# Unemployment Rate (Aged 15 and Over)

# All Persons for the Republic of Korea from 2010 to 2017

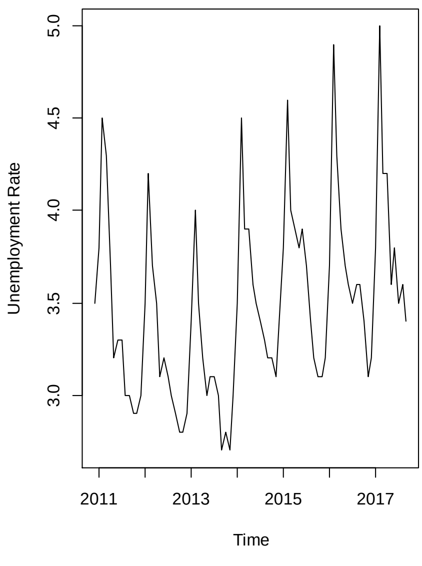
Source: Economic Research

Units:  Units: Percent

Frequency:  Monthly, Not Seasonally Adjusted

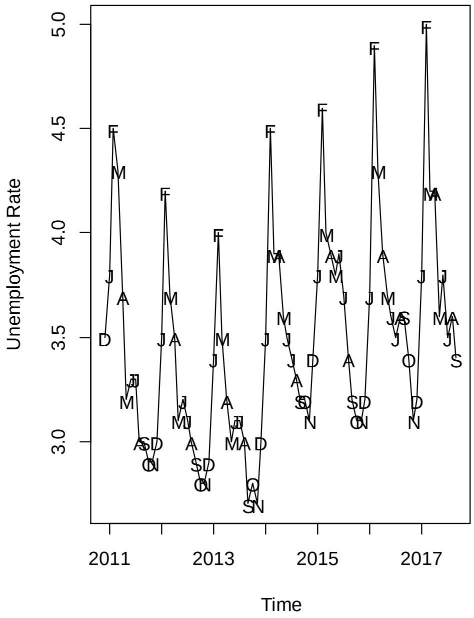
In this paper, we will analyze a data set consisting "unemployment Rate of aged 15 and Over for all Persons for the Republic of Korea" from "December 2010" to "September 2017". Using this data, we hope to first find a model to explain the nature behavior of the dataset and then we test our suggested models for finding the best one among them. By considering ACF of the data, we would be able to catch the strong autocorrelation at different lags in the data that amounts to suggesting appropriate models. We examine the early lags (1, 2, 3, …) to determine non-seasonal terms.  Pins in the ACF (at low lags) demonstrate non-seasonal MA terms.  Pins in the PACF (at low lags) indicate possible non-seasonal AR terms. Also we examine the patterns across lags that are multiples of 12 and for our monthly data, we will look at lags 12, 24, 36, and etc. There isn't any need to look at much more than the first two or three seasonal multiples. We will Judge the ACF and PACF at the seasonal lags in the same way, we do for the earlier lags. The best model is the one that has a more Log likelihood and less AIC criterion. Also we will be able to test our models. We will do this by considering the residuals behavior.

```
> data <- read.csv("E:/SIU/Time Series/Project/Main/korea/korea.csv",header=T)

> data1 <- data$rate

> ae<- ts(data1, frequency=12, start=c(2010,12))

> plot(window(ae,start=c(2010,12)),ylab='Unemployment Rate')
```



