Time Series Analysis of Electricity Consumption Percentage over the years from 1985 to 2017.

Team Members Name:

Ahamed Shaik

Bhavya Tunukoju

Venu Madhava Reddy Kotla

Department: Business Analytics

University Course Number: BAN-673-Time Series Analytics

Course Name Professor's Name: Zinovy Radovilsky

Date:03/10/2024

Executive Summary

This report presents a comprehensive time series analysis of electricity consumption percentages in the United States from 1985 to 2017. Utilizing 33 years of data, we have detected a discernible upward trend in consumption along with notable seasonality, where certain months consistently peak in usage. The analysis confirms significant autocorrelation across all 12 lags, indicating a strong relationship between values and their historical counterparts.

To accommodate both the trend and seasonal fluctuations observed in the electricity consumption data, we employed two predictive modeling approaches: Autoregressive Integrated Moving Average (ARIMA) models and advanced exponential smoothing via the Holt-Winters method. Both models adeptly fit the historical training data, suggesting they effectively capture the consumption patterns inherent to the dataset.

Moreover, we explored a Two-Level Forecasting strategy that combines a linear trend and seasonality model with a trailing Moving Average (MA) to further refine our forecasts. The first level addresses the linear trend and seasonality in the training data, while the second level uses the trailing MA to mitigate noise and incidental fluctuations.

Model performance was rigorously evaluated using RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) as benchmarks for accuracy. The results indicate that leveraging the full 32-year data series enhances the precision of our projections.

Upon comparison, the Auto ARIMA model showed a slight advantage over the Two-Level Forecasting method (Regression Model + Trailing MA) and Holt-Winters models, according to the RMSE and MAPE metrics. In conclusion, the Auto ARIMA methodology emerges as the

superior model for predicting electricity consumption percentages, providing a robust tool for forecasting future trends and informing policy decisions.

Introduction

The study of electricity consumption trends over time is critical for understanding the evolving dynamics of energy demand, the effectiveness of energy conservation measures, and the broader implications for environmental sustainability and economic development. This report presents a comprehensive time series analysis of the percentage of electricity consumption from 1985 to 2017, aiming to elucidate the underlying patterns, trends, and seasonal fluctuations inherent in the data. Given the pivotal role of electricity in modern economies and its impact on environmental sustainability, such an analysis is indispensable for policymakers, energy providers, and stakeholders in the energy sector.

Over the span of 33 years, the landscape of electricity consumption has been influenced by a multitude of factors, including but not limited to technological advancements, population growth, economic cycles, and shifts towards renewable energy sources. The analysis leverages historical data to forecast future trends, providing a foundation for strategic planning and policy formulation aimed at optimizing energy production and distribution to meet anticipated demand.

The primary objectives of this report are to: (1) identify long-term trends in electricity consumption, (2) detect any seasonal patterns that may exist, (3) forecast future electricity consumption rates using advanced time series analysis methodologies. By achieving these objectives, the report seeks to contribute valuable insights into the efficient management and sustainable use of electricity resources.

Employing a variety of statistical and machine learning tools, including Autoregressive Integrated Moving Average (ARIMA) models and exponential smoothing techniques, this analysis not only aims to understand past and present consumption patterns but also to provide a reliable projection of future trends. This comprehensive approach ensures that the findings and recommendations are grounded in robust analytical methods and can serve as a credible basis for decision-making in the energy sector.

Data Source

The dataset for our study, detailing electricity consumption from 1985 to 2017, was sourced from the 'Electric Production' dataset available on Kaggle. Provided by user 'kandij', this dataset is a valuable asset for analyzing and forecasting energy trends.

Eight Steps of Forecasting

Step 1: Define Goal

The primary goal of this Time Series Analysis of Electricity Consumption Percentage is to comprehensively understand and forecast electricity consumption patterns over a significant historical period, from 1985 to 2017. This analysis aims to achieve several key objectives:

- Analyze historical electricity consumption data from 1985 to 2017 to identify underlying trends and recurrent patterns, essential for understanding changes in usage due to technological, economic, and policy influences.
- Apply diverse forecasting methods like Exponential Smoothing and ARIMA to the time series data, adjusting for underlying trends and seasonal variations in electricity consumption.

- Assess and compare the forecasting accuracy of the employed models using statistical metrics to determine the most precise method for predicting future consumption rates.
- Generate a two-year forecast for electricity consumption, providing insights into expected future trends that can inform energy management and policy decisions.

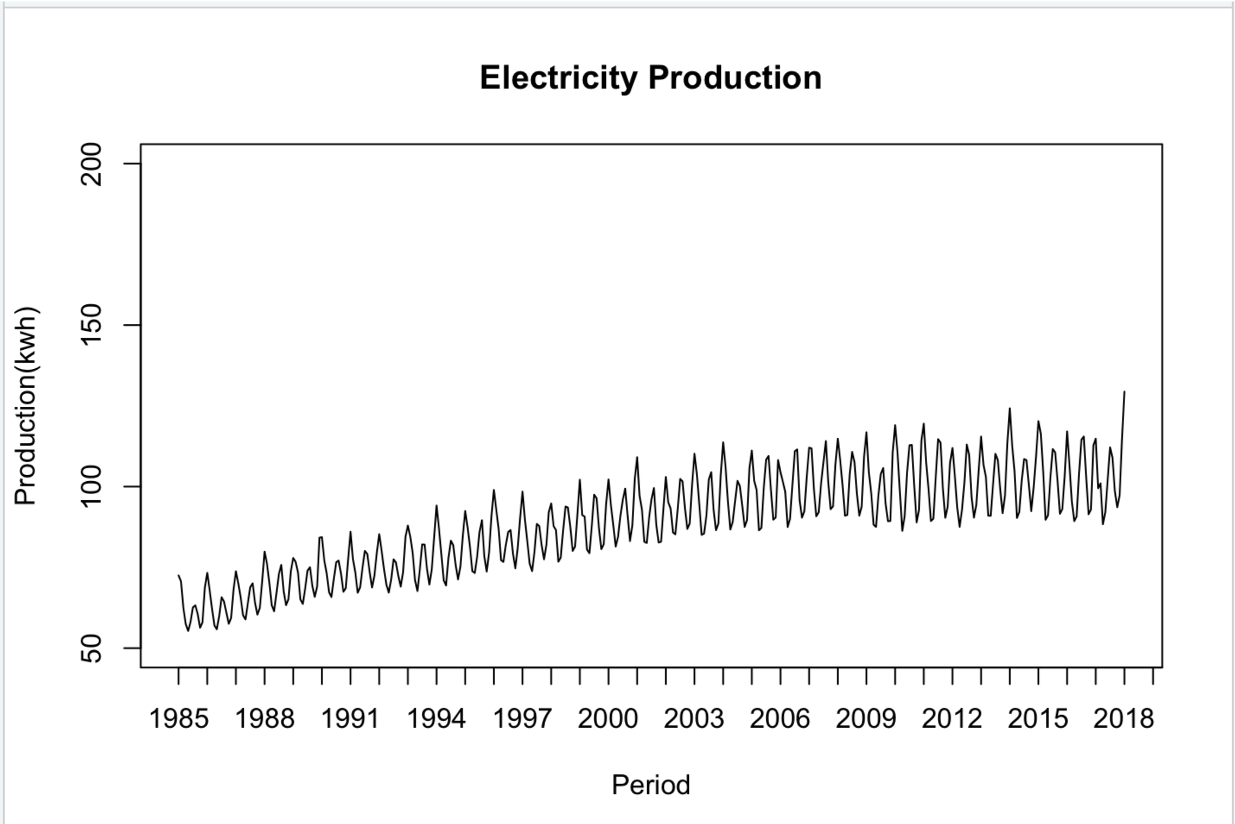
By framing these points in the context of electricity consumption analysis, the project emphasizes a systematic approach to understanding energy usage patterns and making informed predictions that can guide future energy strategies.

