Forecasting of Potato Prices of Hooghly in West Bengal: Time Series Analysis Using SARIMA Model

Debasis Mithiya[1, *], Kumarjit Mandal[2], Lakshmikanta Datta[3]

[1]Department of Business Administration, International School of Hospitality Management, Kolkata, India

[2]Department of Economics, University of Calcutta, Kolkata, India

[3]Department of Statistics, International Institute of Management Sciences, Kolkata, India

Email address:

drdebasis.mithiya@gmail.com (D. Mithiya), kumarjitm@hotmail.com (K. Mandal), lakshmikantadatta@gmail.com (L. Datta)
[*]Corresponding author

Abstract: Potato is one of the most important food crops in India. Potato is also one of the principal cash crop, it gives handsome returns to the farmers. West Bengal is a state in India where potato is one of the major agricultural crops. But potato markets are more volatile due to fluctuation in production rather than Stable. So price of potato is fluctuating in nature. Thus, the time series analysis and price forecast may help producers in acreage allocation and timing of sale of potato. The present study was conducted to know the statistical investigation of price behaviour of potato in Hooghly district of West Bengal. In the present study, Box-Jenkins Seasonal Auto Regressive Integrated Moving Average (SARIMA) modeling is deployed in forecasting of monthly average price of potato in Hooghly of West Bengal up to October 2020 based on data from November 2008 to October 2018 (a period of 120 months). Seasonal indices calculated showed that generally the price is low from January to April and it starts picking up from May and reaches the maximum in November. The best model has been selected based on the Bayesian Information Criteria (BIC), Root Means Square Error (RMSE), Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE) and highest R-Square. The estimated best SARIMA model is $(1,1,0)(4,1,0)_{12}$. Root Mean Square Error (RMSE), Mean Average Percentage Error (MAPE), Mean Absolute Error (MAE) and BIC for Hooghly district are 184.10, 19.08, 118.98 and 10.69 respectively. Short term forecasts based on this model are close to the observed values and the behaviour of forecasted price of potato truly reflected the actual price as well as market tendency.

Keywords: Box-Jenkins, Forecasting, Monthly Average Price, Price Behavior, SARIMA Model, Seasonality

1. Introduction

Potato is one of the principal cash crops, it gives handsome returns to the farmers due to its wide market demand nationally and internationally for different kinds of utilization. Further, it has been reported by the International Food Policy Research Institute (IFPRI) and International Potato Centre (IPC), India is likely to have the highest growth rate of potato production and productivity during 1993 to 2020. During the same period, demand for potato is expected to rise by 40 percent worldwide. This indicates that a picture about a clear opportunity to capture the huge domestic and international market of potato by producing quality potato and its products [1].

Potato is grown in more than 100 countries in the world with a production of around 3768.27million tonnes during the year 2016-17. China (991.224 million tonnes) ranks first while India (437.70 million tonnes) and Russia (311.07 million tonnes) ranks second and third respectively. In India, potato is cultivated in almost all states and under very diverse agro climate conditions. The states of Uttar Pradesh, West Bengal, Bihar, Gujarat and Madhya Pradesh accounted for more than 80 per cent share in total production (https://agmarknet.gov.in/Others/profile-potato.pdf). Uttar Pradesh (155.43 million tonnes) and West Bengal (110.53 million tonnes) ranks first and second respectively during 2016-17 followed by Bihar and Gujarat.

West Bengal produces 110.53 million tonnes of potato

under around 460 thousand hectares area in 2016-17. It is 22.74 percent of Indian production. Under tropical and sub-tropical conditions, the losses due to poor handling and storage are reported to be in between 40-50 per cent. The post-harvest losses of potatoes are defined as qualitative and quantitative losses. The qualitative losses greatly reduce the price of potatoes. Potato prices fluctuate over seasons due to the variations in production and market arrival. Price fluctuations are a matter of concern among consumers, farmers and policy makers and its accurate forecast is extremely important for efficient monitoring and planning. Several attempts have been made in the past to develop price forecast models for various commodities [2, 3]. The prices of potato fluctuate to a great extent mainly because of its supply side and increasing demand at domestic and global level. Thus, the price forecast may help to producers in acreage allocation and time of sale. Naturally, forecasting is one of the main aspects of time series analysis having the art of saying that what will happen in the future. There are various forecasting models is use now-a-days. Analyst can choose their own method of forecasting based on their knowledge and available external information. As the process goes on, this procedure can be modified to meet the conditions and to satisfy the current situation. Different forecasting models may fit more or less equally well to the data, but they forecasts different future values [4].

In West Bengal, sowing time of potato is in November - December. The peak time for arrival is March but it starts in small quantity by January. However, arrivals in the market continue all over the year gradually. Thus, model building and forecasting the monthly price behavior of potato over the years is of much practical importance [5]. In this context, autoregressive integrated moving average (ARIMA) methodology has been successful in describing and forecasting the price dynamics of a potato.

