APPLICATION OF TIME SERIES ANALYSIS ON UNEMPLOYMENT RATES IN MALAYSIA

By

LIM HUI JING

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

TUNKU ABDUL RAHMAN UNIVERSITY COLLEGE

KUALA LUMPUR

ACADEMIC YEAR

2021/2022

APPLICATION OF TIMES SERIES ANALYSIS ON UNEMPLOYMENT RATES IN MALAYSIA

By

LIM HUI JING

Supervisor: Mr. Chee Keh Niang

A project report submmited to the
Faculty of Computing and Information Technology
In partial fulfillment of the requirement for the
Bachelor of Science (Hons.)
Management Mathematics with Computing
Tunku Abdul Rahman University College

Department of Mathematics and Data Science
Faculty of Computing and Information Technology
Tunku Abdul Rahman University College
Kuala Lumpur
2021/2022

DECLARATION

The project report submitted herewith is a result of my own efforts in totality and in every aspects of the works. All information that has been obtained from other sources had been fully acknowledged. I understand that any plagiarism, cheating or collusion of any sorts constitutes a breach of Tunku Abdul Rahman University College rules and regulations and would be subjected to disciplinary actions.

Signature: 

Name: LIM HUI JING

ID No: 20WMR09183

Date: 7/10/2021

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"APPLICATION OF TIMES SERIES ANALYSIS ON UNEMPLOYMENT RATES IN MALAYSIA"** was prepared by **LIM HUI JING** and has met the requored standard for submission in partial fulfillment of the requirements for the award of Bachelor of Science (Hons.) Management Mathematics with Computing at Tunku Abdul Rahman University College.

Approved by,

Signature: ____________________

Supervisor: ____________________

Date: _________________________

Signature: ____________________

Moderator: ____________________

Date: _________________________

ACKNOWLEDGEMENTS

The outcome of the application of times series analysis on unemployment rates in Malaysia, will necessitate advice and assistance from experts in the area. Fortunately, Mr Chee Keh Niang, my final year report supervisor, and Dr Chin Wan Yoke, my lecturer, shared their extensive knowledge, resources, and advice with me. I am learning and watching since my expertise and talents in time series analysis are so extensive. Mr. Chee Keh Niang, on the other hand, is always willing to help and support me when I'm having trouble. I appreciate their knowledge and assistance. The project would not have been accomplished without their assistance.

I should also appreciate the resources given to me by the Tunku Abdul Rahman University College. All the programme structures are well organized. Moreover, students can also easily get the reference book and materials that are needed to complete the research in the college’s library and college’s website. It gives such a convenient environment to conduct this research.

Next, I want to express my gratitude to my pillars of strength, my parents, who, while they may not comprehend what I do, have provided me with unwavering support and understanding throughout my life. Last but not least, my classmates who helped me with extra knowledge during my study; we encouraged each other and helped each other get through every difficult time. For their help and expertise in the field of study.

ABSTRACT

This research investigates the Application of Time Series Analysis on Unemployment Rate in Malaysia. The data of unemployment rate (percentage) in Malaysia were collected from March 2010 to July 2021. These data were analyzed by using time series analysis in this research. After analyzing the data, the forecasting of the unemployment rate in August 2021 of Malaysia also has been done in this research.

Time series analysis has been discussed deeply in this research. First, this research discusses the definition of time series and the definition of unemployment rate. Then, a number of research papers and journals are reviewed in order to have a clear understanding on the application of the time series analysis.

The types of time series have been introduced first in the part of methodology. Basically, time series have three types which are white noise, stationary time series and non-stationary time series. Then, the models of time series also have been investigated. In this research, we only investigated 4 models of time series which are Auto-Regressive Model (AR), Moving Average Model (MA), Auto-Regressive Moving Average Model (ARMA) and Auto-Regressive Integrated Moving Average Model (ARIMA). After that, the flow to generate a time series model has been stated in this research.

Following this, SPSS software was used to do the time series analysis and generate an ARIMA model of the unemployment rate time series. In the result, we get an ARIMA (*0,1,1*) model of the time series. More details about this time series and ARIMA model have been stated in the research. By using the ARIMA model, we can do the forecasting of the unemployment rate in August 2021.

In the end, a conclusion has been made and there are also some limitations of this research that have been conducted. Based on this, some suggestions for future research need to be taken into consideration.

Keywords: time series analysis, unemployment rate, ARIMA model

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

1.1 Introduction and Definition

Unemployment occurs when people are without work and actively seeking a job. Unemployment rate is equal to unemployed divided by employed plus unemployed (Unemployment rate %= Unemployed /(Employed + unemployed)). Unemployment is an important indicator of the capital market and belongs to the category of lagging indicator. Increasing unemployment rate is a signal of economic weakness, which can lead the government to loosen monetary policy and stimulate economic growth. On the contrary, falling unemployment will lead to inflation, which will make the central bank tighten monetary policy and reduce the money supply. ([1] Investopedia.com, 2021)

The unemployment rate has been regarded as an indicator of the overall state of the economy, so it is the focus of all economic indicators. It is the most sensitive monthly economic indicator in the market. How to interpret this indicator? In general, the decline of unemployment rate represents the healthy development of the overall economy, which is conducive to currency appreciation. If the unemployment rate is analyzed with the inflation index of the same period, we can know whether the economic development is overheated, whether there will be pressure to raise interest rates, or whether it is necessary to cut interest rates to stimulate the economic development. ([1] Investopedia.com, 2021)

There are four types of unemployment. First is frictional unemployment. Frictional unemployment means unemployment occurs when people are looking for or changing jobs. The economy is always changing, and it takes time for workers to find the job most suitable for their hobbies and skills. A certain number of frictional unemployment is inevitable, that is, the unemployment caused by the time consumption between workers wanting to work and getting a job. The second type of unemployment is structural unemployment. Structural unemployment is mainly caused by changes in the economic structure, and the regional distribution of knowledge and skills of the existing labor force does not adapt to such changes and does not match the market demand. Structural unemployment is long-term in nature, and usually originates from the demand side of the labor force. Structural unemployment is caused by economic changes that cause the demand for particular types of Labour in particular markets and regions to be relatively lower than the supply. The following type of unemployment is cyclical unemployment. Cyclical unemployment is short-term unemployment caused by insufficient aggregate demand, which generally appears in the depression stage of the economic cycle. The main reason for cyclical unemployment is the decline of the overall economic level. Because it is inevitable, cyclical unemployment is also the last thing people want. Cyclical unemployment is different from structural unemployment and frictional unemployment. The unemployed population of cyclical unemployment is large and widely distributed, which is the most severe situation of economic development and usually takes a long time to recover. Last one is seasonal unemployment. Agriculture, construction and tourism are particularly vulnerable to seasonal factors. Seasonal unemployment is a form of natural unemployment that has two negative effects on society. First, the income of seasonal employees is affected by the short employment period; Second, seasonal unemployment is not conducive to the effective use of labor resources. ([2] Rba.gov.au. 2021)

This research will examine the unemployment rate of Malaysia from March 2010 to July 2021. From 2010 to 2014, the unemployment rate of Malaysia fluctuated between 2.7% to 3.6%. From 2015 to 2019, Malaysia's unemployment rate fluctuated smoothly between 3.0% to 3.5%. However, between 2020 and 2021, Malaysia's unemployment rate fluctuates widely, ranging from 3.2% to 5.3%. Malaysia's unemployment rate was 3.9% in March 2020, but rose to 5.0% in April 2020. It hit 5.3% in May 2020, the highest unemployment rate between 2010 and 2021. The unemployment rate has since fluctuated between 4.5% and 4.9%.

According to the statistics of unemployment rate by Department of Statistics Malaysia, Malaysia's unemployment rate remained at 4.8% in July, with the number of unemployed edging up to 778, 200 persons. The unemployment rate was the same in July as the previous month, but down 0.1% on a year-on-year basis, while the overall Labour force increased 0.04% month-on-month to 16.07 million and the national Labour force participation rate (LFPR) stood at 68.3%. ([3] The Star. 2021)

The time series analysis method was used to analyse the fluctuation of Malaysia's employment market from March 2010 to July 2021 and forecast the situation of the employment market in August 2021. A time series is a series of observations made at a specific time and at the same time interval. There are several different time series models that can be used to forecast the future trends of the employment market in Malaysia. For example, White Noise, AutoRegressive Integrated Moving Average Model (ARIMA) AutoRegressive Model (AR), Moving Average Model (MA), AutoRegressive Moving Average Model (ARMA) and others.

1.2 Problem Statement

The fluctuation of unemployment rate in Malaysia is not consistent, so in this project, the purpose is to study the fluctuation of unemployment rate in Malaysia from 2010 to 2021, and then help determine and forecast the unemployment rate in Malaysia in August 2021. Next, identify the factors and criteria that influence the unemployment rate and calculate the likelihood that future values fall between two specific boundaries.

1.3 Objective

Time series models, such as exponential smoothing models and autoregressive models, including categories of ARIMA, are used to correct for the limitations of the unemployment rate and predict future unemployment rates. This research analyzes the unemployment rate from March 2010 to July 2021, identifying patterns in the unemployment rate by identifying a time series model, extracting meaningful statistics, and predicting future values based on past observations. In addition, the basic structure and function of these observations can also be understood by analyzing the time series analysis of Malaysia's unemployment rate.

1.4 Project Scope

In this project, the study of each time series component and model is included. In addition, the three-stage process of Box-Jenkins is studied to analyze and observe how to find the most suitable model through model recognition, model estimation and diagnostic examination. In the August 2021 unemployment forecast, we used several criteria to calculate and compare to find the best fit model.

CHAPTER 2 Literature Review

2.1 Basic Theory

This research will use time series analysis to analyze and forecast the unemployment rate of Malaysia. The basic theory of time series analysis that is related in this research was learned from a book, Introduction to Time Series and Forecasting, Second Edition by Peter J. Brockwell and Richard A. Davis. This book is for the readers to gain working knowledge of time series and forecasting methods applied to economics, engineering and the natural and social sciences. There are a total of 10 chapters in this book which introduce the objective of Time Series Analysis, Stationary Processes and a few models of time series analysis. This book also provides many examples of different types of time series analysis. There are some appendices to teach the way to do a time series analysis and forecasting. I learned the basic theory of time series analysis and it helps me a lot to do my studies about the time series analysis on the unemployment rate of Malaysia and forecasting the future unemployment rate of Malaysia. ([4] Peter J. Brockwell, Richard A. Davis, 2002)

2.2 Review Recent Literature

There are some research papers and journals done by other researchers. One of the papers which was published in 2012 by Louis Sevitenyi Nkwatoh. The paper studied and forecast the unemployment rate in Nigeria by using the Univariate Time Series Model. Statistics show that unemployment in Nigeria is extremely high, contributing to the country's growing poverty. As a result, policymakers want indicators of future unemployment rates in order to plan ahead and establish ways to avert a protracted rise in national unemployment. To assist cope with the problem of Nigeria's growing unemployment rate, this research investigates alternative univariate time series models, such as trend regression analysis, ARIMA, and GARCH ARIMA/GARCH hybrid models. Quarterly unemployment statistics from the first quarter of 1976 through the fourth quarter of 2011 were utilised in the study. Using predictive accuracy metrics such as root mean square error (RMSE), mean absolute percentage error (Japanese), mean absolute error (MAE), and Searle's coefficient of inequality, this study examined the predictive performance of four competing models (U-statistics). The results reveal that the time series data is influenced by a positive and substantial linear trend component. The autocorrelation function (ACF), augmented Dickey-Fuller (ADF), and

Phillips-Perron tests also reveal that the time series data are non-stationary but are stabilised by difference once. The mixed ARIMA/ARCH model, for example, may be used to forecast the short-term unemployment rate in Nigeria, according to the empirical investigation. Model selection criteria show that ARIMA/ARCH models are preferable than trend regression and ARIMA models, despite the fact that all models may be predicted based on parameter importance and model fitness. The fundamental findings suggest that the ARIMA (1,1,2)/ARCH(1) model may be used to model and forecast the Nigerian unemployment rate. The advice is to continue looking for an appropriate model that can be used to forecast future unemployment rates in order to forecast long-term unemployment. ([5] Louis Sevitenyi Nkwatoh, 2012)

Next, I reviewed a paper by Abouellil, Embareka published in 2011. This paper studied the impact of labor market trends on the unemployment rate in Egypt using a time series analysis model. The goal of this paper is to use a time series analysis of the SPSS (Social Science Statistical Software Package) project to determine the impact of labor market trends on Egypt's unemployment rate from 2004 to 2010, as well as to outline the labour situation in Egypt before and after the global financial crisis. This is critical for macroeconomic and human resource development planning and policy creation, and accurate and timely data, as well as scientific research on labour market indicators, help to achieve these goals. The Egyptian Labour Force Survey is a quarterly survey of about 21,352 homes. Labor market indicators are quantitative or qualitative measures that may be used to better understand labour market circumstances and track the entire state of the economically active population, including unemployment and employment. In this paper, it only used the Seasonal Decomposition procedure to analyze the impact of the labor market on the unemployment rate and also do the forecasting of the unemployment rate.Therefore, the results of the analysis only show some surface phenomenon and then based on the trend to make a prediction on it. ([6] Abouellil, Embareka, 2011)

In the paper by Xinkai Huang, which studied the forecasting of the US unemployment rate with the Job Openings Index in 2015. The emphasis of this study is on the seasonally adjusted national unemployment rate in the United States. The goal is to use seasonally adjusted job openings to forecast unemployment rates. To begin, the benchmark model is built using a synthetic autoregressive moving average model (ARIMA). The integrated autoregressive moving average model (ARIMAX) is effectively created using job openings and initial claims for unemployment insurance as external variables. The Akchi information criteria (AIC), Schwarz-Bayes criterion (BIC), and Hannan-Quinn criterion (HQ) were used to choose the multivariate vector autoregressive model (VAR) for unemployment rate, job openings, and AMP. Both rolling prediction and recursive prediction are explored for out of sample analysis. The mean absolute forecast error (MAFE) and mean square forecast error (MSFE) were determined, and the models were compared using the Diebold-Mariano test. The job openings correlation model outperforms the benchmark model and the initial claims for unemployment insurance correlation model, according to the findings. This suggests that the job openings index might be used as a leading indicator to better anticipate unemployment rates. All of the Job-related VAR models show substantially greater prediction accuracy than the benchmark model, but all of the projected mean absolute forecast error (MAFE) and mean square forecast error (MSFE) are much less. The job-related bivariate VAR model has superior forecasting power than the well-known index initial claims for unemployment insurance related VAR model, and there are lower mean square forecast error (MSFE) in all forecast phases. The best model is a three-variable VAR model, with mean square forecast error (MSFE) being the smallest of the three. Finally, the job openings index may be utilised as a good predictor of unemployment rate. The lower the mean square forecast error (MSFE) when integrating job openings and initial claims for unemployment insurance data, the better the unemployment rate prediction accuracy. Overall, the employment index has a substantial role in improving unemployment rate estimates. ([7] Xinkai Huang, 2015)

