112-2 TIME SERIES ANALYSIS FINAL REPORT

Group 6

Exchange rate of Euro to US dollar

Author：

110304039 CHIU,SHIH-FANG
110304036 SHIH,JING-JIE
110304011 HSU,WEI-BO

Contents

List of Figures

List of Tables

1 Introduction

This research aims to explore the U.S. dollar to Euro exchange rate data by analyzing its trends and relationship with time. We will identify its pattern and determine the best-fitting model(s). The study will begin with data preprocessing. Techniques such as regression detrending and differencing may be employed to remove the trend, based on the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the raw data. Following this, we will select several models that potentially fit the data well. By conducting diagnostics on each model, we will choose the most suitable one for predicting future exchange rates.

Data: U.S Dollars to Euro Spot Exchange Rate
Units: U.S Dollars to one Euro
Frequency: week
Number of data: 366
Time Period: 2017/05   2024/05
Source: Federal Reserve Economic Data

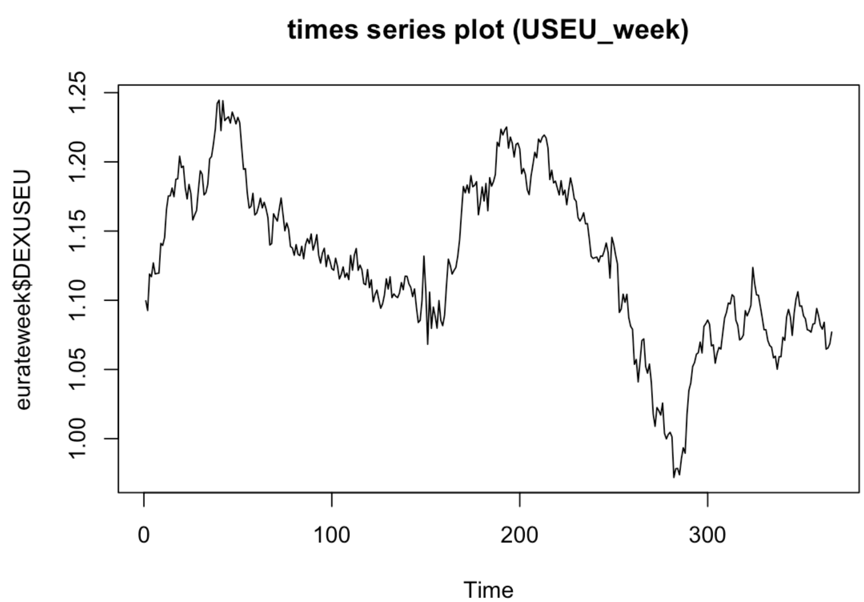

Figure 1-1: Original Data time series plot

Initially, we used the data with a daily unit. However, due to the presence of N/A values, we changed the unit from daily to weekly.

2 Data Processing

2.1 ADF test

Before selecting a model, we first need to perform an ADF test on the time series data to understand the stability of the original data. The alternative hypothesis is that this time series is stationary, while the null hypothesis is that this time series is not stationary. After conducting the ADF test, we obtained a p-value of 0.3863, which is not less than 0.01. Therefore, we do not reject the null hypothesis, and we conclude that this time series is not stationary.

2.2 Regression Detrend

Next, we plotted the ACF and PACF charts as shown below.

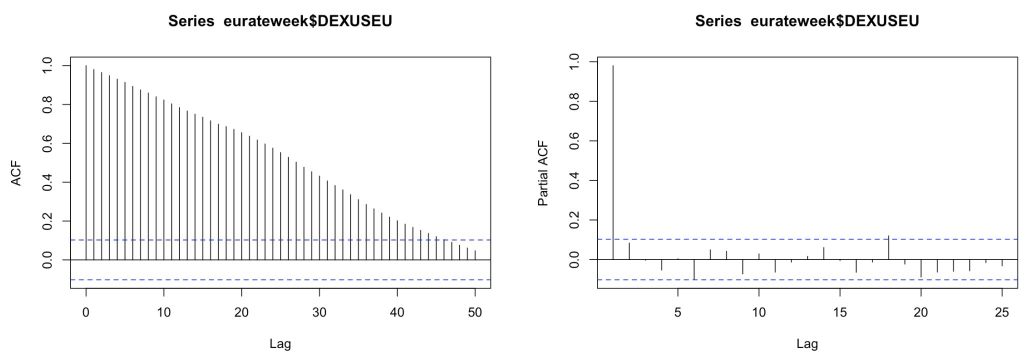

Figure 2-1: ACF and PACF Plots

By directly observing the ACF chart, we can see that as the lag increases, the ACF decline is too slow (greater than 0.8 at lag=20). Therefore, we decided to detrend the data using regression detrend.

At lag = 1, using single-term regression (t^1) for detrending, the ADF test on the residuals shows a p-value of 0.3863, indicating non-stationarity, as shown in the residual plot.

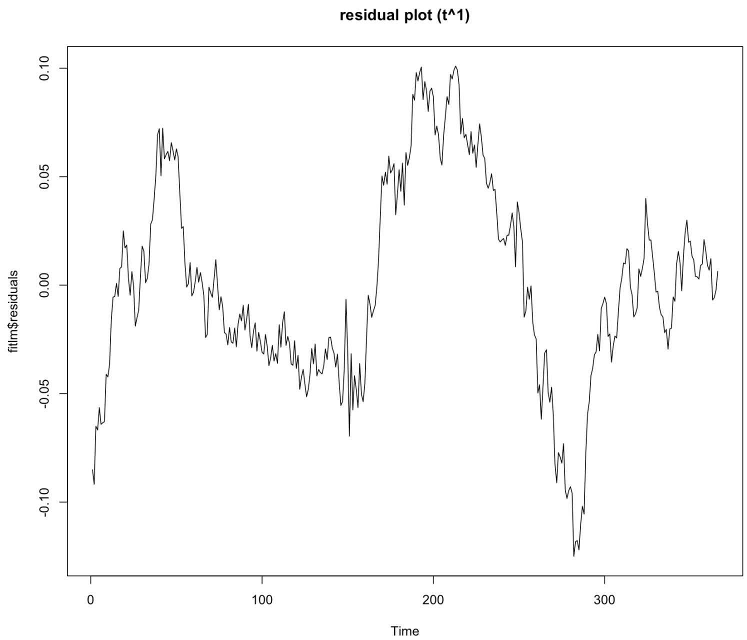

Figure 2-2: Residual plot

2.3 First Order Difference

Finally, we decided to use first-order differencing to detrend. The following plots show the ACF and PACF charts after first-order differencing.

