Mathematical Modeling for Portfolio Optimization

Yun Mai

Advisor: Haiyuan Wang

City University of New York

CUNY School of Professional Studies

yun.mai@spsmail.cuny.edu

Abstract

Portfolio optimization is to find the best portfolio out of a set of assets distribution under certain objective. The objective usually is to maximize the return for a given risk level or to minimize the risk for a expected return. In this report, four mathematical models were constructed to solve two different portfolio optimization problems with expected return under certain constraints.

Keywords: mean variance, minimax, linear programming

Introduction

Optimization is frequently used to model the real world decision making problems. It is also called mathematical programming and its solution is obtained through finding the configuration where a function is minimized or maximized by systematically choosing the values of variables within certain constraints.

Portfolio optimization is to choose a combination of weight for each security in the bucket. This problem is an important problem in the finance theory and has been intensively studied. The traditional configuration of portfolio optimization is to find the investment plan by maximizing the rate of return under a given risk tolerance or minimize the risk while securing the required rate of return level.

The problem

A portfolio manager wants to construct an optimal portfolio for a customer. The manager will do the analysis based on historical returns. The customer only want to invest on Exchange-Traded Fund. So the manager select 52 popular at New York Stock Exchange to begin with. The US central bank return is also selected as a option of low return and low risk. He will not consider all of these ETFs to be included in the portfolio. So, he only choose the ETFs with more than 12% average yearly return in the past 5 years. There are 12 ETFs are chosen: **DFE, DIA, IVV, KBE, KRE, SPY, VTI, XLF, XLI, XLK, XLV, XLY.**

In this report, I will tried four methods to experiment with the dataset and discover things as I went along.

Explore the Data

The historical price of ETFs were obtained from Yahoo Finance and Investing.com. The first three methods uses the adjusted price of the ETFs in the past five years as initial data. The fourth method uses the year-to-date return (YTD return) as initial data. YTD is the amount of profit generated by an investment since the beginning of the current calendar year, commonly used by investors and analysts in the assessment of portfolio performances because of their simplicity.

From Figure 1, we can see that the adjusted price of all the ETFs except KBE in the past five years have not changed a lot. Figure 2 shows the monthly return of ETFs in the past five years. We can see that the monthly return fluctuate around 0 and some big fluctuation happened around 2016 ~ 2017. The Figure 3 shows the difference between YTD and average yearly return based on the past five years adjusted close price.

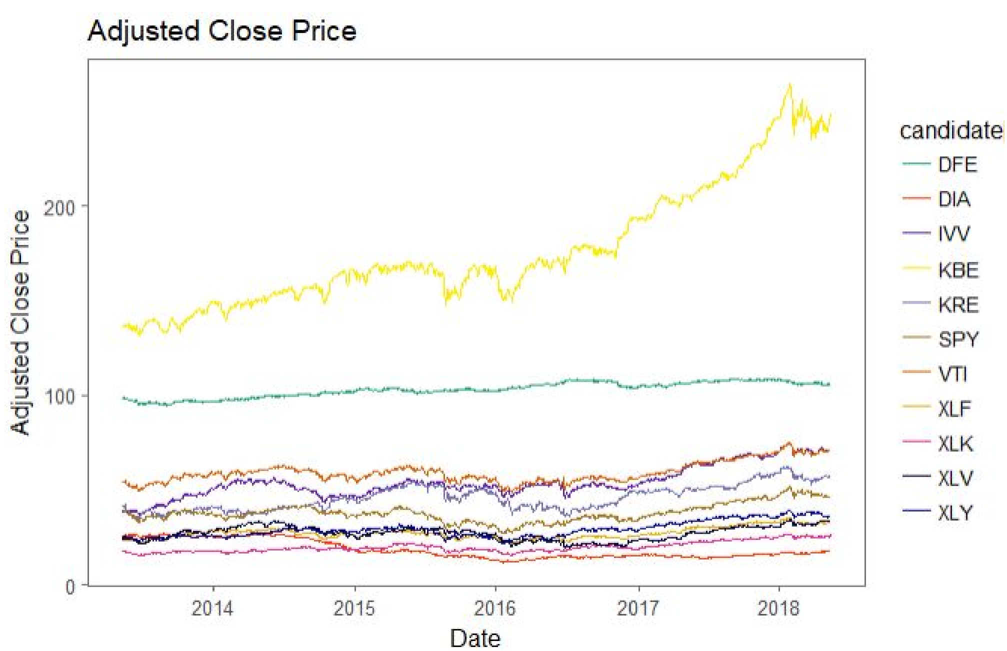

Figure 1. The adjusted close price of the ETFs in the past 5-years.

