

**Progress Report**

**GDP growth Analysis**

Prof., Yazan Roumani | QMM-5520 Forecasting | Group 1

**BY**

Ashreetha Udhayasankar

**Index**

1. **[Introduction](#_Introduction)**
2. [**Data Collection**](#_Data_Collection_1)
3. [**Graph and Summary Statistics**](#_Graph_and_Summary)
4. **[Data Analysis](#_Data_Analysis)**
5. **[Discussion and Conclusion](#_Discussion_and_Conclusion)**

# **Introduction**

This study develops a framework to forecast US GDP (Gross Domestic Product) growth on a quarterly frequency from 1947 – 2022. Real exports of Goods and Services manipulated and released by the U.S. Bureau of Economic Analysis – based on concepts consistent with exports of real Gross Domestic Product — serve as an important tool for measuring the real value of export goods that is adjusted for the effects of price changes.

This time-series data on real exports of Goods and Services are released in the “Research Data” section on the FRED Economic Data website and widely used by those who follow the state of the financial economy (Figure 1).

The models that we use here are based on real sector of the US economy, are forecasted, and estimated by using Holt’ Exponential Smoothing model, Damped Trend Exponential Smoothing model, Linear Regression model (Simple Linear Regression model), ARIMA model and more.

This study seeks to develop an appropriate forecasting model to predict US gross domestic product (GDP) growth for the next few years (5-8 years). We are well known that the economic forecasting of US has historically been in a critical fraught situation. The main objective of this paper is to design a framework that accurately observes and captures the performing and underlying economic structure of US and precise variables that provides information about the indicators.

We would construct a quarterly forecasting framework with given variables as the key variables, with based on economic theory on the real exports of goods and services. By using the different forecasting models, we would be arriving to a conclusion with best model fit by analyzing the MAPE (Mean Absolute Percentage Error) values generated by each model.

Thus, we would be predicting the values by considering the MAPE values of each model fit and will forecast the values for next set of years. We would choose the most effective forecasting model based on estimation of the forecasted values and evaluating the predictive ability of various multivariate models.

# **Data Collection**

The Data source that we use for this paper is obtained from U.S. Bureau of Economic Analysis. It is a source for accurate and objective data. Bureau of Economic Analysis produce some of the world's most closely watched statistics, including U.S. gross domestic product, better known as **GDP.**

For our study we have used the Real Exports of Goods and Services (EXPGSC1) dataset. We would be having the EXPGSC.xls file as the dataset. Also, they have provided a plot for time-series data on real exports of Goods and Services. (Figure 1)

Figure 1. Real Exports for Goods and Services

Chart

Description automatically generated

The above observation is obtained between the period (1947 – 2022) with Date (in years) and Billions of Chained 2012 Dollars, Seasonally Adjusted Annual Rate. It is distributed in Quarterly frequency distribution.

# **Graph and Summary Statistics**

Figure 1. Time series for given Data

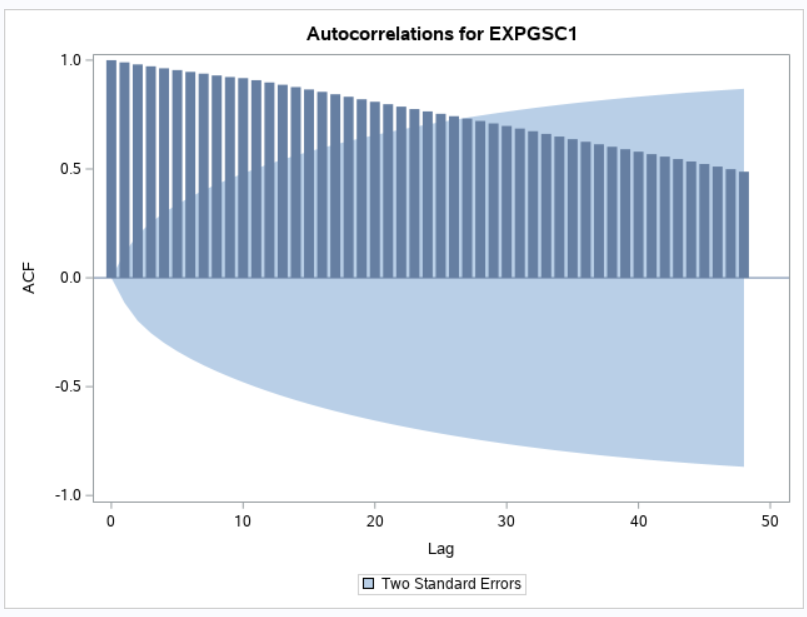
From the time series plot, we can recognize there is an overall positive trend. As there is steady Increase of the exports of goods and services. There is an increase in spike in Exports of goods and services indicates presence of cyclical components.

