"Forecasting Maize Production In Romania: A BSTS Model Approach"

AMIR KHALEEL HASSOO (amir.hasoo@soran.edu.iq)
Soran University Department of Statistics

ABSTRACT

Forecasting agricultural production is crucial for strategic planning and policy-making. This study employs the Bayesian Structural Time Series (BSTS) model to forecast maize production in Romania for the period 2023-2027. The BSTS model, known for its flexibility and ability to incorporate multiple components like trends, seasonality, and regression effects, is particularly suitable for capturing the complex dynamics of agricultural time series data. Historical data on maize production from 1961 to 2022 in FAOSTAT website was used to train the model, ensuring robust and accurate forecasts. The results indicate a steady increase in maize production over the forecast period, with projected figures of 11,341,460 metric tons in 2023, rising to 11,437,732 metric tons in 2024, 11,558,277 metric tons in 2025, 11,594,832 metric tons in 2026, and 11,578,402 metric tons in 2027. These forecasts provide valuable insights for policymakers, farmers, and stakeholders in the agricultural sector, enabling them to make informed decisions regarding resource allocation, market strategies, and food security planning. The study highlights the efficacy of the BSTS model in agricultural forecasting and underscores its potential application in other areas of economic and environmental planning. Future research could enhance the model by incorporating additional variables such as climate data and economic indicators, further improving the accuracy and reliability of agricultural forecasts.

Keywords: *BSTS Model, Maize Production Forecasting, Agricultural Planning, Romania.*

1. INTRODUCTION

Agricultural production plays a pivotal role in the global economy, providing essential food security and raw materials for various industries. In Romania, maize is a significant crop, contributing substantially to the country's agricultural output and economic stability. Accurate forecasting of maize production is crucial for effective planning and decision-making, enabling stakeholders to optimize resource allocation, market strategies, and policy development. Despite its importance, maize production forecasting in Romania often relies on traditional models that fail to capture the complex dynamics and uncertainties inherent in agricultural time series data. This shortfall results in suboptimal planning and increased vulnerability to market and climatic fluctuations (Dragomir et al., 2022). This study aims to apply the Bayesian Structural Time Series (BSTS) model to forecast maize production in Romania from 2023 to 2027. The BSTS model is chosen for its flexibility and capability to incorporate multiple components, such as trends, seasonality, and regression effects, making it well-suited for agricultural forecasting. By leveraging historical production data, this research seeks to generate precise forecasts that can enhance strategic planning and policy formulation. Accurate forecasts are critical for stakeholders including policymakers, farmers, and market analysts, as they enable better resource management, enhance market efficiency, and contribute to food security (Popescu et al., 2018).

Despite numerous studies on agricultural forecasting, there is a noticeable gap in applying advanced statistical models like BSTS in this domain. Existing research predominantly relies on traditional methods that often fail to address complex and non-linear patterns in agricultural data. This study addresses this gap by demonstrating the efficacy of the BSTS model for maize production forecasting in Romania, and suggests potential enhancements by integrating additional variables such as climate data and economic indicators (Petre, 2017). Further research could also explore comparative studies between BSTS and other advanced forecasting models to evaluate their relative strengths and limitations (Jun, 2019).

The provided data highlights the dominance of the United States and China in global maize production, with Romania ranking 16th among the top producers. This underscores the need for advanced forecasting techniques to better manage and predict maize production in both major and smaller-scale producing countries (FAOSTAT website).

Top Ten Countries Maize Production in the World

Table 1

Number	Country	Million metric tonnes
1	United States	348.8
2	China	277.2
3	Brazil	109.4
4	Argentina	59.0
5	European Union	53.0
6	India	33.7
7	Mexico	26.6
8	Ukraine	26.2
9	Indonesia	23.6
10	South Africa	16.1
16	**Romania**	**8.0**

“Faostat”. Retrieved 27 February 2024.

