ISSN 2959-6130

Analysis of Portfolio Optimization Performance: Markowitz Model and Index Model in Capital Markets

Qianqian Yao

Arizona State University
Columbia University
American Scholastic Convention (ASC)
Instructor: Alexei Chekhlov

Abstract:

In the post-pandemic era, marked by economic uncertainty and stock market volatility, investors are turning to Modern Portfolio Theory (MPT) to guide their investment decisions. This study aims to develop a robust mathematical framework for portfolio construction using the Markowitz Model (MM) and the Index Model (IM) with added constraints. The objectives are threefold: (1) to create a portfolio framework that reflects investor preferences using MM and IM with additional constraints, (2) to compare these models against the Gaussian Distribution using Python and Excel, (3) to compare the statistical data and correlation tests of daily logarithmic returns in Python with monthly excess returns in Excel, and (4) to evaluate their performance relative to the traditional Markowitz approach through Monte Carlo simulations.

The methodology involves incorporating five additional constraints into the MM and IM models. The analysis uses 20 years of historical daily return data for ten stocks from various sectors, one equity index (S&P 500) as a risk-free rate proxy (1-month Fed Funds rate). Daily logarithmic returns are analyzed using Python, while Excel Solver and Solver Table are employed for monthly excess return data.

Keywords: Markowitz Model, Index Model, Constraints, Python, Monte Carlo Simulation

1. Introduction

In the period of the post-pandemic era, with the capricious economic situation, the investors consider how to apply the Modern Portfolio Theory (MPT), a pragmatic approach to portfolio selection that aims to maximize overall returns within an acceptable level of risk, in the unprecedented volatility on the stock market. This study is based on the Markowitz Model ("MM") and the Index Model ("IM") with extending five various additional constraints on portfolio selection.

The main idea of this study is (1) to find the mathematical framework to build an investment portfolio t with considering the investor's preference by implementing the ideas of the Markowitz Model ("MM") and the Index Model ("IM") with adding five various additional constraints assumptions, and (2) to give the results of the comparation with Markowitz by supervising a Monte Carlo simulation.

The project is designed by using a recent 20 years of historical daily total return data for ten stocks, which belong in groups to three-four different sectors (according to Yahoo!finance), one (S&P 500) equity index (a total of eleven risky assets) and a proxy for risk-free rate (1-month Fed Funds rate). To reduce the non-Gaussian effects, the project aggregates the daily data to the monthly observations, and based on those monthly observations, with calculating all proper optimization inputs for the full Markowitz Model ("MM"), alongside the Index Model ("IM"). Using these optimization inputs for MM and IM, the projects find the regions of permissible portfolios (efficient frontier, minimal risk portfolio, optimal portfolio, and minimal return portfolios frontier) for the following five cases of the additional constraints. The project presents the results in both the tabular and graphical form with the objective to make inferences and comparisons between the sets of constraints for each optimization problem and between the MM and IM models in general. Notably, the explanations of the observations making the connections to theory is given to predict possible outcomes by conducting a Monte Carlo simulation.

The list of five cases of the additional constraints:

1. This additional optimization constraint is designed to simulate the Regulation T by FINRA (https://www.finra.org/rules-guidance/key-topics/margin-accounts), which al-

lows broker-dealers to allow their customers to have positions, 50% or more of which are funded by the customer's account equity:

$$\sum_{i=1}^{11} |w_i| \leq 2 ;$$

2. This additional optimization constraint is designed to simulate some arbitrary "box" constraints on weights, which may be provided by the client:

$$|w_i| \leq 1, \text{for } \forall i ;$$

3. A "free" problem, without any additional optimization constraints, to illustrate how the area of permissible portfolios in general and the efficient frontier in particular look like if you have no constraints;
4. This additional optimization constraint is designed to simulate the typical limitations existing in the U.S. mutual fund industry: a U.S. open-ended mutual fund is not allowed to have any short positions, for details see the Investment Company Act of 1940, Section 12(a)(3) (https://www.law.cornell.edu/uscode/text/15/80a-12):

$$w_i \geq 0, \text{for } \forall i ;$$

5. Lastly, we would like to see if the inclusion of the broad index into our portfolio has positive or negative effect, for that we would like to consider an additional optimization constraint:

$$w_1 = 0 .$$

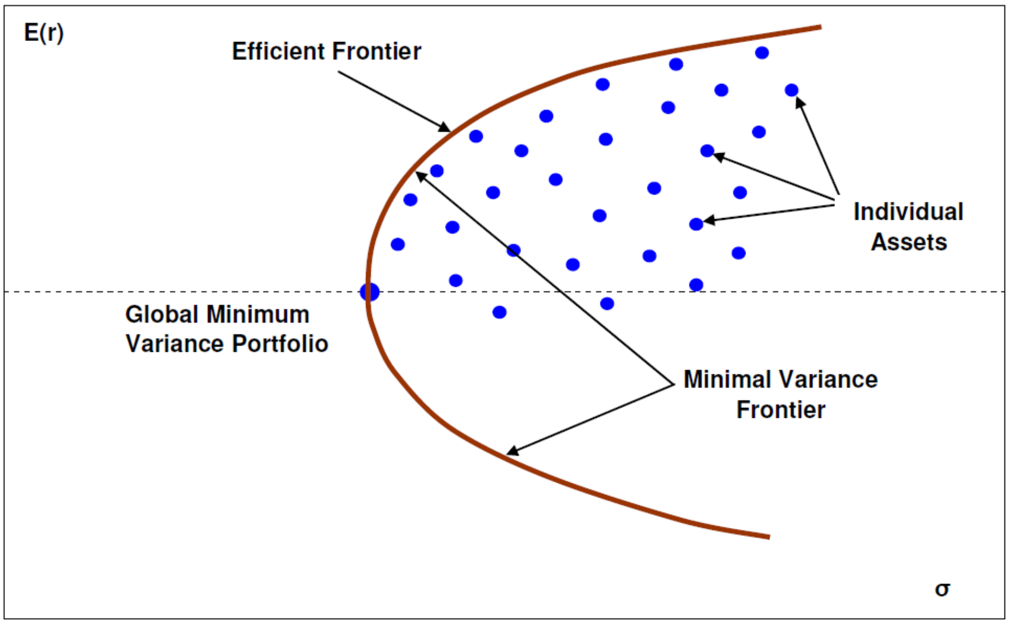
2 Theoretical Model

2.1 Markowitz Model

Markowitz Model, so-called Markowitz Portfolio Optimization Model, is a process of security selection that maximizes overall returns within an acceptable level of risk, proposed by Nobel laureate Harry Markowitz in 1952. The basic of MM is the Markowitz efficient set, also known as the efficient frontier, is a fundamental part of modern portfolio theory (MPT). According to Harry Markowitz, "Portfolio Selection" (Markowitz, 1952), Journal of Finance, March 1952, briefly, the approach to the case of many risky assets and a risk-free asset*.

2.1.1 The process of Markowitz Optimization

The mathematical approach has three parts:
1. Determine the "opportunity set" (minimal variance frontier): allowed risk-return combinations. Minimal variance frontier is a graph of the lowest possible variance that can be attained for a given portfolio expected return.



The Minimum-Variance Frontier of Risky Assets

All individual assets are located to the right of the minimum variance frontier, i.e., portfolios consisting of single assets are sub-efficient. The portion of the minimum-variance frontier that is above the Global Minimum-Variance portfolio is called the efficient frontier of risky assets: they offer the best risk-return combination.
2. Identify the optimal risky portfolio as the steepest Capital Allocation Line (CAL), is the description of all the available risk-return combinations, tangential to the opportunity set. The slope of CAL is reward-to-volatility ratio, so-called the Sharpe ratio after William Sharpe, who first used it extensively. Suppose, investor decided on the composition of risky portfolio, P (with expected return E(rP) and standard deviation

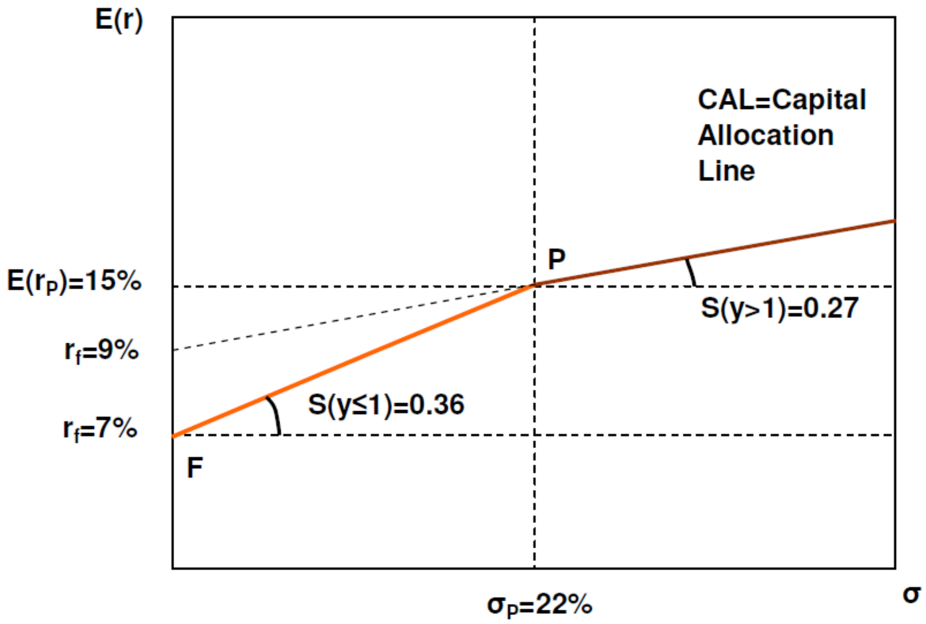
σP) and he wants to know the appropriate proportion y. The remaining part (1-y) will be allocated into risk-free portfolio F. Then, the rate of return of the complete portfolio C is: $r_C = y \times r_P + (1-y) \times r_f$

$$E(r_C) = y \times E(r_P) + (1-y) \times r_f = r_f + y \times [E(r_P) - r_f].$$

The function of CAL:

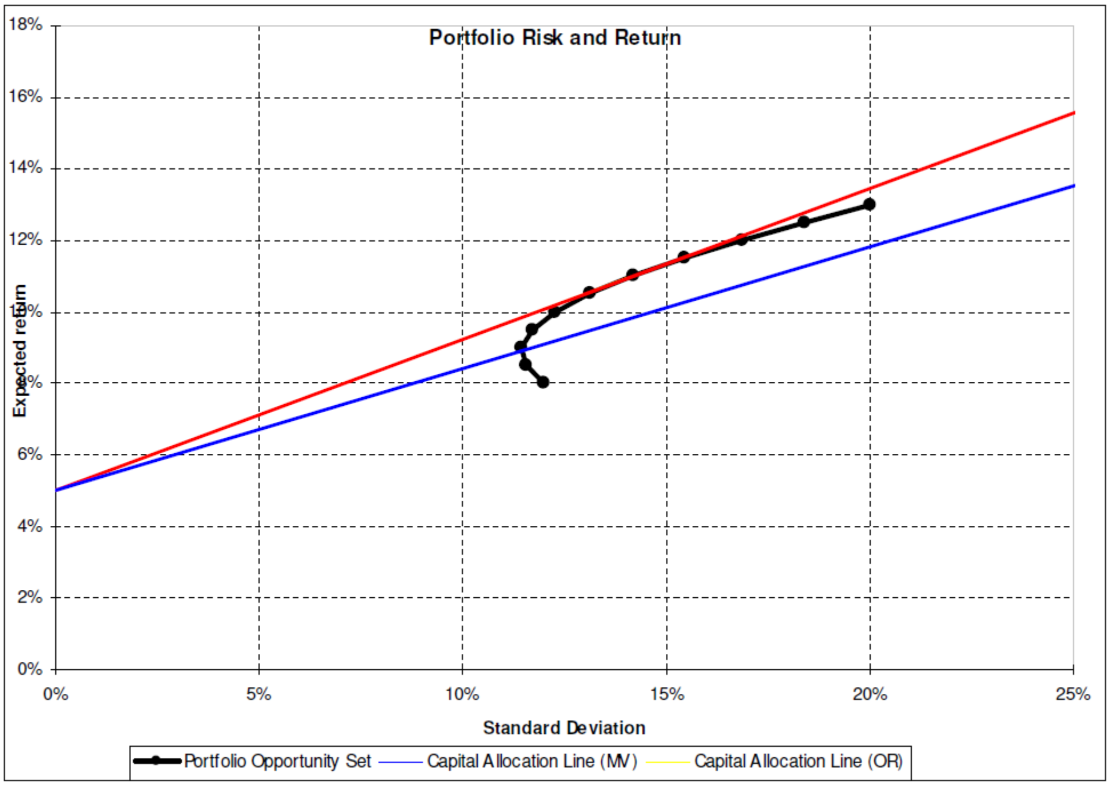
$$E(r_C) = r_f + \sigma_C \times \frac{E(r_P) - r_f}{\sigma_P}.$$

We look for the capital allocation line with the highest return-to-volatility (Sharpe) ratio (i.e., the steepest slope) $S = \frac{E(r_C) - r_f}{\sigma_C}$. The leveraged portfolio has a relatively high expected return and standard deviation, but the return to volatility (Sharpe) ratio is the same $S = \frac{E(r_P) - r_f^B}{\sigma_P}$. Then the leveraged portfolio has a lower reward-to-volatility (Sharpe) ratio. Investors face a kink in the capital allocation line when their borrowing capacity is exhausted, leading to the borrowing rate surpassing the lending rate. This point marks a restriction where an investor can no longer leverage additional funds at the risk-free rate.



The investment opportunity set with different borrowing and lending rates

The optimal risky portfolio P corresponds to tangent CAL on the efficient frontier. Such CAL dominates all other feasible lines.



