J. Appl. Math. & Informatics Vol. 42(2024), No. 6, pp. 1293 - 1306
https://doi.org/10.14317/jami.2024.1293

IMPROVING REGRESSION ANALYSIS WITH BOX-COX TRANSFORMATION IN A BAYESIAN FRAMEWORK

VICTOR ANTHONYSAMY, KHADAR BABU SK*

Abstract. Prediction and forecasting of the Big data is an emerging transformation to build a best linear Model. Regression analysis is widely used for forecasting time series datasets, but non-significant parameters can lead to unreliable predictions. To improve models, the Bayesian technique is used, incorporating prior knowledge, and investigating their superiority. This powerful statistical methodology is essential for accurate crop yield forecasting. The present article proposes to predict and forecast the rice crop production in India using Bayesian linear regression methodology, and the results were compared with simple linear regression. A Box-Cox transformation was also included in the procedure to increase forecast accuracy. The method is obviously perfect for predicting future values. In addition, To evaluate model accuracy, we used squared R, MSE, RMSE, MAE, MAPE, and Theil's U. After applying the Box-Cox transformation, the findings of the result analysis are more accurate and easier to predict. Finally, Comparing simple Linear regression and Bayesian regression models, it was found that the Bayesian framework yielded superior results compared to classical approaches.

AMS Mathematics Subject Classification : 62J05, 62F15.
Key words and phrases : Simple linear regression, Bayesian linear regression, box-cox transformation, Bayesian probability.

1. Introduction

Machine learning faces challenges in creating effective prediction and forecasting models for big data analytic, particularly in agricultural data analysis where explanatory variables are currently unavailable. This necessitates the identification of these variables to accurately predict study variables for forecasting models.

Received April 5, 2024. Revised September 19, 2024. Accepted October 4, 2024.
*Corresponding author.

Traditional statistical methods for predicting rice yield have limitations due to their reliance on influential parameters like cultivation area and production. Advanced machine learning techniques overcome these limitations by analyzing data through different dimensions and identifying diverse patterns. These algorithms can train models better for nonlinear data patterns and reduce forecaster assumptions and bias. Machine learning techniques are crucial in dealing with complex situations and making informed decisions for farmers and decision-makers, making them a valuable tool in predicting rice yield.

The Box-Cox transformation is a crucial technique in linear regression, used when model assumptions are violated. However, previous methods and algorithms are primarily developed for small or moderate data, which are not suitable for big data due to memory and storage capacity barriers. Zhang and Yang [1] propose a method for applying Box-Cox transformation to large data sets, avoiding computer storage before calculations. They also suggest updating the required quantities for least squares regression sequentially. And also, they propose new methods and algorithms that construct and compute a set of summary statistics, known as the Box-Cox information array. The computation of this array is the only issue to consider in data reading, and the optimal power transformation and model parameter estimates can be quickly computed. This is expected to significantly impact the development of statistical methods and algorithms for big data.

Gulzar et al. (2023) [2] has discussed about Bayesian analysis relies on the prior distribution, which represents assumed or known information about a parameter. The choice of prior is crucial, as it affects the reliability of the posterior distribution. Proper priors do not depend on data and have an integral or summation equal to 1, while improper priors do not depend on data and do not integrate or sum to 1.

One of the effective methods for predicting crop yields using agro-metrology is multiple regression analysis [3]. In the majority of cases, the conventional regression model provides a highly accurate estimation of parameter values. However, it is worth noting that there are situations where the parameters of the model may lack statistical significance [4] and Zhang and Yang [1] propose a method for applying Box-Cox transformation to large data sets. So, it gives a scope for additional improvement of regression parameter estimations.

Hence,Using the Box-Cox transformed Bayesian regression approach, we can effectively model and analyze data by addressing issues such as heteroscedasticity and non-normality. This method allows for more flexibility in handling different types of data distributions and can provide more accurate parameter estimates compared to traditional regression methods. By incorporating the Box-Cox transformation within a Bayesian framework, we can improve the overall performance and reliability of our regression models.

2. Literature Survey

Many Researchers have developed traditional and advanced regression models for predicting crop yield in India, including multiple linear regression, kernel ridge, lasso, and elastic net regression models, considering parameters like state, district, season, area, and year.

Yield estimating plays a significant part in agricultural planning and managing, national food supply, international food trade, ecology sustainability discussed [5]. Verma [6] worked on wheat, sugarcane, cotton, and mustard crops for operative yield forecasting purpose under Crop Area Production Estimation (CAPE) project. Mishra [7] discussed the constancy in production performance and forecasted agricultural production in India using ARIMA model. Regression Analysis is a multivariate method used to analyze the ecological factors as explanatory variables and their exaction on crop yield to obtain a decision [8]. Shakila [9] has studied for forecasting of wheat production by using various kind of regression models. For Policymakers to ensure food and nutrition security and forecasting the agricultural production performances of the major crops is important. Hence, they need to be aware of potential production scenarios for the primary crops, according to Mishra's research [7].

