Time Series Analysis of United States 10-Year Bond Yield

PROJECT REPORT SUBMITTED

IN FULFILMENT OF THE REQUIREMENTS FOR COURSE

STAT 497 – Applied Time Series Analysis

DEPARTMENT OF STATISTICS

MIDDLE EAST TECHNICAL UNIVERSITY

BY

Denizcan Bozkurt

Abstract

This study uses a variety of forecasting models, such as ARIMA, ETS, TBATS, NNETAR, and PROPHET, to predict the monthly 10-year bond yield for the United States. The principal tool for this extensive study is R-Studio. The dataset is carefully preprocessed before forecasts are produced. To ensure data integrity, the study includes unit root checking, anomaly cleaning, and outlier analysis. The forecast models are carefully fitted using careful handling of noisy data, and their performances on the training and test datasets are carefully compared concerning several comparison criteria. The results show that NNETAR performs the best out of all the considered models, demonstrating its effectiveness in capturing the complex patterns of the US 10-year Bond Yield. With its insightful analysis of the differences in performance between various forecasting models, this paper offers practitioners and researchers a deeper understanding of financial forecasting.

Key: treasury bonds, forecast, ARIMA and neural network

1. Introduction

The United States 10-year Bond Yield, a crucial financial metric, reflects the annual interest paid by a bond issuer to the bondholder, influencing borrowing costs and acting as a benchmark for various interest rates. Understanding its fluctuations is paramount for investors as it is a barometer for economic health, impacting financial markets, investments, and policy decisions. This investigation aims to enlighten individuals in their investment decisions by providing valuable insights into the intricate patterns and trends of the 10-year Bond Yield. This report predicts the monthly variations in the United States' 10-year Bond Yield from January 1998 to December 2023. The study harnesses the predictive power of diverse models, including classical time series methods such as ARIMA and ETS, TBATS, NNETAR, and PROPHET, all executed within the R-Studio environment. By evaluating model performances on training and test datasets using metrics such as accuracy, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), this study aims to provide investors with information that will improve their comprehension of financial forecasting dynamics and, in the end, help them make well-informed investment decisions.

1.1 Data Description

This dataset explores the dynamics of the US 10-year Bond Yield, covering 312 observations with 5 attributes from January 1998 to December 2023. Date and Close are crucial variables at the center of this investigation. The former bases the dataset chronologically and acts as a temporal guide. However, the close variable stands out, which represents the bond yield's closing prices. Close was carefully chosen to provide a condensed picture of the day's financial activity and remove any potential distortions from intraday fluctuations. The data provided came from investing.com

1.2 Aim of the study

This study aims to predict the monthly variations in the United States' 10-year Bond Yield using forecasting models like ARIMA, ETS, TBATS, NNETAR, and PROPHET. Through rigorous comparison and evaluation in R-Studio, the goal is to provide valuable insights for financial analysis, aid decision-makers, and enhance understanding of patterns in the 10-year Bond Yield.

2. Data Preparation

At the beginning of the analysis, the data set is divided into a test set and a train set. While doing this, the last three years are kept as a test set.

Figure 1 Time Series Plot of Data Set

The plot appears to be nonstationary. The mean term fluctuates. Time has an impact on it. Furthermore, there appears to be a declining trend, but the non-linear trend raises doubts about the stochastic trend. However, it is not possible to draw firm conclusions from that plot.

