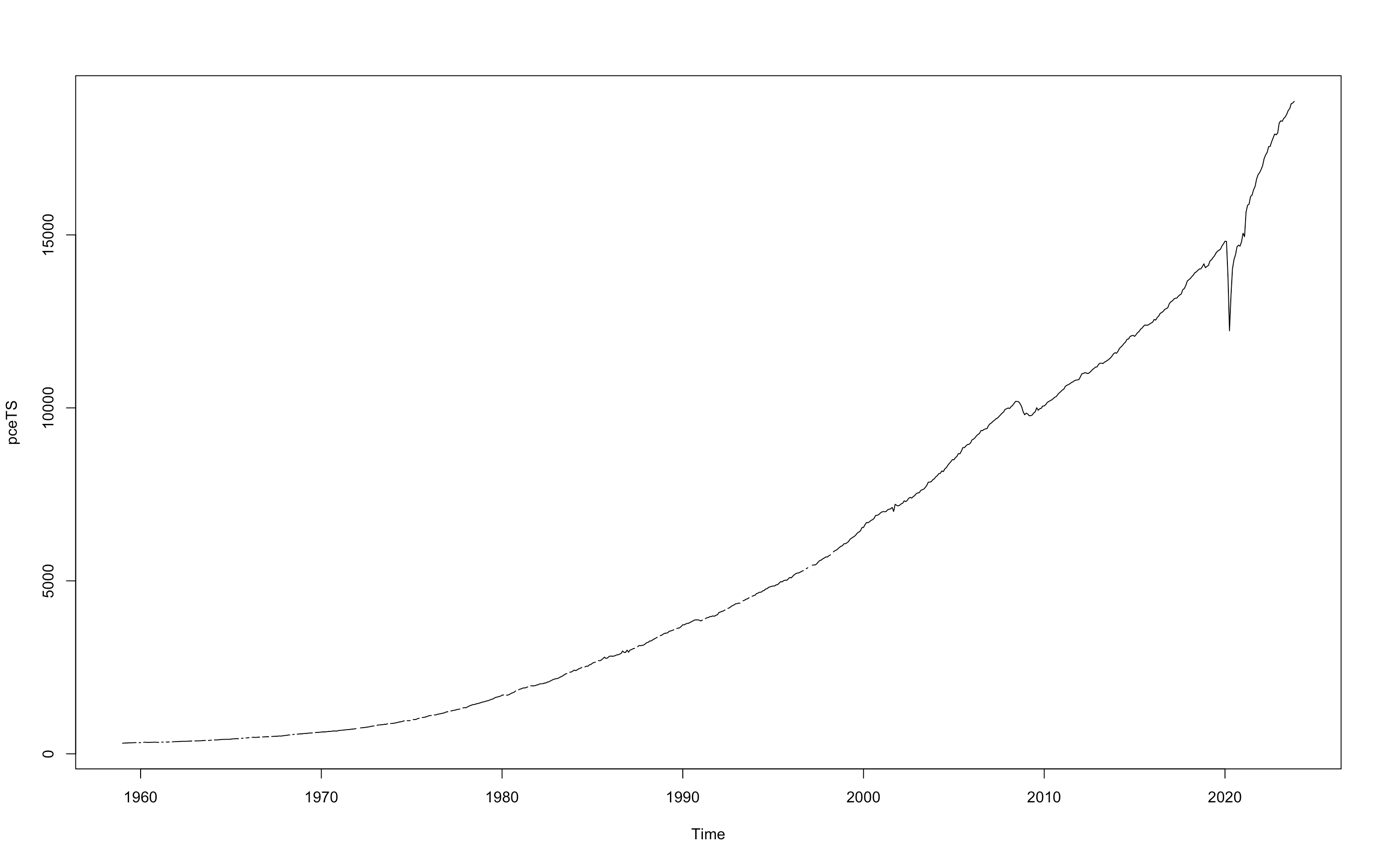
**Time Series Analysis of US personal Consumption Expenditure data**

This report aims to provide a comparison of the predictive capacity of three models including a simple forecasting method, an exponential smoothing model and an Arima model on the US seasonally-adjusted personal consumption expenditures. The objectives are to identify the best model and estimate the personal consumption expenditure for October 2024. The report then proceeds to evaluate the best model for one-step ahead rolling forecasting without re-estimation of the parameters. The following sections of the report delve into the steps involved in processing and analysing the data, selecting the best model and the criteria for selection, estimating expenditure for October of 2024 using the best model, and comparing the models’ predictive performance using one-step rolling ahead forecasting.