Figure 2. ACF Plot for given Data

From the ACF plot we can say that the plot indicates a trend component because the autocorrelations are not declining quickly towards zero.

# **Data Analysis**

Due to presence of Trend and Cyclical components, appropriate models would be as follow:

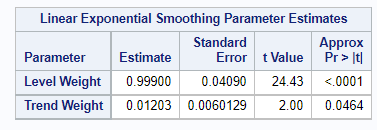
**Holt’s Exponential Smoothing Model:**

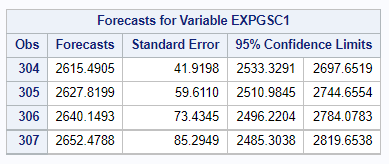
Holt’s Exponential smoothing uses exponential smoothing which will encode large number of values from the past and with the use of this model, we would predict the “typical” values for future. This method is best for data with trend that increases over time. It results in a curved forecast that reproduces the trend changes in the data.

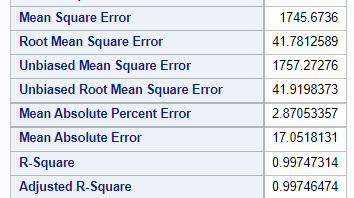
From the ACF plot we can say that the plot indicates a trend component because the autocorrelations are not declining quickly towards zero.

Due to the presence of a trend and possibly cycle components, Holt’s exponential smoothing model would be appropriate.

Parameter Estimates values for Holt’s Exponential



The forecast value for next 4 quarter is as follow:



The level smoothing parameter is 0.99 which indicates that more weight is assigned to the most recent observations. The trend smoothing parameter is 0.012 which indicates that the slope of the time series is easily changing. The model fits the data well as indicated by the MAPE (2.87) and RMSE (41.78)

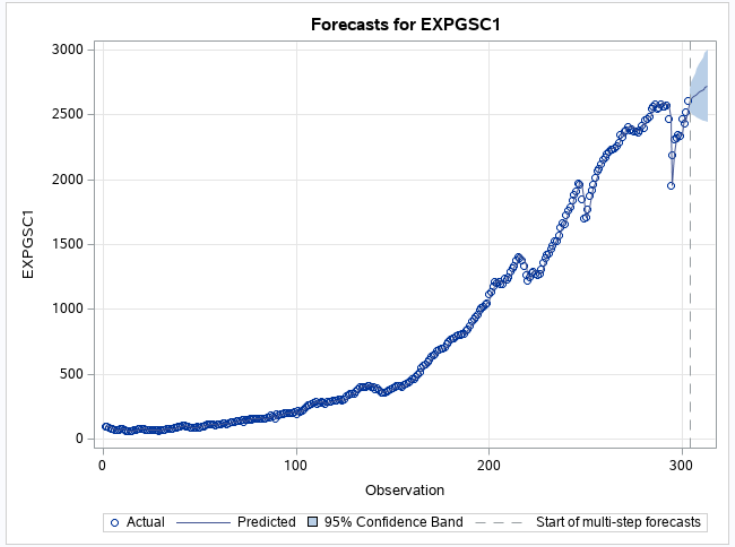
**Damped Trend Exponential Smoothing Model:**

The damped trend method of exponential smoothing is a benchmark that has been difficult to beat in empirical studies of forecast accuracy. It is mostly used for trend components.



