Identifying and Testing Stock Indicators

Austin McGhee

8/20/2021

View the R markdown code

Dollar Volume

In the price data we have dates and daily values for OHLC, adjusted and share volume. We will create a new column called Dollar Volume (DV) which is the adjusted price multiplied by volume. DV can be used to get an idea of money flow. All prices are in US Dollar.

Using the DV we can measure both individual stocks and groups. For example the DV for Didi may be $1 billion, and the sum total of DV for all US listed China Tech (>$3 billion Market Cap) may be $30 billion. We could also calculate the average DV for the top five symbols and analyze it's relationship to the mean DV for all US listed China Tech. We will refer to this filtered US listed China Tech as China Tech moving forward.
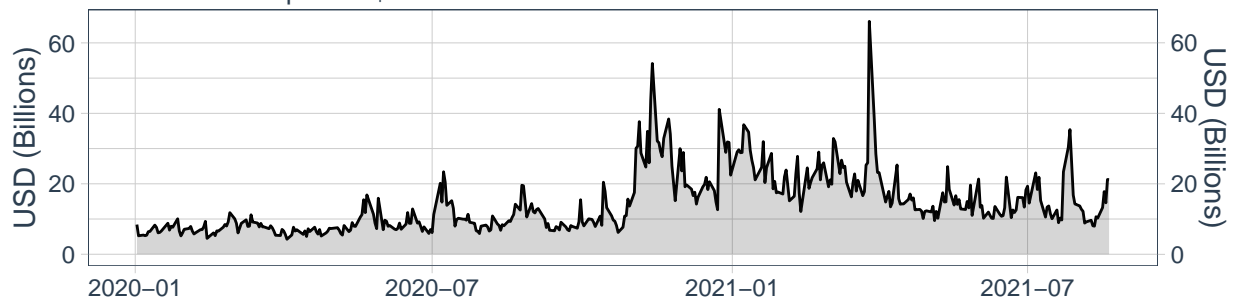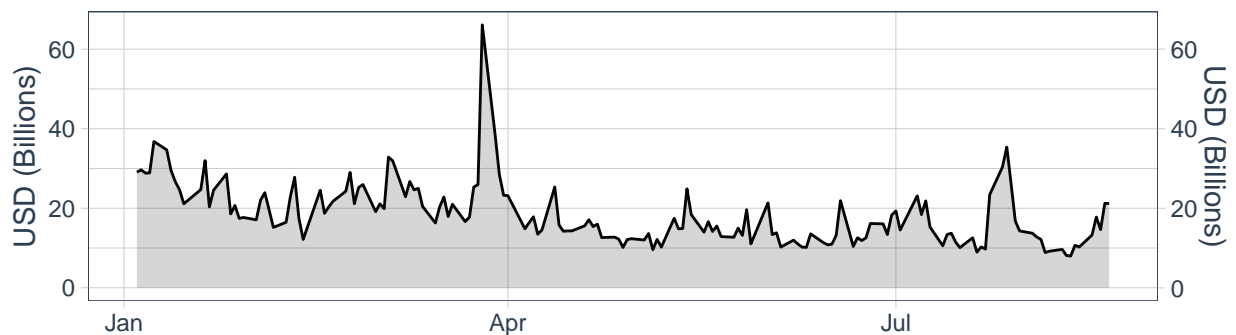
We can see in the figure above that the sum of DV for China Tech is quite volatile but its general value tends to shift on a quarterly or annual basis. From this data alone we can get an idea of the money flow in China Tech, but it doesn't provide insight into the direction of the underlying share price.

Distribution

How is DV distributed among the individual stocks making up China Tech? We could be looking at a basket of stocks with at similar DV, or a minority few making a disproportionate contribution. In the chart below, each column represents a stock and the y axis (US Billions) is a measure of the average daily DV for the year.
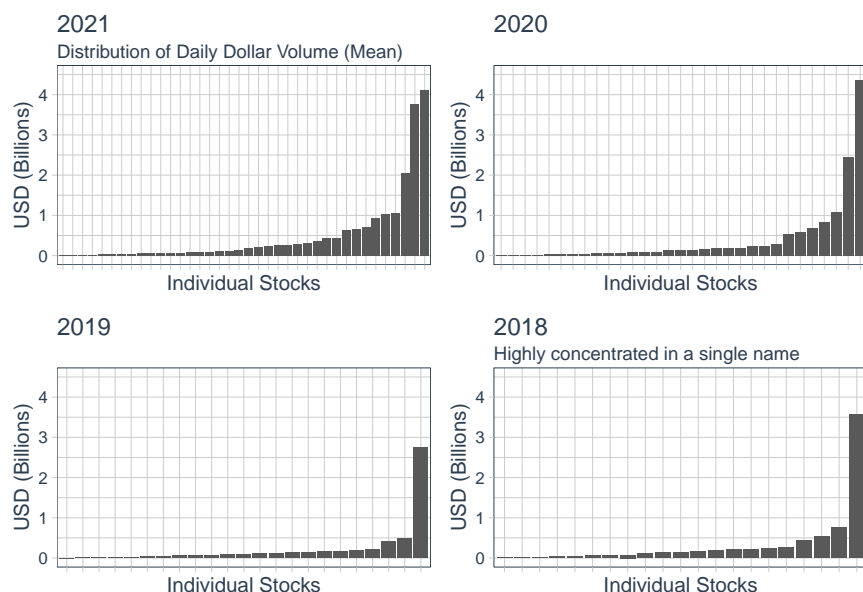
We can see that DV is concentrated in 1-2 stocks each year and have progressively reduced concentration since 2018. In the last four years, the first where there was a close runner-up to top rank was in 2021. The next section identifies the individual stocks in these charts and explores them further.

