

Identifying and Testing Stock Indicators

Austin McGhee

8/20/2021

Calculating the Indicator

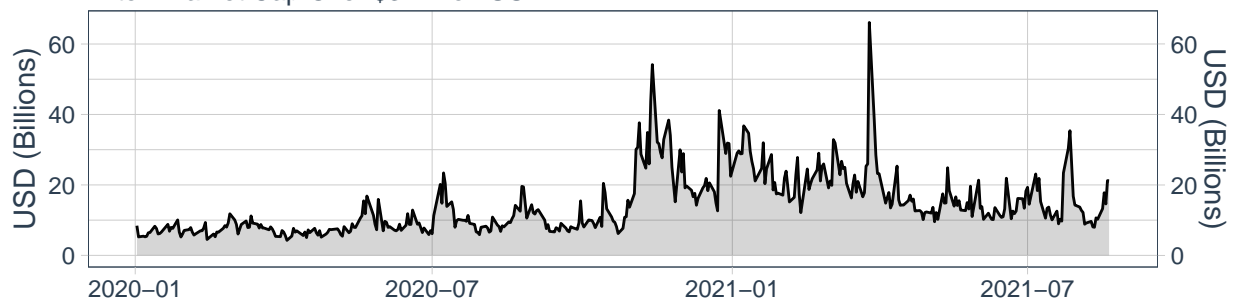
In the price data we have date, OHLC, adjusted and volume. We will create a new column called Daily Dollar Volume which is the adjusted price multiplied by the volume for the day. This links volume data to share prices and can provide an absolute measure of the transaction volume.

View the Rmarkdown file on my Github at: https://github.com/AustinMcg/PortfolioProjects/tree/main/R_Tableau_ChinaTechCrackdown/R-Identifying-and-Testing-Stock-Indicators

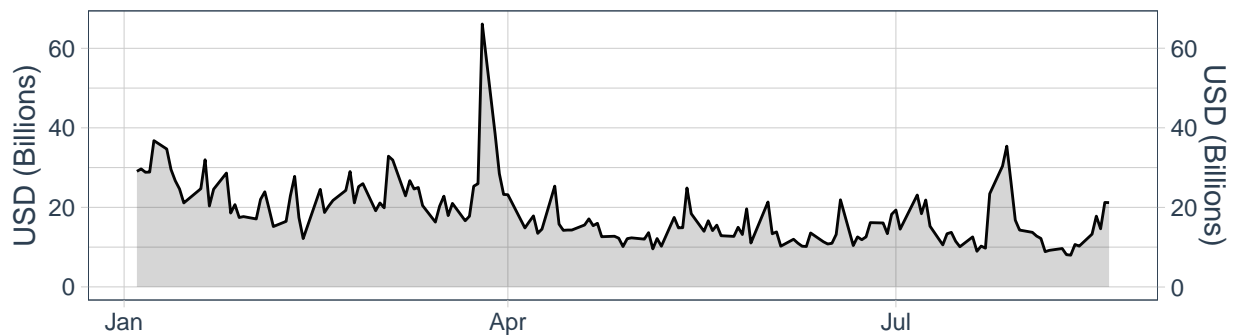
Using the Daily Dollar Volume (abv DDV) we can measure both individual and groups of stocks. For example the DDV for Didi may be 1 billion USD, and the sum of DDV for US listed China Tech (>3 billion Market Cap) may be \$30 billion. We could also calculate the mean DDV for the Top 5 symbols and analyze it's relationship to the mean DDV for all US listed China Tech.