Applying an automatic data cleansing methodology suggested [10], and using that in ARIMA model, Choudhury [11] investigated how to predict the production of rice in the states that produce most of the crop. For forecasting vegetable production in the state of Haryana, Kumari [12] used univariate forecasting models are moving average, simple exponential smoothing, random walk, random walk with drift, and ARIMA models. Sathis [13] studied about the horticultures growth in India Contributed around 33% to agriculture with valid significance of the growth of India economy. They applied ANN architecture for forecasting.

A linear regression model or a non-Bayesian regression model alone typically tends to overfit the data. Bayesian Linear Regression, which incorporates "Predictive Distribution," aids in solving the issue [14]. The Bayesian technique is a subjective method for estimating the linear model's unknown parameters, and this method yields a more accurate estimate of the parameters [15]. The benefit of Bayesian regression estimation can be linked to the fact that one can employ presumed knowledge about the current state of "beliefs" or include a prior distribution to increase the predicted value's accuracy and efficiency. To solve a broad range of issues, the Bayesian regression method can be used using a variety of analytical and numerical techniques [16].and compared the Bayesian regression model to the traditional regression model and concluded to the proper conclusion that the Bayesian model produced better results than the conventional model. Khadar Babu [18] has studied about Growth of the Rice Crop Using Various Smoothing Techniques.

The Generalized Linear Model (GLM) can be successfully improved using the Bayesian approach, and this improvement depends on the prior distribution selected [19]. Bunn [15] suggested that the Bayesian methodology is recommended as a potential tool for combining forecasts. Bayesian regression also represents a unique method of approaching various statistical findings. Instead of the point estimate, we get a range of inferential solutions using the Bayesian framework [20]. The Bayesian regression analysis is that all model parameters are essentially random. To improve forecast values, we can add prior knowledge to the estimated value of the parameters. Comparing Bayesian statistics to classical frequentist statistics, which treats all parameters as constant and illogical amounts, the principle of probability forms the foundation of Bayesian regression analysis. Gandhi [21] describe the use of Bayesian Networks to forecast paddy crop yield in Maharashtra, India.

Various transformation techniques have been employed to meet the Gaussian distribution requirements of the datasets [22]. Power transform methods include log transformation, reciprocal transformation, square-root transformation, arcsine transformation, inverse transformation, and Box-Cox transformation (BCT) [23]. The Box-Cox transformation has been extensively studied and applied in numerous data processing contexts. Research fields such as statistics, economics, econometrics, system dynamics modelling, prediction, medicine and time series forecasting encompass a wide range of academic disciplines [24]. Peng [25] used the Box-Cox transformation in a study to evaluate increased comfort in outside urban public places.He et.al [26] has suggested two Box-Cox transformations the quantile regression techniques to predict how Anhui province in China will use energy in the future and [27] used the Box-Cox transformation to examine soil cores contaminated with certain elements. Atkinson [28] has discussed the Box-Cox transformation and its extensions and given research motivations to extend to Bayesian regression.

Simple Linear Regression Model (SLRM) and Bayesian Linear Regression model (BLRM) were used for prediction and forecasting of rice crop production in India. The main Objective of the present research paper is to apply Bayesian Linear Regression approach for prediction and forecasting using box-cox transformation methodology for efficient model building on rice crop production dataset. After model building the efficient model validation indices are Compared to the Basic Simple Linear Regression Model. For data Transformation applied Box Cox methodology for better statistical forecasting.

In the present study Coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Theil's U Statistic are adopted for model validation. The results revealed that BLRM with SLRM fitted rice crop production. This will be helpful to the farmer of state and Central Government for estimating the rice crop production with possible resources. We use the Bayesian Linear Regression approach to forecast agricultural yields accurately. Since we were unable to identify any comparable research, we believe that this method will add to the knowledge base regarding agricultural yield forecasting models. We discussed the methodology, illustration, results, and conclusions in the follow sections.

3. Materials and methods

3.1. Data sources. The Directorate of Economics and Statistics, Ministry of Agriculture, India, provided data on rice yields for the years 1950–2019. The study used metrics including Area Under Cultivation (Million/Hectares), Production (Million/Tons), and Yield (Kg/Hectare) to analyze the information from all of India. The analysis was carried out to develop and compare with Bayesian Linear Regression with Simple Linear Regression to find the best-fit model.

