Application of ARIMA Model and Exponential Smoothing Method in GDP Prediction of the United States

Gaoyue Lyu[1, *, †], Xiulin Lyu[2, †], Nansong Sun[3, †], Xinrui Zhao[4, †]

[1]Faculty of Humanities and Social Sciences, Beijing Institute of Technology, Beijing, 102401, China

[2]Hangzhou Foreign Language School, Hangzhou, Zhejiang, 310000, China

[3]Dalian No.48 High School, Dalian, Liaoning,116000, China

[4]Suzhou Science Technology Town Foreign Language High School, Suzhou, Jiangsu, 215000, China

*Corresponding author: 1120213111@bit.edu.cn

†These authors contributed equally

Abstract. The GDP of the United States always attracts the global attention. It represents the strength of the world's largest economy and the per capita living standard of developed countries. Therefore, many scholars try to use various methods to predict it. On the basis of previous research, this paper collects the Fred ’ s quarterly U.S. GDP data, uses ARIMA model and exponential smoothing to model the US's GDP from 1950 to 2021. Then, the authors compare the actual values with the predicted values to select a better model. After that, the established model is used to forecast the US's GDP. The results show that the two methods generally maintain a high degree of fit, but the error generated increases with the widening of the expected range. Among these two methods, the average error value of the exponential smoothing method is slightly smaller than ARIMA, but ARIMA's forecasts are more accurate in the early. Therefore, ARIMA should be used preferentially for the forecast in the short term and the exponential smoothing method should be preferred for long-term forecasts. Furthermore, we use the exponential smoothing method to make a short-term forecast for the U.S. GDP by the fourth quarter of 2023. The results show that the U.S. GDP will increase at an average rate of 0.9344%, and the upward trend is stable and cyclical. Finally, the result of the above provides an effective basis for the US economic construction work and plays an inspirational role for China's economic policy formulation.

Keywords: ARIMA; GDP; United States.

1. Introduction

The GDP of a country represents the economic scale and per capita living standard of the country, and it can also be used to further analyze the economic structure and provide guidance for the country's policy formulation [1]. What’s more, CEOs use GDP to make hiring decisions and forecast sales growth. Short-term capital operators improve investment strategies by studying GDP. White House and Fed officials view GDP as a report card on the effectiveness of their own policy implementation. For one reason or another, the prediction of GDP is particularly important and has attracted the attention of all sectors of society. In addition, as the largest economy in the world, the prediction of the US's GDP is of great economic and practical significance, and has reference and inspiration for China's relevant policy formulation.

Since GDP is the best barometer of the overall direction of the economy, many forecasters carefully analyze it, looking for clues about where the economy is headed. For a long time, scholars at home and abroad have conducted extensive research on time series model forecasting of the US’s GDP. In addition, they have also referred to a large number of international environmental factors, trying to further analyze and explain the research results. However, there are still great deficiencies in the verification and identification of models. Most studies fail to compare the models and find the most suitable prediction model for the data based on the characteristics of the data, and there is no comparison between the actual value and the predicted value, which leads to the final predicted value has a large error with the actual value, and the accuracy is not high, which will greatly mislead policy formulation and further analysis [2].

In order to make up for the deficiencies in previous studies and to provide model reference for other GDP forecasts, the author based on summarizing previous research experience and added his own innovative ideas, trying to provide the most accurate way to predict the development trend of US GDP. Initially, this paper collected the Fred's quarterly U.S. GDP data, which is recorded by National Income and Product Accounts of the United States (NIPA) and the Bureau of Economic Analysis, and visualized it in the form of images, which more clearly showed the basic information and main features of the data [3]. However, considering that the data in the early years may have great interference with the prediction of future data, the author extracted the data from the past 20 years for further analysis and prediction. In order to preliminarily select the data prediction model, the author has decomposed the original data, and the results show that the data has obvious seasonality and an upward trend, so the seasonal ARIMA model "ARIMA (p, d, q) (P, D, Q)" and the Holt-Winter method of the exponential smoothing method should be selected to predict the data. Further, in order to test the effectiveness of the two prediction models, the authors divide the data from the first quarter of 1980 to the fourth quarter of 2019 as the training set and the data from the first quarter of 2020 to the second quarter of 2022 as the validation set. And the accuracy of the model is judged by comparing the difference between the predicted value and the actual value.

