Final Group Project – BAN 673

US Gross Domestic Product Measure

MSBA Fall 22, CSU East Bay

BAN 673

Megh Dave (CC3118)

Swati Sharma (XN2486)

Ruhi Sania (LT7551)

Geethika Verma Mudunuri (DK5968)

Krutika Deshpande (LA8465)

Executive Summary

This project analyzes the quarterly gross domestic product (GDP) of the United States for the period from 2003 to 2022(obtained from Kaggle Website). The aim of the analysis is to build a model that can accurately predict future GDP values based on past trends.

The data was preprocessed, and visualized to identify patterns and trends. The analysis showed an overall upward trend with seasonal fluctuations, with the third quarter typically having the highest GDP values and the first quarter having the lowest.

Three different time series models were implemented in the analysis:

1. ARIMA (Autoregressive Integrated Moving Average) models
2. Holt-Winters Model
3. Auto regressive models
4. Moving average Model

The models were optimized by tuning hyperparameters and evaluating their performance based on the mean absolute percentage error (MAPE) and root mean squared error (RMSE).

The results indicated that the ***Auto ARIMA*** provided the best fit for the data. The model was then used to predict future GDP values for the next four quarters, and the results were found to be within a reasonable range of accuracy.

Overall, the analysis suggests that the ***Auto ARIMA*** can be an effective tool for predicting quarterly GDP values in the United States.

Introduction

Gross Domestic Product (GDP) is a critical measure of economic activity and growth in any country. The GDP of the United States is the largest in the world and a key driver of global economic trends. The data available in this project is the modeled quarterly GDP data for the United States from 2003 to 2022.

The United States GDP has consistently grown over the past decades and is valued at over $21 trillion as of 2020, which is approximately 25% of the global GDP. GDP is composed of several sectors, including agriculture, industry, and services.

Therefore, GDP data provides valuable insight into the performance of these sectors and the overall health of the economy. This project aims to analyze and forecast the US GDP using various time series models, taking into account trends, seasonality, and other factors that influence economic growth. The results of this analysis can help policymakers, investors, and businesses in making informed decisions and anticipating future economic trends.

Eight Steps of Forecasting

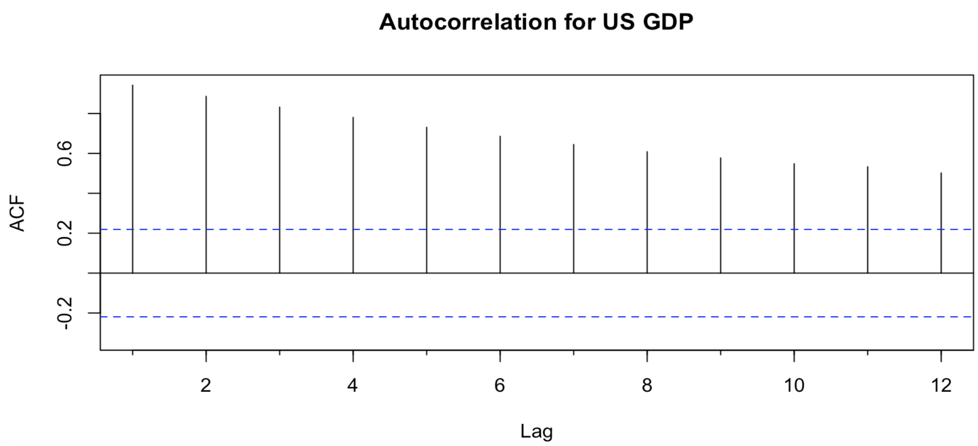
Step 1: Define the Goal

The goal of this project is to create forecasts of the US GDP values for the next 2 years(8 quarters). The aim is to develop a prediction model that can effectively forecast desired quarters by taking into account both the trend and seasonal components of historical data. Clearly, the model of preference will be the one with the maximum accuracy. The forecasting model that is preferred should be analyzed semi-annually in order to incorporate two new quarters when predicting into the future because data for the quarter will be attainable every three months. The forecasting models developed for this project were done in R language.

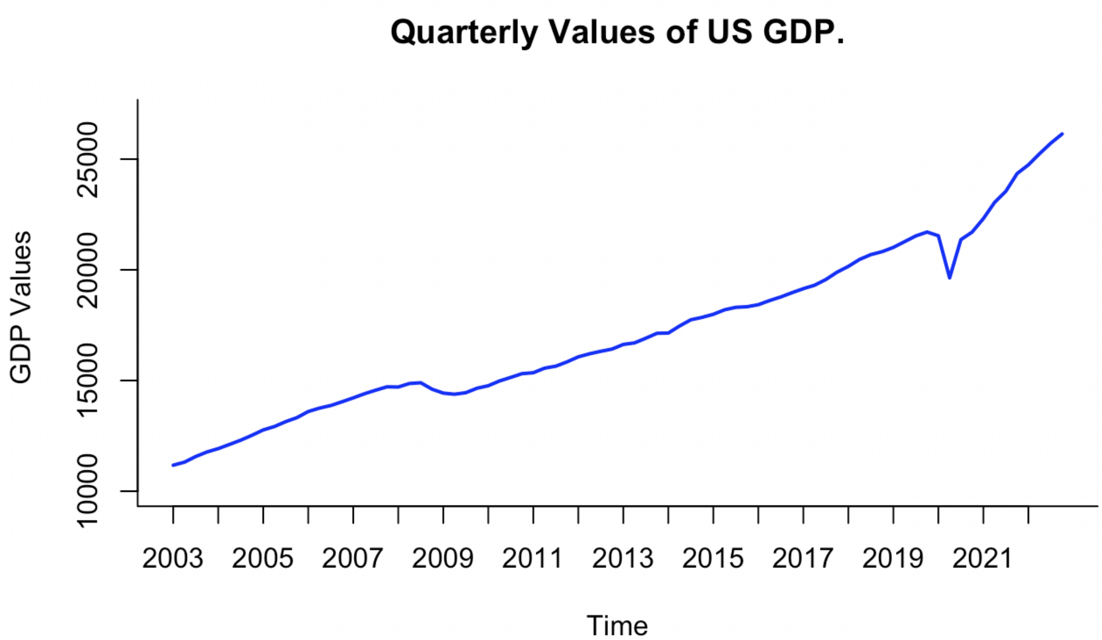
Step 2: Get Data

This report will focus on a time series dataset provided by the Federal Reserve Economic Data (FRED) system, containing quarterly measurements of the seasonally adjusted Gross Domestic Product (GDP) of the United States, measured in billions of dollars. The time period for the dataset ranges from 1947 to 2022, and we will be analyzing the period from 2003 to 2022 (a 20-year series) for the purpose of this project.

Step 3: Explore and Visualize Series



The correlogram for the US GDP time series data with lags 1 to 12 is shown above. We can observe a positive autocorrelation coefficient for all the lags indicating the data is highly correlated.A positive autocorrelation coefficient for lag 1 is substantially higher than the horizontal threshold (significantly greater than zero) showing an upward trend component.A positive autocorrelation coefficient in lag 12, which is also statistically significant (greater than zero), points to monthly seasonality.From the correlogram we can say that it is not just compromised of a level component.



The analysis of the quarterly values of US GDP reveals a consistent increasing trend over the years, indicative of a growing economy. However, the year 2020 stands out as an exception, marked by a slight dip in GDP figures. This decline can be attributed to the unprecedented impact of the COVID-19 pandemic, which caused widespread disruptions across various sectors of the economy. Nevertheless, there has been a noteworthy recovery and resurgence in economic performance since 2021. The GDP figures have shown a significant upward trajectory, indicating a robust rebound and a return to growth.

Step 4: Data Preprocessing

The original dataset provided GDP figures dating back to the year 1947; however, it contained missing values in some quarters. Additionally, considering the changing economic landscape and the focus on recent trends, we decided to focus our analysis on the past 20 years, specifically from 2003 Q1 to 2022 Q4. This served as the foundation for our subsequent analyses and modeling efforts.

Step 5: Partition Series

The data has been partitioned into 65 records for the training period and 16 records for validation period. These partitioned validation and training data sets (2003-2022) are shown below:

Training partition:(2003-2018)

```
> train.ts
         Qtr1     Qtr2     Qtr3     Qtr4
2003 11174.13 11312.77 11566.67 11772.23
2004 11923.45 12112.82 12305.31 12527.21
2005 12767.29 12922.66 13142.64 13324.20
2006 13599.16 13753.42 13870.19 14039.56
2007 14215.65 14402.08 14564.12 14715.06
2008 14706.54 14865.70 14899.00 14608.21
2009 14430.90 14381.24 14448.88 14651.25
2010 14764.61 14980.19 15141.60 15309.47
2011 15351.44 15557.53 15647.68 15842.27
2012 16068.82 16207.13 16319.54 16420.39
2013 16629.05 16699.55 16911.07 17133.11
2014 17144.28 17462.70 17743.23 17852.54
2015 17991.35 18193.71 18306.96 18332.08
2016 18425.31 18611.62 18775.46 18968.04
2017 19148.19 19304.51 19561.90 19894.75
2018 20155.49 20470.20 20687.28 20819.27
```

Validation partition:(2019-2022)

```
> valid.ts
         Qtr1     Qtr2     Qtr3     Qtr4
2019 21013.08 21272.45 21531.84 21706.53
2020 21538.03 19636.73 21362.43 21704.71
2021 22313.85 23046.93 23550.42 24349.12
2022 24740.48 25248.48 25723.94 26137.99
```

Step 6 & 7: Apply Forecasting & Comparing Performance

Trailing ma models

The accuracies we obtained from applying Trailing MA model with window with 4,5,8,10 and 12 have been shown below:

Trailing MA with Windows Width of 4:

```
              ME     RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 287.604 1058.561 822.811 0.946 3.554 0.685     1.249
```

Trailing MA with Windows Width of 5:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	261.849	1005.612	795.972	0.859	3.46	0.669	1.193

Trailing MA with Windows Width of 8:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	832.674	1319.468	1104.078	3.349	4.709	0.687	1.503

Trailing MA with Windows Width of 10:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	1488.33	1962.338	1654.409	6.134	6.98	0.74	2.218

Trailing MA with Windows Width of 12:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	2052.046	2438.097	2149.795	8.604	9.102	0.746	2.772

Comparing the MAPE and RMSE of the 5 forecasts for the validation period – trailing MA with windows width of 4,5,8,10 and 12 – we can see that the most accurate forecast here is the trailing MA forecast with window width of 5, which has the lowest MAPE of 3.46% and the lowest RMSE of 1005.612.