In the present study, the ARIMA model has been used for describing monthly average price of potato in Hooghly district [1] of West Bengal. There are various purposes for which the present analysis of time series is performed. The objectives may include prediction of future prices based on knowledge of the past, control of a process to produce the time series, to obtain an understanding of the mechanism of generating the series. Study of trend helps to compare the long-term behavior of prices. An understanding of periodically moving cycles helps those concerned with planning and development; use of the potato price specifically in the area of short-term and long- term period. The technique of Box-Jenkins model is used to analyze non-stationary and seasonal data [6].

We have used Seasonal Autoregressive Integrated Moving Average (SARIMA) models to forecast monthly average

price of potato of Hooghly in West Bengal (the preliminary understating about the nature of data showed that there is an increasing trend over the time period and presence of seasonality is identified). The forecast accuracy criteria like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) has been used as the selection criteria to determine the best forecasting time series model. The specific objectives of the study are as follows:

I. To analyze the trend in monthly average prices of potato of Hooghly in West Bengal.
II. To forecast the monthly average prices of potato of Hooghly.

2. Material and Method

The time series data has been collected for the monthly average price (Rs./quintal) of potato in Hooghly for the time periods of November 2008 to October 2018. Hooghly is one of the major potato producing districts in West Bengal. It produces nearly 25 percent of total potato output in West Bengal. Hooghly is the main controlling district of the potato market in West Bengal which regulates the potato price. Champadanga is one of the largest potato markets in Hooghly District of West Bengal. The price of potato prevailing in this market has been used as the dummy of price of potato in Hooghly. The data has been collected from the website of Agmarknet (http://agmarknet.gov.in) [7]. Standard computer packages SPSS is used to finding the estimates of relevant parameters and forecasting.

Autoregressive Integrated Moving Average (ARIMA)

The model used in this study is the autoregressive integrated moving average (ARIMA). The ARIMA is an extrapolation (techniques make forecasts using only the past data) method, which requires historical time series data of the underlying variable. The model in specific and general forms may be expressed as follows.

Let Y_t is a discrete time series variable which takes different values over a period of time. The corresponding AR (p) model of Y_t series, which is the generalizations of autoregressive model, can be expressed as: AR (p),

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \qquad (1)$$

Where, Y_t is the response variables at time t,
Y_{t-1}, Y_{t-2}, …… Y_{t-p} is the respective variables at different time with lags;

μ is the constant mean of the series, ϕ_1, ϕ_2,…, ϕ_p are the coefficients; and ε_t is the error factor. ε_t is a white noise process, where $E(\varepsilon_t) = 0$, var $(\varepsilon_t) = \sigma^2 > 0$, cov $(\varepsilon_t, \varepsilon_{t-h}) = 0$, t, h $\neq 0$

Similarly, the MA (q) model which is again the generalization of moving average model may be specified as: MA (q),

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad (2)$$

Where, μ is the constant mean of the series, θ_1, θ_2,… θ_q is

---

[1] Hooghly district is one of the major potato producing districts in West Bengal. It produces nearly 25percent of total potato output in West Bengal. Hooghly is the min controlling district of potato market in West Bengal which regulates the potato price. In Hooghly, the Champadanga is one of the largest potato markets. The price of potato prevailing in this market has been used as the dummy of price of potato in Hooghly.

the coefficients of the estimated error term; ε_t is the error term.

By combining both the models, we get the Autoregressive Moving Average or ARMA models, which have general form as:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad (3)$$

Box and Jenkins argue that a non-stationary series can be transformed either into a stationary or an almost stationary series if it is differenced an appropriate number of times. Thus, if we have a stochastic process $\{Y_t, t = 0, \pm 1, \pm 2,\ldots\}$ which is non-stationary and has a trend, we can find a positive integer 'd' such that the transformed series $W_t = \nabla^d Y_t$ becomes stationary, ∇ being the difference operator, viz. $\nabla Y_t = Y_t - Y_{t-1}$, $\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$ and so on. After transforming the series into a stationary or to an almost stationary series, the model transforms to ARIMA.

If Y_t is stationary at the level or I(0) or at first difference I(1) or at the second difference I(2) determines the order of integration. After the stationary of the series was attained, ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) of the stationary series are employed to select the order p and q of the ARIMA model. The parameters were estimated using the non-linear least square method as suggested by Box and Jenkins, 1976 [8]. ε_t is a white noise process, where $E(\varepsilon_t) = 0$, var $(\varepsilon_t) = \sigma^2 > 0$, cor $(\varepsilon_t, \varepsilon_{t-h}) = 0$, t, h $\neq 0$. Based on the model diagnostic tests and parsimony we obtained the best fitting ARIMA (p, d, q) model. The mathematical equation, involving Y_t and ε_t that summarizes the ARIMA (p, d, q) model as defined in Equation (4):

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) \varepsilon_t \qquad (4)$$