For the application of the ARIMA method, the most important thing is to determine the values of p, d and q. The author first performed a first-order difference on the original data to eliminate its obvious fluctuations, so as not to adversely affect subsequent predictions. Subsequently, the author determined the values of each coefficient by producing the autocorrelation diagram and partial autocorrelation diagram of the data after the first-order difference. Finally, this article inputs the value of each coefficient obtains the corresponding prediction result of the ARIMA model and compares it with the actual value to obtain its error.

For the application of the exponential smoothing method, this paper soothes it for a certain number of times on the basis of the original data, and finally strips off the interference factors, so that it shows a simple time series trend, and compares the result with the actual value to get its error. Finally, this article compares the errors of the two forecasting models and obtains the best forecasting model, which provides experience in model selection for forecasting GDP. Subsequently, the author uses the better forecasting model to predict the short-term GDP trend of the United States, conducts a quantitative analysis of the forecast results, and further interprets the analysis results based on the actual situation. The above results not only have a certain reference value for the future development planning and policy formulation of the United States, but also have important implications for the world's economic situation and international pattern.

2. Method

2.1 ARIMA

ARIMA model, which is an acronym for Autoregressive Integrated Moving Average, is a time series forecasting model proposed by American statistician G.E.P. Box and British statistician G.M. Jenkins in the 1970s. Among them, AR is autoregression, I am summation, and MA is moving average. ARIMA model is divided into non-seasonal ARIMA model "ARIMA (p, d, q) "and seasonal ARIMA model "ARIMA (p, d, q) (P, D, Q)".

And the Autoregressive Moving Average Models is represented by the following equation:

$$y_t = c + \emptyset_1 y_{t-1} + \cdots + \emptyset_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{1}$$

The difference between the two is that the latter takes into account the seasonality of the data when forecasting, and is more suitable for data with seasonal or cyclical changes. The data used in this study is the quarterly GDP data of the United States, which has obvious seasonality and has an upward trend quarter by quarter, so the latter is used. where the orders of AR and MA are p and q, and the degree of differencing is d [4].

2.2 Exponential smoothing

Exponential smoothing method was proposed by American scholar R.G.Brown. This method reflects the essential pattern of the sequence by eliminating the obvious random fluctuations of the data, and then predicts the future data by identifying this pattern.

Compared with the moving average method, the exponential smoothing method makes full use of all-time series data and treats all the data involved in the operation with equal weight and its steps are clearer and the calculation is more convenient. At the same time, the exponential smoothing model is suitable for analyzing time series showing autocorrelation, and it is also suitable for data with seasonal and trend changes, which is consistent with the type of US GDP data to be predicted in this study[4].

2.3 Data

The data used in this study are the Fred's quarterly U.S. GDP data, which is recorded by National Income and Product Accounts of the United States (NIPA) and the Bureau of Economic Analysis [3]. The original data range is from 1950 to 2021, which can be seen in Figure 1. However, in order to remove the interference of various historical factors and make a more accurate forecast of US GDP, the author selected data from 1980 to 2021 for analysis. Further, the data from the first quarter of 1980 to the fourth quarter of 2019 is divided into the training set and the data from the first quarter of 2020 to the second quarter of 2022 is divided into the validation set.

Fig. 1 Original Trend Chart of U.S. GDP

In addition, in order to choose a specific forecasting model, the authors factored the raw data to determine its trend and seasonality.