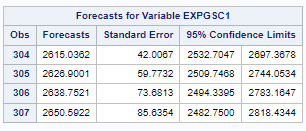
The level smoothing parameter is 0.99 which indicates that more weight is assigned to the most recent observations. The trend smoothing parameter is 0.01 which indicates that the slope of the time series is hardly changing. The damping smoothing parameter is 0.99 which indicates that Slight damping is assigned to observations.

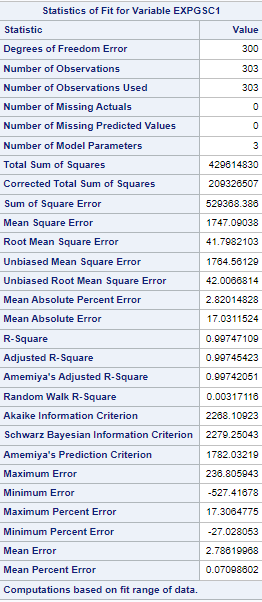
Scatter plot of forecasted values for next 4 quarters:



The above plots shows that the forecasts for next 4 quarters EXPGSC1 has an overall positive trend component.

Forecasted values for next 4 quarters:

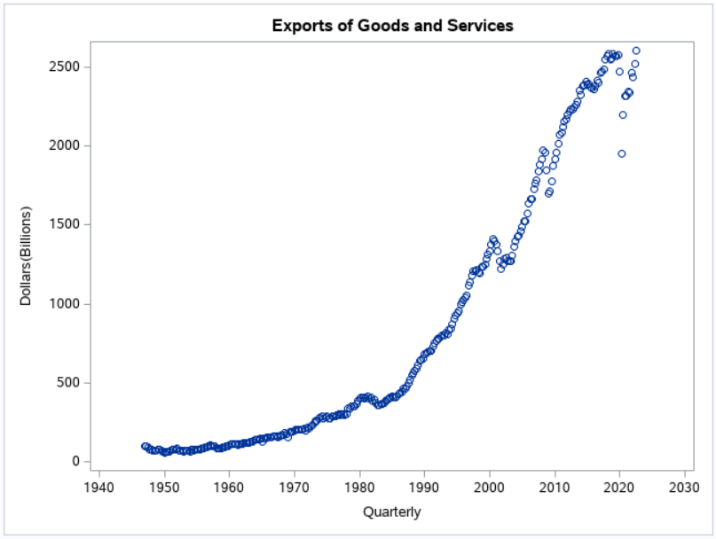
Statistics fit variables for EXPGSC1:

****

MAPE: 2.82, RMSE: 41.8.

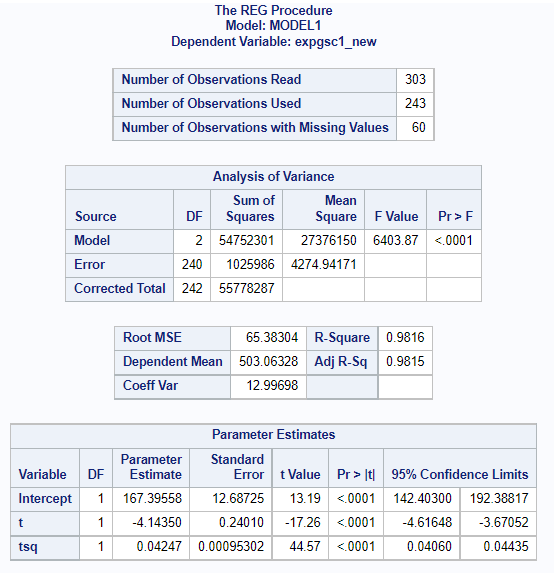
**Nonlinear regression Model:**

In this model, to evaluate the performance of the model, we split the dataset into two subset, train and test set 80-20%. There are total 303 dataset, So 243 observation used as a train set and 60 observation for test set.