Furthermore, a paper that modelling and forecasting unemployment rates in Nigeria using ARIMA Model by M. O. Adenomon in 2016 also had been reviewed. The goal of this study is to find the best ARIMA unemployment rate model for Nigeria from 1972 to 2014, assuming that the current unemployment rate is determined by the unemployment rate from the previous year. It also makes recommendations for lowering Nigeria's unemployment rate. The ARIMA model is used in this research to model and forecast yearly data from Nigeria from 1972 to 2014. Between 2015 and 2018, Nigeria's unemployment rate remained high. The unit root of the unemployment time series variable is stable at the first difference, and is stable at all significance levels only for intercepts and intercepts and trends, according to the augmented Dicki-Fuller (ADF) test. The model agrees with the ARIMA model according to ACF and PACF analyses. Four possible ARIMA models were considered when modeling the Nigerian unemployment rate. The results show that ARIMA(2,1,2) is a suitable model for the Nigerian unemployment rate during the period under consideration. ARIMA(2,1,2) is used to forecast the unemployment rate in Nigeria from 2015 to 2018. Nigeria's unemployment projections show an upward trend from 2015 to 2017, with a slight decline in 2018. Unemployment remained high between 2015 and 2018. ([8] M.O. Adenomon, 2017)

Next, based on the study of the paper about modeling and forecasting unemployment rate in Sweden using various econometric measures by Sune Karlsson and Farrukh Javed in 2016. The major goal of this research is to simulate and forecast Sweden's unemployment rate. SARIMA, SETAR, and VAR time series models were used to model the unemployment rate, establish the excellence of fit and validity of the hypothesis, and choose a suitable and more streamlined model to give beneficial ideas and suggestions in this study. The model's fit is demonstrated using quarterly unemployment data from the OECD from 1983 to 2010. The paper includes several graphical and numerical approaches for evaluating the model's suitability. The tested model fits well and completely, according to the goodness of fit hypothesis. It is demonstrated that some variables are integrals of order 1 using various stationarity tests. The Granger causality test reveals a causative association between unemployment rate, prior period GDP percentage change, and industrial production, but no causal relationship between inflation rate and other variables. In addition, the cointegration vector in the variable was tested for Johansen cointegration, but no cointegration was discovered. For data from 2011 to 2015, a recursive technique was utilised to do out-of-sample prediction performance evaluation. Both the seasonal autoregressive comprehensive moving average model and the self-excited threshold autoregressive model outperform the VAR model in terms of prediction performance in and out of sample, according to the findings. The 8-quarter projected values of the two models were almost identical, and all values were within the SARIMA model's 95 percent confidence zone. Short-term projections are also better than long-term forecasts, according to the findings. Because short-term projections are more accurate, research into appropriate models for projecting future unemployment levels should continue. Because short-term projections are more accurate, a model for projecting future unemployment levels should be validated on a regular basis. Further research into basic and complicated econometric models is required in order to forecast long-term unemployment. Future study might focus on modelling the seasonal character of unemployment rate in a nonlinear framework and employing a structural VAR model to estimate unemployment rate. ([9] Sune Karlsson, Farrukh Javed, 2016)

Furthermore, a research paper which discusses the prediction of the unemployment rate in Malaysia has been reviewed. The goal of this paper is to identify the major factors that influence Malaysia's unemployment rate, as well as to determine the optimal model for predicting Malaysia's unemployment rate. Finally, the best model is used to forecast Malaysia's future unemployment rate. The survey was conducted over the entire country of Malaysia. Inflation rate, population size, and gross domestic product (GDP) were chosen as research factors. To assess the contributing elements, the significance level must be tested, and the unemployment rate must be predicted using ARIMA and Holt exponential smoothing techniques. The findings of regression analysis demonstrate that the rate of inflation and the size of the population have a considerable impact on unemployment. ARIMA is the best model for predicting Malaysia's unemployment rate (2,1,2). Malaysia's jobless rate has risen, according to studies. This suggests that from 2017 to 2026, Malaysia's jobless rate will increase somewhat. The unemployment rate in Malaysia is expected to remain constant in the coming decade and will not be a major concern. Overall, Malaysia's unemployment rate has risen somewhat during the past year. The first goal is to look at the major factors that influence Malaysia's unemployment rate. According to the study's findings, prevalent variables in Malaysia, such as inflation and population growth, have a substantial impact on the unemployment rate. ([10] Ramli, S. F., Fidaus, M., Uzair, H., Khairi, M. & Zharif, A, 2018)

Moreover, a research paper about the forecasting of unemployment rate in Malaysia using exponential smoothing methods also has been reviewed. The purpose of this research is to examine the most effective forecasting methodologies and choose the best timeframe for predicting Malaysia's future unemployment rate in 2016. There are five different sets of unemployment rates in Malaysia. To anticipate unemployment rates, the Naive, Simple exponential Smoothing (SES), and Holt methods are utilised. The prediction model is then chosen based on the lowest accuracy criterion. The Holt model is the strongest predictor of the overall annual unemployment rate, male annual unemployment rate, and overall quarterly unemployment rate, according to the findings. Furthermore, SES is the strongest predictor of women's yearly and total monthly unemployment rates. Different strategies were employed in this case study to properly anticipate various data sets. Both approaches may be used to forecast the unemployment rate in Malaysia. As a result, both techniques are suggested for projecting unemployment rates in Malaysia using various data sets and timeframes. The smaller the scale employed, the more accurate the forecast, according to comparisons of yearly, quarterly, and monthly data. As a result, Malaysia's total jobless rate in 2016 was 3.4%. ([11] Maria E. N., Sabariah S., Lok S. L., Rohayu M. S., Norhaidah M. A., 2018.)

In addition, an initial research on the forecast model for Malaysia's unemployment rate has been examined. The goal of this study is to figure out which strategies are best for creating unemployment rate estimates from data from the Labour Force Survey series. Univariate modelling approaches, such as Naive and trend models, mean change models, double exponential smoothing, and Holt method models, are used to create the models analysed. These algorithms analyse trends like the quarterly unemployment rate to identify short-term forecasts (one quarter in advance). By keeping a portion of the quarterly observations as a persistence sample, the model's performance was confirmed. Comparisons are also done to assess how well historical and forecasted data match and correlate. The selection of the best appropriate model was represented by the minimal mean square error (MSE). The Holt method model is the best model for forecasting the quarterly unemployment rate, according to the data. The Holt technique is the best model for forecasting quarterly unemployment rates, according to one-step forward forecasting research. Each model type has its own set of properties that are best suited to a certain data set. More forecasting approaches must be investigated to handle longer unemployment sequences. Univariate modelling approaches are essentially univariate models that provide prediction values based on prior data. This is predicated on the premise that the anticipated values are completely determined by the data series' previous trends. ([12] Mohd Nadzri M. N., Kon Mee Hwa, Huzaifah M., 2011)

Besides, a paper with the name unemployment rates forecasting with Grey-Based Models in the Post-Covid-19 Period: A case study from Vietnam also reviewed. Because the data set is non-stationary and non-linear, multiple time series models were used to forecast the unemployment rate. The goal of this research is to develop a better forecasting approach for estimating unemployment rates in Vietnam in the face of uncertainty caused by a lack of information and data. GM(1,1), Grey Verhulst (GVM), and ARIMA (Autoregressive Integrated Moving Average) models based on the Grey theory system were suggested in this study to better precisely estimate the unemployment rate. From 2014 to 2019, unemployment rates from six distinct rural and urban locations in Vietnam were utilised to apply the model. The results reveal that the GM(1,1) model produces reduced average percentage error (MAPE) values in both rural and urban regions when compared to other standard techniques (excluding urban upland area). The ARIMA and GVM models have the same accuracy level as the GM(1,1) model. The findings of this research suggest that the model's consequences can aid policymakers in shaping future labour and economic policies. ([13] Phi-Hung Nguyen, Jung-Fa Tsai, Ihsan Erdem Kayral, Ming-Hua Lin, 2021.)

CHAPTER 3 Methodology

3.1 Introduction

A time series is a set of observations X_t, each of which is recorded at a specific time (t). The time interval of a time series can be minutes (such as high-frequency financial data), days, weeks, months, quarters, years, or even larger time units. For example, GROSS domestic product (GDP), consumer price index (CPI), weighted stock price index, interest rate, exchange rate, etc. The picture below is an example of a time series, the x-axis represents the time, and the y-axis represents the observations.

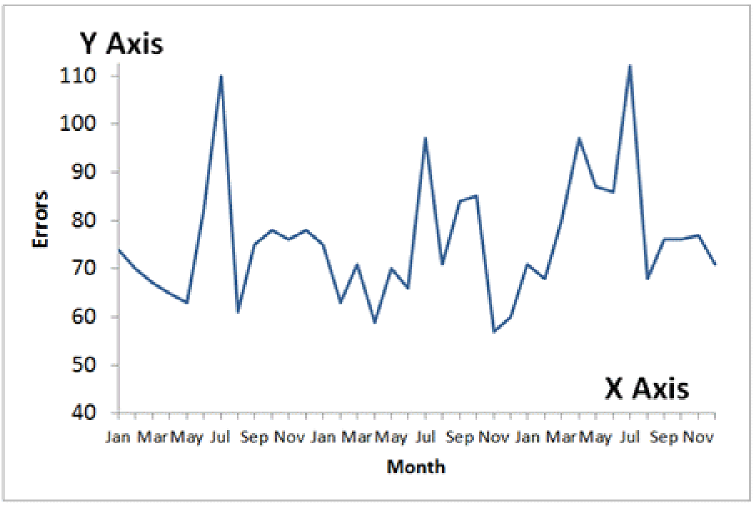

Figure 3.1: Example of Time Series Graph

3.2 Types of Time Series

3.2.1 White Noise

White noise is that the random variables at any two time points are not correlated, and there is no dynamic law that can be used in the sequence. Therefore, historical data cannot be used to forecast and infer the future. The autocorrelation function of white noise sequence is 0. White noise is an extreme case of stationary time series and has a very wide range of applications. Since the values at the time points before and after are not correlated, white noise sequence can be regarded as a new sequence, that is, the white noise value εt at time t is not correlated with the previous white noise sequence $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. If a time series satisfies the following conditions, then it is a white noise time series:

a. The variables are independent and identically distributed with a mean of zero;

$$E(\varepsilon_t) = 0$$

b. All variables have the same variance and each value has a zero correlation with all other values in the series;

$$Var(e) = \sigma^2$$

c. For any s not equal to t, ε_t and ε_s are not correlated.

$$E(\varepsilon_t \varepsilon_s) = 0$$

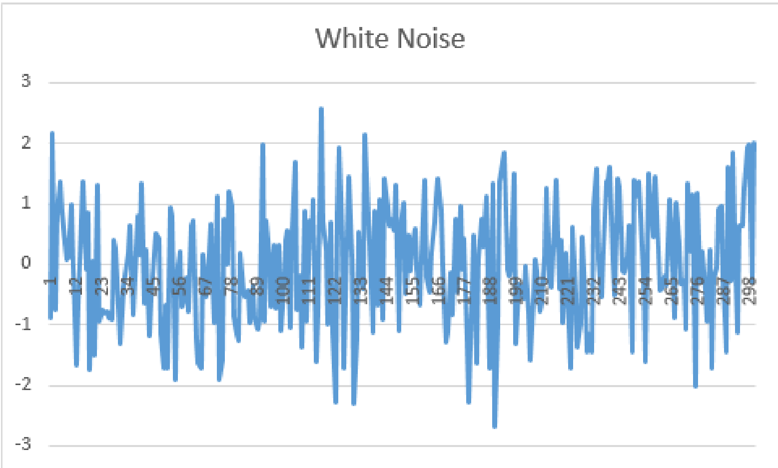

Figure 3.2: Example of White Noise Graph

3.2.2 Stationary Time Series

Stationary time series is the most ideal time series property, which means that the basic characteristics of random variables in a time series remain unchanged in the future for a long period of time, the same as in the past. Assume that a time series is generated by a stochastic process, that is, assume that every value of the time series $\{X_t\}$, and t = 1, 2, ... is randomly obtained from a probability distribution. That is to say, even if a time series has a slip of time T, the joint distribution remains unchanged, which is called a strong stationary time series:

$$Fx_{t_1}, x_{t_2}, \dots, x_{t_n}\,(x_1, x_2, \dots, x_n) = \; Fx_{t_{1+k}}, x_{t_{2+k}}, \dots, x_{t_{n+k}}\,(x_1, x_2, \dots, x_n)$$

Strongly stationary time series is a very ideal situation, in this case, the results of time series analysis will be very good. However, the actual data is difficult to meet this demand to generate a strongly stationary time series. There is another stationary time series which is a weak stationary time series. If a time series meet the following conditions then it is a weak stationary time series:

a. The mean value is a constant independent of time t;

$$E(X_t) = \mu$$

b. The variance Var(Xt)=s^2 is a constant independent of time t;

$$Var(X_t) = \sigma^2$$

c. Covariance Cov(Xt,Xt+k)=gk is a constant that depends only on the interval k, but not on time t.

$$Cov(X_t, X_{t+k}) = \gamma_k$$

Whether strong stationary or weak stationary time series, it describes the statistical properties of time series and the invariance of time stationary. Strong stationary time series require that all statistical properties are invariant with respect to time translation, while weak stationary time series require only that expectation and covariance are invariant.

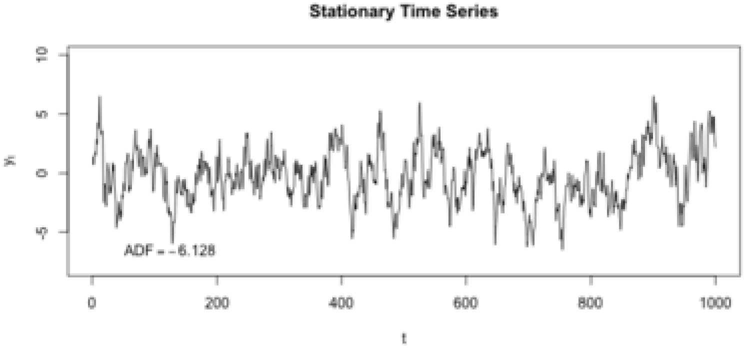

Figure 3.3: Example of Stationary Time Series Graph

3.2.3 Non-stationary Time Series

If a time series does not satisfy stationary time series, but satisfies trend, seasonality, or periodicity, then we can classify it as non-stationary time series. The first is trend, which can be divided into deterministic trend and stochastic trend. Deterministic trend means that the trend of a sequence can be fitted with a trend line, and the sequence becomes a stationary time series after the trend is removed. Random trends can not be analyzed by a single trend line, and the trend changes randomly at different times.

