Portfolio Construction using Machine Learning: Comparison between Monte Carlo Method and LASSO Algorithm

U Kai, Wong

Jinan University, No.601, West Huangpu Avenue, Guangzhou, Guangdong, China

Abstract. Machine Learning are widely used in manifold fields, in which finance is one of the most recurring fields to be used. One of a popular study in finance is Portfolio construction. This essay is attempting to constructing portfolio using Markowitz Portfolio Theory, Monte Carlo Method and Least Absolute Shrinkage and Selection Operator (LASSO) algorithm and comparing the simulation result. We will use the daily data of components of NASDAQ during 2020 to simulate.

Keywords: Machine Learning; Finance; Portfolio Construction.

1. Introduction

Portfolio is composed of holding of serval securities which may be positive which means a long position in security, negative which means a short position in security, or zero. We had to select several objects in given set of securities and assets for constructing the portfolio. Because of the covariance between each individual security in portfolio, the risk of the portfolio can be lower than the risk of individual security. That is also the reason why we rather to diversify our investment. The well-known approach of diversify investments is Markowitz Portfolio Theory (Markowitz, 1952), which can figure out the most efficient portfolio component in the given set, which have the highest return in unit risk. This theory gave a quantitative approach for portfolio selection. Based on such quantitative approach, we can also apply portfolio selection by machine learning method. In this essay we will adopt Monte Carlo Method and LASSO algorithm to comply the portfolio selection.

The concept of Monte Carlo Method (Raychaudhuri, 2008) is using plenty simulation result to ascertain the charecterics of the model and use the ascertained charecterics above to estimate the optimatize result. On the other hand, Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani R, 1996) algorithm is a popular tool in the field of machine learning method in recent years which is an regularisation method with panelty factor. Such method added regularisation penalties in portfolio to overcome the presence of multicollinearity of the data and avoid wrong estimations of the portfolio risk.

The task of this essay is attempting to construction a portfolio using Markowitz Portfolio Theory, Monte Carlo Method and LASSO algorithm based on daily data of components in NASDAQ during 2020 (Data ref. Yahoo Finance API) and comparing the performance of the constructed portfolio.

2. Methodology and Data Choosing

2.1 Markowitz Portfolio Theory

Now we consider there are 2 securities, we can find out the tangency portfolio graphically. First, we find the mean-variance frontier portfolio of two stocks. Then, we find out the tangent line (Capital Market Line, CML) to mean-variance fronter portfolio at risk-free rate. We can get the most efficient portfolio (Tangency Portfolio) which is the tangent points of the mean-variance fronter portfolio. We can expand such concepts to N securities case of portfolio selection.

Suppose there are N securities, R_p is expected return of the portfolio, and Var_p is the risk of portfolio. We can determine R_p and Var_p as following:

$$R_p = w^T r$$

$$Var_p = w^T \Sigma w$$

Where w is the non-negative weight vector of components in portfolio which $w^T 1_n = 1$, r is the return vector of components in portfolio, and Σ is the covariance matrix of components in portfolio. For portfolio selection, we are supposed to find the maxima of R_p and minima of Var_p in constraint $w^T 1_n = 1$.

To measure the efficiency of the unit risk. We need to compare the Sharpe Ratio:

$$SR = \frac{R_p - r_f}{\sqrt{Var_p}} = \frac{w^T r - r_f}{\sqrt{w^T \Sigma w}}$$

Sharpe Ratio is a proportion of the return and volatility. The largest of Sharpe Ratio, the more efficient of unit risk of the constructed portfolio.

However, such concept also has some insufficient part in applications. First, the risk can be mitigated by diversification is specific risk, market risk cannot be mitigated by diversification. Furthermore, Markowitz Portfolio Theory has not considered the multicollinearity between the components of portfolio. It will lead to overfitting. Such result has large volatility. In the other words, the risk of portfolio will be extraordinary high. To prevent overfitting, we need to compute the further statistical tests including autocorrelation, endogeneity, and multicollinearity during using Markowitz Portfolio Theory.

Specifically, portfolio construction is a optimization of a constrained mathematics model. To solve such Mathematical optimization problem of Portfolio construction, we can also employ machine learning method such as Monte Carlo Method or LASSO algorithm etc. In this essay we adopt Monte Carlo Method and LASSO algorithm.

2.2 Monte Carlo Method

Monte Carlo Method is a computational research method which is generating plenty of simulation result using a variety of random number. In this essay, we use Monte Carlo Method to estimate the optimized portfolio weight vector. The procedure of Monte Carlo Method is shown as the following:

Calculate the log return and covariance matrix of the components of portfolio.

