p is the order of the GARCH terms σ^2 and q is the order of the ARCH terms ϵ^2.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \cdots + \alpha_q \epsilon^2_{t-q} + \beta_1 \sigma^2_{t-1} + \cdots + \beta_p \sigma^2_{t-p}$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon + + \sum_{j=1}^{p} \beta_j \sigma \quad (7)$$

E. Change Point Analysis (CPA)

In the red onion price series, it can be observed that there is a considerable price change occurred in the year 2017 with evidences.Thus change point analysis was performedto detect whether any changes have occurred. It determines changes in summary statistics such as mean, variance etc. and anomalous behaviours. In this study change points in the mean of inputs are estimated using Binary Segmentation method which implement the binary segmentation algorithm using C programming language.

F. Forecasting methods

*1) Static Forecasting:*Static forecast uses the actual value for each subsequent forecast. This method is more suitable in one step ahead prediction where actual values are available.

2) Dynamic Forecasting: Dynamic forecast uses the value of the previous forecasted value of the dependent variable to compute the next one. It assumes that actual values of the series are not available.

III. Results

A. Preliminary Analysis

Figure 1 shows that series is not stationary with a trend and a non-constant variance. There are two points in the data set which can be identified as the outliers as consequences of real situations. Thus, to maintain the continuity of the seriesthose values were taken for the analysis.

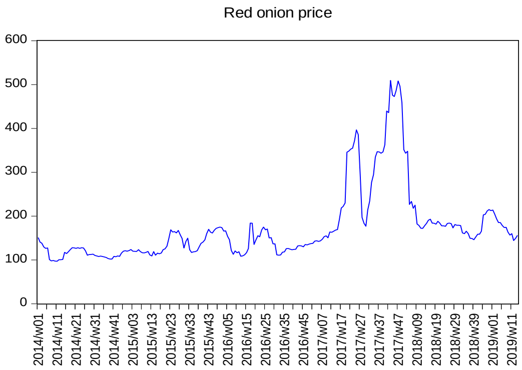


Figure 1. : Time series plot of WARP of red onions

The data set was evaluated with ADF and PP test for the stationary and results are given in Table 3.

Table 3. Stationary test results

Series	Level/ Differenced	p value	
		ADF	PP
Open market weekly average retail prices of red onions	Level	0.9004	0.7464
	Differenced	0.01	0.01

Both tests results are found to be insignificant at 5 percent level of significance, thus confirming the non-stationarity of the series of the WARP of red onions.

B. Univariate Time Series Modelling – ARIMA

Figure 2 shows the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) of the differenced
series. In here ACF cut off at lags 1, 2 while PACF at lag 1.

Figure 2:ACF and PACF of the First Differenced Series

Various combinations of ARIMA models were tested according to the ACF and PACF cut off values.The results of the candidate ARIMA models are given in Table 4.

Table 4. Candidate ARIMA models results

ARIMA Model	**AIC Value**
ARIMA(1,1,0)	8.852958
ARIMA(0,1,1)	8.860000
ARIMA(1,1,1)	**8.851770**

Among all candidate models, ARIMA (1,1,1) was selected as the best model with minimum AIC value.
Estimated equation of the ARIMA (1,1,1) model is:

$$X_t = 0.231994 + X_{t-1} + 0.592755\,(X_{t-1} - X_{t-2}) + 0.0014\,\varepsilon_{t-2} \qquad (8)$$

C. Combined ARIMA – ARCH/GARCH Model

Fitted ARIMA (1,1,1) model was evaluated for the model adequacy. According to the ACF plot of residuals of ARIMA (1,1,1) model in Figure 3, autocorrelationsare within the threshold limits indicating that the residuals areindependently distributed. Further p value of the Ljung Box test is 0.2662 which is greaterthan 0.05 significance level and it suggests that the residualsare independent and uncorrelated.

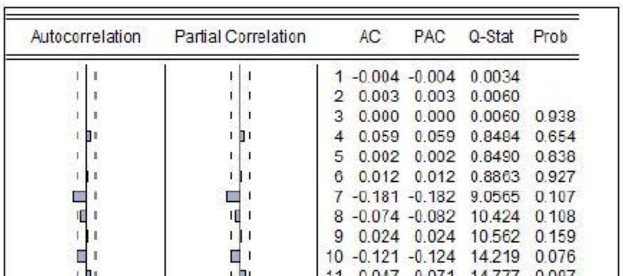


Figure 3.ACF and PACF of the residuals of ARIMA (1,1,1) model

The evidence of the existence or not existence of the heteroscedasticity was tested with the ARCH test and its results are shown in Table 5.