Sum of Daily Dollar Volume for US Listed China Tech

Filter: Market Cap Over \$3 Billion USD



A Closer Look at 2021

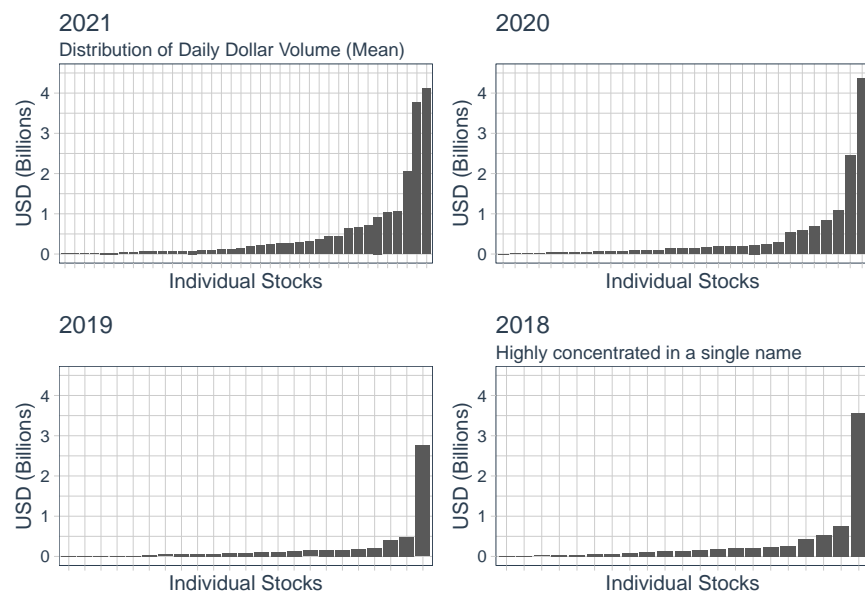


We can see in the figure above that the sum of DDV for China is quite volatile and it's general value tends

to shift slowly on a monthly and quarterly basis. From this data alone we can track transaction volume, but it doesn't provide much insight into the direction of the underlying share price.

Distribution

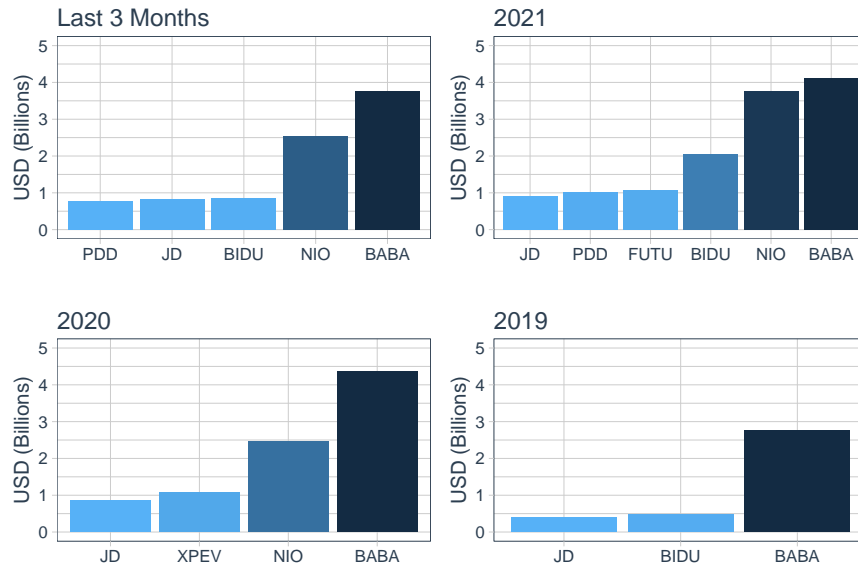
How is DDV distributed among the individual stocks? We could be looking at a basket of stocks with similar DDV or a select few making a disproportionate contribution. In the chart below, each column represents a stock and the y (US Billions) axis is a measure of the Average DDV for the fiscal year in the chart title.



Observing this chart, we can see that returns are highly concentrated but have progressively reduced concentration since 2018. The first year where there was a close runner-up to top rank was in 2021. The next section labels the individual stocks in these charts and explores them further.

