

Forecasting Chickenpox with SARIMA Model

Jianing (Julia) Chen

2/21/2020

Abstract

The present study aims to identify a model best fitting the chickenpox breakout in New York City from 1957 to 1971 in where the chickenpox vaccine has not yet introduced in the United States. Prior to the introduction of the vaccine in 1995 in the United States, every year there was a lot of gets infected. This project utilizes seasonal autoregressive integrated moving average (SARIMA) models to delineate disease patterns in the population over time, examine and quantify the evolution of the infectious diseases, and accurately forecast chickenpox 12 months into the future.

Contents:

1. Introduction:
 - 1.1 History
 - 1.2 About the Dataset
2. Analysis:
 - 2.1 Overview of the Dataset
 - 2.2 Data Transformations
 - 2.2.1 ACF and PACF Analysis of transformed dataset
 - 2.2.2 Stationarity and differencing
 - 2.3 Model Selection
 - 2.3.1 Model Candidates
 - 2.3.2 Determination of P
 - 2.3.3 Determination of q
 - 2.3.4 Determination of Final Model
 - 2.4 Model Diagnostics
 - 2.5 Forecasting and Verification
3. Conclusion
4. Appendix

1.Introduction

1.1 History Chickenpox is a contagious disease, although the two-dose varicella vaccination program started in the United State and possibly could reduce the chances to get infected, the outbreaks still occur even in settings such as schools where most children are vaccinated. So it is worth exploring a model using the past data to predict the future.

1.2 About the Dataset The Dataset I choose for this project is a heath related dataset from the TSDL library. This is a monthly dataset and recorded the reported number of chickenpox in New York City from 1931 to 1972. It originally contains 498 months, but I take the most recent 180 months to analyze, so for the project the dataset covers from 1956 to 1970.

2. Analysis

2.1 Overview of the dataset

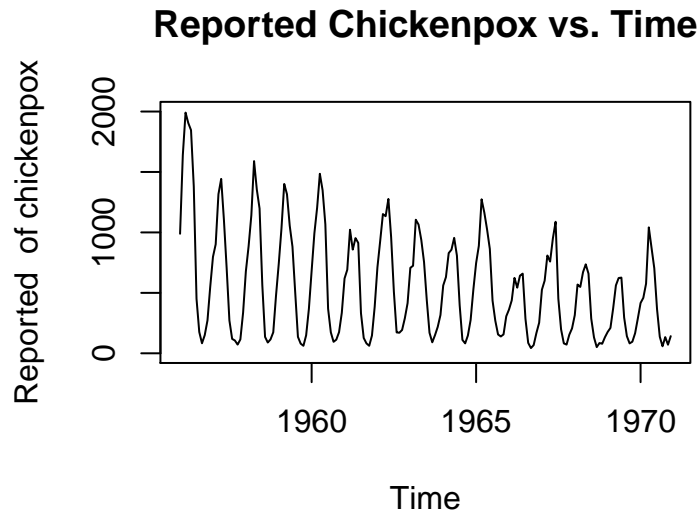


Figure 1: Reported Chickenpox

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      43.0   156.0   451.5   556.9   865.5  1991.0
```

By exploring and analyzing this dataset in Figure 1, there is a seasonal pattern which means some month within the year, which means people may easier get infected by the disease. Besides, there is a clear up (max 1991) and down (min 43) change over time with same time duration, the variance of the original dataset is:

```
## [1] 194816.7
```

Since the range of the reported chickenpox is quite large, and the variance (194816.7). In order to stabilize the variance, I would like to transform the dataset.

2.2 Data Transformations: Box-Cox transformation