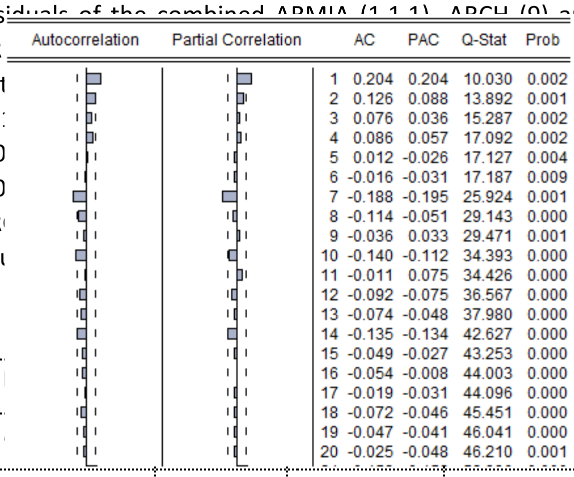
Table 5. Results of the ARCH test

Heteroskedasticity Test: ARCH			
F-statistic	3.919754	Prob. F(1,236)	0.0489
Obs*R-squared	3.888390	Prob. Chi-Square(1)	0.0486

p value of the ARCH LM test is 0.0.0489 which is less than 0.05 significance level and it also suggests that there is ARCH effects in the residuals of ARIMA (1,1,1) model.

Residuals were modelled using the ARCH (9) model and GARCH (9,1) models among many candidate ARCH/GARCH models and theses two models were combined separately with ARIMA (1,1,1) model.Therefore, ARCH/GARCH should be used to model the volatility of the series to reflect more recent changes and fluctuations in the series.

The detailed information about the fitted candidate ARIMA-ARCH/GARCH models are shown in the Table 6. Accordingly, ARCH (9) model has the minimum AIC among the ARCH models while GARCH (9,1) model has the minimum AIC among the GARCH models tested. The residuals of the combined ARMIA (1,1,1) ARCH (9) and
AR he
het IA
(1, is
0.0 an
0.0 no
AR s.
Thu



ARCH(2)	8.237811	GARCH(1,2)	8.010279
ARCH(3)	8.039145	GARCH(1,3)	8.101965
ARCH(4)	8.046735	GARCH(2,2)	8.016527
ARCH(5)	8.029929	**GARCH(9,1)**	**7.925479**

Full models of ARIMA (1,1,1) –ARCH (9) and ARIMA (1,1,1) –GARCH (9,1) can be written as in Equation 9 and Equation 10 respectively.

$$X_t - X_{t-1} = -0.412052 + 0.075465\left(X_{t-1} - X_{t-2}\right) - 0.045738 \; 1.164238\varepsilon_{t-4}^2 \quad (9)$$

$$X_t - X_{t-1} = 0.262520 + 0.691950\left(X_{t-1} - X_{t-2}\right) - 0.566814\,\varepsilon_{t-1} + 6.401871 + 0.094259\varepsilon_{t-1}^2 + 0.096419\varepsilon_{t-2}^2 + 1.164238\varepsilon_{t-4}^2 - 0.162341\varepsilon_{t-7}^2 + 1.095580\varepsilon_{t-8}^2 - 0.008969\,\varepsilon_{t-1} + 10.40185 + 0.212895\varepsilon_{t-1}^2 + 0.654854\varepsilon_{t-3}^2 - 0.117312\varepsilon_{t-4}^2 + 0.787564\varepsilon_{t-8}^2 - 0.097653\varepsilon_{t-9}^2 + 0.133793\,\sigma_{t-1}^2$$

(10)

D. Change point analysis

The data set was again evaluated to identify the change points in the series. The results of the change point analysis of the series are given in Figure 4. Two change points were detected at 3rd week of April 2017 (159th data point) and 4th week of January 2018 (196th data point), hence three windows were identified for the analysis. Separate models were fitted to three windows.

