Modeling and Forecasting the Conditional Prices of Tomato in Distrcit Hyderabad

Ahsan Hayat Khanzada[1], Naeem Ahmed Qureshi[1], Ali Akbar Pirzado[2*], Arman Khan[3], Komal Arain[1], Iftakhar Ahmed[4], Ahmed Khan Memon[1], [5]Lutfullah

[1]Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan
[2]Department of Statistics, Shaheed Benazir Bhutto University, Shaheed Benazirabad, Pakistan
[3]Department of Business Administration, Shaheed Benazir Bhutto University, Shaheed Benazirabad, Pakistan
[4]Department of Economics, Lasbela University of Agriculture Water and Marine Sciences, Uthal, Balochistan, Pakistan
[5]Agriculture Department (Research Wing), Quetta, Government of Balochistan, Pakistan
***Corresponding:** ali.akbar@sbusba.edu.pk

Abstract:
Keeping in view the growing demand of modeling the conditional moments of a distribution, the present study is an attempt to model and forecast the conditional mean prices of the tomato crop in the district Hyderabad by using the sophisticated statistical models such as ***ARIMA(p,d,q)*** and ***SARIMA(p,d,q)(P,D,Q)s***. The weekly prices of the tomato were downloaded from an official website of Sindh Agricultural Marketing Department. The suitability of the data set for time series analysis was checked through Durbin-Watson test and after finding the suitability, the same was checked for stationarity though AFD test. Original prices were found non-stationary while their first difference made them stationary. The differenced prices were then modeled by applying Box-Jenkins methodology. Seasonality in the prices was detected at every 32 weeks through autocorrelation and partial autocorrelation functions (ACFs and PACFs). Different specifications of seasonal ARIMA i.e., **SARIMA(p,d,q)(P,D,Q)s** models were used and based on the AIC and BIC and the white noise property of the residuals, ***SARIMA(1,1,1)(1,1,1)32*** was selected as the best model. The future prices of the tomato were forecasted using the same model. It was concluded that the prices of tomato have great variability i.e., least prices were found in winter (October) while the highest prices were recorded in April. It was also observed that the forecast errors from ***SARIMA(1,1,1)(1,1,1)$_{32}$*** model were of very small magnitude as compared to the other candidates model. The comparison of the forecasted with the real prices shows that our selected model has upward bias of little magnitude that can be easily neglected. However, on overall basis, our selected model performs well in terms of forecasting and can be used to forecast the future prices of tomatoes in the selected study area. Based on the findings of the present study it is recommended many growers who want to take advantage from increases in tomato prices in April should focus on their production during this month. In terms of the sustainability of tomato production to control the price variations, considering consumers requests under good agricultural practices, the production which is qualified and proper to the food safety carries importance. Due to increase in consumption in the country and expansion towards new international markets, it is recommended that the tomato producers produce proper to the good agricultural practices.

Key Words: Tomato prices, Fluctions, Forecasting, ARMA, SARIMA

INTRODUCTION:

Tomato is the most important fruit worldwide. It is a relatively short duration crop and gives a high yield. Its botanical name is *Solanum Lycopersicon*, and it belongs to the family *Solanaceae*. It contains vitamin B, C, iron, and phosphorus. The annual production of tomato is approximately 159 million tonnes. Tomato production in Pakistan was 530 thousand tonnes during 2011. The nine largest producing countries account for 74.2 % of the world's yearly production (GoP, 2011).

Tomato is a fruit of significant economic values in Pakistan. Annual export report of tomato from the country averaged about 9833 tones during the past five years. Lowest export figure was recorded during 2014 - 15 and attributed to the bed crop harvest and rendering export. Per unit export prices are also low which apparently are attributed to produce quality. Pakistan exported tomato to a quantum of 5692 tones and earned rupees 77 million during 2015. Based on the last ten years average, the present national yield of tomato is 10.1 tones/ hectors which is quite low. To obtain a potential yield, high yielding varieties and improved production technology must be adopted (MINFAL, 2013). Sindh tomato production was recorded as 141.586 metric tons during 2015 as compared to 114.771 metric tons in the year 2014. The area under cultivation of tomato in Pakistan from 2000-01 to 2009-10 has increased from 27.9 to 50.0 thousand hectors and the production has increased from 268.8 to 476.8 thousand tones. Among the four provinces of Pakistan, Sindh ranks third in terms of area and production of tomato followed by Balochistan and KPK. It is cultivated in southern region of Sindh which includes Hyderabad, Badin, Thatta, and Karachi while in northern region it is cultivated in Mirpurkhas, Nawabshah, Nowshero Feroz, Larkana, and Sukkur.

There has been observed a big variation in the retail prices of tomato in Pakistan which ranges from Rs. 10-20/kg to Rs.80-120/kg. This is because the prices of agriculture products are determined by supply and demand. During the days of high demand (i.e., during Ramadan, Eid-ul-Fitr and Eid-ul-Azha) the prices hit the ceiling of Rs. 120/kg (DAWN, 2016). There are several factors such as yield/production, quality, and demand and supply etc have already been reported in the literature as the cause of the price variation. Price forecasting is more acute with crops particularly tomato due to its highly perishable nature and seasonality. Forecasting tomato prices can provide critical and useful information to tomato growers making production and marketing decisions. Further, to improve domestic market potential for small holder producers, who are the biggest suppliers in the market and in line with the government's Agriculture Sector Development Strategy (ASDS). Modeling the dynamics of conditional distribution is overall challenging and this study will provide the guidance to other students and researchers who want to model the conditional behavior of prices of any other crop in Pakistan.

1.5 Objectives of the Study:

The specific objectives of the present studies are,

1. To model the time- varying behavior of tomato prices.
2. To review the forecasting techniques used in time series analysis.
3. To forecast the future prices of tomato crop for the area under study.

MATERIALS AND METHODS

3.1 Data Description

The secondary data used in the present study consist of weekly prices of tomato which were collected from the official websites of Sindh Agriculture Marketing, Hyderabad [1]. The data span from the first week of January 2011 to the last week of October 2016. Selection of the district was purely subjective since the researcher belongs to the district Hyderabad. The mean weekly wholesale prices of the tomato were calculated and the same were used for further analysis. The collected data yield nearly 300 observations. MATLAB (MATRIX LABORATORY) 2015a version was used for analyzing the data set.