```
## [1] 0.2626263
```

```
## [1] 20.31288
```

By calling the boxcox function, the lambda is 0.2626263. It is close to 0. Figure 2 shows that it is not within the 95% confident interval, so I will use box cox transformation instead of log transformation. After transformed, the variance is down to the 20.31288. The time series plot of transformed data is shown below.

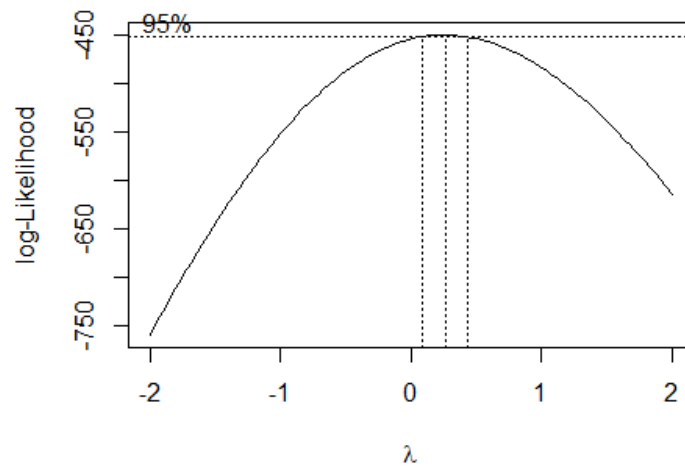
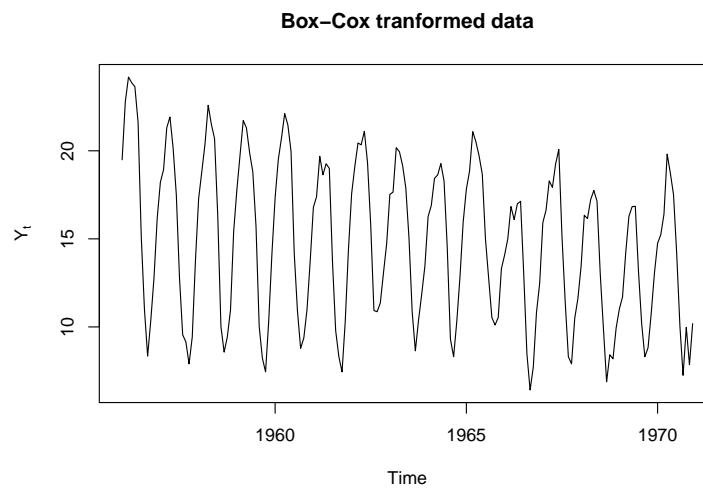
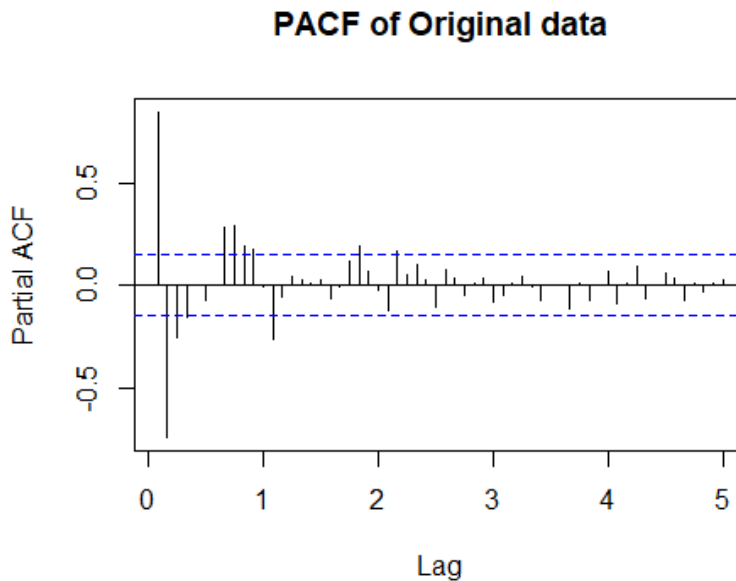
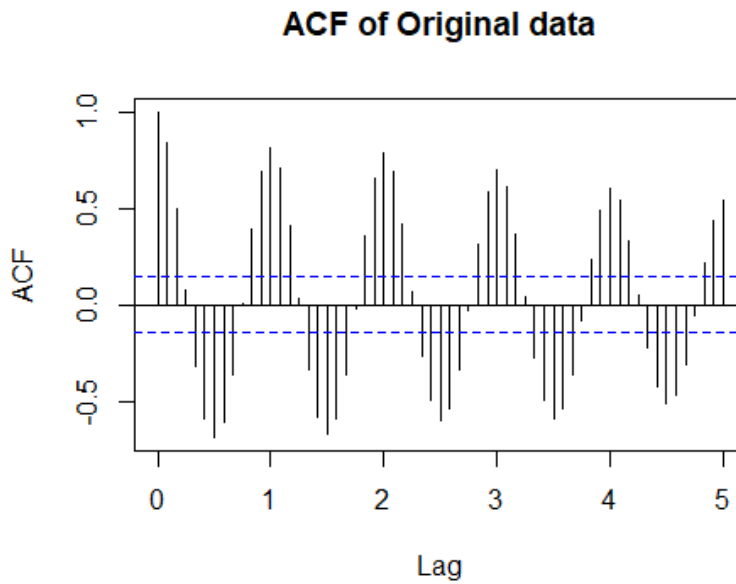


Figure 2: BoxCox Power Transformation



Looking at the graph, the dataset does not change a lot based on the pattern, but the y-axis's range changed. After I get the dataset transformed, it is time to look at ACF and PACF to determine a fit model.

2.2.1 ACF and PACF Analysis of transformed dataset



From ACF, I could see the seasonality is at lag 12, so I will difference it at lag 12 first to remove the seasonal part. Then take difference at lag 1 for unknown times to remove the trend. In order to avoid over difference, each time I calculate the variance.

2.2.2 Stationarity and differencing First of all, I take a difference at lag 12 to remove the yearly seasonality, the variance reduced to:

```
## [1] 3.723613
```

Then I start taking the difference at lag 1, the variance I get at the first time is:

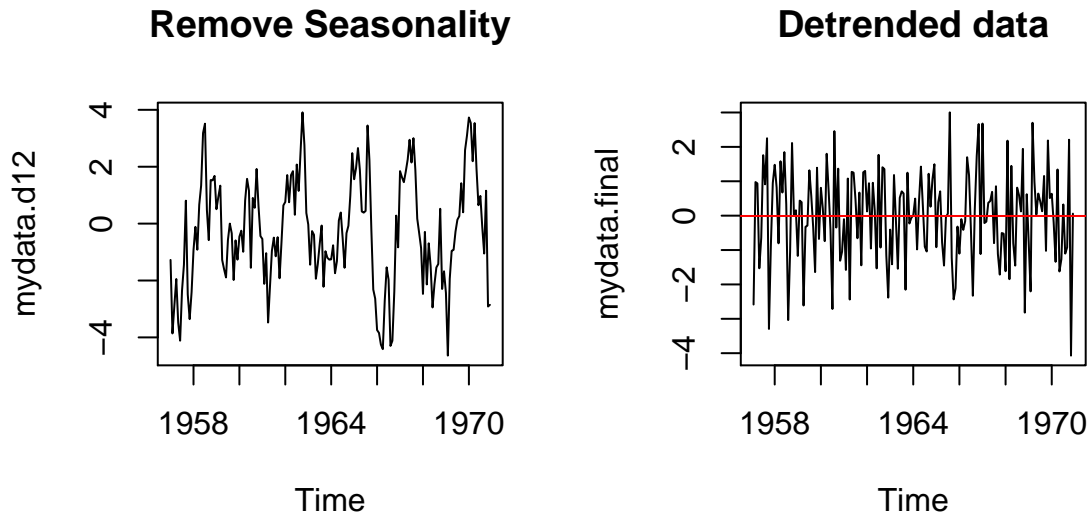
```
## [1] 1.833708
```

When I take the second difference at lag 1, the variance turns out to be:

```
## [1] 4.293459
```

So the second difference at lag 1 is not considered because it gives me a larger variance which is 4.293459, it may cause overdifference. I end up with 1 difference for both seasonal and stational part. (D=1, d=1)

Below will be the removed seasonality and detrend graph:



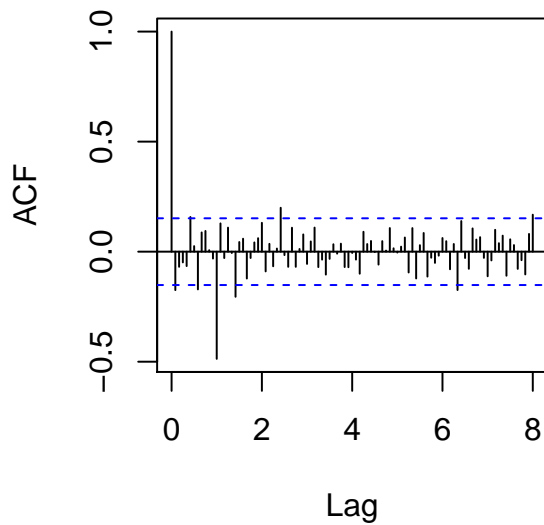
Then call Kwiatkowski-Phillips-Schmidt-Shin (KPSS) to test the stationarity.