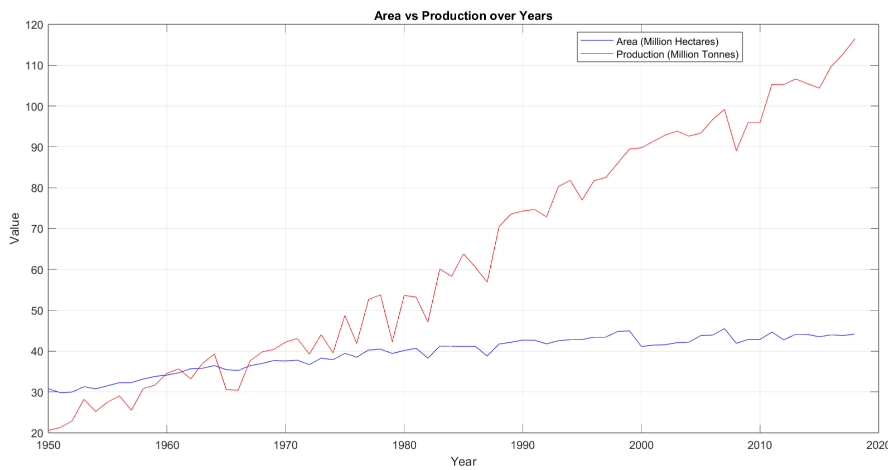

FIGURE 1. Estimates of Area and Production of Rice.

3.2. Methodology.

3.2.1. Simple Linear Regression Model (SLRM). Regression analysis is an essential statistical tool for studying, modelling, and figuring out how the variables relate to one another. There are three parts to the regression model. These are the unknown coefficients, independent variables, and dependent variables. The terms response, outcome, and regressed variables all refer to dependent variables. Predictors and regressors are additional names for independent variables. One dependent variable, one independent variables, and one or more unknown regression coefficients make up the regression model. Y is a function of X and can be used to represent a simple linear regression mode as,

$$Y_i = \tau + \omega X_i + \epsilon_i \tag{1}$$

When τ is an intercept, ω is a slope, and ϵ is individual error when monitoring the real-world data that does not fit perfectly to a straight line. Our focus is on unknown parameters. An unbiased and effective estimation of the model's unknown parameters is required. We try to estimate the model's unknown parameters so that the sum of squared of residuals is as small as possible.

3.2.2. Bayesian Linear Regression Model (BLRM). It is also a curve fitting technique that solely employs the linear regression principle, it uses Bayesian inference to carry out the statistical analysis. Regression works by using the best fit line to predict the values. This approach does not tell us how much our model is certain to predict an output. When we say certain, that means what is the probability of correctness, what we have as predicted as output. To overcome this problem, the concept of Bayesian regression comes into picture.

In Bayesian approach we involve prior knowledge to determine outcome. Prior knowledge makes it possible to incrementally improve the estimate results in new observation or evidence and at last not the least it adds ability into the model to express uncertainty during the time of model prediction.

One of the major Challenges in machine learning is to prevent the model from over fitting specially when we have small dataset, the model will only learn the specific configuration of the given data that does not generalize to the unseen data. However, if we have beliefs that were gained through experience, those could be used to generalize our model in the absence of a sufficiently large datasets. Therefore, with Bayesian Learning, we can use such prior knowledge or belief to provide additional information to the model.

3.2.3. Box-Cox Transformations. Box-Cox Transformation is a transformation technique specifically used when there is a "Non-Normality of Error" present in the response variable. It also fixes problems like non-linearity, and heteroscedasticity.

Step 1: Consider the following equation. $\delta^{\lambda} = \tau x + \epsilon$
Step 2: The parameter lambda is estimated using a Maximum Likelihood Estimation (MLE) function.
Step 3: Then re-estimate the model with the correct lambda, so model assumptions should be improved.
Step 5: Lambda's value ranges from (-5,+5)
Step 6: If the estimated lambda value is close to one, our relationship is just the norm.

$$\delta^{\lambda} = f(x) = \begin{cases} \frac{\delta^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ log\delta, & \lambda = 0 \end{cases}$$

And if 0 the transformation is going to be

$$\log \delta = x\tau + \epsilon$$

Reverse Box-Cox $\left\{[(\rho * \lambda) + 1]^{\frac{1}{\lambda}}\right\}$

Where $\rho = Y$ values, $\lambda =$ Calculate lambda values using MLE.

