Crude Oil Price vs. Airline Stock Price

using Yahoo! Finance data

Authors: Yicheng Fu, Yishan Hou, Drew Laake
Spring 2022

Introduction

The rapid rise in the price of crude oil in the futures market during the early months of 2022 inspired us to observe trends of oil futures contract prices over the last few years, as well as its potential effects on other industries. Specifically, we became interested in the association between daily returns for crude oil prices and daily returns of stock prices for major airline companies, as the air travel industry is heavily reliant on fuel (and current fuel prices) for operation.

We collected our data from Yahoo Finance.[1] We downloaded historical data for each stock, United Airlines (UAL) and American Airlines (AAL), which gives the following values of each trading day on the Nasdaq stock exchange for the last five years: Open, High, Low, Close, Adj Close, and Volume. We did the same for futures contracts for crude oil (CL=F) on the NY Mercantile exchange. From here, we calculate daily returns for each stock and oil price. Finally, we determined two periods in which to evaluate each data set:

1. A period of rising crude oil price: 11/30/2021 - 12/24/2022
2. A period of declining crude oil price: 10/3/2018 - 12/28/2018

This gave us a total of six data frames to analyze.

We proposed five research questions that we were interested in studying:

1. Is there a correlation between the returns of the two airline companies?
2. Is there a difference in the returns of stocks in a rising period of oil price vs. a declining period of oil price?
3. Is the correlation between the returns of the two airline companies the same during the two periods (assessing uncertainty)?
4. Why do we choose to study return instead of stock price? Whose serial correlation is smaller?
5. How can we predict future stock prices?

We were unsure whether the stock and crude oil prices would be positively or negatively related.

1. If the correlation is negative, we suspect that airlines benefit from lower/decreasing fuel prices since their operating costs are lower, thus profit will be greater—ultimately resulting in a more attractive investment for traders, driving the stock price up by increased demand for stocks. The opposite is true when oil prices are higher/increasing.
2. If the correlation is positive, perhaps this is a sign that the overall economy and confidence of investors drive the prices of oil and stocks together (hence whether the economy is in a period of expansion or recession outweighs the effect on airlines' operation costs).

But, of course, we recognize that many variables affect the stock prices of these two companies beyond just the current contracts for the purchase of fuel and the state of economic activity. Industry trends, dynamic market shares, publicity of these airline brands, prices of other necessary operating components

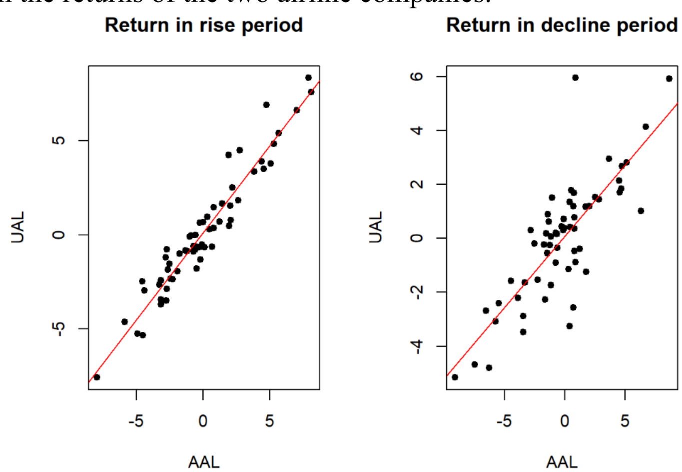
[1] https://finance.yahoo.com/calendar/

like steel or wages, and countless other factors affect each company's wellbeing, its operations, its perception by investors, and ultimately its stock price. Due to these confounding factors, any of our findings must be approached with a fair amount of uncertainty. Indeed, in this project, correlation does not promise causation.

