5-3-2017

A Strategy of Maximizing Profit in the Long Run Using the Concept of Kelly's Criterion

Binyi Xu
bxx9867@rit.edu

Rochester Institute of Technology

A Strategy of Maximizing Profit in the Long Run Using the Concept of Kelly's Criterion

Binyi Xu

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of Master of Science
in Applied & Computational Mathematics

School of Mathematical Sciences
College of Science

Rochester Institute of Technology
Rochester, NY
May 3, 2017

Committee Approval:

Dr. Bernard Brooks Date
Thesis Advisor

Dr. James Marengo Date
Professor

Dr. Zhijian Huang Date
Assistant Professor

ABSTRACT

One of the fundamental problems of portfolio theory is how to rationally optimize the portfolio using diversification. In practice, maximizing the short term interest is not equivalent to maximizing the long term interest. Kelly's criterion is considered to be the best strategy of maximizing profit in the long run. In this paper, we discussed the applications of Kelly's criterion in various scenarios, including binomial cases, univariate stock, uncorrelated and correlated stocks. Different approaches were introduced to construct the model of stocks' behavior. For the first time, we discussed the feasibility of extending Kelly's criterion to option trading.

Contents

1 Introduction

One of the fundamental problems of portfolio theory is how to rationally optimize the portfolio using diversification. In practice, maximizing the short term interest is not equivalent to maximizing the long term interest. Kelly's criterion is considered to be the best strategy of maximizing profit in the long run. In this paper, we discussed the applications of Kelly's criterion in various scenarios, and shows the feasibility of extend Kelly's criterion to option trading.

In chapter 2, a number of basic concepts are introduced, starting with the basic brownian motion to Black-Scholes model, and then to Heston model. Various estimators of volatility are discussed and compared, from constant volatility to stochastic volatility.

In chapter 3, the idea of Kelly's criterion are explained with its properties being proved. The use of Kelly's criterion is discussed by examples of binomial cases, univariate stock, uncorrelated and correlated stocks.

In chapter 4, the daily close price from 1962 to 2015 of General Electronic, DuPont and IBM are collected. The joint distribution of the retruns of these correlated stocks are built using methods including Taylor's expansion, kernel density estimation, and copula.

In chapter 5, we extend our strategy to option trading, for the first time. A portfolio of size 15×15 is built, by solving a 15-dimensional optimization problem.

2 Stochastic Models

In this chapter, a number of basic concepts are explained, including Wiener process, Itô's Lemma, Black-Scholes-Merton model, Dupire's volatility and Heston model. Some definition of terms, theorems and derivations are based on *"Options, Futures, and Other Derivatives"* [9], and *"Introduction to probability models"* [16].

2.1 Wiener processes

Consider a symmetric random walk defined as

$$W(t) = \Delta w(W_1 + W_2 + \cdots + W_{\lfloor t/\Delta t \rfloor})$$

where, $W_i = 1$ or -1 with equal probability and $\lfloor \cdot \rfloor$ is the floor function.

Definition 1. *A stochastic process* $\{W(t), t \geq 0\}$ *is said to be a standard Wiener process if*

(i) $W(0) = 0$*;*

(ii) $\{W(t), t \geq 0\}$ *has stationary and independent increments;*

(iii) *for every* $t > 0, W(t)$ *is normally distributed with mean* 0 *and variance* t

From the definition, we could obtain that $W(t)$ is Markovian, and the increment Δw defined as $W(t + \Delta t) - W(t)$ is itself a normal random variable with

$$\mathrm{E}[\Delta w] = 0$$
$$\mathrm{Var}\,[\Delta w] = \Delta t$$

From these properties, we could define

$$\Delta w = \epsilon \Delta t$$

where ϵ has a standard normal distribution.

This process is also called Brownian motion. British botanist Robert Brown first found the phenomenon of random motion of particle in liquid. Norbert Wiener developed the mathematical theory of this process.

Now we extend our definition of Wiener process to a general case.

Definition 2. *A stochastic process* $\{W(t), t \geq 0\}$ *is a Wiener process with drift coefficient* μ *and variance parameter* σ^2 *if*

(i) $W(0) = 0$*;*

(ii) $\{W(t), t \geq 0\}$ *has stationary and independent increments;*

(iii) *for every* $t > 0, W(t)$ *is normally distributed with mean* μt *and variance* $\sigma^2 t$

While keeping the Markovian property, the mean and variance of change in W is now $\mu \Delta t$ and $\sigma^2 \Delta t$.

2.2 Itô's lemma

As in ordinary calculus, the notation dt, used in this section and further sections, refers to $\Delta t \to 0$.

Lemma 1. *Given a function* $F(x, t)$ *and an Itô process defined as*

$$dx = a(x, t)dt + b(x, t)dW,$$

where dW follows a Wiener process, we have

$$dF = \left(\frac{\partial F}{\partial x} a + \frac{\partial F}{\partial t} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} b^2 \right) dt + \frac{\partial F}{\partial x} b \; dW.$$

Itô's lemma, named after Japanese mathematician Kiyosi Itô, is a very important theorem in stochastic analysis. It indicates the rule of differentiation for a stochastic process.

2.3 Black-Scholes-Merton model

The Black-Scholes-Merton model is a widely used mathematical model for valuation of financial derivatives. It was first published by Fischer Black and Myron Scholes in 1973 [2]. This model was then improved by Robert Merton later in the same year [13].

One of the most common stock derivatives is the option . An option is a contract, with an underlying asset, providing a specific strike price and a specific expiration date. The owner of an European call option has the right to buy a share of the underlying stock at strike price on the expiration date. A put option provides the right to sell. The Black-Scholes-Merton model is the first model providing a closed form solution for option pricing.

The Black-Scholes-Merton model is under the following assumptions:

1. The stock price follows a geometric Brownian motion with constant drift and volatility.

2. The short selling of securities is allowed.

3. There are no transaction cost or taxes.

4. There are no dividends paid, or dividends are continuous.

5. There are no riskless arbitrage opportunities.

6. Security is continuously traded.

7. The risk-free interest rate is constant or deterministic, i.e., the function of risk-free interest rate is known.

Let S be the price of stock. Let μ and σ denotes the annualized expected return and volatility of the stock. Based on the first assumption, we could define

$$dS = \mu S dt + \sigma S dW$$

where dW follows a Wiener process.
To apply Itô's lemma, let $F = \ln S$

$$\begin{aligned} dF &= \left(\frac{\partial F}{\partial S} \mu S + \frac{\partial F}{\partial t} + \frac{1}{2} \frac{\partial^2 F}{\partial S^2} \sigma^2 S^2 \right) dt + \frac{\partial F}{\partial S} \sigma S \ dW \\ &= \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma \ dW \end{aligned}$$

Note $\ln S$ follows a generalized Wiener process. This implies the change of logarithm of stock price has a normal distribution,

$$\ln S(t + \Delta t) \sim \mathcal{N} \left[\ln S(t) + \left(\mu - \frac{\sigma^2}{2} \right) \Delta t, \sigma^2 \Delta t \right]$$

To derive the pricing formula for option, let C be the value of a European call option with strike price K and maturity T. The value of this option at maturity should either

be worthless if the stock price falls below the strike price, or equal to the difference between stock price and strike price, K. This is

$$C(S, t = T) = \max[S(T) - K, 0]$$

By assuming there is no arbitrage opportunity, given r,the risk-free return, the option price at time t should be equal to the discounted value of its expected price at maturity.

$$\begin{aligned} C(S,t) &= e^{-r(T-t)} \int_{-\infty}^{\infty} \max[S(T) - K, 0]\ p(S(T))dS(T) \\ &= e^{-r(T-t)} \int_{K}^{\infty} [S(T) - K]\ p(S(T))dS(T) \end{aligned}$$

where $p(\cdot)$ is the probability density function. Since $S(T)$ follows a log normal distribution as we proved before, the probability density function of $S(T)$, at time t, is given by

$$p((S(T)) = \frac{1}{S(T)\sigma\sqrt{2\pi(T-t)}} e^{-\frac{(\ln S(T) - \ln S(t) - (\mu - \sigma^2/2)(T-t))^2}{2\sigma^2(T-t)}}$$

Let $\phi(\cdot)$ denotes the cumulative distribution function of a standard normal distribution. The solution of the integration above is

$$C(S,t) = S(t)\phi(d_1) - Ke^{-r(T-t)}\phi(d_2)$$

where

$$d_1 = \frac{\ln(S(t)/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = \frac{\ln(S(t)/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t}$$

2.4 Dupire's deterministic volatility

The Black-Scholes-Merton model assumes the volatility of one asset is constant. However, if we input the market price of options into the pricing formula we derived, to calculate *implied* volatility, we will find that the volatility is not constant. If we plot the graph of implied volatility vs. strike price, we will get a "smile" shaped graph.

Dupire developed a method to solve this problem in 1994 [4]. By collecting instantaneous data on market, we can construct a surface of option price $C(K,T)$ with respect to strike price and maturity. Simultaneously, a volatility surface $\sigma(K,T)$ could also be constructed.

The deterministic volatility is usually calculated in the form of partial derivative with respect to K and T. Since the deterministic volatility is seeking the instantaneous behavior in the market, it's natural to express it in term of K and T. However, to be consistent with other parts of this paper, we would like to give an expression in term of S and t.

The Dupire's deterministic volatility is given by

$$\sigma = \frac{b(S,t)}{S}$$

where $b(S,t)$ is the diffusion term, which could be derived from the Black-Scholes equation as follows

$$-\frac{b^2(S,t)}{2}\frac{\partial^2 C}{\partial S^2} = \frac{\partial C}{\partial t}$$

which leads to

$$\sigma^2(S,t) = \frac{\dfrac{\partial C}{\partial t}}{-\dfrac{1}{2}S^2\dfrac{\partial^2 C}{\partial S^2}}$$

2.5 Heston model

Taking one step further than the deterministic volatility, it's natural to assume the volatility, as a measure of randomness of the asset, is itself a random process. Steven Heston has developed a stochastic model with stochastic volatility in 1993 [8].

The Heston model uses two correlated stochastic process to describe the behavior of the stock as follows,

$$dS = \mu S dt + \sqrt{v} S dW_S$$

$$d\sqrt{v} = -\beta\sqrt{v}dt + \delta dW_v$$

By Itô's lemma, the second equation could be transformed into following form,

$$dv = \kappa[\theta - v]dt + \xi\sqrt{v}dW_v$$

The closed form solution for pricing an European call option is given in a similar format as the Black-Scholes-Merton model.

$$C(S,v,t,T) = SP_1 - Ke^{-r(T-t)}P_2$$

with

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\Big[\frac{e^{-\mathrm{i}\ \varphi \log K} f_j(\ln S, v, \tau; \varphi)}{i\varphi}\Big]d\varphi, \quad j = 1,2$$

where f is the characteristic function

$$f_j(\ln S, v, \tau; \varphi_j) = e^{A_j(\tau;\varphi_j) + B_j(\tau;\varphi_j)v + \mathrm{i}\ \varphi \ln S}$$

with

$$A_j(\tau;\varphi_j) = \mu\varphi_j\tau + \frac{a}{\xi^2}\left\{(b_j - \rho\xi\varphi_j\mathrm{i} + d_j)\tau - 2\ln\left[\frac{1 - g_j e^{d_j\tau}}{1 - g_j}\right]\right\}$$

$$B_j(\tau;\varphi_j) = \frac{b_j - \rho\xi\varphi_j\mathrm{i} + d_j}{b_j - \rho\xi\varphi_j\mathrm{i} - d_j}\left[\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}}\right]$$

$$g_j = \frac{b_j - \rho\xi\varphi_j\mathrm{i} + d_j}{b_j - \rho\xi\varphi_j\mathrm{i} - d_j}$$

$$d_j = \sqrt{(\rho\xi\varphi_j\mathrm{i} - b_j)^2 - \xi^2(2u_j\varphi_j - \varphi_j^2)}$$

$$u_1 = 0.5,\ u_2 = -0.5,\ a = \kappa\theta,$$

$$b_1 = \kappa + \lambda - \rho\xi, b_2 = \kappa + \lambda$$

2.6 Estimation of parameters using historical data

Let $C_0, C_1, \ldots, C_n$ denotes the closing price for a stock from $t = 0$ to $t = T$ with a time step Δt. From what we derived in section 1.3, we have

$$\ln\frac{C_i}{C_{i-1}} \sim \mathcal{N}\left[\left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \sigma^2\Delta t\right]$$

Let $u = \mu - \dfrac{\sigma^2}{2}$, denotes the return over a period. Then, for random variables $\ln\dfrac{C_1}{C_0},\ \ln\dfrac{C_2}{C_1}, \ldots,\ \ln\dfrac{C_n}{C_{n-1}}$ we could get a likelihood function,

$$l(u,\sigma) = \prod_{i=0}^{n-1}\frac{1}{\sqrt{2\pi\sigma^2\Delta t}}e^{-\dfrac{(\ln\frac{C_{i+1}}{C_i} - u\Delta t)^2}{2\sigma^2\Delta t}}$$

Taking the log likehood,

$$\ln l(u,\sigma) = -\frac{n}{2}\ln(2\pi\sigma^2\Delta t) - \frac{1}{2\sigma^2\Delta t}\sum_{i=0}^{n-1}(\ln\frac{C_{i+1}}{C_i} - u\Delta t)^2$$

Taking the derivatives with respect to u and σ, we have the maximum likelihood estimators,

$$\hat{u} = \frac{1}{n\Delta t}\sum_{i=0}^{n-1}\ln\frac{C_{i+1}}{C_i}$$

and

$$\hat{\sigma} = \sqrt{\frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\ln\left(\frac{C_i}{C_{i-1}}\right) - \overline{\ln\left(\frac{C_i}{C_{i-1}}\right)}\right)^2}$$

Then, we have the distribution of the sample mean,

$$\hat{u} \sim \mathcal{N}[\mu - \frac{\sigma^2}{2},\ \frac{1}{n}\cdot\frac{\sigma^2}{\Delta t}]$$

Note $n\Delta t = T$, so a $(1-\alpha)\%$ confidence interval for the drift μ could be derived,

$$\begin{aligned}
\hat{u} - z_{\alpha/2}\frac{\sigma}{\sqrt{n\Delta t}} &< u < \hat{u} + z_{\alpha/2}\frac{\sigma}{\sqrt{n\Delta t}} \\
\hat{\mu} - \frac{\sigma^2}{2} - z_{\alpha/2}\frac{\sigma}{\sqrt{n\Delta t}} < \mu - \frac{\sigma^2}{2} &< \hat{\mu} - \frac{\sigma^2}{2} + z_{\alpha/2}\frac{\sigma}{\sqrt{n\Delta t}} \\
\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{T}} &< \mu < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{T}}
\end{aligned}$$

This implies that increasing the number of observations does not help in improving the estimation of the drift. Since as n increases, Δt would correspondingly decrease, and T would remain unchanged. So, a few sampling points over a long period of time would have a better estimation than a lot of sampling points over a short period of time. This result is first mention by Merton in 1980 [14].

