Tourism Industry and Money Supply,

is it a pseudo relation?

詹永裕：H24061066

張孟涵：H24064064

林佳霈：H24061040

Outline

Abstract
I Introduction
II Data Sets Information
1.M2 Money Supply in Taiwan
2. Outbound Passengers of Taiwanese (persons)
3. Inbound Passengers in Taiwan (persons)
III Algorithm flow
IV Establishment of Model
- Time Series Analysis of M2
  1. Remove Trend
  2. Model Specification
  3. Parameter Estimation
  4. Model Diagnostics
- Time Series Analysis of Outbound Passenger
  1. Remove Trend
  2. Model Specification
  3. Parameter Estimation
  4. Model Diagnostics
- Time Series Analysis of Inbound Passengers
  1. Remove Trend
  2. Model Specification
  3. Parameter Estimation
  4. Model Diagnostics

V Correlation
1. Original data correlation
2. Residuals correlation

VI Conclusion
VII Discussion and the Future Work
Reference

Abstract

In many countries, tourism plays a significant role. There are several studies focus on the relations between the tourism industry and economic factors. One previous work uses simple linear regression to formula the correlation between the number of outbound passengers and the money supply without considering time effect. However, due to the number of both outbound and inbound passengers and the money supply are all time-series data. The correlation we observed might be caused by the time effects simply. This is the so-called pseudo relation. To figure out this problem, we use SARIMA models to remove the time effects of our data and then fit simple linear regression again. We found that the regression coefficient between the money supply and the number of outbound passengers is not significant after removing the time effects. Furthermore, we use the cross-correlation function (CCF) to check the potential effects. From the cross-correlation function, we find that there are weak correlations between the money supply and the number of outbound passengers, the money supply and the number of inbound passengers still both in the long term and short term.

I Introduction

The tourism industry has been listed as one of the comprehensive industries in the modern economic structure, which not only shows a country's economy but also the development of the economic structure and the degree of modernization, and its development is also an important part of the economy. Usually, the number of both outbound and inbound passengers can be seen as the indicators that reflect the state of a country's economy. As a result, there have been various researches discussing the correlation between the number of outbound passengers and economic factors. Some previous works formula the relation between the number of outbound passengers and economic factors by simple linear regression; in this way, they ignore time effects [1]. As mentioned in this paper, among all economic factors, the correlation between the money supply and the number of outbound passengers is the most significant.

However, the number of outbound passengers and the money supply are all time-series data, and both of them are linear combinations of present and historical white noise. If we run the simple linear regression directly, the correlation we observed may be caused by the time effects simply. This situation is the so-called **pseudo relation**. Therefore, we want to use the time-series model to remove the time effects of our data and discuss the true relations between the M2 money supply and both the number of outbound and inbound passengers respectively. Furthermore, since the potential long-term effects may still be ignored when applying this method, we will use CCF to check the potential long-term effects of our data sets furthermore.

This project is structured as follows: Section II gives data sets information including Outbound Passengers of Taiwan (persons), Inbound Passengers in Taiwan (persons), and M2 Money Supply in Taiwan. In section III, we visualize the algorithm flow of our analysis, then establishing models in section IV. Section V gives a more reasonable evaluation of the models' performance. Finally, Section VI concludes what we found in this project.

II Data Sets Information

M2 Money Supply

Definition: The **money supply** is the total value of money available in an economy at a point in time.[2] There are several ways to measure money supply and M2 is a measurement of the money supply that has been used widely. This includes all savings, money market, etc. accounts, and also certain short-term investments.[3]

1.M2 Money Supply in Taiwan

Source: 財經 M 平方[4]
Period: from January 2000 through October 2019
Unit of time: monthly data
Observation: 238 observations

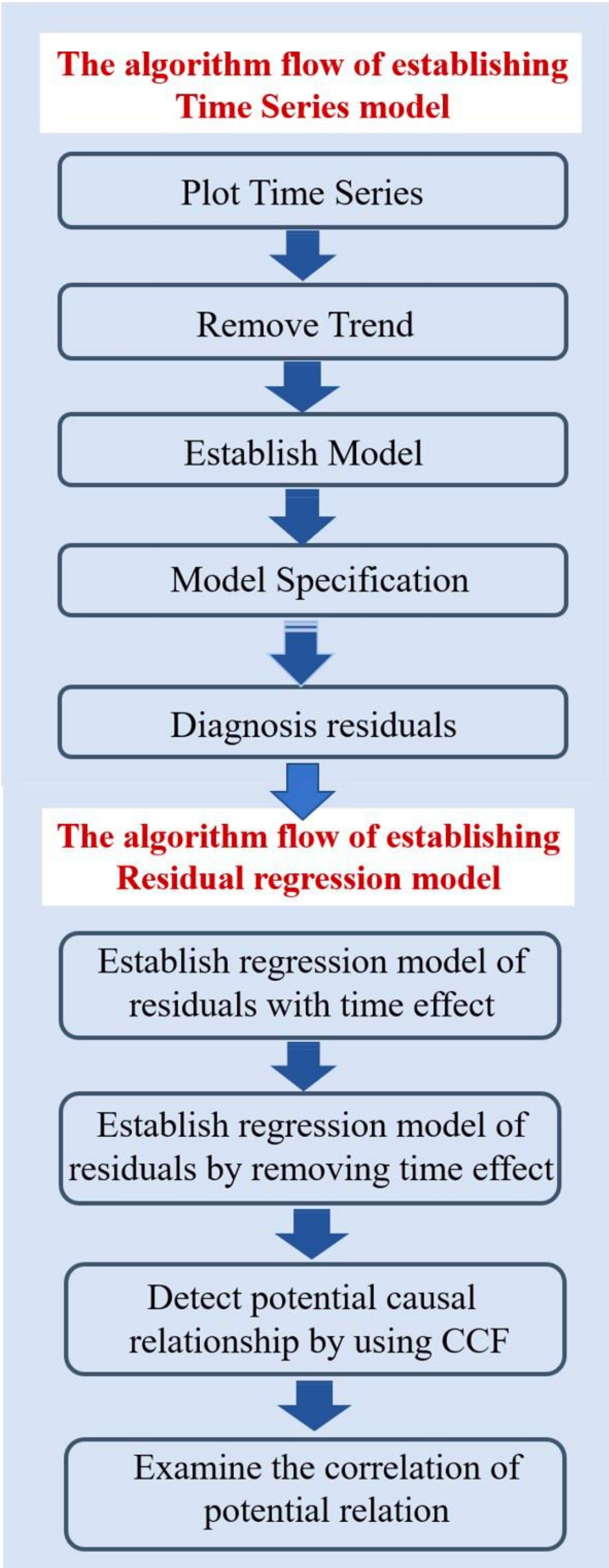
2. Outbound Passengers of Taiwanese (persons)

Source: Ministry of Transportation and Communications (Taiwan) [5]
Period: from January 2000 through October 2019
Unit of time: monthly data
Observation: 238 observations

3. Inbound Passengers in Taiwan (persons)

Source: Ministry of Transportation and Communications (Taiwan) [5]
Period: from January 2000 through October 2019
Unit of time: monthly data
Observation: 238 observations

III Algorithm flow

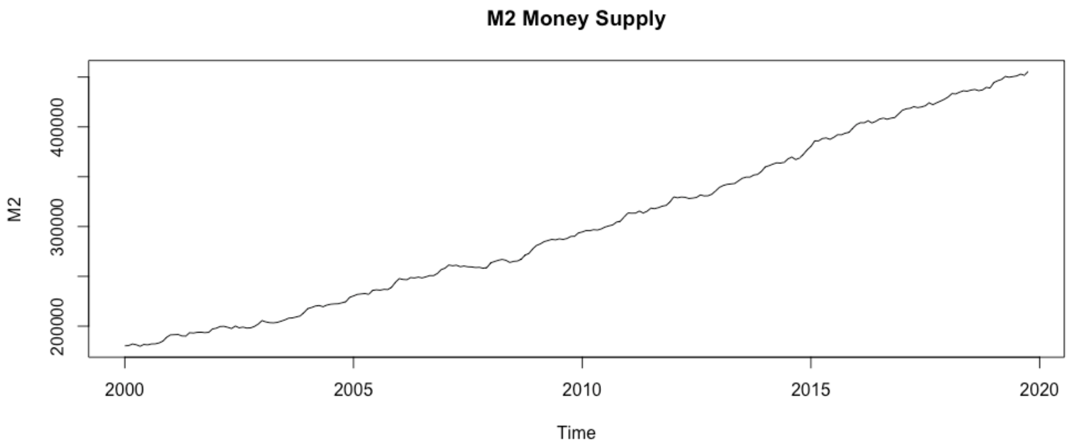


IV Establishment of Model

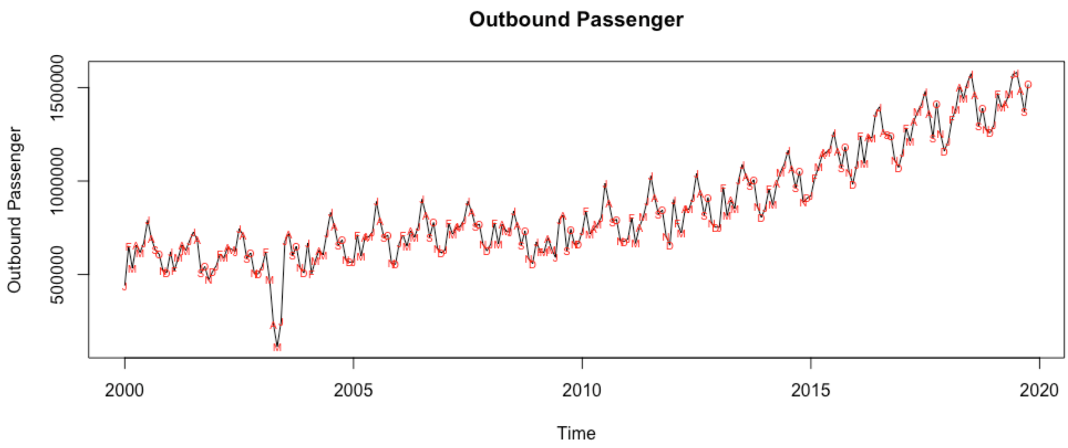
First, we plot the time-series plot of each data set.

Time Series Plot of Original data sets:

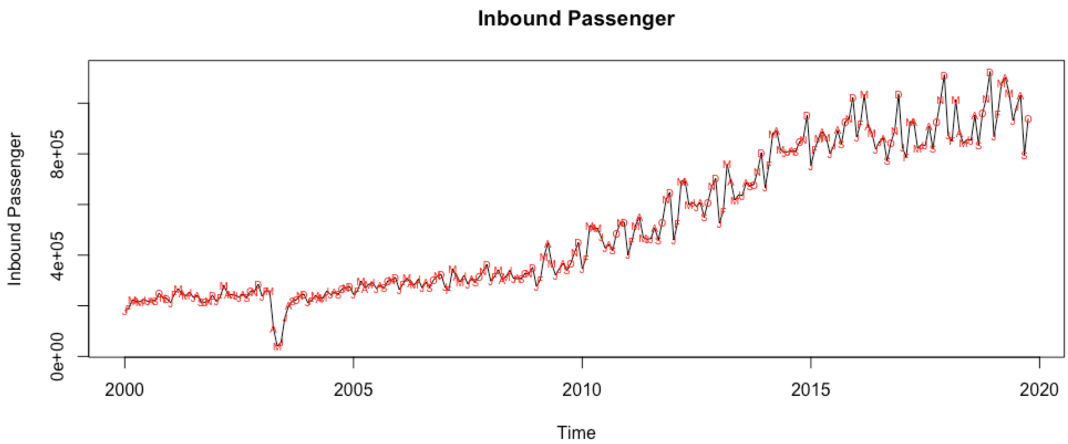
From these three charts, we can observe there exist clear upward trends. As time goes by, the observed values of our data are increasing. And it’s super intuitive that if we just fit a linear regression directly, we will get the outcome indicates that there are strong correlations between our data sets.



