Journal of Food Legumes 32(2): 109-112, 2019

Application of machine learning techniques in time series analysis of prices of pulses

SHRIPAD BHAT, HEMANT KUMAR, RAJESH KUMAR and NP SINGH

ICAR-Indian Institute of Pulses Research, Kanpur, India; Email: shripadsmail@gmail.com
(Received : January 07, 2019 ; Accepted : March 18, 2019)

ABSTRACT

Time series analysis of prices helps to capture the movement, trend and seasonality in price series which is useful to different stakeholders, such as farmers, consumers and policy makers. In order to model the structure of prices of pulses, monthly Wholesale Price Indices (WPI) from January 2005 to March 2019 consisting of 171 time series observations were collected from office of the Economic Adviser, Ministry of Commerce & Industry, Govt. of India. The WPI captures the changes in the price level at the initial stages of transaction and government periodically changes the base year to improve the representativeness. To make the WPI series comparable, linking factor were calculated using the average ratio of overlapping monthly prices. Using R software, time series decomposition was carried out to estimate trend and seasonal components in the price series. Seasonal indices revealed that price indices of pulses were on higher side during July to December months. Further, time series models were built to capture and predict the price indices of pulses.

Key words: ARIMA, Machine learning, Pulses, Wholesale Price Index

Price volatility has been one of problems faced by pulses growers, which directly affects profitability and cropping decisions. Though in recent years, large scale procurements of pulses at the minimum support prices were carried out by the government, still there was the problem of price fluctuations and decline in farm prices during the harvest seasons. It was reported that lack of assured markets and market imperfections affected the Indian pulses sector (Reddy, 2004 and Rahman, 2015). Revenue terms of trade were reported favorable to cereals compared to pulses (Srivastavaa *et al*. 2010). In this study, using the time series data on prices of pulses, an attempt was made to capture the movement, variation, trend and seasonality in price series which will be useful to understand price behaviour. Further, machine learning techniques were employed to decompose the price series and to choose the best model to capture the movements in prices of pulses in India.

MATERIALS AND METHODS

In India, inflation is being measured using two indices: Wholesale Price Index and Consumer Price Index. CPI captures the price change from the perspective of the consumer while the WPI measures the price changes at wholesale stage. The WPI is closer to farm prices and has more time series observation when compared to recently introduced CPI measure. Hence, monthly WPI for all the major pulses from January 2005 to March 2019 consisting of 171 time series observations were collected from Office of the Economic Adviser, Ministry of Commerce & Industry, Govt. of India (http://eaindustry.nic.in/home.asp). As the base year for WPI was changed during 2011-12, to covert the price series to base year 2004-05, linking factors were estimated by averaging the overlapping months (Table 1).

Table 1. Linking factors

	Pulses	Gram	Arhar	Moong	Masur	Urad
Average value (Apr-2012 to March-2017)	2.0	2.0	1.8	2.4	1.7	2.4

After converting the price indices to common base year (2004-05), classical decomposition by moving averages method was carried out to estimate trend, seasonality in prices. Further time series models were developed to capture the movement in prices of pulses. There are different ways with which performance of time series model can be measured: Root Mean Square Error, Mean Absolute Percentage Error, corrected Akaike information criterion. Since AICc measure helps to remove the effect of over fitting, this was used in selecting the best performing model among each category of models. The best performing Exponential Smoothing model and ARIMA models were automatically selected using forecast package in R software. Further, time series cross validation technique was also employed to choose the best performing model.

Holt's linear method with multiplicative errors and damped trend (Hyndman, RJ, and Athanasopoulos, G. 2018) is expressed as

$$y_t = (l_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$$
$$l_t = (l_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$$
$$b_t = \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1})\varepsilon_t$$

Where y_t is the price series

α is smoothing parameter for level

β is smoothing parameter for trend

ϕ is damping parameter

ε_t is white noise error term

This model is also denoted as ETS (M, A_d, N) where E stands for error, T for trend and S for seasonality, M for multiplicative error, A_d for damped trend, N for no seasonality.