Individual Stocks

Which stock is doing most of the heavy lifting? We can see that Alibaba (BABA) has been the DV leader every year. The second spot has been Nio (NIO) or Baidu (BIDU). These charts cover 2019-2021 but have replaced 2018 with a view of the last 3 months. Based on this sample we can see that the composition of stocks in the 3 month view is the same as the annual view, suggesting composition remains the same in shorter time horizons.
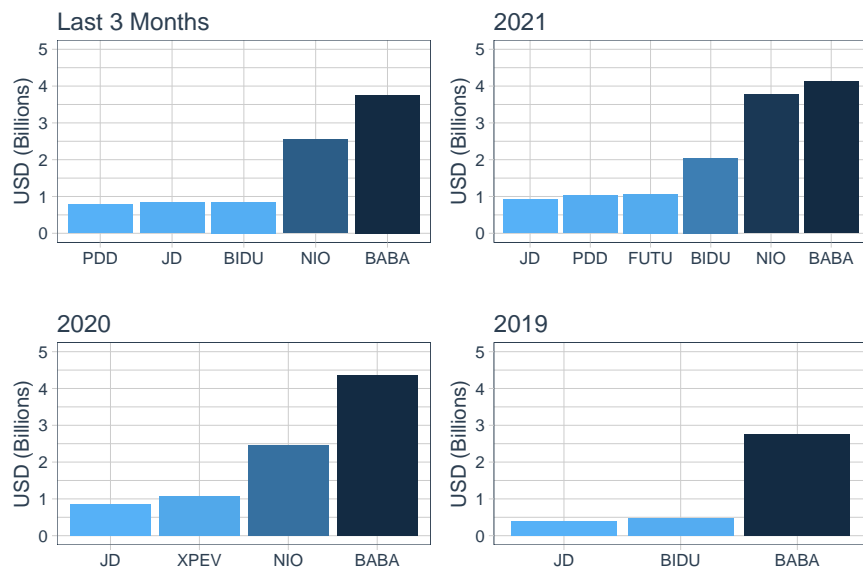

The top stocks were selected by filtering by a minimum DV (Mean) of $500 million and a minimum market capitalization of $3 billion (August, 2021). In 2019 the minimum DV (Mean) filter was reduced to $250 million to show more results, otherwise it would only show Alibaba.

Alibaba Comparisons

With Alibaba having the highest DV of China Tech we will focus our analysis on it. How does Alibaba's performance compare to the rest of China Tech? Can we find relative indicators in the DV by comparing Alibaba to the rest of China Tech?

