Forecasting of Egg Prices data using Time Series models-I

Srikanth.C[1] and Prof. A. Rajendra Prasad[2]

1. Research Scholar, Department of Statistics, Kakatiya University, Warangal, Telangana Email:srikanthc222@gmail.com
2. Professor, Department of Statistics, Kakatiya University, Warangal, Telangana, Email:anchurirp@gmail.com

1. Abstract:

The work presented in this research paper constitutes a contribution to modeling and forecasting the monthly average egg prices of Hyderabad city using a time series models like Simple Exponential Smoothing methodology and Auto Regressive Integrated Moving Average model using Box-Jenkins methodology. The adequate model is selected according to five performance criteria; Akaike Information Criterion, Schwartz Bayesian Criterion, Mean Absolute Error, Mean Absolute and Percentage Error and Root Mean Square Error. The selected model is ARIMA(1,1,1)(0,1,1)12 and it is validated by another historical average egg prices information under the same conditions. The results obtained prove that the model could be utilized to forecast the average egg prices in Hyderabad city.

Keywords: time series, auto regressive integrated moving average, simple exponential smoothing, error measures.

2. Introduction:

Eggs can be enjoyed as part of healthy and balanced diet. Eggs provide us with very high-quality protein that contains all nine essential amino acids in the right amounts needed by the body for optimum growth and maintenance. Now a day's egg is eating by everybody and these are most nutritious food on the earth. It is price most reasonable among all the agriculture commodities. Egg price is generally fluctuating by yearly, monthly and daily also. Forecasting egg prices is a most complex phenomenon, mainly depends on market demand and supply.

A historical data of monthly average egg prices of 100 units from January-2010 to September - 2020 of Hyderabad city are collected from National Egg Co-ordination Committee (NECC). We have discussed various time series models in this paper like Box-Jenkins ARIMA, Simple Exponential Smoothing, Holt's linear trend and Holt-Winters models for forecasting the monthly average egg prices.

3. Literature Review:

Ibina E.O., Igwe N.O., Oyah M.P. and Okonta C.A(2020), used ARIMA models for forecasting the stock market prices of Benne cement and Akshara Cement in Nigeria. Sudeshna Gosh(2017), proposed ARIMA models to forecast the cotton exports of India. Chukwudike C. Nwokike, Bright C. Offorha, Maxwell Obubu, Chukwuma B. Ugoala, Henry I. Ukomah (2020), used SARIMA models for forecasting the monthly rainfall in Nigeria. Kumar Manoj and Anand Madhu (2012), used Box-Jenkins ARIMA models to forecast sugarcane production in India. Jamal Fattah, Latifa Ezzine, Zineb Aman ,Haj El Moussami, and Abdeslam Lachhab (2018), proposed modeling and forecasting the demand in a food company using ARIMA models.

4. Methodology:

4.1. Simple Exponential Smoothing:

This method used past data points to forecast the future data. It gives more weight to recent values. Past values are smoothed like moving averages.

$$S_t = \alpha Y_t + (1-\alpha) S_{t-1} \qquad ; 0< \alpha <1$$

Where α is the smoothing parameter

S_t is the period t's forecast value

Y_t is the actual value in time t

S_{t-1} is the forecast value for t-1

Exponential smoothing is best for short term forecasts without trend or seasonality. A bigger α means more weight is given to recent past data points. Try different values of α and compute Root Mean Square Error (RMSE). Choose the α with lowest RMSE.

4.2. Box-Jenkins ARIMA Model:

An Auto Regressive Integrated Moving Average model is labeled as an ARIMA(p,d,q), Where in

p - is the number of Auto regressive terms

d - is number of differences

q - is the number of moving average terms

The general forecasting equation of ARMA(p,q) model is

$$\hat{y}_t = \mu + \phi_1 y_{t-1} +\ldots+ \phi_p y_{t-p} - \theta_1 e_{t-1} -\ldots- \theta_q e_{t-q}$$

The Box-Jenkins iterative approach is used for constructing the linear time series model. This approach is consisting of 5 steps.

Step1: Stationary of the data

Draw the time series graph and Auto Correlation Function graph to the given time series data. If the trend line is parallel to X-axis and variability is uniform in time series graph and ACF dies out for higher lags then the data is in stationary. If ACF does not dies out for higher lags then the data is in non stationary.

