Forecasting Electricity Supply and Distribution in South Africa

Research Assignment

Pako Kesiamang
BScHons Applied Mathematics
Faculty of Economic and Management Sciences
Stellenbosch University

Supervisor: Prof. H Nel
Date: 11 November 2024

Chapter 1

Summary

The aim of this report is to forecast a time series data set, and interpret which methods results yielded a better results. The data we're using covers the period from January 2002 to August 2024 and was sourced from Stats SA. This dataset reflects current monthly electricity figures but hasn't been adjusted to account for seasonal patterns.

To make accurate forecasts, we used several methods: the Holt-Winters additive and multiplicative models, along with the ARIMA model. These were chosen because they're well-suited for time series data that has both seasonal patterns and trends, which are common in electricity consumption. To figure out which method performs best, we measured accuracy with metrics like root mean squared error (RMSE) and mean absolute percentage error (MAPE).

When we plotted the data, it showed clear seasonal patterns and a trend, which led us to choose methods designed to capture these elements. The results indicated that the Holt-Winters methods were particularly effective at forecasting electricity distribution, especially in handling the seasonality within the data. While the ARIMA model also did well, it didn't quite match the accuracy of the Holt-Winters approach in capturing seasonal changes.

Chapter 2

Introduction

Electricity distribution in South Africa is a time series data that shows seasonal and trend-based patterns. For instance, electricity demand typically increases during colder winter months due to higher heating requirements, while summer months may see reduced demand. Additionally, load shedding—a series of scheduled power cuts implemented by Eskom—affects electricity distribution depending on infrastructure stability and energy availability. Economic conditions also influence electricity consumption; during times of economic downturn or high inflation, demand may decline as businesses and households adjust their energy usage.

The assumptions made in this report are based on the nature of energy consumption and supply dynamics in South Africa. Specifically, it is assumed that seasonal trends, as well as upward or downward shifts in demand, are driven by temperature variations, economic fluctuations, and Eskom's load-shedding schedules. This report employs forecasting models that account for these seasonal patterns and trends, including the Holt-Winters methods, while excluding simpler models that lack seasonality adjustments.

Chapter 3

Background

In October 2007, South Africa experienced its major blackout (load shedding), marking the start of recurring issues known as load shedding. Since then, the country has experienced multiple blackouts. Eskom, South Africa's primary supplier, has struggled to meet demand several times, resulting in scheduled power cuts to prevent grid collapse [1]. Load shedding was supposed to be a temporary measure due to Eskom's inability to meet energy demand, but this situation continued for years and worsened significantly in recent years. In 2023, South Africa recorded 335 days of load shedding, making it the most challenging year regarding energy availability, with outages occurring for approximately 6,700 hours over the year [2].

Load shedding has significantly affected South Africa's economy, resulting in a loss of R12.60 for every kilowatt/hour of electricity that goes unprovided during this load shedding period [3]. Eskom has struggled to meet demand due to aging infrastructure and unexpected breakdowns, often having to implement stage 6 (which means most people will have their electricity turned off for six hours per day) load shedding to manage the grid.

In this report, we aim to forecast electricity available for distribution and the electricity that ends up being distributed in South Africa using historical data and various forecasting techniques. By examining the historical data and applying multiple forecasting techniques like ARIMA, exponential smoothing, and the Holt-Winters method, this report aims to examine future scenarios of Energy Available for Distribution and Energy Distributed, assessing the balance between generation and distribution, and evaluating how changes in energy distribution align with load shedding. The dataset used in this report is obtained from Stats SA, and it contains monthly data from January 2002 to August 2024.

Chapter 4

Methodology

The following section explains the methods that are used to forecast. To forecast energy availability and distribution in South Africa, we utilize time series analysis. The data in this paper will be graphically illustrated. All analyses were conducted in R.

4.1 Data

The dataset, obtained from Stats SA, includes monthly data on "Energy available for distribution" and "Energy Distributed" from January 2002 to August 2024. Figure 4.1 below displays the time series of both graphs. Figure 4.1 below has repeating fluctuations within each year, suggesting the data has a seasonal component and there seems to be an upward trend up until 2010 followed by a gradual decline, suggesting that there is a trend within the data. The data Figure exhibits a strong annual seasonal component. This is expected since electricity is influenced by external factors such as temperature and consumption behaviour.

