P-ISSN: 2618-0723
E-ISSN: 2618-0731

NAAS Rating: 5.04
www.extensionjournal.com

International Journal of Agriculture Extension and Social Development

Volume 7; SP-Issue 3; March 2024; Page No. 140-145

Received: 01-12-2023
Accepted: 03-02-2024

Indexed Journal
Peer Reviewed Journal

Modelling and forecasting of onion prices in Belgaum market of Karnataka

[1]Vinay HT, [2]Pavithra V, [3]Nishtha Pradarshika Rai and [4]Jagadeesh MS

[1-3]Department of Agricultural Statistics, UBKV, West Bengal, India

[2] Division of Agricultural Economics, ICAR-Indian Agricultural Research Institute, New Delhi, Delhi, India

DOI: https://doi.org/10.33545/26180723.2024.v7.i3Sb.413

Corresponding Author: Vinay HT

Abstract

Onions are considered as one of the most sensitive commodities due to their susceptibility to sudden price fluctuations. Therefore, modelling and forecasting commodity prices, particularly for volatile vegetables like onions, are crucial. In our current study, we applied various statistical techniques, including Single Exponential Smoothing, Double Exponential Smoothing, Triple Exponential Smoothing, ARIMA, SARIMA, BATS, and TBATS models, to analyse onion prices from January 2010 to December 2023 in the Belgaum market of Karnataka state. The models were built using an 80-20 split, with 80% of the observations used to the training dataset and the remaining 20% to the testing dataset for model validation. Among the seven models assessed, the BATS model consistently outperformed the others, showcasing lower Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values. Additionally, the residuals displayed a random distribution, indicating the model's robustness. Consequently, the BATS model was selected for forecasting onion prices spanning from January 2024 to December 2024.

Keywords: Exponential smoothing, ARIMA, SARIMA, BATS and TBATS

1. Introduction

Crop cultivation is a major income-driving source for a country. India is well known for its diversity in the types of crops grown in a particular state of the country. Onion (*Alium cepa*) is an important ingredient of Indian cuisine grown widely for its remarkable flavor and health benefits. Onions contributed Rs. 246 billion in 2020 (Statista, 2023) [11] to the Indian economy with Rs. 4522.79 crores worth of onions being exported in 2022 (APEDA) [1]. Indian households consume 13 lakh tons of onion every month (CRISIL, 2021) [4]. Karnataka ranks 3rd in onion area, yield rate and production in India. The state of Karnataka produces 2779.500 thousand tons of onion (Directorate of Economics and Statistics, 2023) [7] in 231.84 thousand hectares of land (Department of Agriculture and Farmer's Welfare, 2021).

Price forecasting plays a major role in managing the economy of the nation. As far as the price of any agricultural commodity is concerned, it is known that it varies within different markets across the state of a country. Belgaum also known as Belagavi is a city with moderate year-round temperatures and adequate rainfall that favors the growth of onion. Onion can be grown in various types of climatic conditions but thrives well in mild weather without heavy fluctuations in environmental conditions. In Karnataka, onion is grown in three seasons, namely, Early Kharif, Kharif and Rabi. It is the role of business analysts, statisticians, data scientists, policymakers and agriculture scientists to analyze the ongoing scenario of an agricultural commodity to build an accurate understanding of the prices of the product and build up strategies and plans accordingly.

The price of a crop is highly volatile and is subject to two concrete factors: seasonality and weather conditions (CRISIL, 2021) [4]. Price drops directly impact the farmers when the price is lower than the cost of cultivation and on the other hand price spikes can disrupt consumer budgets. A forehand estimate can help to adapt to probable changes depending on the trend, seasonality or cyclical pattern. Forecasting using several statistical techniques empowers the analysts with the ability to make forehand preparations in times of prospective predicted changes.