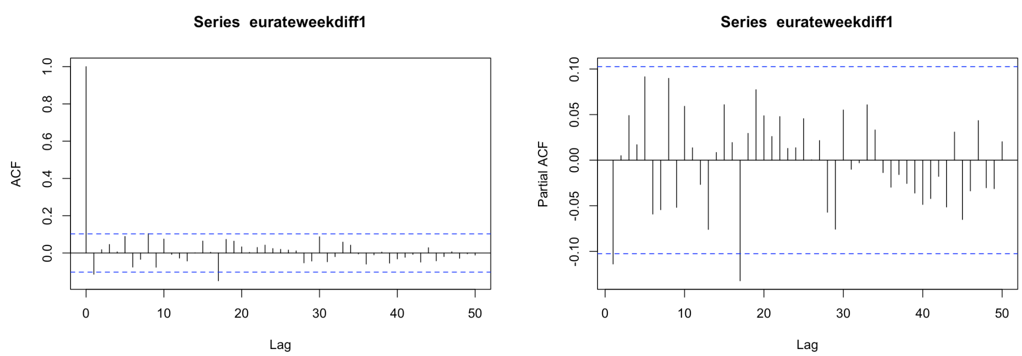

Figure 2-3: ACF and PACF Plots After Differencing

3 Model Selection

The following plots show the ACF and PACF charts after first-order differencing.

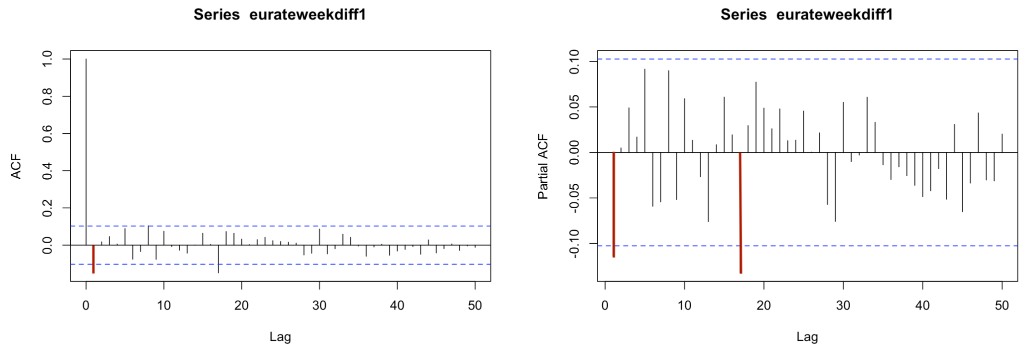

Figure 3-1: ACF and PACF Plots After Differencing

Table 3-1: ACF and PACF Interpretation

	ACF	PACF
Non-seasonal	may CUTS OFF at lag =1	may CUTS OFF at lag = 1,17

From the ACF and PACF charts, we can interpret that the ACF cuts off at lag = 1 and the PACF cuts off at lag = 1 and lag = 17. Therefore, we consider the models ARIMA(1,1,1), ARIMA(17,1,1). Additionally, we can also explain that ACF tails off and PACF cuts off at lag =1 ,so we try AR(1) as our moedel selection. Besides, if ACF cuts off at lag = 1 and PACF tails off, MA(1) may be another choose for us to try.

4 Model fitting

Therefore, we choose four models, ARIMA(1,1,1), ARIMA(17,1,1), ARIMA(0,1,1) and ARIMA(1,1,0), and use residual analysis, that is, the Ljung-Box test and the comparison of AIC and BIC are used to select the final suitable model.

4.1 ARIMA(1,1,1)

First of all, we explain the ARIMA(1,1,1) model.

The figure below shows the results of the Ljung-Box test. It can be found that almost every predicted value falls above the significance line, which means that the model fitting effect may be good. Furthermore, we also performed parameter estimation, and the AIC value was -2264.49.

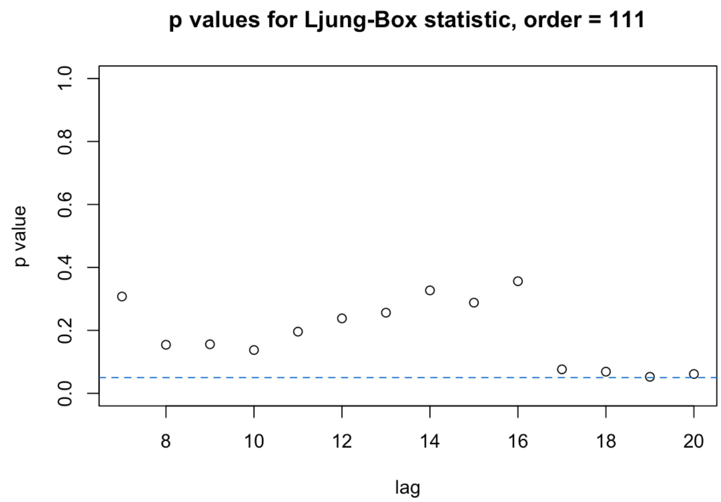

Figure 4-1: Ljung-Box Test Results for ARIMA(1,1,1)

Also , the ARIMA(1,1,1) can be written as

$$(1 + 0.1391B)\nabla X_t = (1 + 0.0254B)Z_t$$

where Zt is white noise, B is backshift operator, Xt is the time series after non-seasonal difference.

4.2 ARIMA(17,1,1)

Then, the next model tried is ARIMA(17,1,1). We also perform the Ljung-Box test to check whether this model is suitable. Since the coefficient of this model is very large, we set the lag to 30 to see the results.

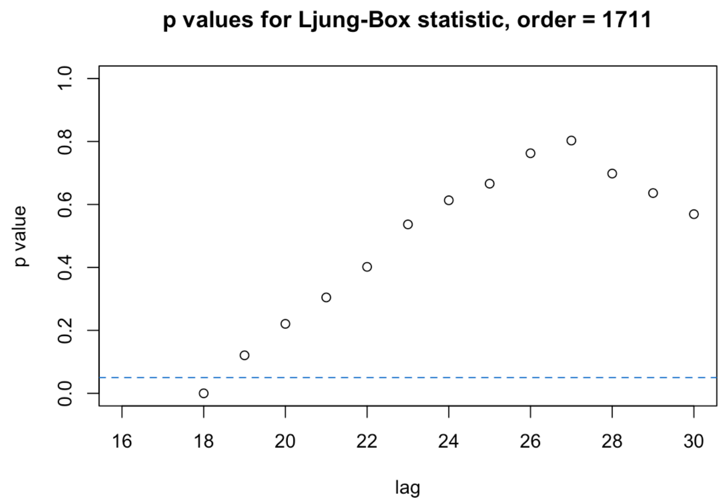

Figure 4-2: Ljung-Box Test Results for ARIMA(17,1,1)

According to the results in the figure 4-2, it can be found that except for the case where lag = 17, the predicted value is less than the significant line, the other predicted values are higher than the significant line. However, there is only one value less than the significant line, and its AIC value is -2254.97. Therefore, we tentatively believe that this model can still fit this data well.

So, the ARIMA(17,1,1) can be written as

$$(1 - 0.2180B - 0.0163B^2 - 0.0558B^3 - 0.0317B^4 - 0.0772B^5 + 0.0602B^6 + 0.0562B^7 - 0.0812B^8 + 0.0158B^9 - 0.0508B^10 - 0.0042B^{11} + 0.0235B^{12} + 0.0799B^{13} - 0.0162B^{14} - 0.0650B^{15} - 0.0139B^{16} + 0.1327B^{17})\nabla X_t = (1 - 0.1184B)Z_t$$

where Zt is white noise, B is backshift operator, Xt is the time series after non-seasonal difference.

