Airline Stock Price and Delays - UAL model1 only

Student

2022-11-15

Libraries

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(caret)
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

Initial Data

There are 3 initial data files. There is a data file for each of 3 chosen airlines with all BTS monthly delay data

```
#dal_data <- read.csv("delta_lbc.csv")
#sw_data <- read.csv("southwest_lbc.csv")
ual_data <- read.csv("united_lbc.csv")
```

United Model1

This is a model of stock price regressed only on summary delay values reported in flight count increments. There are additional delay values reported in minutes that are not used in this model. The full data set of pre and post COVID months are used.

```
ual_model1 <- lm(Open.Stock.Price ~ Sum.of.carrier_ct
                + Sum.of.weather_ct + Sum.of.nas_ct + Sum.of.security_ct
                + Sum.of.late_aircraft_ct + Sum.of.arr_cancelled
                , data = ual_data)

par(mfrow = c(2,2))
plot(ual_model1)
```
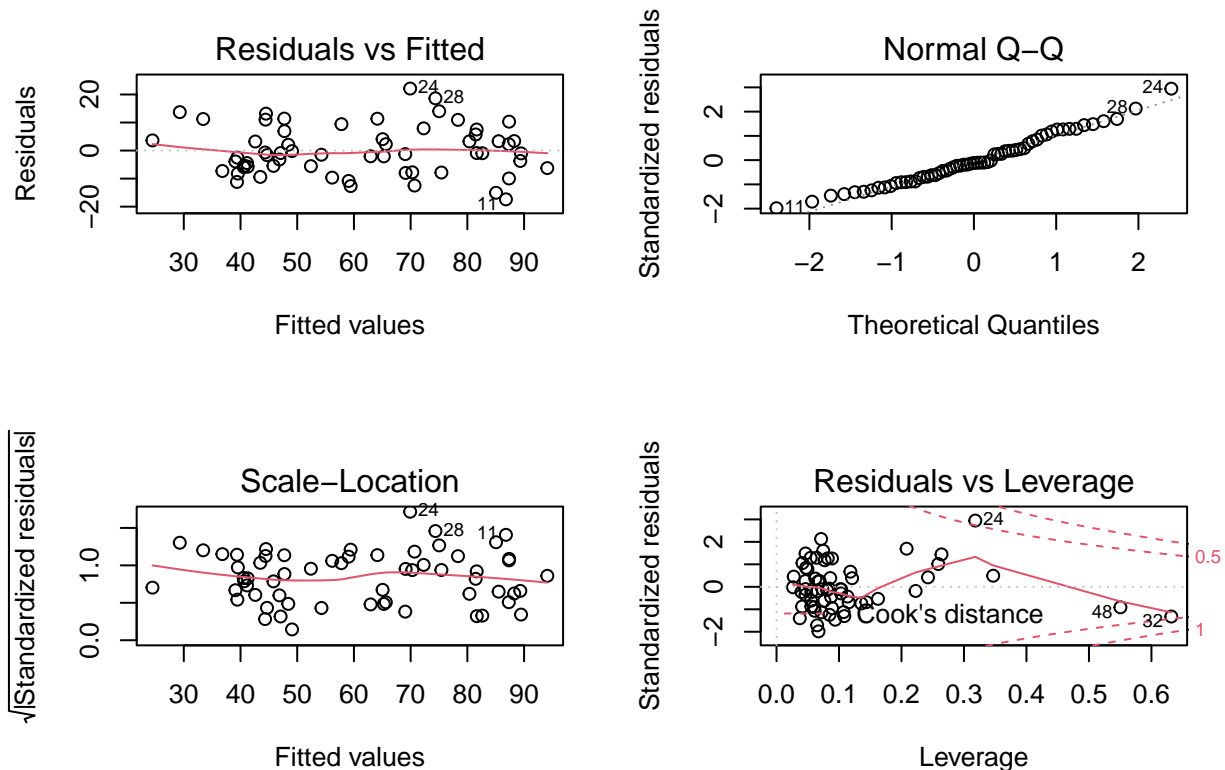


```
par(mfrow = c(1,1))
```

Review of Diagnostic Plots for United Regression Model1

The Residuals vs Fitted plot indicates a linear model absent of non-linear shapes. The Normal Q-Q is good. The Spread-Location plot is good with residuals spread fairly equally. The Residuals vs Leverage plot is a scattered somewhat but not top right or bottom right possible high-influence variables.

