Name: Xuhui Ying    WFU ID: 06648543    Section: C

**1. Results from the stationarity test. Be sure to clearly state the hypotheses and the significance level that you used for conducting the test along with your conclusion.**

$H_0$: Series is non-stationary

$H_a$: Series is stationary

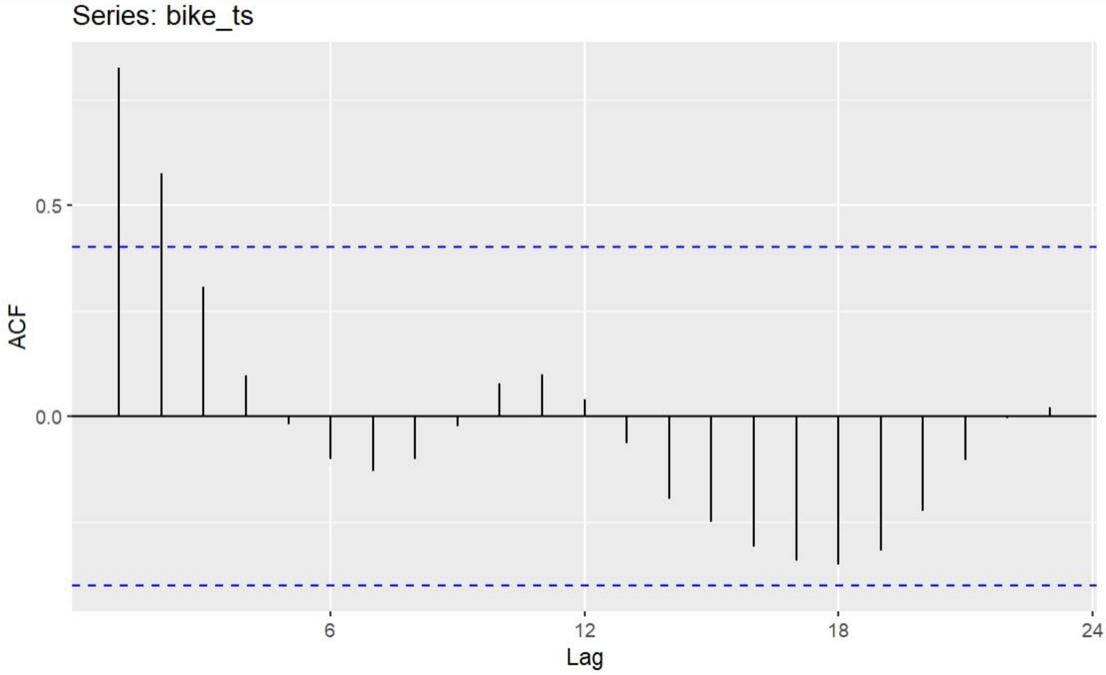
Significance level: 0.05

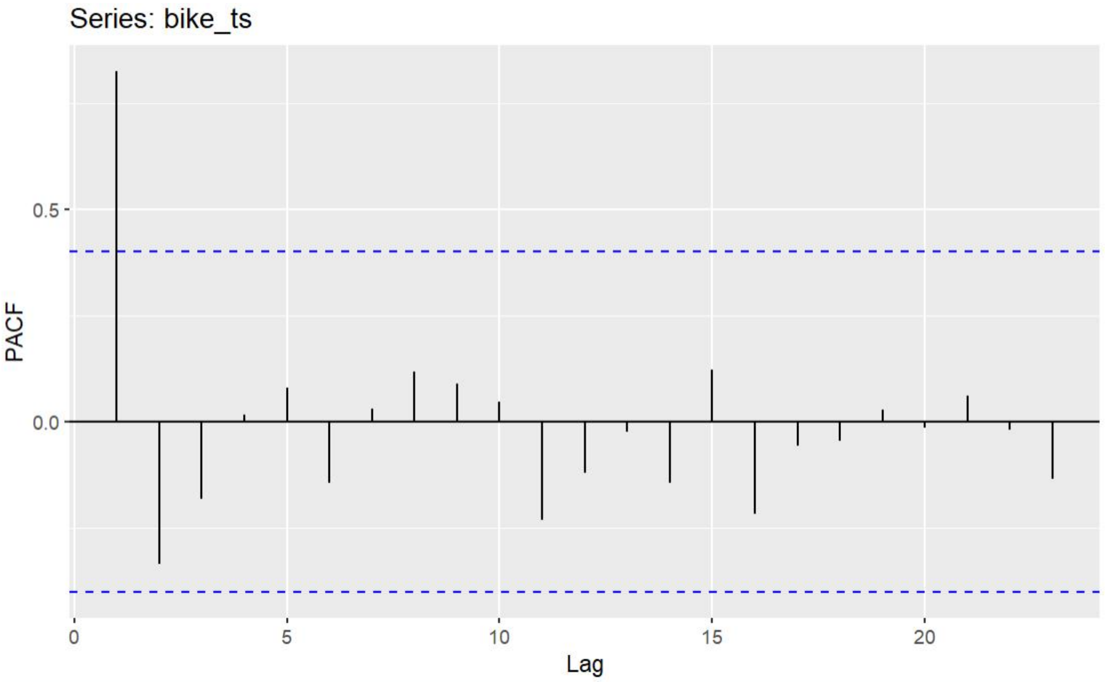
```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
   Min     1Q Median     3Q    Max
-24689  -8605  -3191   7751  49979

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.839e+04  1.204e+04   2.359  0.02920 *
z.lag.1     -1.957e-01  7.982e-02  -2.452  0.02405 *
z.diff.lag   5.557e-01  1.785e-01   3.112  0.00574 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

p-value = 0.02405 < 0.05, reject the null hypothesis. So, the series is stationary.

