Comparison of Forecasting Effect of SARIMA Model and Holt-Winters Smoothing based on GDP

Chenghan Li

University of Sydney, Sydney NSW2050, Australia

Abstract: This study analyses ARIMA Model and Holt-Winters Smoothing. In this report, the GDP of Australia is taken as an example to illustrate the differences between these two models and to decide which model could show a better performance in term of forecasting. RMSE (Root Mean Square Error) is used as an indicator to compare the forecasting results of ARIMA Model and Holt-Winters Smoothing. After analysing, Holt-Winters Smoothing is found that could provide a more accurate result. This report provides people with basic ideas that how to use basic forecasting techniques to explore the future trends of some economic indicators.

Keywords: SARIMA Model; Holt-Winters Smoothing; GDP

1.Introduction

GDP provides a basis for the estimation of economic growth and size. It is used as a measurement of the economic health of a country since it gives an epitome of a country's economy (Landefeld, Seskin & Fraumeni, 2008). In addition, GDP is an important factor that could influence decisions made by economists, policymakers, investors, and so on.

In this report, Holt-Winters Smoothing and SARIMA model are used to predict GDP. The basic idea of Holt-Winters Smoothing is to use the exponential smoothing method to analyze time-series data and forecast the future. ARIMA model is a generalization of an autoregressive moving average (ARMA) model which is wildly applied in different subjects since this model can capture underlying patterns more than trend or seasonality. RMSE is used in this report to determine which model is more suitable to forecast GDP. The data used in this report is Australian quarterly GDP in 60 years. The data between 1959 and 2015 are used as training data and the rest are used as test data. After analyzing, in this report, additive Holt-Winters smoothing model is more suitable to forecast GDP because this method shows the lowest RMSE which is 3.2156.

2. Exploratory Data Analysis

In this report, GDP is presented on quarterly basis, and there are 228 training data in total. This part discusses the pattern of the gross domestic product based on the plot. The trend, seasonal patterns and smoothness will be revealed through the moving average.

2. 1 Original data

According to the database, the data are classified quarterly, so the seasonal pattern of data is 4. The Figure 2.1 shows fluctuation during the whole period, and also increases in this period.

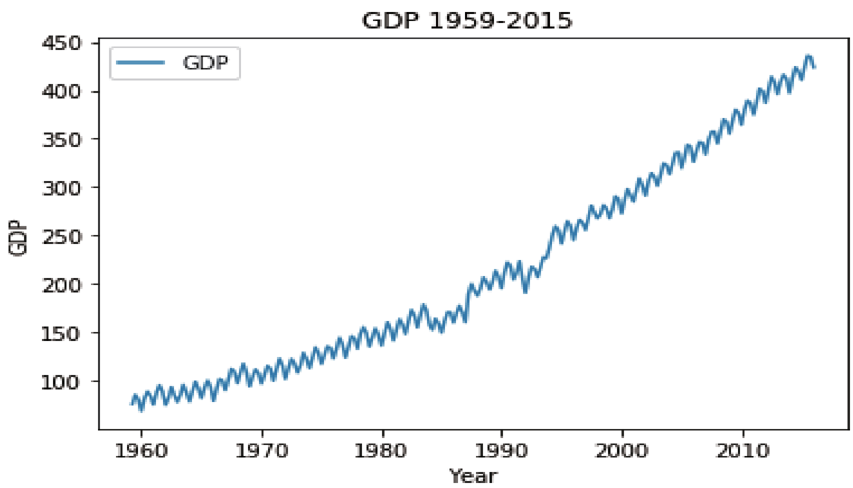


Figure 2.1. GDP between 1959 and 2015.

The log transform of GDP is shown in below Figure.

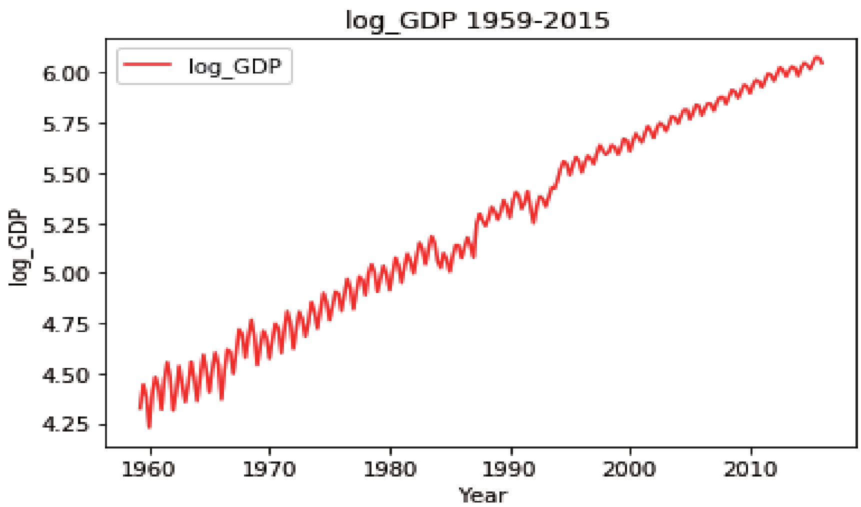


Figure 2.2. Log GDP between 1959 and 2015.