If the data is non-stationary convert it into stationary using appropriate transformation and successive differencing stabilizing variance and mean respectively.

Step2: Model Identification

Determine model parameters p and q using ACF and PACF graphs for the stationary data.

1. If ACF dies out for higher lags and q-spikes in the ACF graph then the model is MA(q) model.
2. If PACF dies out for higher lags and p-spikes in the PACF graph then the model is AR(p) model.
3. If ACF and PACF both dies out for higher lags, q-spikes in ACF and p-spikes in PACF graphs then the model is ARMA(p,q) model.

Step3: Estimation of model parameters

Parameters of the model identified in step2 can be estimated using Lease Square Estimation or Maximum Likelihood Estimation methods.

Step4: Diagnostic Checking or testing the model adequacy

The Ljung-Box test or ACF plot for residuals used to check whether the model is adequate. This test is need for testing for the randomness of the residuals.

If the model is adequate then go to step5 otherwise go to step2 and review the model.

Step5: Forecasting future values

Identify the best model using the error measures like Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and forecast the future values.

5. Results and Discussion:

5.1. Analysis of data

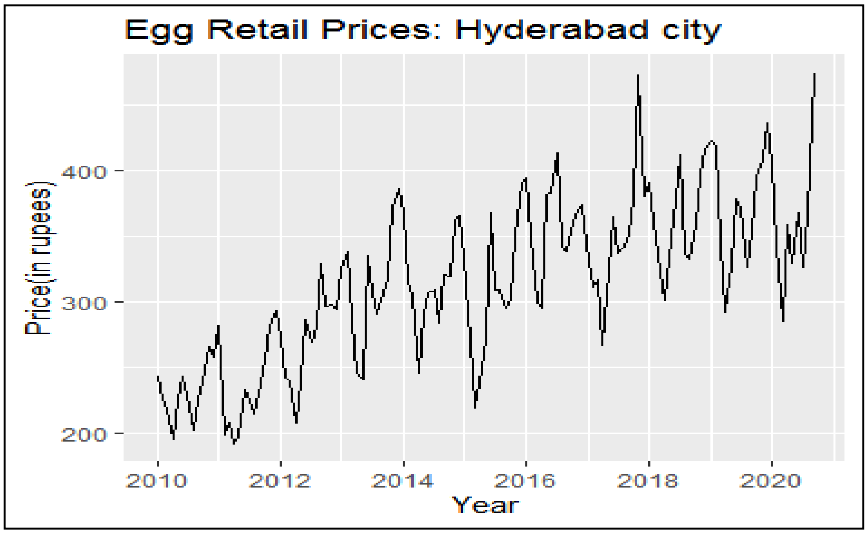

Figure1: Monthly Average egg retail prices in Hyderabad city

Figure1, the time series plot of monthly Average egg retail prices of 100 eggs in Hyderabad city from Jan-2010 to Sep-2020. The data shows an increment in the prices from 2010 to 2020. Any time series data can be decomposed in to trend, seasonal, cyclic and random/irregular components. The additive decomposition of Monthly Average egg retail prices data is shown in figure2.

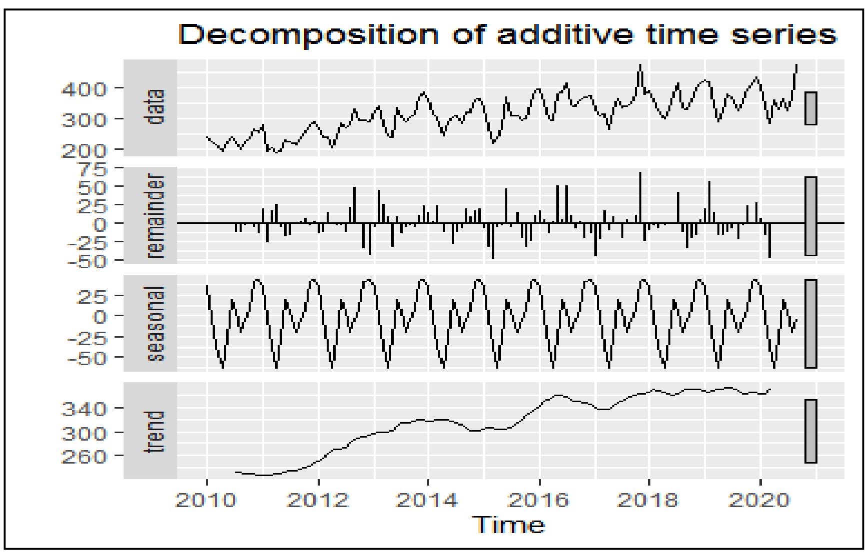

Figure2: Monthly Average egg retail prices (top) and its three additive components

