

# Model Building 3

Lian Morales

4/26/2021

```
# pulling the data from the Los Angeles County GitHub
casedata <- read.csv(text = getURL("https://raw.githubusercontent.com/datadesk/california-coronavirus-d
  filter(county == "Los Angeles") %>%
  mutate(date = date(date), month = month(date)) %>%
  map_df(rev) %>%
  filter(!is.na(new_confirmed_cases) & between(date, date("2020-04-01"),date("2021-03-31")))

# creating the time series
case.ts <- ts(casedata$new_confirmed_cases, start = 1, frequency = 1)

# averaging dec 25th and 26th
case.ts[269] <- 14711
case.ts[270] <- 14712

## april 1 - sep 30
case.ts.1 <- ts(case.ts[1:183], start = 1, frequency = 1)

# april 1 - jul 14
case.ts.1.1 <- ts(case.ts[1:105], start = 1, frequency = 1)
# jul 15 - sep 30
case.ts.1.2 <- ts(case.ts[106:183], start = 1, frequency = 1)

## oct 1 - mar 31
case.ts.2 <- ts(case.ts[184:365], start = 1, frequency = 1)

# oct 1 - dec 20
case.ts.2.1 <- ts(case.ts[184:264], start = 1, frequency = 1)
# dec 21 - mar 31
case.ts.2.2 <- ts(case.ts[265:365], start = 1, frequency = 1)
```

October 1 - December 20

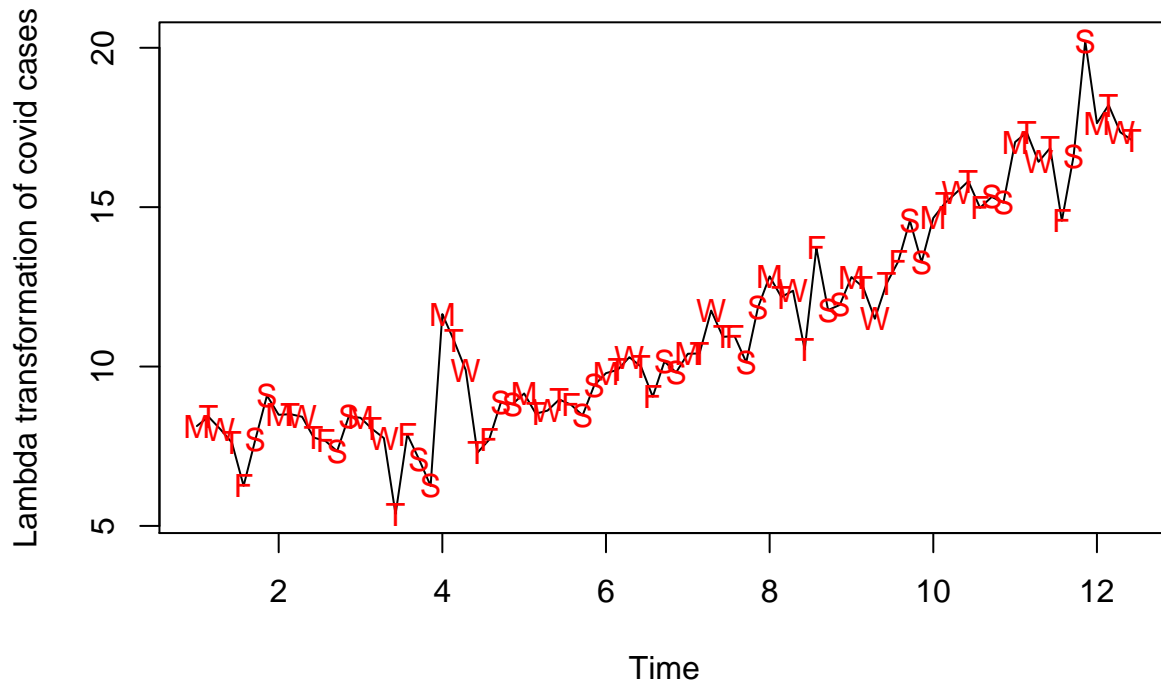
IMA(1,2)X(0,0,1)[7] and IMA(1,2)

```
# transform data October 1 - December 20 to lambda + add seasonality
case.ts.2.1_seas.trans <- ts(case.ts.2.1^0.3, frequency = 7)
case.ts.2.1_trans <- ts(case.ts.2.1^0.3)

##### SEASONAL
```

```
# plot ts of transformed seasonal data
week. <- season(case.ts.2.1_seas.trans)
plot(case.ts.2.1_seas.trans, ylab = "Lambda transformation of covid cases", xlab = "Time", main = "Transformation by Lambda = 0.3",
points(case.ts.2.1_seas.trans, pch = as.vector(week.), col = "red")
```