The efficient frontier of risky assets with the optimal CAL

3. Choose the appropriate complete portfolio by mixing with the risk-free asset given risk-aversion.

This part is investor- (risk-aversion-)dependent. Given risk-aversion parameter A, on the whole (CAL) we now need to choose one optimal risky portfolio allocation y. We construct the optimal risky portfolio P. Now, given the investor's risk aversion, we can calculate the fraction of the complete portfolio invested in the risky and risk-free components (T-Bills). Calculate the shares of complete portfolio invested into each asset and in T-Bills. Optimal ratio to be invested in risky asset for an investor with risk aversion coefficient A:

$$y = \frac{E(r_P) - r_f}{A\sigma_P^2}.$$

The process of the function of risk aversion coefficient:

$$E(r_C) = r_f + y \times [E(r_P) - r_f],$$

$$\sigma_C^2 = y^2 \times \sigma_P^2,$$

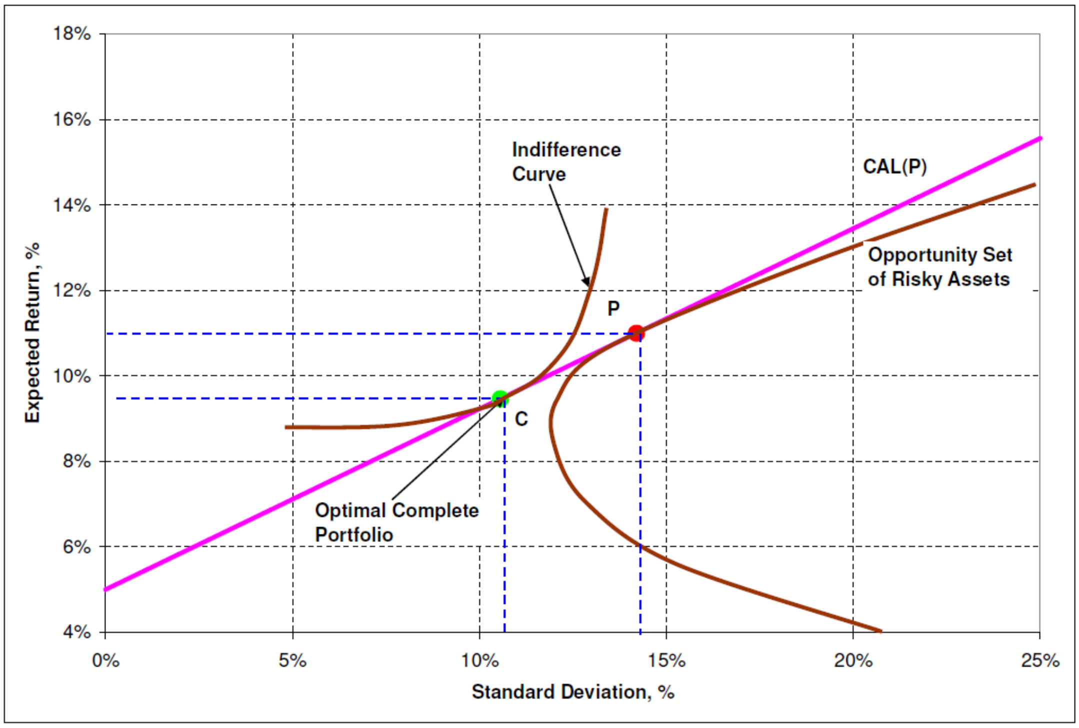
$$U(r) = E(r) - \frac{A\sigma^2}{2}.$$

$$E(r_C) - \frac{A\sigma_C^2}{2} =$$

$$r_f + y \times [E(r_P) - r_f] - \frac{Ay^2\sigma_P^2}{2} \to \max_y$$

$$y^* = \frac{E(r_P) - r_f}{A\sigma_P^2}.$$

U(r) is utility function.



Determination of the optimal complete portfolio

Although the above process seems complete, it requires: a set of estimates of the expected returns of each risky asset and a set of estimates of their covariance matrix. Some combinatorics related to these estimates: The number of estimates of returns $\{E(r_i)\}_{i=1}^{n}$ is n; The number of non-repeated elements in the covariance matrix:

$$\begin{Bmatrix} \sigma_1^2 & c_{1,2} & c_{1,3} & \cdots & c_{1,n-1} & c_{1,n} \\ c_{2,1} & \sigma_2^2 & c_{2,3} & \cdots & c_{2,n-1} & c_{2,n} \\ c_{3,1} & c_{3,2} & \sigma_3^2 & \cdots & c_{3,n-1} & c_{3,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{n-1,1} & c_{n-1,2} & c_{n-1,3} & \cdots & \sigma_{n-1}^2 & c_{n-1,n} \\ c_{n,1} & c_{n,2} & c_{n,3} & \cdots & c_{n,n-1} & \sigma_n^2 \end{Bmatrix} \text{ is } \frac{\mathrm{n}\times(\mathrm{n}+1)}{2},$$

which is a sum of arithmetic progression, which is composed of : diagonal terms, individual assets squared standard deviations, and $\frac{\mathrm{n}\times(\mathrm{n-1})}{2}$ off - diagonal terms, cross - covariances. Thus, the total number of estimates needed is: $n+\frac{n\times(n+1)}{2}$, which for a 50 - asset portfolio is equal to 1,325!

There is the extension of diversification: Power of Diversification:

The basic covariance- related formulas:

$$\sigma_{\mathrm{P}}^2=\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(r_i,r_j).$$

Consider now a simplified case of equally-weighted portfolio, that is when $w_i=\frac{1}{n}$ for any i.

$$\sigma_P^2=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sigma_i^2+\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}\frac{1}{n^2}Cov(r_i,r_j).$$

Introduce the notions of average variance and average covariance:

$$\bar{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \text{, and } \overline{Cov}=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}Cov(r_i,r_j).$$

Then we can re -write the portfolio variance as :

$$\sigma_P^2 = \frac{1}{n}\overline{\sigma}^2 + \frac{n-1}{n}\overline{Cov}.$$

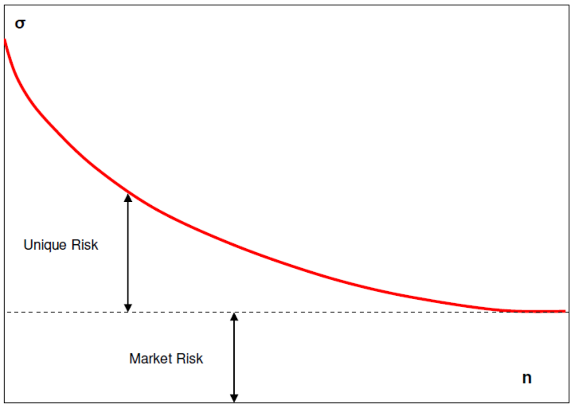
Conclusion:

when all the risk is firm- specific, portfolio variance can be driven to zero! For $n \to +\infty$, we have $\sigma_P^2 \to \overline{Cov}$, which is a function of systematic factors in economy.

Assume, further that for $\forall i: \sigma_i = \sigma$, and for $\forall i, j: Cov(r_i, r_j) = \rho\sigma^2$. Then:

$\sigma_P^2 = \frac{1}{n}\sigma^2 + \frac{n-1}{n}\rho\sigma^2$. We then have for $\rho = 0$ portfolio variance approaches zero for large ; for $\rho = +1$ portfolio variance is σ^2 independently of ; for any ρ portfolio variance approaches $\rho\sigma^2$ for large .

There are two factors affects one stock ("Portfolios"): macro-economic factors (conditions of general economy, business cycle, inflation, interest rates, exchange rates, etc.) and firm-specific influences (advance in research and development, personnel changes, etc.). Although we can't change the economic environment (systematic risk), the unsystematic risk can be degraded. The Naïve diversification, simply including additional securities into such a portfolio, can reduce portfolio risk by reducing firm-specific influences.



Some risk is systematic, and some - firm-specific.

Only considered equally weighted portfolios achieve naïve diversification, efficient diversification is optimal risky portfolios, i.e. having minimal possible risk for any given level of expected return.

Some notations and formulas of Portfolio of two risky are listed:

The Portfolio of two risky assets: a bond fund (long-term debt) D, and a stock fund, E.

Descriptive statistics: Expected Return, E(r);Standard Deviation, σ;Covariance, Cov(rD,rE);Correlation Coefficient, ρDE

w_D - a proportion which is invested into the bond fund;

$w_E = 1 - w_D$ - is invested into the stocks fund.

Rate of return of such portfolio :

$$r_p = w_D \times r_D + w_E \times r_E.$$

The expected return of such portfolio :

$$E(r_p) = w_D \times E(r_D) + w_E \times E(r_E).$$

The variance :

$$\sigma_p^2 = w_D^2 \times \sigma_D^2 + w_E \times \sigma_E^2 + 2 \times w_D \times w_E \times Cov(r_D, r_E).$$

Using the definition of correlation coefficient ρ_{DE} :

$$Cov(r_D, r_E) = \rho_{DE} \times \sigma_D \times \sigma_E$$

we get for the portfolio variance:

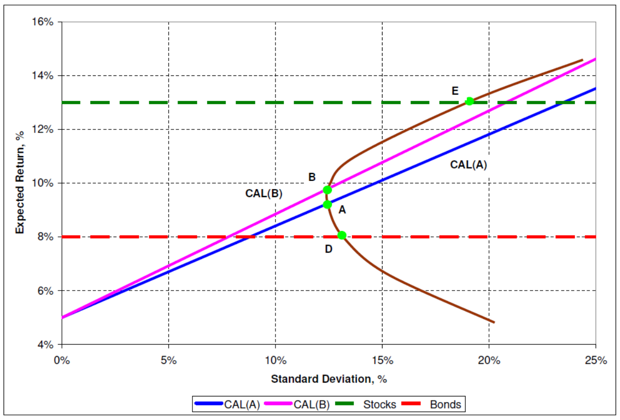
$$\sigma_p^2 = w_D^2 \times \sigma_D^2 + w_E^2 \times \sigma_E^2 + 2 \times w_D \times w_E \times \sigma_D \times \sigma_E \times \rho_{DE}.$$

Case of perfect positive correlation $\rho_{DE} = +1$:

$$\sigma_p = w_D \times \sigma_D + w_E \times \sigma_E.$$

Case of perfect negative correlation $\rho_{DE} = -1$:

$$\sigma_p = |w_D \times \sigma_D - w_E \times \sigma_E|.$$



The opportunity set of the debt and equity funds with the optimal CAL and the optimal risky portfolio

The lowest value of portfolio variance is zero, when ρ=-1, and solution for weights is:

$$w_D = \frac{\sigma_E}{\sigma_D + \sigma_E},$$

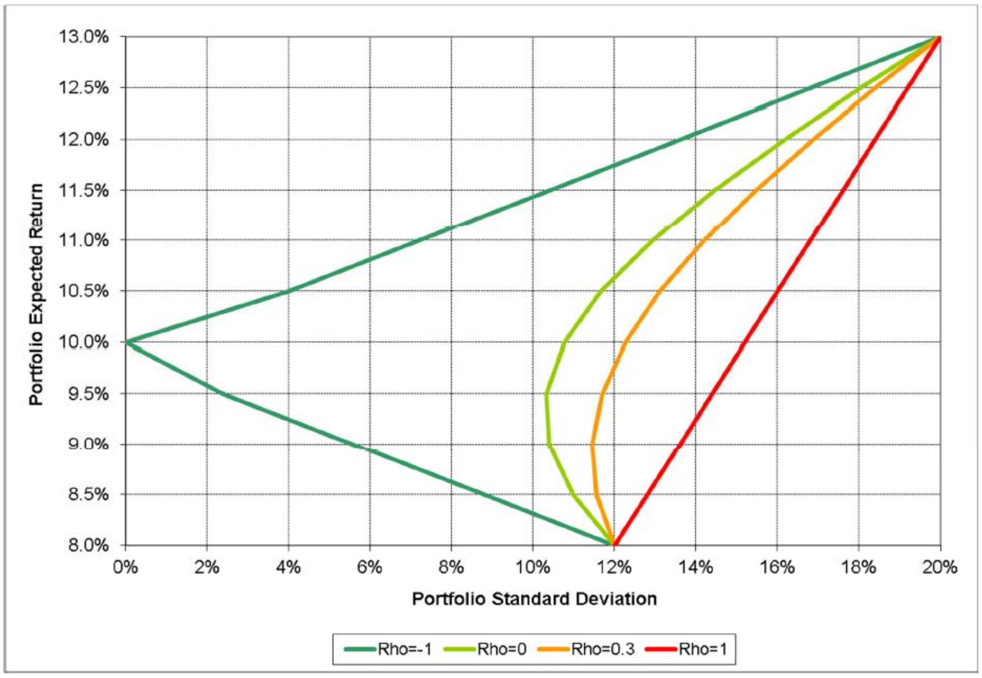
$$w_E = 1 - w_D = \frac{\sigma_D}{\sigma_D + \sigma_E}.$$

How to decide the weights of the debt and equity. The evolution of weights' solution is presented as following. First, there are some effects of varying the stocks weight on Portfolio Standard Deviation

• For any ρ<1, as the portfolio weight in the equities grows from 0 to 1, portfolio risk first falls, then achieves its lowest (optimal) point, but then rises again as the portfolio becomes heavily concentrated in equities.

• Only for ρ=1 portfolio standard deviation monotonically grows from low risk to high risk asset.

• For any ρ<1 the minimal variance portfolio has a standard deviation smaller than that of either of the individual components.



Portfolio expected returns as a function of standard deviation

Second, the illustration of the exact analytic solution of weights calculation.

Introduce easier notations: D →1, E → 2, f→0,

$$\begin{cases} r = w_1 r_1 + w_2 r_2 \\ \sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho \\ w_2 = 1 - w_1 \end{cases}$$

$$\frac{r - r_0}{\sigma} \to \max_{w_1}.$$

$$\frac{d}{dw_1}\left(\frac{w_1 r_1 + w_2 r_2 - r_0}{\sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho}}\right) = 0.$$

Introduce new variables: $R_1 = r_1 - r_0,\ R_2 = r_2 - r_0.$

Then: $w_1 r_1 + w_2 r_2 - r_0 = w_1 R_1 + w_2 R_2.$

Differentiation and use of $\frac{\mathrm{dw}_2}{\mathrm{w}_1} = -1$ gives:

$$(R_1 - R_2)\times\left(w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho\right) =$$
$$= (w_1 R_1 + w_2 R_2)\times\left(w_1\sigma_1^2 - w_2\sigma_2^2 + \sigma_1\sigma_2\rho[w_2 - w_1]\right).$$

w_1^2 – term cancels out and we are left with:

$$w_1 \times \left(R_2\left[\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho\right] + [R_1 - R_2]\sigma_2^2 - [R_1 - R_2]\sigma_1\sigma_2\rho\right) +$$
$$+ R_2\left[-\sigma_2^2 + \sigma_1\sigma_2\rho\right] - [R_1 - R_2]\sigma_2^2 = 0,$$

with the solution:

$$\begin{cases} w_1 = \dfrac{R_1\sigma_2^2 - R_2\sigma_1\sigma_2\rho}{R_2\sigma_1^2 + R_1\sigma_2^2 - (R_1 + R_2)\sigma_1\sigma_2\rho}, \\ w_2 = 1 - w_1. \end{cases}$$

2.2 Index Model

The Index Model is introduced because of some drawbacks of the Markowitz Procedure. First, the model requires a large number of estimates to populate the covariance matrix. Second, it does not provide any guidelines for finding useful estimates of these covariances or risk premiums, which are essential for constructing an efficient frontier for risky assets. Because past returns are noisy guides to expected future returns, this shortcoming is obvious.

The index model was first indicated by Willam Sharpe. According to Sharpe analysis, “This paper describes the advantages of using a particular model of the relationships among securities for practical applications of the Markowitz portfolio analysis technique. A computer program has been developed to take full advantage of the model: 2,000 securities can be analyzed at an extremely low cost-as little as 2% of that associated with standard quadratic programming codes. Moreover, preliminary evidence suggests that the relatively few parameters used by the model can lead to very nearly the same results obtained with much larger sets of relationships among securities. The possibility of low-cost analysis, coupled with a likelihood that a relatively small amount of information need be sacrificed make the model an attractive candidate for initial practical applications of the Markowitz technique.” (Sharpe, 1963) Thus, Sharpe suggested that the extension of Markowitz’s work on portfolio analysis. Index model is to solve the problem of Markowitz’s technique. The Index Model of the relationships between securities, points out the ways in which it allows the portfolio analysis problem to be simplified, and provides evidence on the costs and desirability of using this model for practical applications of Markowitz’s technique.