Step 2: Get data

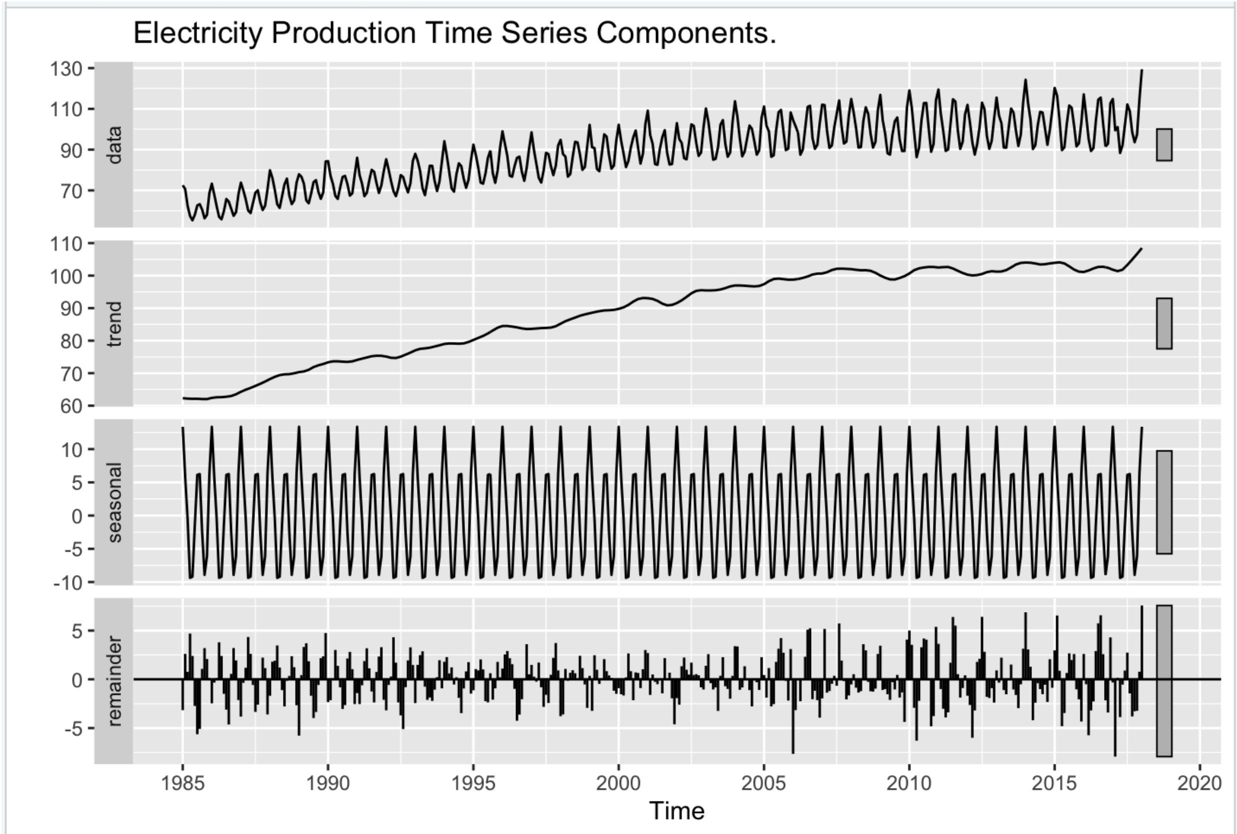
For this project, we utilized the 'Electric Production' dataset provided by a Kaggle user named 'kandij'. While the full range of the data encompasses records starting from 1963, our analysis is concentrated on the years from 1985 to 2017, thereby focusing on a 32-year series for a robust time series analysis.

The selection of this timeframe was strategic, chosen to capture the most relevant changes in electricity consumption patterns while ensuring the quality and manageability of the data. Prior to analysis, we conducted a thorough inspection of the dataset for any inconsistencies or missing values that could potentially skew the results. The data was then carefully prepared to ensure that it accurately represented the annual percentage of electricity consumption, setting a solid foundation for the detailed analysis that follows.