Where, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$$

If a time series is seasonal of period s, Box and Jenkins made a proposal that such a model may be defined as Equation (5)

$$\phi_p(B) \varphi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta(B) \delta(B^s) \varepsilon_t \qquad (5)$$

where, B is the backshift operator (i.e. $BY_t = Y_{t-1}$, $B^2 Y_t = Y_{t-2}$, $B^s Y_t = Y_{t-s}$ and so on),'s' the seasonal lag and 'ε' is a sequence of independent normal error variables with mean 0 and variance σ^2. Φ's and φ's are the non-seasonal and seasonal autoregressive parameters respectively. δ's and θ's are seasonal and non-seasonal moving average parameters respectively. p and q are orders of non-seasonal autoregressive and moving average parameters, whereas P and Q are that of the seasonal auto regressive and moving average parameters. Also 'd' and 'D' denote non-seasonal and seasonal differences, respectively. In its general form, the Seasonal ARIMA (SARIMA) model is characterized by a notation as SARIMA (p, d, q) (P, D, Q)$_s$.

The complete procedure of model building and forecasting are fully described by Box and Jenkins 1976 [8]. In short, they have suggested four basic steps viz., (i) Identification of the model, (ii) Estimation of parameters of the model, (iii) Diagnostic Checking of the model, and (iv) Forecasting. The details of the estimation and forecasting process are discussed below.

2.1. Identification

The first step of applying the Box-Jenkins forecasting model is to identify the appropriate order of SARIMA (p, d, q)(P, D, Q)$_s$ model. Therefore, the identification process is to find the initial values for the orders of seasonal and non-seasonal parameters, p, q, and P, Q. That could be obtained by looking for significant autocorrelation and partial autocorrelation coefficients (Yet another application of the autocorrelation function is to determine whether the data contains a strong seasonal component). This phenomenon is established if the autocorrelation coefficients at lags between t and t-12 are significant). The order of d and D are estimated through I(1) or I(s) process of unit root stationary tests. The model specification and selection of order p, P and q, Q involved plotting of autocorrelations functions (ACF) and partial autocorrelations functions (PACF) or correlogram of the series at different lag length. If the PACF displays a sharp cutoff while the ACF decays more slowly (i.e., has significant spikes at higher lags), we say that the series displays an AR signature. However, if the ACF displays a sharp cutoff while the PACF decay more slowly, we say that the series displays an MA signature [9]. The autocorrelation functions specify the order of moving average process, and partial autocorrelations function to select the order of the autoregressive process.

2.2. Estimation of the Model

At the identification stage one or more Seasonal ARIMA models are tentatively chosen that seem to provide statistically adequate representations of the available data. The next step is to specify an appropriate regression model and estimate it. Seasonal ARIMA models are fitted and accuracy of the model was tested based on diagnostics statistics. Then we attempt to obtain precise estimates of parameters of the model by nonlinear least square method.

2.3. Diagnostic Checking

Now a question may arise that how we know whether the identified model is appropriate. One simple way to figure that out is by diagnostic checking the residual term obtained from SARIMA model by applying the same ACF and PACF functions. First obtaining the ACF and PACF of residual term up to certain lags of the estimated SARIMA model, and then to check whether the coefficients are statistically significant or not. The best model was selected based on the following diagnostics:

1. Low Akaike Information Criteria (AIC): AIC is

estimated by AIC $= -2\log_e (L) + 2m$, where $m = p + q +P+Q$ and L is the likelihood function.

2. Low Bayesian Information Criteria (BIC): The Bayesian information criterion is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to Akaike information criterion (AIC). Sometimes, Bayesian Information Criteria (BIC) is also used and estimated by BIC $= -2\log_e (L) + \log_e (N)\, m$. Where N is number of observation and m is the number of parameters

3. The minimum Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE) are used as a measure of accuracy of the models. RMSE$=\sqrt{\sum_{t=1}^{n}(X_{Actual,t} - X_{Forecast,t})^2 /n}$ and MAPE $= \frac{1}{n}\sum_{t=1}^{n}[\frac{X_{Actual,t} - X_{Forecast,t}}{X_{Forecast,t}}]^2 x\ 100,$

Where, $X_{Actual,t}$ and $X_{Forecast,t}$ are actual and forecast output at time t,

These may also be judged by Ljung-Box Q (LBQ) statistic[2] under null hypothesis that autocorrelation co-efficient up to lag k is equal to zero. LBQ is used to assess assumptions after fitting a time series model (SARIMA), to ensure that the residuals are independent.

2.4. Forecasting

Once the first three steps of seasonal ARIMA model are over, then we can obtain the forecasted values by estimating the appropriate model, which is free from problems. The forecasted values are reported for a maximum of 5 years, as long-term forecasting might not be appropriate.