2.1.1. Identifying trend and seasonal Patterns

The trend is a long term upward or downward direction in the time series, and seasonal pattern is a regular repeating pattern that repeats in a fixed period. The data for the gross domestic product (GDP) is shown in Figure 2.1. From the plot, it is clear that there is an overall increase in the trend, with some seasonality in it. There are some irregular fluctuations from 1985 to 2000 that do not follow the regular pattern. The gradually increasing trend also indicates the additive relationship of the data. Its statistical properties (mean, variance) are changing over time. Therefore, this report needs to run some data preprocessing to get stationary data before this report applies the forecasting techniques. There is a visible seasonal pattern for GDP data, which is suitable for using the decomposition method to train the data. 2.1.2 Identifying Stationarity

Using the Dickey-fuller Test method to identify the stationarity of the original data, this report can recognize it is non-stationary. The test statistic (3.8651) is larger than the critical value (2.5741) at 10% which is shown in Figure 2.3. In our case, the number is larger than the critical value at 10%, which means the data is non-stationary (at least 90% of the data).

Results of Dickey-Fuller Test:	
Test Statistic	3.865121
p-value	1.000000
Lags Used	15.000000
Number of Observations Used	212.000000
Critical Value (1%)	-3.461578
Critical Value (5%)	-2.875272
Critical Value (10%)	-2.574089

Figure 2.3. Results of Dickey-Fuller Test.

2.1.3. Decomposition method

From the data for the gross domestic product (GDP) showed in Figure 2.4. It is clear that there is an overall increase in the trend, with some seasonality in it. There are some irregular fluctuations from 1985 to 2000 that do not follow the regular pattern. The gradually increasing trend also indicates the additive relationship of the data. Therefore, this report needs to use natural logarithm to get stationary data before this report applies the forecasting techniques.

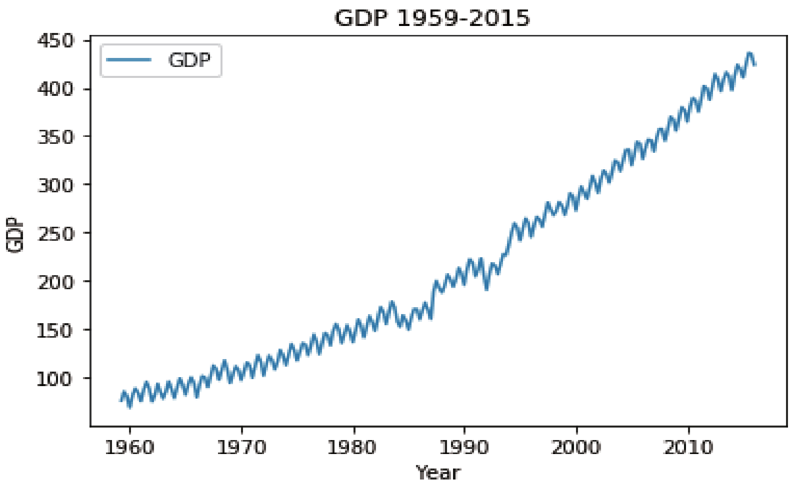


Figure 2.4. Original data.

3.Holt-Winters smoothing

3.1. Methodology

Holt–Winters smoothing is an advanced exponential smoothing method that can be applied to deal with seasonal and trend time series， and this method is regarded as double exponential smoothing (Gelper, Fried, & Croux, 2010). The basic idea of this method is to use exponential smoothing method to analyze trend, level, and seasonal components of past data and then forecast the future (Holt, 2004). In addition, Holt–Winters smoothing can be applied to both Additive seasonality and Multiplicative seasonality.

3.1.1 Additive Holt-Winters Smoothing

According to Hyndman and Athanasopoulos (2018), additive Holt-Winters smoothing is used for time series whose seasonal variation is not changing follow the trend . The model is:

the forecast: $\hat{y}_{t+h|1:t} = l_t + hb_t + s_{t-M+(h \bmod M)}$

the level: $l_t = \alpha(y_t - s_{t-M}) + (1-\alpha)(l_{t-1} + b_{t-1})$ $(0 \le \alpha \le 1)$

the trend: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$ $(0 \le \beta \le 1)$

the seasonal variation: $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-M}$ $(0 \le \gamma \le 1)$

where α ,β , and γ are weight decaying and M is seasonal frequency.

