Testing the Weak Form of Efficient Market Hypothesis with Monthly Tangency Risky Portfolios

Muhammad Furrukh Asif, Yitong Wu, Hanwen Zhang

December 2019, Vassar College

1 Introduction

It all started with a casual conversation with our friends from the Vassar Finance Club. One thing we found really interesting is that, on one hand, you have tens of thousands of people, working day and night, to beat the market, while on the other hand, there is this reigning theory in the world of finance – the Efficient Market Hypothesis – stating that investors can only outperform the market by luck.

The Efficient Market Hypothesis states that the market has priced in all the available information, and consists of three variations in regard to the definition of this "available information." The weak form considers all available information as all technical information about a security – in other words, its prices, volumes, price trends, market capitalization, and etc. The semi-strong form adds fundamental information into the hypothesis, namely, information about the security issuers, and expands the definition into all publicly available information. The strong form goes to extremes by including insider information in the hypothesis.

Literature has primarily tested the Efficient Market Hypothesis using linear regression on investment returns. Our group took another approach by adopting Markowitz model, the model for portfolio optimization, and using generated monthly tangency risky portfolios from 2012 to 2017 to test the weak form of the hypothesis – investors cannot beat the market with only trading data on securities. The Markowitz model is essentially built on linear programming – with a polygon made up of all the possible risky portfolios, the line crossing the risk-free rate and tangent to the polygon would render the intersection point the portfolio with the largest Sharpe ratio – the maximum compensation for a given level of risk.

2 Choices of Stocks

Based on our personal preferences, we first selected a total number of 30 various stocks ranging from Target to Boeing. We did the first round of tests on their data from 2012 to

2019, and took out the stocks with incomplete data. We ended up with the following 14 stocks in our portfolio: American Airlines, Apple, CVS, Ford, Goldman Sachs, IBM, JP Morgan, Netflix, Pepsi, Target, Tesla, United Airlines, Walmart, and Verizon. From these 14 stocks, we would construct our optimal risky portfolios (portfolios consisting of only risky securities) for each month from 2012 to 2017.

3 Data

We used the website https://www.quandl.com/ to obtain the information we needed: date and adjusted closing price for each of our selected stocks. We used the quandl module in python in our code to do this work for us.

4 An Attempt with Linear Programming

The idea of doing linear programming came naturally. Before going into details, we could easily set up the following sketch to obtain a portfolio based on the information of a particular month:

Parameters:

(1) $\{p_1, p_2, \cdots, p_{14}\} \in P$, the monthly averages of the returns of our stocks

(2) $\{r_1, r_2, \cdots, r_{14}\} \in R$, the monthly averages of the risks (standard deviations) of our stocks

Variables:

(1) $w_1, w_2, \cdots, w_{14}$, weights of each stock in our portfolio

Constraints:

(1) Weights of all the stocks should add up to 1: $\sum_{i=1}^{14} w_i = 1$

(2) Weights should all be non-negative: $\forall i \in \{1, 2, \cdots, 14\}, w_i \geq 0$

Objective function: We will have it as a multi-objective version:

$$\text{minimize } c_1\left(-\sum_{i=1}^{14} w_i p_i\right) + c_2\left(\sum_{i=1}^{14} w_i r_i\right),$$

for some positive c_1, c_2 such that $c_1 + c_2 = 1$. We cannot have zero because we do want to take care of both the returns and risks.

Although this seems reasonable, the first challenge, among all, is how we should determine the values of c_1 and c_2. In fact, we were unable to justify our choice of c_1 and c_2 to make this linear programming convincing. When we tried to look up online, wondering if there was any default values people would use, we realized that the Modern Portfolio

Theory has been developed for a long time and there exists an established and generally accepted approach – the Markowitz Model, first introduced in Harry Markowitz's paper "Portfolio Selection" published in 1952. For the rest of our project, we will study this method and try to build the tangency portfolio in this way.

