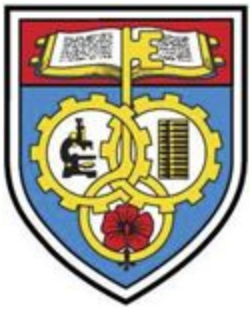

KOLEJ UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY COMPUTING AND INFORMATION TECHNOLOGY

Assignment

BMMS2074 Statistics for Data Science

May / June 2020 Semester

Student Name	Student ID	Signature	Contribution	Marks
Lim Ming Jun	20WMR08879		25%	
Kow Yee Hui	20WMR08868		25%	
Kong Mun Jun	20WMR08867		25%	
Lee Jun Xian	20WMR08873		25%	
Programme: RDS2				
Tutorial Group: 3		Tutor: Dr. Chin Wan Yoke		

Table of Content

Introduction

Alibaba Group Holding Limited, also known as Alibaba Group or as Alibaba.com, is a Chinese Multinational Technology Company specializing in e-commerce, retail, Internet, and technology. It was founded on 4 April 1999 and launched by Jack Ma and a group of 17 friends. The company provides Consumer-to-consumer, Business-to-consumer, and Business-to-business sales services via Web Portals as well as electronic payment services, shopping search engines, and cloud computing services.

In this assignment, my team (Kow Yee Hui, Lim Ming Jun, Lee Jun Xian, Kong Mun Jun) will be studying the E-commerce business of Alibaba Group, which is the Taobao, Tmall, Freshippo, AliExpress, Lazada, and Alibaba.com

Taobao (淘宝) Marketplace was launched in 2003, it provides consumers from both large cities and less developed areas with an engaging, personalized shopping experience. In this platform, consumers are able to learn about products and new trends through highly relevant and engaging content and real-time updates from merchants. This is because the platform is also optimized by big-data analytic and technology. Besides that, Consumers can also interact with their favorite merchants and key opinion leaders. Merchants in Taobao are defined as individuals and small businesses. According to their analysis, Taobao Marketplace was China's largest mobile commerce destination with a large and growing social community.

Tmall (天猫) was the world's largest third-party online and mobile commerce platform for brands and retailers, it was launched in 2008. Due to the high demands for high-quality products and a premium shopping experience. It serves as a platform for consumers worldwide to buy both homegrown and international branded products as well as the products that are not available in traditional retail outlets. With a lot of benefits in the platform, a large number of international and Chinese brands and retailers have established storefronts on Tmall.

In 2016, Freshippo (盒马) was launched as the proprietary grocery retail chain. It exemplifies the creation of new shopping experiences through the convergence of online and offline activities by using retail stores to warehouse and fulfill online orders. Its proprietary fulfillment system enables 30-minute delivery to customers living within a three-kilometer radius of a Freshippo Store. It is also available in a mobile app that allows consumers to search for products and place orders while browsing in-store.

In 2010, Alibaba Group also launched AliExpress as the global retail marketplace by targeting consumers worldwide and enabling them to buy directly from manufacturers and distributors in China and around the world. As it was well-known worldwide, the AliExpress platform is available in 17 other languages for better consumer experience.

When the E-commerce platform is leading and fast-growing, Lazada was launched during 2012 in Southeast Asia with a presence in six countries which are Indonesia, Malaysia, Philippines, Singapore, Thailand, and Vietnam. Lazada provides consumers with access to a broad range of offerings and serves over 70 million unique consumers.

Last but not least, the first business of Alibaba Group, the largest integrated international online wholesale marketplace, Alibaba.com. It connects Chinese and overseas suppliers to overseas wholesalers buyers, who are typically trading agents, wholesalers, retailers, manufacturers, and SMEs engaged in the import and export business and providing sourcing, online transaction, digital marketing, digital supply chain fulfillment, and financial services to them.

Objective

The objective of this time series analysis is to forecast future Alibaba's revenue for financial planning purposes. With this forecast in hand, Alibaba may use this information to make informed business decisions like when to enhance the advertisement online or when to do the promotions in which seasons and so on.

By using time series analysis, data is described to obtain the descriptive measures, such as trends and seasonal fluctuations of the revenue quarterly. Through the descriptive measures, we can know whether there is a seasonal pattern or not. Also, by plotting the time series graph, we are able to look for the outlier or data that does not appear to be consistent with the rest and it also shows the presence of turning points where the upward trend suddenly changed to a downward trend. The observed values of the time series will be used to forecast the future revenue in this case, the revenue of the next 15 quarters.

Methodology

This section discusses the methods used in this study. Gahirwal, M. (2013) indicate that **Quantitative forecasting methods** are used in this study to forecast the future sales data points according to the historical data pattern. Quantitative methods come from two main types, there are Time-series Methods and Explanatory Methods.

In addition, **Qualitative forecasting methods** are also used to refer to the views of exports. Financial report is used in this study to extract the sales data from core-commerce to produce a dataset.

Granger, C (2014) found that **Time-series** is a model that can predict future activities and events based on previously known data. For example, economic forecasting and stock market forecasting. Besides, the contents of the financial report are also used for reference in the data analysis section.

Furthermore, Time-series consist of three components, there is **Trend** (Change in the time series), **Seasonality** (patterns of change over a short period of time), and **Irregular Component** (Random Noise). These three components will be identified in the study also.

Components of Time Series:

$$Y_t = T_t + C_t + S_t + R_t$$

The sales of Alibaba from January 2015 to June 2020 are used in this study and it is analyzed by R language. This study is focused on the forecasting sales of Alibaba in the Year 2021 and the Year 2022. The dataset with quarterly information is used in this study.

In addition, the ARIMA model is used to help with intelligent business decisions.

ARIMA (p, d, q)(P, D, Q)[S] model is classified as:

p = the autoregressive parts of the dataset. (non-seasonal)

d = the integrated parts of the dataset. (non-seasonal)

q = the moving average parts of the dataset. (non-seasonal)

P = the autoregressive parts of the dataset. (seasonal)

D = the integrated parts of the dataset. (seasonal)

Q = the moving average parts of the dataset. (seasonal)

S = the span of seasonality

Besides that, seasonal-differencing and non-seasonal differencing are also tested in this study. The purpose of doing seasonal differencing is to observe the series of Alibaba's sales changes from one season to the next. The purpose of doing non-seasonal differencing is to observe the original series of Alibaba's sales without seasonal effects.

Formula of Seasonal Differencing:

With $S = 4$, which may occur with quarterly data, a seasonal difference is $\left(1 - B^4\right) x_t = x_t - x_{t-4}$.

In this time series, moving average (MA) model and Autocorrelation Average (AR) will be tested to get the values of Akaike information criteria (AIC). AIC is used to quantify the goodness of fit of the model.

Lastly, the best model will be selected according to the results of the AIC to do the forecasting sales of Alibaba in the Year 2021 and Year 2022.