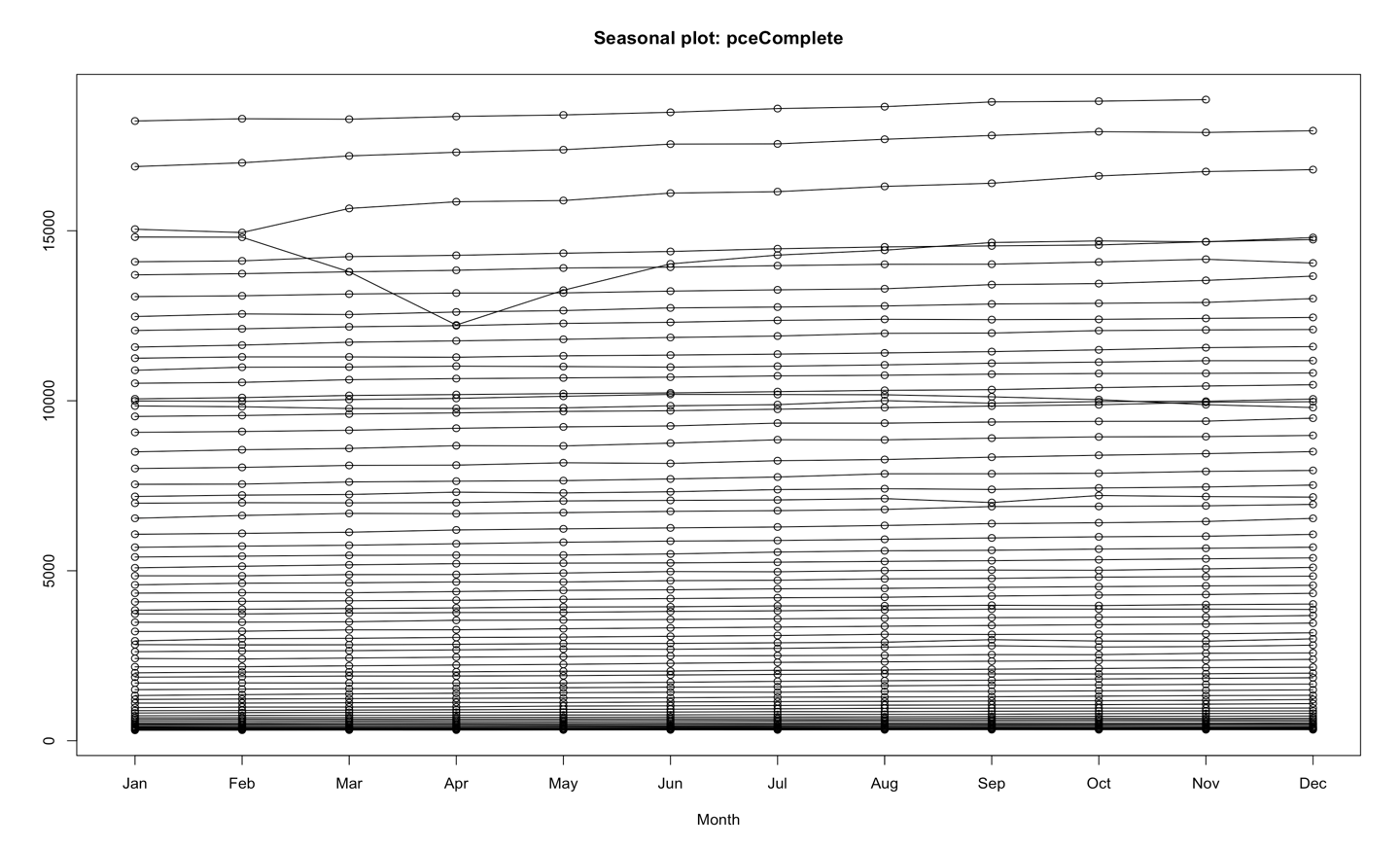
1. **Data Pre-processing:**

The dataset provided has two columns containing the date and the personal consumption expenditure from January 1959 to November 2023. The data is converted to a time series object with a frequency of 12 as the observations are monthly. The data is then checked for any missing observations and imputation is performed. In this case, missing data is handled using a simple linear interpolation method since the data has a notable trend and continuity as observed in Fig-1.1. Interpolation evaluates the missing value by averaging the values on either side of the missing value.

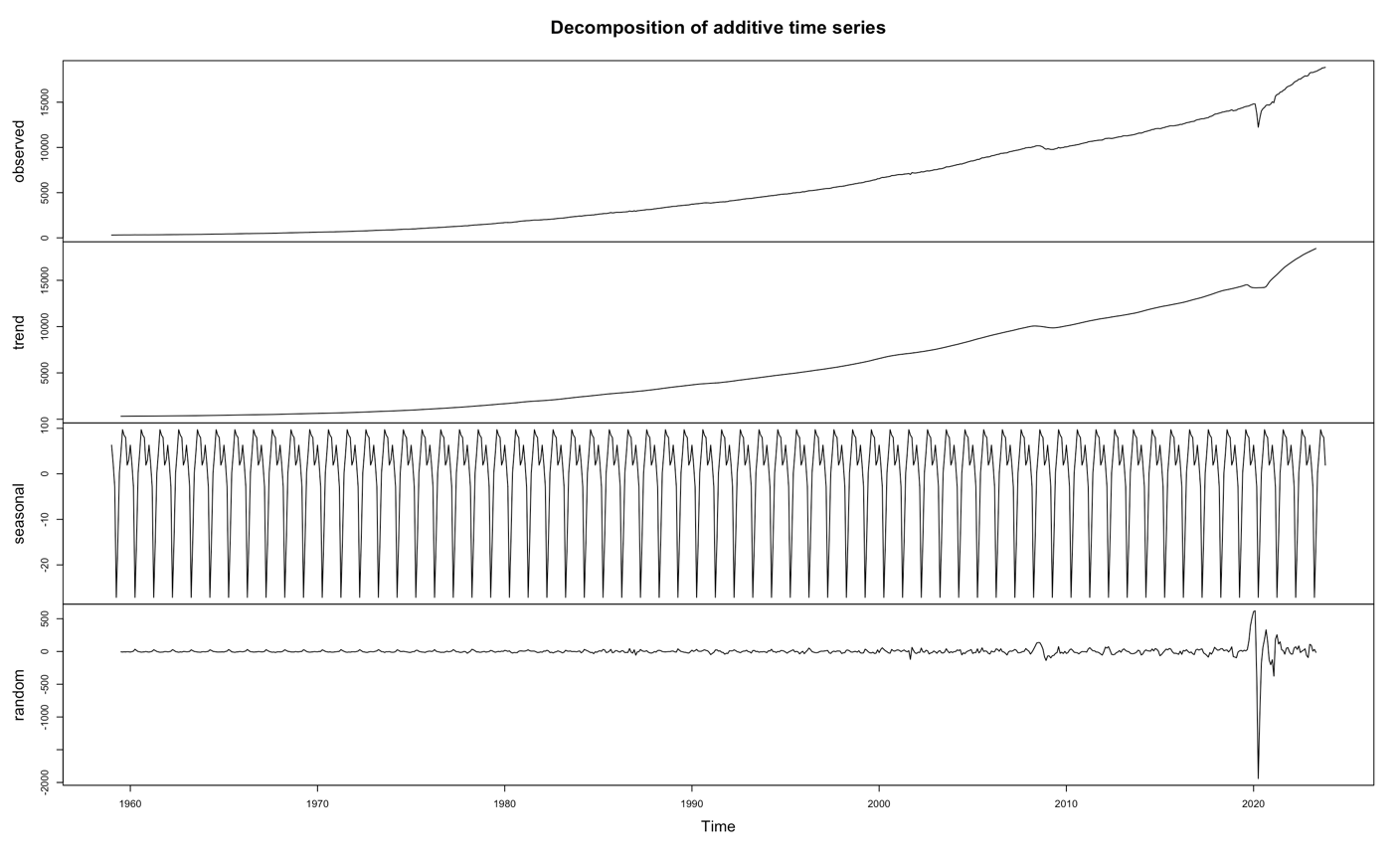


**Fig-1.1 Plot of the original time series before imputation**

The seasonally adjusted data is verified for the absence of seasonality using a season plot. From Fig-1.2, it is clear that there are no calendar-related fluctuations in the time series. The data is further decomposed to identify the trends and irregularities. From Fig-1.3, the expenditure has seen a constant and significantly rising trend and minimal irregularities.



