COMBINING TRADITIONAL CAPM AND SENTIMENT ANALYSIS FOR TIME-BASED PORTFOLIO OPTIMIZATION OF STOCKS AND CRYPTOCURRENCIES

DAVID NNAMDI (113449330)

Abstract

This project applied concepts of capital asset pricing, sentiment analysis and Markowitz portfolio theory in building optimal investment portfolios for a novice investor that aims to combine traditional stock investment with cryptocurrency investments. Utilization of sentiment analysis on tweets as a proxy for cryptocurrency price prediction was investigated resulting in more streamlined approaches for price prediction. Results of portfolio optimization suggests that diversifying investment portfolio to include cryptocurrency may yield significantly higher returns with minor increase in expected risk.

Finally, methods that can be utilized for continuously optimizing portfolio after forecasting crypto price in several timesteps is presented, although it is not utilized in this study due to constraints from the twitter API.

Introduction

Problem Statement

The modern investor prefers a diverse mix of medium earning investments with lower variability (stocks) and higher earing investments with higher variability (cryptocurrencies). To effectively optimize portfolio returns, investors wealth must be shared to both investment types with a target minimum return and maximum risk exposure. This must be effectively estimated assuming the mean variance model and solving the Markowitz for effective portfolio optimization in an iterative manner.

Sentiment Analysis & Utilization in Price prediction

Sentiment analysis which entails the utilization of Natural Language Processing and other machine learning algorithms has become very popular due to the advent of social media through which information is relayed in real time to millions of people across the globe. Sentiment analysis rides off the efficient market hypothesis (EMS) which states that financial market movements depend on news, current events, and product releases, each having a significant impact on a company's stock value – or in our case, cryptocurrency price (Fama, 1965).

The use of twitter as a data source for sentiment analysis and its high correlation with rise and fall in stock prices have been demonstrated by several authors including (Pak & Paroubek, 2010) (Ruiz, Hristidis, Castillo, Gionis, & Jaimes, 2012) (Pagolu, Challa, Panda, & Majhi, 2016).

Capital Asset Pricing Model (CAPM)

The capital asset pricing model was developed primarily by Sharpe, Lintner and Mossin and it is utilized to evaluate the correct price of a risky asset considering its mean variance. The model assumes the following:

- There is market equilibrium (i.e., prices of securities adjust such that aggregate demand equals supply)
- All investors optimally hold risky assets in the same relative proportion
- All investors agree on their estimates of returns and risk correlations

The traditional form of the model can be written as follows:

$$E[R_i - R_f] = \beta_i E[R_M - R_f] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where R_i is the return on investment, R_f is the risk-free rate, β_i is the beta of the investment (a measure of riskiness) and $E[R_M - R_f]$, is the market risk-free premium.

Equation 1 above can be rewritten to perform a time series regression to obtain beta, based on a stock's historical performance. The formulation is given as:

$$\left(R_i - R_f\right)_t = \bar{\alpha} + \beta_i\left(R_M - R_f\right)_t \dots\dots (2)$$

Where $\bar{\alpha} = \alpha + e_i$, and e_i is the error term

The Alpha α represents the excess return for the assumed risk and may be interpreted in one of the following ways:

- $\alpha > 0$ portfolio gives a positive return no matter what the market does
- $\alpha = 0$ investment has earned sufficient return for the risk taken
- $\alpha < 0$ investment has earned too little for the risk

In this project the time series method for the evaluation of beta and alpha will be utilized for all selected stocks.

Selection of Stocks for Analysis

Stocks of eleven companies (including energy companies, technology firms, entertainment-tech companies, investment companies and banks) were assessed using CAPM before a final selection of the top 5 stocks to be used in the portfolio optimization was made. The rationale behind this process was to replicate what a novice investor with no in-depth knowledge of stock trading, company financial management and business would do. The following companies were selected for analysis:

- Apple *(NASDAQ: "AAPL")*
- Tesla *(NASDAQ: "TSLA")*
- Alphabet Corp *(NASDAQ: "GOOGL")*
- Berkshire Hathway *(NASDAQ: "BRK-A")*
- Microsoft *(NASDAQ: "MSFT")*
- JP Morgan Chase & Co. *(NASDAQ: "JPM")*
- Amazon *(NASDAQ: "AMZN")*
- Netflix *(NASDAQ: "NFLX")*
- Johnson and Johnson *(NASDAQ: "JNJ")*
- ExxonMobil *(NASDAQ: "XOM")*
- Coca-Cola co *(NASDAQ: "KO")*

Out of the eleven companies listed above, 6 pay dividends which complement earnings from returns based on stock price, however in the CAPM evaluation, this was not considered.

Choice of Cryptocurrencies

Bitcoin and Dogecoin are the choice cryptocurrencies for this project as they lend themselves perfectly to sentiment analysis. Within the past year, prices have surged and slumped based on tweets by Tesla CEO – Elon Musk indicating adoption of Bitcoin as payment for Tesla cars and later rescinding this decision citing Bitcoin carbon footprint whilst in favour of Dogecoin. Other information detailing uncertainty about adoption as global payment alternative contributed significantly to the prices of these 2 coins.

Data Types & Sources

Data to be used for this project include stock market price data and twitter sentiments. This data will be sourced from yahoo finance, twitter.com and bitcointweets.com (For sentiment analysis comparison).

Portfolio optimization

The objective of portfolio optimization is simply to invest in assets that will yield maximum returns with minimum risk. The way this is done is by evaluating the efficient frontier in a risk-return space and optimizing weights (fraction of wealth invested) with a set objective. For this study, the Sharpe ratio $SR_i = (ER_i - R_f)/\sigma_i$ which is the excess return over risk is maximized. The higher the ratio, the greater the return vs the risk taken.

Methodology

Using CAPM To streamline Stock Selection for portfolio

Using the time series regression method of the CAPM, each of the eleven stocks were evaluated over a 30-month period (January 1, 2019 – June 30, 2021). In the model, the market returns are assumed to be the returns of the S&P 500 stocks and the risk-free rate is obtained from the return of a 3-month US treasury bill for the stated period. The 3-month treasury bill still has inherent risks but due to the short tenure, it is considered by investors to be approximately risk free.

Data for individual stocks were pulled using yahoo finance and all the analysis is automated in Python for easy repeatability of process, a key area of data science.