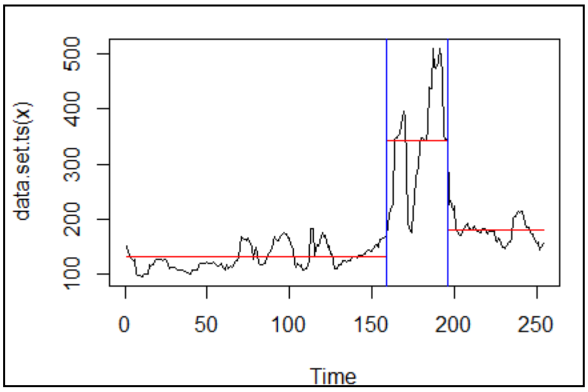


Figure 4.Change point detection – Graphical analysis

The same time series modelling procedure was followed and based on the minimum AIC criterion, ARIMA (0,1,0) model which is known as the random walk model was suggested as the best model for each window.

E. Performance Evaluation of Fitted Models

Four types of models were evaluated in the procedure of finding a suitable model to forecast the open market weekly average retail prices of red onions in Colombo, Sri Lanka. The forecasting results of the models are evaluated in this section.

1) Forecasting with ARIMA (1,1,1) Model: The dynamic and static forecasting results with ARIMA (1,1,1) model are shown in Figure5 and Figure 6 respectively.

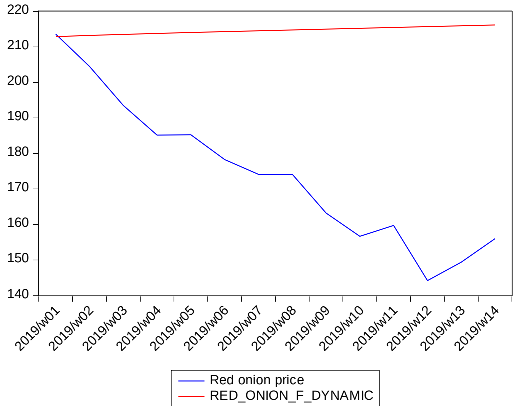


Figure 5.Dynamic forecasting with ARIMA (1,1,1) model

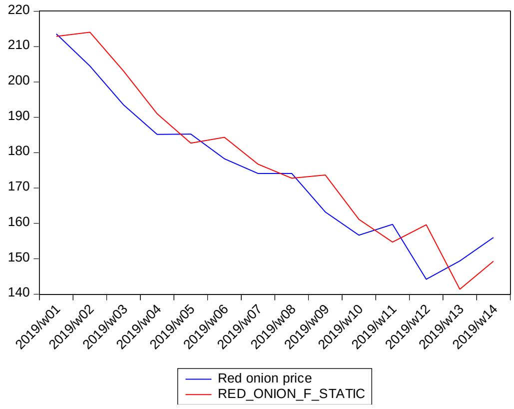


Figure 6.Static forecasting with ARIMA (1,1,1) model

2) Forecasting with ARIMA (1,1,1) – ARCH (9) Model: The dynamic and static forecasting results of ARIMA-ARCH combined model which was compared with actual values of red onion prices are shown in Figure 7 and Figure 8respectively.

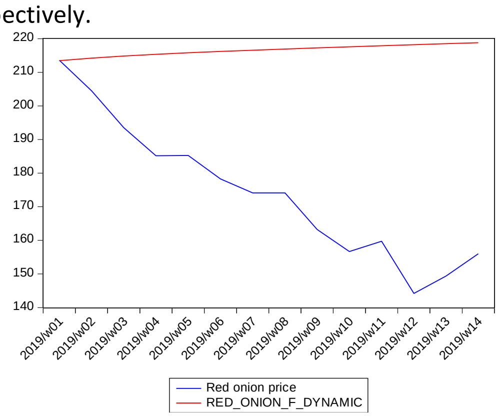


Figure 7.Dynamic forecasting with ARIMA (1,1,1) – ARCH(9) model

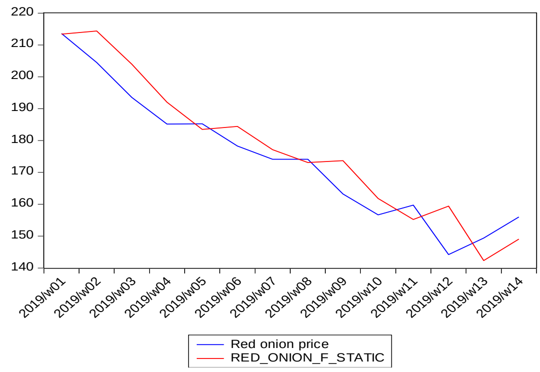


Figure 8.Static forecasting with ARIMA (1,1,1) – ARCH(9) model

3) Forecasting with ARIMA (1,1,1) – GARCH (9,1) Model: Figure 9 and Figure 10 shows the forecasting results of the ARIMA(1,1,1)- GARCH(9,1) model compared to actual values of the red onion retail prices.

