Portfolio Optimization

Dushyant Kumar 150242

Rishu Singh 150588

Mode: Application

Abstract

In this project, we have studied financial portfolio optimization problem. In particular, we have studied modern portfolio optimization theory which provides a mathematical framework for selecting the best portfolios. We took the application based approach and have tried different mathematical formulation based on different objectives of this problem and all these formulations are solved with the help of **cvx software** which in turn uses Interior Point Methods. Best portfolio results for each of formulations are simulated on Stanford dataset(Stanford). We have analysed each of the formulations and compared these formulations based on number of iterations and the time it took cvx to solve it. Based on our analysis and comparisons we have suggested the most efficient formulation for solving this problem.

1. Introduction

A portfolio is a combination of financial assets such as bonds, equities, stocks, and commodities etc. What we dealt here with was the investor's portfolio which describes the combination of assets and distribution of investment on each asset by the investor over certain period of investment. Portfolio optimization or portfolio selection aims to build the combination in a way that would fulfill any specified investment objective such as minimum risk or maximum expected return. We, in this paper aim to apply convex optimization techniques to optimize the portfolio for different objectives and compare them using the run time and iterations required.

The number and assets to invest on is chosen based on the investor's interests and other factors not relevant to our cause. Once the selection of assets is done and total amount of money(which again depends on investor's choice and position) to be invested is specified, portfolio optimization requires us to decide how the investment money would be distributed amongst the assets. Considering there are n assets, and w_i represents the fraction of money invested on asset i, we will have $\sum_{i=1}^{n} w_i = 1$. And $\sum_{i=1}^{n} Vw_i = V$ where V is the total money invested and Vw_i is the money invested on asset i.

The portfolio return and risk form two most important deciding factor in Portfolio Optimization. The portfolio return represents the profit or loss realized through investment in that particular portfolio and the risk accounts for the chance of objective failure of the portfolio.

2. Interior Point Method

The software(cvx) that we are using to solve different formulations of Portfolio Optimization problem uses Interior Point Method(IPM). Therefore, a brief review of IPM will be given in this section. Consider the following convex problem.

$$\begin{aligned} &\min_{x} f_0(x) \\ &\text{s.t. } f_i(x) <= 0 \quad \forall i \in [K] \end{aligned} \tag{1}$$

Interior Point Method can be employed to solve linear/nonlinear constrained/unconstrained convex problems. It first removes the constraints by using barrier functions which will be zero if constraint are satisfied and infinity solution goes outside the feasible region. Note that this barrier function is not differentiable. Equation (2) in the unconstrained version of problem (1).

$$\min_{x} \quad f_0(x) + \sum_{k=1}^{K} I[f_i(x)] \tag{2}$$

It then uses some approximation of barrier function which is differentiable so that newton method can be applied successfully. Most common approxmation is $I_t[f(x)] = \frac{-1}{t}\log(u)$ based on parameter t.

$$x^*(t) = \min_{x} t f_0(x) - \sum_{k=1}^{K} \log(f(x)) \tag{3}$$

IPM iterates over t and solve problem (3) using Newton method in each iteration. KKT conditions have shown that as we increase the value of t, $x^*(t)$ approaches to optimum value x^* of original problem. Also the duality gap in inversely proportion to parameter t so the duality gap is also small for large t. It also handles equality constraints as well.

3. Dataset

We have used the Standford dataset(Stanford) for portfolio optimization. In this dataset, financial data of 10 years, from January 2006 to December 2016 of 53 assets are given on day to day basis. The dataset consists of three tables that contain daily market close prices, daily market volumes, and daily market returns. Last asset is risk free asset with some fixed daily returns which is very small as compared to returns of other assets.

For the purpose of our experimentation, we have randomly choosen 4 assets and one additional risk free assets. We have total of 5 assets that we will be investing in.