Fig 1 below illustrates the analysis done for Apple *(NASDAQ "AAPL")*. In 1(a), the plot of the closing stock price for each day over the 30-month period is shown along with a short and long-term moving average (MA) of 20 & 200 days respectively. Ideally, the short-term MA being above the long-term MA is an indication of positive stock performance and a buy indicator. If the short-term MA is below the long-term MA, an investor may consider selling stocks.

In 1(b), the monthly time series regression to obtain beta and alpha for Apple over the stated period is shown. The red linear line is the linear approximation of equation (2).

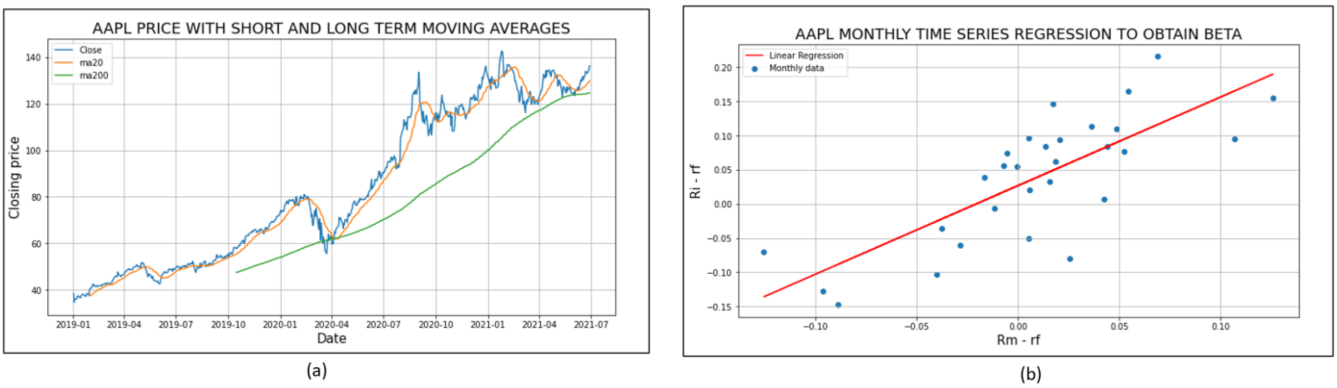


Figure 1: (a) Apple stock prices and moving averages (b) Time series regression of Apple stocks to obtain Alpha and Beta value

This same process was repeated for the 10 other stocks and the results are given in table 1 below:

Stock	Alpha	Beta	Rsquared	Expected Return	Risk
TSLA	0.078	2.820	0.386	0.963	0.703
XOM	-0.017	1.609	0.617	0.019	0.383
AAPL	0.027	1.296	0.589	0.509	0.361
JPM	0.001	1.226	0.600	0.209	0.377
BRK-A	-0.005	0.967	0.676	0.127	0.240
GOOGL	0.012	0.959	0.528	0.338	0.311
AMZN	0.010	0.942	0.385	0.324	0.302
JNJ	-0.004	0.738	0.524	0.127	0.229
KO	-0.007	0.728	0.469	0.088	0.256
MSFT	0.021	0.704	0.532	0.409	0.322
NFLX	0.004	0.531	0.184	0.277	0.402

Table 1: Summary of CAPM results

A visual representation of table 1 is presented in figure 2 a-e below. Analysing Alpha and Beta, the Tesla stock visibly stands out as the highest, thus Tesla has the highest returns over the 30-month period. However, the risk/volatility is also the highest. Stocks of ExxonMobil, Berkshire Hathaway, Johnson and Johnson and Coca-Cola all have negative alpha values and the smallest returns amongst the pack. This indicates that they are not good investments for the kind of risk they pose. The R-squared metric simply evaluates the correlation between stock price movement of each company and the market and does not significantly describe the risk-return for each stock.