Step: 1 Scatter Plot of the data

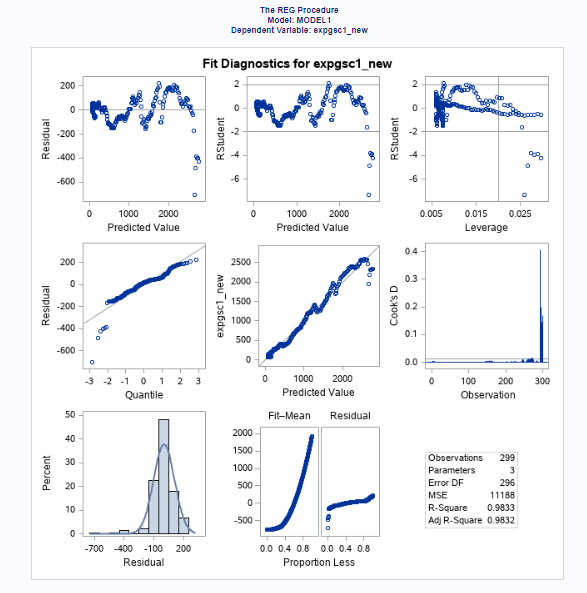
Above plot shows the positive relation between dependent variable and independent variable. And in this plot the graph line is not straight, it is curved line so nonlinear model would be appropriate.

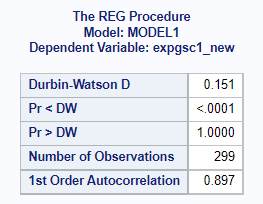
Step: 2 Evaluate the Model:



The relationship between intercept and slope, make sense with positive linearity. The P-value for tsq is very small and less than α, which means that the relation between dependent variable and independent variable is non-linear in the model. Intercept and slope values are in between confidence limits, so the slope is significant. R2 Value is 98% which is showing high level of correlation and it tells that variability in export value, is actually explaining about 98% of the variability in time. So Parameters are statistically significant here.

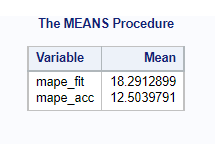
Step: 3 The Model assumption:





Residuals are bell shaped and symmetric so data looks normally distributed. Residual -> Predicted value plot shows random pattern which show assumption is true. QQ plot shows that most of the data points are close to a line so data is normally distributed. DW test shows positive autocorrelation.

Step: 4 Calculate MAPE of model fit:



RMSE: 105.8

Here, MAPE model accuracy is lower than MAPE model fit, so it might shows above created model is able to make accurate predictions on testing data.