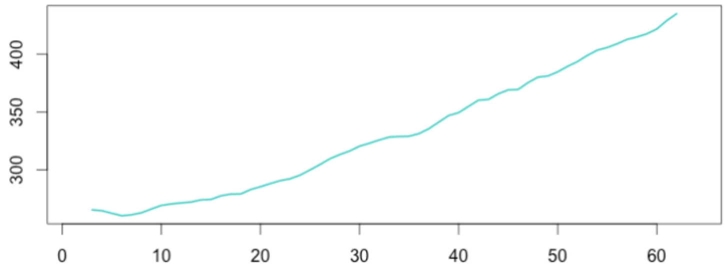

Figure 3.4: Example of Trended Time Series

The second is seasonality, which cycles like the seasons of the year.

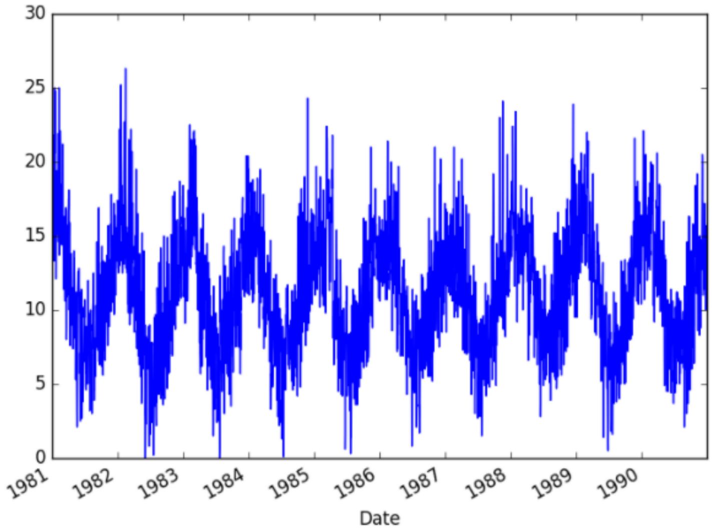

Figure 3.5: Example of Seasonality Time Series

And there's a periodicity, and seasonal cycles of change, but change cycles are not very accurate, can be change, for example, we can say that a sequence once every three loop, it is seasonal, and there is a sequence, sometimes once every three months cycle, cycle time, five months sometimes although can see that it has a cycle, but changing cycle So that's periodicity and for the sake of simplicity, the rest of the discussion is going to focus on seasonality.

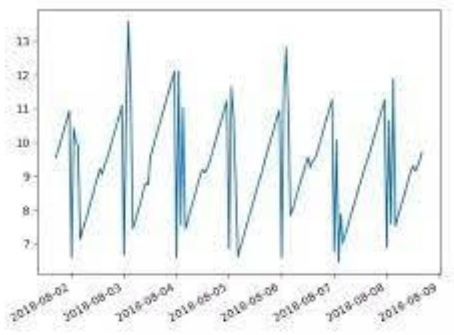

Figure 3.6: Example of Periodicity Time Series

Non-stationary time series tend to be stationary, trend and seasonal superimposed together:

$$Z_t = T_t + S_t + X_t$$

T_t represents the deterministic trend change, S_t represents the deterministic seasonal change, X_t is the time series satisfying the stationarity. Non-stationary time series need to transform to stationary time series to continue the analysis and forecasting. Time series data including time trend and seasonal factors may be transformed into stationary time series after processing. There are ways in which we can get rid of trends and seasonality, let the sequence level off and then analyze it, these will be introduced in the ARIMA model.

3.3 Model of Time Series

3.3.1 Auto-Regressive Model (AR)

In an autoregression model, we use a linear combination of the variable's historical values to anticipate the variable of interest. The word autoregression denotes that the variable is being regressed against itself. Thus, a p-order autoregressive model can be written as follows:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t$$

where εt stands for white noise, c is a constant term. This is similar to multiple regression but with lagged values of yt as predictors. This is known as an AR(p) model, which stands for autoregressive model of order p.

Autoregressive models are surprisingly adaptable when it comes to dealing with a variety of time series patterns.When the parameters ϕ_1, ..., ϕ_p are changed, distinct time series patterns emerge. The error term t's variance will only affect the scale of the series, not the patterns.

For an AR(1) model:

- When $\phi_1 = 0$, y_t is equivalent to white noise;
- When $\phi_1 = 1$ and c = 0, y_t corresponds to a random walk model;
- When $\phi_1 = 1$ and c =/ 0, y_t corresponds to a random walk model with drift;
- When $\phi_1 < 0$, y_t tends to fluctuate between positive and negative values.

We normally restrict the use of autoregressive models to stationary data and place some restrictions on the regression coefficients:

- For an AR(1) Model: $-1 < \phi_1 < 1$.
- For an AR(2) Model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

When $p \geq 3$, the constraints are even more difficult.

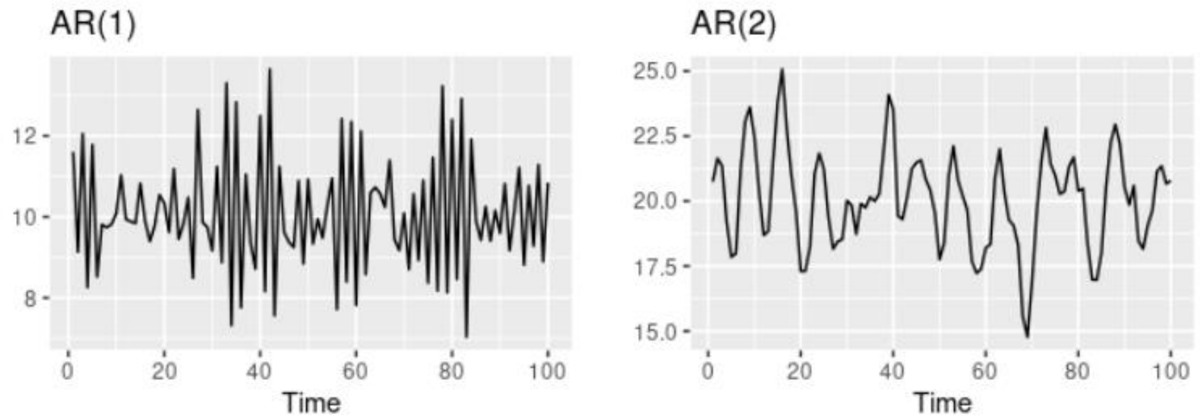

Figure 3.7: Example of AR(1) and AR(2)

Autoregressive models have many limitations:

(1) Autoregressive models use their own data to predict;

(2) Time series data must be stable;

(3) Autoregressive models are only applicable to predict phenomena related to their own early stage (autocorrelation of time series).

3.3.2 Moving Average Model (MA)

Instead of using historical values of prediction variables for regression, the moving Average model uses historical forecast errors to build a regression-like model.

$$y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$y_t = c + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

ε_t in the above formula is white noise. We call this model MA(q) model, namely the q-order moving average model. Of course, since we do not observe the value of ε_t, this is not a linear model in the general sense.

Note that each value of y_t can be considered a weighted moving average of historical prediction errors. Moving average models are used to forecast future values, while moving average smoothing laws are used to estimate cyclical trends in historical values.

Changing θ_1, ..., θ_q will make the data show different time series characteristics. As with the autoregressive model, the variance of the error term ε_t will only change the range of values of the series, but will not change its characteristics.

If we restrict the coefficients of the MA model, we can also express the MA model by the AR model. Such a MA model is said to be invertible. In other words, we can express any reversible MA(q) model as an AR(∞) model. Invertible models not only allow us to convert any MA model into an AR model, they also have a lot of really nice mathematical properties such as, consider the MA(1) model: In the AR(∞) representation of $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, the present error can be expressed as a linear function of present and historical observations：

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j y_{t-j}$$

When $|\theta| > 1$, each weight will increase as the number of delay periods increases, so the farther the observation period is from the current, the greater their influence on the current error will be. When $|\theta| = 1$, the coefficients will remain the same, and each observation will have the same effect. Because few of these scenarios are realistic, So we usually prescribe $|\theta| < 1$, the closer the observation, the more influence it has on the current observation. Therefore, MA model is reversible for $|\theta| < 1$.

Other models have similar limitations on reversibility and stationarity:

- For a MA(1) model: $-1 < \theta_1 < 1$.
- For a MA(2) model: $-1 < \theta_2 < 1$, $\theta_2 + \theta_1 > -1$, $\theta_1 - \theta_2 < 1$.

When $q \geq 3$, the constraints are even more difficult.

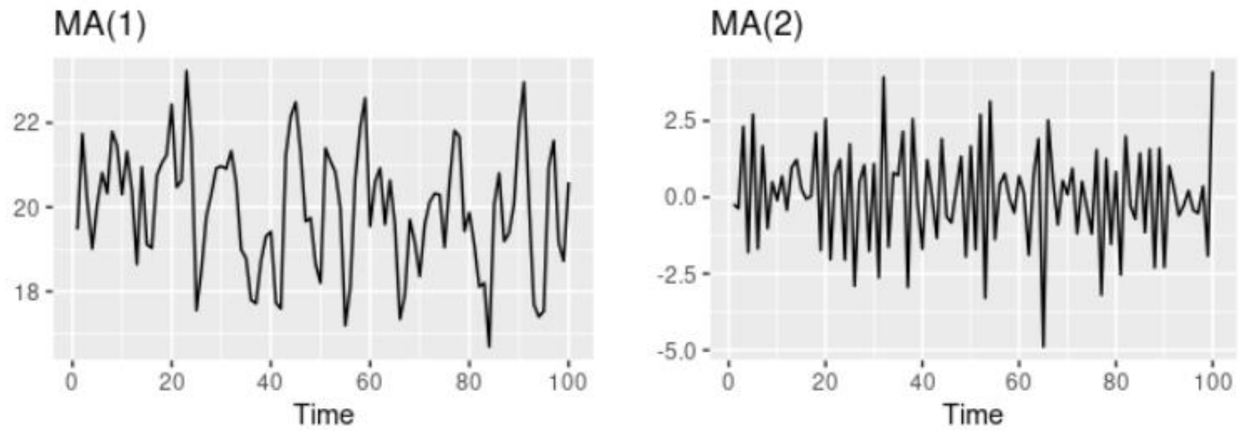

Figure 3.8: Example of MA(1) and MA(2)

The formula shows that:

(1) A random time series by an autoregressive moving average model, namely the sequence can be made by its own past or lag value and random perturbation terms to explain.

(2) If the sequence is smooth, namely its behavior does not change over time, then we can predict the future by the past behavior of the sequence.

3.3.3 Auto-Regressive Moving Average Model (ARMA)

The Auto-Regressive Moving Average model is an important method to study time series, which is composed of an auto-regressive model (AR model) and moving average model (MA model). The ARMA(*p, q*) model contains *p* auto-regression terms and *q* moving average terms. The ARMA(*p, q*) model can be expressed as:

$$y_t = \mathrm{c} + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Different features of time series are extracted from AR and MA respectively to construct a more robust model.

3.3.4 Auto-Regressive Integrated Moving Average Model (ARIMA)

The basic idea of the ARIMA model is to transform non-stationary time series into stationary time series, and then regression the dependent variable only to its lag value and the present value and lag value of random error term. Basically, the ARIMA model is the difference method introduced on the basis of the ARMA model.

First of all, what is the difference method? The difference method is to calculate the difference between adjacent observations and get a new sequence. What does the difference method have to do with time series analysis? As mentioned in the previous non-stationary time series, we always hope that the sequence data is stable, but sometimes, the data may be not stable, but the difference between the data is stable, so a difference method is actually a method that may make the sequence stable. Sometimes, we may need to make multiple differences. The parameter *d* is how many differences are needed to apply on a non-stationary time series to make it become a stationary time series. If the time series itself is stable and the data does not need to make differences, then the ARMIA model is ARIMA(*p, 0, q*), which is equivalent to ARMA(*p, q*). ARIMA(*p, d, q*) model can be expressed as:

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

Where *L* represents the Lag operator.

3.4 Generate a Time Series Model

For the obtained time series data, the data quality should be checked first, such as whether there are any missing or outliers. Stability tests and white noise tests are required to ensure that the data is correct. The time series that can be analyzed and predicted by ARMA model must meet the condition of stationary non-white noise series. Therefore, it is an important step of time series analysis to test the stationarity of data. According to the scatter plot, autocorrelation function and partial autocorrelation function of time series, the variance, trend and seasonal variation rule of ADF unit root are tested to identify the stationarity of the series. Generally speaking, the time series of economic operations are not stationary. Next, the non-stationary sequence is stabilized. If the data sequence is non-stationary and has a certain trend of growth or decline, differential processing is required for the data; if the data has heteroscedasticity, technical processing is required for the data until the values of autocorrelation function and partial correlation function of the processed data are not significantly different from zero. Then, according to the recognition rules of the time series model, the corresponding model is established. If the partial correlation function of a stationary sequence is truncated and the autocorrelation function is trailing, it can be concluded that the sequence is suitable for the AR model. If the partial correlation function of a stationary sequence is trailing and the autocorrelation function is truncated, it can be concluded that the sequence is suitable for the MA model. If the partial correlation function and autocorrelation function of stationary series are trailing, the series is suitable for the ARMA model. After finding a consistent model, we need to perform parameter estimation to test whether it has statistical significance. Then a hypothesis test is performed to diagnose whether the residual sequence is white noise. Finally, the tested model is used for predictive analysis.

3.4.1 Time Series Stationarity and Randomness Test

After getting the time series data, the randomness and stationarity of the data should be tested first, which are important parts of time series prediction. Different analysis methods should be adopted according to different test results. Why does time series require stationarity? Stationarity is the requirement that the curve fitted from the sample can continue to develop in the current shape and trend in the future for a period of time, so that the prediction result will be meaningful. For stationary acoustic sequence, its mean and variance are constant. Now has a very mature steady sequence modeling method Is usually a linear model is established for the development of this sequence Useful information to extract the sequence For nonstationary sequence, because of its mean and variance is not stable, smooth processing method is usually turn it into a sequence, so it can be applied to the stationary time series analysis method, For example, the ARIMA model is established to carry out corresponding research, or the trend and seasonality are decomposed and exponential smoothing model is applied according to the situation. For a pure random sequence, also known as white noise sequence, there is no correlation between the items of the sequence, and the sequence is carrying out completely disordered random fluctuation. The analysis of the sequence can be terminated. White noise sequences are stationary sequences with no information to extract.

Test of stationary time series

a. Time series diagram test

The most basic approaches for stationarity detection rely on visually detecting if the data, or functions of it, show some recognised attribute of stationary or non-stationary data. According to the property that the mean variance of stationary time series is constant, the time series diagram of stationary time series should show that the series always fluctuates randomly near a constant value, and the range of fluctuations is bounded without obvious trend and periodic characteristics. However, non-stationary time series often show different mean values in different time periods, for example, it will continuously rise or decline.