The following estimators are mainly based on the materials in *"Paul Wilmott introduces quantitative finance"* [18]. These estimators are developed to increase the efficiency of estimation. An estimator with low efficiency requires more sampling points over a longer period of time. Then the result from such an estimator will tend to be volatility over a period rather than an estimation close to the instantaneous volatility. In order to keep the timeliness of the estimation, more efficient estimators are developed.

O,H,L,C, in the following estimators, represents opening, high, low, closing price. Again, the following estimators are based on *"Paul Wilmott introduces quantitative finance"*, [18], from where the efficiency is cited.

Parkinson's estimator presented as follows, which is 5.2 times more efficient than the maximum likelihood estimators. However, it does not handle drift.

$$\sigma_p = \sqrt{\frac{1}{4n \ln 2 \ \Delta t} \sum_{i=1}^{n} \left(\ln \left(\frac{H_i}{L_i} \right) \right)^2}$$

Garman & Klass's estimator is an improvement of Parkinson's estimator. It's 7.4 times more efficient than the maximum likelihood estimator, but it still does not consider drift.

$$\sigma_{gk} = \sqrt{\frac{1}{n\Delta t} \sum_{i=1}^{n} \left(0.511 \left(\ln \left(\frac{H_i}{L_i} \right) \right)^2 - 0.019 \ln \left(\frac{C_i}{O_i} \right) \ln \left(\frac{H_i L_i}{O_i^2} \right) - 2 \ln \left(\frac{H_i}{O_i} \right) \ln \left(\frac{L_i}{O_i} \right) \right)}$$

Rogers & Satchell's estimators considers drift, and it's 8 times more efficient than the maximum likelihood estimator.

$$\sigma_{rs} = \sqrt{\frac{1}{n\Delta t} \sum_{i=1}^{n} \left(\ln \left(\frac{H_i}{C_i} \right) \ln \left(\frac{H_i}{O_i} \right) + \ln \left(\frac{L_i}{C_i} \right) \ln \left(\frac{L_i}{O_i} \right) \right)}$$

Yang & Zhang's estimator is a sum of overnight volatility and a weighted average of open-to-close volatility and Rogers & Satchell's estimator. This estimator considers both drift and overnight jumps. It has been proved that their estimator is 14 times more efficient than the maximum likelihood estimator. According to their paper, only two days' data would be enough to estimate the volatility with certain confidence [19].

$$\sigma_{yz} = \sqrt{\sigma_o^2 + k\sigma_c^2 + (1-k)\sigma_{rs}^2}$$

where

$$\sigma_o = \sqrt{\frac{1}{(n-1)\Delta t}\sum_{i=1}^{n}\left(\ln\left(\frac{O_i}{C_{i-1}}\right) - \overline{\ln\left(\frac{O_i}{C_{i-1}}\right)}\right)}$$

$$\sigma_c = \sqrt{\frac{1}{(n-1)\Delta t}\sum_{i=1}^{n}\left(\ln\left(\frac{C_i}{O_i}\right) - \overline{\ln\left(\frac{C_i}{O_i}\right)}\right)}$$

with

$$k = \frac{0.34}{1.34 + \frac{n+1}{n-1}}$$

2.7 Comparison

Usually, the Heston model is considered to be more accurate than the Black-Scholes-Merton model. However, as pricing an financial derivative, the result of the Black-Scholes-Merton model is fairly close to the actual price. The difference between Heston model and BSM model is less than 1%, in most cases. Using Dupire's deterministic volatility with BSM model will help increase the accuracy.

The use of Heston model require the estimation of parameters. There are no closed form estimator for parameters in the Heston model. Calibration with market data is required in this case. However, the goal for this project is to find an optimal strategy

of allocation of capital into multiple assets. Heston model only gives the instantaneous behavior of the market. If we rebalance our portfolio every day in simulation, the computational work required to use Heston model will be too heavy for an individual project.

On the other hand, the stock price has a log normal distribution in Black-Scholes-Merton model. The maximum likehood estimator for a normal distribution is easy to calculate. Since we are seeking long term performance, assuming constant drift and volatility is reasonable.

Although we use the Black-Scholes-Merton model in this paper, the Heston model would be also compatible with the strategy which would be explained in a further chapter. Actually, our strategy will be compatible with most models.

3 Kelly's criterion

One of the fundamental problems of portfolio theory is how to rationally optimize the portfolio using diversification. An investor wants to find a balance between maximizing the return and minimizing the risk. In this chapter, we will begin with a univariate case, with the investor only focusing on one asset. In univariate case, the Kelly's criterion was found to be the optimal solution. Then, we will discuss several properties of the Kelly criterion. Finally, we will extend the criterion to the multivariate case, i.e., a multi-asset model.

3.1 Binomial case

Imagine a simple game with two possible outcomes. A win, with probability $p > \frac{1}{2}$, will double the wager and a loss will lose the entire wager. Let $\Pi_0 = 1$ be the initial value of the portfolio of the gambler. Let f be a constant fraction. On the ith game, the gambler will bet $f\Pi_{i-1}$. After repeatedly playing n games, we have

$$\begin{aligned} E[\Pi_n] &= (1-f)E[\Pi_{n-1}] + 2pfE[\Pi_{n-1}] \\ &= (1-f+2pf)^n\Pi_0 \end{aligned}$$

To maximize the expected return, since $2p > 1$, we want to wager our entire wealth, i.e., set $f = 1$. However, this will leads us to go bankrupt almost surely, since $\lim_{n\to\infty} 1 - p^n = 1$.

Now, let w and l be the number of wins and losses, such that $n = w + l$. Then

$$\Pi_n = (1+f)^w(1-f)^l\Pi_0$$

The expected exponential growth rate is

$$\begin{aligned} G(f) &= E\left[\ln\left[\frac{\Pi_n}{\Pi_0}\right]^{\frac{1}{n}}\right] \\ &= E\left[\frac{w}{n}\ln(1+f)+\frac{l}{n}\ln(1-f)\right] \\ &= p\ln(1+f)+(1-p)\ln(1-f) \end{aligned}$$

Note

$$G''(f) = -\frac{1-p}{(1-f)^2} - \frac{p}{(1+f)^2} < 0 \quad \forall f$$

Solve for critical point, we have a maximum at $f^* = 2p-1$. $G(f^*) = \ln 2 + p\ln p + (1-p)\ln(1-p)$.

The above criterion is developed by Kelly [11] in 1956. Note this criterion has restrictions on no short sell and no leverage. Since when involving short selling and leverage, it is possible for the entire wealth to go negative, while the logarithm of a negative number is not well-defined.

Breiman [3] showed that the Kelly's strategy using f^* beats all other strategy in 1961. We demonstrate the binomial case for fixed fraction strategy as following,

Theorem 1. *For an arbitrary constant $f \in [0,1]$, $E\left[\lim_{n\to\infty}\frac{\Pi_n(f)}{\Pi_n(f^*)}\right] \leq 1$ for all $n \in \mathbb{N}$.*

Proof.

By Borel's law of large number, we have

$$\lim_{n\to\infty} w = np$$

Then,

$$
\begin{aligned}
& E\left[\lim_{n\to\infty}\frac{\Pi_n(f)}{\Pi_n(f^*)}\right] \\
=& E\left[\lim_{n\to\infty}\left(\frac{1+f}{1+f^*}\right)^w\left(\frac{1-f}{1-f^*}\right)^l\right] \\
=& E\left[\left(\frac{1+f}{1+f^*}\right)^{pn}\left(\frac{1-f}{1-f^*}\right)^{n-pn}\right] \\
=& e^{\left[pn\ln\left(\frac{1+f}{1+f^*}\right)+(n-pn)\ln\left(\frac{1-f}{1-f^*}\right)\right]} \\
=& e^{n\left[p\ln\left(\frac{1+f}{1+f^*}\right)+(1-p)\ln\left(\frac{1-f}{1-f^*}\right)\right]} \\
=& e^{n[G(f)-G(f^*)]} \\
\leq& 1
\end{aligned}
$$

□

The above theorem shows the expected wealth using Kelly's strategy is always greater than the other fixed fraction strategy, no matter how many plays we played. Actually, Bellman and Kalaba [1] presented an inductive proof showing that the Kelly's strategy beats, not only fixed fraction betting strategy, but all other betting strategy.

Theorem 2. *For $n > 1$, the optimal strategy to maximize $E[\ln \Pi_n]$ is betting a fixed fraction $f^* = 2p - 1$ at each game.*

Proof. Prove by induction. For $n = 1$, $G(f^*)$ maximizing $E[\ln \Pi_1]$ is true. Assume

the proposition holds for n.

$$\begin{aligned}
\max[E[\ln \Pi_{n+1}]] &= \max_{0\le f\le 1}[p[\ln((1+f)\Pi_0) \\
&\quad + nG(f^*)] + (1-p)[\ln((1-f)\Pi_0) + nG(f^*)]] \\
&= \max_{0\le f\le 1}[p[\ln((1+f)\Pi_0)] + (1-p)[\ln((1-f)\Pi_0)]] + nG(f^*) \\
&= (\ln \Pi_0 + G(f^*)) + nG(f^*) \\
&= \ln \Pi_0 + (n+1)G(f^*)
\end{aligned}$$

□

In addition, Breiman [3] shows Kelly's strategy minimizes the expected time needed to achieve a certain goal.

Theorem 3. *For any number $x > 1$, define the random variable $T(x)$ by*

$$T(x) = \{\text{smallest } n \text{ such that } \Pi_n \ge x\},$$

Then,

$$\lim_{x\to\infty}[E[T(x)] - E[T^*(x)]] \ge 0$$

always holds.

The above two properties shows that Kelly's criterion not only maximizes the return in the long-run, but also minimizes the number of plays required to reach a goal of certain amount of wealth. So, it seems that Kelly's criterion is the optimum solution in this scenario.

When the distribution of the return is heavily skewed, it may be reasonable to seek

the maximum median of the return instead of the geometric mean. Ethier [6] shows that Kelly criterion not only maximizes the expected logarithm of the future wealth, but also the median of the future wealth.

Theorem 4. *For $n \in \mathbb{N}$,* median$[\Pi_n(f)]$ *is uniquely maximized at*

$$\tilde{f}_n = f^* + O(n^{-1})$$

Proof. Note w denotes the number of wins in the first n plays. Then w is binomial(n, p) and

$$\begin{aligned}\Pi_n(f) &= (1+f)^w(1-f)^{n-w}\Pi_0 \\ &= \left(\frac{1+f}{1-f}\right)^w (1-f)^n \cdot 1\end{aligned}$$

Since logarithm is monotonic, we have

$$\ln(\text{median}[\Pi_n(f)]) = \text{w}\ln\left(\frac{1+f}{1-f}\right) + n\ln(1-f)$$

It is proved by Edelman [5] and Hamza [7], that

$$|\gamma_{n,p}| = |\text{median}[w] - E[w]| < \ln 2.$$

Since $E[\ln(\Pi_n(f))] = np\ln(1+f) + n(1-p)\ln(1-f)$, we have

$$\ln(\text{median}[\Pi_n(f)]) = (np + \gamma_{n,p})\ln(1+f) + [n(1-p) - \gamma_{n,p}]\ln(1-f)$$

Note $p + \frac{\gamma_{n,p}}{n}$ is the median of $\frac{w}{n}$. We have the median$[\Pi_n(f)]$ has a maximum at

$$\tilde{f} = 2(p + \frac{\gamma_{n,p}}{n}) - 1 = f^* + 2\frac{\gamma_{n,p}}{n}$$

□

Now consider a stock, which would either raise by 1/3 or fall by 3/10, with equal probability, for each time step. The expected return for one step is,

$$E(R) = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot (-\frac{3}{10}) = \frac{1}{60} \approx 1.67\%$$

For each time step, this stock has a approximately 1.67% return on capital. Then, we derive the behavior of the price of this stock in the long run. Let n be large enough, so we could apply Borel's law of large number.

$$\begin{aligned} S_n &= \left(1 + \frac{1}{3}\right)^{n/2} \left(1 - \frac{3}{10}\right)^{n/2} S_0 \\ &= \left(\sqrt{\frac{14}{15}}\right)^n S_0 \end{aligned}$$

This shows that even the stock has a positive expected return, it would decay by $1 - \sqrt{\frac{14}{15}}$ at each time step. The reason why this would happen is because the expected return is calculated as arithmetic mean or weighted arithmetic mean, while the stock does not follow an additive process. In most models, it's assumed a stock follows a multiplicative process.

Now let's apply Kelly's criterion. A maximum of $G(f)$ could be found at $f = \frac{1}{6}$. Our

portfolio in the long run would be

$$\begin{aligned}\Pi_n &= \left(1+\frac{1}{3}\cdot\frac{1}{6}\right)^{n/2}\left(1-\frac{3}{10}\cdot\frac{1}{6}\right)^{n/2}\Pi_0 \\ &\approx 1.00139^n\Pi_0\end{aligned}$$

The result agrees with our derivation. As long as there is a positive expected return, Kelly's criterion will give a growth in capital in the long run. A simulation of this example is given as Figure 1.