## Transformation by Lambda = 0.3

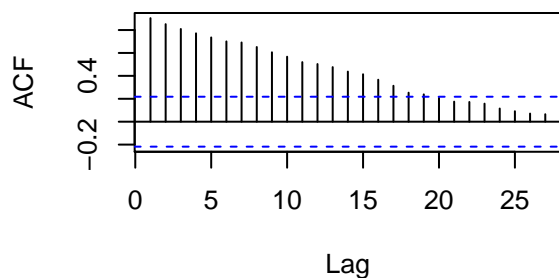


```
par(mfrow = c(2,2))
# acf of transformed data
acf(as.vector(case.ts.2.1_seas.trans), lag.max = 27, main = "ACF of Transformation by Lambda = 0.3")
# pacf of transformed data
pacf(as.vector(case.ts.2.1_seas.trans), lag.max = 27, main = "Transformation by Lambda = 0.3")
# eacf of transformed data
eacf(as.vector(case.ts.2.1_seas.trans)) # seems to show a ARMA(1,1)?
```

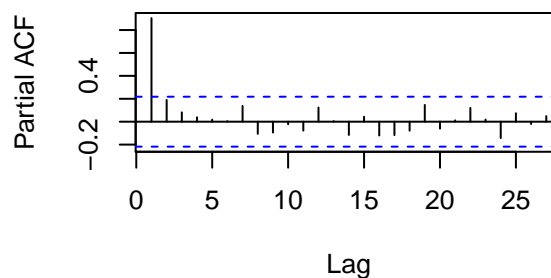
```
## AR/MA
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x x x x x x x x x x x
## 1 x o o o o o o o o o o o o
## 2 x o o o o o o o o o o o o
## 3 x o o x o o o o o o o o o
## 4 x x x x o o o o o o o o o
## 5 o x o x o o o o o o o o o
## 6 o x o x o o o o o o o o o
## 7 x o o x o o o o o o o o o
```

```
# best subsets ARMA approach
par(mfrow = c(1,1))
```

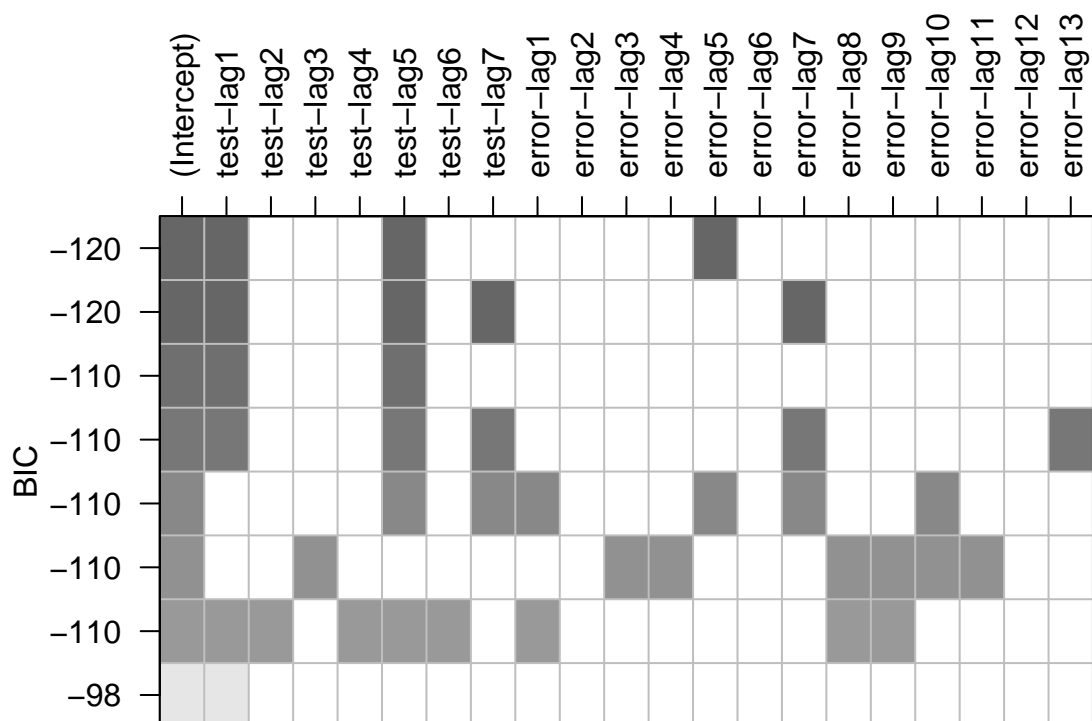
ACF of Transformation by Lambda = 0.