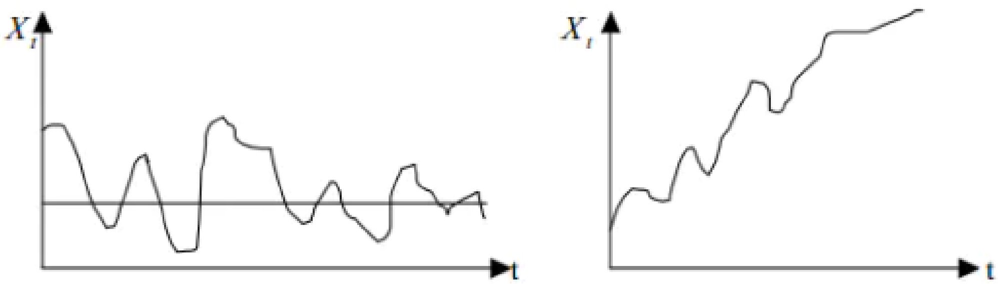

Figure 3.9: Example of Stationary Time Series and Non-stationary Time Series

b. Autocorrelation graph test

The correlation of a signal with a delayed copy or with its own delay as a function of the delay is known as autocorrelation. Stationary sequences generally show short-term correlations when ACF values with incremental lag are displayed (called correlays). This property is described by the autocorrelation coefficient, which states that as the delay period K increases, the stationary sequence's autocorrelation coefficient decays rapidly to zero and fluctuates randomly near zero, whereas the non-stationary sequence's autocorrelation coefficient decays slowly.

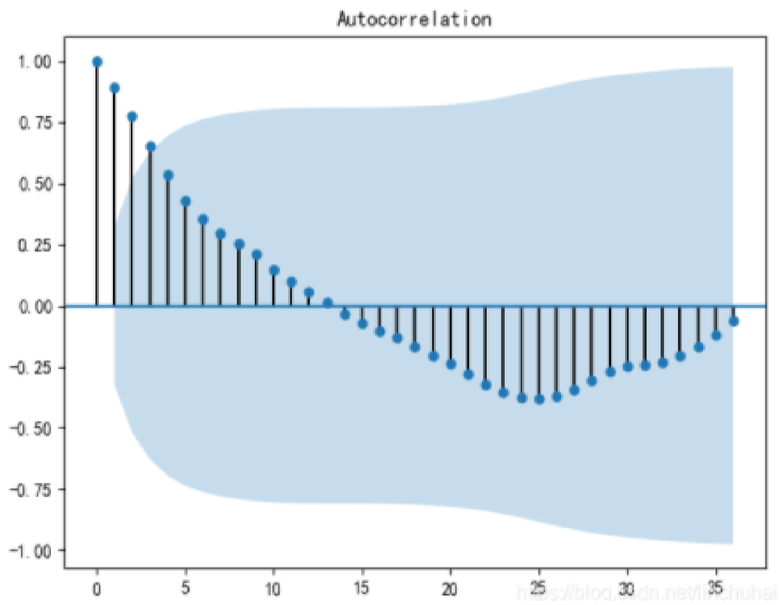

Figure 3.10: Example of Autocorrelation Graph

c. Augmented Dickey–Fuller (ADF) tests

The Dickey Fuller (DF) test is one of the most popular statistical tests that can be used to determine the presence of the unit root in a sequence, thereby helping to determine whether the sequence is stationary. The ADF test is an extension of the DF test. The null hypothesis and alternative hypothesis of this test are as follows:

H_0: Sequence has a unit root (sequence is nonstationary)

H_1: The sequence has no unit root (sequence stationary)

What is the unit root? When an autoregressive process: $y_t = by_{t-1} + a + \varepsilon_t$, if the lag term coefficient b is 1, it is called the unit root. When there is unit root, the relationship between the dependent and independent variables and deceptive, because any error residual sequence will not decay with increasing sample size, that is to say, in the model the influence of the residual error is permanent This return is also called faking regression If a unit root, this process is a random walk ADF test is to judge whether there is unit root in the sequence: if the sequence is stable, there is no unit root; Otherwise, there would be a unit root. The results of ADF mainly look at the following two aspects:

1. If the Value of Test Statistic is smaller than Critical Value (5%), it meets the stability requirements.
2. The lower the p-value (theoretically lower than 0.05), the more stable the sequence.

Test of randomness (White Noise)

For pure random sequence, it is generally tested by constructing statistics. As we know, the autocorrelation coefficients of white noise sequences should all be 0 except for the 0-order autocorrelation coefficient, that is, the variance. Therefore, we can propose the following hypothesis:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0, \forall_m \geq 1$$
$$H_1: At\ least\ one\ \rho_k \neq 0, \forall_m \geq 1, k \leq m$$

Therefore, around this hypothesis, we can construct statistics for testing. Commonly used statistics are Q statistics and Ljung-Box (LB) statistics, and their calculation formulas are as follows:

$$Q = n \sum_{k=1}^{m} \hat{\rho}_k^2$$
$$LB = n(n+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{n-k} \right)$$

Where, n is the number of observation periods of the sequence, m is the number of specified delay periods, K is the delay order, and Box and Pierce prove that the two statistics obey the Chi-square distribution of degree of freedom m. When the statistic is greater than $\chi^2_{1-\alpha}(m)$ or the p value is less than α, the null hypothesis can be rejected, that is, the sequence is considered to be non-random.

3.4.2 Stabilization of Time Series

Difference Method

When we get the data, the data can fluctuate a lot, so how do we get the data to be even, and one of the ways to do that is the difference method. In the difference method, the difference of consecutive terms in a sequence is calculated. The difference operation is usually performed to eliminate changes in the mean. Mathematically, difference can be written as:

$$y_t = y_t - y_{t-1}$$

Where y_t are the values at time *t*, the number of intervals between subtracting values is the order, and the above formula is the first-order difference.

Sometimes the data is still not stationary after the difference, so it may be necessary to differentiate the data again to get a stationary sequence:

$$y_t'' = y_t' - y_{t-1}'$$
$$y_t'' = y_t - 2y_{t-1} + y_{t-2}$$

In this case, the sequence y_t'' of length *T - 2* and then we can model the change in the change in the original data and in real life applications, it's usually not necessary to make a difference of more than second order.

Seasonal difference is the difference between an observation and the corresponding observation of the previous year. Thus there are:

$$y_t' = y_t - y_{t-m}$$

Where *m* equals the number of seasons in a year. Because the time interval between the two subtracted observations is m if the seasonal difference data is white noise, the original data can be fitted with a suitable model:

$$y_t = y_{t-m} + \varepsilon_t$$

If the seasonal difference sequence is represented by $y_t' = y_t - y_{t-m}$, then its second-order difference sequence is:

$$y_t'' = y_t' - y_{t-1}'$$

$$y_t'' = y_t - y_{t-1} - y_{t-m} + y_{t-m-1}$$

When both the seasonal difference and the first difference are used, the order of the two does not affect the result, and the result is still the same after the transformation of the order. However, if the data has strong seasonal characteristics, we recommend seasonal difference first, because sometimes the data after seasonal difference is smooth enough, there is no need to carry out subsequent differences. If the first difference is made first, we will still need to make a seasonal difference. When using a difference, it is important that the difference should be interpretable. The first difference is the difference between adjacent observations, and the seasonal difference is the change in observations from adjacent years. Other differences in delayed periods are difficult to explain and should therefore be avoided at all costs.

3.4.3 Model Identification

After the data is preprocessed and the stationary time series is obtained, model recognition can be carried out. Then determine the parameters related to the ARIMA model. ARIMA takes three parameters, which can be expressed as ARIMA(*p*, *d*, *q*). *p* is the order of autoregression, that is, how many time periods of data do we need to use to do autoregression. *d* is the number of differences made when time becomes stationary, and *q* is the order of moving average.

Autocorrelation functions (ACF) and Partial Autocorrelation functions (PACF)

We select the suitable model fitting observation value sequence by investigating the properties of stationary sequence sample autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF). Therefore, the first step of model fitting is to calculate the sample autocorrelation coefficient and partial autocorrelation coefficient of the observation value sequence according to the value of the observation value sequence.

In autoregression, the ACF Plot depicts the connection between y_t and y_{t-k} for various k values. If y_t already has a correlation with y_{t-1}, then y_{t-1} must also have a correlation with y_{t-2}. However, y_t and y_{t-2} are bound to be correlated because they are both associated with y_{t-1}, not because y_{t-2} includes additional information that can be used to forecast y_t, thus partial autocorrelation can be utilised to alleviate this problem. It calculates the link between y_t and y_{t-k} without taking into account the effects of delays *1, 2, 3, ..., k - 1* on y_t. Because there is no delay to remove, the delay first-order partial autocorrelation coefficient is the same as the delay first-order autocorrelation coefficient. In an autoregressive model, each partial autocorrelation coefficient may be approximated as a final coefficient.

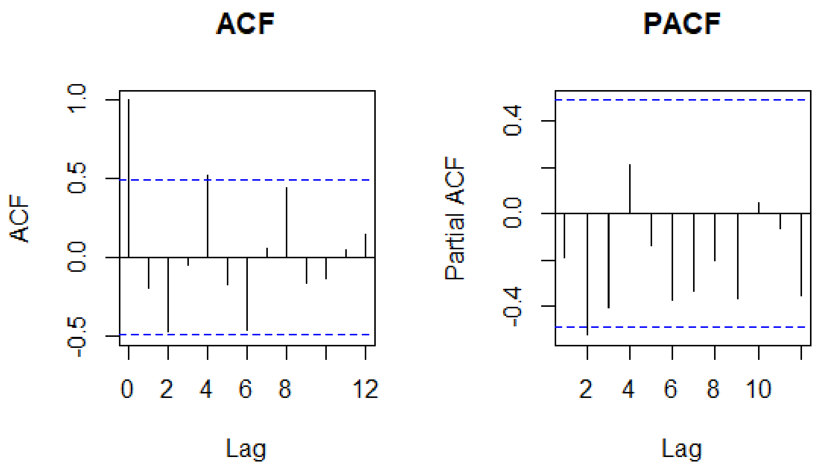

Figure 3.11: Example of ACF and PACF

The ARIMA(*p,d,0*) model may be used if autocorrelation graphs and partial autocorrelation graphs of difference data exhibit the following characteristics:

- The autocorrelation coefficient fell exponentially or varied in a sinusoidal pattern.
- There is a considerable hump in the delay P in the partial autocorrelation graph, but no such spike when the delay is bigger.

The ARIMA(*0,d,q*) model may be used if the autocorrelation graphs and partial autocorrelation graphs of difference data exhibit the following characteristics:

- The partial autocorrelation coefficient fell exponentially or varied in a sinusoidal pattern.
- In the autocorrelation graph, there are visible bumps in the delay Q, but no such spikes when the delay is higher.

According to the properties of ACF and PACF, we need to select the appropriate ARMA model to fit the observed value sequence. In fact, this process is to estimate the order p of the autocorrelation and the order Q of the moving average according to the properties of the autocorrelation coefficient and partial autocorrelation coefficient of the sample. Therefore, the model recognition process is also called the model ordering process.

Terms	ACF	PACF
AR(p)	Tails off	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off
ARMA(p,q)	Tails off	Tails off

Table 3.1: Example of ACF and PACF of AR(p), MA(q) and ARMA(p,q)

To characterise the process, the ACF and PACF plots should be studied together. After p substantial delays, we predict the ACF plot to progressively decline whereas the PACF plot would sharply fall for the AR process. The ACF and PACF plots should indicate the reverse of an MA process, i.e., the ACF should show a sudden decrease after a given q number of lags, but the PACF should show a geometric or steady declining trend. If both the ACF and PACF plots show a steady decrease in pattern, then the ARMA process should be considered for modelling.

3.4.4 Model Checking

The process of model checking is performed to check for residuals. Model independence and normal distribution should be satisfied by residuals. If these assumptions are not satisfied, a new model for the series is chosen. One of the factors to evaluate is the model's goodness of fit. The residual graph should first be examined to see if it is randomly distributed.

Stationary R-squared

A measurement method that compares the constant part of a model to a simple mean model. With trend or periodic patterns, this measure is better than ordinary R-squared. Stationary R-squared can be negative numbers in the range from negative infinity to 1. A negative value means that the model being considered is worse than the baseline model and a positive value means that the model being considered is better than the baseline model.

R-Squared

The fitness measures of linear models are sometimes called judgment coefficients. Proportion of dependent variable variation as explained by the regression model. Its range value is between 0 and 1. The smaller the value, the less suitable the model is for the data.

Normalized Bayesian Information Criterion (BIC)

Under incomplete information, some unknown states are estimated by subjective probability, and then the occurrence probability is modified by Bayes formula. Finally, the optimal decision is made by using expected value and modified probability. The model with the lowest BIC is the best.

Root-Mean-Square Error (RMSE)

It is a commonly used measurement of the difference between values. Its value is usually the quantity predicted by the model or the observed estimator square mean deviation, which represents the sample standard deviation of the difference between the predicted value and the observed value. When these differences are estimated with data samples, they are usually called residual. When these differences are not calculated with the sample, it is usually called prediction errors square mean root offset, which is mainly used to gather the size of the errors in the prediction, usually in different time, with a magnitude to show its prediction ability Square mean root offset is a good measure of accuracy, but because it is related to the range of values, it is limited to comparing the prediction error of a particular variable between different models. If there is a predicted value that is significantly different from the true value, the RMSE will be large.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(x_{1,t} - x_{2,t}\right)^2}{n}}$$

Where the $x_{1,t}$ is the actual value,

$x_{2,t}$ is the forecasted value.

Mean Absolute Percentage Error (MAPE)

The actual online and offline sales forecast has a very important evaluation significance. However, in the actual project process, it is found that sometimes the indicators cannot represent the effect of model fitting very well. The MAPE is easier to grasp than the other accuracy measure data since it is expressed as a percentage. For example, if the MAPE is 10, then means the forecast is off by 10%. The formula of MAPE is as below:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|x_{1,t} - x_{2,t}\right|}{x_{1,t}}$$

Where the $x_{1,t}$ is the actual value,

$x_{2,t}$ is the forecasted value.

CHAPTER 4 Main Result

4.1 Model Identification

Table 4.1 below is a model description and the case processing summary of the unemployment rate (percentage) in Malaysia from March 2010 to July 2021 before any differencing. There are no missing values therefore we can continue to process the graph.

