Forecasting Median Home Sale Prices in NJ, USA

1. About the Data

About

Zillow see's listings nationwide. Taking advantage of the vast amount of listing data, Zillow has been able to produce monthly indexes for various data points of interest. For this mid-term, we will look at the median home prices for House Listings in New Jersey.

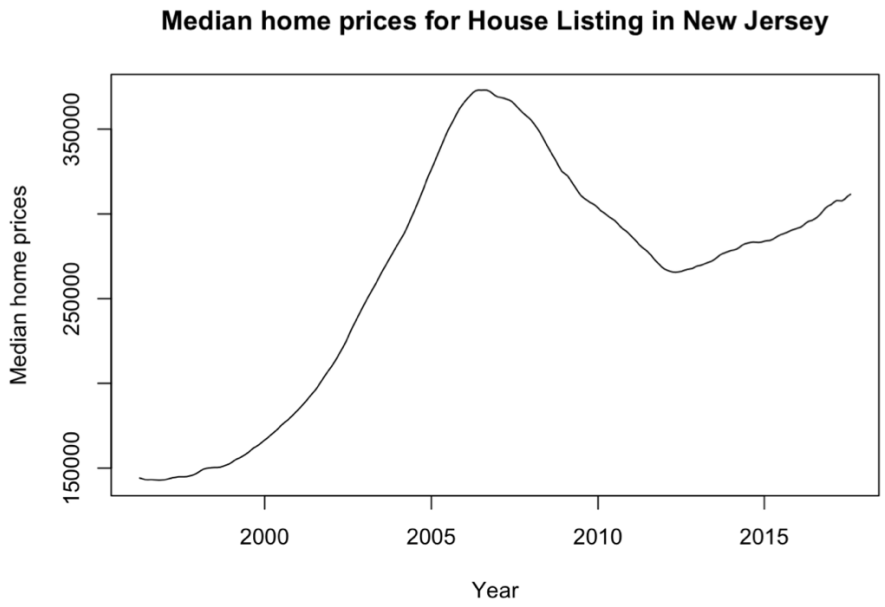
Data Source

Link: https://www.zillow.com/research/data/#median-home-value

Data Dictionary

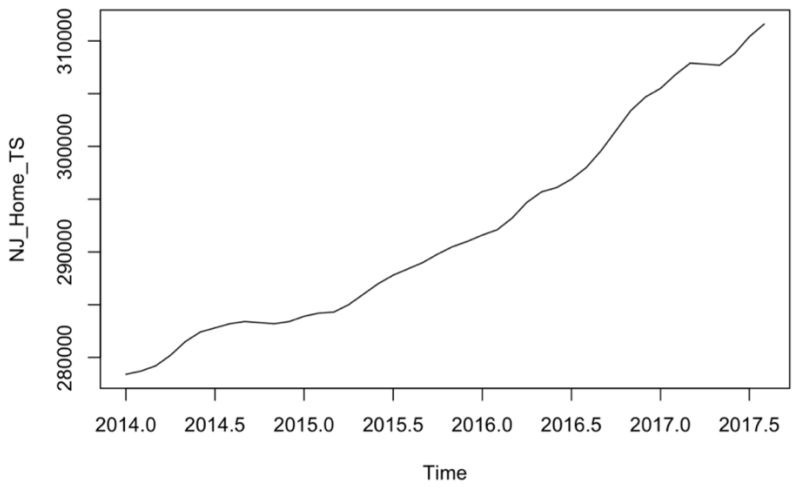
YYYY-MM: Year and Month during with the data was recorded
Value: Median Listing Price of properties in New Jersey, USA