4.3 ARIMA(1,1,0)、ARIMA(1,1,0)

Finally, we try two models: ARIMA(1,1,0) and ARIMA(0,1,1).

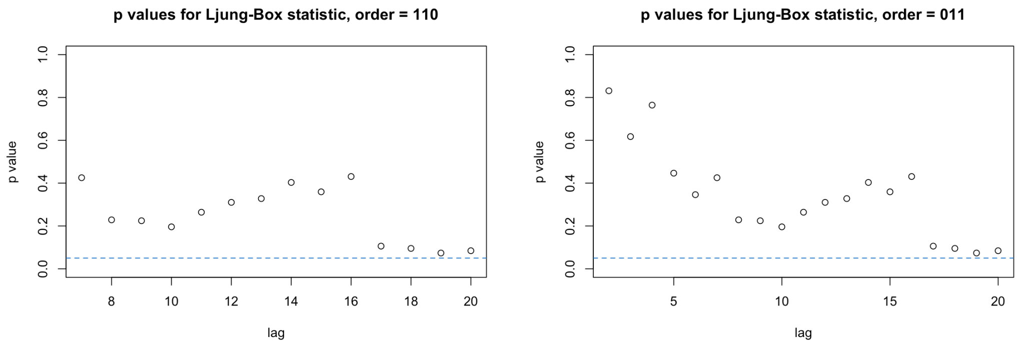

Figure 4-3: Ljung-Box Test Results for ARIMA(1,1,0) and ARIMA(0,1,1)

The left picture is the result of ARIMA(1,1,0) and its model can be written as

$$\nabla X_t = (1 - 0.1094B)Z_t$$

where Zt is white noise, B is backshift operator, Xt is the time series after non-seasonal difference.

The right picture is the result of ARIMA(0,1,1) and the model can be written as

$$(1 + 0.1138B)\nabla X_t = Z_t$$

where Zt is white noise, B is backshift operator, Xt is the time series after non-seasonal difference.

From these two figures, we find that the distribution of predicted values is very similar, and all predicted values are above the significance line. In addition, their AIC values are -2266.49 and -2266.33 respectively, which are also very close.

Since there is no significant rejection in the Ljung-Box test of the four models, it also means that these four models can fit the data we selected well. However, in order to select the final adaptation model, we also conduct a BIC value test. The following figure is a table summary of the comparison of AIC and BIC among these four models.

Table 4-1: AIC and BIC of our four candidate models

Model	AIC	BIC
ARIMA(1,1,1)	-2264.49	-2250.792
ARIMA(17,1,1)	-2254.97	-2178.875
ARIMA(1,1,0)	-2266.49	-2256.686
ARIMA(0,1,1)	-2266.33	-2256.527

From the above table, we can find that ARIMA(1,1,0) has the minimum value in both AIC and BIC, and its Ljung-Box test results are all greater than the significant line, so we finally selected ARIMA(1,1,0) as our adaptation model.

To further confirm that this model fits our data, we perform a coefficient t test.

H0: The coefficient has no significant impact on the model
H1: The coefficient has a significant impact on the model

Since P-value = 0.02860, almost close to zero, we reject the null hypothesis. Therefore, the coefficient of determination has a significant impact on the model, which also means that the model we chose can be well adapted.

5 Prediction

To proceed with our model fitting results, we will now employ the ARIMA(0,1,1) model for prediction. For the one-step prediction, we have 366 data points in total. We use the first 360 data points to predict the values of the 361st to 366th data points.

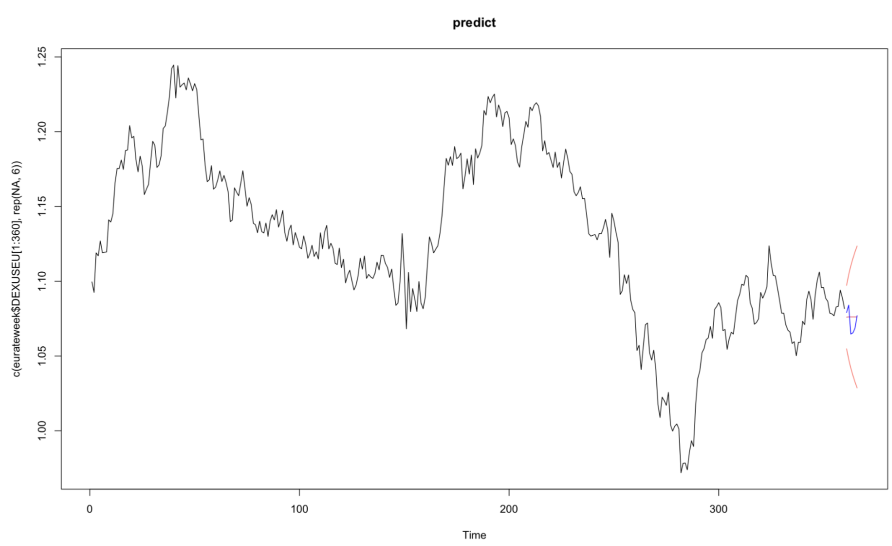

Figure 5-1: One-step Prediction

6 New part (regression detrend t^9)

6.1 Another data processing

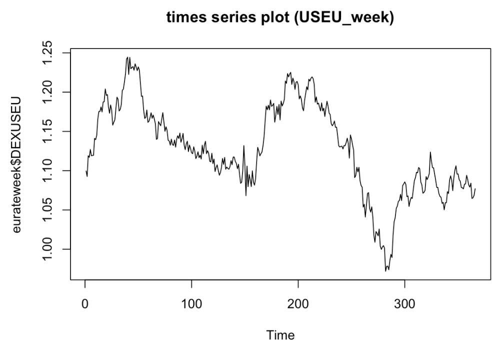

Figure 6-1: Time sereis Plot of original data

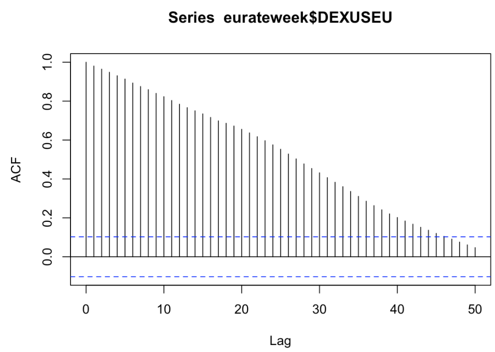

Figure 6-2: ACF of original data

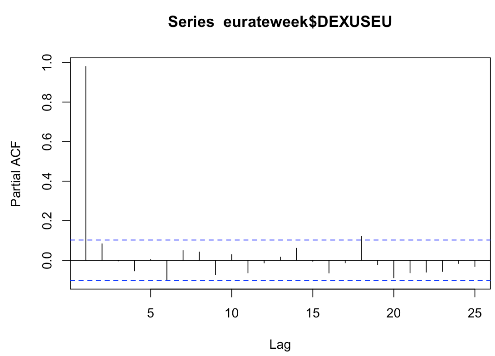

Figure 6-3: PACF of original data

We can see from Figure 6-1 that there is a trend in the original data and from Figure 6-2 we can see that ACF decreases slowly as lag increases. Hence, we employ regression detrend to deal with the trend.

