

Analyzing stock market with Python



By Tom Tang

Project Introduction

In this project, I will apply data analytic tools to the stock market. More specifically, I have two objectives:

1. **Explore** the trends and relationships between Intel and AMD share prices (INTC, AMD) in the past year.
2. **Test** the effectiveness of analytic indicator “bollinger bands” in forecasting stock price movements.





Introduction to Data Collection Tool

Unfortunately, all mainstream API for acquiring stock quotes (NASDAQ etc.) are proprietary and I was not able to access one for the time being.

As a result, for this project, I will use a python module called yfinance, which simulates the Yahoo Finance API before Yahoo discontinued its service.



Roadmap|

Context

Part I: AMD and INTC

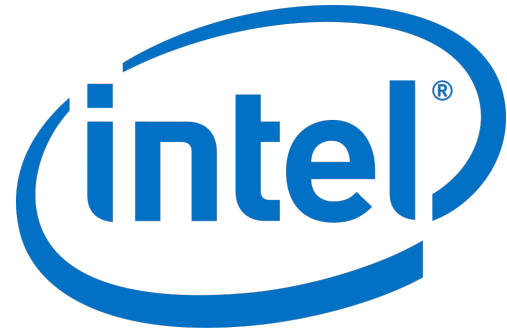
Part II: Bollinger Band Test

Conclusion



AMD and Intel Rivalry

For the past decade, AMD and Intel has been the two largest contender in the semiconductor industry, with both firms releasing their lines of powerful chips. Intel has always had a larger market share than AMD, due to their partnerships with other tech firms and earlier entry into the market. However, the dynamic between these two changed in recent years, when AMD started to output more powerful chips at cheaper prices.





Bollinger Band

Bollinger Band is a very commonly used stock analysis tool. According to Investopedia, "a Bollinger Band® is a technical analysis tool defined by a set of trendlines plotted two standard deviations (positively and negatively) away from a simple moving average (SMA) of a security's price". It can be effective to predict stock movement.



Roadmap|

Context

Part I: AMD and INTC

Part II: Bollinger Band Test

Conclusion



Question

Would stock price of AMD and Intel showcase their competitive relationship in the market?



Hypothesis I

AMD and INTC will have **opposite** stock price movements due to their **competitive relationship**.

Plotting

- Using yfianance, we retrieved the historic stock data of these two stocks. After we prepared the data, we plot the closing stock price of each day against time.
- From the plot, we can see that these two stocks did move in opposite direction after around mid-to-late July. However, they actually had very similar movements up until that point, which raised the important questions:
 - **When and why did this happen?**
 - **What caused this divergence?**



Figure 1: Stock Prices of AMD and INTC from 2020-01-01 to 2021-01-01



Further Exploration

- To tackle the question brought up in the previous plot, I will hypothesize that **the direct opposite price movements** started due to new technology released from AMD, instead of the bullish market run.
- To achieve this, I will focus in on these two stocks from 2020-07-01 to 2020-09-01, and use index fund SPY, a reliable indicator of the market, as a benchmark to compensate for the gain/loss due to market performance. Since a **good** market performance has a **positive** influence on blue chip stocks like AMD and INTC

Plotting

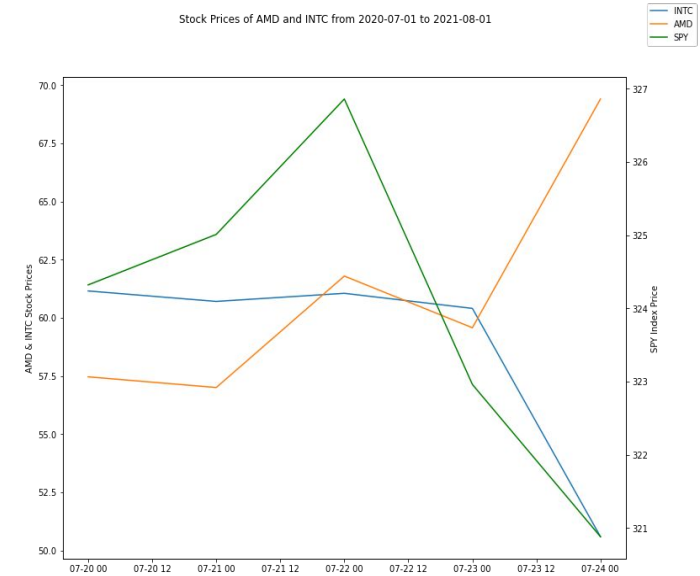
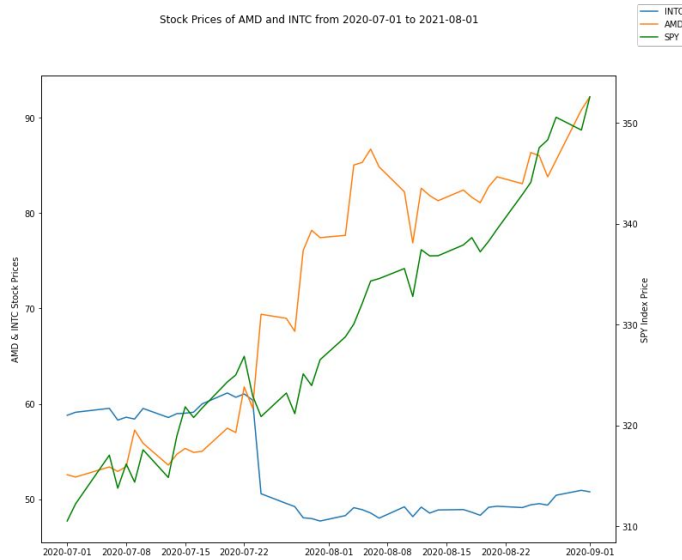