Generate the weight vector using random number and calculate the corresponding expected return and volatility.

Iterate step 2. We can get a plenty of weight vector and the expected return and volatility.

Calculate the Sharpe ratio using the expected return and volatility and find the maximum Sharpe ratio in the iterated results.

Such method is a trial-and-error process which calculated by computer, so the sampling is time-consuming if the model is complicated. Yet, the quantity of the components of portfolio, the more time of the model simulation and the quality of the result. In this essay, we will iterate the model 20000 time and sketch in the graph. We will discover that the approximate characteristics of the mean-variance fronter portfolio.

2.3 LASSO Algorithm

Assume we have N assets in portfolio. To estimate the optimal weight of the assets in portfolio, we need to solve the following minimum variance optimization problem:

$$w^* = \arg\min_{w \in \mathbb{R}^N} \{w^T \hat{\Sigma} w\}$$
$$s.t.\ 1_N^T w = 1$$

Where w^* is optimal weight vector, w is assets weights and $\hat{\Sigma}$ is estimated covariance matrix. Solving the problem above we can have the solutions:

$$w^* = \left(1_N^T \hat{\Sigma}^{-1} 1_N\right)^{-1} \hat{\Sigma}^{-1} 1_N$$

Such optimized weights are highly sensitive to estimation errors and to existence of multicollinearity in the inputs. We can regularize the minimum variance problem as following:

$$w^* = \arg\min_{w \in \mathbb{R}^N} \{w^T \hat{\Sigma} w + \lambda \sum_{i=1}^{N} \rho(w_i)\}$$
$$s.t.\, 1_N^T w = 1$$

Where λ controls the impact of the summation of l_1-norm: $\sum_{i=1}^{N} \rho(w_i)$, the intensity of the shrinkage applied to the weights vector and the level of sparsity. When $\lambda = 0$, the problem will become an OLS regression problem. Such penalty factor can mitigate the overfitting due to multicollinearity. On the other words, it can mitigate the volatility of the constructed portfolio. Notwithstanding it can mitigate the overfitting, the penalty factor can lead to the lower return of the portfolio.

3. Date

The stock chosen in this essay is the stock in the components of NASDAQ which include among 100 stocks (Data ref. Yahoo Finance API). We choose the stock which have the most return to construct the portfolio. The return and the volatility of the chosen stock is shown in Table. 1.

Table 1. The return and the volatility of the chosen stock in NASDAQ

Stocks	Return	Variance	Standard Deviation
TSLA	208.03%	80.41%	89.67%
MRNA	174.72%	97.49%	98.74%
ZM	163.10%	68.26%	82.62%
PDD	146.30%	44.41%	66.64%
CRWD	144.20%	42.40%	65.11%
ZS	144.19%	45.58%	67.52%
DOCU	108.63%	35.98%	59.98%
MELI	102.81%	33.86%	58.19%
DDOG	99.15%	43.03%	65.60%
OKTA	79.47%	25.97%	50.96%
NVDA	78.32%	33.87%	58.20%
PYPL	73.44%	27.24%	52.20%

The range of the data is forming the start of 2020 to the end of 2020. There are 252 trading days in NASDAQ during 2020. We will find out the weight of the portfolio and comparing the Sharpe ratio of both portfolios to compare the performance of these two approaches. The risk-free rate is chosen the average yield of 1yr US treasure bill (0.37%). In this essay, we use excel solver to solve the optimal portfolio with highest Sharpe ratio and compare with the result of Monte Carlo Method and LASSO simulated by python. We also compare the Sharpe Ratio of the constructed portfolio to estimate the efficiency of unit risk.

4. Evaluation and Analysis

According to the result of using excel solver and python, we obtain weight of the optimized portfolio in Table. 2.

Table 2. The simulated weight of the optimized portfolio

Stocks	Weight by Excel Solver	Weight by Monte Carlo Method	Weight by LASSO Algorithm
TSLA	0.2203	0.1974	0
MRNA	0.1833	0.1538	0.1247
ZM	0.1740	0.1251	0.1131
PDD	0.2556	0.1859	0.1501
CRWD	0.1027	0.0446	0
ZS	0.0641	0.0605	0
DOCU	0	0.0712	0
MELI	0	0.0216	0.0426
DDOG	0	0.0330	0.0542
OKTA	0	0.0143	0.1735
NVDA	0	0.0616	0
PYPL	0	0.0308	0.3418

According to the simulation result in Table 2, we discover that the weight of TSLA is zero in LASSO algorithm, which has the largest return. It because the LASSO algorithm has a penalty term that will mitigate the risk but sacrifice the expected return. Furthermore, since the target of LASSO algorithm is minimum the risk, the maximum of the return is not the consideration of portfolio construction. On the contrary, the weight of TSLA is biggest in the portfolio. It because Markowitz Portfolio Theory does not consider the effect of multicollinearity, the overfitting of the constructed portfolio needs to be checked by further statistical tests.