```
## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo
##
## KPSS Test for Level Stationarity
##
## data:  mydata.final
## KPSS Level = 0.041834, Truncation lag parameter = 4, p-value = 0.1
```

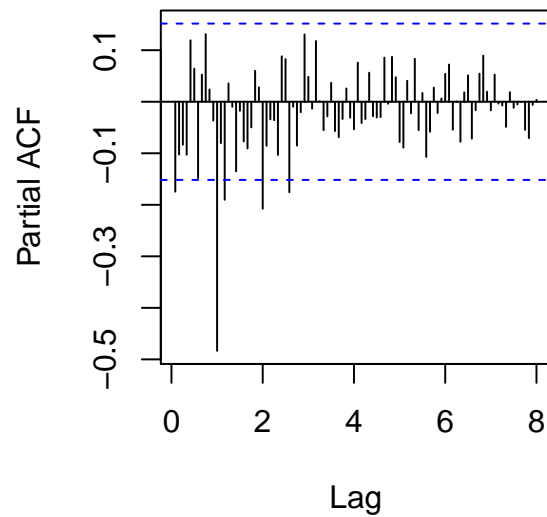
The null hypothesis for the KPSS test is that the data are stationary. p-value = 0.1 (less than 0.5) greater than printed p-value, so the null hypothesis of stationarity around a trend is not rejected and I will use it as the stationary part to do further analysis.

2.3 Model Selection Now that the time series is stationary, I will analyze its ACF and PACF plots to identify the AR, MA, SAR, and SMA orders in the SARIMA model.

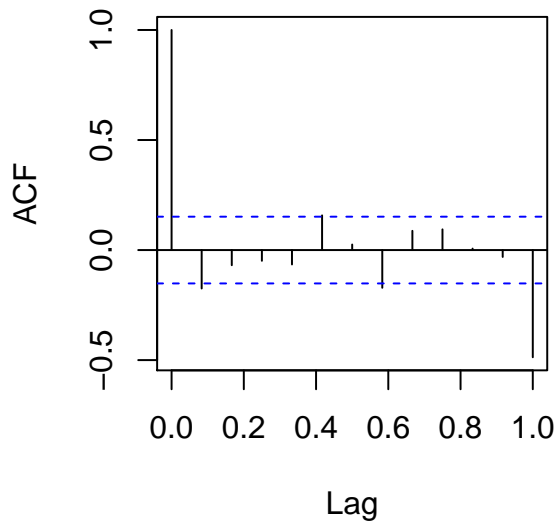
ACF of detrended data



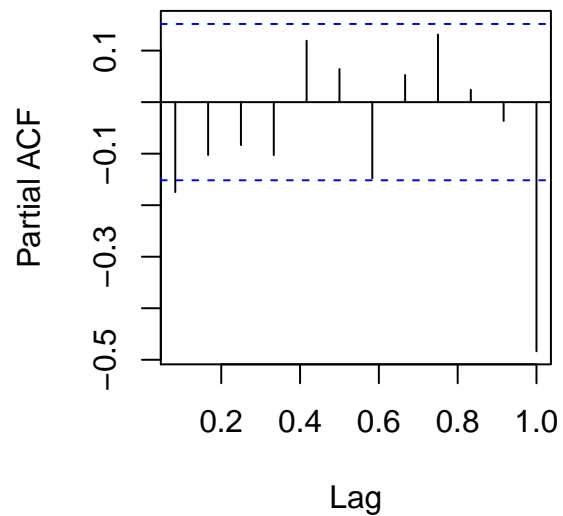
PACF of detrended data



ACF of detrended data



PACF of detrended data



2.3.1 Model Identification For seasonal part (P, D, Q): focus on the seasonal lags $h = 1s, 2s, 3s$ and so on:

- One seasonal differencing so $D = 1$ at lag $s = 12$
- The ACF shows a strong peak at $h = 1s$ then cuts off, so $Q = 1$ (SMA=1).
- The PACF exponentially decays at lags that are multiples of 12. This leads us to consider could consider P either 1 or 2 (SAR=1 or SAR=2)

In ARIMA (p, d, q), the character “p” refers to the order of the autoregressive component, “d” the number of differences needed to arrive at a stationary ARMA (p, q) process, and “q” the order of the moving average component:

- ACF outside confidence interval at lags 0, 1, 4, 6, so $q = 0, 1, 4, 6$
- PACF outside confidence intervals at lags 0, $p = 0$
- Skip the combination of both p and $q = 0$

2.3.2 Determination of P SARMA model with season $P = 1$ tried $Q = 1$ $q = 1, 4, 6$ $p = 0$

- When $P = 1$, $q = 1$, the AIC is

```
## [1] 545.1609
```

- When $P = 1$, $q = 4$, the AIC is

```
## [1] 536.4832
```

- When $P = 1$, $q = 6$, the AIC is

```
## [1] 539.8604
```

SARMA model with season $P = 2$ tried $Q = 1$ $q = 1, 4, 6$ $p = 0$

- When $P = 2$, $q = 1$, the AIC is

```
## [1] 544.9754
```

- When $P = 2$, $q = 4$, the AIC is

```
## [1] 533.6927
```

- When $P = 2$, $q = 6$, the AIC is

```
## [1] 537.1475
```

Lists of the AIC:

- If the $p = 0$, $q = 1$ AIC of $Q = 1$ is 545.1609, AIC of $Q = 2$ is 544.98
- If the $p = 0$, $q = 4$ AIC of $Q = 1$ is 536.4832, AIC of $Q = 2$ is 533.69
- If the $p = 0$, $q = 6$ AIC of $Q = 1$ is 539.8604, AIC of $Q = 2$ is 537.15

So I end up using $Q = 2$ since it has a relatively lower AIC. Two models I will be considered are also based on the AIC. SARIMA (0,1,4) x (2,1,1)₁₂ and SARIMA (0,1,6) x (2,1,1)₁₂. Further, I will calculate the significance interval for each model to see if there any insignificant coefficient could be removed from the model.

2.3.4 Determination of Final Model

First Model: SARIMA (0,1,4) x (2,1,1)₁₂

- MA2, MA3 and MA4 are not significant
- With fixed MA2, MA3 and MA4, AIC becomes 544.98 (increase), the coefficient of MA goes to -1.
- With fixed MA3 and MA4, AIC = 532.4 (decrease), the coefficient of SMA1 goes to -1.
- With fixed MA4, AIC = 532.04 (decrease), gives SMA1 = -0.9991, MA part is not invertible.