Model Description

Model Name		MOD_1
Series or Sequence	1	Unemployment Rate (Percentage)
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		12
Horizontal Axis Labels		Date_
Intervention Onsets		None
Reference Lines		None
Area Below the Curve		Not filled

Applying the model specifications from MOD_1

Case Processing Summary

		Unemployment Rate (Percentage)
Series or Sequence Length		137
Number of Missing Values in the Plot	User-Missing	0
	System-Missing	0

Table 4.1: Model description and case processing summary of Unemployment Rate (percentage) in Malaysia before differencing

The Figure 4.1 sequence chart below shows the trend of the unemployment rate (percentage) in Malaysia from March 2010 to July 2021. As we can see from the figure, the sequence has no obvious seasonal component, but there is an obvious change, so there is no need to do seasonal decomposition. However the figure also shows that the sequence is a non-stationary sequence.

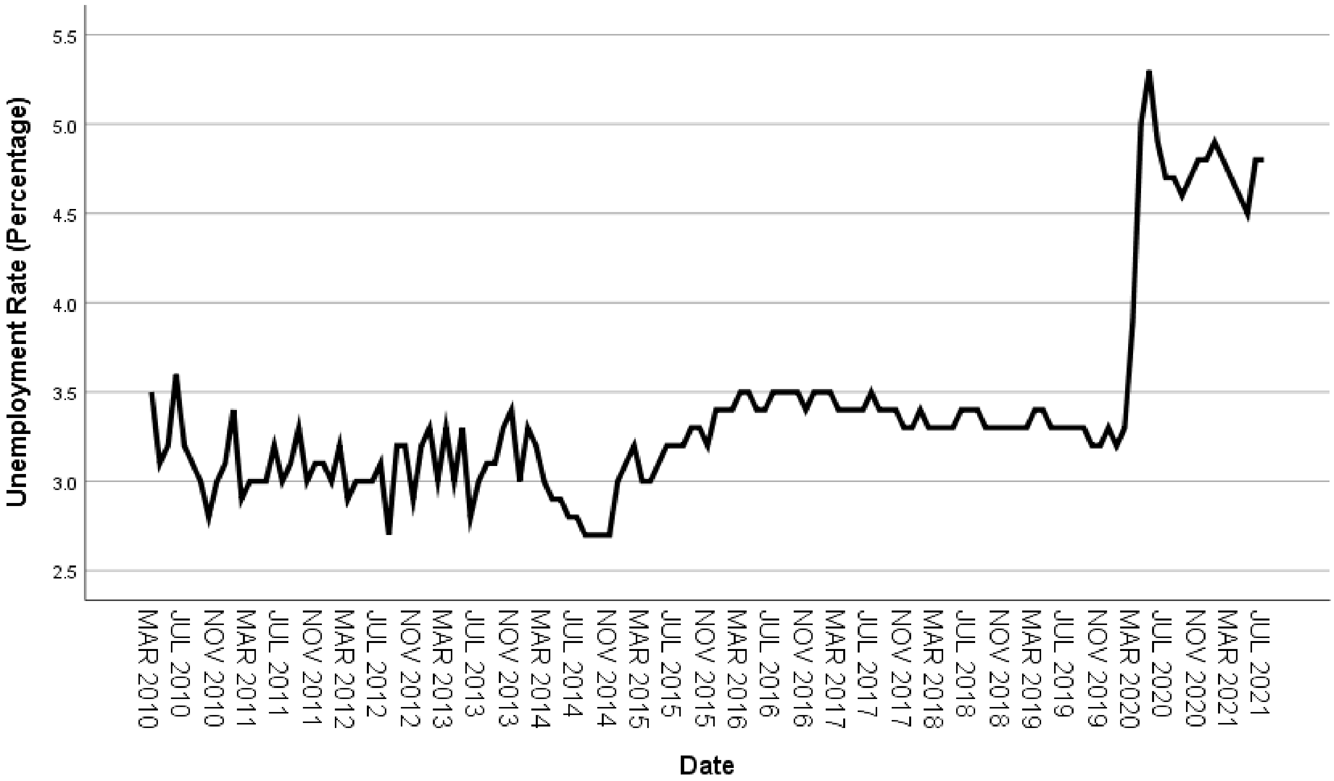

Figure 4.1: Sequence chart of the Unemployment Rate (percentage) in Malaysia before differencing

It can be seen from the Figure 4.2 and Figure 4.3 below that the correlation does not converge to zero and also both the autocorrelation graph (ACF) and partial autocorrelation graph (PACF) of the sequence are trailing. In addition, from the autocorrelation of the original sequence in Figure 4.2, it does not disappear rapidly when the lag is high, confirming the non-stationary behavior of the sequence, which also indicates that differencing is needed to achieve the stationary behavior.

Due to the slow decay of the ACF correlation graph, the difference transformation method was adopted to stabilize the original sequence, so as to make the first-order difference (d =1). There is no obvious peak in the autocorrelation graph within a constant period, indicating that there is no seasonality in the model.

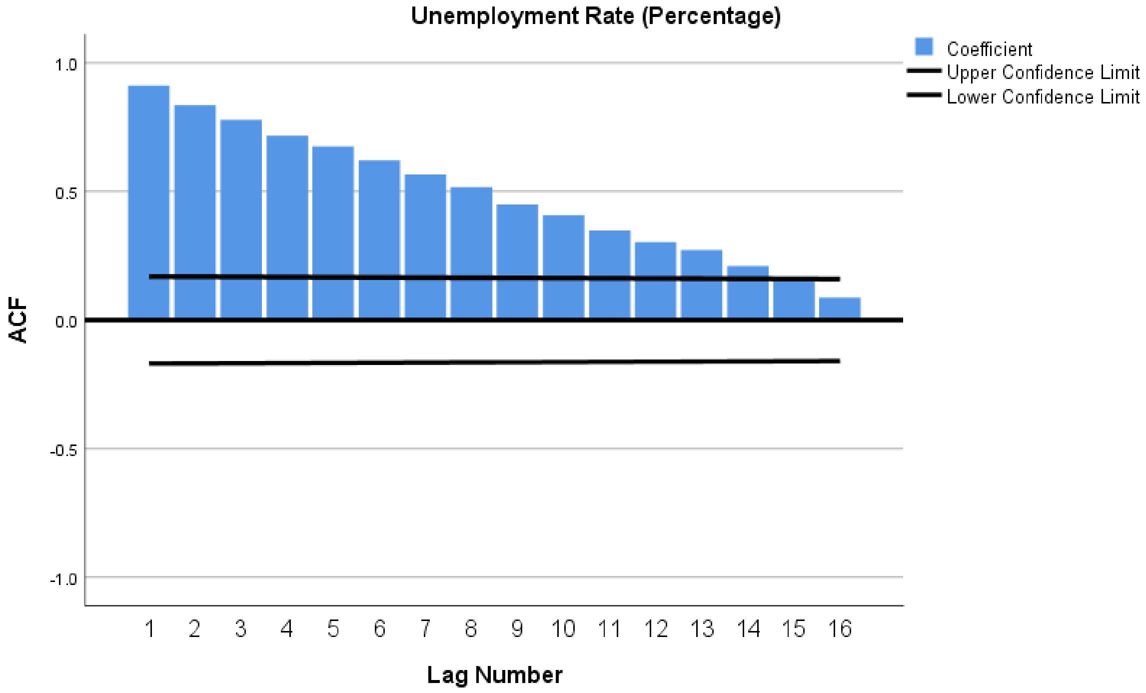

Figure 4.2: ACF correlogram of Unemployment Rate (percentage) in Malaysia before differencing

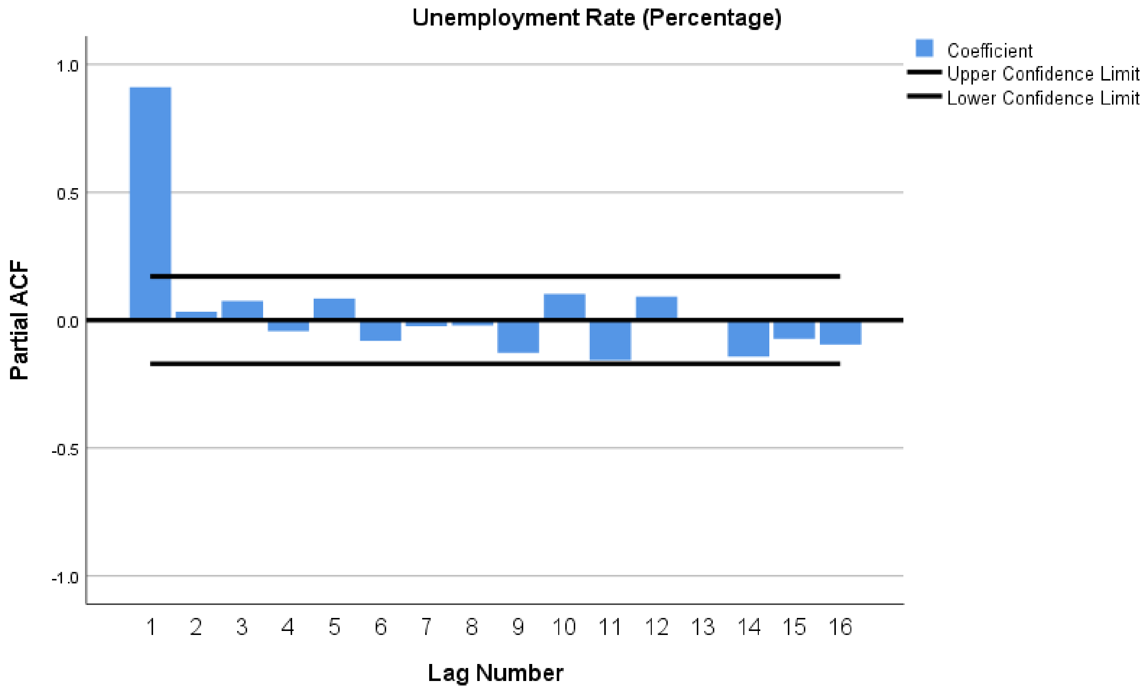

Figure 4.3: PACF correlogram of Unemployment Rate (percentage) in Malaysia before differencing

Table 4.2 and 4.3 below show the ACF and PACF plot (Correlogram) for the data.

Autocorrelations

Series: Unemployment Rate (Percentage)

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.911	.085	116.244	1	.000
2	.836	.084	214.777	2	.000
3	.779	.084	301.040	3	.000
4	.718	.084	374.811	4	.000
5	.675	.083	440.582	5	.000
6	.622	.083	496.752	6	.000
7	.567	.083	543.815	7	.000
8	.517	.082	583.290	8	.000
9	.449	.082	613.244	9	.000
10	.408	.082	638.211	10	.000
11	.348	.081	656.486	11	.000
12	.304	.081	670.523	12	.000
13	.272	.081	681.905	13	.000
14	.211	.080	688.781	14	.000
15	.151	.080	692.321	15	.000
16	.086	.080	693.497	16	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Table 4.2: Table of ACF of Unemployment Rate (percentage) in Malaysia before differencing

Partial Autocorrelations

Series: Unemployment Rate (Percentage)

Lag	Partial Autocorrelation	Std. Error
1	.911	.085
2	.033	.085
3	.075	.085
4	-.044	.085
5	.084	.085
6	-.080	.085
7	-.023	.085
8	-.021	.085
9	-.129	.085
10	.103	.085
11	-.156	.085
12	.091	.085
13	-.005	.085
14	-.143	.085
15	-.073	.085
16	-.096	.085

Table 4.3: Table of PACF of Unemployment Rate (percentage) in Malaysia before differencing

After the first modeling for the sequence of unemployment rates in Malaysia, we know that the series is not stationary. Therefore, the next step is to do the differencing on the original series.

Table 4.4 below shows the model description and case processing summary of the series after the first differencing.

Model Description

Model Name		MOD_2
Series or Sequence	1	Unemployment Rate (Percentage)
Transformation		None
Non-Seasonal Differencing		1
Seasonal Differencing		0
Length of Seasonal Period		12
Horizontal Axis Labels		Date_
Intervention Onsets		None
Reference Lines		None
Area Below the Curve		Not filled

Applying the model specifications from MOD_2

Case Processing Summary

		Unemployment Rate (Percentage)
Series or Sequence Length		137
Number of Missing Values in the Plot	User-Missing	0
	System-Missing	0

Table 4.4: Model description and case processing summary of Unemployment Rate (percentage) in Malaysia after first differencing.

It can be seen from the figure 4.4 below that the differencing sequence is basically evenly distributed on both sides of the 0 scale line, so it can be considered as stationary. However, the variability seems to be decreasing. Logarithmic transformation is considered for the sequence below variance over time and whose variance does not stabilize the variance.

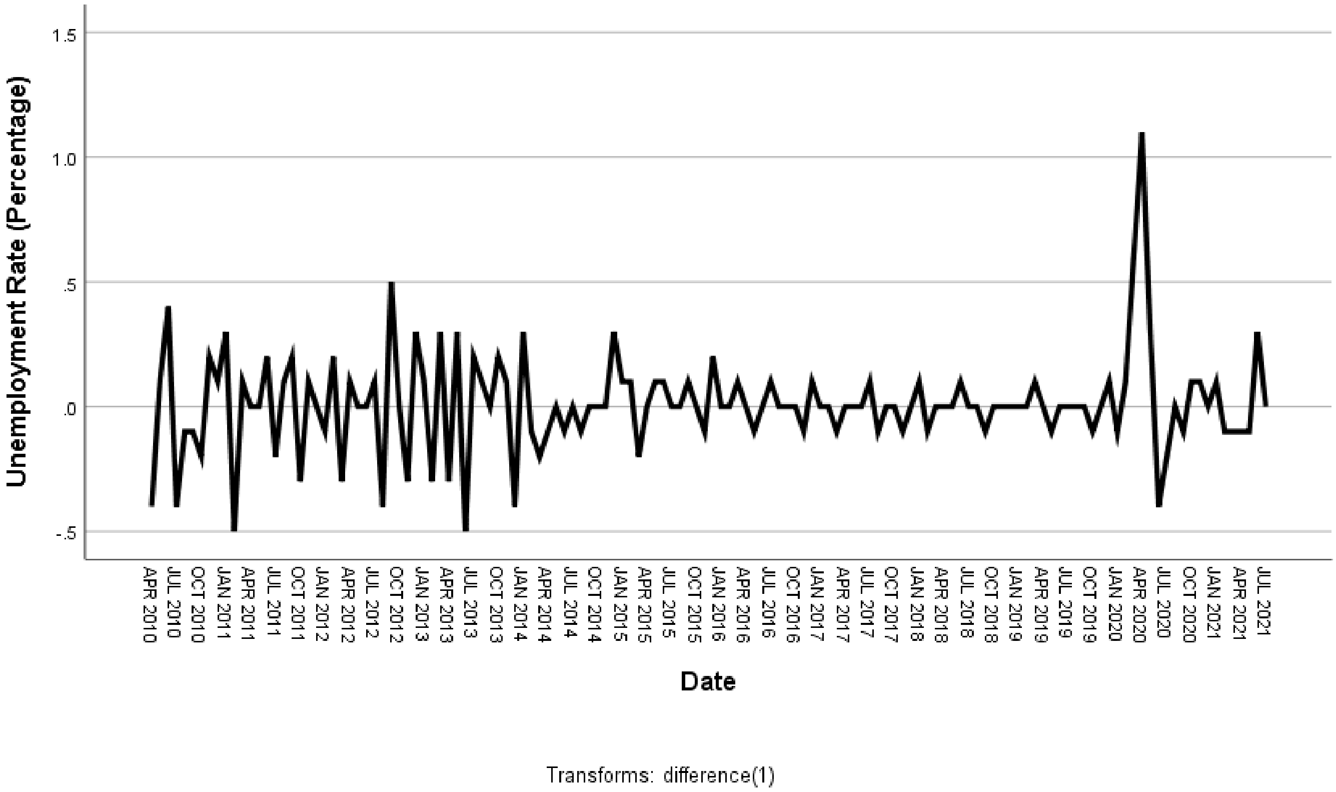

Figure 4.4: Sequence chart of the Unemployment Rate (percentage) in Malaysia after first differencing

The Figure 4.5 and Figure 4.6 below show that both the autocorrelation graph (ACF) and partial autocorrelation graph (PACF) of the sequence show that the sequence is stationary now and no further variance is required. This is because the ACF and PACF correlators are failing faster than the original sequence. Also, from the below figures, it gives no indication of seasonality as there is no repeated lag, which is a multiple of the number of cycles per season.