Figure 2: Stock Prices of AMD and INTC with respect to SPY from 2020-07-01 to 2020-09-01

Figure 3: Stock Prices of AMD and INTC with respect to SPY from 2020-07-20 to 2020-07-24

Conclusion for Hypothesis I

From these two figures, we can see that the jump of AMD and plummet of INTC coincided on 2020-07-22, which is when AMD announced its plan to release a new line of processors based on its 7-nanometer architecture, while Intel announced to push back the release date of its new chip. This event marked the end of Intel's competitive advantage over AMD, and is well reflected in the share prices.

Before this event, these two stocks moved quite consistently with respect to each other, as we can see in Figure 2: the movement of AMD and INTC had the same shape as SPY, the index fund. And months after the event, this is no longer the case. As we see in Figure 4, even when the market is very bullish, INTC still experienced a lot of horizontal movements, indicating that it has fallen out of investors' favor.

Therefore, it can be concluded that the share prices of AMD and INTC successfully reflected their competition in the market.

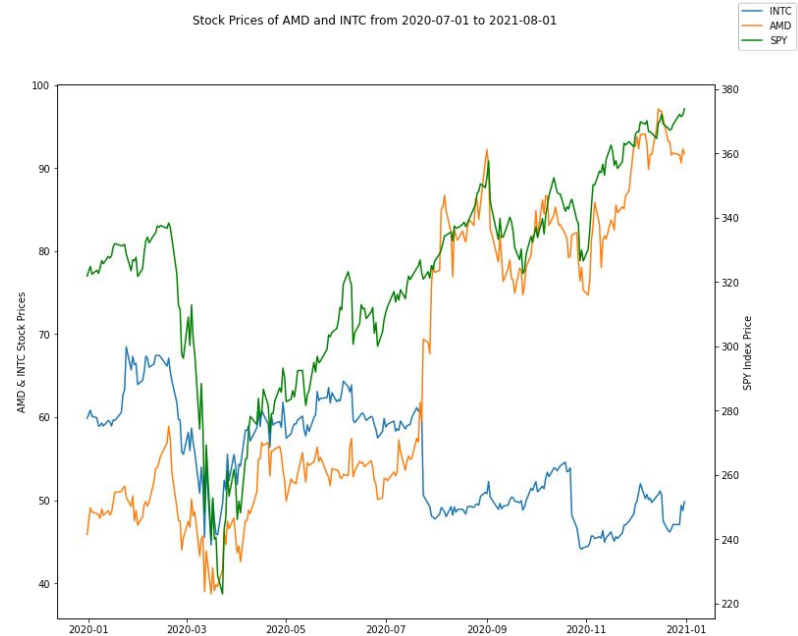


Figure 4: Stock Prices of AMD and INTC with respect to SPY from 2020-01-01 to 2021-01-01

Roadmap|

Context

Part I: AMD and INTC

Part II: Bollinger Band Test

Conclusion



Graphing Bollinger Band for INTC

As formerly introduced, bollinger band is a technical tool for stock analysis, and its equation is as follows:

The calculation of Bollinger Band is divided into 4 parts:

1. Standard Deviation of the stock prices in a given period of time.
2. Mean of the stock prices in a given period of time, called a moving average.
3. Upperband = Moving average + n * standard deviation
4. Lowerband = Moving average - n * standard deviation

Note: n can be tweak by the users, for the sake of consistency, I would use $n = 1.5$ for this project.