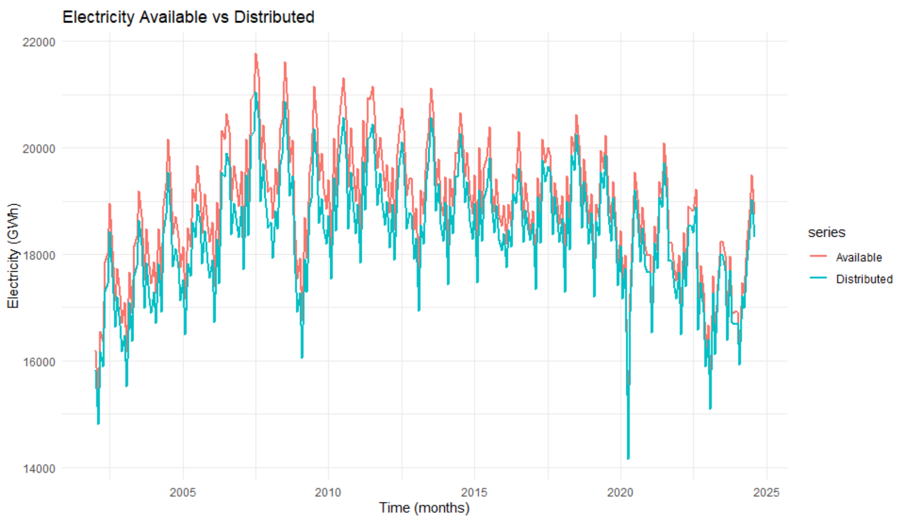

Figure 4.1: Time series of energy availability and distributed

4.2 Forecasting techniques

The section below explains the forecasting technique that will be used to provide assumptions made about the data. The forecasting includes:

4.2.1 Holt-Winters' Additive Method

The seasonal component is expressed in absolute terms in the scale of the observed series when using the additive method. The seasonal component is then subtracted from the level equation to seasonally adjust the series.
The component form for the additive method is:

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

The trend equation is identical to Holt's linear method. The seasonal equation shows a weighted average between the current seasonal index, $(y_t - \ell_{t-1} - b_{t-1})$, and the seasonal index of the same season last year (i.e., m time periods ago).
The equation for the seasonal component is often expressed as:

$$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m}$$

If we substitute ℓ_t from the smoothing equation for the level of the component form above, we get:

$$s_t = \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1-\gamma^*(1-\alpha)]s_{t-m}$$

which is identical to the smoothing equation for the seasonal component we specify here, with $\gamma = \gamma^*(1-\alpha)$. The usual parameter restriction is $0 \leq \gamma^* \leq 1$, which translates to $0 \leq \gamma \leq 1-\alpha$.

4.2.2 Holt-Winters' Multiplicative Method

Using the multiplicative method, the series is seasonally adjusted by dividing through by the seasonal component, which is given in relative terms (percentages). The seasonal component will add up to around m within each year. The multiplicative equation is:

$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma\frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}$$

4.2.3 ARIMA

Auto-regressive integrated moving average is referred to as ARIMA. To put it briefly, it is a common statistical model for analyzing and forecasting time series. These models are often used interchangeably and are a type of Box-Jenkins model. The authors, Box and Jenkins, provide a procedure for
finding, calculating, and verifying models for a given collection of time series data. An explanation of the ARIMA model and the procedure proposed by Box and Jenkins will be provided below. The three components of ARIMA are I (integrated), which uses raw observation differencing to create a stationary time series, MA (moving average), which uses the dependent relationship between an observation and several lag observations, and AR (autoregression), which uses the dependent relationship between an observation and a number of lag observations.

Chapter 5

Results and Discussion

This section presents the evaluation of each model for energy available for distribution and energy distributed. The methods used in this section include Holts-winter additive, Holts-winter multiplicative and seasonal ARIMA. The models were trained on the data from January 2002 to December 2021 and tested from January 2022 to August 2024.

5.1 Electricity Available for Distribution

5.1.1 Holts-winters additive

The Table 5.1 display the accuracy of the Holts-winter additive method. Based on the results, the model archived an RMSE of 365.2417 and MAPE of 1.32% for the train data set, indicating a lack of seasonal component. The Theil's U of 1.60 indicates that this model is worse than guessing.

Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-20.70718	365.2417	247.9353	-0.136788	1.327965	0.4822168	0.03824198	NA
Test set	-1384.64641	1601.1921	1390.7299	-7.936305	7.969956	2.7048718	0.80862600	1.602925

Table 5.1: Accuracy of Holts-winter additive

The Figure 5.1 illustrates the performance of the Holt-winter additive model, comparing the forecast values against the test data set and the forecast of 12 months more. This visualization helps assess how well the model performs compared to the Historical data. Based on the results in the figure, the forecast does not fit the training and testing data well by looking at the plot.