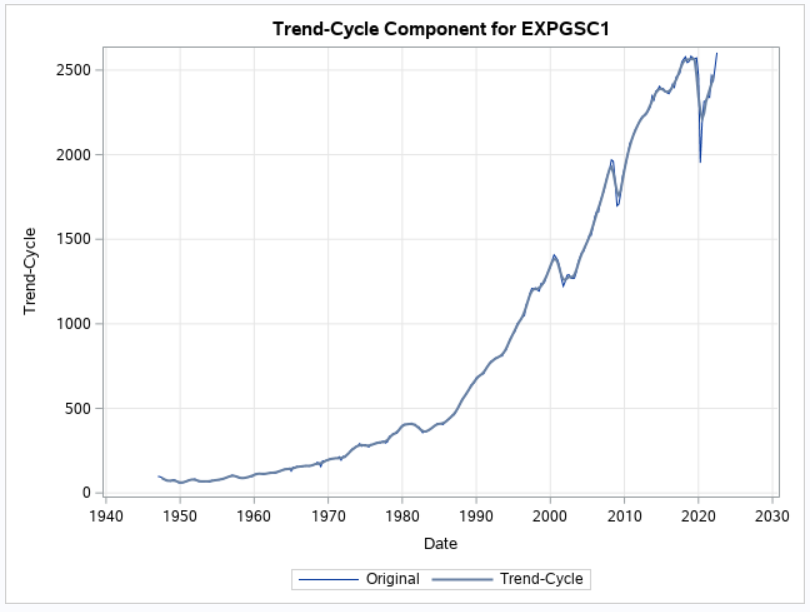
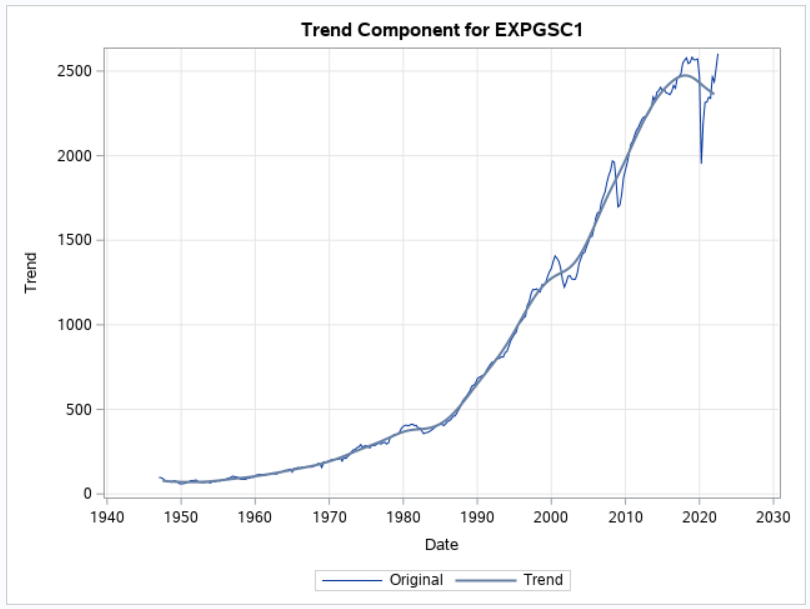
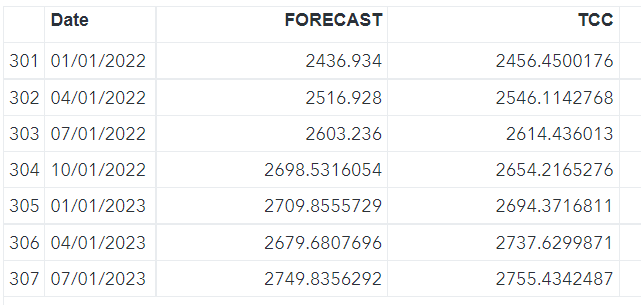
**X11 Decomposition Model:**