Individual Stocks

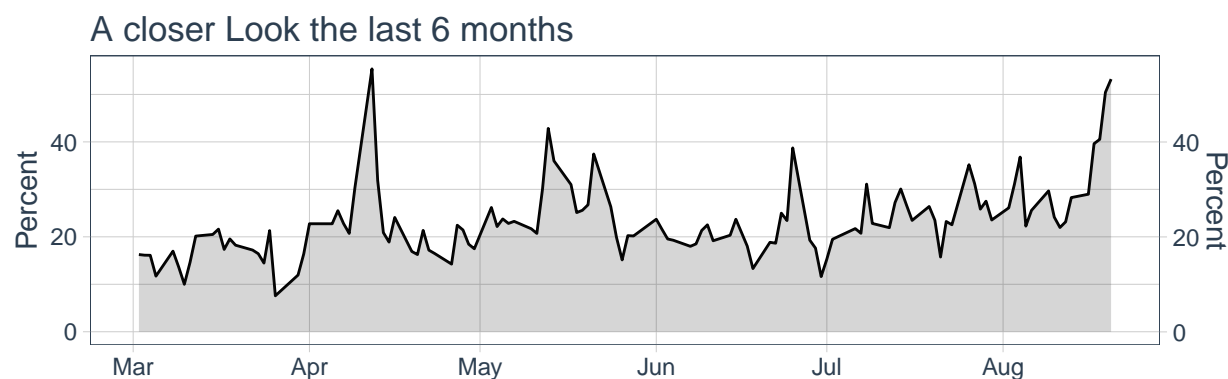
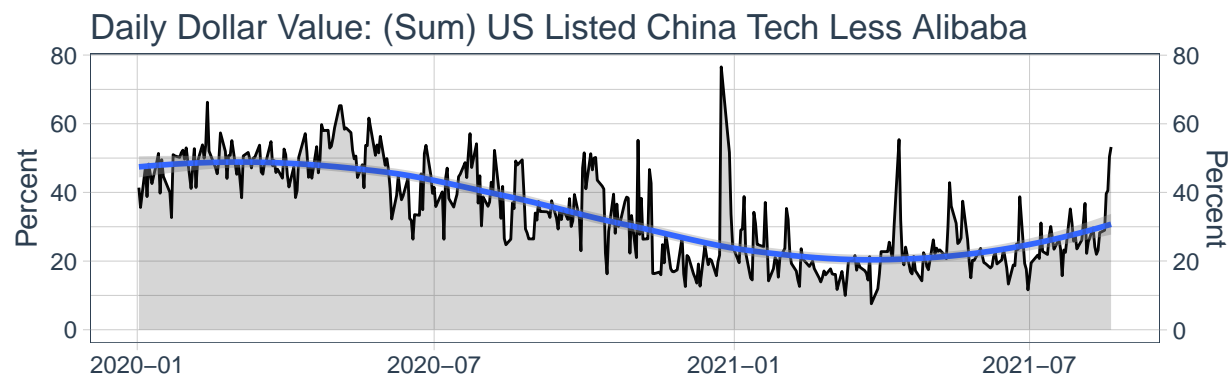
Which stock is doing most of the heavy lifting? We can see that for every time period Alibaba (BABA) has the highest DDV (Mean).



We can see that for every time period Alibaba (BABA) has the highest DDV (Mean). The top stocks were selected by running a filter of DDV (Mean) with minimum 500MM USD over US listed China Technology stocks with a minimum market cap of \$3 billion USD. In 2019 the filter was lowered to minimum 250MM USD to accommodate more results, otherwise it would only show Alibaba.