First we will isolate Alibaba's DV and divide it by the China Tech DV (Sum). This tells us the percent of China Tech (Sum) that Alibaba contributes on a daily basis. We will focus on dates in the period January 2020 to August 2021.
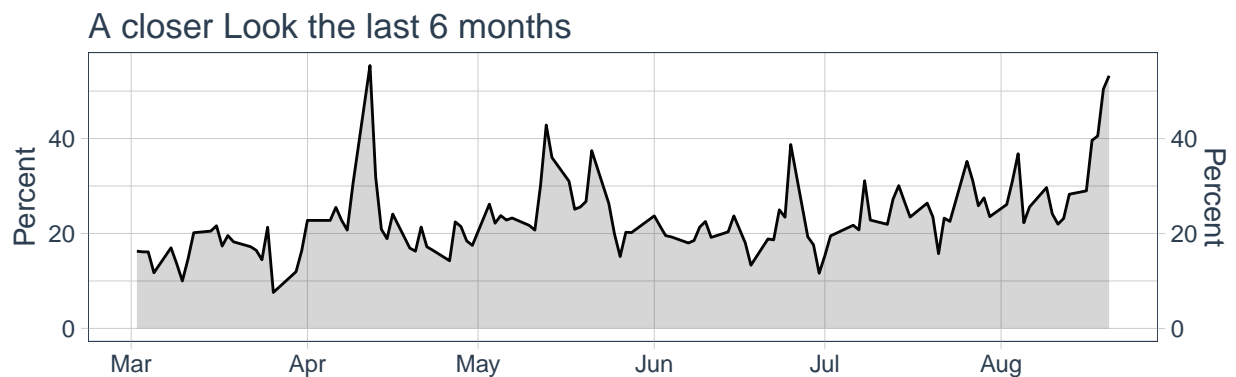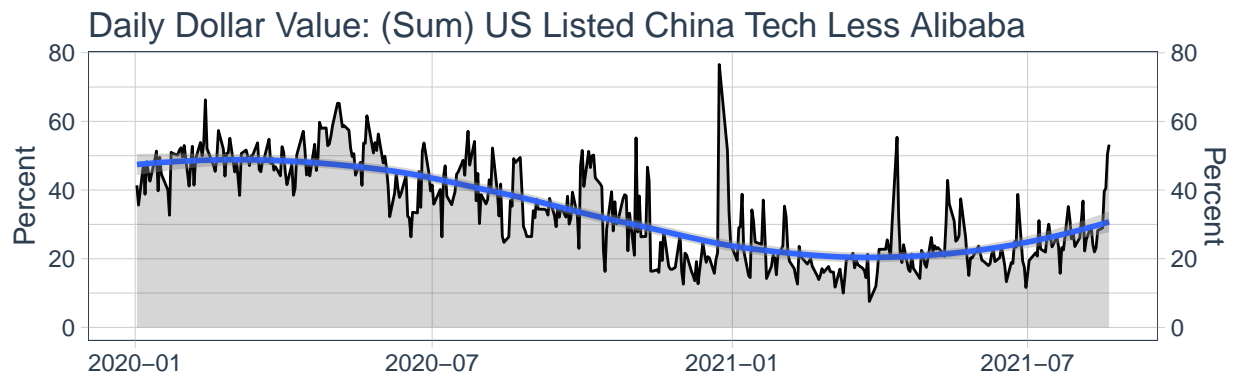
Looking at the data we can see that Alibaba is key contributor to the China Tech DV. The relationship is volatile, but it's trendline ranges from 20-50%. It's important to note that this an absolute measure, which is why I think it is important to reference shorter time periods. Initial public offerings for instance will result in another stock that increases China Tech DV and will reduce the contribution of Alibaba to the total.

Spread Comparison
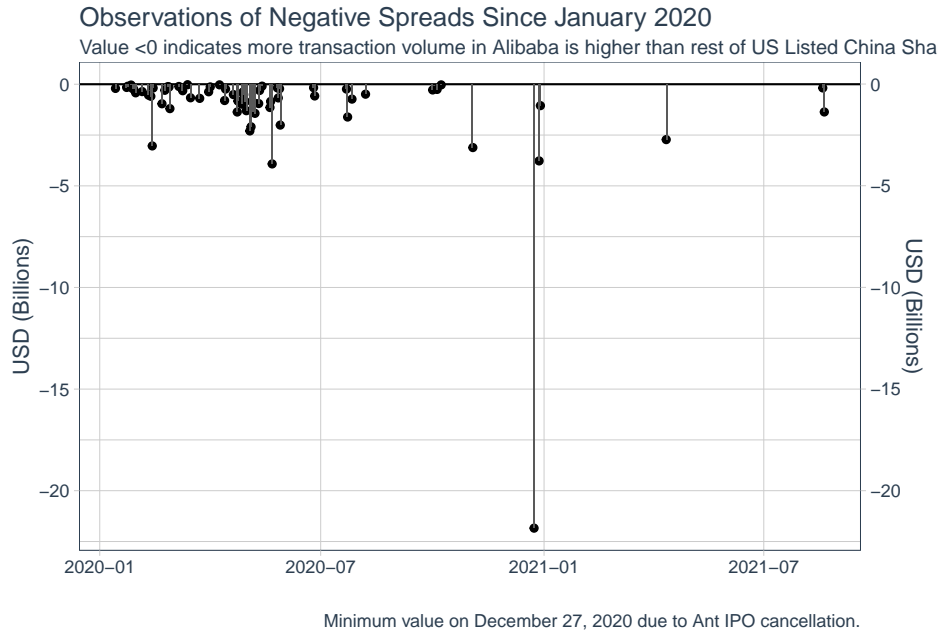
Next we will look at the DV (Sum) spread between Alibaba and China Tech. To calculate this we will first isolate Alibaba DV and deduct Alibaba from China Tech DV. Now we are working with Alibaba DV and China Tech without Alibaba. Next we take China Tech without Alibaba DV (Sum) and subtract Alibaba DV to create the spread.
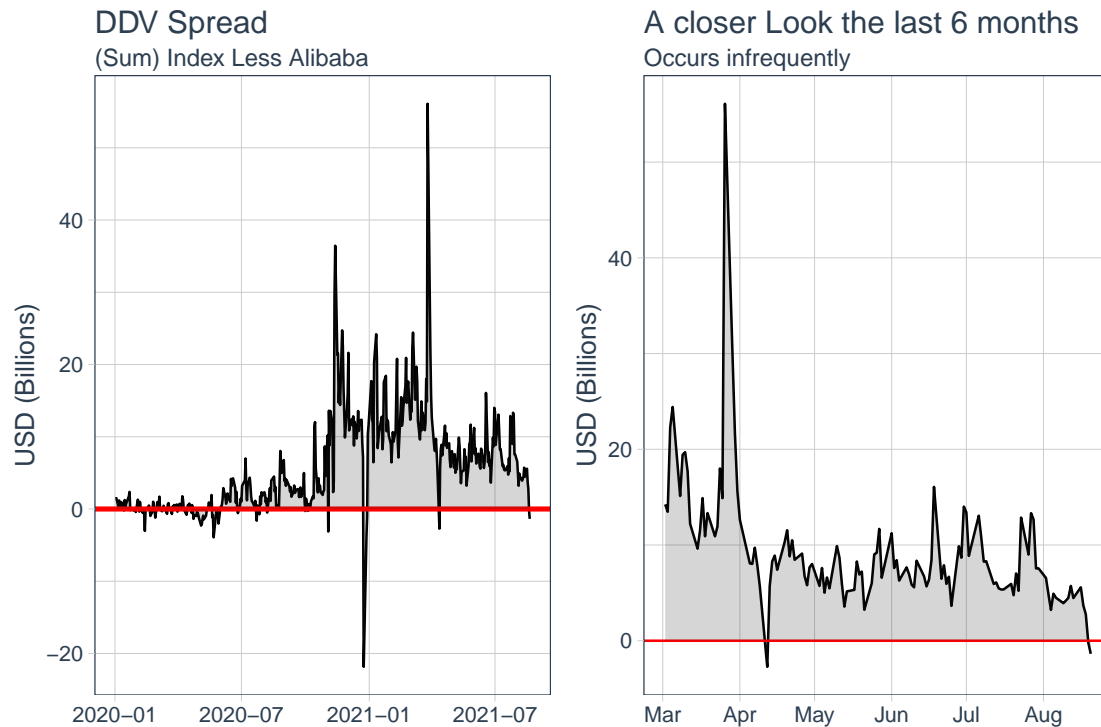
We can see that the is volatile, but positive most of the time. What does it mean when it's negative or positive? Can be looked at as Alibaba having higher or lower DV compared to China Tech. The values are absolute, we can reference the spread value with the y axis.