Now since the data is in day to day basis and this doesn't make much sense as rates can fluctuate rapidly on a daily basis. We need to look into the value of assets over a period of time say, after one year because investment are made not in daily basis but over a period of time. So, We need to convert it into annual basis. If there are 't' trading days in a year. Therefore annualized returns and risk for a particular asset is calculated as

$$\text{Annualized returns} = t * (\text{Average daily returns})$$
$$\text{Annualized risk} = \sqrt{t} * (\text{Standard deviation})$$

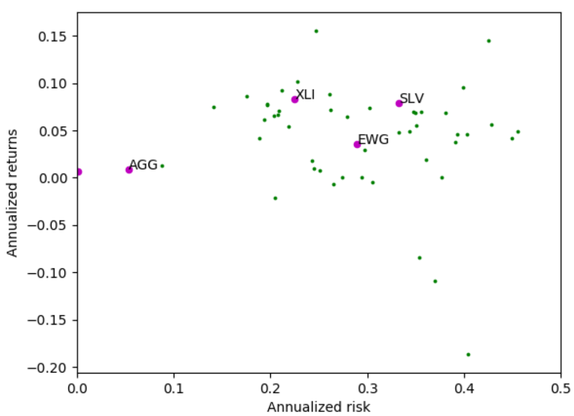

Figure 1. Annualized (t=250) risk-return curve for all the 53 portfolios. Bold points represent assets that we have adopted for experimentation. 'AGG', 'XLU', 'EWG', 'SLV' are just the names of assets. The point on the y-axis close to 0 represent return of risk-free asset('USDOLLAR'). Assets with low risk and high returns (i.e. top-left points) are better for investment.

In this dataset, there are total 8 assets that have annualized means negative. We won't be considering those assets because anyway fractional investment in those assets would be zero as they will always reducing our portfolio (or investment) value. Portfolio or investment value is the value of our investment that we have invested on certain portfolio after some period of time. We will be frequently using this term. Let's have a look at the plot of daily returns of selected assets. From this plot 2 we can get a sense about the volatility(risk) of each assets.

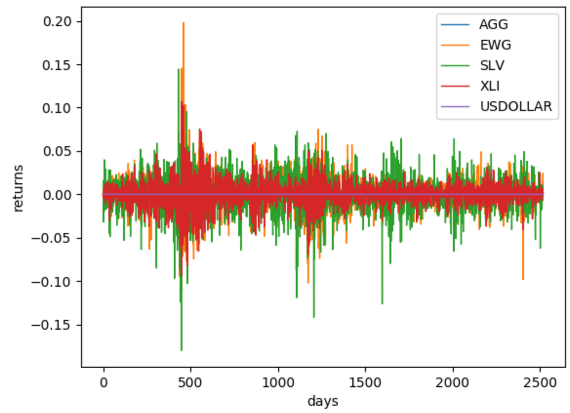

Figure 2. Curve of daily returns of selected assets. All the selected assets have negative spikes except the risk free one('USDOLLAR') which is parallel to x-axis with very less constant annual return. From this curve, 'SLV' (green curve) seems to be the most volatile with some large negative spikes and 'AGG' seems to be the most stable one.

4. Mathematical Formulation (Boyd, 2017; Cornuejols & Ttnc, 2006)

Consider r_i be the daily time series of returns $\{r_{it}\}_{t=1}^{T}$ of asset S_i. Let $\mu = [\mu_1, \ldots, \mu_n]^T$ be a n x 1 matrix that denotes the expected returns and $\Sigma = (\sigma_{ij})$ be a n x n symmetric covariance matrix of returns of assets.

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} \tag{4}$$

$$\sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \mu_i)(r_{jt} - \mu_j) \tag{5}$$

Standard deviation i.e. σ_i of asset S_i is $\sqrt{\sigma_{ii}}$. To convert μ_i & σ_{ij} in annual basis, we have to multiply them with t. Remember t is total no of trading days in a year.

Let x_i be the fraction of total funds invested in asset S_i, the resulting portfolio is $x = (x_1, \ldots, x_n)$ with $\sum_{i=1}^{n} x_i = 1$. Then the expected return and the risk of the portfolio is given as:

$$E[x] = \sum_{i=1}^{n} x_i \mu_i = \mu^T x \tag{6}$$

$$\text{Risk}[x] = \sum_{i,j} \sigma_{ij} x_i x_j = x^T \Sigma x \tag{7}$$

Here one thing to note that Risk is always positive i.e. $x^T \Sigma x \geq 0 \quad \forall\, x$. This follows the definition of positive semi definiteness, therefore Σ is a **positive semidefinite** matrix. Different values of expected returns and risk for random portfolios are visualized in figure 3.