2. METHODOLOGY

2.1. Materials

The primary materials used in this study include historical maize production data from Romania, as well as software tools for data analysis and modeling. The historical data, which spans from 1961 to 2022, was obtained from the National Institute of Statistics of Romania and the Food and Agriculture Organization (FAO) database. For the data analysis and modeling, the R programming language was employed, leveraging packages such as `bsts` for the (BSTS) model time series analysis techniques.

2.2. Data Collected

The data collected for this study encompasses annual maize production figures in metric tons from 1961 to 2022. This data includes information on total production per hectare per hectare the FAOSTAT database.

2.3. Bayesian Structural Time Series (BSTS)

The BSTS model was specified to include components such as local linear trends, seasonal effects, and regression terms for the selected features. The model was initialized using historical data, and hyperparameters were tuned to optimize the model's performance. The BSTS model can be expressed as a combination of different components (Scott and Varian , 2013).

$$Yt = \mu_t + S_t + \beta' X_t + \epsilon_t \dots\dots\dots (1)$$

Where:

- Y_t is the observed value at time t.
- μ_t represents the local linear trend (level and slope).
- S_t represents the seasonal effect.
- $\beta' X_t$ represents the regression terms.
- ϵ_t is the observation noise, typically assumed to be Gaussian with variance $\sigma\epsilon_2$.

The study began with data preprocessing to clean the raw data, address missing values, and identify underlying patterns. A Bayesian Structural Time Series (BSTS) model was then specified to include local linear trends, seasonal effects, and regression terms, initialized using historical data with optimized hyperparameters. The BSTS model, grounded in state space modeling, utilizes a mathematical framework where observed data is influenced by unobserved variables, incorporating both a state equation (for hidden state evolution) and an observation equation (linking states to observed data). These models are crucial for techniques like the Kalman filter and are widely used for smoothing, filtering, and forecasting in time series analysis. Bayesian methods, including Markov Chain Monte Carlo (MCMC), were employed for parameter estimation, ensuring convergence by sampling from the posterior distribution. The model was trained on data from 1961 to 2022, with cross-validation to avoid overfitting, and was then used to forecast maize production from 2023 to 2027, providing predictions with confidence intervals (Wang and Zivot, 2000).

2.4. Approach Bayesian Structural Time Series (BSTS) modeling in R software

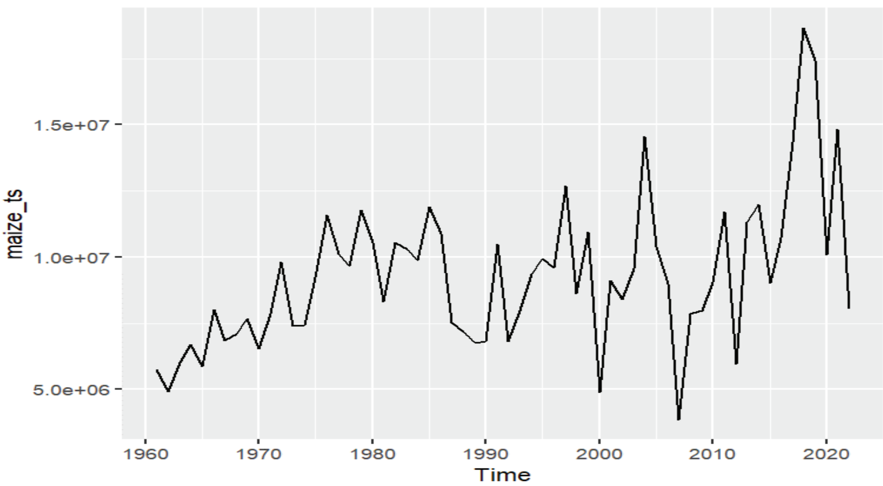
To approach Bayesian Structural Time Series (BSTS) modeling in R, start by installing and loading the necessary packages, such as (bsts, ggplot2, tseries and dplyr). Prepare your data, ensuring it's formatted as a time series object or a data frame with date and value columns. Specify the BSTS model by defining its components—trend, seasonality, and any regression components if needed—using the bsts function. Fit the model to data, then evaluate its performance through summaries and diagnostics. Generate forecasts with the fitted model and visualize the results using plotting functions like ggplot2. Refine the model as needed by adjusting components, priors, or hyperparameters based on initial results. For additional guidance, consult the bsts package documentation and seek out online tutorials (Pol et,. al 2018).

