A Univariate Time Series Analysis for U.S. Imports of Goods by Customs Basis from World

Yusuf Kağan ÖZKAN

2429215

Abstract

The univariate time series analysis of the U.S. Imports of Goods by Customs Basis from World was the main emphasis of this project. Several procedures were used in this project's time series analysis. Following the identification of a trend component with the plots and related tests, the appropriate transformation was used to stabilize the variance of the series. Several models were tested, and the best model to utilize in the forecast was selected. Diagnostics were verified using a few plots and the relevant tests prior to forecasting. Ultimately, the purpose of forecasting was to comprehend and analyze future findings.

1. INTRODUCTION

This project focused on the univariate time series analysis of U.S. Imports of Goods by Customs Basis from World. This data contains 2 columns which are date and monthly price of goods in terms of millions of dollars. Our data is starting from January 1989 to September 2023. This study aims to forecast future values and making interpretations about U.S. Imports of Goods by Customs Basis from World.

2. TIME SERIES PLOT

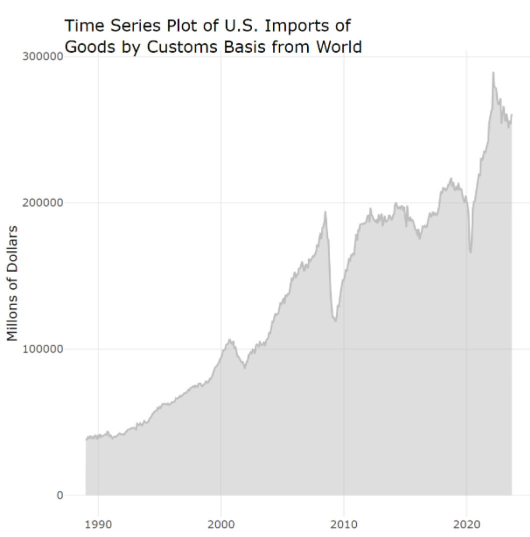

Figure 1: Time Series Plot of US Imports

The time series plot of U.S. Imports of Goods by Customs Basis from World increases over time, so we can say that there is an increasing trend, this might cause non-stationary process. Also, from the plot we do not observe seasonality, since we are working on seasonality adjusted data. However, we should apply statistical tests to be sure that which patterns does the series have. Before statistical tests let's have a look some descriptives about data.

1	**1.01.1989**	**37513.2**
2	1.02.1989	38561.7
3	1.03.1989	39724.7
4	1.04.1989	38664.7
5	1.05.1989	40909.6
413	1.05.2023	254134.3
414	1.06.2023	251039.7
415	1.07.2023	256218.4
416	1.08.2023	253667.5
417	1.09.2023	260612.7

Table 1: First and Last 5 Observations

We can see from the figure the first and last five observations of our data. We can observe a significant change in amounts from 1990's to 2020's.

Min	1st Q.	Median	Mean	3rd Q.	Max
37513	73623	139189	135539	189809	2895568

Table 2: Summary Statistics of Data

From above table we can see that median and mean values are close each other.

There is a huge difference between minimum and maximum values the time series.

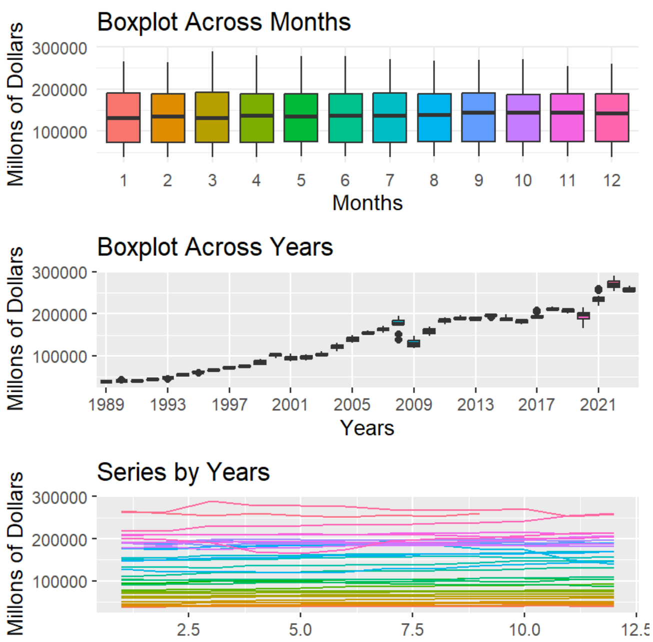

Figure 2: Seasonality Plots

From above plots, the first one shows that the median values for each month for each year are almost equal, so our series does not show seasonal patterns.