We can observe a clear upward trend in this plot. Based on this plot, we guess this time-series data is nonstationary.



There are a seasonal trend and a clear upward trend since 2010 in this plot. Based on this plot, we guess we need to fit seasonal trends and this time-series data is nonstationary. Moreover, we can see a sharp decline in the number of outbound passengers in March 2003 and several months thereafter was triggered by the outbreak of SARS in Taiwan in 2003.



Like the time-series plot of the number of outbound passengers, we observe a seasonal trend and a clear upper ward trend since 2010 in this plot. Based on this plot, we guess we need to fit seasonal trends and this time-series data is nonstationary. Moreover, we can see a sharp decline in March 2003 and several months thereafter was triggered by the outbreak of SARS in Taiwan in 2003.

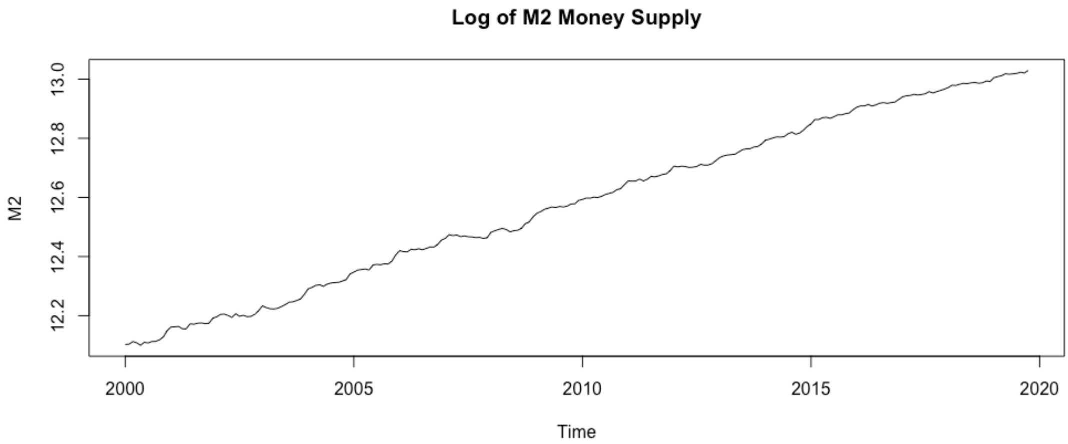
Then, we start to fit time-series to each data set.

Time Series Analysis of M2

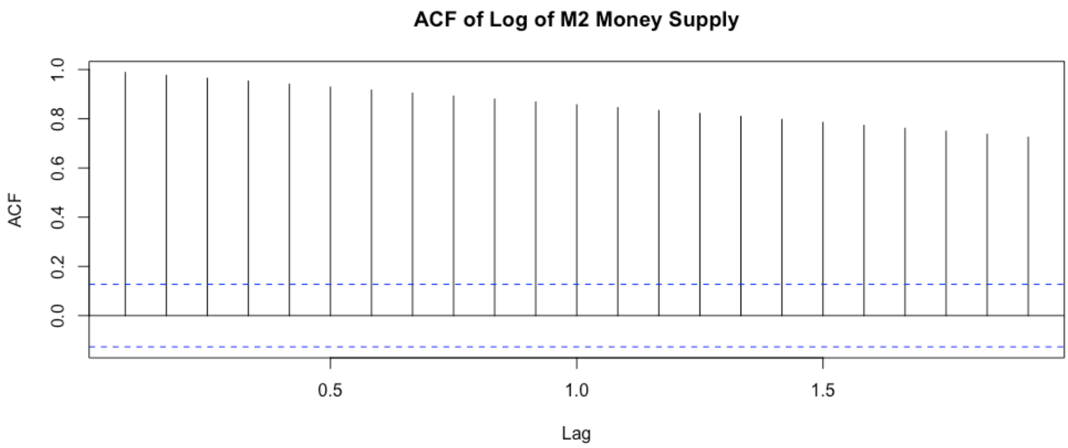
1. Remove Trend

Log

First, we notice the value of M2 Money Supply is too large. So, we did a log transformation.



Furthermore, we plot ACF of the log of M2 Money Supply. This time-series data is nonstationary.



Test

Then, we performed the Dickey-Fuller Test and the KPSS test to check if this data is nonstationary and requires differencing or not.

H0: Nonstationary

Ha: Stationary

Dickey-Fuller Test	
Dickey-Fuller test statistic	p-value
-2.6222	0.3143

H0: Stationary

Ha: Nonstationary

KPSS Test	
KPSS test statistic	p-value
4.8601	<0.01

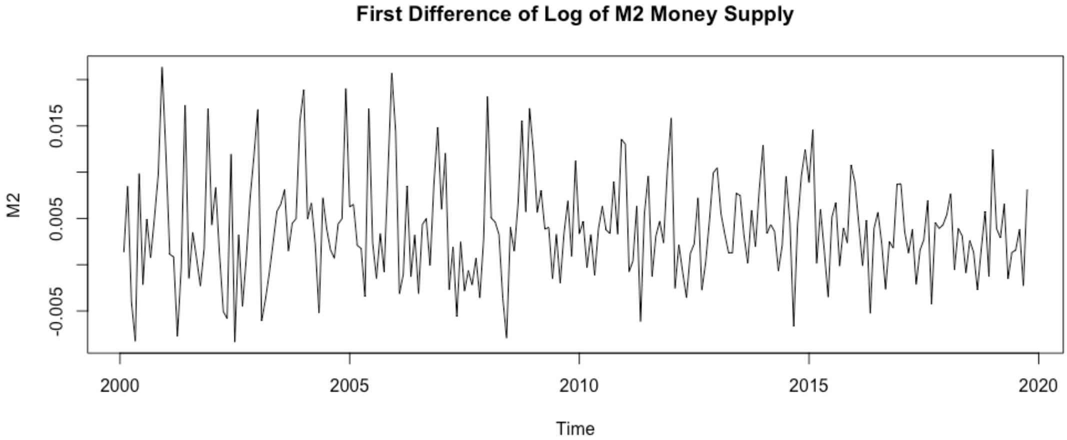
From the Dickey-Fuller test, we can find p-value = 0.3143 > 0.05, do not reject H0. So, we need to take a difference according to the Dickey-Fuller test.

From the KPSS test, we can find p-value < 0.05, reject H0. So, we need to take a difference according to the KPSS test.

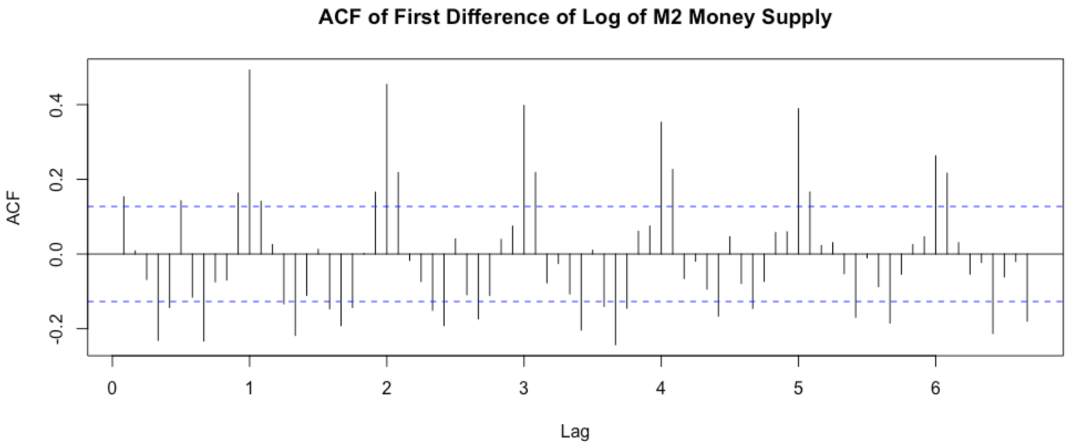
The conclusions of the Dickey-Fuller Test and the KPSS test are the same, therefore we choose to take the first difference next.

First Difference

The time-series data is more stationary after the first difference.

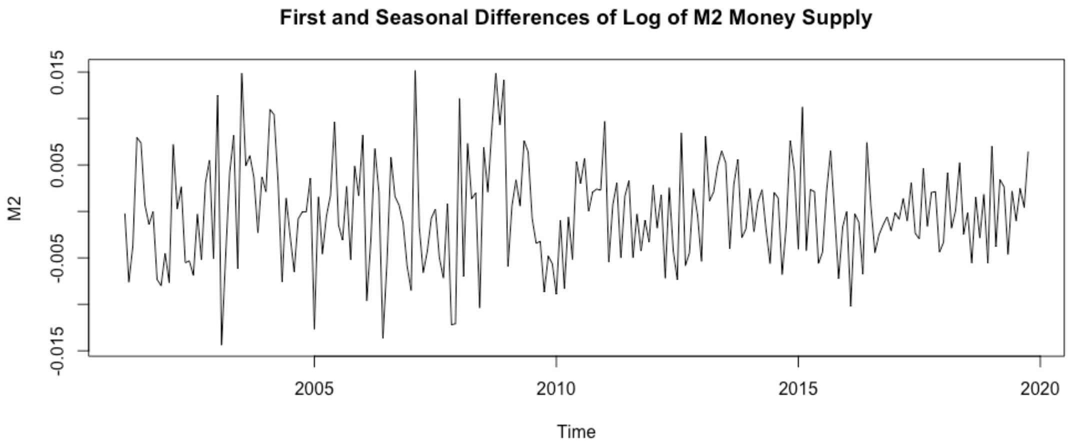


However, from the ACF plot, we observed that the correlation is still significant after 72 steps. It seems we need to take the seasonal difference again.



Seasonal Difference

Through the first difference and the seasonal difference, it seems that this time-series data is stationary already.



Test

After that, we still performed the Dickey-Fuller test and the KPSS test to check if this data is nonstationary and requires differencing or not.

H0: Nonstationary

Ha: Stationary

Dickey-Fuller Test	
Dickey-Fuller test statistic	p-value
-4.4221	<0.01

H0: Stationary

Ha: Nonstationary

KPSS Test	
KPSS test statistic	p-value
0.028562	>0.1

From the Dickey-Fuller test, we can find p-value < 0.05, reject H0. So, we don't need to take the second difference according to the Dickey-Fuller test.

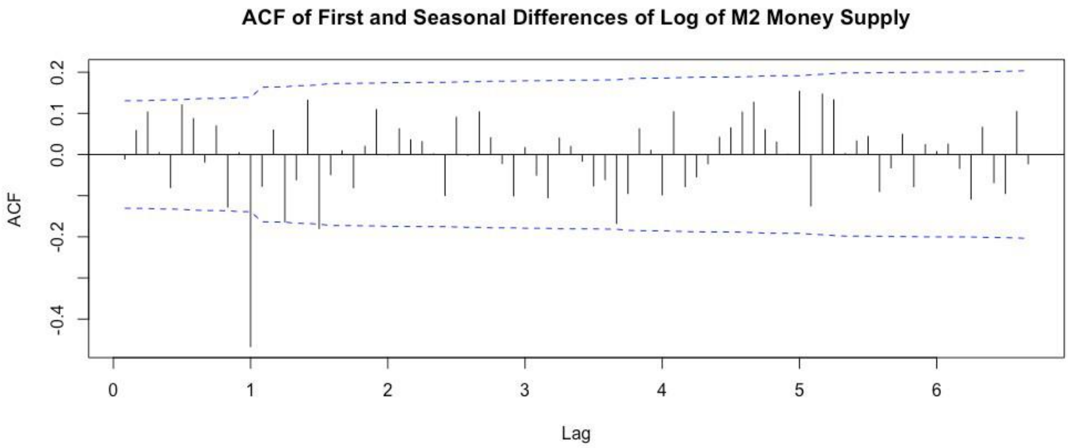
From the KPSS test, we can find p-value > 0.05, do not reject H0. So, we don't need to take the second difference according to the KPSS test.