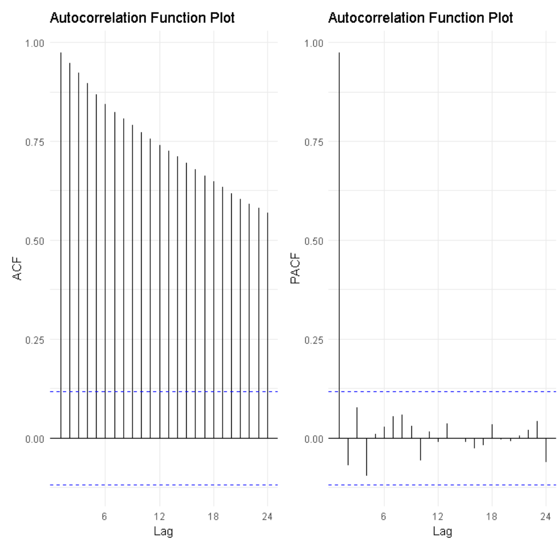

Figure 2 ACF&PACF Plot of Data Set

The ACF plot exhibits a slow, linear decay, supporting the analysis presented in the first plot. As a result, our process is referred to as nonstationary. It can be observed that PACF stops after the initial lag; however, since it is already decided that the process is nonstationary by looking at ACF and the time series plot of the data, there is no need to interpret the PACF plot of the data set.

Next, we use stl decomposition to examine the data anomalies and find that the series has no anomalies.

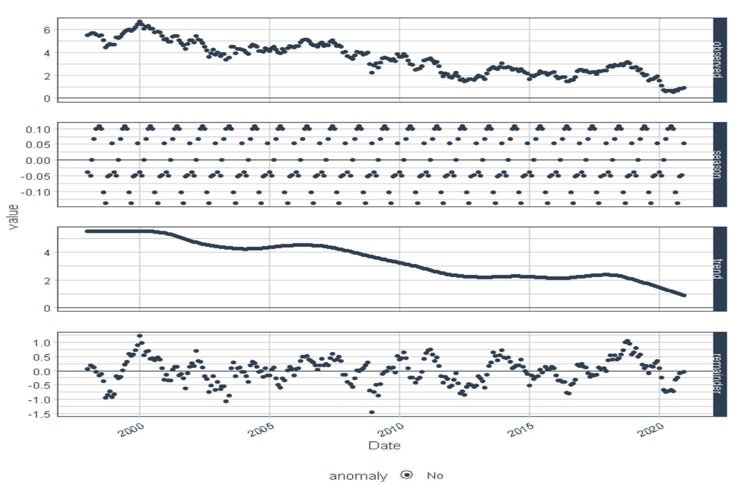

Figure 3 Anomaly Detection Plot

It is usually necessary to have stationarity before starting the modeling process. Thus, looking for a variance-stabilizing transformation that gets the data closer to satisfying this requirement becomes necessary. Even though the Box-Cox transformation, data does not change in variation. Then, continued without transformation.

Figure 4 Data and Log transformed data

To assess the stationarity of the training dataset, the KPSS test is applied, taking into account both the trend and level components. P-values less than 0.05 in the results show that the data is nonstationary about both the level and trend. Furthermore, it is discovered that the training data's mean is not zero. We then use the augmented Dickey-Fuller (ADF) tests to evaluate stationarity and find a p-value greater than $\alpha=0.05$ in the first ADF test results in non-stationarity. The ADF test results indicate a nonstationary system with a deterministic trend when the p-value falls below $\alpha=0.05$ when a trend term is included. However, since the KPSS test is more potent than ADF, the series has a stochastic trend. To address this, differencing is applied to the dataset.

The HEGY test is employed to investigate the presence of unit roots and seasonal unit roots in the training dataset. The statistical results reveal that there needs to be one differentiating for unit roots.