Analysis

1. *Is there a correlation between the returns of the two airline companies?*

Investors may want to know the association between two stocks when trying to diversify their investments or evaluate the risk of certain ETFs, mutual funds, etc. For instance, if there is a strong positive association between the returns of these companies, then the investor may be taking on more risk. Evaluating this correlation is a **bivariate** analysis involving a *numerical x* (returns of AAL) and a *numerical y* (returns of UAL) variable. We use linear regression to explore whether there is a correlation between the returns of the two airline companies.



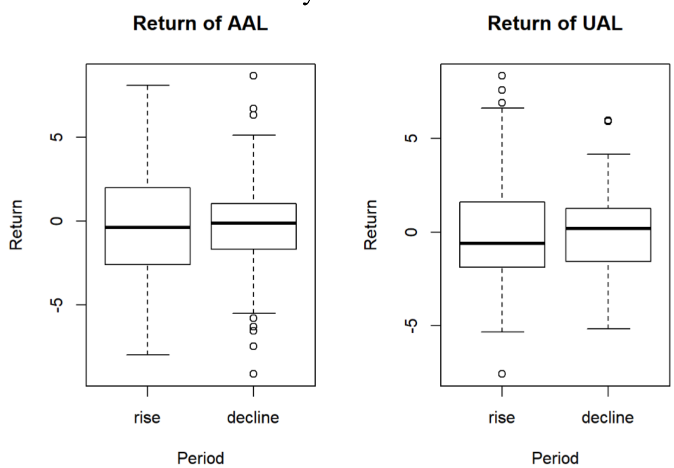

```
##
## Call:
## lm(formula = UAL_rise$ret ~ AAL_rise$ret)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4617 -0.5660 -0.2034  0.5364  2.3811
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.12751    0.11638   1.096    0.278
## AAL_rise$ret  0.92344    0.03471  26.603   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9015 on 58 degrees of freedom
## Multiple R-squared:  0.9243, Adjusted R-squared:  0.923
## F-statistic: 707.7 on 1 and 58 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = UAL_decline$ret ~ AAL_decline$ret)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5559 -0.7283  0.1185  0.6819  5.4016
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.08915    0.17454   0.511    0.611
## AAL_decline$ret  0.52784    0.05065  10.422 6.57e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.348 on 58 degrees of freedom
## Multiple R-squared:  0.6519, Adjusted R-squared:  0.6459
## F-statistic: 108.6 on 1 and 58 DF,  p-value: 6.567e-15
```

The p-value of the coefficient is very small in both the rise and decline periods. At a 5% level of significance, we conclude that there is indeed a strong correlation in each period.

2. *Is there a difference in the returns of stocks in a rising period of oil price vs. a declining period of oil price?*

Evaluating the historical returns of each stock may provide investors with insight that they can use for future trades. For instance, if an investor is unsure whether to buy (or sell) United or American stock, but they have a prediction about future oil prices, then past returns of the two stocks in a rising/declining period of oil prices can help them decide what stock they should buy (or sell) for this period. We can calculate the uncertainty of these results and make our final conclusion about whether these results are significant. This is a **bivariate** analysis involving a *categorical x* (time period) and *numerical y* (returns of AAL/UAL) variable. We use a boxplot to visualize the return of the two companies and use a t-test to assess the uncertainty.




```
t.test(AAL_rise$ret, AAL_decline$ret)
```

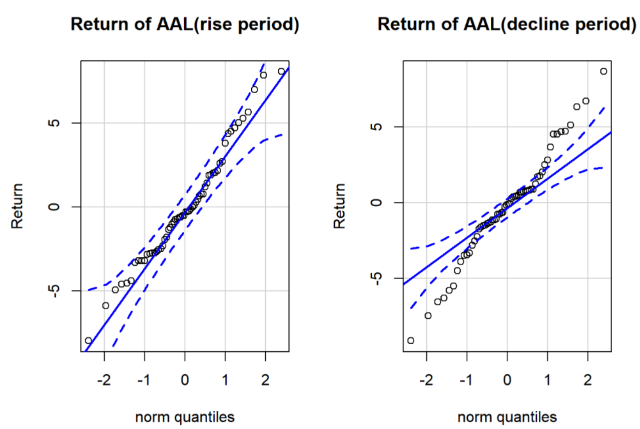
```
##
##  Welch Two Sample t-test
##
## data:  AAL_rise$ret and AAL_decline$ret
## t = 0.37363, df = 117.93, p-value = 0.7093
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.004298  1.471401
## sample estimates:
##   mean of x   mean of y
## -0.01911137 -0.25266311
```

```
t.test(UAL_rise$ret, UAL_decline$ret)
```

```
##
##  Welch Two Sample t-test
##
## data:  UAL_rise$ret and UAL_decline$ret
## t = 0.30139, df = 105.44, p-value = 0.7637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8595361  1.1676902
## sample estimates:
##   mean of x   mean of y
## 0.10986337 -0.04421367
```