3.3. Model optimization Indices. The predicted accuracy was assessed. There are three metrics for a model fit that are most frequently used are R-squared, RMSE, and MAE.

a. Coefficient of determination (R^2)

The coefficient of determination (R^2), which ranges from 0 to 1, expresses how well a statistical model forecasts a result. The (R^2) may be seen as the percentage of variance in the dependent variable that the statistical model forecasts.

b. Root mean squared error (RMSE)

The deviation of the residuals is measured by the Root Mean Square Error (RMSE) (prediction errors). The distance between the data points and the regression line is measured by residuals, and the spread of these residuals is measured by RMSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Actual_i)^2}{n}}$$

c. Mean absolute error (MAE)

The average of all absolute errors is known as the Mean Absolute Error (MAE).

$$MAE = \frac{\sum_{i=1}^{n}|Predicted_i - Actual_i|}{n}$$

d. Mean absolute percentage error (MAPE)

To determine the mean absolute percentage error (MAPE), the absolute error for each period is subtracted from predicted values then as follows the procedures.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Predicted_i - Actual_i}{Actual_i}\right|$$

e. Theils U statistics

Theil's U statistic is a measure of relative accuracy that contrasts the outcomes of forecasting with incomplete historical data with the outcomes of forecasting. Additionally, it squares the deviations to accentuate and give more weight to significant errors, which can help in the elimination of approaches that have large errors.

$$Theil - U = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{\hat{Y}_{t+1}-Y_t}{Y_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{Y_{t+1}-Y_t}{Y_t}\right)^2}}$$

Where Y_t is the observed values at time t, $\hat{Y}_t$ is forecasted value at time t and n is the total number of data.

4. Results and discussion

4.1. Overview of Rice Parameter Statistics. The Rice yield data were applied to each of the Simple Linear regression and Bayesian Linear Regression to evaluate their relationship, and the efficacy of each model was assessed using MSE and R^2. In Figure 1 shows All India rice production and Area. The annual paddy yield for each agricultural term from 1950 to 2019 expanding the overall figures of main statistical indicators are used to analyzed.

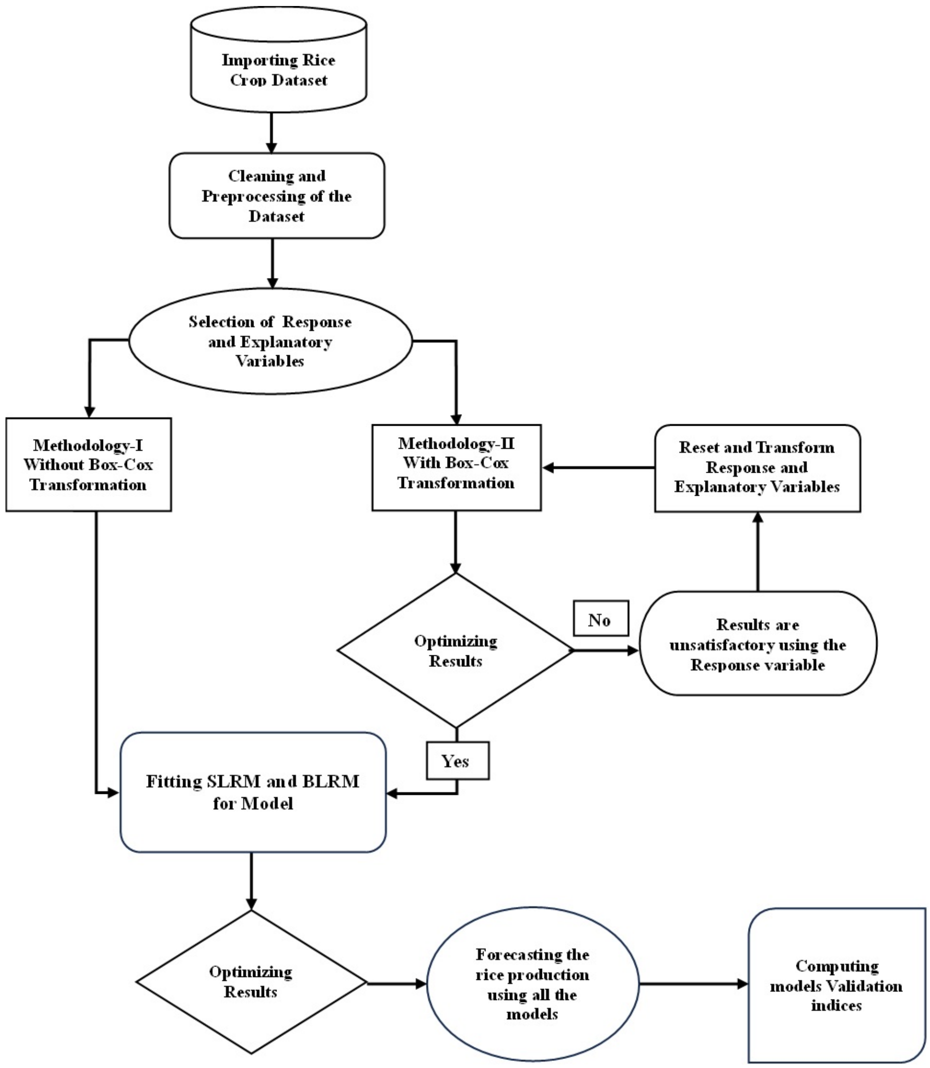

Figure 2. Flow Diagram for Methodology.