From the second plot we can see an increase in different years, so we can observe increasing trend patterns for long run.

From the last plot each year's observations listed, since they do not show overlapping patterns, we can suggest that our data do not contains seasonality patterns.

3. CROSS VALIDATION

To check our forecasts, we should divide our data into 2 parts which are train and test.

Since we have monthly data, we should divide last 12 observations for test set.

Now we have 405 observations for train set and 12 observations for test set. After this step we will continue our analysis on train set.

4. ANOMALY DETECTION

Anomalies may affect the behavior of our time series data, so we should detect anomalies and remove them from data. We used STL decomposition for anomaly detection.

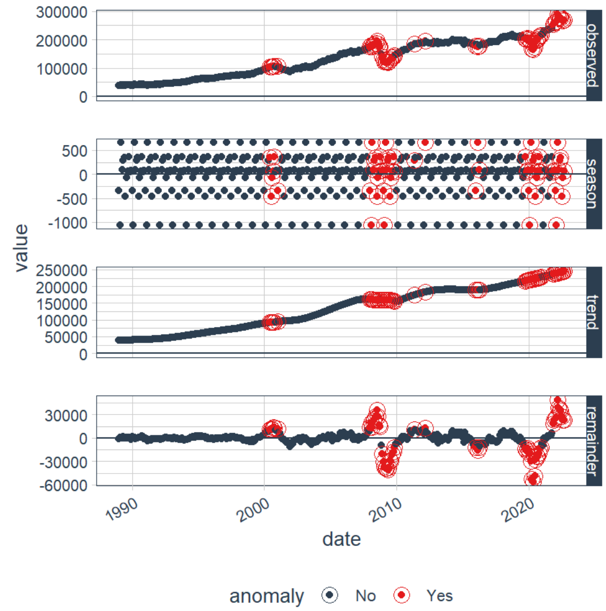

Figure 3: Anomaly Decomposition

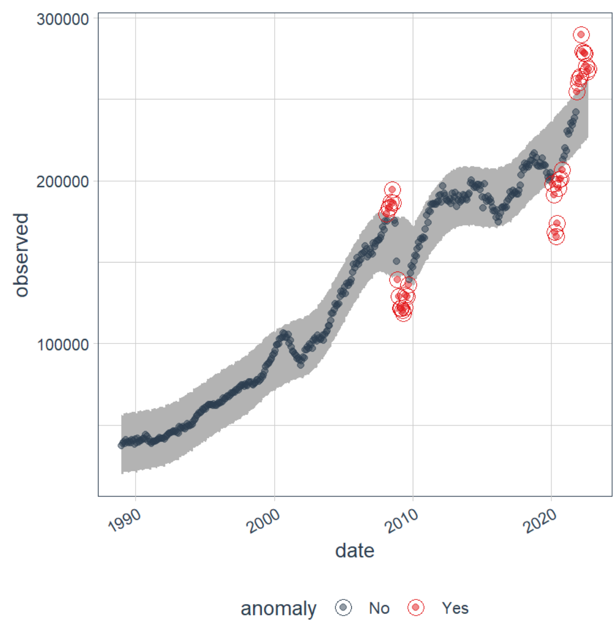

Figure 4: Anomaly Graph

We checked that the data includes any anomalies or not. It can be seen from the plot we have some anomalies, and we need to remove them. We can also observe anomalies from decomposition plot.