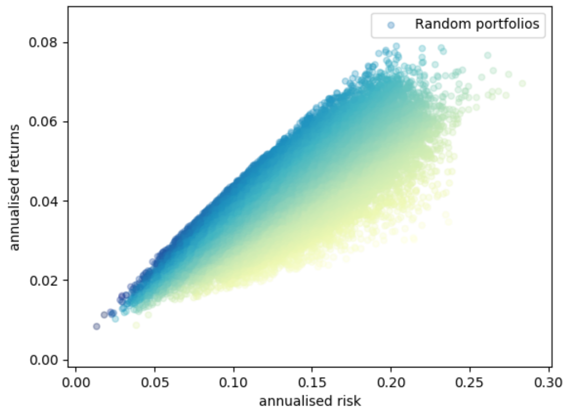

Figure 3. Annualized risk-return curve for 50,000 random simulated portfolios. X-axis represent the annualized risk and y-axis represent annualized returns. Each of the point denotes the risk and return of some random portfolio. Colour coding is based on Sharpe ratio(Sharpe) which basically measure the performance of portfolio.

5. Efficient Frontier

Efficient frontier is the upper boundary of the plot between expected returns and risk involved or in other words a set of efficient portfolios for different risk. Therefore, solving the portfolio optimization problem would require us to get the efficient frontier for the given class of assets. The investor would then be able to choose an efficient portfolio from the set defined by the efficient frontier based on the risk tolerance and required minimum expected return.

From fig 3, we can see that the top line forms the efficient frontier because it provides the lowest risk of a given target return and vice-versa. This can be obtained through formulating it as a constraint optimization problem and solving it using cvx software to get the desired efficient frontier.

5.1. Formulation 1 (Cornuejols & Ttnc, 2006)

We will pose the problem with the objective to minimize the variance of portfolio securities with different least target values of expected returns. Mathematically, the formulation is as follows

$$
\begin{aligned}
&\min_x \; x^T \Sigma x \\
&\text{s.t. } \mu^T x >= R \\
&\quad \mathbf{1}^T x = 1 \\
&\quad x \geq 0
\end{aligned} \tag{8}
$$

This formulation produces a **quadratic programming** problem. In first constraint R denoted the least expected returns. Last constraint makes sure that all the fractional investment i.e. x_i's are positive. Solving this optimization problem for different of R, fig 4 represent efficient frontier.

Analysis : This formulation is just a convex problem and can be solved efficiently for different values of R. One thing to note that whatever be the combination of assets one can consider for investing, each of the them have a definite upper limit on the maximum expected returns one can get regardless of portfolios selected. The assets we have selected have an upper bound in between 0.082 & 0.083. So, while trying out different values of R one have to be little careful otherwise the value of R that one is trying out can go outside the feasible region. One thing we have noticed that as we increase the value of R towards the uppper limit, solving time increases to 100 folds and no of iterations increases to 10 folds.

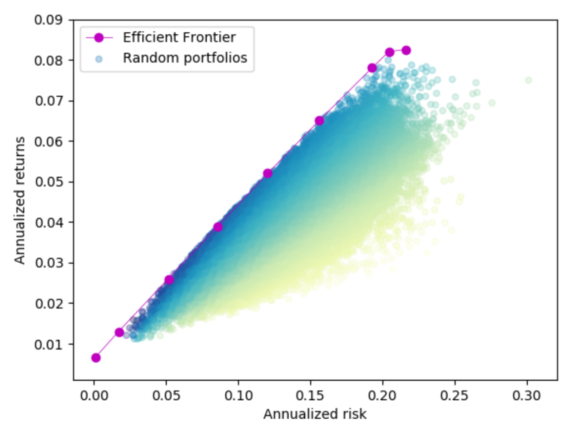

Figure 4. Efficient Frontier for formulation 1 plotted against the random simulated portfolios. Clearly efficient frontier curve gives the minimum risk for a fixed value of return.

5.2. Formulation 2 (Cornuejols & Ttnc, 2006)

Here the objective function is to maximize the expected returns of assets while the risk is not exceeding a particular fixed value. Mathematically it is represented as follows

$$\begin{aligned} &\max_x \ \mu^T x \\ &\text{s.t. } x^T \Sigma x <= \sigma^2 \\ &\mathbf{1}^T x = 1 \\ &x \geq 0 \end{aligned} \tag{9}$$

In first constraint σ^2 represent the maximum permissible value of risk. This is a quadratically constraint problem. This quadratic constraint can be coverted into second order cone constaint because Σ is a positive semidefinite matrix that we have shown earlier. The **SOCP** variant of this formulation is

$$\begin{aligned} &\max_x \ \mu^T x \\ &\text{s.t. } ||\sqrt{\Sigma} x||_2 <= \sigma \\ &\mathbf{1}^T x = 1 \\ &x \geq 0 \end{aligned} \tag{10}$$

This problem can be feed into cvx for different values of σ to obtain efficient frontier plotted in fig 5.