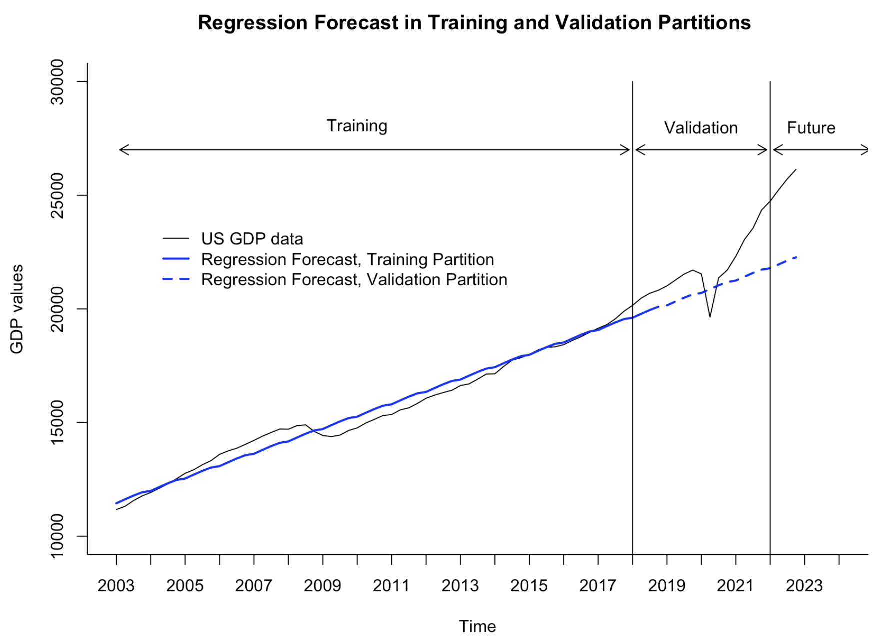
<u>Regression model with linear trend and seasonality</u>

```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-603.07 -309.08  -58.79  314.18  737.73 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11313.608    129.347  87.467   <2e-16 ***
trend         136.044      2.685  50.671   <2e-16 ***
season2        35.341    140.050   0.252    0.802    
season3        65.153    140.127   0.465    0.644    
season4        73.992    140.256   0.528    0.600    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 396 on 59 degrees of freedom
Multiple R-squared:  0.9777,	Adjusted R-squared:  0.9761 
F-statistic: 645.2 on 4 and 59 DF,  p-value: < 2.2e-16
```

The regression model contains 12 predictors including the trend *(t)* and 11 seasonal dummy variables (*D2 – D11*) for February (season 2) through December (season12). The regression equation is:

$$y^t = 181.80 + 2.04\, t - 39.41\, D2 - 69.25\, D3 + \ldots + 161.47\, D12$$

The plot below shows the linear regression line being fit through the training and validation data.



Two- level forecast by combining regression model and trailing ma for residuals

Regression Model with Linear Trend and Seasonality:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	0	798.117	627.139	-0.049	3.594	0.842	2.14

Trailing MA for residuals:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	17381.17	17755.94	17381.17	100.229	100.229	0.956	51.612

Two-Level Model – Regression and Trailing MA for Residuals:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	30.272	242.62	144.379	0.096	0.782	0.379	0.614

Out of the three forecasting models above in the forecast model, the most accurate model is the two-level model, which produced the lowest values for MAPE (0.782%) and RMSE (242.62).

<u>Holt-Winter's model</u>

Next, we used Holt-Winter's exponential smoothing model (ZZZ) with automatic selection of error, trend and seasonality options. Presented below is the summary of the model:

```
ETS(M,Ad,N)

Call:
 ets(y = train.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.9999
    beta  = 0.4895
    phi   = 0.9556

  Initial states:
    l = 10954.6743
    b = 222.3858

  sigma:  0.0063

     AIC     AICc      BIC
859.8876 861.3613 872.8409
```

The output shows the result of applying the Holt-Winter's exponential smoothing method to the training dataset using the ets() function in R with model = "ZZZ". This model selection option enables the automatic selection of error, trend, and seasonality options. The output shows the model type as ETS(M,Ad,N).

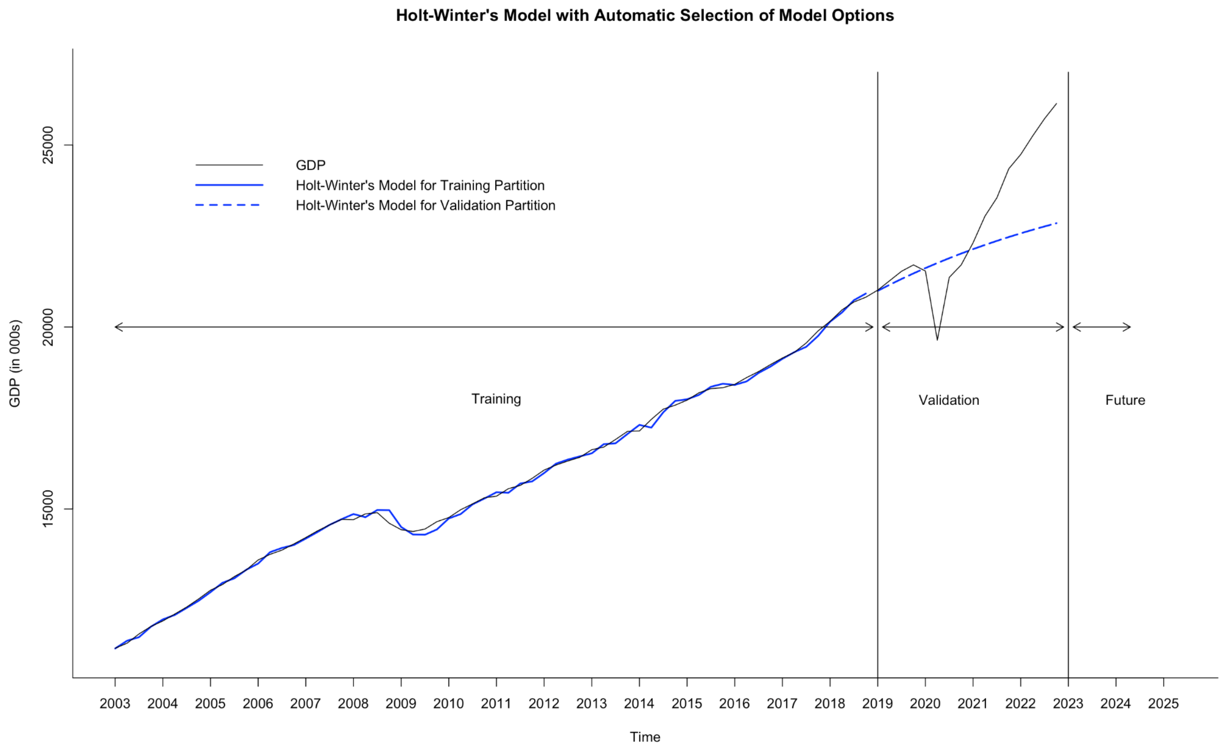
The output also shows the smoothing parameters for the model. The alpha value is 0.9999, which indicates that the method heavily weighs the most recent observation in the time series. The beta value is 0.4895, which indicates that the method assigns moderate importance to the trend component of the time series. The phi value is 0.9556, which indicates that the method assigns high importance to the seasonal component of the time series.

The output also shows the initial states of the model, which are the level and trend estimates at the beginning of the time series. The level estimate is 10954.6743, and the trend estimate is 222.3858. The sigma value is 0.0063, which represents the residual standard deviation.

Lastly, the output shows the information criteria values for the model, which can be used for model selection. The AIC, AICc, and BIC values are 859.8876, 861.3613, and 872.8409, respectively. Lower values of these criteria indicate a better model fit.

Forecast based on the Holt-Winters model (ETS(M,Ad,N)) that was fitted to the validation data:

	Point Forecast	Lo 0	Hi 0
2019 Q1	20993.95	20993.95	20993.95
2019 Q2	21160.88	21160.88	21160.88
2019 Q3	21320.39	21320.39	21320.39
2019 Q4	21472.83	21472.83	21472.83
2020 Q1	21618.50	21618.50	21618.50
2020 Q2	21757.70	21757.70	21757.70
2020 Q3	21890.73	21890.73	21890.73
2020 Q4	22017.86	22017.86	22017.86
2021 Q1	22139.34	22139.34	22139.34
2021 Q2	22255.44	22255.44	22255.44
2021 Q3	22366.38	22366.38	22366.38
2021 Q4	22472.40	22472.40	22472.40
2022 Q1	22573.71	22573.71	22573.71
2022 Q2	22670.53	22670.53	22670.53
2022 Q3	22763.05	22763.05	22763.05
2022 Q4	22851.46	22851.46	22851.46



According to the forecast, the time series is expected to continue to increase over the validation period, with quarterly values ranging from 20993.95 to 22851.46. However, it's important to note that the prediction interval is quite narrow, indicating a high level of confidence in these forecasts. Therefore the model may be a good fit for the data.