**Fig-1.2 Season plot for the Time series**

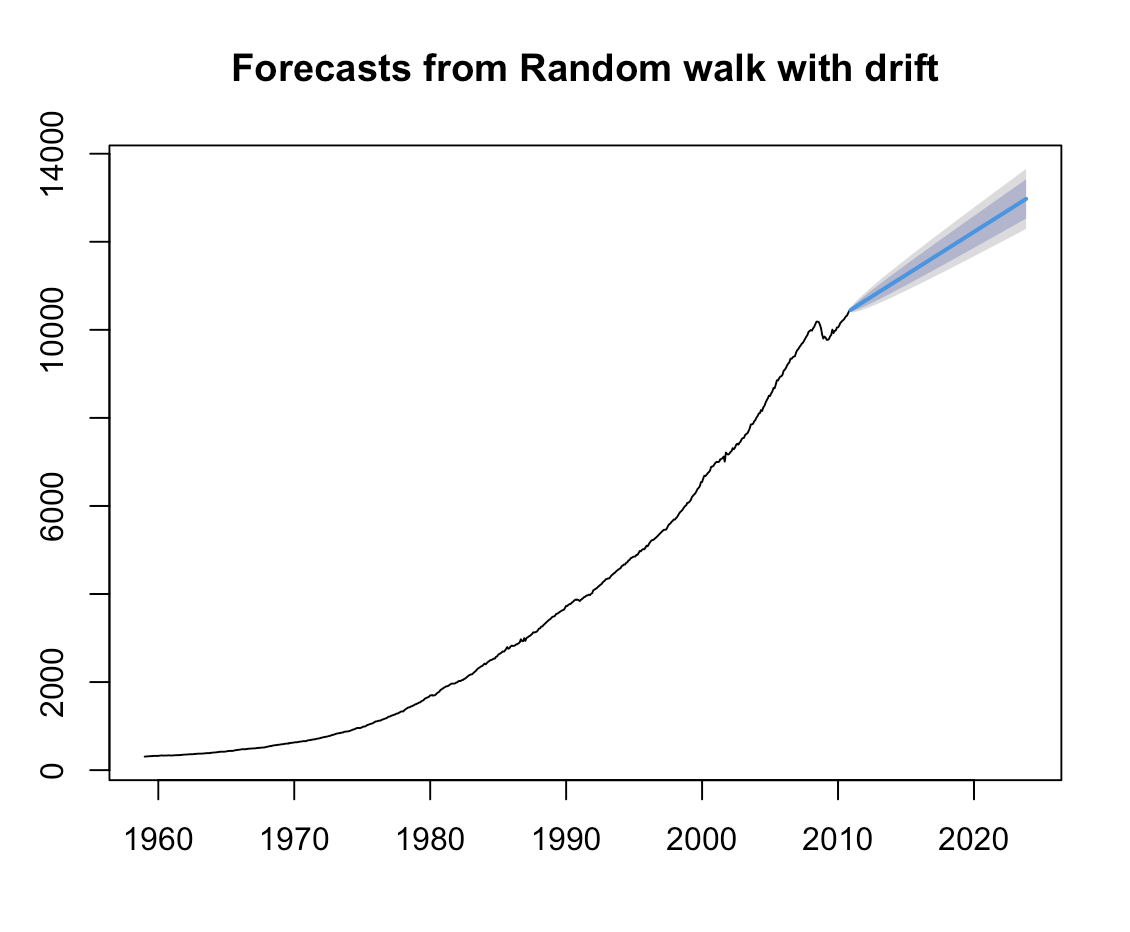


**Fig-1.3 Decomposed components of time series**

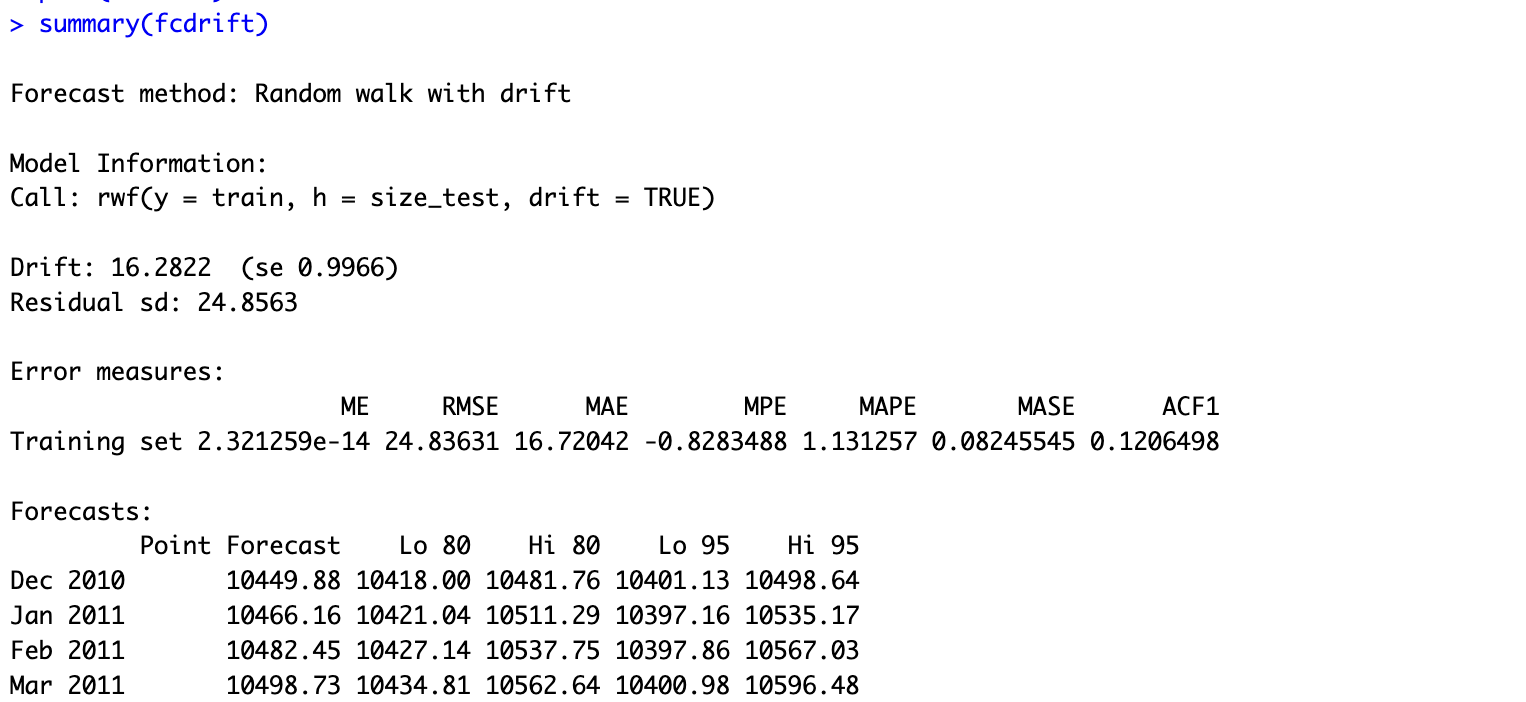
Further, the data is split to create train and test sets to train the forecasting models and evaluate their performance. There are a total of 779 records in the data out of which 80% is taken as the training data and the remaining 20% is taken as the test set.

1. **Model Evaluation**
2. **Simple forecasting methods**

Of the four simple forecasting methods, the drift model is used to forecast the data as the time series displays a stable trend with no seasonality and no repeating patterns with little irregularities making the drift model a more suitable choice. The model is trained using the training set and predicted for the test set to produce forecasts as in Fig-2.a.1. From the summary of the drift model in Fig-2.a.2, the error measures are quite low for the training set.

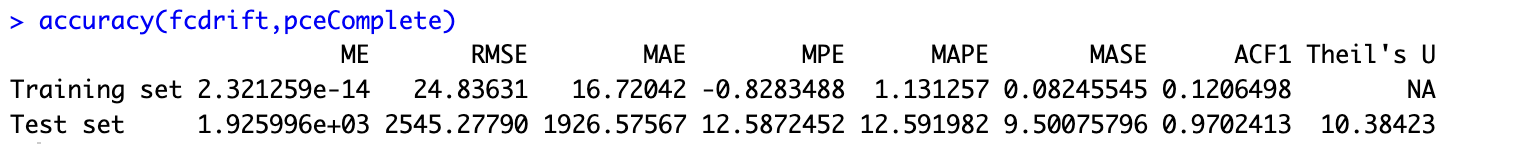


**Fig-2.a.1 Plot of forecasts from drift method.**



**Fig-2.a.2 Summary of the drift model**

The accuracy of the drift model is checked by comparing the forecast to the original time series. From Fig-2.a.3 It is observed that the RMSE is 2545.28 and MAE is 1926.58.