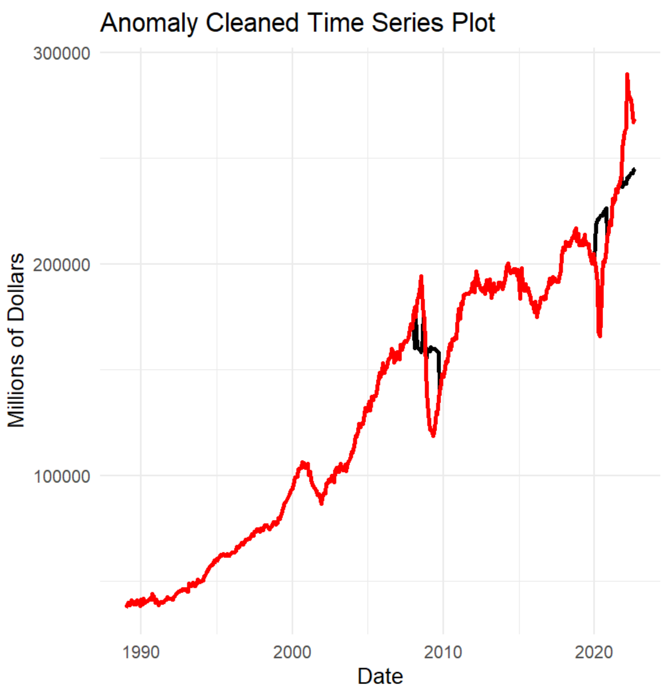

Figure 5: Anomaly Removed Time Series Plot

After removing the anomalies from train set, we can see the difference between train set and anomaly removed series which is with the black lines. We can see that after removing anomalies our outliers and the points which do not follow the regular pattern of the data, pointed in more reasonable places. We will conduct our analysis with this anomaly removed series.

5. BOX- COX TRANSFORMATION

To stabilize the variance, we can apply Box-Cox transformation, when we check the lambda value of our series it gives us the result 0.16, so we should apply the transformation to our series.

6. ACF-PACF PLOTS AND TESTS STATIONARITY

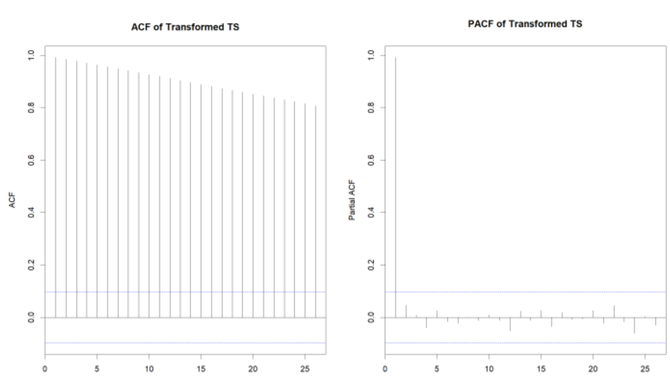

Figure 6: ACF-PACF plots of series

From figure 6 we can observe a slow decay at ACF which shows that our series are non-stationary. Therefore, we should apply statistical test to check stationarity of our series.

<u>KPSS Test:</u>

For Level component the p-value is: 0.01.

For Trend Component the p-value is: 0.01.

<u>ADF Test:</u>

For regression with an intercept (constant) but no time trend (c) the p value is: 0.5251.

For a regression with an intercept (constant) and a time trend (ct) the p value is: 0.9008.

Both tests suggest that we have nonstationary stochastic trend.

6.1 REMOVING TREND

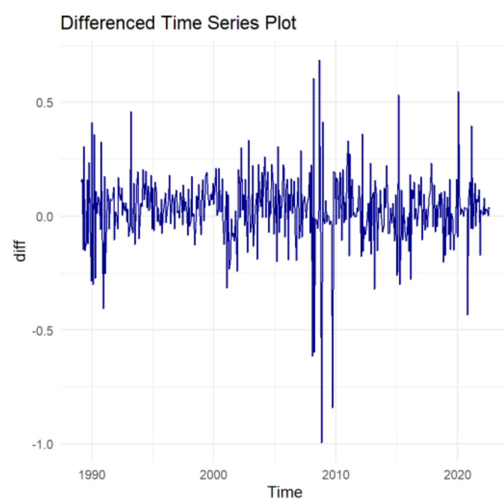

Figure 7: Plot of Differenced Series

As we can see that we have stationary series around mean 0, we still have some outlier values.

<u>KPSS Test:</u>

For Level component the p-value is: 0.2168.

<u>ADF Test:</u>

For a regression with no intercept (constant) nor time trend the p- value is: 0.01.

So, after taking one difference each test suggests that we have stationary process.