First, we conduct the regression detrend with (t^1) ; however, the p-value is bigger than 0.01, which means that residual is not stationary. It is until t^9 that the residual becomes stationary.

Table 6-1: P-value of t in each power

t^n power	P-value
t^1	0.3863
t^2	0.4415
t^3	0.4437
t^4	0.26
t^5	0.4418
t^6	0.1195
t^7	0.03577
t^8	0.01757
t^9	<0.01

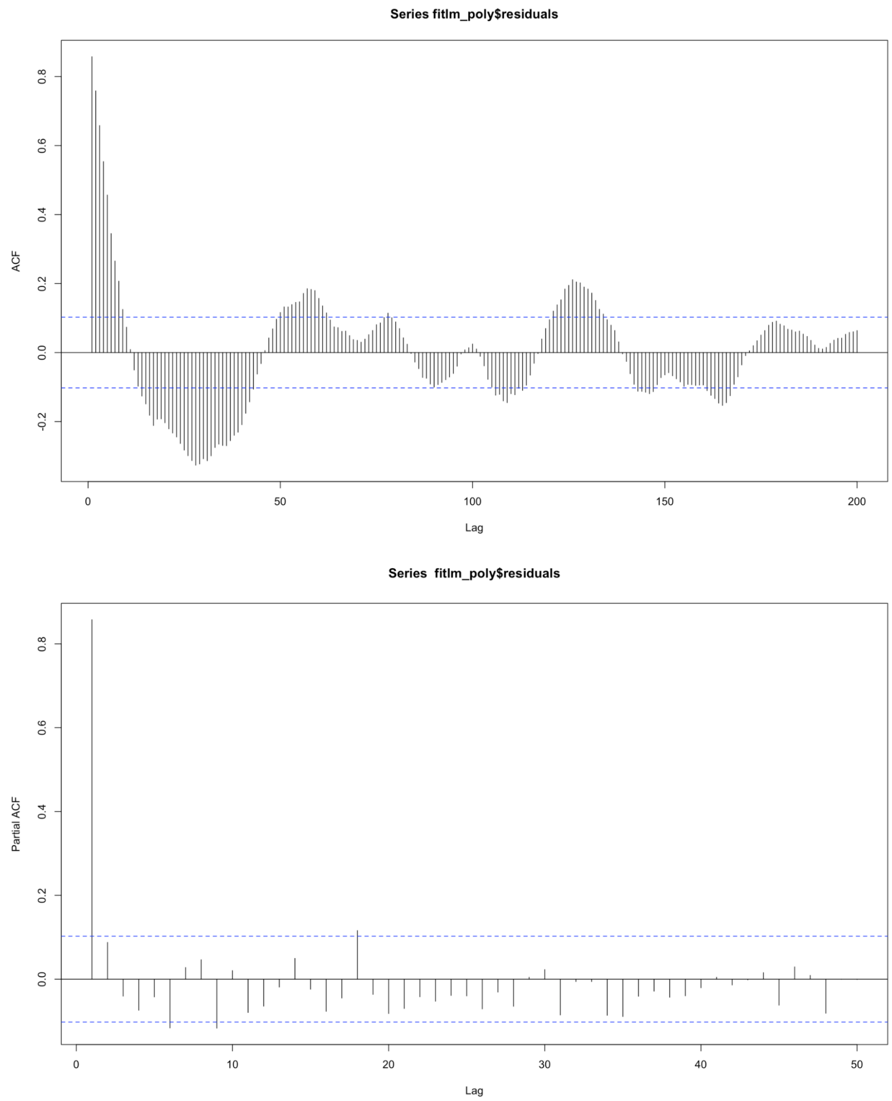

Figure 6-4: ACF and PACF of residual after conduction regression detrend when t^9

We can see from figure 6-4, there exists a seasonal trend in ACF. Since it does not show a steady period, we tried three kinds of seasonal periods. The first one is period = 50, the second one is period = 25, and the third one is period = 29. Because the original data is recorded in a weekly unit, we can roughly consider period = 25 and period = 29 as half-year-period and period = 50 as one-year period. On the other hand, the nonseasonal part of ACF tails off. For the PACF plot, it cuts off at lag = 1, 6, 9, 18

6.2 Model Selection

From what we observe from the chart, we will tried the following models which are

- SARIMA$(1,0,0)(0,0,1)_{50}$
- SARIMA$(1,0,0)(0,0,1)_{25}$
- SARIMA$(1,0,0)(0,0,1)_{29}$
- SARIMA$(6,0,0)(0,0,1)_{50}$
- SARIMA$(6,0,0)(0,0,1)_{25}$
- SARIMA$(6,0,0)(0,0,1)_{29}$
- SARIMA$(9,0,0)(0,0,1)_{50}$
- SARIMA$(9,0,0)(0,0,1)_{25}$
- SARIMA$(9,0,0)(0,0,1)_{29}$
- SARIMA$(18,0,0)(0,0,1)_{50}$
- SARIMA$(18,0,0)(0,0,1)_{25}$
- SARIMA$(18,0,0)(0,0,1)_{29}$

6.3 SARIMA$(1,0,0)(0,0,1)_{50}$

```
Coefficients:
         ar1    sma1  intercept
      0.8600  0.0271     0.0002
s.e.  0.0266  0.0505     0.0039

sigma^2 estimated as 0.0001077:  log likelihood = 1151.96,  aic = -2297.92
```

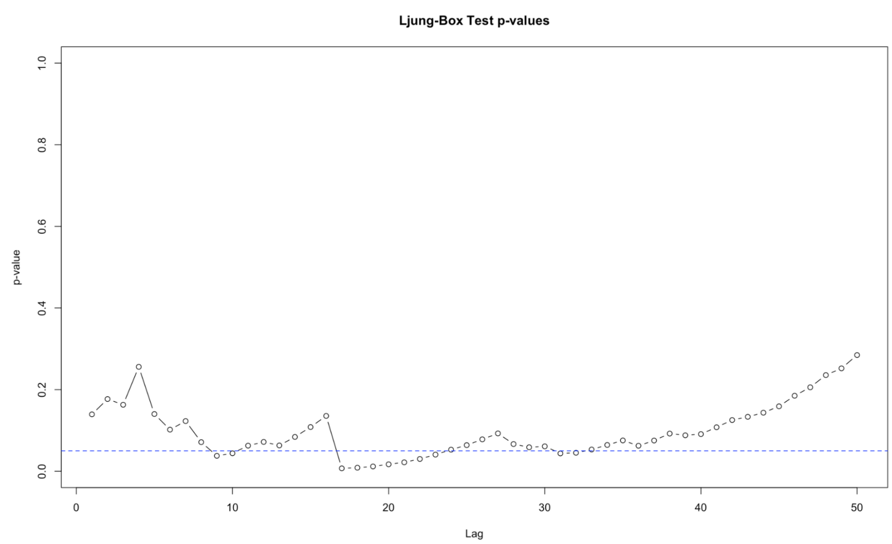

Figure 6-5: Ljung-Box Test Results for SARIMA$(1,0,0)(0,0,1)_{50}$

6.4 SARIMA$(1,0,0)(0,0,1)_{25}$

```
Coefficients:
         ar1     sma1  intercept
      0.8586  -0.0284     0.0002
s.e.  0.0270   0.0523     0.0037

sigma^2 estimated as 0.0001077:  log likelihood = 1151.96,  aic = -2297.93
```