2.2.1 The process of Index Model

Decompose security returns into: $r_i = E(r_i) + e_i$,

where the unexpected return has zero mean : $E(e_i) = 0$

and a standard deviation of σ_i The uncertainty is firm-specific :

$$E(e_i \cdot e_j) = 0 \text{ for all } i \neq j.$$

Further, assume that e_i are normally- distributed.

Next, assume that there is a common, stock-independent “macroeconomic” random factor m that affects all stocks equally: $r_i = E(r_i) + m + e_i$, such that m is also normally- distributed, its’ standard deviation is σ_m and $E(r_i \cdot m) = 0$.

$$\sigma_i^2 = \sigma_m^2 + \sigma^2(e_i).$$

Then: $Cov(r_i, r_j) = \sigma_m^2$.

Finally, we need to consider that some companies are more dependent on macroeconomic factors and some are less dependent on them:

$$r_i = E(r_i) + \beta_i \cdot m + e_i.$$

The risk and covariance are determined by the stock’s beta coefficient: β

$$\sigma_i^2 = \beta_i^2 \cdot \sigma_m^2 + \sigma^2(e_i),\ Cov(r_i, r_j) = \beta_i \cdot \beta_j \cdot \sigma_m^2.$$

It is most convenient to choose a broad index (S&P 500) as a broad macroeconomic factor. It has a considerable amount of past data available for estimation. If denotes m market index, then, its excess return is

$$R_M = r_M - r_f$$

and standard deviation σ_m. The factor β can be estimated using linear regression between observations of $R_i(t)$ and $R_M(t)$:

$$R_i(t) = \alpha_i + \beta_i \cdot R_M(t) + e_i(t).$$

If we take the expected value of both sides, we get :

$$E(R_i) = \alpha_i + \beta_i \cdot E(R_M),$$

where the first term, α_i ,is non -market risk - premium. (Sharpe)

Risk and covariance

Total risk = Systematic risk + Firm-Specific risk

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma^2(e_i);$$

Covariance = Product of betas * Market index risk

$$Cov(r_i, r_j) = \beta_i \beta_j \sigma_M^2$$

Correlation = Product of correlations with the market index

$$Corr(r_i, r_j) = \frac{\beta_i \beta_j \sigma_M^2}{\sigma_i \sigma_j} = Corr(r_i, r_M) \times Corr(r_j, r_M)$$

All of these are determined by the security’s beta and the properties of the market index.

To further assume, for simplicity, an equally - weighted portfolio $w_i = \frac{1}{n}$:

$$R_P = \frac{1}{n}\sum_{i=1}^{n}(\alpha_i + \beta_i \cdot R_M + e_i) = \frac{1}{n}\sum_{i=1}^{n}\alpha_i + \left(\frac{1}{n}\sum_{i=1}^{n}\beta_i\right) \cdot R_M + \frac{1}{n}\sum_{i=1}^{n} e_i.$$

from which follows:

$$\beta_P = \frac{1}{n}\sum_{i=1}^{n}\beta_i,\ \alpha_P = \frac{1}{n}\sum_{i=1}^{n}\alpha_i \text{ and } e_P = \frac{1}{n}\sum_{i=1}^{n} e_i.$$

The firm - specific risk is diversifiable:

$$\sigma^2(e_P) = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \sigma^2(e_i) = \frac{1}{n} \cdot \overline{\sigma}^2(e).$$

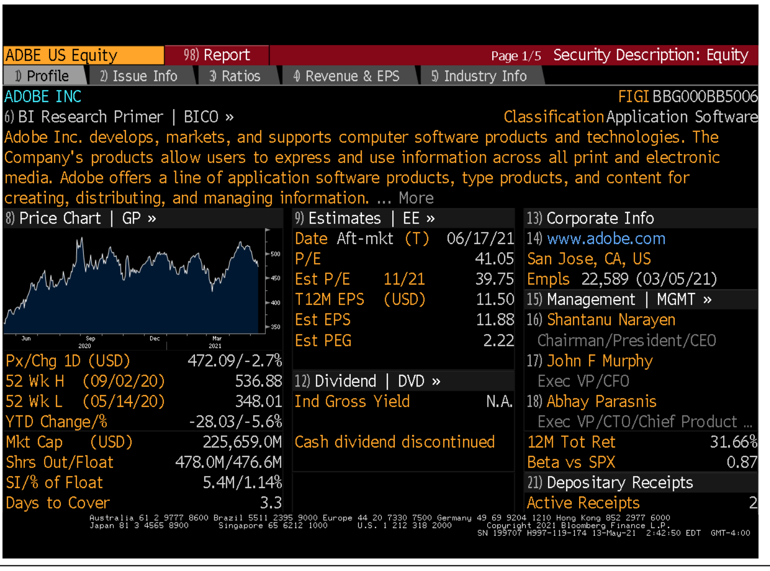
The index model is a very useful abstraction: it reduces the number of estimates required from O(n2) to O(n). This is crucial for the specialization of security analysis work: analysts can specialize by industry. The covariance between securities is due to a single common factor, the effect of the market index. The price of the index model's simplification is the restrictions it places on the statistics of asset returns. The assumption that asset returns can be perfectly decomposed into macro and micro components is an oversimplification of the real world. For example, this will ignore industry-specific events that do not affect the macro environment. If stocks with correlated residuals have high alpha, then the index model can lead to a worse portfolio than the full Markowitz model. For example: if the excess returns of BP (British Petroleum) and RDS/A (Royal Dutch Shell) are correlated, the index model will ignore this. The Markowitz model will take this correlation into account. The two models lead to completely different portfolios for a small number of instruments. If the residual return correlation of two stocks is positive, the Markowitz model will give both stocks a smaller weight; if it is negative, the index model will underweight both stocks, resulting in higher than Markowitz variance.

3 Research Design

According to the theory, I design the test with 10 stocks. First, this test lists the historical daily total return data of 10 stocks in 3-4 different industries for nearly 20 years (according to Yahoo Finance), the S&P 500 stock index (which contains a total of 11 risky assets), and a proxy for the risk-free rate (the 1-month federal funds rate). Second, to test 10 stocks through IM and MM calculations as well as diminishing the non-Gaussian effects, the data are dealt with Excel and Python. Finally, the results are conducted a Monte Carlo simulation and scenario analysis.

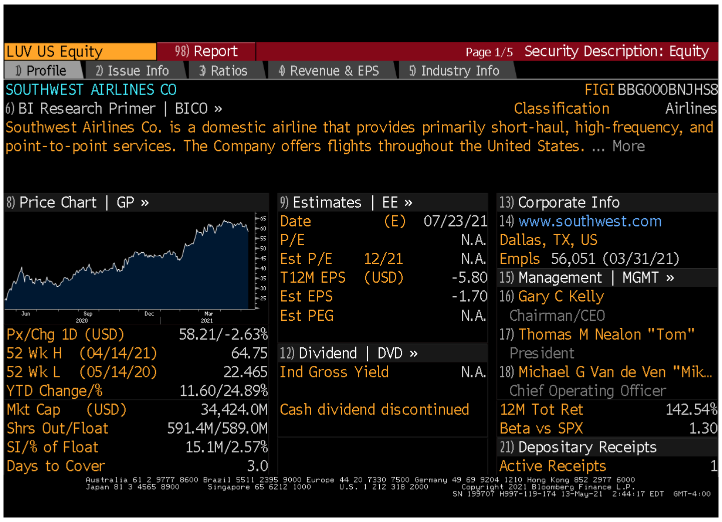
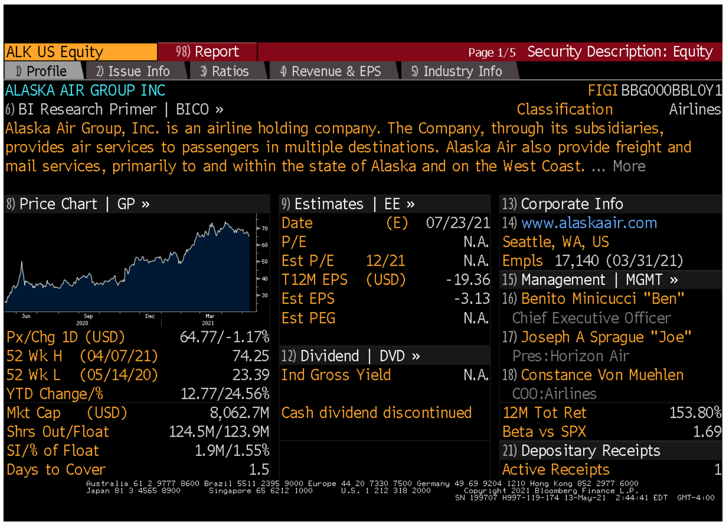
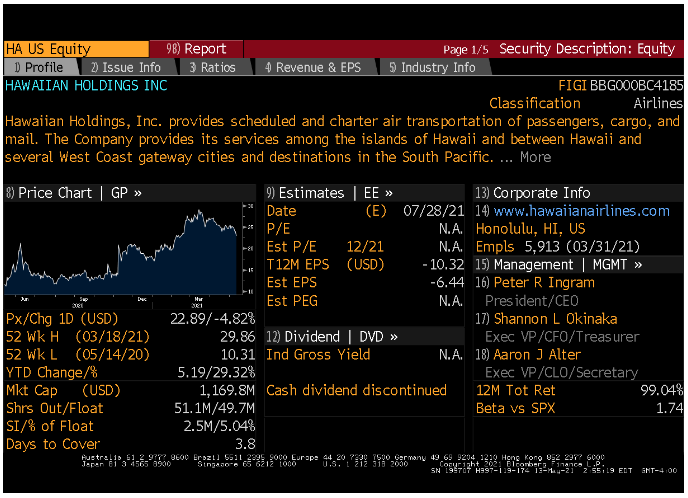
3.1 Stock Description and Analysis

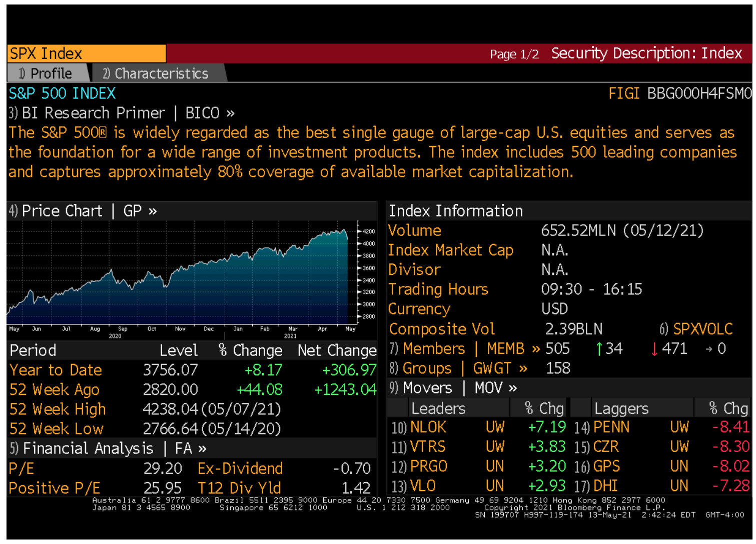
The chat illustrates the raw data of 10 stocks, the S&P 500 stock index and a proxy for the risk-free rate (the 1-month federal funds rate) from Bloomberg.

<table>
<tr><th>Stock Code</th><th>Stock Name</th><th>Sector</th><th>Equity Description</th></tr>
<tr><td>Technology</td><td>ADBE</td><td>Adobe Inc.</td><td>
</td></tr>
</table>

<table>
<tr>
<td></td>
<td>IBM</td>
<td>International Business Machines Corporation</td>
<td>
</td>
</tr>
<tr>
<td></td>
<td>SAP</td>
<td>SAP SE</td>
<td>
</td>
</tr>
<tr>
<td>Financial Services</td>
<td>BAC</td>
<td>Bank of America Corporation</td>
<td>
</td>
</tr>
</table>

<table>
<tr><td></td><td>C</td><td>Citigroup Inc.</td><td>
</td></tr>
<tr><td></td><td>WFC</td><td>Wells Fargo Company</td><td>
</td></tr>
<tr><td></td><td>TRV</td><td>The Travelers Companies, Inc.</td><td>
</td></tr>
</table>

<table>
<tr><td rowspan="3">Industrials</td><td>LUV</td><td>Southwest Airlines Co.</td><td>
</td></tr>
<tr><td>ALK</td><td>Alaska Air Group, Inc.</td><td>
</td></tr>
<tr><td>HA</td><td>Hawaiian Holdings, Inc.</td><td>
</td></tr>
</table>

Index	SPX	S&P 500 Index	

4 Data Test in Python and Excel

The test part is separated into two parts: using daily logarithm returns in Python and calculating monthly excess returns in Excel. Using monthly data is conventional but I observe daily data of stocks in Python to compare with using monthly data excess return in Excel. First, I test the simple return and logarithm return on every 10 stocks and the raw data of 10 stocks, the S&P 500 stock index and a proxy for the risk-free rate (the 1-month federal funds rate) from 2001 to 2021 by Python. It is important to be consistent in the way we calculate returns. If we choose to calculate simple returns, we must do so for all further financial calculations. Similarly, if we decide to calculate log returns, we should only use log returns. There is no universal rule for the method we should use, but most econometricians agree that simple returns are preferable when you must deal with multiple assets over the same time period, and logarithm returns are preferable when you calculate a single asset over a period of time. Thus, we compare the simple return for each asset with the logarithm return.