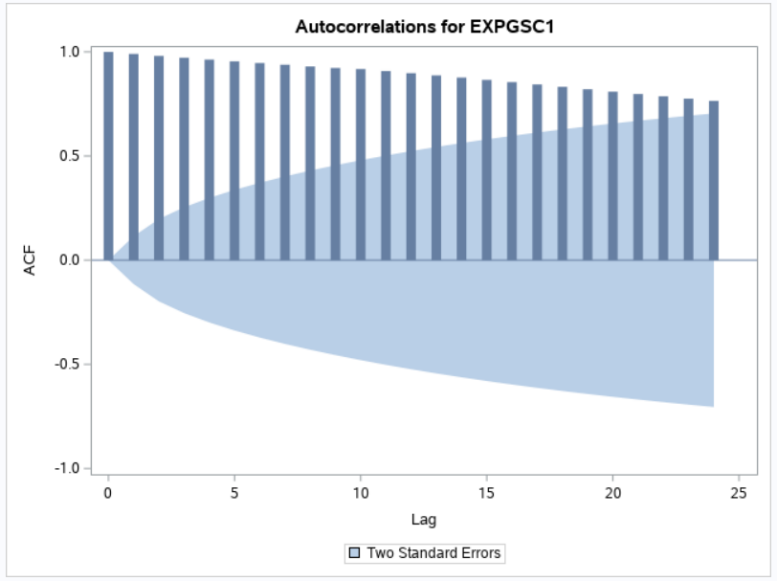
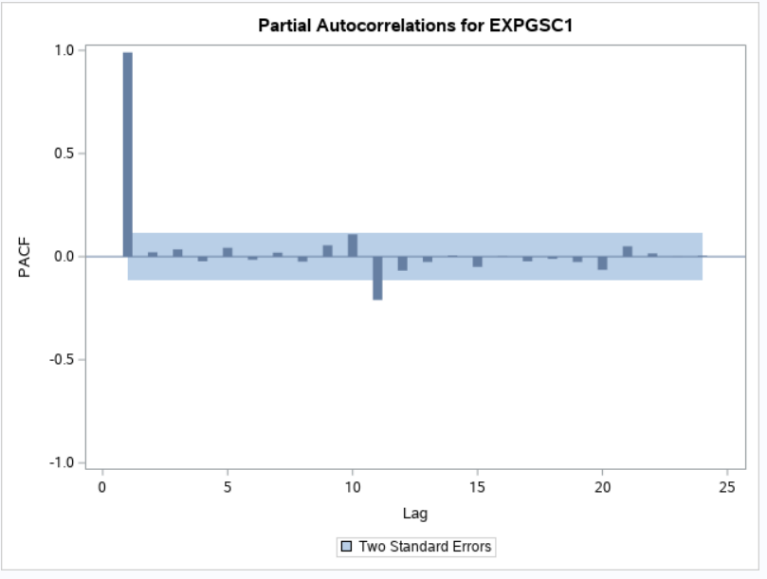
Figure: 1.2 the original data (blue), the trend component

Figure: 1.1 the original data (blue), the trend-cycle component

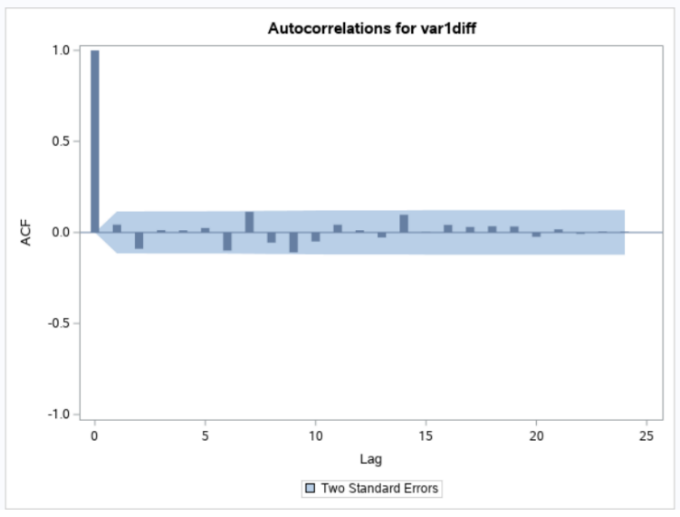
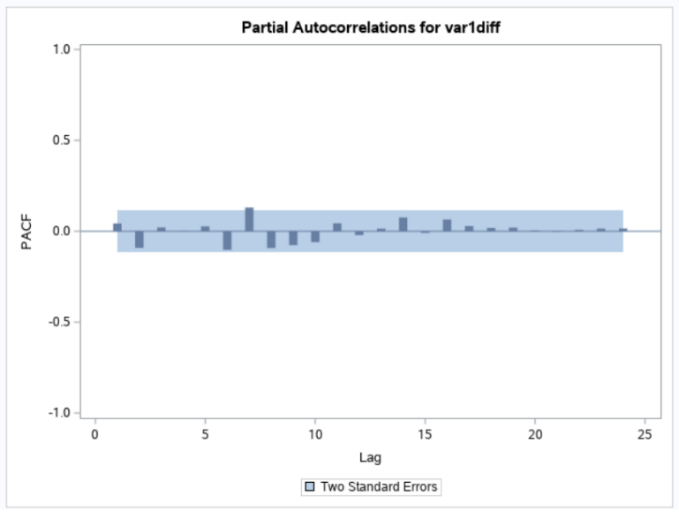
It can be useful to plot trend and cycle component. These help to visualize the variation in the trend component over time. Above two plots shows that X11 decomposition, can capture the cycle and trend components very well. And the forecast value for next 4 quarter is as follow:



**ARIMA Model:**

Step: 1 - Plot the data (ACF, PACF)

ACF and PACF plot shows that given data is non stationary and differencing method is needed.

Step: 2 – Apply differencing method and plot the data (ACF, PACF)

Above both plots shows the data is white noise and there is no pattern in data that ARIMA model can capture.

# **Discussion and Conclusion**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **RMSE** | **MAPE** | **AIC** | **BIC** |
| Holt’s Exponential Smoothing | **41.0** | **2.90** | **1305.4** | **1312.4** |
| Damped Trend Exponential smoothing | **41.0** | **2.90** | **1305.0** | **1315.5** |
| Non Linear regression model | **105.8** | **18.3** | **2034.6** | **2036.7** |

[Figure: 1 - Comparison between Holt’s exponential, Damped Trend Exponential and Nonlinear regression model]

Figure 1 summarizes the performance of the various models. RMSE, MAPE, AIC and BIC values for Holt’s and Damped trend model are lower than nonlinear regression model. Which indicates that best fit model for our data is Holt’s or Damped trend model. Both models are significant in terms of all measurement, there is no much difference in AIC and BIC values for those two models so, according to parsimony principle, Holt’s exponential smoothing model is best fit for given real exports of Goods and Services dataset.

|  |  |  |
| --- | --- | --- |
| **Model Fit** | | |
|  | **Holt’s Exponential smoothing** | **Damped trend method** |
| **RMSE** | **41.2** | **41.2** |
| **MAPE** | **2.90** | **2.82** |
| **Model Accuracy** | | |
|  | **Holt’s Exponential smoothing** | **Damped trend method** |
| **RMSE** | **152.73** | **154.55** |
| **MAPE** | **5.65** | **5.75** |

[Figure: 2 - Comparison between models fit and model accuracy between Holt’s Exponential and Damped Trend Exponential model]

Based on the measures, showed in figure 2 we can say that Holt’s Exponential smoothing model is better because it generates less MAPE for Model accuracy so it will generate more accurate forecast for analysis.

# **Future Scope And Ideas**

The forecasted values from different forecasting methodologies have suited well to the given dataset and it works in a spectacular way. Although the study has some caveats such as, when we use some datasets with huge amounts of data, we should make our model to be fit in that scenario as well. Hence, for the benefit of data users, it may be useful for compiling agencies to provide sufficient background information on the nature of the revisions that they have done to past data when disseminating economic statistics such as GDP as this would facilitate and improve the analysis and use of the statistics.

# **References**

[1] FRED – Economic data, Federal Reserve Bank of St. Louis Retrieved from <https://fred.stlouisfed.org/series/EXPGSC1>.