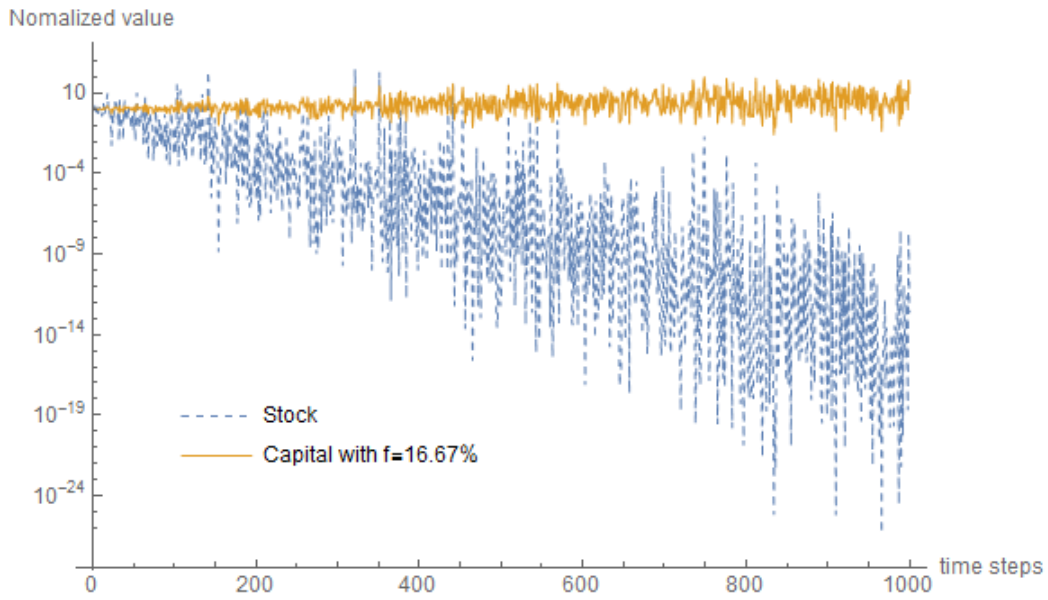


Figure 1: The value of both stock price and our portfolio are normalized in the plot, i.e., they are calculated as S_t/S_0 and Π_t/Π_0. The plot is in log scale.

3.2 Univariate stock

In this section, we assume the stock follows a geometric Brownian motion with a drift μ and a volatility σ.

Our goal is to find the maximum of the expected exponential growth rate $G(f)$ with

respect to f. Let R be the return on capital, we have

$$G(f) = E[\ln(1 + Rf)]$$

To get the expectation, we need to derive the distribution of R.

When R is small, we could simply use the discretized the stochastic differential equation we have in section 1.3, i.e.,

$$dS = \mu S dt + \sigma S dW$$

By discretization,

$$\Delta S = \mu S \Delta t + \sigma S \epsilon \sqrt{\Delta t}$$

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \epsilon \sqrt{\Delta t}$$

Then,

$$R = \frac{\Delta S}{S} \sim \mathcal{N}[\mu \Delta t, \sigma^2 \Delta t]$$

It shows R has a normal distribution. However, this is only an approximation when ΔS is small. We should notice ΔS has a lower boundary at $-S$, so R should be bounded by -1. Otherwise, it's possible that $1 + Rf$ will be negative. The logarithm of a negative number is not well defined. This implies that there is a part of the domain of the integral, for calculating the expectation, are not well defined.

Another approach would eliminate the risk of having non-well-defined domain. Let u be the continuously compounded growth rate over a period of time, which we derived

in section 1.6. The return on capital could be defined in terms of u,

$$R = e^{u\Delta t} - 1$$

where

$$u \sim \mathcal{N}\left[\mu - \frac{\sigma^2}{2}, \frac{\sigma^2}{\Delta t}\right]$$

Now the problem is

$$\begin{aligned} G(f) &= E[\ln(1 + Rf)] \\ &= \int_{-1}^{\infty} \ln(1 + Rf)\ p(R)dR \\ &= \int_{-\infty}^{\infty} \ln[1 + f(e^{u\Delta t} - 1)]\ p(u)du \end{aligned}$$

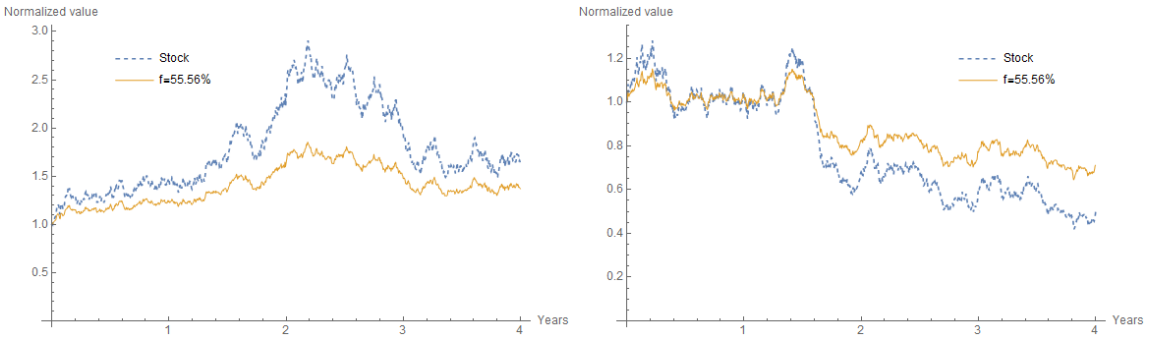
where $p(u)$ is the probability density function of the normal random variable u.

This integration seems that there is no closed form analytic solution, however, it possible to be solved numerically. Note the second derivative of this expectation with respect to f is

$$-\int_{-\infty}^{\infty} \frac{(e^{u\Delta t} - 1)^2}{[1 + f(e^{u\Delta t} - 1)]^2}\ p(u)du < 0 \quad \forall f$$

So this is a convex problem. For any level of accuracy goal, there is a finite number of points to be sampled. If we were seeking to two significant figures, there would be at most 101 points to be examined since $f \in [0, 1]$. At each point, a numerical integral could be done to evaluate the value of the expectation.

Consider a stock with $\mu = 0.05$ and $\sigma = 0.3$. The maximum of $G(f)$ could be found at $f = 0.5556$. Two simulations are given in Figure 2.



(a) A stock with upward trend (b) A stock with downward trend

Figure 2: Two simulations with different trends using identical parameters

Note in Figure 2, the simulations are under identical parameter. It shows that Kelly's strategy tends to reduce the risk while it also give up some return. When a stock has an upward trend, investing only a fraction will result in an slower growth. However, when a stock has a downward trend, investing only a fraction will result in an slower decay.

However, based on our theory, our strategy will eventually beats all other strategy in the long run. Although, in Figure (a), the performance of our strategy falls behind the market in 4 years, it will eventually overtake the market performance in long run, as shown in Figure 3.

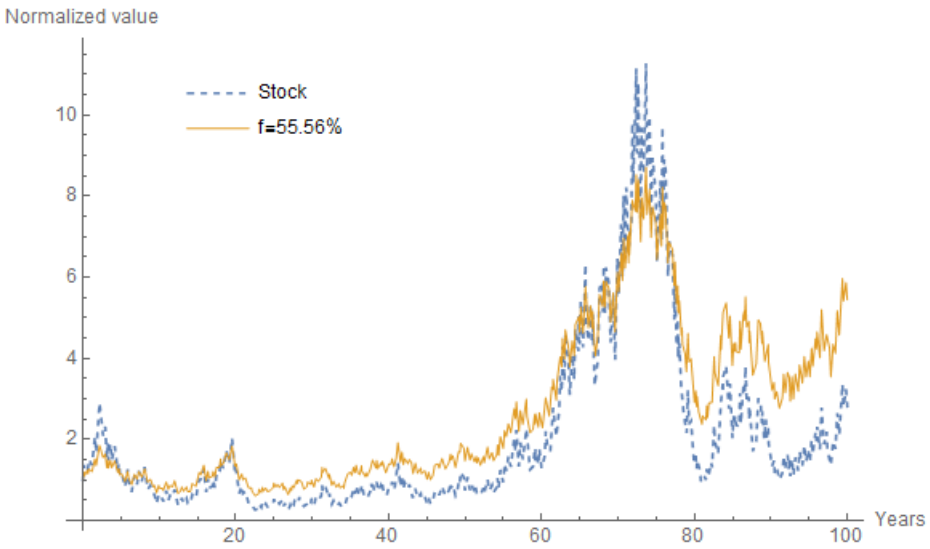


Figure 3: subsequent simulation of Figure 2(a)

3.3 Uncorrelated multivariate stocks

If we have a number of uncorrelated stocks, the criterion could be extended in a simple way. Let $S_1, S_2, \ldots, S_n$ be n different independent stocks, where S_i follows a geometric Brownian motion with μ_i and σ_i for $i = 1, \ldots, n$. Then the expected exponential growth rate of the portfolio would be

$$G(f_1, f_2, \ldots, f_n) = E[\ln[1 + \sum_{i=1}^{n} f_i R_i]]$$

Result in

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\ln[1 + \sum_{i=1}^{n} f_i (e^{u_i \Delta t} - 1)]] \prod_{i=1}^{n} p(u_i) du_1 du_2 \ldots du_n$$

Note

$$\frac{\partial^2 G}{\partial f_i^2} = -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{(e^{u_i\Delta t}-1)^2}{[1+\sum_{i=1}^{n} f_i(e^{u_i\Delta t}-1)]^2}\prod_{i=1}^{n} p(u_i)du_1du_2\ldots du_n$$
$$< 0$$

This implies the second derivative of G with respect to each f_i is always negative. So this is a convex optimization problem. Let the vector $\mathbf{f} = [f_1, f_2, \ldots, f_n]$ represents the our portfolio, where f_i is the fraction invested on ith stock.

$\mathbf{f}$ is constrained by $f_i > 0,\ \forall i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n} f_i \leq 1$. These two constrains ensure the assumption of no leverage and no short selling.

Now a simple hill climbing algorithm would be good to solve this problem. The hill climbing algorithm may be the optimal choice for this case, since we are lacking the necessary information such as the gradient or the analytic solutions. The only thing we could do is to do numerical integrations to give $\mathbf{f}$. Start with some initial guess $\mathbf{f}_0$. For each iteration, one entry of $\mathbf{f}$ would be changed to create a new vector. Then the vector with a greater value of evaluation would be kept. The algorithm would keep climbing until no improvement could be found. Then a global maximum is found.

If the global maximum is within the constrains, we are done. If it's located outside the constrains, then a local maximum within the constrains must be found on the boundaries.

Suppose $G(f)$ is convex with a closed boundary, and the global maximum of $G(f)$ was found outside the boundary. Assume the maximum within the boundary is an interior point. Then draw a straight line connecting that point and the global maximum. The intersection of the boundary and the line must has a greater value since

$G(f)$ is convex, which leads to a contradiction. Therefore, if the global maximum of a convex function is found outside the boundary, then the maximum within the boundary must be found on the boundary.

Then, if the climber found the global maximum outside the boundary, we should run the algorithm along the boundary to find the local maximum.

Consider an example with three stocks, where $\mu_1 = 0.1, \sigma_1 = 0.4,\ , \mu_2 = 0.07, \sigma_2 = 0.25$ and $\mu_3 = 0.04, \sigma_3 = 0.1$. The global maximum was found at $\mathbf{f} = [0.6228,\ 1.1180,\ 3.9903]$, and the maximum within the boundary was found at $\mathbf{f} = [0.1301,\ 0.6069,\ 0.2630]$. A simulation is shown in Figure 4.

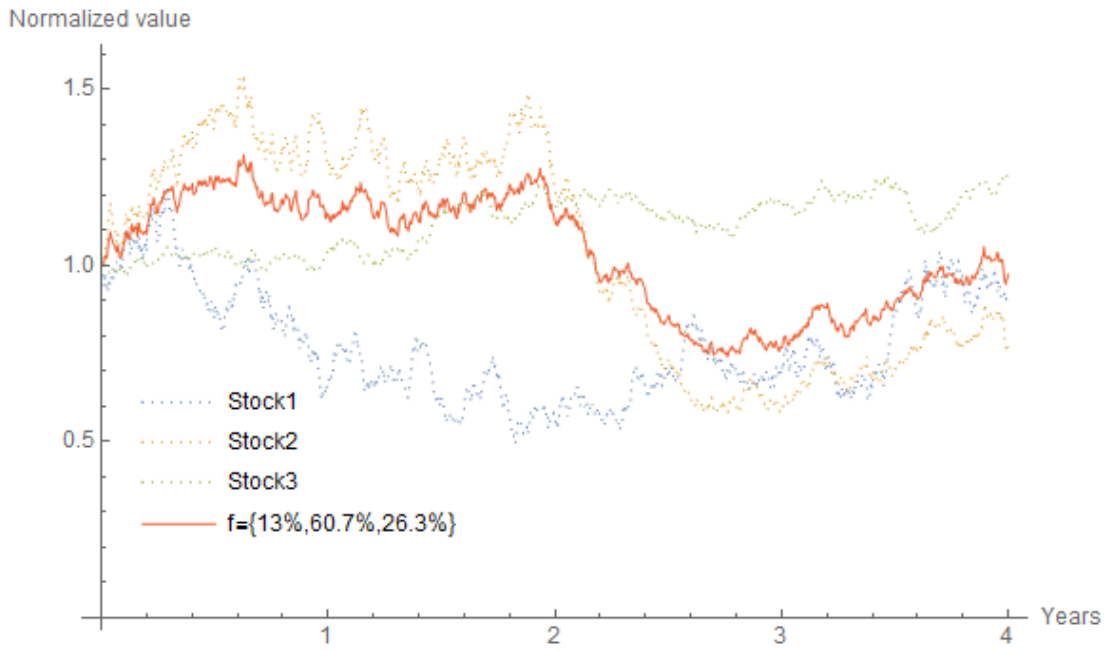


Figure 4: simulation of 3 stocks, with given parameters, and the portfolio build on Kelly's criterion

As shown in the simulation, our strategy usually performs at an average level. It gives up some returns to keep a reasonable risk level, in the short run.

In the long run, as the previous univariate case, our strategy is much better than buying and holding a single stock, as shown in Figure 5.