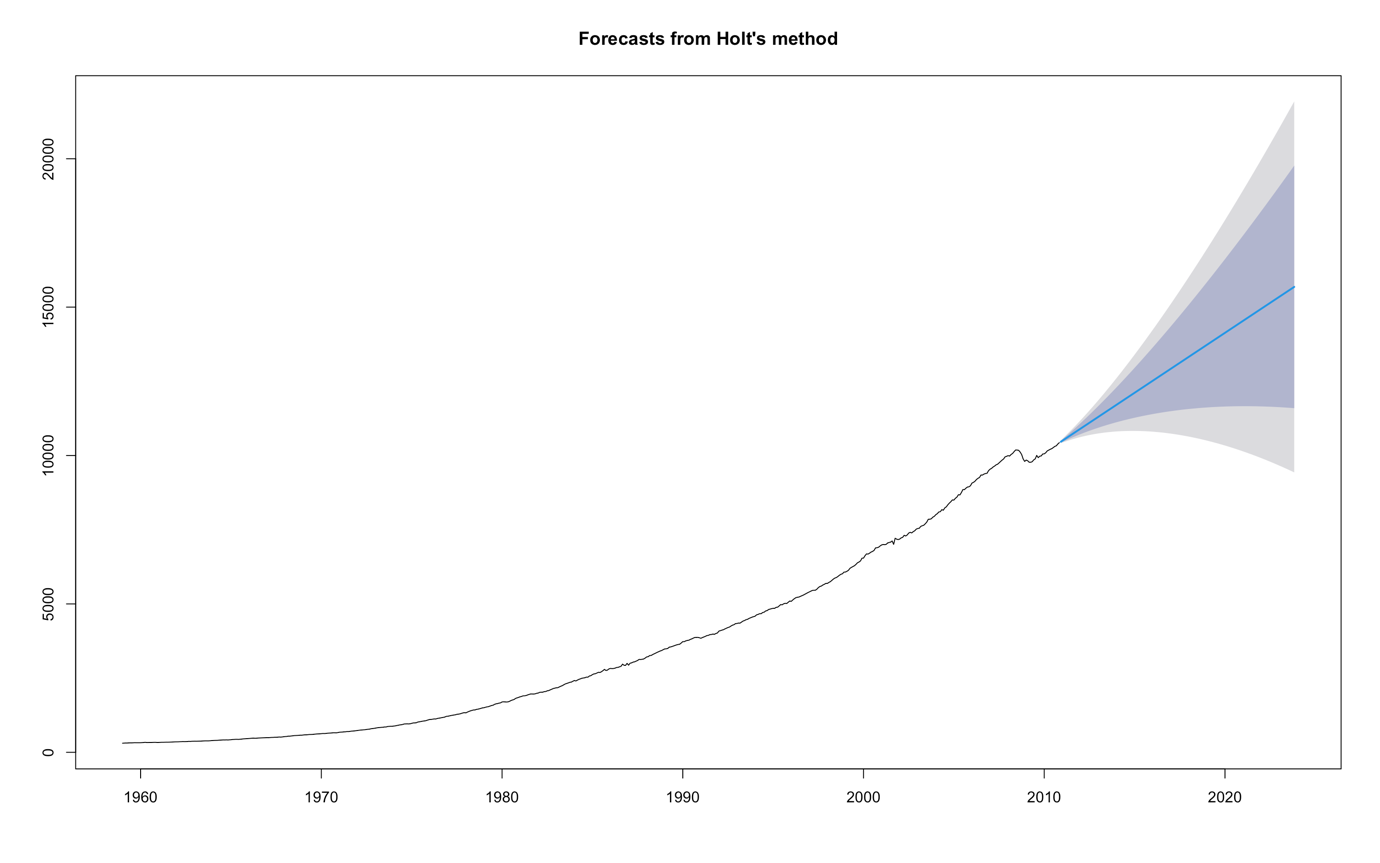


**Fig-2.a.3 Accuracy measure for the drift model**

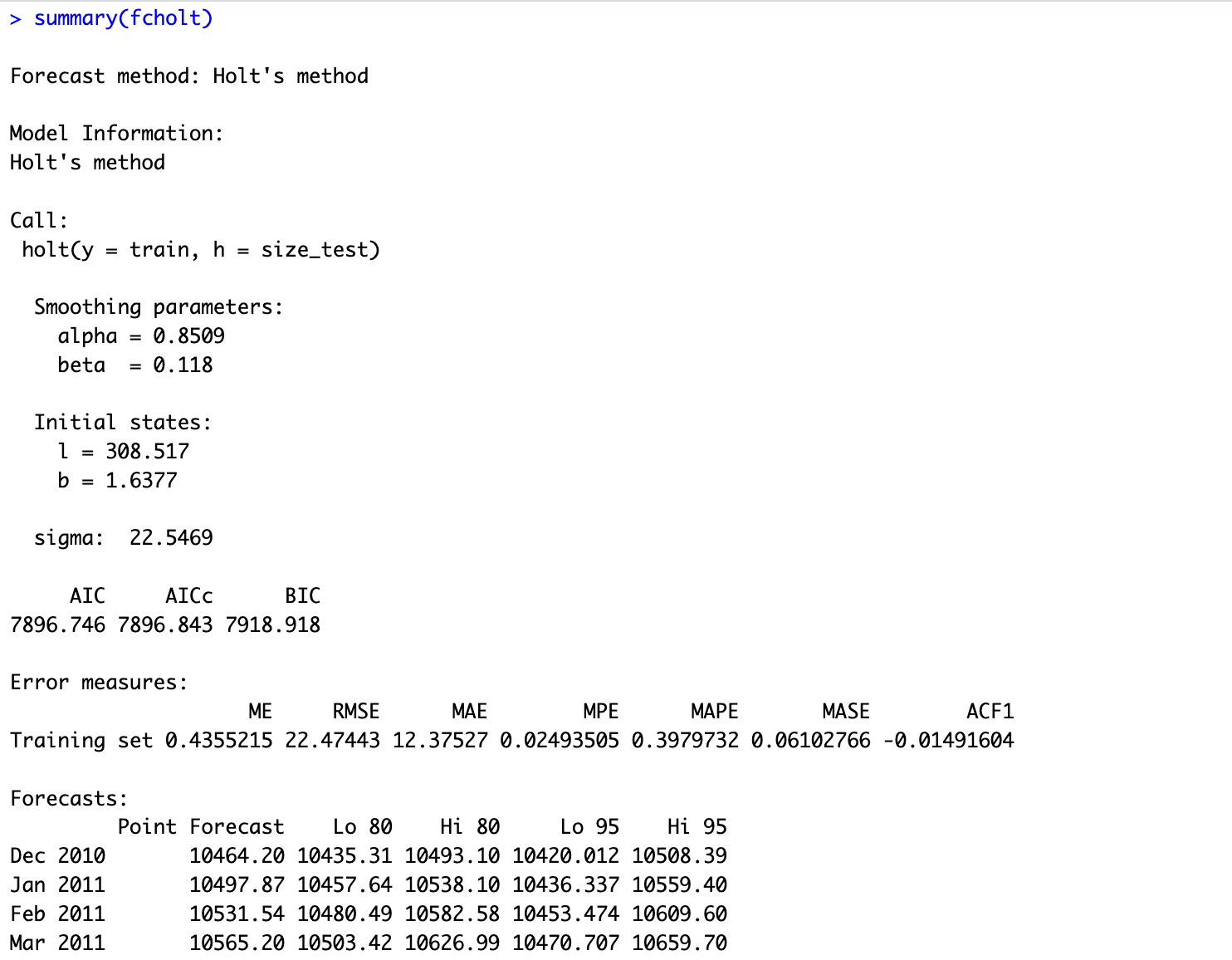
The RMSE or Root Mean Squared Error is the square root of the average of squared errors between the forecasted and actual values and provides the measure of average errors between the forecasted and actual values. The MAE or Mean absolute error is the average of the absolute error between the forecasted and actual value. The lower the value of RMSE and MAE, the better the accuracy of the model and the better the forecasts.