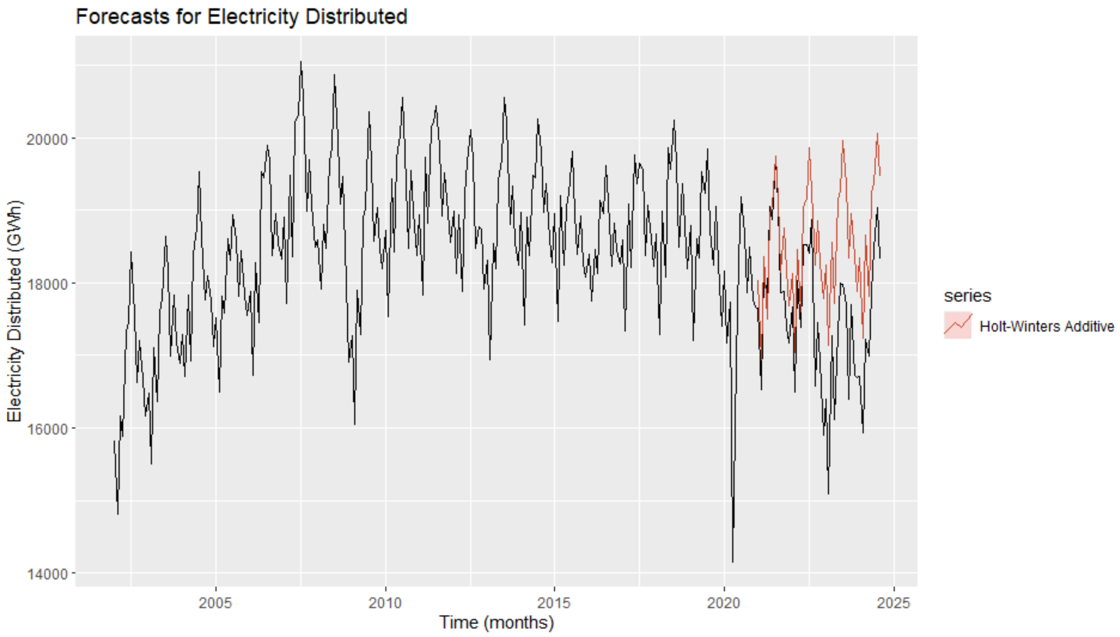

Figure 5.1: Performance of the Holt-Winter additive Model – Comparison of Forecasted Values with Training and Test Datasets for the Next 12 Months

5.1.2 Holts-winter multiplicative

The Table 5.2 below displays the accuracy of the Holts-winter multiplicative method. Based on the results in the table below, the model archived an RMSE of 382.76 and MAPE of 1.41% for the train data set, indicating strong performance in capturing seasonal trends. For the Test data set, the model has a higher RMSE of 867.50 and MAPE of 4.01%, indicating that the model does underperform on the new data set. The Theil's U of 0.87 indicates that this model provides relatively accurate forecasting results.

Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-15.93899	382.7590	264.2710	-0.1104901	1.412951	0.5139885	0.175963	NA
Test set	-645.41633	867.4983	692.5648	-3.7548687	4.007600	1.3469898	0.636466	0.8688902

Table 5.2: Accuracy of Holt-winter Multiplicative

Figure 5.2 below illustrates the performance of the Holt-winter multiplicative model, comparing the forecast values against the test data set and the forecast of 12 months more. This visualization helps assess how well the model performs compared to the Historical data. Based on the results of the figure, the model does indicate strong performance in capturing seasonal trends and it does provide accurate forecasting results when compared to the test data set.