The conclusions of the Dickey-Fuller test and the KPSS test are the same; therefore, we choose not to take the second difference.

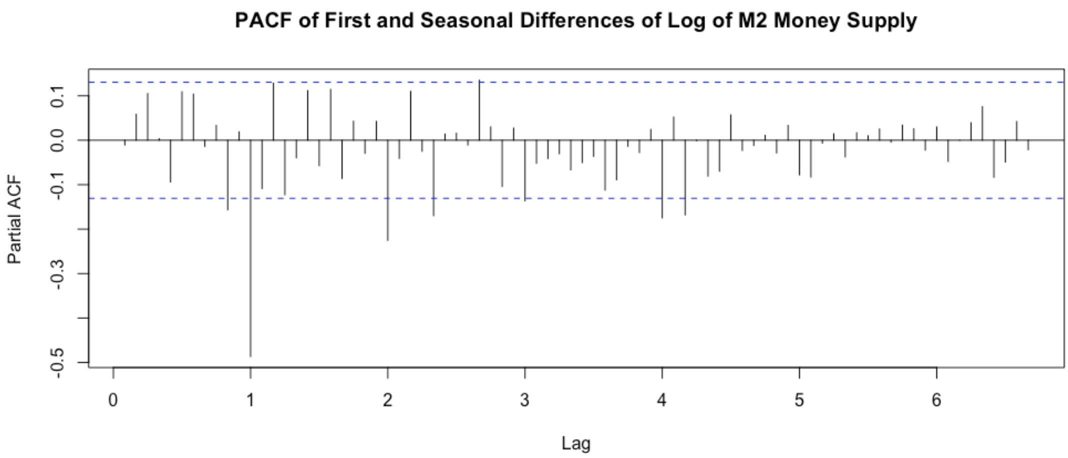
2. Model Specification

Next, we can fit a SARIMA model.
We plot the ACF & PACF charts to help us with model specifications.

From the ACF plot, we conjecture that model
$SARIMA(0,1,0) * (0,1,1)_{12}$can fit this data well.



From the PACF plot, we can observe there are several seasonal steps at lag 1,2,3,4 that out of the confidence interval. It seems that $SARIMA(0,1,0) * (4,1,0)_{12}$can fit this data well. However, fitting a model indicating that the current value will influence the future value after four years is strange in the data set like M2. Therefore, we decide not to implement this.



3. Parameter Estimation: $\boldsymbol{SARIMA(0,1,0)*(0,1,1)_{12}}$

Following, we use maximum likelihood to estimate the coefficients of our model. According to the table below, we find the coefficients are all significant.

M_t: M2 money supply

$(1-\mathrm{B})(1-\mathrm{B}^{12})log(M_t) = e_t - \Theta e_{t-12}$

$e_t \sim white\ noise(0, \sigma^2 = 0.00001932)$

	Θ
Estimate	0.7372
Standard error	0.0504

AIC= -1792.3

So, the model for the M2 money supply is:

$(1-\mathrm{B})(1-\mathrm{B}^{12})log(M_t) = e_t - 0.7372e_{t-12}$

$e_t \sim white\ noise(0, \sigma^2 = 0.00001932)$

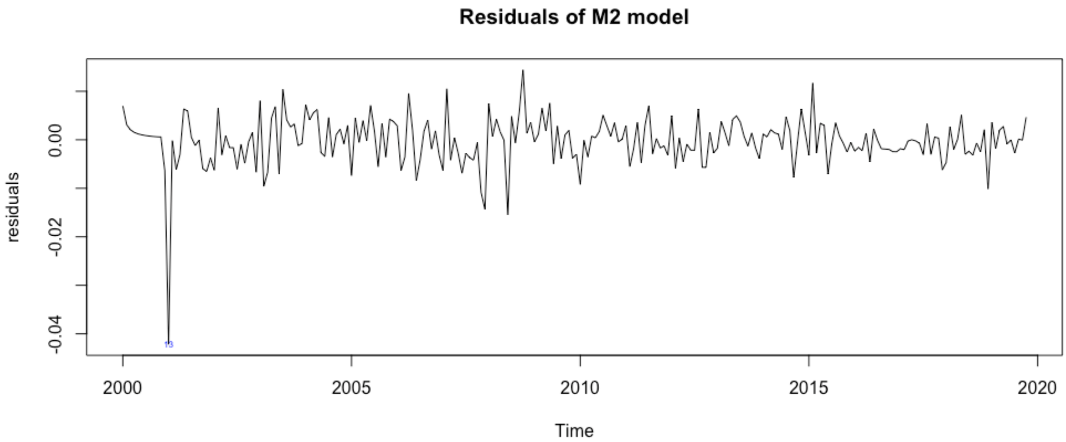
4. Model Diagnostics

After fitting the model, we started to check the model adequacy.

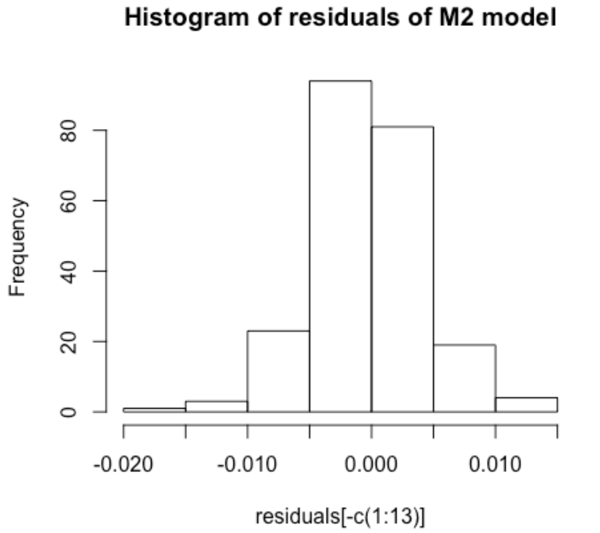
First, we plot the time-series plot of residuals.

From this plot, we observe an outlier that is significantly lower than average at the 13^{th} step. However, we had taken the first difference and the seasonal difference, the first 13 steps of the residuals should be unable to calculate. Thus, we believe this situation is just a numerical error while computing.

To avoid this numerical error affecting our model adequacy, we will remove the first 13 steps of the residuals and perform other tests later.

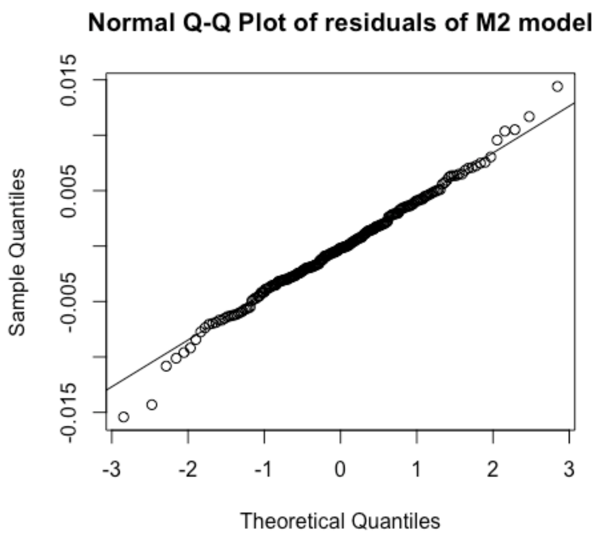


After that, we plot histogram to have an overview of the residuals and perform the t-test to check if the mean of the residuals is zero or not. According to the right table below, we can find the mean of the residuals is zero.



One Sample t-test	
T test statistic	p-value
-0.5954	0.5522

Next, we use the Shapiro-Wilk test and QQ plot to check if the residuals are normally distributed or not. According to the figure and table below, we can find the residuals are normally distributed.

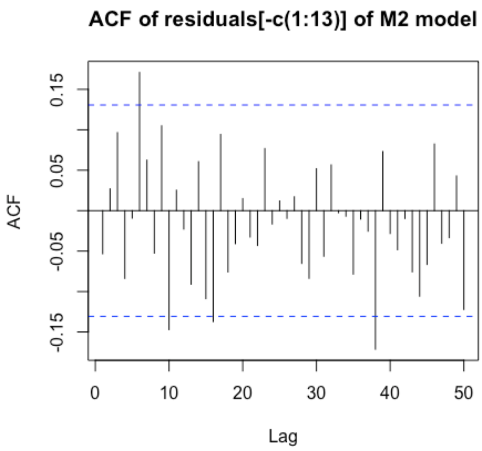


Shapiro-Wilk test	
Shapiro-Wilk test statistic	p-value
0.9921	0.2714

At last, we plot the ACF plot of the residuals, but unfortunately, there are some steps out of the confidence interval.

To solve this problem, we had tried to fit different models to our data. However, after fitting various models, this problem can't be solved still! Therefore, we decide to choose the original model, which is also the simplest model.

Although, according to the ACF plot of the first difference and the seasonal difference of M2 money supply, there are few steps out of the confidence interval after the 12th step. We believe it's counterintuitive, for the data like M2 money supply, to fit a time-series model indicating that the current value will influence future value after more than a year instead of recent months. Therefore, to the best of our knowledge, we did our best to fit the model!



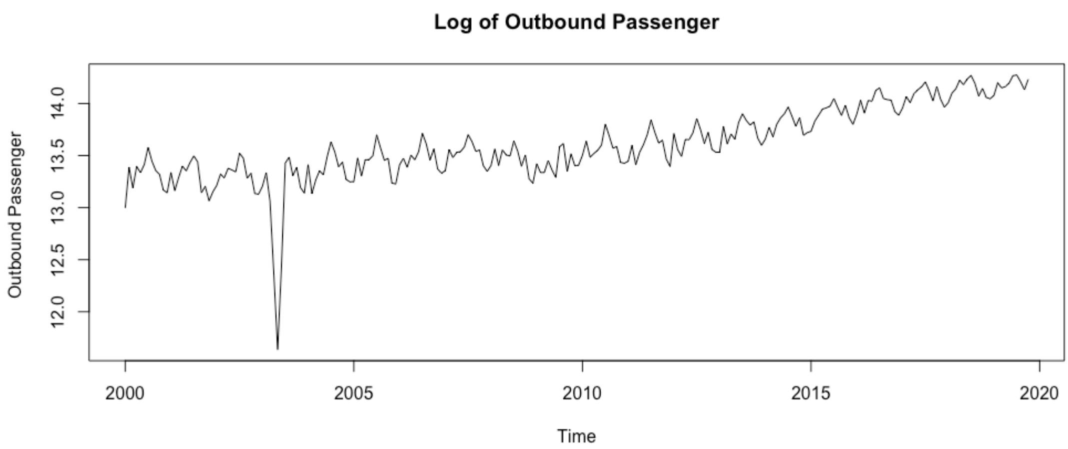
Ljung Box test	
Lag	p-value
lag 6	0.0764
lag 10	0.0231
lag 16	0.0121

Time Series Analysis of Outbound Passenger

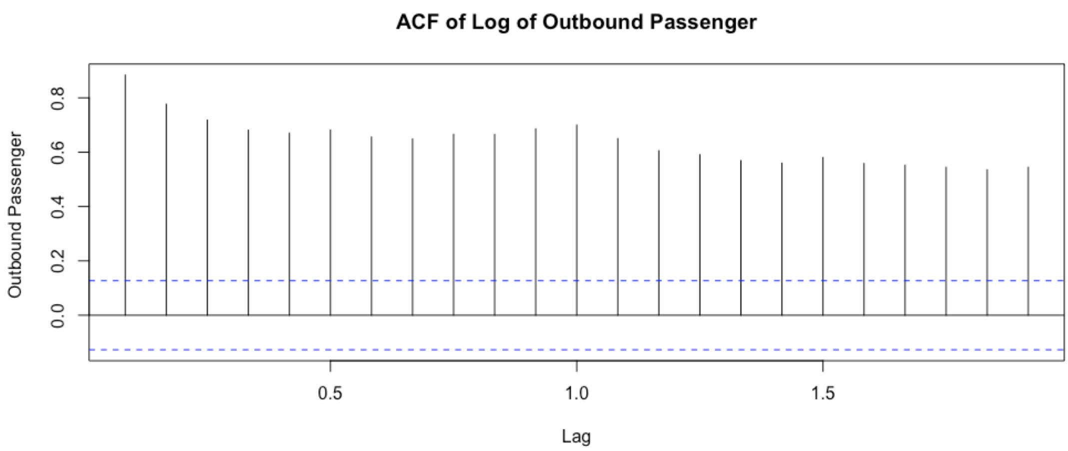
1. Remove Trend

Log

First, we notice the values of outbound passengers are too large. So, we did a log transformation.