Transformation by Lambda = 0.3



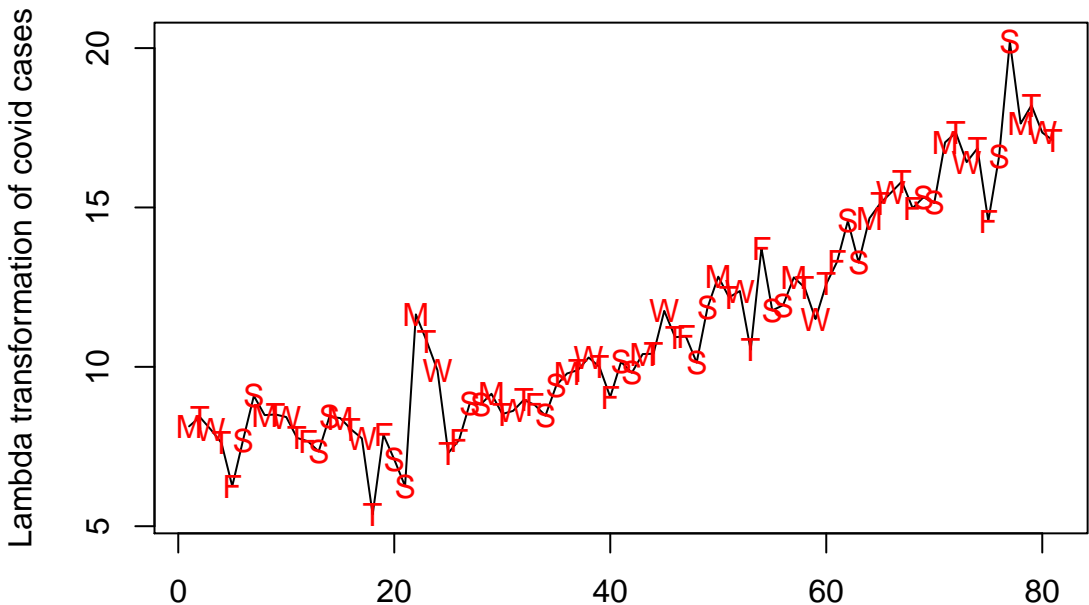
```
res = armasubsets(y=case.ts.2.1_seas.trans, nar= 7, nma = 13, y.name = "test", ar.method = 'ols')
plot(res) # not sure how to analyze this
```



##### TRANSFORMATION (no seasonality)

```
par(mfrow = c(1,1))
plot(case.ts.2.1_trans, ylab = "Lambda transformation of covid cases", xlab = "Time", main = "Transformation of covid cases")
points(case.ts.2.1_trans, pch = as.vector(week.), col = "red")
```

## Transformation by Lambda = 0.3

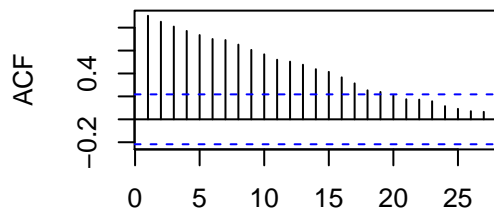


```
par(mfrow = c(2,2))
# acf of transformed data
acf(as.vector(case.ts.2.1_trans), lag.max = 27, main = "ACF of Transformation by Lambda = 0.3")
# pacf of transformed data
pacf(as.vector(case.ts.2.1_trans), lag.max = 27, main = "Transformation by Lambda = 0.3")
# eacf of transformed data
eacf(as.vector(case.ts.2.1_trans)) # seems to show a ARMA(1,1)?
```

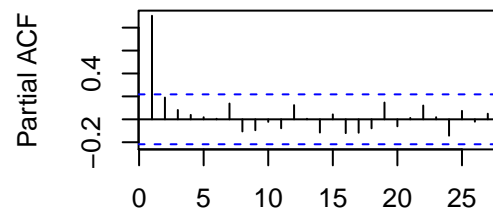
[illegible]

```
# best subsets ARMA approach
par(mfrow = c(1,1))
```