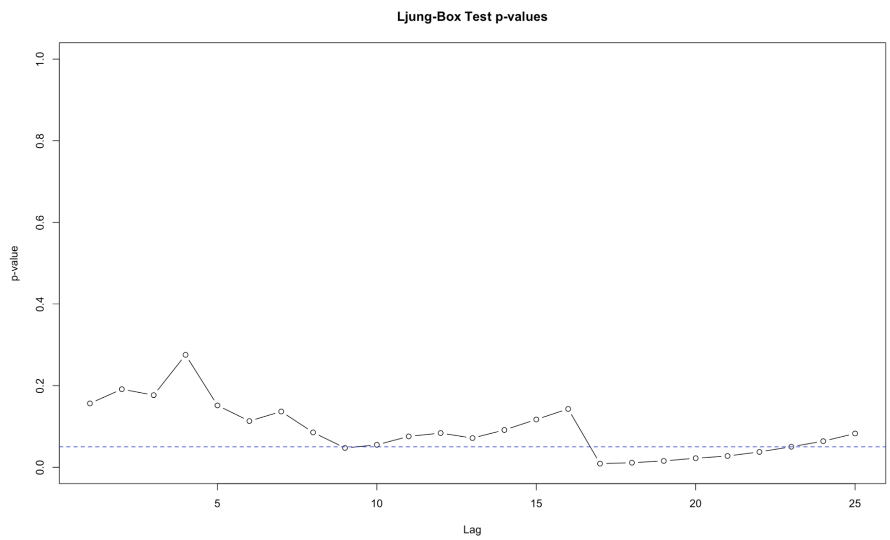

Figure 6-6: Ljung-Box Test Results for SARIMA$(1,0,0)(0,0,1)_{25}$

6.5 SARIMA$(1,0,0)(0,0,1)_{29}$

```
Coefficients:
         ar1     sma1  intercept
      0.8555  -0.0757     0.0001
s.e.  0.0272   0.0539     0.0034

sigma^2 estimated as 0.0001071:  log likelihood = 1152.81,  aic = -2299.61
```

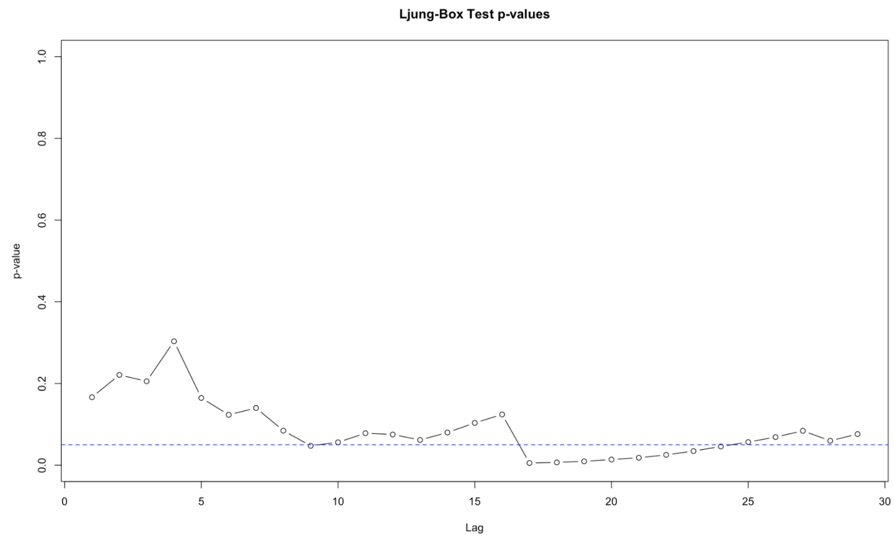

Figure 6-7: Ljung-Box Test Results for SARIMA$(1, 0, 0)(0, 0, 1)_{29}$

6.6 SARIMA$(6, 0, 0)(0, 0, 1)_{50}$

```
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6    sma1  intercept
      0.7760  0.1153  0.0416  -0.0252  0.0571  -0.1346  0.0426     0.0001
s.e.  0.0521  0.0663  0.0667   0.0666  0.0669   0.0525  0.0513     0.0032

sigma^2 estimated as 0.0001039:  log likelihood = 1158.32,  aic = -2300.64
```

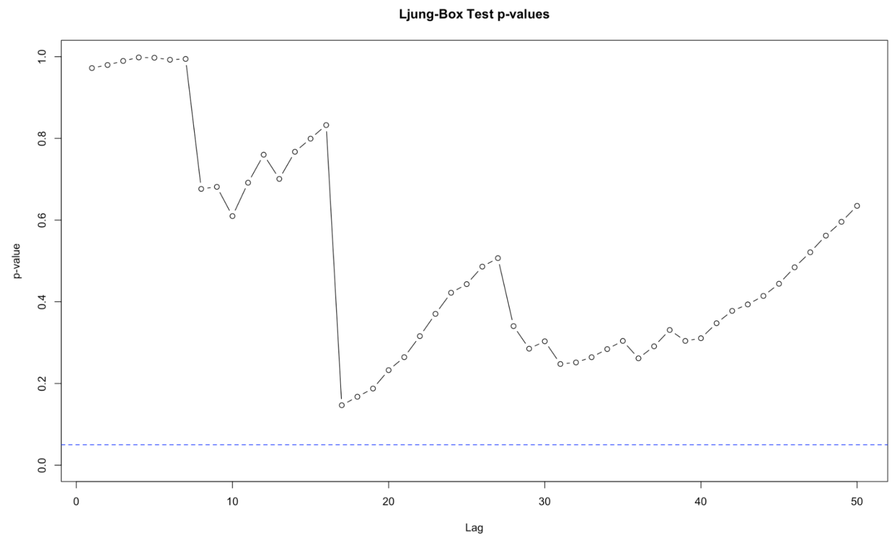

Figure 6-8: Ljung-Box Test Results for SARIMA$(6,0,0)(0,0,1)_{50}$

6.7 SARIMA$(6,0,0)(0,0,1)_{25}$

```
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6     sma1  intercept
      0.7786  0.1124  0.0389  -0.0263  0.0579  -0.1346  -0.0383     0.0001
s.e.  0.0519  0.0663  0.0668   0.0667  0.0671   0.0526   0.0523     0.0029

sigma^2 estimated as 0.000104:  log likelihood = 1158.24,  aic = -2300.48
```