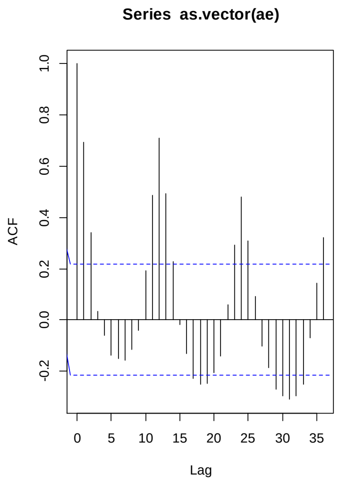

This plot demonstrates the monthly unemployment rates from December 2010 through September 2017. From December 2013 to September 2017, there is a strong upward trend and also a seasonality but this is not the case from December 2010 to December 2013. This trend alone would lead us to specify a nonstationary model. We can see better and with the more details in the next plot.

```
> Month=c('D','J','F','M','A','M','J','J','A','S','O','N')

> points(window(ae,start=c(2010,12)),pch=Month)
```



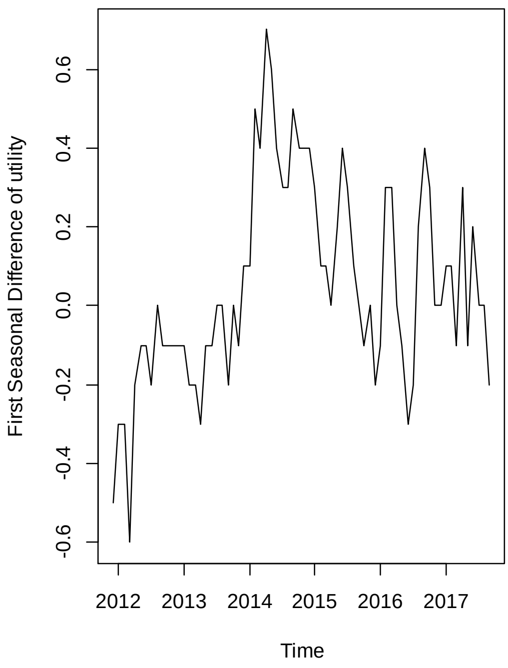

As we see in the plot, unemployment rates are higher during the winter months (especially February) and much lower in the summer (especially September). We can consider deterministic seasonal models such as linear model but I found out that such models do not explain the behavior of this time series. I considered this models and wrote the codes but for avoiding describing long story I didn't bring the plots here. In the rest of this article, I will consider that the stochastic seasonal models developed here do work well for this series.

```
>  acf(as.vector(ae),lag.max=36)
```



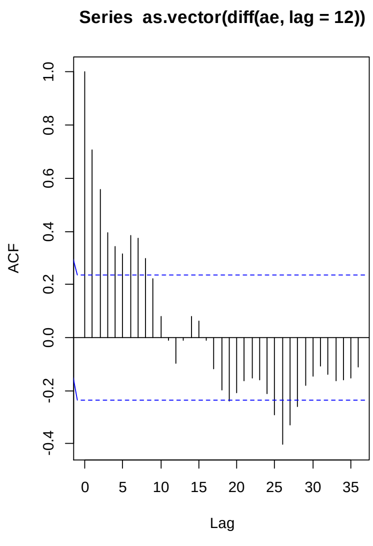

The plot demonstrates the sample ACF for our series. In this plot, the seasonal autocorrelation relationships are eye-catching, especially the strong correlation at lags 12, 24, 36, and etc. Also, maybe there is other correlation that needs to be modeled. So, we start doing our series Stationary through differencing.

```
> plot(diff(ae,lag=12),ylab='First Seasonal Difference of utility',xlab='Time')
```



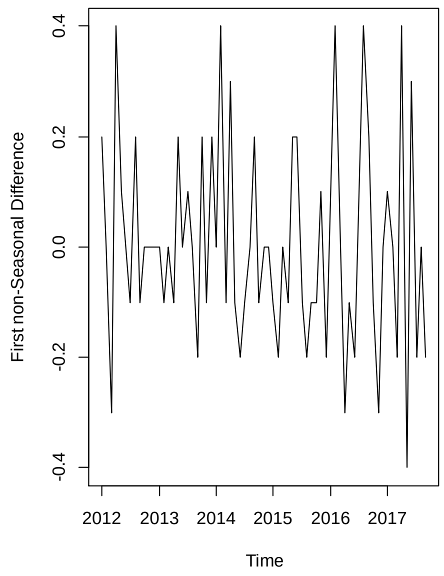

This plot demonstrates the time series plot of the unemployment rates after we take a first seasonal difference. We can see that after 2015 the upward trend has been disappeared but the strong seasonality is still persist. This is while that it is not the case for the before 2015.
 We can see this behavior also in the ACF of the first seasonal difference at below.

```
> acf(as.vector(diff(ae,lag=12)),lag.max=36)
```