Using HW exponential smoothing - ZZZ model for the entire GDP data:

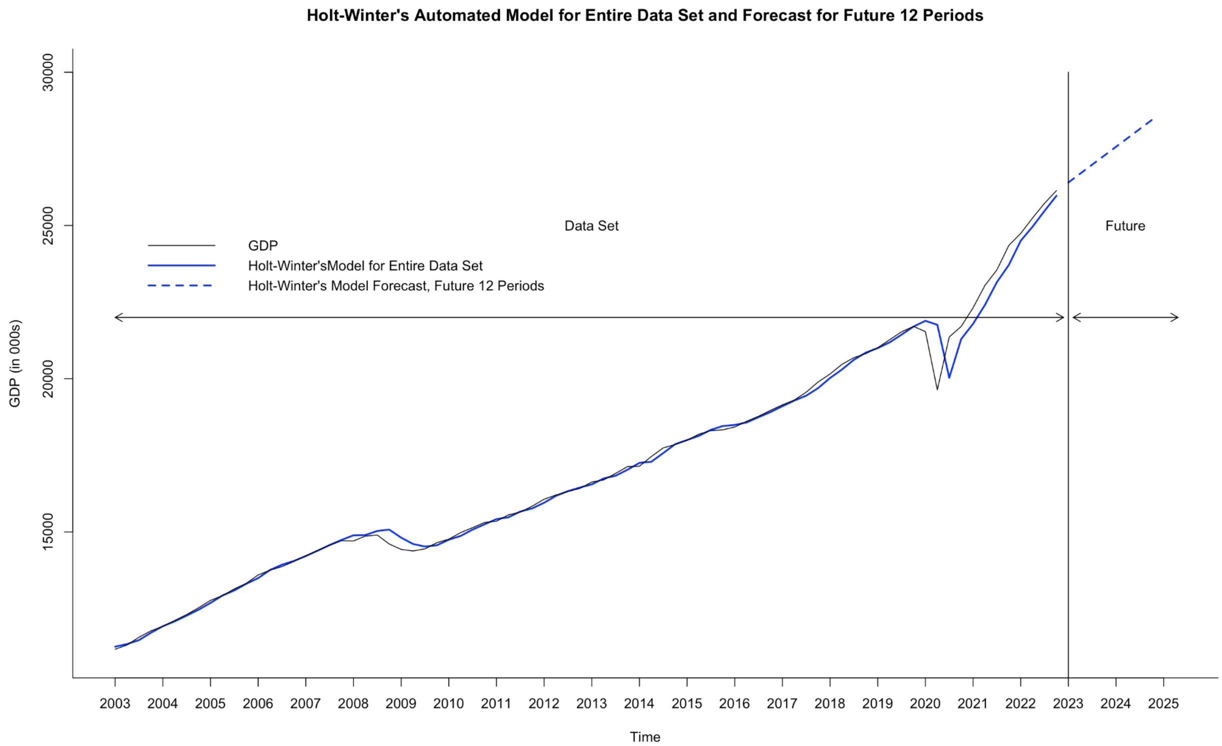
```
ETS(M,A,N)

Call:
 ets(y = gdp.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.8486
    beta  = 0.0447

  Initial states:
    l = 11098.4053
    b = 160.083

  sigma:  0.0165

     AIC     AICc      BIC
1256.392 1257.203 1268.303
```

The model appears to be (M, A, N), which means that it includes additive errors, an additive trend, and no seasonality.

The smoothing parameters are estimated to be alpha = 0.8486 and beta = 0.0447. These values represent the level of weight given to the current observation and the trend in the forecast, respectively. The initial values for the level and trend are 11098.4053 and 160.083 respectively. The estimated value for the standard deviation of the errors (sigma) is 0.0165.

Future forecast for 8 quarters of GDP using this model is presented below:

```
> # Use forecast() function to make predictions using this HW model for
> # 8 quarters into the future.
> HW.ZZZ.pred <- forecast(HW.ZZZ, h = 8 , level = 0)
> HW.ZZZ.pred
        Point Forecast     Lo 0     Hi 0
2023 Q1       26403.96 26403.96 26403.96
2023 Q2       26695.27 26695.27 26695.27
2023 Q3       26986.58 26986.58 26986.58
2023 Q4       27277.89 27277.89 27277.89
2024 Q1       27569.20 27569.20 27569.20
2024 Q2       27860.51 27860.51 27860.51
2024 Q3       28151.83 28151.83 28151.83
2024 Q4       28443.14 28443.14 28443.14
```




```
> # Identify performance measures for HW forecast.
> round(accuracy(HW.ZZZ.pred$fitted, gdp.ts), 3)
             ME   RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 36.673 332.98 156.635 0.131 0.832 0.026     0.835
> round(accuracy((naive(gdp.ts))$fitted, gdp.ts), 3)
              ME   RMSE    MAE   MPE  MAPE   ACF1 Theil's U
Test set 189.416 384.99 255.14 1.057 1.413 -0.091         1
> round(accuracy((snaive(gdp.ts))$fitted, gdp.ts), 3)
              ME     RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 737.172 1002.398 819.383 4.067 4.547 0.732     2.592
```

Comparing the performance measures of the Holt-Winter's model with the naive and seasonal naive models, we can see that the Holt-Winter's model outperforms both of them in terms of all measures. In particular, the Holt-Winter's model has a lower RMSE and MAPE, indicating that it is more accurate in predicting the future values of the GDP time series.

Regression Models

We developed the 5 regression models identified

1. **Regression Model with Linear Trend**

The model using the summary() function is presented below:

```
Call:
tslm(formula = train.ts ~ trend)

Residuals:
    Min      1Q  Median      3Q     Max 
-581.03 -323.53  -47.24  345.33  756.45 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11354.232     98.012  115.84   <2e-16 ***
trend         136.136      2.622   51.92   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 387.5 on 62 degrees of freedom
Multiple R-squared:  0.9775,	Adjusted R-squared:  0.9772 
F-statistic:  2696 on 1 and 62 DF,  p-value: < 2.2e-16
```

The regression model with linear trend contains a single independent variable: period index (t).

The model's equation is:

$$y^{t} = 11354.232 + 136.136t$$

According to the model summary, the regression model with linear trend is statistically significant. It has a high R-squared of 0.9775 and adjusted R_squared of 0.9772, which represents a good fit for the training data. The intercept and coefficient for the trend (t) variable are statistically significant (p-values are much lower than 0.05 or 0.01). Therefore, this model may be used for time series forecasting.

The forecast for the training period is the following:

```
        Point Forecast     Lo 0     Hi 0
2019 Q1       20203.10 20203.10 20203.10
2019 Q2       20339.24 20339.24 20339.24
2019 Q3       20475.37 20475.37 20475.37
2019 Q4       20611.51 20611.51 20611.51
2020 Q1       20747.65 20747.65 20747.65
2020 Q2       20883.78 20883.78 20883.78
2020 Q3       21019.92 21019.92 21019.92
2020 Q4       21156.06 21156.06 21156.06
2021 Q1       21292.19 21292.19 21292.19
2021 Q2       21428.33 21428.33 21428.33
2021 Q3       21564.47 21564.47 21564.47
2021 Q4       21700.60 21700.60 21700.60
2022 Q1       21836.74 21836.74 21836.74
2022 Q2       21972.87 21972.87 21972.87
2022 Q3       22109.01 22109.01 22109.01
2022 Q4       22245.15 22245.15 22245.15
```

2. **Regression Model with Quadratic Trend**

The model using the summary() function is presented below:

```
Call:
tslm(formula = train.ts ~ trend + I(trend^2))

Residuals:
   Min     1Q Median     3Q    Max 
-662.5 -238.2 -116.0  231.5  736.5 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.173e+04  1.368e+02  85.768  < 2e-16 ***
trend       1.016e+02  9.713e+00  10.457 3.14e-15 ***
I(trend^2)  5.318e-01  1.448e-01   3.672 0.000508 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 353.5 on 61 degrees of freedom
Multiple R-squared:  0.9816,	Adjusted R-squared:  0.981 
F-statistic:  1626 on 2 and 61 DF,  p-value: < 2.2e-16
```

The regression model with quadratic trend contains two independent variables: period index (t) and squared period index (t 2).

The model's equation is:

$$y^{t} = 1.173e^{04} + 1.016e^{02}\mathrm{t} + 5.318e^{-01}\, t^{2}$$

According to the model summary, the regression model with quadratic trend is statistically significant. It has a high R-squared of only 0.9816 (adj. R_squared is 0.981), which is a good fit for the training data.The coefficients for the trend (t) and quadratic trend (t^{2}) variables are statistically significant (p-values for both coefficients are much lower than 0.05 or 0.01). This model may also be used for time series forecasting.

The forecast for the validation period is the following:

	Point Forecast	Lo 0	Hi 0
2019 Q1	20583.36	20583.36	20583.36
2019 Q2	20754.60	20754.60	20754.60
2019 Q3	20926.90	20926.90	20926.90
2019 Q4	21100.27	21100.27	21100.27
2020 Q1	21274.70	21274.70	21274.70
2020 Q2	21450.19	21450.19	21450.19
2020 Q3	21626.75	21626.75	21626.75
2020 Q4	21804.37	21804.37	21804.37
2021 Q1	21983.05	21983.05	21983.05
2021 Q2	22162.80	22162.80	22162.80
2021 Q3	22343.61	22343.61	22343.61
2021 Q4	22525.48	22525.48	22525.48
2022 Q1	22708.42	22708.42	22708.42
2022 Q2	22892.42	22892.42	22892.42
2022 Q3	23077.49	23077.49	23077.49
2022 Q4	23263.62	23263.62	23263.62

3. **Regression Model with Seasonality**

The model using the summary() function is presented below:

```
Call:
tslm(formula = train.ts ~ season)

Residuals:
    Min      1Q  Median      3Q     Max 
-4389.6 -1547.5  -462.1  1986.0  4819.1 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  15531.0      655.1  23.708   <2e-16 ***
season2        171.4      926.5   0.185    0.854    
season3        337.2      926.5   0.364    0.717    
season4        482.1      926.5   0.520    0.605    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2620 on 60 degrees of freedom
Multiple R-squared:  0.005029,  Adjusted R-squared:  -0.04472 
F-statistic: 0.1011 on 3 and 60 DF,  p-value: 0.9591
```

The regression model with seasonality contains 3 independent seasonal dummy variables for Q2 (season2 – D2), Q3 (season3 – D3) and Q4 (season4 – D4).