The flowchart (Figure 2) describes the methodology of the current study. It addresses the simple procedures for applying the Box-Cox transformation to the SLRM and BLRM. In Figure 3 shows the Pearson correlation coefficient matrix which is the correlation between all characteristics of the dataset. In our dataset, we have three characteristics i.e., Area, production, and yield. According to the matrix, the correlation between the production and yield is equal to 1, which is very highly positively correlated that is because the yield is the production

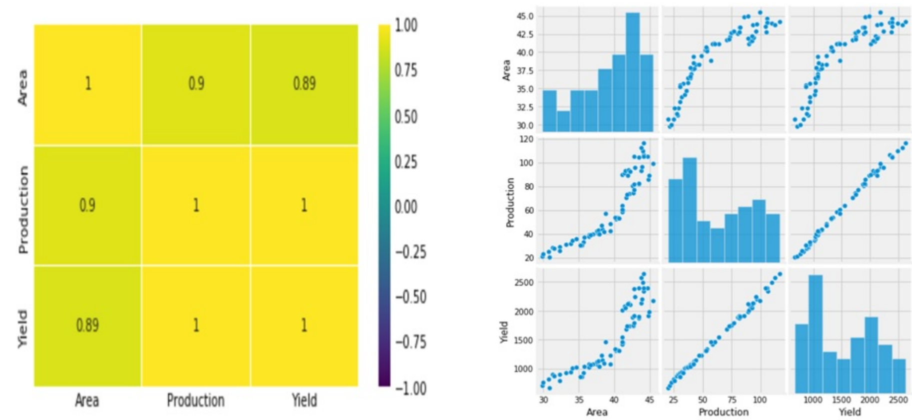

FIGURE 3. The correlation Between the Area, Production and Yield

upon Area. The correlation between the area and production is 0.9, which shows that the relationship between the exploratory and response variables is highly positively correlated.

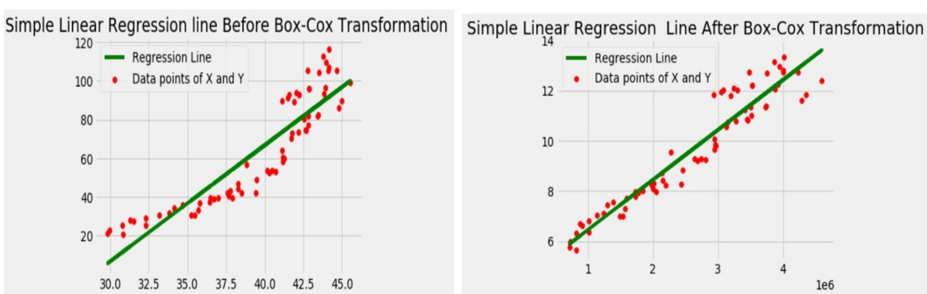

FIGURE 4. Simple regression Model line of before and after Transformed data.

Figure 4 shows the best fit line of the model before and after box-Cox Transformation. Figure 5 and Figure 6 shows that graph compares the model optimization indices of SLRM and BLRM before and after transformation.

4.2. Bayesian Linear Regression procedure. First, the response variable "Y" was transformed. The outcome is not very satisfying. Then explanatory variable "X" was transformed as well. The outcome is now fair and satisfactory. Bayesian Linear Regression as involving following steps

Step 1: From the prior and likelihood function, determine the joint posterior.