**2. The values of p, d, and q for the chosen model along with a series plot, plots of the ACF and PACF, and some justification for why you chose these values of p, d, and q *** NOTE: You may present an auto.arima model as your final model. However, you need to include the results from at least one other model that you built on your own - and explain why you felt the auto.arima model was better than the model that you built.**

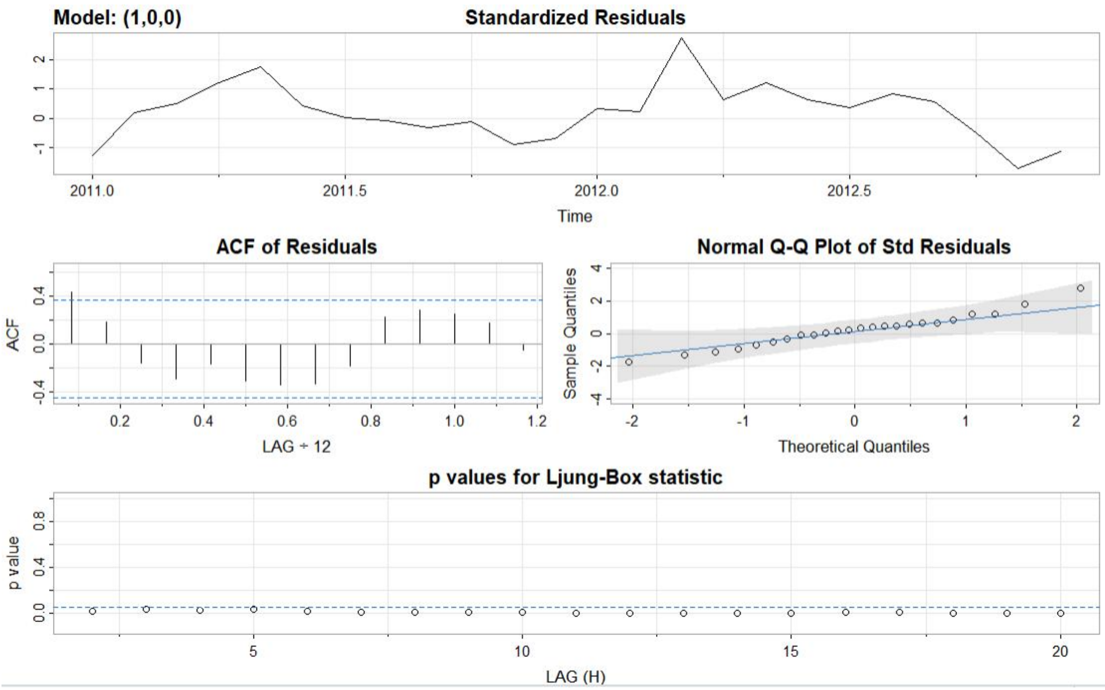
ACF and PACF plots:



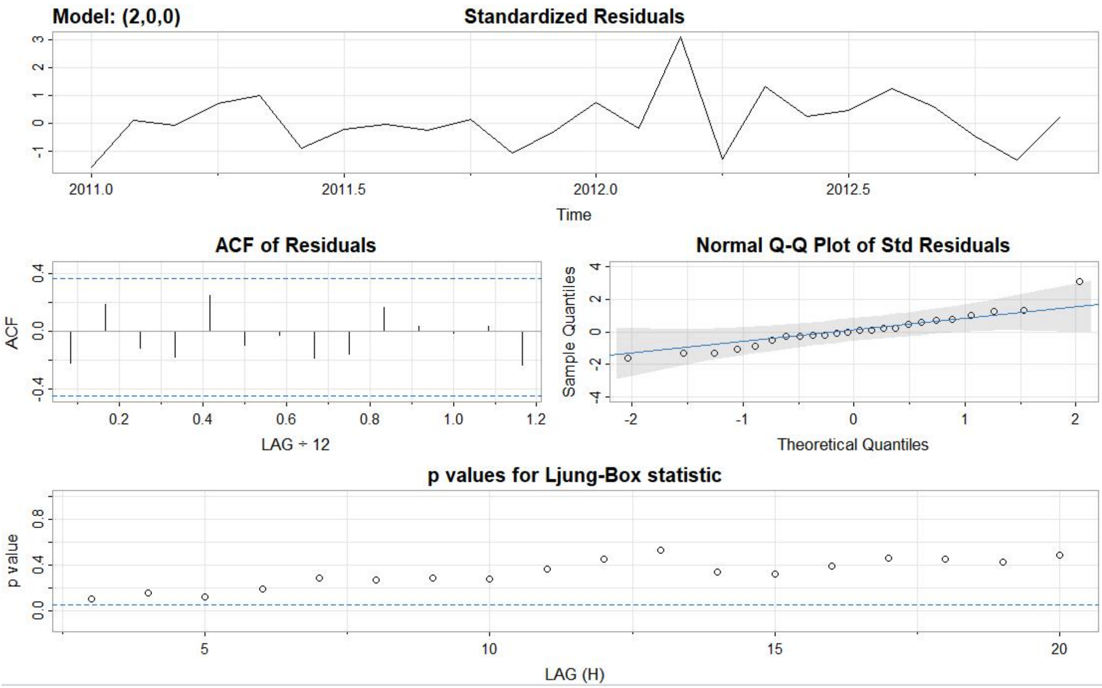

Since the model is stationary, I used d = 0.
Since ACF decays, PACF cuts off quickly, I used significant ACF.
So I started with AR(1) and iterated until residuals are white noise.
When I used ARIMA(1,0,0), in the Ljung-Box test, p-value = 0.03545 < 0.05, reject the null hypothesis. So, residuals are not white noise. This is not a good model.