Check for multicollinearity

```
vif(ual_model1)
```

```
##        Sum.of.carrier_ct       Sum.of.weather_ct             Sum.of.nas_ct
##                 4.210240                2.388060                  7.633752
##       Sum.of.security_ct Sum.of.late_aircraft_ct      Sum.of.arr_cancelled
##                 1.193354               15.271865                  1.053773
```
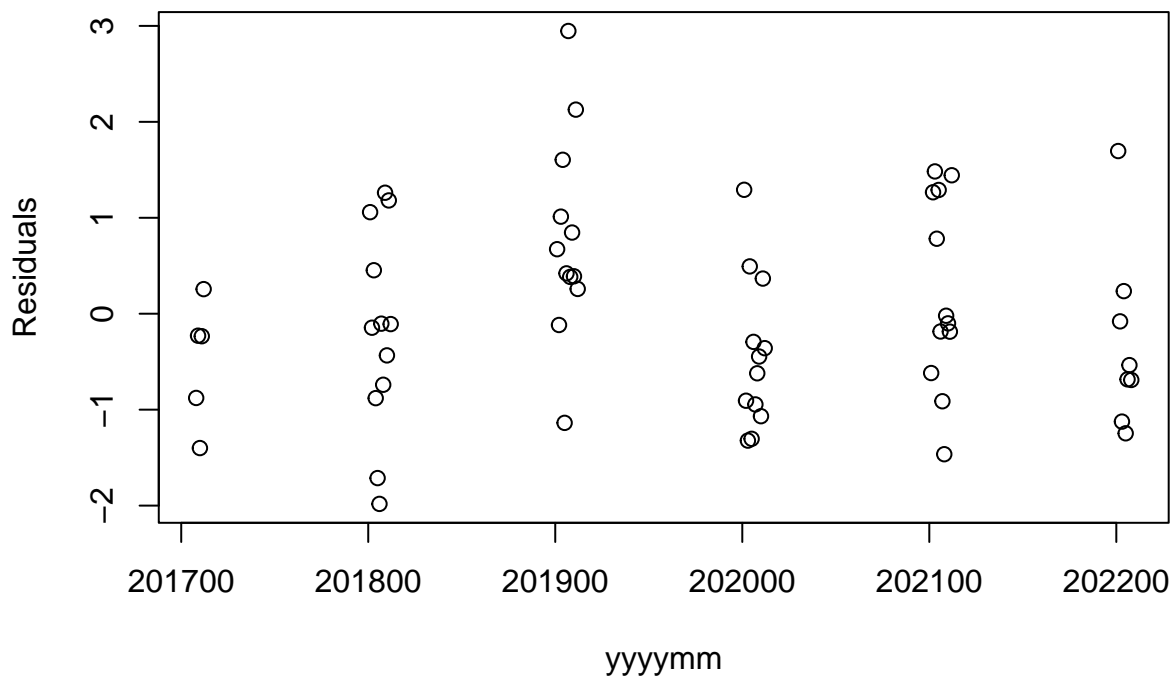
```
max(10, 1/(1-summary(ual_model1)$r.squared))
```

```
## [1] 10
```