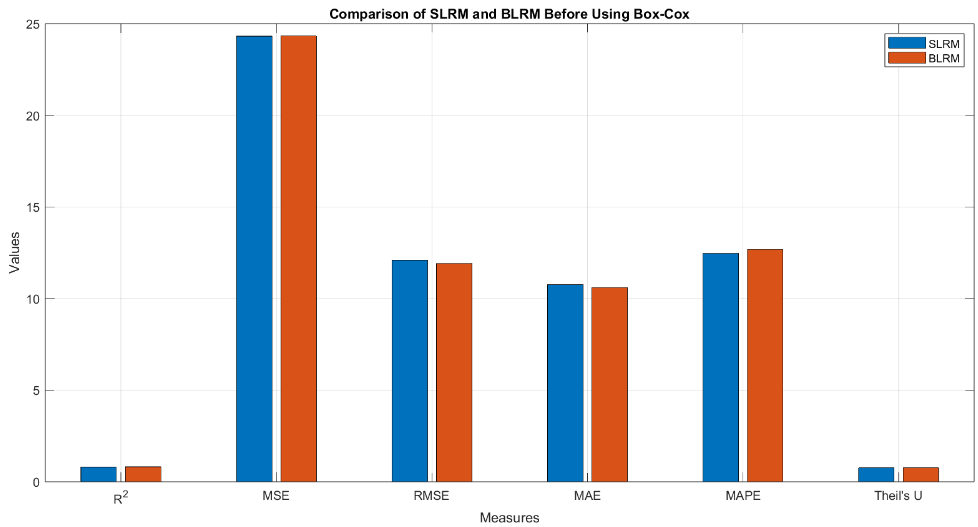

FIGURE 5. Result analysis Linear Regression and Bayesian Analysis Before Transformation of the data.

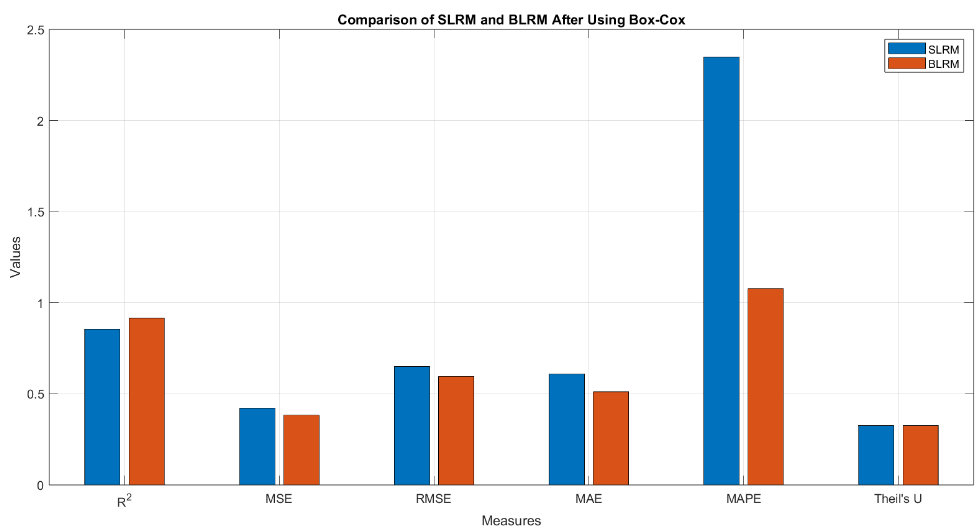

FIGURE 6. Result analysis Linear Regression and Bayesian Analysis Before Transformation of the data.

Steps 2: Find the Marginal Posterior distribution using the join posterior.

Steps 3: For the provided variable, determine the condition probability of the unknown parameter.
Steps 4: Determine the posterior distribution's mean. The posterior distribution's mean is thus unbiased For the unknown parameter $\hat{\beta}$.
Steps 5: In the posterior predictive distribution, change the values of X and Y to make a prediction. and the complete calculation procedures are given in the flow chart.

TABLE 1. Result of Models using Evaluating Criteria

	Before using Box-Cox		After using Box-Cox	
Measures	SLRM	BLRM	SLRM	BLRM
R^2	0.8032	0.8180	0.8554	0.9154
MSE	24.3173	24.3239	0.4222	0.3823
RMSE	12.0962	11.90657	0.6498	0.5963
MAE	10.7563	10.6033	0.6077	0.5116
MAPE	12.456	12.6838	2.3487	1.0777
Theils U	0.763	0.762	0.3260	0.3260

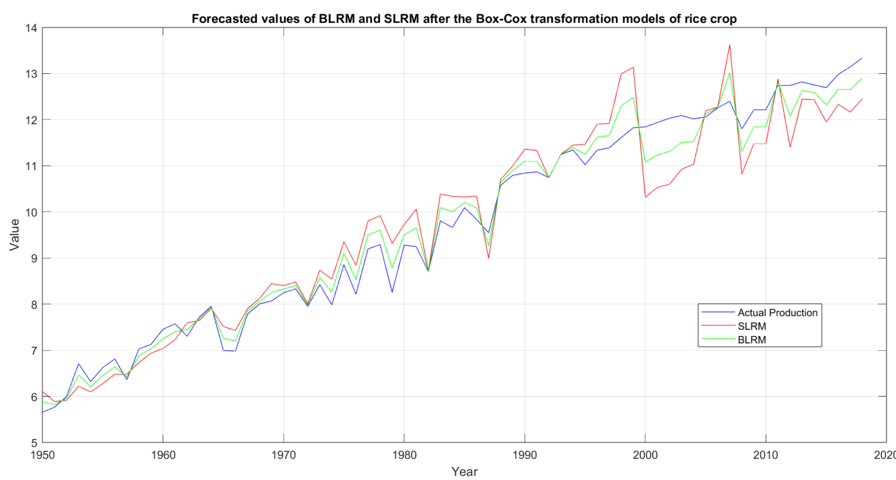

FIGURE 7. Forecasted values of BLRM and SLRM after the Box-Cox Transformation models of rice crop .