1. **Exponential Smoothing methods**

Exponential smoothing methods forecast the future values wherein larger weights are assigned to recent observations and the weights gradually reduce for the older observations. This emphasizes the more recent values to predict future values without ignoring the older observations. The method of choice for exponential smoothing is the Holt linear method since the data has trend only and no seasonal components. The forecasts for the test set period are as in Fig-2.b.1.

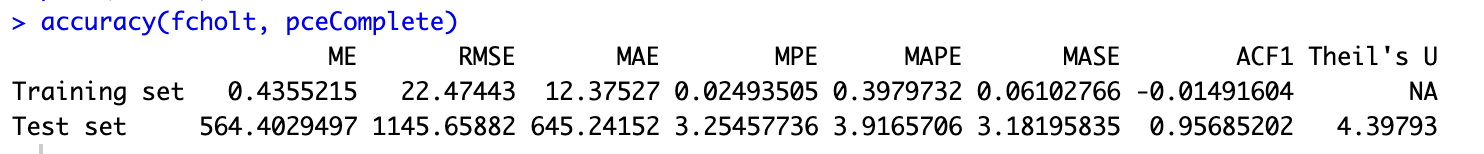


**Fig-2.b.1 Plot of Forecasts from Holt method**



**Fig-2.b.2 Summary statistics for the Holt’s method**

From the summary of the Holt method in Fig-2.b.2, it is seen that alpha, the smoothing parameter, is high with a value of 0.8509, which indicates that recent observations are heavily weighted. The AIC, AICc and BIC provide the measure of fit of the model and lower values indicate better fit. The ME is 0.435, showing a small positive bias in the forecasts from the actual values. It can be seen that the values for RMSE, and MAE are quite low, hence the model has good performance. The value of ACF1 is very close to zero signifying that residuals are uncorrelated.

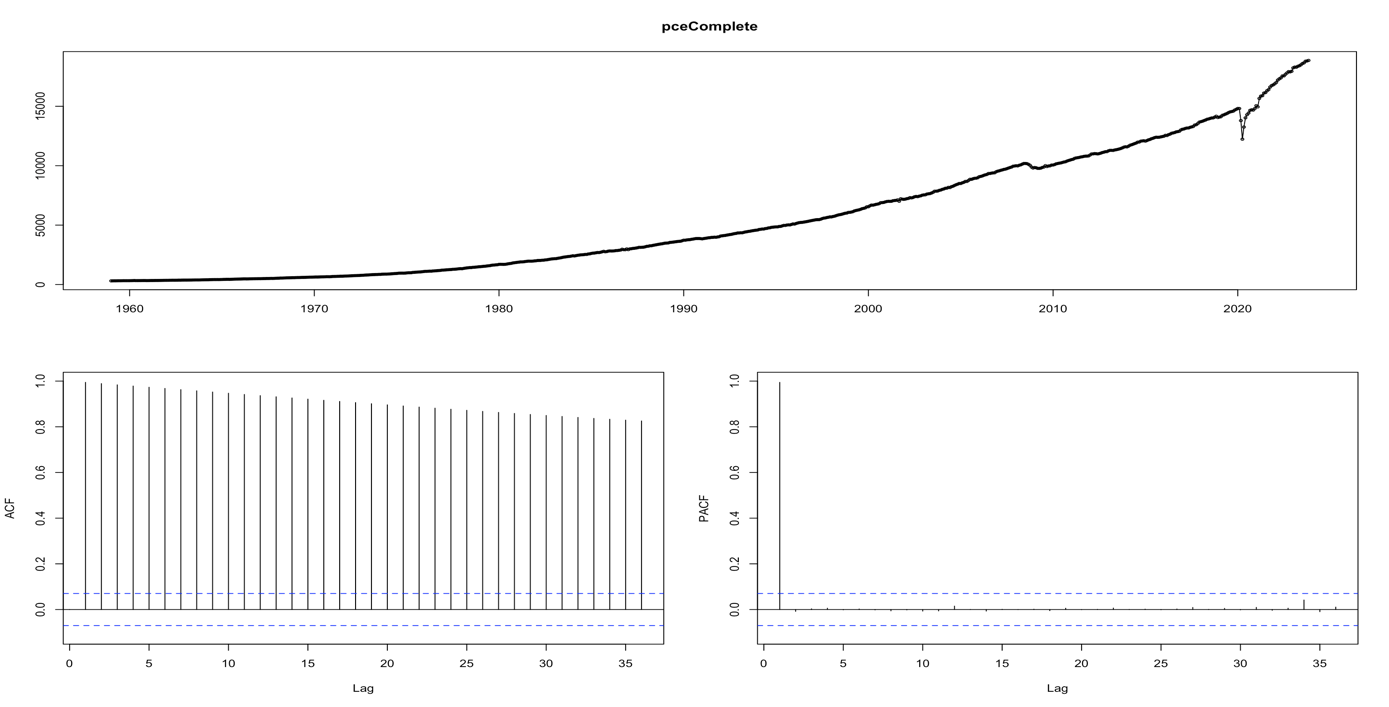


**Fig-2.b.3 Accuracy measure for the Holt’s method**

The accuracy of Holt’s method is checked by comparing the forecast to the original time series. From Fig-2.b.3 It is observed that the RMSE is 1145.65 and MAE is 645.241.