3.1.2 Multiplicative Holt-Winters Smoothing

Multiplicative Holt-Winters smoothing is used for seasonal variation is not constant along the trend (Hyndman & Athanasopoulos, 2018), and the model is:

the forecast: $\hat{y}_{t+h|1:t} = (l_t + hb_t)S_{t-M+(h \bmod M)}$

the level: $l_t = \alpha \frac{y_t}{s_{t-M}} + (1-\alpha)(l_{t-1} + b_{t-1})$ $((0 \le \alpha \le 1))$

the trend: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$ $(0 \le \beta \le 1)$

the seasonal variation:

$$s_t = \gamma \frac{y_t}{(l_{t-1} + b_{t-1})} + (1-\gamma)s_{t-M}$$

where α ,β , and γ are weight decaying and M is seasonal frequency.

Above-mentioned models are two kinds of Holt-Winters smoothing methods used for different seasonal variation. However, before using these models in reality, there are some initial values ($l_0, b_0, s_0, s_{-1}, \ldots, s_{-M+1}$) and 3 parameters (α ,β , and γ) need be determined. According to Hyndman and Athanasopoulos (2018), firstly, a linear regression based on the data ($y_1, y_2, \ldots, y_t$) can be used to find l_0 and b_0 and the linear regression model is $\hat{y}_t = l_0 + b_0 t$. In addition, $s_0, s_{-1}, \ldots, s_{-M+1}$ are the averages of $\hat{s}_t$ and $\hat{s}_t$ is equal to $y_t - \hat{y}_t$.

Secondly, the minimizing SSE/MSE method can be used to select parameters (α ,β , and γ). In terms of forecasting, the forecasting equation of additive Holt-Winters smoothing is $\hat{y}_{t+h|1:t} = l_t + h b_t + s_{t-M+(h \bmod M)}$.

The variance for interval forecasts:

When $h \le$M,

$$Var_{(y_{t+h}|y_{1:t})} = \sigma^2(1 + \alpha^2 \sum_{i=1}^{h-1} (1 + i\gamma)^2)$$

When $h >$ M,

$$Var_{(y_{t+h}|y_{1:t})} = \sigma^2(1 + \sum_{i=1}^{h-1} [\alpha(1 + i\gamma) + l_{i,M}\delta(1-\alpha)]^2)$$

If i is an integer, $l_{i,M}$ is equal to 1. Otherwise, $l_{i,M}$ is equal to 0.

The forecasting equation of Multiplicative Holt-Winters smoothing is:

$\hat{y}_{t+h|1:t} = (l_t + h b_t)s_{t-M+(h \bmod M)}$

However, if the trends are extrapolated indefinitely into the future, there cloud be some problems. In order to solve such problems, dampened trend exponential smoothing can be used. The model of damped trend exponential smoothing is:

$l_t = \alpha(y_t - s_{t-M}) + (1-\alpha)(l_{t-1} + \varphi b_{t-1})$

$b_t = \gamma(l_t - l_{t-1}) + (1-\gamma)\varphi b_{t-1}$

$S_t = \delta(y_t - l_t) + (1-\delta)S_{t-M}$

$y_{t+1} = l_t + \varphi b_t + S_{t-M+1} + \varepsilon_{t+1}$

Where φ is the dampening and $(0 \le \varphi \le 1)$.

3.2. Forecast processing and analysis

Firstly, we use the training data to draw the plot.

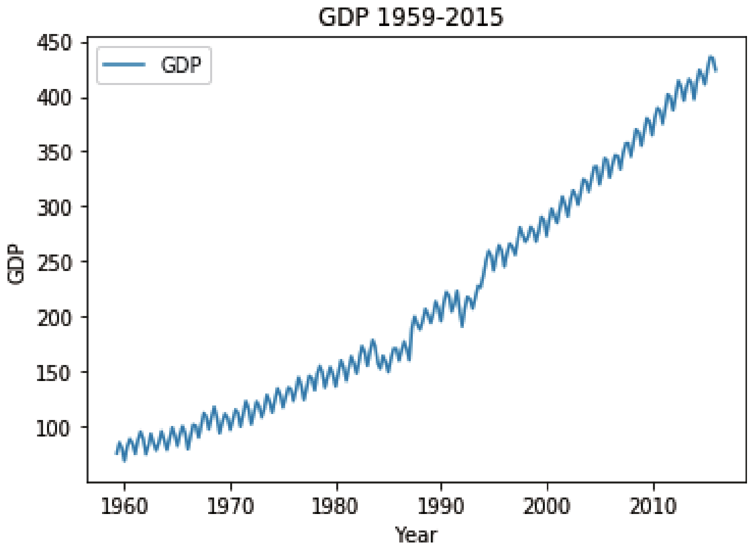


Figure 3.1. GDP from 1959 to 2015.

