Ha Doan - 32592035, Ha Pham - 32820367
Final Project, INFO348, Professor Matthew Rattigan

MULTI-FACTOR MODEL

1. Introduction

- Stock volatility is often time a concerning problem for investors and traders. Although the central limit theorem shows that stocks are normally distributed over time, sometimes they behave in very distinct manners that no one would able to predict such shortcomings.

- Thus, we found another way to spot stock volatility by looking at the multi-factor-based model! By taking a closer look at significant events that might effect the company, we can spot the abnormalities of these stock behaviors.

2. Data Collection

Our Data file consists of one Excel file and each sheet contains three columns:

- Date: the date of the stock price
- Adj Close: The stock price which is adjusted by the bid, ask, and volume of the stock throughout the day.
- Earning: a dummy variable denoting whether it is the date of an earning release.

We use the following websites to find our data:

- Yahoo finance: to find the historical stock price of the data

URL: https://finance.yahoo.com/quote/AMZN?p=AMZN&.tsrc=fin-srch

- EDGAR: to find the earning release dates of each company

URL: https://www.sec.gov/edgar/search-and-access

3. Data Cleaning

a. *Excel cleaning*

- We merged all data into an Excel file and did the first round of data cleaning so that every sheet is similar in date
- Changing date formats into "yyyymmdd"
- As well as apply some simple Excel functions to create the dummy variable of Earning for each sheet

b. *Python data cleaning*

- We first do some necessary calculations by calculating the return, variance, and implied volatility (annual volatility) of the stock.

Where:

$$Return = \frac{P_t}{P_{t-1}} - 1$$

$$Var_t = (r_t - \bar{r})^2$$

$$Vol = \sqrt{var * 252}$$

Note: this resulted in 3 new columns for each sheet of stock_return, stock_var, and stock_vol

- Next, we set the index as date and changed the name Price and Earning to stock_price and stock_earning in this stage.
- Finally, we merge all data into a master data frame for easy access to information.

4. Analysis

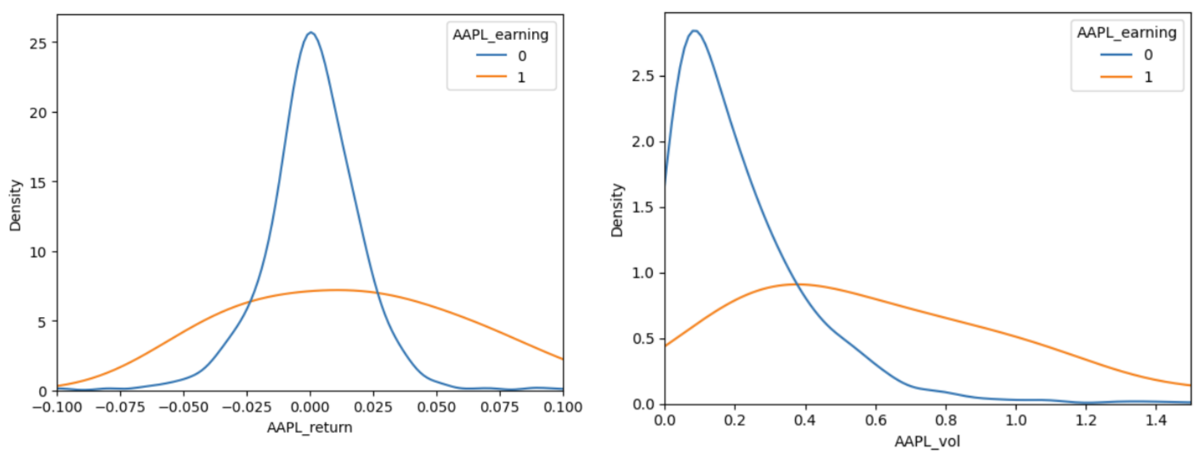
a. *We divided the data frame into 2 file for comparison:*

- one is the large-cap (denoted as large) stock data where their capitalization is at or over 1 trillion dollars in valuation and the other is the small-cap (denoted as small) stock data where their capitalization is in the range of 1 to 10 billion dollars in valuation.