The formula of simple return:

$$\frac{End\ Price\ -\ Beginning\ Price}{Beginning\ Price}$$

$$=\frac{End\ Price}{Beginning\ Price}-1$$

The formula of logarithm return:

$$ln\left(\frac{End\ Price}{Beginning\ Price}\right)$$

4.1 Data Test in Python and Excel

ADBE:

```
Date
2001-05-11         NaN
2001-05-14   -0.017812
2001-05-15   -0.009195
2001-05-16    0.059294
2001-05-17    0.061085
                ...
2021-05-05   -0.008172
2021-05-06   -0.006328
2021-05-07    0.010587
2021-05-10   -0.019131
2021-05-11    0.012120
Name: simple_return, Length: 5031, dtype: float64
Date
2001-05-11         NaN
2001-05-14   -0.017973
2001-05-15   -0.009238
2001-05-16    0.057602
2001-05-17    0.059292
                ...
2021-05-05   -0.008206
2021-05-06   -0.006349
2021-05-07    0.010531
2021-05-10   -0.019317
2021-05-11    0.012047
Name: log_return, Length: 5031, dtype: float64
```

There are 5031 returns, excluding holidays, for each of the companies. A small difference is between simple returns and log returns. I will use logarithm returns throughout all testing. In addition, the daily return output is a very small number, much smaller than 1%, which makes it difficult to interpret. I calculate a close approximation of the average annual rate of return by multiplying the average daily return by 250. The number of trading days actually ranges from 250 to 252 because it excludes non-trading days, such as Saturdays, Sundays, and bank holidays. This value will be easier to understand than the previous one. The statistical information of daily data is as follows:

The stock's variance formula: $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

The stock's standard deviation formula: $s = \sqrt{s^2}$

Sample skewness formula: $\text{Skewness} \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i-\bar{X})^3}{s^3}$

Sample excess kurtosis (KE): $K_E \approx \left[\left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i-\bar{X})^4}{s^4}\right]-3$

```
The annual mean of log return ADBE
1586.6373939846867%
The mean of logreturn ADBE is 0.0006
The median of logreturn ADBE is 0.0011
The standard deviation of logreturn ADBE is 0.0237
The annual mean of logreturn ADBE is 15.8664
The annual standard deviation of logreturn ADBE is 0.3750
The skewness of logreturn ADBE is -0.7167
The kurtosis of logreturn ADBE is 15.7644
```

```
The annual mean of log return IBM
126.51039826330876%
The mean of logreturn IBM is 0.0001
The median of logreturn IBM is 0.0002
The standard deviation of logreturn IBM is 0.0153
The annual mean of logreturn IBM is 1.2651
The annual standard deviation of logreturn IBM is 0.2425
The skewness of logreturn IBM is -0.2757
The kurtosis of logreturn IBM is 7.8145
```

```
The annual mean of log return SAP
660.1197815557717%
The mean of logreturn SAP is 0.0003
The median of logreturn SAP is 0.0007
The standard deviation of logreturn SAP is 0.0207
The annual mean of logreturn SAP is 6.6012
The annual standard deviation of logreturn SAP is 0.3276
The skewness of logreturn SAP is -0.2767
The kurtosis of logreturn SAP is 13.2839
```

```
The annual mean of log return BAC
218.92172863047622%
The mean of logreturn BAC is 0.0001
The median of logreturn BAC is 0.0004
The standard deviation of logreturn BAC is 0.0290
The annual mean of logreturn BAC is 2.1892
The annual standard deviation of logreturn BAC is 0.4581
The skewness of logreturn BAC is -0.3521
The kurtosis of logreturn BAC is 27.4656
```

```
The annual mean of log return C
-936.3539669754759%
The mean of logreturn C is -0.0004
The median of logreturn C is 0.0000
The standard deviation of logreturn C is 0.0311
The annual mean of logreturn C is -9.3635
The annual standard deviation of logreturn C is 0.4921
The skewness of logreturn C is -0.5765
The kurtosis of logreturn C is 39.7781
```

```
The annual mean of log return WFC
346.90393409638125%
The mean of logreturn WFC is 0.0001
The median of logreturn WFC is 0.0000
The standard deviation of logreturn WFC is 0.0243
The annual mean of logreturn WFC is 3.4690
The annual standard deviation of logreturn WFC is 0.3845
The skewness of logreturn WFC is 0.7073
The kurtosis of logreturn WFC is 26.0160
```

```
The annual mean of log return TRV
603.2516707950169%
The mean of logreturn TRV is 0.0002
The median of logreturn TRV is 0.0005
The standard deviation of logreturn TRV is 0.0181
The annual mean of logreturn TRV is 6.0325
The annual standard deviation of logreturn TRV is 0.2865
The skewness of logreturn TRV is -0.3548
The kurtosis of logreturn TRV is 22.8523
```

```
The annual mean of log return LUV
596.5702694988171%
The mean of logreturn LUV is 0.0002
The median of logreturn LUV is 0.0000
The standard deviation of logreturn LUV is 0.0225
The annual mean of logreturn LUV is 5.9657
The annual standard deviation of logreturn LUV is 0.3563
The skewness of logreturn LUV is -0.6211
The kurtosis of logreturn LUV is 9.7802
```

```
The annual mean of log return ALK
1103.431470357093%
The mean of logreturn ALK is 0.0004
The median of logreturn ALK is 0.0006
The standard deviation of logreturn ALK is 0.0291
The annual mean of logreturn ALK is 11.0343
The annual standard deviation of logreturn ALK is 0.4605
The skewness of logreturn ALK is -0.3531
The kurtosis of logreturn ALK is 13.2950
```

```
The annual mean of log return HA
1019.862755498049%
The mean of logreturn HA is 0.0004
The median of logreturn HA is 0.0000
The standard deviation of logreturn HA is 0.0429
The annual mean of logreturn HA is 10.1986
The annual standard deviation of logreturn HA is 0.6779
The skewness of logreturn HA is -4.3875
The kurtosis of logreturn HA is 138.6983
```

```
The annual mean of log return ^SPX
598.3800600681963%
The mean of logreturn SPX is 0.0002
The median of logreturn SPX is 0.0007
The standard deviation of logreturn SPX is 0.0124
The annual mean of logreturn SPX is 5.9838
The annual standard deviation of logreturn SPX is 0.1953
The skewness of logreturn SPX is -0.4421
The kurtosis of logreturn SPX is 12.0776
```

Statistics in 10 stocks and S&P500 in Python

As is the statistics information of each stock shown above, the highest annual mean logarithm return is ADBE, 15.8663. which is in the technology sector. The securities of 10 stocks in the technology sector and industries sector have higher annual mean logarithm returns than the securities in the finance sector. The annual mean logarithm return of Citigroup Inc. is -9.3635. If skewness is positive, the average magnitude of positive deviations is larger than the average magnitude of negative deviations. Other than that, the skewness of Wells Fargo Company is 0.7072, and the rest stocks' skewness is negative. It means that the mean logarithm return of securities is less than the median and mode of securities. The average magnitude of negative deviation is larger than the average magnitude of positive deviations. The most securities with negative skewness have the most extreme values and are found further to the left. In addition, the kurtosis of a normal distribution is 3.0, so a fat-tailed distribution has a kurtosis above 3 and a thin-tailed distribution has a kurtosis below 3.0. A return distribution with positive excess kurtosis, or fat-tailed return distribution, has extremely large deviations from the mean more frequently than a normal distribution. Most stock return series have been found to have fat tails, with a high probability of very bad or very good results. Although 10 stocks all have fat tails, the kurtosis above 3, the highest kurtosis of Hawaiian Holdings, Inc (HA) is 138.6983. Thus, Hawaiian Holdings, Inc. is easy to appear a high probability of very bad or very good results. The annual standard deviation of Hawaiian Holdings, Inc. is 0.6779, the highest of 10 stocks.

4.2 Comparison with Gaussian distribution

The test has used empirical data to compare to the Gaussian distribution, which is so-called the Normal distribution, a "bell-shaped" curve. It is a probability distribution that is symmetric about its mean, meaning that values close to the mean occur more frequently than values far from the mean.

The Gaussian probability density function:

$$G(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Here x is a random variable, which is a price change $p(t+\tau)-p(t)$ in our case, and μ and σ are parameters, for which, as you can check the following equalities are true:

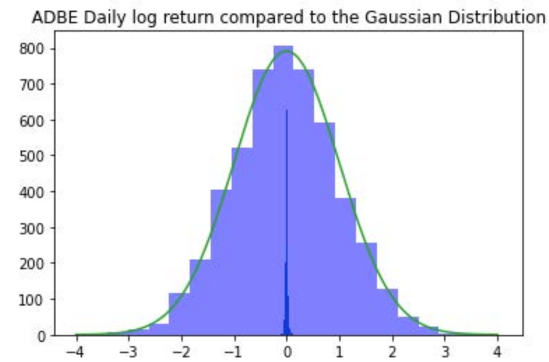
$\langle x\rangle=\mu$ and $\langle x^2\rangle=\sigma^2$ whereby $\langle\ldots\rangle$, I have denoted averaging over a PDF.

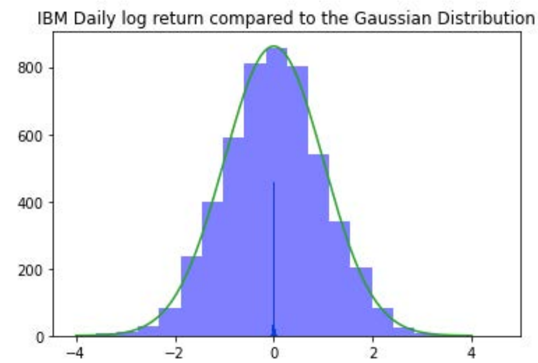
I chose to histogram an empirically given to random variable y (for example, using an

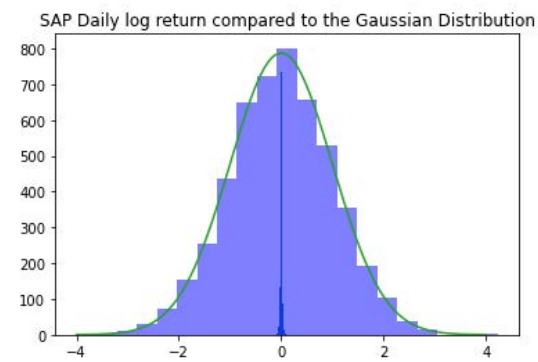
Histogram tool in Python). For that I need to select an equidistant range of bins, separated with step Δ. Then the PDF value obtained from the numerical histogram measurements can be calculated as follows:

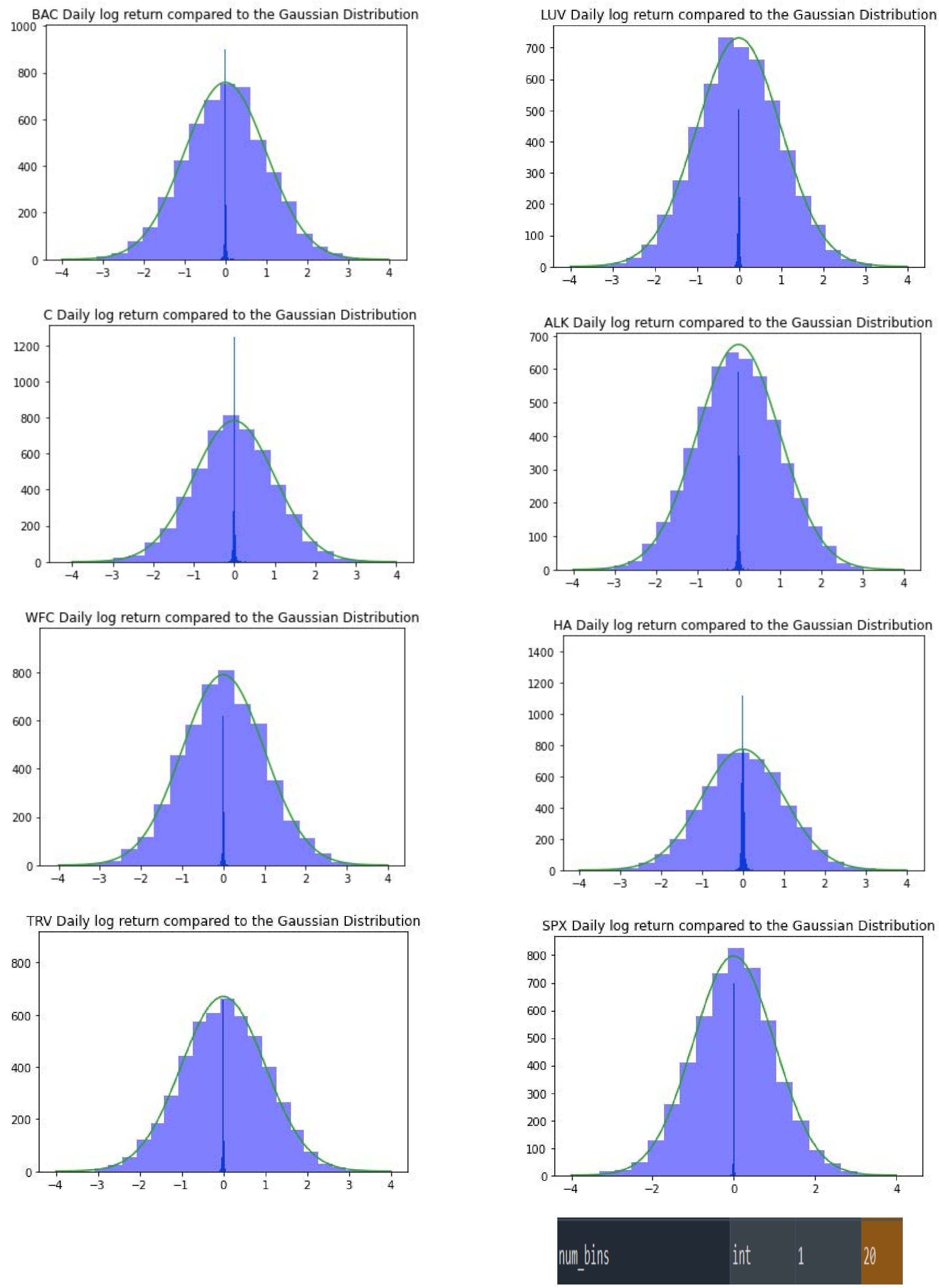
$$P(x_i)=\frac{1}{\Delta}\frac{N_i}{N}$$

where i is the bin number, $N\ i$ is the count of values in bin i, and $N=\sum_{i=1}^{M}N_i$, where M is the number of bins. This empirical PDF needs to be calculated as a function of $\hat{x}_i=x_i-\frac{\Delta}{2}$, as the center of bin having $x\ i$ as its larger endpoint.





