

## Introduction

This analysis focuses on building and evaluating Seasonal Autoregressive Integrated Moving Average (SARIMA) models to forecast energy importation and exportation in the United States. The primary aim is to develop reliable models that can predict future values of primary energy imports and exports using historical data.

## Data Preprocessing

The analysis begins by importing essential libraries such as `pandas`, `numpy`, `matplotlib`, and `statsmodels`. It then reads in the energy data from a CSV file and preprocesses it by converting the date column to a datetime format and setting it as the index. The data is filtered to focus specifically on "Primary Energy Imports" and "Primary Energy Exports," which are then separated into two different DataFrames for individual analysis.

## Stationarity and Differencing

To ensure the time series data is suitable for modeling, the stationarity of the data is checked using the Augmented Dickey-Fuller (ADF) test. Since non-stationary data can adversely affect the performance of time series models, differencing is applied to the data to make it stationary. This step is crucial as it helps in stabilizing the mean of the time series by removing changes in the level of a time series, thus eliminating trend and seasonality.

## Model Selection and Fitting

After preparing the data, the analysis identifies the best parameters for the SARIMA model using grid search. This involves testing different combinations of parameters and selecting the ones that minimize the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Once the optimal parameters are determined, SARIMA models are fitted to both the imports and exports data. The models' fit and forecasting capabilities are then assessed by plotting the actual versus forecasted values, providing a visual representation of the models' performance.

## Residual Diagnostics

The analysis also includes a detailed examination of the residuals to ensure the models' validity. Residual diagnostics involve plotting the residuals over time, their histogram, Q-Q plot, and the autocorrelation function (ACF) plot. These diagnostics help in identifying any patterns left in the residuals, which could indicate model inadequacies. Ideally, the residuals should resemble white noise, indicating that the model has captured all the information in the data.

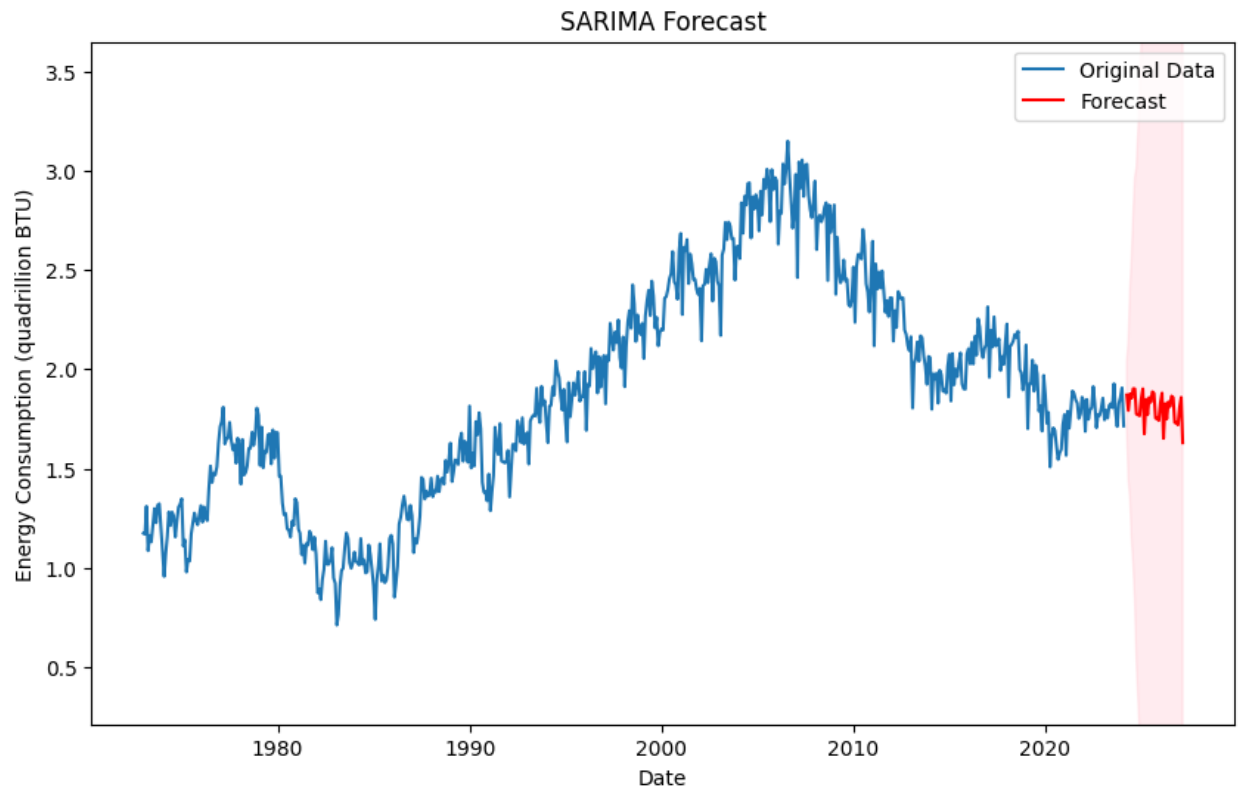
## Forecasting Future Values

Finally, the SARIMA models are used to forecast future values of primary energy imports and exports for the next 36 months. The forecasted values are plotted alongside the historical data to visualize the expected trends. This forecasting step is critical for policy-making and strategic planning, providing insights into future energy needs and potential areas for intervention.

