*SKUAST Journal of Research* 25(2): 235-341; (2023)
(*10.5958/2349-297X.2023.00032.6*); NASS rating: 4.17

**Research Paper**



# Utilization of parametric and non-parametric models for trends of castor seeds in India

M. M. Kachroo*[1], M. H. Wani[1], S. H. Baba[1], Nageena Nazir[2], F. A. Lone[3], F. A. Shaheen[1], A. R. Malik[4], K. Gautam[1] and Tehleel Nazir[4]

[1]*School of Agricultural Economics and Horti-Business Management;* [2]*Division of Agricultural Statistics;* [3]*Division of Environmental Sciences,* [4]*Division of Fruit Sciences;* [4]*Department of Linguistics, University of Delhi; Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir*
*e-mail*: mubashirkachroo@gmail.com*

(Received November 20, 2022; accepted January 11, 2023)

ABSTRACT

**Castor is an important non-edible oilseed, basically a cash crop, cultivated around the world owing to the commercial importance, of its oil. Gujarat is the largest Castor producing state followed by Andhra Pradesh and Rajasthan. The present study is an attempt to find past trends of Castor seed in India using parametric, nonparametric and semi-parametric regression methods. The performance of each method is compared using higher values of $R^2$ and lower values of residual criteria. It is found that the parametric regression comes out to be good fit for trends in Castor seed in comparison to the nonparametric/semi-parametric regression. In Comparison to nonparametric and semi-parametric, semi-parametric spline regression was selected as the best fitted model for trend analysis. The study advocates for researchers technological breakthrough in Castor seed production in India.**

**Key words:** Non-parametric, production, regression, spline, trend

Castor (*Ricinus communis* Linn.) is an important non-edible oilseed crop grown almost entirely in the arid and semi-arid regions in the country. The crop, an important cash crop is grown for its beans from which oil is extracted. India is the world's largest producer of castor beans. Brazil and China are the other major producers of castor beans and castor oil in the world. The global demand for castor oil is estimated to be about 1 billion pounds worth US$ 500 million (Rao et al. 2003). The area sown to castor largely depends on the area sown to its competing crops such as cotton, sunflower, sorghum, pigeon pea, etc which are also grown under rainfed conditions. The crop faced competition from cotton in particular, in the Telangana region of Andhra Pradesh (Rao, 2000). The growth rates of crops are mostly estimated by the linear regression models. However, it might be the case that these models may not fit the data well. Under such conditions it becomes essential to apply nonparametric and semi-parametric regression, which is based on fewer assumptions. In last few years, nonparametric regression and semi-parametric regression technique for functional estimation has become increasingly popular as a tool for data analysis. These techniques impose only few assumptions about shape of function and therefore it is more flexible than usual parametric regression approaches. Smoothing techniques are commonly used to estimate the function non-parametrically (Hardle, 1990).

Nonparametric regression models avoid restrictive assumptions of the functional form of the regression function. Semi-parametric regression model combines the components of parametric and nonparametric regression models, by keeping the easy interpretability of the former and retains some of the flexibility of the latter. Various scientists viz., (Chandran, 2004) has applied nonparametric regression to study the growth rates of total food grain production of India during the period 1987 to 2001. Teczan (2010) has studied the nonparametric regression technique to find out the growth rate trends of various crops. Sahu and Pal (2004) used nonparametric regression (Lowess) and semiparametric (spline) for modelling of pest incidences. (Dhekale *et al.,* 2017) employed the nonparametric regression model to study the trends of tea in India. The current study is aimed to develop appropriate parametric and nonparametric regression models to fit the trends in various variables of castor seed in India.

MATERIALS AND METHODS

To study the growth rates and trends, the monthly data of 16 years from 2004-2020 for Prices, Area, Production, and Yield was collected for the CMIE database. The descriptive measures of central tendency and dispersions along with the compound growth rates were used to explain the features of the data (Mishra *et al.* 2012).

**Trend Models**

**Parametric regression models:** To find out the path of the production process different parametric trend models are fitted. Among the fitted models, the best model is selected on the basis of their goodness of fit ($R^2$) value and significance of the coefficients. The dependent variable Y is prices, export, import, area, production, yield and arrivals and independent variable X is the time points (years).