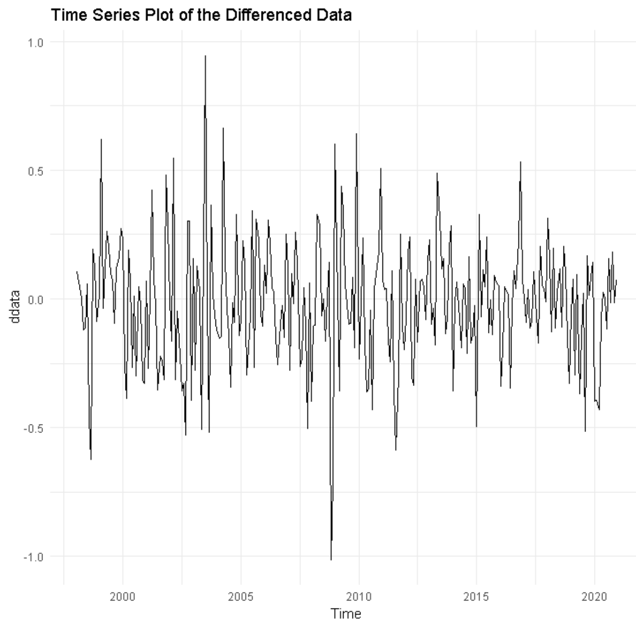

Figure 5 Time Series Plot of Differenced Data Set

After differentiating and obtaining stationarity, ACF and PACF were checked. The process's ACF and PACF cut off after lag 2. Thus, by examining the ACF and PACF plots, ARIMA(2,1,2) is the model recommended for this data set.

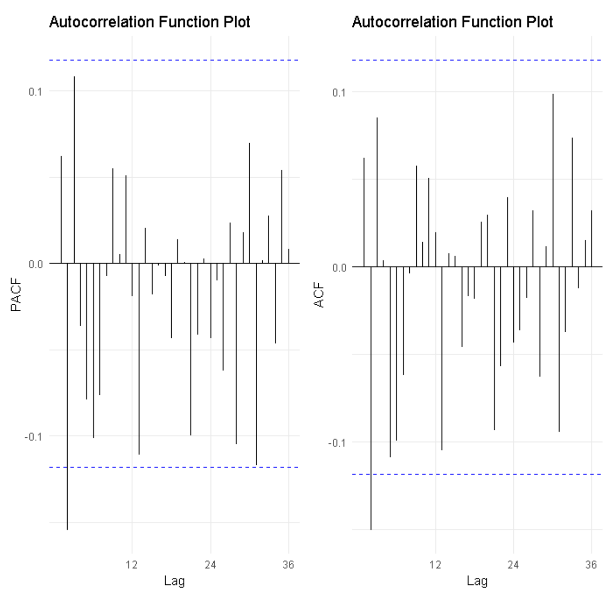

Figure 6 ACF&PACF Plot of Stationary Data Set

Moreover, the ESACF method can infer probabilistically the orders of a stationary or nonstationary ARMA process. By creating a triangle with "o" terms, the ESACF method seeks to identify the model with the fewest parameters. As a result, the table recommends using the ARIMA (1,2,1) model with the data set.

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o x o o o o o o o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x o o o o o o o o o  o  o  o
4 x o o o o o o o o o o  o  o  o
5 x x o o o o o o o o o  o  o  o
6 x x o o o o o o o o o  o  o  o
7 o x o o x o o o o o o  o  o  o
```

Figure 7 EASCF

Next, the model was identified using the MINIC table. To determine which is optimal, the SBC values are taken into account. The model that best fits the data set is indicated by the order with the lowest SBC value. It is claimed that the recommended models are ARIMA (0,1,0), ARIMA(0,1,2), and ARIMA(2,1,0) based on the data.

p	q	sbc
0	0	-761.49
9	2	-757.87
2	0	-756.24
1	0	-756.16
0	1	-755.54
0	3	-753.01
3	0	-752.89
1	2	-752.58
2	1	-751.58
1	1	-749.93

Table 1 EASCF

Once the models that could fit the series have been identified, their performances and significance are compared based on criteria. As a result, it is observed that ARIMA(4,1,2) and ARIMA(2,1,4) emerge as significant models. Subsequently, the best model is determined by examining the AIC values, with the selection criterion favoring a smaller AIC value. Accordingly, ARIMA(2,1,4) is chosen as the best model due to its lowest AIC value.