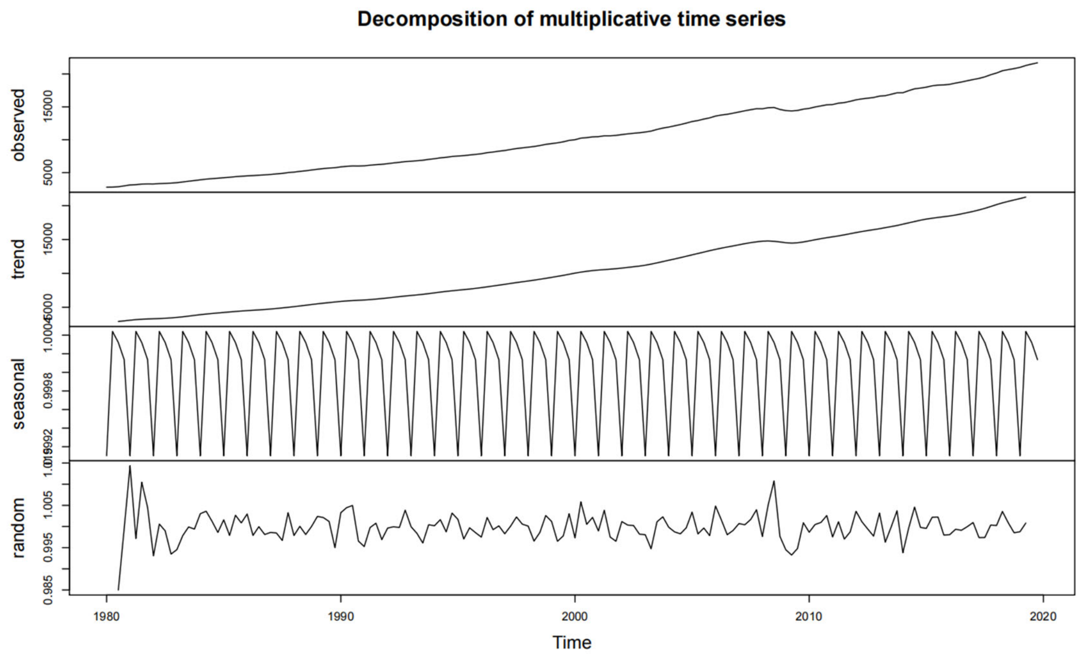

Fig. 2 Decomposition of multiplicative time series

It can be seen from Figure 2 that the data has obvious seasonality and upward trend, so this study plans to use the seasonal ARIMA model "ARIMA (p, d, q) (P, D, Q)" and the Holt-Winter method for forecasting [4].

3. Results

3.1 Establishment of ARIMA

3.1.1 Differentiate the raw data and determine the d value

Since the initial data of this study have some obvious random fluctuations, these fluctuations may interfere with the subsequent forecast results, so it is necessary to perform a differential on the initial US GDP data before forecasting to achieve the effect of smoothing the data[5],[6],[7].

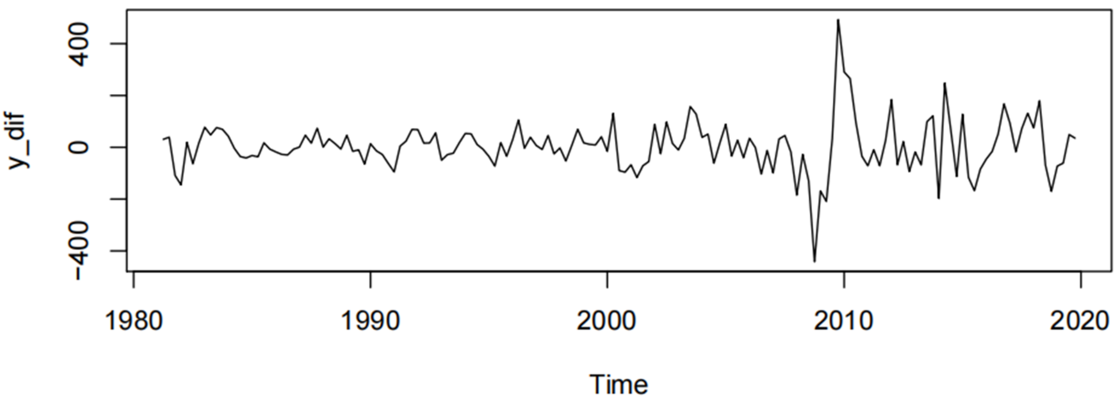

Fig. 3 First-difference plot of raw data

It can be seen intuitively from Figure 3 that the curve after the first difference of the original data is more stable except that the data around 2010 has large fluctuations, and the data fits well with the original data. Therefore, the subsequent ARIMA model predictions are based on one-time smoothed data. So the authors determined the d value in the ARIMA model to be 1.

3.1.2 Generate autocorrelation plots and determine q-values

For the determination of the q value, it is necessary to draw and observe the ACF plot of the series after one difference (as shown in Figure 4).