And we plot ACF of the log of outbound passengers. This time-series data is nonstationary.



Test

Then, we performed the Dickey-Fuller test and the KPSS test to check if this data is nonstationary and requires differencing or not.

H0: Nonstationary
Ha: Stationary

Dickey-Fuller Test	
Dickey-Fuller test statistic	p-value
-5.3916	<0.01

H0: Stationary
Ha: Nonstationary

KPSS Test	
KPSS test statistic	p-value
3.9929	<0.01

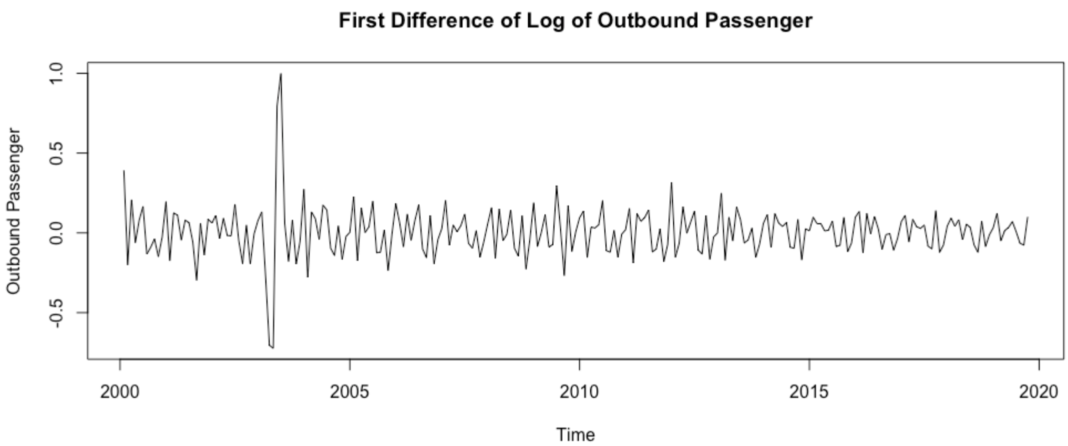
From the Dickey-Fuller test, we can find p-value = 0.01 < 0.05, reject H0. So, we don't need to take a difference according to the Dickey-Fuller test.

From the KPSS test, we can find p-value < 0.05, reject H0. So, we need to take a difference according to the KPSS test.

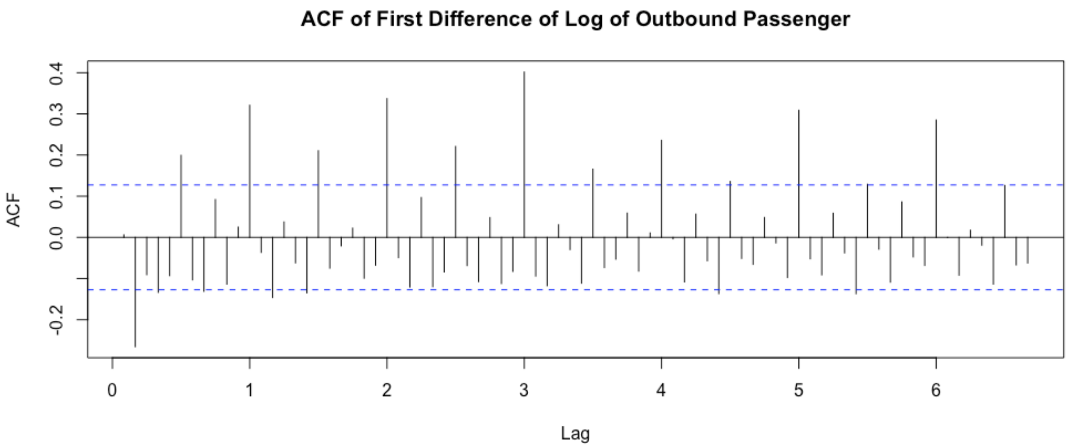
Although the conclusions of the Dickey-Fuller test and the KPSS test are different, the KPSS test is more suitable. Because comparing the meaning behind these tests, in the Dickey-Fuller test, there is no evidence to show that we need to take a difference; in the KPSS, there is evidence to support us to take a difference. Therefore, we choose to do the first difference next.

First Difference

The time-series data is more stationary after the first difference.



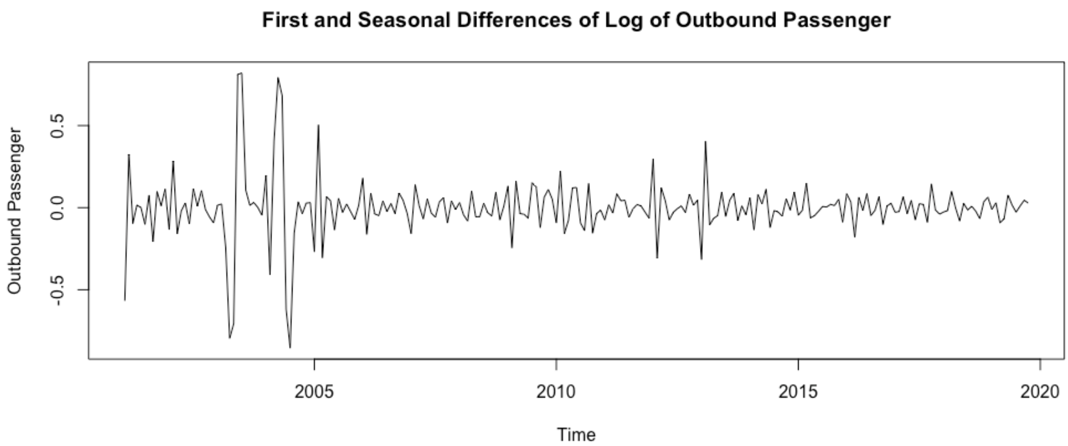
However, from the ACF plot, we observed that the correlation is still significant after 72 steps. It seems we need to take the seasonal difference again.



Seasonal Difference

Through the first difference and the seasonal difference, it seems that this time-series data is stationary already. Next, we can fit a SARIMA model now. However, we observe two sharp fluctuations in 2003 and 2004 at the same time. The sharp fluctuation in 2003 is caused by the outbreak of SARS. And the sharp fluctuations in 2004 is caused by the seasonal difference we had taken.

It's obvious we need to consider an intervention event in 2003 in our model.



Test

After that, we still performed the Dickey-Fuller test and the KPSS test to check if this data is nonstationary and requires differencing or not.

H0: Nonstationary
Ha: Stationary

Dickey-Fuller Test	
Dickey-Fuller test statistic	p-value
-8.0572	< 0.01

H0: Stationary
Ha: Nonstationary

KPSS Test	
KPSS test statistic	p-value
0.019953	> 0.1

From the Dickey-Fuller test, we can find p-value < 0.05, reject H0. So, we don’t need to take the second difference according to the Dickey-Fuller test.

From the KPSS test, we can find p-value > 0.05, do not reject H0. So, we don’t need to take the second difference according to the KPSS test.

The conclusions of the Dickey-Fuller test and the KPSS test are the same, therefore we choose not to take the second difference.

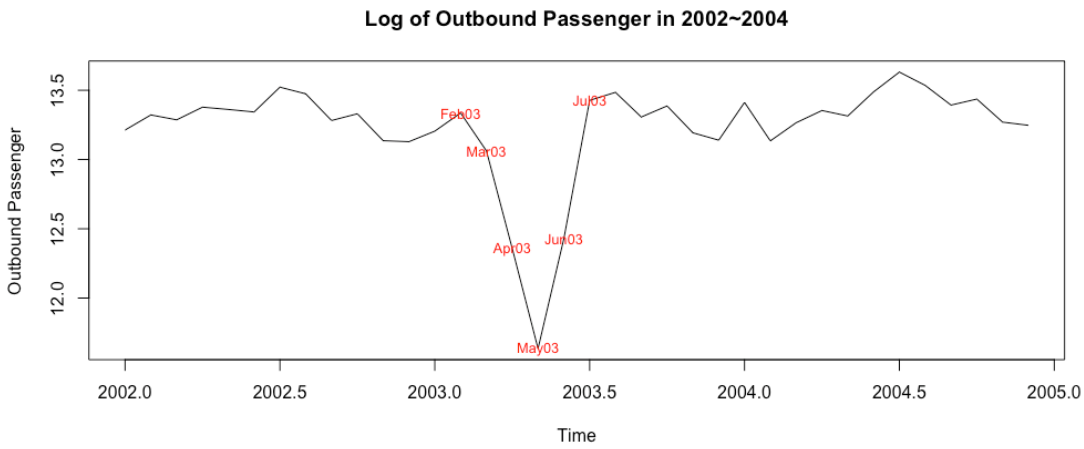
.

Intervention Analysis

From the previous analysis, we know the sharp fluctuation in 2003 is caused by the outbreak of SARS and we need to consider an intervention event in 2003 in our model.

According to the time-series plot below, the decline process can be split into three parts, from February to March, from March to April and from April to May. And the speed of return is also quick.

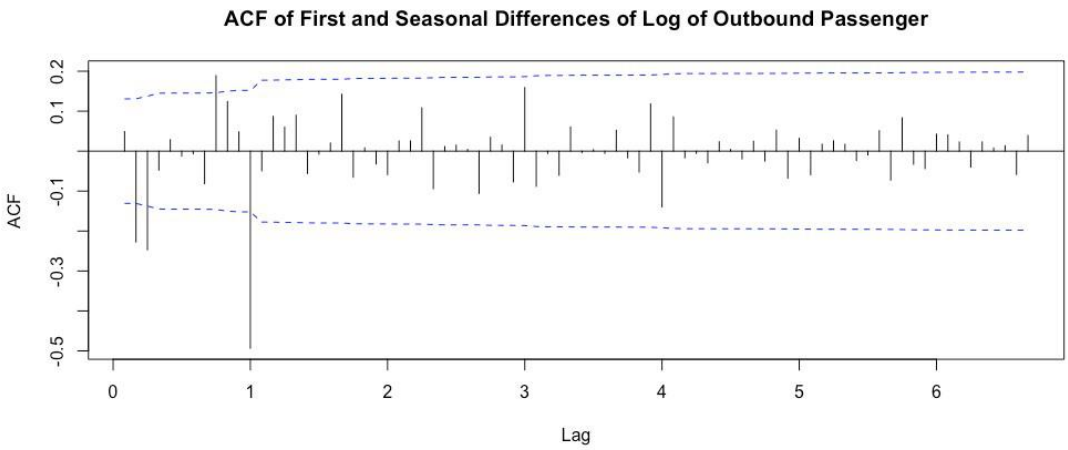
Therefore, it seems fitting an ARMA (1,2) filter at T=March03 is reasonable for this time-series data.