```
##
```

```
## Call:
```

```
## arima(x = mydata.final, order = c(0, 1, 4), seasonal = list(order = c(2, 1,
##      1), period = 12), fixed = c(NA, NA, NA, NA, NA, NA, NA), method = "ML")
```

```
##
```

```
## Coefficients:
```

```
##          ma1      ma2      ma3      ma4      sar1      sar2      sma1
##      -1.3245  0.1702  0.2043 -0.0500 -0.717  -0.2069 -0.9999
## s.e.   0.0905  0.1394  0.1413  0.0832  0.090  0.0913  0.2352
##
## sigma^2 estimated as 1.124:  log likelihood = -258.85,  aic = 533.69
```

- I keep the model without any fixed 0: SARIMA (0,1,4) x (2,1,1)₁₂ $\nabla_1 \nabla_{12} Y_t (1 + 0.7153B^{12} + 0.2069B^{24}) = (1 - 1.3245B + 0.1702B^2 + 0.2043B^3 - 0.0500B^4)(1 - 0.9999B^{12})Y_t$

This model is stational because it is a pure SMA model and The absolute value of SAR part roots is greater than 1. But the root of MA part contains 1, it is not invertible, then the innovations can still be represented by observations of the future, but is not helpful at all for forecasting purpose.

```
## [1] -1.732721+1.353119i -1.732721-1.353119i
## [1] 1.000000+0i 1.771213-0i -4.151360-0i
```

Second model: SARIMA (0,1,6) x (2,1,1)₁₂

- MA part only MA1 is significant
- With fixed MA2, MA3, MA4, and MA5 AIC = 539.74 (increase).
- With fixed MA3, MA4, MA5 and MA6 AIC = 532.4 (decrease), the coefficient of SMA1 goes to -1.

```
##
## Call:
## arima(x = mydata.final, order = c(0, 1, 6), seasonal = list(order = c(2, 1,
##      1), period = 12), fixed = c(NA, 0, NA, NA, NA, NA, NA, NA, NA), method = "ML")
##
## Coefficients:
##          ma1  ma2      ma3      ma4      ma5      ma6      sar1      sar2      sma1
##      -1.2409   0  0.2969 -0.0817  0.0074  0.0201 -0.7139 -0.1967 -0.9999
## s.e.   0.0663   0  0.1241  0.1106  0.1430  0.1112  0.0956  0.0921  0.3789
##
## sigma^2 estimated as 1.138:  log likelihood = -259.61,  aic = 537.23
```

- With fixed MA2, AIC = 537.23, I get $\nabla_1 \nabla_{12} Y_t (1 + 0.7139B^{12} + 0.1967B^{24}) = (1 - 1.2409B + 0.2969B^3 - 0.0817B^4 + 0.0074B^5 - 0.0201B^6)(1 - 0.9999B^{12})Y_t$

This model is stational because it is a pure SMA model and The absolute value of SAR part roots is greater than 1. The absolute value of SMA1 is less than 1. The roots of SAR are all greater than 1.

```
## [1] -1.814692+1.338199i -1.814692-1.338199i
## [1] 1.003484+0.000000i -2.450894+0.306863i 1.483253+0.000000i
## [4] 1.023446-2.105046i 1.023446+2.105046i -2.450894-0.306863i
```

2.4 Model Diagnostics

Model 1 $\nabla_1 \nabla_{12} Y_t (1 + 0.7153B^{12} + 0.2069B^{24}) = (1 - 1.3245B + 0.1702B^2 + 0.2043B^3 - 0.0500B^4)(1 - 0.9999B^{12})Y_t$

We begin by plotting the residuals of this model, the mean of it is very close to 0, and the variance is very close to 1.

```
mean(res)
```

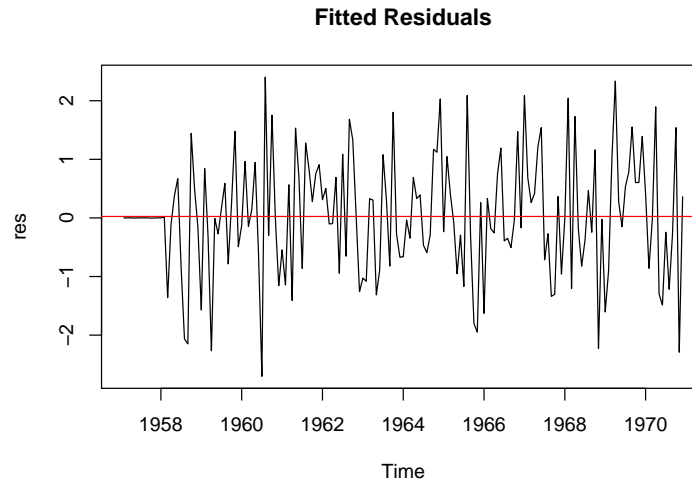
```
## [1] 0.02512893
```



```
var(res)
```

```
## [1] 1.041784
```

No trend, no visible change of variance, and no seasonality.