The model's equation is:

$$y^t = 15531 + 171.4\ D2 + 337.2\ D3 + 482.1\ D4$$

The model's summary shows a very low R-squared of 0.005029 and even lower adjusted R_squared of -0.04472, and all regression coefficients for the seasonal variables are statistically insignificant. Overall, this regression model is not a good fit, and thus cannot be applied for time series forecasting.

4. **Regression Model with Linear Trend and Seasonality**

The model using the summary() function is presented below:

```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-603.07 -309.08  -58.79  314.18  737.73 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11313.608    129.347  87.467   <2e-16 ***
trend         136.044      2.685  50.671   <2e-16 ***
season2        35.341    140.050   0.252    0.802    
season3        65.153    140.127   0.465    0.644    
season4        73.992    140.256   0.528    0.600    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 396 on 59 degrees of freedom
Multiple R-squared:  0.9777,	Adjusted R-squared:  0.9761 
F-statistic: 645.2 on 4 and 59 DF,  p-value: < 2.2e-16
```

The regression model contains 4 independent variables: trend index (t) and 3 seasonal dummy variables for Q2 (season2 – D2), Q3 (season3 – D3) and Q4 (season4 – D4).

The model’s equation is:

$$y^{t} = 11313.608 + 136.044\ t + 35.341\ D2 + 65.153\ D3 + 73.992\ D4$$

The model’s summary shows a very high R-squared of 0.9777 and adj. R-squared of 0.9761, which is a very good fit for the training data, statistically.The regression coefficient for D2,D3 and D4 are statistically insignificant for p-value > 0.05, overall, this regression model is a good fit and thus can be applied for time series forecasting.

The forecast for the validation period is the following:

```
        Point Forecast     Lo 0     Hi 0
2019 Q1       20156.48 20156.48 20156.48
2019 Q2       20327.87 20327.87 20327.87
2019 Q3       20493.72 20493.72 20493.72
2019 Q4       20638.61 20638.61 20638.61
2020 Q1       20700.66 20700.66 20700.66
2020 Q2       20872.04 20872.04 20872.04
2020 Q3       21037.90 21037.90 21037.90
2020 Q4       21182.78 21182.78 21182.78
2021 Q1       21244.84 21244.84 21244.84
2021 Q2       21416.22 21416.22 21416.22
2021 Q3       21582.08 21582.08 21582.08
2021 Q4       21726.96 21726.96 21726.96
2022 Q1       21789.01 21789.01 21789.01
2022 Q2       21960.40 21960.40 21960.40
2022 Q3       22126.25 22126.25 22126.25
2022 Q4       22271.14 22271.14 22271.14
```

5. **Regression Model with Quadratic Trend and Seasonality**

The model using the summary() function is presented below:

```
Call:
tslm(formula = train.ts ~ trend + I(trend^2) + season)

Residuals:
   Min     1Q Median     3Q    Max 
-622.1 -246.1 -113.0  217.5  705.5 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.169e+04  1.584e+02  73.830  < 2e-16 ***
trend       1.015e+02  9.927e+00  10.222 1.37e-14 ***
I(trend^2)  5.319e-01  1.480e-01   3.594 0.000672 ***
season2     3.640e+01  1.277e+02   0.285 0.776670    
season3     6.622e+01  1.278e+02   0.518 0.606375    
season4     7.399e+01  1.279e+02   0.578 0.565249    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 361.2 on 58 degrees of freedom
Multiple R-squared:  0.9817,	Adjusted R-squared:  0.9801 
F-statistic:   623 on 5 and 58 DF,  p-value: < 2.2e-16
```

The regression model with quadratic trend and seasonality contains 5 independent variables: trend index (t), squared trend index (t^2), and 3 seasonal dummy variables for Q2 (season2 – D2), Q3 (season3 – D3) and Q4 (season4 – D4).

The model's equation is:

$$y^{t} = 1.169e^{04} + 1.015e^{02}\ \mathrm{t} + 5.319e^{-01}\ t^{2} + 3.640e^{+01}\ \mathrm{D2} + 6.622e^{+01}\ \mathrm{D3} + 7.399e^{+01}\ \mathrm{D4}$$

The model summary shows a very high R-squared of 0.9817 and adj. R-squared of 0.9801. The regression coefficients for D2,D3 and D4 are statistically insignificant which are not below p-value of 0.05. Overall, this regression model is a very good fit and thus can be applied for time series forecasting.

The forecast for the validation period is the following:

	Point Forecast	Lo 0	Hi 0
2019 Q1	20536.27	20536.27	20536.27
2019 Q2	20743.82	20743.82	20743.82
2019 Q3	20945.85	20945.85	20945.85
2019 Q4	21126.90	21126.90	21126.90
2020 Q1	21227.25	21227.25	21227.25
2020 Q2	21439.07	21439.07	21439.07
2020 Q3	21645.35	21645.35	21645.35
2020 Q4	21830.66	21830.66	21830.66
2021 Q1	21935.26	21935.26	21935.26
2021 Q2	22151.33	22151.33	22151.33
2021 Q3	22361.86	22361.86	22361.86
2021 Q4	22551.43	22551.43	22551.43
2022 Q1	22660.29	22660.29	22660.29
2022 Q2	22880.61	22880.61	22880.61
2022 Q3	23095.40	23095.40	23095.40
2022 Q4	23289.22	23289.22	23289.22

The accuracy measures for the 5 regression models are presented below:

```
> round(accuracy(train.lin.pred$mean, valid.ts),3)
               ME     RMSE      MAE   MPE  MAPE  ACF1 Theil's U
Test set 1580.689 2068.301 1736.571 6.513 7.307 0.748     2.345
> round(accuracy(train.quad.pred$mean, valid.ts),3)
              ME     RMSE      MAE   MPE  MAPE  ACF1 Theil's U
Test set 899.937 1484.151 1172.117 3.571 4.937 0.725     1.686
> round(accuracy(train.season.pred$mean, valid.ts),3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 7026.147 7261.899 7026.147 30.367 30.367 0.796     8.488
> round(accuracy(train.lin.trend.season.pred$mean, valid.ts),3)
               ME     RMSE      MAE   MPE  MAPE ACF1 Theil's U
Test set 1584.378 2067.905 1738.793 6.532 7.318 0.75     2.343
> round(accuracy(train.quad.trend.season.pred$mean, valid.ts),3)
              ME     RMSE      MAE   MPE  MAPE  ACF1 Theil's U
Test set 903.528 1482.724 1179.928 3.589 4.974 0.726     1.684
```

Based on the MAPE and RMSE values, the three most accurate regression models for forecasting are:

(1)**Regression Model with Quadratic Trend and Seasonality**

RMSE: 1482.724

MAPE: 4.974

(2)**Regression Model with Quadratic Trend**

RMSE: 1484.151

MAPE: 4.937

(3)**Regression Model with Linear Trend and Seasonality**

RMSE: 2067.905

MAPE: 7.318

These three models demonstrate the lowest MAPE and RMSE values, indicating higher accuracy in their forecasting performance compared to the other models.

The entire data set to make a time series forecast.

Applying the three most accurate regression models identified from above to make the forecast in the four quarters (Q1-Q4) of 2023 and 2024.

1. **Regression Model with Quadratic Trend and Seasonality**

For forecasting US GDP in 2023-2024, we apply for the entire data set the regression model with quadratic trend and seasonality, which was identified to be the first best model.

The regression model with quadratic trend and seasonality for the entire data set is presented below:

```
Call:
tslm(formula = gdp.ts ~ trend + I(trend^2) + season)

Residuals:
     Min       1Q   Median       3Q      Max
-2484.78  -250.49   -48.55   141.23  1474.36

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 12082.6808   232.2632  52.021  < 2e-16 ***
trend          62.1044    11.6393   5.336 9.99e-07 ***
I(trend^2)      1.1692     0.1392   8.397 2.29e-12 ***
season2       -37.4035   187.8052  -0.199    0.843
season3        86.6764   187.8713   0.461    0.646
season4       129.9234   187.9813   0.691    0.492
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 593.8 on 74 degrees of freedom
Multiple R-squared:  0.9763,    Adjusted R-squared:  0.9747
F-statistic: 610.2 on 5 and 74 DF,  p-value: < 2.2e-16
```

This regression model with quadratic trend and seasonality contains 5 independent variables: trend index (t), squared trend index (t^2), and 3 seasonal dummy variables for Q2 (season2 – D2), Q3 (season3 – D3) and Q4 (season4 – D4).

The regression equation is:

y^{t} = 12082.6808 + 62.1044 t - 1.1692 $t2$ - 37.4035 D2 + 86.6764 D3 + 129.9234 D4

The model's summary shows a very high R-squared of 0.9763 and adj. R-squared of 0.9747, and all regression coefficients being statistically significant (p-value < 0.01 or 0.05) with exception for regression coefficient for D2, D3 and D4. Overall, this regression model is a very good fit for the historical data set, and thus can be used for forecasting US GDP in 2023-2024.