The three components are shown separately in the bottom three panels of Figure2. These components can be added together to reconstruct the data shown in the top panel. Notice that the seasonal component changes slowly over time, so that any two consecutive years have similar patterns, but years far apart may have different seasonal patterns. The remainder component shown in the panel is what is left over when the seasonal and trend-cycle components have been subtracted from the data.

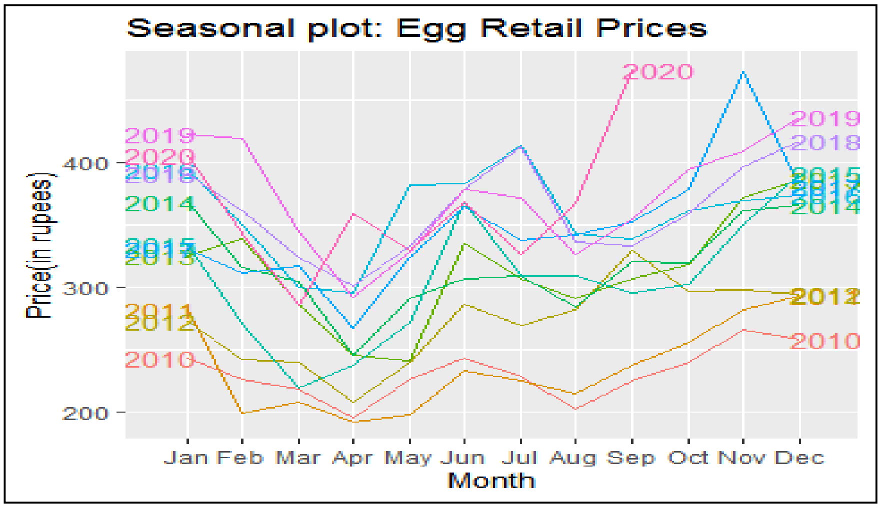

Figure3: Seasonal plot of monthly average egg retail prices in Hyderabad city

In Seasonal plot the data are plotted against the individual seasons. A seasonal plot allows the underlying seasonal pattern to be seen more clearly, and is especially useful in identifying years in which the pattern changes.

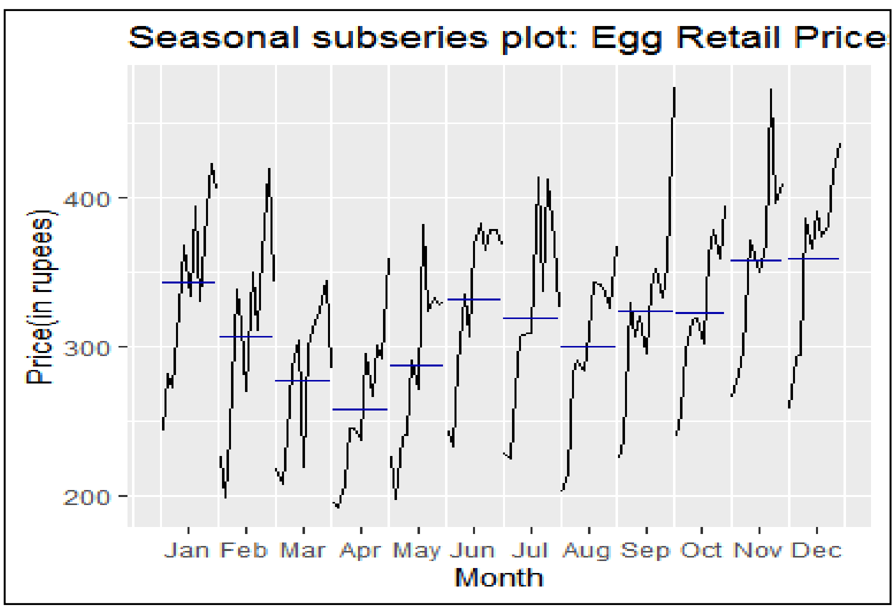

Figure4: Seasonal subseries plot of monthly average egg retail prices in Hyderabad city

The horizontal lines indicate the means for each month. This form of plot enables the underlying seasonal pattern to be seen clearly, and also shows the changes in seasonality over time. It is especially useful in identifying changes within particular seasons. From figure3 and figure4, we can see that the average egg retail prices are least in April month, moderate in June to October months and high in November, December months.