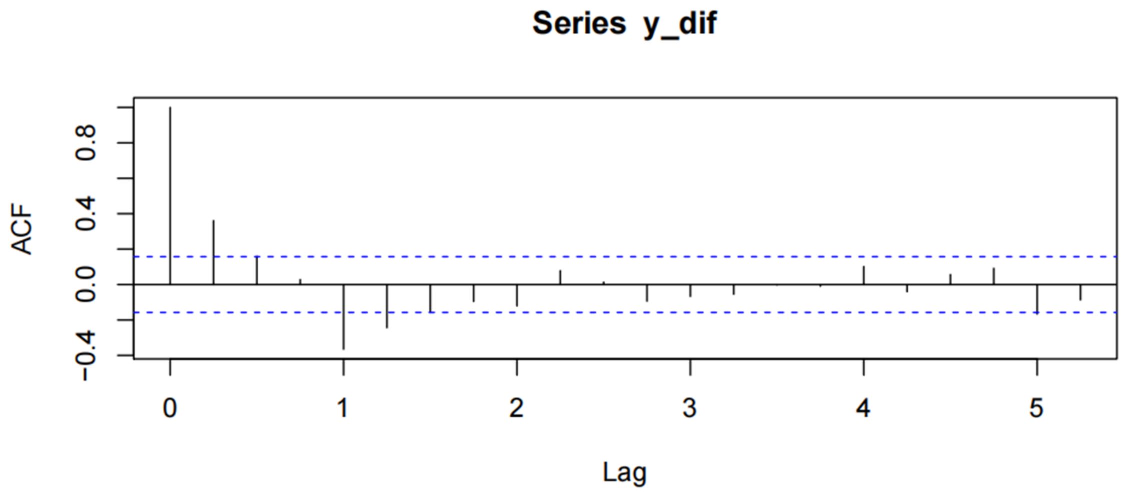

Fig. 4 Autocorrelation plots of the first-order difference data

Figure 4 is the autocorrelation diagram generated after the original data is smoothed once. Except for the first-order function value, the subsequent function values tend to 0 and show tailing, so the value of q is taken as 1.

3.1.3 Generate partial autocorrelation plots and determine p values

For the determination of the p value, it is necessary to draw and observe the PACF plot of the series after one difference (as shown in Figure 5).

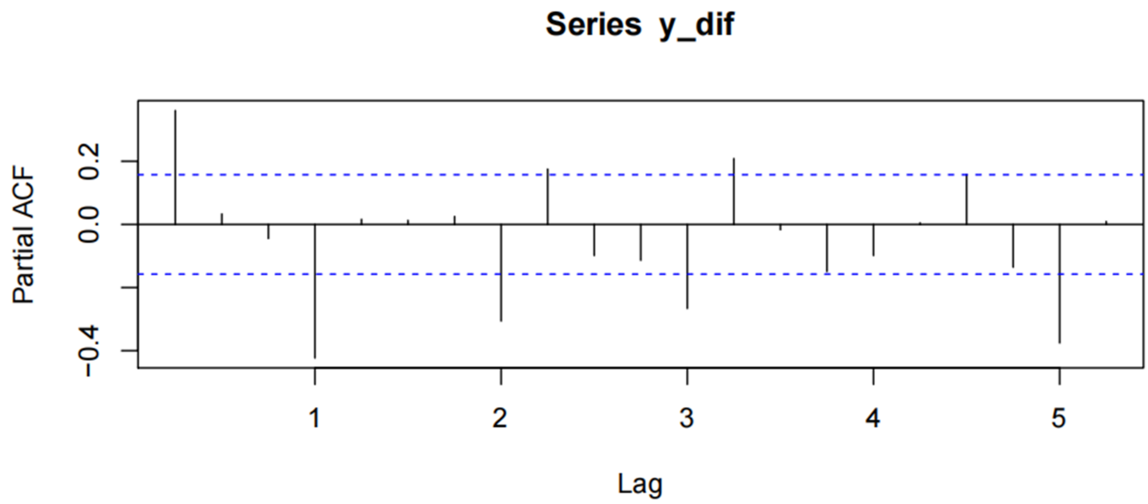

Fig. 5 Partial autocorrelation plots of the first-order difference data

Figure 5 is the partial autocorrelation graph generated after the original GDP data is smoothed once. Since many function values in this graph exceed the baseline and have no obvious trend to 0, so the p value is taken as 0.