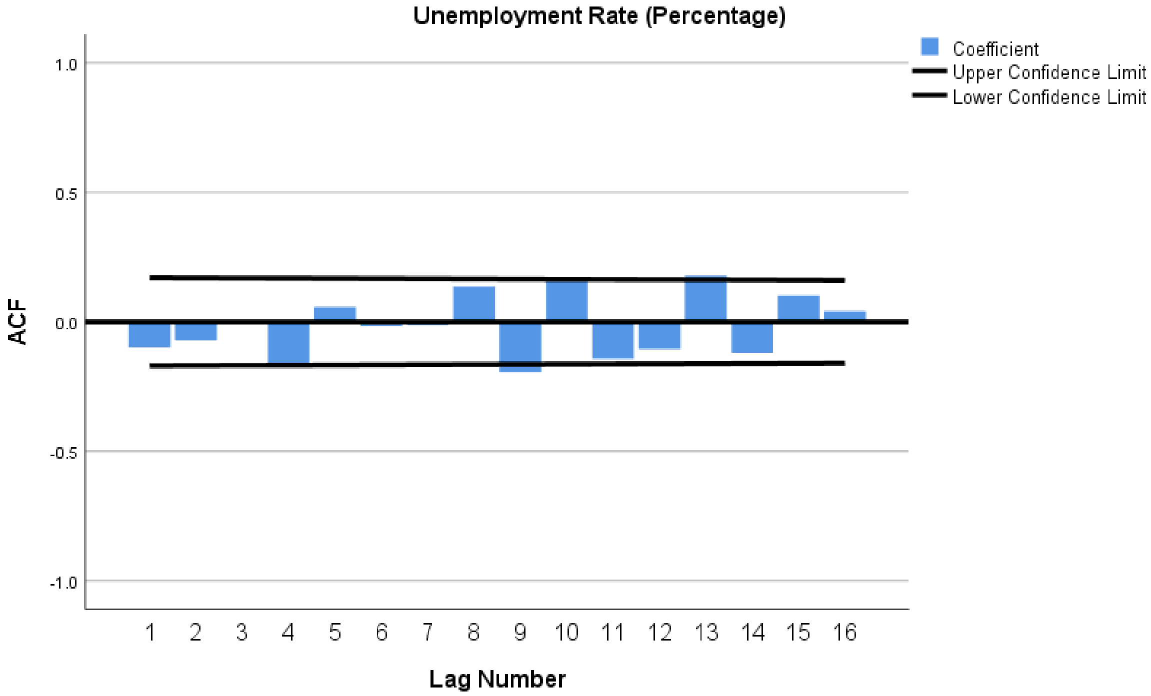

Figure 4.5: ACF correlogram of Unemployment Rate (percentage) in Malaysia after first differencing

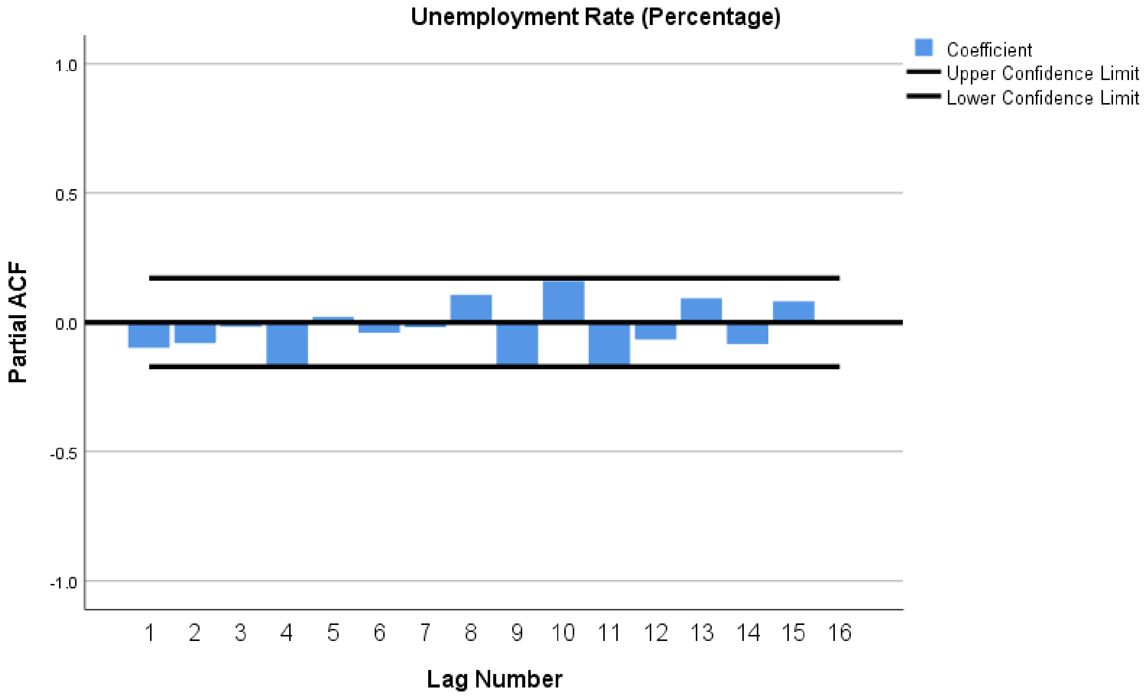

Figure 4.6: PACF correlogram of Unemployment Rate (percentage) in Malaysia after first differencing

Table 4.5 and 4.6 below show the ACF and PACF plot (Correlogram) for the data.

Autocorrelations

Series: Unemployment Rate (Percentage)

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	-.099	.085	1.363	1	.243
2	-.071	.084	2.061	2	.357
3	.000	.084	2.061	3	.560
4	-.165	.084	5.939	4	.204
5	.058	.084	6.418	5	.268
6	-.018	.083	6.466	6	.373
7	-.013	.083	6.490	7	.484
8	.137	.083	9.241	8	.322
9	-.193	.082	14.751	9	.098
10	.169	.082	19.025	10	.040
11	-.144	.082	22.123	11	.023
12	-.105	.081	23.783	12	.022
13	.177	.081	28.580	13	.008
14	-.119	.081	30.749	14	.006
15	.102	.080	32.356	15	.006
16	.041	.080	32.614	16	.008

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Table 4.5: Table of ACF of Unemployment Rate (percentage)in Malaysia after first differencing

Partial Autocorrelations

Series: Unemployment Rate (Percentage)

Lag	Partial Autocorrelation	Std. Error
1	-.099	.086
2	-.081	.086
3	-.016	.086
4	-.175	.086
5	.021	.086
6	-.040	.086
7	-.018	.086
8	.107	.086
9	-.168	.086
10	.161	.086
11	-.167	.086
12	-.066	.086
13	.093	.086
14	-.084	.086
15	.081	.086
16	-.004	.086

Table 4.6: Table of PACF of Unemployment Rate (percentage) in Malaysia after first differencing

Table 4.7 below shows the model description and case processing summary of the series after the first differencing and logarithmic transformation.

Model Description

Model Name		MOD_3
Series or Sequence	1	Unemployment Rate (Percentage)
Transformation		Natural logarithm
Non-Seasonal Differencing		1
Seasonal Differencing		0
Length of Seasonal Period		12
Horizontal Axis Labels		Date_
Intervention Onsets		None
Reference Lines		None
Area Below the Curve		Not filled

Applying the model specifications from MOD_3

Case Processing Summary

		Unemployment Rate (Percentage)
Series or Sequence Length		137
Number of Missing Values in the Plot	Negative or Zero Before Log Transform	0
	User-Missing	0
	System-Missing	0

Table 4.7: Model description and case processing summary of Unemployment Rate (percentage) in Malaysia after first differencing and logarithmic transformation

As can be seen from Figure 4.7, the scale has changed and the difference after transformation also fluctuates around 0, which looks more stable than the sequence graph after the first differencing. Therefore, the difference of the logarithm transformation will be considered for the following prediction.

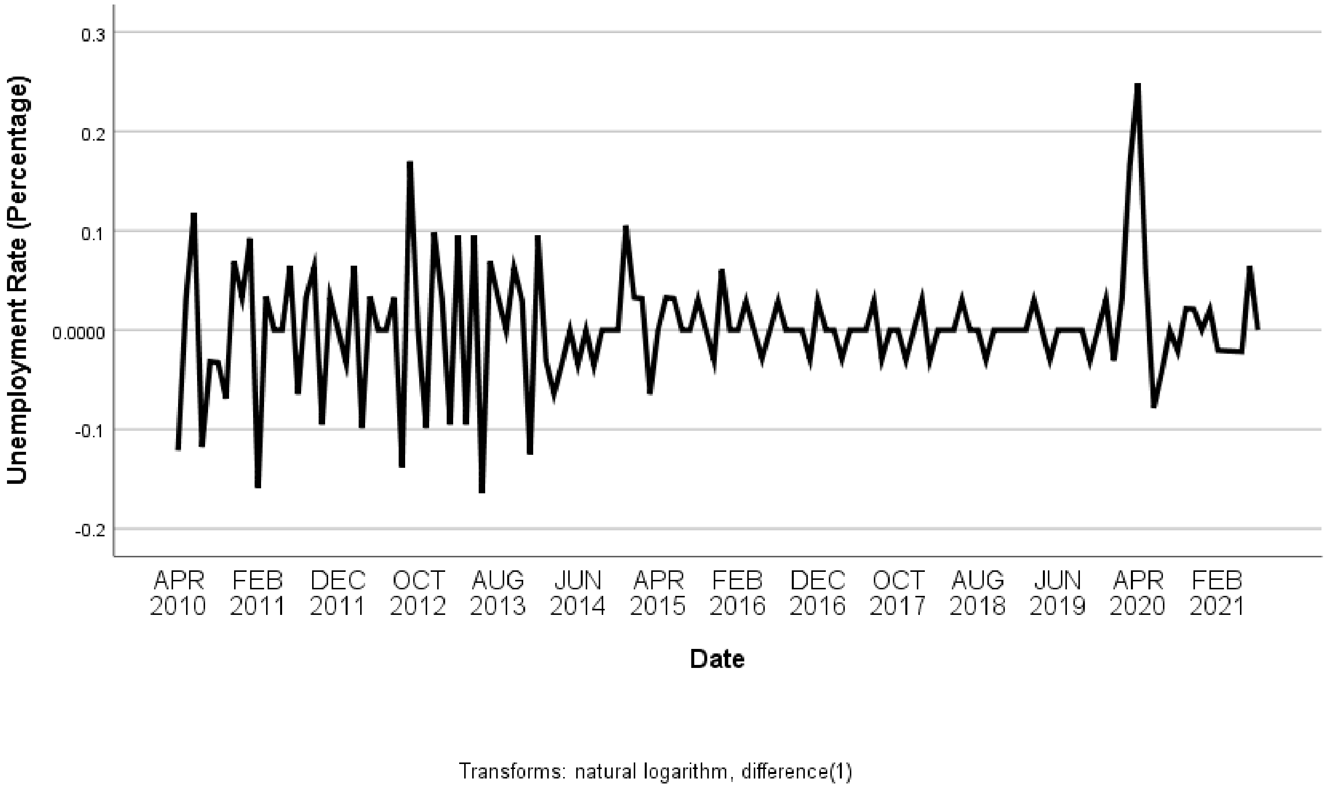

Figure 4.7: Sequence chart of the Unemployment Rate (percentage) in Malaysia after first differencing and logarithmic transformation

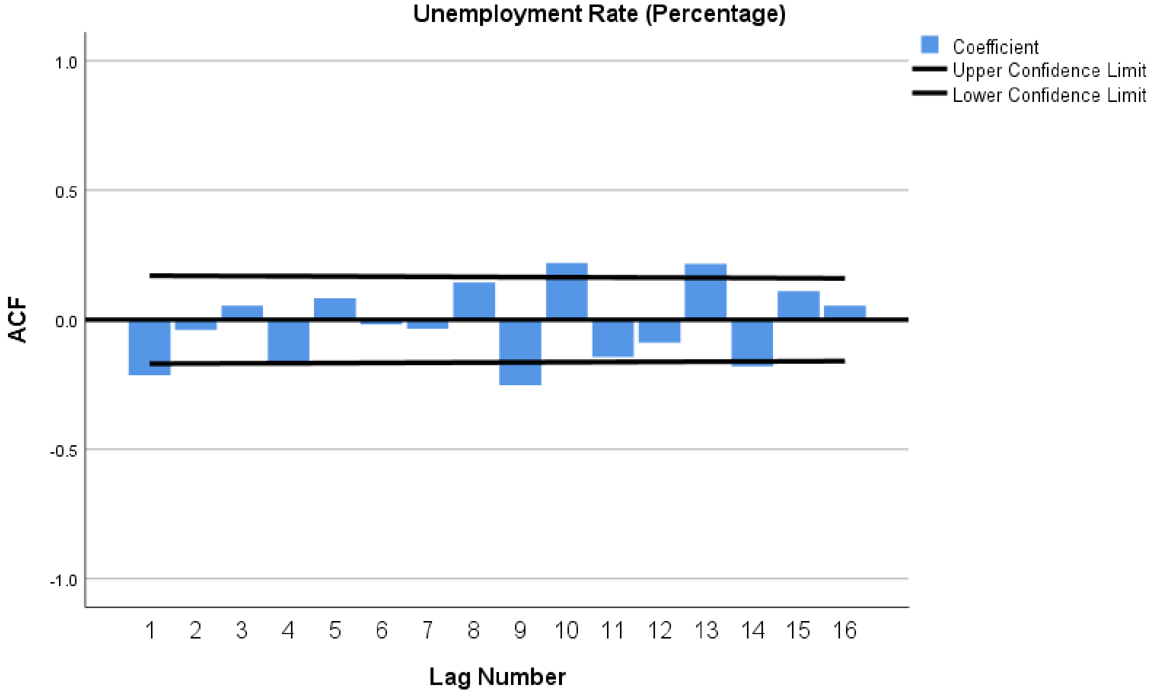

Figure 4.8: ACF correlogram of Unemployment Rate (percentage) in Malaysia after first differencing and logarithmic transformation