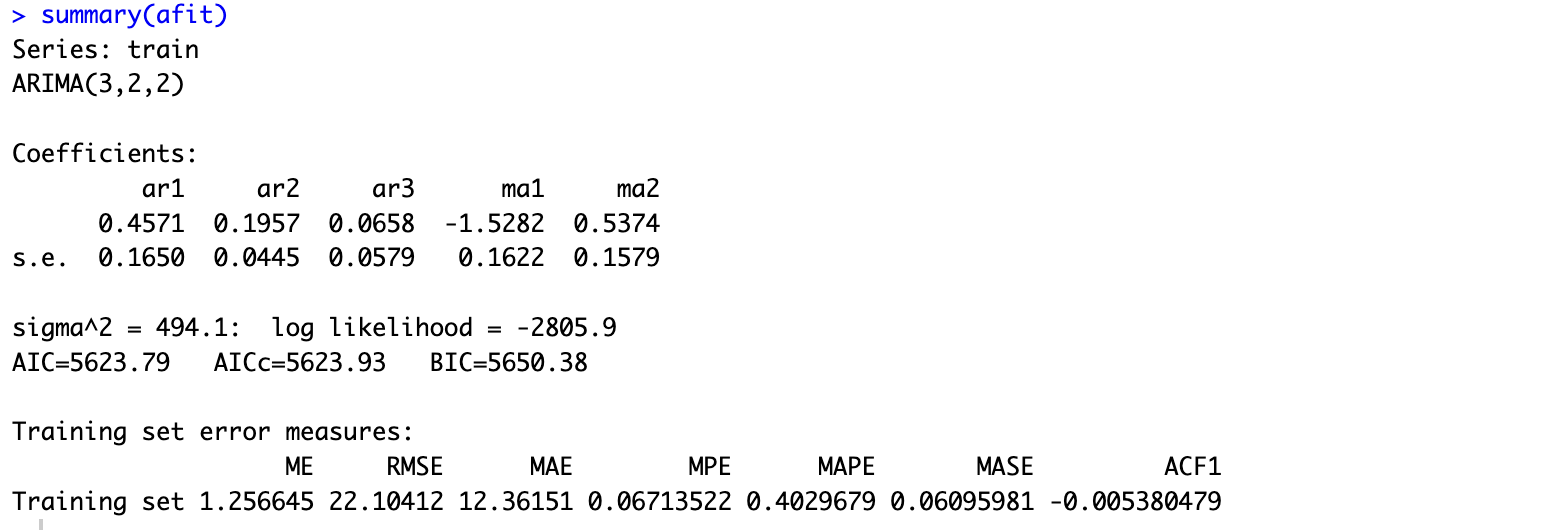
1. **ARIMA model**

ARIMA (Auto Regressive Integrated Moving Average) forecasts future values by taking into account the autoregressive, differencing and moving average components, thus combining the effect of past observations and residual errors on the current values. The integrated component is used to convert a non-stationary series into a stationary one. From ACF and PACF plots in Fig-2.c.1, it can be observed that there is significant autocorrelation at lags indicating the order of MA and AR components.



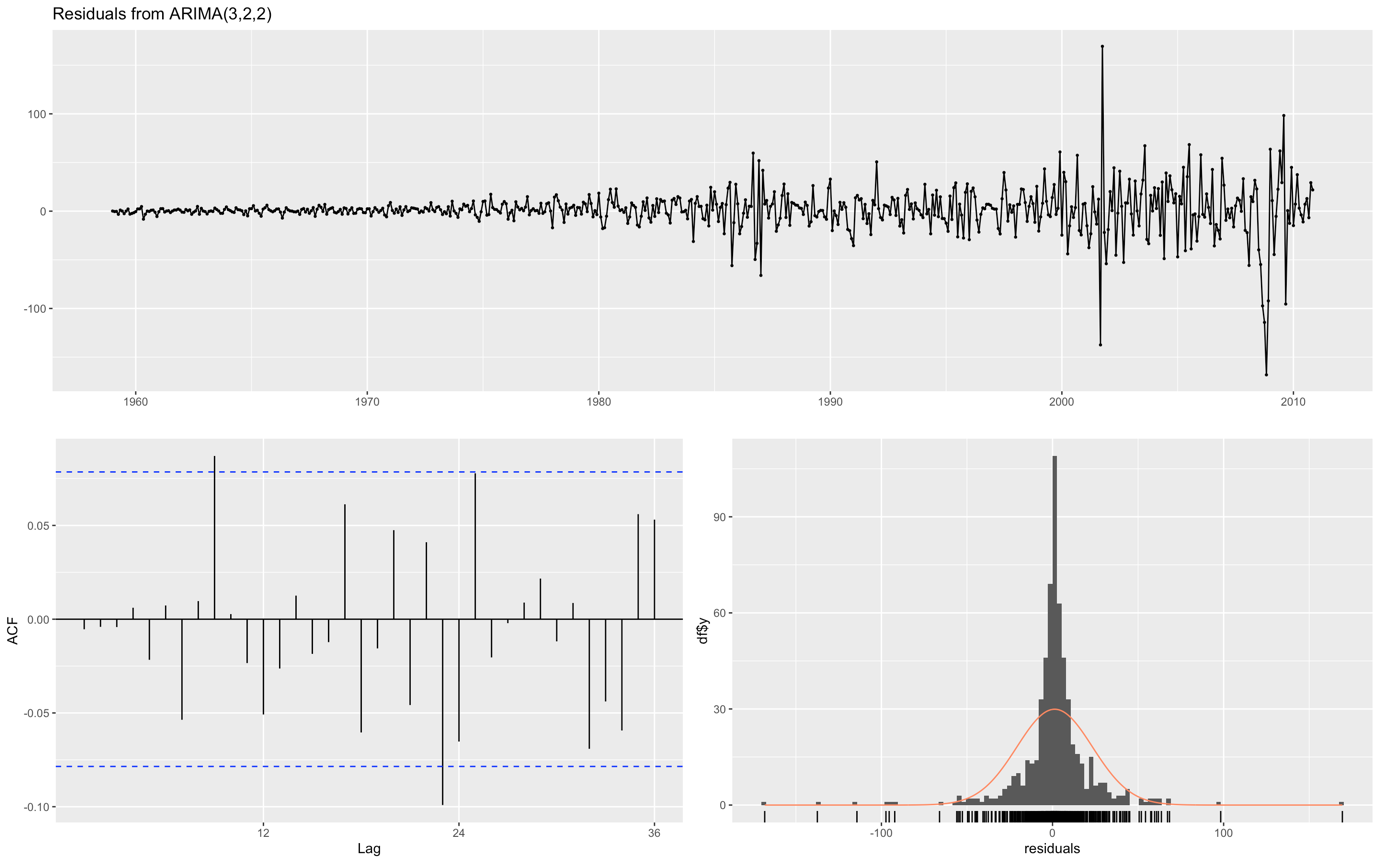
**Fig-2.c.1 ACF and PACF plots along with the original time series**

To select the best arima model based on the data, the auto.arima function is employed and trained using the train set. This produces a model with an order of (3,2,3). Fig-2.c.2 indicates an autoregressive order of 3, differencing of order 2 and moving average order of 2. The AIC, AICc, BIC are comparatively lower than the Holt method indicating a better fit and less complex model.