2. Model Specification

We plot the ACF & PACF charts to help us with model specifications.

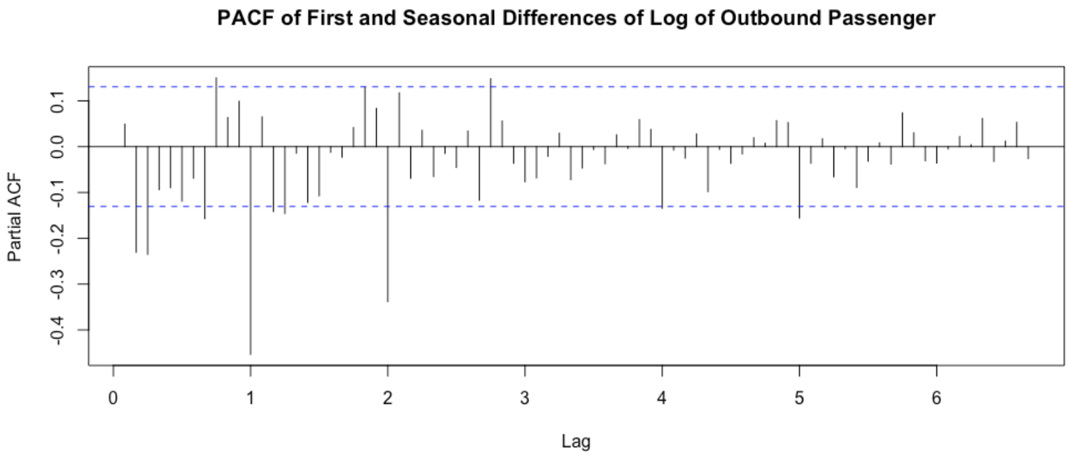
From the ACF plot, we conjecture that model：
$SARIMA(0,1,3) * (0,1,1)_{12}$+ intervention can fit this data well.



However, the coefficients of θ_2(ma2) and θ_3(ma3) aren't significant in $SARIMA(0,1,3) * (0,1,1)_{12}$

Thus, we tried to fit $SARIMA(0,1,1) * (0,1,1)_{12}$+ intervention+ outlier. The coefficients of this model are all significant and the performance seems better.

From the PACF plot, we can observe there are several seasonal steps at lag 1,2,4,5 that out of the confidence interval. It seems that $SARIMA(0,1,0) * (5,1,0)_{12}$ can fit this data well. However, fitting a model indicating that the current value will influence the future value after five years is strange in the data set like outbound passengers. Therefore, we decide not to implement this.



3. Parameter Estimation: $\boldsymbol{SARIMA(0,1,1)*(0,1,1)_{12}}$**+ intervention+ outlier**

Following, we use maximum likelihood to estimate the coefficients of our model. According to the table below, we find the coefficients are significant.

O_t: Outbound Passengers

$\log(O_t) = m_t + N_t \qquad N_t \sim SARIMA(0,1,1)*(0,1,1)_{12}$

$(1 - \mathrm{B})(1 - B^{12})\, N_t = (1-\theta_1 B)(1-\Theta B^{12})\varepsilon_t' + m_t + W_{AO\_50} {P_t}^{(50)}$

$(1 - \delta\mathrm{B})\, m_t = W_0 {P_t}^{(39)} + W_1 {P_{t-1}}^{(39)} + W_2 {P_{t-2}}^{(39)}$

$\varepsilon_t' = \varepsilon_t + W_{IO\_43} P_t^{(43)}$

$\varepsilon_t \sim white\ noise(0, \sigma^2 = 0.003657)$

	θ_1	Θ	δ	W_0	W_1	W_2	W_{IO_43}	W_{AO_50}
Estimate	0.6705	0.8031	0.4953	-0.1729	-0.9080	- 1.2301	0.2682	-0.2565
Standard error	0.0602	0.0437	0.0264	0.0562	0.0576	0.0684	0.0752	0.0533

So, the outbound passenger model is:

$\log(O_t) = m_t + N_t \qquad N_t \sim SARIMA(0,1,1)*(0,1,1)_{12}$

$(1-\mathrm{B})(1-B^{12})N_t = (1-0.6705B)(1-0.8031B^{12}) + m_t - 0.2565{P_t}^{(50)}$

$(1 - 0.4953\mathrm{B})\, m_t = -0.1729\, {P_t}^{(39)} - 0.9080{P_{t-1}}^{(39)} - 1.2301{P_{t-2}}^{(39)}$

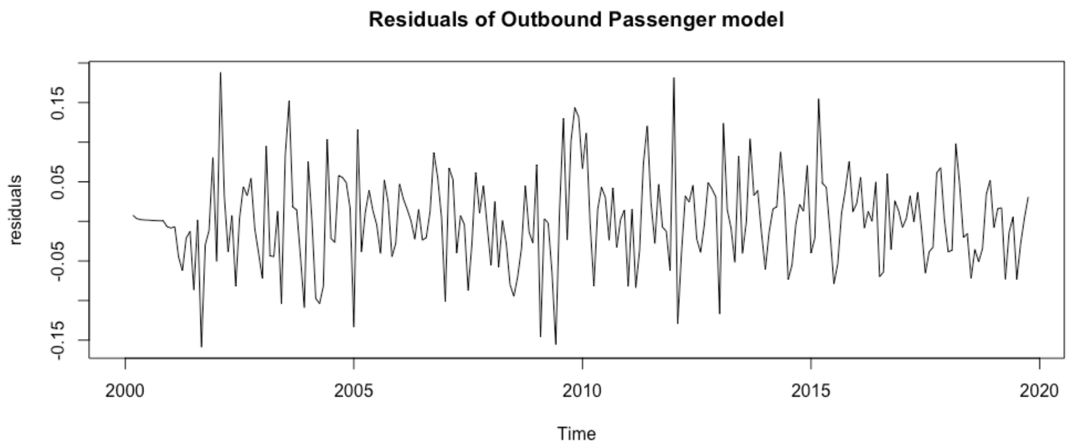
$\varepsilon_t' = \varepsilon_t + 0.2682 P_t^{(43)}$

$\varepsilon_t \sim white\ noise(0, \sigma^2 = 0.003657)$

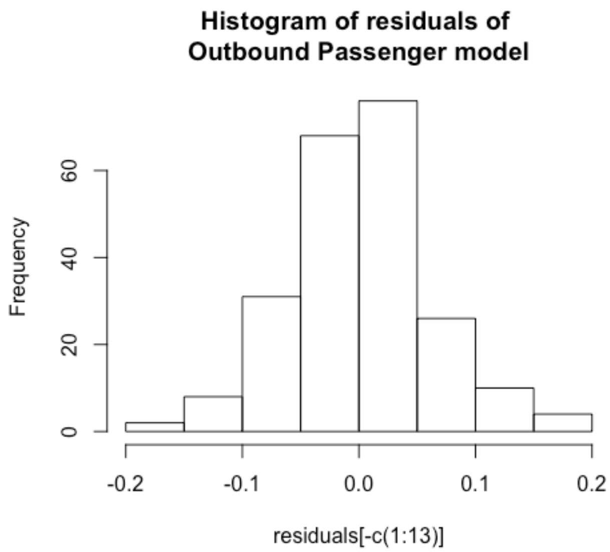
AIC= -589.38

4. Model Diagnostics

After fitting the model, we started to check the model adequacy.
First, we plot the time-series plot of residuals.

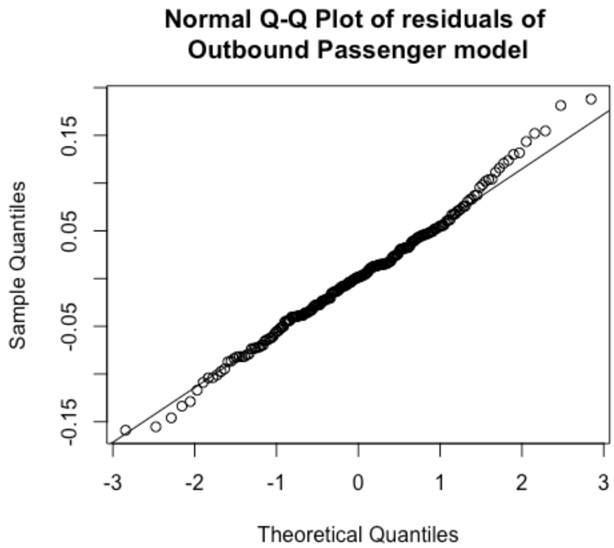


After that, we plot the histogram to have an overview of the residuals and perform the t-test to check if the mean of the residuals is zero or not. According to the right table below, we can find the mean of the residuals is zero.



One Sample t-test	
t test statistic	p-value
0.40337	0.6871

Next, we use the Shapiro-Wilk test and QQ plot to check if the residuals are normally distributed or not. According to the figure and table below, we can find the residuals are normally distributed.

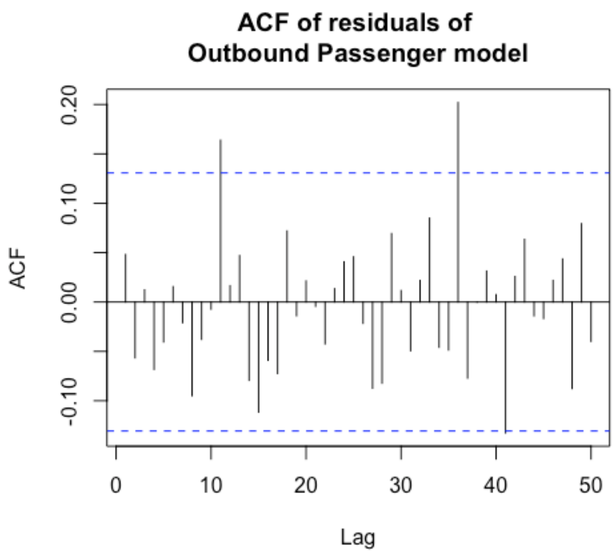


Shapiro-Wilk test	
Shapiro-Wilk test statistic	p-value
0.99229	0.2867

At last, we plot the ACF plot of the residuals. There are only three steps out of the confidence interval.

To check are these steps truly significant or this situation may only be caused by multiple testing simply, we performed the Ljung-Box test later.

The conclusion of the test is accepting the null hypothesis. We don't have enough evidence to show these three steps are significant.