Analysis : This formulation is just another convex problem and can be solved efficiently for different values of σ. Similar to formulation 1, one have to little careful while chosing the minimum value of σ as there would be lower limit on the minimum risk that one can get. As we decrease the value of σ towards the lower limit, solving time increases to 100 folds and no of iterations increases to 3 folds.

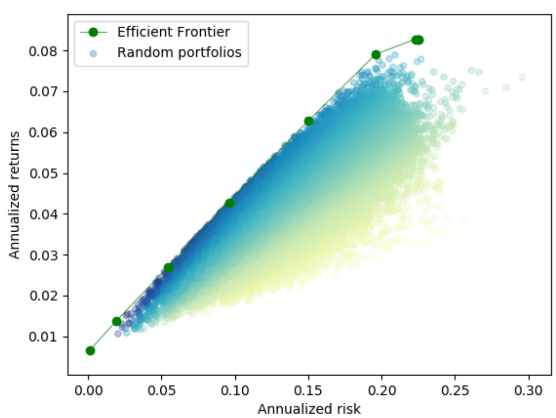

Figure 5. Efficient Frontier for formulation 2 plotted against the random simulated portfolios. Clearly efficient frontier curve gives the maximum returns for a fixed risk.

5.3. Formulation 3 (Cornuejols & Ttnc, 2006)

In this formulation, our objective function is a risk-adjusted return function where it trades off between risk and return simultaneously. Mathematically, it is represented as follows

$$\begin{aligned} &\min_x \ -\mu^T x + \delta x^T \Sigma x \\ &\mathbf{1}^T x = 1 \\ &x \geq 0 \end{aligned} \tag{11}$$

δ is a risk-aversion constant. Note that this is again a **quadratic programming** problem. Efficient portfolios is plotted in fig 6 which is obtained by solving it with different values of δ.

Analysis : This formulation is also a convex problem and can be solved efficiently for different values of δ. **This formulation is better** than the previous two as here one don't have to worry about the values of δ unlike the previous two where we have to carefully choose R or σ. This formulation takes also equal time and also run for almost same no. of iterations when one try different value of δ. So, in a way it covers the drawback that previous two have.

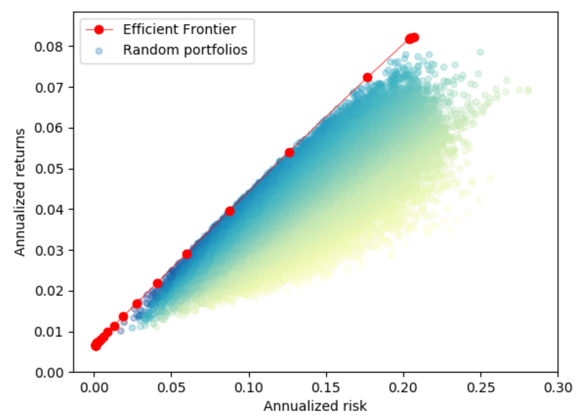

Figure 6. Efficient Frontier for formulation 3 plotted against the random simulated portfolios. Clearly efficient frontier curve gives the optimum portfolio.

5.4. Formulation 4 (Boyd, 2017)

Covariance matrix $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \mu_i)(r_{jt} - \mu_j)$. In a vectorized way, Σ can be written as $\Sigma = \frac{1}{T}\tilde{R}^T\tilde{R}$ where $\tilde{R} = R - \mathbf{1}\mu^T$. R is T x n return matrix, $\mathbf{1}$ is a column vector of length T & μ is also a column vector of length n. Don't get confused with R of formulation 1, here R means return matrix and it has no link with formulation 1. Therefore

$$\begin{aligned} x^T \Sigma x &= \frac{1}{T} x^T \tilde{R}^T \tilde{R} x \\ &= \frac{1}{T} x^T (R - \mathbf{1}\mu^T)^T (R - \mathbf{1}\mu^T) x \\ &= \frac{1}{T} ||(Rx - \mu^T x)||^2 \end{aligned} \tag{12}$$

This formulation is same as formulation 1 but now risk is written in different way.

$$\begin{aligned} &\min_x ||(Rx - \rho\mathbf{1})||^2 \\ &\text{s.t. } \mu^T x = \rho \\ &\quad \mathbf{1}^T x = 1 \\ &\quad x \geq 0 \end{aligned} \tag{13}$$

The objective function might look different from eq (7) but it's the same. We have just taken $\mu^T x$ into first constraint. This is a **least square** problem. Efficient portfolios is plotted in fig 7 for different values of ρ.