**Fig-2.c.2 Summary statistics for the ARIMA model**

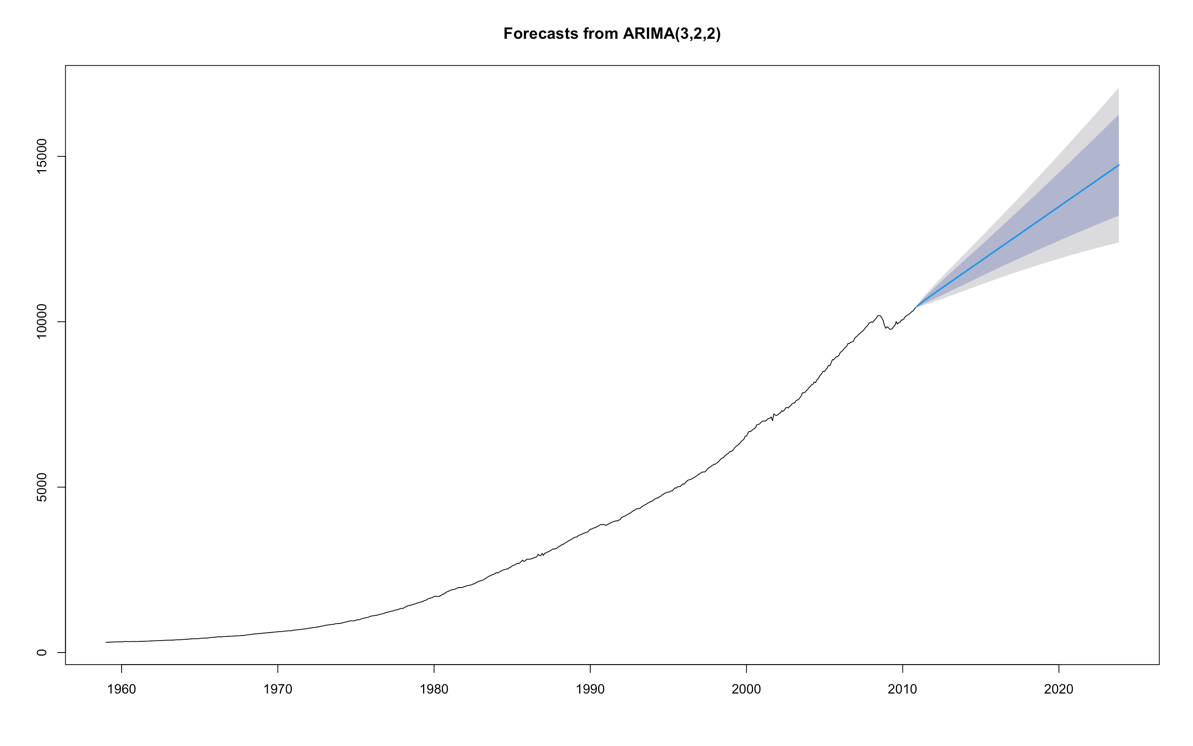
To check the goodness of the model, the residuals are examined. Fig-2.c.3 below shows the graph of residuals, the distribution of residuals and the autocorrelation of the residuals. From the graphs, it can be observed that the residuals have a zero mean, there is no significant autocorrelation between the residuals and the residuals have a normal distribution. Since all three properties are satisfied, it can be concluded that the residuals are white noise. The Ljung-Box test has a p-value of 0.08 which is greater than 0.05, thus the null hypothesis that there is no correlation in the residuals cannot be rejected and the residuals are indeed white noise. This suggests that the ARIMA model with order (3,2,2) is a good fit to the data.





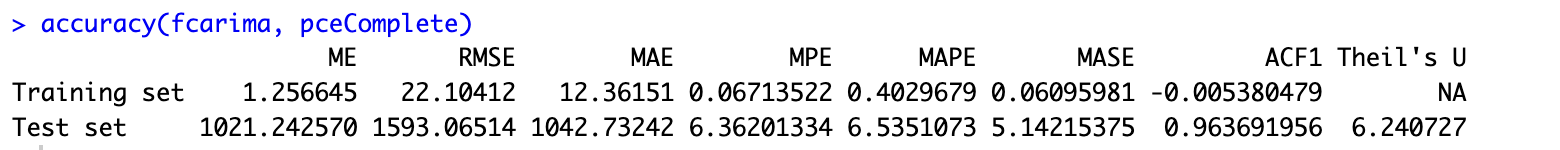
**Fig-2.c.3 Plots from checking residuals and Ljung-Box test**

The forecasts for the test set are as in Fig 2.c.4.



**Fig-2.c.4 Plot of Forecasts from ARIMA model**

The accuracy of the ARIMA model is checked by comparing the forecast to the original time series. From Fig-2.c.5 It is observed that the RMSE is 1593.06.65 and MAE is 1042.73.