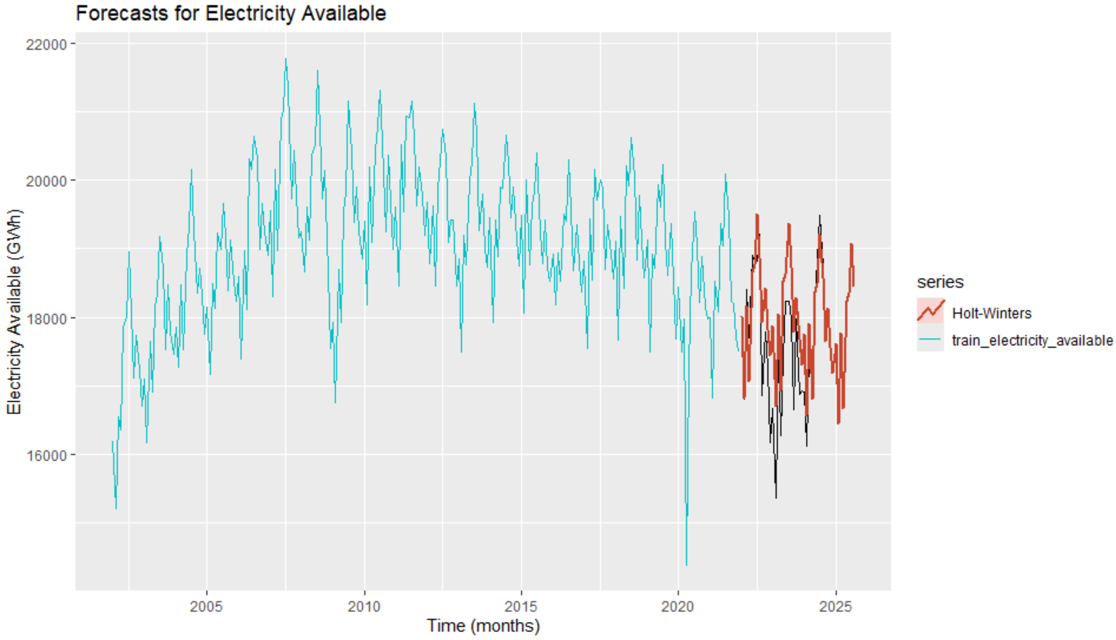

Figure 5.2: Performance of the Holt-Winter Multiplicative Model – Comparison of Forecasted Values with Training and Test Datasets for the Next 12 Months

5.1.3 SARIMA

Table 5.3 below displays the accuracy of the seasonal ARIMA method. The model archived an RMSE of 365.01, indicating a strong fit and MAPE of 1.41% for the train data set, which also indicates strong performance in capturing seasonal trends. For the Test data set, the model yields a higher RMSE of 902.69 and MAPE of 4.22%, indicating that the model does underperform on the new data set. The Theil's U of 0.91 indicates that this model still provided some reliable predictions.

Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-39.2367	365.0109	248.9927	-0.2379028	1.327216	0.4842733	0.0004737419	NA
Test set	-656.5017	902.6899	731.0625	-3.8253814	4.223565	1.4218652	0.6860115225	0.9059363

Table 5.3: Accuracy forecast of Arima

Figure 5.3 illustrates the performance of the seasonal ARIMA, comparing the forecast values against the test data set and the forecast of 12 months more. This visualization helps assess how well the model performs compared to the Historical data. Based on the plot, the forecast seems to fit the training and testing data well by looking at the plot.

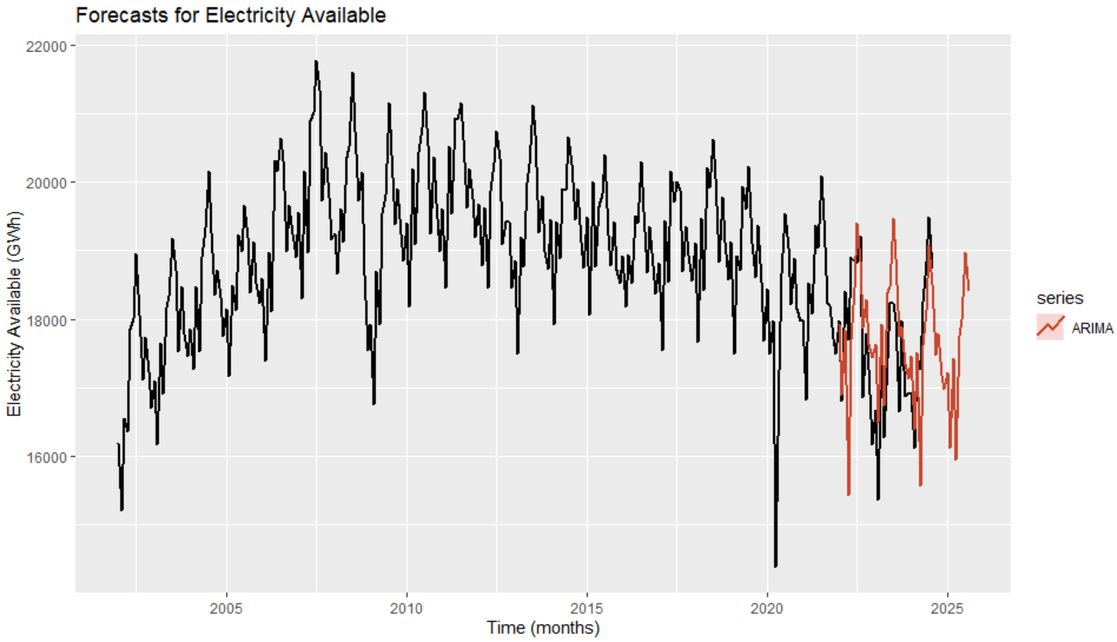

Figure 5.3: Performance of the seasonal ARIMA of Forecasted Values with Training and Test Datasets for the Next 12 Months

5.2 Electricity distributed

5.2.1 Holts-winters additive

Table 5.4 displays the accuracy of the Holts-winter additive method. Based on the results, the model archived an RMSE of 354.1797 and MAPE of 1.310185% for the train data set, indicating a lack of seasonal component. The Theil's U of 1.214276 indicates that this model is worse than guessing.

Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-28.5438	354.1797	237.7054	-0.180510	1.310185	0.4800468	0.03629764	NA
Test set	-991.7332	1180.2322	1007.0335	-5.795875	5.880081	2.0337073	0.75034508	1.214276

Table 5.4: Accuracy of Holts-winters additive method

Figure 5.4 illustrates the performance of the Holt-winter additive model, comparing the forecast values against the test data set and the forecast of 12 months more. This visualization helps assess how well the model performs compared to the Historical data. Based on the results in the figure, the forecast does not fit the training and testing data well by looking at the plot.

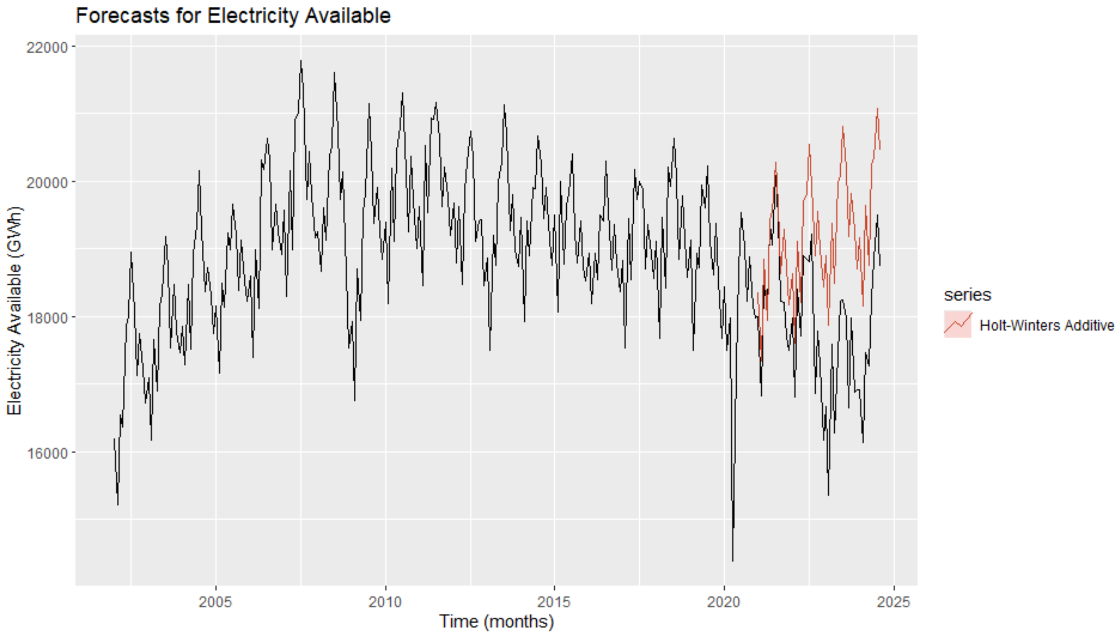

Figure 5.4: Performance of the Holt-Winter additive Model – Comparison of Forecasted Values with Training and Test Datasets for the Next 12 Months.

5.2.2 Holt-winter multiplicative

The Table 5.5 below displays the accuracy of the Holt-winter multiplicative method. Based on the results in the table below, the model archived an RMSE of 369.74 and a low MAPE of 1.34% for the train data set, indicating robust performance in capturing seasonal trends. For the Test data set, the model has a higher RMSE of 1061.33 and MAPE of 5.23%, indicating that the model does underperform on the new data set. The Theil's U of 1.09 indicates that the model has a higher degree of error. Figure 5.5 below illustrates the

Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-25.93835	369.7497	243.7145	-0.1718902	1.344252	0.4921822	0.04708855	NA
Test set	-880.56571	1061.3365	894.4520	-5.1520190	5.227856	1.8063486	0.72385017	1.092036

Table 5.5: Accuracy of Holts-winter multiplicative method

performance of the Holt-winter multiplicative model, comparing the forecast values against the test data set and the forecast of 12 months more. This visualization helps assess how well the model performs compared to the Historical data. Based on the results of the figure, the forecasting does not fit the historical data.

