Data

The Standard and Poor's 500, S&P 500, is a stock market index tracking the stock performance of 500 large companies listed on exchanges in the United States. This paper selects the top 6 representative stocks according to the market capitalization of the S&P 500—Apple (AAPL): 7.14%, Microsoft (MSFT): 6.1%, Amazon (AMZN): 3.8%, Tesla (TSLA): 2.5%, Berkshire Hathaway Class B (BRK.B): 1.7%, Meta (META), formerly Facebook, Class A: 1.4%, [1]. Adjusted closing prices from August 30th, 2021, to August 30th, 2022, are used for calculating the average return and covariance matrices to construct the efficient frontier. The selected stocks' performance is summarised by descriptive statistics, including mean, volatility, Sharpe ratio, Var, CVar, and Maximum Drawdown. This information on the ten chosen stocks is presented in Table 1, Table 2, and Figure 1. According to Table 1, TSLA has the highest average return and the highest Sharpe ratio, risk-adjusted return. Figure 1 reveals that presents TSLA the highest cumulative return, whereas the cumulative return of META decreased with fluctuation.

	Mean	Vol	Sharpe
TSLA	1.67%	13.82%	12.07%
AAPL	0.47%	6.69%	6.97%
BRK-B	0.10%	4.21%	2.28%
MSFT	-0.41%	6.65%	-6.14%
AMZN	-0.89%	9.64%	-9.28%
META	-3.47%	11.84%	-29.28%

```
Asset with Max Mean Return: TSLA
Asset with Max Sharpe: TSLA
```

Table 1. Descriptive statistics of the daily return of the ten stocks

	VaR (0.05)	CVaR (0.05)	Max Drawdown
AAPL	-3.31%	-4.15%	-28.35%
MSFT	-3.72%	-4.09%	-29.08%
AMZN	-4.34%	-6.47%	-44.64%
BRK-B	-1.89%	-2.67%	-25.60%
META	-5.08%	-7.86%	-59.22%
TSLA	-6.68%	-8.75%	-48.93%

Table 2. Tail metrics of the daily return of the ten stocks

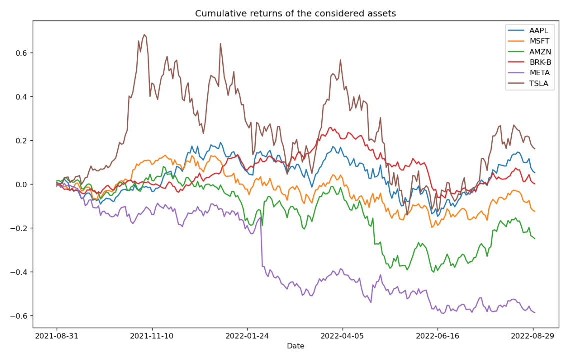


Figure 1: Cumulative Returns of the considered stocks

Methods

3.1 Return Prediction:

ARIMA Model is used to forecast returns. An ARIMA model requires time series to be differenced at least once to make it stationary and combine the autoregressive and moving average terms.

Auto-Regressive (AR only) model is one where Yt depends only on its own lags.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1 \quad (1)$$

Since Auto-Regressive Model, a linear regression model, uses its own lags as predictors and works best when the predictors are not correlated and are independent of each other, the time series data must be made stationary by differencing. The value of d in the ARIMA Model is the minimum number of differencing needed to make the series stationary. ADF test is used to find this d term that makes the series stationary. The null hypothesis of the ADF test is that the time series is non-stationary. So, if the p-value of the test is less than the significance level (0.05), then the null hypothesis is rejected and inferred that the time series is indeed stationary. [2]

The moving Average (MA only) model is one where Yt depends only on the lagged forecast errors, where the error terms are white noise errors of the autoregressive models of the respective lags.

$$Y_t = \alpha + \epsilon_t + \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + .. + \phi_q\epsilon_{t-q} \qquad (2)$$

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p}\epsilon_t + \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + .. + \phi_q\epsilon_{t-q} \qquad (3)$$

The Akaike information criterion (AIC) is a criterion for model selection among a finite set of models; the order of the ARIMA model with the lowest AIC will be applied to forecast returns for the next month. To do out-of-time cross-validation, the training and testing dataset is created by splitting the time series into two contiguous parts in approximately a 75:25 ratio [2].

Portfolio Optimization:

To optimize the portfolio by altering the weight allocation, one needs to initialize the weights, calculate the initial metrics, and use a random generator with a random state of 42 to generate weights for 100,000 portfolios. The weights must be uniform from 0 to 1 (as weights denote percentage holdings). The Monte Carlo Simulation generates 100,000 portfolios with different weights and draws the efficient frontier. Since Sharpe Ratio is a good metric that provides how efficient a portfolio return is, concerning how risky its composition is, the optimal choice is the portfolio with the largest Sharpe Ratio on the efficient frontier.