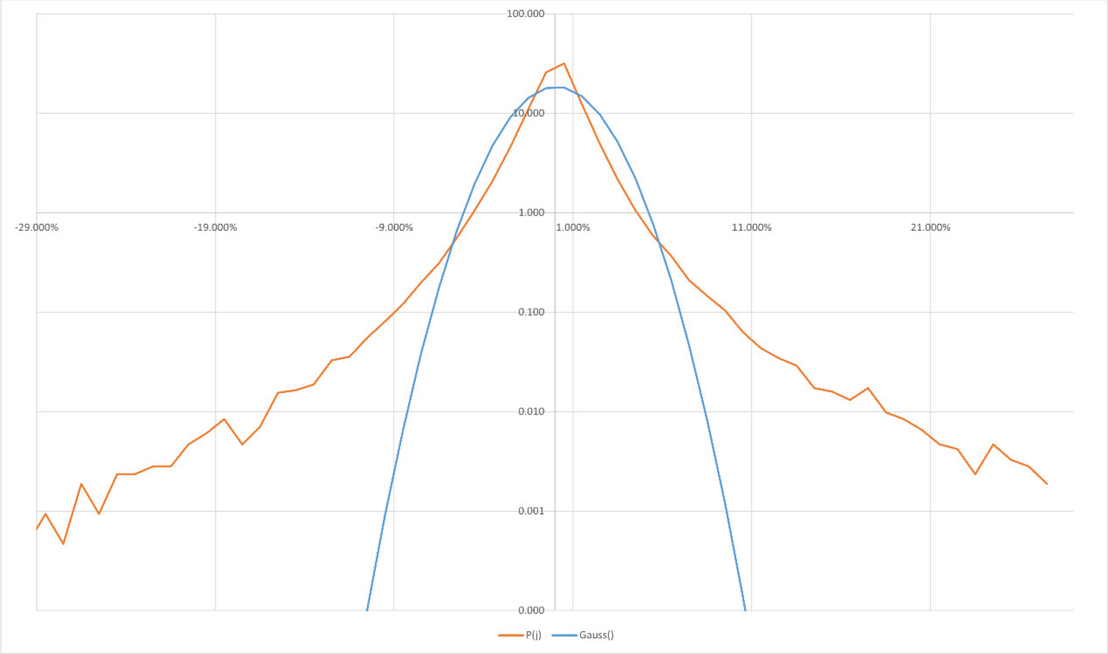




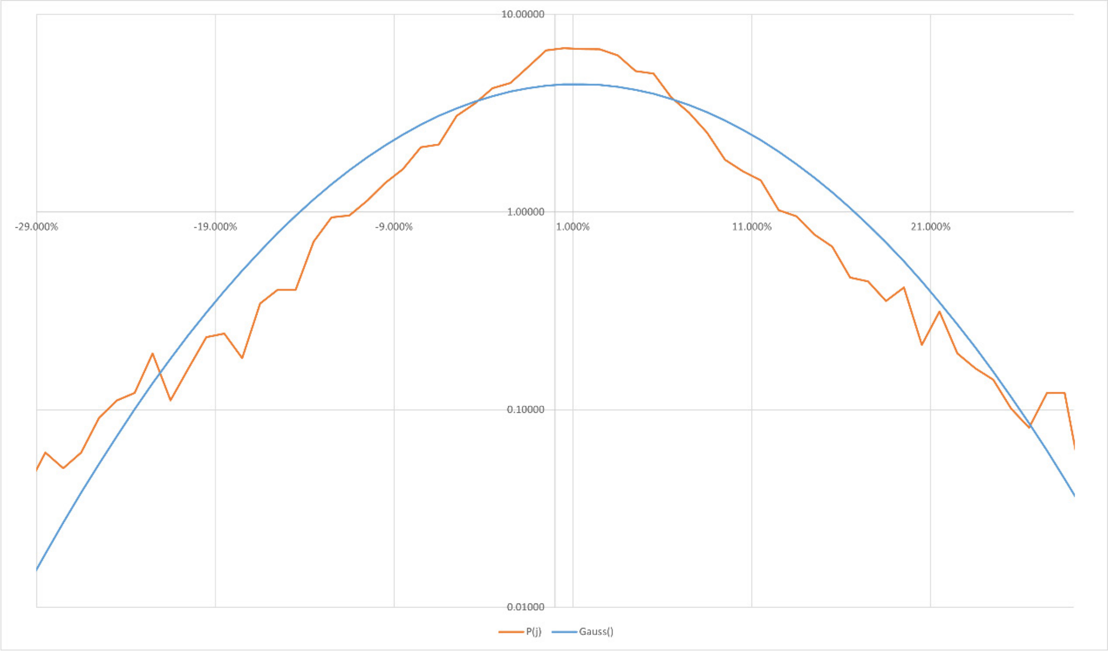
Daily logarithm return compared to the Gaussian in Python

The logarithmic return of securities are dark blue histogram. The purple histogram is described as an abnormal distribution. I set 20 bins in the histogram and the range is (-4,4) (Figure). The line is followed by a normal distribution. From the figure, the stocks aren't suitable for the normal distribution. The highest frequency of stocks occurs in the center of the distribution.

Here is the Daily Data compared to the Gaussian and Monthly Data compared to the Gaussian in Excel.



Daily Data compared to the Gaussian in Excel

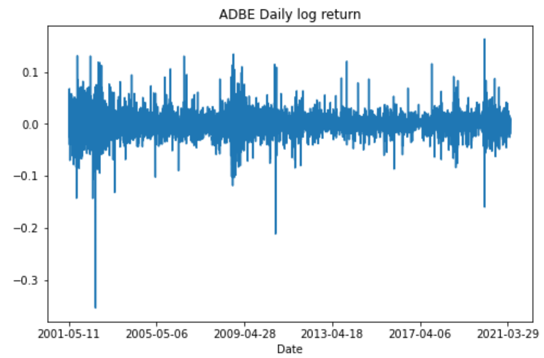


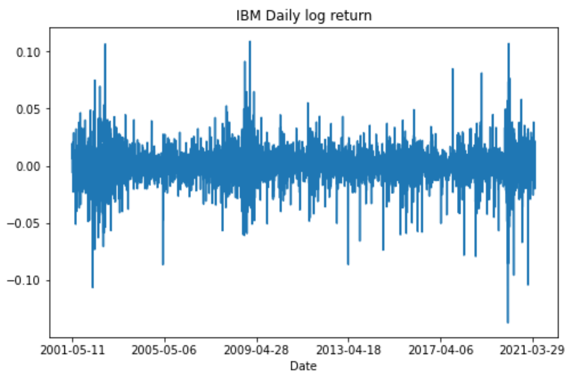
Monthly Data compared to the Gaussian in Excel

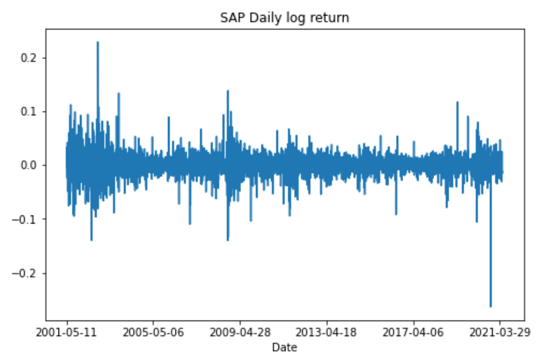
Monthly data is more suitable for the Gaussian function than daily data. Therefore, using monthly data as an analysis tool in the future is a good choice.

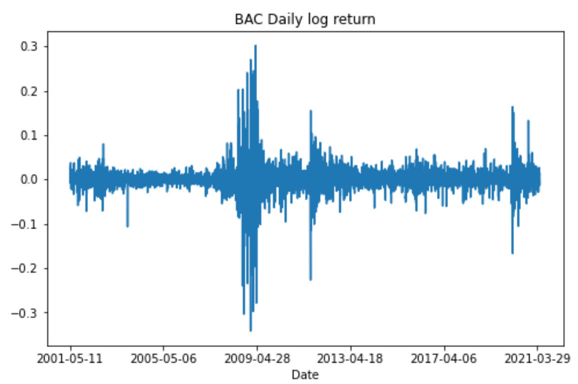
4.3 Illustration of Systematic risk

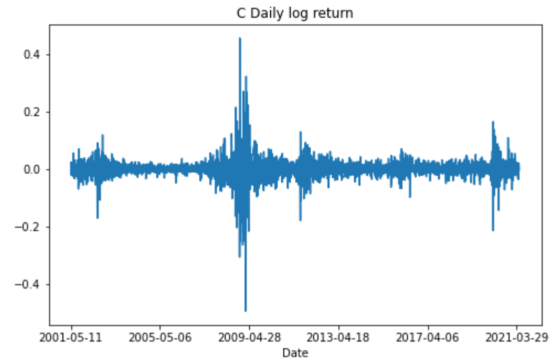
I calculate the daily logarithmic yield of a security to show the plots of each security and compare them the Market Index (S&P 500). The corresponding graph of the daily logarithmic yield of securities is listed below:

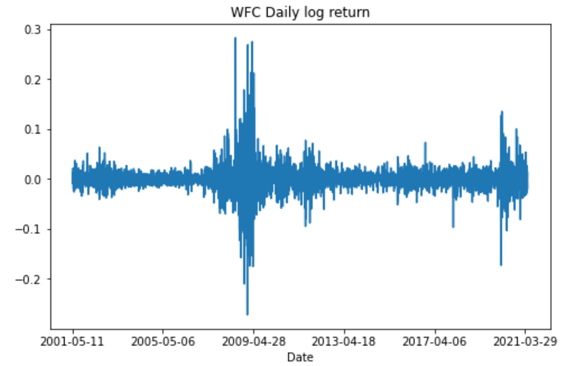


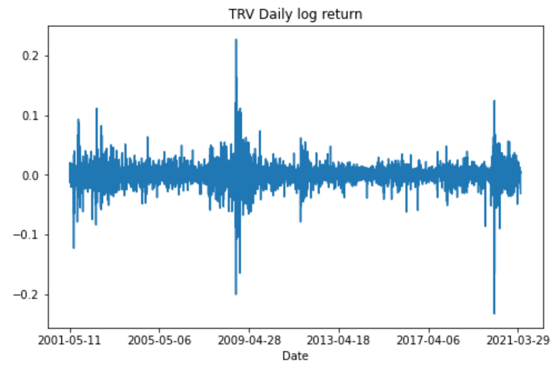


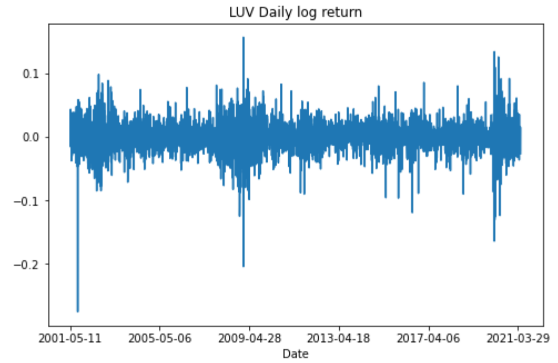


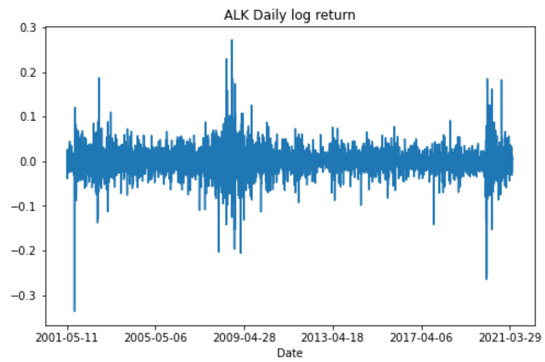


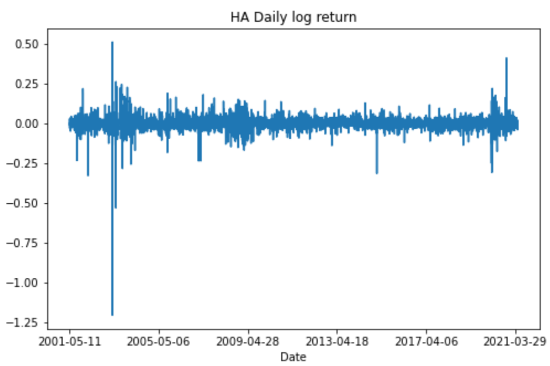


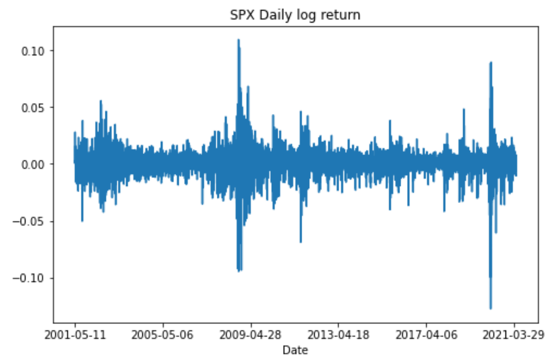












The daily logarithmic yield of securities in Python

According to the graphs above, I observe that the SPX experienced significant fluctuations in certain years, such as 2008 and 2020. Building on this point, I examined the companies (SAP, BAC, C, WFC, TRV, LUV, ALK) that are influenced by market risk, also known as systematic risk. The macroeconomic factors include conditions of the general economy, the business cycle, inflation, interest rates, exchange rates, and so on. The financial crisis of 2008 and the COVID-19 pandemic of 2020 led to a significant wave of understanding in the market during that time.

4.4 Correlation Test

After testing individual stocks' statistics, first, I check the correlation of the 10 stocks and one risk-free rate's daily logarithm return. The numbers 0-10 correspond to these stocks in order:['ADBE','IBM','SAP','BAC','C','WFC', 'TRV','LUV','ALK',' HA','SPX']

The heat map of the correlation test by daily logarithm return:

	0	1	2	3	4	5	6	7	8	9	10
0	1	0.461725	0.500249	0.375422	0.389489	0.360836	0.391822	0.332149	0.32644	0.198226	0.638609
1	0.461725	1	0.524637	0.425697	0.451489	0.448228	0.46372	0.417044	0.388465	0.220351	0.693668
2	0.500249	0.524637	1	0.378174	0.410604	0.382047	0.410099	0.371286	0.353285	0.193822	0.636594
3	0.375422	0.425697	0.378174	1	0.808558	0.820309	0.528836	0.402877	0.421363	0.286942	0.683411
4	0.389489	0.451489	0.410604	0.808558	1	0.73392	0.51653	0.40221	0.424032	0.295426	0.683201
5	0.360836	0.448228	0.382047	0.820309	0.73392	1	0.553328	0.431437	0.454254	0.306541	0.687247
6	0.391822	0.46372	0.410099	0.528836	0.51653	0.553328	1	0.426178	0.431098	0.238431	0.68503
7	0.332149	0.417044	0.371286	0.402877	0.40221	0.431437	0.426178	1	0.655426	0.358664	0.555237
8	0.32644	0.388465	0.353285	0.421363	0.424032	0.454254	0.431098	0.655426	1	0.382687	0.543892
9	0.198226	0.220351	0.193822	0.286942	0.295426	0.306541	0.238431	0.358664	0.382687	1	0.325914
10	0.638609	0.693668	0.636594	0.683411	0.683201	0.687247	0.68503	0.555237	0.543892	0.325914	1

Correlations in Python

The heat map of the covariances test

	0	1	2	3	4	5	6	7	8	9	10
0	0.000562605	0.000168006	0.000245884	0.000258007	0.00028756	0.000208178	0.000168407	0.000177548	0.000225539	0.000201604	0.000187162
1	0.000168006	0.000235329	0.000166778	0.000189212	0.000215584	0.000167248	0.000128903	0.000144179	0.000173582	0.000144941	0.000131483
2	0.000245884	0.000166778	0.00042942	0.000227061	0.000264847	0.000192567	0.000153992	0.000173393	0.000213246	0.000172219	0.000162998
3	0.000258007	0.000189212	0.000227061	0.000839495	0.000729209	0.000578108	0.000277651	0.000263065	0.000355616	0.000356484	0.000244665
4	0.00028756	0.000215584	0.000264847	0.000729209	0.000968864	0.000555652	0.000291337	0.00028214	0.000384455	0.000394291	0.00026276
5	0.000208178	0.000167248	0.000192567	0.000578108	0.000555652	0.000591624	0.000243879	0.000236495	0.000321838	0.000319704	0.000206546
6	0.000168407	0.000128903	0.000153992	0.000277651	0.000291337	0.000243879	0.000328351	0.000174037	0.000227541	0.000185254	0.000153376
7	0.000177548	0.000144179	0.000173393	0.000263065	0.00028214	0.000236495	0.000174037	0.000507882	0.00043025	0.000346582	0.000154611
8	0.000225539	0.000173582	0.000213246	0.000355616	0.000384455	0.000321838	0.000227541	0.00043025	0.000848459	0.000477964	0.000195753
9	0.000201604	0.000144941	0.000172219	0.000356484	0.000394291	0.000319704	0.000185254	0.000346582	0.000477964	0.00183854	0.000172671
10	0.000187162	0.000131483	0.000162998	0.000244665	0.00026276	0.000206546	0.000153376	0.000154611	0.000195753	0.000172671	0.000152672

Covariances in Python

All correlations and covariances are positive. It illustrates that 10 stocks move in the same direction. It's hard to offset relative risk.

portfolio_return_annual	float64	1	0.07101057502531007
portfolio_std	float64	1	0.27899633668222246
portfolio_variance	float64	1	0.07783895588210002
portfolio_stocks_AVGcorrc…	float64	1	0.5072257581380388
portfolio_stocks_MAXcorrc…	float64	1	1.0
portfolio_stocks_MINcorrc…	float64	1	0.19382167619793036

Statistics of completed portfolio in Python

Second, I use the monthly excess return to calculate the correlation and some statistics. Excess Return = simple return – average return. The raw data is total return which means the price is adjusted by dividends.