Analysis : This formulation is also a convex problem and can be solved efficiently for different values of ρ. This formulation suffers from the same problem of carefully choosing the value ρ so that one doesn't land in infeasible region. Like formulation 3 this takes equal time and also run for almost same no. of iterations for different value of ρ. But the time it takes if thousand times more than all the previous three formulations. More about this in section 5.

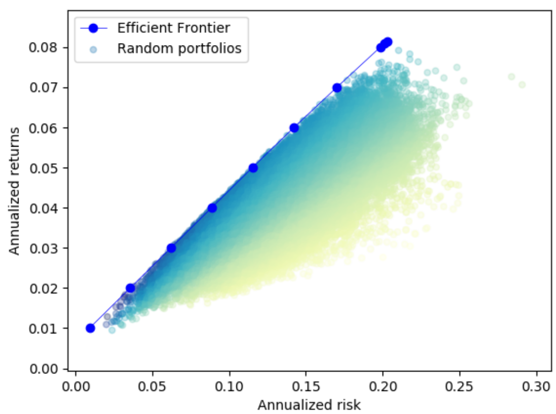

Figure 7. Efficient Frontier for formulation 4 plotted against the random simulated portfolios. Clearly efficient frontier gives the minimum risk for a fixed value of expected return.

6. Comparison

As we can see in fig 8 efficient frontiers for all the formulations turn out to be same which should be the case as all formulations are the equivalent ways of writing the same problem.

Here we will be comparing the different formulations based on two factors - average time and the average no. of iterations that the cvx software takes to solve each formulation.

The way we have computed average time is that we have solve the formulation for different values of parameter p (i.e. p=R in formulation 1) and take the average of run times and we have run this procedure for few(e.g. 100) times take the average of average run time of each procedure. This is done just to get stable time estimate. Mathematically,

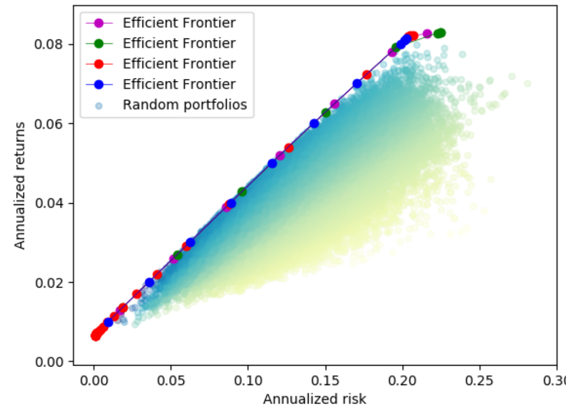

Figure 8. Efficient Frontier for all formulations plotted together. We can observe that efficient frontier of all the formulations exactly overlaps each other which does make sense.

$$\text{avg-time} = \frac{1}{100}\sum_{i=1}^{100}\Big(\frac{1}{\text{no. of p values}}\sum_{p=p_1}^{p_{\text{last}}} \text{run-time}(p)\Big) \tag{14}$$

100 is just a number we could have taken some other value also. Average no. of iterations are calculated by just running it one time over all values of parameter p there is no fluctuation in no. of iterations. Mathematically,

$$\text{avg-no of iters} = \frac{1}{\text{no. of p values}}\sum_{p=p_1}^{p_{\text{last}}} \text{no. of iters}(p) \tag{15}$$

Formulation	Time(in μsec)	No. of Iterations
1	71.75	173
2	86.49	12
3	46.75	52
4	25971.084	178

Table 1. Table showing the average time and the average number of iterations of different formulations when solved using cvx software. Clearly formulation 4 leads the table with far greater avg. time as compared to other formulations which means that formulation 4 is highly inefficient way of solving this problem

It is clear from the table 1 and our analysis of all the formulations that **formulation 3 is best** for solving portfolio optimization because of primarily two reasons

- Average time is less all than all the other formulations
- One doesn't need to worry about the parameter(i.e. δ) values in this formulation whereas other formulation require ones to be little careful in choosing one end of parameter value.