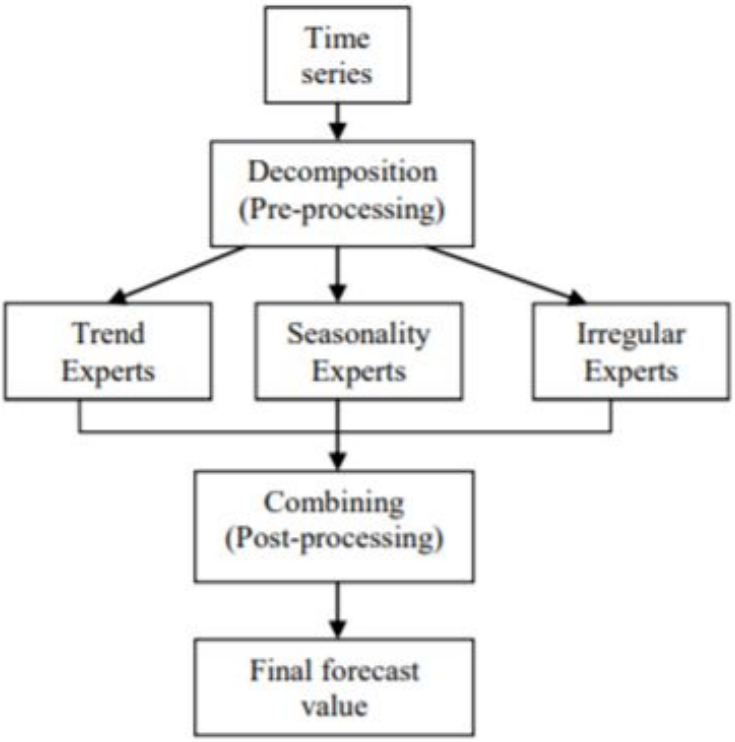

Figure 1: Time series forecasting model (Gahirwal, M. 2013)

Data Sources

The dataset used is collected from the Alibaba official website. In this dataset, it contains the **quarterly sales** data from different **core commerce** such as China commerce retail business, China commerce wholesale business, International commerce retail business, International commerce wholesale business, Cainiao logistics services, and Local consumer services.

Core commerce can be defined as the primary activity that a company was founded on or focuses on in its business operations. In Alibaba's case, its core commerce is a web-based e-commerce shopping cart software that enables you to launch and maintain an online store.

So, the total core commerce of each quarter is the quarterly sales that we are collecting to form our dataset and there are **4 quarters in a year**. Therefore, we recorded the quarterly sales from 2015 to 2020 in our dataset, in which the **features (independent variable)** of our dataset is **quarter (time)** and the **dependent variable** is **revenue (RMB)** of each quarter in each year. In total, we have recorded 22 samples in our dataset, **every 3 months are considered as 1 quarter**. Thus, we have recorded 20 samples from the year 2015 to 2019 and 2 samples in the year 2020 (March & June). Each sample contains a revenue amount of 1 quarter.

	A	B
1	**Revenue (in RMB millions)**	**Year (Quarterly)**
2	15,532	31-Mar-15
3	18,538	30-Jun-15
4	20,231	30-Sep-15
5	32,111	31-Dec-15
6	21,455	31-Mar-16
7	27,241	30-Jun-16
8	28,493	30-Sep-16
9	46,576	31-Dec-16
10	31,570	31-Mar-17
11	43,027	30-Jun-17
12	46,462	30-Sep-17
13	73,244	31-Dec-17
14	51,287	31-Mar-18
15	69,188	30-Jun-18
16	72,475	30-Sep-18
17	102,843	31-Dec-18
18	78,894	31-Mar-19
19	99,544	30-Jun-19
20	101,220	30-Sep-19
21	141,475	31-Dec-19
22	93,865	31-Mar-20
23	133,318	30-Jun-20