```
        Ljung-Box test

data:  Residuals from ARIMA(1,0,0) with non-zero mean
Q* = 10.315, df = 4, p-value = 0.03545

Model df: 1.   Total lags used: 5
```



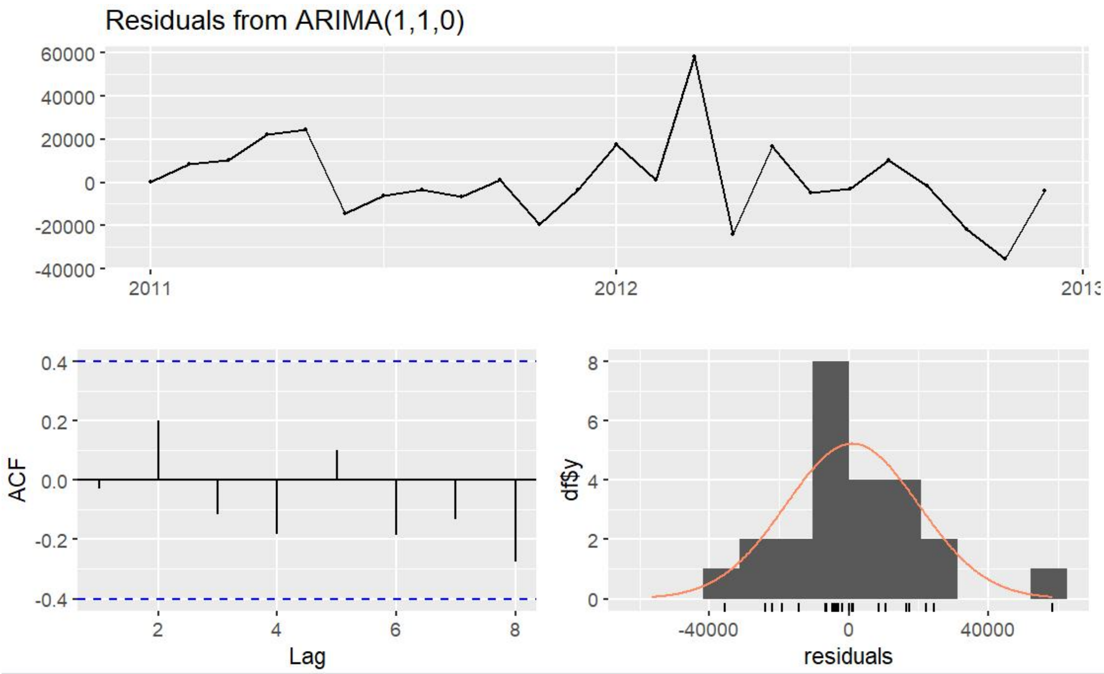

Then I used ARIMA(2,0,0), Since all terms are significant, residuals are white noise, and have low forecast error. Thus, this is a good model.