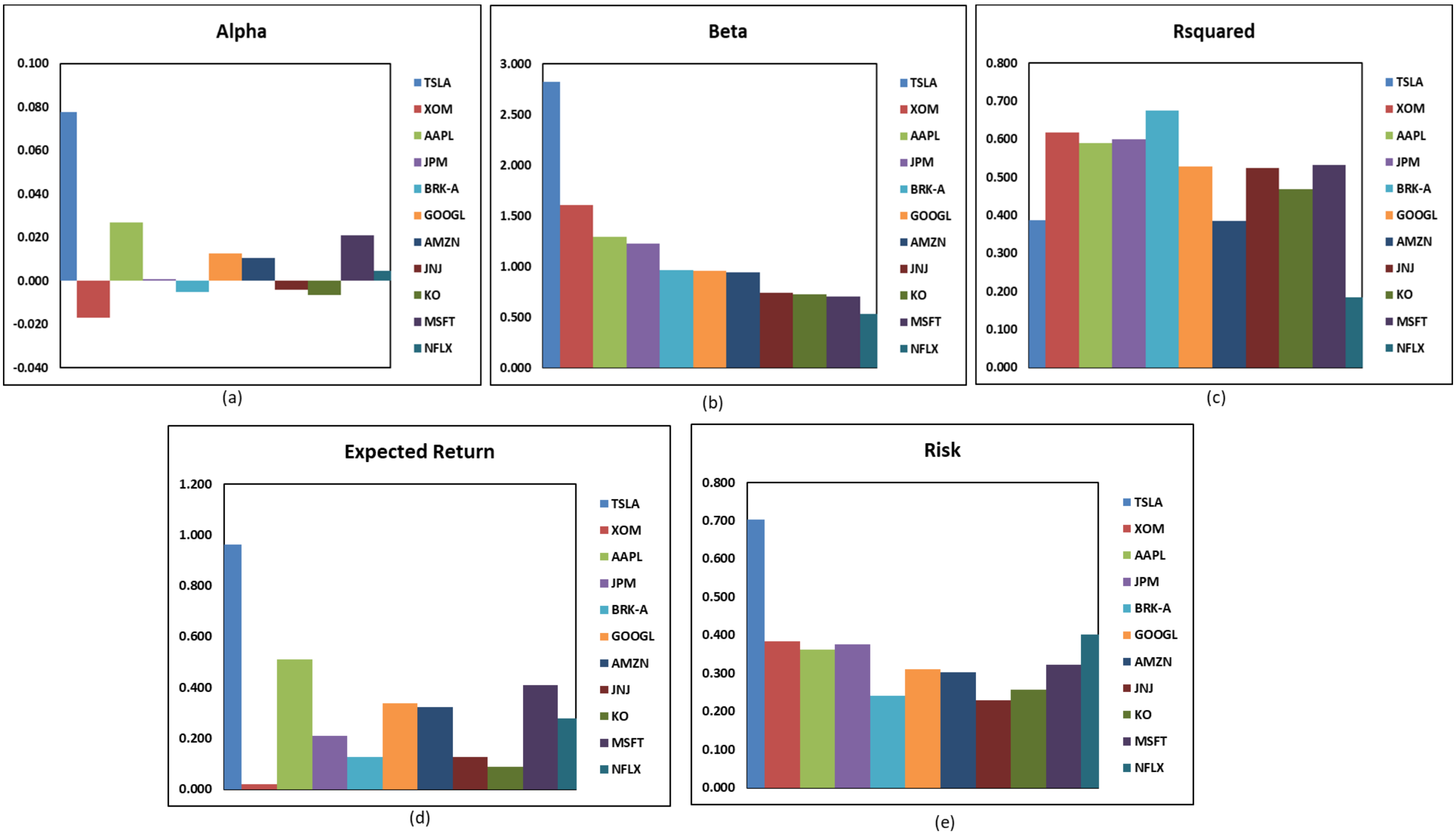


Figure 2: Comparison plots of CAPM statistics for different stocks

Stock Ranking & Selection

To better rank all the stocks for final selection, the following methodology was defined:

- Consider Alpha, Beta, Expected Return and Risk as qualifiers
- Rank each of the qualifiers as follows:
  - Alpha, Beta & Expected Return: Highest to Lowest (rank 1-11)
  - Risk: Lowest to Highest (rank 1-11)
- Sum all qualifiers scores and rank stocks based on this total score. Least score is the highest rank.

The ranking results are illustrated in table 2 and the top 5 selected stocks for portfolio optimization are: Apple, Tesla, Alphabet Corp., Amazon, and Microsoft.

Table 2: Ranking of stocks

Stock	Alpha	Beta	Expected Return	Risk	TOTAL SCORE	RANK
AAPL	2	3	2	7	14	1
TSLA	1	1	1	11	14	2
GOOGL	4	6	4	5	19	3
AMZN	5	7	5	4	21	7
MSFT	3	10	3	6	22	5
BRK-A	9	5	9	2	25	4
JNJ	8	8	8	1	25	8
JPM	7	4	7	8	26	6
KO	10	9	10	3	32	10
XOM	11	2	11	9	33	9
NFLX	6	11	6	10	33	11

The CAPM results plots for remaining top 4 stocks (like that of Apple illustrated in figure 1 is presented in the appendix).

Sentiment Analysis of Cryptocurrencies

Sentiment analysis is a natural language processing (NLP) technique used to determine whether data is positive, negative, or neutral. For this study, the relationship between twitter sentiments and cryptocurrency prices were investigated to evaluate price predictive approaches. A Python-based NLP package called **Textblob** was used to analyse random tweets generated from the twitter platform API.

The process for this analysis is as follows:

- Create a twitter developer account to have access to its API required for pulling tweets
- Using a Python package called **Tweepy**, connect to the twitter developer API platform using user specific access tokens.
- Pull required number of tweets using Tweepy up to application number of tweets and time limits (free developer apps on twitter have a maximum of 900 tweet requests every 15 minutes)
- Pass the pulled tweets into Textblob and get the number of positive, negative, and neutral tweets
- Calculate a sentiment index which is cumulative in its formulation. The proposed sentiment index utilized for this study is given as:
  - ➢ Calculate the cumulative sum of positive and negative tweets per time step as filtered by Textblob
  - ➢ Add the positive and negative cumulative sums to get the total cumulative sum
  - ➢ Calculate sentiment index as (positive / total) - (negative / total). Because values are cumulative, they can be trended with changes in positive or negative user sentiments
- Plot price of crypto currency vs sentiment index to evaluate trend match. Prices pulled from yahoo finance

Based on the above process, analysis of Dogecoin and Bitcoin was done for a 7-day period and the results are presented in figure 3 below.

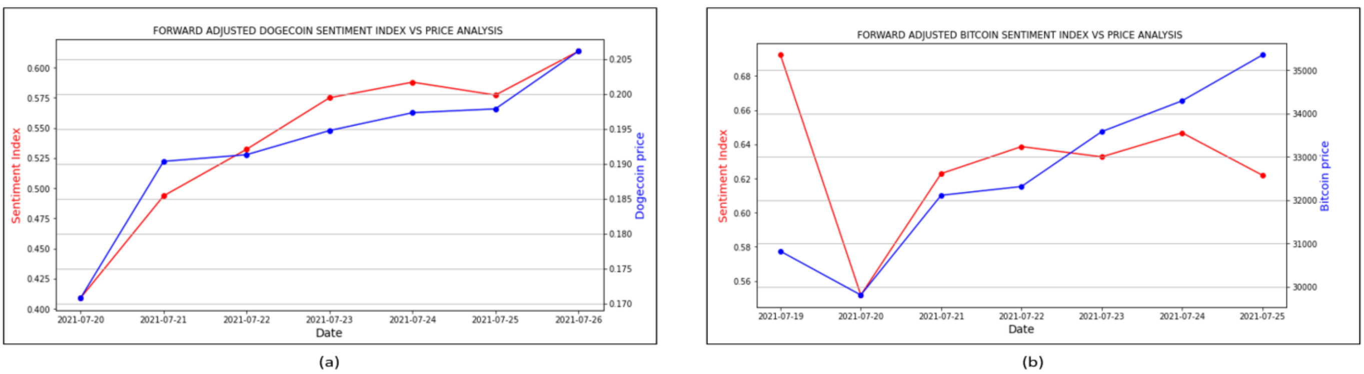
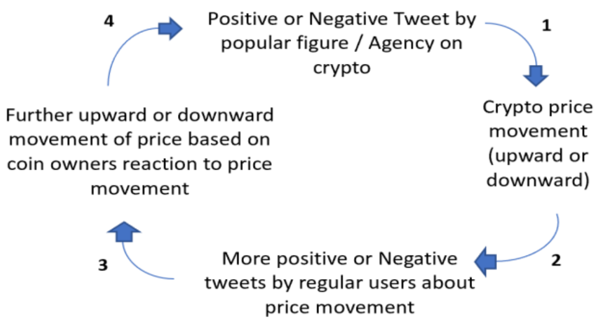


Figure 3: (a) Dogecoin sentiment index and price vs time (b) Bitcoin sentiment index and price vs time

In 3a, Dogecoin was analysed between August 20 – 26, 2021. Examining the plot, one can easily tell a strong correlation between the price of Dogecoin and twitter sentiments. This is not surprising considering the origins of Dogecoin.