2. Plot & Inferences

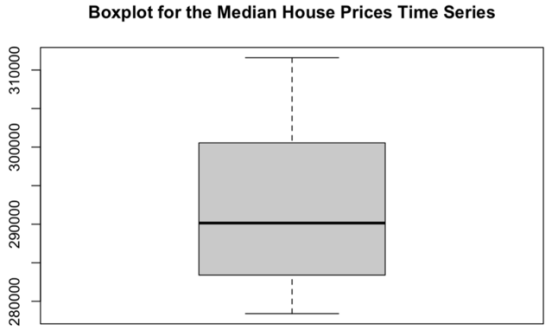


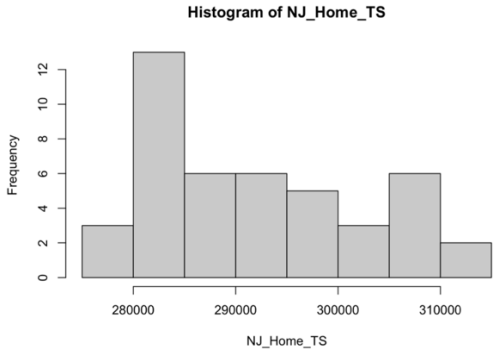
- The plot shows that there is an increasing trend in the median home prices starting from 1996 till around 2006.
- From 2006 till 2012, there has been a decreasing trend in home prices.
- From 2012, there has been a steady trend till 2017.
- The data, however, doesn't appear to show any seasonal variation.

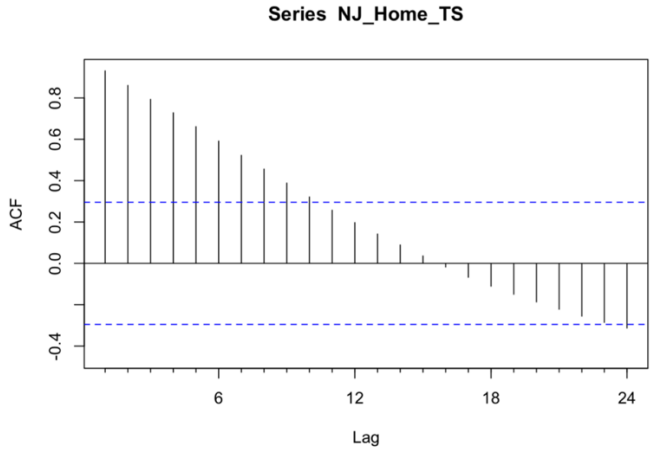
- If we were to forecast the data, we should be considering the window from 2014.



- The window function has been used since 2014 to forecast the data better.
- If we consider the whole data, we might have a different forecast.
- From 2014 it will be more than 3 years of data that we are considering, which should be good enough to be considered for forecasting.

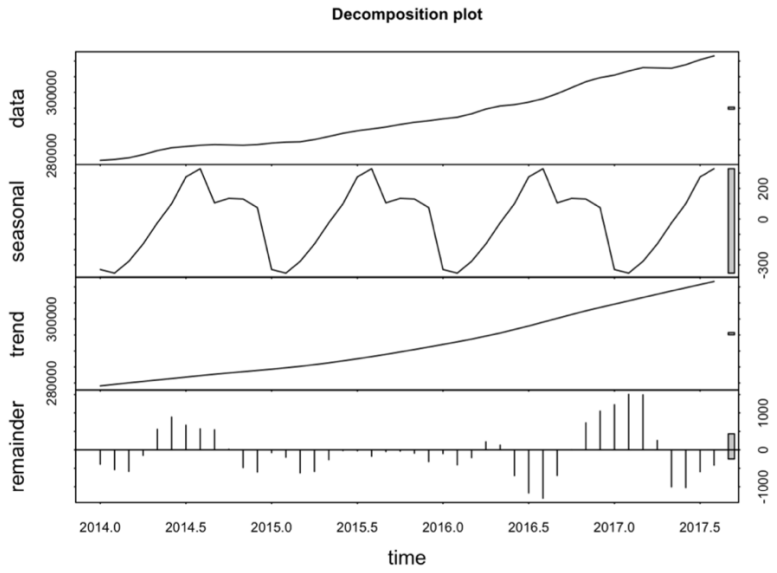






- The boxplot shows that there are no outliers in the data.
- The Median is more towards the first quartile.
- This means that the data is right-skewed. This can be justified by seeing the histogram above as well.
- The ACF plot shows that many of the values crossed the confidence intervals, stating there is a trend component in the data.
- Also, we can see that after the 15th lag period, the ACF plot is dipping into the negative values stating that seasonality also exists in the data.