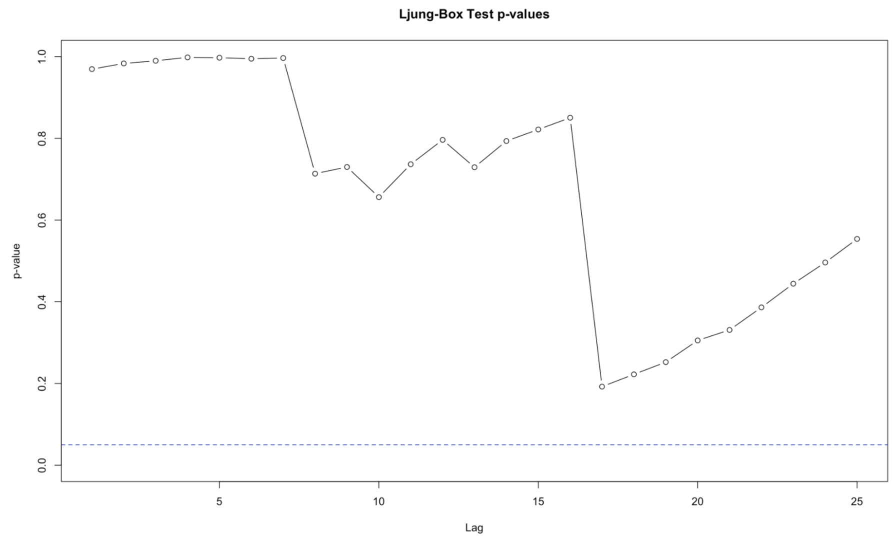

Figure 6-9: Ljung-Box Test Results for SARIMA$(6, 0, 0)(0, 0, 1)_{25}$

6.8 SARIMA$(6, 0, 0)(0, 0, 1)_{29}$

```
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6     sma1  intercept
      0.7769  0.1063  0.0424  -0.0184  0.0540  -0.1389  -0.0847     0.0001
s.e.  0.0519  0.0663  0.0666   0.0668  0.0666   0.0526   0.0540     0.0027

sigma^2 estimated as 0.0001034:  log likelihood = 1159.21,  aic = -2302.41
```

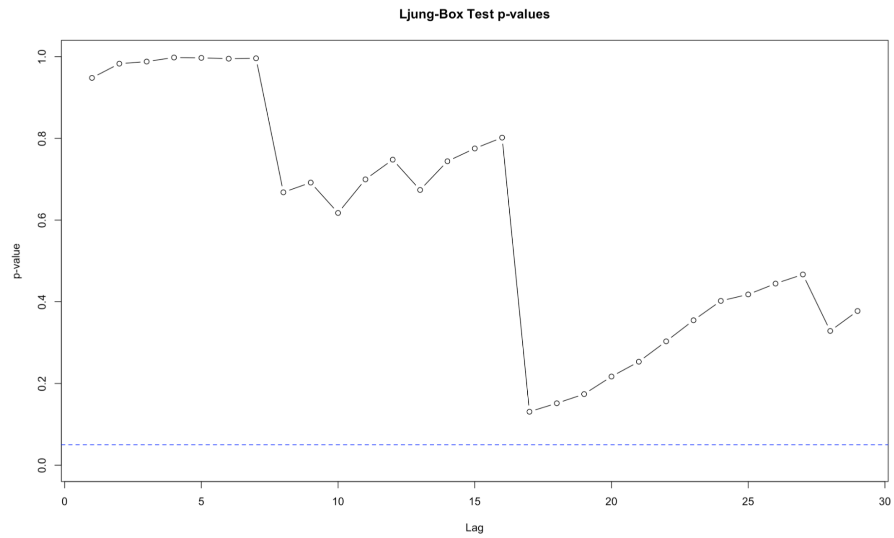

Figure 6-10: Ljung-Box Test Results for SARIMA$(6,0,0)(0,0,1)_{29}$

6.9 SARIMA$(9,0,0)(0,0,1)_{50}$

```
\begin{minted}[baselinestretch=1, fontsize=\small]{R}
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6     ar7     ar8      ar9
      0.7820  0.1213  0.0207  -0.0178  0.0477  -0.1522  0.0064  0.1286  -0.1129
s.e.  0.0522  0.0664  0.0668   0.0663  0.0666   0.0669  0.0668  0.0668   0.0527

        sma1   intercept
      0.0416      0.0001
s.e.  0.0513      0.0031

sigma^2 estimated as 0.0001024:  log likelihood = 1161.01,  aic = -2300.02
```

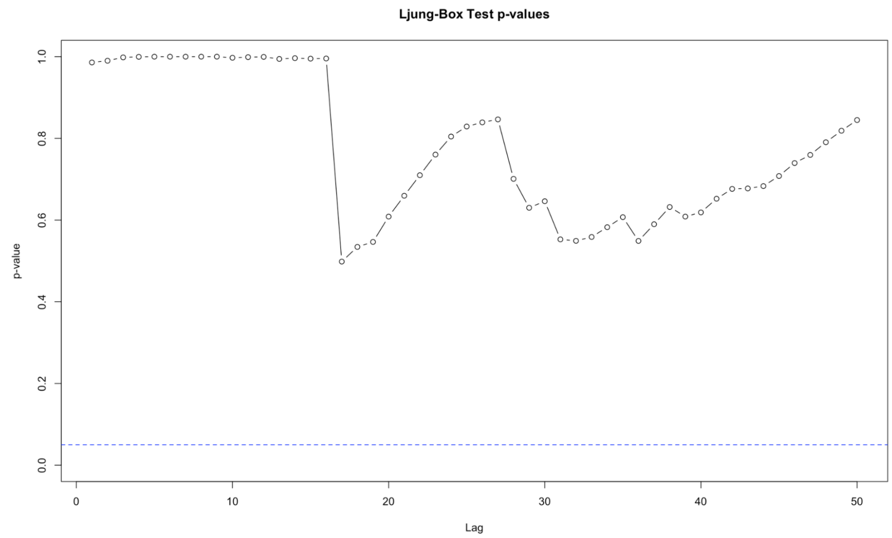

Figure 6-11: Ljung-Box Test Results for SARIMA$(9,0,0)(0,0,1)_{50}$

6.10 SARIMA$(9,0,0)(0,0,1)_{25}$

```
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6     ar7     ar8      ar9
      0.7842  0.1177  0.0195  -0.0187  0.0481  -0.1485  0.0062  0.1244  -0.1132
s.e.  0.0521  0.0664  0.0670   0.0663  0.0669   0.0666  0.0669  0.0671   0.0527

         sma1   intercept
      -0.0298      0.0001
s.e.   0.0530      0.0028

sigma^2 estimated as 0.0001025:  log likelihood = 1160.84,  aic = -2299.68
```