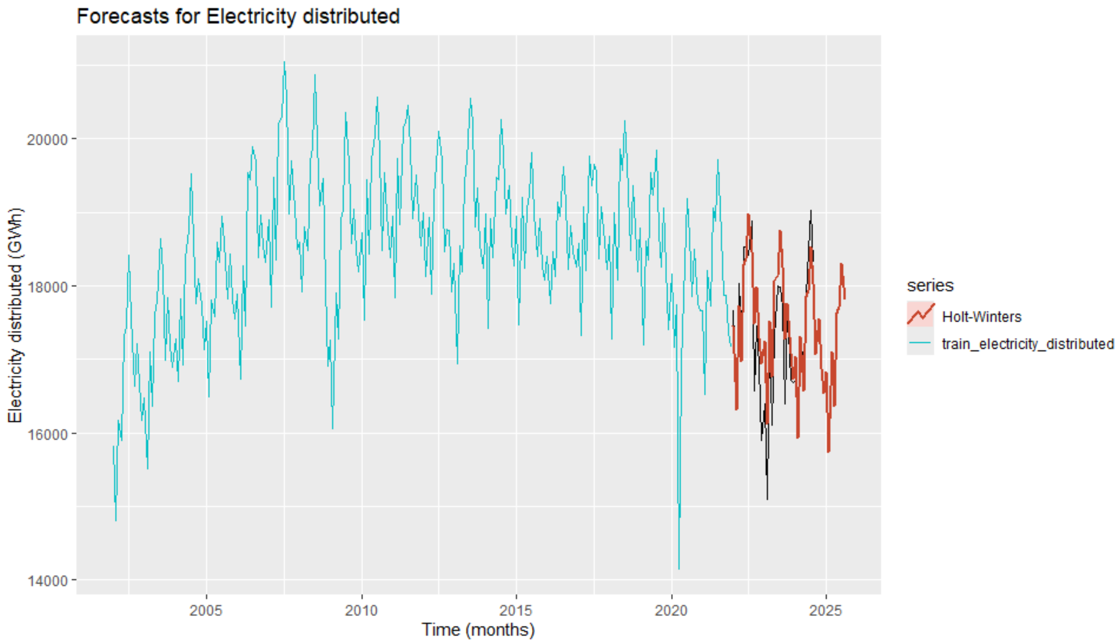

Figure 5.5: Performance of the Holt-Winter Multiplicative Model – Comparison of Forecasted Values with Training and Test Datasets for the Next 12 Months.

5.2.3 SARIMA

The Table 5.6 below displays the accuracy for the seasonal ARIMA method. The model archived an RMSE of 354.24, indicating a strong fit and MAPE of 1.31% for the train data set, which also indicates a strong fit. For the Test data set, the model yields a higher RMSE of 853.94 and MAPE of 4.09%, indicating slight underperformance on the new data set. The Theil's U of 0.87 indicates that the model provides some reliable prediction. Figure 5.6

Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-33.42269	354.2442	239.4832	-0.2119381	1.314343	0.483637	0.006420608	NA
Test set	-617.60612	853.9438	697.7694	-3.6561684	4.091000	1.409147	0.691158368	0.8796977

Table 5.6: Accuracy of ARIMA for Electricity distributed

illustrates the performance of the seasonal ARIMA, comparing the forecast values against the test data set and the forecast of 12 months more. This visualization helps assess how well the model performs compared to the Historical data. Based on the plot, the forecast seems to fit the training and testing data well by looking at the plot.

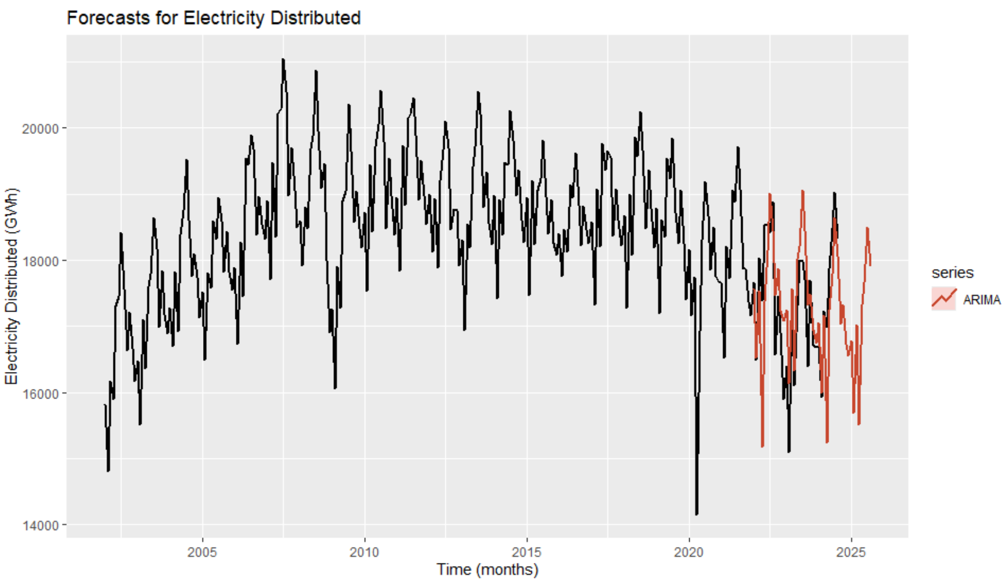

Figure 5.6: Performance of the seasonal ARIMA Model – Comparison of Forecasted Values with Training and Test Datasets for the Next 12 Months

Chapter 6

Comparision

In this chapter, we compare the performance of all model. We compare the model using key metric for both training data set and test data set. Table 6.1 summarise the key metric of each model for the toatl Electricity distributed and Electricity available in South Africa.