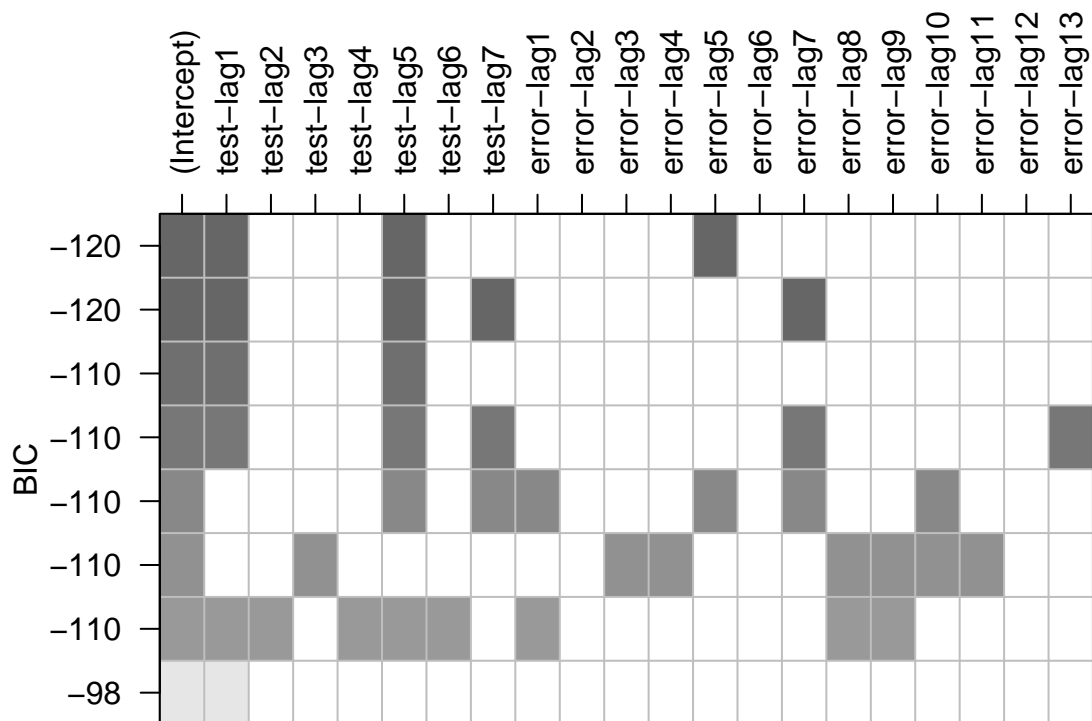
## ACF of Transformation by Lambda = 0.



## Transformation by Lambda = 0.3



```
res = armasubsets(y=case.ts.2.1_trans, nar= 7, nma = 13, y.name = "test", ar.method = 'ols')
plot(res) # not sure how to analyze this
```



```
# simulated model prediction
auto.arima(case.ts.2.1_seas.trans)
```

```
## Series: case.ts.2.1_seas.trans
## ARIMA(0,1,2)(0,0,1)[7] with drift
##
## Coefficients:
##          ma1          ma2          sma1      drift
##      -0.5695   -0.2049    0.1970    0.1270
## s.e.    0.1128    0.1093    0.1245    0.0344
##
## sigma^2 estimated as 1.249:  log likelihood=-120.86
## AIC=251.73   AICc=252.54   BIC=263.64
```

```
auto.arima(case.ts.2.1_trans)
```

```
## Series: case.ts.2.1_trans
## ARIMA(0,1,2) with drift
##
## Coefficients:
##          ma1          ma2      drift
##      -0.5567  -0.1929   0.1271
## s.e.    0.1126   0.1061   0.0326
##
## sigma^2 estimated as 1.274:  log likelihood=-122.03
## AIC=252.06   AICc=252.59   BIC=261.59
```