**Non-parametric and Semi-parametric model:** The model considered here is of the form:

$$Y_i = m(X_i) + E_i, X_i = \frac{i}{n}, i = 1,2, \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where, $Y_i$ is observation of $i^{th}$ time point, $m(.)$ is trend function which is assumed to be smooth and $E_i$ are random errors with mean zero and finite variance. Since there is no assumption of parametric form of function $m(.)$ This approach is flexible and robust to deviations from an assumed model form. To obtain an estimate of the mean response value at a point $X$, most of the observations having predictor values closer to the target value $X$. The averaging is done in neighbourhoods around target value. The main decision to be made in any of the smoothing techniques is to fix the size of neighbourhood which is typically expressed in terms of an adjustable smoothing parameter or bandwidth. Intuitively, large neighbourhoods will provide an estimate with low variance but potentially high bias, and conversely for small neighbourhoods. Lowes regression, introduced by (Cleveland, 1979), is obtained on the basis of the data points around it within a band of certain width. The point $X_i$ is the midpoint of the band. The data points within the band are assigned weights in a way so that $X_i$ has the highest weight. The weights for the other data points decline with their distance from $X_i$ according to a weight function. The weighted least squares method is used to find the fitted value corresponding to $X_i$, which is taken as the smoothed value. The procedure is repeated for all the data points. The spline method of estimation make use of the penalized least squares method (Simonoff, 2012), which balances the fitting of the data closely. The objective is to estimate m by means of a function that fits the data well and is as smooth as possible. A measure of smoothness of $m$ is the integral of the square of its second derivative as;

$$\sum_{i=1}^{n}(Y_i - m(x_i))^2 + \lambda\int_a^b (m''(x))^2 dx \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

Where $\lambda > 0$ is a fixed constant; $x_i$ [a,b], $i$ =1,2,...n

The first term is the sum of squares of the residuals; it provides a measure of how well the function m fits the data. The integral of the above equation is a measure for the $m$. The function which are highly curved will result in a large value of the integral; straight lines result in the integral being zero. The roughness penalty λ, controls the emphasis which one wishes to place on smoothness. By increasing the value of λ, one places more emphasis on smoothness; as λ becomes large the function approaches a straight line. On the other hand, a small value of λ emphasizes the fit of m to the data points: as λ approaches a function that interpolates the data points. The analysis for different commodities was performed using econometric software SAS version 9.3.

RESULTS AND DISCUSSION

The maximum growth rate was observed in yield of Castor seed over the years, whereas the minimum growth rate was exhibited by prices of the castor seed. The positive compound growth of yield (1.93 percent per annum) shown in Table 1 reveals that there is no decrease in the yield of castor over the years with a maximum of 1.90 million kilogram and minimum of 1.06 million kilogram. Similarly, the compound growth rate (0.65 per cent

**Table.1 Performance of castor seed in India during 2004-2020**

| Particulars | Prices | Area | Production | Yield |
|---|---|---|---|---|
| Mean | 3052.70 | 900.97 | 1356.17 | 1483.02 |
| SD | 1057.70 | 216.17 | 450.09 | 246.12 |
| Minimum | 1253.40 | 628.40 | 762.30 | 1067.80 |
| Maximum | 5119.20 | 1470.90 | 2294.90 | 1902.10 |
| Skewness | -0.15 | 1.36 | 0.48 | -0.30 |
| Kurtosis | -1.02 | 1.74 | -0.57 | -0.94 |
| CV% | 34.65 | 23.99 | 33.19 | 16.60 |
| CAGR% | 0.41 | 0.55 | 0.65 | 1.93 |

per annum) is observed in production indicates an increase in the production of castor in India over the years. This is due to the fact that a large area of land is being brought under agriculture. It was noticed that a compound growth rate of area (0.55 per cent per annum) under castor cultivation indicating that a large portion of the land is being utilized for the latter.