We have divided whole data as training data (Jan-2010 to Sep-2018) and test data (Oct-2010 to Sep-2020). The model is to be developed on training data and validate on the test data.

5.2. Fitting Simple Exponential Smoothing model:

From the R-output, the fitted time series model from training data of Simple Exponential Smoothing method is

$S_t = 0.9501 * Y_t + 0.0499 * S_{t-1}$

Where S_t is the period t's forecast value

Y_t is the actual value in time t

S_{t-1} is the forecast value for t-1

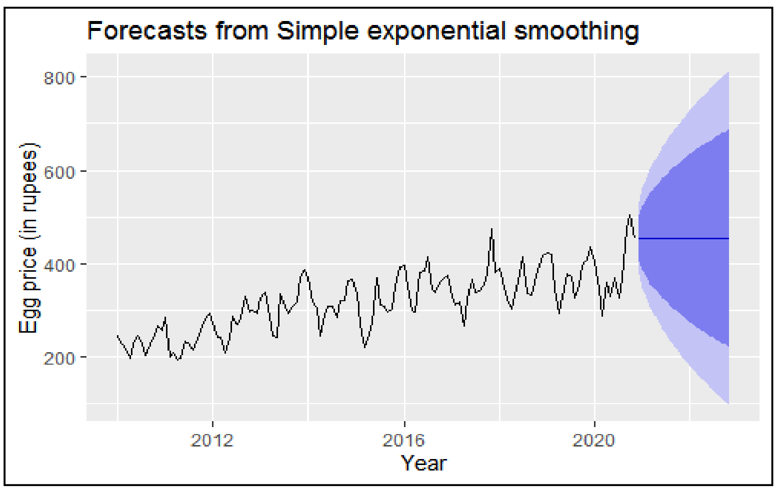

Figure5: Simple exponential smoothing method with forecasts

R-output:

```
Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = ts_train, h = 24)

  Smoothing parameters:
    alpha = 0.9501

  Initial states:
    l = 242.4408

  sigma:  37.2965
```

```
     AIC      AICc       BIC
1252.615 1252.853 1260.577
```

5.3. Fitting ARIMA model:

5.3.1. Stationarity of Data

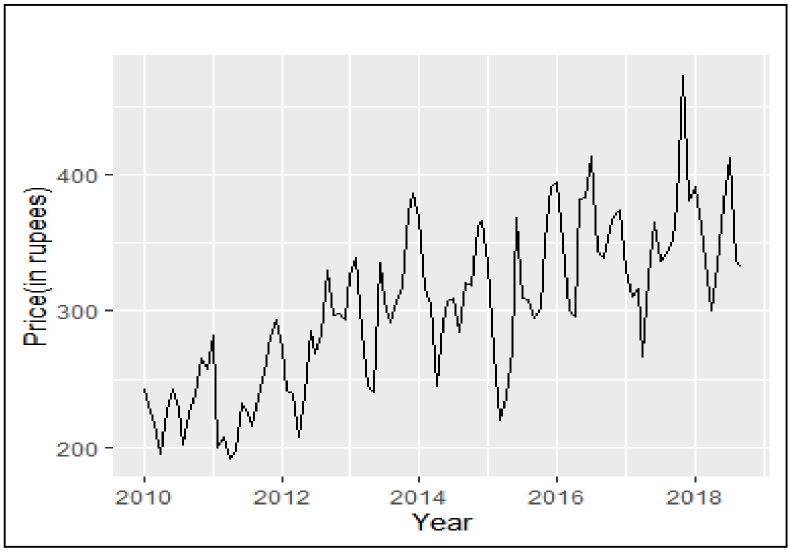

Figure6: Monthly retail prices of 100 eggs in Hyderabad from January-2010 to September-2018

Augmented Dicky-Fuller Test:

Null Hypothesis(H_0): Data is not in Stationary at p-value.