Result

First, we perform the ADF test on ten stocks' daily returns; all the p-value are insignificant, so we difference the series (d=1), and now the series is stationary.

```
AAPL  p-value: 0.312440  AAPL  p-value: 0.000000
MSFT  p-value: 0.495753  MSFT  p-value: 0.000000
AMZN  p-value: 0.607599  AMZN  p-value: 0.000000
BRK-B  p-value: 0.583605 BRK-B  p-value: 0.000000
META  p-value: 0.656557  META  p-value: 0.000000
TSLA  p-value: 0.215577  TSLA  p-value: 0.000000
```

Original series (d=0)   Difference once (d=1)

AAPL: AIC (1, 0) has the minimum score, and thus the order of the ARIMA model is (1,1,0).

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  253
Model:                 ARIMA(1, 1, 0)   Log Likelihood                -636.603
Date:                Tue, 30 Aug 2022   AIC                           1277.205
Time:                        19:46:09   BIC                           1284.264
Sample:                             0   HQIC                          1280.045
                                - 253
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0082      0.061      0.135      0.892      -0.111       0.127
sigma2         9.1570      0.765     11.968      0.000       7.657      10.657
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 2.48
Prob(Q):                              1.00   Prob(JB):                         0.29
Heteroskedasticity (H):               2.05   Skew:                            -0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.45
===================================================================================
```

MSFT: AIC (3, 2) has the minimum score, and thus the order of the ARIMA model is (3,1,2).

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  253
Model:                 ARIMA(3, 1, 2)   Log Likelihood                -782.262
Date:                Tue, 30 Aug 2022   AIC                           1576.524
Time:                        20:09:15   BIC                           1597.701
Sample:                             0   HQIC                          1585.045
                                - 253
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6700      0.079    -21.196      0.000      -1.824      -1.516
ar.L2         -0.9383      0.127     -7.397      0.000      -1.187      -0.690
ar.L3          0.0022      0.068      0.032      0.974      -0.132       0.136
ma.L1          1.6446      0.048     34.294      0.000       1.551       1.739
ma.L2          0.9470      0.045     20.878      0.000       0.858       1.036
sigma2        29.0296      2.521     11.513      0.000      24.088      33.972
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 1.10
Prob(Q):                              0.99   Prob(JB):                         0.58
Heteroskedasticity (H):               1.50   Skew:                            -0.12
Prob(H) (two-sided):                  0.06   Kurtosis:                         3.23
===================================================================================
```

TSLA: AIC (3, 2) has the minimum score, and thus the order of the ARIMA model is (3,1,2)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  253
Model:                 ARIMA(3, 1, 2)   Log Likelihood                -982.059
Date:                Tue, 30 Aug 2022   AIC                           1976.119
Time:                        20:17:41   BIC                           1997.295
Sample:                             0   HQIC                          1984.640
                                - 253
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.3009      0.117    -11.127      0.000      -1.530      -1.072
ar.L2         -0.8568      0.134     -6.415      0.000      -1.119      -0.595
ar.L3         -0.0631      0.068     -0.921      0.357      -0.197       0.071
ma.L1          1.3005      0.090     14.495      0.000       1.125       1.476
ma.L2          0.8830      0.083     10.685      0.000       0.721       1.045
sigma2       141.8853     10.499     13.514      0.000     121.307     162.464
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                30.57
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               0.78   Skew:                            -0.11
Prob(H) (two-sided):                  0.27   Kurtosis:                         4.69
===================================================================================
```

AMZN: AIC (3, 3) has the minimum score, and thus the order of the ARIMA model is (3,1,3)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  253
Model:                 ARIMA(3, 1, 3)   Log Likelihood                -695.684
Date:                Tue, 30 Aug 2022   AIC                           1405.367
Time:                        20:23:00   BIC                           1430.073
Sample:                             0   HQIC                          1415.308
                                - 253
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1840      2.210     -0.083      0.934      -4.515       4.147
ar.L2          0.0155      2.363      0.007      0.995      -4.615       4.646
ar.L3         -0.2009      1.353     -0.148      0.882      -2.853       2.451
ma.L1          0.1664      2.233      0.075      0.941      -4.209       4.542
ma.L2          0.0236      2.331      0.010      0.992      -4.545       4.592
ma.L3          0.1578      1.353      0.117      0.907      -2.493       2.809
sigma2        14.6338      0.995     14.712      0.000      12.684      16.583
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):               232.66
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               2.14   Skew:                            -0.15
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.70
===================================================================================
```