## Completing analysis with suggested models

```
# simulate aimra models for seasonal and transformed data
arima.seas_trans <- arima(case.ts.2.1_seas_trans,
                          order = c(0,1,2),
                          seasonal = list(order = c(0,1,2), period = 7))
arima.trans <- arima(case.ts.2.1_trans, order = c(0,1,2))
```

```
# residual tests
```

```
# check for correlation of error terms
```

```
Box.test(rstandard(arima.seas_trans), type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data:  rstandard(arima.seas_trans)
## X-squared = 0.33756, df = 1, p-value = 0.5612
```

```
Box.test(rstandard(arima.trans), type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data:  rstandard(arima.trans)
## X-squared = 0.48346, df = 1, p-value = 0.4869
```

```
# check for independence of error terms
```

```
runs(rstandard(arima.seas_trans))
```

```
## $pvalue
## [1] 0.991
##
## $observed.runs
## [1] 42
##
```

```
## $expected.runs
## [1] 41.44444
##
## $n1
## [1] 39
##
## $n2
## [1] 42
##
## $k
## [1] 0
```

```
runs(rstandard(arima.trans))
```

```
## $pvalue
## [1] 0.676
##
## $observed.runs
## [1] 42
##
## $expected.runs
## [1] 39.71605
##
## $n1
## [1] 32
##
## $n2
## [1] 49
##
## $k
## [1] 0
```

```
# check for normality of error terms
shapiro.test(rstandard(arima.seas_trans))
```

```
##
## Shapiro-Wilk normality test
##
## data:  rstandard(arima.seas_trans)
## W = 0.93447, p-value = 0.0004514
```

```
shapiro.test(rstandard(arima.trans))
```

```
##
## Shapiro-Wilk normality test
##
## data:  rstandard(arima.trans)
## W = 0.93084, p-value = 0.0002935
```

## Ljung-Box

*Seasonal Transformed Model*

For the seasonal transformed data, the residuals from the Ljung-Box test have a high p-value of 0.5612, which indicate that the residuals are uncorrelated.

#### *Transformed Model*

Similarly, in the transformed data with no seasonality, the p-value is very high, 0.4869, therefore we do not have evidence against the null hypothesis that the error terms are uncorrelated.

### **Runs Test**

#### *Seasonal Transformed Model*

For the runs test, we notice that the seasonal transformed model has observed runs of 42 versus expected runs of 41.44. The p-value of this test is large, equalling 0.991, which means we do not have statistically significant evidence against independence of the error terms in the model.

#### *Transformed Model*

For the transformed model, the model has an observed runs of 42 and expected of 39.7. Similar to the seasonal model, this runs test has a high p-value of 0.676. Although this is not as high as the seasonal model, it still indicated that we do not have statistically significant evidence against independence of the error terms in the model.

### **Shapiro Wilks Test**

#### *Seasonal Transformed Model*

For the seasonal model, we notice that the p-value for this test is very small, with 0.0004514, indicating that we do have evidence against the null hypothesis that the data are normal.

#### *Transformed Model*

Similar to the seasonal model, the transformed data have a p-value of 0.0002935 indicating that we do have evidence against the null hypothesis that the data are normal.

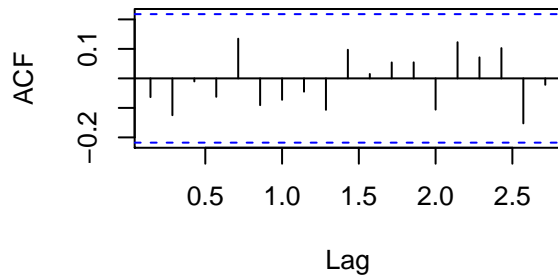
## **Visualize residuals**

```
par(mfrow = c(2,2))
# acf of seasonal and transformed residuals
acf(rstandard(arima.seas_trans))
acf(rstandard(arima.trans))

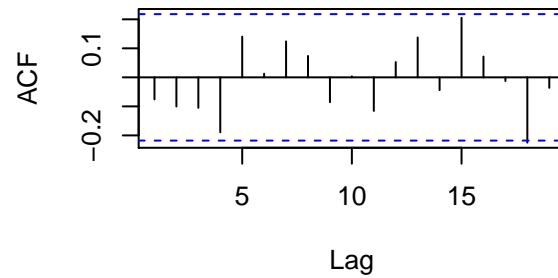
# pacf of seasonal and transformed residuals
pacf(rstandard(arima.seas_trans))
pacf(rstandard(arima.trans))
```