A non-seasonal ARIMA model can be represented as

$$(1-\phi_1 B - \cdots - \phi_p B^p)(1-B)^d y_t = c + (1+\ \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

Where y_t denotes the price series (log of price series in case of log transformation)

B is backward shift operator,

p denotes the order of autoregressive terms

q denotes order of moving average terms

d denotes of order of non-seasonal differencing

ε_t is white noise error term

RESULTS AND DISCUSSION

Descriptive statistics for wholesale price indices of pulses from January 2005 to March 2019 with base year 2004-05 are given in Table 2. The coefficient of variation for price indices was greater than 30% across all major pulses during the study period. The price indices reached the maximum values during 2016-17 (Fig. 1). During this period, price levels of major pulses were about four to five times higher than the price levels during the base year 2004-05. After 2016-17, price indices across all major pulses showed declining trend.

The wholesale price indices for pulses were subjected to classical multiplicative decomposition method using moving average method. The time series is decomposed in to trend-cycle component, seasonal component and remainder component (Fig. 2). The trend component depicted the increasing trends in prices till 2016-17, which

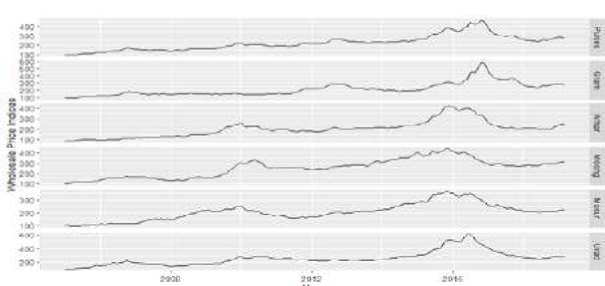

Figure 1. Monthly Wholesale Price Indices of major pulses in India (Jan-05 to Mar-19) (Base 2004-05)

later started to decline. The seasonal component showed the seasonal variation in prices of pulses across months. Using these seasonal component values, seasonal indices were also constructed.

Box plot representation of wholesale price indices of pulses in India (Jan-5 to Mar-19) across months is depicted

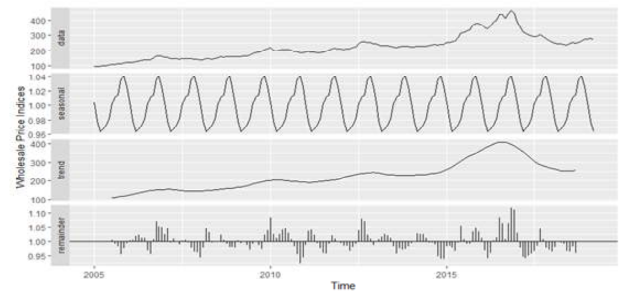

Fig. 2. Decomposition of multiplicative time series-WPI of pulses in India (Jan-5 to Mar-19)

in Figure 3. The box plot showed that the variation in prices were higher during January to May while it was comparatively lower during June to September months.

Seasonal indices capture the seasonal variations in prices across months. Using the seasonal components from

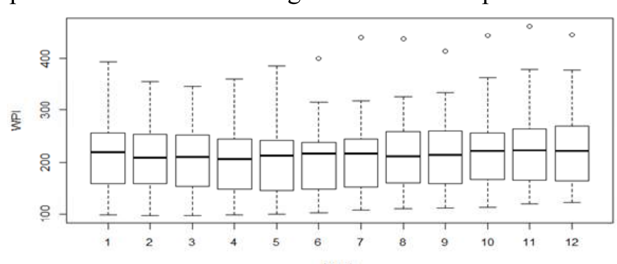

Fig 3. Wholesale Price Indices of pulses in India (Jan-5 to Mar-19)-Box plot

classical multiplicative decomposition, seasonal indices were calculated (Fig. 4). The seasonal indices were higher during August to December while during February to June, seasonal indices were lower.