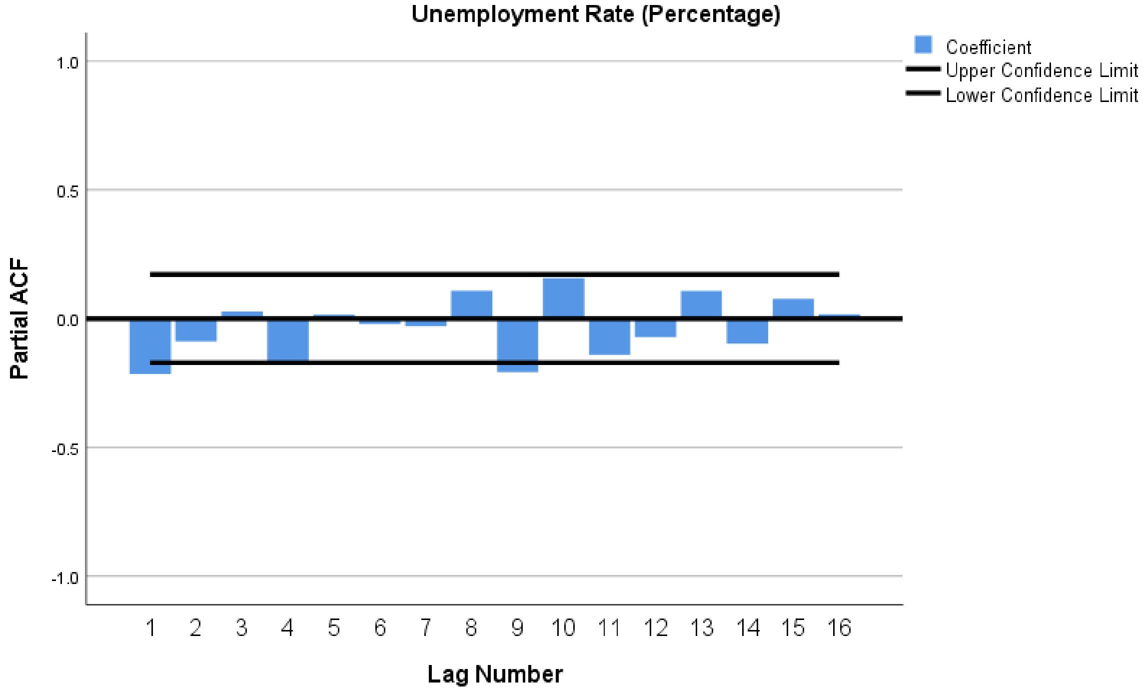

Figure 4.9: ACF correlogram of Unemployment Rate (percentage) in Malaysia after first differencing and logarithmic transformation

Table 4.8 and 4.9 below show the ACF and PACF plot (Correlogram) for the data.

Autocorrelations

Series: Unemployment Rate (Percentage)

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	-.215	.085	6.423	1	.011
2	-.039	.084	6.635	2	.036
3	.054	.084	7.043	3	.071
4	-.170	.084	11.168	4	.025
5	.083	.084	12.157	5	.033
6	-.019	.083	12.207	6	.058
7	-.035	.083	12.384	7	.089
8	.144	.083	15.429	8	.051
9	-.254	.082	24.984	9	.003
10	.219	.082	32.159	10	.000
11	-.144	.082	35.290	11	.000
12	-.089	.081	36.484	12	.000
13	.216	.081	43.592	13	.000
14	-.180	.081	48.566	14	.000
15	.110	.080	50.443	15	.000
16	.054	.080	50.897	16	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Table 4.8: Table of ACF of Unemployment Rate (percentage) in Malaysia after first differencing and logarithmic transformation

Partial Autocorrelations

Series: Unemployment Rate (Percentage)

Lag	Partial Autocorrelation	Std. Error
1	-.215	.086
2	-.089	.086
3	.027	.086
4	-.164	.086
5	.015	.086
6	-.021	.086
7	-.030	.086
8	.109	.086
9	-.207	.086
10	.158	.086
11	-.141	.086
12	-.072	.086
13	.108	.086
14	-.098	.086
15	.077	.086
16	.017	.086

Table 4.9: Table of PACF of Unemployment Rate (percentage) in Malaysia after first differencing and logarithmic transformation

4.2 Model Estimation

We can use the SPSS Expert Modeler to help us to find a suitable ARIMA model of the unemployment rate (percentage) in Malaysia. The Table 4.10 below shows that the suitable model is ARIMA (0,1,1) and does not have seasonal estimation of the series because there is no seasonal variation.

Time Series Modeler

Model Description

			Model Type
Model ID	Unemployment Rate (Percentage)	Model_1	ARIMA(0,1,1)(0,0,0)

Table 4.10: ARIMA Model that generated by SPSS Expert Modeler

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.049	.	.049	.049	.049	.049	.049	.049	.049	.049	.049
R-squared	.868	.	.868	.868	.868	.868	.868	.868	.868	.868	.868
RMSE	.202	.	.202	.202	.202	.202	.202	.202	.202	.202	.202
MAPE	3.696	.	3.696	3.696	3.696	3.696	3.696	3.696	3.696	3.696	3.696
MaxAPE	25.130	.	25.130	25.130	25.130	25.130	25.130	25.130	25.130	25.130	25.130
MAE	.126	.	.126	.126	.126	.126	.126	.126	.126	.126	.126
MaxAE	1.256	.	1.256	1.256	1.256	1.256	1.256	1.256	1.256	1.256	1.256
Normalized BIC	-3.164	.	-3.164	-3.164	-3.164	-3.164	-3.164	-3.164	-3.164	-3.164	-3.164

Table 4.11: Model Fit of the ARIMA(0,1,1) Model generated by SPSS Expert Modeler

The next step is to estimate the parameters of the model. According to the ARIMA model parameters in Table 4.12, we can know that this ARIMA model is with the natural logarithm to perform the forecasting. This ARIMA model consists only of first differencing and MA Lag 1, which d =1 and q =1. In the ARIMA model parameters table, the estimated value of the first coefficient value of the MA model is 0.246. Hence, the model can written in the following form:

$$\hat{Z}_t = 0.246\hat{\varepsilon}_{t-1}$$

The MA coefficient value 0.246 is with the standard error value of 0.084 and the t-value is 2.937. The t-value means the estimated value of a parameter divided by the standard error. The Sig. represents the significance level of the parameter estimation. If the value is above 0.05 are considered to have no significant statistical significance. In this model, we have the significance level of 0.04, so this model parameter is considered significant.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Unemployment Rate (Percentage) -Model_1	Unemployment Rate (Percentage)	Natural Logarithm	Difference		1			
			MA	Lag 1	.246	.084	2.937	.004

Table 4.12: ARIMA Model Parameters of ARIMA(0,1,1) Model

Figure 4.9 reports the residuals of the ARIMA(0,1,1) model for ACF and PACF, respectively. Viewed from the bottom up, neither number shows a correlation pattern between residuals, nor any correlation beyond the vertical 95% confidence interval contained in the graph. Therefore, the ARIMA(0,1,1) model appropriately simulates the dynamics of this time series.

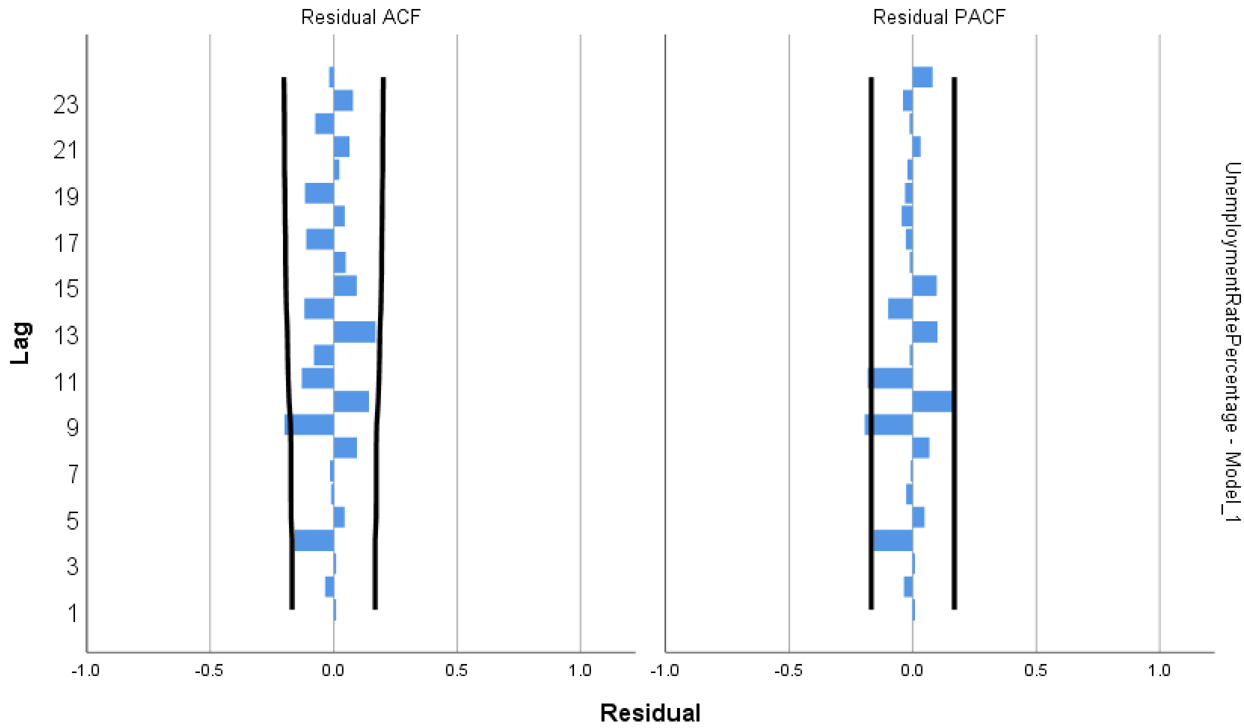

Figure 4.10: The ACFs and PACFs of the Residual

The Table 4.13 below shows the forecasting table of the model. The forecasting value of the unemployment rate of Malaysia in September 2021 is 4.7% and the value of 5.2 and 4.2 for the Upper Control Limit (UCL) and the Lower Control Limit (LCL) respectively. The actual value of unemployment rate in Malaysia for August 2021 is 4.6%, the forecasted value is still in the range of acceptance.

Forecast

Model		Aug 2021
Unemployment Rate (Percentage) -Model_1	Forecast	4.8
	UCL	5.4
	LCL	4.3

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Table 4.13: Forecast of Unemployment Rate (Percentage) in Malaysia generated by SPSS Expert Modeler in September 2021

From the Figure 4.10 below, we can see that the fitting forecasting is carried out and the acceptance range also shown in this graph. It can be seen that the fitting effect is considered good.

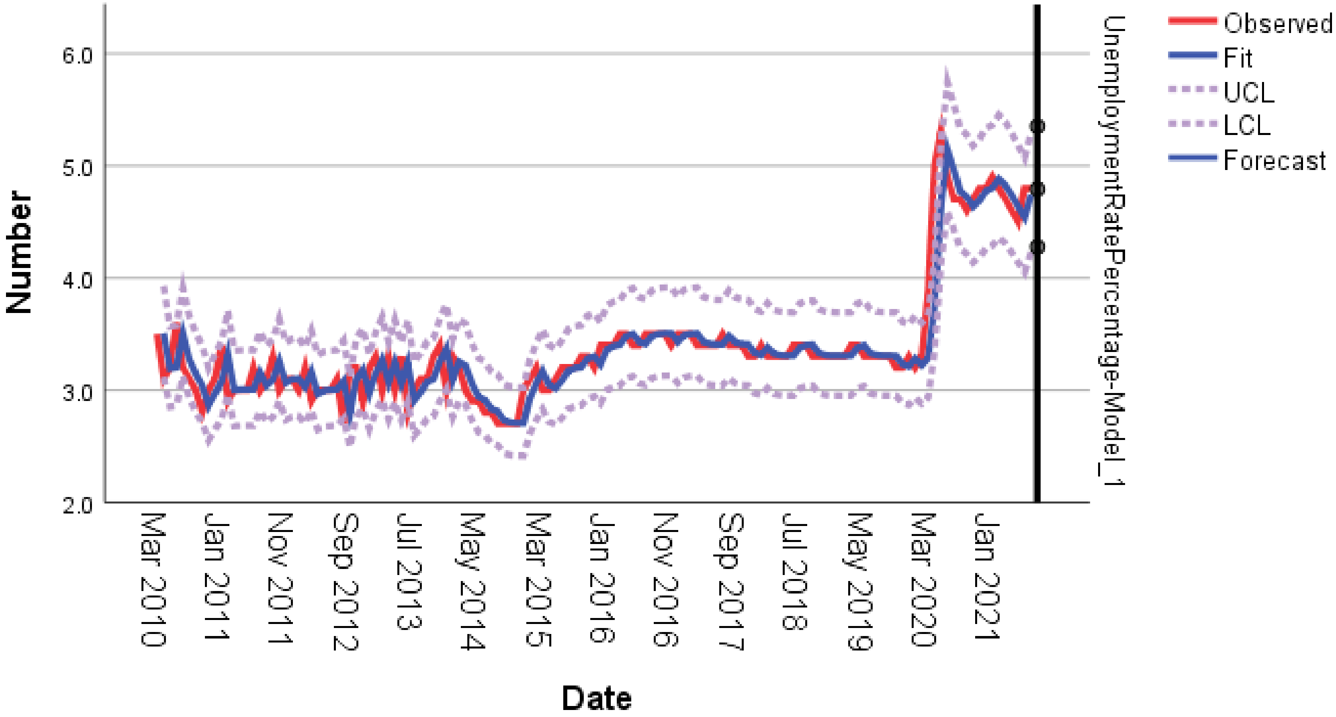

Figure 4.11: Plot Graph of Unemployment Rate (percentage) in Malaysia with observed, fitted value, UCL and LCL

4.3 Diagnostic Checking

From Table 4.14 below, we can read the values of Stationary R-squared value, R-squared, Root-Mean-Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Maximum Absolute Percentage Error (MaxAPE), Maximum Absolute Error (MaxAE) and Normalized Bayesian Information Criterion (BIC) from the model fit statistics. We can see that the R-squared value is 0.868, which means that 86.8% of the model fits the data and the normalized BIC is -3.164, the smaller the value of normalized BIC, the better the model.