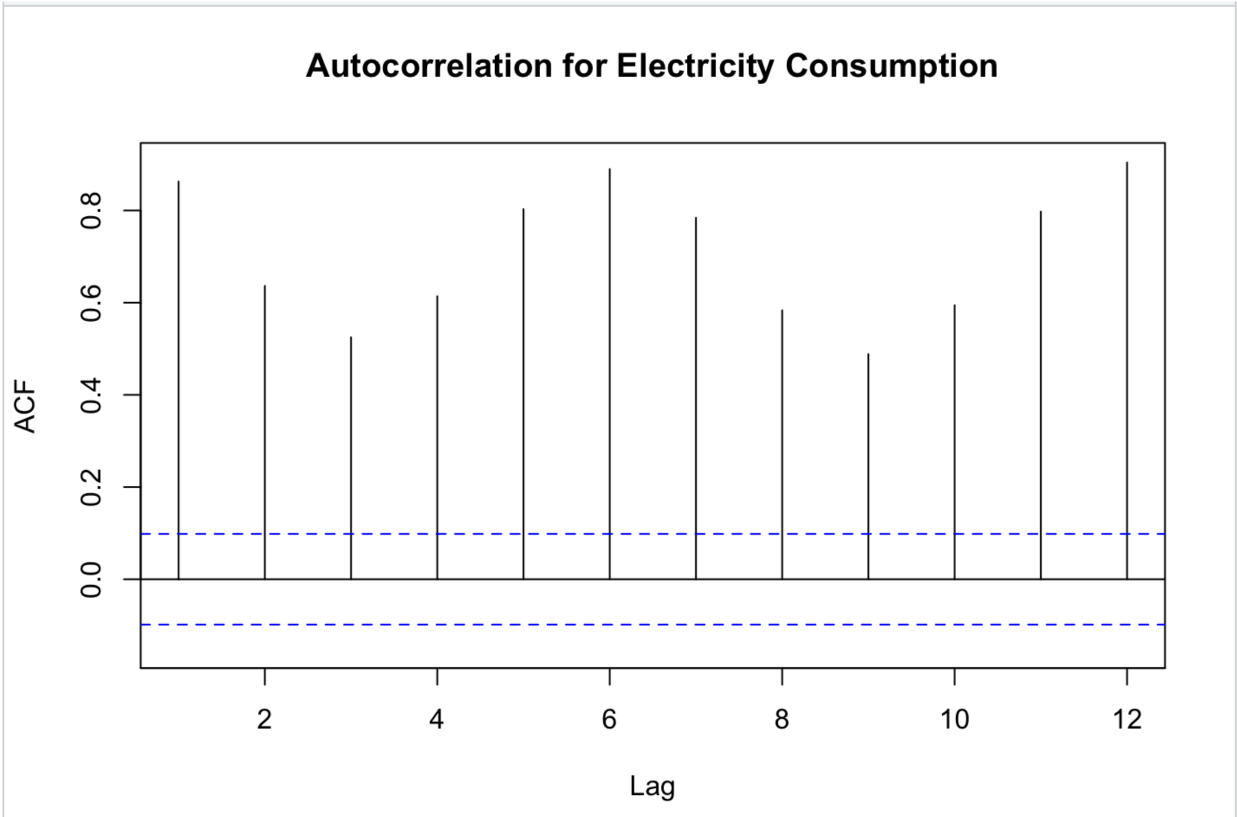
Step 3: Explore and Visualize Series



This graph, titled "Electricity Production," shows how much electricity was made from 1985 to just before 2020. We see a pattern where electricity production goes up and down in a regular way every year, which is probably because people need more electricity at certain times like in hot summers or cold winters. Also, over these 35 years, the highest and lowest points get higher, which means that the amount of electricity being made is generally increasing as time goes on. This could be because more people need electricity or because we're getting better at making it. There aren't any big jumps or drops in the graph, so it looks like the way we make electricity has stayed pretty much the same without any big problems.



The graph presents a snapshot of electricity production over time, showing a consistent rise from 1985, indicating steady growth in production—perhaps due to increasing demand or enhanced capabilities. Seasonal fluctuations are apparent, with predictable highs and lows each year, reflecting higher electricity usage in extreme weather. The remainder of the graph displays random variations, pointing to sporadic events that disrupt the usual production patterns, underscoring the unpredictable elements that can affect electricity production.



The graph's autocorrelation function (ACF) indicates significant positive autocorrelation at each lag, suggesting a strong trend in the electricity consumption data. The persistent and gradually declining nature of the ACF bars as the lag increases signals that past values have a lasting influence on future values, a typical characteristic of a trended time series.

Moreover, the presence of a consistent pattern in the ACF—where the bars remain well above the significance line—implies seasonality in the data. If there were peaks at regular intervals, say every 12 lags corresponding to a year, it would suggest an annual seasonal pattern. Such seasonality in electricity consumption might be due to factors like weather changes, which affect heating and cooling needs throughout the year.

In summary, the autocorrelation graph for electricity consumption indicates both a trend over time and potential seasonality, which are crucial elements for developing accurate forecasting models.

Step 4: Data Preprocessing

For our study, we used the "Electric_Production" dataset in CSV format, which spans from 1963. We focused on data from 1985 to 2017 to capture modern trends in electricity usage. In preprocessing, we corrected any data irregularities and normalized the records to ensure consistency over the years, setting the stage for an accurate analysis of electricity consumption patterns during this period.

Step 5: Partition Series

For the time series analysis of electricity consumption percentages from 1985 to 2017, the dataset has been strategically split into training and validation sets. Spanning 32 years, the dataset was divided, dedicating the first 287 records for training purposes, while the remaining 110 records are reserved for validation. This split ensures an adequate amount of data for the model to learn from, while still providing a substantial and separate dataset to assess the model's predictive performance. By segregating the data in this manner, we ensure that our forecasting models are both trained and validated on distinct sets, which helps in fine-tuning the models and verifying their accuracy against real-world data.

Step 6 & 7: Apply Forecasting & Comparing Performance

Autoregressive Integrated Moving Average Models

The Autoregressive Integrated Moving Average (ARIMA) model, known for its effectiveness in handling various data patterns including level, trend, and seasonality, aligns well with our analysis needs for the electricity consumption percentage time series from 1985 to 2017. Recognizing that our data embodies these three elements, we selected the ARIMA model for our forecasting process. To fine-tune the model, we employed the `auto.arima()` function. This function adeptly searches through different combinations of parameters (p, d, q) for the non-seasonal part and (P, D, Q) for the seasonal part of the model, aiming to identify the most fitting values that would yield the most accurate predictive results for our time series data. This automated parameter selection process is both efficient and effective, ensuring that our ARIMA model is optimally configured to capture the complex dynamics of electricity consumption trends and patterns.