In summary, the author considers establishing an ARIMA (1,1,0) (2,1,1) model for subsequent predictions.

3.2 Results of ARIMA

On the basis of determining the p, d, and q values respectively and inputting the smoothed data, the author tested and predicted the GDP of the United States through the ARIMA model, and the results are shown in Figure 6.

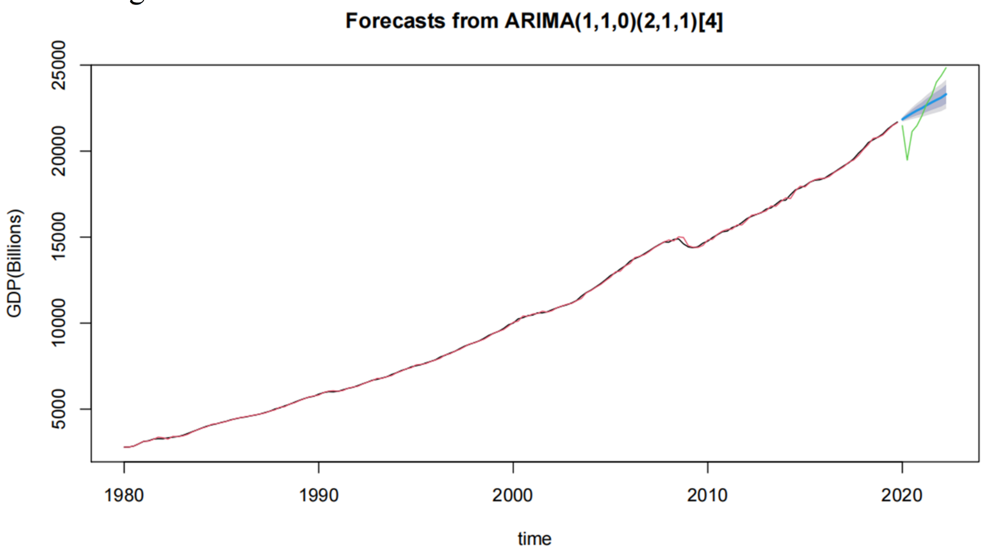

Fig. 6 Forecast from ARIMA (1,1,0) (2,1,1)

As shown in Figure 6, in this test, the author selected the first quarter of 1980 to the fourth quarter of 2019 as the training set, and the first quarter of 2020 to the second quarter of 2022 as the validation set. The red curve is the fitted value generated by the ARIMA model, and the black curve is the real U.S. GDP data. The red and black curves have a very high degree of overlap, indicating that the difference between the ARIMA prediction results and the real value is very small. In addition, through the white noise test on the residuals of the ARIMA prediction results, it can be concluded that the residual sequence is white noise, and the model fits well [4]. For the data in the validation set, the green curve represents the real data, and the blue shading represents the forecast interval, due to the sudden outbreak of COVID-19 in 2020, which wreaked havoc on the global economy, and the U.S. GDP fell sharply as a result. This sudden change cannot be predicted in advance, so ARIMA's prediction interval does not well cover the real GDP from 2020 to 2022, but this unexpected result deviation can be properly tolerated. To sum up, the ARIMA model has a small error in the recent forecast, and is suitable for forecasting and analyzing the future GDP of the United States. However, the choice of the final forecasting method needs to be further compared with the exponential smoothing method.

3.3 Establishment of Exponential Smoothing

From the Figure 2, it is clear that the seasonality can be found in the data, and it shows a linear increasing trend for the US GDP. As a result, the authors use the exponential smoothing method to analyze the US GDP data. Besides, because the seasonal variations are changing proportional to the level of the series, Holt-Winters multiplicative method is chosen [4]. As for the exponential smoothing (Holt-Winters) model, its component form is presented by the following equation [4]:

$$\widehat{y_{t+h|t}} = (\ell_t + hb_t)st + h - m(k+1) \tag{2}$$

$$\ell_t = \alpha \frac{\mathrm{y_t}}{\mathrm{S_{t-m}}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{3}$$

$$b_t = \beta^*(\ell_\mathrm{t} - \ell_{\mathrm{t-1}}) + (1-\beta^*)b_{t-1} \tag{4}$$

$$s_t = \gamma \frac{\mathrm{y_t}}{\ell_{\mathrm{t-1}} + \mathrm{b_{t-1}}} + (1-\gamma)\mathrm{S_{t-m}} \tag{5}$$

k is integer part of (h-1)/m. And with multiplicative method St is in relative terms: within each year ≈m. So m terms vary about 1.