3.2 Methodology

Suitability of Data Set for Time Series Analysis

After collecting data, it was tested for its suitability for time series analysis. For this purpose, Durbin-Watson test was carried out to understand the nature of the data.

1. www.sindhagrimarketing.com.pk.

3.6 Augmented Dickey Fuller (ADF) Test

This is a unit root-test which is used to check whether the time series is stationary. The null hypothesis in the ADF test is "there is a unit root" i.e., in case of ***AR(1)*** model;

$$Y_{t-1} = \alpha y_{t-1} + \varepsilon_t$$

$$H_0; \alpha = 1$$

While the alternate hypothesis states that "The time series has no unit root". The alternate hypothesis can be formulated as under:

$$H_A; \alpha < 1$$

The dickey fuller (DF) test is simply the *t*-test for *H₀*

$$\hat{T} = \frac{\hat{\theta} - 1}{se(\hat{\theta})}$$

The asymptotic distribution of $\hat{T}$ is not normal.

NOTE all the time series process can be well represented by the first order Autoregressive process i.e., $\Delta y_t = \alpha_\circ + Yy_{t-1} + \alpha_2 t + \varepsilon_t$. It is possible to use the Dickey-Fuller tests in higher order equations. Consider the P^{th} order autoregressive process.

$$y_t = \alpha_\circ + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + ... + \alpha_{p-2} y_{t-p+2} + \alpha_{p-1} y_{1-p+1} + \alpha_p y_{t-p} + \varepsilon_t$$

To best understand the methodology of the augmented dickey-fuller test, add and subtract $\alpha_p y_{t-p+1}$ to obtain:

$$y_t = \alpha_\circ + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + ... + \alpha_{p-2} y_{t-p+2} + (\alpha_{p-1} + \alpha_p) y_{t-p+1} - \alpha_p \Delta y_{t-p+1} + \varepsilon_t$$

Next add and subtract $(\alpha_{p-1} + \alpha_p) y_{t-p+2}$ to obtain

$$y_t = \alpha_\circ + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + ... - (\alpha_{p-1} + \alpha_p) \Delta y_{t-p+2} - \alpha_p \Delta y_{t-p+1} + \varepsilon_t$$

Continuing in this fashion, we get:

$$\Delta y_t = \alpha_\circ + Yy_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

$$r = -\{1 - \sum_{i=1}^{p} \alpha_i\}$$

$$\beta_i = \sum_{j=i}^{p} \alpha_j$$

Note that the dickey-fuller tests assume that the errors are independent and have constant variance. This raises four important problems related to the fact that we do not know the true data-generating process. First, the true data-generating process may contain both autoregressive and moving averages components. We need to know how to conduct the test order of the moving average terms. Second, we cannot properly estimate y and its standard error unless all the autoregressive terms are included in the estimating equation, clearly, the simple regression $\Delta y_t = \alpha_\circ + \gamma y_{t-1} + \varepsilon_t$ is inadequate to this task if the true data-generating process. However, the true order of the autoregressive process.

As test statistics indicates that the data is stationary, we can proceed estimate and forecast the model in order to fulfill our objectives of our proposed study using ***ARIMA(p,d,q)*** or ***SARIMA(p,d,q)(P,D,Q)*** model. A brief about these models is given as under:

3.7 Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA(p,d,q) is better known as a time series forecasting techniques for short run, which is widely used in today's world since the evolution of sophisticated statistical software packages. ***ARIMA(p,d,q)*** has four major steps in model building, identification, estimation, diagnostics & forecast. With these four steps first tentative model parameters are identified through graphs *ACF* and *PACF* then coefficient are determined and find out the likely model, next steps involves is to validate the model and at the end use simple statistics and confidence intervals to determine the validity of the forecast and track model performance.

ARIMA(p,d,q) model use the historic data and decomposes it into autoregressive ***(AR(p))*** indicates autoregressive lags and ***(MA(q))*** indicates weighted moving average lags over past errors. Therefore, it has three model parameters ***AR(p)***, ***I(d)*** and ***MA(q)*** all combined to form ***ARIMA(p,d,q)*** model. Where

p = order of autoregression
d = order of integration (differencing)
q = order of moving average
A non-seasonal stationary time series can be modeled as a combination of past values and the errors which can be denoted as ***ARIMA(p,d,q)*** or can be expressed as follows:

$$X_t = \varphi_\circ + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + + \varphi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_q e_{t-q}$$

$\varphi_0, \varphi_1, ..., \varphi_p$ are the autoregressive parameters, $\theta_1, \theta_2, ..., \theta_q$ are the moving average parameters, $X_{t-1}, X_{t-2}, ..., X_{t-p}$ are the lagged values of the dependent variable $e_{t-1}, e_{t-2}, ..., e_{t-q}$ are the lagged values of the stochastic error term.

3.8 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

Once the seasonal indices are obtained, each observation is divided by its seasonal index to deseasonalize the data. The reason for deseasonalizing the price series is to remove the seasonal fluctuations so that the trend and cycle can be studied (Lind *et al.*, 2009). For example, if you see a time series of sales that has not been deseasonalized, and it shows a large increase from November to December, you might not be sure whether this represents a real increase in sales or a seasonal phenomenon. However, if this increase is really just a seasonal effect, the deseasonalized version of the series will show no such increase in sales (Albright, *et al.*, 2011). In this study, we used ***SARIMA*** (seasonal ***ARIMA*** or seasonal autoregressive integrated moving average) model to forecast one-period ahead of the weekly tomato price series by applying Box-Jenkins approach. ***SARIMA*** model is useful in situations when the time series data exhibit seasonality-periodic fluctuations that recur with about the same intensity each year (Garcia-Martinez, *et al.*, 2011).