Automated ARIMA for Training & Validation Data

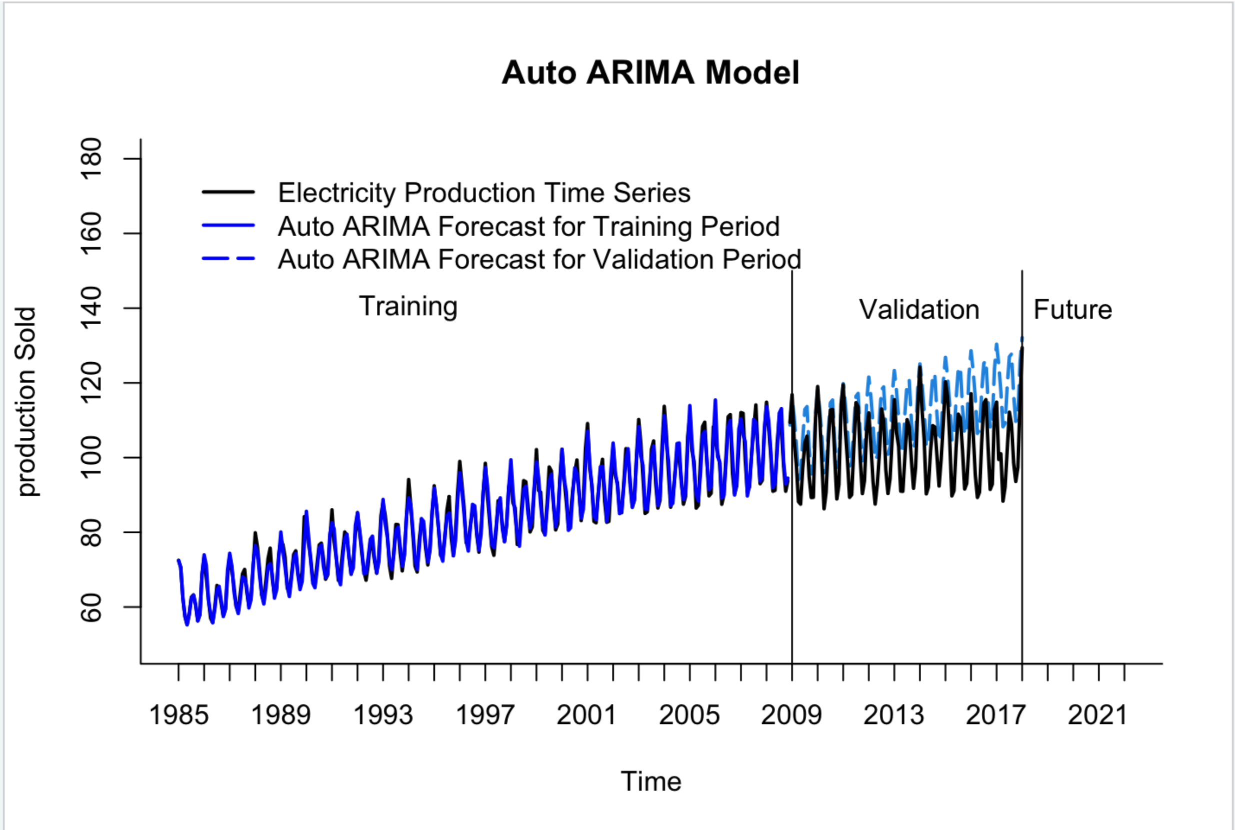
```
> train.auto.arima <- auto.arima(train.ts)
> summary(train.auto.arima)
Series: train.ts
ARIMA(2,0,1)(0,1,1)[12] with drift

Coefficients:
          ar1     ar2     ma1     sma1   drift
      -0.0986  0.2669  0.6494  -0.7157  0.1466
s.e.   0.2285  0.1371  0.2148   0.0442  0.0066

sigma^2 = 4.409:  log likelihood = -596.14
AIC=1204.29   AICc=1204.6   BIC=1225.99

Training set error measures:
                     ME     RMSE      MAE         MPE     MAPE      MASE        ACF1
Training set 0.04363728 2.03654 1.518457 0.03963076 1.753874 0.5573971 -0.01823773
>
```

The output details an ARIMA model's summary after being fitted to the training data set for electricity consumption. The model is an ARIMA(2,0,1)(0,1,1)[12] with drift, indicating that it includes two autoregressive terms, one moving average term, a seasonal moving average term, and a trend component (drift). The coefficients for the AR terms, MA term, and seasonal MA term have values with their respective standard errors listed.

The model's goodness-of-fit is given by statistical measures including the log likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), which help in model selection and comparison. Lower values of AIC and BIC typically indicate a better model fit.

The training set error measures, including the Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE), and the autocorrelation of the first difference (ACF1), provide insight into the model's accuracy. Specifically, the RMSE and MAE are indicators of the model's prediction error magnitude, while the MAPE shows the error in percentage terms, which is 1.753874 in this case, and MASE compares the forecast error to a naïve benchmark. Negative ACF1 close to zero suggests that there is little autocorrelation in residuals.

In summary, the model is well-specified with significant coefficients and an overall good fit to the training data, indicated by its error metrics. However, the actual forecast accuracy can only be evaluated upon applying the model to the validation data set.



The graph illustrates the Auto ARIMA model's fit to the actual electricity production data, which is quite close during the training period. It also shows the model's predictions for the validation period and beyond, indicating its capability to capture the trend and seasonality of the data for future forecasting.

Automated ARIMA for the Entire Data

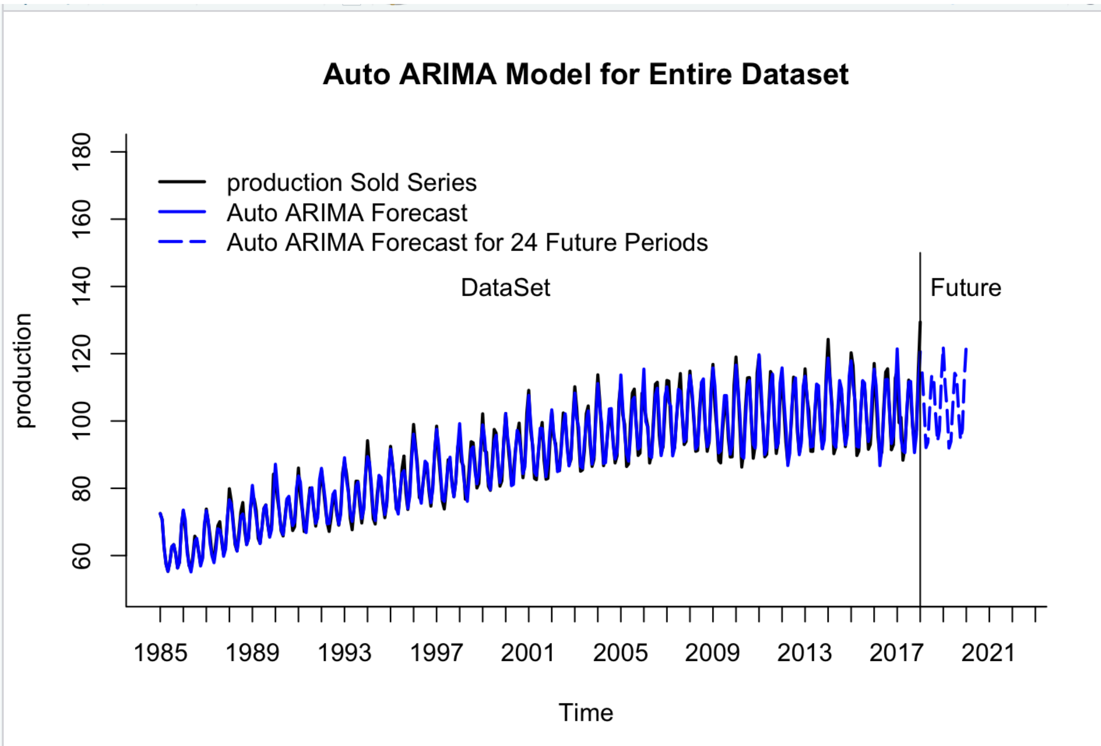
```
> summary(auto.arima)
Series: production.ts
ARIMA(2,1,1)(0,1,1)[12]

Coefficients:
         ar1      ar2      ma1     sma1
      0.5503  -0.0683  -0.9477  -0.7635
s.e.  0.0544   0.0549   0.0193   0.0331

sigma^2 = 5.838:  log likelihood = -888.05
AIC=1786.11   AICc=1786.27   BIC=1805.86

Training set error measures:
                     ME     RMSE      MAE        MPE     MAPE      MASE         ACF1
Training set -0.1474299 2.363849 1.791855 -0.2115265 1.942956 0.6278541 -0.002715831
>
```