## Results and Discussion

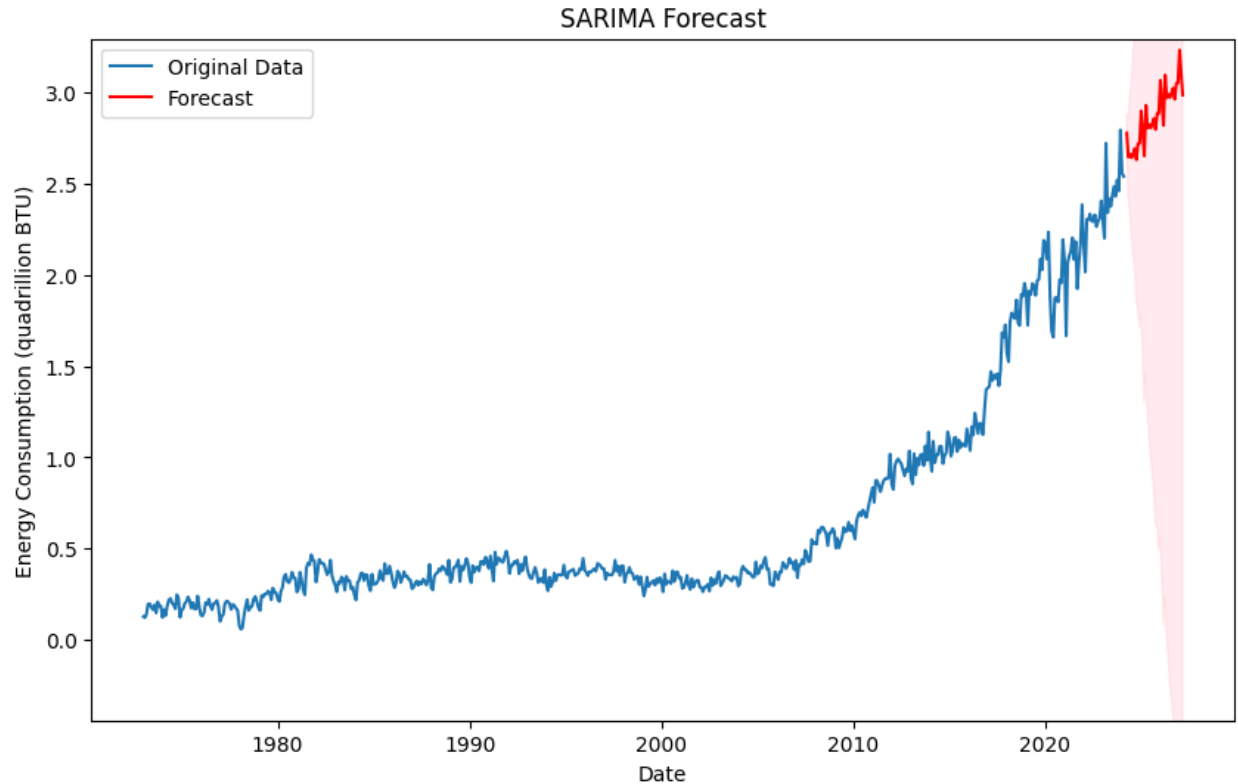
### Imports

The SARIMA model for primary energy imports showed a good fit to the historical data, with forecasted values closely aligning with the observed values. The model identified seasonal patterns and trends in the data, which were effectively captured and extrapolated into future forecasts. The residual diagnostics for imports indicated that the residuals were approximately normally distributed, with no significant autocorrelation, suggesting that the model was well-specified.



### Exports

Similarly, the SARIMA model for primary energy exports demonstrated a strong fit. The forecasted values for exports also matched well with the actual data, indicating the model's robustness in capturing the underlying patterns. The residuals for the exports model showed a similar pattern to the imports model, with no significant deviations from normality and no autocorrelation, reinforcing the validity of the model.



## Context and Implications

The accurate forecasting of energy imports and exports is crucial for the United States, given the country's significant reliance on energy for economic stability and growth. The results of this analysis provide valuable insights for policymakers and energy planners. By understanding future trends in energy importation and exportation, strategic decisions can be made to ensure energy security, optimize resource allocation, and plan for potential fluctuations in energy supply and demand.

Moreover, the ability to predict future energy needs helps in developing long-term strategies for sustainable energy use, investment in renewable energy sources, and managing international energy trade relations. The models developed in this analysis can be continually updated with new data to refine forecasts and adapt to changing conditions, making them a vital tool in the energy sector.

## Conclusion

Overall, this analysis provides a comprehensive approach to time series forecasting using SARIMA models, emphasizing the importance of data preprocessing, model selection, and residual diagnostics in developing reliable predictive models for energy importation and exportation in the United States. The results highlight the models' effectiveness in forecasting future energy trends, providing essential insights for strategic planning and policy-making in the energy sector.