R^2, MAPE, and Theil's U are compared from Table.1 as done for the research. The discrepancy between SLRM and BLRM was not all that noticeable before the Box-Cox Transformation. Therefore, the dataset requires Transformation. Using a Box-Cox transformation to make forecasts from a dataset slight adjustment were made to the SLRM and BLRM after the Box-Cox transformation so that more accurate results could be drawn. We found that after Box-Cox, the MAPE of the SLRM and BLRM significantly decreased from $10\% - 15\%$ to $5\% - 6\%$. When compared to R^2 in both scenarios, there was an increase in it to 91% of the total values.

In Figure 7 show that the Forecasted values of BLRM and SLRM after the Box-Cox Transformation models of rice crop.

5. Conclusions

The goal of the current study was to develop, forecast, and predict production (million/tons) of the rice crop from 1950 to 2019 using SLRM and BLRM. However, we noticed that the dataset had normality of error, so we chose to transform the data using the Box-Cox methodology and evaluate the model's performance. Mainly focus on SLRM and BLRM for better prediction when predicting the production before and after using the Box-Cox Transformation. we discover that the model following the Box-Cox transformation produces superior outcomes than the model used initially. Overall, our findings suggest that incorporating the Box-Cox transformation can significantly improve the predictive performance of the models and enhance their overall accuracy.

Conflicts of interest : The authors confirm that there is no conflict of interest to declare for this publication.

Data availability : All data generated and analyzed availed from The Directorate of Economics and Statistics, Ministry of Agriculture, India, provided data on rice yields for the years 1962–2019.
$https : //eands.dacnet.nic.in/APY_96_To_06.htm$

References

1. Tonglin Zhang, Baijian Yang, *Box-Cox Transformation in Big Data*, Technometrics **59** (2017), 189-201. DOI: 10.1080/00401706.2016.1156025
2. R. Gulzar, I. Sajjad, M.Y. Bhat, and S.U. Rehman, *Simple Ranked Sampling Scheme: Modification and Application in The Theory of Estimation of Erlang Distribution*, J. APPL. MATH. and INFORMATICS **41** (2023), 449-468.
3. R. Gommes, *Agrometerological Crop Yield Forecasting Methods*, The Netherlands : International Statistical Institute, 1998.
4. S. Chatterjee, A.S. Hadi, *Regression analysis by example*, John Wiley and Sons, 2015.
5. W.W. Guo, H. Xue, *Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models*, Mathematical Problems in Engineering, 2014.
6. U. Verma, D.S. Dabas, M.S. Grewal, J.P. Singh, R.S. Hooda, M. Yadav, *Crop yield forecasting in Haryana: 1986 to 2010*, Summary Report, Department of Soil Science, CCS HAU, Hisar (Haryana), 2011, 1-148.
7. P. Mishra, A.M.G. Al Khatib, I. Sardar, J. Mohammed, K. Karakaya, A. Dash, M. Ray, L. Narsimhaiah, A. Dubey, *Modeling and Forecasting of Sugarcane Production in India*, Sugar Tech. **23** (2021), 1317-1324. DOI. https://doi.org/10.1007/s12355-021-01004-3