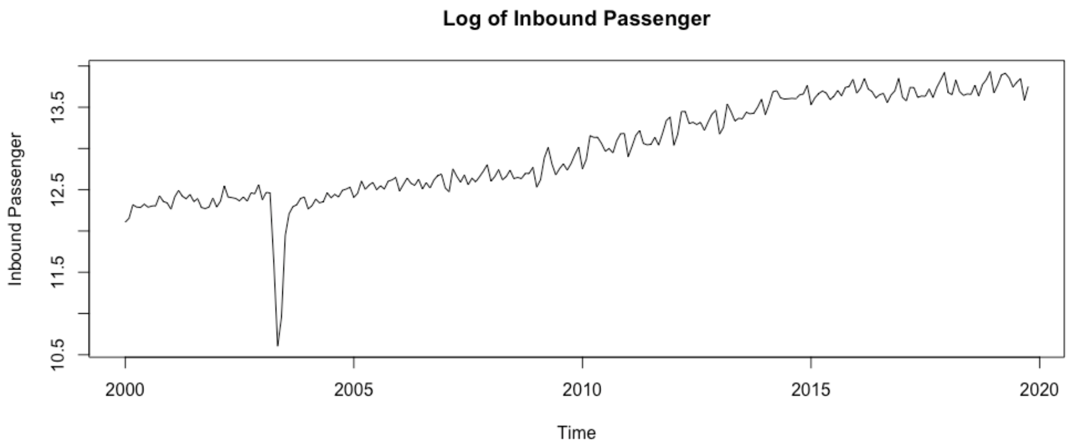
Ljung Box test	
Lag	p-value
lag 11	0.3782
lag 36	0.225
lag 41	0.1869

Time Series Analysis of Inbound Passengers

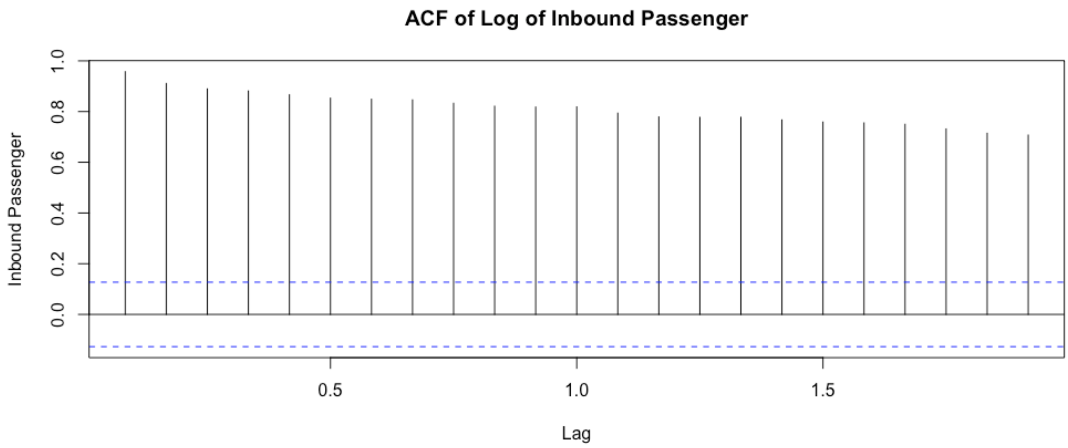
1. Remove Trend

Log

First, we notice the values of inbound passengers are too large, so we did a log transformation.



And we plot ACF of the log of inbound passengers. This time-series data is nonstationary.



Test

Then, we performed the Dickey-Fuller test and the KPSS test to check if this data is nonstationary and requires differencing or not.

H0: Nonstationary
Ha: Stationary

Dickey-Fuller Test	
Dickey-Fuller test statistic	p-value
-3.4954	0.04392

H0: Stationary
Ha: Nonstationary

KPSS Test	
KPSS test statistic	p-value
4.619	<0.01

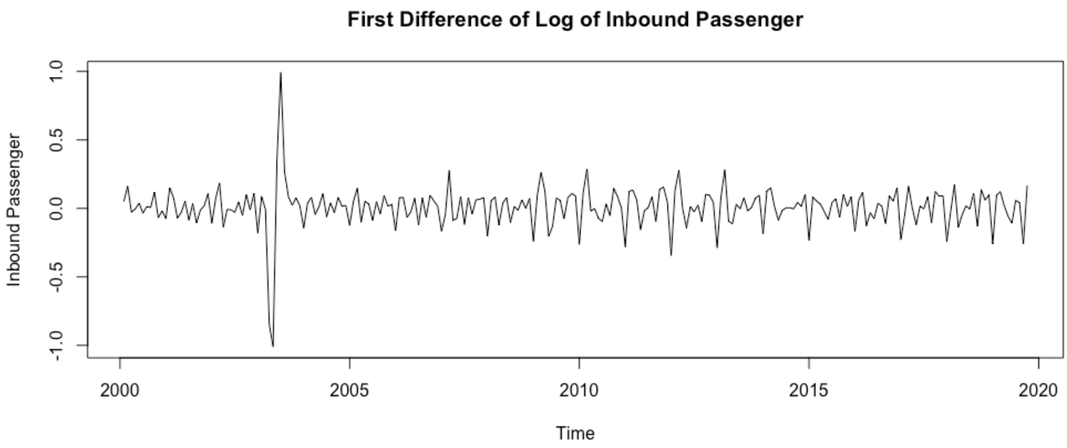
From the Dickey-Fuller test, we can find p-value = 0.04392 < 0.05, reject H0. So, we don't need to take a difference according to the Dickey-Fuller test.

From the KPSS test, we can find p-value < 0.05, reject H0. So, we need to take a difference according to the KPSS Test.

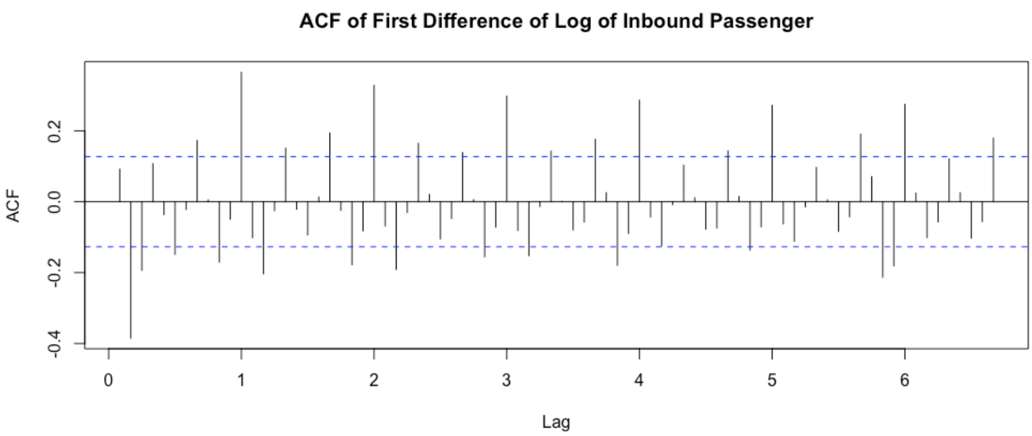
Although the conclusions of the Dickey-Fuller test and the KPSS test are different, the KPSS test is more suitable. Because comparing the meaning behind these tests, in the Dickey-Fuller test, there is no evidence to show that we need to take a difference; in the KPSS, there is evidence to support us to take a difference. Therefore, we choose to do the first difference next.

First Difference

The time-series data is more stationary after the first difference.



However, from the ACF plot, we observed that the correlation is still significant after 72 steps. It seems we need to take the seasonal difference again.

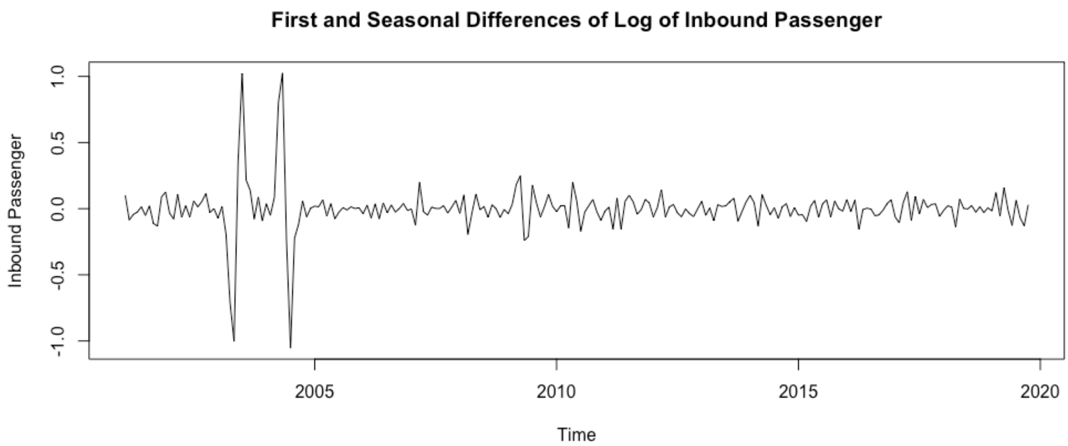


Seasonal Difference

After the first difference and the seasonal difference, it seems that this time-series data is stationary already. Next, we can fit a SARIMA model.

However, we observe two sharp fluctuations in 2003 and 2004 at the same time. The sharp fluctuation in 2003 is caused by the outbreak of SARS. And the sharp fluctuations in 2004 is caused by the seasonal difference we had taken.

It's obvious we need to consider an intervention event in 2003 in our model.



Test

After that, we still performed the Dickey-Fuller test and the KPSS test to check if this data is nonstationary and requires differencing or not.

H0: Nonstationary
Ha: Stationary

Dickey-Fuller Test	
Dickey-Fuller test statistic	p-value
-8.0038	<0.01

H0: Stationary
Ha: Nonstationary

KPSS Test	
KPSS test statistic	p-value
0.01257	>0.1

From the Dickey-Fuller test, we can find p-value < 0.05, reject H0. So, we don't need to take the second difference according to the Dickey-Fuller test.

From the KPSS test, we can find p-value > 0.05, do not reject H0. So, we don't need to take the second difference according to the KPSS test.

The conclusions of the Dickey-Fuller test and the KPSS test are the same; therefore, we choose not to take the second difference.

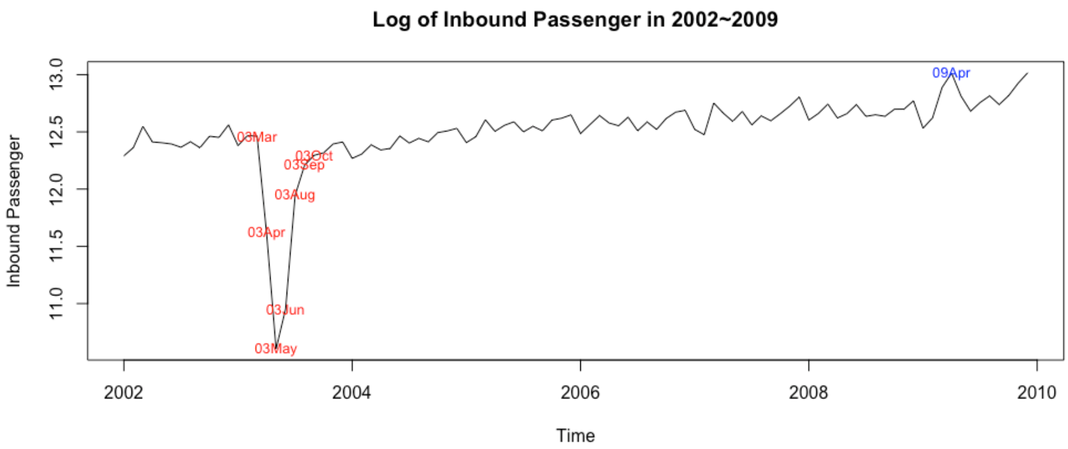
Intervention Analysis

From the previous analysis, we know the sharp fluctuation in 2003 is caused by the outbreak of SARS and we need to consider an intervention event in 2003 in our model.

According to the time-series plot below, the decline process can be split into two parts, from March to April and from April to May. And the speed of return is also quick.

From the time series plot, we conjecture that fitting an ARMA (1, 2) filter at T=March03 is reasonable for this time-series data. However, we found that AIC of ARMA (1, 3) at T=March03 filter is better than ARMA (1, 2) filter at T=March03 and all the regression coefficients are all significant.