7. IDENTIFYING MODELS

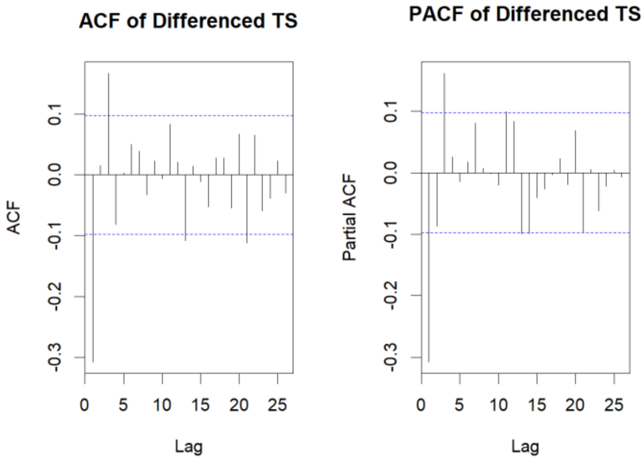

Figure 8: ACF-PACF Plots of Differenced Series

Proper Models (ACF-PACF):

- ARIMA (3,1,0)
- ARIMA (3,1,3)
- ARIMA (3,1,1)

Auto-Arima Model:

- ARIMA (0,1,3)

Arma-select Model:

- ARIMA (0,1,1)

EACF Model:

- ARIMA (0,1,3)

7.1 MODELLING

After we determine possible models, we should check their parameters significance and with the models that has significant parameters, we will choose the model with minimum BIC value.

After checking the significance of each model parameters, we will see that all models' parameters are significant, so let's check the BIC values.

MODEL	BIC
ARIMA (3,1,0)	-366.7123
ARIMA (3,1,3)	-333.9621
ARIMA (3,1,1)	-335.9201
ARIMA (0,1,1)	-326.3946
ARIMA (0,1,3)	-333.4594

Table 3: Model BIC Values

Since ARIMA (3,1,0) has the lowest BIC value we will continue to our forecast with this model.

	AR1	AR2	AR3
Coefficients	-0.27	0.03	0.21
Standard Error	0.04	0.05	0.049

Table 4: ARIMA Model Summary

7.2 DIAGNOSTIC CHECKING

To continue our ARIMA forecasting we must check the normality, serial correlation, and heteroscedasticity of the residuals of our model. Therefore, we will check the assumptions visually and statistically.

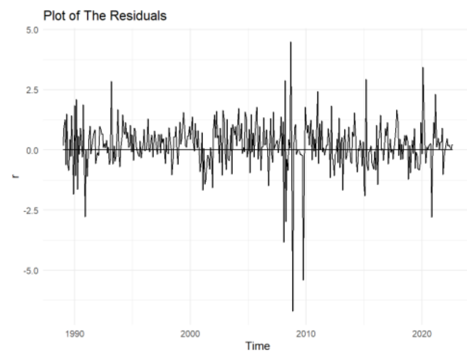

Figure 9: Residual Plot

We can see that residuals are deviate around zero.

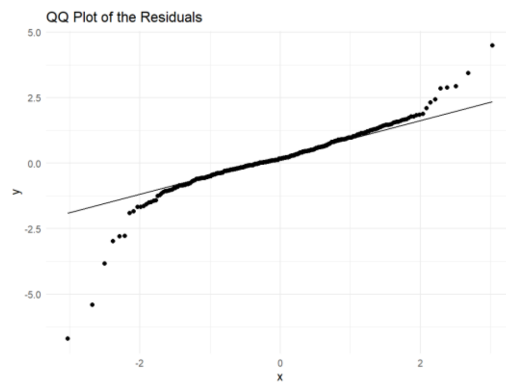

Figure 10: QQ Plot of Residuals

We can see that we have some outlier values, and from the plot we can say that our residuals follow heavy tailed distribution.

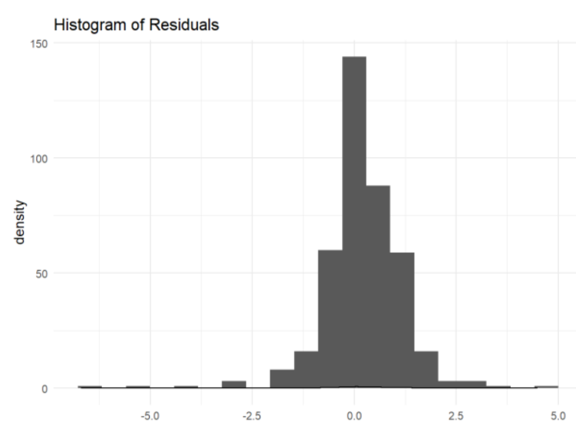

Figure 11: Histogram of Residuals

We can see that most of the residuals deviates around zero.