**Table.2 Trends in prices, area, production and yield of castor seed in India**

| | $R^2$ | Constant $b_O$ | $b_1$ | $b_2$ | $b_3$ | RMSE | MAPE | MAE | Max APE | Max AE |
|---|---|---|---|---|---|---|---|---|---|---|
| **Prices** | 0.806 | -40.1751 | 12.16112 | -12.1448 | 14.4003 | 0.078 | 0.018 | 0.061 | 1.768 | 0.199 |
| **Area** | 0.189 | 3.005131 | 0.000179 | 1.000864 | -1.00269 | 0.088 | 0.016 | 0.069 | 1.571 | 0.223 |
| **Production** | 0.586 | -3.00252 | 0.999127 | -0.00018 | 1.001827 | 0.106 | 0.017 | 0.082 | 1.738 | 0.251 |
| **Yield** | 0.869 | 2.996939 | -0.99721 | 0.000211 | 0.998089 | 0.027 | 0.005 | 0.021 | 0.520 | 0.060 |

**Trend analysis of prices, area, production and yield**

**Parametric techniques:** The linear models used here are the cubic model or the third-degree polynomial model $Y = b_0 + b_1t^1 + b_2t^2 + b_3t^3$ and the quadratic model or second-degree polynomial model $Y = b_0 + b_1t + b_2t^2$ are shown in Table 2. The value of $b_2$ for prices were negative which indicates that prices of castor seed decreased in the middle period and the value of $b_1$ and $b_2$ being positive clearly indicates that there was an increase in the castor prices. Further, the negative value of $b_2$ for production and $b_3$ for area is an indication of the decrease in the production and area during the middle and final period of the study and the positive values of $b_2$ and $b_3$ indicates an increase in the yield in castor seed over the years.

**Table.3 Trends in prices of castor seeds in India using non-parametric and semipara metric regression**

| | Loess | Spline |
|---|---|---|
| **AICc** | 16.92 | 4.44 |
| **RMSE** | 895.30 | 563.52 |
| **MAPE** | 0.35 | 0.22 |
| **MAE** | 25.60 | 17.79 |
| **MaxAPE** | 37.78 | 18.86 |
| **MaxAE** | 2010.69 | 1802.20 |
| **$R^2$** | 0.73 | 1.28 |

**Nonparametric and semi-parametric regression:** Trend analysis of prices, area, production and yield using nonparametric (Loess) and semi parametric (spline) regression are presented in the tables 3, 4, 5 and 6. In Table 3 the value of $R^2$ is 0.73 for Loess and 1.28 for Spline regression. The AICc, RMSE, MAPE, MAE, MaxAPE and MaxAE values comes out be small for Spline regression for the all the variables except in case of production. The area under the castor seed has increased over the years of study and is shown in figure 2.

**Table.4 Trends in area of castor seeds in India using non-parametric and semi-parametric regression**

| | Loess | Spline |
|---|---|---|
| **AICc** | 7.30 | 14.07 |
| **RMSE** | 17.57 | 17.41 |
| **MAPE** | 0.26 | 0.26 |
| **MAE** | 3.58 | 3.60 |
| **MaxAPE** | 25.13 | 24.03 |
| **MaxAE** | 49.76 | 48.01 |
| **$R^2$** | 0.45 | 0.46 |

The increasing trend in the production and yield over the years of study is shown in the fig. 3 and 4. The values of prices, area, production and yield are initially fitted at the smoothing parameters in order to obtain the best fit of the data points in the neighbourhood of the smoothing parameters and look for the curve which covers all the points of the data. The one which covers maximum points is the best fit of the data points. In Figure 1 the smooth curve fits are obtained for prices of Castor seed in the neighbourhood of smoothing parameters i.e., at 0.3, 0.44, 1.00 and 1.50. It was observed that the best fit is obtained at smooth=0.44, similarly the smoothing parameters were obtained for other variables also. Pal (2011) observed similar results where the spline gave the better results than the Loess smoothing.