Model	Dataset	ME	RMSE	MAE	MPE (%)	MAPE (%)	MASE	ACF1	Theil's U
Electricity Available for Distribution									
Holt-Winters Multiplicative Model	Training Set	-15.94	382.76	264.27	-0.11	1.41	0.51	0.18	NA
	Test Set	-645.42	867.50	692.56	-3.75	4.01	1.35	0.64	0.87
SARIMA Model	Training Set	-39.24	365.01	248.99	-0.24	1.33	0.48	0.00	NA
	Test Set	-656.50	902.69	731.06	-3.83	4.22	1.42	0.69	0.91
Holt's Additive Model	Training Set	-20.71	365.24	247.94	-0.14	1.33	0.48	0.04	NA
	Test Set	-1384.65	1601.19	1390.73	-7.94	7.97	2.70	0.81	1.60
Electricity Distributed									
Holt-Winters Multiplicative Model	Training Set	-25.94	369.75	243.71	-0.17	1.34	0.49	0.05	NA
	Test Set	-880.57	1061.34	894.45	-5.15	5.23	1.81	0.72	1.09
SARIMA Model	Training Set	-33.42	354.24	239.48	-0.21	1.31	0.48	0.01	NA
	Test Set	-617.61	853.94	697.77	-3.66	4.09	1.41	0.69	0.88
Holt's Additive Model	Training Set	-28.54	354.18	237.71	-0.18	1.31	0.48	0.04	NA
	Test Set	-991.73	1180.23	1007.03	-5.80	5.88	2.03	0.75	1.21

Table 6.1: Comparison of Model Accuracy Metrics for Electricity Availability and Distribution Data

Based on Table 6.1 it is evident that, for Electricity Available for distribution Holt-winter Multiplicative method performed better with lower RMSE, MAE, and MAPE values compared to seasonal ARIMA and additive, followed by seasonal ARIMA. For Visualization, refer to the Figure 6.1. The Figure 6.1 illustrate the prediction of each model compared to the actual testing data set. The plots revels that the Holt-winter multiplicative prediction follows closely to the actual data trend and followed by the SARIMA.
For Electricity distributed, it is evident that, SARIMA method better compared to other models. It has lower values of RMSE, MAE, and MAPE. The Figure 6.2 shows that SARIMA align closely with the actual data patterns, followed by Holt-winter multiplicative.