These **p-values are all very large**, certainly relatively large to a significance level of 0.05. We conclude that the data does not provide sufficient evidence that there is a difference in the returns of stocks in a rising period of oil price vs. declining period of oil price.

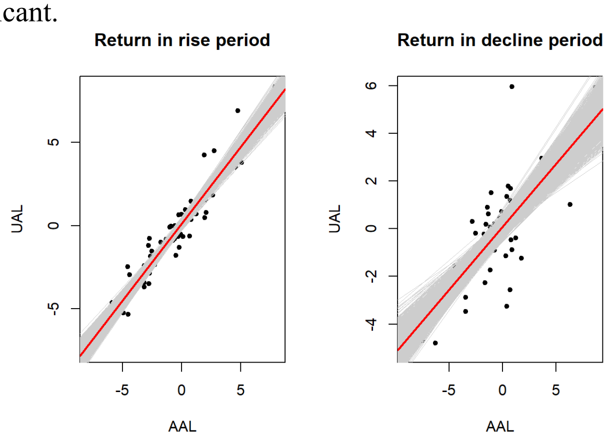
However, we can only use classical procedures such as a t-test to quantify uncertainty around mean return of different periods if the set of returns is roughly normally distributed. We use qqplot to examine this:



For the decline period, many points don't fall within the 95% confidence band around the straight line in each Q-Q plot, showing that the returns are not normally distributed in the decline period. From the central limit theorem we know that this problem can be solved if we have adequate data. Our analysis would be more accurate if we collect more data, however, outstanding trends in oil prices only exist for so long, which imposes a restriction on the data we are able to collect.

3. *Is the correlation between the returns of the two airline companies the same during the two periods (assessing uncertainty)?*

In general, investors, economists, and others may be interested in whether the returns of two stocks in the same sector are more correlated in periods of growth or recession. Our problem is slightly more complex: we did not see a significant difference in the returns of stocks during the period of rising oil prices vs falling oil prices. Thus, we cannot say with certainty which of these periods is indicative of higher or lower returns for stocks. Still, we can observe whether we see a higher correlation in stocks in an oil price rise period vs. oil price decline period, or vice versa. This is a **multivariate** analysis involving *numerical x* (returns of AAL), *numerical y* (returns of UAL), and *categorical z* (period) variables. We generate a bootstrapped data set 10,000 times and store the values of correlations and the predicted line for each data set. The following figure shows the original fitted line in red and the new fitted lines in gray. Then we use a t-test to see whether the difference in correlation is statistically significant.

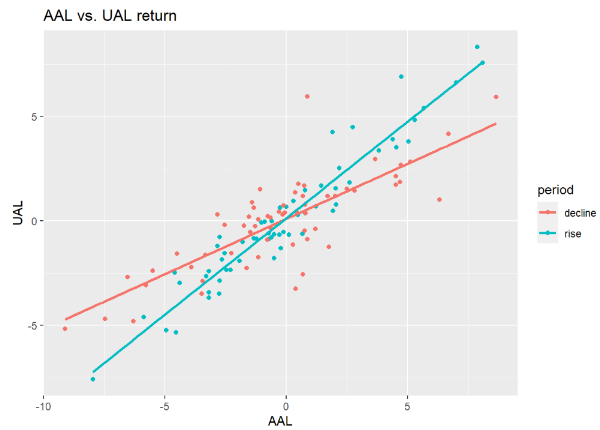



```
t.test(corr_rise.boot, corr_decline.boot)
```

```
##
##  Welch Two Sample t-test
##
## data:  corr_rise.boot and corr_decline.boot
## t = 252.66, df = 10589, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1536111 0.1560132
## sample estimates:
## mean of x mean of y
## 0.9604188 0.8056066
```