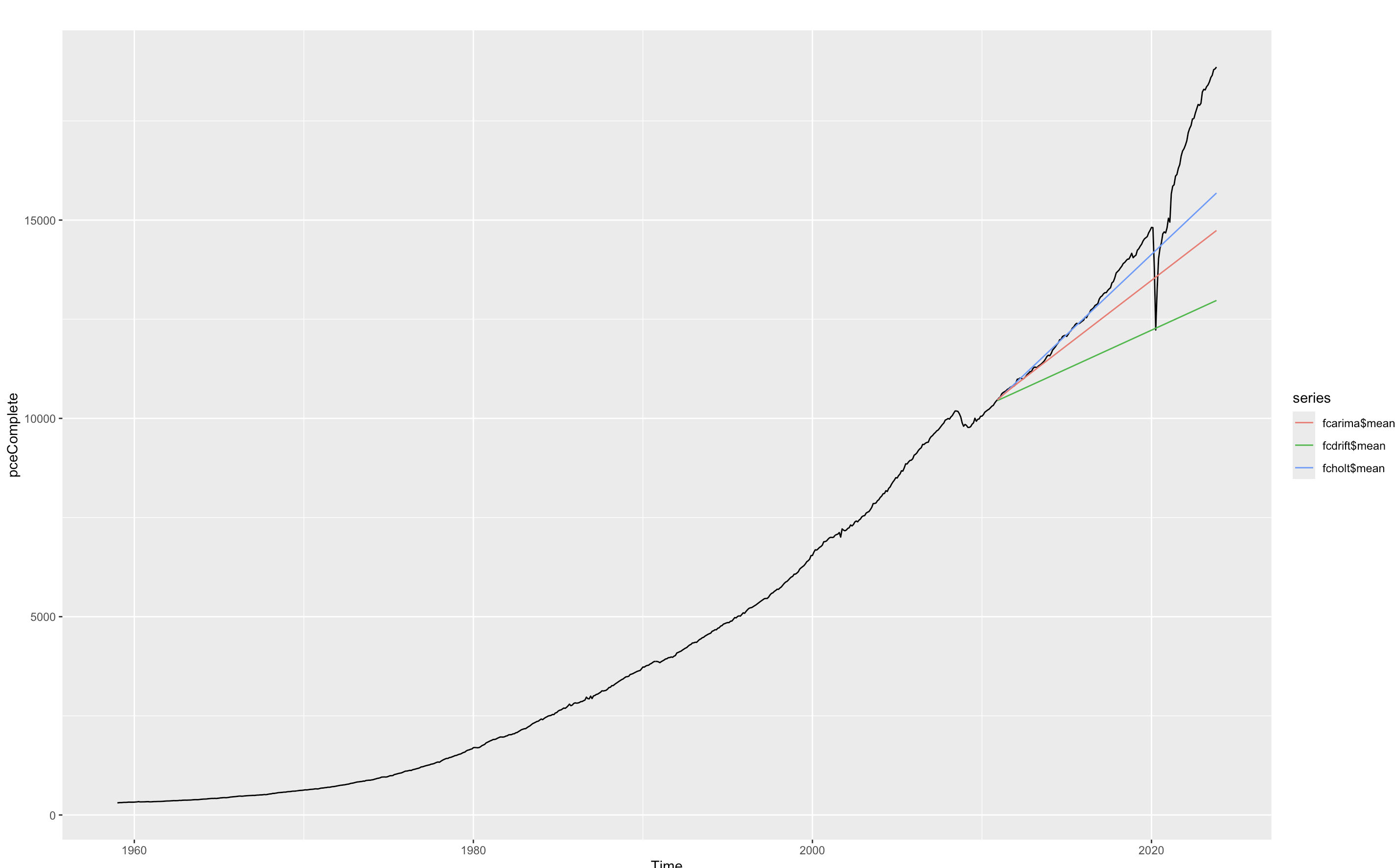


**Fig 2.c.5 Accuracy measure for the ARIMA model**

Comparing the accuracy of the drift, holt and Arima models, it is evident that Holt’s method has the lowest values for RMSE at 1145.65 and MAE at 645.241. The MPE, MAPE and MASE are also lower compared to the drift and Arima models. However, there is a notable difference in errors between the train and test set which indicates poor performance of models for the test set. Using a different split of data might prevent possible overfitting of the model. Moreover, Holt’s method has higher information criteria values when compared to the ARIMA model, indicating that Arima model produces a better fit and effectively explains the data. Considering the accuracy as a criterion for selecting the best model, it can be concluded that Holt’s method is the best of the three.

1. **Plotting Predictions**

Plotting the predictions of the models with actual data in Fig-3. 1 show that the forecasts for Holt’s method are closest to the actual test data.

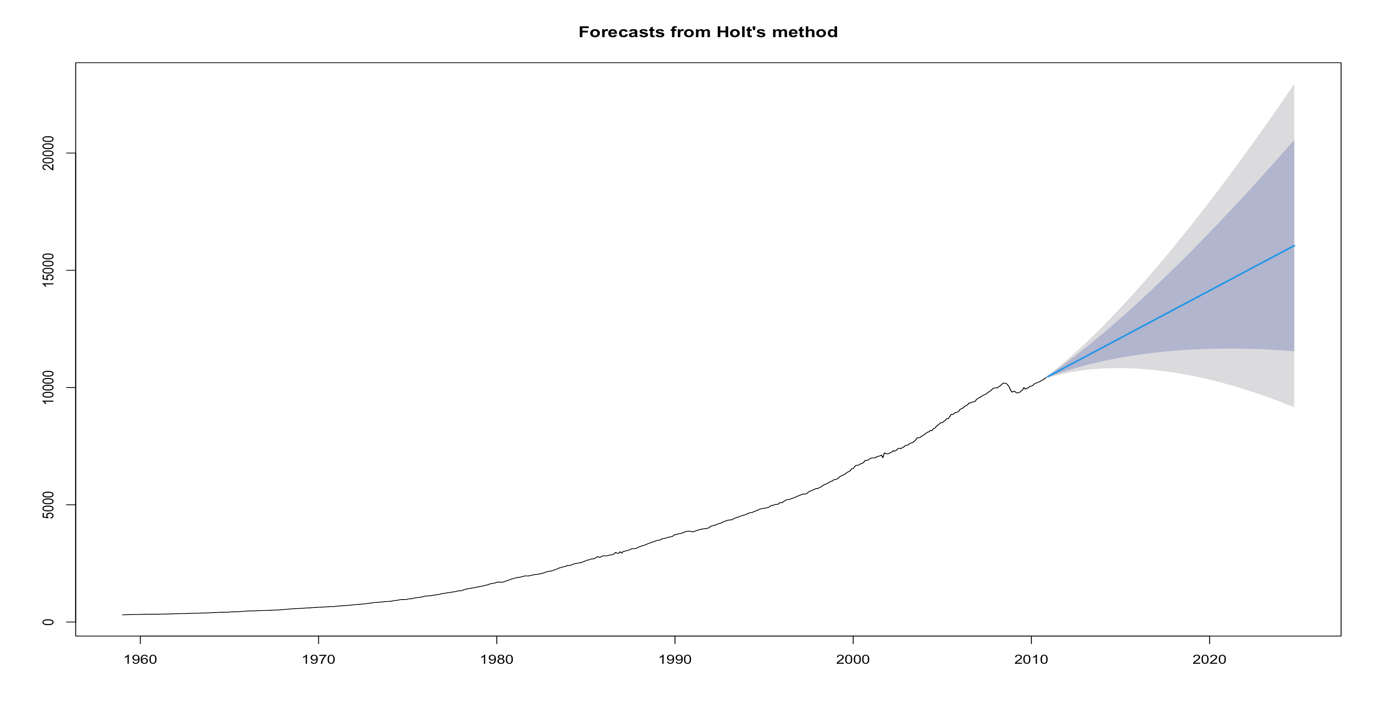


**Fig-3.1 Plot of predictions against actual data**.

1. **Estimation of PCE for October 2024**

The best of the three, i.e., holt’s method is used to estimate the personal consumption expenditure for October of 2024. The PCE for October 2024 is 16052.81 as seen in fig below.

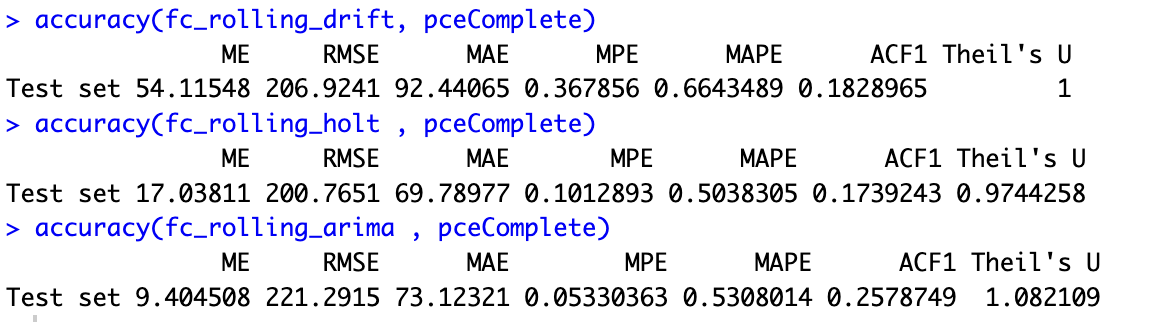




**Fig-4.1 Plot for October 2024 prediction**

1. **One-step ahead rolling forecasting**

Rolling forecasting is a technique where the forecasting model is constantly updated by iteratively training the model using new sets of data and predicting the next set of observations ahead. This enables the model to continuously update and adapt to provide better forecasts. In this case, the test set is predicted in steps of one time period ahead by fitting the model for the training data (assuming train set as observations till 2010).



**Fig-5.1 Accuracy comparison of the models for one-step ahead rolling forecasting**

Comparing the accuracy of the models from Fig-5.1 above shows that Holt’s method has the lowest RMSE at 200.76 and MAE at 69.78 indicating that Holt’s model performs the best in comparison to drift model and Arima model when using one-step ahead rolling forecasting without re-estimation of parameters.

In conclusion, Holt’s method performs best and provides better accuracy while forecasting the test data directly and also while employing one-step ahead rolling forecasting. Comparing the models, it is very evident that one-step ahead rolling forecasting provides much better accuracies than the direct forecasting models since the model is constantly trained and updated. For providing future forecasts of this data, Holt’s method with rolling forecast would be the best choice.