From Figure 3.1, it shows that although there are obvious fluctuations between 1980 and 2000, the trend of GDP increases from 1960 to 2010. In addition, above plot shows that the general seasonal variation is not change follow the trend and the model, which suggests that Additive Holt-Winters smoothing model should be used. However, subjective observation does not accurate. As a result, SSE is used to select smoothing methods.

Based on Python, SSE of additive Holt-Winters smoothing is 2480.0602 and SSE of multiplicative Holt-Winters smoothing is 3056.7672. The SSE of multiplicative Holt-Winters smoothing is higher than additive Holt-Winters smoothing. As a result, additive Holt-Winters smoothing method is used in this report.

Secondly, original data is split into train and test, where train data is used to train and validation, and additive Holt-Winters smoothing is used to do in-sample forecast and calculate RMSE. In this report, 10% of train data is used as validation data.

Based on Python, RMSE is 3.2156. And then this RMSE is applied to compare with other models in this report to select which one is the suitable to do further forecast. In-sample forecast plot as follow (red line shows the results of forecast):

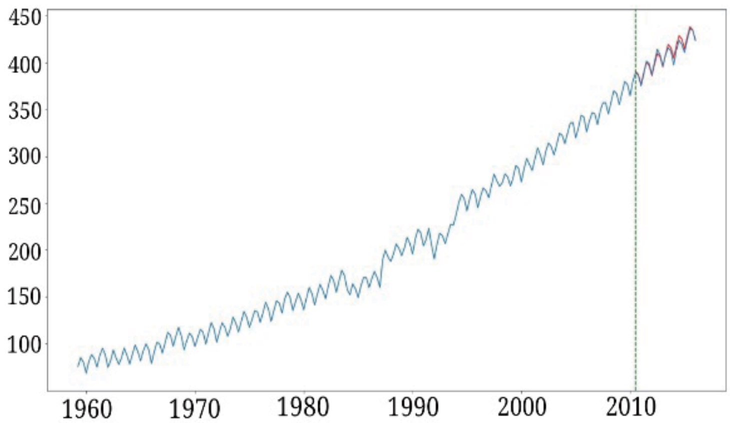


Figure 3.2. In sample forecast.

Above graph shows that the result of forecast is close to the true value.

Finally, additive Holt-Winters smoothing is used to do out-of-sample forecast and then test error is calculated. Below chart shows the result of out-of-sample forecast,

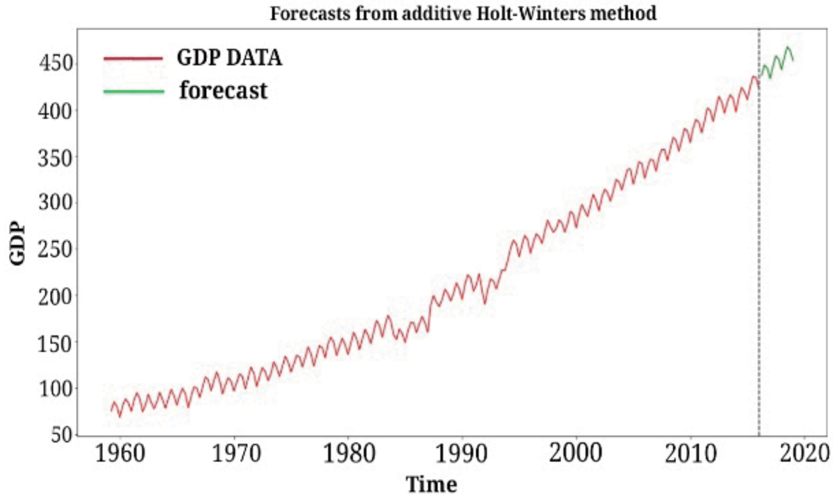


Figure 3.3. "Forecasts from additive Holt-Winters method".

The test error is calculated by follow formula:

$$test\_error = \frac{1}{12}\sum_{1}^{12}(\hat{y}_{T+h|1:T} - y_{T+h})^2$$

And the test error is 197216.4868.

3.3 Limitation

Additive Holt-Winters smoothing method contains l_t(the level or the smoothed value), b_t (the trend) and s_t(seasonal component) which is a useful method to analyse time series data. This method gives greater weight for the more recent observations and decays exponentially as the observations fall at an older time. Although it contains level, trend and seasonal

component, it is very vulnerable to outliers and the weight for α, γ and δ. Sometimes, using this method on the analysis and forecast of frequently changing time series data may also lead to high RMSE.