3. Results and Discussion

The descriptive statistics like maximum price, minimum price, mean price, of the time series of monthly average price of potato in Hooghly has been found as Rs.1836.67 per quintal, Rs.230.19 per quintal, Rs.872.09 per quintal respectively. The standard deviation of the series is 433.20. In time series analysis one of the major tasks is to identify the patterns that exist in the data. These patterns may be trend, seasonality, cyclicality and random variability. Trend can be identified and confirmed by fitting a trend line using the data set.

Trend can be judge whether the value of variable may increase or decrease in a particular time. The seasonality, defined as a structured pattern of changes within a year, is the subject matter. Seasonality in a time series is a regular pattern of changes that repeats over S time periods, where S defines the number of time periods until the pattern repeats again.

---

[2]The Ljung-Box Q statistic to test whether a series of observations over time are random and independent. If observations are not independent, one observation can be correlated with a different observation k time units later, a relationship called autocorrelation. Autocorrelation can decrease the accuracy of a time-based predictive model, such as time series plot, and lead to misinterpretation of the data.

The preliminary understating about the nature of data showed that there is an increasing trend over the time period (Figure 1). The presence of seasonality in time series is identified by the analysis of auto correlation function values after the trend has been removed.
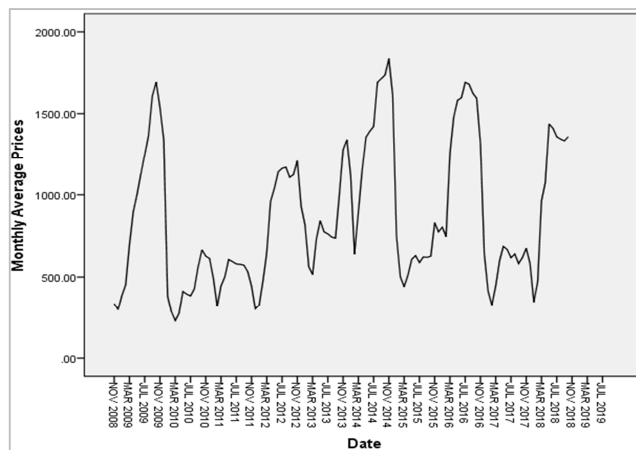

Figure 1. *Time series plots of potato price in Hooghly.*

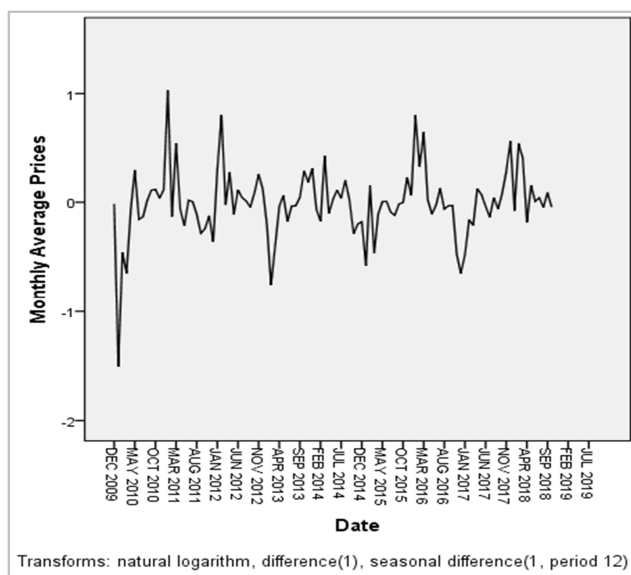Transforms: natural logarithm, difference(1), seasonal difference(1, period 12)

Figure 2. *Differenced time series plots of potato price in Hooghly.*

The values of auto correlation function presented in Table 1 have been derived from our observations. It clearly indicates that the series has seasonality.

Table 1. *The Autocorrelation Function of Monthly Average Prices after non seasonal differences.*

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistics		
			Value	df	Sig.[b]
1	0.399	0.91	19.464	1	0.000
2	0.078	0.90	20.206	2	0.000
3	-0.129	0.90	22.271	3	0.000
4	-0.169	0.89	25.861	4	0.000
5	-0.071	0.89	26.497	5	0.000
6	-0.023	0.89	26.567	6	0.000
7	-0.064	0.88	27.094	7	0.000
8	-0.218	0.88	33.279	8	0.000

| | | | Box-Ljung Statistics | | |
Lag	Autocorrelation	Std. Error[a]	Value	df	Sig.[b]
9	-0.293	0.87	44.485	9	0.000
10	-0.175	0.87	48.518	10	0.000
11	0.130	0.87	50.755	11	0.000
12	0.223	0.86	57.459	12	0.000
13	0.111	0.86	59.141	13	0.000
14	-0.199	0.85	64.592	14	0.000
15	-0.277	0.85	75.229	15	0.000
16	-0.147	0.85	78.234	16	0.000

a. The underlying process assumed is independence (white noise).
b. Based on the asymptotic chi-square approximation.