Test	Jarque- Bera	Shapiro- Wilk
P-value	2.2e-16	2.2e-16

Table 5: Normality Tests

From this table we can see that both tests suggests that our residuals do not distribute as normal.

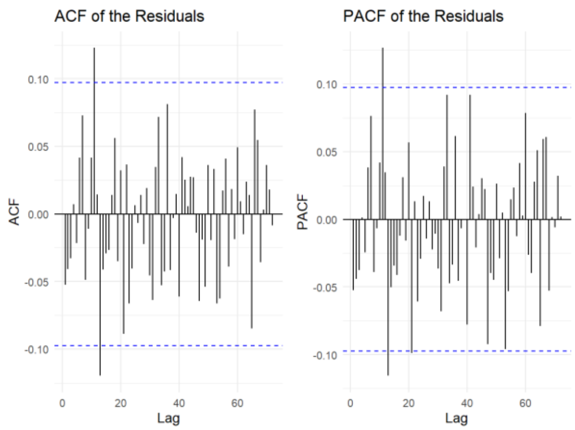

Figure 12: ACF-PACF Plot of Residuals

From above plots we can see that almost all the lags are inside of the white noise band, so that might show residuals do not have serial correlations.

Tests	Breusch- Godfrey	Box-Ljung
P-value	0.1747	0.1477

Table 6: Serial Correlation tests

From above table we can see that both tests suggest that our residuals do not have serial correlation.

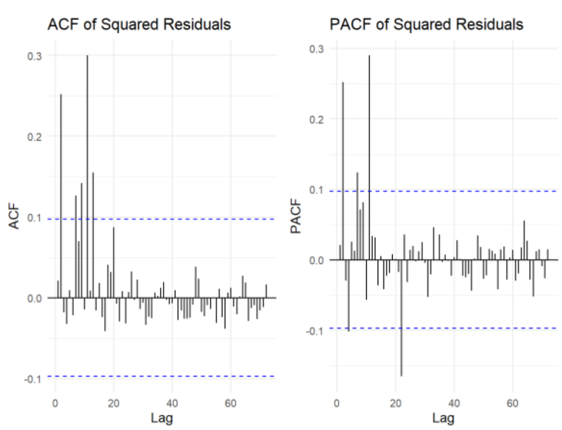

Figure 13: ACF-PACF Plot of Squared Residuals

From above plots we observe some lags that outside of the white noise band, this might cause heteroscedasticity problem. Let's check this with appropriate test.

<u>Arch Test:</u>

P-value: 2.445e-10

So, we can see from plots and the test result, our residuals have heteroscedasticity problem which is they do not have constant variance. To overcome this problem, we should apply GARCH model.

GARCH Model:

Conditional Variance Dynamics	
GARCH Model	sGARCH(1,1)
Mean Model	ARFIMA(3,0,0)
Distribution	norm

Table 7: Garch Model Output

We can see that we use SGarch model and model distribution is identified as rnorm.

	Statistic	Shape	Scale	P-Value
ARCH Lag[3]	0.01145	0.500	2.000	0.9148
ARCH Lag[5]	0.02144	1.440	1.667	0.9985
ARCH Lag[7]	0.04456	2.315	1.543	0.9999

Table 8: Weighted ARCH LM Tests

As we can see from the table, we solve the non-constant variance problem with GARCH (1,1) model.

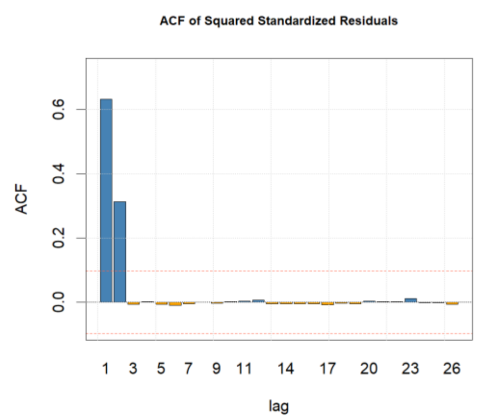

Figure 14: ACF of Squared residuals

Also, from the plot we can see that after GARCH, almost all the lags are inside of the white noise band, which is indication of non-heteroscedastic pattern.

<u>Jarque-Bera Test</u>: P-value is 2.2e-16

Garch model does not fix the normality, but we will assume that our residuals normally distributed.

8. FORECASTING

After these processes we will continue with forecasting. We will start with the ARIMA forecast which we get help from GARCH modeling to get rid of non-constant variance problem.