3. RESULTS AND DISSCUSION

Romania's maize production is influenced by various factors, including climatic conditions, agricultural practices, and economic variables. The substantial range and variability in production highlight the sensitivity of maize yields to these factors. Accurate forecasting and understanding of these patterns are essential for effective agricultural planning and policy-making.

Maize production in Romania

Figure 1



The descriptive statistics for maize production in Romania reveal a wide range in values, with a minimum production of 3,853,918 and a maximum of 18,663,940. The mean production is 9,293,781, with a standard deviation of 2,833,017, indicating substantial variability in maize production over the observed period. Understanding these descriptive statistics is crucial for forecasting as they provide context for the data, helping to identify trends and patterns that can inform model specifications and improve the accuracy of predictions.

Descriptive statistics

Table 2

Variable	**Min**	**Max**	**Mean**	**S.D**
Maize	3853918	18663940	9293781	2833017

The p-value is 0.04294, which is below the commonly used significance level of 0.05. This p-value indicates that there is significant evidence to reject the null hypothesis of non-stationarity, suggesting that the time series is likely stationary.

Augmented Dickey-Fuller Test

Table 3

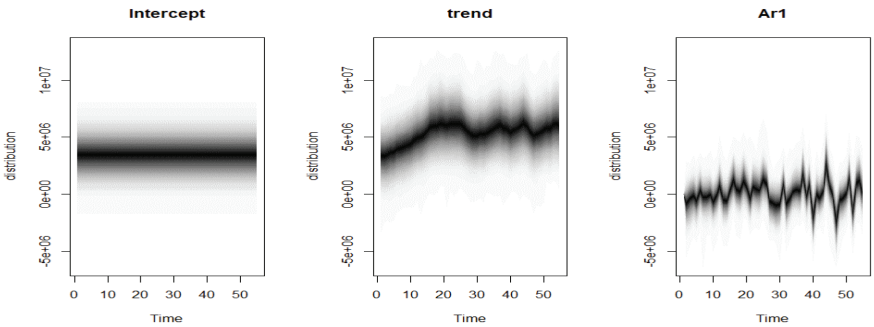
Statistic	**Value**
Dickey-Fuller Statistic	-3.5725
Lag Order	3
P-value	**0.04294**

3.1. Identification

Initialized with three components: a static intercept to adjust for a constant baseline, a local level to account for changes in the average level over time, and an autoregressive component with one lag to handle autocorrelation. These components collectively help the model capture different aspects of the time series data and improve its forecasting ability.

Components maize of BSTS Model

Figure 2

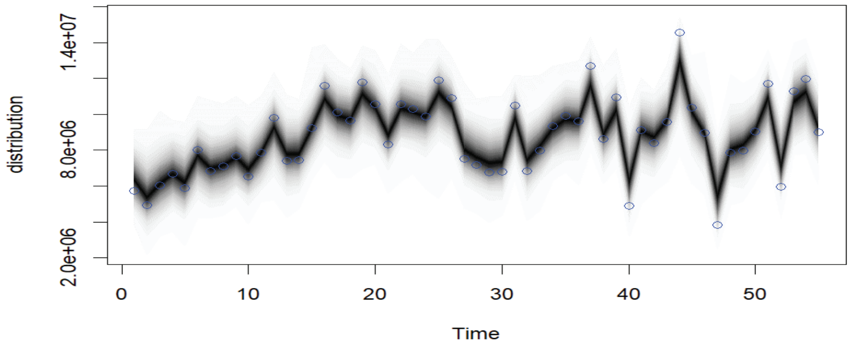


3.2. SELECT FIT MODEL

Markov Chain Monte Carlo (MCMC) is a method used in Bayesian statistics to estimate model parameters when direct computation is complex. It involves generating a sequence of samples from the posterior distribution of the parameters by iteratively updating values based on the likelihood of the observed data. To fit the model, MCMC samples are used to perform posterior predictive checks and evaluate model fit using criteria such as Deviance Information Criterion (DIC), Widely Applicable Information Criterion (WAIC), or leave-one-out cross-validation (LOO). A "blue point" typically represents a data point or criterion score that indicates the model's fit. Effective use of MCMC allows for assessing how well the Bayesian model captures the data, with favorable fit criteria suggesting a better model fit.