3.1. Model Selection

ARIMA model is generally applied for stationary time series data. Stationary vs. non-stationary can check through correlogram or autocorrelation functions. The general procedure to convert a non-stationary series to a stationary series is through difference. The series has both trend and seasonality present, the study may need to apply both a non-seasonal first difference and a seasonal first difference. A non-seasonal (d=1) and seasonal (D=1) first difference will "de-trend" and "de-seasonal" the data respectively. The sequence of the differenced series is presented in Figure 2.

The differenced time series data plotted in Figure 2 shows a fluctuation over constant mean, It implies that series is stationary. A seasonal (i.e. 12-point) differencing of monthly average price (Figure 2) shows a constant variance, a negative spike in 2009 being counterbalanced by a comparably positive one in 2010. Similarly, the negative spikes in 2012, 2014 and 2016 have been counter balanced by 2013, 2015 and 2017 respectively. The auto-correlation coefficients are dying out slowly after a non-seasonal first difference and a seasonal first difference, so the series is stationary. The figure confirms that the de-trend and deseasonalized price are nearly stable and the SARIMA model (p, 1, q) (P, 1, Q)$_{12}$ could be identified for further analysis [10].

The identification step is to tentatively choose one or more SARIMA model(s) using the estimated ACF and PACF plots. The ACF plot of the AR (Auto-Regressive)/ SAR (Seasonal Auto-Regressive) process shows an exponential decay while its PACF plot truncates at lag p /seasonal lag P and diminishes to zero afterward. The ACF plot of the MA process truncates to zero after lag q / seasonal lag Q while its PACF decays exponentially to zero. The two processes: AR (p)/SAR (P) and MA (q)/SMA (Q), could be combined to form the ARMA (p, q)/SARMA (P, Q) process which has ACF and PACF that decays exponentially to zero. The non-linear least square estimation method could be used to estimate the parameters of the identified model(s) in the identification stage. The last diagnostic checking stage involves assessing the adequacy of the identified and fitted models through a possible statistically significant test on the residuals to verify its consistency with the white noise process e.g. the Ljung-Box test [11]. The ACF and PACF plots of the SARIMA model (p, 1, q)(P, 1, Q)$_{12}$ with first

order non seasonal and seasonal differencing (Figure 3 & 4) suggested that at the initial stage the tentative model has been identified as (2,1,0)(1,1,0)$_{12}$ in Hooghly (the coefficients of ACF shows exponential decay, whereas coefficients of PACF spike at lag 2 after that values tend to zero and thereafter a spike at seasonal lag 1).
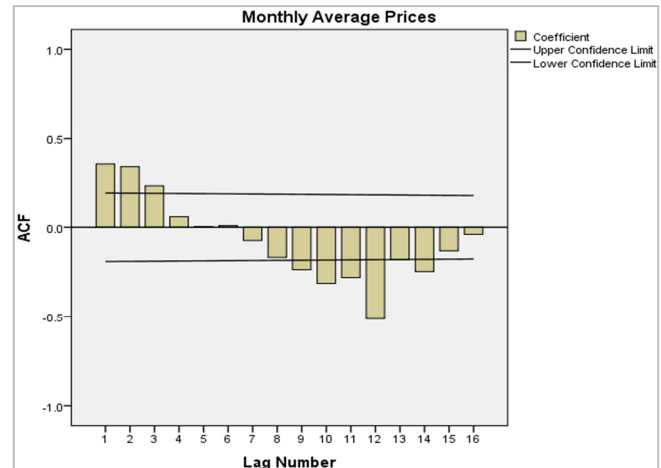

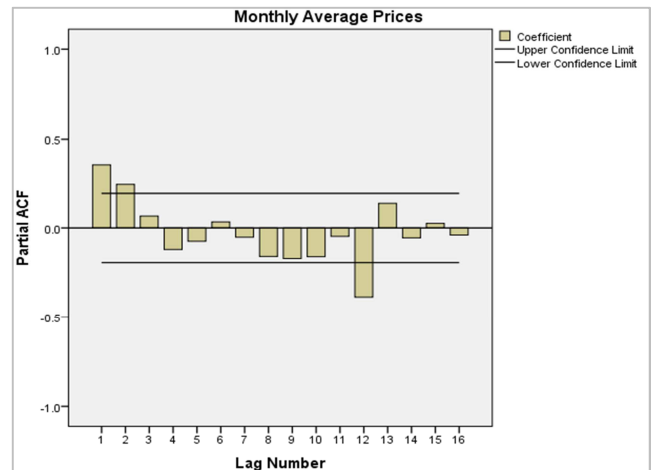*Figure 3. ACF plots after differentiation in Hooghly.*

Figure 4. PACF plots after differentiation in Hooghly.

The alternative models are also selected by inspecting and considered for the principle of parsimony. The diagnostic checks were applied in order to determine whether the residual of alternative models were independent, homoscedasticity, and normally distributed.