8.1 GARCH FORECASTING

Our ARIMA (3,1,0) model forecast from GARCH(1,1) plot is as follows.

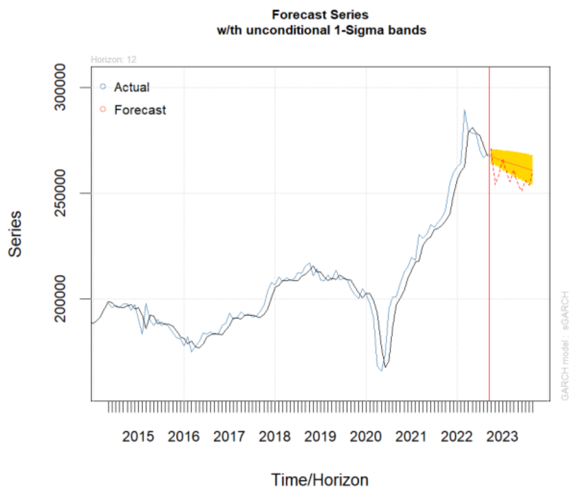

Figure 15: Forecast Plot of ARIMA (GARCH)

We can see that the black lines are the fitted values of model which are close the observed values at some points, also we can see the forecasted values are close to real values at some time points, but it did not capture the fluctuation. When we check our prediction intervals, they are narrow but the real observations outside of the intervals which is not a good indication.

We can see that in above plot we do not observe all the variables, they start from 2015, but when plotting the GARCH model it did not allow us to plot all of the variables, so we continue to visualize our model in this way.

8.2 ETS FORECASTING

ETS(M,A,N)		
Smoothing Parameters		
Alpha	0.7911	
Beta	0.2411	
Phi	0.8	
AIC	AICc	BIC
8906.914	8907.125	8930.937

Table 9: ETS Model Output

Our ETS model build on multiplicative trend and additive error components.

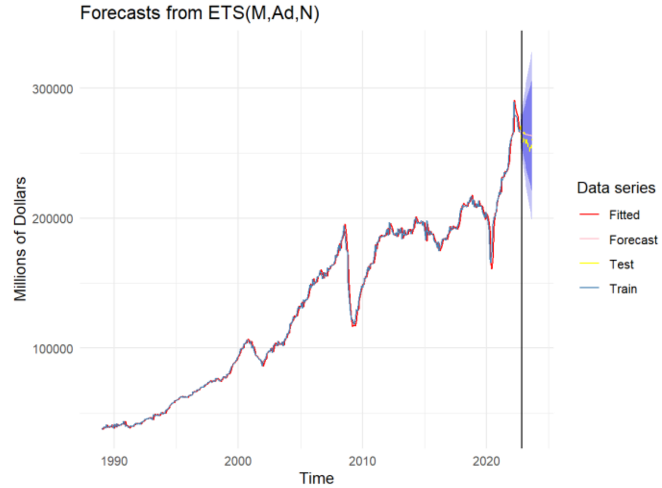

Figure 16: Forecast Plot of ETS

From above plot we can see that our fitted values and original data close to each other. Forecast values are close to real values at some points, but in general they are far away from them. Also, our prediction interval is highly wide when we look at the values.

<u>Jarque-Bera Test</u>: P-value is 2.2e-16

So, residuals of ETS model do not normally distributed.

8.3 NNETAR FORECASTING

Series: train
Model: NNAR(1,1,2)[12]
Call: nnetar(y = train, lambda = 0)
Average of 20 networks, each of which is a 2-2-1 network with 9 weights options were - linear output units
sigma^2 estimated as 0.000647

Table 10: Summary Table of NNETAR

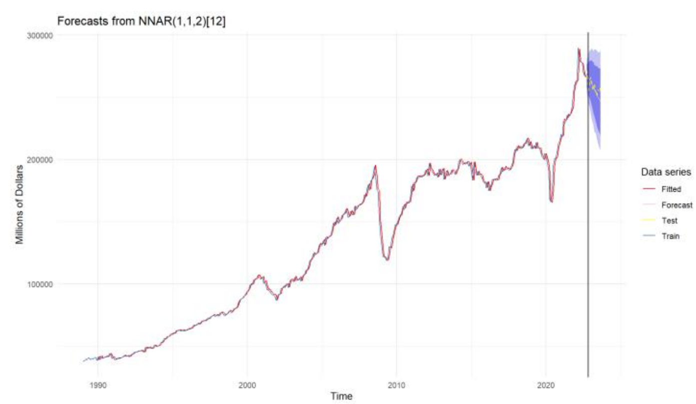

Figure 18: Forecast Plot of NNETAR

From above plot we can see that our fitted values and original data close to each other. Forecast values and test values are close to each other for most of the points, and it seems that forecast values catch the pattern of test values. Prediction intervals are seeming reasonable, they are not very wide and contains real values.