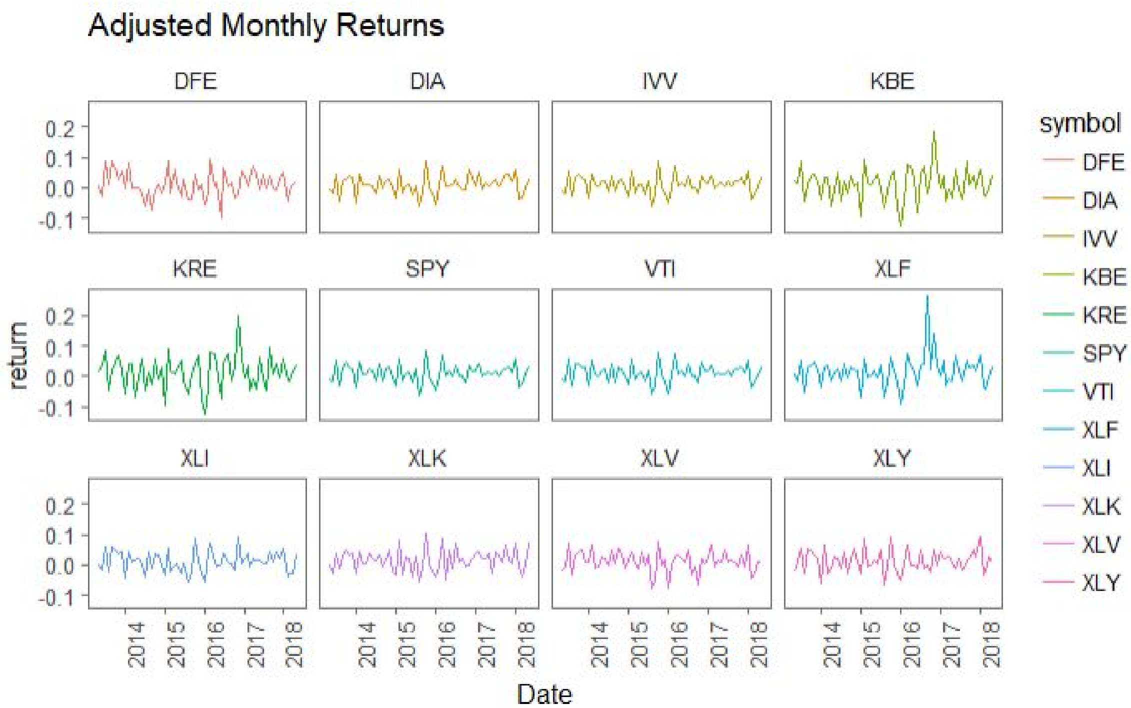

Figure 2. The monthly return of ETFs in the past five years.