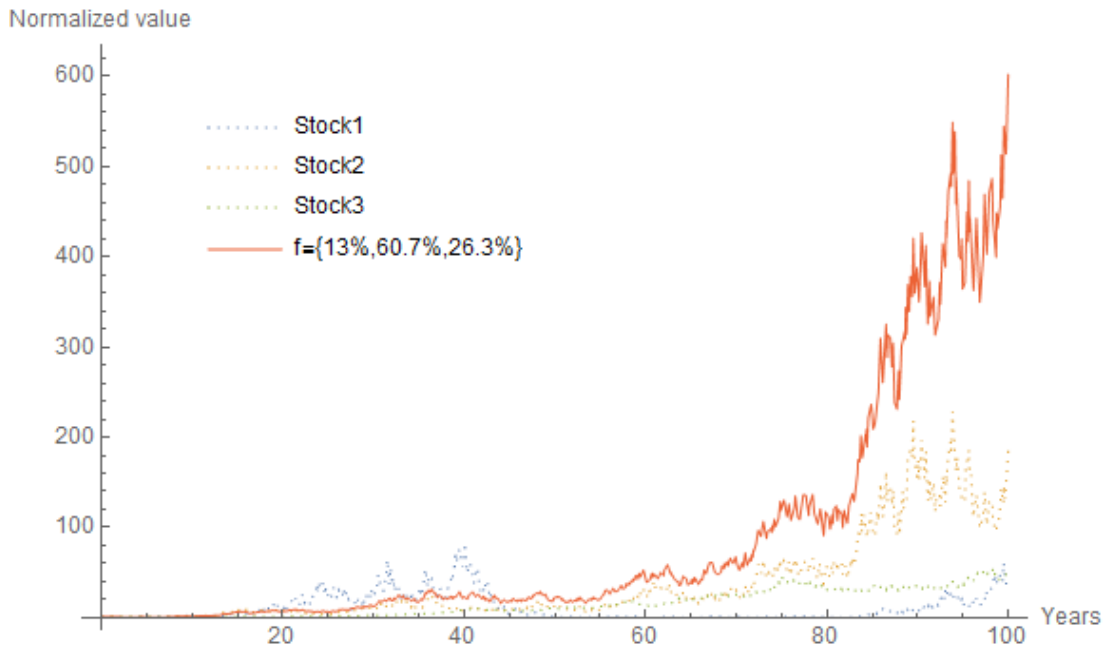


Figure 5: subsequent simulation of Figure 4

So, as the figure above, we will hit the highest expected return, in 100 years. It's possible that even after 100 years, we are still behind some other strategy, but, as t goes to infinity, i.e., in the long run, we will eventually beat all other strategies.

Consider the expected return per unit of risk, which has been defined as Sharpe's ratio:

$$\frac{E[R_\Pi] - r}{\sigma_\Pi}$$

where r is the risk-free rate, which we set to be zero. So in this paper, the Shape's ratio is defined as

$$\frac{E[R_\Pi]}{\sigma_\Pi}$$

Sharpe's ratio is a measure of expected return per unit of risk. Figure 6 is the graph

of Shape's ratio calculated by the simulated data from Figure 5.

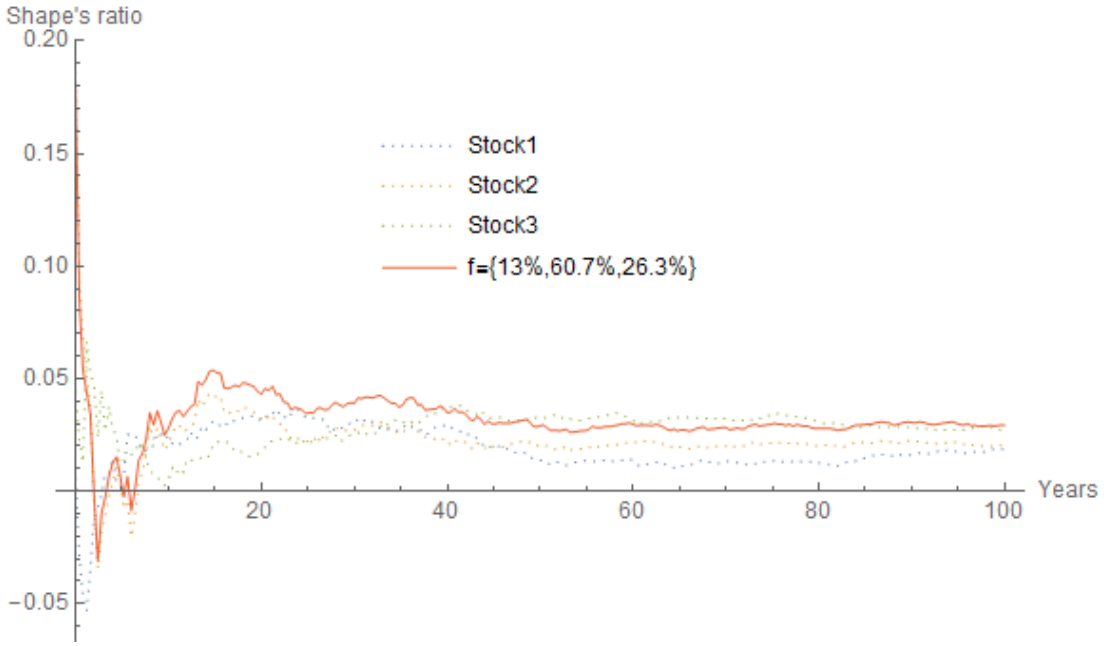


Figure 6: Shape's ratio calculated using the data in Figure 5

The Shape's ratio tends to converge as the number of iterations increases. At the end of the 100th year, our strategy achieves a higher Shape's ratio then buying and holding single stocks.

Therefore, our strategy not only has the highest expected return, but also has the highest expected return per unit of risk, in the long run.

3.4 Trading frequency

Although our strategy is correct and the best in the long run, the "long run" is probably too long which makes the strategy not very practical.

Recall the result in section 1.6. When estimating μ, the drift, based on the Black-Scholes-Merton Model, the number of sampling points does not matter, but the length

of period of time matters. T, the length of time period is the only parameter helping improve the maximum likelihood estimator of μ.

Enlightened by this result, it's possible that, for identical games, it doesn't matter how long you played the game, but how many times you played. For a fixed length of time period, to increase the number of times, the frequency must increases.

All the previous simulations are made on a basis that we rebalanced our portfolio once a day. In the next simulation, a univariate case was tested and all the parameter were kept constant, except the trading frequency. Three values of trading frequency were tested, once a day, once half an hour, and once per minute. The simulation result was shown in Figure 7.

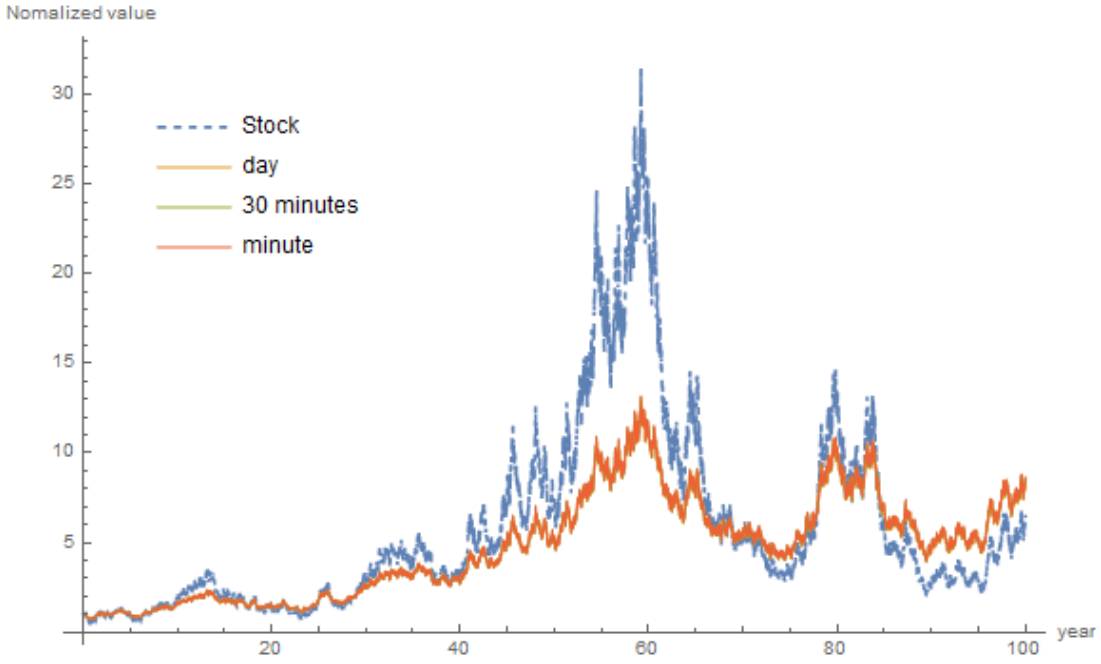


Figure 7: Three different trading frequency compare to the stock performance using the same μ and σ from figure 2.

It turns out that the lines of different frequency, are almost completely overlapping each other. There is actually some differences between the result, however, they are too tiny to be seen on the graph. The numerical results of final normalized capital

are 8.41, 8.46, and 8.50, corresponding to trading every day, every half an hour, and every minute. The improvement made by increasing the trading frequency is so tiny, over 100 years. If the strategy involves the transaction cost, the improvement would not be able to cover the extra cost.

4 Correlated multivariate stocks

In previous chapter, we assume there is no correlation between different stocks, which is not quite true in reality. It's easier to assume the independence between different assets, to construct the joint probability density function as the product of the marginal probability density function. However, when assets are correlated, which is true most times, we probably need some other technology to build our joint probability density function, or avoid using it.

In this chapter, we will use the data from real market. In order to have enough data, those stocks with long history were chosen. General Electric, DuPont, and IBM are chosen, because they have long history, and they are large enough, which we could assume there are enough liquidity. We collect their adjusted daily close price from Jan. 2nd 1962 to Dec. 31st 2015.

By calculating the continuously compounded return of each of stock, among 13594 trading days, heteroskedasticity are found. We could use *Generalized Autoregressive Conditional Heteroskedasticity* model to eliminate its influence. However, it's possible that the existence of heteroskedasticity is reasonable because the volatility itself could be stochastic as well. To keep our model simpler, we assume volatility is constant.

In this chapter, we will introduce 3 methods, the Taylor's expansion, nonparametric kernel density estimation, and copula.

4.1 Taylor's expansion

It is relatively difficult to build joint distribution from marginal distributions. So there exist some work to trying to avoid making assumption on joint distributions. Vasily Nekrasov has done a model free approach by not making any assumption on the distribution of returns. He simply starts with the expression of portfolio's capital

after one time step:

$$\sum_{i=1}^{n} f_i(1+r_i) = 1 + \sum_{i=1}^{n} f_i r_i$$

where f_i and r_i are the fractions and returns on the ith asset. The expected exponential growth rate is:

$$E\Big[\ \ln\left(1 + \sum_{i=1}^{n} f_i r_i\right)\Big]$$

Then, without making any assumption on the distribution of r_i, he takes Taylor's expansion of the expression above around $\mathbf{f} = [0, 0, \cdots, 0]^T$. The first two terms of the expansion is given as follows,

$$E\Big[\sum_{i=1}^{n} f_i r_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j r_i r_j\Big]$$

Rewrite the expression in term of vectors,

$$E\Big[(\mathbf{r} - \mathbf{1}r)^T\mathbf{f} - \frac{1}{2}\mathbf{f}^T\mathbf{\Sigma}\mathbf{f}\Big]$$

where Σ is the $n \times n$ covariance matrix of returns.

The problem have been turned into a quadratic problem, the unconstrained solution is given by,

$$\mathbf{f}_{opt} = \hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{r}}$$

where $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{r}}$ are the estimations of covariance matrix and returns.

Vasily Nekrasov has done an error analysis for this approach for some special cases. The result is that, when the return for a single asset is between -20% and 20%, the error will be less than or equal to 0.7%. This result is based on the assumption that r_i is fixed for all i.

The advantage of this approach is that the result is in closed form, which provides a

fast way for computing.

The disadvantage is that, the result is based on the estimation of expectation of returns. Most of the times, the distribution of return will have a fat tail. The expectation is probably not a better statistic than median, in these case. Second, the result depends on the covariances, which could only describe the linear correlationship. Third, the result is unconstrained, which means it allows leverage and short. It leads to the risk of bankruptcy, which is on the contrary to the motivation of Kelly's criterion.

Paolo Laureti et al. gives another approach using Taylor's expansion. Let $g(u)$ be a function of u, the random vector of the realized return over time for n assets. Take Taylor's expansion around $u = E(u) = \mu - \frac{\sigma^2}{2}$.

$$g(u) = g(\mu-\frac{\sigma^2}{2})+\nabla g(\mu-\frac{\sigma^2}{2})\cdot(u-(\mu-\frac{\sigma^2}{2}))+\frac{1}{2}(u-(\mu-\frac{\sigma^2}{2}))^T V(\mu-\frac{\sigma^2}{2})(u-(\mu-\frac{\sigma^2}{2}))+\cdots$$

where $V(\mu-\frac{\sigma^2}{2})$ is an $n\times n$ matrix. The ijth entry is set to be $\frac{\partial^2}{\partial u_i \partial u_j} g(u)$ calculated at $\mu - \frac{\sigma^2}{2}$.

Then the expectation could be approximately calculated without using the distribution function as follows,

$$\begin{aligned}
E(g(u)) \approx & g(\mu - \frac{\sigma^2}{2}) \int p(u)du + \sum_{i=1}^{n} \partial_i g(\mu - \frac{\sigma^2}{2}) \int (u_i - (\mu_i - \frac{\sigma_i^2}{2}))p(u)du \\
& + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} V_{ij} \int (u_i - (\mu_i - \frac{\sigma_i^2}{2}))(u_j - (\mu_j - \frac{\sigma_j^2}{2}))p(u)du \\
= & g(\mu - \frac{\sigma^2}{2}) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}V_{ij} \\
= & g(\mu - \frac{\sigma^2}{2}) + \frac{1}{2}\mathrm{Tr}(SV)
\end{aligned}$$

where S is the covariance matrix. $p(u)$ is the joint probability density function, which has been cancel at the end.

This approach has taking the distribution of returns into account. However, there is no close form result, and the result is again not constrained.

4.2 Nonparametric kernel density estimation

In previous section, the approach is completely based on parametric analysis. We need to make some prior assumptions such as what process the stock follows. The model is decided first, then the data is used to calibrate the parameters. The approach is highly sensitive to the choice of model.

In this section, we will introduce a method called *kernel density estimation*. This method is completely data-driven. It needs no assumption in advance.

Based on my understanding, the kernel density estimation is a generalization of histogram. We could consider histogram as a special "rectangular kernel".

The following introduction of kernel density estimation is mainly based on "*Nonparametric Econometrics Theory and Practice*" written by Qi Li and Jeffrey Racine [12].