Dogecoin was created as a joke crypto currency by an IBM software engineer with its logo fashioned after a popular internet meme. The coin remained flat for nearly 2 years until billionaire founder of Tesla, Elon musk took an interest in it and started tweeting about it in late 2020 (or early 2021) after which the price skyrocketed and has since fluctuated based on his and other prominent persons tweets. Simply put, dogecoin price rise is a function of speculation and since an unlimited number of coins are available for mining, it lends itself perfectly to sentiment analysis.

In addition, the price of Dogecoin and the resulting sentiment index is implicit on a daily timescale. The cycle below illustrates this concept:



To be able to use sentiment analysis to predict dogecoin price, one must take advantage of a very short time window (which is a lag period for coin holders to react to sentiment) between 1 & 2, and 2 & 3 to extract twitter sentiments and trade coin based on forecast. This time window will typically be between 1 minute to 1 hour. This requires near real-time automated analysis which was not possible for this study due to twitter API free apps limitation.

In 3(b), Bitcoin was analysed between August 20 – 26, 2021. While there is agreement between sentiment index trend and Bitcoin price for the first 3 days, there is some deviation towards the end. This indicates that consumer sentiments alone do not affect price of bitcoin. Again, this is expected as bitcoin has been around for much longer than crypto, the number of coins available in circulation are limited, and it is also affected by global adoption by central banks and other financial institution as payment means, it is also subject to complex demand and supply. In the past 5 years, bitcoin has seen 3 cycles of price rise and fall. The most recent being in 2021, was significantly bumped up by user sentiments and new interests due to newly unemployed persons (due to COVID 19) seeking means to generate income.

Figure 4 illustrates the above discussions for plots 3a and 3b.

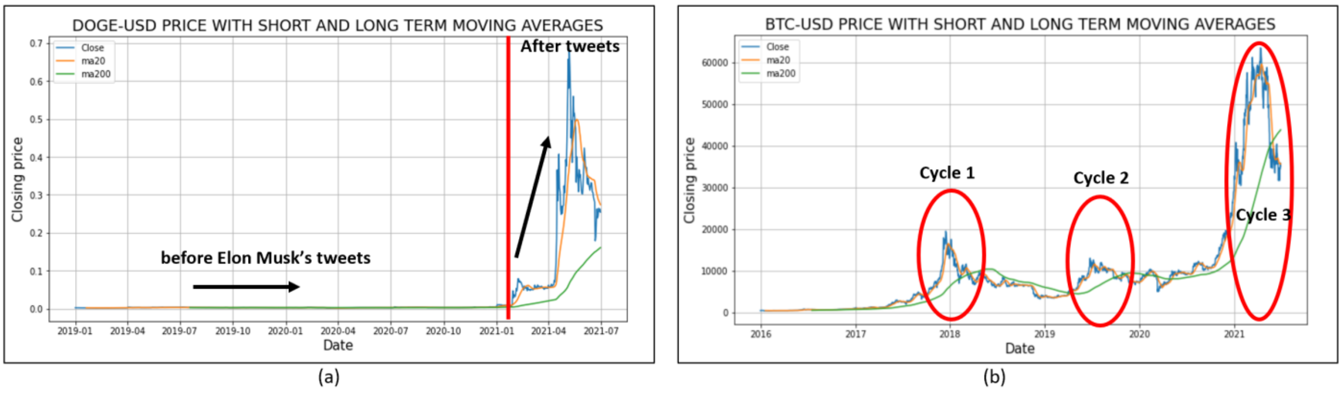


Figure 4: (a) Dogecoin price history showing movement before and after Elon Musk adoption and tweets (b) Bitcoin price history showing cycles of rise and fall, with most prominent in 2021 fueled by COVID-19

Clearly, sentiment analysis from consumer tweets alone cannot be used to predict price of bitcoin, one must consider other factors including regulatory adoption and global financial environment to adequately predict price.

Portfolio Optimization

To generate the optimal investment portfolios, the top 5 stocks and crypto currency were combined in 3 different cases. In each case, the risk return space was, and efficient frontier was generated using Monte-Carlo simulation. The following steps outline how the portfolio optimization for each case was implemented in Python:

1. First each stock average return and risk over the specified period was collected. Note that average return was the daily log returns annualized over 252 trading days in a year.
2. Generate random weights using a random number generator, ensuring the sum of weights equal 1
3. Calculate the portfolio expected return using $ER_p = \sum_{i=1}^{n} w_i ER_i$
    a. Where ER_i is the expected return on stock i and w_i is the weight attached to stock i
4. Calculate the portfolio variance using $\sigma_p^2 = \sum w_i^2 \sigma_i^2 + \sum\sum w_i w_j \sigma_{ij}$
    a. Where $\sum w_i^2 \sigma_i^2$ represents the summation of the variance of each stock i multiplied by the weight
    b. σ_{ij} is the covariance term. A covariance matrix was utilized for each stock
5. Steps 2 – 5 is repeated 5000 times to generate the risk return space using Monte-Carlo and to show the efficient frontier
6. Using a nonlinear optimizer, the Sharpe ratio is maximized, and the linear capital market line (CML) is drawn

The above is done for both a 30-month (Jan 2019 – June 2021) and 66-month (Jan 2016 – June 2021) period to evaluate the effect of historical data used in portfolio building and optimization. Note that the risk-free rate utilized here is also the average 3-month treasure bill returns over specified periods.

The 3 Cases that were considered in this study are:

- No crypto currency in portfolio (Stocks only)
- Stocks + BTC
- Stocks + BTC + DOGE

Figure 5 – 7 illustrates the result of the portfolio optimization in each case. The red line is the CML obtained after optimization.