Augmented Dickey-Fuller Test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test were employed to check for the stationary of price series. Both these tests indicated that WPI for pulses was not stationary. Hence log transformation was carried out and then first order differencing was applied. Results indicated that

Table 2. Descriptive statistics of wholesale price indices of pulses (Jan-05 to Mar-19)

Particulars	Pulses	Gram	Arhar	Moong	Masur	Urad
Minimum	97.1	97.7	88.5	99.7	92.8	102.0
First Quartile (Q_1)	159.2	150.5	142.3	160.7	161.1	179.9
Median	214.0	194.3	203.4	267.4	207.6	243.3
Mean	222.6	216.7	201.1	251.3	205.8	257.3
Third Quartile (Q_3)	256.6	268.2	225.2	305.9	236.0	291.4
Maximum	463.0	582.8	421.0	441.2	369.7	611.4
Standard deviation	80.1	91.7	76.6	87.6	67.0	104.7
Coefficient of variation (%)	36.0	42.3	38.1	34.9	32.6	40.7

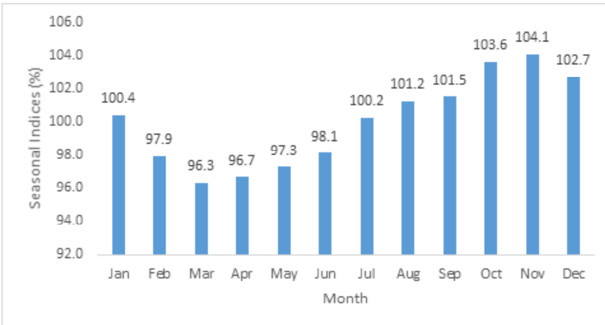

Fig 4. Seasonal Indices (%) for prices of pulses in India

transformed and differenced price series was stationary.

To capture the movement in prices of pulses, three types of models were fitted. Using the forecast package in R software, Holt's linear method with multiplicative errors and damped trend [ETS (M, Ad, N)], ARIMA model and ARIMA model with log transformation (Table 4) were fitted on price indices. Model parameters were automatically selected to fit the best performing models in each category. On log transformed price series, 192 ARIMA models were fitted and ARIMA (1, 1, 0) was found to be the best performing model as measured by corrected Akaike information criterion. This model performed better in terms of minimum Root Mean Square Error when compared to other two models. Ljung-Box test which tests the presence of autocorrelation in residuals also indicated that model 3 was the better model as test statistic (33.94) was not significant at 5 per cent level of significance. For this model 3, the residuals plot, Autocorrelation plot (ACF) and histograms are depicted in Fig. 5.

Further using the time series cross validation, these

Table 3. Tests for stationary of price series

Series	Augmented Dickey-Fuller Test		KPSS test	
	Test statistic	P- value	Test statistic	P- value
WPI Pulses	-2.5969	0.3270	2.6624	0.01
Log (WPI pulses)	-2.5623	0.3415	2.9646	0.01
Log (WPI pulses) with first order differencing (d=1)	-4.5526	0.01	0.20262	0.1

Table 4. Summary of the models fitted to wholesale price indices of pulses (Jan-5 to Mar-19)

Particulars	Model 1: ETS (M,A_d,N)	Model 2: ARIMA (0,1,2)	Model 3: ARIMA (1,1,0) with log transformation
Model coefficients	Smoothing parameters:	ma1	ar1
	alpha = 0.9999	0.6283	0.5163
	beta= 0.6388	(s.e. 0.0749)	(s.e. 0.0653)
	phi= 0.8	ma2	
	Initial states:	0.1576	
	l = 92.593	(s.e. 0.0768)	
	b = 2.3812		
AICc	1527.97	1217.33	-709.45
RMSE	9.08	8.49	8.72
Ljung-Box test (Q*)	51.61(p- value <0.001)	35.10 (p- value 0.0378)	33.94 (p- value 0.066)