This study is an endeavor to use statistical models to make forecasts and select the model that best explains the variations. Modelling is a crucial analytical tool in modern statistical analysis. Exponential smoothing models are forecasting methods that work by weighting the observed time series according to the incidence of observations. More weights are given to the more recent observations and less weight to past observations. In this analysis, several exponential smoothening methods such as Simple Exponential Smoothing, Double Exponential Smoothing and Triple Exponential Smoothing are employed. The classical time series Autoregressive Integrated Moving Average Model is used where the present observations are regressed on the past observations and error terms. Along with this, the seasonal modification of the ARIMA model: the Seasonal Autoregressive Integrated Moving Average Model has also been compared with other techniques for forecasting. These models are simple and easy to understand algebraically. Rajput *et al.*, 2022 [10] have compared ARIMA, ANN and exponential methods for forecasting cotton prices in different markets of Gujarat. In this study, monthly data of onion price in Rs. from 2010 to 2023 is investigated to predict the sale price in the future. This analysis compares the performance of traditional time series models.

2. Materials and Methods

2.1 Data description

The prices of onions from January 2010 to December 2023 were collected from the AGMARKNET website (http://agmarknet.gov.in), with a specific focus on the Belgaum market in Karnataka state. The data on prices refer to modal prices in a month, which are considered superior to monthly average prices as they represent the major proportion of the commodity marketed during the month.

2.2 Methods

The prices of onion were modelled using Single Exponential Smoothing, Double Exponential Smoothing, Triple Exponential Smoothing, ARIMA model, SARIMA model, BATS model and TBATS model.

2.21 Single exponential smoothing (SES)

In 1963, Brown [3] introduced a method to estimate future values using a single weight or parameter. This technique assigns more weightage to recent observations and less weightage to distant observations.

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

Where, α is a smoothing parameter taking values in the interval (0, 1).
F_t = the forecasted price at time t.
Y_t =actual price at time t.

2.22 Double exponential smoothing (DES)

It is also known as Holt's linear method. Simple exponential smoothing (SES) is not effective in predicting time series with a local linear trend to address this limitation, Holt [8] proposed an extension of SES called Holt's method. This method includes an additional updating equation for the slope (trend), resulting in improved forecasts for time series with a local linear trend.

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

$$F_{t+m} = L_t + mb_t$$

Where,
L_t= Level at time t
b_t= Trend at time t
F_{t+m}= Forecast value for m period ahead
α, β = Smoothing parameters ranging from 0 to 1. The combination of these parameters is selected based on minimum RMSE value.

2.3 Triple exponential smoothing (TES)

It is also known as Holt-Winter's Exponential Smoothing (H-WES) method. The Holt- Winter's method is based on three smoothing equations, one for level, one for trend, and one for seasonality. It is similar to Holt's method, with an additional equation to deal with seasonality. Holt-Winter's Exponential Smoothing (H-WES) methods are widely used when the data shows trend and seasonality. In this study the multiplicative model is used as the seasonal variation over time is observed. The four equations for the model are given as follows:

$$L_t = \alpha \frac{Y_t}{S_t} + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

$$S_t = \gamma \frac{Y_t}{L_t - Y_t} + (1-\gamma)S_{t-s}$$

$$F_{t+m} = L_t + mb_t + S_{t-s+m}$$

Where,
s= length of seasonality
L_t= Level at time t
b_t= Trend at time t
S_t= Seasonal component at time t
F_{t+m}= Forecast value for m period ahead
$\alpha,\ \beta\ and\ \gamma$ are level, trend and seasonal smoothing constants or the weights respectively, which lies between 0 and 1. The combination of these parameters is selected based on minimum RMSE and MAPE value.

2.2.4 ARIMA model

In an autoregressive integrated moving average model, the future value of a variable is assumed to be a linear function of several past observations and random errors (Box and Jenkins) [2]. Theoretically ARIMA model includes three components: Auto-Regressive (AR), Moving-Average (MA), and Integrated (I) terms. The first two components are expressed in equation

$$\nabla^d y_t = \underbrace{\emptyset_1 \nabla^d y_{t-1} + \cdots + \emptyset_p \nabla^d y_{t-p}} + \underbrace{\varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_p \varepsilon_{t-p}}$$

AR terms MA terms

Where ϕ is a number strictly between −1 and +1, and θ are the weights, and *p* is the order of the AR model, and *q* is the order of the MA model. Here, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 $\forall$ t=1,2, …, *n*.