Comparing Alibaba

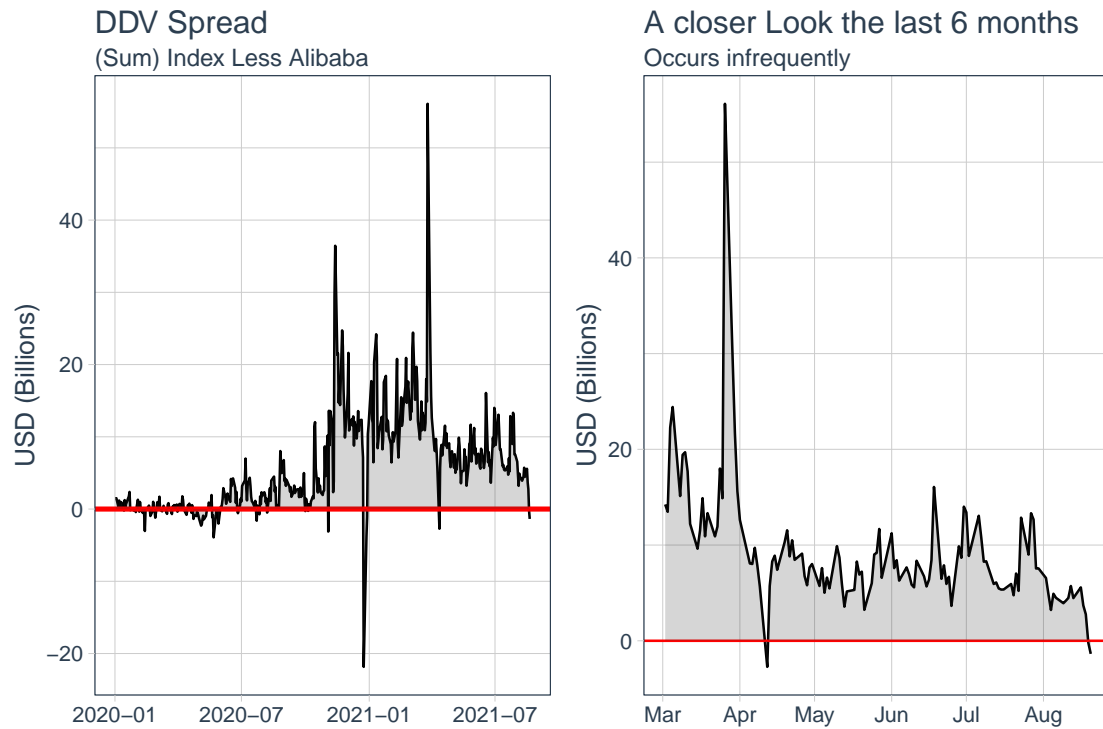
After taking the Sum of the index we will look at the largest stock, Alibaba, against the set to look for potential indicators. we take the sum of the daily dollar volume for the sector and deduct alibaba from the calculation. One dataframe is the sum of daily dollar volume for the china sector (think transaction volume) and the other is alibaba daily dollar volume. We deduce alibaba from the index total to calculate the difference. when the difference is negative, it means that more daily dollar volume is trading in alibaba then the rest of the china tech sector combined. We are looking at this indicator specifically in the 2020 and 2021 environments.