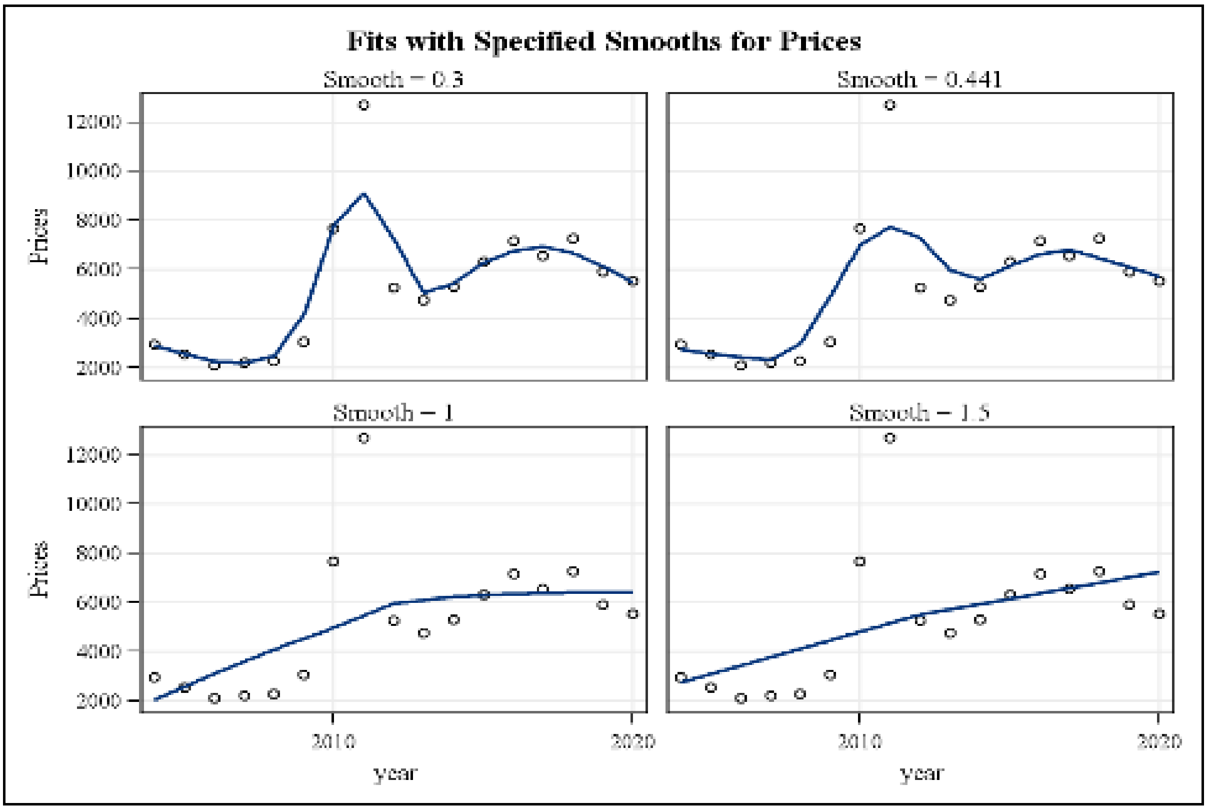


**Fig.1: Fits with specified smooths for prices**

**Table 5: Trends in production of castor seeds in India using non-parametric and semi-parametric regression**

| | Loess | Spline |
|---|---|---|
| **AICc** | 10.99 | 13.85 |
| **RMSE** | 57.04 | 61.35 |
| **MAPE** | 0.22 | 0.24 |
| **MAE** | 6.54 | 7.12 |
| **MaxAPE** | 15.25 | 13.23 |
| **MaxAE** | 102.38 | 121.27 |
| **$R^2$** | 0.78 | 0.89 |

The values of RMSE, MAE, MAPE, MaxAE and MaxAPE for price, area, production and yield in India for nonparametric regression has observed higher values than the parametric regression (Tables 3, 4, 5). This is clear indication of the superiority of parametric models over the non-parametric models. These models perform very well in visualizing the past trends where the non-parametric/semi-parametric models fails to. From the above study, it was observed that there is dramatic increase in the price, area, production and productivity of castor seed in India. In order to maintain the trend more and more land is to be brought under the castor seed cultivation. Non-Parametric and semi-parametric regression usually utilized in studying the trend seems not to perform better than the Parametric regression and out of the nonparametric and semi-parametric regression methods the semi-parametric regression (spline) is the best fit for the trend analysis of the castor seed in India