Jarque-Bera Test: P-value is 2.2e-16

So, residuals of NNETAR model do not normally distributed.

8.4 TBATS FORECASTING

TBATS(0.002, {0,3}, 1, -)
Call: tbats(y = train)
Parameters **Lambda: 0.002488** **Alpha: 1.603217** **Beta: 0.007593507** **Damping Parameter: 1** **MA coefficients: -0.630339 0.534197 -0.111417**

Table 11: Summary Table of TBATS

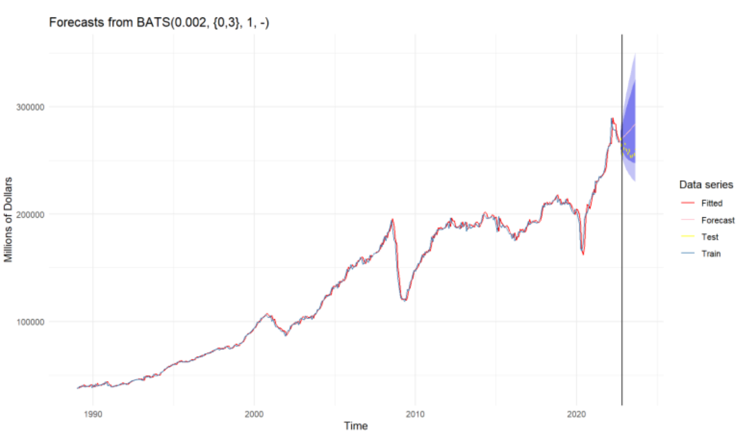

Figure 19: Forecast plot of TBATS

From above plot we can see that our fitted values and original data highly close to each other. When we look at the forecast values and test data, we can see that our forecast values are far away from the test values, also while test values are showing a pattern like going down, our forecast values showing an upward trend. When we look at the prediction interval it for 95 percent blue line is almost edge of the real values and they are respectively wide.

Jarque-Bera Test: P-value is 2.2e-16

So, residuals of TBATS model do not normally distributed.

8.5 PROPHET FORECASTING

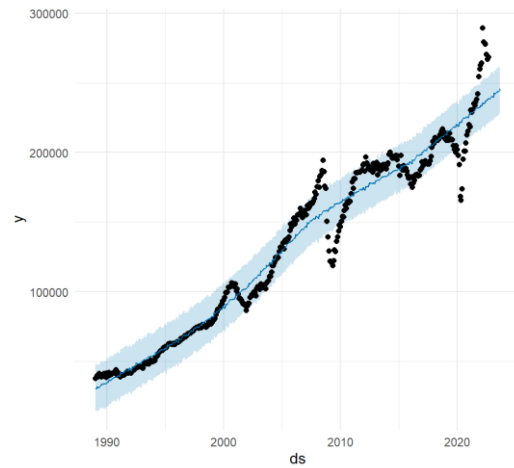

Figure 20: Forecast Plot of Prophet

In this plot block dots are representing the real values and blue line represent the predicted values. Predicted values are close to real values until 2000, but after that it did not catch the fluctuation of the data. Light-blue lines represent prediction intervals, and they are highly wide, this model does not show seem to show precise forecast.

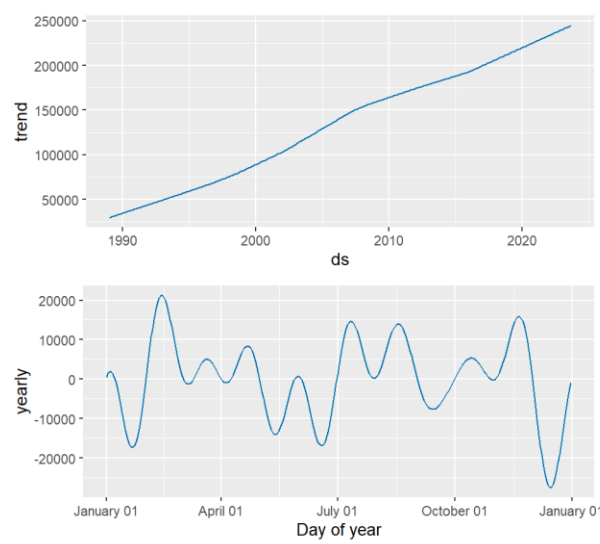

Figure 21: Trend and Seasonality

From above plot we can see the trend and seasonality components from prophet.