Alternative Hypothesis(H_1): Data is in Stationary at p-value.

R-output:

```
        Augmented Dickey-Fuller Test

data:  ts_train
Dickey-Fuller = -3.3219, Lag order = 4, p-value = 0.07125
alternative hypothesis: stationary
```

Interpretation: Here p-value is 0.07125 which is higher than significance level 0.05, so there is no evidence for rejecting the null hypothesis. Therefore data is not in stationary.

From figure6 and Augmented Dicky-Fuller test results, the data are clearly non-stationary with increasing trend and some seasonality. Apply differencing to covert the data into stationary form, so we will take seasonal difference.

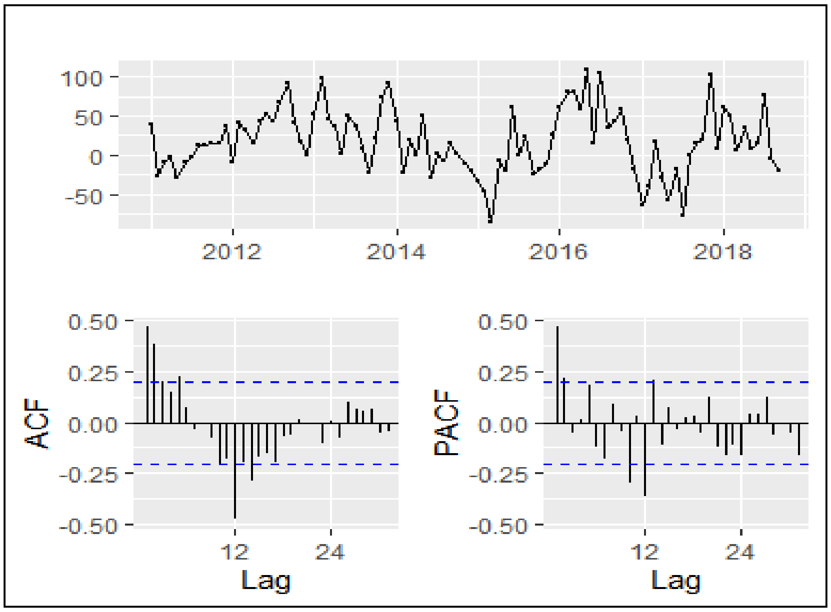

Figure7: First order seasonally differenced hyderabad egg retail price data

Augmented Dicky-Fuller Test for differenced data:

Null Hypothesis(H_0): Data is not in Stationary at p-value.

Alternative Hypothesis(H_1): Data is in Stationary at p-value.

R-output:

Augmented Dickey-Fuller Test

data: diff(ts_train, lag = 12)
Dickey-Fuller = -2.6795, Lag order = 4, p-value = 0.296
alternative hypothesis: stationary

Interpretation: Here p-value is 0.3303 which is higher than significance level 0.05, so there is no evidence for rejecting the null hypothesis. Therefore data is not in stationary.

From figure-7 and Augmented Dicky-Fuller test results, the data are still appears to be non-stationary, so we take an additional non seasonal first difference shown in figure6.

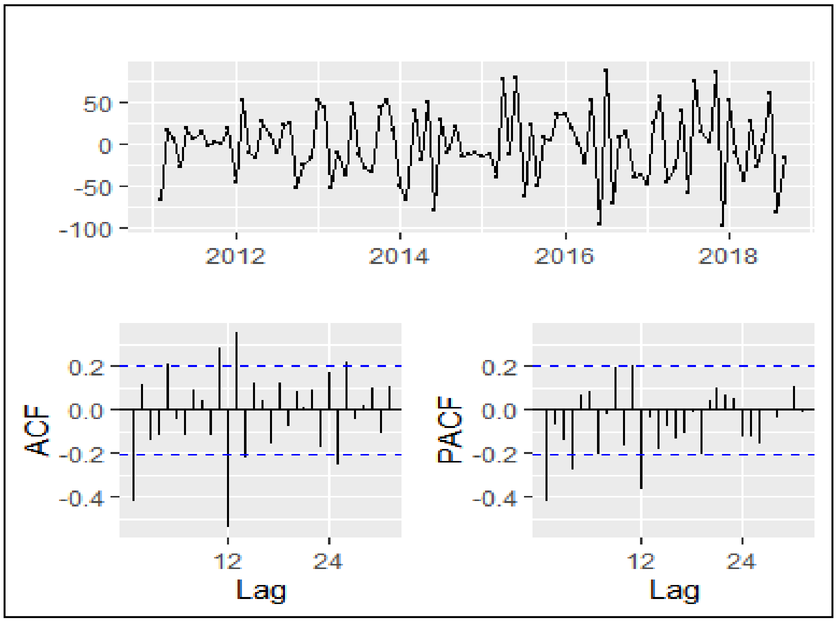

Figure8: First order non seasonal and first order seasonally differenced hyderabad egg retail price data