4. Seasonal Autoregressive Integrated Moving Average (SARIMA)

4.1 Methodology

Autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model, which is wildly applied in statistics, econometrics and time series analysis, since ARIMA can capture underlying patterns more than trend or seasonality. However, ARIMA requires strong assumption on stationarity, which means the mean, variance and covariance do not variate over time for the given time series data. In order to use this method, we can transform time series into a stationary data by taking first or second order of differencing.

The ARIMA (p, d, q) model can be presented in the formula as

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon'_{t-1} + \dots + \theta_q \varepsilon'_{t-q} + \varepsilon_t$$

Where y'_t is the differenced data in order of d, on the right side of the equation, the part with lagged value upon lag p represents autoregressive model with lag p or AR(p), the rest part with lagged value upon lag q represents moving average model with lag q or MA(q). ARIMA is essentially the combination of AR and MA. In a special case of p = q = 0, ARIMA turns to be a white noise process.

According to the special characteristic of MA and AR model, we can use ACF and PACF plot to assess time series stationary and find the appropriate value for p and q. ACF measure the autocorrelation between lagged observations, while PACF is the partial autocorrelations which measures the linear correlation between lagged values by removing the common effect between two lagged values. Under the condition of stationary, ACF for MA would cut off after lag q while PACF would die down exponentially. Similarly, ACF for AR would die down exponentially while PACF would cut off to zero after lag p.

However, Akaike's Information Criterion (AIC) is also a popular method in selecting the optimal number of orders for an ARIMA model as AIC place a penalty parameter into the function, which helps to select the optimal order and simplify the model. An improved method is corrected AIC (AICc), which is similar to AIC but penalized extra order or parameter more heavily, thus, it often produces a simpler model than AIC. Both AIC and AICc are based on the assumption of normal distribution of residuals. There is also another advanced measurement model called Schwarz's Bayesian Information Criterion (BIC), it penalizes the model complexity more heavily than AIC, with enough data, BIC is in advantage of higher probability to select the true model.

Furthermore, in order to apply ARIMA to seasonal time series, seasonality should be considered. Therefore, this report uses SARIMA to analyse and forecast GDP. Seasonal ARIMA can be expressed as ARMA (p, q) (P, Q) $_m$ which has mathematical form of

$$(1-\sum_{i=1}^{P} \emptyset_i B^{im})(1-\sum_{i=1}^{P} \emptyset_i B^{i})Y_t=c+(1+\sum_{i=1}^{Q} \theta_i B^{im})(1+\sum_{i=1}^{q} \theta_i B^{i})\varepsilon_t$$

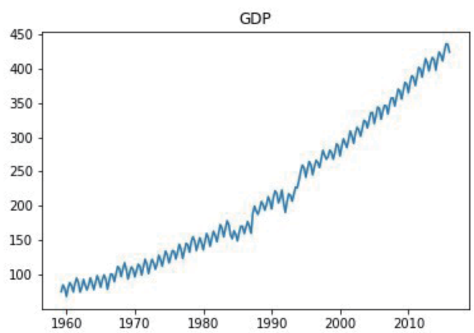
The backshift operator is B which can be used to show lagged observed value,

$$B^k Y_t = Y_{t-k}$$

Seasonal ARIMA shares similar component to ARIMA but add the elements of seasonal components of ARMA (P, Q)$_m$. Seasonal ARIMA also requires stationarity, we can assess this by checking the ACF and PACF plot, the rational is same as regular ARIMA.

4.2 Data processing

4.2.1 Log transformation



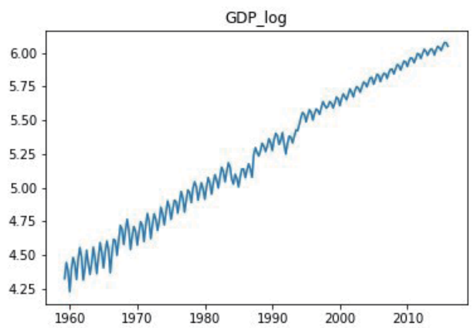
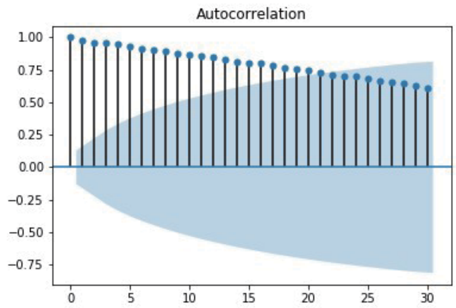


Figure 4.2.1. original GDP data. **Figures 4.2.2. log GDP data.**

Figure 4.2.1 shows the variation of original GDP data, which illustrates a significant increasing trend and constant stable seasonal pattern, meanwhile the peak within each period is quite constant in the datasets. Since the original GDP data has large increase in the time window, we applied log transformation to stabilize the magnitude. After comparing the log plots with the original plots, the variability of data got smaller but seemed have become non-stationary as shown in figures 4.2.2.