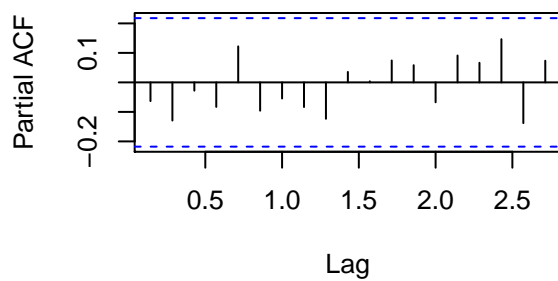
**Series rstandard(arima.seas\_trans)**



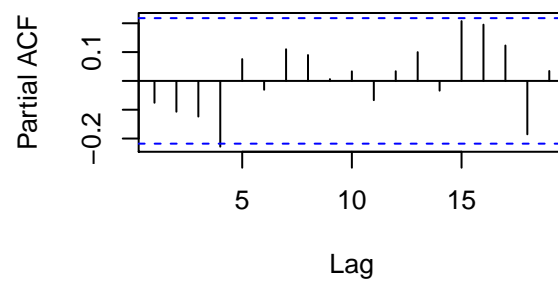
**Series rstandard(arima.trans)**



**Series rstandard(arima.seas\_trans)**

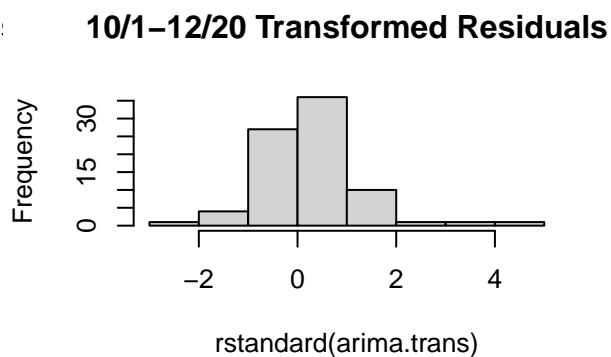
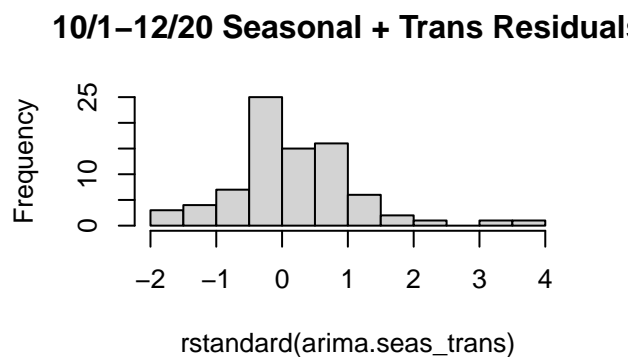
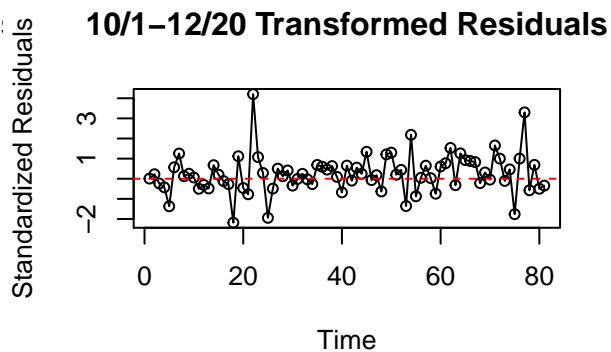
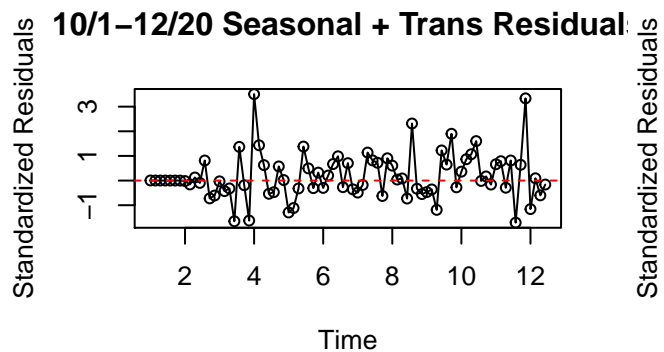