```
##
```

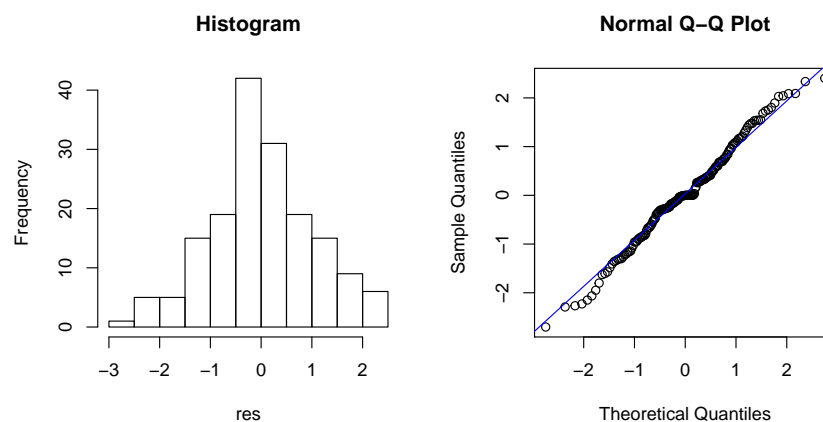
```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: res
```

```
## W = 0.99094, p-value = 0.3704
```

The residuals also passed the Shapiro-Wilks normality test as $p=0.3704$, then I plot the Histogram and the QQ-Plot.



From these plots, the residuals appears to be approximately normally distributed. We now perform the Box-Pierce, Ljung-Box, and McLeod-Li tests on the residuals.

```
## [1] "Box-Pierce"
```

```
##
```

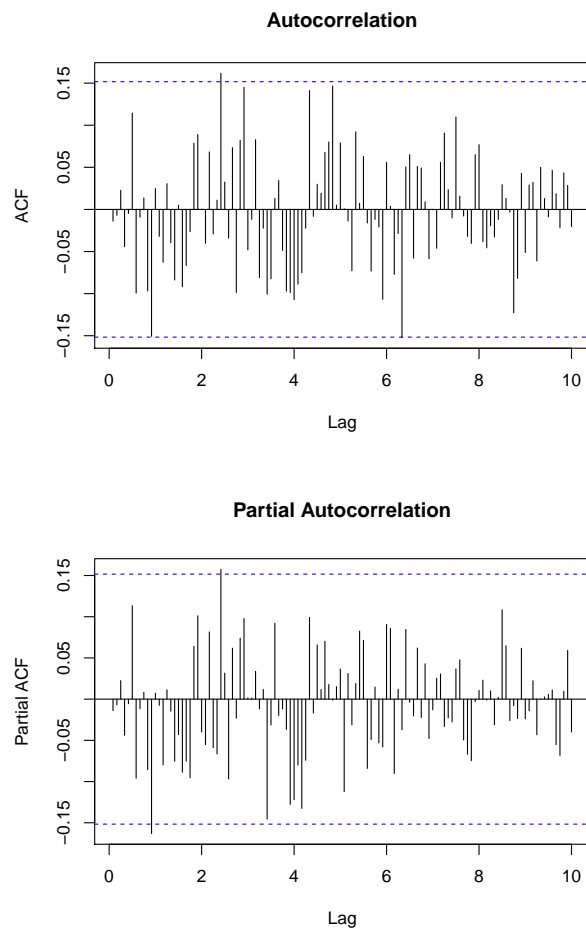
```
## Box-Pierce test
```

```
##
```

```
## data: res
```

```
## X-squared = 9.9595, df = 6, p-value = 0.1264
## [1] "Ljung-Box"
##
## Box-Ljung test
##
## data:  res
## X-squared = 10.641, df = 6, p-value = 0.1001
## [1] "McLeod-Li"
##
## Box-Ljung test
##
## data:  (res)^2
## X-squared = 10.569, df = 13, p-value = 0.6469
```

This model passes all tests at the $\alpha=0.05$ significance level. We now plot the ACF and PACF of the residuals to ensure they resemble white noise.



Even though there are few ACFs and PACFs are little bit out of the confident interval, but I can still assume they are white noise.

Model 2

- $\nabla_1 \nabla_{12} Y_t (1 + 0.7139B^{12} + 0.1967B^{24}) = (1 - 1.2409B + 0.2969B^3 - 0.0817B^4 + 0.0074B^5 - 0.0201B^6)(1 - 0.9999B^{12})Z_t$

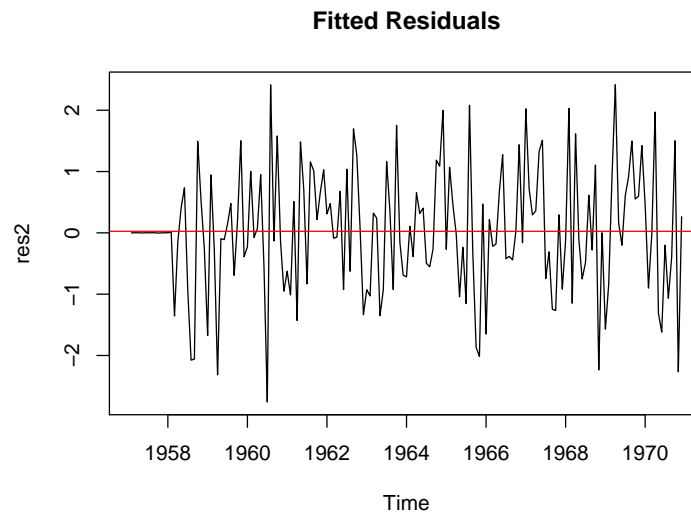
We begin by plotting the residuals of second model. Again, the mean of it is very close to 0, variance is close to 1.