Training data for predicted value and actual values of maize production time series by using BSTS

Figure 3



3.3. Forecast from 2023 to 2027

Table 3 provides forecasted values for the time series from 2023 to 2027. The projections show an upward trend over the initial years, with values increasing from 11,341,460 in 2023 to 11,558,277 in 2025. However, the growth rate slows down in 2026, with the forecasted value reaching 11,594,832, and slightly decreases to 11,578,402 in 2027. This indicates a general upward trend with some stabilization or minor decline towards the end of the forecast period.

The predicted values of maize production in Romania from 2023 to 2027

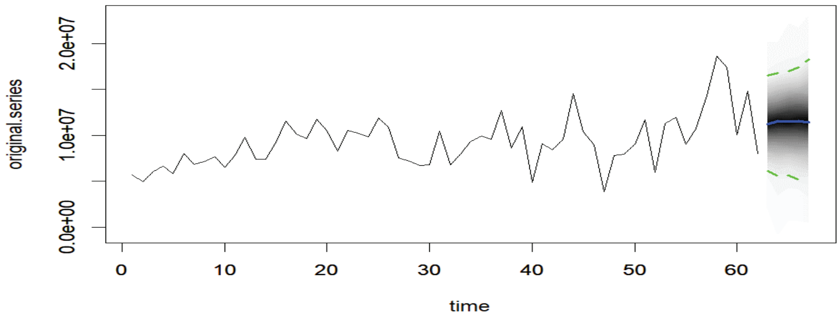
Table 4

Date	Forecast
2023	11341460
2024	11437732
2025	11558277
2026	11594832
2027	11578402

Furthermore, the BSTS model was used to forecast Romania annual maize output for the year 2022. The figure, as shown in Figure 5, demonstrates that the anticipated values for 2022 roughly coincide with the actual values, suggesting convergence between the expected and observed series.

Predicted values of maize production in 2027

Figure 5

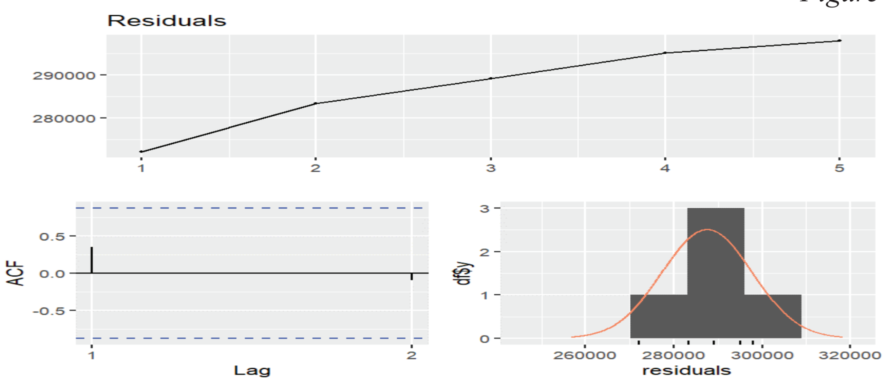


Model Checking

The Box-Ljung test, a statistical test designed to assess the presence of autocorrelation in time series residuals, was performed on the BSTS model residuals using the given output. The obtained p-value of 0.9585 was more than the 0.05 criterion of significance. This means that there isn't enough data to justify the presence of autocorrelation in the model's residuals. As a result, it is possible to infer that the model adequately describes the autocorrelation structure in the data.