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from some univariate distribution. A very intuitive way to estimate the distribution function is called *empirical distribution function*. The empirical distribution function is defined as follows,

$$F(x) = \frac{\#X_i \leq x}{n}$$

Note this empirical function is a step function, which is not differentiable. To estimate the probability density function, it's nature to use the original definition of derivative, like following

$$f(x) = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

Now, the estimation of $f(x)$, derived from the empirical distribution function, could be write in the following form,

$$\hat{f}(x) = \frac{1}{2h}\frac{\#X_i \in [x-h, x+h]}{n}$$
$$\frac{1}{2hn}\sum_{i=1}^{n} I(X_i)$$

where

$$I(X_i) = \begin{cases} 1 & \text{if } x-h \leq X_i \leq x+h \\ 0 & \text{otherwise} \end{cases}$$

This function is similar to a smoothed histogram, with the smooth factor h. This estimation is also called kernel density estimation, where the kernel function is rectangular kernel, defined as

$$k_{rec}(x) = \frac{1}{2} I(|x| < 1)$$

Rewrite the estimation using kernel function, we have

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k_{rec}\left(\frac{x - X_i}{h}\right)$$

To check if the estimation, $\hat{f}(x)$, is a probability density function, take the integral,

$$\begin{aligned} \int \hat{f}(x)dx &= \int \frac{1}{nh}\sum_{i=1}^{n} k_{rec}\left(\frac{x-X_i}{h}\right) dx \\ &= \frac{1}{nh}\sum_{i=1}^{n}\int k_{rec}\left(\frac{x-X_i}{h}\right) dx \\ &= \frac{1}{n}\sum_{i=1}^{n}\int k_{rec}\left(x - X_i\right) dx \\ &= \frac{1}{n}\sum_{i=1}^{n}\int k_{rec}\left(x\right) dx \end{aligned}$$

From the result above, it's clear that, as long as the kernel function itself is a probability density function, the estimation, $\hat{f}(x)$, will also be a probability density function. The estimation using rectangular kernel is not smooth enough to be differentiable. A common choice is to use the probability density function of standard normal distribution as the kernel function, called *Gaussian kernel*. Rewrite the estimation of $f(x)$ using Gaussian kernel,

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k_{Gauss} \left(\frac{x - X_i}{h} \right)$$

where

$$k_{Gauss}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Actually the choice of kernel function is not very important. Consider a random sample of size 10,000 was drawn from the standard normal distribution. Apply the kernel density estimation on these samples using the following kernels,

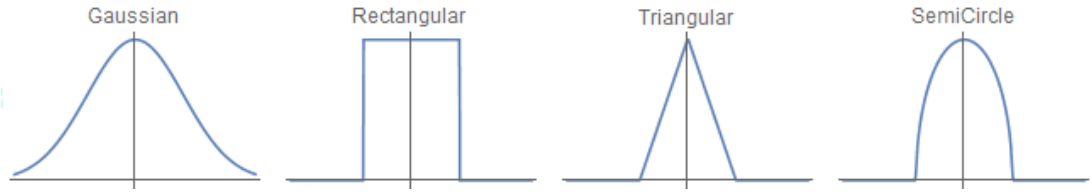


Figure 8: the shape of different kernels

The definition of Gaussian and rectangular kernels are given before. The triangular kernel and semicircle kernel are defined as following,

$$k_{Tri}(x) = 1 - |x|, \qquad k_{semi}(x) = \frac{2}{\pi} \sqrt{1 - x^2}$$

Applying these kernel to the samples, to estimate the probability density function, we get

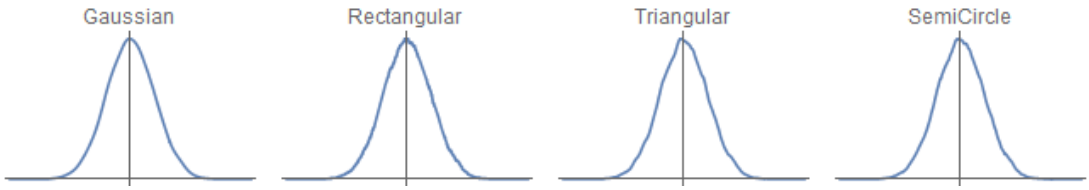


Figure 9: the graph of estimated probability density function of 10,000 random samples, drawn from standard normal distribution, using different kernel functions

The shapes of the graphs of resulting estimations do not differ too much. Although some properties are different when different kernels are applied, like differentiability, the graphs are very close.

Actually, the crucial factor is how to choose h, the smoothing factor. The figure below shows the difference caused by choice of h.

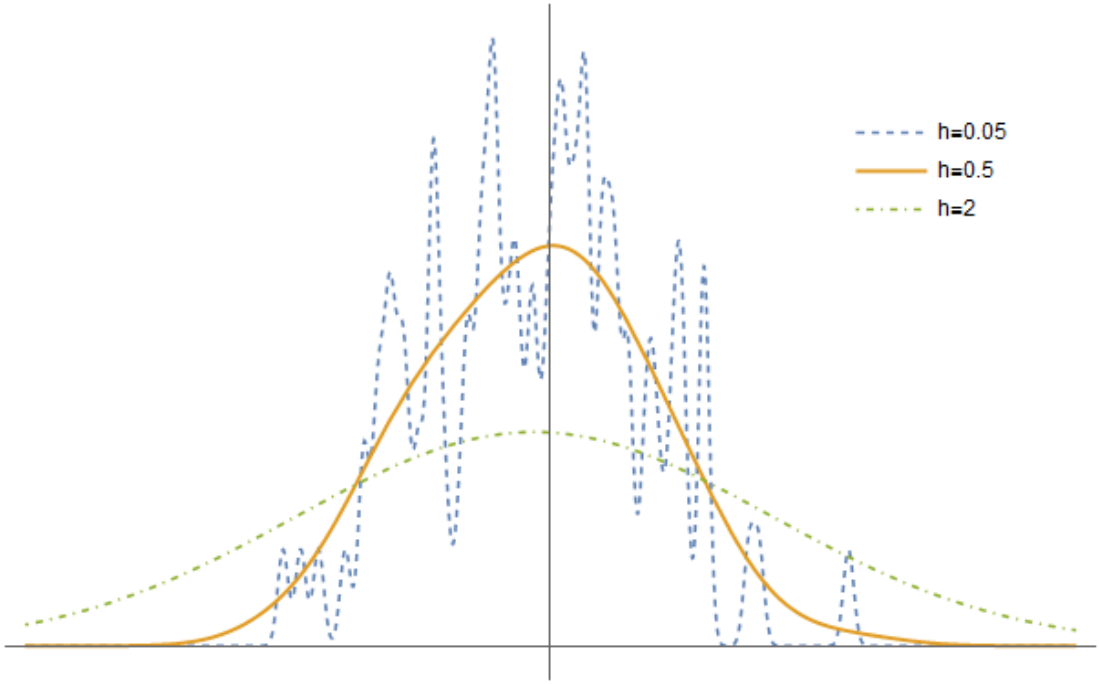


Figure 10: A random sample of size 100 was drawn from standard normal distribution. Apply kernel density estimation using Gaussian kernel with different h. When h is too small, the variance is large. When h is too large, the bias is large.

The choice of h, is a trade-off between variance and bias. The essential requirement of this method is $h \to 0$. However, as the figure above shows, when h is too small,

there will be a big variance.

A common method is to choose an h, which minimized the mean integrated square error of the estimation. One approach is called *least square cross-validation*.

The integral of the square error of the estimation give by

$$\int [\hat{f}(x) - f(x)]^2 dx = \int \hat{f}(x)^2 dx - 2\int \hat{f}(x) f(x) dx + \int f(x)^2 dx$$

Note the last term is constant. So, to minimize the expression, we only need to consider the first two terms, as following

$$\int \hat{f}(x)^2 dx - 2\int \hat{f}(x) f(x) dx$$

The second term could be treated as the expectation of $\hat{f}(h)$, $E[\hat{f}(h)]$. We could use the sample mean to replace the expectation,

$$\begin{aligned} \int \hat{f}(x) f(x) dx &= E[\hat{f}(h)] \\ &= \frac{1}{n} \sum_{i=1}^{n} \hat{f}_{-i}(X_i) \end{aligned}$$

where $\hat{f}_{-i}(X_i)$ is called "leave-one-out kernel estimator" of $f(X_i)$. Note it's an estimator of $f(X_i)$, not $\hat{f}(X_i)$. The estimator is defined as following,

$$\hat{f}_{-i}(X_i) = \frac{1}{(n-1)h} \sum_{j=1, j\neq i}^{n} k\left(\frac{X_i - X_j}{h}\right)$$

The first term called be estimated as following,

$$\begin{aligned}\int \hat{f}(x)^2 dx &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int k\left(\frac{X_i - x}{h}\right)k\left(\frac{X_j - x}{h}\right)dx \\ &= \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}\left(\frac{X_i - X_j}{h}\right)\end{aligned}$$

where

$$\bar{k}(v) = \int k(u)k(v-u)du$$

Now, to minimize the integrated square error using cross validation method is equivalent as minimizing the following expression with respect to h.

$$\begin{aligned}\mathrm{CV}(h) =& \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}\left(\frac{X_i - X_j}{h}\right) \\ &- \frac{2}{n(n-1)h}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}k\left(\frac{X_i - X_j}{h}\right)\end{aligned}$$

Now, we could extend this method to a higher dimension. To extend the kernel density estimation to d dimensions, a multivariate probability density function could be used as kernel, and a $d \times d$ matrix, say $\mathbf{H}$, is now the bandwidth, the smoothing factor.

An easier way is to use product of univariate kernels as multivariate kernel. Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample collected. Let d be the number of dimensions. Let $\mathbf{X}_{ij}$ be the jth entry in the random vector $\mathbf{X}_i$.

The estimation of multivariate probability density function is given as following

$$\hat{f}(\mathbf{x}) = \frac{1}{nh_1h_2\cdots h_d}\sum_{i=1}^{n}\mathbf{k}\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)$$

where

$$\mathbf{k}\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right) = \prod_{j=1}^{d} k\left(\frac{\mathbf{X}_{ij} - \mathbf{x}_j}{h_j}\right)$$

The way to choose bandwidth is similar as the univariate case. The function to minimize is

$$\begin{aligned}\mathrm{CV}(h_1, h_2, \cdots, h_d) =& \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{\mathbf{k}}(\mathbf{X}_i, \mathbf{X}_j) \\ &- \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\mathbf{k}(\mathbf{X}_i, \mathbf{X}_j)\end{aligned}$$

where

$$\mathbf{k}(\mathbf{X}_i, \mathbf{X}_j) = \prod_{s=1}^{d}\frac{1}{h_s}k\left(\frac{\mathbf{X}_{is} - \mathbf{X}_{js}}{h_s}\right)$$

$$\bar{\mathbf{k}}(\mathbf{X}_i, \mathbf{X}_j) = \prod_{s=1}^{d}\frac{1}{h_s}\bar{k}\left(\frac{\mathbf{X}_{is} - \mathbf{X}_{js}}{h_s}\right)$$

As we mentioned in the beginning of this chapter, we collect the adjusted close price in the past 54 years of 3 company,General Electric, DuPont, and IBM. We want to get an estimation of the joint probability density function of the annualized continuously compounded return from the daily closing price of these company. There is a total of 13,594 trading days from 1962 to 2015. Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{13594}$ be the closing price of these three company. $\mathbf{X}_i$ is a 3 dimensional vector. The annualized continuously compounded return, $\mathbf{u}$ is given by

$$\mathbf{U}_i = \frac{1}{\Delta t}\ln\left(\frac{X_{i+1}}{X_i}\right)$$

where ln is element-wise operation, and $\Delta t = \dfrac{2015 - 1962 + 1}{13594}$.
We use the probability density function of standard Gaussian distribution as our

kernel. To determine the bandwidth, we minimize the following function,

$$\begin{aligned}\mathrm{CV}(h_1,h_2,h_3) =& \frac{1}{13593^2}\sum_{i=1}^{13593}\sum_{j=1}^{13593}\prod_{s=1}^{3}\frac{1}{h_s}\bar{k}\left(\frac{\mathbf{u}_{is}-\mathbf{u}_{js}}{h_s}\right)\\ &-\frac{2}{13593\cdot 13592}\sum_{i=1}^{13593}\sum_{j=1,j\neq i}^{13593}\prod_{s=1}^{3}\frac{1}{h_s}k\left(\frac{\mathbf{u}_{is}-\mathbf{u}_{js}}{h_s}\right)\end{aligned}$$

We found the minimum of $\mathrm{CV}(h_1,h_2,h_3)$ at $h_1=0.455, h_2=0.427, h_3=0.5$.
Then, the estimation of joint probability density function is

$$\hat{f}(\mathbf{x}) = \frac{1}{13593\ h_1h_2h_3}\sum_{i=1}^{13593}\prod_{j=1}^{3}k_{Gauss}\left(\frac{\mathbf{U}_{ij}-\mathbf{u}_j}{h_j}\right)$$

By Kelly's criterion, the goal is to maximize the expected exponential return, $G(f_1,f_2,f_3)$, where f_i is the fraction spend on the ith asset.

$$\begin{aligned}G(f_1,f_2,f_3) =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\ln[1+\sum_{i=1}^{3}f_i(e^{\mathbf{u}_i\Delta t}-1)]]\\ &\frac{1}{13593\ h_1h_2h_3}\sum_{i=1}^{13593}\prod_{j=1}^{3}k_{Gauss}\left(\frac{\mathbf{U}_{ij}-\mathbf{u}_j}{h_j}\right)d\mathbf{u}_1d\mathbf{u}_2d\mathbf{u}_3\end{aligned}$$

The maximum of $G(f_1,f_2,f_3)$ was found at $f_1=0.6471, f_2=0.1765, f_3=0.1744$. The computing time to find the maximum took about 9 hours.
The result suggests to spend 64.71%, 17.65%, and 17.44% on GE, DuPont, and IBM. If the portfolio starts with an initial capital of 1 dollar at 1962, it will end up with 214.5 dollar in the end of 2015. The annual return is about 10.45%. GE, DuPont, and IBM have raised to 213 times, 54 times, and 59 times against their initial price, over 54 years. A graph of normalized capital has shown below.

If the initial capital is 1 dollar, the suggested portfolio has beats "buying and holding GE" by 1 dollar, after a 54-year investment.