```
mean(res2)
```

```
## [1] 0.02508063
```

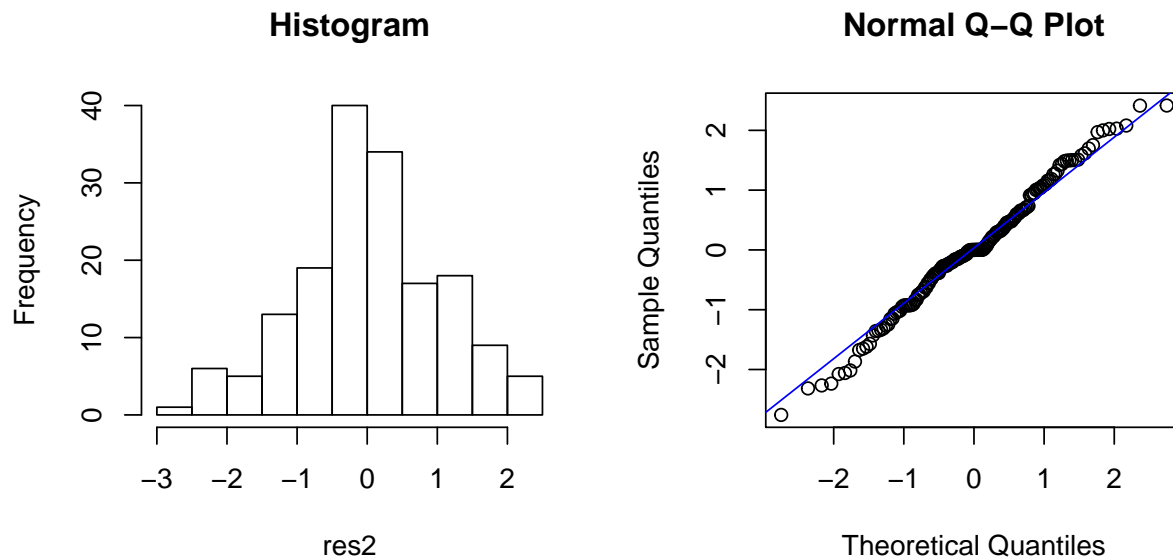
```
var(res2)
```

```
## [1] 1.038785
```



```
##
## Shapiro-Wilk normality test
##
## data:  res2
## W = 0.99099, p-value = 0.3756
```

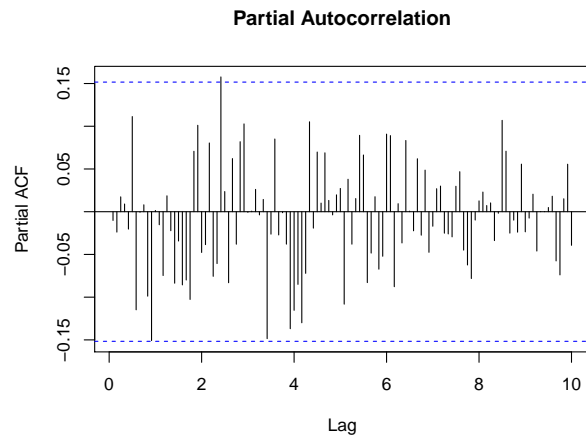
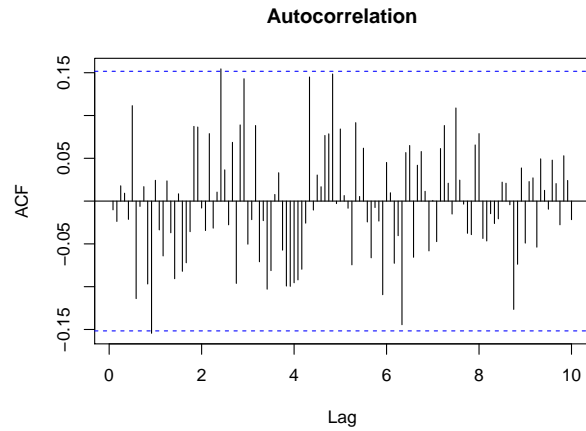
The residuals passed the Shapiro-Wilks normality test as $p=0.3756$.



Both histogram and QQ-Plot show that the residuals are normally distributed. We now can perform the test on residuals.

```
## [1] "Box-Pierce"
##
## Box-Pierce test
##
## data:  res2
## X-squared = 10.39, df = 6, p-value = 0.1092
## [1] "Ljung-Box"
##
## Box-Ljung test
##
## data:  res2
## X-squared = 11.106, df = 6, p-value = 0.08517
## [1] "McLeod-Li"
##
## Box-Ljung test
##
## data:  (res2)^2
## X-squared = 11.303, df = 13, p-value = 0.5855
```

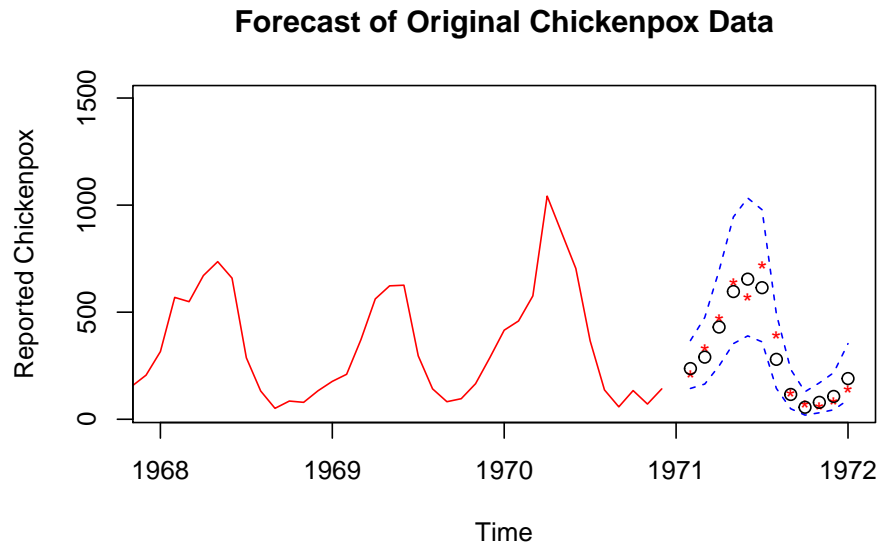
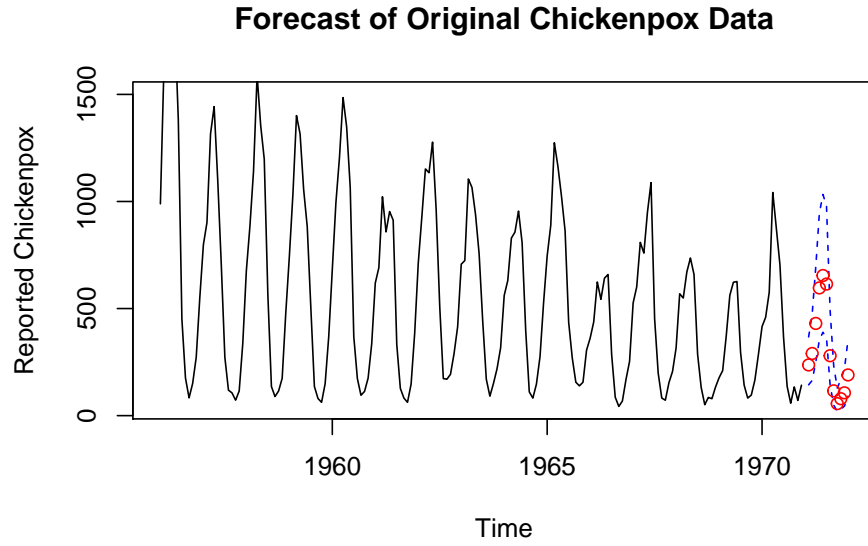
This model passes all tests. Finally, I would check that the ACF and PACF of the residuals.



The ACFs and PACFs are all within the interval.

Based on the results of these tests, both models passed all tests. All in all, I prefer to choose the second model, despite the first model is more simple, but it is not invertible which is not helpful for future forecasting. The invertible process is important because the most recent observations have a higher weight than observations from the past.

2.5 Forecasting Forecasts on original data, the blue dotted lines represent the 95% confidence interval of the forecasts. It is hard to tell if all of the predictions are in the interval.



To zoom the graph. Forecasts are plotted below in red, and the actual observations are in red. We could see that the first 4 months and the last 5 months is very close to the realized observations. The model should be appropriate.

3. Conclusion

The best model fit this dataset will be: $-\nabla_1 \nabla_{12} Y_t (1 + 0.7139B^{12} + 0.1967B^{24}) = (1 - 1.2409B + 0.2969B^3 - 0.0817B^4 + 0.0074B^5 - 0.0201B^6)(1 - 0.9999B^{12})Z_t$

Since this data is only collected until 1972, I am not very sure if it still useful to predict chickenpox currently due to the progress of medical knowledge and technology, the reported number of chickenpox in recent years may be much lower than this data. But it is worth seeing the disease-related time series project, and it still practical in other related contagious disease analyses.

4. Appendix

The R code used to create the analysis in this report is provided in its entirety below.