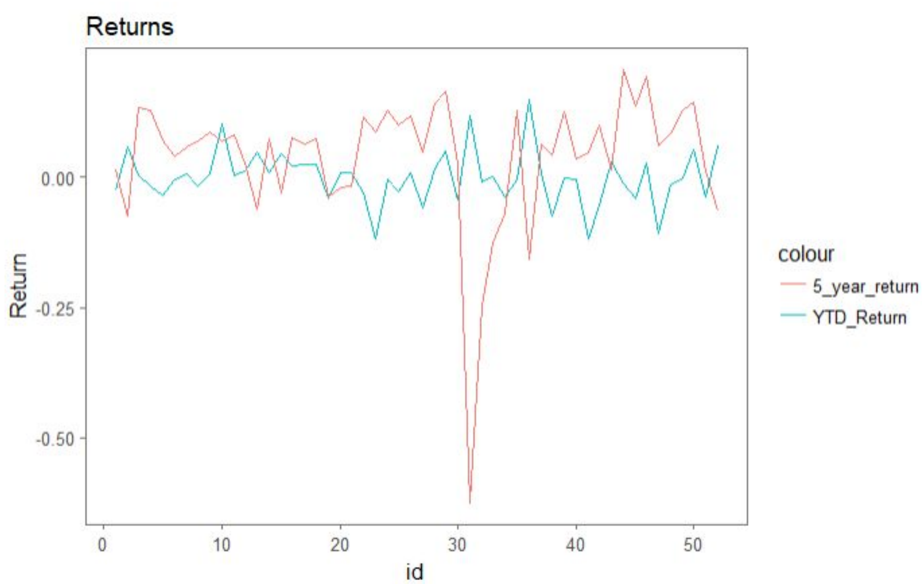

Figure 3. The difference between year-to-date return and the average annual return for the past 5 years.

Method

The yearly return is calculated as follows:

$$y_{jt} = (P_{j,t+1} - P_{jt}) / P_{jt}.$$

P_{jt} is the price of asset j at day t, $P_{j,t+1}$ is the price of asset j at day t+1.

The monthly or yearly return is calculated as:

monthly %=(P_{jt}+1)*($P_{j,t+1}$+1)....the last day

yearly %=(P_{jt}+1)*($P_{j,t+1}$+1)....the last day

Or

yearly %= [(return_year_1 + 1) (return_year_2 + 1)... (return_year_n + 1)] ^1/n

Game Theory

First, I will use the game theory to build a model on these 12 ETFs by considering the investor playing against the state of the economy.The best and the worst average yearly return is derived from adding or subtract the standard deviation from the average yearly return, as shown in the following table.

	worst <dbl>	average <dbl>	best <dbl>
DFE	-0.01468432	0.012497980	0.03968028
DIA	-0.01283273	0.015418400	0.04366953
IVV	-0.01198667	0.012256929	0.03650053
KBE	-0.02588422	0.013832204	0.05354863
KRE	-0.02725780	0.014741812	0.05674143
SPY	-0.01191148	0.012281038	0.03647356
VTI	-0.01063364	0.012198345	0.03503033
XLF	-0.01859376	0.016003434	0.05060062
XLI	-0.01781980	0.011018654	0.03985710
XLK	-0.01398581	0.021170469	0.05632675
XLV	-0.01979990	0.008889583	0.03757907
XLY	-0.02011010	0.013301258	0.04671262

The problem is as follows:

Suppose our investor has $1 to allocate among the 12 assets with the unknown amounts y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12} respectively. That is,

$$y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 + y_{10} + y_{11} + y_{12} = 1$$

Then we can view y as weight. If the smallest return is R , the returns should be:

$$-0.0295y_1 - 0.0387y_2 - 0.0334y_3 - 0.0255y_4 - 0.0143y_5 - 0.0332y_6 - 0.0304y_7$$

$$- 0.0399y_8 - 0.043y_9 - 0.0281y_{10} - 0.0406y_{11} - 0.0363y_{12} \geq R \quad \text{\{if worst return\}}$$

$$0.0044y_1 + 0.0027y_2 + 0.0057y_3 + 0.0111y_4 + 0.0172y_5 + 0.0059y_6 + 0.0061y_7$$

$$+ 0.004y_8 - 0.001y_9 + 0.0194y_{10} + 0.0023y_{11} + 0.0138y_{12} \geq R \quad \text{\{if average return\}}$$

$$0.0384y_1 + 0.0441y_2 + 0.0447y_3 + 0.0476y_4 + 0.0486y_5 + 0.0449y_6 + 0.0427y_7$$

$$+ 0.0478y_8 + 0.0409y_9 + 0.0669y_{10} + 0.0451y_{11} + 0.064y_{12} \geq R \quad \text{\{if best return\}}$$

$$y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12} \geq 0$$

Solving this problem by Linear Programming solution algorithm, the optimal solution is $y_1 = 0$, $y_2 = 0$,$y_3 = 0$, $y_4 = 0$,$y_5 = 0$, $y_6 = 0$,$y_7 = 0$, $y_8 = 0$, $y_9 = 0$, $y_{10} = 0$,$y_{11} = 1$, $y_{11} = 0$, $y_{12} = 0$, and $R = 0.1941$. That is, the investor must put all the money in the XLV account with the accumulated return rate at 2.117%.