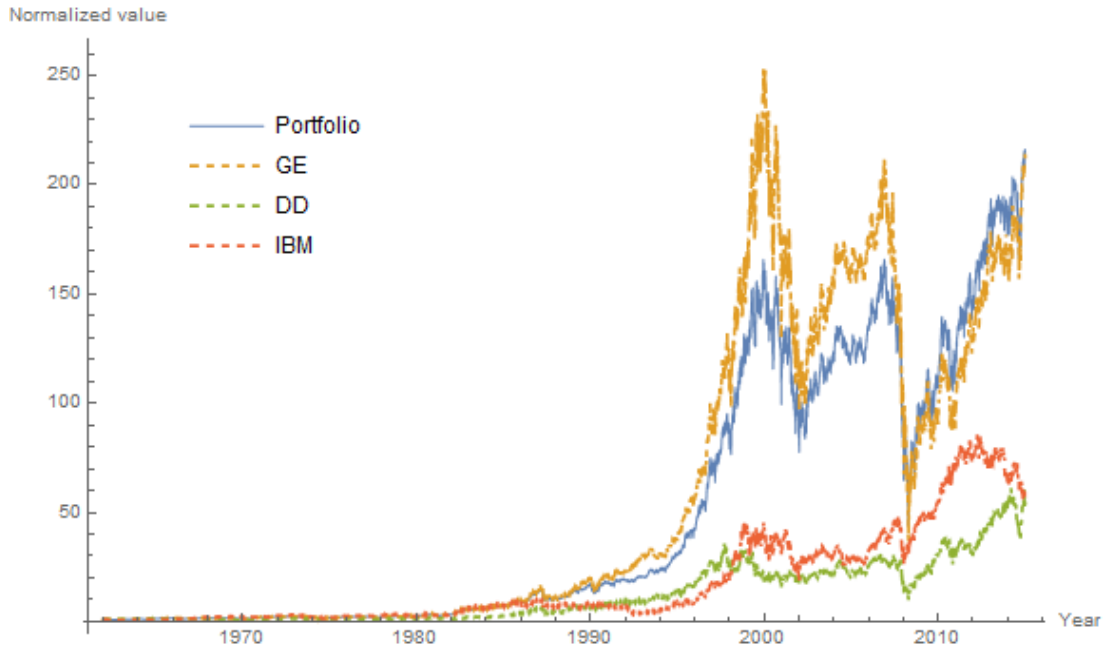


Figure 11: the normalized capital of GE, DuPont, IBM, and our portfolio

4.3 Copula

In previous section, a non-parametric method is used, which is completely data-driven. However, the non-parametric method is computationally expensive. The method using Taylor's expansion are limited and relatively inaccurate. The goal of this section is to drive our strategy using a method which is computationally less expensive with certain accuracy. The method being used is *copula*. It's a semi-parametric method.

A copula is a function which builds an connection between the marginal cumulative distribution functions and joint cumulative distribution function. The mathematical definition of copula is given by Nelson [15], as following

Definition 3. *C is an d-dimensional* **copula** *if,*

1. *The domain of C is $[0,1]^d$.*

2. *C is grounded and d-increasing.*

3. *The marginal distribution C_i satisfies $C_i = C(1, \cdots, 1, U_i, 1, \cdots, 1) = U_i$, where $U \in [0, 1]$ for all $i = 1, 2, \cdots, d$.*

Mostly, marginal cumulative distributions are used as variables in copula. Let $U = F_X(X)$, where X is some continuous distribution. Then U will has a uniform distribution. A brief proof is provided as follows.

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P(F_X(X) \leq u) \\ &= P(X \leq F_X^{-1}(u)) \\ &= F_X(F_X^{-1}(u)) \\ &= u \end{aligned}$$

Then, we could say $C(U_1, \cdots, U_n)$ is a n-dimensional distribution which its marginal distributions following uniform distribution on $[0, 1]$.
The existence and uniqueness of copula is guaranteed by Sklar's theorem [17] given below,

Theorem 5. *Let F be the joint cumulative distribution function of a multivariate distribution where it has $F_1, F_2, \cdots, F_d$ as marginal cumulative distributions. Then, there exist a copula C such that*

$$F(x_1, x_2, \cdots, x_d) = C(F_1(x_1), F_2(x_2), \cdots, F_d(x_d))$$

If $F_1, \cdots, F_d$ are continuous, then C is unique. Conversely, given $C : [0, 1]^d \rightarrow$

$[0,1]$ *and marginals* $F_1, \cdots, F_d$*, then* $C(F_1(\cdot), \cdots, F_d(\cdots))$ *defines a d-dimensional cumulative distribution function.*

Next, some common copula are introduced based on the Nelson's book, "*An introduction to copulas*" [15].

The n-dimensional Gauss copula is defined as

$$C_{gauss}(u_1, u_2, \cdots, u_d; \Sigma) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \cdots, \Phi^{-1}(u_d))$$

where Σ is a positive definite matrix, Φ_Σ is an n-dimensional Gaussian distribution function with correlation matrix Σ, and $\Phi^{-1}(\cdot)$ is a inverse function of Gaussian distribution function. Note there could be infinitely many Gauss copula. To keep the model simple, we use the standard Gaussian distribution function in the definition above.

The n-dimensional t-copula is defined as

$$C_T(u_1, u_2, \cdots, u_d; \Sigma, v) = T_{\Sigma,v}(T_v^{-1}(u_1), T_v^{-1}(u_2), \cdots, T_v^{-1}(u_d))$$

where Σ is a positive definite matrix, and v is the degree of freedom. $T_{\Sigma,v}$ is an n-dimensional t-distribution function with scale matrix Σ and degree of freedom v. T_v^{-1} is the inverse function of t-distribution function. Again, there could be infinitely many t-copula. We use the standard t-distribution function in the definition above.

An important family of copula is Archimedean copula, which defined as

$$C(u_1, u_2, \cdots, u_d; \alpha) = \phi^{-1}(\sum_{i=1}^{n} \phi(u_i; \alpha))$$

where ϕ is called generator, which is convex, strictly decreasing and $\phi(1;\alpha) = 0$. ϕ^{-1} is the inverse of generator.

There are three most commonly used Archimedean copulas which are Gumbel copula, Clayton copula, and Frank copula. The definition is based on Nelson's book, "*An introduction to copulas*" [15].

$$C_{Gumbel}(u_1, u_2, \cdots, u_d; \alpha) = \exp\left\{ -\left[\sum_{i=1}^{d} (-\ln u_i)^{\frac{1}{\alpha}} \right]^{\alpha} \right\}$$

where $\alpha \in (0, 1]$.

$$C_{Clyaton}(u_1, u_2, \cdots, u_d; \alpha) = \left(\sum_{i=1}^{d} u_i^{-\alpha} - d + 1 \right)^{-\frac{1}{\alpha}}$$

where $\alpha \in (0, \infty)$.

$$C_{Frank}(u_1, u_2, \cdots, u_d; \alpha) = -\frac{1}{\alpha} \ln\left(1 + \frac{\prod_{i=1}^{d} (e^{-\alpha u_i} - 1)}{(e^{-\alpha} - 1)^{d-1}} \right)$$

where $\alpha \in (0, \infty)$ for $n \geq 3$.

The advantage of Archimedean copula is the associativity. Also, no matter how high the dimension is, there is still only one parameter α needed to be estimated, which is impossible for Gauss copula and t-copula.

Note the product copula or independence copula is also an Archimedean copula which could be defined as following,

$$C_{Product}(u_1, u_2, \cdots, u_d) = \exp\left(-\sum_{i=1}^{n} -\ln(u_i) \right) = \prod_{i=1}^{n} u_i$$

where the generator is $-\ln(\cdot)$.

One common method of parameter calibration is maximum likelihood estimation. To

get the joint probability density function, using the definition from Sklar's theorem, we have

$$F(x_1, x_2, \cdots, x_d) = C(F_1(x_1), F_2(x_2), \cdots, F_d(x_d))$$
$$f(x_1, x_2, \cdots, x_d) = c(F_1(x_{1i}), F_2(x_{2i}), \cdots, F_d(x_{di})) \prod_{i=1}^{n} f_i(x_i)$$

where

$$c(F_1(x_1), F_2(x_2), \cdots, F_d(x_d)) = \frac{\partial^d (C(F_1(x_1), F_2(x_2), \cdots, F_d(x_d)))}{\partial F_1(x_1), \partial F_2(x_2) \cdots F_d(x_d)}$$

Again, like in previous section, let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample collected. Let d be the number of dimensions. Let x_{ij} be the jth entry in the random vector $\mathbf{X}_i$. The log likelihood function could be written in the following form,

$$\begin{aligned} l(\theta_1, \cdots, \theta_d, \alpha) = &\sum_{i=1}^{n} \ln c\left(F_1(x_{i1}; \theta_1), F_2(x_{i2}; \theta_2), \cdots, F_d(x_{id}; \theta_d); \alpha\right) \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{d} \ln f_j(x_{ij}; \theta_j) \end{aligned}$$

where θ_j is the set of parameters for jth marginal distribution, α is the set of parameters for copula. The maximum likelihood estimators are found by

$$(\hat{\theta_1}, \cdots, \hat{\theta_d}, \hat{\alpha}) = \text{argmax} \;\; l(\theta_1, \cdots, \theta_d, \alpha)$$

Clearly, this maximization problem has $d + 1$ dimensions, or higher. This problem is computationally expensive in most cases.

Another method is introduced by Joe and Xu [10], called inference for margins method. It breaks the problems to $d + 1$ separate problems.

First estimate the parameters for margins, using MLE method,

$$\hat{\theta}_j = \text{argmax}_{\theta_j} \sum_{i=1}^{n} \ln f_j(x_{ij}; \theta_j)$$

Then, given $(\hat{\theta}_1, \cdots, \hat{\theta}_d)$, the estimation of parameters for copula is given by

$$\hat{\alpha} = \text{argmax}_{\alpha} \sum_{i=1}^{n} \ln c(F_1(x_{1i}), F_2(x_{2i}), \cdots, F_d(x_{di}); \alpha, \hat{\theta}_1, \cdots, \hat{\theta}_d)$$

Note the MLE method and IFM method are not equivalent. Although the IFM method is not as expensive as MLE method, it's still expensive comparing to the method we are going to introduced below.

The pseudo maximum likelihood method is based on MLE method with empirical distribution function. The problem will be simplified to only solving a set of parameters for copula. For Archimedean copula, this problem will be reduced one dimensional.

$$\hat{\alpha} = \text{argmax}_{\alpha} \sum_{i=1}^{n} \ln c(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}), \cdots, \hat{F}_d(x_{di}); \alpha)$$

where $\hat{F}_j(\cdot)$ is the empirical distribution function of the jth marginal distribution. Note the term $\sum_{i=1}^{n} \sum_{j=1}^{d} \ln f_j(x_{ij}; \theta_j)$ is eliminated because the empirical distribution is discrete.

This maximization is computationally much cheaper than MLE method and IFM method. We will use the pseudo maximum likelihood method in the later computation.

We have introduced some common methods for estimating parameters for a given copula. Next, some common goodness-of-fit tests used to compare among different copulas would be introduced.

Three goodness-of-fit tests would be introduced below, including Kolmogorov-Smirnov

test, Anderson-Darling test, and least distance test.

$$T_{KS} = \max|\hat{F}(x_i) - F(x_i)|$$

$$T_{AD} = \max \frac{|\hat{F}(x_i) - F(x_i)|}{\sqrt{F(x_i)[1 - F(x_i)]}}$$

$$T_{ld} = \sqrt{\sum_{i=1}^{n} \left(\hat{F}(x_i) - F(x_i)\right)}$$

where $\hat{F}(\cdot)$ is the empirical distribution function. All three tests are testing the difference between observed values and theoretical values. The KS test looks at the maximum distance, while the AD test is similar but putting more weights on the tails. For all three test, a smaller statistic means a better goodness of fit.

Applying all these methods to the same data used in the previous section, i.e., the data collected from GE, DuPont, and IBM. Estimate the parameter for different copulas using pseudo maximum likelihood method, we have

Copula	parameter
Frank	0.045731796
Clayton	1.5553922
Gumbel	1.4548229

The same method is also applied on multivariate Gaussian copula and multivariate t-copula, each of which including a 3×3 covariance or scale matrix as parameters. Then the goodness of fit tests are applied on different copulas.

Copula	Log-likelihood	Kolmogorov-Smirnov	Anderson-Darling	Least distance
Product	-377515	0.123772	116.589	8.04062
Frank	-373373	0.0230184	36.0705	0.988535
Clayton	-373350	0.0541942	0.110593	2.39594
Gumbel	-373280	0.0254937	1.83818	1.27237
Multi-Gauss	-372981	0.0126158	0.266305	0.557138
Multi-t	-371745	0.0079978	0.0287249	0.348389

As the table above shows, multivariate t-copula is the copula which has the best goodness of fit among all the copula tested by all the tests we used.

The marginal distribution of realized return should be normal distribution by the Black-Scholes-Merton model. However, since we are constructing the joint distribution without prior assumption, we would like to find the most likely distribution on real data. First, estimate the parameters for different distributions using maximum likelihood method. The results is given below,

	Parameters		
Distributions	GE	DuPont	IBM
Student-t	(0.479, 2.761,2.388)	(0.179, 2.483, 2.781)	(0.136, 2.701, 2.344)
Normal	(0.608973,6.17403)	(0.0439971, 5.03209)	(0.223546, 6.76036)
Laplace	(0.0993256,2.84451)	(0.0739564, 2.85368)	(0.223546, 3.53709)
Cauchy	(-0.216721,1.82377)	(0.181595, 1.84384)	(0.0146807, 1.89978)
Logistic	(0.495667,2.53624)	(0.180254, 2.10459)	(0.248098, 2.55526)

Then we compare the log-likelihood of the marginal distributions with parameters estimated. The results are given in the following table.

	Log-likelihood		
Distributions	GE	DuPont	IBM
Student's t	-2.74776	-2.74377	-2.72572
Normal	-2.96065	-2.85418	-3.00632
Laplace	-2.73854	-2.74176	-2.74977
Cauchy	-2.82782	-2.8363	-2.80925
Logistic	-2.77879	-2.75019	-2.76427

As the table above shows, for all three companies, the annualized realized returns are most unlikely to be normal distribution among all five distribution we chose. It seems like that GE and DuPont follows a Laplace distribution and IBM follows a Student's t-distribution.