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is (Chatfield, 2012): ***ARIMA(p,d,q)×(P,D,Q)_S***, with *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, and *S* = time span of repeating seasonal pattern (in a monthly data s = 12).
Without differencing operator, the model could be written more formally as;

$$\left(\Phi\left(B^S\right)\varphi\left(B\right)x_t - \mu\right) = \Theta\left(B^S\right)\theta\left(B\right)w_t$$

The non-seasonal components are:

$$AR : \varphi\left(B\right) = 1 - \varphi_1 B - ..... - \varphi_p B^p$$

$$MA : \theta\left(B\right) = 1 + \theta_1 B + ... + \theta_q B^q$$

The seasonal components are:

$$\text{Seasonal } AR : \Phi\left(B^S\right) = 1 - \Phi_1 B^S - .... - \Phi_P B^{PS}$$

$$\text{Seasonal } MA : \Theta\left(B^S\right) = 1 + \Theta_1 B^S + ... + \Theta_Q B^{QS}$$

3.9 Box-Jenkins Methodology

Box - Jenkins Analysis refers to a systematic method of identifying, fitting, diagnostic checking, and then forecasting using autoregressive integrated moving average (ARIMA) time series models. The method is appropriate for time series of medium to long length (at least 50 observations). There are four steps of Box-Jenkins methodology.

Step # 1. Identification
Step # 2. Estimation
Step # 3. Model Validation/ Diagnostics
Step # 4. Forecasting
These steps are briefly described as under;

Step # 1. Identification

The identification step involves fitting the autoregressive component (variable "*p*"), the moving average component (variable "*q*") of the *ARIMA* model as well as any required differencing (variable "*d*") to make the time series stationary or to remove seasonal effects. The identification process is accomplished with the help of plotting the Autocorrelation and Partial Autocorrelation functions i.e., *ACF* and *PACF* of the data with their lagged values or by using the AIC (Akaike Information Criterion) and SBIC (Shwarz Bayesian Information Criterion) criteria. Details of these techniques are given as under:

1. Autocorrelation Function *(ACF)*

Autocorrelation is defined as the correlation of the data series along with its own lagged values. For example, if $\{Y_t\}$ is the time series then first order autocorrelation is the correlation of Y_t with Y_{t-1} i.e., $Corr(Y_t, Y_{t-1})$. In general, the autocorrelation between Y_t and its *ith* lagged value is $Corr(Y_t, Y_{t-i})$. It value ranges from -1 to +1.

2. Partial Autocorrelation Function *(PACF)*

Yet another important characteristic is a partial auto-correlation function (PACF) which is conditional correlation of Y_{t-k} with $Y_{t.}$ after removing the effects of Y_{t+k-1}. PACF is defined for positive lag only; their values also lie between -1 to +1. Both the characteristics, ACF & PACF are equally important, but ACF is relatively easier to calculate than PACF. Table 1 shows how ACF and PACF help in identifying the lagged values of the time series data?

Table 1. Properties of ACF and PACF for AR, MA and ARMA models

Properties	***AR (p)***	***MA (q)***	***ARMA (p,q)***
ACF	Decay	Cuts after *q* legs	Decay
PACF	Cuts after *p* legs	Decay	Decay

3. Akaike's Information Criterion (**AIC**)

It is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Hence, AIC provides a means for model selection.

The following is the formula for calculating the value of AIC;

$$AIC = \ln \sigma_k^2 + \frac{n+k}{n-k-2}$$

Where *k* is the number of parameters in the model, *n* is the sample size and *ln* is the natural logarithm.

4. Bayesian Information Criterion (**BIC**) or **Schwarz Bayesian Information Criterion** (**SBIC**)

It is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike's information criterion (AIC).

The BIC was developed by Gideon E. Schwarz and published in a 1978 paper, where he gave a Bayesian argument for adopting it. It can be calculated as under;

$$SIC = \ln \sigma_k^2 + \frac{k \ln n}{n}$$

Where *k* is the number of parameters in the model, *n* is the sample size, and *ln* is the natural logarithm.

Step # 02. Estimation

The estimation procedure involves estimating the model parameters for different values of *p*, *d* and *q* orders to fit the actual time series. We allow the software to fit the historical time series, while the user checks that there is no significant signal from the errors using an ACF for the error residuals, and that estimated parameters for the autoregressive or moving average components are significant. Shortly, we will select the model that is parsimonious (a model having all significant parameters after excluding the redundant parameters). Usually, the following estimation methods are used to estimate the parameters.

1. Ordinary Least Square (OLS)
2. Maximum Likelihood Estimation (MLE)
3. Yule-Walker Equation

Since the estimation of *ARIMA(p,d,q)* is mostly done by using the MLE method, so here we will describe only this method briefly.

Maximum Likelihood Estimation (**MLE**): It is a method of estimating the parameters of a statistical model for the given observations. It works in the manner by finding the parameter values that maximize the likelihood function which simply tells you about the likelihood (most likely chances) that with these parametric values the model generates the data set. MLE be a special case of the Maximum Posteriori Estimation (MAP) that assumes a uniform prior distribution of the parameters or as a variant of the MAP that ignores the prior and which therefore is un-regularized.

Step # 03. Model validation/ diagnostics

Once a model has been fit, the final step is the diagnostic checking of the model. The checking is carried out by studying the autocorrelation plots of the residuals to see if further structure (large correlation values) can be found. If all the autocorrelations and partial autocorrelations are small i.e., less than *2s.e*, the model is considered adequate, and forecasts are generated. If some of the autocorrelations are large, the values of *p* and/or *q* are adjusted, and the model is re-estimated.

Ljung-Box Test

This test is used to check if the residuals from the estimated *ARMA (p,q)* model behave like a white noise (Enders, 2010). In this study, we applied this test in a univariate fashion. The test statistic is formulated as follows;

$$Q_{MK} = \frac{T(T+2)\sum_{m=1}^{M} \hat{\rho}_{ek}(m)}{T-m}$$

Where $\hat{\rho}_{e_k}(m)$ is the estimated sample autocorrelation at lag *m* and *T* is the sample size. We reject the null hypothesis of no significant autocorrelations i.e.,

$$\mathrm{H}_0;\ \hat{\rho}_e(1) = \hat{\rho}_e(2) = \dots\dots = \hat{\rho}_e(m) = 0$$

against the alternative that at least one of these autocorrelations is not equal to 0 i.e.,

$$H_A;\ \hat{\rho}_e(1),\dots\dots\hat{\rho}_e(m) \neq 0,$$

is non zero at a conventional level of significance α = 0.05 & 0.01. The test statistics follows chi-square distribution with *m-p-q* degrees of freedom, i.e., $Q_m \sim \chi^2(m-p-q)$, where *p* and *q* refer to the autoregressive and the moving average lag of the *ARIMA(p,d,q)* process respectively. Selecting an appropriate lag length is crucial when applying this test as the number of selected lags (*m*) affects the power of the test. If *m* is too small, the test will not detect higher-order autocorrelations. If it is too large, the test will lose power when significant correlation at one lag is washed out by insignificant correlation at other lags. The default value of *m*=20 has been suggested by Box, Jenkins, and Reinsel (1994), while Brockwell and Davis (1991) showed with simulation evidence that a value approximating *ln(T*) provides better power performance.