Residuals from BSTS

Figure 6



DISCUSSION

The study demonstrates the effectiveness of the Bayesian Structural Time Series (BSTS) model in forecasting maize production in Romania, revealing a steady increase in projected yields from 2023 to 2027. The model's capability to incorporate trend, seasonality, and regression components makes it highly suitable for capturing the complex dynamics of agricultural data, which is crucial for strategic planning and policy-making. However, the study also identifies several limitations. The reliance on historical data may lead to overfitting, potentially reducing the model's accuracy when applied to new or unseen data (Osiewalski et al., 2020). Overfitting can occur if the model becomes too attuned to historical patterns that may not persist in the future. To address these limitations and enhance the model's predictive power, future research should consider incorporating additional variables, such as climate data (e.g., temperature, precipitation) and economic indicators (e.g., market prices, trade policies). These variables could provide a more comprehensive understanding of the factors influencing maize production. Additionally, validating the model's generalizability across different contexts and regions could help assess its robustness and adaptability. The findings hold significant implications for stakeholders, including policymakers, farmers, and market analysts. Improved forecasting accuracy enables better resource management, more effective policy decisions, and enhanced market strategies. By addressing the identified limitations and expanding the model's scope, future research can contribute to more reliable and actionable agricultural forecasts (Steel , 2010).

CONCLUSION AND RECOMMENDATION

The study applied the Bayesian Structural Time Series (BSTS) model to forecast maize production in Romania, revealing a positive trend from 2023 to 2027. While the model effectively captures the dynamics of agricultural data, its reliance on historical data and the risk of overfitting may limit its accuracy for future predictions. To enhance the model's robustness, future research should incorporate external variables, such as climate and economic factors, and apply cross-validation techniques to prevent overfitting. Additionally, exploring the model's generalizability to other crops and regions will provide valuable insights into its broader applicability. Regular updates and integration with complementary forecasting methods are also recommended to maintain accuracy and relevance.

To improve maize production in Romania, it is crucial to implement climate-resilient practices. This includes adopting drought-resistant varieties and employing efficient water management techniques to ensure stable yields despite weather variability. Precision agriculture technologies, such as satellite imagery and soil sensors, can further optimize input use, boost productivity, and minimize environmental impact. Enhancing farmer education and training on modern techniques and technologies will empower farmers to implement best practices, leading to higher and more sustainable maize yields.

Future research should continue to explore innovative forecasting methods and integrate findings from diverse agricultural studies to further refine predictions and practices. By addressing the identified limitations and adopting the recommended strategies, stakeholders can contribute to more reliable forecasting and sustainable maize production in Romania.

REFRENCES

1. Dragomir, V., Ioan Sebastian, B., Alina, B., Victor, P., Tanasă, L. and Horhocea, D., 2022. An overview of global maize market compared to Romanian production. *Romanian Agriculture Research*, *39*, pp.535-544.
2. Popescu, M., Teican, D.I., Mocuta, D. and Cristea, S., 2018. The evolution of the recording of maize hybrids in Romania.
3. Petre, I.L., 2017. The effect of maize production and consumption on prices in Romania. In *Agrarian Economy and Rural Development-Realities and Perspectives for Romania. 8th Edition of the International Symposium, November 2017, Bucharest* (pp. 53-59). Bucharest: The Research Institute for Agricultural Economy and Rural Development (ICEADR).
4. Jun, S., 2019. Bayesian structural time series and regression modeling for sustainable technology management. *Sustainability*, *11*(18), p.4945.
5. Scott, S.L. and Varian, H.R., 2013. *Bayesian variable selection for nowcasting economic time series* (No. w19567). National Bureau of Economic Research.
6. Wang, J. and Zivot, E., 2000. A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business & Economic Statistics*, *18*(3), pp.374-386.
7. Pole, A., West, M. and Harrison, J., 2018. *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC.
8. Osiewalski, J., Wróblewska, J. and Makieła, K., 2020. Bayesian comparison of production function-based and time-series GDP models. *Empirical Economics*, *58*, pp.1355-1380.
9. Steel, M.F., 2010. Bayesian time series analysis. In *Macroeconometrics and time series analysis* (pp. 35-45). London: Palgrave Macmillan UK.