We can see that negative spreads are uncommon, especially since 2021. However since 2021 they have become more erratic. The chart below isolates the negative spreads to get a better visual representation. The largest negative spread occurred on December 24, 2020 when the Ant Financial IPO was cancelled. On that day Alibaba has over $20 billion more dollar volume than the rest of China Tech combined.

Observations of Negative Spreads Since January 2020

Value <0 indicates more transaction volume in Alibaba is higher than rest of US Listed China Sha

Minimum value on December 27, 2020 due to Ant IPO cancellation.

Testing Negative DV Indicators

We can see that the number of occurrences where the spread is negative have decreased in 2021. It may be worth exploring whether these occurrences can now serve as an indicator for next day price return. If we take the occurrences (negative spread, date), match with the same day returns and then calculate next day returns we can test whether negative DV occurrences could be a useful indicator of returns in 2021.

Same day return refers tp the return of Alibaba on the date the negative spread occured. Next day return refers to the next trading day. It is worth noting that some same day returns with negative spreads will be positive and some will be negative. Alibaba can have a negative spread of $2 billion and a positive same day return of 5%. Returns are sorted by highest negative spread.

```
##         date   Spread returns
## 1 2020-12-24 -21.8421 -0.1432
## 2 2020-05-22  -3.9210 -0.0605
## 3 2020-12-28  -3.7723  0.0016
## 4 2020-11-03  -3.1157 -0.0848
## 5 2020-02-13  -3.0348 -0.0178
## 6 2021-04-12  -2.7237  0.0886
```