Step # 04. Forecasting

After a time series is assured to be stationary, and fitted a model in such a way that there is no autocorrelation information in the residuals, we can proceed to forecasting. Forecasting assesses the performance of the model against real data. There is an option to split the time series into two parts, using the first part to fit the model and the second part to check model performance. Usually, the utility of a specific model or the utility of several classes of models to fit actual data can be assessed by minimizing a value such as mean square forecast error (MSFE), root mean square forecast error (RMSFE), mean absolute percentage error (MAPE), mean absolute error (MAE) etc.

1. Mean Square Forecast Error (MSFE)

In statistics, the mean squared forecast error of a curve fitting procedure is the expected value of the squared difference between the forecasted values implied by the forecasted function and the observed values of the data set. Mathematically;

$$MSFE = \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{n}$$

Where y_t is the observed value and $\hat{y}_t$ is the forecasted value and *n* is the total number of observations forecasted.

2. Root Mean Square Forecast Error (RMSFE)

It is defined as the square root of mean square (the arithmetic means of the squares of a set of numbers) the RMS is also known as the quadratic mean and is a particular case of generalized mean. RMS can also be defined for a continuously varying function in terms of an integral of the squares of the instantaneous values during a cycle. Mathematically.

$$RMSFE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}$$

3. Mean Absolute Percentage Error (MAPE)

Also known as mean absolute percentage deviation (MAPD) is a measure of prediction accuracy of a forecasting method in statistics, example in trend estimation. It usually expresses accuracy as a percentage. Mathematically;

$$MAPE = 100 \times \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

RESULTS

4.1 Descriptive Statistics:

Figure 1: Weekly Wholesale Prices of Tomato: January, 2011 – September, 2016.

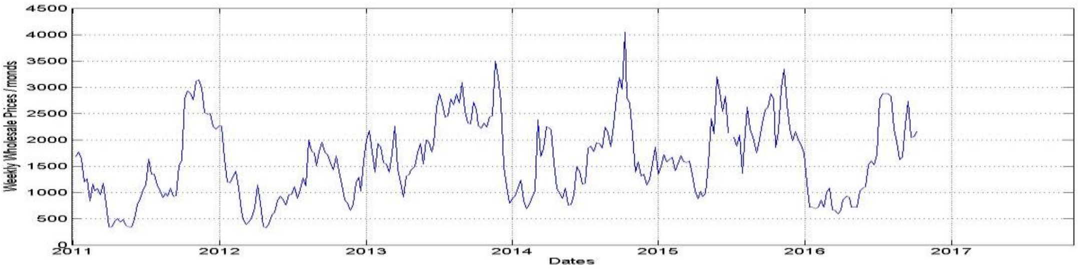


Figure 1 shows the time series plot of weekly prices of tomato (in monds) from 1st January 2011 to 30th September 2016. It can be clearly observed that these prices show some seasonal pattern however, this seasonal cycle is very difficult to examine here. Tomato price increases and decreases due to many reasons whether it is annual festivals like (Eid-ul-Fitr, Eid-ul-Azha, and Moharram-ul-Haram) and also due to seasonality. There are several months in which tomato have sky touching prices due to the unavailability of the tomato i.e., in the months of March-April, August-September, and October. During these months, the suppliers import tomato from the other countries like India, and China (MINFAL, 2015).

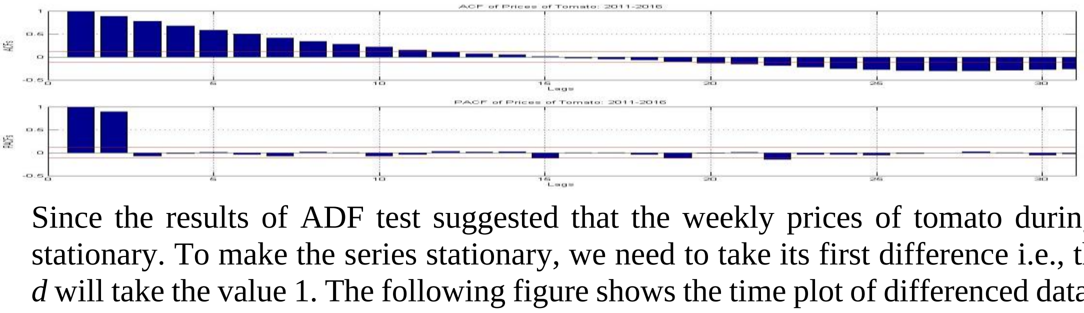
Table 2. Summary Statistics of Weekly Prices of Tomato along with different tests

Minimum	325.7143	
Maximum	4050.000	
Mean	1617.052	
Variance	568582.600	
Skewness	0.398476	
Kurtosis	5.417782	
	Test Statistic	***p-value***
Ljung-Box test	1146.226	0.000
Durbin-Watson test	0.048127	0.000
ADF test	-1.62806	0.09775

Above table describe the descriptive results of the data set. The minimum of weakly wholesale prices of tomato /mounds is found Rs. 325.71 whereas the maximum of prices was attained RS 4050.00. The mean of the prices is RS 16117.05 with the variance of RS 568582.600 the large value of variance shows the greater variability in the prices of tomato. Similarly, the value of skewness (0.39876) and kurtosis (2.417782) shows that the data are away from normality.