**Series rstandard(arima.trans)**



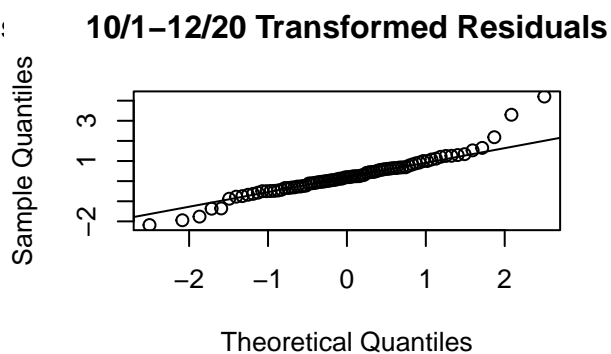
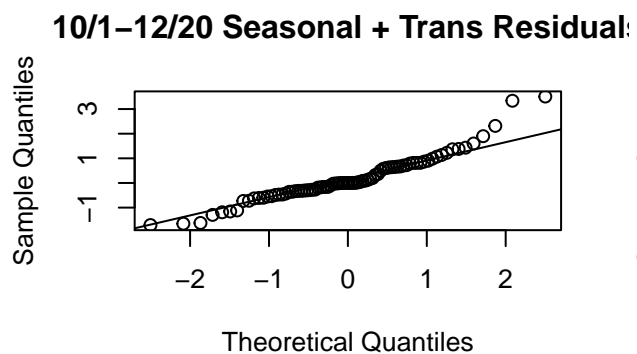
```
# plot residuals
plot(rstandard(arima.seas_trans),
     type = "o",
     main = "10/1-12/20 Seasonal + Trans Residuals",
     ylab = "Standardized Residuals")
abline(h=0,lty=2,col="red")
plot(rstandard(arima.trans),
     type = "o",
     main = "10/1-12/20 Transformed Residuals",
     ylab = "Standardized Residuals")
abline(h=0,lty=2,col="red")

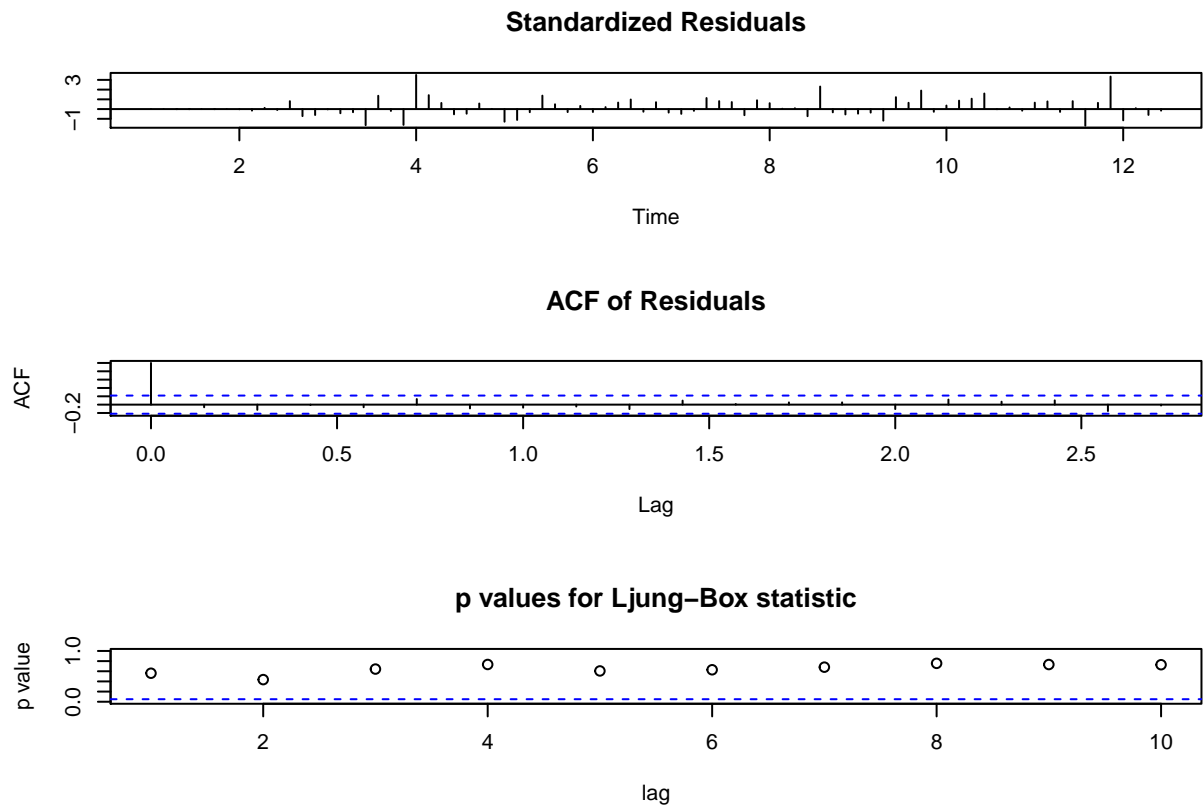
# Histogram of residuals
hist(rstandard(arima.seas_trans),
     main = "10/1-12/20 Seasonal + Trans Residuals")
hist(rstandard(arima.trans),
     main = "10/1-12/20 Transformed Residuals")
```