3. Decomposition



- The decomposition plot shows a trend and seasonality in the time series data.
- Based on this analysis, we can develop a question and hypothesis to start our forecast.

4. Question and Hypothesis

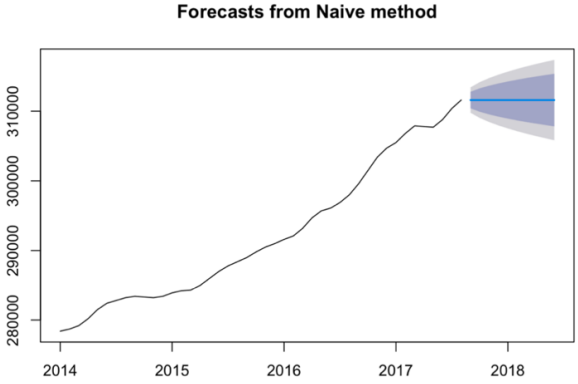
Question

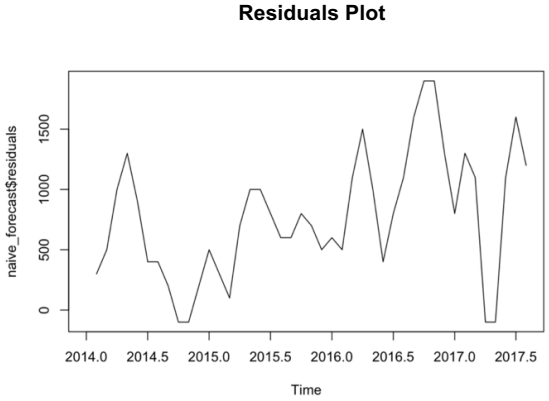
What will be the best method to forecast the given time series data?

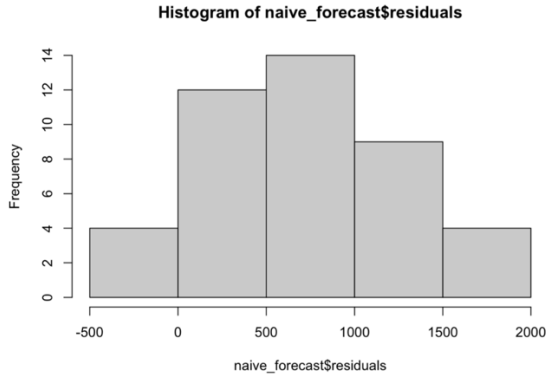
Hypothesis

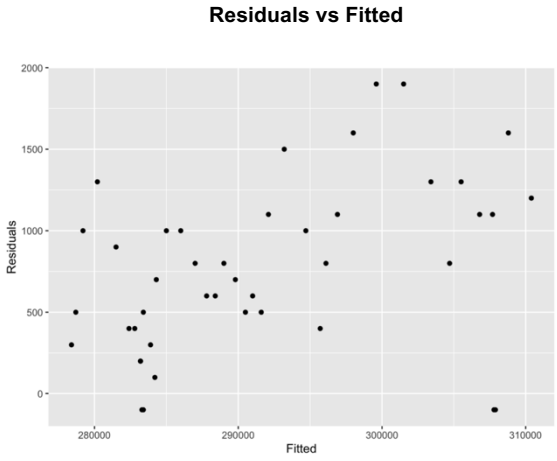
As we know, House pricing increases with time. It can be a time series with an increasing trend. HoltWinters can be the best method to forecast this type of data.
We can check this hypothesis based on the accuracy of each model that we can check below.