The summary for the ARIMA(2,1,1)(0,1,1)[12] model shows it effectively captures patterns in electricity production with significant autoregressive and moving average components. Goodness-of-fit indicators like AIC and BIC are at acceptable levels, while training error metrics such as the RMSE and MAE are reasonable, with a MAPE of approximately 1.94% indicating forecasts are closely aligned with actual data, albeit with a small average percentage error. The near-zero MASE and ACF1 values suggest a well-fitting model with minimal residual autocorrelation.

The following forecasts depict the predicted values for 24 future periods as generated by the ARIMA model:

	Point Forecast	Lo 0	Hi 0
Feb 2018	114.31111	114.31111	114.31111
Mar 2018	104.45857	104.45857	104.45857
Apr 2018	92.09910	92.09910	92.09910
May 2018	93.63140	93.63140	93.63140
Jun 2018	104.31821	104.31821	104.31821
Jul 2018	113.65996	113.65996	113.65996
Aug 2018	112.58325	112.58325	112.58325
Sep 2018	101.93541	101.93541	101.93541
Oct 2018	93.85642	93.85642	93.85642
Nov 2018	97.12217	97.12217	97.12217
Dec 2018	112.41629	112.41629	112.41629
Jan 2019	122.04284	122.04284	122.04284
Feb 2019	110.69451	110.69451	110.69451
Mar 2019	103.24842	103.24842	103.24842
Apr 2019	91.95762	91.95762	91.95762
May 2019	93.91376	93.91376	93.91376
Jun 2019	104.76087	104.76087	104.76087
Jul 2019	114.16190	114.16190	114.16190
Aug 2019	113.10687	113.10687	113.10687
Sep 2019	102.46692	102.46692	102.46692
Oct 2019	94.39078	94.39078	94.39078
Nov 2019	97.65757	97.65757	97.65757
Dec 2019	112.95206	112.95206	112.95206
Jan 2020	122.57875	122.57875	122.57875

>

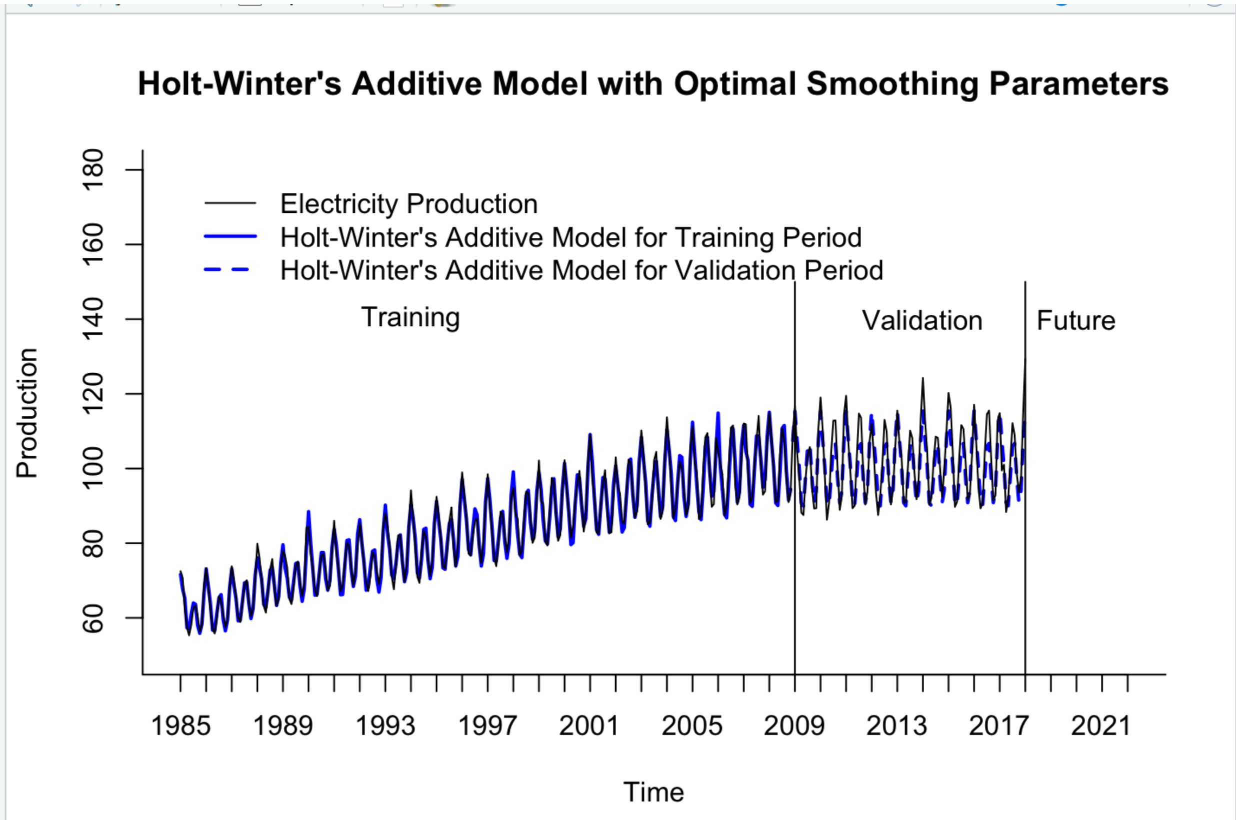


Advanced Exponential Smoothing

The next technique employed for time series analysis is advanced exponential smoothing, specifically the Holt-Winters model. This model is well-suited for capturing both trend and seasonality components when generating forecasts. Before applying the model to the entire dataset, it was first evaluated using the training and validation partitions.