5 The Markowitz Model and the Efficient Frontier

The fundamental assumptions for the Modern Portfolio Theory boil down to four words – "high risk, high reward." Given a certain amount of risk, the tangency optimal risky portfolio needs to provide investors with the highest compensation in return. Exactly the same as we find our optimal solutions on the boundaries in the diet problem, the Markowitz model states that when we plot all possible portfolios (coming from all possible choices of weights), with x-axis being the risk (standard deviation) and y-axis being the expected return, we will only need to look for the tangency risky portfolio (highest Sharpe ratio) on the boundary of our polygon. This boundary is made up of optimal portfolios either with the same return but a lower risk, or with the same risk but a higher return. In other words, all the portfolios on the efficient frontier are "efficient." This boundary curve exists because of Markowitz's theory of diversification. For the purpose of this project, we will not go into details to prove this.

We followed the Markowitz model to use covariance to calculate our portfolio risk (72). Theoretically, our portfolio is constructed on the expected returns, but as Markowitz points out, "the law of large numbers will ensure that the actual yield of the portfolio will be almost the same as the expected yield." Therefore, our trading data would suffice to build our portfolios.

We define the risk to be the standard deviation of return. With the existing data on our selected stocks' trading information, we could turn our portfolio construction into the Markowitz model using his way of calculation:

Parameters:

(1) $\{p_1, p_2, \cdots, p_{14}\} \in P$, the monthly averages of the returns of our stocks

(2) $\{r_1, r_2, \cdots, r_{14}\} \in R$, the monthly averages of the risks (standard deviations) of our stocks

(3) $\{\rho_{i,j}\} \in Q$, $i = \{1, \cdots, 14\}, j = \{1, \cdots, 14\}$, the volatility correlation coefficients among any two stocks of our stocks

(4) r_f, the risk-free return rate in the market

Variables:

(1) $w_1, w_2, \cdots, w_{14}$, weights of each stock in our portfolio

Constraints:

(1) Weights of all the stocks should add up to 1: $\sum_{i=1}^{14} w_i = 1$

(2) Weights should all be non-negative: $\forall i \in \{1, 2, \cdots, 14\}, w_i \geq 0$

(3) The portfolio return is bounded by the minimum and maximum return of a single stock among all our selected stocks: $min[p_1, \cdots, p_{14}] \leq p_k \leq max[p_1, \cdots, p_{14}]$.

Computations:

(1) p_k, our monthly portfolio return,

$$p_k = \sum_{i=1}^{14} w_i p_i.$$

(2) r_k, our monthly portfolio risk (standard deviation),

$$r_k = \begin{bmatrix} w_1 r_1 & \cdots & w_{14} r_{14} \end{bmatrix} \times \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,14} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,14} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{14,1} & \cdots & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} w_1 r_1 \\ \vdots \\ w_{14} r_{14} \end{bmatrix}$$

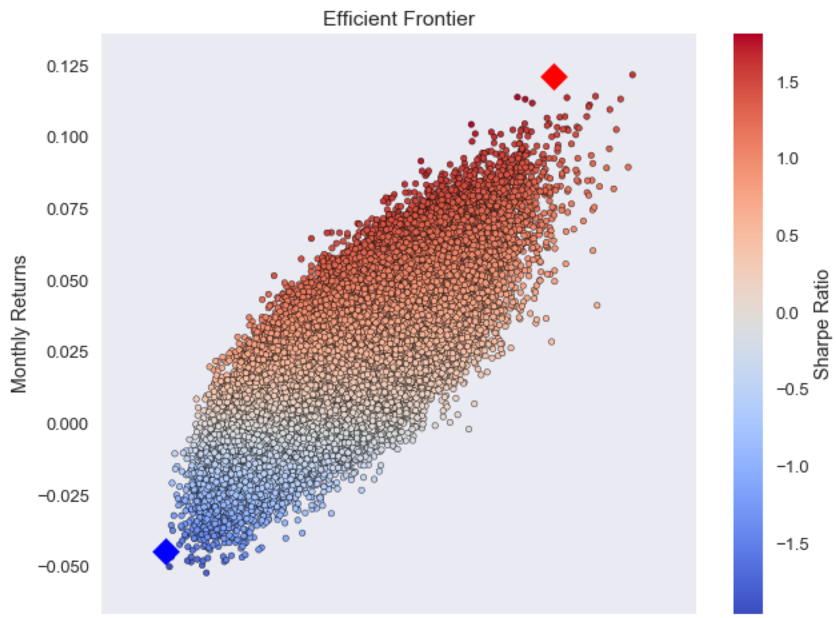
For our convenience, we define a set that contains all possible tuples of length 14, where each 14-tuple can act as weights of our 14 stocks to make a portfolio:

$$W = \left\{ (w_1, w_2, \cdots, w_{14}) \mid \sum_{i=1}^{14} w_i = 1, \text{and } \forall i \in \{1, 2, \cdots, 14\}, w_i \geq 0 \right\}.$$

We observe that each element in W has a corresponding pair of (r_k, p_k), where r_k and p_k are the monthly portfolio risk and return defined as above. However, the size of W is not finite, because $[0, 1]$ is not countable.

6 Monte Carlo Simulation

As we noted above, since we have infinitely many possibilities for weights, it is not practical to generate all the results even with a computer. Therefore, we decided to use Monte Carlo simulation to approximate our polygon with a large amount of simulated portfolios with randomized weights for each stock. In this project, for each monthly portfolio, we generated 50000 simulated portfolios to approximate our polygon.



7 Notes for Python

In our simulation process, we chose a random seed number 101 for reproducibility. We tested several numbers and they all look very similar. We chose 101 in the end because of divine inspiration. For the number of random samples, we picked 50000 for each month. Since we are getting dots to approximate the polygon, the more samples we have, the better the result should be. However, we are also calculating the portfolio for 72 months between 2012 and 2017, and thus we do not want the sample size to be too large. We think 50000 is not too big or too small.

8 Finding Optimal Risky Portfolio

How can we find the tangency risky portfolio on our efficient frontier? If we do not care about the risk, can we figure out if a portfolio on the efficient frontier performs better than others? An answer to this question, after research, is the idea of Sharpe Ratio given by the Nobel laureate William F. Sharpe. The formula is the following:

$$\frac{R_p - R_f}{\sigma_p},$$

Where R_p is the return of portfolio, R_f is the risk-free rate, and σ_p is the standard deviation of the portfolio's excess return. The Modern Portfolio Theory uses the Sharpe Ratio as a measure of a portfolio's overall performance by taking both the risk and the return into account. For simplicity, in this project we assume our risk-free rate is 0%. Since the standard deviation of the portfolio's excess return can represent the stability of this portfolio (or its volatility), this allows us to understand our Sharpe Ratio as simply the ratio of the expected return to the risk. Our objective becomes finding the portfolio with the maximum Sharpe Ratio. Recall that we plot our random samples with x-axis being the risk (standard deviation) and y-axis being the expected return. This means, in our simplified case, a fixed Sharpe Ratio would be a curve, just as we plot the lines of different costs in the diet problem. We can thus color our plot to visualize where we have better portfolio overall.

9 Testing the Efficient Market Hypothesis

The weak form of Efficient Market Hypothesis states that we cannot build our portfolio with past trading data to beat the market in the future. As we derived the tangency risky portfolios with weights for our 14 selected stocks for each month from 2012 to 2017, we could test the Weak-form Efficiency by regressing with lagged variables. For this project, we created two lagged variables, one with a one-month lag, and the other with a two-month one. For example, for the weights of the American Airline over time, as denoted by the variable w_1, we wrote the linear regression function as