```
$ttable
         Estimate         SE t.value p.value
ar1        1.4699     0.1569  9.3707  0.0000
ar2       -0.6156     0.1648 -3.7345  0.0012
xmean 124399.0388 24664.5847  5.0436  0.0001

        Ljung-Box test

data:  Residuals from ARIMA(2,0,0) with non-zero mean
Q* = 5.8338, df = 3, p-value = 0.12

Model df: 2.   Total lags used: 5
```




Finally, I used auto.arima model. I found all terms are significant, residuals is white noise, and the model has low forecast error. Therefore, this is also a good model.

```
	Ljung-Box test

data:  Residuals from ARIMA(1,1,0)
Q* = 2.9228, df = 4, p-value = 0.5708

Model df: 1.   Total lags used: 5
```




By comparison, in the ARIMA(2,0,0) model, MAPE = 15.01429; while in the auto.arima model, MAPE = 11.18161. Thus, auto.arima is the better model.

```
Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE      ACF1
Training set 4550.782 22222.04 17143.87 0.6122979 15.01429 0.255094 0.4353424
```

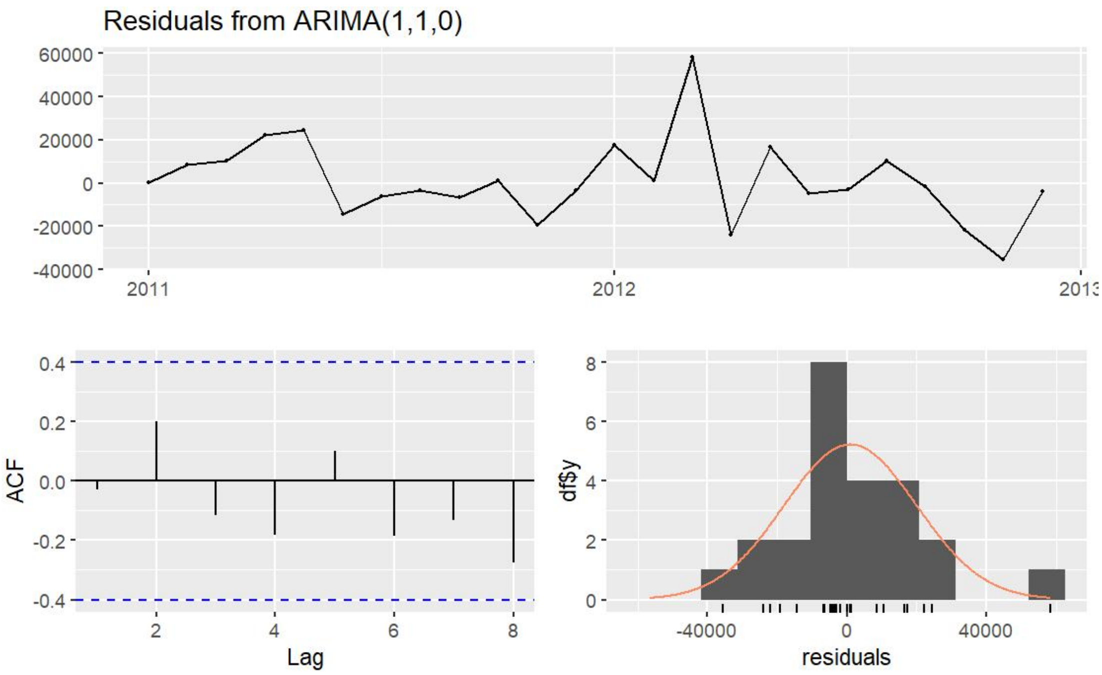
```
Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 1500.392 17645.09 12915.86 -1.691429 11.18161 0.1921829 -0.2194434
```

Therefore, I chose auto.arima model as my final model.

**3. A table of parameter estimates for the chosen model**