**Table 6: Trends in Yield of Castor seeds in India using non-parametric and semi-parametric regression**

| | Loess | Spline |
|---|---|---|
| **AICc** | 12.52 | 13.71 |
| **RMSE** | 232.53 | 222.94 |
| **MAPE** | 0.20 | 0.19 |
| **MAE** | 13.35 | 13.00 |
| **MaxAPE** | 11.80 | 10.92 |
| **MaxAE** | 504.61 | 479.50 |
| **$R^2$** | 0.39 | 0.43 |

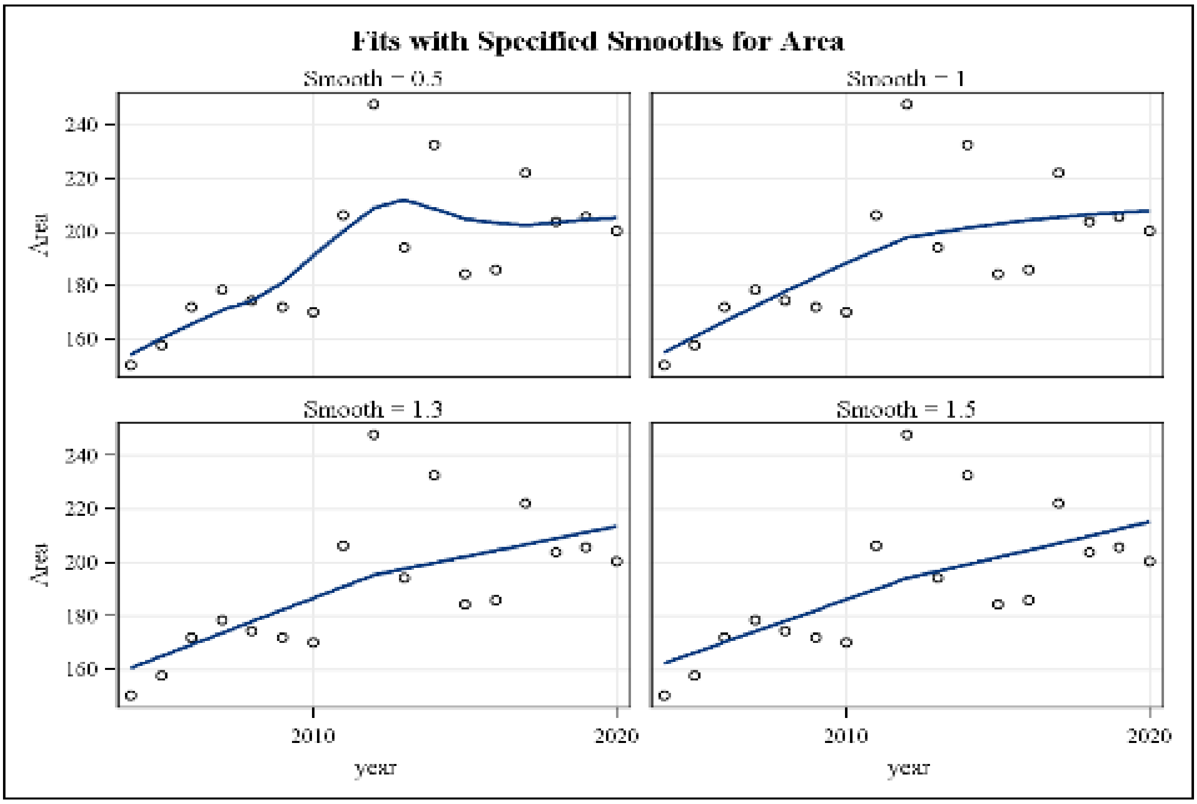


**Fig. 2: Fits with specified smooths for area**

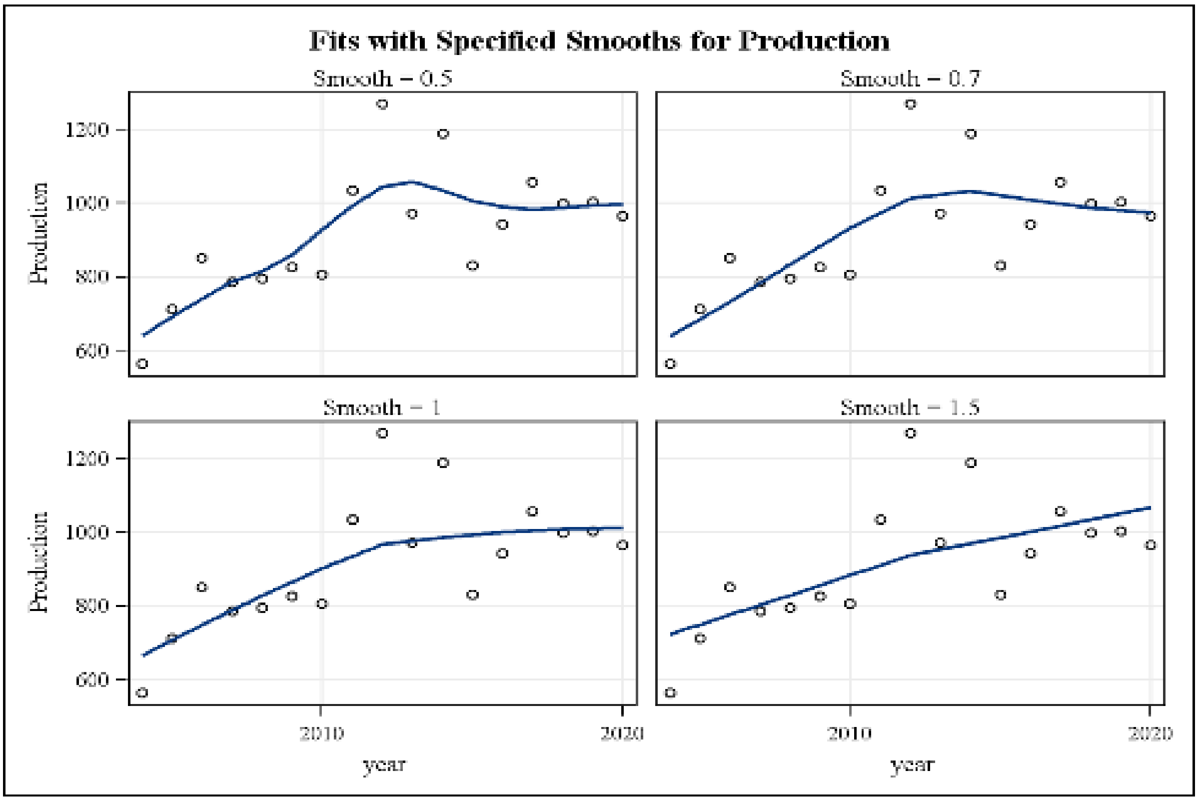


Fig.3: Fits with specified smooths for production

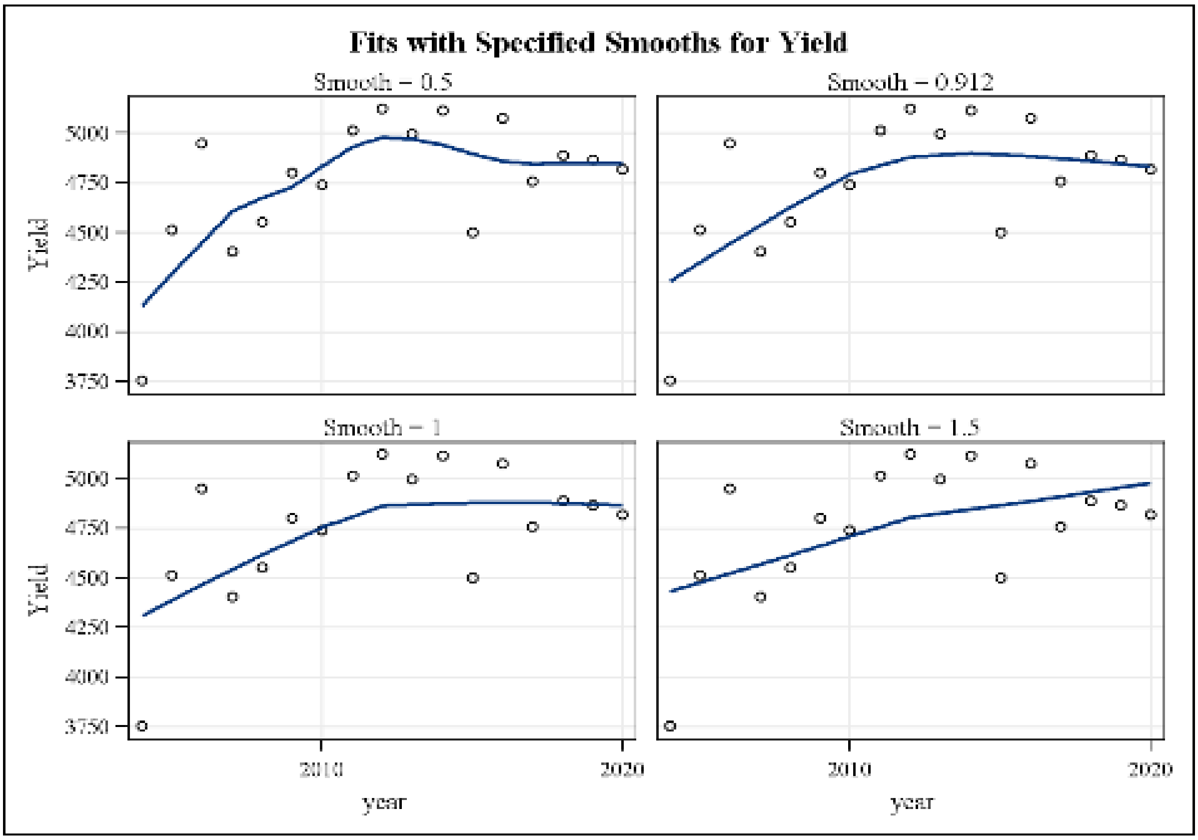


Fig. 4: Fits with specified smooth's for yield

## CONCLUSION

It was observed that the prices of Castor seed were decreased in the middle period of the study which later on gets increased. The AICc, RMSE, MAPE, MAE, MaxAPE and MaxAE values comes out be small for Spline regression for the all the variables except in case of production. It was observed that the best fit is obtained at smooth=0.44, similarly the smoothing parameters were obtained for other variables