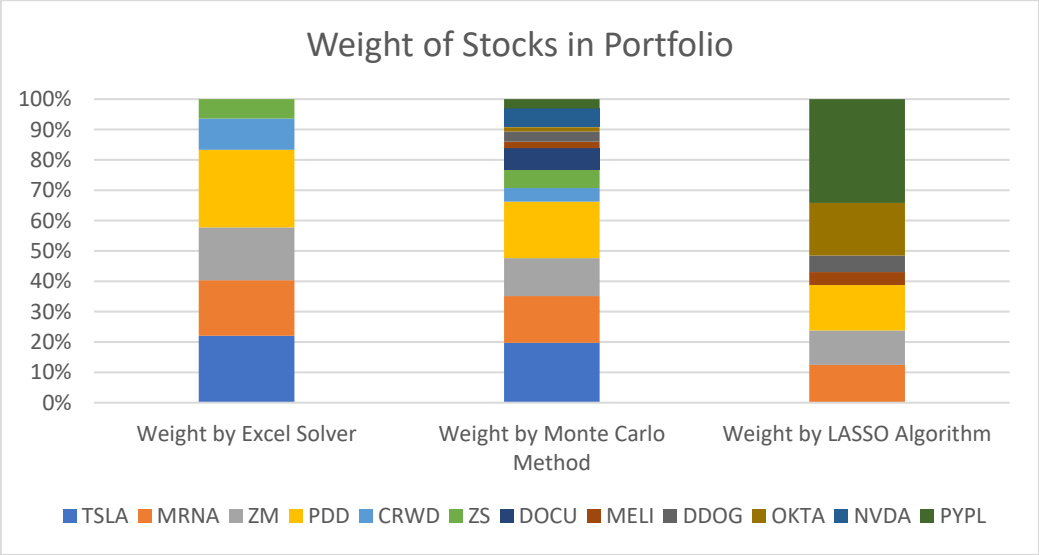


Figure 1. Weight of stocks in portfolio

Table 3. The simulated expected returns, standard error, and Sharpe Ratio

	Excel Solver	Monte Carlo Method	LASSO Algorithm
Expected Return	167.68%	153.06%	111.64%
Volatility	47.76%	45.12%	40.03%
Sharpe Ratio	3.501	3.382	2.777
Var	22.81%	20.36%	16.02%
Risk Free Rate	0.47%	0.47%	0.47%

The corresponding expected returns, standard error, and Sharpe Ratio are shown in Table. 3. According to the simulation result of expected returns, standard error, and Sharpe Ratio in Table. 3, we can discover that the result by Monte Carlo Method could approach the result if we choose the sampling time is large enough. However, the simulation will be time-consuming if the sampling time

is too long. In this essay we have chosen the sampling time n = 20000. On the other hand, the result by LASSO algorithm has the smallest variance. Notwithstanding that the result is not the most efficient one.