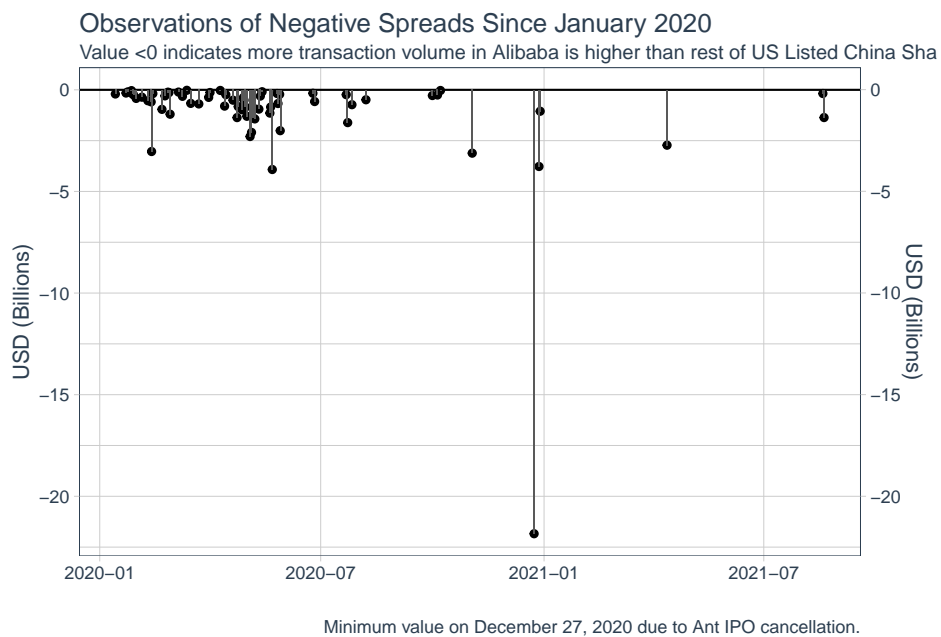
I am skeptical of how useful it will be as an indicator for farther back periods, because we are looking at it as an absolute indicator, perhaps it would be better to look back at past periods when using a mean indicator. we would likely need to factor in changes to the basket of stocks. This was chosen on market capitalization above a threshold, there would be stocks coming and going from that category that may change results. New market entrants could confuse results as well.

Spread Comparison

Calculating and comparing Alibaba spread to the rest of the market.



We can see that the number of instances where the spread is less than 0 are low. These occurrences are visualized in the chart on the next page.



Indicator Testing

After observing that the number of occurrences where the spread is below zero have gone down in 2021, it may be worth exploring whether these occurrences can now serve as an indicator. If we take the occurrences and the dates they happened we can then reference the stock returns for that date, the next trading day, and for 5 trading days in the future.

Some dates with large spreads will be negative and some will be positive. We will filter on their +/- status and analyze them separately. Statistics will also be provided on the group as a whole.

Now taking a look at the results

```
##           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
```

Looking at the results of the high spreads filter.

```
##           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
## 7 2020-05-04  -2.2974 -0.0173
## 8 2020-05-05  -2.0940  0.0200
## 9 2020-05-29  -2.0100  0.0388
## 10 2020-07-23  -1.6134  0.0007
## 11 2020-05-08  -1.4306  0.0236
## 12 2020-05-06  -1.3774  0.0008
## 13 2020-04-23  -1.3663 -0.0227
## 14 2021-08-20  -1.3642 -0.0163
## 15 2020-05-01  -1.3009 -0.0412
## 16 2020-02-28  -1.1996  0.0144
## 17 2020-05-20  -1.1448 -0.0019
## 18 2020-12-29  -1.0532  0.0606
```

Looking at the negative returns

```
##           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 the positive returns

```
##           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
```

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 dataframe 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
```

We will keep this date.

```
##           date returns
## 1 2020-05-04 -0.0173
```

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
```

Renaming the returns to next.day.return for easier reference when we combine with the spreads and same day return dataframe.

```
##  next.day.return
## 1           0.0016
## 2           0.0349
## 3           0.0101
## 4          -0.0216
## 5           0.0200
## 6          -0.0173
## 7          -0.0043
## 8          -0.0033
```



```
##          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 2020-05-01  -1.3009 -0.0412
## 8 2020-05-20  -1.1448 -0.0019
```

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
```

Percentage

```
##    mean_next_day_return
## 1                0.251825
```

Summary

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.021588 -0.007540 -0.000849  0.002518  0.012559  0.034892
```

Maximum

```
## [1] 0.03489201
```

Minimum

```
## [1] -0.02158838
```

Conclusion

Based on these summary statistics it appears that using the occurrence of a negative spread on a negative return date as a transaction point and seeking a return the next day is not an effective strategy. As shown in the data above, by purchasing on these large negative spread days and selling at the close the following day the average return is 0.25%.

Thank you for taking the time to read my report.