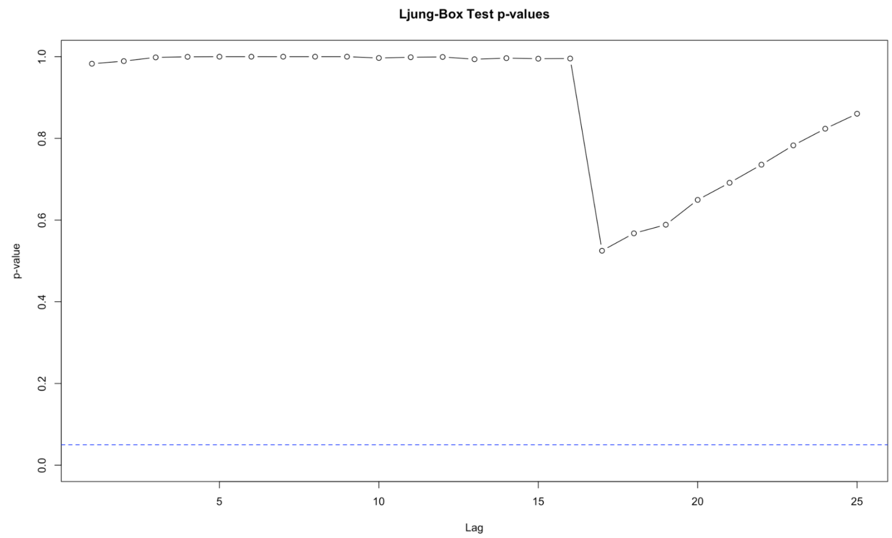

Figure 6-12: Ljung-Box Test Results for SARIMA$(9,0,0)(0,0,1)_{25}$

6.11 SARIMA$(9,0,0)(0,0,1)_{29}$

```
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6     ar7     ar8      ar9
      0.7820  0.1213  0.0207  -0.0178  0.0477  -0.1522  0.0064  0.1286  -0.1129
s.e.  0.0522  0.0664  0.0668   0.0663  0.0666   0.0669  0.0668  0.0668   0.0527

         sma1   intercept
      -0.0852      0.0001
s.e.   0.0547      0.0026
sigma^2 estimated as 0.0001024:  log likelihood = 1161.01,  aic = -2300.02
```

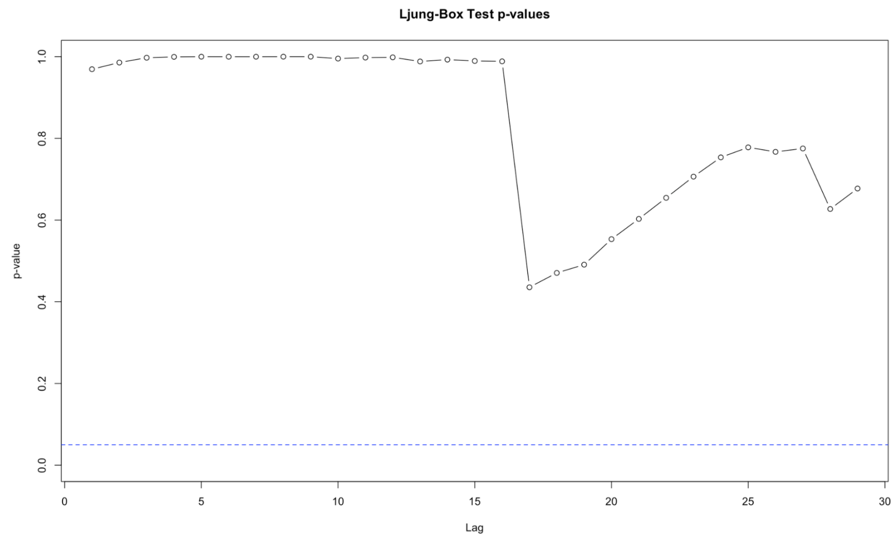

Figure 6-13: Ljung-Box Test Results for SARIMA$(9,0,0)(0,0,1)_{29}$

6.12 SARIMA$(18,0,0)(0,0,1)_{50}$

```
Coefficients:
        ar1     ar2     ar3      ar4     ar5      ar6     ar7     ar8      ar9
      0.783  0.1281  0.0179  -0.0305  0.0478  -0.1537  0.0047  0.1304  -0.1126
s.e.  0.052  0.0662  0.0666   0.0666  0.0670   0.0671  0.0669  0.0672   0.0674
       ar10     ar11     ar12     ar13    ar14    ar15     ar16     ar17
      0.0752  -0.0423  -0.0248  -0.0624  0.0986  0.0241  -0.0549  -0.1474
s.e.  0.0676   0.0677   0.0675   0.0669  0.0674  0.0673   0.0669   0.0666
       ar18    sma1   intercept
      0.1308  0.0628     0.0001
s.e.  0.0529  0.0529     0.0029

sigma^2 estimated as 9.825e-05:  log likelihood = 1168.17,  aic = -2296.35
```

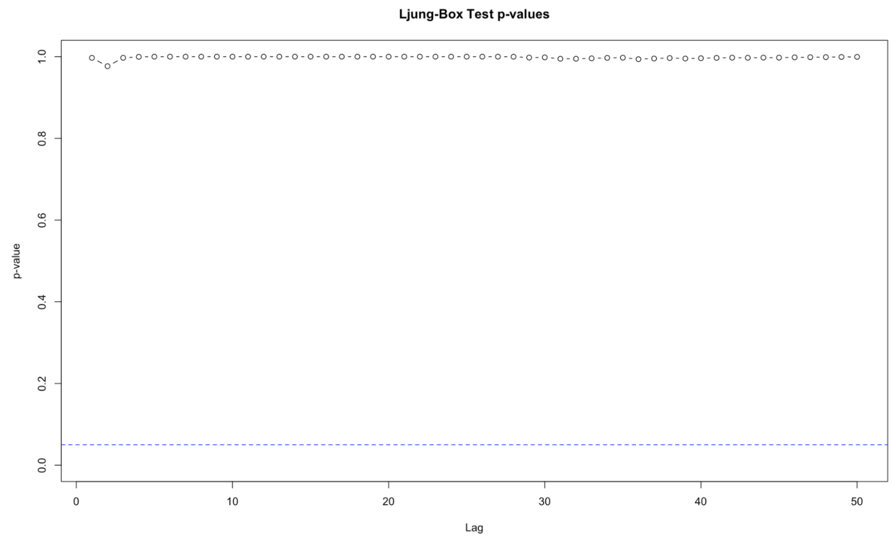

Figure 6-14: Ljung-Box Test Results for SARIMA$(18, 0, 0)(0, 0, 1)_{50}$

6.13 SARIMA$(18, 0, 0)(0, 0, 1)_{25}$

```
Coefficients:
         ar1     ar2     ar3      ar4     ar5      ar6     ar7     ar8      ar9
      0.7863  0.1223  0.0182  -0.0303  0.0432  -0.1453  0.0039  0.1265  -0.1090
s.e.  0.0520  0.0663  0.0669   0.0667  0.0671   0.0668  0.0670  0.0675   0.0675
        ar10     ar11     ar12     ar13    ar14    ar15     ar16     ar17
      0.0745  -0.0495  -0.0232  -0.0596  0.0878  0.0336  -0.0547  -0.1408
s.e.  0.0676   0.0676   0.0676   0.0670  0.0669  0.0670   0.0671   0.0671
        ar18     sma1  intercept
      0.1220  -0.0230     0.0001
s.e.  0.0524   0.0532     0.0026

sigma^2 estimated as 9.864e-05:  log likelihood = 1167.56,  aic = -2295.12
```