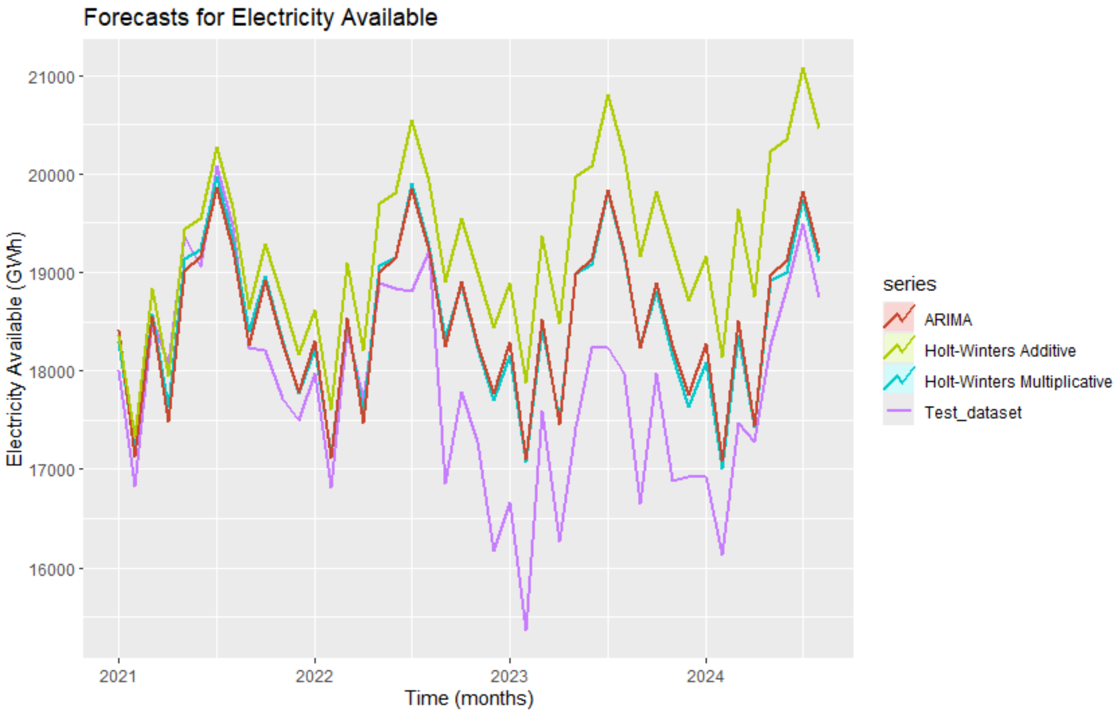

Figure 6.1: Comparison of Model Fits on Training Data: Forecasting Electricity Available using Holt-Winters Multiplicative, SARIMA, and Holt's Additive Models

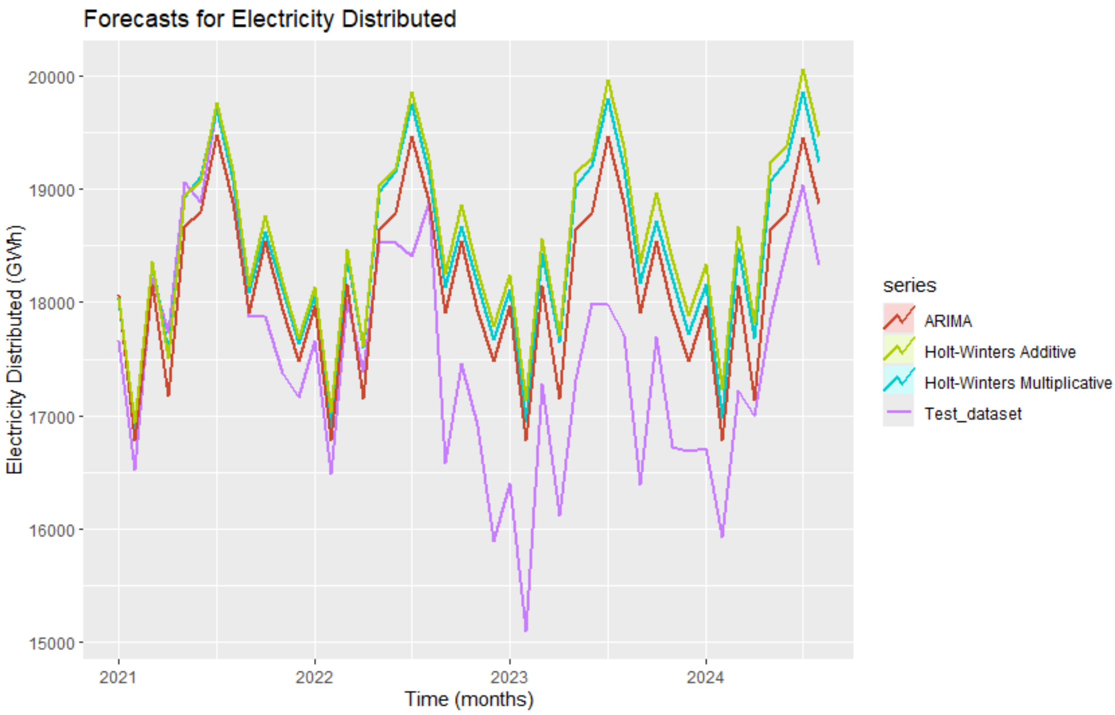

Figure 6.2: Comparison of Model Fits on Training Data: Forecasting Electricity distributed using Holt-Winters Multiplicative, SARIMA, and Holt's Additive Models

Chapter 7

Conclusion

In conclusion, the Holt-Winter multiplicative method was found to be the most effective for forecasting the total energy available for distribution in South Africa, aligning with Pagel's classification of the data. For forecasting electricity distribution, the SARIMA model proved to be the most suitable technique, also aligning with Pagel's classification. While SARIMA performed well on both datasets, the multiplicative method demonstrated higher efficiency for total energy availability. Therefore, for accurately forecasting both energy availability and distribution, using the Holt-Winter method for the energy available data and SARIMA for the electricity distribution data is recommended.

Bibliography

[1] Energy Central. (2019). South Africa's Load Shedding Crisis. Retrieved from https://energychamber.org/understanding-south-africas-energy-crisis/?form=MG0AV3

[2] Stats SA. (2023). Annual Load Shedding Data. Retrieved from https://www.statssa.gov.za/?s=load+sheddingform=MG0AV3

[3] Mkhwanazi, T. (2023). Economic Impact of Load Shedding in South Africa. Retrieved from https://www.statssa.gov.za/?s=load+sheddingform=MG0AV3