candidate	Weight
DFE	0
DIA	0
IVV	0
KBE	0
KRE	0
SPY	0
VTI	0
XLF	0
XLI	0
XLK	1
XLV	0
XLY	0

The result is not ideal as it has a maximized return but at the same time also give a maximum risk. Putting all money into one ETF is too risky. This method do not necessary lead to a efficient portfolio as it formulate the optimal portfolio either to minimize the risk given some minimum return or maximize return given a maximum risk.

Mean Variance

The mean-variance model of Markowitz is proposed to choose the portfolio which can achieve a specified mean return with the minimum risk[1].

Suppose data are observed for N securities, over T time periods. Let

y_{jt} = Return on one dollar invested in security j in time period t.

$\bar{y}_j$ =Average Return on security j

w_j = Portfolio allocation to security j.

y_{pt} = Return on portfolio in time period t

E_p =Average Return on portfolio

M_p = Minimum return on portfolio

The problem is expressed as follows:

The objective function:

$$min \sum_{j=k}^{N} \sum_{j=1}^{N} w_j w_k s_{jk}$$

subject to:

$$\sum_{j=1}^{N} w_j \bar{y}_j \geq G$$

With:

$$s_{jk} = \frac{1}{T-N} \sum_{t=1}^{T} (y_{jt} - \bar{y}_j)(y_{kt} - \bar{y}_k)$$

The average monthly return of the portofolio at the evenly distributed allocation is 2.506%. After optimization, the average monthly return of the portfolio is 1.273% when the global variance is at minimum 0.0231. The weight of each ETF is shown in the table as follows:

	Weight <dbl>	Ave.Return <dbl>	Stdev <dbl>	Sharp <dbl>
DFE	0.3269	0.012497980	0.02718230	0.4597837
DIA	0.1152	0.015418400	0.02825113	0.5457623
IVV	0.0358	0.012256929	0.02424360	0.5055738
KBE	0.0055	0.013832204	0.03971643	0.3482741
KRE	0.0461	0.014741812	0.04199962	0.3509987
SPY	0.2576	0.012281038	0.02419252	0.5076379
VTI	0.1608	0.012198345	0.02283198	0.5342656
XLF	0.0002	0.016003434	0.03459719	0.4625646
XLI	0.0107	0.011018654	0.02883845	0.3820820
XLK	0.0001	0.021170469	0.03515628	0.6021817
XLV	0.0272	0.008889583	0.02868948	0.3098551
XLY	0.0139	0.013301258	0.03341136	0.3981059

It make sense that the results show that the securities with higher average return ,such as XLK, get more weight and those with lower average return get less weight, such as XLV.

Minimax model

The Minimax model is to minimize the maximum losses with a restriction on the minimum acceptable average return over a historical period.

The problem is expressed as follows:

The objective function:

$$min \sum_{j-k}^{N} \sum_{j-1}^{N} w_j w_k s_{jk}$$

subject to:

$$\sum_{j-1}^{N} w_j \bar{y}_j \geq G$$

With:

$$s_{jk} = \frac{1}{T-N}\sum_{t=1}^{T}(y_{jt} - \bar{y_j})(y_{kt} - \bar{y_k})$$

The average monthly return of the portofolio at the evenly distributed allocation is 1.712%. After optimization, the minimum average losses f the portfolio is 1.422% when the variance is at 0.0250578. The weight of each ETF is shown in the table as follows:

	Weight <dbl>	Worst.Return <dbl>	Ave.Return <dbl>	Stdev <dbl>	Sharp <dbl>
DFE	0.0995	-0.01468432	0.012497980	0.02718230	0.4597837
DIA	0.1453	-0.01283273	0.015418400	0.02825113	0.5457623
IVV	0.1463	-0.01198667	0.012256929	0.02424360	0.5055738
KBE	0.0024	-0.02588422	0.013832204	0.03971643	0.3482741
KRE	0.0008	-0.02725780	0.014741812	0.04199962	0.3509987
SPY	0.0022	-0.01191148	0.012281038	0.02419252	0.5076379
VTI	0.2654	-0.01063364	0.012198345	0.02283198	0.5342656
XLF	0.0803	-0.01859376	0.016003434	0.03459719	0.4625646
XLI	0.1117	-0.01781980	0.011018654	0.02883845	0.3820820
XLK	0.0297	-0.01398581	0.021170469	0.03515628	0.6021817
XLV	0.0907	-0.01979990	0.008889583	0.02868948	0.3098551
XLY	0.0257	-0.02011010	0.013301258	0.03341136	0.3981059

It make sense that the results show that the securities with higher maximum average losses get more less weight, such as KBE and those with lower average return get less weight, such as SPY.

To see the difference return of the portfolio between the Mean Variance and MiniMax, the return over the time is compared in Figure 4. The returns of these two methods are different . Mean Variance is more stable while MiniMax allows more risk and thus has higher rate of return (1.422 vs. 1.273%).

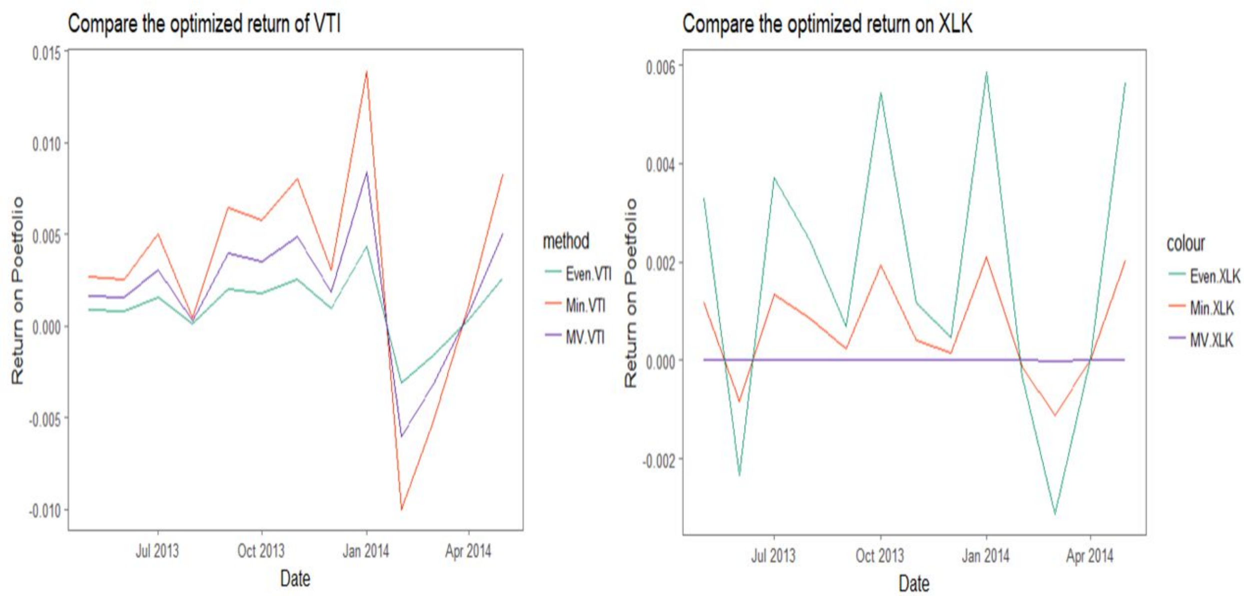

Figure 4. Compare the return of Mean Variance and MiniMax. Even: the weight for each ETF is the same. MV: Mean Variance. Min: MiniMax.

Linear Programming

Actually linear programming has been used in the above three methods to solve the problem. Instead of considering variance as risk, the fourth method will calculate the failure rate or risk using beta value of the securities. The data in the fourth method is different from the previous three ones. The year-to-day return and the beta value were obtained from Yahoo Finance instead of calculated from the past 5 years adjusted close price.