Figure 2: Alibaba Dataset

Data Analysis

Observation, Trend, and Seasonality

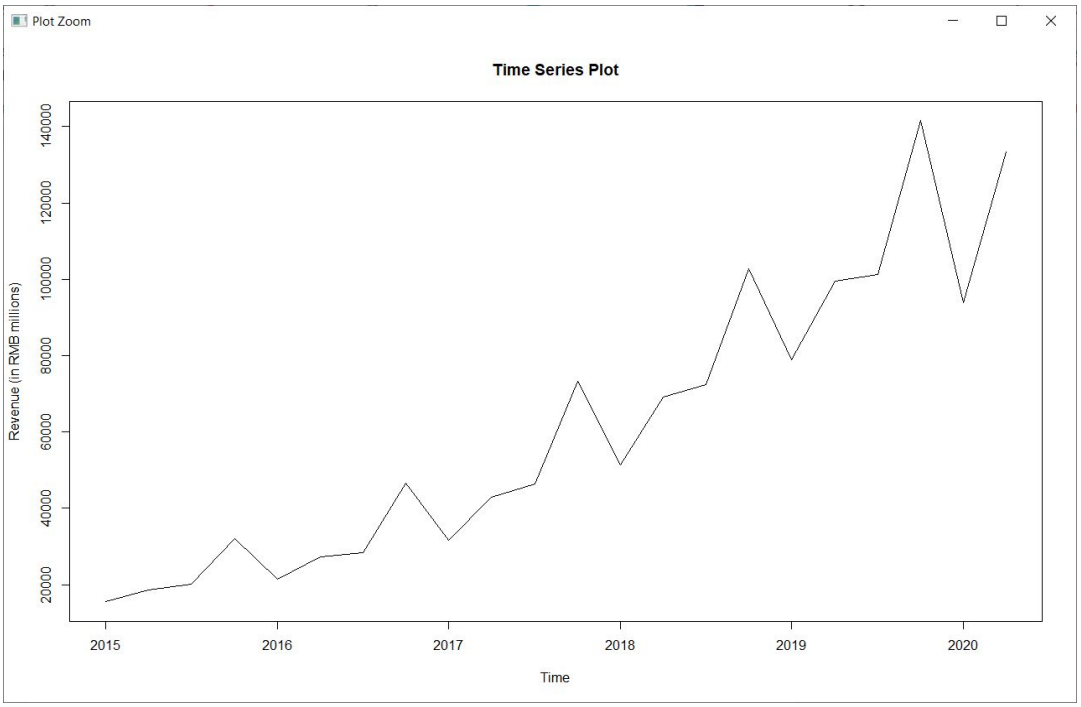

Figure 3: Observation graph

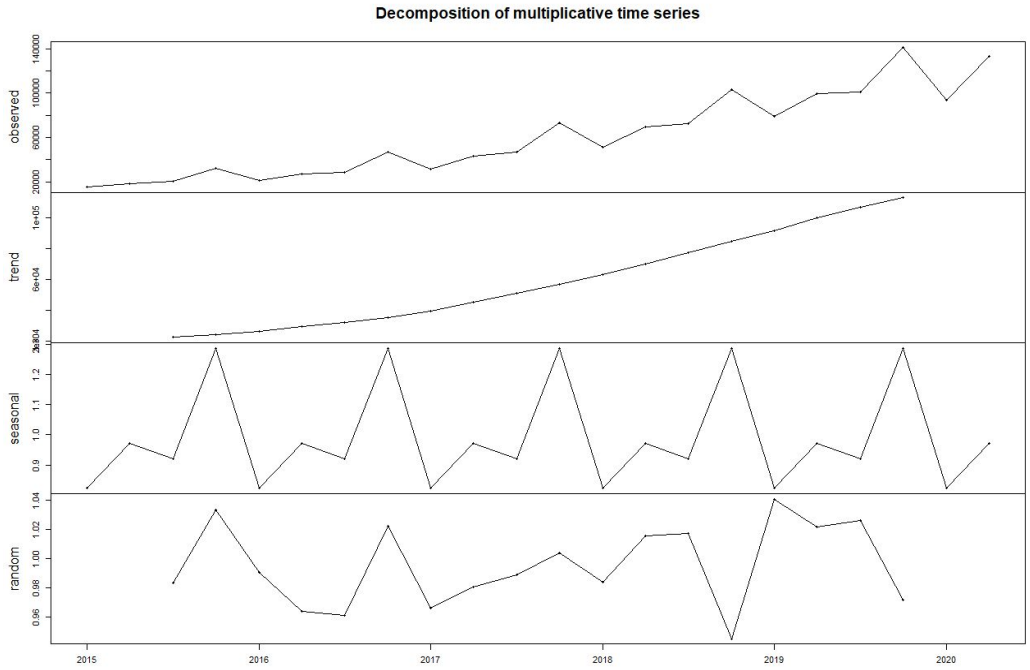

Figure 4: Details of decomposition

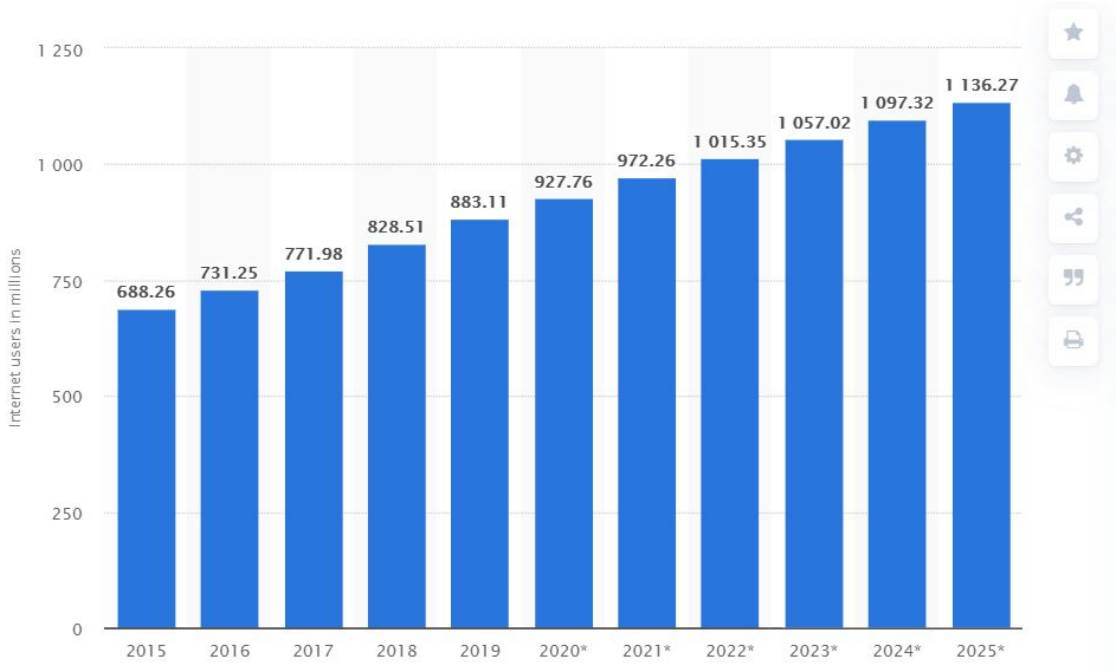

Figure 5: Number of Internet users in China

Based on the time series plot below (Figure 3), we can see that Alibaba's quarterly revenue seems to have seasonal variation as there is a peak every 4th quarter. Besides, it seems that this time series could probably be described using a **multiplicative model**, and conclude that this is a **non-stationary time series** as the **seasonal fluctuations** and **random fluctuations** seem to **increase with the level of the time series** and there is no constant in both mean and variance over time.

After decomposing the time series data (Figure 4), we can easily see the trend and seasonality of this time series data. Based on the decomposition result below, we can clearly see that this time series data is having a **multiplicative trend** as the graph is growing in an inclined line from left to right. For seasonality, we also can clearly see that there is a **multiplicative seasonal effect** in our time series data as the **seasonal fluctuations are increasing over time within each year**.

From the time series plot above (Figure 3), we actually can obviously see that Alibaba's **revenue trend is increasing annually**. There must be some reason behind this positive scenario. But one of the biggest reasons is **the growth of internet users in China.** According to the statistics

carried out by Statista (Figure 5), internet users from 2015 to 2019 have increased from 688.26 million to 883.11 million and are expected to reach 927.76 million by 2020 and this large number of internet users are the backbone of the sales growth of Alibaba. As proof, Taobao.com was launched as an online market when the number of Internet users in China reached 80 million in 2003. Shortly after that, Alipay and Aliwangwang were launched to complete the purchasing process in Taobao. Then, Alimama was launched as an advertisement transaction platform as the number of Internet users in China increased to 210 million in 2007. Besides, the main reason why Jack Ma founded Alibaba is that he discovered the power and potential of the Internet early on and his Alibaba is the only biggest e-commerce platform in Asia nowadays. As a result, Alibaba has dominated the field of e-commerce in Asia and we can see that as the number of Internet users increases, more projects and plans will be implemented by Alibaba to meet the demands of the market. In short, the fast-growing Internet users in China have contributed a lot to the revenue of Alibaba and caused the revenue to rise annually.

Besides, we can also see that there is a peak in revenue every fourth quarter. This is mainly due to the **11.11 Global Shopping Festival**. The 11.11 shopping festival, which falls and also culminates on November 11. This festival was started in 2009 with participation from just 27 merchants as an event for merchants and consumers to raise awareness about the value of online shopping. During this festival, there are many promotions and discounts that will be offered to consumers and any other customers who buy items through an e-commerce website. Thus, many customers will do planning for this festival. They will plan well what to buy by listing down all their favorite items before this festival and wait for the festival to arrive. These customers will then instantly buy all their favorite items at once when the festival has arrived to avoid facing the problem of running out of stock for their items. This is because huge discounts are given to the customers provided that the customers buy products in bulk. So, some customers worry about the problem of no stock available. But this scenario somehow will cause **a positive social influence** which affects or reminds most of the customers better to plan what to buy early and place the orders immediately during the festival so they can enjoy shopping with the least amount of money to buy all their favorite items. A win-win situation for customers and retailers. In 2019,

this quarter's annual 11.11 Global Shopping Festival has successfully broken last year's record, generating RMB268.4 billion in GMV compared to RMB213.5 billion in 2018. A total increase of RMB54.9 billion and 25.7% over the previous year. During this year's festival, over 200,000 brands, including 22,000 international brands, participated in this year's event, among which 15 brands each generated GMV of over RMB1 billion and 299 brands each generated GMV of over RMB100 million. As a result, the seasonal effect in quarter 4 of every year is at peak and can be seen obviously as the 11.11 Global Shopping Festival helps Alibaba to achieve revenue growth in a skyrocket that could not be achieved in other quarters.

Lastly, based on figure 3, we can see an increase in sales in the first quarter of 2020 compared to the first quarter of the previous year, but if we look at the changes in details, the sales in the first quarter of 2020 are lower than sales in the third quarter of 2019, which has never happened before. This is mainly due to the **coronavirus outbreak in China** in early 2020. On 23 January 2020, China imposed a lockdown in Wuhan and other cities in Hubei in order to quarantine the center of an outbreak of COVID-19. This lockdown has brought much of economic activity in China to a halt and affecting many types of businesses in the country. Most of the businesses are being seriously hit especially the food and beverage, retail, real estate, and travel business types are the four biggest losers in China. Consequently, there are about a total of 80 million Chinese out of work in March, and the official urban unemployment rate rose to 5.9% in March which has barely moved beyond 4% to just over 5% for years. As a result, the purchasing power of most people in China has fallen, causing Alibaba's first-quarter revenue to be lower than the third quarter of the previous year. But the overall revenue in the first quarter is still performing well as it is higher than the previous first quarter. Then, the revenue performance bounced back in the second quarter as the **COVID-19 crisis is under control and recovering** in China. The lockdown in Hubei and Wuhan was also ended on 25 March 2020 and 8 April 2020 respectively. As a result, most of the economic activities will undergo rejuvenation, people who lost their job at least can look for jobs outside instead of being quarantined inside the house and people who owned a small or medium business at least can run their business back to generate income instead of remain closed. Due to the effectiveness in tackling the COVID-19 pandemic, China's economy is able to recover much faster than other countries although the economy is not as

strong as before COVID-19 but at least China succeeded in minimizing economic deterioration. According to Alibaba's Chief Financial Officer, Maggie Wu said as China's economy recovers from the COVID-19 pandemic with the spread of the coronavirus has been basically controlled, Alibaba's domestic core commerce business has fully recovered to pre-COVID-19 levels across the board. The only area of disappointment was that the number of annual active customers in the Chinese retail market didn't grow as much as predicted by analysts. In short, China's high efficiency in responding to the Covid-19 pandemic helped Alibaba's revenue rebound better and stronger in the second quarter.