3.4 Results of Exponential Smoothing

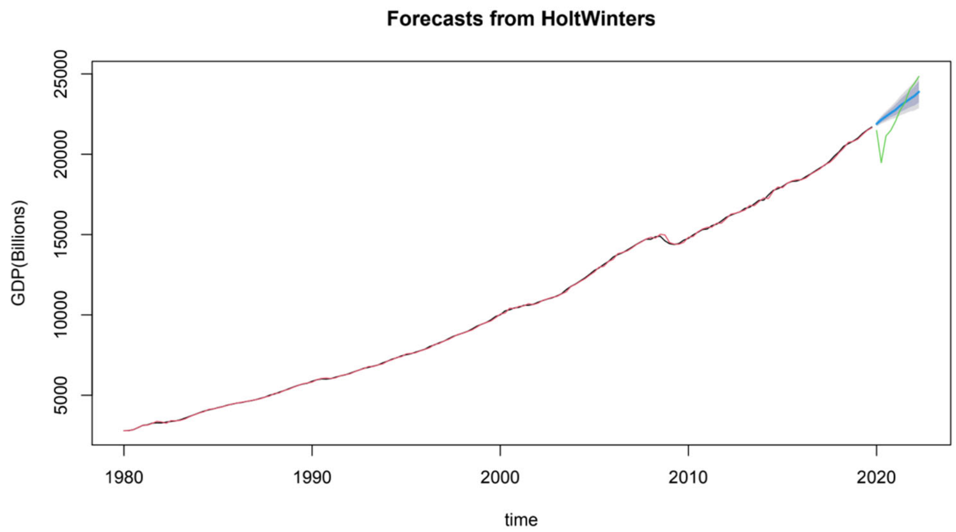

Fig. 7 Forecasts from HoltWinters

In Figure 7 above, the forecasts from Holt-Winters model shares the similar situation with that of ARIMA model, high overlap between the fitted value and the real value. Furthermore, due to the detrimental effect of COVID-19 on the economy, the green line, which indicates the true value for the data set, dropped in 2020 but climbed to 25000 billion in the next few years. It is an unexpected variation for the GDP forecasts because the Corona Virus Disease destroyed a lot of economic activities like the international trade. Thus the GDP declined and there was a gap between the true value and the forecast value.

4. Further Discussion

4.1 Comparison of ARIMA and Exponential Smoothing

In order to more accurately predict the U.S. GDP and thus play a certain enlightening role in the global economic situation, the author made a detailed comparison of the forecast errors generated by the above two forecast models and arranged them in chronological order. The results are shown in Table 1.

Table. 1 Comparison with two different models

Quarterly（year）	Actual Value	SARIMA Predicted value	Holt-Winters Predicted value	SARIMA Error	Holt-Winters Error
1stquarter（2020）	21481.37	21844.87	21887.88	363.50	406.51
2ndquarter（2020）	19477.44	22026.22	22147.38	2548.78	2669.94
3rd quarter（2020）	21138.57	22200.20	22354.38	1061.63	1215.81
4th quarter（2020）	21477.60	22360.16	22561.85	882.56	1084.25
1st quarter（2021）	22038.23	22492.11	22754.34	453.88	716.11
2ndquarter（2021）	22740.96	22670.30	23015.52	-70.66	274.56
3rd quarter（2021）	23202.34	22824.71	23222.13	-377.63	19.79
4th quarter（2021）	24002.81	22972.44	23429.23	-1030.37	-573.58
1st quarter（2022）	24386.73	23110.09	23620.80	-1276.64	-765.93
2ndquarter（2022）	24851.81	23309.65	23883.66	-1542.16	-968.15

As can be seen from the above table, the prediction results of the two models increase with the increase of the number of forecasting periods. In terms of the absolute value of the total error, the SARIMA model (9607.81) is larger than the three-parameter Holt-winters model (8694.63), so its average error is larger than the three-parameter Holt-winters model. Therefore, the prediction accuracy of the three-parameter Holt-winters model is better than that of the SARIMA model, so the three-parameter Holt-Winters model can be used for prediction.