```
Coefficients:
         ar1
      0.5383
s.e.  0.1755
```

**4. Residual plots for the chosen model**




**5. A test of white noise for the residuals (either a single p-value from a Ljung-Box test or a plot of the p-values from the Ljung-Box test that is automatically generated by R)**

```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,0)
Q* = 2.9228, df = 4, p-value = 0.5708

Model df: 1.   Total lags used: 5
```

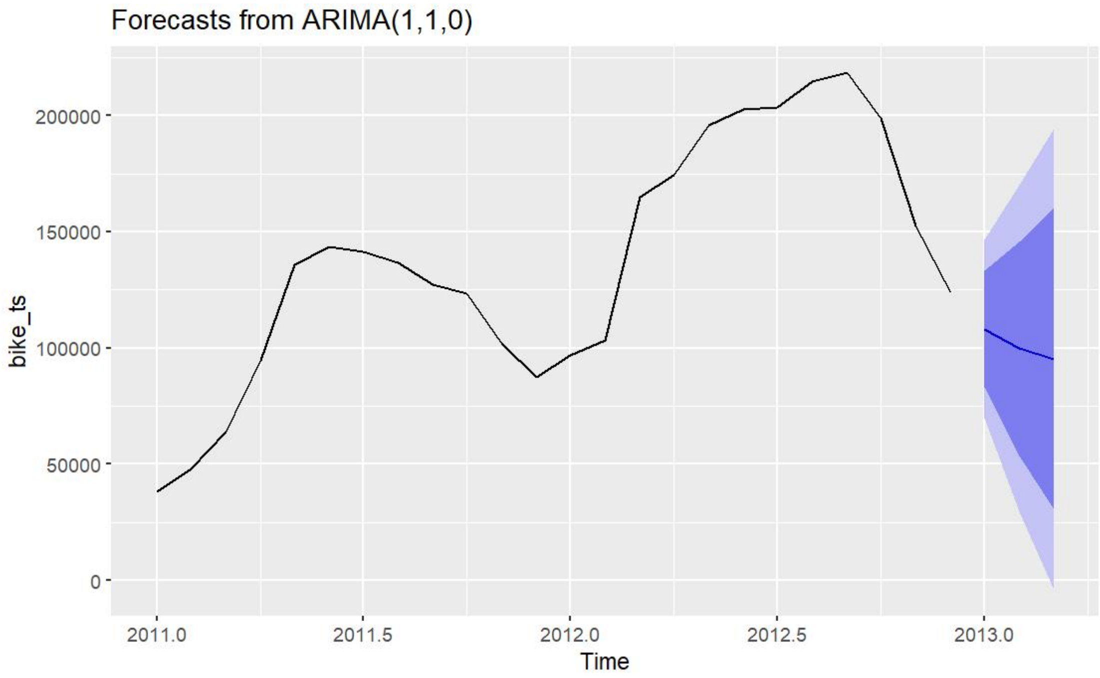
In the Ljung-Box test, since p-value = 0.5708 > 0.05, fail to reject the null hypothesis. So, residuals is white noise.

## 6. The values of the RMSE and MAPE for the chosen model

```
Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE   MASE        ACF1
Training set 931.7514 18722.18 13300.08 1.804659 10.21507 0.1979 -0.02802576
```

RMSE = 18722.18, MAPE = 10.21507

## 7. A forecast plot and the forecast values for the 3 time periods requested above.




| | Point Forecast <dbl> | Lo 80 <dbl> | Hi 80 <dbl> | Lo 95 <dbl> | Hi 95 <dbl> |
|---|---|---|---|---|---|
| Jan 2013 | 108128.27 | 83067.95 | 133188.6 | 69801.813 | 146454.7 |
| Feb 2013 | 99738.80 | 53758.67 | 145718.9 | 29418.257 | 170059.3 |
| Mar 2013 | 95222.63 | 30315.20 | 160130.0 | -4044.712 | 194490.0 |