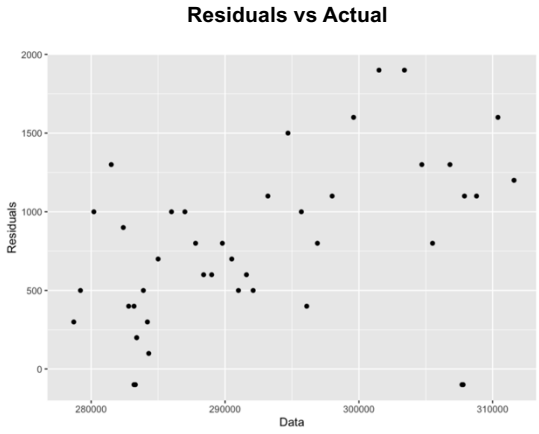
5. Forecast - Naïve Method

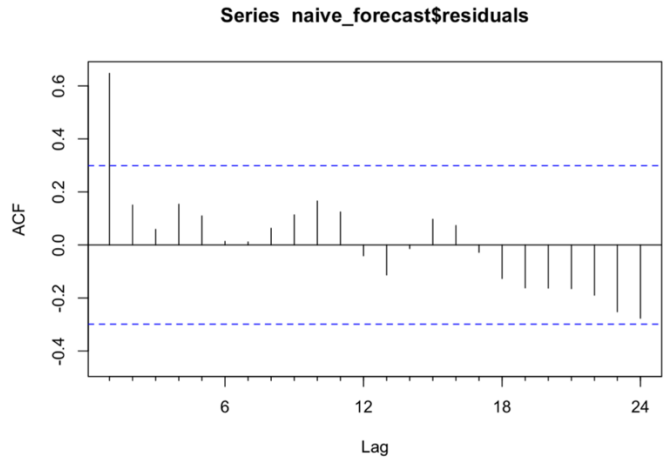












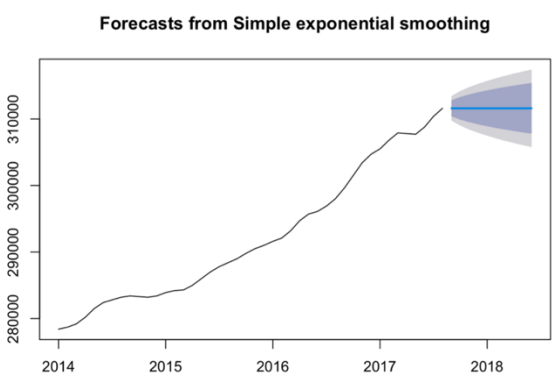
Accuracy

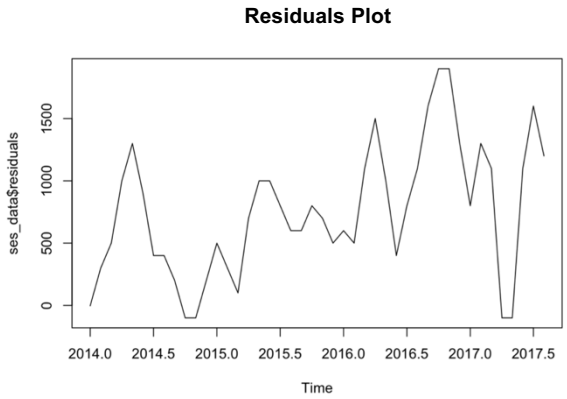
```
accuracy(naive_forecast)
```

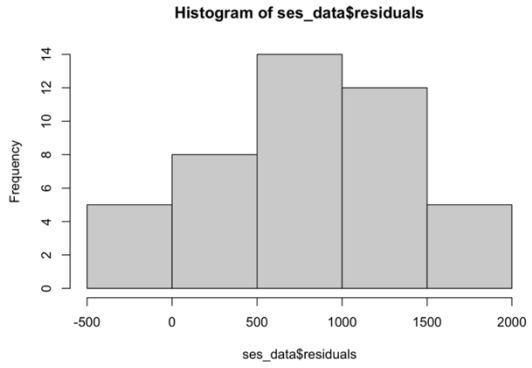
```
##                    ME     RMSE      MAE       MPE      MAPE       MASE      ACF1
## Training set 772.093 929.6161 790.6977 0.2615133 0.2678201 0.08548083 0.6470755
```