```
library(tsd1)
mydata=ts(tsd1[[241]][301:480], start = c(1956,1), frequency = 12)
next12 = tsdl[[241]][481:492]
ts.plot(mydata, ylim = c(0,2000),
        ylab = "Reported of chickenpox",
        main = "Reported Chickenpox vs. Time")
#abline(h=mean(mydata), col="blue")
#abline(lm(as.numeric(mydata)~as.numeric(1:length(mydata))),col='red')
summary(mydata)
var(mydata)

library(MASS)
t=1:length(mydata)
fit=lm(mydata~t)
bcTransform = boxcox(mydata ~ t, plotit = TRUE)
lambda = bcTransform$x[which(bcTransform$y == max(bcTransform$y))] # calculate lambda
lambda
mydata.bc = (1/lambda)*(mydata^lambda-1)
var(mydata.bc) # calculate variance of transformed data
ts.plot(mydata.bc,main = "Box-Cox tranformed data", ylab = expression(Y[t]))

acf(mydata.bc, main='ACF of Original data',lag.max = 12*5)
pacf(mydata.bc, main='PACF of Original data',lag.max = 12*5)

#remove seasonality
mydata.d12=diff(mydata.bc,12)
#variance after remove seaonality
var(mydata.d12)
#take diff at lag 1 to remove trend
mydata.d12.lag1=diff(mydata.d12,1)
#Variance after diff at lag1
var(mydata.d12.lag1)
#take diff at lag 1 on the previous Mydata.d12.lag1
mydata.d12.lag1.lag1=diff(mydata.d12.lag1,1)
#variance after diff at lag 1 again
var(mydata.d12.lag1.lag1)
###This is not considered because the variance increases, it may causes overdifference

par(mfrow = c(1,2))
#stationary part
mydata.final=mydata.d12.lag1
plot.ts(mydata.d12, main='Remove Seasonality')
plot.ts(mydata.final, main='Detrended data')
abline(h=mean(mydata.final, na.rm = T), col = "red")

# KPSS Test for stationarity
library(tseries)
kpss.test(mydata.final)

par(mfrow = c(1,2))
# determine P and Q
```

```

acf(mydata.final, main='ACF of detrended data',lag.max = 12*8)
pacf(mydata.final, main='PACF of detrended data',lag.max = 12*8)
# determine p and q
acf(mydata.final, main='ACF of detrended data',lag.max = 12)
pacf(mydata.final, main='PACF of detrended data',lag.max = 12)

modeltest1 = arima(mydata.final, order = c(0,1,1),
                   seasonal = list(order = c(1,1,1),
                                   period = 12),
                   method="ML")
modeltest1$aic # aic = 545.1609
modeltest2 = arima(mydata.final, order = c(0,1,4),
                   seasonal = list(order = c(1,1,1),
                                   period = 12),
                   method="ML")
modeltest2$aic # aic = 536.4832
modeltest3 = arima(mydata.final, order = c(0,1,6),
                   seasonal = list(order = c(1,1,1),
                                   period = 12),
                   method="ML")
modeltest3$aic # aic = 539.8604
modeltest4 = arima(mydata.final, order = c(0,1,1),
                   seasonal = list(order = c(2,1,1),
                                   period = 12),
                   method="ML")
modeltest4$aic # aic = 544.9754
modeltest5 = arima(mydata.final, order = c(0,1,4),
                   seasonal = list(order = c(2,1,1),
                                   period = 12),
                   method="ML")
modeltest5$aic # aic = 533.6927
modeltest6 = arima(mydata.final, order = c(0,1,6),
                   seasonal = list(order = c(2,1,1),
                                   period = 12),
                   method="ML")
modeltest6$aic # aic = 537.1475