The value of Durbin-Watson (DW) was 0.048127 for the sample data of the tomato price from January, 2011 to September, 2016 which indicates that the data is suitable for time series analysis. As $DW \approx 2[1-\rho(1)]$ = 0.954 which indicates that tomato prices show high 1st order autocorrelation which can be translated as the data under consideration are suitable for time series analysis. The p-value (0.09775) of the test statistic is greater than the selected level of significance (α = 0.05) shows that the null hypothesis is accepted. This shows that the series have a unit root which means it is non-stationary and needs to be differenced to make it stationary.

Figure 2. Autocorrelations and Partial Autocorrelations of weekly wholesale prices of tomato



Since the results of ADF test suggested that the weekly prices of tomato during the sample span are not stationary. To make the series stationary, we need to take its first difference i.e., the differencing parameters *d* will take the value 1. The following figure shows the time plot of differenced data. It can be easily seen from the above figure that the first difference of the data makes the weekly prices stationary. In simple words, now mean and variance of the data under study are now time invariant i.e., these measures are now not a function of time but the function of a lagged value. This statement can be validated from the ADF test again (the results of which are shown in the following table.

Figure 3. Plot of differenced weekly prices of tomato: January, 2011 to September, 2016.

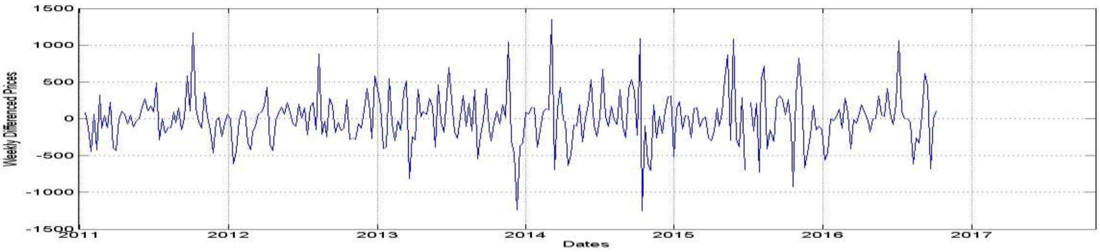
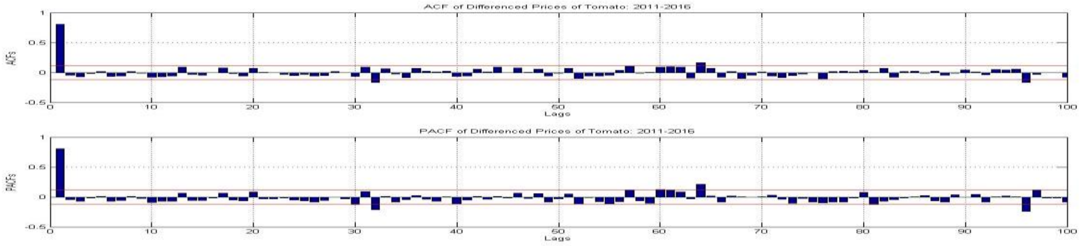


Table 3. Summary Statistics of Differenced Prices of Tomato along with ADF-test

Minimum	-1254.290	
Maximum	1350.000	
Mean	1.600	
Variance	124044.300	
Skewness	0.303	
Kurtosis	7.130	
	Test Statistic	***p-value***
Ljung-Box test	21.716	0.864
ADF test	-17.059	0.001

Table 3 divulges that the Augmented Dickey Fuller (ADF) test rejects the null of unit root which means that the differenced time series in now stationary. In the similar way, the Ljung-Box test statistic shows the autocorrelations are not significant.

Figure 4. Autocorrelations and Partial Autocorrelations of weekly differenced prices of tomato



The ACFs and PACFs of the differenced series are plotted in the Figure 04. ACF and PACF values were found to be high at specific lags for the series. These values were determined as making sudden peaks and not disappearing especially at period of 32 weeks (32, 64, 96 etc). This demonstrates that the series has a seasonal structure. The seasonal spikes of ACF and PACF at lag 32, 64, 96 and so on are observed as being cut off after taking the difference. This also indicates that the seasonal model of ***AR(1)*** and ***MA(1).*** Therefore, to include the model of (1,1,1) to the part (P,D,Q) of the model will be formed can be considered as one of the best possibilities among the alternative choices. So far as at the non-seasonal part of the model *(p, d, q)*, the discontinuation of ACF and PACF after lag 1 indicates the addition of ***AR(1)*** may be appropriate. On the other hand, even the discontinuation occurs after 1 lag at the ACF and PACF value, these values are observed to be increased after a certain lag. Therefore, there is no clarity for the ***MA(q)*** term at the non-seasonal part of the model. In this situation, different alternatives are to be considered in order to account for the non-seasonal part of the model.

Table 4. SARIMA(p,d,q)(P,D,Q)S model estimates for differenced weekly prices

Parameters	(1,1,1)(1,1,1)$_{32}$	(1,1,1)(1,1,2)$_{32}$	(1,1,1)(1,1,3)$_{32}$	(1,1,1)(2,1,1)$_{32}$	(1,1,1)(3,1,1)$_{32}$
α_1	0.898	0.850	0.824	0.988	-0.138
t-value	17.038	11.090	8.886	17.378	-0.286
α_2				-0.110	-0.047
t-value				-2.030	-0.778
α_3					-0.129
t-value					-2.610
SAR(32)	-0.295	-0.342	-0.351	-0.308	-0.387
t-value	-6.391	-7.036	-7.142	-6.330	-8.136
β_1	-0.969	-0.836	-0.801	-0.965	0.186
t-value	-32.119	-9.224	-7.488	-29.873	0.382
β_2		-0.108	-0.091		
t-value		-1.784	-1.377		
β_3			-0.042		
t-value			-0.716		
SMA(32)	-0.765	-0.755	-0.754	-0.768	-0.744
t-value	-17.755	-16.657	-16.455	-17.779	-16.136
LL	-2191.700	-2190.300	-2190.100	-2189.900	-2192.200
AIC	4393.400	4394.700	4396.300	4393.900	4400.300
SBIC	4411.900	4420.600	4425.900	4419.800	4430.000
Q-STAT	34.060	41.549	43.775	39.119	45.880
p-value	0.278	0.0782	0.050	0.012	0.032