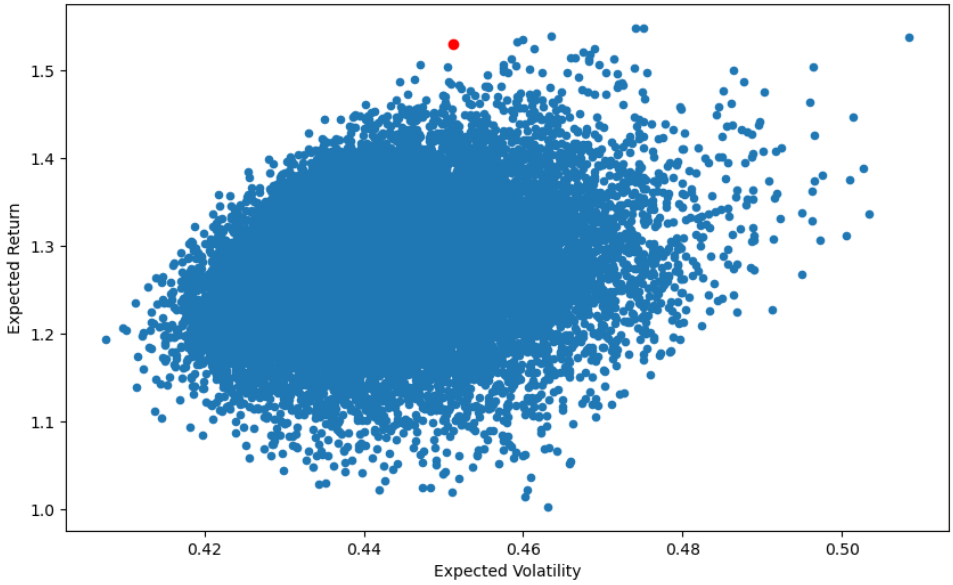


Figure 2. The result of the Monte Carlo Method. The red point represents the simulate result which has maximum Sharpe ratio

The sampling result has been demonstrated in Fig. 2. We can recover that the scatters points shaped the mean-variance fronter portfolio approximately. The red point represents the portfolio which has the largest return in unit risk. On the other words, the portfolios which has largest Sharpe Ratio in the simulation result.

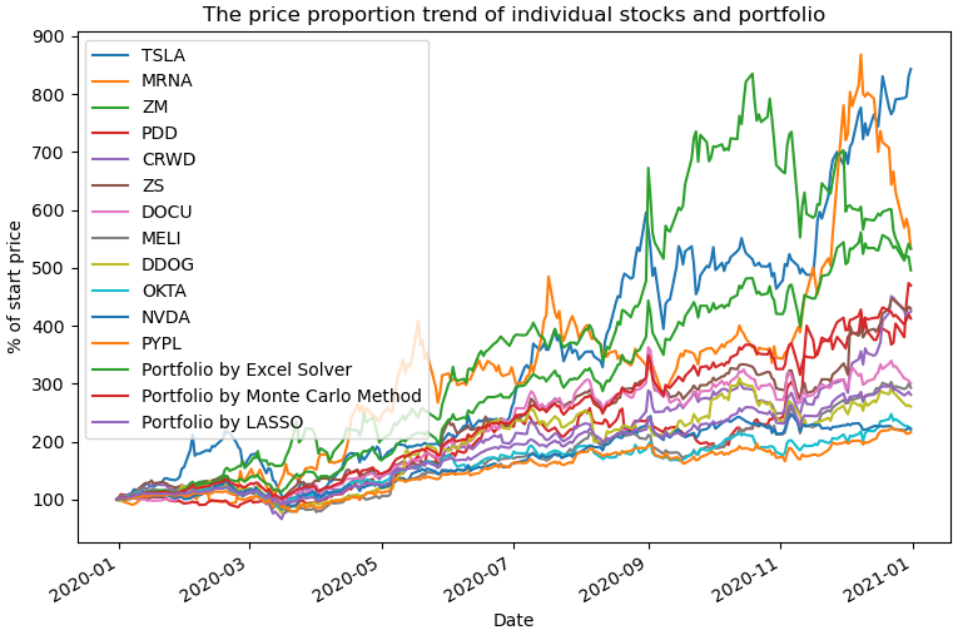


Figure 3. The price trend of individual stocks and constructed portfolio

According to the price trend of the price proportion trend of individual stocks and portfolio chart in Fig. 3, we can discover that the trend of the constructed portfolio is increase steady compare with the high volatility stocks like TSLA or ZM etc. Moreover, the return of the constructed portfolio is higher than the low volatility stock like OKTA or DDOG etc.