Augmented Dicky-Fuller Test for differenced data:

Null Hypothesis(H_0): Data is not in Stationary at p-value.

Alternative Hypothesis(H_1): Data is in Stationary at p-value.

R-output:

Augmented Dickey-Fuller Test

data: diff(diff(ts_train, lag = 12))
Dickey-Fuller = -5.4075, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Interpretation: Here p-value is 0.01 which is lesser than significance level 0.05, so reject null hypothesis. Therefore first order non seasonal and first order seasonally differenced hyderabad egg retail price data is in stationary.

5.3.2. Identification of the Model

We have to find an appropriate ARIMA model based on ACF and PACF shown in figure3. The significant spike at lag1 in the ACF suggests a non-seasonal MA(1) component, and the significant spikes at lag12,lag13 in the ACF suggests a seasonal MA(2) component s. The significant spikes at lag1 in the PACF suggests a non seasonal AR(1) component, and in the significant spikes at lag4,lag12 in the PACF suggests a seasonal AR(2) component . The pattern in the ACF and PACF is not indicative of any simple model.

We have computed error measures for training and test data of adequate models (satisfies L-jung Box test) using all the possible combinations of the above parameters shown in table1, table2.

S.No	Model	Error Measures					L-jung Box	
		RMSE	MAE	MAPE	AIC	BIC	Q-Statistic	P-Value
1	ARIMA(1,0,1)(2,0,0)12	28.51	22.27	7.45	1022.34	1038.27	25.99	0.054
2	ARIMA(1,0,1)(0,1,1)12	25.79	18.79	6.09	901.13	911.26	26.23	0.095
3	ARIMA(1,0,1)(1,1,0)12	28.25	21.27	7.03	909.51	919.64	26.29	0.093
4	ARIMA(1,1,1)(1,0,1)12	25.65	19.23	6.35	1000.91	1014.13	25.30	0.088
5	ARIMA(1,1,1)(1,0,2)12	26.16	19.50	6.44	1000.29	1016.16	24.37	0.082
6	ARIMA(1,1,1)(2,0,0)12	27.64	21.05	6.97	1004.77	1017.99	21.59	0.201
7	ARIMA(1,1,1)(2,0,1)12	26.07	19.39	6.41	1000.08	1015.95	23.49	0.101
8	ARIMA(1,1,1)(2,0,2)12	26.01	19.36	6.40	1002.06	1020.57	23.13	0.082
9	ARIMA(1,1,1)(0,1,1)12	23.25	16.50	5.21	887.56	897.64	24.99	0.125
10	ARIMA(1,1,1)(0,1,2)12	24.74	17.39	5.49	887.54	900.15	22.63	0.162
11	ARIMA(1,1,1)(1,1,0)12	27.33	19.73	6.24	897.45	907.54	23.68	0.166
12	ARIMA(1,1,1)(1,1,1)12	24.63	17.27	5.46	887.29	899.90	21.82	0.192
13	ARIMA(1,1,1)(1,1,2)12	24.47	17.20	5.44	889.20	904.33	21.24	0.170
14	ARIMA(1,1,1)(2,1,0)12	25.93	18.35	5.80	892.09	904.70	22.54	0.165
15	ARIMA(1,1,1)(2,1,1)12	24.36	17.11	5.41	889.19	904.32	21.10	0.175
16	ARIMA(1,1,1)(2,1,2)12	24.52	17.23	5.45	891.20	908.85	21.31	0.127
17	Simple Exponential Smoothing	36.94	28.85	9.61	1252.62	1260.58	31.45	0.174

Table1: Error Measures for various time series models applied to training data

From table1, we can see that ARIMA(1,1,1)(0,1,1)12 has smallest error measures as RMSE(23.25), MAPE(5.21), AIC and BIC values as compared with other models. Therefore ARIMA(1,1,1)(0,1,1)12 is best model.