This table has 66 total rows. Next we will explore the relationship between the negative spread values and returns in more depth.
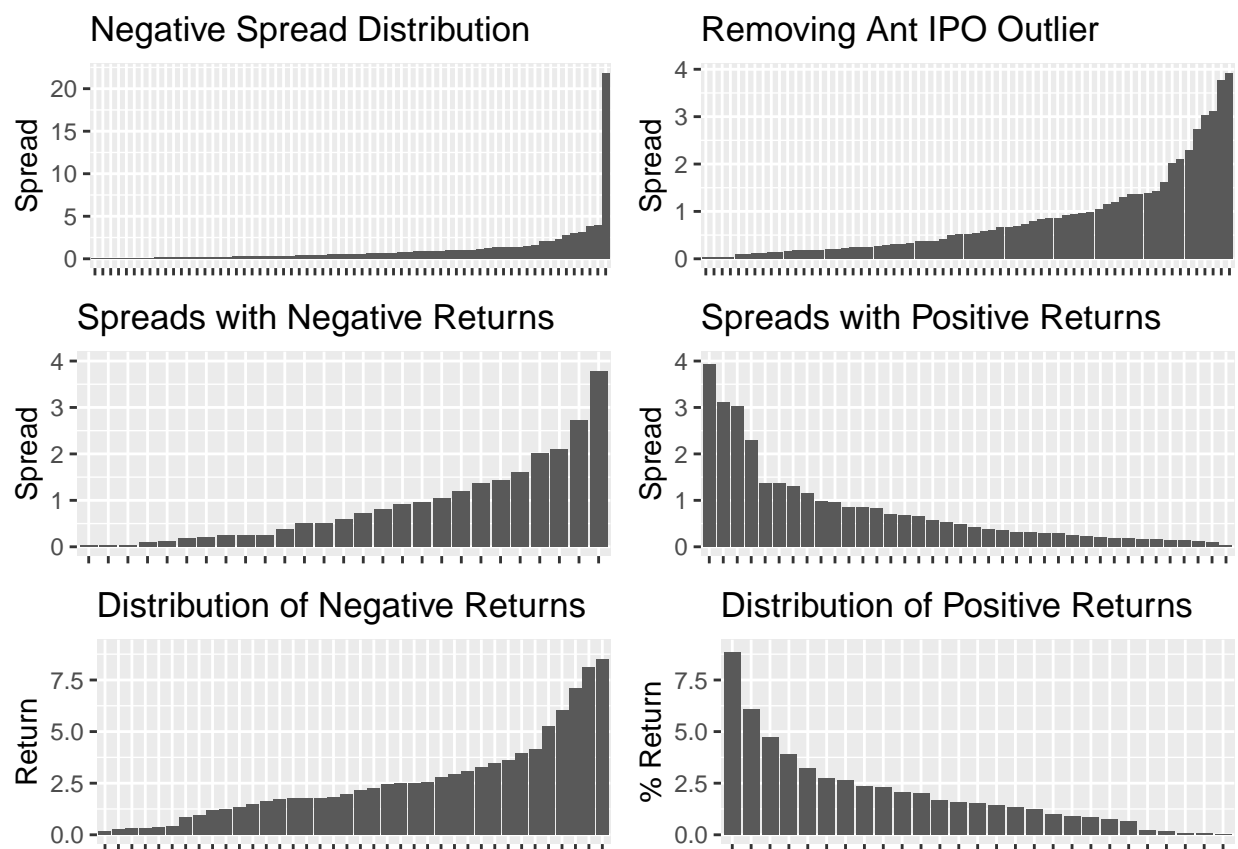
Next we will explore the relationship between a negative spread and the same day return. Although we have a small number of data points we can roughly show that higher absolute returns are associated with large (negative) DV spreads.
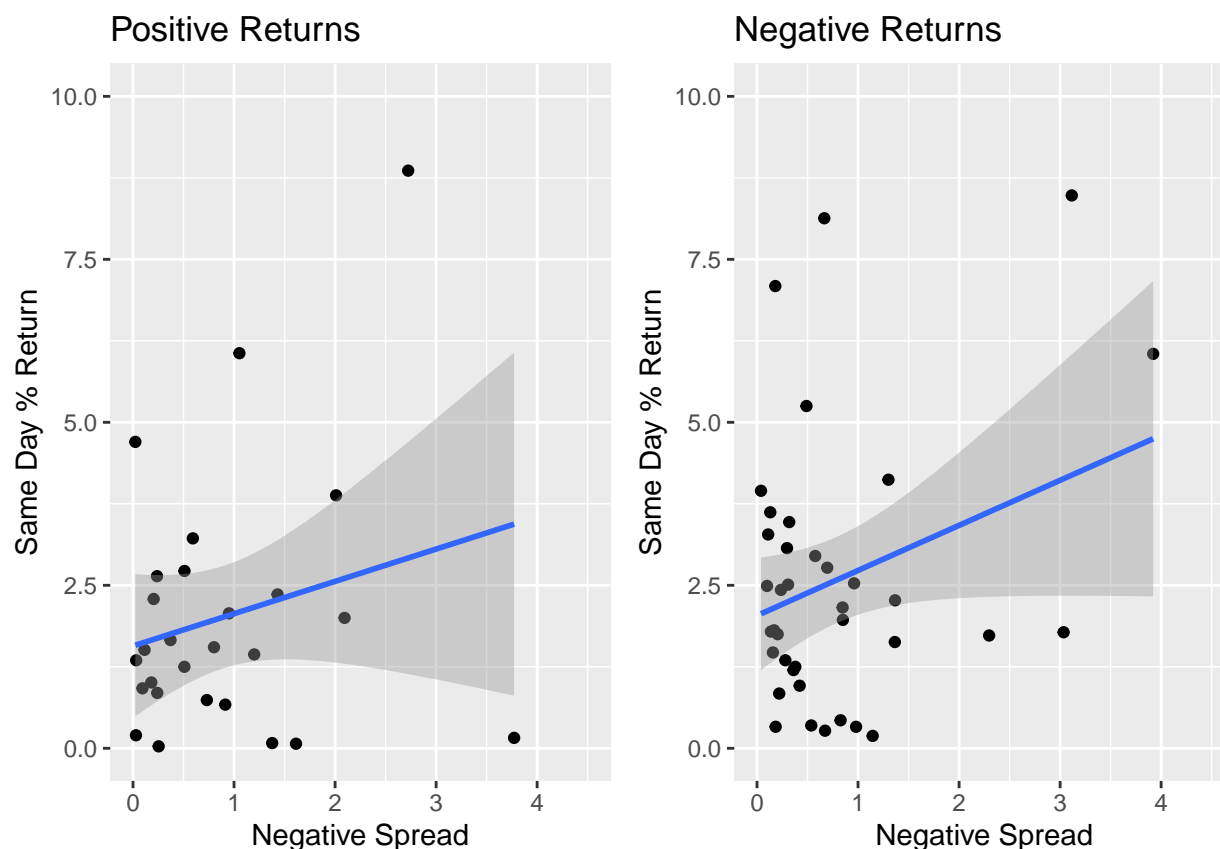
Now we will filter for negative same day returns with negative spreads greater than 1. We selected this negative filter value based on the the spreads for negative/positive return charts above. It appears 1 is a level where there are fewer results and more volatility in returns. The Ant IPO cancellation on December 24, 2020 is a major outlier.