Another residual plot

```
residuals<- rstandard(ual_model1)
plot(ual_data$yyyymm, residuals, xlab= "yyyymm", ylab = "Residuals")
```
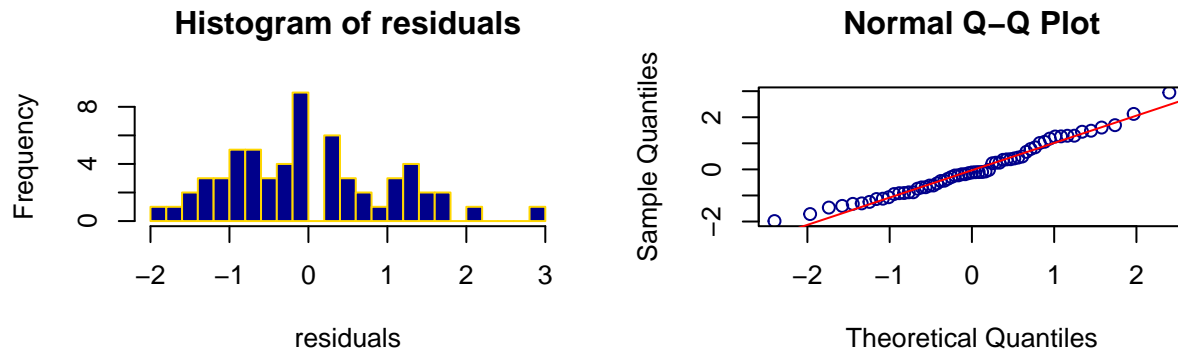

coorelation

```
cor(ual_data[, c('Open.Stock.Price','Sum.of.arr_cancelled')])
```

```
##                      Open.Stock.Price Sum.of.arr_cancelled
## Open.Stock.Price            1.0000000           -0.3057296
## Sum.of.arr_cancelled       -0.3057296            1.0000000
```

3

```r
par(mfrow=c(2,2))
hist(residuals,nclass=20, col="darkblue", border="gold",main="Histogram of residuals")
qqnorm(residuals,col="darkblue")
qqline(residuals,col="red")
```

Histogram of residuals

Normal Q–Q Plot

View model summary

```r
summary(ual_model1)
```

```
##
## Call:
## lm(formula = Open.Stock.Price ~ Sum.of.carrier_ct + Sum.of.weather_ct +
##     Sum.of.nas_ct + Sum.of.security_ct + Sum.of.late_aircraft_ct +
##     Sum.of.arr_cancelled, data = ual_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.400  -5.901  -1.010   5.740  22.119
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             40.2938833  3.2669519  12.334  < 2e-16 ***
## Sum.of.carrier_ct       -0.0097907  0.0021966  -4.457 4.23e-05 ***
## Sum.of.weather_ct       -0.0168770  0.0082410  -2.048  0.04544 *
## Sum.of.nas_ct            0.0154198  0.0020535   7.509 6.15e-10 ***
## Sum.of.security_ct      -0.5181368  0.3495411  -1.482  0.14406
## Sum.of.late_aircraft_ct  0.0009860  0.0030146   0.327  0.74487
## Sum.of.arr_cancelled    -0.0016752  0.0006269  -2.672  0.00995 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.089 on 54 degrees of freedom
## Multiple R-squared:  0.8269, Adjusted R-squared:  0.8076
## F-statistic: 42.99 on 6 and 54 DF,  p-value: < 2.2e-16
```

Review of the Fit and Significance United Model1

Sum.of.carrier_ct, Sum.of.weather_ct, and Sum.of.arr_cancelled are all significant at a greater than 95% level and have a correct negative value coefficient. The Sum.of.nas_ct coefficient is positive and not in line with field knowledge, this is identical to the Delta Airlines and SW Airlines models that had low R-squared values. The R-sqared value is above the financial industry standard of above 0.7 for a high level of significance at 0.8269. The null hypothesis that the coefficients are not significant can be rejected.

Overview Conclusion

Delay counts for Southwest do not model for fit or significance with the data. Delay counts for Delta are are closer for fit and significance for the Sum.of.arr_cancelled coefficient but the R-squared value is still too low. United models well on Sum.of.carrier_ct, Sum.of.weather_ct, and Sum.of.arr_cancelled for fit and signifcance of the model and correct sign of the coefficients.