The appropriate forecast is shown below:

```
        Point Forecast     Lo 0     Hi 0
2023 Q1       24784.05 24784.05 24784.05
2023 Q2       24999.33 24999.33 24999.33
2023 Q3       25378.43 25378.43 25378.43
2023 Q4       25679.03 25679.03 25679.03
2024 Q1       25808.80 25808.80 25808.80
2024 Q2       26033.43 26033.43 26033.43
2024 Q3       26421.88 26421.88 26421.88
2024 Q4       26731.83 26731.83 26731.83
```

2. **Regression Model with Quadratic Trend**

For forecasting the US GDP in 2023-2024, we apply for the entire data set the regression model with quadratic trend, which was identified to be the second best model.

The regression model with quadratic trend for the entire data set is presented below:

```
Call:
tslm(formula = gdp.ts ~ trend + I(trend^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-2570.57  -192.23   -34.56   131.14  1554.63 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.212e+04  2.016e+02  60.133  < 2e-16 ***
trend       6.222e+01  1.149e+01   5.416 6.70e-07 ***
I(trend^2)  1.169e+00  1.374e-01   8.509 1.06e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 586.1 on 77 degrees of freedom
Multiple R-squared:  0.976,	Adjusted R-squared:  0.9754 
F-statistic:  1565 on 2 and 77 DF,  p-value: < 2.2e-16
```

The regression model with quadratic trend contains two independent variables: period index (t) and squared period index (t 2).

The model's equation is:

$$y^{t} = 1.212e^{+04} + 6.222e^{+01}\mathrm{t} + 1.169e^{+00}\, t^{2}$$

According to the model summary, the regression model with quadratic trend is statistically significant. It has a high R-squared of only 0.976 (adj. R_squared is 0.9754), which is a good fit for the training data.The coefficients for the trend (t) and quadratic trend (t^2) variables are statistically significant (p-values for both coefficients are much lower than 0.05 or 0.01). This model may also be used for time series forecasting US GDP in 2023-2024.

The forecast is shown below:

	Point Forecast	Lo 0	Hi 0
2023 Q1	24833.83	24833.83	24833.83
2023 Q2	25086.63	25086.63	25086.63
2023 Q3	25341.78	25341.78	25341.78
2023 Q4	25599.26	25599.26	25599.26
2024 Q1	25859.09	25859.09	25859.09
2024 Q2	26121.25	26121.25	26121.25
2024 Q3	26385.75	26385.75	26385.75
2024 Q4	26652.58	26652.58	26652.58

3. **Regression Model with Linear Trend and Seasonality**

For forecasting the US GDP in 2023-2024, we apply for the entire data set the regression model with linear trend and seasonality, which was identified to be the third most accurate model.

The regression model with linear trend and seasonality for the entire data set is presented below:

```
Call:
tslm(formula = gdp.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-2089.6  -593.4  -167.6   528.2  2673.9 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10789.580    241.361  44.703   <2e-16 ***
trend         156.807      3.996  39.245   <2e-16 ***
season2       -39.742    260.695  -0.152    0.879    
season3        84.338    260.787   0.323    0.747    
season4       129.923    260.940   0.498    0.620    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 824.3 on 75 degrees of freedom
Multiple R-squared:  0.9538,	Adjusted R-squared:  0.9513 
F-statistic: 386.7 on 4 and 75 DF,  p-value: < 2.2e-16
```

This regression model with linear trend and seasonality contains 4 independent variables: trend index (t) and 3 seasonal dummy variables for Q2 (season2 – D2), Q3 (season3 – D3) and Q4 (season4 – D4).

The regression equation is:

y^t = 10789.58 + 156.807 t - 39.742 D2 + 84.338 D3 + 129.923 D4

The model’s summary shows a high R-squared of 0.9538 and adj. R-squared of 0.9513, and all regression coefficients being statistically insignificant (p-value <0.01 or 0.05). Overall, this regression model is a very good fit and can be used for forecasting US GDP in 2023-2024.

The forecast is shown below:

```
        Point Forecast     Lo 0     Hi 0
2023 Q1       23490.95 23490.95 23490.95
2023 Q2       23608.02 23608.02 23608.02
2023 Q3       23888.90 23888.90 23888.90
2023 Q4       24091.30 24091.30 24091.30
2024 Q1       24118.18 24118.18 24118.18
2024 Q2       24235.25 24235.25 24235.25
2024 Q3       24516.13 24516.13 24516.13
2024 Q4       24718.52 24718.52 24718.52
```

The accuracy measures for the three forecasts for future 8 quarters specified above are listed below along with the accuracy for the respective naïve and seasonal naïve forecast:

```
> round(accuracy(quad.trend.season.pred$fitted, gdp.ts),3)
         ME    RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set  0 571.119 378.043 -0.098 2.273 0.776     1.634
> round(accuracy(quad.trend.pred$fitted, gdp.ts),3)
         ME    RMSE     MAE    MPE MAPE  ACF1 Theil's U
Test set  0 574.988 369.452 -0.099 2.22 0.772     1.636
> round(accuracy(lin.trend.season.pred$fitted, gdp.ts),3)
         ME    RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set  0 798.117 627.139 -0.049 3.594 0.842      2.14
> round(accuracy((naive(gdp.ts))$fitted, gdp.ts), 3)
              ME   RMSE    MAE   MPE  MAPE   ACF1 Theil's U
Test set 189.416 384.99 255.14 1.057 1.413 -0.091         1
> round(accuracy((snaive(gdp.ts))$fitted, gdp.ts), 3)
              ME     RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 737.172 1002.398 819.383 4.067 4.547 0.732     2.592
```

Based on a comparison of the MAPE and RMSE values of the five forecasts, we have:

Quadratic Trend Seasonal Forecast: MAPE = 2.273, RMSE = 571.119

Quadratic Trend Forecast: MAPE = 2.22, RMSE = 574.988

Linear Trend Seasonal Forecast: MAPE = 3.594, RMSE = 798.117

Naive Forecast: MAPE = 1.413, RMSE = 384.99

Seasonal Naive Forecast: MAPE = 4.547, RMSE = 1002.398

From these results, we can see that the naive forecast has the smallest MAPE and RMSE values, indicating that it is the most accurate forecast for the US quarterly GDP in Q1-Q4 of 2023-2024.

Auto Regressive models

Time series predictability

The output of the AR(1) model for gdp.ts time series data is presented below. ARIMA(1, 0, 0) is an autoregressive (AR) model with order 1, no differencing, and no moving average model.

```
Series: gdp.diff
ARIMA(1,0,0) with non-zero mean

Coefficients:
          ar1      mean
      -0.0902  189.2335
s.e.   0.1117   34.4819

sigma^2 = 114301:  log likelihood = -571.13
AIC=1148.26   AICc=1148.58   BIC=1155.36

Training set error measures:
                     ME     RMSE   MAE      MPE    MAPE      MASE        ACF1
Training set -0.0551876 333.7767 153.6 3.918045 118.315 0.7010081 0.008290091
```

The model's equation is:

$$y^{t} = 189.2335 - 0.1117 \text{ Yt-1}$$

The coefficient of the ar1 (Yt-1) variable, β1 = -0.902, and standard error of estimate, s.e. = 0.1117. We will use these two parameters for hypothesis testing about the value of the AR(1) regression coefficient.

Hypothesis Testing: Z- Test

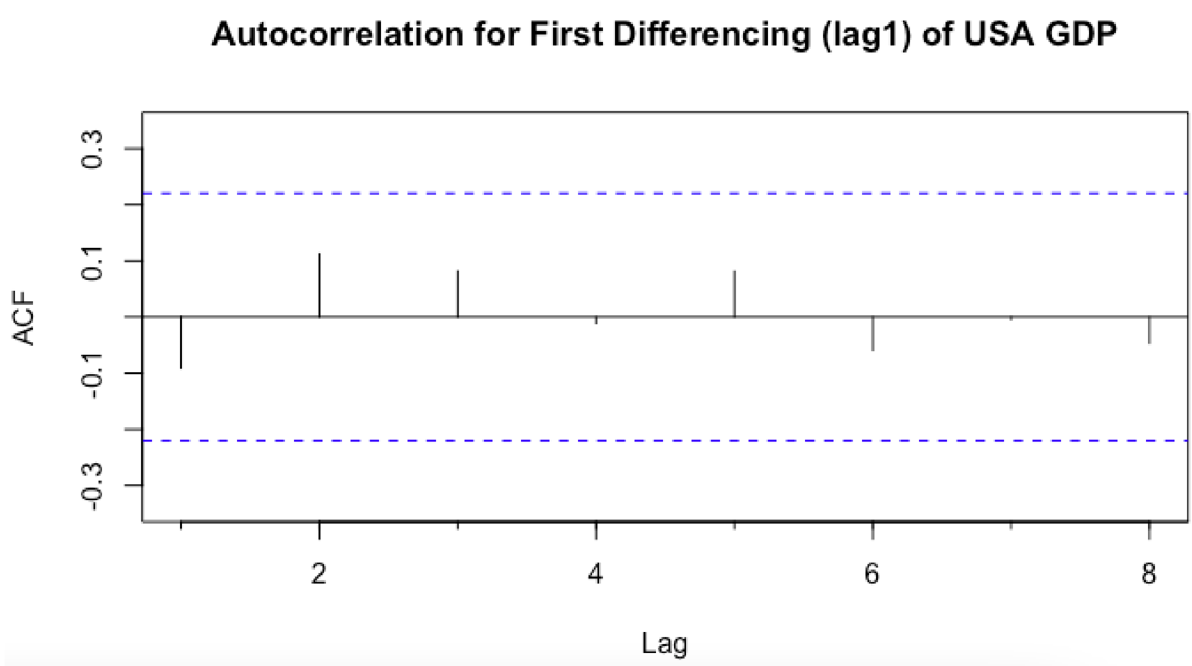
Null hypothesis Ho: β1 = 1

Alternative hypothesis H1: β1 ≠ 1

```
> z.stat <- (-0.0902 - 1)/0.1117
> z.stat
[1] -9.760072
> p.value <- pnorm(z.stat)
> p.value
[1] 8.351861e-23
```

Based on the p-value of 8.251861e-23, that is much less than 0.05, we can reject the null hypothesis that β1 = 1. Therefore, the time series data for US GDP, gdp.ts, is not random walk and predictable.