8. S. Veera, P. Eswaran, *Prediction of Crop Yield using Regression Analysi*, Indian Journal of Science and Technology **9** 2016. DOI. 10.17485/ijst/2016/v9i38/91714.
9. C.V. Shakila, S.K. Khadar Babu, *Wheat Crop Yield Forecasting Using Various Regression Models*, Ratio Mathematica **46** 2023. DOI. http://dx.doi.org/ 10.23755/rm.v46i0.1085
10. A. Natale, S. Antognelli, E. Ranieri, E. Cruciani, A. Boggia, *A novel cleaning method for yield data collected by sencors: A case study on winter cereals*, 20th International Conference, Cagliari, Italy, July, 2020, 1-4.
11. N. Choudhury, S. Saurav, R.R. Kumar, N. Budhlakoti, *Modelling and Forecasting of Total Area, Irrigated Area, Production and Productivity of Important Cereal Crops in India towards Food Security*, International Journal of Current Microbiology and Applied Sciences **6** (2017), 2591-2600. DOI. https://doi.org/10.20546/ijcmas.2017.610.305
12. P. Kumari, D.J. Parmar, K. Sathish, A.B. Mahera, Y.A. Lad, *Prediction of area, production, and productivity of total fruit crops in Gujarat*, The Pharma Innovation Journal **11** (2022), 750-754.
13. D.J. Prity Kumari, Parmar, M. Sathish Kumar, A.B. Mahera and Y.A. Lad, *Comparison of Statistical Models for Prediction Area, Production and Yield of Citrus in Gujarat*, Biological Forum-An International Journal **14** (2022), 540-545.
14. P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*, Cambridge University Press, 2005.
15. D. Bunn, *A Bayesian Approach to the Linear Combination of Forecasts* Operational Research Quarterly, **26** 1975. DOI. 10.1057/jors.1975.67.
16. J.M. Bernardo, A.F. Smith, *Bayesian theory*, Wiley Series, New York, 2001.
17. M.D. Yeasin, K. SINGH, A. Lama, B. Gurung, *Improved weather indices based Bayesian regression model for forecasting crop yield*, MAUSAM **72** (2021), 879-886. DOI. 10.54302/mausam.v72i4.670
18. S.K. Khadar Babu, C. Chesneau, V. Anthonysamy, S. Ravichandaran, C.V. Shakila, *Efficacy of the Rice Crop Growth Using Different Smoothing Methods*, Journal of Applied Analysis and Computation **12** (2022), 2593-2599. DOI. 10.11948/20220116.
19. S. Das, *Generalized linear models and beyond: An innovative approach from Bayesian perspective*, Dissertation, University of Connecticut, 2008.
20. J.M. Marin, N.S. Pillai, C.P. Robert, J. Rousseau, *Relevant statistics for Bayesian model choice*, Journal of the Royal Statistical Society : Series B (Statistical Methodology) **76** (2014), 833-859.
21. N. Gandhi, L.J. Armstrong, O. Petkar, *Predicting Rice crop yield using Bayesian networks*, International Conference on Advances in Computing. Communications, and Informatics (ICACCI), 2016, 795-799. DOI. 10.1109/ICACCI.2016.7732143
22. R.J. Howarth, S.A.M. Earle, *Application of a generalized power transformation to geochemical data*, Journal of the International Association for Mathematical Geology **11** (1979), 45-62. DOI. https://doi.org/10.1007/bf01043245
23. J.W. Osborne, *Improving your data transformations: Applying the Box-Cox transformation Practical Assessment*, Research and Evaluation **15** (2010), 1-9.
24. M. Bicego, S. Baldo, *Properties of the Box–Cox transformation for pattern classification*, Neurocomputing **218** (2016), 390-400.
25. Y. Peng, T. Feng, H.J.P. Timmermans, *Expanded comfort assessment in outdoor urban public spaces using Box-Cox transformation*, Landscape and Urban Planning **190** 2019. DOI. 10.1016/j.laplan.2019.103594
26. Y. He, Y. Zheng, Q. Xu, *Forecasting energy consumption in Anhui province of China through two Box-Cox transformation quantile regression probability density methods*, Measurement **136** (2019), 579-593.
27. M. Meloun, M. Sanka, P. Nemec, S. Kritkova, K. Kupka, *The analysis of soil cores polluted with certain metals using the Box–Cox transformation*, Environmental Pollution **137** (2005), 273-280.

28. A.C Atkinson, M. Riani, A. Corbellini, *The Box-Cox Transformation: Review and Extensions*, Statistical Science **36** (2021). DOI. https://doi.org/10.1214/20-sts778

Victor Anthonysamy received the B.Sc. and M.Sc. degrees in Mathematics from the Sacred Heart College (Autonomous), Tirupattur, Tamil Nadu, India. He is currently pursuing the Ph.D. degree with the Vellore Institute of Technology (VIT), Vellore, India. He has nine years of experience in teaching and research.He has published about 5 research articles in reputed journals. His research interests include stochastic processes, time series analysis, forecasting models, and advanced statistical modeling.

Department of Mathematics, Vellore Institute of Technology (VIT), Vellore;Tamil Nadu, India-632014.
`e-mail: avictorarun@gmail.com`

Khadar Babu SK is a Professor in the Department of Mathematics, School of Advanced Sciences, Vellore Institute Technology, Vellore, Tamil Nadu, India. He has 25 years of teaching and research experience. His areas of research are statistical quality control, statistical modeling, time series analysis, computer networks, and digital image processing. He is a member of different societies. He has published about 94 research articles in various national and international journals and presented about 120 articles at national and international conferences, seminars, symposiums, and workshops.

Department of Mathematics, Vellore Institute of Technology (VIT), Vellore;Tamil Nadu, India-632014.
`e-mail: khadarbabu.sk@vit.ac.in`