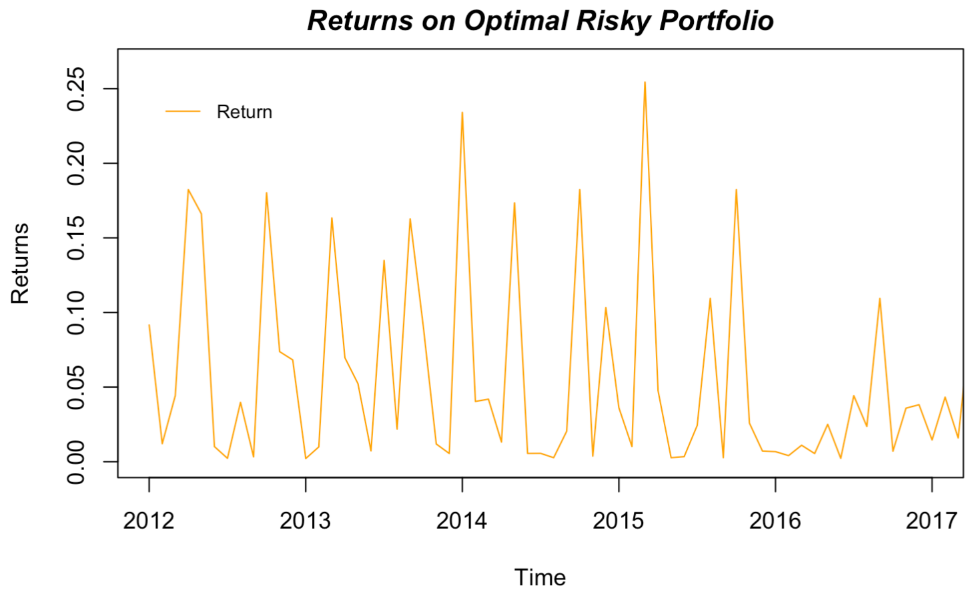
$$w_{1,t} = \beta_0 + \beta_1 w_{1,t-1} + \beta_2 w_{1,t-2} + \epsilon_t$$

and used the Frequentist approach to test the statistical significance of our β's. The Efficient Market Hypothesis is our null hypothesis, namely,

$$\beta_1 = \beta_2 = 0$$

since we would be testing the predictive power of our entire lagged model. In order to do this, we would be checking the p-values, F-statistics, and R-squared in our regression results.

Before we go into the section of our regression results, we would first present some exploratory data analysis on our time series data. We plotted the time series data on returns of our monthly tangency portfolios, and we found little stickiness or sign of autocorrelation in our data. This preliminary analysis provides evidence in favor of the Weak-Form Efficiency. We would now proceed to testing it with the formal statistical method.



For returns, risks, and weights, we created a distributed lag model for each, and the collection of our linear regression models follows

$$\text{Portfolio risk: } r_{k,t} = \beta_0 + \beta_1 r_{k,t-1} + \beta_2 r_{k,t-2} + \epsilon_t$$

$$\text{Portfolio return: } p_{k,t} = \beta_0 + \beta_1 p_{k,t-1} + \beta_2 p_{k,t-2} + \epsilon_t$$

$$\text{Stock weights: } w_{i,t} = \beta_0 + \beta_1 w_{i,t-1} + \beta_2 w_{i,t-2} + \epsilon_t, i \in \{1, \cdots, 14\}.$$

Our regression results indicate no statistical significance for any of the β's, and the F-statistics are also small enough to reject the claim that our lag model has predictive power on the current value of our variables based on their lagged values. That being said, our regression results show evidence in favor of the Weak-Form Efficiency, which states that trading data alone would not allow investors to beat the market.

Here is an example of our results from autocorrelation in American Airlines Weights Time Series.

	Estimate	P-value	R-Squared	F-Statistic
Lag 1	-0.033	0.784	0.008	0.2777
Lag 2	-0.085	0.485		

10 Conclusion

For our project, we used trading data on our selected 14 stocks alone to construct a tangency risky portfolio – portfolio with the highest Sharpe ratio lying on the Efficient

Frontier – for each month from 2012 to 2017. With the construction of these 72 monthly portfolios come the time series data on their returns, risks, and also weights on each individual stock. We then applied a distributed lag model with one-month and two-month lag for each of our variables, and our regression results show little statistical significance in the predictive power of our model. That said, we failed to reject the Efficient Market Hypothesis in its weak form and arrived at the conclusion that investors cannot beat the market with technical analysis alone.

11 Reference

Bodie, Zvi, Alex Kane, and Alan J. Marcus. *Investments and portfolio management.* McGraw Hill Education (India) Private Limited, 2013.

Markowitz, Harry. "Portfolio selection." *The journal of finance* 7.1 (1952): 77-91.

Wooldridge, Jeffrey M. *Introductory econometrics: A modern approach.* Nelson Education, 2016.