2.2.4 SARIMA model

When time series data have seasonal component, SARIMA model is employed. SARIMA model is characterized by SARIMA (*p*, *d*, *q*) $(P, D, Q)_s$ Here, *p* and *q* are orders of non-seasonal autoregressive and moving average parameters respectively, whereas *P* and *Q* are the seasonal autoregressive and moving average parameters respectively. Also 'd' and '*D*' denote non-seasonal and seasonal differences respectively (Makridakis *et al.*, 2008) [9] and it is given by

$$(1-\varphi_p B)(1-\Phi_P B^s)(1-\mathrm{B})(1-B^s)y_t = (1-\theta_q B)(1-\Theta_Q B^s)\varepsilon_t$$

Where, B=backshift operator,
s =seasonal lag,
ε_t= sequence of error ~ N (0, σ^2),
Φ's and φ's = the seasonal and non-seasonal autoregressive parameters respectively
Θ's and θ's = the seasonal and non-seasonal moving average parameters.

2.2.5 BATS (B: Box-Cox transformation A: ARIMA errors T: Trend S: Seasonal components) model

Box-Cox transformation is a power transformation that helps make the series stationary, by stabilizing the variance and mean over time. BATS model is developed by extension of Double-Seasonal Holt-Winter's (DSHW) method with Box-Cox transformtion, ARMA errors, Trend, and multiple seasonal patterns (De Livera, 2010[6]).

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}; & \omega \neq 0 \\ \log y_t & \omega = 0 \end{cases}$$

$$y_t^{(\omega)} = l_{t-1} + \emptyset b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + d_t$$

$$l_t = l_{t-1} + \emptyset b_{t-1} + \alpha d_t$$

$$b_t = (1-\emptyset)b + \emptyset b_{t-1} + \beta d_t$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Where, $y_t^{(\omega)}$ represents Box–Cox transformed observations with a parameter ω at time t, $m_{1,\dots,}$ m_T denote the seasonal periods, l_t is the local level at time t, b is the long-run trend and b_t is the short-run trend at time t, $s_t^{(i)}$ indicates the i^{th} seasonal component at time t, d_t represents an ARMA (p, q) process, ε_t is a Gaussian white-noise process with zero mean and constant variance σ^2, and the smoothing parameters are given by α, β, and γ_i for i= 1, T. The model was represented by BATS (ω,(p, q), $\emptyset$, m_1, m_2, …, m_T), where, ω is the Box–Cox transformed value, (p, q) is ARMA components, $\emptyset$ dampening parameter, m_i represents i^{th} seasons.

2.2.6 TBATS (T: Trigonometric B: Box-Cox transformation A: ARIMA errors T: Trend S: Seasonal components) model: For high frequency and non-integer seasonality BATS model are not efficient, therefore, to overcome this problem, TBATS was introduced as an extension of BATS model by adapting the following equations (De Livera *et al.*, 2011) [5]:

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos\lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin\lambda_j^{(i)} + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin\lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos\lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

where, $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$ are the smoothing parameters, $\lambda_j^{(i)} = \frac{2\pi j}{m_i}$, $s_{j,t}^{(i)}$ describe the stochastic level of the i^{th} seasonal component, $s_{j,t}^{*(i)}$ describe the stochastic growth of the i^{th} seasonal component, k_i is the number of harmonics required for the i^{th} seasonal component, $k_i = \frac{m_i}{2}$ for even values of m_i, and $k_i = \frac{(m_i-1)}{2}$ for odd values of m_i.

2.2.6 Model evaluation criteria

The model performance was evaluated on two important criteria like Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

$$\text{RMSE}= \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right| * 100$$

where, n is number of observations, Y_t and $\hat{Y}_t$ are the actual and forecasted price at time t. lower value for these metrics denotes a more accurate prediction from the model.