As mentioned earlier, our time series data is a **non-stationary time series** as the seasonal fluctuations and random fluctuations are increasing with the level of the time series and it is having a **multiplicative trend** & **multiplicative seasonal effect** after identifying through the decomposition process. Therefore, it is necessary for us to **conduct non-seasonal differencing** to **remove the non-stationarity** in our time series data and **transform** it into stationary time series data.

Non-Seasonal Differencing

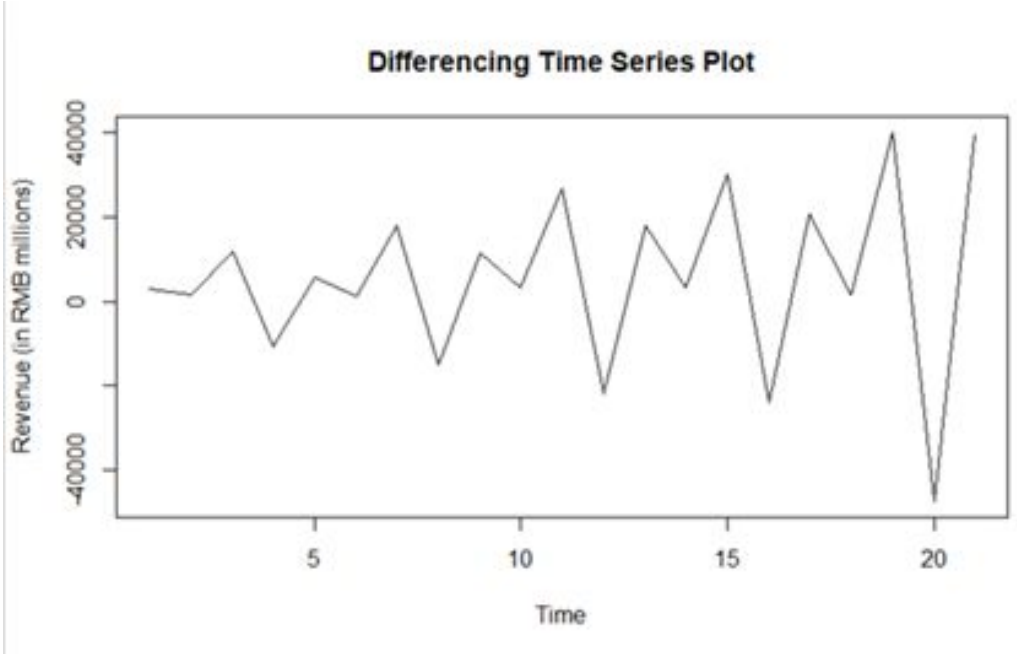

Figure 6: Time Series Plot of Non-Seasonal Differencing

Figure 7: Consumer Price Index (CPI) monthly in China. Source: Moody's Analytics

Furthermore, it showed that the sales of Alibaba were rising every year. In addition, it also clearly showed that the sales of Alibaba will **fall in every Quarter 3**. This was because there

will be fewer promotions and special festivals in Quarter 3 than in other quarters. It could be proved in Figure 7, the graph reported that the CPI was low from August to January.

At the same time, it also showed that **Alibaba sales will rise significantly in Quarter 4**. This was because there were more promotions at the end of the year, such as the **11.11 Global Shopping Festival and 12.12 Global Shopping Festival**. In addition, the rise in sales also had some external factors such as **Christmas Day**. It was because Christmas Day is a festival that both the West and East participate in and a lot of people started to buy New Year's goods on Christmas Day. Thus, the sales of Alibaba will be increased in every Quarter 4. Sheng Guoqing, a senior data analyst with the city statistics department of the National Bureau of Statistics (NBS), said the high demand and consumption boom during the Spring Festival affected food and service prices. As a result, the CPI rose sharply in February.

ACF and PACF of Non-Seasonal Differencing

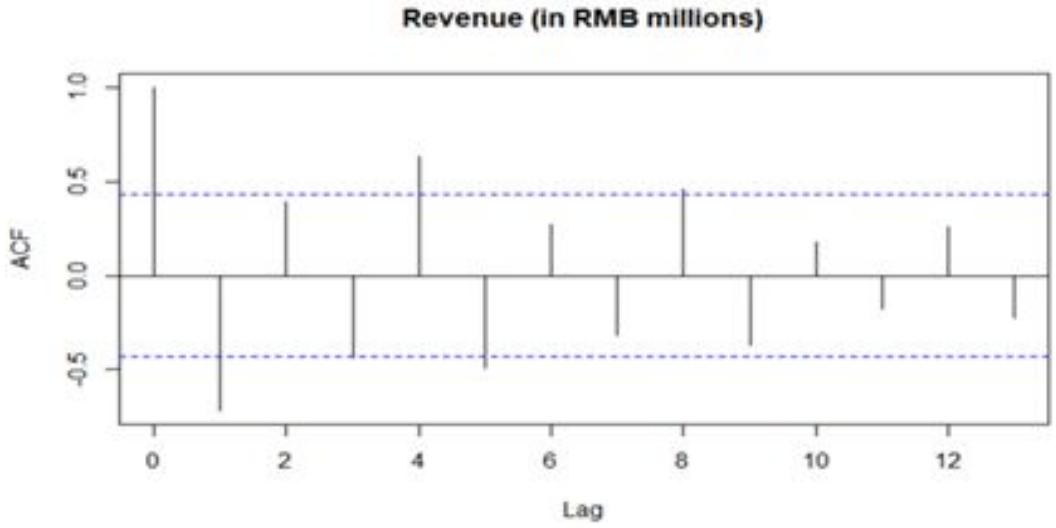

Figure 8: ACF of Non-Seasonal Differencing

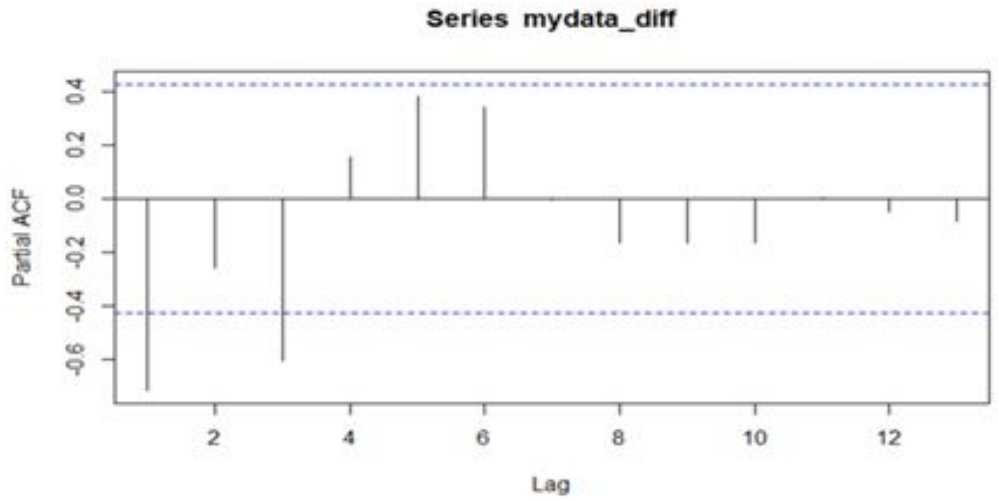

Figure 9: PACF of Non-Seasonal Differencing

Based on Figure 9, it showed that lag 1 (negative) had a large spike and lag 3 (negative) had a significant spike in PACF. A large spike was observed at lag 1 which meant it showed a strong correlation with each value of the time series. According to Bowerman et al. (2005), the spike at lag 1 could be ignored to simplify the initial tentatively identified model.