We have 9 instances of negative returns with negative spreads greater than 1.

```
##         date   Spread returns
## 1 2020-12-24 -21.8421 -0.1432
## 2 2020-05-22  -3.9210 -0.0605
## 3 2020-11-03  -3.1157 -0.0848
## 4 2020-02-13  -3.0348 -0.0178
## 5 2020-05-04  -2.2974 -0.0173
## 6 2020-04-23  -1.3663 -0.0227
## 7 2021-08-20  -1.3642 -0.0163
## 8 2020-05-01  -1.3009 -0.0412
## 9 2020-05-20  -1.1448 -0.0019
```

Now we will filter same day returns for positive returns. We have 9 instances here as well. We will remove the Ant IPO outlier for 8 instances.

```
##         date  Spread returns
## 1 2020-12-28 -3.7723  0.0016
## 2 2021-04-12 -2.7237  0.0886
## 3 2020-05-05 -2.0940  0.0200
## 4 2020-05-29 -2.0100  0.0388
```

```
## 5 2020-07-23 -1.6134  0.0007
## 6 2020-05-08 -1.4306  0.0236
## 7 2020-05-06 -1.3774  0.0008
## 8 2020-02-28 -1.1996  0.0144
## 9 2020-12-29 -1.0532  0.0606
```

Outside of the major outlier in the negative

Now we will take the dates from the previous dataframe and add a day to it. Then we will take the dates and run them against the table of daily returns we have.

```
##          date
## 1 2020-02-14
## 2 2020-04-24
## 3 2020-05-02
## 4 2020-05-05
## 5 2020-05-21
## 6 2020-05-23
## 7 2020-11-04
## 8 2020-12-25
## 9 2021-08-21
```

Matching the dates with returns, some are NA because the next business day comes on a weekend. So we will isolate those and sim to the next day, and repeat that until we can get the next days return.

```
##          date returns
## 1 2020-02-14 -0.0033
## 2 2020-04-24 -0.0043
## 3 2020-05-02      NA
## 4 2020-05-05  0.0200
## 5 2020-05-21 -0.0216
## 6 2020-05-23      NA
## 7 2020-11-04  0.0349
## 8 2020-12-25      NA
## 9 2021-08-21      NA
```

These are the dates that we will keep and merge into a dateframe after we get the rest. To get the rest we will push the dates forward another date and run against the returns table. We will keep doing this until we get the next return for all dates.

```
##          date returns
## 1 2020-02-14 -0.0033
## 2 2020-04-24 -0.0043
## 3 2020-05-05  0.0200
## 4 2020-05-21 -0.0216
## 5 2020-11-04  0.0349
```

Now we try for the next day. There were no business days that fell on this day. On to the next day then. As we can see from the NA values.

```
##          date returns
## 1 2020-05-03      NA
## 2 2020-05-24      NA
## 3 2020-12-26      NA
## 4 2021-08-22      NA
```

Looks like we got one more value by repeating the same process to the next one.

```
##         date returns
## 1 2020-05-04 -0.0173
## 2 2020-05-25      NA
## 3 2020-12-27      NA
## 4 2021-08-23      NA
```

Here are the last 2 dates we need. Omitting 2021-08-24 as that date was in the future when this report was generated.

```
##         date returns
## 1 2020-05-26  0.0101
## 2 2020-12-28  0.0016
## 3 2021-08-24      NA
```