3. Results and Discussion

3.1 Descriptive statistics

The descriptive statistics like Mean, Median, Maximum, Minimum Standard deviation, Coefficient of Variation, Skewness and Kurtosis are depicted in Table 1. This table reveals that the average price of the onion crop is 1463 Rs/Quintal and the onion prices in Belgaum market ranges between 444 to 8450 Rs/ Quintal. Since, CV is 68.98% we can say that so much of variation is there in the price data. Time plot of the average monthly onion price for the original series is presented in Figure 1.

Table 1: This table reveals that the average price of the onion crop

Statistics	Price
Observations	168
Mean (Rs/quintal)	1463
Minimum	444
Median	1130
Maximum	8450
SD	77.9
CV	68.98
Skewness	2.9
Kurtosis	14.2

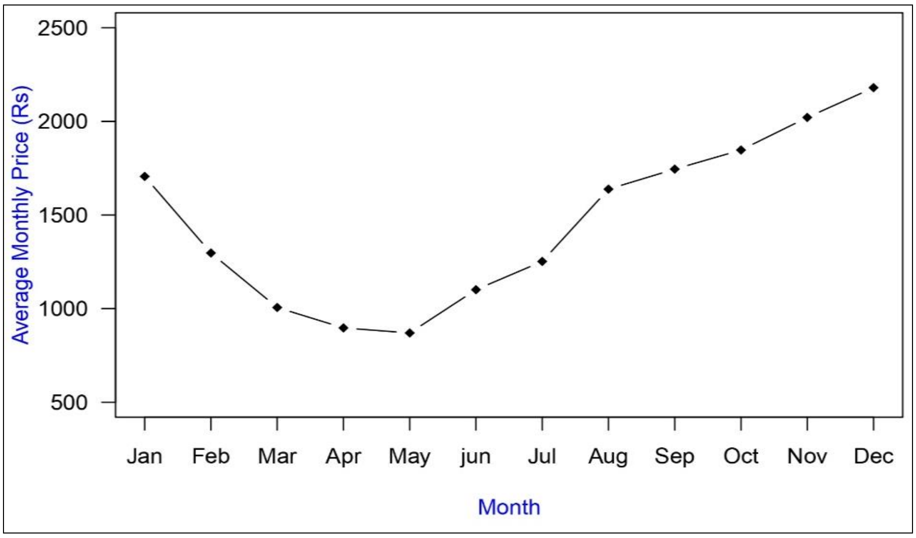

Fig 1: Monthly average wholesale price of Onion in Belgaum market, Karnataka

3.2 Fitting of Exponential Smoothing models

All the three methods like Single, Double and Triple Exponential Smoothing were considered for modelling the price data. The combination of estimates of smoothing parameters are given in Table 2.

Table 2: Estimated parameters of the Exponential Smoothing models

Model	Parameter	Estimate	AIC
SES	Alpha (Level)	0.9999	2598.548
DES	Alpha (Level)	0.9999	2602.869
	Beta (Trend)	0.0001	
TES	Alpha (Level)	0.9999	2423.002
	Beta (Trend)	0.0109	
	Gamma (Seasonal)	0.0001	

3.3 Fitting of ARIMA model and SARIMA model

Table 3 reveals the results of ADF test which was performed to test the stationarity of the price data and the data was found to be non-stationary hence we used first order differencing to make the data Stationary. ARIMA model is fitted using the 80% of the data set in R software and it reveals that ARIMA (1, 1, 2) was best model with less AIC value. The estimated parameters of the ARIMA model along with standard error is given in Table 4. We employed Kruskal Wallis test to test the seasonality of the price data and the test results reveals the test statistic is 47.17 with a p-value, 0.01 it indicates the presence of seasonality, further we proceed to fit SARIMA model and the most suitable model was selected based on AIC values. SARIMA (1,1,2) (2,0,0) $_{[12]}$ was found to be the best model with AIC value of 2241.08 and the estimated parameters of the SARIMA model along with standard error is depicted in Table 5.