	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA
SPX	100.0%	66.5%	64.9%	64.9%	60.2%	70.2%	55.5%	59.8%	53.7%	46.4%	39.0%
ADBE	66.5%	100.0%	45.5%	53.4%	42.3%	46.3%	29.8%	45.2%	38.8%	23.3%	18.0%
IBM	64.9%	45.5%	100.0%	58.5%	31.3%	42.0%	26.7%	38.2%	34.7%	35.7%	24.6%
SAP	64.9%	53.4%	58.5%	100.0%	33.1%	43.4%	29.8%	37.5%	31.8%	28.2%	14.4%
BAC	60.2%	42.3%	31.3%	33.1%	100.0%	82.6%	76.1%	39.3%	42.8%	27.5%	33.8%
C	70.2%	46.3%	42.0%	43.4%	82.6%	100.0%	70.3%	51.2%	42.8%	30.4%	34.3%
WFC	55.5%	29.8%	26.7%	29.8%	76.1%	70.3%	100.0%	34.5%	40.6%	34.7%	35.8%
TRV	59.8%	45.2%	38.2%	37.5%	39.3%	51.2%	34.5%	100.0%	40.7%	36.0%	24.0%
LUV	53.7%	38.8%	34.7%	31.8%	42.8%	42.8%	40.6%	40.7%	100.0%	51.9%	42.2%
ALK	46.4%	23.3%	35.7%	28.2%	27.5%	30.4%	34.7%	36.0%	51.9%	100.0%	40.4%
HA	39.0%	18.0%	24.6%	14.4%	33.8%	34.3%	35.8%	24.0%	42.2%	40.4%	100.0%

Correlations in Excel

	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA
Annual Average Return	7.5%	19.6%	4.8%	12.0%	11.1%	1.0%	8.9%	9.1%	9.8%	17.4%	26.9%
Annual StDev	14.9%	31.8%	23.2%	33.9%	39.3%	42.5%	28.1%	20.0%	31.8%	37.7%	62.1%
beta	1.0000	1.4228	1.0137	1.4829	1.5955	2.0067	1.0519	0.8033	1.1496	1.1779	1.6293
alpha	0.0000	0.0885	-0.0289	0.0081	-0.0093	-0.1410	0.0095	0.0301	0.0118	0.0855	0.1458
residual Stdev	0.0%	23.8%	17.6%	25.8%	31.4%	30.3%	23.4%	16.0%	26.8%	33.4%	57.2%

Statistics of completed portfolio in Excel

4.5 Markowitz Model with Constraints

1) Constraint 1

The so-called constraints are the weight restrictions. Constraint 1 is that the absolute value of the sum of the stock's weights is less than or equal to 2. This additional optimization constraint is intended to permit broker-dealers to allow their customers to hold positions where 50% or more of the position is funded by the customer's account equity:

$$\sum_{i=1}^{11} |w_i| \le 2$$

MM (Constr1):	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA	Return	StDev	Sharpe			
MinVar	110.97%	-9.68%	5.14%	-9.90%	0.35%	-22.54%	14.06%	19.45%	-0.13%	-4.86%	-2.85%	6.72%	11.75%	0.572	CAL:	0.0%	0.0%
MaxSharpe	39.54%	28.26%	-0.10%	0.29%	16.67%	-48.07%	19.80%	30.14%	-1.80%	7.70%	7.57%	17.59%	17.70%	0.994	4.5	79.2%	79.6%

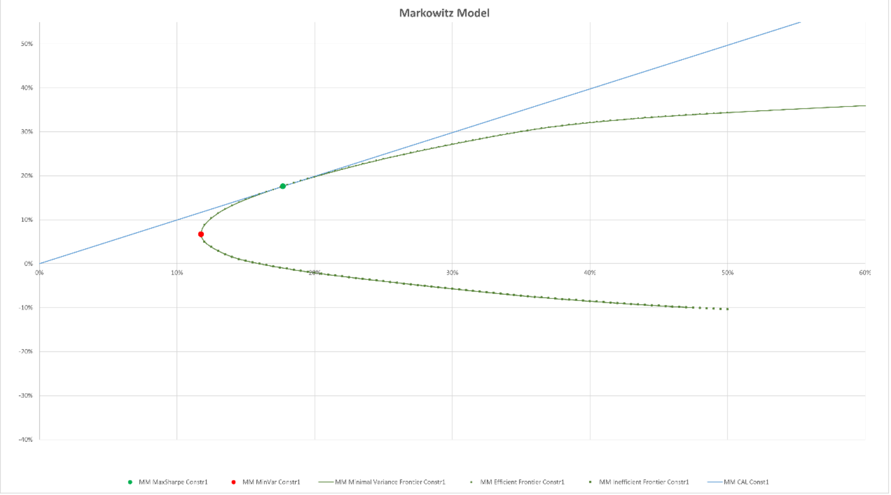


Figure 1

Figure 1 above states that the stocks' weights in consideration of constraint 1. There are two considerations: Minimum Variance (MinVar) and Maximum Sharpe (MaxSharpe). In MinVar condition, the minimum standard deviation is 11.75% with a return of 6.72%. The Sharpe ratio is 0.572. The lowest risk is the red point on the MM efficient Frontier in Figure 1. In MaxSharpe condition, the maximum Sharpe ratio is 0.994 with a return of 17.59% and standard a deviation of 17.7%. In Figure 1, the highest Sharpe ratio is the blue point passing the capital allocation line (MM CAL Constr1).

2) Constraint 2

Constraint 2 is that the absolute value of the sum of the stock's weights is less than or equal to 1. This additional optimization constraint is designed to simulate some arbitrary "box" constraints on weights, which may be provided by the client: $|w_i| \le 1, \text{ for } \forall i$

MM (Constr2):	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA	Return	StDev	Sharpe			
MinVar	100.00%	-8.53%	8.25%	-9.24%	0.66%	-23.03%	16.19%	22.93%	0.28%	-4.76%	-2.75%	6.97%	11.79%	0.591	CAL:	0.0%	0.0%
MaxSharpe	50.09%	35.81%	-22.32%	7.19%	29.74%	-66.14%	20.56%	35.57%	-14.83%	13.45%	10.87%	22.07%	21.33%	1.035	4.5	99.3%	96.0%

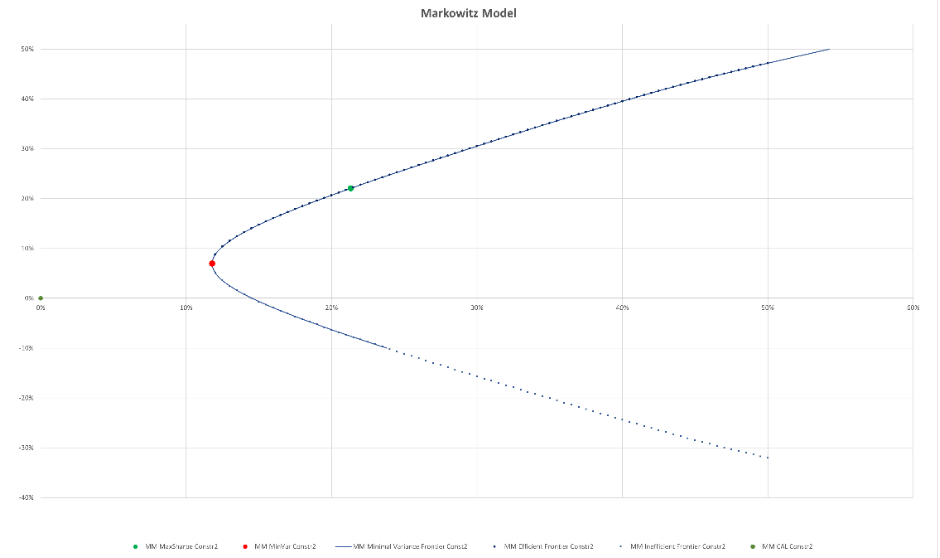


Figure 2

In MinVar condition, the minimum standard deviation is 11.79% with a return of 6.97%.
The Sharpe ratio is 0.591. The lowest risk is the red point on the MM efficient Frontier in Figure 2. In MaxSharpe condition, the maximum Sharpe ratio is 1.035 with a return of 22.07% and a standard deviation of 21.33%. In Figure 1, the highest Sharpe ratio is the blue point.

3) Constraint 3

The constraint 3 is the Markowitz Model because it has no constraints, which solves a “free” problem, to illustrate how the area of permissible portfolios in general and the efficient frontier in particular.

MM (Constr3):	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA	Return	StDev	Sharpe			
MinVar	111.46%	-9.91%	5.27%	-10.07%	0.60%	-22.81%	14.06%	19.53%	-0.26%	-4.98%	-2.88%	6.69%	11.75%	0.570	CAL:	0.0%	0.0%
MaxSharpe	50.09%	35.81%	-22.32%	7.19%	29.74%	-66.14%	20.56%	35.57%	-14.83%	13.45%	10.87%	22.07%	21.33%	1.035	4.5	99.3%	96.0%

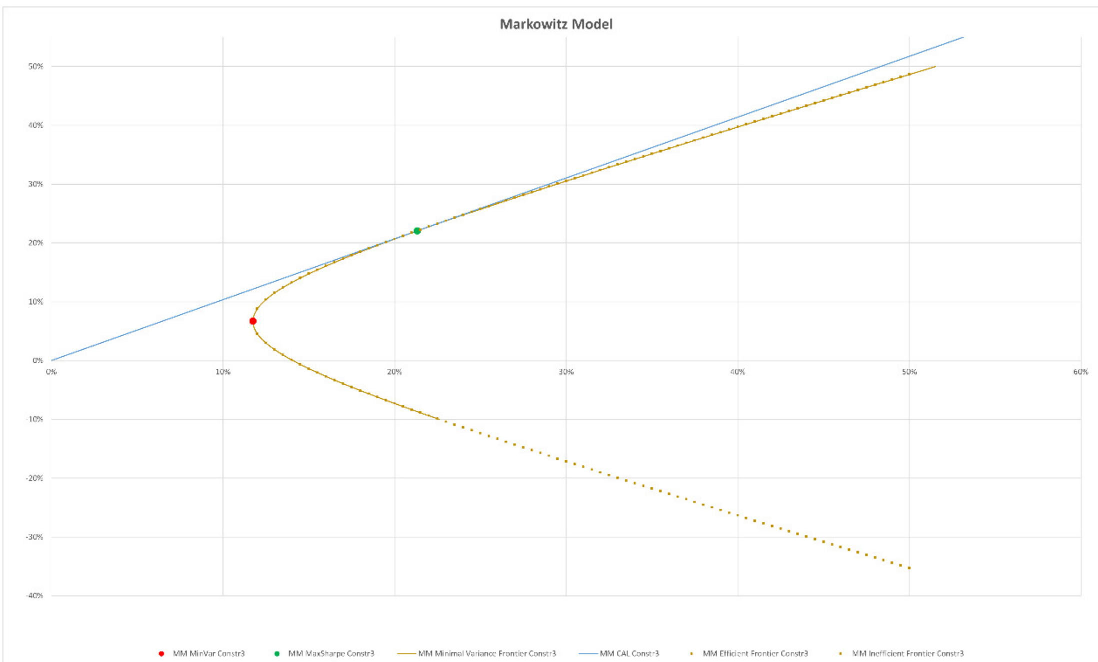


Figure 3

In MinVar condition, the minimum standard deviation is 11.75% with a return of 6.69%.
The Sharpe ratio is 0.57. The lowest risk is the red point on the MM efficient Frontier in Figure 3. In MaxSharpe condition, the maximum Sharpe ratio is 1.035 with a return of 22.07% and a standard deviation of 21.33%. In Figure 3, the highest Sharpe ratio is the blue point. The results MaxSharpe in Constraint 3 are the same as in Constraint 2.

4.6 Index Model with Constraints

The Index Model with constraints is different from the MM with constraints in three constraints.

1) Constraint 1

In MinVar condition, the minimum standard deviation is 11.96% with a return of 6.07%. The Sharpe ratio is 0.508. The lowest risk is the red point on the MM efficient Frontier in Figure 4. In MaxSharpe condition, the maximum Sharpe ratio is 0.898 with a return of 18.90% and a standard deviation of 21.04%. In Figure 4, the highest Sharpe ratio is the blue point passing the capital allocation line (IM CAL Constr1).