Now we will join all the dates and returns into a single dataframe.

```
##         date returns
## 1 2020-02-14 -0.0033
## 2 2020-04-24 -0.0043
## 3 2020-05-05  0.0200
## 4 2020-05-21 -0.0216
## 5 2020-11-04  0.0349
## 6 2020-05-04 -0.0173
## 7 2020-05-26  0.0101
## 8 2020-12-28  0.0016
```
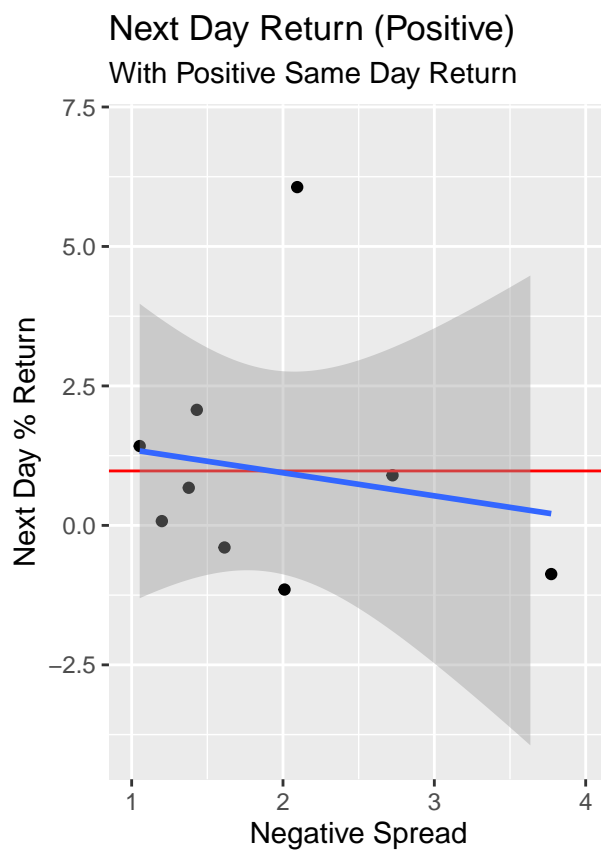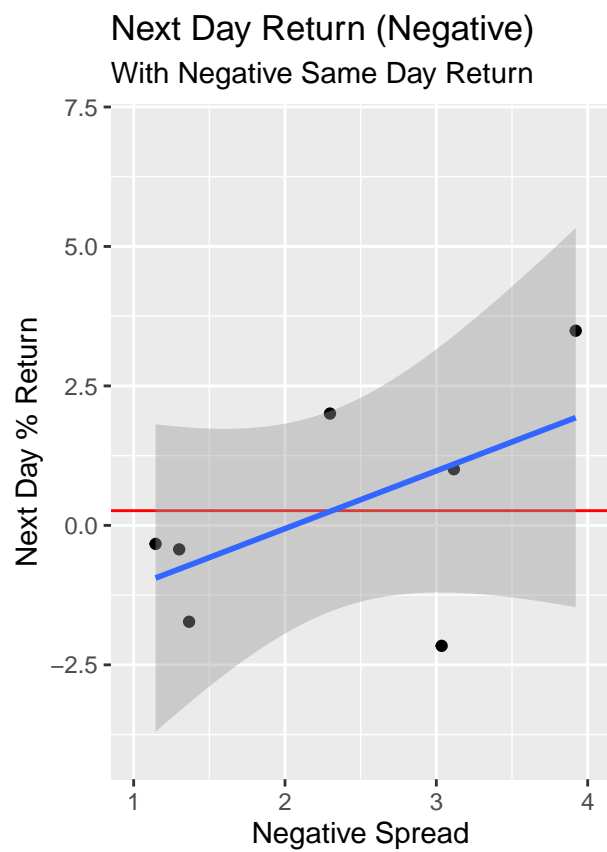
Now we put it all together in a dataframe. We are looking at date, spread, same day, and next trading day return.

```
##         date   Spread same.day.return next.day.return
## 1 2020-12-24 -21.8421         -0.1432          0.0016
## 2 2020-05-22  -3.9210         -0.0605          0.0349
## 3 2020-11-03  -3.1157         -0.0848          0.0101
## 4 2020-02-13  -3.0348         -0.0178         -0.0216
## 5 2020-05-04  -2.2974         -0.0173          0.0200
## 6 2020-04-23  -1.3663         -0.0227         -0.0173
## 7 2020-05-01  -1.3009         -0.0412         -0.0043
## 8 2020-05-20  -1.1448         -0.0019         -0.0033
```

Now we will do the same for positive returns. The table is below

```
##         date   Spread same.day.return next.day.return
## 1 2020-12-28 -3.7723          0.0016         -0.0087
## 2 2021-04-12 -2.7237          0.0886          0.0090
## 3 2020-05-05 -2.0940          0.0200          0.0606
## 4 2020-05-29 -2.0100          0.0388         -0.0115
## 5 2020-07-23 -1.6134          0.0007         -0.0040
## 6 2020-05-08 -1.4306          0.0236          0.0207
## 7 2020-05-06 -1.3774          0.0008          0.0067
## 8 2020-02-28 -1.1996          0.0144          0.0008
## 9 2020-12-29 -1.0532          0.0606          0.0142
```

Looking at the return charts again for the next day.