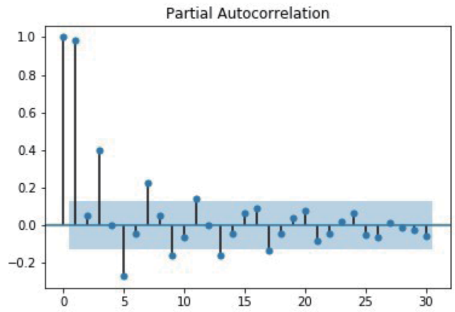
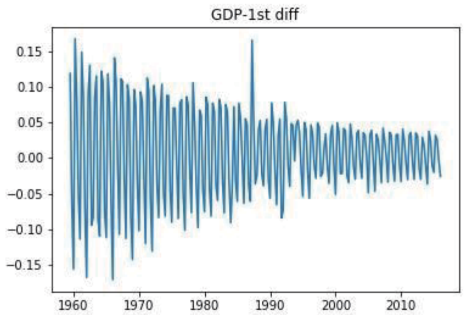


Figure 4.2.3. ACF of log GDP. **Figure 4.3.4. PACF of log GDP.**

ACF and PACF plot of log GDP are made of the log GDP data, plots are shown in Figures 4.2.3 and 4.3.4 respectively. From Figure 4.2.3 it is noted that the log data needs transformation for it dies down extremely slowly.

4.2.2 1-st order differencing



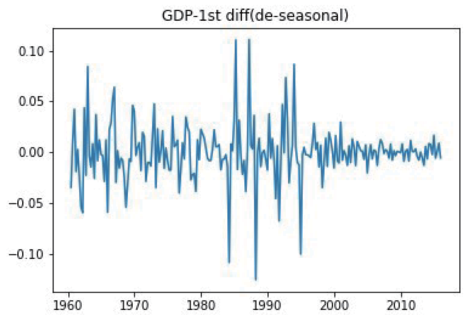


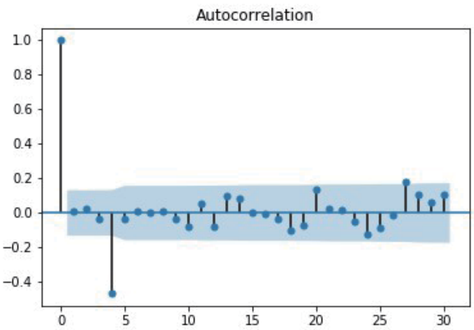
Figure 4.2.5. 1st order differencing. **Figure 4.2.6. 1st order differencing by 4-step shift.**

Therefore, we need to apply differencing method until we decide on an appropriate order. The outputs of 1st order differencing for GDP and the one without seasonal patterns by 4-step differencing are shown as above respectively. Figure 4.2.5 shows that the 1st order differencing for log GDP is still non-stationary after the first differencing. Then after differencing by 4-step shift, it becomes stationary based on the figure 4.2.6. So, d of GDP in SARIMA model is equal to 1 and D = 1.

4.3 Model fitting and Interpretation

4.3.1 ACF & PACF

Here, ACF and PACF plots are drawn based on the 1st order differenced data to obtain a suitable order for SARIMA process. The outputs are shown as below:



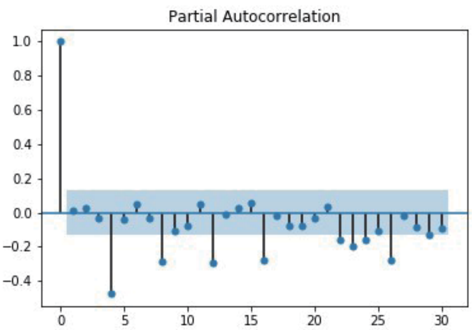


Figure 4.3.1. ACF for differencing data. **Figure 4.3.2. PACF for differencing data.**

From figure 4.3.1, it indicates that the data can be considered stationary after differencing since the ACF of this time series cut off reasonably quickly. Moreover, there exists obvious seasonal pattern since lag (4) largely out of the confidence interval, SARIMA model should be used. Then we could select m = 4 as the seasonal pattern is shown at lag (4) and D = 1 since the series has a stable seasonal pattern. From the plots above, we decided to select q = 4 or 0 from the ACF plot and p = 4 from the PACF plot which are out of the confidence interval and cut off at the next point. And P = 0 and Q = 1 as in PACF lag (4) < 0 and in ACF lag(4) < 0. So, the SARIMA model can be built as SARIMA (4, 1, 4) (0, 1, 1)4 or SARIMA (4,1, 0) (0, 1, 1)4.