The **autocorrelation plot of the first differencing** for the gdp.ts data is presented below.



The autocorrelation chart above of the Autocorrelation for first differencing(lag1) of USA GDP shows that all autocorrelations are random (within horizontal thresholds), which means that the model absorbed significant autocorrelations for First Differencing (lag1).

The **output** for the regression model with **linear trend and seasonality** for the training period and forecast for the validation period are shown below

```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-603.07 -309.08  -58.79  314.18  737.73 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11313.608    129.347  87.467   <2e-16 ***
trend         136.044      2.685  50.671   <2e-16 ***
season2        35.341    140.050   0.252    0.802    
season3        65.153    140.127   0.465    0.644    
season4        73.992    140.256   0.528    0.600    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 396 on 59 degrees of freedom
Multiple R-squared:  0.9777,    Adjusted R-squared:  0.9761 
F-statistic: 645.2 on 4 and 59 DF,  p-value: < 2.2e-16
```

	Point Forecast	Lo 0	Hi 0
2019 Q1	20156.48	20156.48	20156.48
2019 Q2	20327.87	20327.87	20327.87
2019 Q3	20493.72	20493.72	20493.72
2019 Q4	20638.61	20638.61	20638.61
2020 Q1	20700.66	20700.66	20700.66
2020 Q2	20872.04	20872.04	20872.04
2020 Q3	21037.90	21037.90	21037.90
2020 Q4	21182.78	21182.78	21182.78
2021 Q1	21244.84	21244.84	21244.84
2021 Q2	21416.22	21416.22	21416.22
2021 Q3	21582.08	21582.08	21582.08
2021 Q4	21726.96	21726.96	21726.96
2022 Q1	21789.01	21789.01	21789.01
2022 Q2	21960.40	21960.40	21960.40
2022 Q3	22126.25	22126.25	22126.25
2022 Q4	22271.14	22271.14	22271.14

The autocorrelation chart (correlogram) of the residuals from the regression model with linear trend and seasonality is provided below.

The chart shows that there is significant autocorrelation coefficient of residuals in lags 1 through 6 but there is a gradual decrease in the lags gradually indicating a trend in training residuals. Thus, modeling these residual autocorrelations with an AR model and developing a two-level model may, overall, improve the forecast.

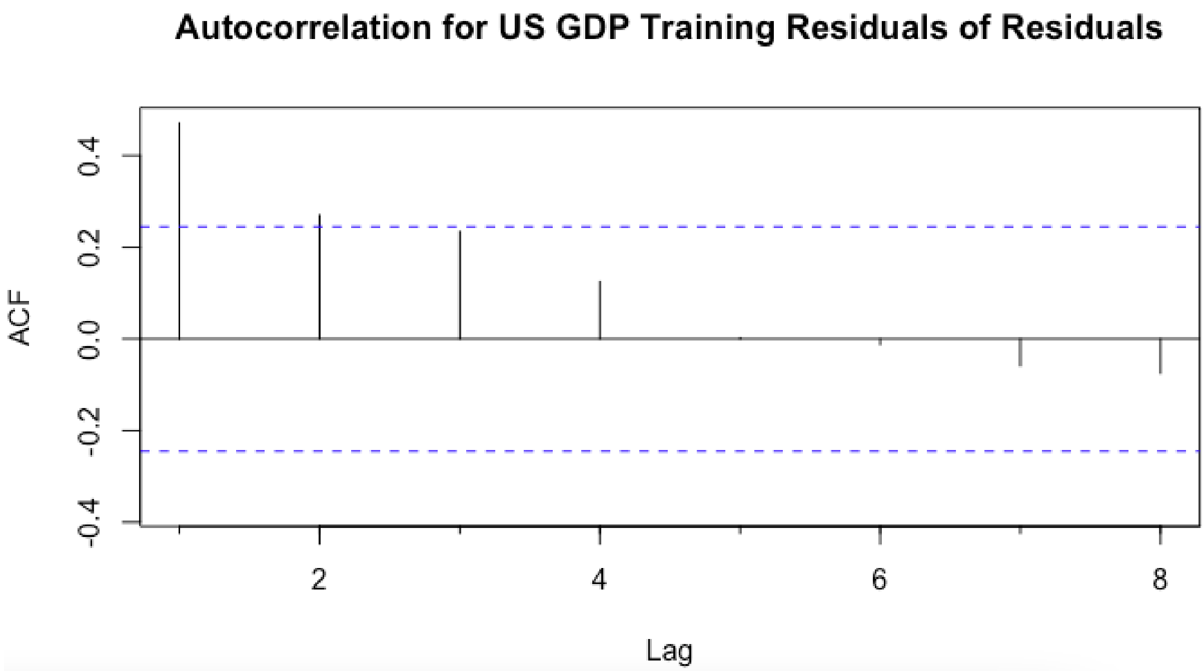
The **output** of the **AR(1) model for regression residuals** is presented below. ARIMA(1, 0, 0) is an autoregressive (AR) model with order 1, no differencing, and no moving average model.

```
Series: train.trend.season$residuals
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1      mean
      0.9697  113.9779
s.e.  0.0274  317.5020

sigma^2 = 11839:  log likelihood = -391.34
AIC=788.68   AICc=789.08   BIC=795.16

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE      ACF1
Training set 10.40842 107.0947 76.89547 34.46999 70.03505 0.3351336 0.4709673
```

The AR(1) model's equation is:

et = 113.9779 + 0.9697 et-1

An autocorrelation chart for the AR(1) model's residuals (residuals of residuals) is presented below.



As can be seen from the chart (correlogram), all autocorrelations of AR(1) model's residuals (residuals of residuals) are random except for lags 1 and 2 which is only weakly significant.This means that the *AR(1)* model absorbed significant autocorrelations in the regression residuals.

ARIMA Models

Built Auto ARIMA model which will give us the best ARIMA model for the dataset.

The summary of Auto ARIMA is as given below

```
> summary(train.auto.arima)
Series: train.ts
ARIMA(1,1,0) with drift

Coefficients:
         ar1     drift
      0.4848  152.5684
s.e.  0.1083   21.8402

sigma^2 = 8484:  log likelihood = -373.45
AIC=752.9   AICc=753.31   BIC=759.33

Training set error measures:
                    ME     RMSE    MAE        MPE      MAPE       MASE        ACF1
Training set 0.3117995 89.9236 64.835 -0.0011387 0.4143769 0.09989838 -0.02632458
>
```

- The model has selected an ARIMA(1,1,0) model with drift for the training data. This means that the model includes one autoregressive term, one differencing term, and no moving average terms, with a drift term added to the model.

  The equation of the model is as follows:

  $$y_t = 152.5684 + 0.4848\, y_{t-1} + 0.1083\, \rho_{t-1}$$

- The coefficient for the autoregressive term (ar1) is 0.4848, which indicates that there is a moderate positive correlation between the current observation and the previous observation.
- The drift coefficient is 152.5684, which means that the model has a positive drift trend, indicating that the GDP has increased over time.
- The model's log-likelihood value is -373.45, which indicates that the model is a good fit for the training data.
- The MAPE value for the training set is 0.4143769 or 41.44%. This means that the average absolute percentage error of the model's forecasts is around 41.44% of the actual GDP values in the training set.

- The model's RMSE (Root Mean Squared Error) value for the training set is 89.9236, which indicates that the model has an average error of about $89 billion in predicting GDP values.

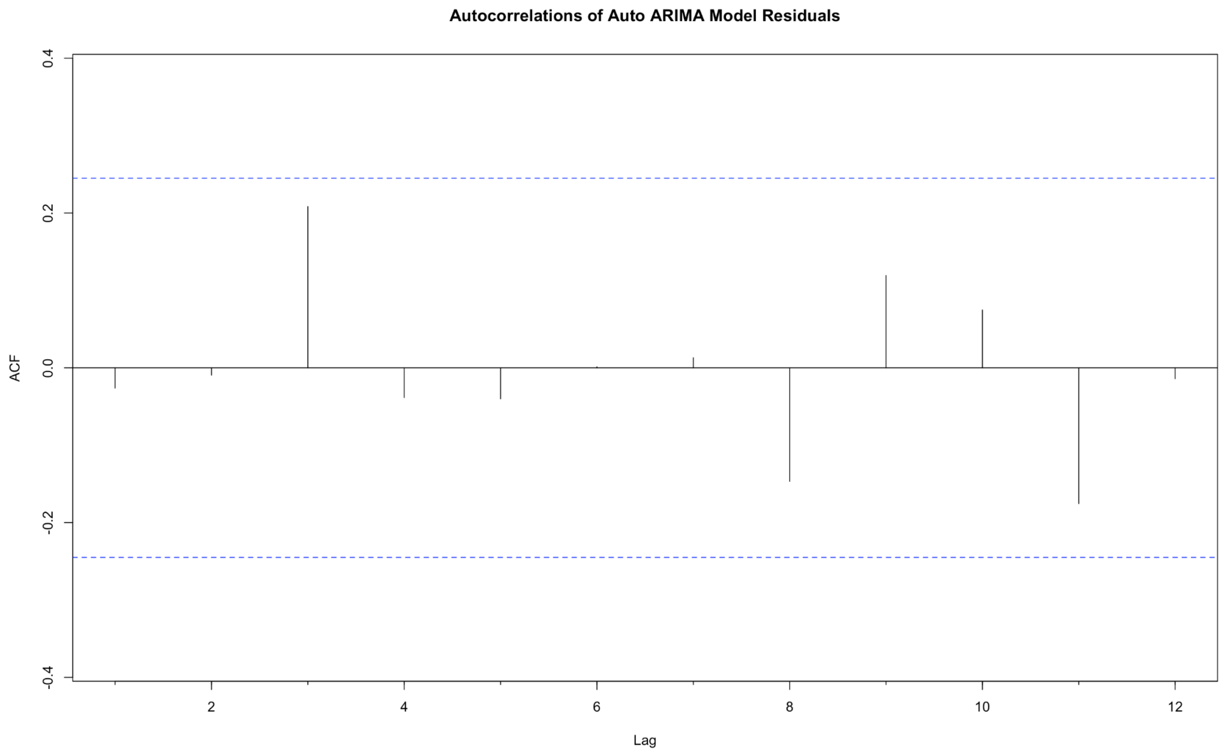
Forecast of the Auto ARIMA model

```
> train.auto.arima.pred
        Point Forecast     Lo 0     Hi 0
2019 Q1       20961.86 20961.86 20961.86
2019 Q2       21109.59 21109.59 21109.59
2019 Q3       21259.82 21259.82 21259.82
2019 Q4       21411.25 21411.25 21411.25
2020 Q1       21563.26 21563.26 21563.26
2020 Q2       21715.57 21715.57 21715.57
2020 Q3       21868.00 21868.00 21868.00
2020 Q4       22020.51 22020.51 22020.51
2021 Q1       22173.05 22173.05 22173.05
2021 Q2       22325.60 22325.60 22325.60
2021 Q3       22478.16 22478.16 22478.16
2021 Q4       22630.73 22630.73 22630.73
2022 Q1       22783.29 22783.29 22783.29
2022 Q2       22935.86 22935.86 22935.86
2022 Q3       23088.43 23088.43 23088.43
2022 Q4       23241.00 23241.00 23241.00
```