```
ARIMA(2,1,4)

Coefficients:
         ar1      ar2      ma1     ma2     ma3      ma4
      1.2129  -0.6997  -1.1421  0.4701  0.3331  -0.2388
s.e.  0.2149   0.1430   0.2099  0.1457  0.0845   0.0586

sigma^2 = 0.06002:  log likelihood = -0.52
AIC=15.03   AICc=15.45   BIC=40.35
```

Figure 8 ARIMA model

The next step involves proceeding with the diagnostic checks of the selected model.

Once the time series model has been identified and estimated, verifying its goodness of fit and the accuracy of the underlying assumptions is essential. The ARIMA forecast is created if we have the ideal fit model. Model checking in time series analysis relies on residual analysis and is comparable to conventional regression analysis.

Next, formal, and visual inspections were conducted to assess the presence of serial correlation. As seen, all spikes are in the White Noise.

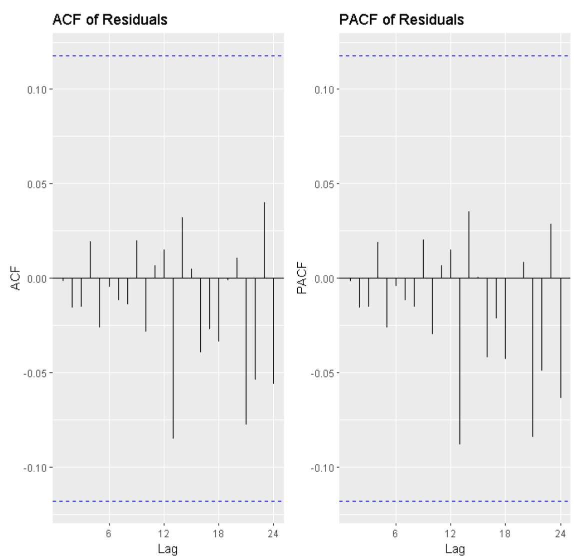

Figure 9 ACF&PACF Plot of the standard residuals

Consequently, it can be inferred that there is no correlation among the residuals. Formal tests should be applied. Box-Ljung is used in this process; it is not deemed indispensable. They are, therefore, uncorrelated.

```
> Box.test(r2,lag=15,type = c("Ljung-Box"))

        Box-Ljung test

data:  r2
X-squared = 3.3918, df = 15, p-value = 0.9991
```

Figure 10 Box-ljung test

The visual inspection tool Q-Q plot and the formal Shapiro-Wilk and Jarque-Bera tests verify the normalcy assumption.

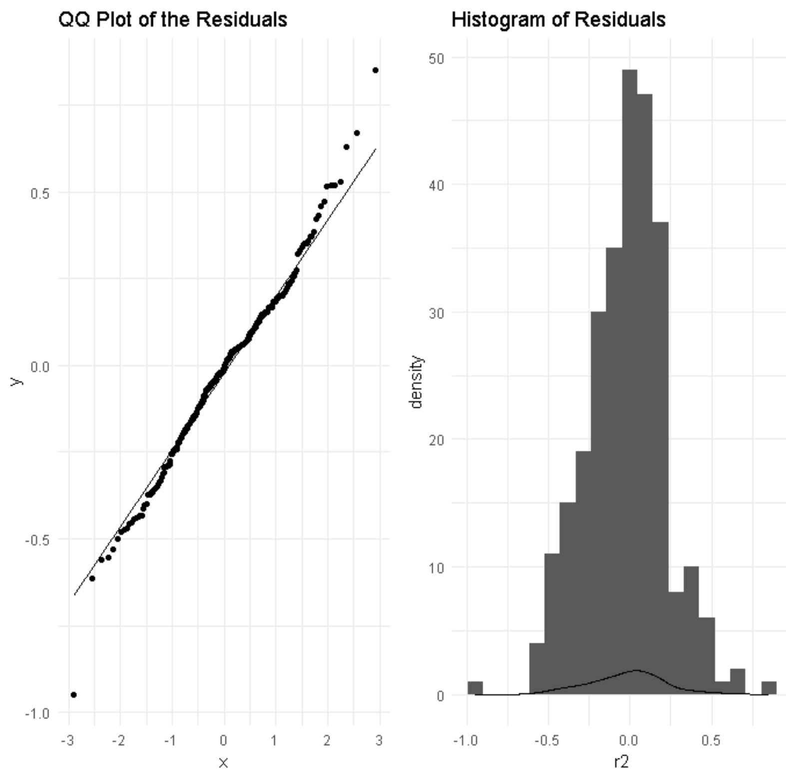

Figure 11 QQ plot and histogram of the standard residuals

Relatives may not follow a normal distribution at the Q-Q plot. Shapiro-Wilk and Jarque-Bera should be used to confirm non-normality. The residuals do not follow normality, and the test was significant.

The residuals' heteroscedasticity is the final assumption to be verified. The bresuch-Pagan test was used to test this assumption.