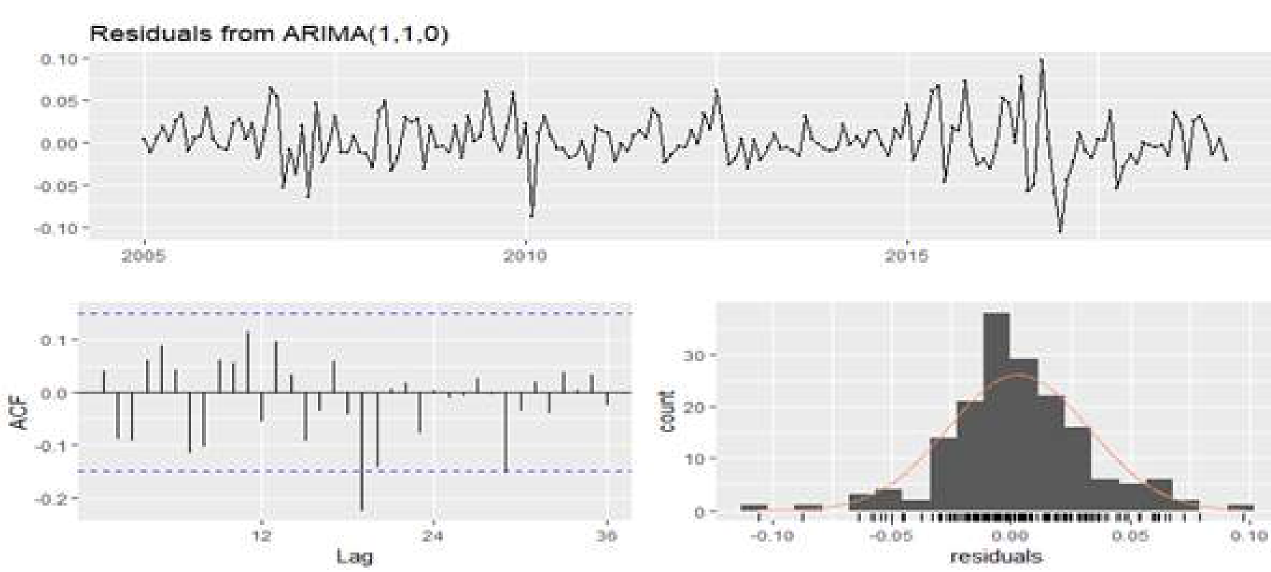

Fig. 5. Residuals from model 3 (ARIMA (1, 1, 0) with log transformation)

Table 5. Performance comparison of time series models using time series cross validation

Models	Mean squared error (h=1)	Mean squared error (h=2)	Mean squared error (h=3)	Overall Mean squared error
Model 1	89.59	321.01	608.92	339.84
Model 2	94.78	337.65	590.03	340.82
Model 3	85.08	311.23	590.03	328.78

three models were compared for three forecast horizons (h) and results are presented in Table 5. Overall mean squared error was lowest for model 3 (328.78) compared to model 1 (339.84) and model 2 (340.82).

The present study revealed that price indices of major pulses were on the declining trend after 2016-17. Results indicated that seasonal indices for price of pulses were higher during August to December months whereas seasonal indices were lower during February to June months. The variation in prices was higher during January to May while it was lower during June to September months. Out of 192 Autoregressive Integrated Moving Average (ARIMA) models fitted, ARIMA (1,1,0) model fitted on log transformed price indices performed better on corrected Akaike information criterion, residual diagnostic tests and time series cross validation.

REFERENCES

Hyndman RJ and Athanasopoulos G. 2018. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. O Texts.com/fpp2.

Rahman A. 2015. Is Price Transmission in the Indian Pulses Market Asymmetric? Journal of Quantitative Economics **13(1):** 129-146.

Reddy AA. 2004. Consumption pattern, trade and production potential of pulses. Economic and Political Weekly **9(44):** 4854-4860.

Srivastava SK, Sivaramane N and Mathur V.C. 2010. Diagnosis of pulses performance of India. Agricultural Economics Research Review **23(347-2016-17026):** 137.