Stock only

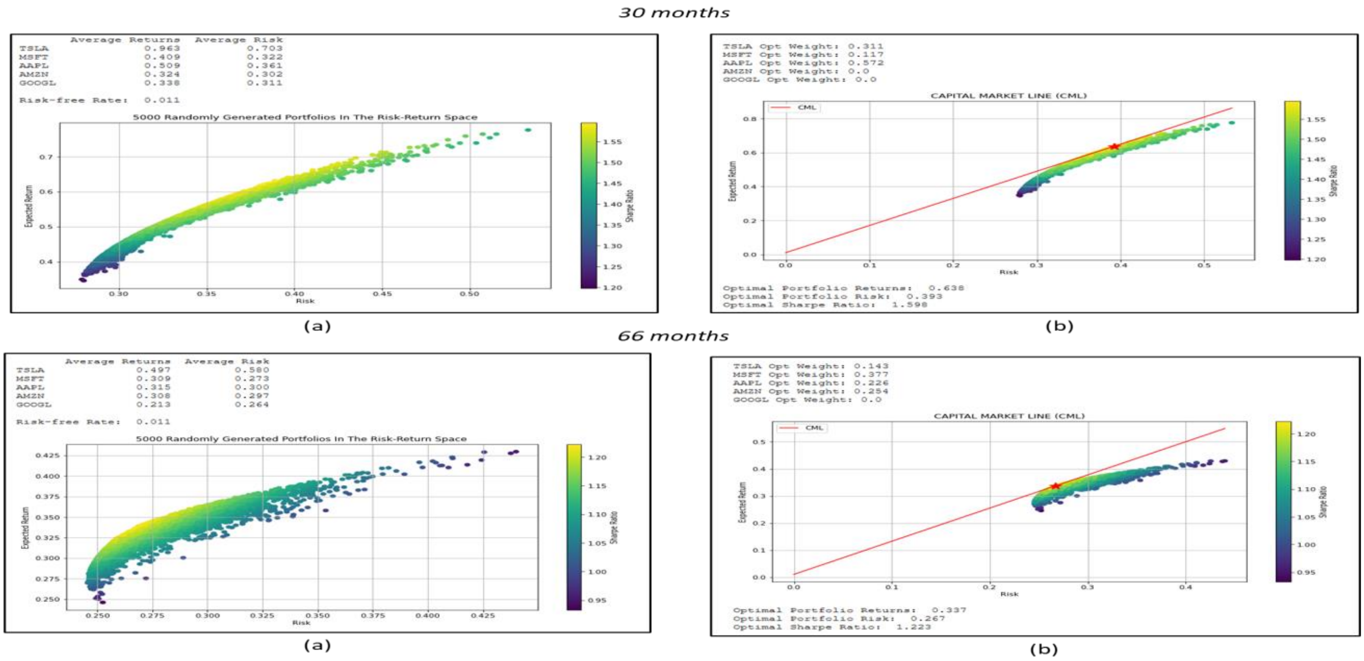


Figure 5: Risk return space for stock only portfolio. Capital market line illustrated in plots b for the 30 and 66-month period