IM (Constr1):	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA	Return	StDev	Sharpe			
MinVar	138.69%	-9.85%	0.03%	-9.68%	-8.19%	-15.39%	-0.28%	11.23%	-2.17%	-1.77%	-2.62%	6.07%	11.96%	0.508	CAL:	0.0%	0.0%
MaxSharpe	48.18%	35.88%	-14.54%	2.04%	-0.61%	-34.85%	3.97%	28.36%	3.60%	17.73%	10.24%	18.90%	21.04%	0.898	2.5	47.3%	52.6%

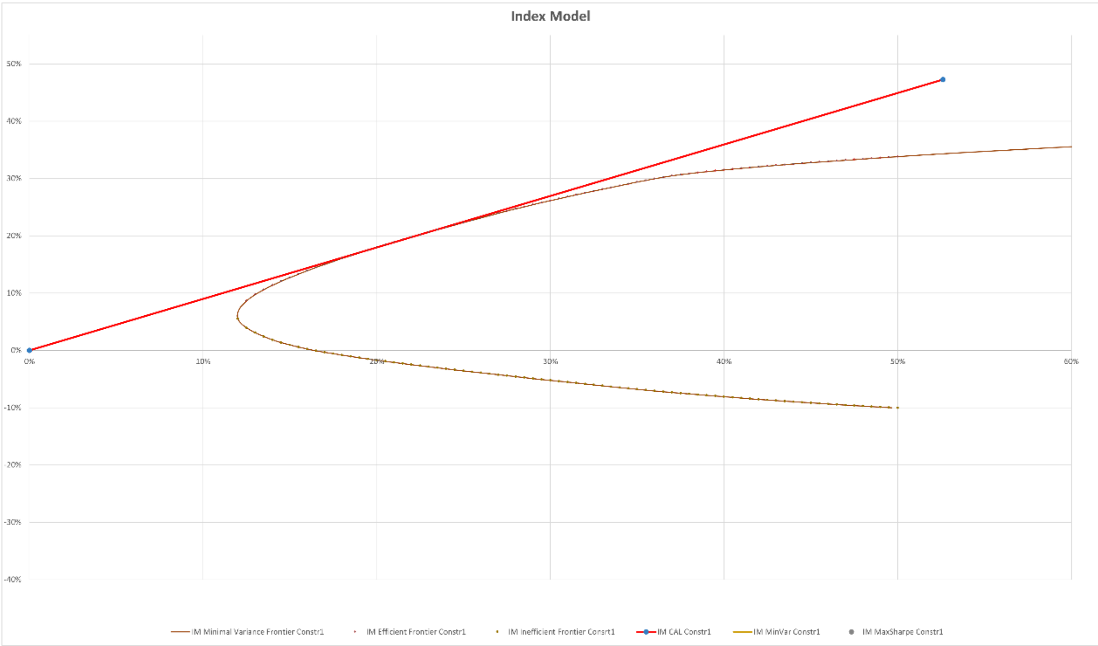


Figure 4

2) Constraint 2

In MinVar condition, the minimum standard deviation is 12.47% with a return of 6.90%. The Sharpe ratio is 0.553. The lowest risk is the red point on the MM efficient Frontier in Figure 5. In MaxSharpe condition, the maximum Sharpe ratio is 0.901 with a return of 19.81% and a standard deviation of 21.99%. In Figure 1, the highest Sharpe ratio is the blue point.

IM (Constr2):	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA	Return	StDev	Sharpe			
MinVar	100.00%	-6.58%	8.50%	-6.99%	-6.49%	-13.97%	3.74%	23.09%	0.74%	0.08%	-2.12%	6.90%	12.47%	0.553	CAL:	0.0%	0.0%
MaxSharpe	54.94%	38.26%	-22.71%	2.98%	-2.31%	-37.58%	4.25%	28.63%	4.00%	18.66%	10.89%	19.81%	21.99%	0.901	2.5	49.5%	55.0%

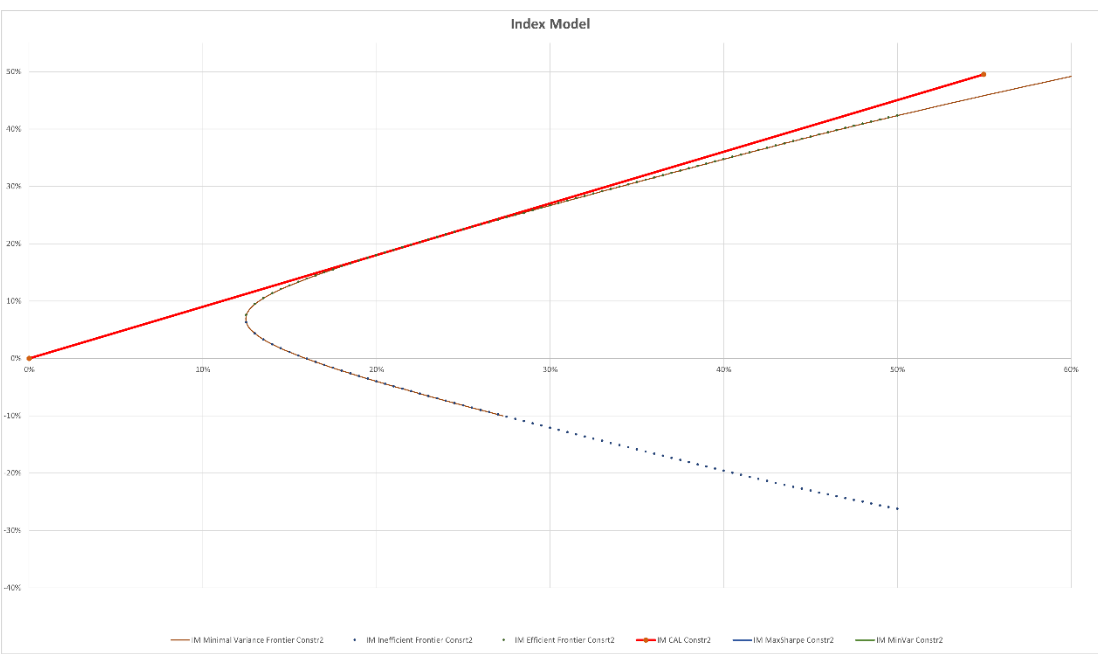


Figure 5

3) Constraint 3

In MinVar condition, the minimum standard deviation is 11.95% with a return of 5.85%. The Sharpe ratio is 0.49. The lowest risk is the red point on the MM efficient Frontier in Figure 3. In MaxSharpe condition, the maximum Sharpe ratio is 0.901 with a return of 19.81% and a standard deviation of 21.99%. In Figure 6, the highest Sharpe ratio is the green point. The results MaxSharpe in Constraint 3 are the same as in Constraint 2.

IM (Constr3):	SPX	ADBE	IBM	SAP	BAC	C	WFC	TRV	LUV	ALK	HA	Return	StDev	Sharpe			
MinVar	144.39%	-10.70%	-0.63%	-10.37%	-8.62%	-15.69%	-1.35%	10.96%	-2.97%	-2.27%	-2.75%	5.85%	11.95%	0.490	CAL:	0.0%	0.0%
MaxSharpe	54.94%	38.26%	-22.71%	2.98%	-2.31%	-37.58%	4.25%	28.63%	4.00%	18.66%	10.89%	19.81%	21.99%	0.901	2.5	49.5%	55.0%

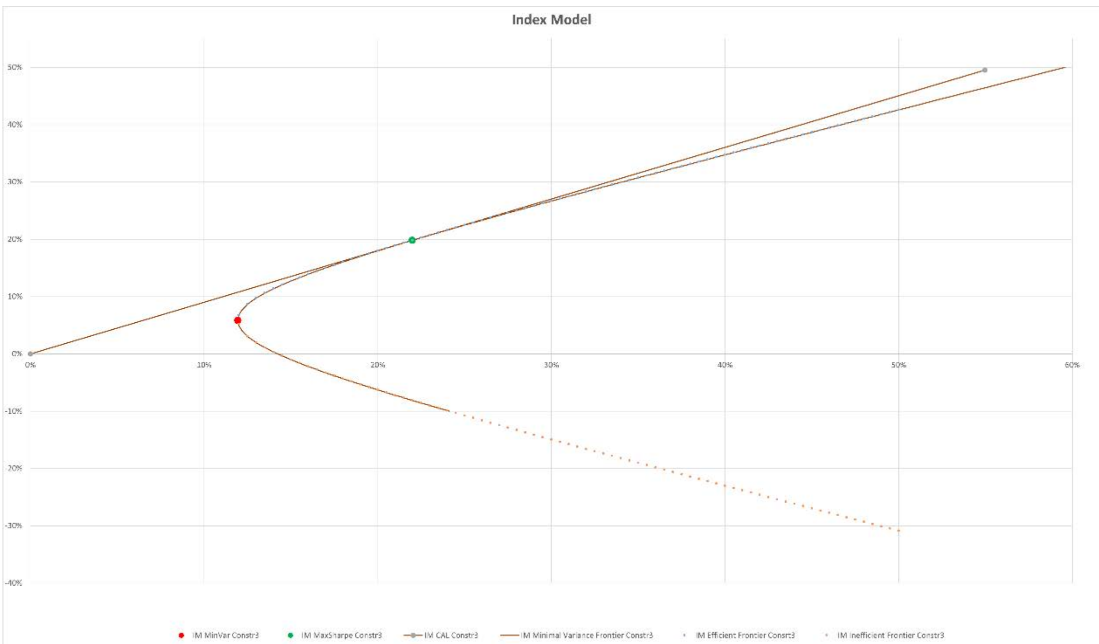


Figure 6

4.7 Constraints Comparison in MM and IM

4.7.1 MM

The optimal CAL is the portfolio with Constraint 3. If the investor is risk-aversion, Constraint 1 will be shown in the future and the Sharpe ratio will be low. However, Constraint 1 has a lower risk than two other constraints. If the investors are eager to maximize the Sharpe ratio, I suggest helping clients choose Constraints 2&3. Constraints 2&3 have the same results in MaxSharpe. In MinVar, Constraint 2 is the best. I think the Constraints 2&3 are suitable for investors who can accept the risks. Thus, what kind of performance depends on the investor's risk preference and pursuit of returns.

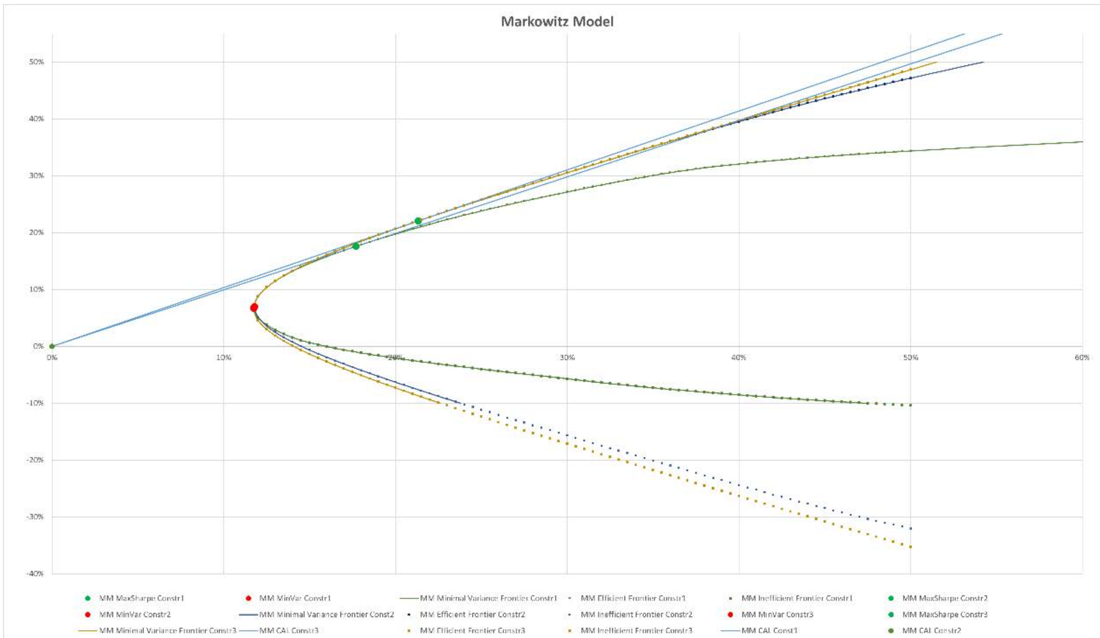


Figure 7

4.7.2 IM

The three CALs almost overlap in Figure 9. The Efficient Frontier of Constraints 2&3 is very close. The investors with risk-averse will be suggested to choose Constraint 1. The Efficient Frontier of Constrain 1 gradually tends to be constant. Constraints 2&3 has the same results in MaxSharpe. However, Constraint 3 in MinVar performs the worst in all conditions. Under similar risks (11.95%-11.96%) in MinVar of Constraint 3, I recommend choosing Constraint 1, which has a higher return. In MinVar, I think Constraint 2 is better than two other constraints for the highest Sharpe Ratio in three conditions.

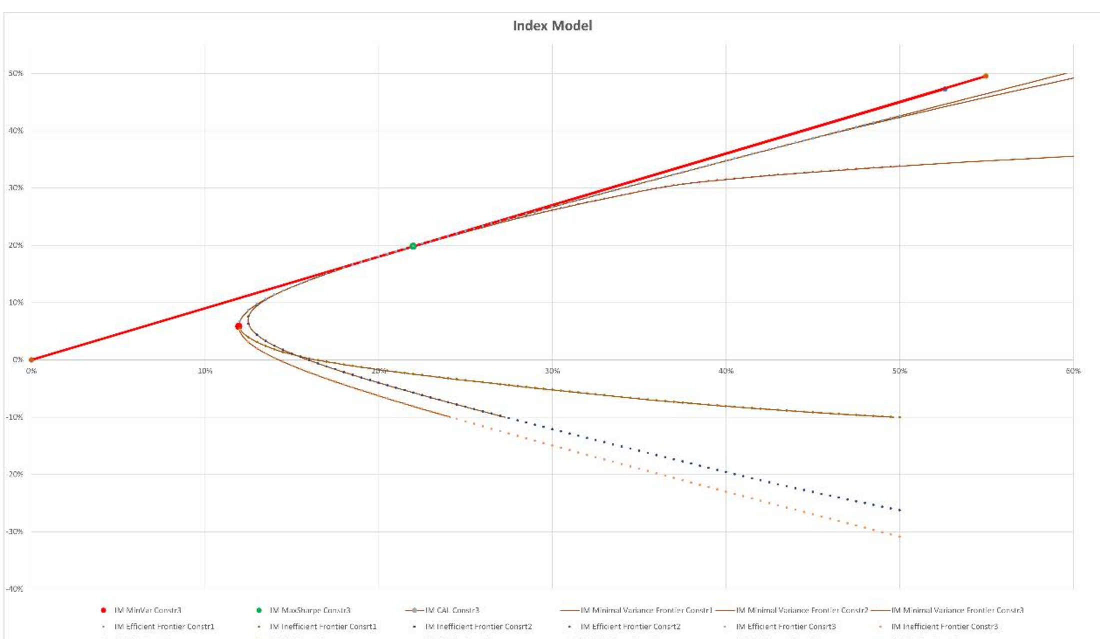


Figure 9

4.7.3 MM & IM

In the Figures below, the figures are contrasted in MM and IM with Constrain 1&2&3. There is no obvious difference between MM and IM with Constrain 1. On the contrary, the visible difference between MM and IM with Constrain 1&2. The IM has a lower risk than MM. MM has a higher return and Sharp Ratio in three conditions. According to Figure 12, If stocks with correlated residuals have high alphas, then the index model may lead to a worse portfolio than the full Markowitz model.