Table 3: Augmented Dicky Fuller Test Results

Data	ADF test	Lag order	P-value
Original	-3.18	12	0.0936
First Differenced	-4.83	12	0.01

Table 4: Estimated parameters of ARIMA (1, 1, 2) model

Parameters	Estimate	S.E.	p-value
AR1	0.53	0.09	<0.01 ***
MA1	-0.38	0.08	<0.01 ***
MA2	-0.57	0.07	<0.01 ***

***: Significant at 0.1%

Table 5: Parameter estimates of SARIMA (1, 1, 2) (2,0,0) $_{[12]}$ model

Parameters	Estimate	S.E.	p-value
AR1	0.51	0.10	<0.01 ***
MA1	-0.38	0.09	<0.01 ***
MA2	-0.57	0.08	<0.01 ***
SAR1	0.03	0.08	0.70 NS
SAR2	0.15	0.09	0.10 NS

***: Significant at 0.1%; NS: Non-Significant

3.4 Fitting of BATS and TBATS models

Various combinations of smoothing parameters and Box-Cox transformed values were explored for BATS models using a grid search method. Ultimately, BATS (0.003, {1, 2}, -, {12}) was selected as the best-fitted model due to its lower AIC value. Similarly, different combinations of smoothing parameters, Box-Cox transformed values, and damping parameters were tested for TBATS models. Finally, TBATS (0, {0, 0}, 0.8, {<12, 2>}) emerged as the best model based on AIC value. The parameter estimates for both the BATS and TBATS models are provided in Table 6.

Table 6: Parameter estimates of BATS and TBATS model

Model	*Box-Cox transformation (Omega)	Smoothing parameter			Phi	ARMA coefficients		Prediction error	
		Alpha	Beta	Gamma		AR coefficients	MA coefficients	Sigma	AIC
BATS (0.003, {1,2}, -, {12})	0.003	0.032	-	-0.112	-	0.796	-0.230 -0.106	0.213	2344.93
TBATS (0, {0,0}, 0.8, {<12,2>})	0	1.446	-0.366	Gamma 1 0.001 Gamma 2 -0.001	0.8	-	-	0.228	2350.72

3.5 Model Evaluation, Residual Diagnostic and forecasting

The models are evaluated based on the two criteria *viz* Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) for both the training and testing data sets which contains 80 and 20% observations respectively. The results were presented in Table 7. The BATS model demonstrated superior performance in both the training and testing datasets, consistently yielding higher accuracy compared to other models. Ultimately, it was selected as the best model based on its performance on the testing dataset, which exhibited the lowest Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values. This indicates that the BATS model outperforms its counterparts in terms of predictive accuracy and reliability. Ljung-Box test was used to diagnose the residuals of BATS model and the findings reveals that the errors are random in nature. Hence BATS was used to forecast the monthly price of onion in Belgaum market of Karnataka state and the plot showing observed and fitted values of the BATS model is depicted in Figure 2. The forecasted prices are depicted in Table 9 and the results revealed that onion price is expected to be high in January 2024 followed by November and December 2024. In general, the price of onion will be more during the winter season.

Table 7: Model accuracy evaluation

Models	Training set		Testing set	
	RMSE	MAPE (%)	RMSE	MAPE (%)
SES	570.28	39.97	348.39	52.07
DES	570.61	40.77	363.37	51.05
TES	493.49	37.20	355.34	38.45
ARIMA	479.17	34.51	305.40	42.59
SARIMA	479.12	34.47	305.31	42.49
BATS	384.35	26.31	233.29	30.06
TBATS	430.13	30.56	256.01	33.54