7. Portfolio value (Boyd, 2017)

Remember portfolio or investment value is the value of our investment that we have invested on certain portfolio after some period of time. Suppose V_t is the portfolio value, $\tilde{r}_t$ is the n x 1 rate vector of portfolios and w_t is the n x 1 vector denoting the fractional investment in each of the portfolios at time t. So the portfolio return is $\tilde{r}_t^T w_t$ and therefore profit over period t is $V_t \tilde{r}_t^T w_t$. The value of portfolio becomes $V_t(1 + \tilde{r}_t^T w_t)$ at the end of time t.

Suppose the investment is made over a period of time T, then the value of portfolio (or investment value) at the end of period T is :

$$V_T = V_1 \prod_{t=1}^{T} (1 + \tilde{r}_t^T w_t) \quad (16)$$

Here in eq (16) V_1 is the initial investment value. Since $\tilde{r}_t^T w_t$ is very small as compared to 1. Therefore $\prod_{t=1}^{T}(1 + \tilde{r}_t^T w_t) \approx 1 + \sum_{t=1}^{T} \tilde{r}_t^T w_t$. Hence V_T becomes

$$V_T = V_1 (1 + \sum_{t=1}^{T} \tilde{r}_t^T w_t) \quad (17)$$

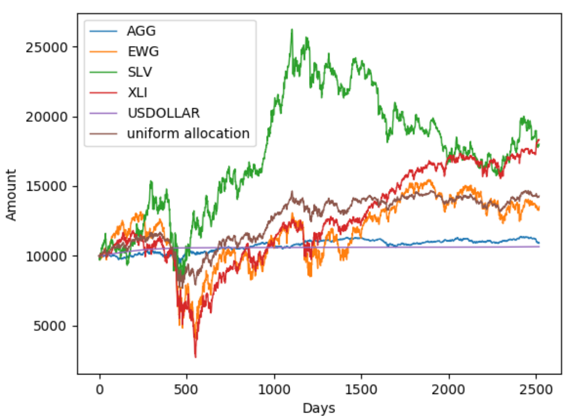

Figure 9. Plots showing the investment values over a period of time. Each of the assets curve [e.g. AGG, XLI, etc] shows the variation in investment values if the total budget is invest in that asset only. Uniform allocation curve shows the investment value when budget is invested equally in all the assets.

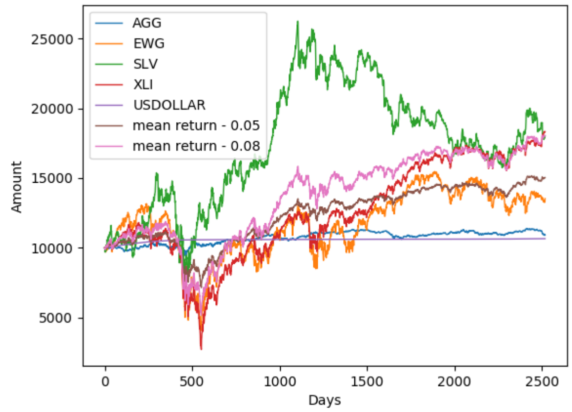

Figure 10. Plots of investment values over a period of time. Each of assets denoted the same as that of fig 9. Mean return curve denotes the investment values of optimum portfolio at different return values.

Figure 10 is a **little interesting**. It tells us that in some parts of graph optimum portfolio value is less than the portfolio value that we have if we have invested all our budget in asset 'SLV' only. This is because we have kept fractional weights $w_t = w$ whole time. If we keep changing w_t each trading day then we would have got same or higher returns than asset 'SLV' only. But that is not the ideal thing to do because investors doesn't change their investment on each trading days.

8. Conclusion

From the study of our different formulations to solve portfolio selection problem, we conclude that formulation 3 is the most efficient way to solve this problem.

References

Boyd, S. Portfolio optimization, 2017. URL http://stanford.edu/class/ee103/lectures/portfolio_slides.pdf.

Cornuejols, G. and Ttnc, R. (eds.). *Optimization Methods in Finance*. Carnegie Mellon University, Pittsburgh, PA 15213 USA, 2006.

Sharpe, W. F. The sharpe ratio. URL https://web.stanford.edu/~wfsharpe/art/sr/sr.htm.

Stanford. Data for finance and portfolio optimization. URL http://stanford.edu/class/ee103/portfolio.html.