5.3.3. Test set evaluation:

We will compare the models fitted in on training data using a test set consisting of the last two years data. Thus we fit the models using data from Jan-2010 to May-2018, and forecast the retail prices for oct-2018 to sept-2020. The results are summarized in table2.

S.No	Model	Error Measures		
		RMSE	MAE	MAPE
1	ARIMA(1,0,1)(0,1,1)12	38.64	25.78	7.39
2	ARIMA(1,1,1)(1,0,1)12	36.29	28.14	7.47
3	ARIMA(1,1,1)(0,1,1)12	37.47	29.68	7.76

4	ARIMA(1,1,1)(2,1,1)12	38.32	29.52	7.80
5	ARIMA(1,1,1)(1,1,2)12	38.46	29.66	7.84
6	ARIMA(1,1,1)(2,1,2)12	38.59	29.78	7.87
7	ARIMA(1,1,1)(2,0,2)12	38.10	29.31	7.87
8	ARIMA(1,1,1)(1,1,1)12	38.63	29.93	7.91
9	ARIMA(1,1,1)(2,0,1)12	38.18	29.53	7.94
10	ARIMA(1,1,1)(0,1,2)12	38.70	30.13	7.97
11	ARIMA(1,1,1)(1,0,2)12	38.27	29.76	8.02
12	ARIMA(1,1,1)(2,0,0)12	42.08	30.33	8.45
13	ARIMA(1,1,1)(2,1,0)12	41.28	32.54	8.59
14	ARIMA(1,1,1)(1,1,0)12	44.23	33.85	8.88
15	ARIMA(1,0,1)(1,1,0)12	46.90	33.15	9.55
16	ARIMA(1,0,1)(2,0,0)12	53.95	40.52	12.17
17	Simple Exponential Smoothing	36.93	47.49	14.26

Table2: Error Measures for various time series models applied to test data

From tabble2, ARIMA(1,1,1)(0,1,1)12 has the second lowest error measures (RMSE,MAPE) on test data, so ARIMA(1,1,1)(0,1,1)12 is the best model for forecasting the monthly retail egg prices.

5.3.4. Diagnostic checking or Residual Analysis:

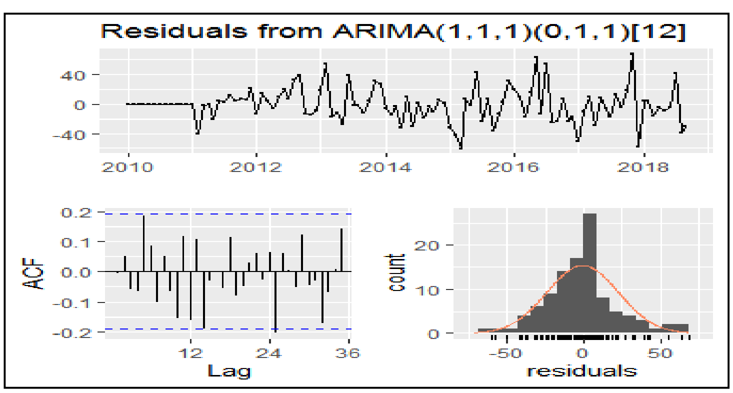

Figure9: Residual plots

From Figure9, We can observe that

1. Year vs Residual plot, Residuals are identical with mean zero and constant variance
2. Lag vs ACF plot, Residuals are uncorrelated
3. Residuals vs count plot, Residuals follows normal distribution

R-output:

```
Series: ts_train
ARIMA(1,1,1)(0,1,1)[12]

Coefficients:
         ar1      ma1     sma1
      0.4278  -0.9112  -0.9988
s.e.  0.1481   0.1061   0.9624

sigma^2 estimated as 637.5: log likelihood=-439.78
AIC=887.56 AICc=888.02 BIC=897.64
```

The ARIMA(1,1,1)(0,1,1)12 model equation is

$(1-\phi_1 B)(1-B^{12})(1-B)Y_t = (1-\theta_1 B)(1-\psi_1 B^{12})\,e_t$; Where B is the back shift operator: $BY_t = Y_{t-1}$

The Fitted ARIMA(1,1,1)(0,1,1)12 model equation is

$$\mathbf{(1-(0.4278)B)(1-B^{12})(1-B)Y_t = (1-(-0.9112)B)(1-(-0.9988)B^{12})\,e_t}$$

6. Conclusion:

The point forecasts of 24 months, 80% and 90% confidence interval values of point forecasts obtained using Seasonal ARIMA models are presented below.