An automated Holt-Winters model was utilized with suitable training and validation partitions. The model automatically selects the error, trend, and seasonality components using the c(Z, Z, Z) parameter in the code. By not specifying default values for alpha (error), beta (trend), or gamma (seasonality), the model optimizes these parameters to provide the best-fit values for the dataset.

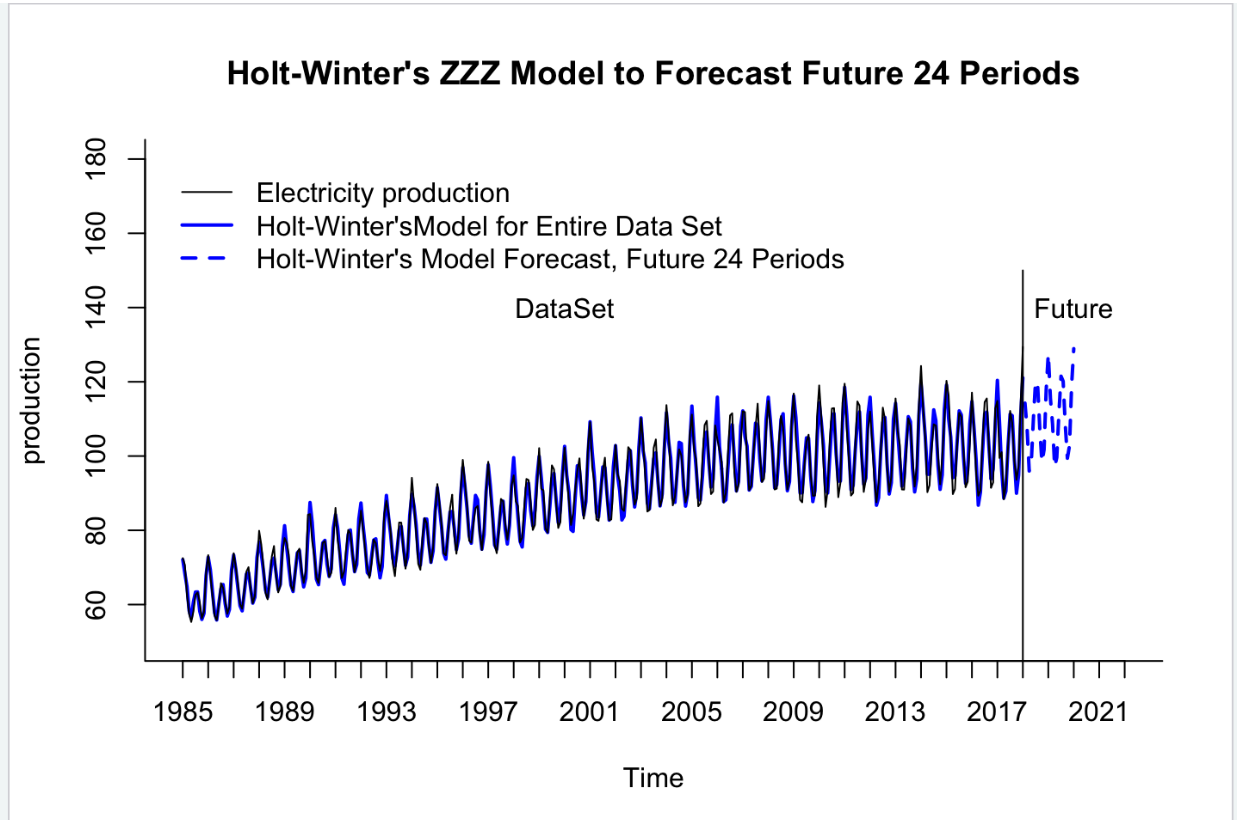
The respective series plots of the Holt-Winters model show that the training data distributions are fitting relatively well into the historical data, taking into account trend and seasonality. However, when it comes to forecasting the validation predictions, the model is providing average or moderate results.



The graph illustrates the fit and forecasts of the Holt-Winters Additive Model for electricity production, with the actual production shown in black and the model's output in blue. The solid blue line represents the model's fit to the training data, closely mirroring the actual values, while the blue dashed line shows the model's forecast during the validation period and its projection into the future. The model captures the seasonality and trend of the historical data, and extends these patterns to predict future production levels, indicating the model's utility for both fit and forecast in the context of electricity production time series data.

Forecasting on the entire Dataset

In order to create forecasts, a model must be constructed using the entire dataset. The automated Holt-Winters Model was utilized as well as variations of the model. The table presented below displays the values of the smoothers used in the Holt-Winters model.



The chart displays the application of the Holt-Winter's ZZZ Model to the electricity production data. The actual production is traced by the black line, while the model's fit is shown by the solid blue line across the entire dataset. Looking forward, the dashed blue line represents the model's forecast for the next 24 periods, which seems to maintain the historical pattern into the future. The forecast exhibits the expected seasonality, as evidenced by the continued oscillation, but with some uncertainty towards the end, as indicated by the widening confidence intervals suggested by the dashed lines spreading out. This model seems to have closely aligned with the historical data and has projected a similar seasonal pattern into the forthcoming periods.