In the first six periods, the prediction accuracy of SARIMA model was significantly higher than that of the three-parameter Holt-Winters model. However, from the seventh period, the prediction accuracy suddenly decreased, indicating that SARIMA model was more suitable for short-term prediction. Although the prediction accuracy of the three-parameter Holt-Winters model is lower than that of the SARIMA model in the first six periods, the prediction accuracy of the three-parameter Holt-Winters model is significantly better than that of the SARIMA model from the seventh period, indicating that the three-parameter Holt-Winters model is suitable for the data with obvious linear trend (increasing or decreasing) and seasonality. To sum up, the appropriate prediction model should be selected according to the specific data characteristics and prediction length.

4.2 Forecast based on exponential smoothing

From the above analysis, it can be concluded that the average accuracy of the exponential smoothing method is slightly higher than that of ARIMA, so the author uses the exponential smoothing method to predict the US GDP index by the fourth quarter of 2023. The results are shown in Table 2:

Table. 2 US GDP index

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020 Q1	21887.88	21781.51	21994.25	21725.20	22050.56
2020 Q2	22147.38	21987.72	22307.05	21903.20	22391.57
2020 Q3	22354.38	22138.49	22570.27	22024.20	22684.55
2020 Q4	22561.85	22307.33	22816.36	22172.59	22951.10
2021 Q1	22754.34	22427.44	23081.24	22254.39	23254.29
2021 Q2	23015.52	22627.13	23403.92	22421.52	23254.29
2021 Q3	23222.13	22768.80	23675.46	22528.82	23915.44
2021 Q4	23429.23	22923.64	23934.83	22655.99	24202.48
2022 Q1	23620.80	23030.90	24210.70	22718.63	24522.97
2022 Q2	23883.66	23220.94	24546.39	22870.12	24897.21
2022 Q3	24089.88	23352.24	24827.52	22961.76	25218.00
2022 Q4	24296.62	23496.69	25096.55	23073.23	25520.01
2023 Q1	24487.26	23591.79	25382.73	23117.75	25856.77
2023 Q2	24751.81	23773.03	25730.59	23254.89	26248.72
2023 Q3	24957.63	23894.70	26020.56	23332.02	26583.24
2023 Q4	25164.01	24030.47	26297.55	23430.41	26897.61

In order to reflect the changing trend of the forecast results more clearly, and analyze the forecast results based on the actual situation, the author conducts a quantitative analysis on the forecast results.The results are shown in Table 3:

Table. 3 Quantitative analysis on the US GDP index

Point	Forecast	Difference	Proportion
2020 Q1	21887.88	—	—
2020 Q2	22147.38	259.50	1.186%
2020 Q3	22354.38	207.00	0.935%
2020 Q4	22561.85	207.47	0.928%
2021 Q1	22754.34	162.49	0.853%
2021 Q2	23015.52	261.18	1.148%
2021 Q3	23222.13	206.61	0.898%
2021 Q4	23429.23	207.10	0.892%
2022 Q1	23620.80	191.57	0.818%
2022 Q2	23883.66	262.86	1.113%
2022 Q3	24089.88	206.22	0.863%
2022 Q4	24296.62	206.74	0.858%
2023 Q1	24487.26	190.64	0.785%
2023 Q2	24751.81	264.55	1.080%
2023 Q3	24957.63	205.82	0.832%
2023 Q4	25164.01	206.38	0.827%

Based on the above analysis results, it can be concluded that in the next year, the GDP of the United States will maintain a steady upward trend, and is expected to reach about 25164.01 in the fourth quarter of 2023, with an average growth rate of about 0.9344%. The highest growth rate is 1.080% in the second quarter of 2023, and the lowest growth rate is 0.785% in the first quarter of 2023. The overall growth trend is relatively stable. In addition to this, the growth rate is cyclical, with a peak every four forecast periods, followed by three successive declines until the next peak. The peaks are also decreasing round by round.