3.2. Estimation of the Model

After the identification process has completed, the possible selected model is identified SARIMA (2,1,0)(1,1,0)$_{12}$. However, the coefficient of AR (2) is not statistically significant. Hence in addition to this model, the study also attempts to estimate SARIMA (1,1,0)(2,1,0)$_{12}$, SARIMA (1,1,0)(3,1,0)$_{12}$, SARIMA (0,1,1)(4,1,0)$_{12}$ and SARIMA (1,1,0)(4,1,0)$_{12}$ models. All the estimated parameters of AR are statistically significant. The results of these SARIMA models are summarized in Table 2.

***Table 2.** Coefficients of estimated values of fitted ARIMA models.*

Sl. No	Variable	Model	Constant	AR (1)	AR (2)	MA (1)	SAR (1)	SAR (2)	SAR (3)	SAR (4)
		SARIMA $(2,1,0)(1,1,0)_{12}$	-3.85	0.236	0.106	-	-0.563	-	-	-
1	Price	SE	19.248	0.099	0.098	-	0.085	-	-	-
		t- value	-0.2	2.389	1.078	-	-6.598	-	-	-
2	Price	SARIMA $(1,1,0)(2,1,0)_{12}$	-0.939	0.299	-	-	-0.729	-0.304	-	-
		SE	13.856	0.096	-	-	0.100	0.103	-	-
		t- value	-0.068	3.103	-	-	-7.284	-2.941	-	-
		SARIMA $(1,1,0)(3,1,0)_{12}$	-2.588	0.282	-	-	-0.796	-0.498	-0.308	-
3	Price	SE	10.719	0.098	-	-	0.100	0.125	0.117	-
		t- value	-0.241	2.891	-	-	-7.939	-3.970	-2.634	-
4	Price	SARIMA $(0,1,1)(4,1,0)_{12}$	-2.505	-	-	-0.212	-0.990	-0.778	-0.735	-0.524
		SE	5.801	-	-	0.101	0.091	0.131	0.127	0.108
		t- value	-0.432	-	-	-2.09	-10.908	-5.580	-5.80	-4.83
5	Price	SARIMA $(1,1,0)(4,1,0)_{12}$	-2.515	0.252	-	-	-0.967	-0.774	-0.717	-0.514
		SE	6.462	0.101	-	-	0.091	0.132	0.126	0.110
		t- value	-0.389	2.491	-	-	-10.566	-5.869	-5.678	-4.666

Source: Authors' Calculation

Among the selected models, only in the case of SARIMA $(1,1,0)(4,1,0)_{12}$, ACF and PACF of residuals are within the range shown in Figure 5.
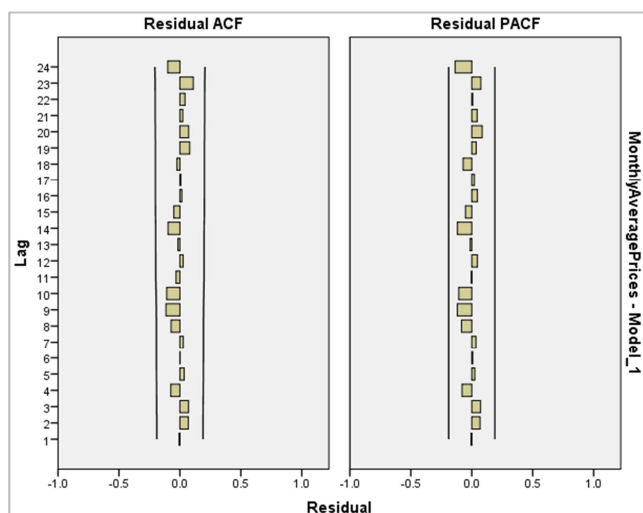
***Figure 5.** The Residual ACF and Residual PACF of SARIMA $(1,1,0)(4,1,0)_{12}$.*

The study has analyzed the statistical significance of these possible models. The best model is selected based on the diagnostics checking.

3.3. Diagnostic Checking

Now the question may arise that how we know whether these identified model is appropriate. After the estimation of parameters, we test the adequacy of the model based on Box-Pierce (Q) and Ljung-Box (LB) statistics. The statistics are calculated from the ACF of residual term up to 16 lags of the estimated SARIMA model. We also check the statistical significance of the parameters. An adequate model does not always generate good forecasts. Further, we select the model having low Bayesian Information Criteria (BIC), the lowest root means square error (RMSE), the lowest mean absolute percent error (MAPE), and the highest R-Square.

Comparing these five models, the SARIMA $(1,1,0)(4,1,0)_{12}$ model is found to be the best for potato price in Hooghly district. In this model, the estimated coefficients are statistically significant. LB and Q statistics of the model is also statistically significant. At the same time, RMSE, MAPE, MAE and BIC of SARIMA $(1,1,0)(4,1,0)_{12}$ have shown a value lower than that of other models. The summary of the statistical data of all these SARIMA models is given in Table 3.