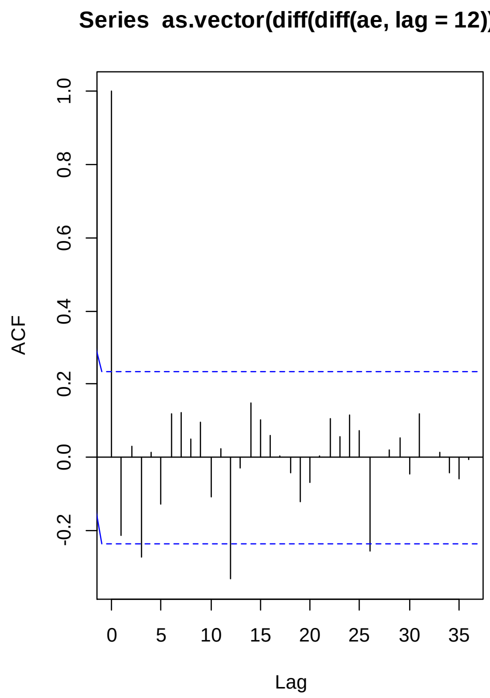

Maybe the non-seasonal difference can make our series stationary. We are going to try it here:

plot(diff(diff(ae,lag=12))),ylab='First non-Seasonal Difference,xlab='Time')



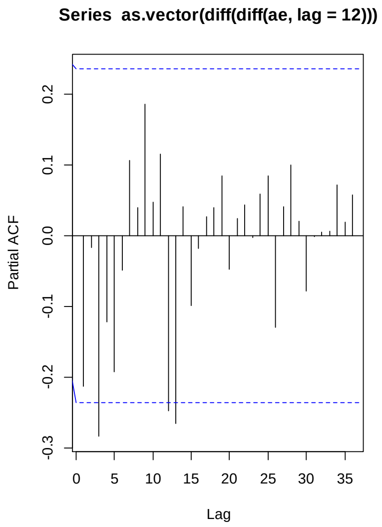

The plot demonstrates the time series plot of the unemployment rates after taking both a first difference and a seasonal difference. It seems that almost seasonality and upward and downward trend is disappeared and we have a better series now in compare with the previous plot.

> acf(as.vector(diff(diff(ae,lag=12))),lag.max=36)




The plot of ACF (after two differencing) demonstrates that very small autocorrelation leftover in the series after these two differences. According to the plot of ACF, we can see strong correlation at lags 1,3, 12 and 24. Pins in the ACF (at 3 low lags) indicate non-seasonal MA (3) terms and also, pins in the ACF (at lags 12 and 24) indicates seasonal MA (2) terms.

```
> pacf(as.vector(diff(diff(ae,lag=12))),lag.max=36)
```




According to the plot of PACF (after two differencing), we can see strong correlation at lags 3, 12 and 13. Pins in the PACF (at 3 low lags) indicate non-seasonal AR (3) terms and also, pin in the ACF (at lags 12) indicates seasonal AR (1) terms. So, We will consider specifying the multiplicative, SARMA (3,1,3) $\times$ $(1,1,2)_{12}$ model. Also, since the second lag in ACF and PACF are not strong and also for reducing parameter dimensions and as a result parameter reduction, we suspect to another SARMA model that is the multiplicative, SARMA (1,1,1) $\times$ $(1,1,2)_{12}$ model. Thus, we will consider tow following models:

Model 1: SARMA (3,1,3) $\times$ $(1,1,2)_{12}$

Model 2: SARMA (1,1,1) $\times$ $(1,1,2)_{12}$

Model 1 is shown first below. Before fitting our suggested models, we can test the stationary of our series after two seasonal and non-seasonal differences by "Dicky Fuller Unit-Root Test".

```
> CADFtest(diff(diff(ae,lag=12)))
```

ADF test

data: diff(diff(ae, lag = 12))

ADF(1) = -6.3814, p-value =

5.709e-06

alternative hypothesis: true delta is less than 0

sample estimates:

```
   delta 
-1.263304 
```

According to the Dicky Fuller Unit-Root Test, with p-value very close to 0 we can be sure that oure series is stationary.

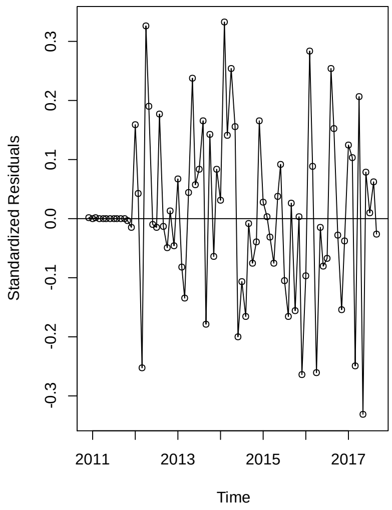
With having suggested model SARMA $(3,1,3) \times (1,1,2)_{12}$, we can proceed to parameter estimation of our model.

```
> sarma=arima(ae,order=c(3,1,3),seasonal=list(order=c(1,1,2),period=12))
> sarma

Call:
arima(x = ae, order = c(3, 1, 3), seasonal = list(order = c(1, 1, 2), period = 12))

Coefficients:
        ar1     ar2      ar3      ma1
     0.4777  0.1346  -0.6992  -0.8399
s.e.  0.2388  0.2743   0.1871   0.2858
        ma2     ma3     sar1    sma1
     0.0536  0.5071  -0.5231  0.1744
s.e.  0.3867  0.2923   0.3607  0.3589
       sma2
     0.0839
s.e.  0.2485

sigma^2 estimated as 0.02148:  log likelihood = 32.46,  aic = -44.93
```

The coefficient estimates are all highly significant, and the next step is to check the model.

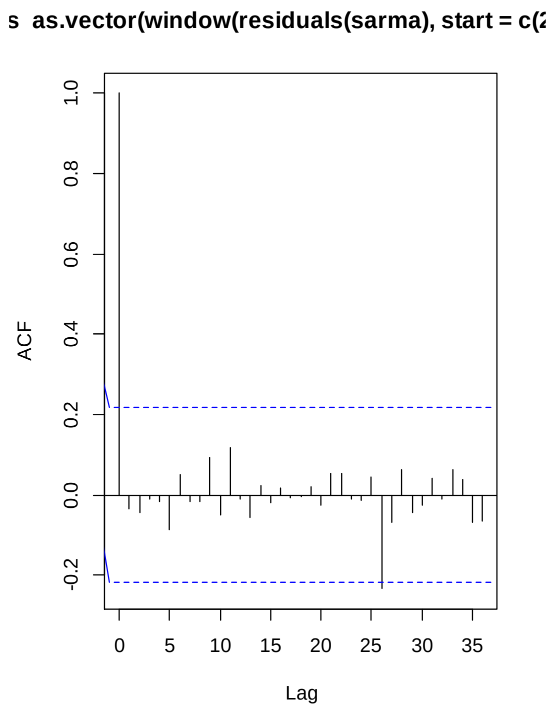
To check the estimated monthly SARMA $(3,1,3) \times (1,1,2)_{12}$ model, we first look at the time series plot of the residuals.

```
> plot(window(residuals(sarma),start=c(2010,12)),ylab='Standardized Residuals',type='o')
> abline(h=0)
```