The estimates of the models along with their AICs, BICs, and the results of Ljung-Box test are presented in the Table 4. The model selection is purely based on the smallest AIC and SBIC values (Wang and Lim, 2005). Besides these two criteria, the residuals of the selected model should behave like a white noise process (without having any significant autocorrelations in the residuals of the selected model). Among all the models, only the parameters from ***SARIMA(1,1,1)(1,1,1)$_{32}$*** and ***SARIMA(1,11)(2,1,1)$_{32}$*** models were found significant at conventional level of significance (**α = 0.05**). Besides these models, the parameters of the remaining models were found to be non-significant (since t-values are less than **1.95**) which indicates that these models contain redundant parameters and their selection leads us to non-parsimonious model. Based on the estimates presented in the Table 5, log-likelihood (***LL***) selects the model with large number of parameters i.e., ***SARIMA(1,1,1)(3,1,1)$_{32}$***. It is well documented in the literature that ***LL*** always selects the model with large number of parameters, so it does here. Similarly, ***AIC*** and ***SBIC*** select ***SARIMA(1,1,1)(1,1,1)$_{32}$*** model. Their values were found to be least as compared to the other alternative models.

Figure 5. ACF and PACF of the residuals from *SARIMA(1,1,1)(1,1,1)$_{32}$* model

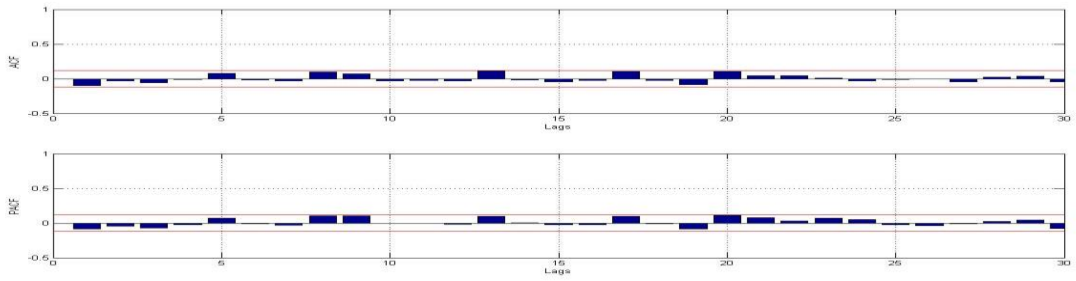


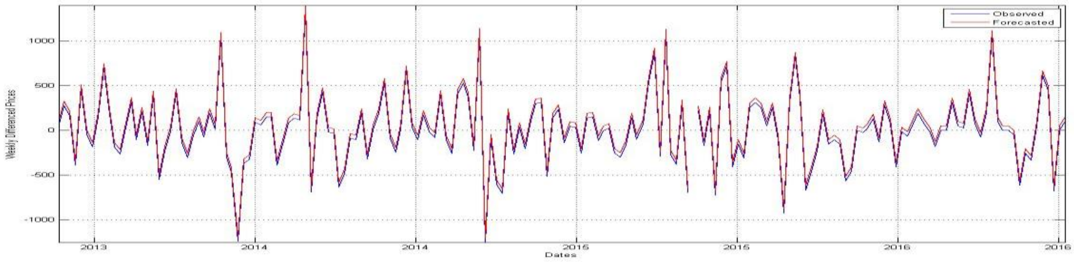
Table 5. Out-of-sample forecast evaluation of selected *SARIMA* models

	(1,1,1)(1,1,1)$_{32}$	***(1,1,1)(1,1,2)$_{32}$***	***(1,1,1)(1,1,3)$_{32}$***	***(1,1,1)(2,1,1)$_{32}$***	***(1,1,1)(3,1,1)$_{32}$***
MSFE	581419.2	798714.7	741129.7	641153.1	585998.5
RMSFE	762.5085	893.7084	860.8889	800.7204	765.5054
MAPE	12.52	23.65	31.84	21.97	34.47

It can be clearly seen form the table above that all the out-of-sample forecasts criterion select ***SARIMA(1,1,1)(1,1,1)$_{32}$*** model. In the present study, in case of in-sample-fitting and out-of-sample forecast evaluation, the same model is selected. The outcome shows that the proposed model can forecast the real tomato prices with an accuracy of MAPE value 12.52. MAPE is 12.52%, meaning that the forecasts are off by about 12% on average. The following figure shows the plot of observed and the forecasted differenced prices from the selected model i.e., ***SARIMA(1,1,1)(1,1,1)$_{32}$*** for the last 180 observations (out-of-sample period).

It can be clearly observed that our selected model produces forecasts which are greater than the observed values which mean that our model has upward bias. The uncertainty associated with each estimated values is greater than the expected and hence upward biased is introduced in the estimated model.

Figure 6. Plot of Observed and Forecasted Prices of Tomato from SARIMA(1,1,1)(1,1,1)$_{32}$ Model: 1st week of May, 2013 to 4th week of September, 2016



The following table shows some of the real prices and their forecasted values respectively. However, on overall basis, our selected model performs well in terms of forecasting and can be used to forecast the future prices of tomato in the selected study area.

Table 6. Out-of-sample Real and Forecasted Tomato Prices from S*ARIMA(1,1,1)(1,1,1)$_{32}$*: 1st May, 2013 to 30th December, 2013

Weeks	**Observed**	**Forecasted**	**Weeks**	**Observed**	**Forecasted**
1	57.14286	77.22176	13	28.57143	46.32175
2	274.2857	311.4321	14	314.2857	355.9831
3	165.7143	201.6234	15	-107.143	-125.6294
4	-388.5710	-401.7830	16	207.1429	241.1958
5	462.8571	498.6341	17	-171.429	-198.5589
6	-45.7143	-62.9263	18	392.8571	422.4582
7	-184.286	-202.176	19	-550.000	-590.1633
8	148.5714	163.4591	20	-228.571	-256.6221
9	697.1429	725.2271	21	-14.2857	-26.3671
10	261.4286	301.2749	22	414.2857	432.3417
11	-182.857	-227.4569	23	-140.000	-156.1532
12	-265.714	-296.4391	24	-302.857	-345.1032

DISCUSSIONS:

The results found and presented in the previous chapter are discussed in the following paragraphs.