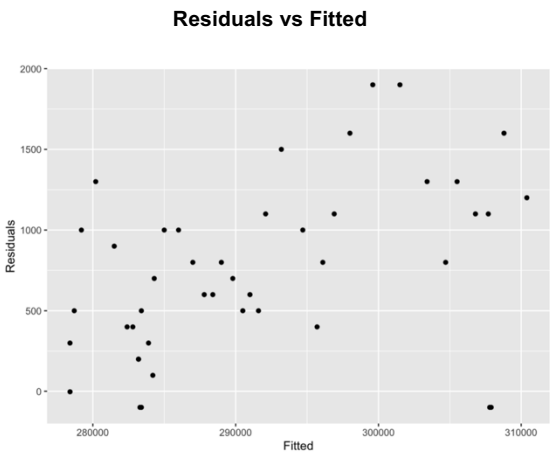
- The current forecasting method shows significant issues: positive residuals with increasing values, biased towards overestimation, and skewed residuals, indicating a need for more accuracy.
- The model exhibits heteroscedasticity, with varying variance in the errors, and misses essential variables and seasonality, leading to incorrect forecasts.
- High Mean Error (ME) and Root Mean Squared Error (RMSE) values highlight the poor performance of the current method, making it necessary to consider alternative forecasting approaches.