We observe higher correlation between the stocks in a period of rising oil prices. The p-value of our t-test is less than 0.05, so at 5% level of significance the data provides sufficient evidence that the correlation in different trends of oil prices are different.

We can also use advanced graphics with ggplot() to visualize the two correlations with color representing each period. This color-coded scatterplot shows a clear difference between the slopes of the two lines.



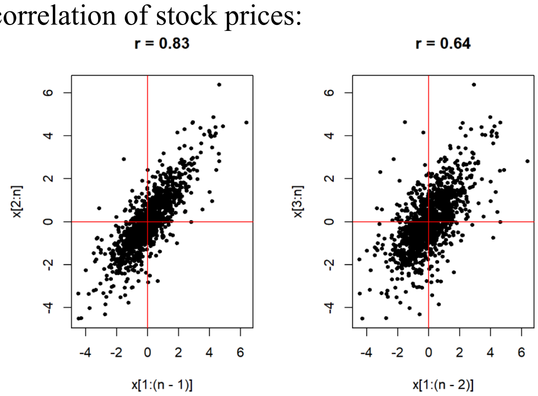
4. *Why do we choose to study return instead of stock price? Whose serial correlation is smaller?*

    We check the memory (serial correlation) in the random component of our time series data after decomposing by calculating the autocorrelation $cor(x_t,\ x_{t-1})$ and $cor(x_t,\ x_{t-1})$.

    In fact, we can just study the original stock price and use it to do analysis like linear regression. So why bother to calculate the return first? Serial correlation plays an important role here.

    Serial correlation is the relationship between a given variable and a lagged version of itself over various time intervals. It measures the relationship between a variable's current value given its past values. A variable that is serially correlated indicates that it may not be random, so the quantitative techniques we applied to the dataset are not reliable. We test whether our series has serial correlation by calculating the autocorrelation and use the Ljung-Box test.

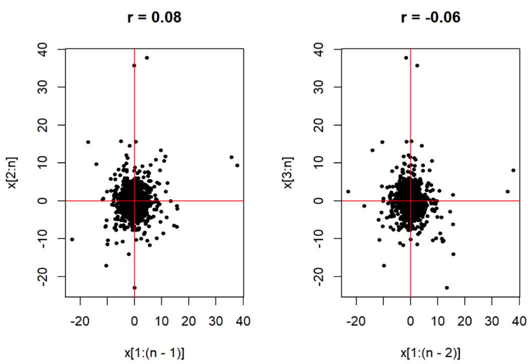
    Autocorrelation of stock prices:




```
Box.test(AAL.ts.comp_full$random, type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  AAL.ts.comp_full$random
## X-squared = 839.05, df = 1, p-value < 2.2e-16
```

    Autocorrelation of returns:



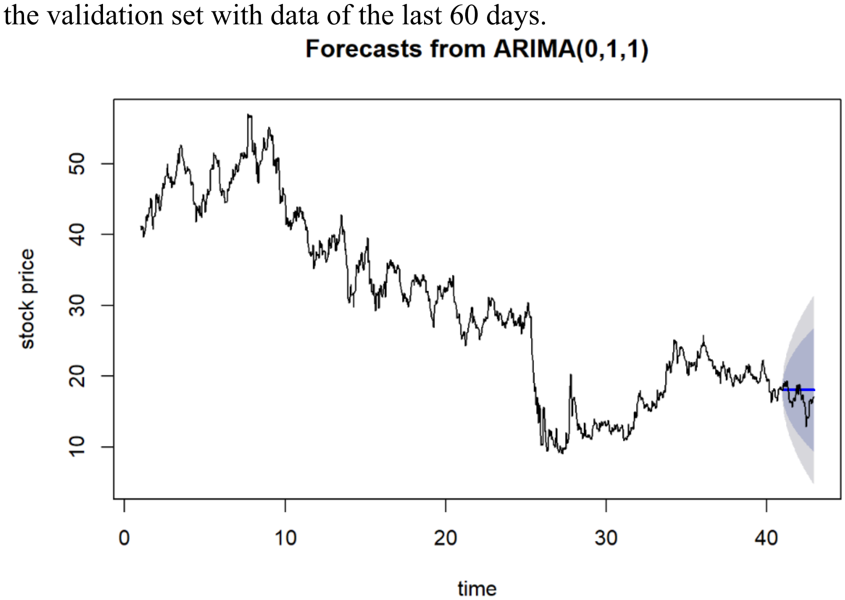

```
Box.test(AAL_return.ts.comp$random, type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  AAL_return.ts.comp$random
## X-squared = 8.7129, df = 1, p-value = 0.00316
```

The returns have a much smaller correlation compared to stock prices, indicating that return is less serial correlated and is a more appropriate feature to study. However if we do a Ljung-Box test to see whether there exists serial correlation, **both p-values are less than 0.05**. This means the variable return also has the problem of serial correlation and our analysis can be wrong because of this problem. Time can be a confounding factor when we are looking for correlation in our previous analysis.

5. How can we predict future stock prices?

It is always difficult to predict the future prices of a stock. In this project we try to use the Autoregressive Integrated Moving Average Model (ARIMA) to predict the stock price. We split our data into a training set and a validation set and fit the model using the training set with 1200 days of data, and then evaluate our model using the validation set with data of the last 60 days.



```
p_d_q <- auto.arima(AAL.ts)
p_d_q
```

```
## Series: AAL.ts
## ARIMA(0,1,1)(1,0,0)[30]
##
## Coefficients:
##          ma1     sar1
##       0.0739  -0.0147
## s.e.  0.0292   0.0292
##
## sigma^2 estimated as 0.6694:  log likelihood=-1459.7
## AIC=2925.39   AICc=2925.42   BIC=2940.66
```

```
Box.test(fit$residuals, type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  fit$residuals
## X-squared = 0.0028998, df = 1, p-value = 0.9571
```

We use the function auto.arima to fit ARIMA models and select model order. The result of auto.arima shows that the best (p,d,q) of the ARIMA model should be (0,1,1). We can see that the true values are all within the 95% confidence interval of our prediction.

Also, we do Box.test for the residuals of the model and see that p-value is very close to 1. We fail to reject the null hypothesis and say that the autocorrelations of the residuals are very small, so we can conclude that the model does not exhibit significant lack of fit.

However, this kind of model may not be very helpful in trading. What matters most in trading is whether the price will increase or decrease tomorrow, not whether the predicted price will fall in the confidence interval. Knowing that the data will fall in an interval does not increase the probability you will earn money if you buy this stock. So if we want to make more money by predicting the stock prices, maybe we should instead predict the probability of seeing a higher price tomorrow instead of the precise price.

Summary

Our findings indicate that there's a strong positive correlation between the stock prices of American Airlines and United Airlines, as expected.