The weekly prices of the tomato show a non-stationary structure due to the presence of the trend in the prices. It is well reported in the literature that the prices are non-stationary in the nature as reported by Lewbel and Serena (2002), Chang et al.(2015). The trend effects were removed by taking the first difference of the prices. This differencing makes it stationary. Most of the time series becomes stationary at their first difference as reported by Hyndman (2008) and Wooldridge (2006), the same was observed and reported in the present research. The results of ADF-test also suggested that the time series in trend stationary. The stationary structure of Real prices can be considered as negative regarding the sustainability of tomato production while the increase (Lundell *et al.*, 2004) in real prices of input is considered.

Once stationarity is achieved, the next step was to find the data generating process with the help of ACF and PACF of the differenced series. Since the exponential decay of the ACF showed that process is *AR(1)* and the behavior of PACF showed that it also includes the *MA(1)* process, so the expected model on the basis of ACF and PACF was *ARMA(1,1)* model. Besides, the significance of every 32 lag showed the seasonality pattern in the prices of tomato which was around 8-month seasonality in the prices of tomato. The same results were also reported by Adanacioglu and Yercan (2012) in their research regarding the modeling of tomato prices in Turkey. On the basis of all these results the proposed model was *SARIMA(p,d,q)(P,D,Q)S*. The seasonality in the behavior of tomato prices were also reported by Adanacioglu and Yercan (2012). The estimation of the parameters of the ***SARIMA(p,d,q)(P,D,Q)S*** model was done through maximum likelihood method. The estimation results showed that all the parameters of ***SARIMA(1,1,1) (1,1,1)*** $_{32}$ was the best model. This decision was also confirmed with the help of AIC and BIC criteria. These two criterion also selected the ***SARIMA(1,1,1) (1,1,1)*** $_{32}$ model. Adanacioglu and Yercan (2012) in their research also selected the ***SARIMA(1,0,0) (1,1,1)*** $_{12}$ model for modeling the tomato price behavior in Turkey which clearly indicates that the price behavior of tomato has clear seasonal effects. This difference in the seasonality behavior between Turkey and Pakistan might be due to the change in sowing and harvesting period, due to difference in supply and demand period and due to the climatic conditions of the two countries. After estimating the model, the next step was to forecast the future values of the time series under study. One-step-ahead forecasts were generated using the best selected model. The forecasted values were found quite close to the real values of the data under study. The forecast errors were of very small values as reported by the Adanacioglu and Yercan (2012) in their study. It was clearly observed that our selected model produces forecasts which are greater than the observed values which mean that our model has upward bias. The uncertainty associated with each estimated value is greater than the expected and hence upward biased is introduced in the estimated model. However, on overall basis, our selected model performs well in terms of forecasting and can be used to forecast the future prices of tomatoes in the selected study area.

CONCLUSION:

This study concludes that the there is a great variability in the prices of tomato in the selected district. The prices of tomato are changing after every thirty two weeks which shows the seasonality pattern in the prices i.e., the prices are very low during winter while very high during summer. The results obtained from this study shows that the prices of tomatoes in Hyderabad district have not showed any trend towards an increase or a decrease. In fact, the decrease of income of tomato growers may bring out the farmer group who is unwilling to continue to produce tomato. It can already be stated that the tomato producers work away from the profitability. Based on the findings, it can be concluded that the conditional price behavior of tomato can be best modeled by using the ***SARIMA(1,1,1)(1,1,1)*** $_{32}$ model in the selected district. The forecasts predicted from the model which has chosen to determine the course of the prices of next few years show that any significant changes will not occur in real tomato prices by the coming years and the past behavior of tomato will continue in the future with the same pattern of seasonality. However, the price forecasts put forward that the tomato growers face with a price risk caused by the uncertainty of the market.

FUTURE OUTCOMES:

The following recommendations are purely based on the major findings of the current study.

Many growers who want to take advantage from increases in tomato prices in April should focus on their production during this month.
In terms of the sustainability of tomato production in order to control the price variations, considering consumers requests under good agricultural practices, the production which is qualified and proper to the food safety carries importance.
Regarding the low tomato production in the selected district during summer requires the government to provide credit facilities that will enable households to access such credit at a reasonable cost.
To provide a reasonable price which would be accepted by the growers and to measure against risk factors of the price to be taken are required.
For this purpose, considering consumers requests under good agricultural practices, the production which is qualified and proper to the food safety carries importance.
Due to increase in consumption in the country and expansion towards new international markets will be possible. Lately, arriving to the awareness of this situation of the tomato growers is thought. Thus, the survey in the Antalya province of Turkey in 2011 conducted by Yercan *et al.* has showed that the tomato producers produce proper to the good agricultural practices.

Literature Cited:

[1]. Adams, S. O., Awujola, A., and Alumgudu, A. I. (2014). Modeling Nigeria’s Consumer Price Index Using ARIMA Model. *International Journal of Development and Economic Sustainability,* 2(2): 37-47.

[2]. Adanacioglu, H and Yercenm, M. (2012). An analysis of tomato prices at wholesale level in Turkey: an application of SARIMA model. Custos e @gronegócio *on line*. 8(4): 52-75.

[3]. Albright, S. C., Winston, W., and Zappe, C. (2011). Data Analysis and Decision Making, 4th Edition, Time Series Analysis and Forecasting, ISBN-10: 0538476125 ISBN-13: 9780538476126, 669-743.

[4]. Amir, M. and Ani, S.(2015). Modeling and forecasting monthly crude oil price of Pakistan: A comparative study of ARIMA, GARCH and ARIMA Kalman model. ADVANCES IN INDUSTRIAL AND APPLIED MATHEMATICS: Proceedings of 23rd Malaysian National Symposium of Mathematical Sciences (SKSM23).

[5]. Anwar, J., Shah, S., and Saif, H. (2016). Business Strategy and Organizational Performance: Measures and Relationship. *Pakistan Economic and Social Review,* 54(1): 97-122.

[6]. Archibong, O. O., George, O. A., Jude, O., Ifeyinwa, M. H., Andrew, I. I. (2014). Application of Sarima Models in Modeling and Forecasting Nigeria’s Inflation Rates." *American Journal of Applied Mathematics and Statistics*, 2(1): 16-28.

[7]. Armstrong, J. S. and Grohman, M. C. (1972). A Comparative Study of Methods for Long-Range Market forecasting. *Management Science,* 9(2): 211-221.