First, the ETFs with positive annual return were selected. 26 ETFs were selected: **DBC, DFE, EFA, EWH, EWI, EWT, EWU, EWW, EWY, EWZ, EZU, FEZ, FXI, GLD, IAU, IWM, KBE, KRE, OIL, SH, USO, VGK, XLE, XLK, XLY, XOP**.

To calculate the risk, some assumptions and known conditions are needed:

1) Assuming there is only two outcomes for each investment, success and failure.
2) Typically, market risks for individual stocks range from 0.5 to 2.5 or 3.0[2].
3) Assuming 2% is a point due to the minimal, yet apparent, system risk.
4) Assuming the general risk for investing stock range from 10% to 50%.
5) If there is a linear relationship between beta value and risk, we can estimate risk based on the beta value.
6) A negative beta indicates an inverse relation to the market.Some stock like gold and gold stocks should have negative betas because they tended to do better when the stock market declines. The smaller the negative beta value, the more volatility. As such, negative beta should be converted to positive when calculate the risk as risk can not be negative.

The first order linear relationship between failure rate and beta value is tested and there is no investment will generate the positive return. Then the quadratic relationship between failure rate and beta value is tested and there are two investment will bring the positive return. The failure rate will be used in the optimization computation.

We will select these two stocks, EWI and USO, with expected return as 2.88% and 3.28% respectively.

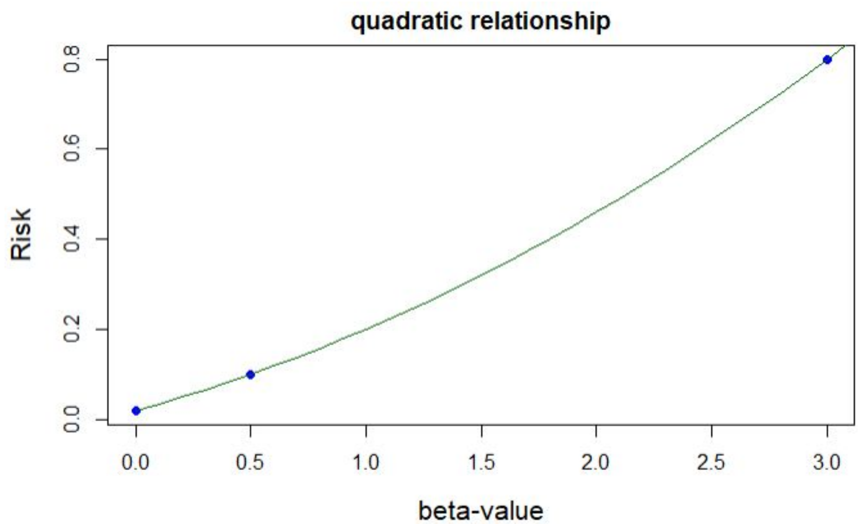

Figure 4. The quadratic relationship between risk and beta value.

To diversify the portfolio, we can invest bonds. U.S. savings bonds. U.S. savings bonds are endorsed by the federal government, they are considered risk free and considered one of the safest types of investments. Suppose we purchase Series I Bonds, the current interest rate is 2.52%. The expected return for Series EE Bonds is: 100%*2.52% = 2.52%. The rate of return, risk and expected return are summarized as the following table:

	Investment <chr>	Rate of Return <dbl>	Beta <dbl>	Risk <dbl>	Expected Return <dbl>
10	EWI	0.1028	1.16	0.06709306	0.02880978
36	USO	0.1490	1.90	0.10112802	0.03280391
3	Series I Bonds	0.0252	NA	0.00000000	0.02520000

Suppose the manager have $50,000 and wish to invest. If we invest x, y and z amount of money in EWI, USO and Series I Bonds respectively, the expected net gain will be $0.0288x+0.0328y+0.0252z$.

Assuming there is only two outcomes for each investment, success and failure, there are eight different scenarios that may occur with our investments. The expected net gain and likelihood for each situation to occur is summarized in the table below (A,B and C represents EWI, USO and Series I Bonds respectively).