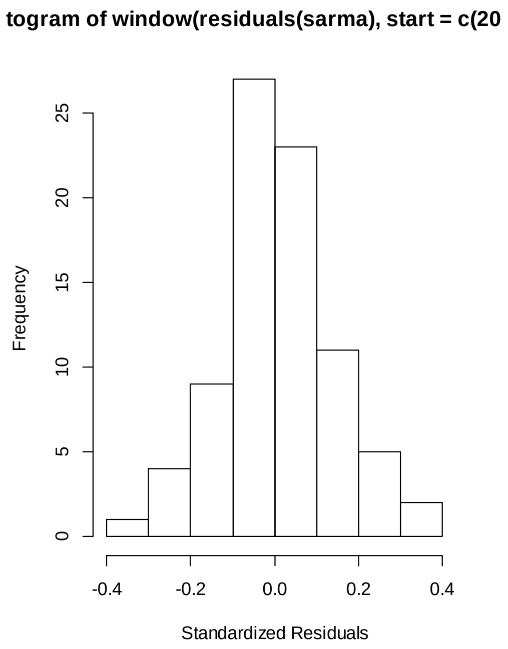

This plot gives us standardized residuals. The plot does not suggest any main disorder with the model (except some strange behavior in the middle and end of the series). Also, we may need to consider the model further for outliers.

> acf(as.vector(window(rstandard(sarma),start=c(2010,12))),lag.max=36)



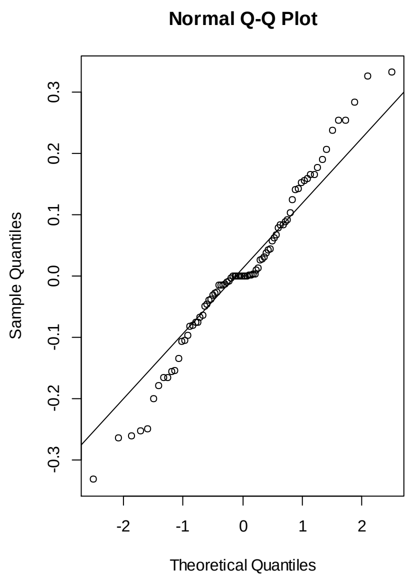

Here, we plot the sample ACF of the residuals. There isn't any “statistically significant” correlation here except at lag 26. Also, we can say there isn't any reasonable answer for this dependence at lag 26. All in all, we should not be shocked that one autocorrelation out of the 36 demonstrated is strong. Also, we can consider normality of the error terms by the residuals.

> hist(window(rstandard(sarma),start=c(2010,12)),xlab='Standardized Residuals')




This plot demonstrates the histogram of the residuals. The shape of the plot is to some extant bell-shaped but definitely not exactly bell-shaped. Also we can get more information about residuals from a quantile-quantile plot.

```
> qqnorm(window(residuals(sarma),start=c(2010,12)))
> qqline(window(residuals(sarma),start=c(2010,12)))
```




According to the quantile-quantile plot, again we can see some outlier (not very far from the straight line) in the lower tail. As a statistical test, we can use the Shapiro-Wilk test of normality:

```
> shapiro.test(residuals(sarma))
```