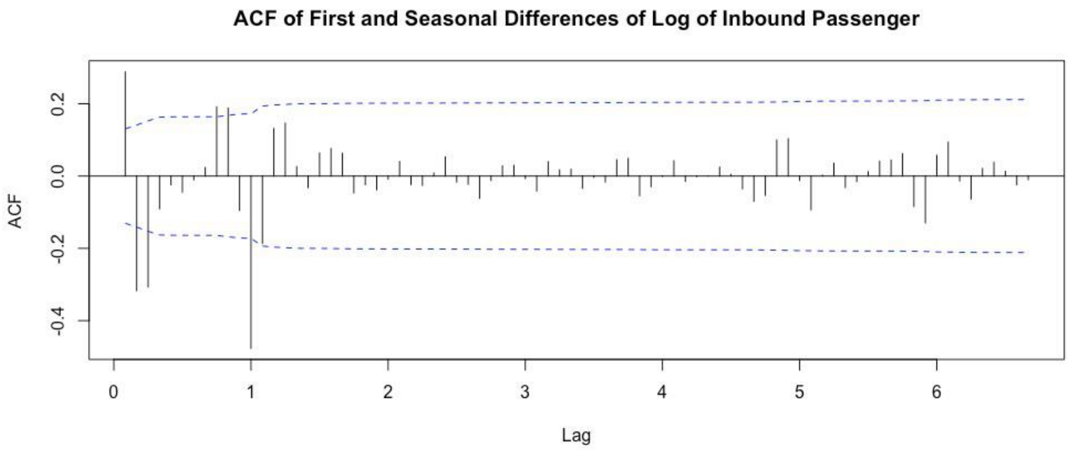
Therefore, we choose to an ARMA (1, 3) filter at T=March03 rather than ARMA (1, 2) filter at T=March03 for this time-series data.



2. Model Specification

Next, we can fit a SARIMA model.
We plot the ACF & PACF charts to help us with model specifications.

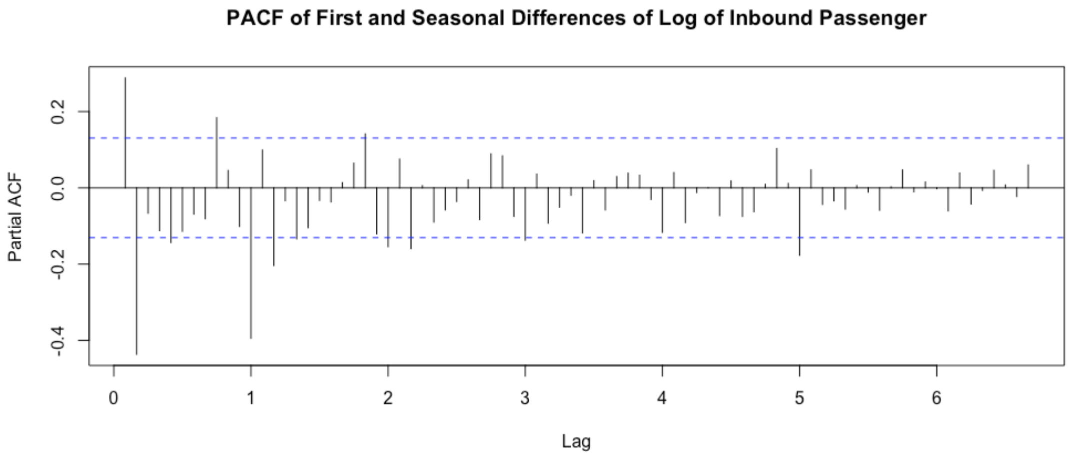
From the ACF plot, we conjecture that model
$SARIMA(0,1,3)*(0,1,1)_{12}$+ intervention can fit this data well.



However, the coefficients of and θ_3(ma3) isn't significant in $SARIMA(0,1,3)*(0,1,1)_{12}$

Thus, we tried to fit $SARIMA(0,1,2)*(0,1,1)_{12}$ + intervention. The coefficients of this model are all significant and the performance seems better.

From the PACF plot, we can observe there are several seasonal steps at lag 1,2,3,5 that out of the confidence interval. It seems that $SARIMA(1,1,0) * (5,1,0)_{12}$can fit this data well. However, fitting a model indicating that the current value will influence the future value after five years is strange in the data set like inbound passengers. Therefore, we decide not to implement this.



3. Parameter Estimation: $\boldsymbol{SARIMA(0,1,2)*(0,1,1)_{12}}$**+ intervention+ outlier**

Following, we use maximum likelihood to estimate the coefficients of our model. According to the table below, we find the coefficients are all significant but W_0 is just barely significant.

I_t: Inbound Passengers

$$\log(I_t) = m_t + N_t \qquad N_t \sim SARIMA(0,1,2)*(0,1,1)_{12}$$

$$(1-\mathrm{B})(1-B^{12})\,N_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta\, B^{12})\tau_t'$$

$$(1-\delta\mathrm{B})m_t = W_0 P_t^{(39)} + W_1 P_{t-1}^{(39)} + W_2 P_{t-2}^{(39)} + W_3 P_{t-3}^{(39)}$$

$$\tau_t' = \tau_t + W_{IO\_112} P_t^{(112)}$$

$$\tau_t \sim white\ noise(0, \sigma^2 = 0.002865)$$

	θ_1	θ_2	Θ	δ	W_0	W_1	W_2	W_3	W_{IO_112}
Estimate	0.3125	0.2431	0.6893	0.2779	-0.0982	-0.7914	-1.5739	-0.9851	0.2670
Standard error	0.0666	0.0688	0.0587	0.0357	0.0480	0.0493	0.0531	0.0672	0.0541

AIC= -643.79

So, the model for inbound passengers is:

$$\log(I_t) = m_t + N_t \qquad N_t \sim SARIMA(0,1,2)*(0,1,1)_{12}$$

$$(1-\mathrm{B})(1-\mathrm{B}^{12})\,N_t = (1-0.3125B-0.2431\mathrm{B}^2)(1-0.6893\mathrm{B}^{12})\tau_t'$$

$$(1-0.2779\mathrm{B})m_t = -0.0982P_t^{(39)} - 0.7914P_{t-1}^{(39)} - 1.5739P_{t-2}^{(39)}$$

$$-0.9851P_{t-3}^{(39)}$$

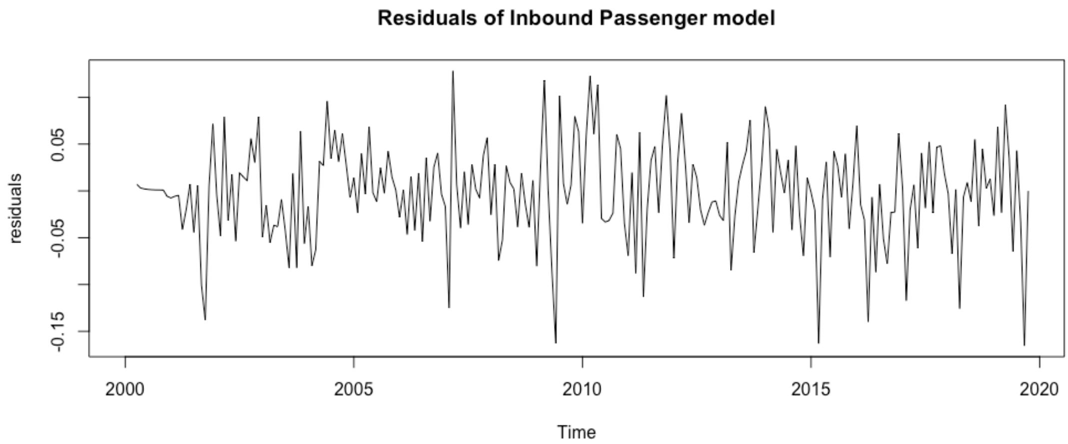
$$\tau_t' = \tau_t + 0.2670P_t^{(112)}$$

$$\tau_t \sim white\ noise(0, \sigma^2 = 0.002865)$$

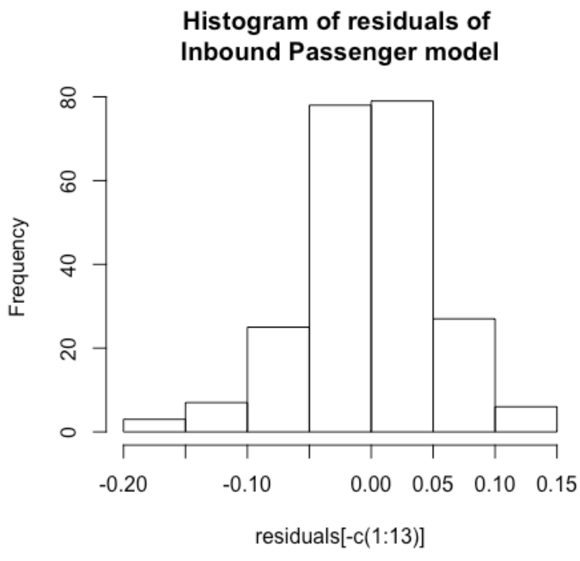
4. Model Diagnostics

After fitting the model, we started to check the model adequacy.

First, we plot the time-series plot of residuals.

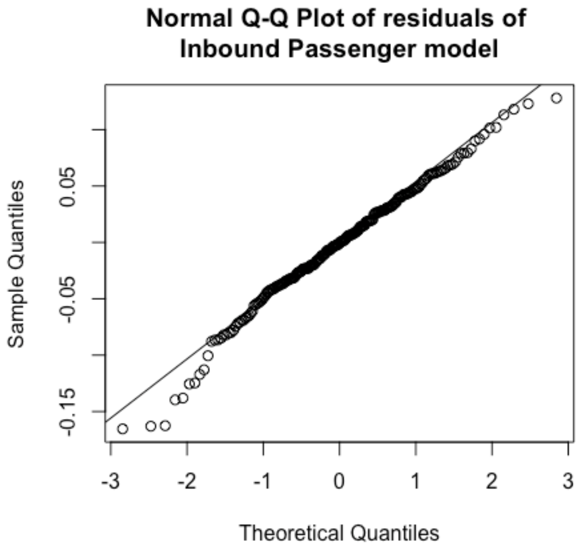


After that, we plot the histogram to have an overview of the residuals and perform the t-test to check if the mean of the residuals is zero or not. According to the right table below, we can find the mean of the residuals is zero.



One Sample t-test	
t test statistic	p-value
-0.6368	0.5249

Next, we use the Shapiro-Wilk test and QQ plot to check if the residuals are normally distributed or not. According to the figure and table below, we can find the residuals aren't normally distributed.

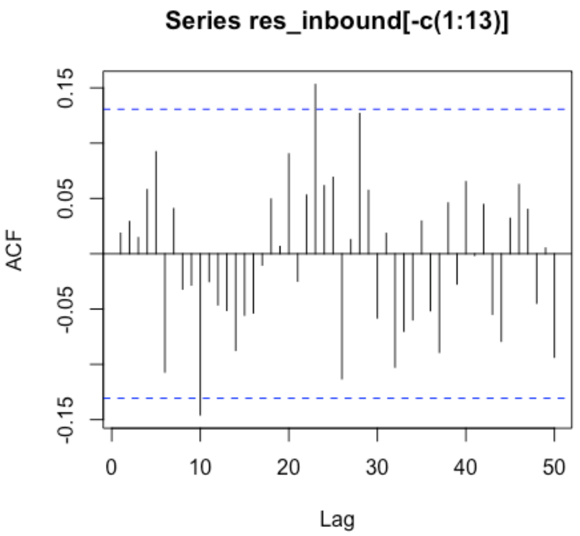


Shapiro-Wilk test	
Shapiro-Wilk test statistic	p-value
0.9869	0.0366

At last, we plot the ACF plot of the residuals. There are only two steps out of the confidence interval.

To check are these steps truly significant or this situation may only be caused by multiple testing simply, we performed the Ljung-Box test later.

The conclusion of the test is accepting the null hypothesis. We don't have enough evidence to show these two steps are significant.



Ljung Box test	
Lag	p-value
lag 10	0.3111
lag 23	0.316

V Correlation

1. Original data correlation

Before we start analyzing the true relations between our data sets, we fit simple linear regression models to our data without removing the time effect directly.