also. The values of RMSE, MAE, MAPE, MaxAE and MaxAPE for price, area, production and yield in India for nonparametric regression has observed higher values than the parametric regression. The parametric regression model was found to be the best fit model for trend analysis of castor seed in India.

## REFERENCES


Chandran, K. 2004. *A Study on Some Aspects of Nonlinear Statistical Models in Agriculture*, Indian Agricultural Statistics Research Institute Indian Agricultural Research Institute: New Delhi.

Cleveland WS 1994. *The elements of graphing data (revised edition).* Hobart Press, Summit, NJ, USA.

Dhekale, B., Sahu, P., Vishwajith, K., Mishra, P., and Narsimhaiah, L. 2017. Application of Parametric and Nonparametric Regression Models for Area, Production and Productivity Trends of Tea (Camellia Sinensis) in India, *Indian Journal of Ecology*, **44**, 192-200.

Hardle, W. 1990. *Applied Nonparametric Regression*, Cambridge university press.

Mishra, P., Sahu, P., Bajpai, P., and Nirnjan, H. 2012. Past Trends and Future Prospects in Production, and Export Scenario of Tea in India, *International Review of Business and Finance*, **4**: 25-33.

Pal S. 2011. *Modeling Pest Incidences in Some Agricultural Crops*, Thesis Submitted to the Bidhan Chandra Krishi Viswavidyalaya under Department of Agricultural Statistics.

Rao, C. R., Rao, M. S., Ramakrishna, Y. S., & Reddy, Y. V. R. 2003. *Economic Analysis of Castor Seed Multiplication*. Central Research Institute for Dryland Agriculture Santoshnagar, Hyderabad-500, 59.

Rama Rao C.A. 2000. *Growth and Efficiency in Crop Production in Andhra Pradesh*. Ph.D. Thesis submitted to ANGR University, Hyderabad.

Sahu PK and Pal S 2004. *Modelling Pest Incidence-Nonparametric Approach.* Paper presented in 24th International Symposium on Forecasting, Sydney.

Simonoff, J. S. 2012. *Smoothing Methods in Statistics*, Springer Science & Business Media.

Teczan N 2010. *Analysis of the Factors Affecting Productivity Using Non-Parametric Regression Method.* 10th Global Conference on Business & Economics, Rome, Italy.