***Table 3.** RMSE, MAPE, BIC values and Q statistics of fitted SARIMA models.*

Sl. No	Variable	Model	RMSE	MAPE	MAE	BIC	Stationary R^2	R^2	Ljung Box Q Statistics	Df
1	Price	SARIMA $(1,1,0)(4,1,0)_{12}$	184.10	19.08	118.98	10.69	0.49	0.82	7.43	13
2	Price	SARIMA $(0,1,1)(4,1,0)_{12}$	185.90	19.47	119.92	10.71	0.48	0.82	8.698	13
3	Price	SARIMA $(1,1,0)(3,1,0)_{12}$	196.127	20.347	125.57	10.77	0.42	0.79	10.48	14
4	Price	SARIMA $(1,1,0)(2,1,0)_{12}$	200.620	21.17	130.91	10.78	0.39	0.78	8.03	15
5	Price	SARIMA $(2,1,0)(1,1,0)_{12}$	206.46	21.51	132.50	10.84	0.35	0.77	8.38	15

Source: Authors' calculation

Based on the parameter estimates in table 2 and model statistics presented in the table 3, we have selected the SARIMA $(1,1,0)(4,1,0)_{12}$ as the best model for the potato price of Hooghly district in West Bengal. The model is as follows:

$$\phi_1(B)\ \phi_4\ (B^{12})(1\text{-}B)^1(1\text{-}B^{12})^1 Y_t = \varepsilon_t \qquad (6)$$

This model is a special case of the SARIMA model, which is called a *Seasonal Integrated Auto Regressive Model*.

3.4. Forecasting

Using the identified model SARIMA $(1,1,0)(4,1,0)_{12}$, the study has forecasted the price of potato in Hooghly for period of November 2018 to October 2020. The results of forecasted price presented in table 4. The behaviour of potato price has shown in Figure 6. In this figure, the time is measured along the horizontal axis and the vertical axis measures the level of monthly average prices (Rs./quintal).

Table 4. Forecasting monthly average price (Rs./quintal) of fitted SARIMA models.

Month	Forecast	UCL	LCL	Month	Forecast	UCL	LCL
Nov-18	1451.48	1762.42	1140.55	Nov-19	1335.56	2802.58	-131.47
Dec-18	1284.66	1782.96	786.35	Dec-19	1136.32	2664.74	-392.10
Jan-19	928.30	1573.09	283.51	Jan-20	617.96	2205.58	-969.66
Feb-19	611.46	1377.82	-154.90	Feb-20	478.72	2123.45	-1166.02
Mar-19	788.22	1659.95	-83.51	Mar-20	647.73	2347.66	-1052.21
Apr-19	968.64	1934.45	2.84	Apr-20	756.12	2509.53	-997.29
May-19	1120.35	2171.88	68.82	May-20	896.31	2701.61	-908.98
Jun-19	1129.97	2260.75	-0.81	Jun-20	894.70	2750.43	-961.04
Jul-19	1120.47	2325.30	-84.37	Jul-20	869.00	2773.83	-1035.84
Aug-19	1259.19	2533.78	-15.40	Aug-20	877.20	2829.91	-1075.51
Sep-19	1245.43	2586.15	-95.29	Sep-20	849.36	2848.79	-1150.07
Oct-19	1255.10	2658.84	-148.65	Oct-20	844.22	2889.31	-1200.87

Source: Authors' calculation

The results reveal that the average price of potato is the highest in the month of November and the lowest in February. In general, average prices of potato are comparatively lower in January, February, and March.

However, the price climbs up during the month of May and gradually move up during the remaining months. The behaviour of the forecasted price of potato truly reflected the actual price as well as a market tendency.
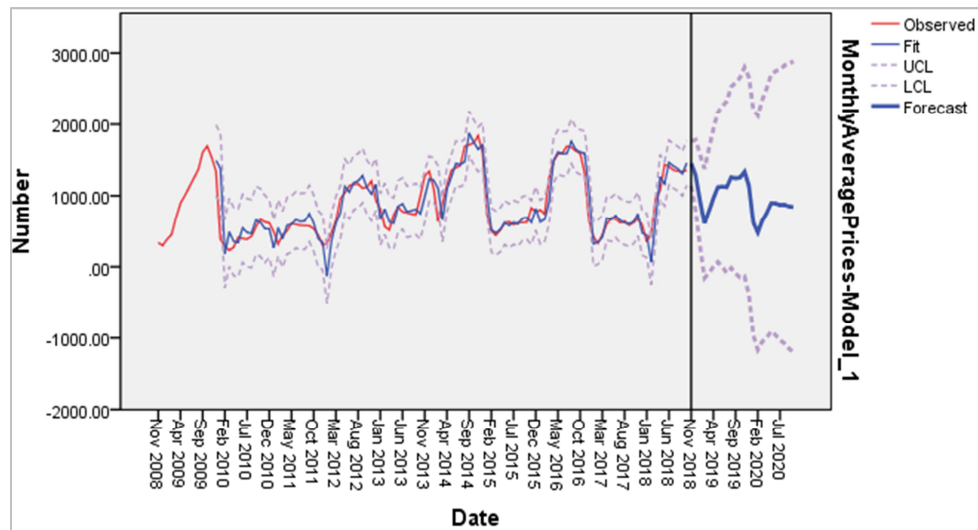

Figure 6. Series Chart of Observed, Fit and Forecasted value.