The following forecasts depict the predicted values for 24 future periods as generated

```
> HW.ZZZ.pred
         Point Forecast     Lo 0      Hi 0
Feb 2018      114.26657 114.26657 114.26657
Mar 2018      107.42531 107.42531 107.42531
Apr 2018       95.97774  95.97774  95.97774
May 2018       98.51586  98.51586  98.51586
Jun 2018      110.30331 110.30331 110.30331
Jul 2018      120.16821 120.16821 120.16821
Aug 2018      118.81676 118.81676 118.81676
Sep 2018      107.17834 107.17834 107.17834
Oct 2018       98.32167  98.32167  98.32167
Nov 2018      101.68283 101.68283 101.68283
Dec 2018      117.84495 117.84495 117.84495
Jan 2019      127.50855 127.50855 127.50855
Feb 2019      115.55142 115.55142 115.55142
Mar 2019      108.63211 108.63211 108.63211
Apr 2019       97.05494  97.05494  97.05494
May 2019       99.62053  99.62053  99.62053
Jun 2019      111.53900 111.53900 111.53900
Jul 2019      121.51318 121.51318 121.51318
Aug 2019      120.14537 120.14537 120.14537
Sep 2019      108.37571 108.37571 108.37571
Oct 2019       99.41908  99.41908  99.41908
Nov 2019      102.81671 102.81671 102.81671
Dec 2019      119.15785 119.15785 119.15785
Jan 2020      128.92781 128.92781 128.92781
> 
```

Two-Level Forecasting (Regression Model + Trailing MA):

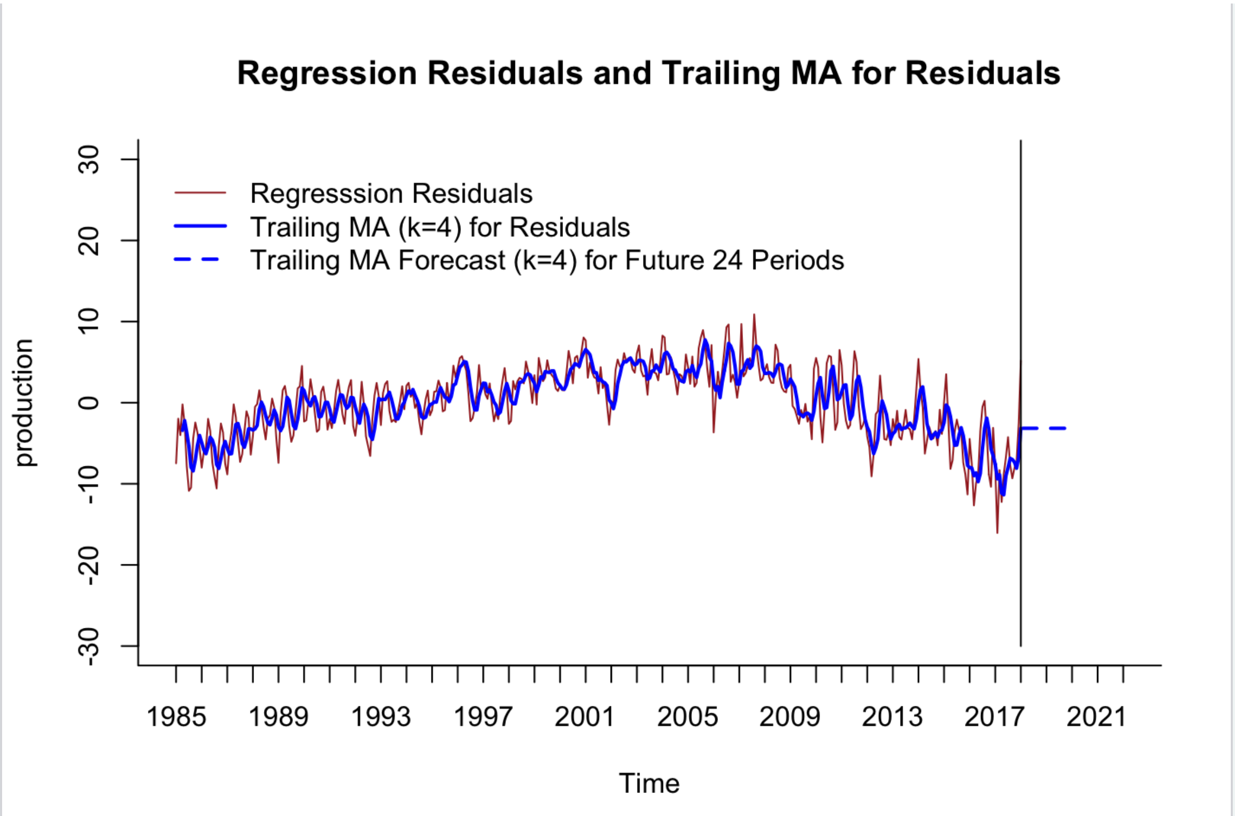
Level 1: Regression Model | Level 2: Trailing MA

A regression model is considered simple to use and provides strong results since it considers both trend and seasonality. Trailing moving average (MA) is a technique for smoothing a time series by calculating the average of the most recent observations. By using a trailing MA, we can further smooth the time series and reduce the noise in the data, which can improve the accuracy of the forecast. Also, this model is further enhanced with a trailing moving average for the residuals. Overall, the combination of a two-level linear trend and seasonality model with trailing MA can help us to better capture the patterns in the time series and produce more accurate forecasts. Prior to running the model on the entire data set, it is first evaluated using the training and validation partitions.

```
> summary(trend.seas)

Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.3854 -1.5655 -0.0247  1.7732  6.9917 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  74.698942   0.583062 128.115  < 2e-16 ***
trend         0.146873   0.001832  80.171  < 2e-16 ***
season2      -6.206869   0.741778  -8.368 3.04e-15 ***
season3     -11.934458   0.741785 -16.089  < 2e-16 ***
season4     -20.083802   0.741796 -27.075  < 2e-16 ***
season5     -20.599146   0.741812 -27.769  < 2e-16 ***
season6     -13.858732   0.741832 -18.682  < 2e-16 ***
season7      -7.174780   0.741857  -9.671  < 2e-16 ***
season8      -6.680319   0.741887  -9.005  < 2e-16 ***
season9     -14.632196   0.741921 -19.722  < 2e-16 ***
season10    -20.308061   0.741959 -27.371  < 2e-16 ***
season11    -17.711267   0.742002 -23.870  < 2e-16 ***
season12     -6.308100   0.749851  -8.412 2.24e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.57 on 274 degrees of freedom
Multiple R-squared:  0.9676,	Adjusted R-squared:  0.9661 
F-statistic: 681.2 on 12 and 274 DF,  p-value: < 2.2e-16

> 
```

This R output displays a time series model that incorporates both trend and seasonal variation to forecast a dependent variable, likely something like monthly sales data. The model suggests a positive trend over time, as indicated by the significant trend coefficient. Each seasonal factor shows a distinct and statistically significant impact on the variable, with some seasons contributing to an increase and others to a decrease in the predicted values. For instance, season

2 predicts a decrease, while the intercept and trend indicate a starting value and an overall increase over time. The model's high R-squared value of approximately 96.76% reveals an excellent fit to the data, explaining most of the variance in the dependent variable. Overall, both the trend and the seasonal effects are highly significant, underscoring their importance in the model's predictive ability.