[8]. Ayodele A. Adebiyi., and Aderemi O. Adewumi. (2014). Stock Price Prediction Using the ARIMA Model. *16th International Conference on Computer Modelling and Simulation*, USA.

[9]. Brockwell, P. J., and R. A. Davis (1995). *Time Series. Theory and Methods*. 2nd ed. Springer, New York.

[10]. Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and Methods*. 2nd ed. Springer Science + Business Media, LLC, New York.

[11]. Brockwell, P. J. and Davis, R. (2014). Modeling and Forecasting of Gold Prices on Financial Markets. *American international journal of Contemporary Research*, 4(3): 56-73.

[12]. Box, G. E. P., Jenkins, G. M., Reinsel, C. (1994). Time Series Analysis, Forecasting and Control. Englewood Cliffs: Prentice Hall.

[13]. Box, G. E. P.; Pierce, D. A. (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models". *Journal of the American Statistical Association*. 65: 1509–1526.

[14]. Box, G. E. P. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, USA.

[15]. Chandran, K. P. and Pandey, N. K. (2007). Potato Price Forecasting Using Seasonal ARIMA Approach. Central Potato Research Institute, Shimla-171 001, HP, *Indian Potato Journal*, 34(1): 137-138.

[16]. Chang, J., et al. (2015). An ancestral role in peroxisome assembly is retained by the divisional peroxin pex11 in the yeast Yarrowia lippolytica. *Journal of Cell Sceinces*, 128(7): 1327-40.

[17]. Chatfield, C. (2000). Time-Series Forecasting. CHAPMAN & HALL/CRC. Boca Raton London New York Washington, D.C.

[18]. DAWN (2016). Tomato Crash. Article published in Daily DAWN on 18.03.2016.

[19]. Dieng Alioune (2014). Alternative Forecasting Techniques for Vegetable Prices in Senegal. *Revue Senegalese recherché agncoles et agroalimentaires*, 1(3): 21-32.

[20]. Ender, W. (2010). Applied Time Series Econometrics. 3rd ed. John & Wiley Sons, USA.

[21]. Garcia-Martinnez. C. and Martinez, E. L. (2011). Price trends in greenhouse tomato and pepper and choice of adoptable technology. *Spanish Journal of Agricultural Research,* 6(3), 320-332.

[22]. Gathondu, E. K. (2001). Modeling of Wholesale Prices for Selected Vegetables Using Time Series Models in Kenya. University of Nairobi College of Biological and Physical Sciences School of Mathematics.

[23]. GoP. (2011). Ministry of Food, Agriculture and Livestock, Pakistan Agricultural Research (PAR), Islamabad, Pakistan.

[24]. Guha, B. and Bandyopadhyay, G. (2016). Gold Price Forecasting Using ARIMA Model. *Journal of Advanced Management Science,* 4(2): 117-121.

[25]. Hamilton, J. D. (2014). Time Series Analysis. 1st ed. Princeton University Press, USA.

[26]. Harrison E. E., Aboke, I. A., Edema, U. V., Dimkpa, M. Y. (2014). An additive seasonal Box-Jenkins model for Nigerian monthly savings deposit rates. *Issues in Business Management and Economics,* 2(3): 054-059.

[27]. Hossain, M. M. and Abdullah, F. (2015). On the production, behaviors, and forecasting the tomatoes production in Bangladesh. Department of Statistics, Jahangir Agar University, Savar, Dhaka-1342, Bangladesh. Department of Statistics, Islamic University, Kushtia-7003, Bangladesh.

[28]. Hyndman, R. J. (2008). Forecasting: Principles and Practice. 3rd ed. Springer, Australia.

[29]. Lewbell, A. and Serena, Ng. (2002). Demand Systems with Non-stationary Prices. Revised Edition, John & Wiley Sons, USA.

[30]. Lind, D., Marchal, W., Wathen, S., and Waite, C.A. (2009). Basic Statistics for Business and Economics: 3rd ed. Time Series and Forecasting, Chapter 16, ISBN: 0070980357.

[31]. MINFAL (2013) Agriculture Statistics of Pakistan. Govt. of Pakistan, Ministry of Food, Agriculture and Livestock. Economic Wing, Islamabad.

[32]. Ljung, G. M. and G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". *Biometrika*. 65 (2): 297–303.

[33]. Mondal, P. et al. (2014). Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications* (IJCSEA), 4(2): 36-49.

[34]. Nochai, R. and Nochai, T. (2006). ARIMA Model For Forecasting Oil Palm Price. Department of Agribusiness Administration, Faculty of Agricultural Technology, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, 10520 Thailand

[35]. Phillips, P. C. B.; Perron, P. (1988). "Testing for a Unit Root in Time Series Regression". *Biometrika*. 75 (2): 335–346.

[36]. Shahwan, T. and Lemke, F. (2005) Forecasting Commodity Prices for Predictive Decision Support Systems. Humboldt-Universität zu Berlin, School of Business and Economics, Institute of Banking, Stock Exchanges and Insurance, D-10178 Berlin, Germany.

[37]. Shukla, M. and Jharkharia, S. (2011). Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation. Proceedings of the International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia. January 22 – 24, 2011.

[38]. Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and Its Applications With R Examples*. 2nd ed. Springer Science + Business Media, LLC, New York.

[39]. Singh, M., Singh, R. and Shinde, V. (2011). Application of software packages for monthly stream flow forecasting of Kangsabati River in India. *International Journal of Computer Applications*, 20 (3), 7-14.

[40]. Swanson, N. and White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13(4): 439-461.

[41]. Wang, S., Lim, G., Yang, L., Zeng, Q., Sung, B., Martin,.J. J. A., and Mao, J. (2005). A rat model of unilateral hind paw burn injury: slowly developing rightwards shift of the morphine dose response curve. *Pain*, 116(1-2): 87-95.

[42]. Wooldridge, J. (2006). Introductory Econometrics Paperback- International Edition. Springer Science + Business Media, LLC, New York.

[43]. XIN1 W and Can2 W (2016) Empirical Study on Agricultural Products Price Forecasting based Internet-based Timely Price Information. *International Journal of Advanced Science and Technology* Vol.87, pp.31-36.