**SAS Code:**

**/\* Final Project \*/**

**/\*Import data file\*/**

**proc import out=proj datafile="/home/u59596291/sasuser.v94/Data/Project.xls"**

**dbms=xls replace;**

**run;**

**/\* Creating a timeseries plot \*/**

**proc sgplot data=proj;**

**series x=date y=expgsc1;**

**run;**

**/\* Creating an ACF plot \*/**

**proc timeseries data=proj plots=acf out=\_null\_;**

**var expgsc1; /\* y variable - dependetnt variable\*/**

**corr acf/nlag=48;**

**run;**

**/\* Holt's model \*/**

**proc esm data=proj plot = forecasts print=all outfor=projout out=\_null\_ lead=4 back=4;**

**forecast expgsc1/model=linear;**

**run;**

**/\*Damp trend method\*/**

**proc esm data=proj plot = forecasts print=all outfor=airpout1 out=\_null\_ lead=4 back=4;**

**forecast expgsc1/model=damptrend;**

**run;**

**/\* Non linear regression model \*/**

**proc sgplot data=proj;**

**scatter x=date y=expgsc1;**

**title "Exports of Goods and Services";**

**xaxis label="Quarterly";**

**yaxis lable="Dollars(Billions)";**

**run;**

**proc timeseries data=proj plots=acf out=\_null\_;**

**var expgsc1; /\* checking trend in y variable \*/**

**corr acf/nlag=36;**

**run;**

**/\* Using first 243 observation \*/**

**data proj1;**

**set proj;**

**t=\_n\_; /\* observation number \*/**

**tsq=t\*t;**

**expgsc1\_new=expgsc1;**

**if t>243 then expgsc1\_new=.;**

**run;**

**proc reg data=proj1 outest=projout11;**

**model expgsc1\_new=t tsq/dwprob clb aic bic adjrsq;**

**output out=projout predicted=projhat residual=projres;**

**run;**

**proc reg data=proj1;**

**model expgsc1\_new=t tsq;**

**output out=projout predicted=projhat residual=projres;**

**run;**

**/\* Calcualting MAPE for model fit and model accuracy \*/**

**data projout1;**

**set projout;**

**mape\_fit=(abs(projres)/expgsc1\_new)\*100; /\* residual is equal to difference between original value and pridected value \*/**

**if t>243 then mape\_acc=(abs(projhat-expgsc1)/expgsc1)\*100; /\* this row only calculate mape for remaining last four rows\*/**

**run;**

**proc means data=projout1 mean;**

**var mape\_fit mape\_acc;**

**run;**

**/\*Classical Decomposition Method\*/**

**proc import out=proj datafile="/home/u59596291/sasuser.v94/Data/Project.xls"**

**dbms=xls replace;**

**run;**

**proc timeseries data=proj plots=(tc cc tcc ic) outdecomp=decomp1 printdetails;**

**id date interval=qtr;**

**var expgsc1;**

**decomp orig tc tcc cc ic;**

**run;**

**/\* X11 Decomposition Method \*/**

**proc import out=proj datafile="/home/u59596291/sasuser.v94/Data/Project.xls"**

**dbms=xls replace;**

**run;**

**proc x11 data=proj noprint outextrap;**

**quarterly date=date;**

**var expgsc1;**

**arima forecast=1;**

**output out=proj\_out a1=original d12=tcc a15=forecast;**

**run;**

**data merged;**

**merge proj\_out proj;**

**by date;**

**run;**

**proc sgplot data=merged;**

**series x=date y=expgsc1;**

**series x=date y=forecast;**

**proc sgplot data=merged;**

**series x=date y=expgsc1;**

**series x=date y=tcc;**

**proc print data=merged; var forecast; run;**

**/\*ARIMA\*/**

**proc timeseries data=proj plots=(acf pacf) out=\_null\_;**

**var expgsc1;**

**run;**

**data proj ;**

**set proj ;**

**var1diff=dif(expgsc1);/\*help to make stationary\*/**

**run;**

**proc timeseries data=proj plots=(acf pacf) out=\_null\_;**

**var var1diff;**

**run;**

**proc arima data=proj;**

**identify var=expgsc1(1)nlag=36 whitenoise=ignoremiss; /\*Ljung box test apply- data is not white noise\*/**

**/\*first order differecnces\*/**

**run;**