Based on Figure 8, the **ACF graph has 2 spikes at lag 4 (positive)** and **lag 8 (positive)** which exceeds the confidence interval in the ACF graph. As we are **having quarterly time-series data,** Spikes at lag 4 and lag 8 means that our time-series data still having seasonal effects. So, there is a need for us to **implement seasonal differencing** to eliminate seasonal effects contained in our time-series data.

Seasonal Differencing

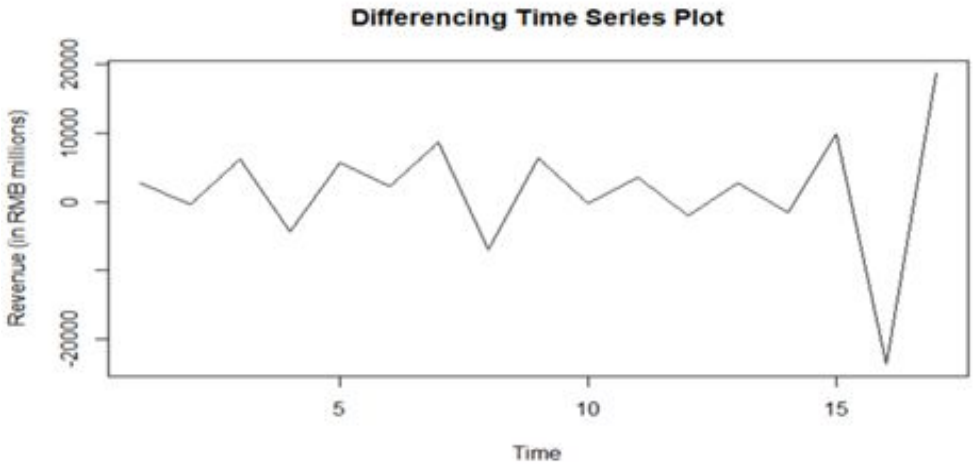

Figure 10: Time Series Plot of Seasonal Differencing

Based on Figure 10, it shows the mean varies (increases) with time and the result is in an upward trend. It indicated that the sales of Alibaba are gradually rising. Thus, the time series plot was non-stationary. It indicated the seasonal effects, trends and other structures depend on the time index. Thus, the **seasonal differencing was needed by removing trends and stationary effects**. After removing the seasonal effects from the time series plots, they showed that the sales trend of Alibaba which was not affected by seasonality. It used to make the non-stationary time series data become stationary. The reason for removing trends by time series plot was that trends will make a time series non-stationary by increasing the level. According to Abish and Mohanakumar (2013) found that the time series is non-stationary when it did not have a constant mean, variance, and autocorrelation.

The entire sales of Alibaba were relatively rising, except in the Quarter 1 2020. In **February 2020, there was a large spread of the COVID-19**. COVID-19 had a huge impact on Alibaba's core commerce sales. This was caused by the number of deaths around the world, especially China. Based on Figure 11, it also showed that **many people unemployed** in February 2020. It was because many industrial, tourism, retail, and transportation companies went bankrupt during the COVID-19 period. As a result of many city lockdowns, it caused inflation and the price level

increased greatly. All the above reasons, it greatly affects the sales of Alibaba. But after the Quarter 1, the sales of Alibaba began to increase again.

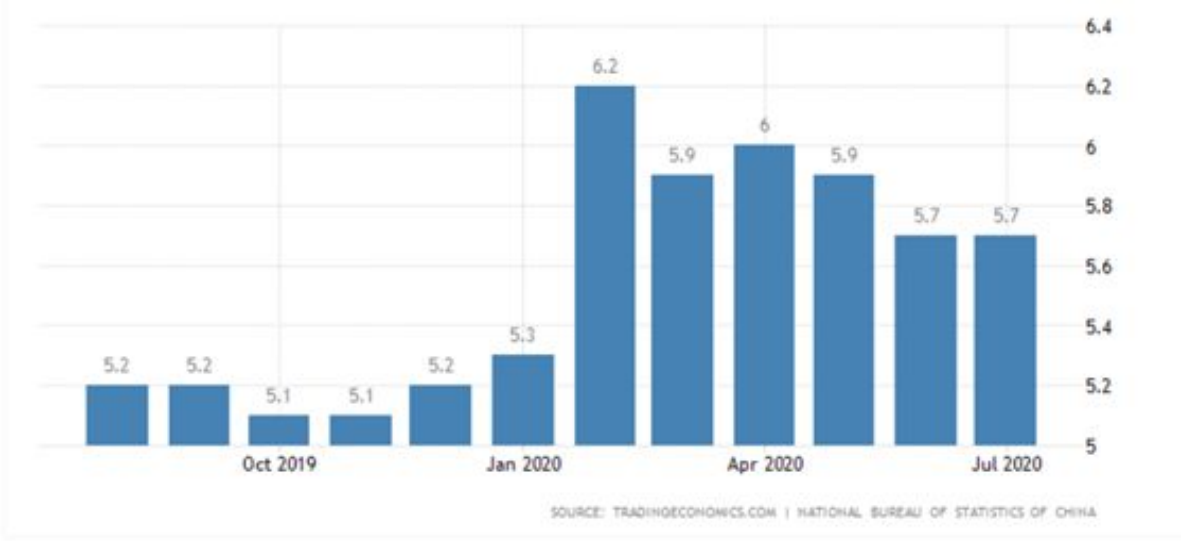

Figure 11: China Urban Survey Unemployed rate in China. Source: Trading Economics

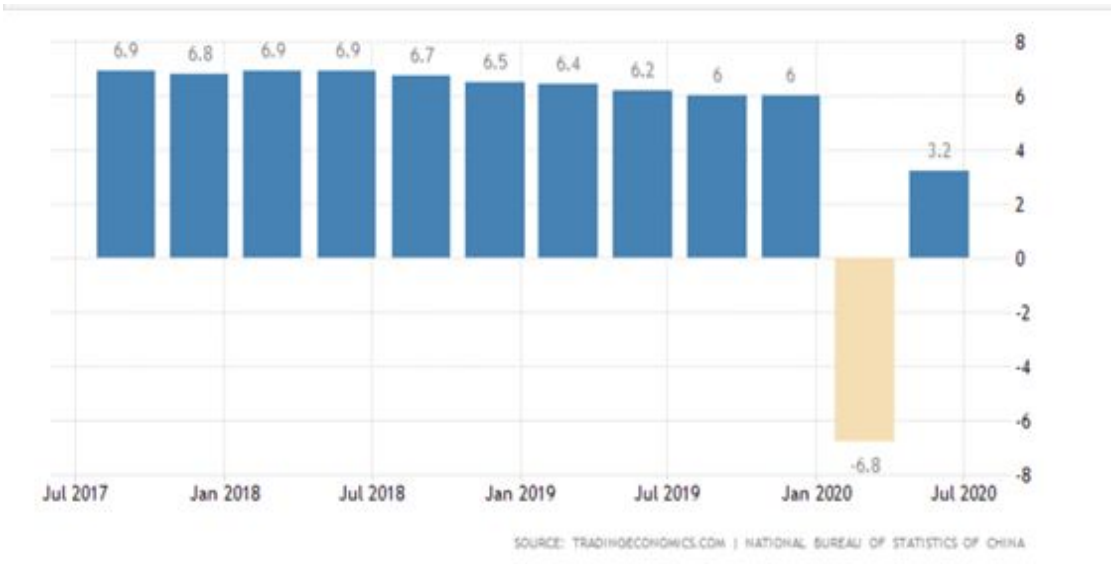

Figure 12: China GDP (Gross Domestic Products) Annual Growth Rate. Source: Trading Economics

Furthermore, COVID-19 had triggered a series of national and urban lockdowns, social distance requirements, and travel restrictions. These circumstances had a **significant impact on the global economy**. Besides that, declining productivity will also affect the economy of China. Based on Figure 13, it had indicated that China's production index fell sharply in February.