```
> bgtest(m,order=15)

        Breusch-Godfrey test for serial correlation of order up to 15

data:  m
LM test = 3.7668, df = 15, p-value = 0.9984
```

Figure 12 BG test

The Bresuch-Pagan test results show that the model's residuals are uncorrelated.

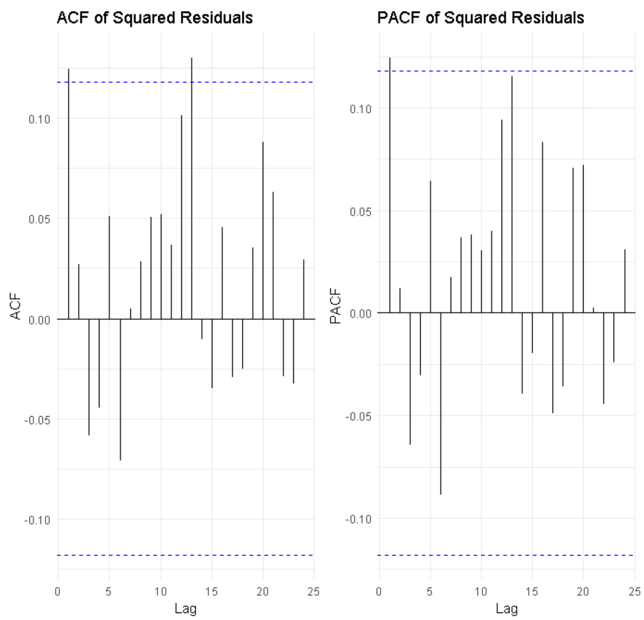

Figure 13 ACF&PACF of the squared residuals

Both plots show that some spikes are out of the white noise bands, and there is a correlation between them, which indicates the heteroscedasticity problem and usage of the (G)ARCH model. However, Engle's Arch test was used to test it formally.

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  r2
Chi-squared = 11.806, df = 12, p-value = 0.4614
```

Figure 14 Arch test

Since the results show there is no arch effect. There is no need to continue with GARCH models. The ARIMA model is selected since normality is not satisfied with transformation, and other assumptions are satisfied.

3. Modelling

After ARIMA model, the optimal exponential smoothing model will be pursued using the ets function in the R forecast package following the ARIMA model. It is observed that Holt's exponential smoothing model demonstrates the most effectiveness for the series.

```
holt(y = train, h = 36, beta = 0.221)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 0.221

  Initial states:
    l = 5.5627
    b = -0.0696

  sigma:  0.2688

     AIC     AICc      BIC
829.9892 830.1368 844.4709
```

Figure 15 Holt's model

After fitting the model, the residuals of the ETS model is checked by Shapiro-Wilk test and seen that they do not follow normal distribution.

After ETS model, TBATS model is fitted to the series. The model details are given below.

```
Call: tbats(y = train)

Parameters
  Lambda: 0.806817
  Alpha: 1.098635

Seed States:
         [,1]
[1,] 3.59363
attr(,"lambda")
[1] 0.8068169

Sigma: 0.1988661
AIC: 787.0865
```

Figure 16 TBATS model

After fitting the model, the residuals of the ETS model are checked by Shapiro-Wilk test and seen that they do not follow normal distribution.

Other model is Neural Network Model, model is fitted to the series. In model, size,repeats and decay parameters selected for the best performance with a loop. The model details are given below.

```
Series: train
Model:  NNAR(1,1,1)[12]
Call:   nnetar(y = train, size = 1, repeats = 1, lambda = NULL, decays = 0,
    PI = TRUE)

Average of 1 networks, each of which is
a 2-1-1 network with 5 weights
options were - linear output units

sigma^2 estimated as 0.06235
```

Figure 17 NNETAR model

After fitting the model, the residuals of the NNETAR model are checked by Shapiro-Wilk test and seen that they do not follow normal distribution.