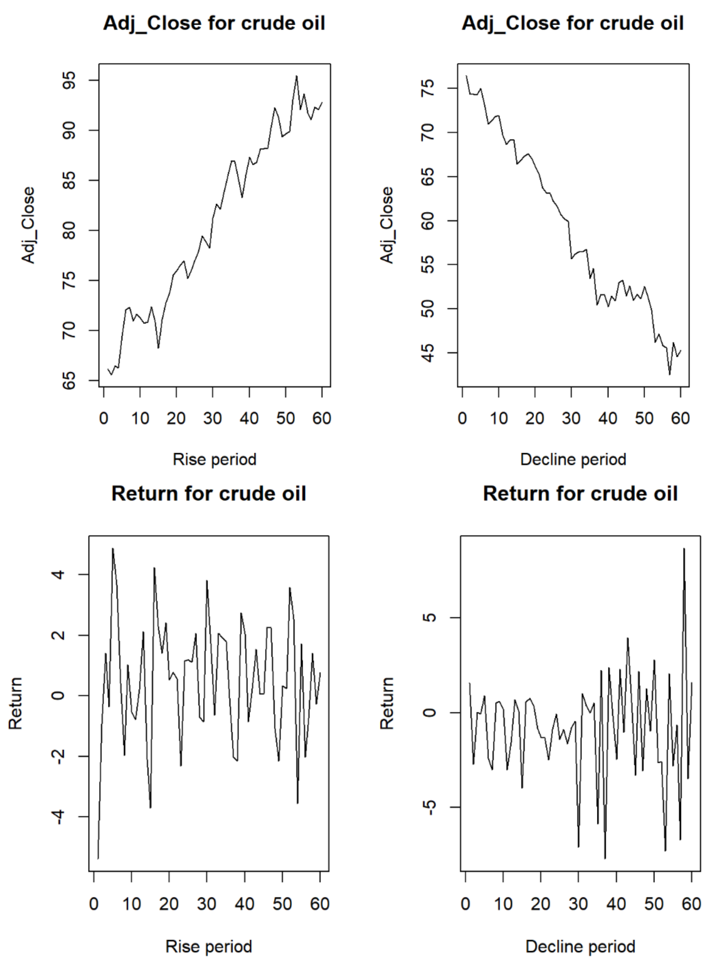
However, we fail to prove our hypothesis that there's a difference in the returns of stocks in a rising period of oil price vs. a declining period of oil price. Ample research exists on "correlation breakdowns," where stocks are more correlated during periods of crises with low market returns and high volatility. In our study, we did not find strong enough evidence to conclude that the stock returns were different for different periods of oil prices. So we cannot assess which period of oil prices would be the "crisis" period for airline stocks. We found a higher Pearson correlation between stocks in the period when oil price was rising (further depicted in appendix #2), indicating by reverse logic that rising oil prices are perhaps a crisis period for airline stocks, but there are many factors that were not taken into account which determine the state of the market, and the higher correlation during this time could be the result of any number of these other factors.

As previously addressed, more confounding factors specifically relating to the operation of airline companies and the appeal of investments include steel and other material prices which relate to the cost of buying/leasing aircrafts, economic and political policies, customers' outlooks on travel, seasonal periods of high travel and low travel and how investors react to the possibility of future dividends paid by the firms.

Finally, with only the data from two airline companies, our study has some limitations that might contain uncertainty and bias: 1) confounding factors specifically relating to the operation of airline companies and the appeal of investments, 2) the result of the Ljung-Box test shows the return is still serially correlated. To better visualize the relationship between oil and airline stock prices, more data from the industry is needed. We would like to (1) collect more data so that even if our returns are not normally distributed, the t test can be reliable (2) find a better way to process the dataset to relieve the problem of serial correlation (3) try to predict the probability of seeing a higher price tomorrow instead of the precise price, making our prediction more practical.

Appendix

1. The following figures show how the stock prices and returns of crude oil change during the rising and declining period. We can clearly see the increasing and decreasing trend of the stock price.



2. From this R calculation, we can see that the stocks have a higher correlation in the rise period.

```
93  cor(AAL_rise$ret, UAL_rise$ret)
94  cor(AAL_decline$ret, UAL_decline$ret)
95  ```

[1] 0.9613824
[1] 0.8073956
```

If we know that stocks are more correlated when the market is worse, then the logical reasoning would be to say that the market is worse for airline stocks when oil prices are rising. However, because we did not see a statistically significant difference in returns of stocks for each period, this may not be a reasonable conclusion. Likely, the differences in correlation have to do with other market forces than the price of oil.