Based on Figure 12, it showed that GDP (Gross Domestic Products) of China was negative in Quarter 1 which is -6.8%. This proved that the economy of China was very declining in Quarter 1. When the economy is weak, people's living standards will be greatly reduced. Thus, it will affect the online shopping rate.

Figure 13: China Manufacturing Production Index.Source: YCHARTS

Besides, Alibaba's financial report also mentioned that in Quarter 1, the **logistic services Cainiao was broken** due to the lockdown of the city. Many goods could not be delivered. Thus, many orders were canceled because of the backlog of goods. These kinds of problems had caused a lot of goods transactions in Cainiao cannot work properly. During the Quarter 1, sales of Alibaba experienced negative year-over-year growth due to the pandemic.

Based on Figure 10, it showed that after COVID-19 gradual control, the China economy began to recover. The **economic capacity of the people was also beginning to improve**. According to Figure 10, it showed that sales of Alibaba began to rise sharply. This was because when the epidemic was effectively controlled, the lockdown measure began to slacken, and many activities gradually returned to normal. In addition, after the outbreak, many consumers began to **change their lifestyle** because many people were working at home in that season. Besides, people also

began to prefer to buy daily necessities such as fresh through online platforms like Taobao and Tmall.

Figure 14: Consumer Prices depend on the prices of foodstuff and non-foodstuff in 2019 and 2020 (China). Source: National Bureau of Statistics Of China

In addition, Alibaba's sales rose in Quarter 2 because the prices of the goods began to rise during the COVID-19 period. Based on Figure 14, the graph indicated the **prices had started to rise, people's consumption will become lower** and vice versa. From the chart, we could observe that prices had risen by 5.4 percent and it was falling in February. It was due to the outbreak of Covid-19, the lack of supplies caused many businesses to price their goods high. This had also caused many consumers to pay attention to daily necessities storage. Thus, according to Figure 14, it showed that the prices of products started to rise down even though the CPI could be considered a high rate.

ACF and PACF of Seasonal Differencing

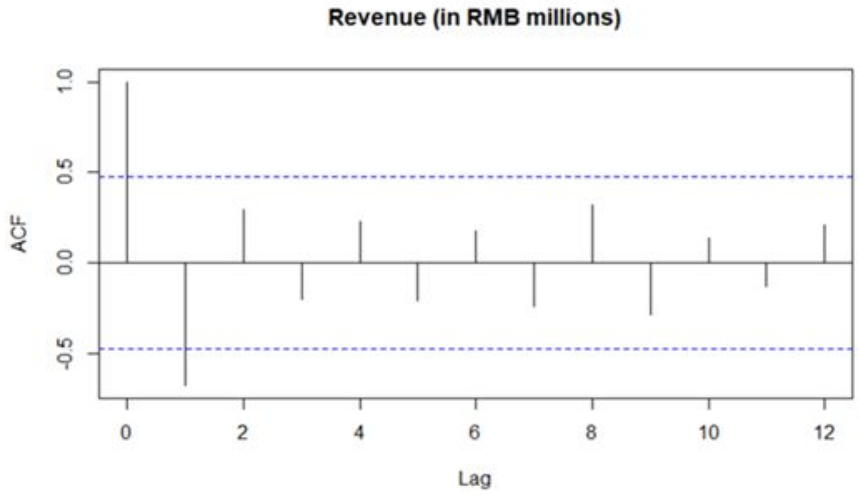

Figure 15: ACF of Seasonal Differencing

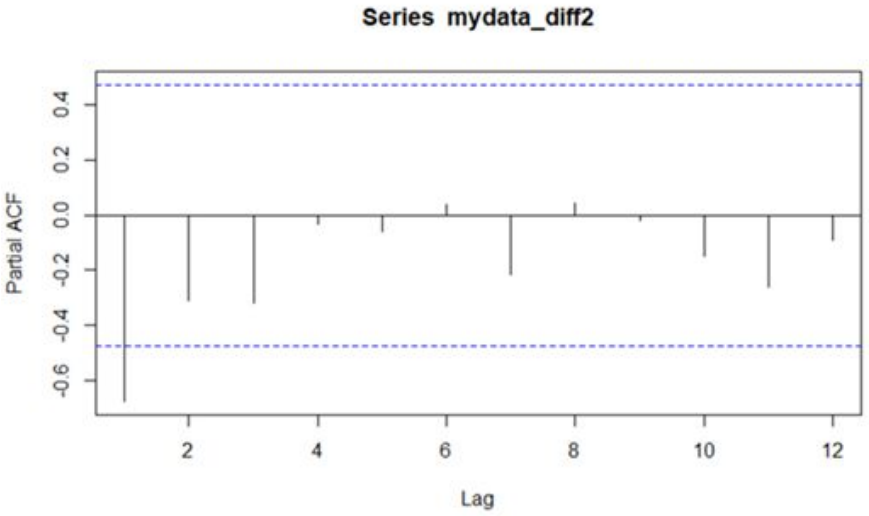

Figure 16: PACF of Seasonal Differencing

Based on Figure 15 and Figure 16, the ACF and PACF showed a single spike at lag 1. Thus, the AR (1) and MA (1) model was indicated to be used for forecasting. So, we can conclude that the **best fit model** for forecasting is either AR(1) or MA(1). We also could conclude that the ARIMA model cannot make predictions successfully because of the **seasonal effects still contained in the time-series** data.

AR & MA Model

Among these models, AIC is most often used for model selection. By calculating and comparing the AIC scores of several possible models, analysts are able to choose the one that is the best fit for the data.

```
> # AR(1) Compare AIC
> fit_ar = arima(mydata, order = c(1,1,0), seasonal=list(order=c(0,1,0), period = 4))
> fit_ar

Call:
arima(x = mydata, order = c(1, 1, 0), seasonal = list(order = c(0, 1, 0), period = 4))

Coefficients:
          ar1
      -0.8099
s.e.   0.1544

sigma^2 estimated as 32963165:  log likelihood = -171.8,  aic = 347.6
```

Figure 17: Comparing AIC with AR Model

```
> # MA(1) Compare AIC
> fit_ma = Arima(mydata, order = c(0,1,1), seasonal=list(order=c(0,1,0), period = 4))
> fit_ma
Series: mydata
ARIMA(0,1,1)(0,1,0)[4]

Coefficients:
          ma1
      -0.5281
s.e.   0.1440

sigma^2 estimated as 51879691:  log likelihood=-174.77
AIC=353.54   AICc=354.39   BIC=355.2
```