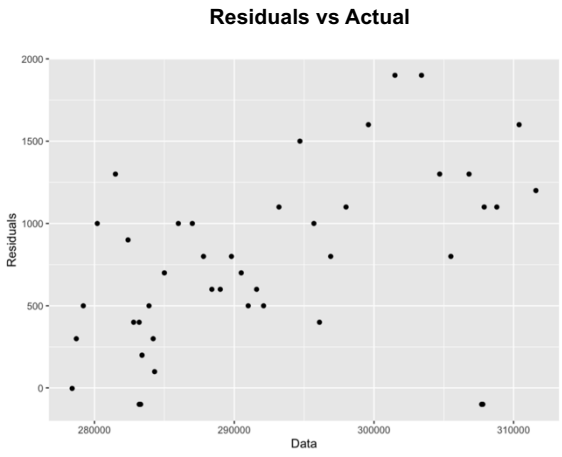
6. Forecast - Simple Smoothing

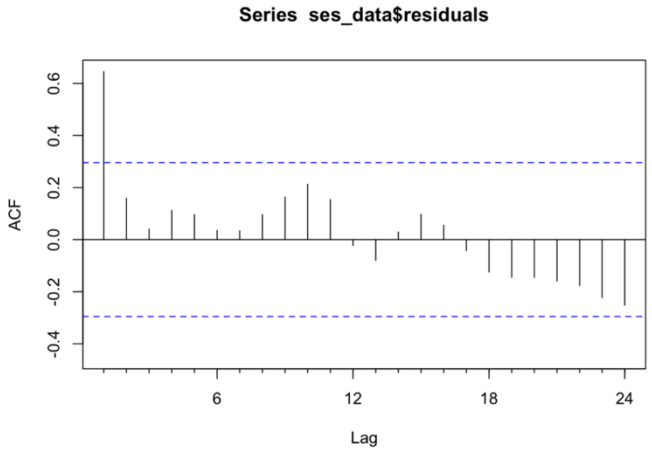












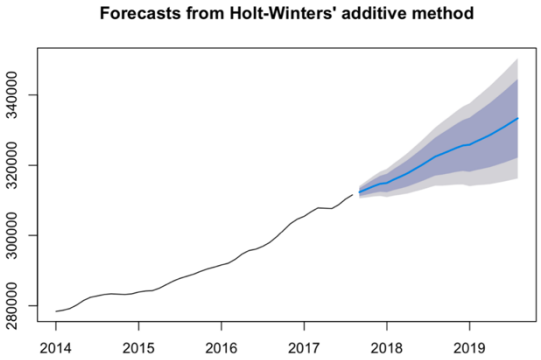
Accuracy

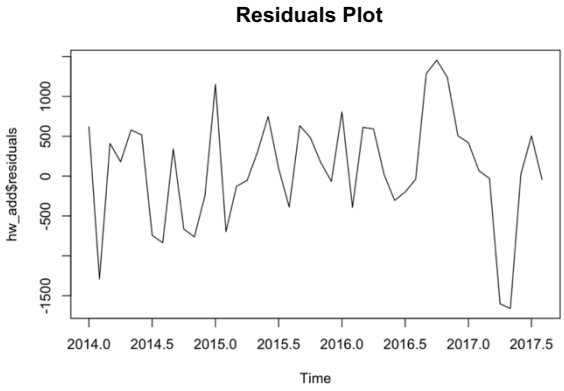
```
accuracy(ses_data)
```

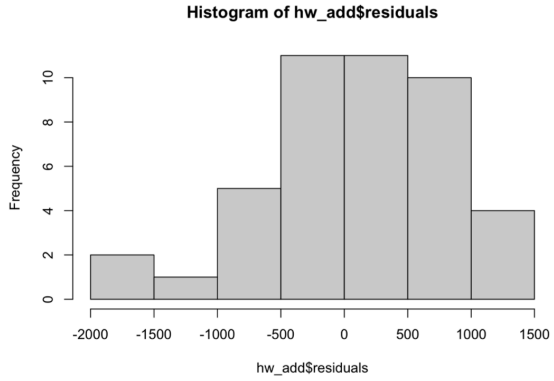
```
##                     ME     RMSE      MAE       MPE      MAPE       MASE      ACF1
## Training set 754.5426 919.0724 772.871 0.2555673 0.2617836 0.08355362 0.6452862
```

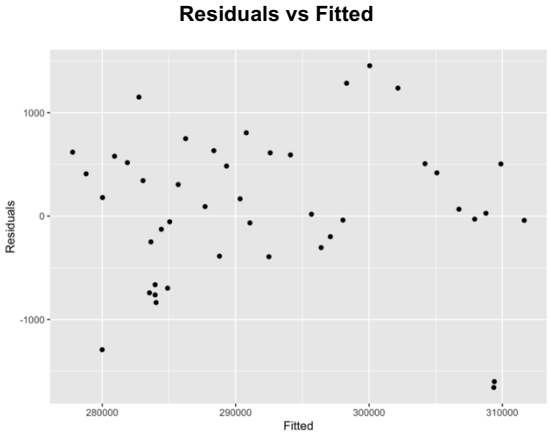
- The residuals show an increasing positive trend until the third quarter of 2016, followed by a dip. Most residuals are positive, indicating a bias towards overestimation. The histogram appears skewed, not having a mean of zero, suggesting biased data.
- The Fitted vs. Residuals plot and Actual vs. Residuals plot both exhibit trends, indicating heteroscedasticity, where the variance of residuals is not constant over time. The ACF values crossing the confidence level suggest missing variables and seasonality in the forecast.
- The high Mean Error (ME) and Root Mean Squared Error (RMSE) values indicate poor forecasting accuracy. Alternative techniques like Holt-Winters, suitable for trend + seasonal time series, may provide better results.
- In summary, the current forecasting method has several issues, including bias, missing variables, and incorrect modelling of trends and seasonality. Exploring more appropriate forecasting techniques like Holt-Winters is recommended to improve accuracy.