4.3.2 Model Fitting

The training set after log transformation has been used to fit the two models above. First, the training data has been split to sample training set which covers 90% and sample testing set which covers 10% of the total training data.

The output of comparison between the two model's predictions in the training set is shown as below. Figures 4.3.3 and 4.3.4 shows that the prediction from SARIMA (4, 1, 4) (0, 1, 1)4 are even higher than the prediction from SARIMA (4, 1, 4) (0, 1, 1)4 compared with the actual GDP data.

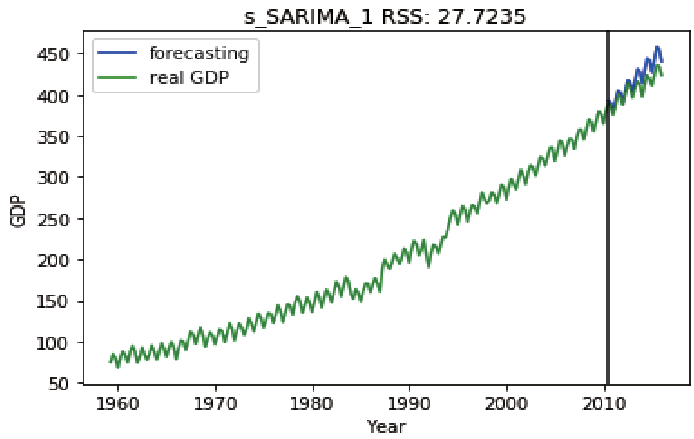


Figure 4.3.3. SARIMA (4,1, 4) (0, 1, 1)4.

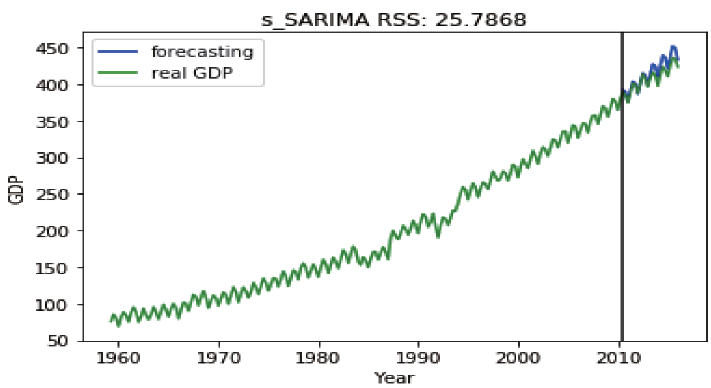


Figure 4.3.4. SARIMA (4,1, 0) (0, 1, 1)4.

The coefficients are all insignificant expect for ma.S.L4. The p-value of Ljung-Box for SARIMA (4, 1, 0)(0, 1, 1)4 is less than 5% which means that we reject the null hypothesis and we can conclude that at least one of the sample auto-correlations in the GDP data are significantly different to 0, while SARIMA (4, 1, 4)(0, 1, 1)4 has p-value greater than 5%, so we may conclude that all of the sample auto-correlations in the GDP data are significantly different to 0. Thus, we may need to include more lags to SARIMA (4, 1, 0) (0, 1, 1)4 in order to solve this auto-correlation problem to make data stationary with better prediction performance. Both SARIMA (4, 1, 4) (0, 1, 1)4 and SARIMA (4, 1, 0) (0, 1, 1)4 have lower than 5% p-value for Jarque-Bera test which implies that these two models may produce normal distributed

GDP data. Then, AIC is also implemented to select the order (p, q) of the SARIMA model. Comparing to the AIC score, SARIMA (4, 1, 0) (0, 1, 1)4 produces lower AIC score, it shows that SARIMA (4, 1, 0) (0, 1, 1)4 shows better performance in prediction and is chosen finally.