Figure 18: Comparing AIC with MA Model

From the above figure, we can see that the results of the AIC value in the AR model are 347.6 and 354.39 in MA model which shows that AR model has the AIC value lesser than MA model. This is able to conclude that the **AR(1) model is the best fit model** for forecasting.

$$AR(1): \gamma_t = \emptyset_1 \gamma_{t-1} + \varepsilon_t$$
$$(1 - \emptyset_1 \beta)(1 - \beta)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$(1 - (-0.8099)\beta)(1 - \beta)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$(1 + 0.8099\beta)(1 - \beta)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$(1 + 0.8099\beta)(1 - \beta)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$(1 - \beta + 0.8099\beta - 0.8099\beta^2)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$(1 - 0.1901\beta - 0.8099\beta^2)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$(1 - 0.1901\beta - 0.8099\beta^2)\gamma_t = (1 - \beta^4)\varepsilon_t$$
$$\gamma_t - 0.1901\beta\gamma_t - 0.8099\beta^2\gamma_t = \varepsilon_t - \beta^4\varepsilon_t$$
$$\gamma_t - 0.1901\gamma_{t-1} - 0.8099\gamma_{t-2} = \varepsilon_t - \varepsilon_{t-4}$$
$$\gamma_t = \varepsilon t - \varepsilon_{t-4} + 0.1901\gamma_{t-1} + 0.8099\gamma_{t-2}$$

Figure 18: Comparing AIC with MA Model

Forecasting

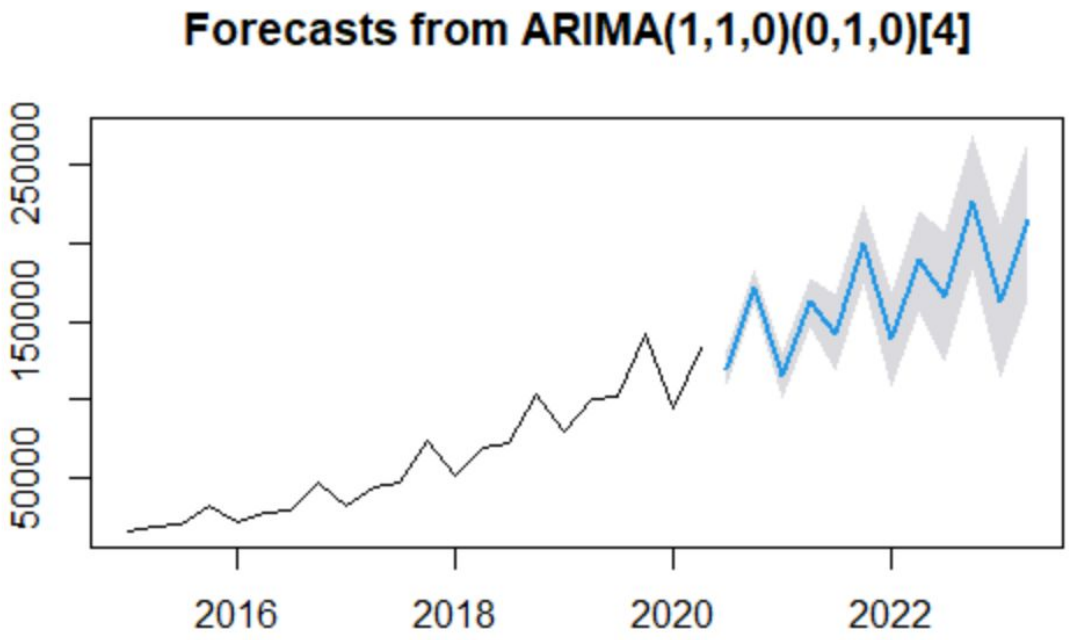

Figure 13: Alibaba's Quarterly Revenue Forecasting using AR(1) Model

Figure 13 shows the result of the time series analysis which is the forecasting of Alibaba's quarterly revenue where it predicts the revenue 3 years or 12 quarters in the future. In the figure, the blue line indicates the forecast value where the grey areas are the confidence intervals. Based on the forecast value in the figure, it seems that the revenue will have an increasing seasonal trend. As we forecast further into the future, the confidence intervals generated by the model are growing larger, which is the nature for the value to be less accurate as the forecasting goes on.

By looking at the journal and the current actions from Alibaba, the forecast might turn out to be the real graph for their future. One of the factors that bring them these advantages is society. In the current year, the 80's society started to use e-commerce to settle all their expenses from daily supplies to furniture. The most obvious we can realize is from the 90's and 00's generation, as this group of teenagers are very familiar with technology and involve themselves in e-commerce a lot. Streetwear is the fashion clothes they bought a lot in e-commerce.

As time goes on, 80's will turn to be elderly, 90's and 00's will be categorized as the middle-age adults, also the 10's and 20's will be teenagers. At that time, it will turn out that all of society will use e-commerce which indirectly increases the revenue of Alibaba as the number of consumers increases.

Another impact that has forecast results is the evolution of business. In 2009, Jack Ma once said that E-commerce will eliminate retail shopping and turn out to be real after 10 years when none believe what he said. At this moment, a lot of retail shops changed their sales methods and brought themselves into e-commerce business. Although, new retailing became a trend suddenly but still will not affect Alibaba as the technology used in Alibaba is the strongest support for their business.

With the recent research of ET Brain Technology and AliOS, Alibaba will out pass all the e-commerce business in the future as ET Brain brings a lot of impact towards the Alibaba Business. With Artificial Intelligence, ET Brain Technology apply into the business, the time of purchasing the items for consumers, the time for replying to consumers about inquiries, the time to make the payment, and lesser cost of providing delivery fees as the delivery man will be replaced by AI with an auto-driving car. With all the convenience and time-saving in every payment voucher and order made, it will indirectly increase the revenue of their business too.

Conclusion

In conclusion, we have performed the decomposition process to better identify and observe time-series data. Through decomposition, we have identified that time-series data is having an increasing trend and it is a multiplicative series as there are seasonal effects every year. Besides, we also have successfully identified that our time-series data is non-stationary as the seasonal fluctuations changes over time if we observed through the Observed Graph.

In order to have better prediction or forecasting results, we have to convert the non-stationary time-series data to stationary time-series data. Thus, we first conduct non-seasonal differencing to stationarize our time-series data. But there are still seasonal effects available in the dataset after conducting non-seasonal effects after looking through the ACF plot. Therefore, we then conduct seasonal differencing to remove the seasonal effects.

From seasonal differencing, the ACF and PACF plots show that there are no seasonal effects left in our time series data. Therefore, we have successfully stationarized our time-series data. Besides, we also have identified the possible models that we should use for prediction and forecasting during this process which are AR(1) and MA(1) models through looking through ACF and PACF plots. Then, we will compare the AIC from these 2 models and choose the one who provides lower AIC values. This means that we have chosen the right model for forecasting.

In the end, our forecasting shows that there will be an increasing seasonal trend in the future which probably can be explained by the popularity of e-commerce towards the younger generation. It is safe to say that the further into the future, the more popular e-commerce will be, thus making the revenue of Alibaba increase significantly.

Reference

zone.ni.com. (n.d.). Introduction to Time Series Analysis (Advanced Signal Processing Toolkit) - LabVIEW 2010 Advanced Signal Processing Toolkit Help - National Instruments. [online] Available at:

https://zone.ni.com/reference/en-XX/help/371419D-01/lvasptconcepts/tsa_intro/#:~:text=The%20objectives%20of%20time%20series%20analysis%20are%20as%20follows%3A&text=Prediction%E2%80%94You%20can%20use%20observed [Accessed 2 Sep. 2020].