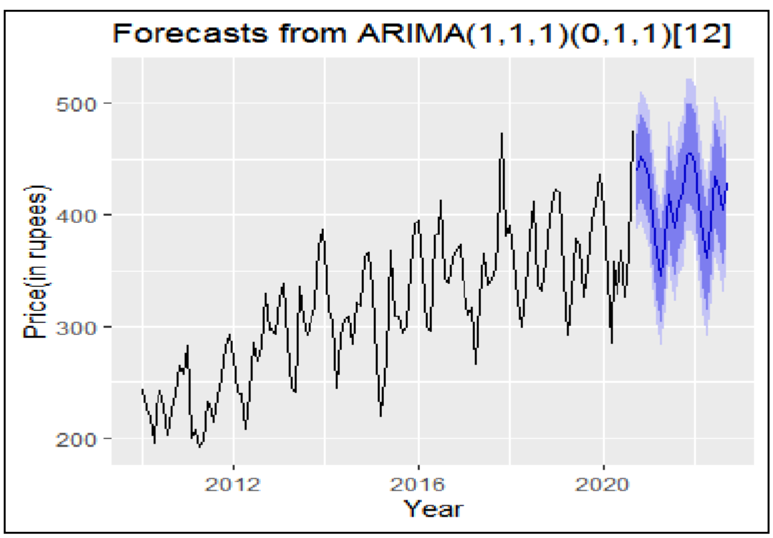

Figure10: ARIMA model with forecasts

Time	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Oct-20	438.37	404.44	472.30	386.48	490.26
Nov-20	452.56	414.36	490.76	394.14	510.98
Dec-20	.445.11	405.48	484.74	384.50	505.71
Jan-21	433.50	393.20	473.80	371.87	495.13
Feb-21	396.29	355.52	437.07	333.93	458.65
Mar-21	365.55	324.39	406.72	302.59	428.51
Apr-21	345.97	304.45	387.49	282.47	409.47
May-21	375.71	333.85	417.57	311.69	439.73
Jun-21	419.45	377.26	461.64	354.92	483.98
Jul-21	406.98	364.46	449.49	341.95	472.00
Aug-21	387.83	344.98	430.67	322.31	453.35
Sep-21	412.27	369.11	455.44	346.26	478.29
Oct-21	420.76	376.79	464.73	353.51	488.01
Nov-21	454.13	409.61	498.65	386.05	522.21
Dec-21	454.88	409.92	499.85	386.11	523.65
Jan-22	446.79	401.43	492.14	377.43	516.15
Feb-22	411.08	365.36	456.80	341.16	481.00
Mar-22	380.98	334.91	427.05	310.52	451.44
Apr-22	361.67	315.25	408.09	290.68	432.66
May-22	391.53	344.76	438.29	320.01	463.05
Jun-22	435.32	388.21	482.43	363.28	507.36
Jul-22	422.87	375.42	470.31	350.30	495.43
Aug-22	403.73	355.94	451.51	330.65	476.81
Sep-22	428.18	380.06	476.30	354.58	501.78

Table: 3 Forecasts of monthly average egg prices in Hyderabad city using Seasonal ARIMA model

7.References:

1. Aarti Mehta Sharma and Saina Baby-2015 : Gold price forecasting in India using ARIMA modeling
2. Banhi Guha and Gautam Bandyopadhyay, 2016 :Gold Price Forecasting Using ARIMA Model
3. Box, G. E. P., Jenkins, G.M. and Reinsel,G.C. 1994. Time Series Analysis,Forecasting and Control. , Prentice,Hall, 3rd edition.
4. Ette Harrison Etuk, Tariq Mahgoub Mohamed,2014:Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by Sarima Methods
5. K. Ranga Rao, Naveen Kumar and M. Krishna Reddy,2012: A new method of forecasting south-west monsoon rainfall
6. Mohammad Almasarweh & S. AL Wadi,2018: ARIMA Model in Predicting Banking Stock Market Data