4. Policy Implication

The findings of the study have been suggested that the farmer suffers from various constraints, which must be removed if their financial position is to be strengthened. Some of the measures that could be adopted to achieve this result are indicated below [12].

(a) The wider and frequent fluctuations in wholesale prices, wide variation in arrivals etc., affected the returns to the potato grower. In order to encourage the farmers to continue in potato product, the price should be stabilized by the potato marketing cooperatives and minimum and maximum prices for the potato to be fixed.

(b) The wide seasonal fluctuation in arrivals of the produce has a consequential unfavourable impact on prices in a

regulated market over different months of the year. The huge quantity of arrivals during post-harvest months of the year lead to a decline in prices. The effective use of warehousing facilities and credit to the producer-seller against warehouse receipts would go a long way in avoiding seasonal variation in arrivals and prices.

(c) The regulated market should take necessary steps to oversee the dissemination of the market information regarding the prices. It should reach to the farmers of the remote places.

(d) With the help of the SARIMA model, the future price will be available. The forecast prices show an increasing trend, with due consideration to seasonality. In this regard, farmers may be advised to plan the production process and decide when to sell the produce, so that they would get a higher price for their produce. *The prices during September, October, November and December months has been* observed to be high and farmers can plan to sell their produce during these months.

(e) Since the potato is mainly used in making 'chips' which is a processed product. So the establishment of processing units may provide value addition to potato as indicated a lower number of processing industries. This would help the farmer to get better income, reduce the price fluctuation, and alternatively trigger the interest of the farmer to produce a good quality product.

5. Conclusion

The univariate time series model developed by Box-Jenkins whether Seasonal or non-seasonal is found to be good. In this study, an ARIMA model incorporates the seasonality of time series. Using the time series data of monthly average price of potato of Hooghly in West Bengal, the study build a Seasonal ARIMA $(1,1,0)$ $(4,1,0)_{12}$ model. It could be successfully used for modeling as well as forecasting of average monthly price of potato in Hooghly. The model demonstrated a good performance in terms of explained variability and predicting power. The behaviour of the forecasted price of potato truly reflected the actual price as well as market tendency. The relevant forecast interval for the average price can help both the potato farmers as well as the planners for future planning. Though we have used the univariate model to predict the future values, researches can use bivariate or multivariate model taking policy variables as an exogenous variable to improve the prediction of the price of potato.

References

[1]   U. Singh, C. S. Praharaj, S. S. Singh and N. P. Singh, *Biofortification of Food Crops*, Ummed Singh, CS Praharaj, SS Singh, NP Singh Edited; Spinger; 2016.

[2]   H. Ghosh and Prajneshu, "Non-linear time series modelling of volatile onion price data using AR (p)-ARCH (q)-in- mean", *Calcutta Statistical Association Bulletin*, 54, 231-47, 2003.

[3]   T. L. Mohankumar, H. B. Mallikarjuna, N. M. Singh, C. S. Satish Gowda, "Comparative study of univariate time series technique for forecasting of onion price", *International Journal of Commerce and Business Management*, 4 (2), 304-308, 2011.

[4]   Hemant Sharma and S. S. "Burark, Bajra Price Forecasting in Chomu Market of Jaipur District: An Application of SARIMA Model", *Agricultural Situation in India*, 61 (11), 2015.

[5]   K. P. Chandran and N. K. Pandey, "Potato price forecasting using seasonal ARIMA approach", *Potato Journal*, 34 (1-2), 137-138, 2007.

[6]   D. S. Dhakre and D. Bhattacharya, "Price Behaviour of Potato in Agra Market A Statistical Analysis", *Indian Research Journal Extension Education*, 14 (2), 2014.

[7]   Agricultural Marketing Information Network-Agmarknet, http://agmarknet.gov.in, downloaded on 29.11.2018.

[8]   G. E. P. Box, and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco. 1976,

[9]   M. O. Nasiru and S. O, "Olanrewaju Forecasting Airline Fatalities in the World Using a Univariate Time Series Model", *International Journal of Statistics and Applications*, 5 (5), 223-230, 2015.

[10] S. C. Hillmer and G. C. Tiao, "An ARIMA model based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, 72 (377), 1982.

[11] A. C. Akpanta, I. E. Okorie, N. N. Okoye, "SARIMA Modelling of the Frequency of Monthly Rainfall in Umuahia Abia State of Nigeria", *American Journal of Mathematics and Statistics*, 5 (2), 82-87, 2015.

[12] H. Gangadharappa, "Statistical study of variation in arrivals and prices of potato in the selected markets of Karnataka". M. Sc. (Ag.) Thesis, University of Agricultural Science, Dharwad, India, 2005.