Based on the output you provided, it appears that the Auto ARIMA model has been used to generate point forecasts for the US GDP for the next several quarters, starting from Q1 2019 to Q4 2022.

The point forecasts indicate the predicted values for each quarter, along with the upper and lower bounds of the 95% confidence interval. The point forecasts suggest that the US GDP is expected to continue to grow over the forecast period, with each successive quarter showing an increase in GDP from the previous quarter.

Autocorrelation of the Auto ARIMA model:



All the autocorrelation values of the residuals fall within the threshold limit, indicating that there are no significant remaining patterns or dependencies in the data. This suggests that the Auto ARIMA model has adequately captured the underlying patterns and trends in the data, leaving minimal residual correlations. This provides confidence in the reliability of the model's forecasts and suggests that the model has effectively accounted for the autocorrelation structure of the time series.

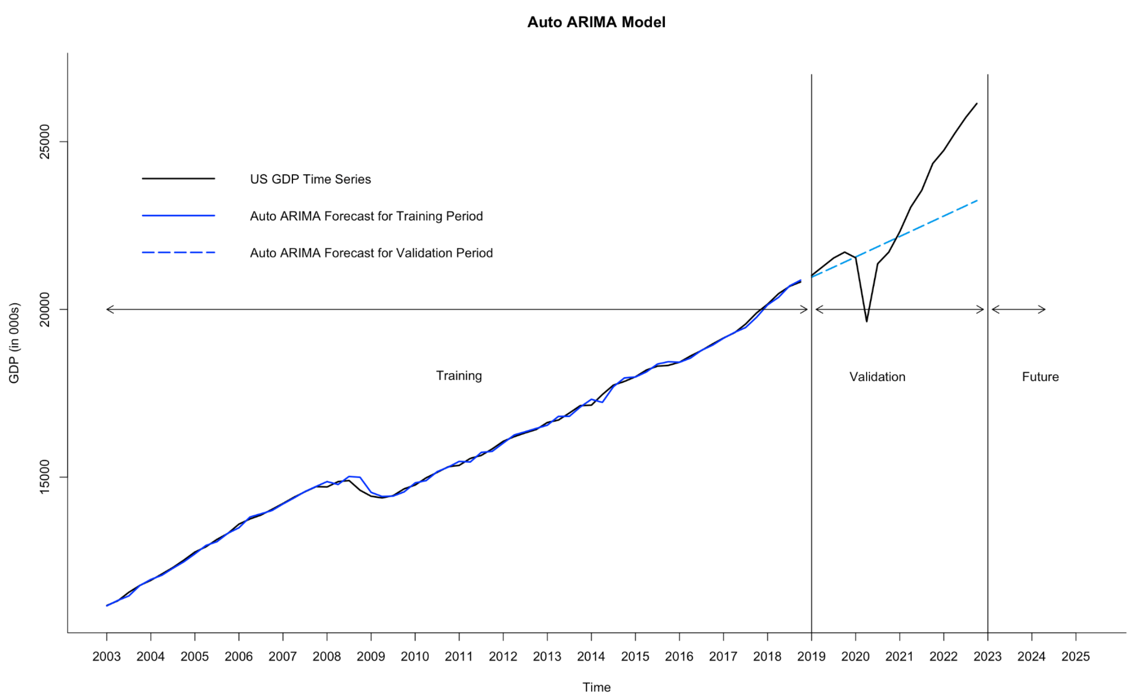
Accuracy of the Auto ARIMA model:

```
> round(accuracy(train.auto.arima.pred$mean, valid.ts), 3)
             ME     RMSE     MAE  MPE  MAPE ACF1 Theil's U
Test set 706.94 1458.335 1072.62 2.68 4.496 0.74     1.662
>
```

- MAPE (Mean Absolute Percentage Error): The MAPE value is 4.496%, indicating that, on average, the model's forecasts deviate from the actual values by approximately 4.496% of the actual values. Lower MAPE values are generally desirable as they indicate a smaller percentage difference between the forecasts and the actual values which we can see from the MAPE value.
- RMSE (Root Mean Squared Error): The RMSE value is 1458.335, which represents the square root of the average squared errors between the forecasts and the actual values. RMSE provides an estimate of the typical size of the forecast errors. Smaller RMSE values indicate lower overall forecast errors.

  Lower MAPE and RMSE values indicate that the model is a good fit.

  Plot of the Auto ARIMA with Validation and Predicted values is depicted below



Built model for the Entire data set using Auto ARIMA.

Summary of the Auto ARIMA model for Entire Data set

```
> summary(auto.arima)
Series: gdp.ts
ARIMA(0,1,0) with drift

Coefficients:
         drift
      189.4160
s.e.   37.7096

sigma^2 = 113781:  log likelihood = -571.45
AIC=1146.91   AICc=1147.06   BIC=1151.64

Training set error measures:
                    ME     RMSE     MAE         MPE      MAPE      MASE        ACF1
Training set 0.1373088 333.0709 147.097 -0.08752415 0.7777912 0.1795216 -0.09087878
>
```

- The **auto.arima** function has selected an ARIMA(0,1,0) model with a drift term. This means that the model includes differencing of order 1 to make the time series stationary and a drift term to account for a linear trend in the data.
- Drift Coefficient: The drift coefficient is 189.4160, indicating a positive linear trend in the data. This means that the GDP is expected to increase over time.
- Model Evaluation: The log-likelihood of the model is -571.45, and the corresponding AIC, AICc, and BIC values are provided. These metrics are used to evaluate the goodness of fit and compare different models. Lower values of AIC, AICc, and BIC generally indicate a better fit.
- The RMSE (Root Mean Squared Error) is 333.0709, representing the typical magnitude of the forecast errors
- The MAPE (Mean Absolute Percentage Error) is 0.7777912, representing the average absolute percentage difference between the forecasts and the actual values.

Forecast of the Auto ARIMA model for Entire Data set:

```
> auto.arima.pred
        Point Forecast     Lo 0     Hi 0
2023 Q1       26327.41 26327.41 26327.41
2023 Q2       26516.82 26516.82 26516.82
2023 Q3       26706.24 26706.24 26706.24
2023 Q4       26895.66 26895.66 26895.66
2024 Q1       27085.07 27085.07 27085.07
2024 Q2       27274.49 27274.49 27274.49
2024 Q3       27463.90 27463.90 27463.90
2024 Q4       27653.32 27653.32 27653.32
2025 Q1       27842.74 27842.74 27842.74
2025 Q2       28032.15 28032.15 28032.15
2025 Q3       28221.57 28221.57 28221.57
2025 Q4       28410.98 28410.98 28410.98
```

The following are the insights of the forecast:

- Stable Growth: The forecasted GDP values show a consistent pattern of gradual growth over time. Each subsequent quarter's forecasted value is higher than the previous quarter, indicating a steady upward trend in the GDP.
- Confidence Intervals: The "Lo 0" and "Hi 0" columns provide the lower and upper bounds of the 95% confidence interval for each forecasted value. These intervals represent the range within which the actual GDP values are likely to fall with 95% confidence. The narrow width of the confidence intervals suggests a relatively high degree of confidence in the forecasts.
- Incremental Increases: The point forecast values show incremental increases from one quarter to the next. The differences between consecutive quarters are relatively consistent, suggesting a steady growth rate.