Gather everything we've done to construct the joint cumulative distribution function. The cumulative distribution function of the annualized realized return for GE, DuPont, and IBM are given below

$$\begin{aligned}
F_{GE}(x) =& \frac{1}{2} + \frac{1}{2}\left(\left\lfloor \frac{x - 0.0993256}{|x - 0.0993256| + 1} \right\rfloor - \left\lfloor \frac{0.0993256 - x}{|0.0993256 - x| + 1} \right\rfloor\right) \\
& \cdot \left[1 - \exp\left(-\frac{|x - 0.0993256|}{2.84451}\right)\right] \\
F_{DD}(x) =& \frac{1}{2} + \frac{1}{2}\left(\left\lfloor \frac{x - (0.0739564}{|x - (0.0739564| + 1} \right\rfloor - \left\lfloor \frac{(0.0739564 - x}{|(0.0739564 - x| + 1} \right\rfloor\right) \\
& \cdot \left[1 - \exp\left(-\frac{|x - (0.0739564|}{2.85368}\right)\right] \\
F_{IBM}(x) =& 1 - \frac{1}{2}\ I\left(\frac{2.34382}{\left(\frac{x - 0.136051}{2.70129}\right)^2 + 2.34382}, \frac{2.34382}{2}, \frac{1}{2}\right)
\end{aligned}$$

where $I(\cdot, \cdot, \cdot)$ is the regularized incomplete beta function, defined as following.

$$I(x,a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{\Gamma(a)\Gamma(b)}\Gamma(a+b)$$

Remember the formula for t-copula is given as,

$$C_T(u_1,u_2,\cdots,u_d;\Sigma,v) = T_{\Sigma,v}(T_v^{-1}(u_1),(T_v^{-1}(u_2),\cdots,(T_v^{-1}(u_d))$$

By maximum likelihood method, we have the estimations of parameters,

$$\Sigma = \begin{pmatrix} 8.13382 & 3.87702 & 3.65086 \\ 3.87702 & 8.257 & 3.31765 \\ 3.65086 & 3.31765 & 7.66551 \end{pmatrix}, \quad v = 3.82477$$

For the inverse function used in copula above, the definition of quantile function of Student's t distribution with degree of freedom v is give as

$$T_v^{-1}(x) = \begin{cases} -\sqrt{\nu}\sqrt{\frac{1}{I^{-1}\left(2x,\frac{\nu}{2},\frac{1}{2}\right)}-1} & \text{if } 0<x<\frac{1}{2} \\ 0 & \text{if } x=\frac{1}{2} \\ \sqrt{\nu}\sqrt{\frac{1}{I^{-1}\left(2(1-x),\frac{\nu}{2},\frac{1}{2}\right)}-1} & \text{if } \frac{1}{2}<x<1 \end{cases}$$

where $I^{-1}(\cdot,\cdot,\cdot)$ is the inverse beta function. $I^{-1}(x,a,b)$ gives the solution for y in $I(y,a,b)=x$, where I is the regularized incomplete beta function.

By gathering everything we have, we could build a joint distribution using t-copula. Then we could trying to maximize the expected exponential return of the portfolio,

as following.

$$G(f_1, f_2, f_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\ln[1+\sum_{i=1}^{3} f_i(e^{u_i\Delta t}-1)]] \\ \frac{\partial^d(T_{\Sigma,v}(T_v^{-1}(F_1(u_1)),(T_v^{-1}(F_2(u_2)),(T_v^{-1}(F_3(u_3)))))}{\partial F_1(u_1)\partial F_2(u_2)F_d(u_3)}\prod_{i=1}^{3} f_i(u_i)du_1du_2du_3$$

However, when I put everything together, the algorithm is extremely slow. There are mainly two problems.

First, the multivariate t-distribution does not have an analytic expression for its cumulative distribution function. If we still use numerical integration for the above expression, for every time we try to get the result numerically, we have to do a Monte Carlo integration for to estimate $T_{\Sigma,v}$. This means we are putting a Monte Carlo integration in a Monte Carlo integration. The computational time will be multiplied.

Second, to numerical integrate the above expression, we have to randomly choose samples. For each sample vector we chosen, we have to get the result of T_v^{-1}, which contains the inverse of the regularized incomplete beta function. I didn't find any other ways of solving this problem but numerical method.

Finally, it takes more than an hour to finish one single integral. To maximize the expression with respect to f_1, f_2, f_3. It seems like that it will take more than a week to compute.

Since these two problems are both caused by the use of t-copula, it's probably better to use another copula until we could solve these problem. By the goodness of fit tests we have done before, it seems like that the second best copula is Gaussian copula. The definition of Gaussian copula is given as following, as a reminder.

$$C_{gauss}(u_1, u_2, \cdots, u_d; \Sigma) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \cdots, \Phi^{-1}(u_d))$$

The estimation of parameter is given as following, by maximum likelihood method.

$$\Sigma = \begin{pmatrix} 16.6162 & 8.10574 & 7.59306 \\ 8.10574 & 16.1752 & 6.4075 \\ 7.59306 & 6.4075 & 16.0932 \end{pmatrix}$$

For the inverse function used in copula above, the definition of quantile function of standard normal distribution is give as

$$\Phi^{-1}(x) = -\sqrt{2}\text{erfc}^{-1}(2x)$$

where the erfc^{-1} is the inverse complementary error function. Gathering everything together, we want to maximize the expected exponential growth rate given as the following expression.

$$G(f_1, f_2, f_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\ln[1 + \sum_{i=1}^{3} f_i(e^{u_i \Delta t} - 1)]] \\ \frac{\partial^d(\Phi_\Sigma(\Phi^{-1}(F_1(u_1)), \Phi^{-1}(F_2(u_2)), \Phi^{-1}(F_3(u_3))))}{\partial F_1(u_1)\partial F_2(u_2)F_d(u_3)} \prod_{i=1}^{3} f_i(u_i)du_1du_2du_3$$

Note there is also no analytic expression for cumulative distribution function of multivariate normal distribution or inverse complementary error function. However, it seems that Mathematica does a much better job a normal distribution than t-distribution. The maximization is pretty fast in this case. The computation takes about 5 hours in total. The maximum of $G(f_1, f_2, f_3)$ was found at $f_1 = 0.8189, f_2 = 0.0694, f_3 = 0.1117$.

The result suggests to spend 81.89% , 6.94% , and 11.17% on GE, DuPont, and IBM. If the portfolio starts with an initial capital of 1 dollar at 1962, it will end up with 224.2 dollar in the end of 2015. The annual return is about 10.54%. Again, as a

reminder, GE, DuPont, and IBM have raised to 213 times, 54 times, and 59 times against their initial price over 54 years. A graph of normalized capital has shown below.

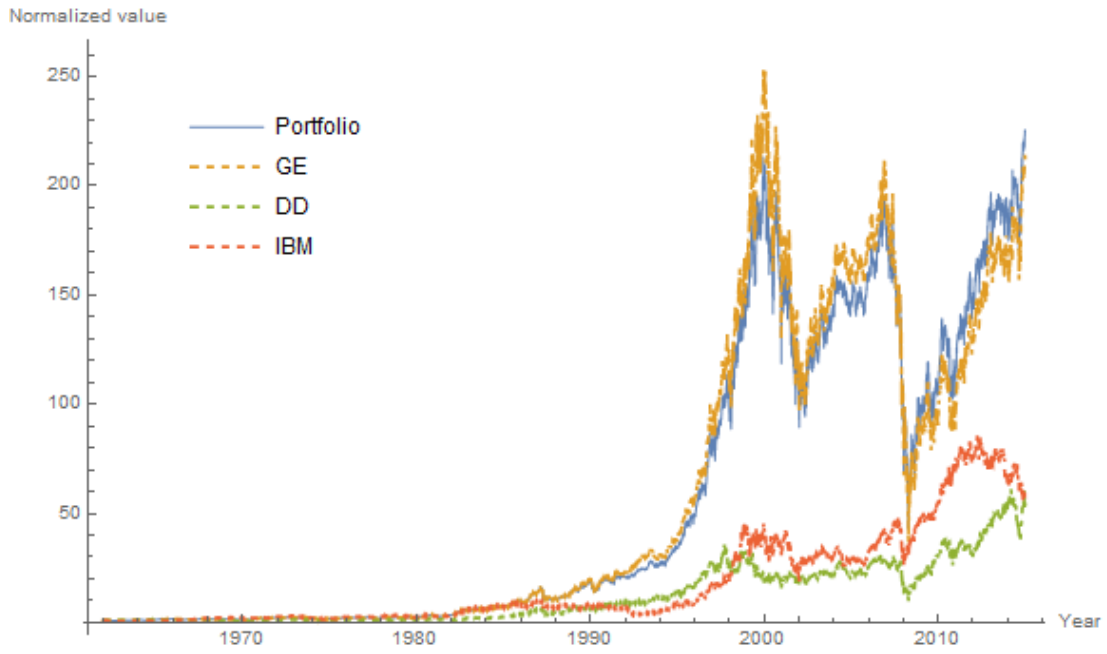


Figure 12: the normalized capital of GE, DuPont, IBM, and our portfolio

The final result is pretty close to what we have gotten with the use of kernel function in previous section.

5 Extension to option

As we have seen in the previous section, building a portfolio with only stocks would cause a significant decay during the bear market, such as 2008 to 2009. It's natural to extend the portfolio with the use of options, as hedging tools.

Let $f_1,\ f_2,\ f_3$ be the fraction of capital spent on stock Ge, DD, and IBM. Let $f_{11},\ f_{21},\ f_{31}$ be the fraction of capital spent on the call options of corresponding stock, and let $f_{12},\ f_{22},\ f_{32}$ be the fraction of captical spent on the put options.

Let k_{11} be the strike price of call option of GE, such that

$$k_{11} = c_{11} S_1(t)$$

where c_{11} is a constant ratio of strike price of call to spot price of GE. Likewise, let c_{12} be the ratio of strike price of put to spot price of GE, and so as $c_{21},\ c_{22},\ c_{31},\ c_{32}$. Now the goal is to maximize the expected exponential growth rate of our portfolio, with respect to $f_1,\ f_2,\ f_3,\ f_{11},\ f_{21},\ f_{31},\ f_{12},\ f_{22},\ f_{32},\ c_{11},\ c_{21},\ c_{31},\ c_{12},\ c_{22},\ c_{32}$. This is a 15-dimensional optimization problem. In order to speed up the optimization, it would be better to derive analytic expression in the process.

5.1 Joint distribution

Note no matter what copula we use, the joint probability density function is derived by

$$f(x_1, x_2, \cdots, x_d) = \frac{\partial^d (C(F_1(x_1), F_2(x_2), \cdots, F_d(x_d)))}{\partial F_1(x_1), \partial F_2(x_2) \cdots F_d(x_d)} \prod_{i=1}^{d} f_i(x_i)$$

The probability density function of each stock is given as following.

$$
\begin{aligned}
f_{GE} &= 0.175777e^{-0.351555|x-0.0993256|} \\
f_{DD} &= 0.175212e^{-0.350424|x-0.0739564|} \\
f_{IBM} &= 0.956126\left(\frac{1}{0.136375(x+0.196066)^2+2.80129}\right)^{1.90064}
\end{aligned}
$$

The cumulative distribution function of each stock is given by,

$$
\begin{aligned}
F_{GE}(x) =& \frac{1}{2}+\frac{1}{2}\mathrm{sgn}(\mathrm{x}-0.0993256)\left(1-e^{-0.351555|\mathrm{x}-0.0993256|}\right) \\
F_{DD}(x) =& \frac{1}{2}+\frac{1}{2}\mathrm{sgn}(\mathrm{x}-0.0739564)\left(1-e^{-0.350424|\mathrm{x}-0.0739564|}\right) \\
F_{IBM}(x) =& \frac{1}{2}+\frac{1}{2}\mathrm{sgn}(\mathrm{x}+0.196066) \\
& \cdot\left(1-I\left(\frac{2.80129}{0.136375(\mathrm{x}+0.196066)^2+2.80129}, 1.40064, \frac{1}{2}\right)\right)
\end{aligned}
$$

where $I(\cdot,\cdot,\cdot)$ is the regularized incomplete beta function, defined as following.

In order to improve the speed of the optimization algorithm, it's probabily better to use Archimedean copula rather than Gaussian copula or Student-t copula. According to the goodness-of-fit test done in previous chapter, the Clayton copula is chosen to be used.

The generator function of clayton copula is

$$g(t;\theta) = \frac{1}{\theta}\left(t^{-\theta}-1\right)$$

And the generator inverse is given by

$$g^{-1}(t;\theta) = (1+\theta t)^{-1/\theta}$$

By simple algebra, we get the copula desenty function as

$$
\begin{aligned}
&\frac{\partial^3}{\partial F_1(x_1)\partial F_2(x_2)F_3(x_3)}C(F_1(x_1),F_2(x_2),F_3(x_3);\theta)\\
=&\Big(-\frac{1}{\theta}-2\Big)\Big(-\frac{1}{\theta}-1\Big)\theta^2\left(\frac{1}{2}\Big(1-e^{-0.351555|x_1-0.0993256|}\Big)\mathrm{sgn}(x_1-0.0993256)+\frac{1}{2}\right)^{-\theta-1}\\
&\cdot\left(\frac{1}{2}\Big(1-e^{-0.350424|x_2-0.0739564|}\Big)\mathrm{sgn}(x_2-0.0739564)+\frac{1}{2}\right)^{-\theta-1}\\
&\cdot\left(\frac{1}{2}\mathrm{sgn}(x_3+0.196066)\Big(1-I\Big(\frac{2.80129}{0.136375(x_3+0.196066)^2+2.80129},1.40064,\frac{1}{2}\Big)\Big)\right.\\
&\left.+\frac{1}{2}\right)^{-\theta-1}\cdot\left[\left(\left(\frac{1}{2}\Big(1-e^{-0.351555|x_1-0.0993256|}\Big)\mathrm{sgn}(x_1-0.0993256)+\frac{1}{2}\right)^{-\theta}\right.\right.\\
&+\left(\frac{1}{2}\Big(1-e^{-0.350424|x_2-0.0739564|}\Big)\mathrm{sgn}(x_2-0.0739564)+\frac{1}{2}\right)^{-\theta}\\
&+\left(\frac{1}{2}\mathrm{sgn}(x_3+0.196066)\Big(1-I\Big(\frac{2.80129}{0.136375(x_3+0.196066)^2+2.80129},1.40064,\frac{1}{2}\Big)\Big)\right.\\
&\left.\left.\left.+\frac{1}{2}\right)^{-\theta}\right)-2\right]^{-\frac{1}{\theta}-3}
\end{aligned}
$$

By maximizing the log likelihood function of the joint probability density function, we get

$$\theta = 0.650734$$

Thus, the final joint probability density function could be easily build up as the product of the copula densety function and marginal probability density functions.