Plotting

After implementing the equation from the previous slide with pandas, we have:

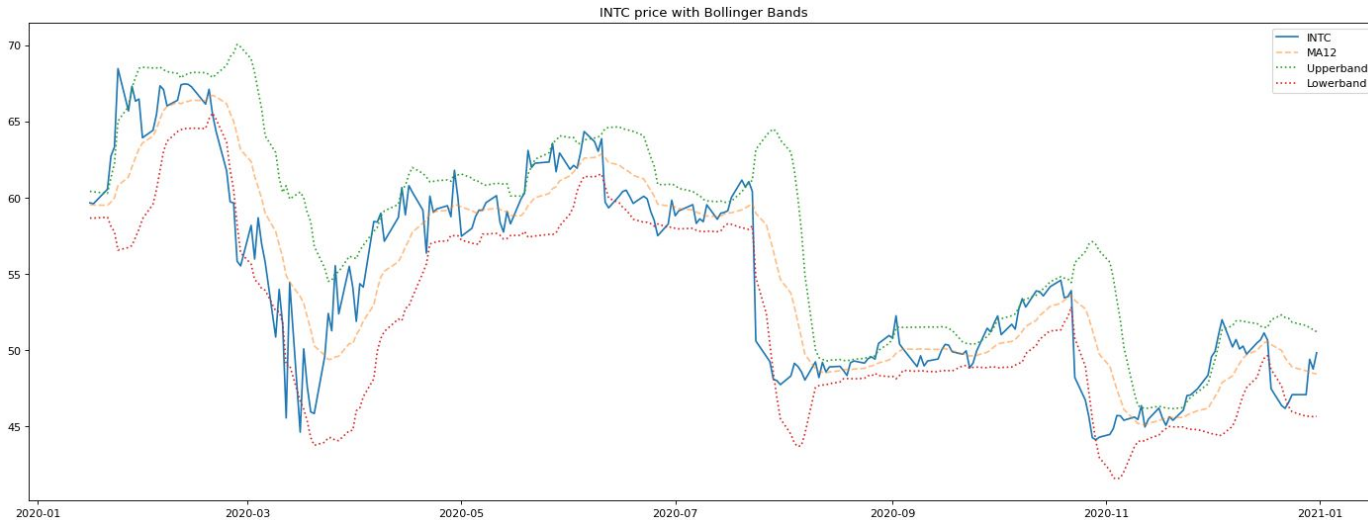


Figure 5: INTC stock price with bollinger bands

Question

As we can see from Figure 5, the Upper and lower bollinger bands actually created a range in which the stock price can oscillate. This property creates a scenario where if the stock price goes outside of the range between the two bands, one can expect it to return to the range within a period of time.

But really, how reliable is it?

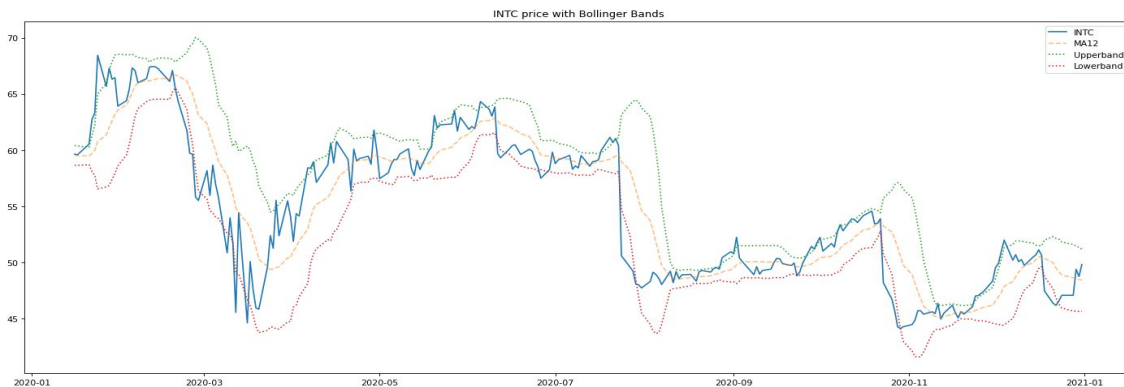


Figure 5: INTC stock price with bollinger bands



Testing Bollinger Band

To verify how reliable the bollinger band is, I have come up with a method:

1. I am going to assume that once stock price goes outside the range provided by the bollinger band, it would return within a grace period, g. For the sake of consistency, I will make $g = 3$
2. Count for every day in which the stock price is outside of bollinger band, call it D
3. Count every instance where the stock price is outside of bollinger band for 3 days in a row, call it d
4. The accuracy would be d/D

Code using INTC as example:

```
def checkFalse(booList):
    lengthList = []
    count = 0
    for i in range(len(booList)):
        if booList[i] == False:
            count += 1
        elif booList[i] == True and count > 0:
            lengthList.append(count)
            count = 0
        if i==len(booList)-1 and count>0:
            lengthList.append(count)
    return lengthList

INTC["Inside"] = ((INTC["Close"] >= INTC['Lower']) &
                 (INTC["Close"] <= INTC['Upper']))

exceptions = checkFalse(INTC["Inside"])
falsePrediction = 0
for item in exceptions:
    if item > 3:
        falsePrediction += 1
accuracy = (1 - float(falsePrediction/ len(exceptions))) * 100
```



Hypothesis II

Companies that have **larger** market cap will have a **higher accuracy** in the test, since they tend to grow more consistently, thus making them more predictable.