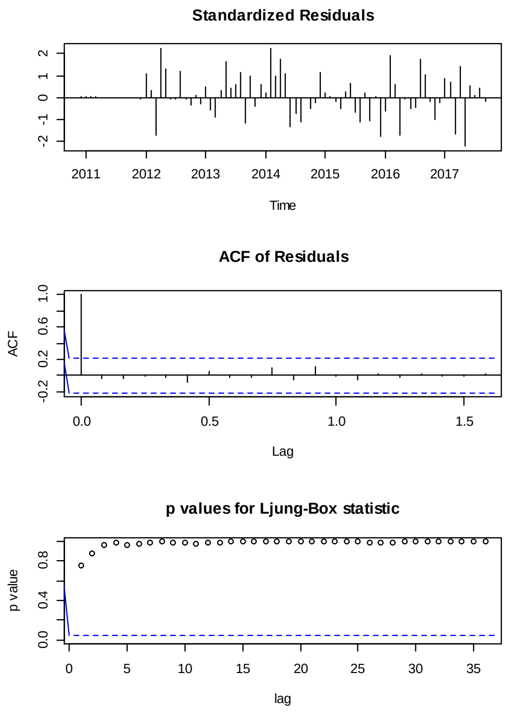
Shapiro-Wilk normality test

data: residuals(sarma)

W = 0.97799, p-value = 0.1712

According to the result of output we can see, it has a test statistic of W = 0.97799, leading to a p-value of 0.1712. Thus normality is not rejected at any of the usual significance levels. Also, we can use Ljung-Box test for this model.

> tsdiag(sarma,gof=36,omit.initial=F)




According to the Ljung-Box test for this model easily we can see a further indication that the model SARMA (3,1,3) × $(1,1,2)_{12}$ has grabbed the dependence in the time series.

For another suggested model SARMA (1,1,1) × $(1,1,2)_{12}$, we can do all the steps and for parameter estimation of the model we have:

> sarma2=arima(ae,order=c(1,1,1),seasonal=list(order=c(1,1,2),period=12))

> sarma2

Call:

arima(x = ae, order = c(1, 1, 1), seasonal = list(order = c(1, 1, 2), period = 12))

Coefficients:

ar1 ma1 sar1 sma1

0.3438 -0.6730 -0.3603 -0.1056

s.e. 0.2108 0.1587 0.5547 0.5702

sma2

-0.0029

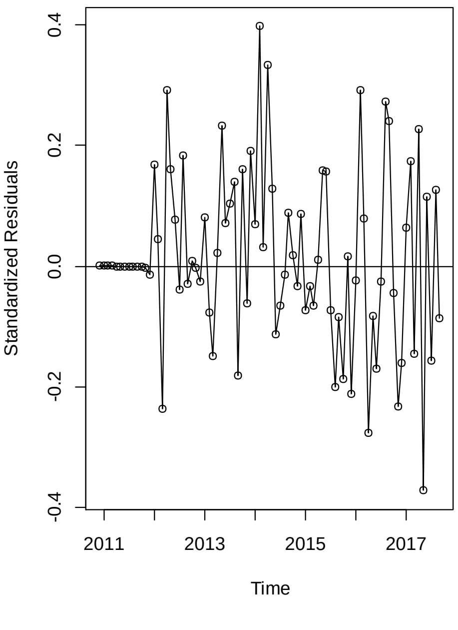
s.e. 0.2957

sigma^2 estimated as 0.02475: log likelihood = 28.26, aic = -44.52

Again, the coefficient estimates are all highly significant. Compare with the first model, the log likelihood of the first model (SARMA (3,1,3) × $(1,1,2)_{12}$) more than second model (SARMA (1,1,1) × $(1,1,2)_{12}$) with rate 32.46 to 28.26, respectively. Another usefull criterion is AIC that the AIC of the first model (SARMA (3,1,3) × $(1,1,2)_{12}$) less than second model (SARMA (1,1,1) × $(1,1,2)_{12}$) with rate -44.52, -44.93 to 28.26, respectively. However, according to the AIC criteria, they are very close to each other.