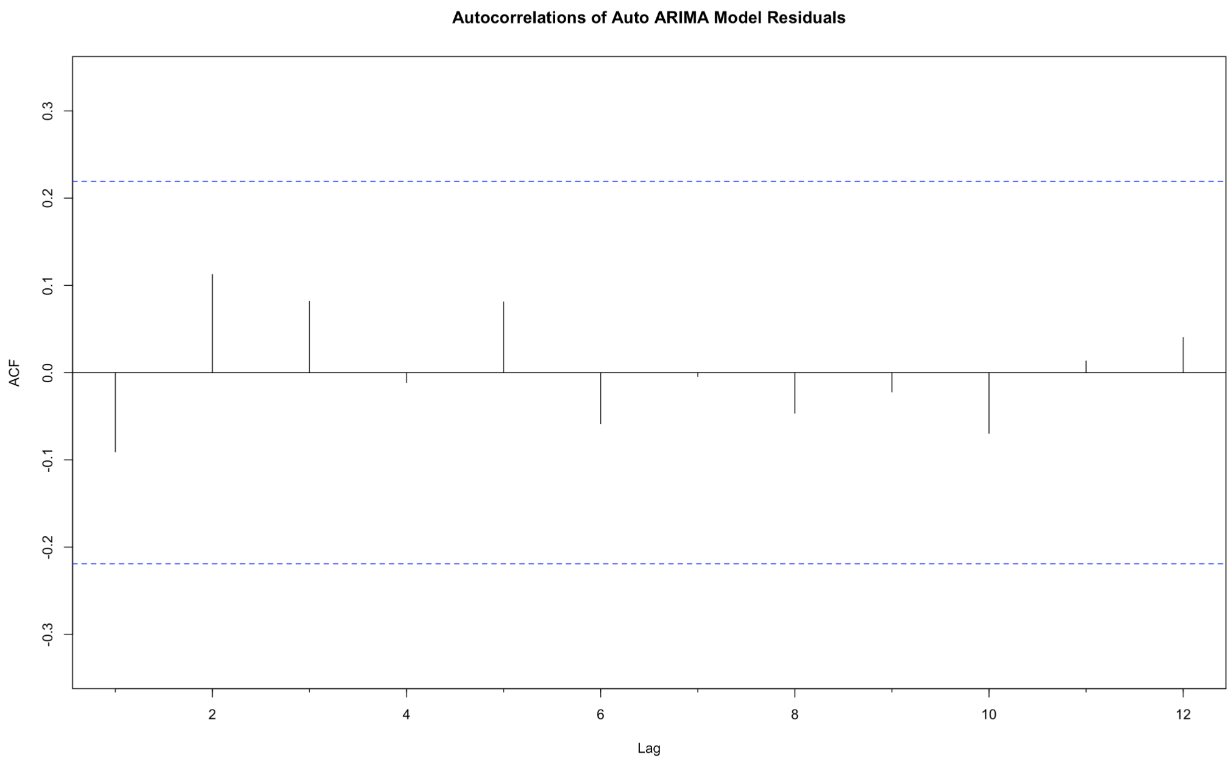
- Long-Term Projection: The forecast extends up to 2025 Q4, providing an outlook for the US GDP over the next few years. According to the forecast, the GDP is expected to continue growing steadily throughout this period.

Accuracy measure of Auto ARIMA for entire data set

```
> round(accuracy(auto.arima.pred$fitted, gdp.ts), 3)
            ME    RMSE     MAE    MPE  MAPE   ACF1 Theil's U
Test set 0.137 333.071 147.097 -0.088 0.778 -0.091      0.84
```

- The MAPE value of 0.778 indicates a relatively high average percentage difference between the forecasted and actual values. This suggests that the model's predictions may have a significant deviation from the true values, on average, in percentage terms.
- The RMSE value of 333.071 suggests that, on average, the forecasted values differ from the actual values by approximately 333.071 units. This magnitude of error can be considered relatively large, depending on the scale of the GDP data

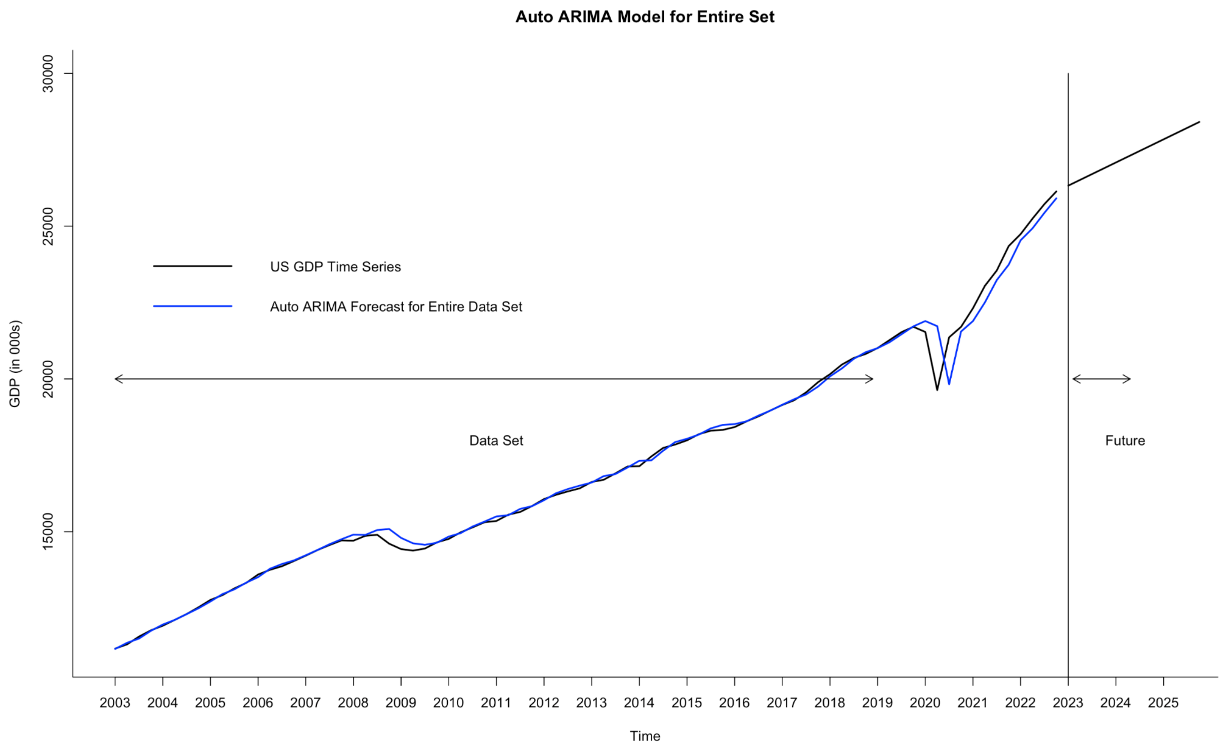
Autocorrelation graph of Auto ARIMA model for Entire dataset



The autocorrelation values for various lags were found to be within the threshold limit or dotted blue lines. This suggests that the residuals of the Auto ARIMA model exhibit little to no significant autocorrelation. The absence of notable autocorrelation indicates that the model has adequately captured the temporal dependencies in the data. It implies that the model's residuals do not contain any systematic information that the model has failed to account for.

Overall, the autocorrelation chart confirms the satisfactory performance of the Auto ARIMA model, demonstrating its ability to capture and explain the autocorrelation patterns in the data. This provides confidence in the model's ability to generate accurate and reliable forecasts.

Plot of the Auto ARIMA model for Entire Dataset



Step 8: Implement Forecast

```
> round(accuracy(res.ar1.pred$fitted, gdp.ts), 3)
                ME     RMSE      MAE     MPE    MAPE ACF1 Theil's U
Test set 15793.67 15993.71 15793.67 100.159 100.159 0.95    83.136
> round(accuracy(tot.trend.seas.pred$fitted+tot.ma.trail.res, GDP.ts), 3)
              ME   RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 30.272 242.62 144.379 0.096 0.782 0.379     0.614
> round(accuracy(HW.ZZZ.pred$fitted, gdp.ts), 3)
              ME   RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 36.673 332.98 156.635 0.131 0.832 0.026     0.835
> round(accuracy(auto.arima.pred$fitted, gdp.ts), 3)
             ME    RMSE     MAE    MPE  MAPE   ACF1 Theil's U
Test set 0.137 333.071 147.097 -0.088 0.778 -0.091      0.84
> round(accuracy(quad.trend.season.pred$fitted, gdp.ts),3)
         ME    RMSE     MAE    MPE  MAPE  ACF1 Theil's U
Test set  0 571.119 378.043 -0.098 2.273 0.776     1.634
> round(accuracy((naive(gdp.ts))$fitted, gdp.ts), 3)
               ME   RMSE    MAE   MPE  MAPE   ACF1 Theil's U
Test set 189.416 384.99 255.14 1.057 1.413 -0.091         1
> round(accuracy((snaive(gdp.ts))$fitted, gdp.ts), 3)
               ME     RMSE     MAE   MPE  MAPE  ACF1 Theil's U
Test set 737.172 1002.398 819.383 4.067 4.547 0.732     2.592
> round(accuracy(res.ar1.pred$fitted, gdp.ts), 3)
                ME     RMSE      MAE     MPE    MAPE ACF1 Theil's U
Test set 15793.67 15993.71 15793.67 100.159 100.159 0.95    83.136
```

Based on these measures, all three models show reasonably similar performance, hence it is recommended to use the Auto ARIMA, Holt-Winters and two-level model with moving average component models for GDP forecasting. The Auto ARIMA model demonstrates good overall performance with low errors and captures the autocorrelation structure well. The Holt-Winters model also performs reasonably well, though it exhibits a slight bias in the forecasts.

After conducting analysis through all the above models and comparing the accuracy measures for future period , we can finally see from above that the Auto ARIMA model has lowest MAPE(0.778) value ,hence making it the best model for our time series data.

Conclusion

From analyzing several regression and forecasting models, it has been determined that the most suitable model for predicting US GDP in the future is the Auto ARIMA Model. This model demonstrated superior performance in terms of the Mean Average Percentage Error (MAPE) and Root Mean Squared Error (RMSE) with values of 0.778 and 333.08, respectively.

The Holt-winter's model and two-level model with moving average component are the second best models with comparable MAPE and RMSE values .However, the Auto ARIMA model is better than either of the models. Hence, we suggest the Auto ARIMA model for predicting/forecasting the future US GDP values.