Stocks + BTC only

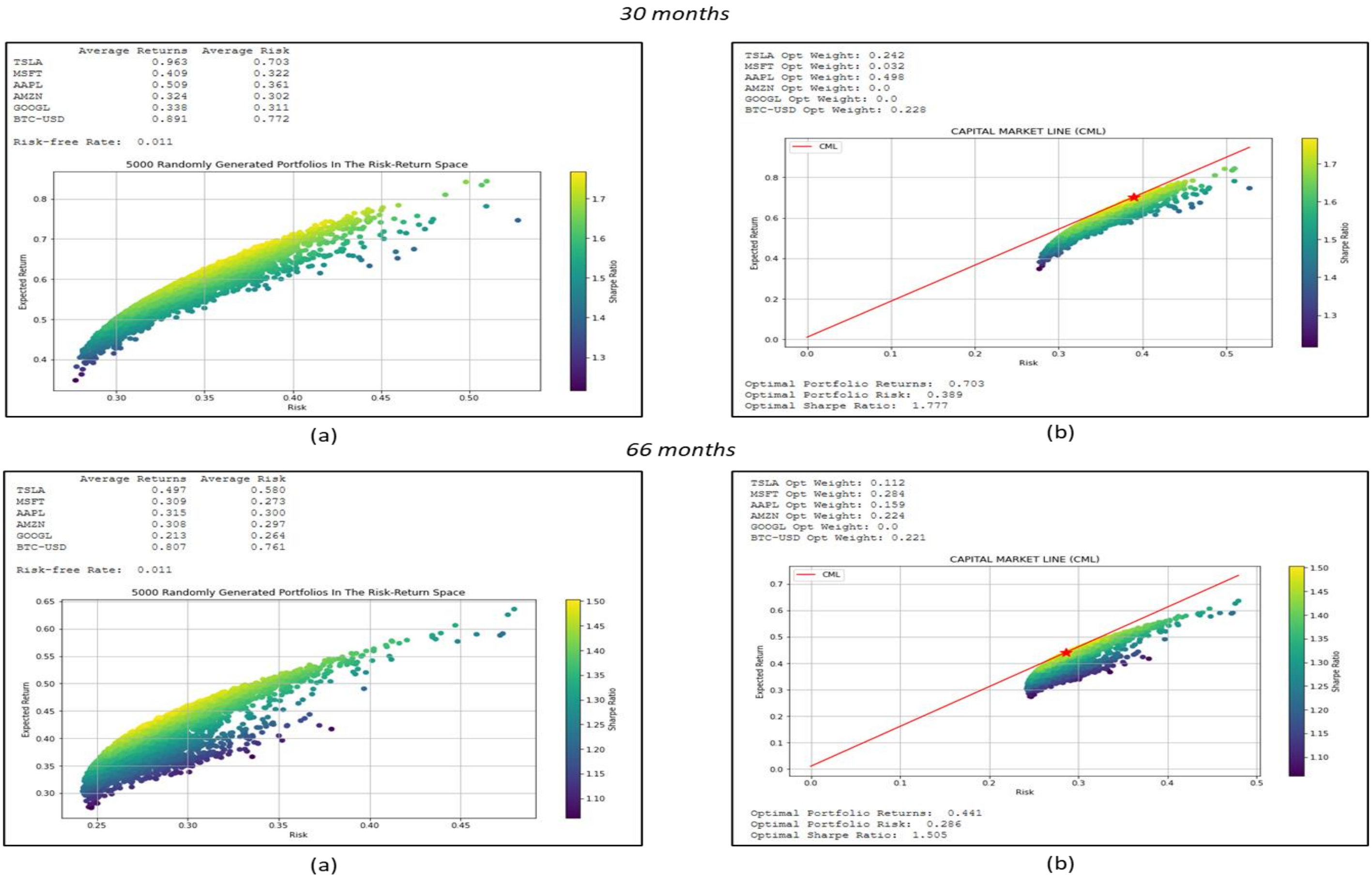


Figure 6: Risk return space for stock + BTC portfolio. Capital market line illustrated in plots b for the 30 and 66-month period

Stock + 2 coins

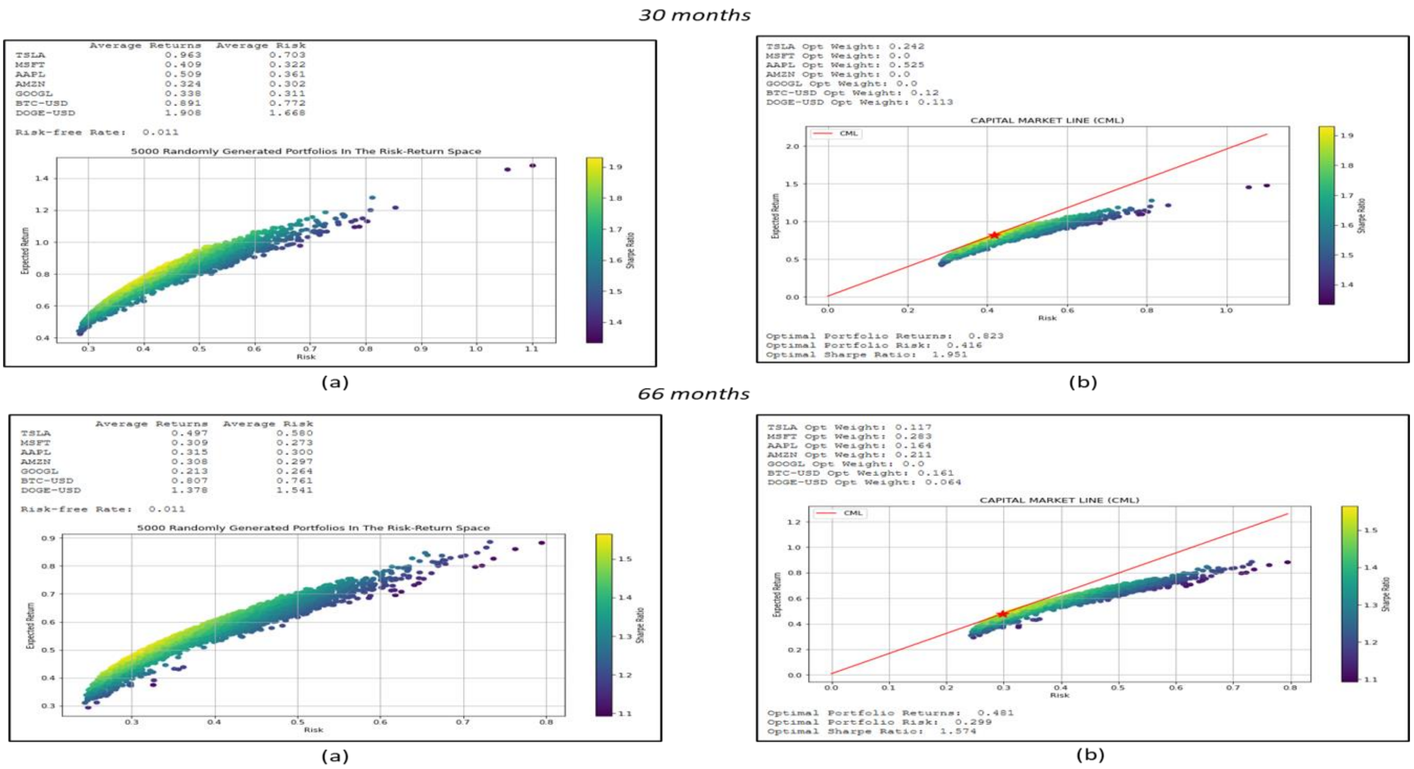


Figure 7: Risk return space for stock + 2 Cryptos (Dogecoin & Bitcoin). Capital market line illustrated in plots b for the 30 and 66-month period

Portfolio Analysis

The resulting weights from the portfolio optimization is summarized in figure 8 and tables 3a-c below. One key thing stands out, in all cases, the weight attached to the Alphabet (GOOGL) stock is 0. This indicates that while the Alphabet stock ranked 3rd in the grouping present in table 2, to maximize returns while minimizing risk (the objective of portfolio optimization), we need not invest in the company. The same goes for the Amazon stock in all cases but 66-month stock + BTC and stock + 2 coins cases.

In all cases, the apple stock takes majority of the available wealth for investment, this aligns with it being first in the ranking, it has the second highest return of all stocks evaluated with lower risk than the Tesla stock which has the highest risk and returns.

Finally, in almost all cases, the addition of crypto currencies bolsters the investors earnings by between 6 – 18 percent whilst only increasing risk by a maximum of 3 percent. This would excite the modern investor who aims to diversify their investment portfolios by including cryptocurrency.