Either return

Combined Next Day Returns to Spread
With Both Positive and Negative Same Day Returns
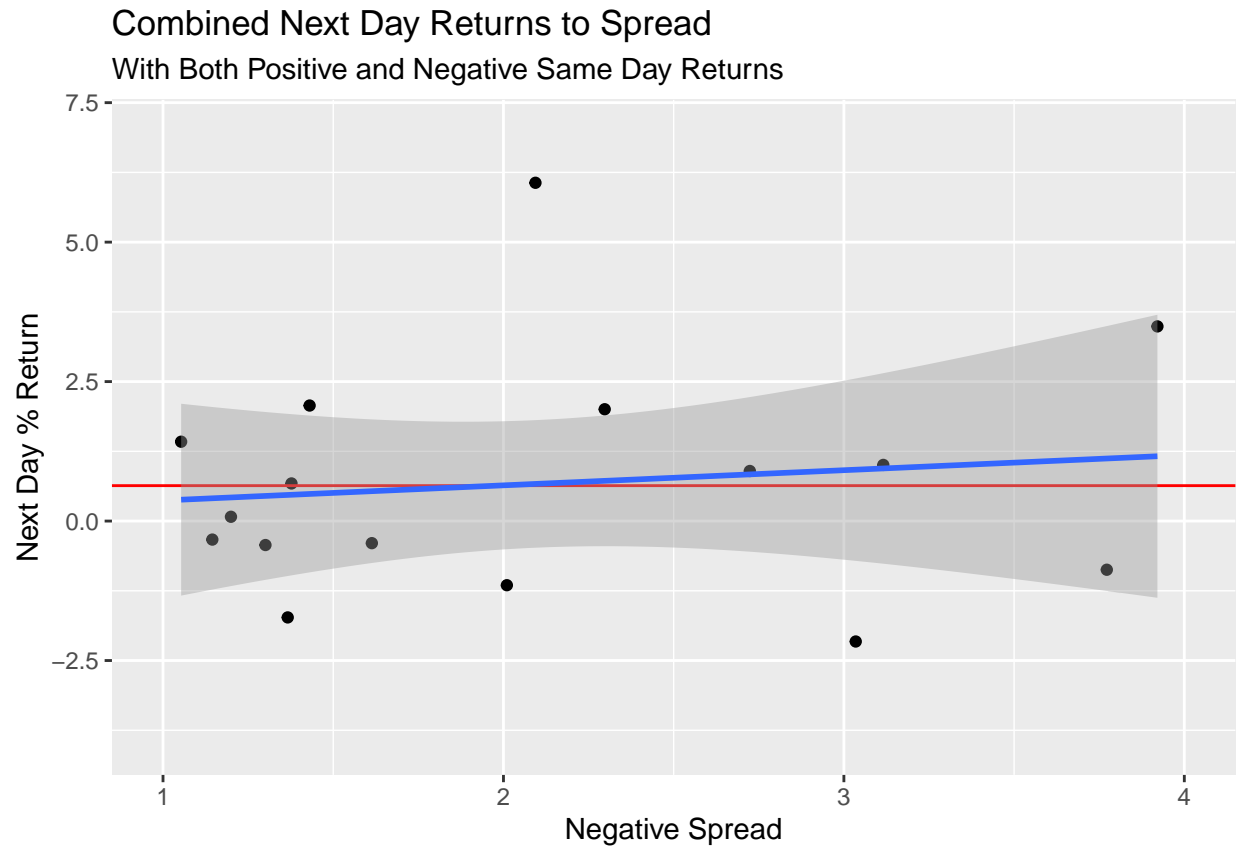
Negative Same Day Return

All of the following values are in Percentage.

```
##   mean_next_day_return
## 1            0.2646526
```

Summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.1588 -1.0784 -0.3318  0.2647  1.5054  3.4892
```

Maximum

```
## [1] 3.489201
```

Minimum

```
## [1] -2.158838
```

Positive Same Day Return

Percentage

```
##   mean_next_day_return
## 1            0.9762973
```

Summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.1500 -0.3962  0.6741  0.9763  1.4225  6.0635
```

Maximum

```
## [1] 6.063518
```

Minimum

```
## [1] -1.149991
```

Combined (Positive/Negative) Same Day Return

Percentage

```
##   mean_next_day_return
## 1            0.6353692
```

Summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.1588 -0.4297  0.1620  0.6354  1.4225  6.0635
```

Maximum

```
## [1] 6.063518
```

Minimum

```
## [1] -2.158838
```

Conclusion

Looking at the summary statistics we can see that for Alibaba on days where it's DV spread to China Tech (Sum) is over 1 billion and the same day return is positive, the average next day return is 0.97%. With the same parameters, on days where the same day return is negative the average next day return is 0.26%. The Ant IPO occurrence with a spread value of billion has been excluded in the summary statistics.

Based on this data, if one were to employ the strategy for positive return days they would do the following. Wait for a negative spread occurrence greater than 1 to occur on a date that has a positive return. To execute, this will require intraday volume and price data as close to the closing price as possible. When an occurrence is observed, the purchase will be made as close to closing price as possible on that day. Then the stock will be held until as close as possible the following day to be sold.

Thank you for taking the time to read my report.