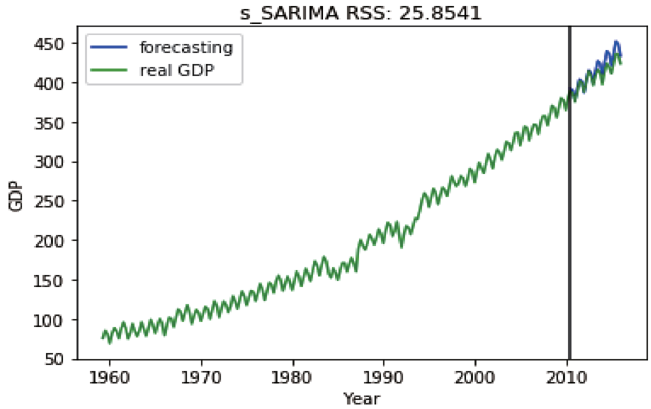


Figure 4.3.5. SARIMA (4, 1, 1) (1, 1, 1)4.

After that, three other models are tested and compared in the same way as above. SARIMA (4, 1, 1) (1, 1, 1)4 has the best performance among all, therefore the model's parameters are selected as the final model. The prediction of SARIMA (4, 1, 1) (1, 1, 1)4 and performance summary are represented in figure 4.3.5.

4.3.3 Prediction of SARIMA

In this way the whole training set is used to fit the final model SARIMA (4, 1, 1) (1, 1, 1)4 and made forecast about the time window in testing set. The result of the forecast is shown as below:

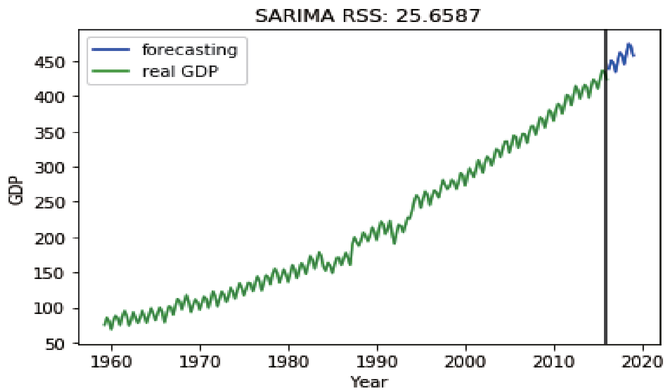


Figure 4.3.6. Final SARIMA model with forecasting.

4.4 Limitation

SARIMA is also one useful method to analyze many types of time series data. However, SARIMA model requires the observations to be stationary (Vijayakumar & Vennila, 2016) and it is the reason why this report uses the natural logarithm of original data to fit the model.

Moreover, SARIMA model cannot be applied to estimate the instantaneous relationship between two time series data (Chamlin, 1988). SARIMA model also requires the utilization of rather long time series to produce reliable parameter estimates (Vijayakumar & Vennila, 2016).

5. Conclusion and Recommendation

5.1. RMSE

Model	RMSE
HoltWinters	3.2156
SARIMA	85.2736

Table 6.1. Test Error.

The above table shows RMSE of different models when doing in-sample forecast. It is found that the RMSE of Holt-Winters smoothing method is the smaller. As a result, Holt-Winters smoothing is chosen to do the out-of-sample forecasting and the test error is 197216.4868.

5.2 Conclusion and Recommendation

After checking the RMSE of the remaining 20% data, additive Holt-Winters smoothing model shows the lowest RMSE, which means this model is more accurate in predictions. Therefore, this report chooses the forecasting value of additive Holt-Winters smoothing model as the result. However, due to the data (quarterly GDP) is very flexible, this model may be vulnerable to errors. If the test error between actual quarterly GDP and predicted quarterly GDP is in an acceptable area, this model can be proved to be useful for forecast. Although there should be many well-performed models which can predict more accurately, this report cannot use them for analysis and forecast due to the limited ability of author. Finally, this report gives the opportunity for researchers to discover and use models to solve real world problems.

Reference

1. Landefeld, J. S., Seskin, E. P., & Fraumeni, B. M. Taking the pulse of the economy: Measuring GDP [J]. Journal of Economic Perspectives 2008; 22(2): 193-216.
2. Gelper, S., Fried, R., & Croux, C. Robust forecasting with exponential and Holt–Winters smoothing [J]. Journal of forecasting 2010; 29(3): 285-300.
3. Holt, C. C. Forecasting seasonals and trends by exponentially weighted moving averages [J]. International journal of forecasting 2004; 20(1): 5-10.
4. Hyndman, R. J., & Athanasopoulos, G. Forecasting: principles and practice [J]. 2018.
5. Vijayakumar, N., & Vennila, S. A comparative analysis of forecasting reservoir inflow using ARMA model and Holt winters exponential smoothening technique [J]. International Jour. of Innovation in Science and Mathematics 2016; 4(2): 85-90.
6. Chamlin, M. B. Crime and arrests: An autoregressive integrated moving average (ARIMA) approach [J]. Journal of Quantitative Criminology 1988; 4(3): 247-258.