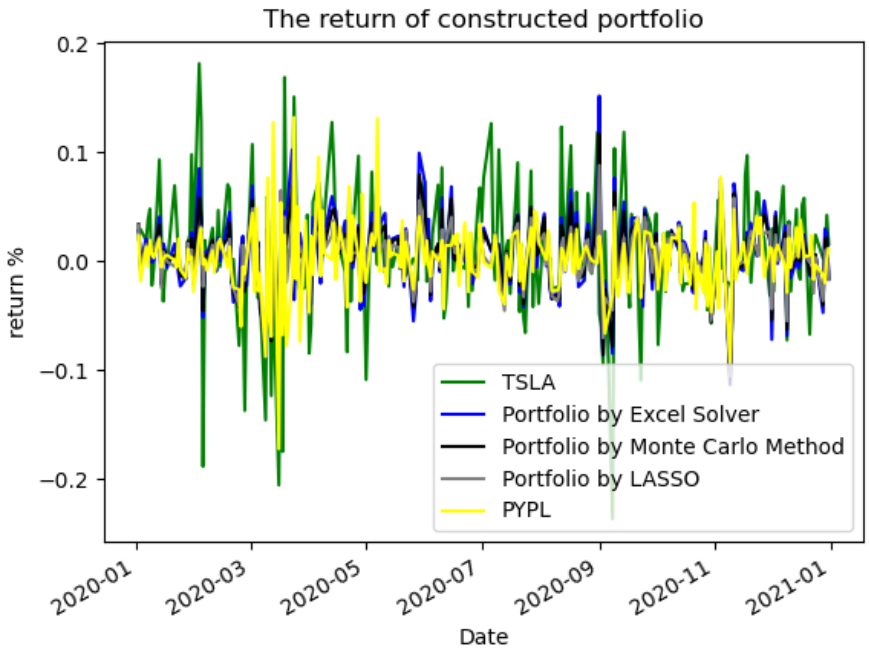


Figure 4. The return of Constructed portfolio

The return of the constructed portfolios is demonstrated in Fig. 4. The stocks with maximum return (TSLA) and minimum volatility (PYPL) are also demonstrated in the chart for reference. It is obvious that the volatility of constructed portfolios is smaller than TSLA. It can be inferred that the volatility of constructed portfolios is mitigated.

5. Conclusion

This essay compares the Markowitz Portfolio Theory, Monte Carlo Method, and LASSO algorithm. According to the estimation, we discover that such algorithm can mitigate the risk when we construct the portfolio. However, the Markowitz Portfolio do not consider the multicollinearity that we need to compute the further statistical tests to prevent the overfitting. Monte Carlo Method is generating plenty of result by random number to approach the optimize solution. Yet, we need to consider the sampling time when modeling a complicated model. LASSO algorithm can mitigate the overfitting due to multicollinearity because of the penalty l_1-norm factor. Notwithstanding such penalty factor LASSO algorithm have, the return of the constructed portfolio is lower than the other two approach. However, since the target of LASSO algorithm is minimized the covariance matrix rather than maximized the return, the result has small variance and volatility. Despite the different characteristics of the portfolio construction approach mention in this essay, each of them can mitigate specific risk.

References

[1] Stephen F. LeRoy, Jan Werner. (2014) Principles of Financial Economics 2ed. Cambridge: Cambridge University Press.

[2] Bartram, S. M., Branke, J., & Motahari, M. (2020). Artificial intelligence in asset management. United States: CFA Institute Research Foundation.

[3] Harry Markowitz. (1952). Portfolio Selection. The Journal of Finance, Vol. 7, No. 1. (Mar., 1952), pp. 77 - 91.

[4] Dunis, C., Middleton, P., Karathanasopolous, A., & Theofilatos, K. (2016). Artificial intelligence in financial markets. London: Palgrave Macmillan.

[5] USTAOĞLU, E. (2022). PORTFOLIO OPTIMIZATION WITH GRAPHICAL LASSO AND AN APPLICATION IN BORSA ISTANBUL. Öneri Dergisi, 17(58), 760 - 765.

[6] Dunis, C., Middleton, P., Karathanasopolous, A., & Theofilatos, K. (2016). Artificial intelligence in financial markets. London: Palgrave Macmillan.

[7] Radlak, M (2008). Artificial Intelligence with Modern Portfolio Theory. MSc Thesis, The University of Birmingham, Business School, England.

[8] Volta, V. Portfolio Optimization during Financial Crisis: Markowitz, LASSO and RIDGE.

[9] Yuan, X., Yu, W., Yin, Z., & Wang, G. (2020). Improved Large Dynamic Covariance Matrix Estimation with Graphical Lasso and Its Application in Portfolio Selection. IEEE Access, 8, 189179 - 189188.

[10] Meyers, B. (2011). Principles of Corporate Finance, 10th editions, the MG-Grow- Hill companies.

[11] Awoye, O. A. (2016). Markowitz minimum variance portfolio optimization using new machine learning methods Doctoral dissertation, (UCL) University College London, England.

[12] Roques, F. A., Newbery, D. M., & Nuttall, W. J. (2008). Fuel mix diversification incentives in liberalized electricity markets: A Mean–Variance Portfolio theory approach. Energy Economics, 30 (4), 1831 - 1849.

[13] Koné, N. G. (2021). Efficient mean-variance portfolio selection by double regularization (No. 1453). Queen's Economics Department Working Paper.

[14] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58 (1), 267 - 288.

[15] Raychaudhuri, S. (2008, December). Introduction to monte carlo simulation. In 2008 Winter simulation conference (pp. 91-100). IEEE.