Jarque-Bera Test: P-value is 2.2e-16

So, residuals of Prophet model do not normally distributed.

9. MODEL SELECTION

TRAIN SET	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
NNETAR	47.50	3953.54	2430.09	-0.033	1.774	0.185	0.033
PROPHET	-1.359	13229.52	9214.769	-0.6028	7.011	-	0.951
TBATS	-211.3894	3848.676	2358.607	-0.1812	1.773	0.1803	0.0624
ETS	314.89	3883.74	2383.85	0.259	1.807	1.8222	0.051
GARCH	-110.4542	9178.769	3353.22	-	3.752	1.347	-

Table 12: Train Set Accuracy Measures

When we look at train accuracies, we can observe that TBATS model has the lowest accuracy measures. For TBATS model we can say that observed values and fitted values are close to each other, so TBATS is a good model to fitting the train set.

TEST SET	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
NNETAR	**1102.2**	**6691.995**	**5137.792**	**0.410**	**1.988**	**0.392**	**0.231**
PROPHET	221244.3	221337.5	221244.3	85.541	85.54	-	0.157
TBATS	-17967.90	20098.46	18392.95	-7.016	7.1729	1.406	0.434
ETS	-6056.26	7875.335	6962.084	-2.3836	2.718	0.5322	-0.1611
GARCH	5366.36	7267.37	6183.99	-	2.415	0.048	-

Table 13: Test Set Accuracy Measures

From above table we can see accuracy measures of different models. As we can see, some of the measures are not available for some models, but as we know when we are comparing models the most important values are RMSE and MAPE which we have these values for all the models.

When we try to compare test accuracy measurements, we can observe that NNETAR has the lowest RMSE, MAE, MPE and MAPE values. Therefore, we can say that NNETAR is the best model for predicting the dataset. We will choose the NNETAR model while forecasting.

10. DISCUSSION AND CONCLUSION

This study aimed that forecasting the U.S. Imports of Goods by Customs Basis from World between 1989-2023. The first step was showing how our data behaves and what are the summary statistics of our data. Then, we checked whether it has seasonality or trend components, and first we visually observed that our data has trend but not seasonality. After that, we divide our data as train and test set. Since our data is monthly data, we divide last 12 observations to test set. Then we checked the anomalies of our data with STL decomposition and remove the anomalies from the data. After that to stabilize the variance we applied box-cox transformation to our data which has the lambda value 0.1599589. To continue our analysis our series must be stationary, but when we check ACF/PACF plot of the data, KPSS test, and ADF test we observed that it has stochastic trend which means it is not stationary. To make it stationary we take one regular difference, after that we checked plots and applied the same tests, then we observed that it became stationary. After that we suggested some models by ACF&PACF plots, ESACF table, Auto-arima, and armaselect functions. Then, using t statistics and a BIC comparison, the optimal model with the fewest and most significant parameters was selected from these preliminary models. After we decide that best model is ARIMA (3,1,0), we go through the diagnostic checking part. At this point, we observed that our residuals are not normally distributed, our residuals do not have autocorrelation, and residuals do not have constant variance which implies there is a heteroscedasticity problem. To solve this problem, we applied GARCH model with parameters (1,1), after that we observed that our constant variance problem solved, and we continue to forecast from this model. After ARIMA forecasting we obtain forecast from 4 different models which are NNETAR, TBATS, PROPHET, and ETS. To check the normality of these models we applied Jarque Bera test to each of them, and the result was same for all of them which says that they are not showing normality. Also, we interpret forecast plots of each model, and we checked their summary tables. Finally, we compared accuracy measurements of different models, and we decide that NNETAR is the best model to forecast the dataset since it has lowest test set accuracy measures.

11. REFERENCES

U.S. Imports of Goods by Customs Basis from World. (2024, January 9). Https://fred.stlouisfed.org/series/IMP0004