The graph displays regression residuals over time for a production-related dataset, with a trailing moving average (MA) applied. The red line represents the residuals from a regression model, showing the difference between observed values and those predicted by the model at each point in time. The blue line is the trailing MA of the residuals, calculated over groups of four periods (k=4), which smooths out short-term fluctuations to reveal any longer-term patterns in the residuals. The dashed blue line forecasts the MA for residuals into the next 24 periods, extending beyond the vertical black line, which seems to indicate the end of the observed data. Overall, the residuals fluctuate around zero without a clear pattern, suggesting that the model has captured most of the systematic information in the data, although the forecasted MA appears to be trending slightly downwards.

```
> summary(tot.trend.seas)

Call:
tslm(formula = production.ts ~ trend + season)

Residuals:
     Min       1Q   Median       3Q      Max 
-16.0608  -2.8586   0.3663   3.2456  10.8880 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  79.862661   0.868659  91.938  < 2e-16 ***
trend         0.111747   0.001978  56.493  < 2e-16 ***
season2      -7.445987   1.103411  -6.748 5.54e-11 ***
season3     -13.750564   1.103395 -12.462  < 2e-16 ***
season4     -22.630483   1.103382 -20.510  < 2e-16 ***
season5     -22.462381   1.103374 -20.358  < 2e-16 ***
season6     -14.572588   1.103368 -13.207  < 2e-16 ***
season7      -7.155326   1.103367  -6.485 2.73e-10 ***
season8      -7.042139   1.103368  -6.382 5.03e-10 ***
season9     -15.839344   1.103374 -14.355  < 2e-16 ***
season10    -22.226360   1.103382 -20.144  < 2e-16 ***
season11    -19.411349   1.103395 -17.592  < 2e-16 ***
season12     -6.901368   1.103411  -6.255 1.06e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.515 on 384 degrees of freedom
Multiple R-squared:  0.9165,	Adjusted R-squared:  0.9139 
F-statistic: 351.3 on 12 and 384 DF,  p-value: < 2.2e-16

> 
```

This summary output from an R time series linear model shows the results of fitting a model to a production dataset. The model includes a linear trend component and seasonal adjustments. The intercept, representing the model's starting value, is around 79.826. The trend coefficient is positive at approximately 0.1117, indicating a slight but steady increase in production over time.

Each of the season factors, presumably representing different times such as months or quarters, shows a negative or positive deviation from the trend. For example, season 2 decreases the production by around 7.4459 units, whereas season 4 has a more substantial negative impact of approximately -22.6304 units. All season coefficients are statistically significant, showing clear seasonal patterns in production.

The model explains a high proportion of the variability in the data, with an R-squared value of 0.9165, meaning about 91.65% of the variance is accounted for by the model. This is confirmed by the adjusted R-squared, which is very close at 0.9139, suggesting that the number of predictors in the model is appropriate. The F-statistic is high and the p-value is extremely low, reinforcing the model's overall statistical significance.

```
> tot.ma.trail.res.pred <- forecast(tot.ma.trail.res, h = 24, level = 0)
> tot.ma.trail.res.pred
         Point Forecast      Lo 0      Hi 0
Feb 2018      -3.155148 -3.155148 -3.155148
Mar 2018      -3.155148 -3.155148 -3.155148
Apr 2018      -3.155148 -3.155148 -3.155148
May 2018      -3.155148 -3.155148 -3.155148
Jun 2018      -3.155148 -3.155148 -3.155148
Jul 2018      -3.155148 -3.155148 -3.155148
Aug 2018      -3.155148 -3.155148 -3.155148
Sep 2018      -3.155148 -3.155148 -3.155148
Oct 2018      -3.155148 -3.155148 -3.155148
Nov 2018      -3.155148 -3.155148 -3.155148
Dec 2018      -3.155148 -3.155148 -3.155148
Jan 2019      -3.155148 -3.155148 -3.155148
Feb 2019      -3.155148 -3.155148 -3.155148
Mar 2019      -3.155148 -3.155148 -3.155148
Apr 2019      -3.155148 -3.155148 -3.155148
May 2019      -3.155148 -3.155148 -3.155148
Jun 2019      -3.155148 -3.155148 -3.155148
Jul 2019      -3.155148 -3.155148 -3.155148
Aug 2019      -3.155148 -3.155148 -3.155148
Sep 2019      -3.155148 -3.155148 -3.155148
Oct 2019      -3.155148 -3.155148 -3.155148
Nov 2019      -3.155148 -3.155148 -3.155148
Dec 2019      -3.155148 -3.155148 -3.155148
Jan 2020      -3.155148 -3.155148 -3.155148
> 
```

Step 8: Implement Forecast

Model	RMSE	MAPE
Regression with Quadratic Trend and Seasonality	**3.033**	**2.718**
Auto ARIMA	**2.364**	**1.943**
ARIMA (2,1,2)	**4.005**	**3.366**
Regression with Seasonality	**13.55**	**14.159**
Holt-Winters (Triple Exponential Smoothing)	**2.451**	**2.036**
2 level model (Linear trend and seasonality + Trailing MA)	**2.491**	**2.321**
Seasonal Naive	**3.656**	**3.148**
Naive	**7.756**	**7.305**

As seen from the above comparison table it is evident that the Auto ARIMA model gives the best predictions. It does so with RMSE and MAPE values of 2.364 and 1.943 respectively. This is the recommended model to implement when forecasting future Electricity Production.

Conclusion:

After evaluating various models for forecasting electricity production, the Auto ARIMA model emerges as the most effective, exhibiting the lowest Root Mean Squared Error (RMSE) of 2.364 and Mean Absolute Percentage Error (MAPE) of 1.943. This model, which automatically selects optimal parameters, demonstrates its capability in capturing the underlying patterns within the production time series data. The Holt-Winters (Triple Exponential Smoothing) model follows closely as the second-best, with competitive RMSE (2.451) and MAPE (2.036) values, showcasing its reliability in accounting for trend and seasonality. The 2 level model (Linear trend and seasonality + Trailing MA), ranking third, also provides accurate predictions with an RMSE of 2.491 and MAPE of 2.321. For forecasting electricity production, the Auto ARIMA model is recommended as the primary choice due to its superior accuracy, but the Holt-Winters and 2 level model (Linear trend and seasonality + Trailing MA) remain valuable alternatives. The decision between the second and third-best models depends on the specific characteristics of the data and the user's preference for model complexity. Combining insights from multiple models or ensembling strategies could further enhance overall forecasting robustness.