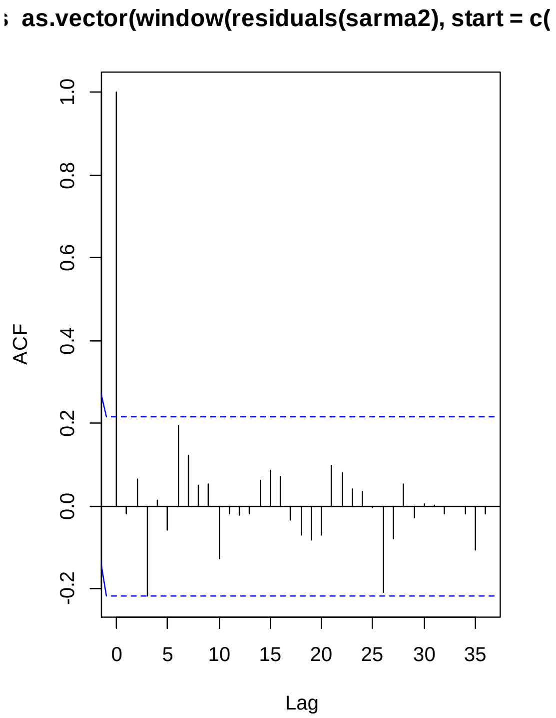
the next step is to check the second model (SARMA (1,1,1) × $(1,1,2)_{12}$). To check the estimated the monthly SARMA (1,1,1) × $(1,1,2)_{12}$ model, we first look at the time series plot of the residuals.

```
> plot(window(residuals(sarma2),start=c(2010,12)),ylab='Standardized Residuals',type='o')
> abline(h=0)
```



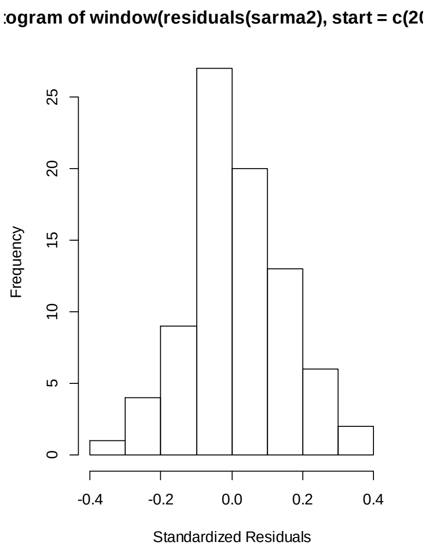
This plot gives us plot for standardized residuals. The plot does not suggest any main disorder with the model (except some strange behavior in the middle and end of the series). Also, we may need to consider the model further for outliers.

```
> acf(as.vector(window(rstandard(sarma2),start=c(2010,12))),lag.max=36)
```




Here, we plot the sample ACF of the residuals. There are two strong "statistically significant" correlation here. Also, we can consider normality of the error terms by the residuals.

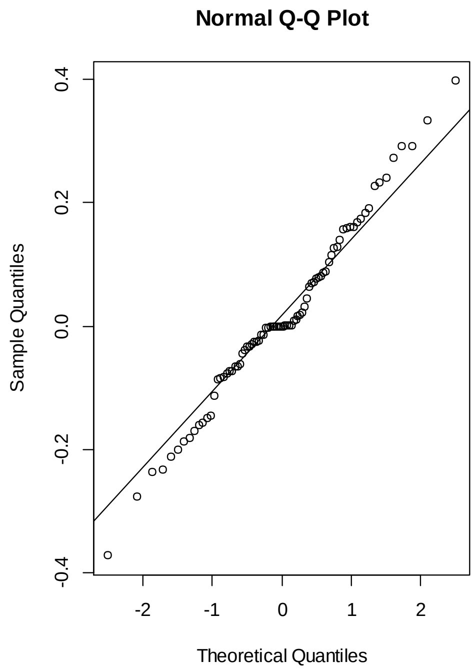
```
> hist(window(rstandard(sarma2),start=c(2010,12)),xlab='Standardized Residuals')
```




This plot demonstrates the histogram of the residuals. Also we can get more information about residuals from a quantile-quantile plot.