The Sig. column gives the significance value of Ljung-box statistics, which is a random test for residual errors in the model. Indicates whether the specified model is reliable. A significance value less than 0.05 means that the residual error is not random, and a higher value means that the data is more likely to be random. By reading Ljung-box significance value 0.043, less than 0.05, we reject the null hypothesis and conclude that residual is not white noise.

Model Statistics

		Model Fit statistics								Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
Unemployment Rate (Percentage) -Model_1	0	.049	.868	.202	3.696	.126	25.130	1.256	-3.164	28.126	17	.043	0

	Model Fit statistics							
Number of Predictors	Stationary R-squared	R-squared	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC
0	.049	.868	.202	3.696	.126	25.130	1.256	-3.164

Ljung-Box Q(18)			
Statistics	DF	Sig.	Number of Outliers
28.126	17	.043	0

Table 4.14: Model Statistic of ARIMA(0,1,1) Model generated by SPSS Expert Modeler

ARIMA (*p,d,q*) Model	Logarithmic transformation	R^2	RMSE	MAPE	Normalized BIC	Forecast value (%)
ARIMA (*0,1,1*)	Yes	0.868	0.202	3.696	-3.164	4.8
ARIMA (*0,1,1*)	No	0.873	0.199	3.873	-3.116	4.8
ARIMA (*1,1,1*)	Yes	0.871	0.202	3.871	-3.054	4.9
ARIMA (*1,1,1*)	No	0.877	0.197	3.930	-3.101	4.9
ARIMA (*0,1,2*)	Yes	0.871	0.202	3.869	-3.052	4.9
ARIMA (*0,1,2*)	No	0.875	0.199	3.868	-3.082	4.8
ARIMA (*1,1,2*)	Yes	0.868	0.205	3.921	-2.990	4.9
ARIMA (*1,1,2*)	No	0.875	0.200	3.870	-3.040	4.8

Table 4.15: ARIMA (p,d,q) model generated by SPSS custom ARIMA model

There are some of the custom ARIMA models generated to do some comparison. The R-squared value, RMSE, MAPE, Normalized BIC and forecasted value of those custom ARIMA models are shown as above Table 4.15. The highest value of R-squared is 0.877 which is ARIMA (*1,1,1*) model without logarithmic transformation and the lowest value of Normalized BIC is -3.164 which is ARIMA (*0,1,1*) with logarithmic transformation. All the forecasted values are between 4.8 and 4.9. The values of RMSE are in the range from 0.197 to 0.205 and the values of MAPE are in the range of 3.696 to 3.930.

In a nutshell, the higher R-squared values the model is more fitted, the smaller the values of RMSE, MAPE and normalized BIC the better the forecasting model. The ARIMA(*0,1,1*) model with logarithmic transformation has the lowest MAPE and the lowest normalized BIC. Therefore, the ARIMA (*0,1,1*) with logarithmic transformation provided by SPSS Expert Modeler is sufficient enough.

CHAPTER 5 Conclusion

5.1 Conclusion

According to the above time series analysis, we can obtain several results from the above analysis. First, we can see from the sequence chart that the unemployment rate of Malaysia fluctuated between 2.7% and 3.6% from March 2010 to April 2020, and the fluctuation range was not large. But in May 2020, Malaysia's unemployment rate peaked at 5.3%. In the years that followed, Malaysia's unemployment rate fluctuated between 4.5% and 5.0%.

After time series analysis, we can know that Malaysia's unemployment rate is an unstable time series. So we performed the difference method to convert this sequence into a stationary sequence, and then continued the following test. We can also judge whether a time series is stable by observing the ACF and PACF of the time series.

Next, SPSS was used to find the ARIMA model that best fits this time series. The ARIMA model given by SPSS is ARIMA(0,1,1) model and we can see that the R-squared value in the model fit statistics is 0.872 which means the fitting effect of ARIMA(0,1,1) model is 87.2% fit. In the model fit table, we can get many reference values such as stationary R-squared, R-squared, RMSE, MAPE, MaxAPE, MAE,MaxAE and normalized BIC. After that, we can know the model parameters from the ARIMA model parameters table. The first coefficient of the MA model is 0.248. Finally is the forecasted value of the unemployment rate of Malaysia. The ARIMA model can only do a short-term forecasting for the value of unemployment rate in Malaysia, so only the value of August 2021 has been forecasted which is 4.8% and the UCL and LCL values are 5.4% and 4.3% respectively.

5.2 Limitations

There are a few limitations when this research has been done. One of the limitations is the accuracy of the data set. The data set of the unemployment rate in Malaysia is only in 1 decimal place. Thus, it will affect the accuracy of the analysis and forecasting result. Next is that this research considers the time as the only factor that affects the unemployment rate in Malaysia. It did not consider other factors such as the gross domestic product (GDP) of Malaysia, the impact of Covid-19 pandemic. Time series analysis also has the problem of single generalization, which makes it difficult to obtain appropriate measures and accurately identify the correct model to represent data.

5.3 Suggestions for future research

Time series analysis and forecasting method is to predict the future development according to the change trend of the market in the past. Its premise is to assume that the past of things will continue in the future. The reality of things is the result of historical development, and the future of things is an extension of reality, the past and future of things are connected. The time series analysis method of market prediction is precisely based on the continuous regularity of the development of objective things, and further predicts the future development trend of the market by using the past historical data and statistical analysis. In market forecasting, the past also continues into the future, meaning that the market will not change in the future by sudden leaps, but by gradual changes.

It should be pointed out that the development of things is not only characterized by continuity, but also by complexity and diversity. Therefore, attention should be paid to the future development and change law and development level of the phenomenon when using time series analysis method for forecasting, which may not be completely consistent with its history and current development and change law. With the development of market phenomena, there will also appear some new characteristics. Therefore, in time series analysis and forecasting, it is impossible to extend outward mechanically according to the law of phenomena in the past and present. It is necessary to study and analyze the new characteristics and manifestations of phenomenon change, and take these new characteristics and manifestations into full consideration in the predicted value. Only in this way can we make reliable prediction results of market phenomena which not only continue their historical change rules but also conform to their realistic performance.

The time series analysis and forecasting method highlights the role of time factor in the prediction and temporarily ignores the influence of external specific factors. Time series is at the core of time series analysis and forecasting. Without time series, there would be no such method. Although, the development and change of the forecast object is affected by many factors. However, when time series analysis is used for quantitative prediction, all the influencing factors are actually attributed to time, and only the comprehensive effect of all the influencing factors is recognized, which will still play a role on the predicted object in the future. The causal relationship between the predicted object and the influencing factors is not analyzed and discussed. Therefore, in order to reflect changes in the future market of accurate prediction, the use of time series analysis method to forecast, quantity analysis method and qualitative analysis method must be combined, fully from the aspects of qualitative research, the relationship between various factors and market research analyzed the various factors influencing the market changes on the basis of predicted value is determined.

It should be pointed out that the time series forecasting method has the defect of prediction error because it emphasizes that the time series does not consider the influence of external factors temporarily. When the external environment changes greatly, there will often be a large deviation. The time series prediction method has a better effect on the short and medium term prediction than the long-term prediction. Because objective things, especially economic phenomena, are more likely to change in a longer period of time, they are bound to have a significant impact on market economic phenomena. In this case, when making predictions, only considering the time factor without considering the influence of external factors on the predicted objects, the prediction results will be seriously inconsistent with the actual situation.

REFERENCE/BIBLIOGRAPHY

[1] Investopedia. 2021. *Unemployment Definition and Types*. [online] Available at: <https://www.investopedia.com/terms/u/unemployment.asp > [Accessed 25 October 2021].

[2] Rba.gov.au. 2021. [online] Available at: <https://www.rba.gov.au/education/resources/explainers/pdf/unemployment-its-measurement-and-types.pdf > [Accessed 25 October 2021].

[3] The Star. 2021. *July unemployment rate remains at 4.8%*. [online] Available at: <https://www.thestar.com.my/business/business-news/2021/09/10/july-unemployment-rate-remains-at-48 > [Accessed 25 October 2021].

[4] [4] Peter J. Brockwell, Richard A. Davis. 2021. *Introduction to Time Series and Forecasting, Second Edition.* [online] Available at: <http://home.iitj.ac.in/~parmod/document/introduction%20time%20series.pdf> [Accessed 25 October 2021].

[5] Louis Sevitenyi Nkwatoh, 2012. *Forecasting Unemployment Rates in Nigeria Using Univariate Time Series Models.* Viewed on 27th March 2021. Available at: <https://www.ijbcnet.com/1-12/IJBC-12-11202.pdf>

[6] Abouellil, Embareka, 2011. *The impact of Labor Market Trends on Unemployment rates in Egypt using time series analysis model.* Viewed on 27th March 2021. Available at: <http://2011.isiproceedings.org/papers/950728.pdf>

[7] Xinkai Huang, 2015. *Forecasting the US Unemployment Rate with Job Opening Index.* Viewed on 27th March 2021. Available at: <https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1716&context=theses>

[8] M.O. Adenomon, 2017. *Modelling and Forecasting Unemployment Rates in Nigeria Using ARIMA Model.* Viewed on 27th March 2021. Available at: <https://www.researchgate.net/publication/318280014_MODELLING_AND_FORECASTING_UNEMPLOYMENT_RATES_IN_NIGERIA_USING_ARIMA_MODEL>

[9] Sune Karlsson, Farrukh Javed, 2016. *Modeling and Forecasting Unemployment Rate In Sweden using various Econometric Measures.* Viewed on 27th March 2021. Available at: <https://www.diva-portal.org/smash/get/diva2:949512/FULLTEXT01.pdf>

[10] Ramli, S. F., Fidaus, M., Uzair, H., Khairi, M. & Zharif, A, 2018. *Prediction of the Unemployment Rate in Malaysia.* Viewed on 27th March 2021. Available at: <http://www.ijmtss.com/PDF/IJMTSS-2018-04-12-04.pdf >

[11] Maria E. N., Sabariah S., Lok S. L., Rohayu M. S., Norhaidah M. A., 2018. *Forecasting of Unemployment Rate in Malaysia Using Exponential Smoothing Methods*. Viewed on 27th March 2021. Available at: < https://www.researchgate.net/publication/332679570_Forecasting_of_Unemployment_Rate_in_Malaysia_Using_Exponential_Smoothing_Methods >

[12] Mohd Nadzri M. N., Kon Mee Hwa, Huzaifah M., 2011. *An Initial Study on the Forecast Model for Unemployment Rate*. Viewed on 27th March 2021. Available at: < http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.630.9195&rep=rep1&type=pdf >

[13] Phi-Hung Nguyen, Jung-Fa Tsai, Ihsan Erdem Kayral, Ming-Hua Lin, 2021. *Unemployment Rates Forecasting with Grey-Based Models in the Post-COVID-19 Period: A Case Study from Vietnam*. Viewed on 27th March 2021. Available at: < https://www.mdpi.com/2071-1050/13/14/7879/pdf >

APPENDICES

1. Dataset of Monthly Unemployment Rate (percentage) in Malaysia from March 2010 to July 2021.

Unemployment Rate (percentage) in Malaysia

Month	Percentage
Mar-10	3.5
Apr-10	3.1
May-10	3.2
Jun-10	3.6
Jul-10	3.2
Aug-10	3.1
Sep-10	3.0
Oct-10	2.8
Nov-10	3.0
Dec-10	3.1
Jan-11	3.4
Feb-11	2.9
Mar-11	3.0
Apr-11	3.0
May-11	3.0
Jun-11	3.2
Jul-11	3.0
Aug-11	3.1
Sep-11	3.3
Oct-11	3.0
Nov-11	3.1
Dec-11	3.1
Jan-12	3.0
Feb-12	3.2
Mar-12	2.9
Apr-12	3.0
May-12	3.0
Jun-12	3.0
Jul-12	3.1
Aug-12	2.7
Sep-12	3.2
Oct-12	3.2
Nov-12	2.9
Dec-12	3.2
Jan-13	3.3
Feb-13	3.0
Mar-13	3.3
Apr-13	3.0
May-13	3.3
Jun-13	2.8
Jul-13	3.0
Aug-13	3.1
Sep-13	3.1
Oct-13	3.3
Nov-13	3.4
Dec-13	3.0
Jan-14	3.3
Feb-14	3.2
Mar-14	3.0
Apr-14	2.9
May-14	2.9
Jun-14	2.8

Jul-14	2.8
Aug-14	2.7
Sep-14	2.7
Oct-14	2.7
Nov-14	2.7
Dec-14	3.0
Jan-15	3.1
Feb-15	3.2
Mar-15	3.0
Apr-15	3.0
May-15	3.1
Jun-15	3.2
Jul-15	3.2
Aug-15	3.2
Sep-15	3.3
Oct-15	3.3
Nov-15	3.2
Dec-15	3.4
Jan-16	3.4
Feb-16	3.4
Mar-16	3.5
Apr-16	3.5
May-16	3.4
Jun-16	3.4
Jul-16	3.5
Aug-16	3.5
Sep-16	3.5

Oct-16	3.5
Nov-16	3.4
Dec-16	3.5
Jan-17	3.5
Feb-17	3.5
Mar-17	3.4
Apr-17	3.4
May-17	3.4
Jun-17	3.4
Jul-17	3.5
Aug-17	3.4
Sep-17	3.4
Oct-17	3.4
Nov-17	3.3
Dec-17	3.3
Jan-18	3.4
Feb-18	3.3
Mar-18	3.3
Apr-18	3.3
May-18	3.3
Jun-18	3.4
Jul-18	3.4
Aug-18	3.4
Sep-18	3.3
Oct-18	3.3
Nov-18	3.3
Dec-18	3.3

Jan-19	3.3
Feb-19	3.3
Mar-19	3.4
Apr-19	3.4
May-19	3.3
Jun-19	3.3
Jul-19	3.3
Aug-19	3.3
Sep-19	3.3
Oct-19	3.2
Nov-19	3.2
Dec-19	3.3
Jan-20	3.2
Feb-20	3.3
Mar-20	3.9
Apr-20	5.0
May-20	5.3
Jun-20	4.9
Jul-20	4.7
Aug-20	4.7
Sep-20	4.6
Oct-20	4.7
Nov-20	4.8
Dec-20	4.8
Jan-21	4.9
Feb-21	4.8
Mar-21	4.7
Apr-21	4.6
May-21	4.5
Jun-21	4.8
Jul-21	4.8

2. Actual Value of Unemployment Rate (percentage) in Malaysia in August 2021.

Month	Percentage
Aug-21	4.6