-1.3245-1.96*0.0905
-1.3245+1.96*0.0905

0.1702-1.96*0.1394
0.1702+1.96*0.1394

0.2043-1.96*0.1413
0.2043+1.96*0.1413
#ma3 is not significant

-0.0500-1.96*0.0832
-0.0500+1.96*0.0832
#ma4 is not significant

-0.717-1.96*0.090
-0.717+1.96*0.090

```



```

-0.2069 -1.96*0.0913
-0.2069 +1.96*0.0913

-0.9998-1.96*0.2352
-0.9998+1.96*0.2352

0.1877-1.96*0.1464
0.1877+1.96*0.1464

arima(mydata.final, order=c(0,1,4),
      seasonal = list(order = c(2,1,1), period = 12),
      fixed =c(NA, NA, NA, NA, NA, NA, NA), method="ML")

polyroot(c(1, 0.717 ,0.2069))
polyroot(c(1, -1.3237 , 0.1877 , 0.1360 ))
#plot.roots(NULL,polyroot(c(1, 0.7153,0.2062)),main='Root for SAR')

-1.3283-1.96*0.0905
-1.3283+1.96*0.0905

0.2034-1.96*0.1450
0.2034+1.96*0.1450
#ma2 is not significant

0.1647-1.96*0.1561
0.1647+1.96*0.1561
#ma3 is not significant

-0.1020-1.96*0.1188
-0.1020+1.96*0.1188
#ma4 is not significant

0.1092-1.96*0.1671
0.1092+1.96*0.1671
#ma5 is not significant

-0.0471-1.96*0.1206
-0.0471+1.96*0.1206
#ma6 is not significant

-0.713-1.96*0.097
-0.713+1.96*0.097

-0.2059-1.96*0.0919
-0.2059+1.96*0.0919

-0.9999-1.96*0.2696
-0.9999+1.96*0.2696

arima(mydata.final, order=c(0,1,6),
      seasonal = list(order = c(2,1,1), period = 12),
      fixed =c(NA,0, NA, NA, NA, NA, NA, NA), method="ML")

```

```

polyroot(c(1, 0.7139,0.1967))
polyroot(c(1, -1.2409,0,0.2969,-0.0817,0.0074,0.0201))

#TEST MODEL 1
model = arima(mydata.final, order = c(0,1,4),
              seasonal = list(order = c(2,1,1),
                              period = 12,
                              method="ML"))

model
res=residuals(model)
mean(res)
var(res)
ts.plot(res,main = "Fitted Residuals")
abline(h = mean(res), col = "red")
shapiro.test(res)
#The residuals passed the Shapiro-Wilks normality test as p=0.3704
par(mfrow = c(1,2))
# Histogram
hist(res,main = "Histogram")
# q-q plot
qqnorm(res)
qqline(res,col ="blue")
paste("Box-Pierce")
Box.test(res, lag = 13, type = "Box-Pierce", fitdf=7)
paste("Ljung-Box")
Box.test(res, lag = 13, type = "Ljung-Box", fitdf=7)
paste("McLeod-Li")
Box.test((res)^2, lag = 13, type="Ljung-Box")
# acf
acf(res,main = "Autocorrelation", na.action=na.pass,lag.max = 12*10)
# pacf
pacf(res,main = "Partial Autocorrelation", na.action=na.pass,lag.max = 12*10)

#TEST MODEL 2
model2 = arima(mydata.final, order = c(0,1,6),
              seasonal = list(order = c(2,1,1),
                              period = 12,
                              method="ML"))

model2
res2=residuals(model2)
mean(res2)
var(res2)
ts.plot(res2,main = "Fitted Residuals")
abline(h = mean(res2), col = "red")
shapiro.test(res2)
par(mfrow = c(1,2))
# Histogram
hist(res2,main = "Histogram")
# q-q plot
qqnorm(res2)
qqline(res2,col ="blue")
paste("Box-Pierce")
Box.test(res2, lag = 13, type = "Box-Pierce", fitdf=7)

```

```

paste("Ljung-Box")
Box.test(res2, lag = 13, type = "Ljung-Box", fitdf=7)
paste("McLeod-Li")
Box.test((res2)^2, lag = 13, type="Ljung-Box")
# acf
acf(res2,main = "Autocorrelation", na.action=na.pass,lag.max = 12*10)
# pacf
pacf(res2,main = "Partial Autocorrelation", na.action=na.pass,lag.max = 12*10)

fit.A = arima(mydata.final, order=c(0,1,6),
              seasonal = list(order = c(2,1,1), period = 12),
              fixed =c(NA,0, NA, NA, NA, NA, NA, NA, NA), method="ML")

library(forecast)

# To produce graph with 12 forecasts on transformed data:
pred.tr = predict(fit.A, n.ahead = 12)
U.tr = pred.tr$pred + 1.96*pred.tr$se # % upper bound
L.tr = pred.tr$pred - 1.96*pred.tr$se # % lower bound
ts.plot(mydata.bc, ylim=c(min(mydata.bc),30),
        xlim = c(1956, 1972),
        ylab = 'Reported box cox transformed Chickenpox',
        main = 'Forecast of Transformed Chickenpox Data')
lines(1971+(1:12)/12,U.tr, col="blue", lty="dashed")
lines(1971+(1:12)/12,L.tr, col="blue", lty="dashed")
points(1971+(1:12)/12, pred.tr$pred, col="red")

# To produce graph with forecasts on original data:
pred.orig = InvBoxCox(pred.tr$pred,lambda)
U = InvBoxCox(U.tr,lambda)
L = InvBoxCox(L.tr,lambda)
ts.plot(mydata, ylim=c(min(mydata),1500),
        xlim = c(1956, 1972),
        ylab = 'Reported Chickenpox',
        main = 'Forecast of Original Chickenpox Data')
lines(1971+(1:12)/12, U, col="blue", lty="dashed")
lines(1971+(1:12)/12, L, col="blue", lty="dashed")
points(1971+(1:12)/12, pred.orig, col="red")

#To plot zoomed forecasts and true values
ts.plot(mydata, ylim=c(min(mydata),1500),
        xlim = c(1968, 1972),
        ylab = 'Reported Chickenpox',
        main = 'Forecast of Original Chickenpox Data',
        col='red')
lines(1971+(1:12)/12,U, col="blue", lty="dashed")
lines(1971+(1:12)/12,L, col="blue", lty="dashed")
points(1971+(1:12)/12,next12, pch = "*", col="red")
points(1971+(1:12)/12, pred.orig, col="black")

```