> qqnorm(window(residuals(sarma2),start=c(2010,12)))

> qqline(window(residuals(sarma2),start=c(2010,12)))




As a statistical test, we can use the Shapiro-Wilk test of normality.

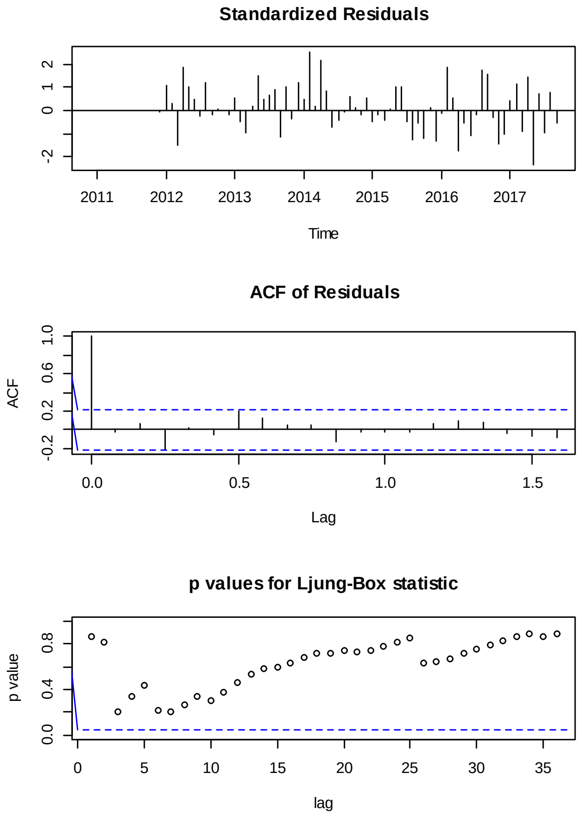
> shapiro.test(residuals(sarma2))

Shapiro-Wilk normality test

data: residuals(sarma2)
W = 0.98624, p-value = 0.532

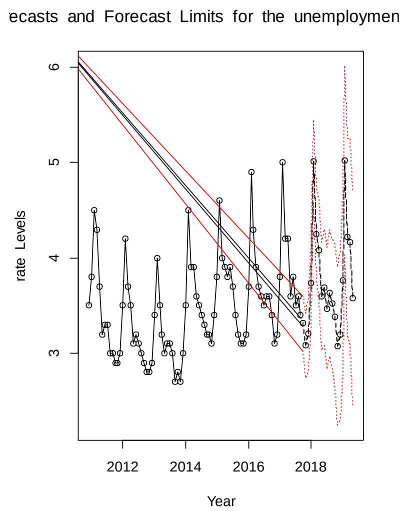
According to the result of output we can see, it has a test statistic of W = 0.98624, leading to a p-value of 0.532. Thus normality is not rejected at any of the usual significance levels. Also, we can use Ljung-Box test for this model.

tsdiag(sarma2,gof=36,omit.initial=F)




According to the Ljung-Box test for this model easily we can see a further indication that the model SARMA (1,1,1) × $(1,1,2)_{12}$ has grabbed the dependence in the time series. As a result, we can say that the SARMA (1,1,1) × $(1,1,2)_{12}$ model did well but in compare with SARMA (3,1,3) × $(1,1,2)_{12}$ model, the later one has a better performance and cached more dependencies in our dataset. The next and last step is seasonal forecasting.

```
plot(sarma, n1 = c(2010,12), n.ahead = 20, col = "red", xlab = "Year", type = "o",ylab =
expression(rate~~Levels),

 main = expression(Forecasts~~and~~Forecast~~Limits~~"for"~~the~~unemployment~~Model))
```



This plot demonstrates the forecasts and 95% forecast limits for a lead time of 20 months for the SARMA $(3,1,3) \times (1,1,2)_{12}$ model that we fit at before. The forecasts follow the stochastic periodicity in the data very well, and the forecast limits give a good feeling for the accuracy of the forecasts.