Last model is Prophet. In model, seasonality.prior.scale, changepoint.prior.scale and changepoint.range parameters selected for the best performance with a loop. The result of fitted model accuracy given below. The residuals of the Prophet model are checked by Shapiro-Wilk test and seen that they do not follow normal distribution.

After the models have been fitted, forecast values are obtained from each method using the **forecast** function, and their accuracy is subsequently calculated. The accuracy of the models is provided below.

MODELS	ME	RMSE	MAE	MAPE	MASE	ACF1
ARIMA	-0.013	0.184	0.140	12.066	0.289	-0.014
ETS	0.001	0.353	0.206	7.227	0.321	-0.022
TBATS	-0.015	0.249	0.188	6.535	0.294	-0.020
NNETAR	**0.000**	**0.25**	**0.190**	**6.770**	**0.296**	**0.069**
PROPHET	0.000	0.567	0.458	19.03	-	0.907

Table 2 The train accuracy of model

MODELS	ME	RMSE	MAE	MPE	MAPE	ACF1
ARIMA	1.543	1.552	1.543	105.661	105.661	0.281
ETS	0.005	0.175	0.14	-1.035	9.913	0.414
TBATS	0.537	0.563	0.537	35.772	35.772	0.281
NNETAR	**0.001**	**0.17**	**0.132**	**-1.421**	**9.491**	**0.338**
PROPHET	1.241	1.758	1.38	30.018	40.73	0.915

Table 3 The forecasting performance of models

Both tables show that the Neural Network model outperforms the other methods compared to the train and test sets concerning all measures. Also, it can be seen that ETS Holt has the second-best forecasting performance when compared to another model. Lastly, the prophet model has the lowest forecasting and train accuracy, as shown in the tables above.

The forecasting performance of the models can be also observed from following plots.

Figure 18 ARIMA model forecasting

In figure 18, the ARIMA confidence interval does not seen well. It can be caused by it’s performance.

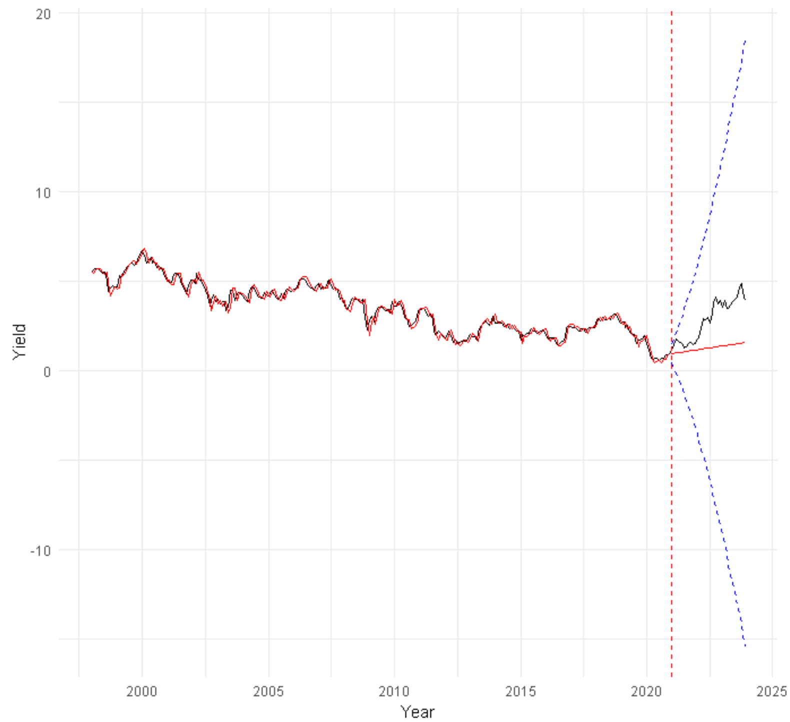

Figure 19 Holt's model forecasting

In figure 19, ETS model confidence interval' does not seen well. However, model forecasting means more accurate then ARIMA model.