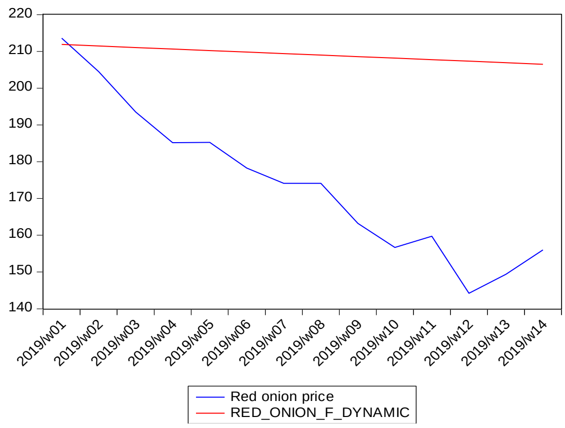


Figure 9.Dynamic forecasting with ARIMA(1,1,1)- GARCH(1,9) model

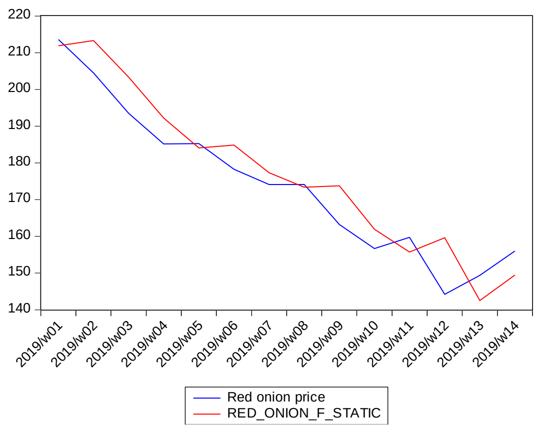


Figure 10.Static forecasting with ARIMA(1,1,1)- GARCH(1,9) model

4) Forecasting under the change point analysis: Under the change point analysis, three windows wereanalysed. The models built for three windows were all ARIMA (0,1,0) models. The dynamic and static forecasting results of ARIMA (0,1,0) model of the last window are shown in Figure 11 and Figure 12 respectively.

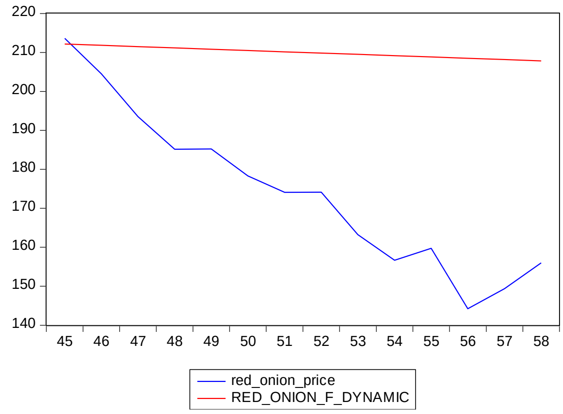


Figure 11.Dynamic forecasting under CPA – Window 3

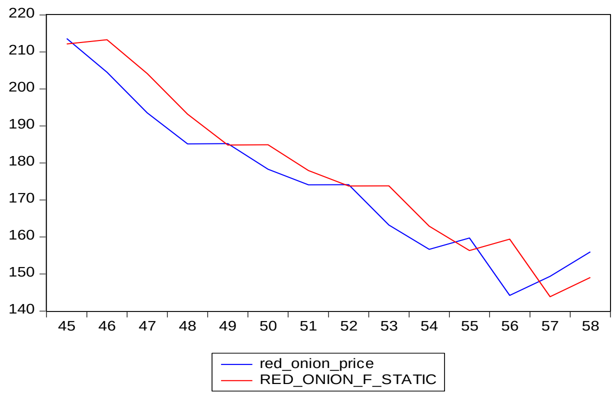


Figure 12.Static forecasting under CPA – Window 3

F. Model Comparisons

All the fitted models were used in forecasting under two techniques and compared to each other. A detailed explanation of the comparison of the models are given in Table 7 and Table 8.