Case <fctr>	A <fctr>	B <fctr>	C <fctr>	Probability <dbl>	Net Gain <fctr>
1	S	S	S	0.8386	0.1028x+0.1490y+0.0252z
2	F	S	S	0.0603	-x+0.1490y+0.0252z
3	S	F	S	0.0943	0.1028x-y+0.0252z
4	S	S	F	0.0000	0.1028x+0.1490y-z
5	F	F	S	0.0068	-x-y+0.0252z
6	F	S	F	0.0000	-x+0.1490y-z
7	S	F	F	0.0000	0.1028x-y-z
8	F	F	F	0.0000	-x-y-z

If we hope the gain being greater than or equal to 90%, there are three different combinations: 1+2+3,1+2+5,1+3.

Optimize the portfolio by linear programing. The objective function is:

$$\text{Max } E(G) = 0.0288x+0.0328y+0.0252z$$

subjected to:

$$x, y, z > 0$$

$$x + y + z <= 50000$$

$$0.1028x+0.1490y+0.0252z >=0$$

$$-x+0.1490y+0.0252z >= 0$$

$$0.1028x-y+0.0252z >= 0$$

or we solve this:

$$\text{Max } E(G) = 0.0288x+0.0328y+0.0252z$$

subjected to:

$$x, y, z > 0$$

$$x + y + z <= 50000$$

$$0.1028x+0.1490y+0.0252z >=0$$

$$-x+0.1490y+0.0252z >= 0$$

$$-x-y+0.0252z >= 0$$

or we solve this:

$$\text{Max } E(G) = 0.0288x+0.0328y+0.0252z$$

subjected to:

$$x, y, z >0$$

$$x + y + z <= 50000$$

$$0.1028x+0.1490y+0.0252z >=0$$

$$0.1028x-y+0.0252z >= 0$$

For the first scenario, we got the maximum net gain $1275.164 by put $1390.149 in EWI, $1334.252 USO in and $47275.599 in Series I Bonds. For the second scenario, we got the maximum net gain $1269.345 by put $0 in EWI, $1229.028 USO in and $48770.972 in Series I Bonds. For the third scenario, we got the maximum net gain $1459.105 by put $45339.137 in EWI, $4660.863 USO in and $0 in Series I Bonds. The results for the three scenarios are summarized in the following table:

EWI <fctr>	USO <fctr>	Series.I.Bonds <fctr>	Net.Gain <fctr>
1390.149	1334.252	47275.60	1275.164
0.000	1229.028	48770.97	1269.345
45339.137	4660.863	0.00	1459.105

So the third scenario was the winner.

Conclusion

Game theory model where the investor plays against the state of economy will not necessary lead to efficient portfolio management.

Mean Variance model aim for safer investment plan while MiniMax aim for relative higher return for the investors who can tolerant relative higher risk.

When the size of the portfolio is small, enumerate the combination of successful and failure status could be used in building the linear programming for the optimization problem.

Discussion

The constraints in the linear programming in the fourth model can be changed, such as the required rate of positive return/ risk tolerance of the investor, will change the distribution of the money.

The time period of the historical data will change the input of the model and give a different result. According to Figure 3, we will get more ETFs with positive annual return for model-4 if we use the 5-years data. Similarly, the failure rate calculated from different period of the historical data will be different, thus produce different coefficients of the model.

Each model is built upon different assumptions and accounts for those assumptions in different manners. Taking the carefully analysis of historical data will guide people to decide how to best diversify the portfolio. This report built four two simple models and two complex models. Only consider risk or return will not necessary lead to best solution. Complex methods to find the best trade-off between risk and return, like efficient front theory, need to be studied in the future learning of the modeling for portfolio optimization problem.

References

1. Markowitz H (1952). Portfolio selection, *Journal of Finance*, 7 77-91.

2.YOUNG, M. R. (1998). A Minimax Portfolio Selection Rule with Linear Programming Solution, Management Science, v. 44, p. 673-683.

Appendix

R code:

https://raw.githubusercontent.com/YunMai-SPS/DATA609_homework/master/DATA609_finalproject/YunMai_DATA609_final_project_v2.Rmd

or

https://github.com/YunMai-SPS/DATA609_homework/blob/master/DATA609_finalproject/YunMai_DATA609_final_project_v2.md