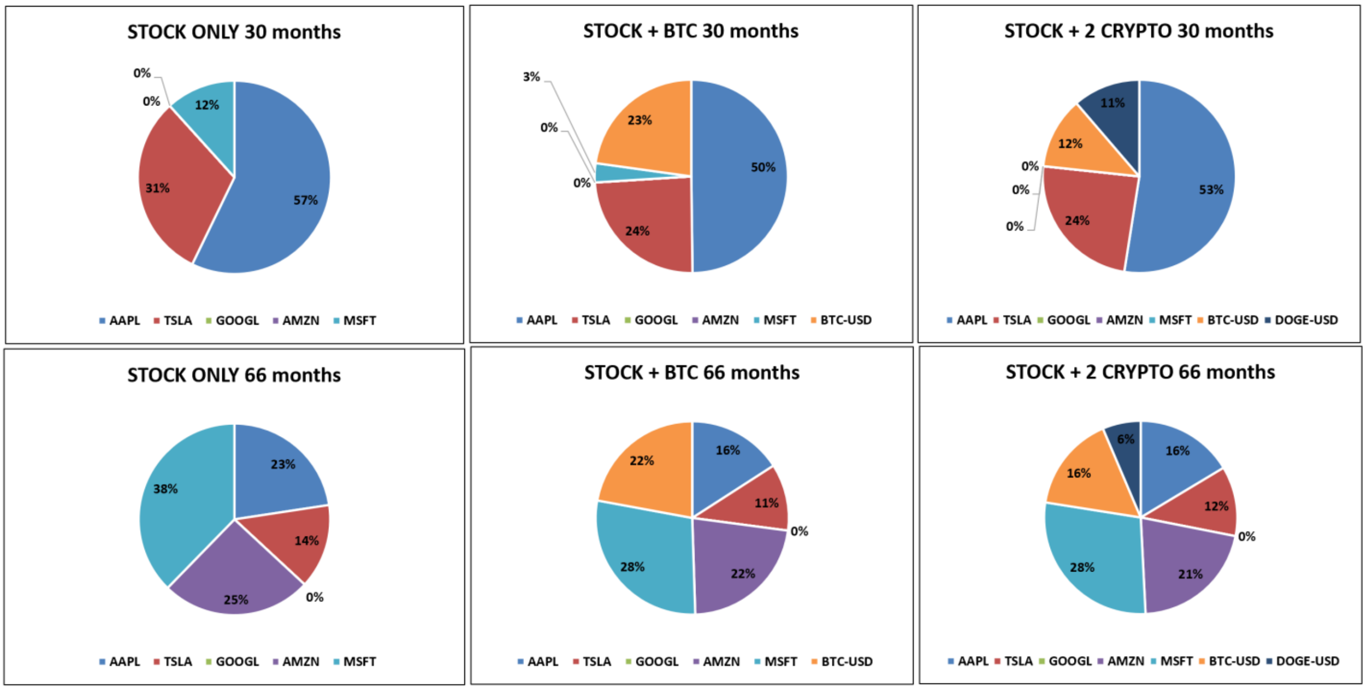


Figure 8: Pie-chart showing weight distribution of stocks in each optimize portfolio case for both 30 and 66-month period

STOCK ONLY PORTFOLIO		
	Weights	
Stock	30 months	66 months
AAPL	57%	23%
TSLA	31%	14%
GOOGL	0%	0%
AMZN	0%	25%
MSFT	12%	38%
TOTAL	100%	100%

Risk free Rate	1%	1%
Expected Return	64%	34%
Risk	39%	27%
Sharpe Ratio	1.598	1.223

(a)

STOCK + BTC PORTFOLIO		
	Weights	
Stock	30 months	66 months
AAPL	50%	16%
TSLA	24%	11%
GOOGL	0%	0%
AMZN	0%	22%
MSFT	3%	28%
BTC-USD	23%	22%
TOTAL	100%	100%

Risk free Rate	1%	1%
Expected Return	70%	44%
Risk	39%	29%
Sharpe Ratio	1.777	1.505

(b)

STOCK + 2 CRYPTO PORTFOLIO		
	Weights	
Stock	30 months	66 months
AAPL	53%	16%
TSLA	24%	12%
GOOGL	0%	0%
AMZN	0%	21%
MSFT	0%	28%
BTC-USD	12%	16%
DOGE-USD	11%	6%
TOTAL	1	1

Risk free Rate	1%	1%
Expected Return	82%	48%
Risk	42%	30%
Sharpe Ratio	1.951	1.574

(c)

Table 3: Portfolio optimization statistics

Overall, depending on the time horizon been considered for historical data extraction, the expected portfolio returns, and risks varies. Looking at only recent years in the 30-month period suggests expected returns between 64-82% annually and for the 66-month lookback period, the expected return for the optimal portfolio ranges from 34 – 48% annually. The risk-return for each case is summarized in figure 9.

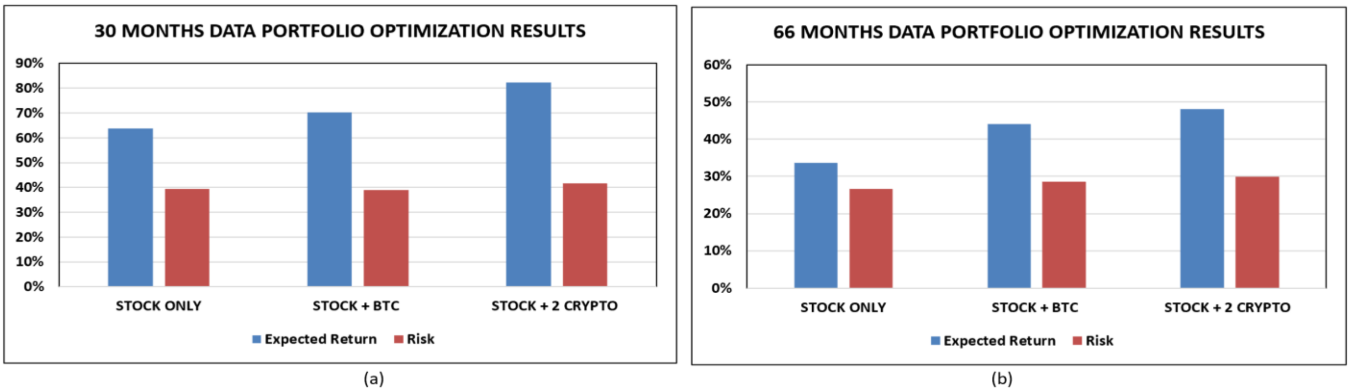


Figure 9: Expected return and risk for each portfolio case for the 30 and 66-month period

Discussion

Difficulty Encountered

The major difficulty encountered in this study revolves chiefly around the sentiment analysis and price prediction. Due to restrictions by twitter free API, the initial set objective of automatically optimizing portfolios based on price forecasts was hindered.