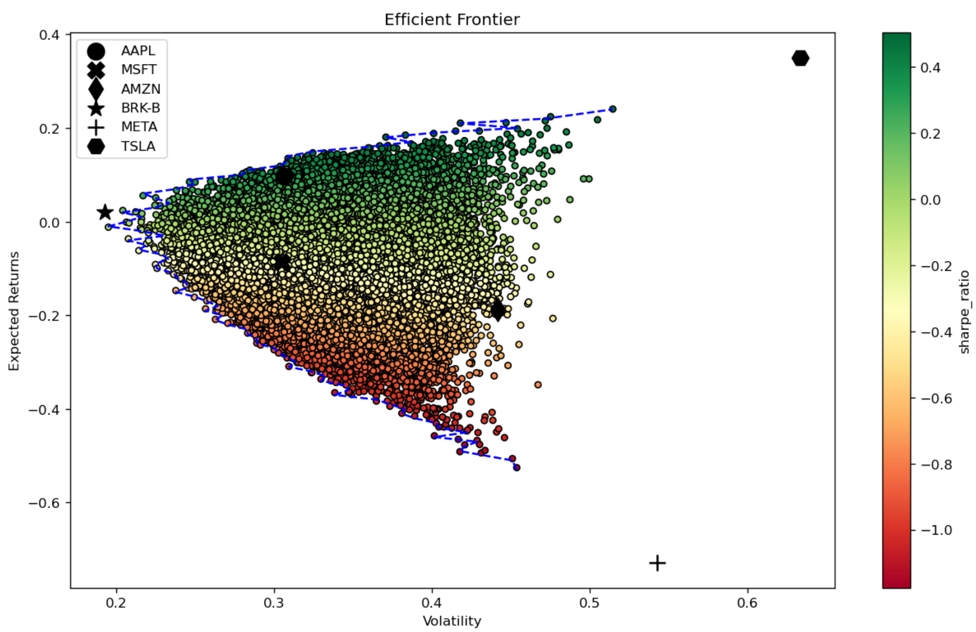
BRK.B: AIC (4, 3) has the minimum score, and thus the order of the ARIMA model is (4,1,3)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  253
Model:                 ARIMA(4, 1, 3)   Log Likelihood                -680.701
Date:                Tue, 30 Aug 2022   AIC                           1377.402
Time:                        20:25:16   BIC                           1405.637
Sample:                             0   HQIC                          1388.763
                                - 253
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5243      0.317     -1.655      0.098      -1.145       0.097
ar.L2          0.6228      0.460      1.353      0.176      -0.279       1.525
ar.L3          0.8365      0.246      3.400      0.001       0.354       1.319
ar.L4         -0.0600      0.068     -0.878      0.380      -0.194       0.074
ma.L1          0.6467      0.330      1.960      0.050       0.000       1.293
ma.L2         -0.5390      0.509     -1.059      0.290      -1.536       0.458
ma.L3         -0.9515      0.329     -2.895      0.004      -1.596      -0.307
sigma2        12.8799      1.407      9.152      0.000      10.122      15.638
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 1.56
Prob(Q):                              0.96   Prob(JB):                         0.46
Heteroskedasticity (H):               2.57   Skew:                            -0.02
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.38
===================================================================================
```

META: AIC (3, 2) has the minimum score, and thus the order of the ARIMA model is (3,1,2)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  253
Model:                 ARIMA(3, 1, 2)   Log Likelihood                -895.230
Date:                Tue, 30 Aug 2022   AIC                           1802.461
Time:                        20:21:20   BIC                           1823.638
Sample:                             0   HQIC                          1810.982
                                - 253
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.6760      0.108     15.570      0.000       1.465       1.887
ar.L2         -0.8893      0.181     -4.901      0.000      -1.245      -0.534
ar.L3         -0.0500      0.102     -0.492      0.622      -0.249       0.149
ma.L1         -1.7488      0.054    -32.304      0.000      -1.855      -1.643
ma.L2          0.9836      0.056     17.569      0.000       0.874       1.093
sigma2        70.5670      3.515     20.076      0.000      63.678      77.456
===================================================================================
Ljung-Box (L1) (Q):                   0.04   Jarque-Bera (JB):             13240.38
Prob(Q):                              0.84   Prob(JB):                         0.00
Heteroskedasticity (H):               1.00   Skew:                            -3.57
Prob(H) (two-sided):                  0.99   Kurtosis:                        37.79
===================================================================================
```

Maximum Sharpe Ratio portfolio has a performance with returns: 21.05% , volatility: 41.81%, sharpe ratio: 50.34%. The portfolio allocation weights are AAPL: 5.15% MSFT: 3.44% AMZN: 0.33% BRK-B: 31.49% META: 0.55% TSLA: 59.04%.



Predicted Portfolio Return

By doing the dot product of the optimal portfolio weight with the forecasting data of the next 30 days of the six stocks, one gets the portfolio's return for the next month. The cumulative return for next month is expected to be 0.001206%.

REFERENCES

1. Where to Find a List of the Top Stocks in the S&P 500. (2022, April 5). Investopedia. https://www.investopedia.com/ask/answers/08/find-stocks-in-sp500.asp

2. Prabhakaran, S. (2022, March 8). *ARIMA Model - Complete Guide to Time Series Forecasting in Python | ML+*. Machine Learning Plus. https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/