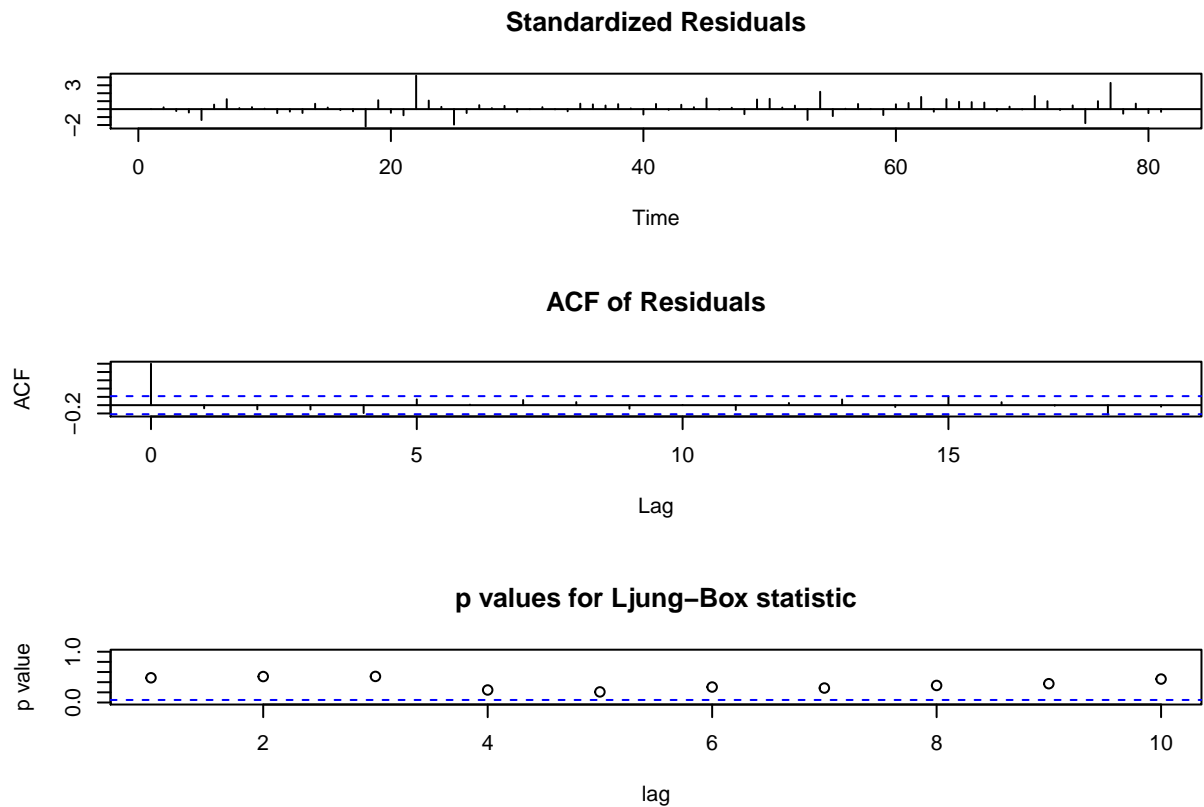
```
# qqnorm plots of residuals
qqnorm(rstandard(arima.seas_trans),
       main = "10/1-12/20 Seasonal + Trans Residuals")
qqline(rstandard(arima.seas_trans))
qqnorm(rstandard(arima.trans),
       main = "10/1-12/20 Transformed Residuals")
qqline(rstandard(arima.trans))

# Diagnostic Plot
tsdiag(arima.seas_trans)
```





```
tsdiag(arima.trans)
```



We notice in the residual visualization, they are very similar, and show signs of no correlation and independence.

We can see this in the acf and pacfs, where they show a zero mean white noise process. In the plot of the residuals, we see that they are generally centered around zero, however there are some outliers that are beyond 3, which we should look out for in our analysis. Both histograms are centered around zero, however we notice in the seasonal model some outliers that are around 3,4, which indicate the outliers and also the non normality of the residuals. Furthermore, in the normality plots, we notice that there is a general linear trend for both data sets, however, the tail ends of data points are very far away from the line, which indicate the data is not normal. In the `tsdiag()`, Diagnostic Plots for Time-Series Fits function it is further reiterating our points of zero mean, white noise, independent process of residuals that are not exactly normal. In choosing one transformation over the other, we will go for the seasonal model, as it is capturing a more accurate representation of the data shown by these statistics and visualizations.