Figure 20 HBATS model forecasting

In figure 20, both performance and confidence interval does not seems to fit well in HBATS model.

Figure 21 NNETAR model forecasting

In figure 21, best performance and confidence interval is in this model. NNETAR model is best fit for this data.

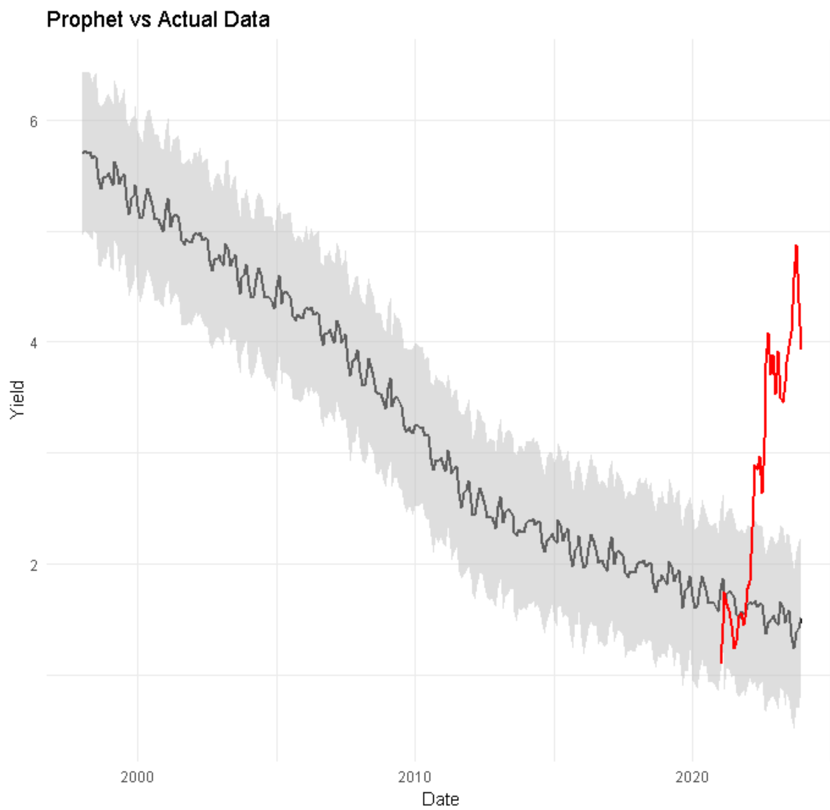

Figure 22 Prophet model forecasting

In figure 22, the worst performance is belonging to Prophet model. Future prediction and actual data are different. It can be easily seen in the plot.

4. Discussion

The stationary check of the series was the first step in this study after the data was cleaned and divided; however, it was discovered that this requirement was not satisfied because of the stochastic trend in the series after examining time series, ACF&PACF plots, and the outcomes of the KPSS and ADF tests. This problem was solved by applying differencing techniques. Following process stabilization, a few preliminary models were proposed using particular techniques like ACF&PACF plots, the ESACF table, and the MINIC table. Then, in AIC comparison, the best model with the fewest and most significant parameters was selected from these preliminary models. On residuals, diagnostic checks are applied after the optimal model has been fitted. At this point, the non-normality of the errors became an issue, and transformation was used; however, the series residuals did not provide a satisfactory solution. Nevertheless, formal tests and visual inspection tools confirm that the errors are homoscedastic and uncorrelated. Four distinct forecasting methods were taken into consideration in addition to the best ARIMA models, and forecasts were generated using these methods. In the end, when compared to other models, NNETAR performs the best in both series modelling and value prediction. Overall, to make future predictions for US 10-year treasury bonds, people can use NNETAR model because NNETAR outperforms other forecasting models.

References

1. **Investing.com. (n.d.). United States 10-Year Bond Yield. https://www.investing.com/rates-bonds/u.s.-10-year-bond-yield**