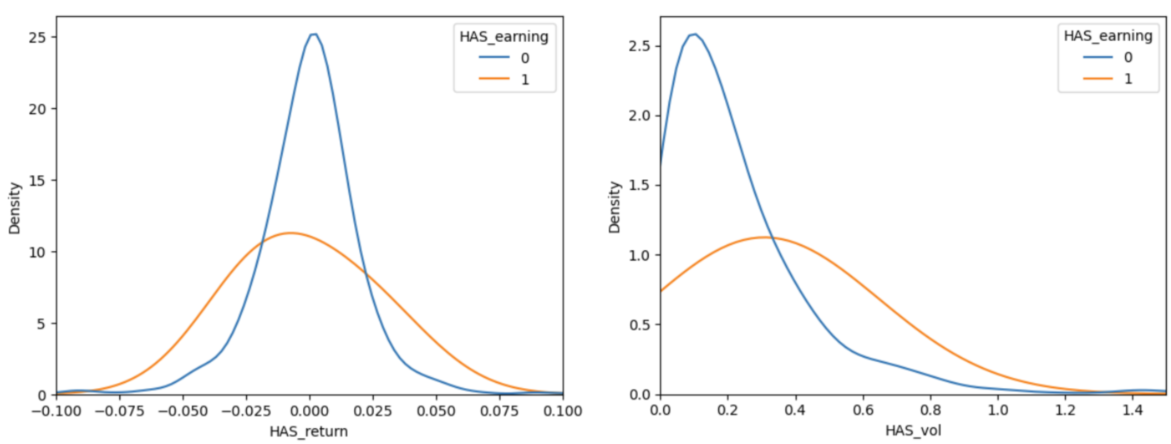
b. *Distribution*

In the first analysis, we will take a look at the distribution of returns and volatility of one stock in each category on earning and not earning dates:

- Large Cap: AAPL - Apple Inc (tech company)



- Small Cap: Hasbro (toys company)



- Through observation, we can see that there exist an large gap between the earning dates and the overall normal distribution of returns for each stock. An assumption for this is that at the earning release dates, investors will have different opinions about the company which change their position on the stock accordingly. Thus, leads to the high variance of the stock making the volatility distribution shift right.

c. *Linear Regression:*
- To test the relationship between the factor and the stock we used linear regression for all stocks in large and small-cap companies which yielded the following tables:

	Stock	T-Statistic	P-Value	Significant
0	AAPL	7.470743	1.488742e-13	True
1	AMZN	13.284285	8.908639e-38	True
2	COST	1.853583	6.403343e-02	False
3	JPM	-0.694990	4.871900e-01	False
4	BAC	-0.776703	4.374806e-01	False
5	CRM	4.106570	4.275586e-05	True
6	CVX	-0.670132	5.028972e-01	False
7	KO	0.565683	5.717107e-01	False
8	MA	2.809441	5.039778e-03	True
9	NVDA	-0.391217	6.957035e-01	False

	Stock	T-Statistic	P-Value	Significant
0	HAS	2.977765	2.959303e-03	True
1	SEAS	1.449190	1.475345e-01	False
2	JEF	0.956553	3.389769e-01	False
3	W	7.010274	3.874479e-12	True
4	WING	9.326124	4.795587e-20	True
5	CROX	11.969374	2.430163e-31	True
6	HXL	-1.101092	2.710677e-01	False
7	ALLY	0.138332	8.900003e-01	False
8	UHS	-1.338038	1.811265e-01	False
9	HLNE	-1.178793	2.387042e-01	False

- There is no particular patterns in this table, however, after a while of research, we found out that companies with no significant relation with earning release usually tend to be companies under the financial market which are funds, PE firms, and banks.
- One hypothesis given is that these firms often hedged their risks against the market and earning release dates is already incorporated into their models

d. *Volatility difference:*
- Since our graph is only able to denote the density of each category but totally omitted the comparison between large and small cap stock to see the difference thus we did the following calculations to test out the volatility level of each stock:

For each stock category:
- First we calculated the volatility of each stock at the earning dates

where:

$$Vol = \sqrt{\frac{\Sigma var}{n-1} * 252}$$

- Secondly, we take the difference between that figure and the volatility of each stock over the period.

- Finally, we averaged these differences to a number that yielded: 0.152 for large-cap stocks, and 0.338 for small-cap stocks

5. Conclusion

- In this research, we experimented with stock volatility levels at the earning release dates of the company.
- There is a clear relation between these two variables which could be shown through the distribution of the model as well as a simple linear regression for each stock's annualized volatility and the dummy variable of its earning release dates.
- However, companies that within the financial market are usually not affected by this particular factor. One hypothesis is that they usually hedge against market risk and this factor is already incorporated into their model.
- Small-cap stocks are much more volatile than large-cap stocks which can be seen that their annualized volatility difference between even and non-event is nearly double compared to large-cap stocks.

Note: We also have prepared the codes for other factors, but due to the amount of time that we have we have to reduce it into one. The longer version of the research could also be able to implement on other factors not just earning but also, FOMC, GDP, conc_conf, housing, etc. Since getting and cleaning data requires a rigorous amount of time we can't find a fitted solution to incorporate all the information into our research. However, we will just send in our code compiling file for other factors for you to examine.

Although reducing the data is not a Multi-factor model anymore, we believe that this could be a great foundation for it. By thoroughly looking through all possible corners of one factor we can easily create a streamlined process for other factors as well!