Linear Regression

$$O_t = \alpha_0 + \alpha_1 M_t$$

M2 money supply vs. the number of outbound passengers				
	Estimate	Standard error	t value	p-value
α_0	-80144	30620	-2.617	<0.05
α_1	3.113	0.0977	31.87	<0.05

Adjusted R-squared: 0.8107

$$I_t = \beta_0 + \beta_1 M_t$$

M2 money supply vs. the number of inbound passengers				
	Estimate	Standard error	t value	p-value
β_0	-462036	19252	-24	<0.05
β_1	3.227	0.0614	52.55	<0.05

Adjusted R-squared: 0.9209

According to the table above, we can see these coefficients are all significant without removing the time effect! Moreover, there are positive relations between the money supply and both the number of outbound and inbound passengers respectively.

Furthermore, the R-squared of these two models is incredibly high. It seems the fitting models are great.

After that, we fit simple linear regression models to our data by removing the time effect.

Linear Regression

e_t: $the\ residuals\ of\ the\ M2\ model.$

ε_t: $the\ residuals\ of\ \ model\ for\ outbound\ passengers$

$$\varepsilon_t = \mathrm{A}_0 + \mathrm{A}_1 e_t + \gamma_t$$

$\gamma_t \sim white\ noise$

M2 money supply vs. the number of outbound passengers				
	Estimate	Standard error	t value	p-value
A_0	0.0018	0.0040	0.449	0.654
A_1	1.0430	0.9161	1.138	0.256

Adjusted R-squared: 0.00132

e_t: $the\ residuals\ of\ the\ M2\ model.$

τ_t: $the\ residuals\ of\ \ model\ for\ inbound\ passengers$

According to the table above, we can see all the coefficients of the number of outbound passengers on the M2 money supply are not significant after removing the time effect! As a result, we can conclude that the number of outbound passengers on the M2 money supply we had observed is a pseudo relation.

$$\tau_t = \mathrm{B}_0 + \mathrm{B}_1 e_t + \eta_t$$

$\eta_t \sim white\ noise$

M2 money supply vs. the number of inbound passengers				
	Estimate	Standard error	t value	p-value
B_0	-0.0025	0.0035	-0.720	0.4724
B_1	-1.6030	0.8039	-1.994	0.0474

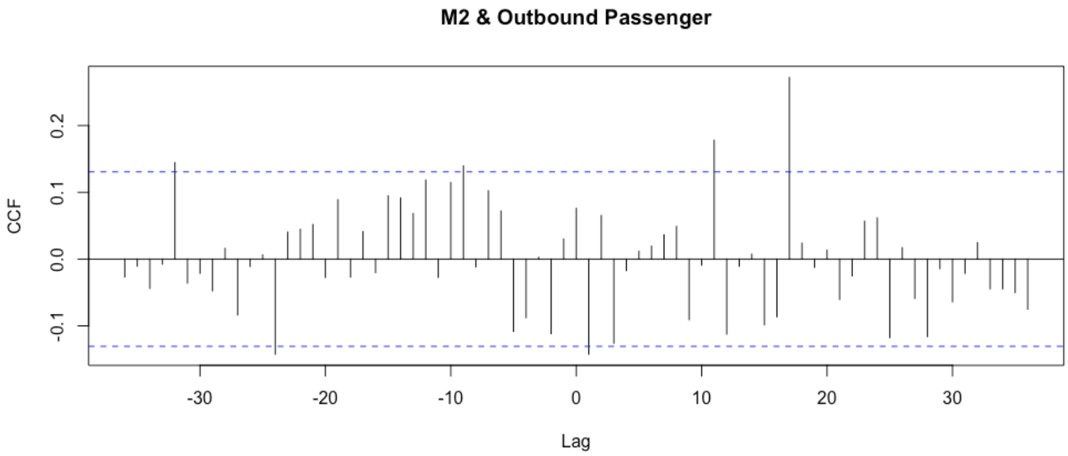
Adjusted R-squared: 0.0131

However, the intercept of the number of inbound passengers on the M2 money supply is not significant after removing the time effect, but the regression coefficient of the number of inbound passengers on the M2 money supply is barely significant after removing the time effect! As a result, we can conclude that there are certain effects between these two data sets.

2. Residuals correlation

Then, we fit simple linear regression models to our residuals again, which represent the true value after removing the time effect. Moreover, we calculate the cross-correlation function (CCF) of the residuals to detect the potential short-term and long-term effects.

By calculating the cross-correlation function of the residuals between the M2 money supply and the number of outbound passengers, we find that there are some lags out of bound. Therefore, we examine the relation of those lags next.

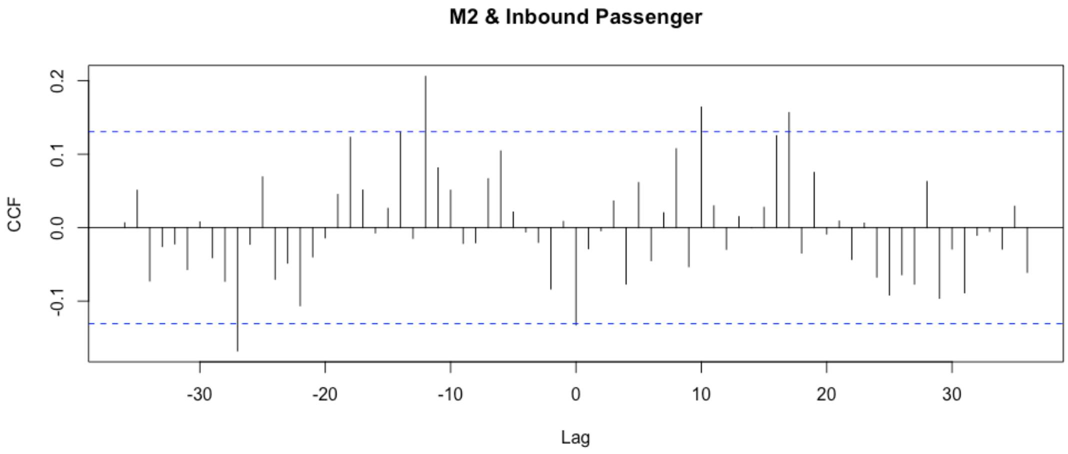


Although the correlation of the M2 money supply to the number of outbound passengers at lag -32, -24, -9, 1, 11, 17 are significant, we can find that the correlation coefficients of those lags are weak below the table. Therefore, we can conclude that there isn't a strong relationship between the M2 money supply and the number of outbound passengers.

Correlation coefficient of $M2$ vs. Outbound Passengers	
Lag	Corr (e_{t+i}, ε_t)
i=-32	0.1674
i=-24	-0.1587
i=-9	0.1443

Correlation coefficient of $M2$ vs. Outbound Passengers	
Lag	Corr (e_{t+i}, ε_t)
i=1	-0.1426
i=11	0.1842
i=17	**0.2870**

Furthermore, we calculate the cross-correlation function of the residuals between the M2 money supply and the number of inbound passengers, we find that there are some lags out of bound. Therefore, we examine the relation of those lags next.



Although the correlations of the M2 money supply to the number of inbound passengers at lag 0, 12, 14, 27 are significant, we can find that the correlation coefficients of those lags are weak below the table. Therefore, we conclude that there isn't a strong relationship between the M2 money supply and the number of inbound passengers.

Correlation coefficient of $M2$ vs. $Inbound$ Passengers	
Lag	Corr(e_{t+i}, τ_t)
i=-27	-0.1844
i=-14	0.1377
i=-12	**0.2164**

Correlation coefficient of $M2$ vs. $Inbound$ Passengers	
Lag	Corr(e_{t+i}, τ_t)
i=10	0.175
i=17	0.1703

VI Conclusion

In conclusion, we try to remove the time effect of our data and run the simple linear regression.
And we can get the following conclusions:

1. The regression coefficients of the simple linear regression between the number of outbound passengers and the M2 money supply aren't significant anymore after removing the time effect. And the regression coefficients between the number of inbound passengers and the M2 money supply are barely significant. The effects we observed in simple linear regression is caused by the time effect, so this is the so-called **pseudo relation**.

2. We also use the cross-correlation functions to examine potential effects between our data sets. From the CCF, we can observe there are long-term and short-term effects between these two data sets. However, the long-term and short-term effects between our data sets are weak.

VII Discussion and the Future Work

In this project, we performed logarithmic transformation and used the Augmented Dickey-Fuller test and KPSS test to help us examine if these time-series data are stationary or not. Next, we plot the ACF plot to assist us in establishing models. After that, we check the models' adequacy.

Through these analyses, we use SARIMA models to remove the time effects of our data sets and fitting simple linear regression again. Consequently, we found, the regression coefficients between the number of outbound passengers and the M2 money supply aren't significant anymore. And the regression coefficients between the number of inbound passengers and the M2 money supply are barely significant. Therefore, we conclude that the relationships between the relations between the M2 money supply and both the number of outbound and inbound passengers respectively we had seen are caused by time-effect simply.

discuss.

Furthermore, we suspected there are still exist potential long-term and short-term effects. For this reason, we use the cross-correlation function to check the potential effects.

From the CCF, we observe there are several steps out of the confidence interval. However, the correlation coefficients are all less than 0.3. This evidence indicates that the long-term effects between the number of outbound passengers and the M2 money supply, the number of inbound passengers and the M2 money supply are weak correlation.

Clearly, among all economic factors, the M2 money supply may not be the most significant one correlated with the number of outbound passengers. Therefore, the next question should be which of all economic factors is the most relevant. In that case, we will leave this problem as future work!

Reference

https://hdl.handle.net/11296/bja5rr [1]

https://www.quora.com/Money-Supply-What-is-M0-M1-M2-and-M3[2]

https://en.wikipedia.org/wiki/Money_supply[3]

https://www.macromicro.me/charts/12879/tai-wan-M1B-M2-huo-bi-zong-ji-shu[4]

https://stat.motc.gov.tw/mocdb/stmain.jsp?sys=100&funid=a7101[5]

心得:

從期末報告中可以發現實證資料不會像課本中練習的例子一樣模型配適良好，在當中遇到很多不同的困難，舉例來說當我們透過時間序列圖配適 outlier，但是當我們考慮 outlier 的效應卻又發現模型不顯著，又或者在 CCF 中檢測特定 lags 有顯著效應，卻是難以解釋的弱相關短期效應與長期效應…等，還有一個更大的困難是原來用英文寫報告需要花這麼多的心力! 希望老師不會覺得看我們的報告更困難～ 雖然做這份報告真的有點辛苦，但是遇到困難可以和同組的同學們一起討論，上課沒聽懂的部分可以透過實作更釐清觀念，從討論中也能學到不少東西。本堂課從理論、模型配適練習到實作，每一階段都有扎實的訓練，覺得老師很用心!

建議 ：

1. [配適偶發事件]

請老師上課時能詳細說明如何用 R 套件配適偶發事件，舉例來說 ：

xtransf=data.frame(I911=1*(seq(airmiles)==69), I911=1*(seq(airmiles)==69)),
transfer=list(c(0,0),c(1,0))

其中 transfer 的第一個 c(0,0)，指的是對Ｐt 而非為Ｓt，可參照課本Ｐ.452 說明。

2. [選定合適模型]

另外因為實證資料常常不會像課本中練習的例子一樣容易選擇模型，希望老師能夠分享如何透過 acf 與 pacf 的圖，快速選定合適的 sarima 模型。

3. [課程編排順序]

建議能夠先教第四章 Stationary Time Series 後再教第三章 Trends，能夠對模型先有診斷的概念。另外還有使用 CCF 檢測潛在因果效應也可以再配適完模型後就先教，如此以來，能夠對模型解釋先有初步概念。