It was also discovered that for bitcoin, pure twitter sentiments may not be sufficient for price prediction, other data sources including news reports on regulatory and business adoption as payment solution and general stock market performance may be required for a more robust analysis. The former may be quantified using NLP and utilized to generate supplementary sentiments for further analysis.

Major Contributions

The earlier proposed Algorithm is shown in figure 10 below and while this was not utilized for this study, a template of this algorithm has been written in Python and is available on GitHub as open-source code for utilization by anyone who has paid access tokens for twitter developer API for further investigation.

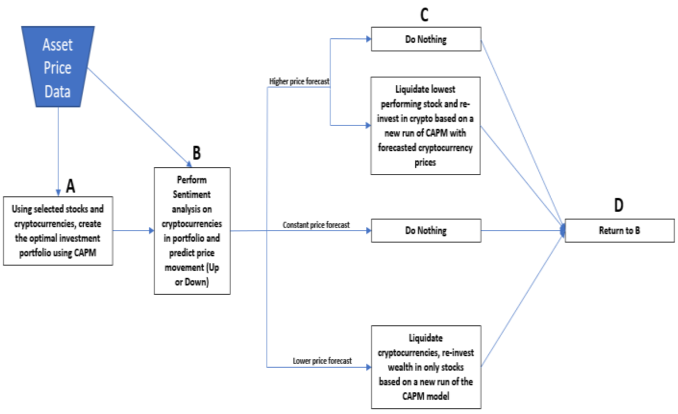


Figure 10: Continuous portfolio optimization with embedded sentiment analysis based priced forecast algorithm

All code used in this paper can be accessed on GitHub [here]

Conclusions

The project aimed explore portfolio diversification to include stocks and cryptocurrency. Utilizing CAPM and Markowitz portfolio optimization strategy, it was shown that adding crypto currency to investment portfolio will yield greater returns while only slightly increasing the risk.

The study sought to analyse twitter sentiments as a means of predicting cryptocurrency price, and while dogecoin price was observed to have strong correlation with twitter sentiments, it was determined that prediction can only be done with near real time data, that enables investor anticipate price movements. Bitcoin required more than just twitter sentiments to predict prices.

Finally, the study aimed to combine traditional portfolio optimization and sentiment analysis to improve returns generated by investment in crypto and stocks, and this algorithm has been developed, but require paid twitter access for full utilization.

References

Fama, E. F. (1965). The behavioir of stock-market prices. *The journal of Business*. doi:http://dx.doi.org/10.2307/2350752

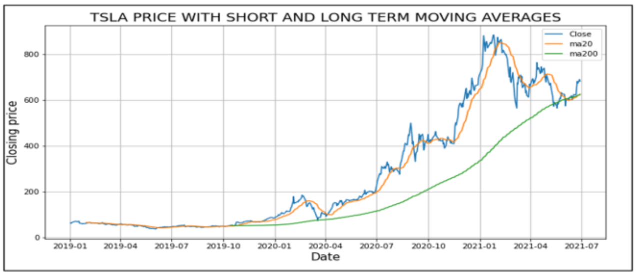
Pagolu, V. S., Challa, K. N., Panda, G., & Majhi, B. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *International conference on Signal Processing, Communication, Power and Embedded Systems (SCOPES).*

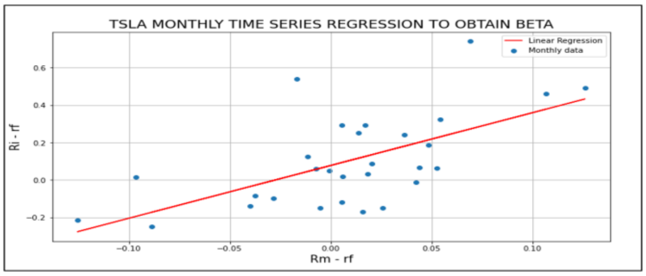
Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Seventh International Conference on Language Resources and Evaluation.*

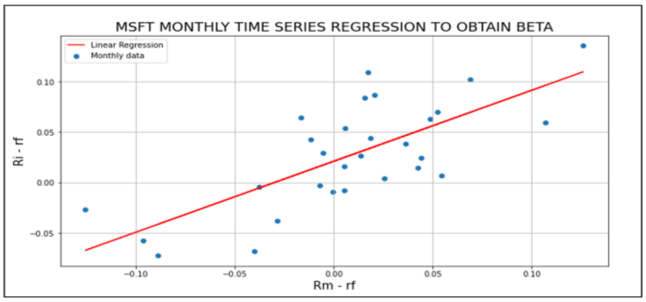
Ruiz, E., Hristidis, V., Castillo, C., Gionis, A., & Jaimes, A. (2012). Correlating financial time series with micro-blogging activity. *Fifth ACM International Conference on Web Search and Data Mining*, (pp. 492-499).

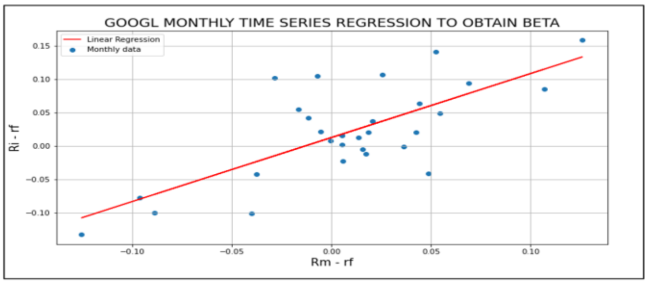
Appendix

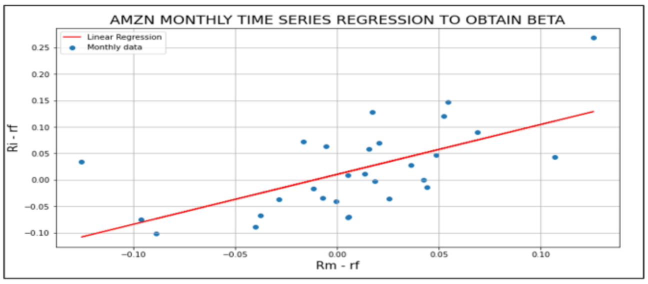
CAPM Results of 4 other selected stocks











CAPM of crypto currencies