Model	RMSE	MAE
ARIMA(1,1,1)	45.56832	40.57850
ARIMA(1,1,1) – ARCH(9)	47.43516	42.33871
ARIMA(1,1,1) – GARCH(9,1)	**39.58652**	**35.30699**
ARIMA(0,1,0) under CPA	40.43767	36.06731

Table 7. Comparison of the models under dynamic forecasting

Table 8. Comparison of the models under static forecasting

Model	RMSE	MAE
ARIMA(1,1,1)	7.420611	6.312117
ARIMA(1,1,1) – ARCH(9)	7.477184	6.323610
ARIMA(1,1,1) – GARCH(9,1)	**7.392217**	**6.267603**
ARIMA(0,1,0) under CPA	7.499486	6.284767

According to Table 7 and Table 8, ARIMA (1,1,1)-GARCH (9,1) model has the minimum RMSE and MAE values under both types of forecasting techniques which can be considered as the most suitable model to forecast WARP of red onions.

IV. DISCUSSION AND CONCLUSION

This study mainly aims at identifying the behavior of price variations and modelling weekly average retail price of red onions in Colombo, Sri Lanka for the forecasting purpose. Time series approach was used to analyze the data set. As the level is not stationary, differencing technique was used to make the series stationary. After testing several models which are mentioned in Table 4, using AIC model selection criterion ARIMA (1,1,1) model was chosen as the best model.But the assumption of the constant variance of the residuals was violated in the ARIMA (1,1,1) model. Thus several ARCH and GARCH models were tested and finally ARCH (9) model and GARCH (9,1) model were chosen as the possible models. ARIMA(1,1,1) model and combined ARIMA-GARCH models were used for the forecasting purposes.To capture the change occurred in this series change point analysis was done and as a result, ARIMA (0,1,0) model was fitted to all three windows. Only window 3 was used to forecast WARP of red onions for 2019.

Four types of models were used in forecasting and compared with the original retail prices of red onions. Other than this, two forecasting techniques;dynamic and static forecasting were also compared. As the static forecasting use subsequent actual values to forecast one step ahead, Figure 6, 8,10 and 12shows that the predicted line is much closer to actual line. Dynamic forecasting uses the lagged forecasted value assuming that actual values are unavailable. As represented in Figure 5,7,9 and 11, the forecasted values do not closer to the actual values properly.

A numerical comparison of all four models is represented in Table 7 and 8. From this result, ARIMA (1,1,1)-GARCH (9,1) model has the minimum RMSE and MAE values under both types of forecasting techniques. Therefore, ARIMA (1,1,1) – GARCH (9,1) model was selected as the best model with highest accuracy to forecast the retail prices of red onions in Colombo, Sri Lanka. When compared to norm, the achieved forecasting accuracy from time series models were not at satisfactory level. Therefore, as future studies, machine learning techniques such as artificial neural network models are supposed to be used in achieving more accurate forecasting.

References

Bogahawatta, C. (1987) 'Seasonal Retail and Wholesale Price Analysis of Some Major Agricultural Commodities', *AGRARIAN RESEARCH AN D TRAININ G INSTITUTE .*, (July).

Department of Agriculture. (no date) *Red onion*. Available at: https://www.doa.gov.lk/FCRDI/en/enhome/75-red-onion-e (Accessed: 10 May 2019).

Institute, H. K. A. R. and T. (2018a) *Food information bulletin January 2017*.

Institute, H. K. A. R. and T. (2018b) 'Food information bulletin May 2017', 10(05).

Jalikatti,V.N,&Patil, B. . (2015) 'Onion price forecasting in Hubali market of Northern Karnataka using ARIMA technique', *Karnataka J. Agric. Sci*, 28(2), pp. 2–5.

Sivarajasingham, S. and Mustafa, A. M. M. (2011) 'Dynamic Behaviour of Red Onion Prices in Jaffna District', in *proceedings of 1st International Symposium,South Eastern University of Sri Lanka*, p. 32.