Ullah, M.I. (2014). Objectives of Time Series Analysis. [online] Basic Statistics and Data Analysis. Available at: https://itfeature.com/time-series-analysis-and-forecasting/objectives-of-time-series-analysis [Accessed 2 Sep. 2020].

Analytics Vidhya. (2015). Time Series Analysis | Time Series Modeling In R. [online] Available at: https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/ [Accessed 2 Sep. 2020].

Blystone, D. (n.d.). *Understanding the Alibaba Business Model*. [online] Investopedia. Available at: https://www.investopedia.com/articles/investing/062315/understanding-alibabas-business-model.asp#:~:text=of%20June%202019.- [Accessed 2 Sep. 2020].

Alibabagroup.com. (2019). *Alibaba Group*. [online] Available at: https://www.alibabagroup.com/en/about/businesses.

Wikipedia Contributors (2019). *Alibaba Group*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Alibaba_Group.

Medium. 2020. Understanding SARIMA. [online] Available at: <https://towardsdatascience.com/understanding-sarima-955fe217bc77> [Accessed 12 September 2020].

Chew, D., 2020. *How To Create Your Ecommerce Forecast In 3 Steps - Fullmoon Digital*. [online] Fullmoon Digital. Available at: <https://fullmoondigital.com/e-commerce-forecasting-in-3-steps/#:~:text=Forecasting%20sales%20or%2 revenue%20projection,in%20mind%2C%20it%27s%20a%20forecast.&text=For%20most%20e Commerce%20businesses%20this,in%20your%20digital%20marketing%20mix.> [Accessed 13 September 2020].

Li, S., 2020. *An End-To-End Project On Time Series Analysis And Forecasting With Python*. [online] Medium. Available at: <https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b> [Accessed 13 September 2020].

Peixeiro, M., 2020. *The Complete Guide To Time Series Analysis And Forecasting*. [online] Medium. Available at: <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775> [Accessed 13 September 2020].

Srivastava, T., 2015. *Time Series Analysis | Time Series Modeling In R*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/> [Accessed 13 September 2020].

Roos, D., 2020. *How Sales Forecasting Works*. [online] HowStuffWorks. Available at: <https://money.howstuffworks.com/sales-forecasting1.htm#:~:text=Sales%20forecasting%20hel

ps%20retailers%20decide,of%20a%20product%20to%20stock.&text=Sales%20are%20the%20lifeblood%20of,products%20and%20attract%20more%20investors.> [Accessed 13 September 2020].

Granger, C. W. J., & Newbold, P., 2014, 'Forecasting economic time series. ', pp. 1.

Gahirwal, M., 2013, 'Inter time series sales forecasting.', pp. 1.

Tradingeconomics.com. 2020. *China Urban Survey Unemployed Rate | 2002-2020 Data | 2021-2022 Forecast | Historical*. [online] Available at: <https://tradingeconomics.com/china/unemployment-rate> [Accessed 13 September 2020].

Tradingeconomics.com. 2020. *China GDP Annual Growth Rate | 1989-2020 Data | 2021-2022 Forecast | Calendar*. [online] Available at: <https://tradingeconomics.com/china/gdp-growth-annual> [Accessed 13 September 2020].

Yicaiglobal.com. 2020. *Affected By Chinese New Year Holiday, China's CPI In February Ups Nearly 3%*. [online] Available at: <https://www.yicaiglobal.com/news/affected-by-chinese-new-year-holiday-china-cpi-in-february-ups-nearly-3--> [Accessed 14 September 2020].

Ycharts.com. 2020. *China Manufacturing Production Index*. [online] Available at: <https://ycharts.com/indicators/china_manufacturing_production_index> [Accessed 14 September 2020].

Faculty.fuqua.duke.edu. 2020. *Seasonal Differencing*. [online] Available at: <https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/Class10notes.htm> [Accessed 14 September 2020].

Statista. (2017). *China: number of internet users 2017-2023 | Statista*. [online] Available at: https://www.statista.com/statistics/278417/number-of-internet-users-in-china/.

kuaibao.qq.com. (n.d.). *马云再次预言，未来10年网购将被淘汰，这个行业将取代电商-看点快报*. [online] Available at: https://kuaibao.qq.com/s/20191007A0EWF600?refer=spider [Accessed 15 Sep. 2020].

Retail News Asia. (2016). *The seven reasons for Alibaba's success ; Alibaba's development and framework*. [online] Available at: https://www.retailnews.asia/seven-reasons-alibabas-success-alibabas-development-framework/.

www.businesswire.com. (2019). *Alibaba Kicks Off 2019 11.11 Global Shopping Festival*. [online] Available at: https://www.businesswire.com/news/home/20191021005241/en/Alibaba-Kicks-2019-11.11-Glob
al-Shopping-Festival [Accessed 15 Sep. 2020].

China Briefing News. (2020). *The Social and Economic Impact of COVID-19 on China*. [online] Available at: https://www.china-briefing.com/news/social-economic-impact-covid-19-china-recovery-potential/.

Business, L.H. and N.G., CNN (n.d.). *80 million Chinese may already be out of work. 9 million more will soon be competing for jobs, too*. [online] CNN. Available at: https://edition.cnn.com/2020/05/08/economy/china-unemployment-intl-hnk/index.html.

Johnston, M. (n.d.). *Alibaba Earnings: What Happened*. [online] Investopedia. Available at: https://www.investopedia.com/alibaba-earnings-4777079 [Accessed 15 Sep. 2020].

Appendix

```
# Import data from xlsx file
mydata <- read_excel("Alibaba_dataset.xlsx")

# Display the original dataset
mydata

# Format dataset
mydata <- ts(mydata[,1], frequency = 4, start = c(2015,1))

# Display the time series data
mydata

# Plot the Observed Time Series Data
plot.ts(mydata, main = "Time Series Plot")

# Transformation using log()
# mydata <- log(mydata)
# plot.ts(mydata, main = "Time Series Plot", ylab="Log(Revenue in RMB millions)")

# Decomposition of multiplicative time series
mydatacomponents <- decompose(mydata, type = "multiplicative")
mydatacomponents
mydatacomponents$seasonal
plot(mydatacomponents, col="black", type = 'o')

# Non - seasonal Differencing
mydata_diff <- ts(diff(mydata, differences = 1, lag = 1))
plot(mydata_diff, main = "Differencing Time Series Plot")
acf(mydata_diff) # The reason why I need to do seasonal differencing, alternating (non-stationary)
pacf(mydata_diff)

# Seasonal Differencing
mydata_diff2 <- ts(diff(mydata_diff, differences = 1, lag = 4))
plot(mydata_diff2, main = "Differencing Time Series Plot")
acf(mydata_diff2)
pacf(mydata_diff2)
Box.test(mydata_diff2, lag = 20, type = 'Ljung-Box')

# AR(1) Compare AIC
fit_ar = arima(mydata, order = c(1,1,0), seasonal=list(order=c(0,1,0), period = 4))
fit_ar
fit_resid = residuals(fit_ar)
Box.test(fit_resid, lag = 20, type = 'Ljung-Box')
### Since p-value > 0.05, so my residuals is random, means no more patterns is remaining in the residuals (Residual is stationary already)
### because I have identified all the possible patterns in my time series data

# MA(1) Compare AIC
fit_ma = arima(mydata, order = c(0,1,1), seasonal=list(order=c(0,1,0), period = 4))
fit_ma
fit_resid = residuals(fit_ma)
Box.test(fit_resid, lag = 20, type = 'Ljung-Box')

# Forecast
predict(fit_ar, n.ahead = 5)
forecast_sales <- forecast(fit_ar, h = 3 * 4, level = c(95.0))
plot(forecast_sales)
```