To sum up, as the world's largest economy, the economic foundation of the United States is still solid. In the absence of major changes, the future GDP and development trend will remain stable. However, because the current economic volume of the United States is very large and the basic economic construction has been perfected, the speed of economic improvement is declining, which is also in line with the famous theory of diminishing returns[8]. In addition, the first quarter of each year has the fastest GDP growth. What’s more, considering that the economic damage caused by the COVID-19 outbreak in 2019 is gradually easing, the economic growth rate will increase slightly in the coming year [9].

5. Conclusion

From the comparison of the forecast results from the two models, the authors have reached the following conclusions. Admittedly, both the ARIMA and Holt- winters models preserve high degree of fitting and smaller difference compared with real-world time series. However, with the extension of the evaluated period of time, the accuracy of both models declines. When comparing the deviation of the two models from the perspective of the whole series, one can conclude that Simple Exponential Smoothing (Holt-Winters trend) method actually has smaller mean error than ARIMA. In contrast, when focusing on the first six periods within the prediction interval, one can safely conclude that ARIMA is much more precise and fits the trend of the seasonal change of the research object US gross domestic product better.

Based on the previous achievement and this research, the authors also conclude several suggestions for other researchers who are analyzing the Gross Domestic Product time series. This article asserts that ARIMA model should be given utmost attention during the first stage of a research, for its effectiveness in forecasting the US Gross Domestic Product in short run, regardless of the model’s relatively complexity of model construction. Exponential Smoothing methods, (Holt-Winters in this research) on the other hand, should be profoundly utilized when forecasting the US Gross Domestic Product in long run. The Exponential Smoothing methods also have the advantage that the models are much easier to build and analyze because the fact that the basis of Exponential Smoothing is the historical data that has been already obtained [10].

This research has revealed the internal relationship and given a practical situation that evaluates the efficiency and effectiveness between two of the most repeatedly used models. The methodology, data sets selection, discussion parts in this research can contribute to other similar researches, serving as a form of model and instruction in the field of economical time series analysis. However, this research, does not present further and more general analysis, which may become the research motivation for other researchers.

Studying the time series of gross domestic product of the US in this research provides basis of the creation, revision and innovation for the US fiscal and monetary policies. This research can also play an encouraging role in the formulation of Chinese economic policies by giving an US reference to the Chinese economic growth.

References

[1] Xiao Shan. Research on the Influencing Factors of China's Gross National Product Based on Statistical Regression. Economic Management Digest, 2021(03):197-198.

[2] Ximin Wang & Jianwu Qi. Application of ARIMA model and exponential smoothing method in GDP prediction of Tianshui City. Journal of Lanzhou University of Arts and Sciences (Social Science Edition) 2017(02), 54-58.

[3] Fred.stlouisfed.org. 2022. Gross Domestic Product. https://fred.stlouisfed.org/series/GDP

[4] Hyndman R J , Athanasopoulos G . Forecasting: Principles and Practice. London: Bowker-Saur. Pharo, 2014.

[5] Yulong Shan, Dingqi Yan. Application of ARIMA and Exponential Smoothing in GDP Forecasting. 2014.

[6] An Xiaodan, Li Xiaoxia. Prediction and analysis of Yuncheng GDP based on ARIMA model. Journal of Yuncheng University, 2022, 40(03): 69-73+90.

[7] Fan Hengrui, Ren Lixiu. Application of ARIMA and Exponential Smoothing in GDP Forecasting in Jiangsu Province. Jiangxi Agricultural Journal, 2011, 23(02): 187-189.

[8] Wang Jingcai, Fang Jiaqin, Qiao Lili. The law of diminishing marginal returns. Inner Mongolia Petrochemical, 2003(04):205.

[9] Xu Shilin. ARIMA model research and its application in US GDP forecasting. Proceedings of the 2016 International Symposium on Smart City and Informatization Construction III. 2016:44-45.

[10] Zhang Denan, Zhang Xinyan. Determination of Smoothing Coefficient in Exponential Smoothing Forecasting Method. Journal of Dalian Railway Institute, 2004(01):79-80.