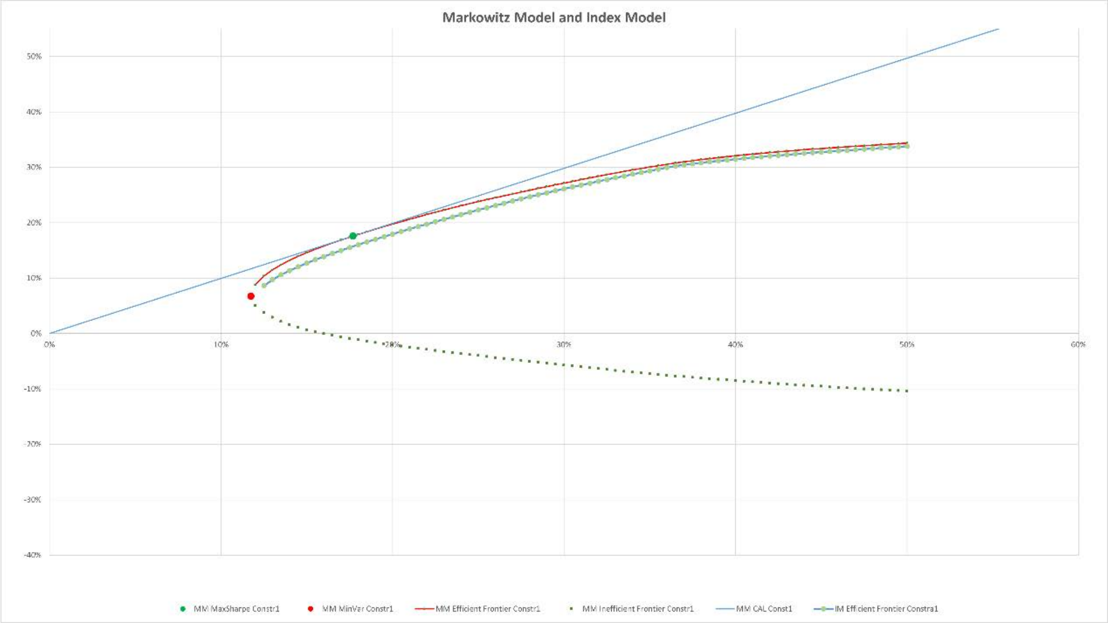


Figure 10

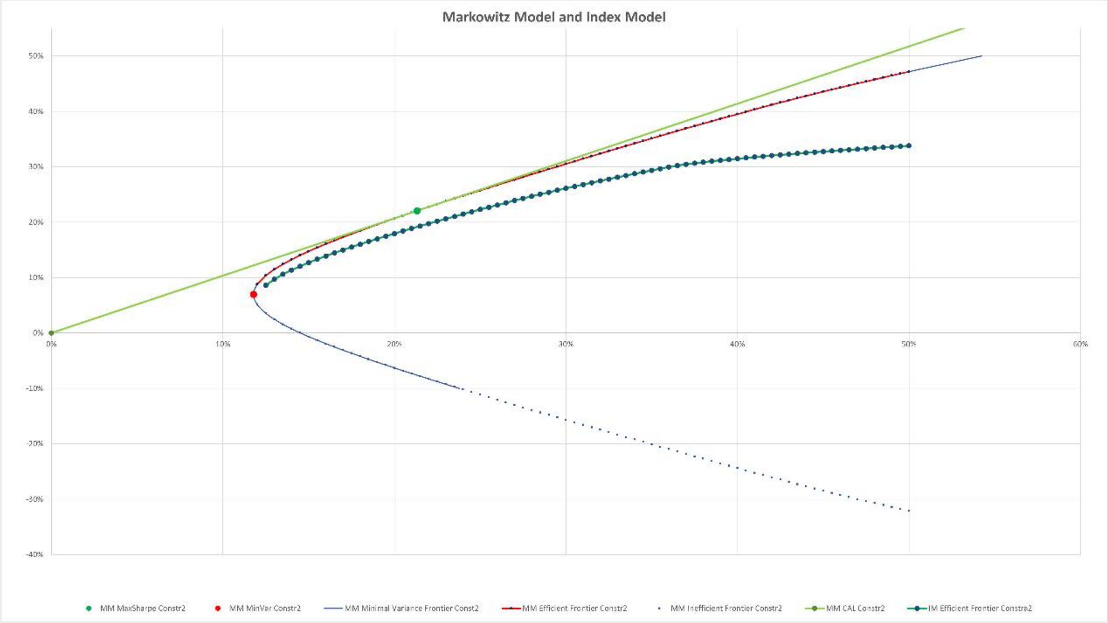


Figure 11

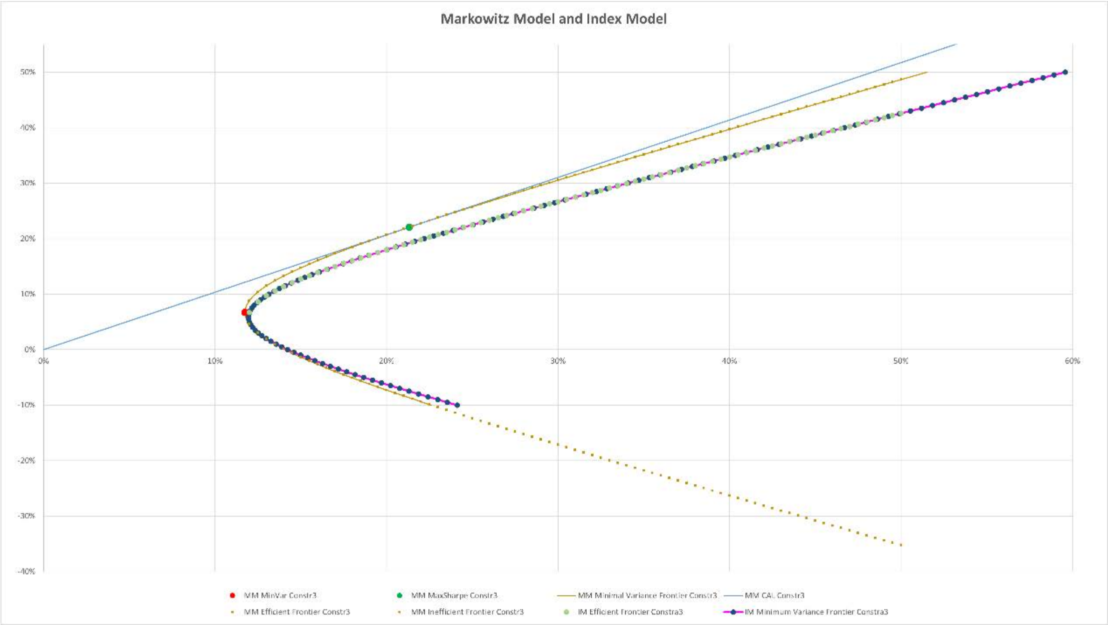
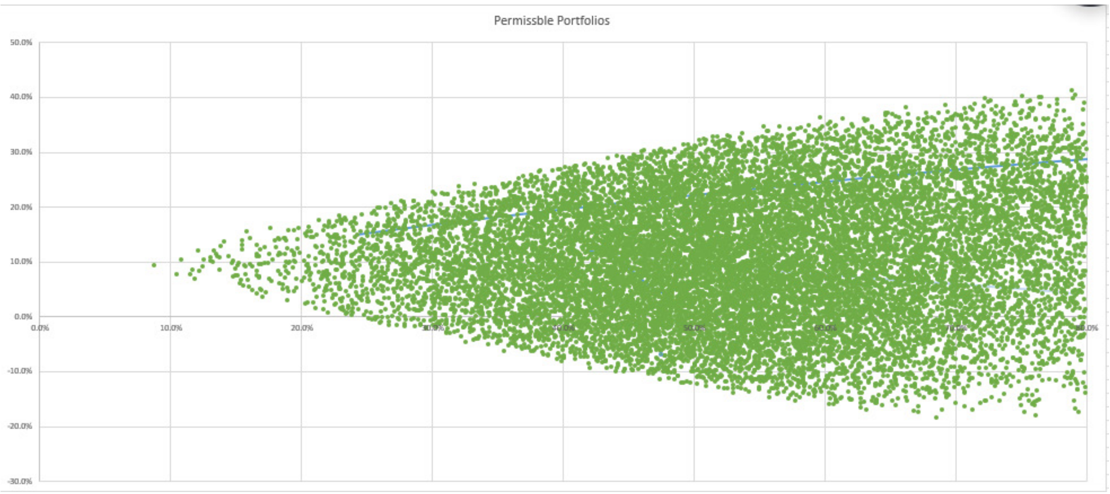


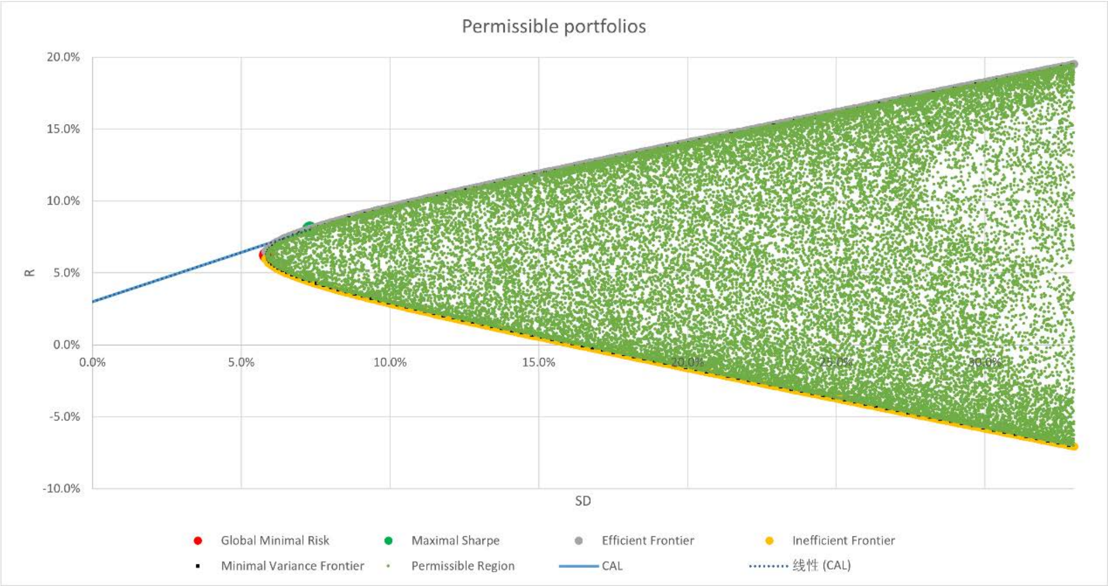
Figure 12

5. Monte Carlo Simulation

Monte Carlo simulation is an important tool with a wide range of applications in business and finance. When we run a Monte Carlo simulation, we are interested in observing different possible realizations of future events. These realizations are generated by analyzing the distribution of historical data and calculating its mean and variance. Monte Carlo simulations are used in corporate finance, investment valuation, asset management, risk management, insurance liability estimation, option pricing, and other derivatives. The significant uncertainty in finance makes Monte Carlo simulations a valuable tool for improving the decision-making process when several random variables are at play.

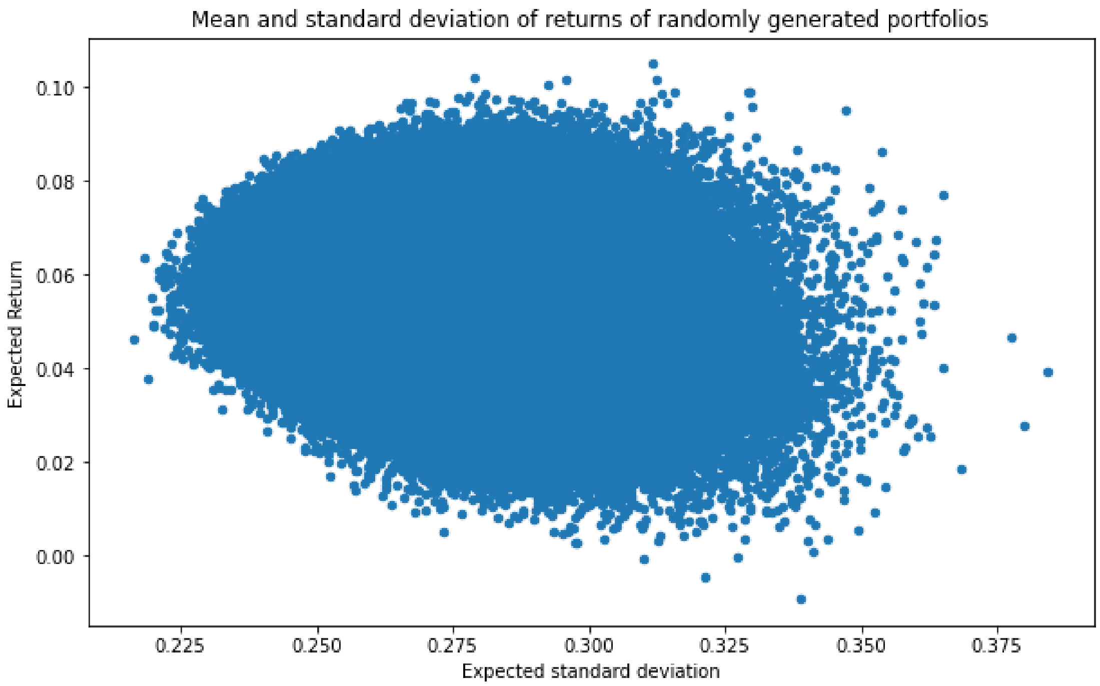
In Excel, I random weights of 10 stocks and give the 50,000 standard deviations and returns. By pulling down tables to 50,000 rows, I get the Permissible Portfolios.





Permissible Portfolios in Excel

In Python, I put all the data of optimal portfolio of stocks and the daily logarithm return of securities. I give the *for* loop and random weights in 1000000. It forms the graph of “Mean and standard deviation of returns of randomly generated portfolios”. The Minimum Variance Frontier of the graph gives Max return and Min risk in randomly generated portfolios in Python below.



Permissible Portfolios in Python

MM_Maxreturn	float64	1	0.10837011552275459
MM_Minstd	float64	1	0.21323895723059832

Max return and Min risk in random generated portfolios in Python

6. Conclusion

From the development of model theories to data testing, the data selection problem is addressed under two conditions. When comparing it with the Gaussian distribution, the monthly excess return is considered conventional data for analysis in Excel. However, when I use daily data in Python, it is more extensive and convenient for users to operate on the Monte Carlo with 1,000,000 random daily logarithmic return.

Perform hypothesis testing on the MM (Markowitz Model) and IM (Index Model) with constraints using Excel or Python, and I can conclude that there is no significant difference in the average returns between the Index Model and the Markowitz Model. Conservative investors tend to avoid risks (risk aversion) and can invest their funds in the optimal portfolio of stocks formed using the Index Model, as the given risk level is low. In contrast, aggressive investors, who have a high-risk, high-return profile, are willing to take risks and can invest their funds in the optimal portfolio of stocks formed by the Markowitz model if there are no constraints.

Since these 10 stocks are from different industries, it is challenging to make a comprehensive comparison of companies within the same industry. Therefore, future researchers should focus on research subjects that include companies in the same industry and give priority to those with strong liquidity. Additionally, future studies are expected to utilize different analytical tools, applying the concept of two models.

References

[1] Markowitz H. portfolio selection [J]. The journal of finance, 1952: 7 (1). (Markowitz, 1952)

[2] Sharpe W F. A simplified model for portfolio analysis[J]. Management science, 1963, 9(2): 277-293. (Sharpe, 1963)