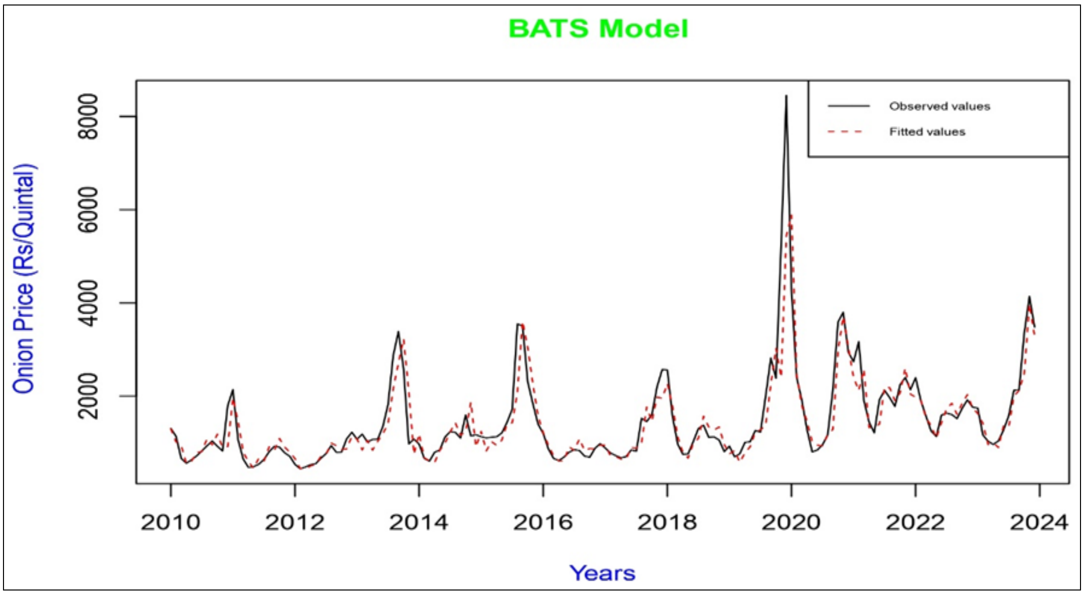

Fig 2: Plot showing observed vs fitted values by BATS model

Table 8: Residual diagnostics test

Model	Ljung-Box test	
	Test statistic	p-value
BATS	0.017	0.896

Table 9: Forecasted monthly wholesale prices of Onion

Month-year	Price (Rs/quintal)
Jan-24	2877
Feb-24	2107
Mar-24	1699
Apr-24	1449
May-24	1291
Jun-24	1562
Jul-24	1653
Aug-24	1945
Sep-24	2008
Oct-24	2331
Nov-24	2665
Dec-24	2360

4. Conclusion

In our current study, we utilized Single Exponential Smoothing, Double Exponential Smoothing, Triple Exponential Smoothing, ARIMA, SARIMA, BATS, and TBATS models to analyse onion prices in the Belgaum market of Karnataka state. Among these models, the BATS model outperformed the others, exhibiting lower RMSE and MAPE values, with residuals demonstrating a random distribution. Consequently, we employed the BATS model to forecast onion prices for the year 2024. Our findings indicate that onion prices are expected to be high in January 2024, followed by November 2024 and December 2024. This insight provides valuable guidance to onion growers not only within the region but also to growers in other areas. By strategically planning their sowing dates based on this information, growers can align their harvests with peak price periods. This strategic approach allows them to maximize profitability by capitalizing on favourable market conditions.

5. References

1. Apeda RN. Agricultural and Processed Food Products Export Development Authority.
2. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time

series analysis: forecasting and control. John Wiley & Sons; c2015 May 29.
3. Brown RG. Smoothing, forecasting and prediction of discrete time series. Courier Corporation; c2004.
4. CRISIL the Onion Paradox. Market Intelligence & Analytics; c2021. Retrieved from https://www.crisil.com/content/dam/crisil/our-analysis/reports/Research/documents/2022/11/the-onion-paradox.pdf
5. De Livera AM, Hyndman RJ, Snyder RD. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American statistical association. 2011;106(496):1513-27.
6. De Livera AM. Automatic forecasting with a modified exponential smoothing state space framework. Monash Econometrics and Business Statistics Working Papers. 2010;10(10):6.
7. Directorate of economics & statistics. 2021
8. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting. 2004;20(1):5-10.
9. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting methods and applications. John wiley & sons; c2008.
10. Rajput S, Sharma H, Sharma S, Pandya A. Univariate Time series Methods for Forecasting Cotton Prices: A Comparative Analysis.
11. Statista (August 2023). Gross value output of onions in India FY 2012-2020. Retrieved from https://www.statista.com/statistics/1080571/india-economic-contribution-of-onions/