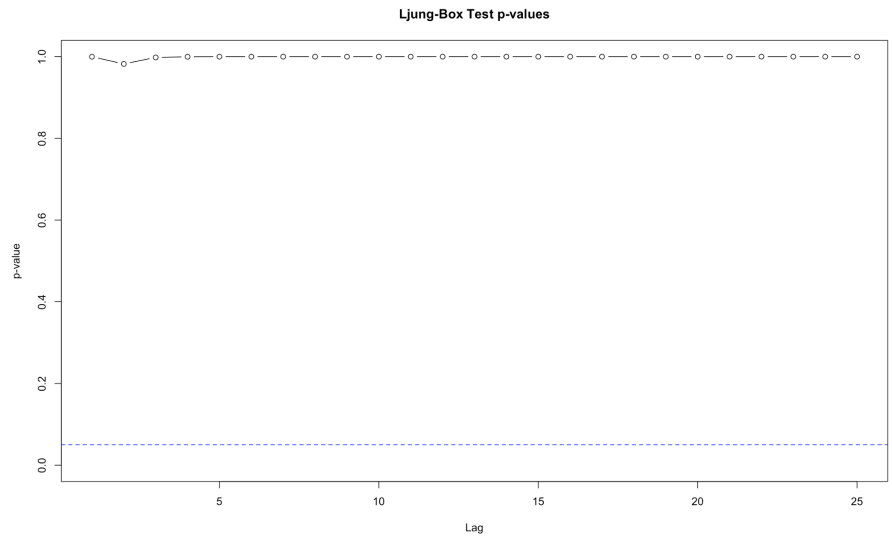

Figure 6-15: Ljung-Box Test Results for SARIMA$(18, 0, 0)(0, 0, 1)_{25}$

6.14 SARIMA$(18, 0, 0)(0, 0, 1)_{29}$

```
Coefficients:
        ar1     ar2     ar3      ar4     ar5      ar6      ar7     ar8      ar9
      0.781  0.1184  0.0174  -0.0224  0.0411  -0.1458  -0.0011  0.1298  -0.1085
s.e.  0.052  0.0660  0.0664   0.0665  0.0664   0.0664   0.0666  0.0668   0.0670
        ar10     ar11     ar12     ar13    ar14    ar15     ar16     ar17
      0.0714  -0.0475  -0.0416  -0.0515   0.0934  0.0356  -0.0625  -0.1425
s.e.  0.0672   0.0670   0.0678   0.0666   0.0665  0.0663   0.0667   0.0663
        ar18     sma1   intercept
      0.1307  -0.1111      0.0001
s.e.  0.0524   0.0562      0.0023

sigma^2 estimated as 9.755e-05:  log likelihood = 1169.42,  aic = -2298.85
```

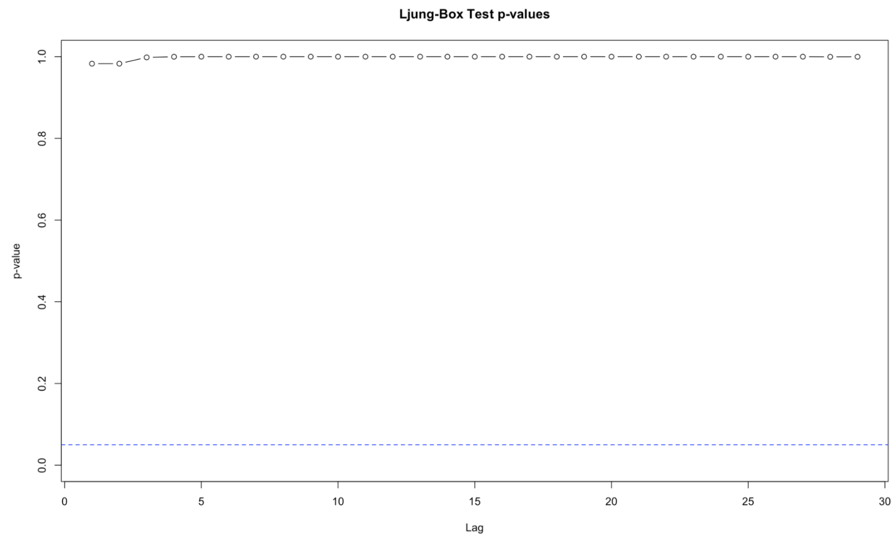

Figure 6-16: Ljung-Box Test Results for $\text{SARIMA}(18,0,0)(0,0,1)_{29}$

From figure 6-5 to 6-16, we can see that the Ljung-Box test of $\text{SARIMA}(1,0,0)(0,0,1)_{50}$, $\text{SARIMA}(1,0,0)(0,0,1)_{25}$ and $\text{SARIMA}(1,0,0)(0,0,1)_{29}$ do not perform well as some of the points are under the significance line. As a result, here we compare the rest of the model with their AIC and BIC. The information are shown in table 6-2.

Table 6-2: AIC and BIC comparison

Model	AIC	BIC
$\text{SARIMA}(6,0,0)(0,0,1)_{50}$	-2300.64	-2263.513
$\text{SARIMA}(6,0,0)(0,0,1)_{25}$	-2300.48	-2263.358
SARIMA$(6,0,0)(0,0,1)_{29}$	**-2302.41**	**-2265.288**
$\text{SARIMA}(9,0,0)(0,0,1)_{50}$	-2300.02	-2251.187
$\text{SARIMA}(9,0,0)(0,0,1)_{25}$	-2299.68	-2250.844
$\text{SARIMA}(9,0,0)(0,0,1)_{29}$	-2301.79	-2252.959
$\text{SARIMA}(18,0,0)(0,0,1)_{50}$	-2296.35	-2212.392
$\text{SARIMA}(18,0,0)(0,0,1)_{25}$	-2295.12	-2211.166
$\text{SARIMA}(18,0,0)(0,0,1)_{29}$	-2298.85	-2214.891

We can see that $\text{SARIMA}(6,0,0)(0,0,1)_{29}$ performs the best as it has the smallest AIC and BIC. Then we conduct a coefficient t test on this model

H0: The coefficient has no significant impact on the model
H1: The coefficient has a significant impact on the model

Table 6-3: Coefficient Testing of Highest-order Terms - SARIMA$(6,0,0)(0,0,1)_{29}$

	Coefficients	StdError	t-value	p-value
ar1	7.769148e-01	0.05189058	14.97217476	0.000000000
ar2	1.063088e-01	0.06633028	1.60271896	0.108996716
ar3	4.235612e-02	0.06659829	0.63599415	0.524780234
ar4	-1.840275e-02	0.06675721	-0.27566688	0.782803934
ar5	5.397994e-02	0.06664345	0.80998117	0.417950998
ar6	-1.389028e-01	0.05258071	-2.64170736	0.008248929
sma1	-8.467555e-02	0.05399169	-1.56830717	0.116809464
intercept	9.458139e-05	0.00273756	0.03454952	0.972438953

- ar1 : p-value = $0.000000000 < 0.05 \rightarrow$ reject H0
- ar2 : p-value = $0.108996716 > 0.05$
- ar3 : p-value = $0.524780234 > 0.05$
- ar4 : p-value = $0.524780234 > 0.05$
- ar5 : p-value = $0.417950998 > 0.05$
- ar6 : p-value = $0.008248929 < 0.05 \rightarrow$ reject H0
- sma1 : p-value = $0.116809464 > 0.05$
- intercept : p-value = $0.972438953 > 0.05$

We can see from the result that the p-values of ar1 and ar6 are smaller than 0.05, which means they have better contribution to the model.