Result

After running this test on 100 stocks on Nasdaq, I formulated Figure 6. The result really surprised me.

1. The mean accuracy for these 4 market cap categories is **83.92%**
2. As seen in the figure, the companies in all four market cap categories displayed **approximately the same level of accuracy** in the bollinger test

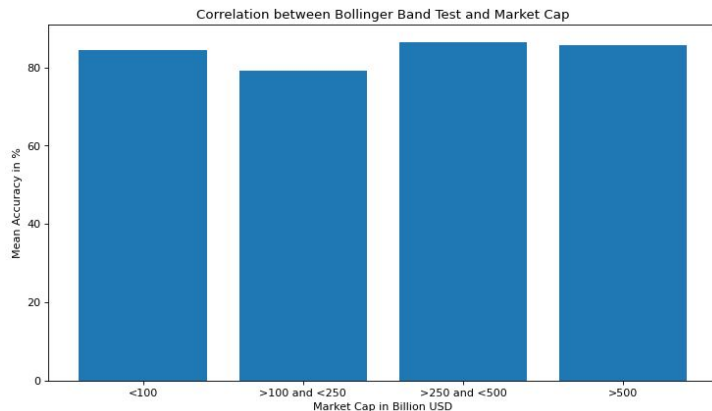


Figure 6: Correlation between bollinger band test and market cap

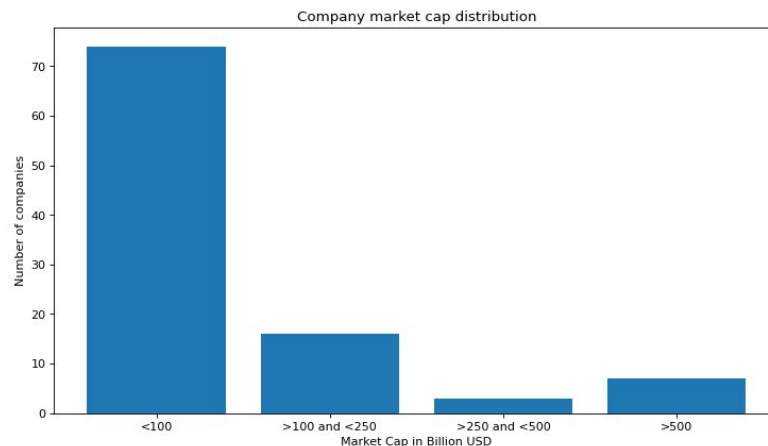


Figure 7: Company market cap distribution



Conclusion for Hypothesis II

Though I was pleasantly surprised, the result made sense. As discussed in the previous slide, the different market caps did not show different results in the bollinger band test. As a result, the bollinger bands would be a consistently reliable tool for companies of all different sizes.

Roadmap|

Context

Part I: AMD and INTC

Part II: Bollinger Band Test

Conclusion



Conclusion

Stock market is a very unpredictable place, where immense fortunes and financial ruins can be made in the same day. This is why trying to understand it in a higher level is very advantageous for investors to secure a decent return on their investments.

In this project, we explored many things (still just a tip of the iceberg) that could affect the stock prices: market performance, delay of a new product, and the rise of a competitor, but to be fair, no one can say for sure to what degree these events are going to affect the share prices.

On the other hand, analytic tools allow investors to take a break from being anxious due to news in the market, but rather, use mathematics to view the stocks on a macro level. At the end of the day, money is still a number. Thus, picking the right tool for the job is crucial to investors. Proven by this project, the bollinger band can consistently forecast stock movements in companies across different market caps.

However, there are so many other tools waiting to be created and explored.



Future Improvements:

1. **Introduce** more factors that could affect movements of stocks and analyze their influence.
 - 1) Mentions on twitter, reddit and other similar platforms
 - 2) Whale in fish tank scenario (When a big investor buy a whale share of a penny stock)
 - 3) Institution involvement
2. **Introduce** more stocks into the test for consistency.
3. **Optimize** my algorithm.



Thank You