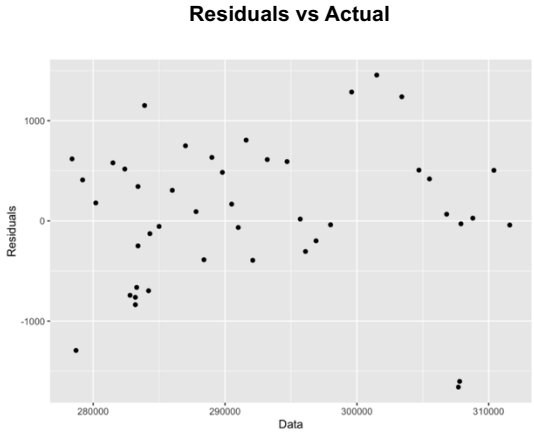
7. Forecast – Holt-Winters

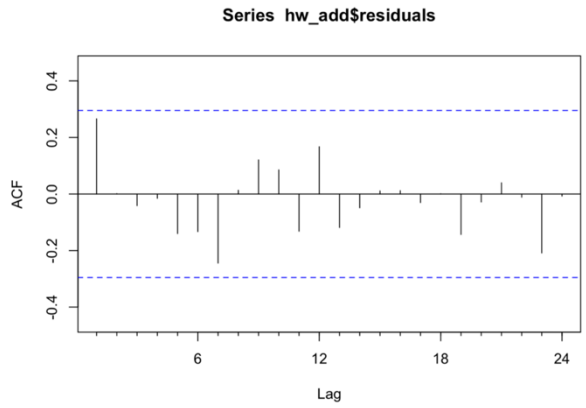












Accuracy

```
accuracy(hw_add)
```

```
##                     ME      RMSE       MAE         MPE      MAPE       MASE
## Training set 82.11107 700.8931 543.4792 0.02774417 0.1859478 0.05875451
##                    ACF1
## Training set 0.2657767
```

- The residuals appear random, and the mean is near zero, as confirmed by the normally distributed histogram. This indicates that the data is not biased, making this forecast the best among the previous attempts.
- The Fitted vs Residuals plot shows no trend, suggesting no heteroscedasticity in the errors, and the Actual vs Residuals plot also appears random, further supporting the validity of the forecast.
- The ACF plot shows no values crossing the confidence levels, indicating white noise and good forecast performance.
- The ME and RMSE values are low compared to previous forecasts, reinforcing the superiority of the Holt-Winters method over naive and simple smoothing.
- Holt-Winters appears to be the best forecast among the previous methods tested, but there is room for improvement by trying ARIMA models for further refinement.

8. Conclusion

Based on the analysis and comparison of the three forecasting methods (naive, simple smoothing, and Holt-Winters), it has been determined that Holt-Winters provides the best forecast for the given data, considering the presence of trend and seasonality.

The reasons supporting Holt-Winters as the better forecasting method are as follows:

- Holt-Winters effectively captures the trend and seasonality in the data, as confirmed by the ACF analysis.
- The Holt-Winters forecast fits well with the actual data, resulting in low error values compared to the other methods.
- The residuals in the Holt-Winters forecast appear random, indicating that the model captures the underlying patterns well.
- All ACF values of the residuals fall within the confidence interval, signifying the model's accuracy.

Based on the forecast, the time series is expected to exhibit an increasing trend over the next year and the next two years.

Overall, the analysis and the selected Holt-Winters forecast support the hypothesis that the time series has a discernible trend and seasonality. Holt-Winters is the best-suited method for accurate predictions in this case.

9. Future Scope

- ARIMA (AutoRegressive Integrated Moving Average) models have effectively captured complex time series patterns. As a future scope, ARIMA can be explored for more advanced forecasting tasks, mainly when the data exhibits non-linear trends or more intricate seasonality patterns.