5.2 Results

As long as we have the joint distribution function, we could simply build up the function of expected exponential growth rate, our target function.

The return of a call option, denoted as r_{call}, is a function of drift u, spot/strike ration c, volitility σ, expeiration date T, current time t, and time length of hold the option

Δt. It's simply the ratio of the future value of option to current value. Then, the expected exponential growth rate is

$$
\begin{aligned}
&G(f_i, f_{i1}, f_{i2}, c_{i1}, c_{i2}; m) \\
&= \int\int\int_{-\infty}^{\infty} \ln\left(1 + \sum_{j=1}^{2}\sum_{i=1}^{3}\left(f_i e^{u_i \Delta t} + f_{ij}\Big(r_{call}(u_i, c_{ij}, \sigma_i, m\Delta t) - 1\Big)\right)\right)\mathbf{jointpdf}(\mathbf{u})\ d\mathbf{u}
\end{aligned}
$$

Again, f_i is the fraction of capital spent on stock i; f_{i1}, is the fraction of capital spent on the call options of corresponding stock; f_{i2} is the fraction of captical spent on the put options and m is the number of days to maturity with Δt set to be one day.

To be consistent, the maturity will always be the second soonest Fridays, and the difference of strike price to spot price is set to be less than 25%. The possible number of days to maturity is from 3 to 17, since 1962. So we need to solve the maximum of our target function 15 times with $m = 3, \cdots, 17$.

Note this is a set of 15-dimensional optimization problem, I did try several methods, such as simulatied annealing, evolution, Nelder-Mead, but none of them returns a feasible solution. However, with Bayesian optimization method, the problem has been easily solved, and the result shows as following,

f_1	f_2	f_3	f_{11}	f_{21}	f_{31}	f_{12}	f_{22}	f_{32}	c_{11}	c_{21}	c_{31}	c_{12}	c_{22}	c_{32}
0.07	0.04	0.02	0.42	0.05	0.01	0.04	0.09	0.03	0.94	1.01	1.13	0.89	1.	1.08
0.18	0.04	0.04	0.	0.33	0.05	0.01	0.09	0.19	1.21	0.91	1.19	1.25	0.96	1.
0.05	0.43	0.05	0.19	0.06	0.01	0.04	0.02	0.07	0.86	0.81	1.16	0.97	1.03	0.85
0.12	0.18	0.	0.03	0.09	0.03	0.06	0.06	0.03	1.02	0.76	1.06	1.1	0.88	0.78
0.1	0.	0.12	0.01	0.1	0.09	0.15	0.35	0.01	1.16	1.05	1.05	1.15	1.09	0.75
0.04	0.3	0.24	0.16	0.06	0.01	0.	0.01	0.02	1.	1.07	1.19	1.22	1.18	0.76
0.13	0.31	0.02	0.05	0.02	0.12	0.05	0.	0.04	1.19	1.18	1.07	1.2	1.15	0.79
0.02	0.13	0.03	0.02	0.18	0.04	0.02	0.23	0.07	1.18	0.89	1.18	0.96	1.06	0.76
0.05	0.03	0.4	0.18	0.08	0.04	0.01	0.02	0.02	0.9	1.09	1.19	1.02	1.1	0.76
0.27	0.04	0.16	0.08	0.07	0.08	0.02	0.08	0.07	0.84	0.94	1.11	0.88	1.15	0.77
0.12	0.04	0.09	0.36	0.09	0.02	0.02	0.03	0.04	0.89	1.05	0.92	1.01	0.84	0.76
0.12	0.24	0.1	0.07	0.23	0.07	0.02	0.01	0.06	1.18	0.92	1.	1.21	0.76	0.77
0.15	0.07	0.36	0.03	0.14	0.07	0.02	0.07	0.09	1.23	0.78	1.23	0.85	1.03	0.76
0.01	0.16	0.28	0.32	0.08	0.03	0.03	0.02	0.04	0.97	0.92	1.21	1.19	1.2	0.77
0.12	0.3	0.	0.22	0.04	0.07	0.02	0.01	0.16	0.8	1.11	1.24	1.21	0.99	0.87

The table above shows the feasible results of $m = 3, \cdots, 17$ from top to bottom.

Run the simulation throught the historical data with the portfolio above, our capital will rise 1528 times against the initial capital, over 54 years. It is also 7 times better than the pure stock strategy we did in chapter 4.

However, when we plotted the graph out, we found the performence of our strategy is highly oscillating.

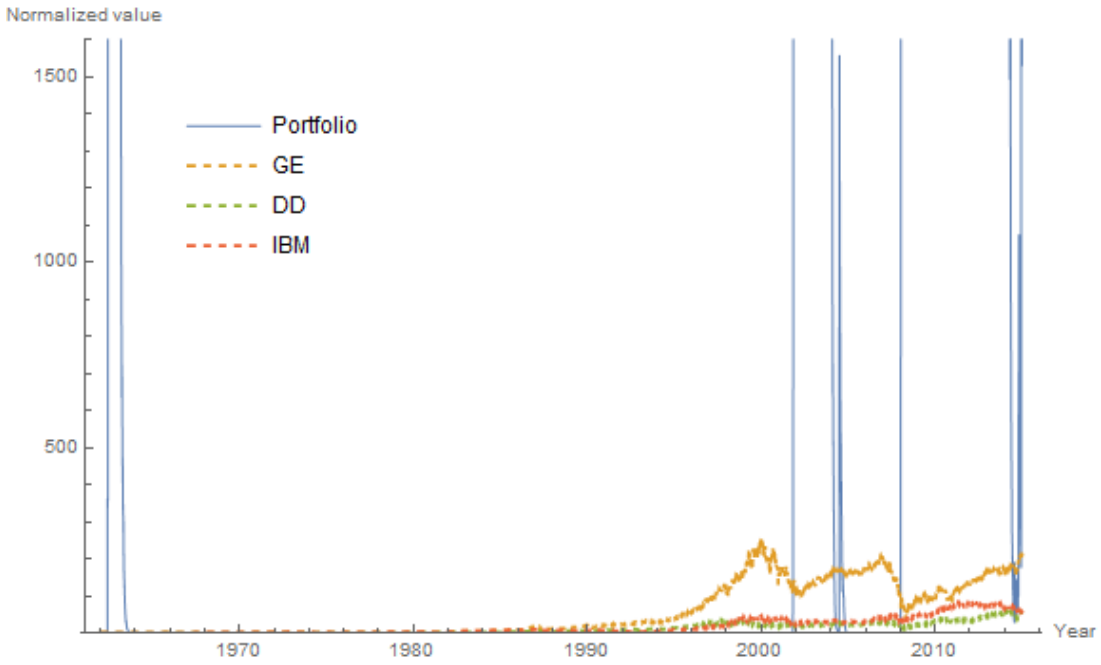


Figure 13: the normalized capital of GE, DuPont, IBM, and our option strategy

The reason might be that the fraction spent on options is too heavy, which makes it no longer a hedge tool, but a high risk speculation.

To fix this problem, we set the maximum fraction investing on options at 2%, then run the simulation again. The result shows that, at the end of 2015, our portfolio has raised 1.94×10^{64} times against the initial capital over 54 years.

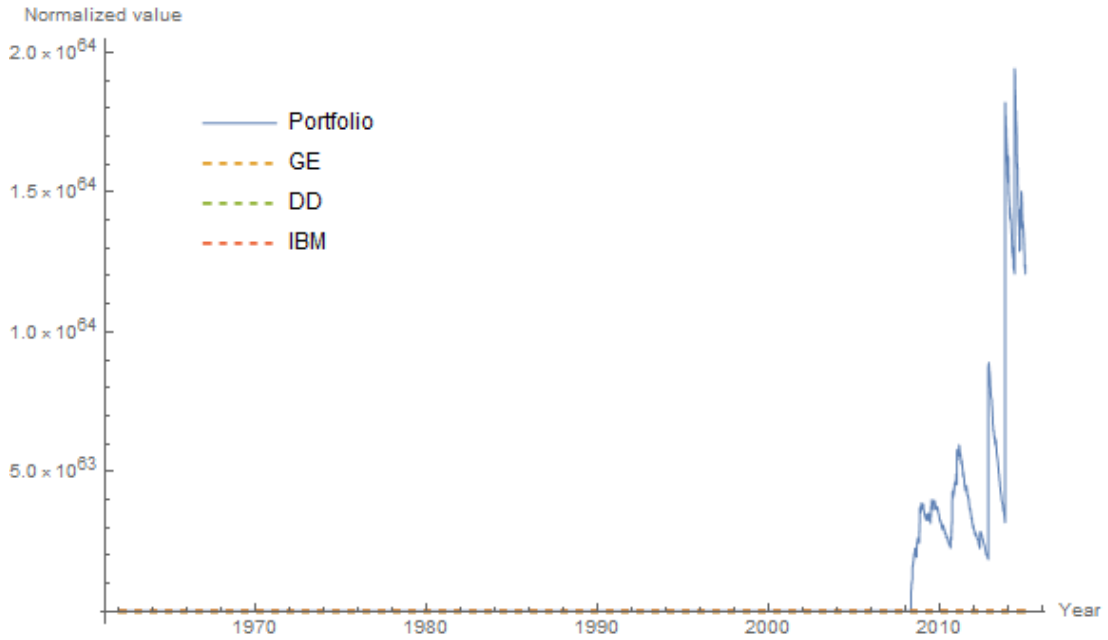


Figure 14: the normalized capital of GE, DuPont, IBM, and our new option strategy from 1962 to 2015

The maximum daily return occurs on Oct 19th, 1987. On that day, IBM's share price falled 23.52%, and our portolio spent .028% of our capital on a IBM's put which is 22.45% out-the-money. The portfolio made 2.29×10^{11}% profit on the day. However, it is not practical in reality. We use adjusted price for historical price to eliminate the effects of split and dividend, which makes it being hard to set an minimum currency unit. So,in our simulation, the price of put we bought on Oct. 16th, 1987 is just 3.62×10^{-14} dollar, while in practice it will be at least 1 cent.

So far, we always test our strategy on the same set of data we trained it. Next we tried to run the simulation on the data of 2016.

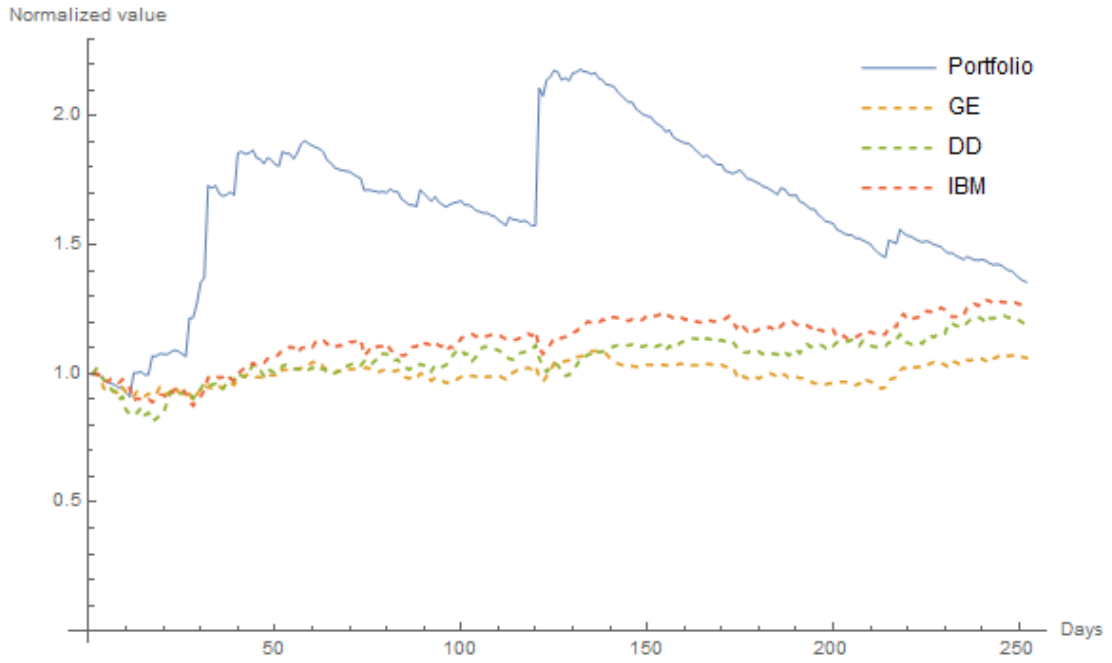


Figure 15: the normalized capital of GE, DuPont, IBM, and our new option strategy during 2016

The maximum daily return occurs on Jun 24th, 2016. On that day, IBM's share price falled 5.6%, and our portolio spent .31% of our capital on a IBM's put which is 8.4% out-the-money. The portfolio made 33.79% profit on the day. In our simulation, the price of put we bought on one day before is just 0.254 cent, while in practice it will be at least 1 cent. So it's less practial in reality.

Our portfolio prefered out-the-money option rather than in-the-money option. So, on the graph, we are constantly lossing about 1% every day for options end up worthless. This is probabily caused by the inefficient of optimization algorithm we used to solve our 15-dimensional target problem.

6 Conclusion

In this paper, we attempt to derive a strategy of maximizing profit in the long run. We introduced some basic concept on Mathematical Finance and Kelly's criterion. We tested Kelly's criterion on some ideal cases, then extend its concept to more practical case, which helps us testing its feasibility.

In the simulation of our strategy, we tried to build distributions on historical return of correlated stocks. A moedl-free approach, a nonparametric approach and copulas were used in building the joint distribution. A possible further improvement is to turn the model from still distribution to stochastic process, such as Hidden Markov Process, which could be used to tell the market status over time.

The difficulties during the derivation of the strategy are building up a more accurate model and find a more efficient algorithm to sovle the optimal strtegy. It is clear that there is still plenty of room of improving the optimization algorithm.

The final result shows there exists feasibility of applying the concept of Kelly's criterion on large and complex portfolio.

Appendices

A Mathematica codes

A.1 Binomial case

```
Table[{Times @@ #, Times @@ (1 + 1/6 (# - 1))} &@
  Part[RandomChoice[{4/3, 0.7}, 999], 1 ;; i], {i, 0, 999}]
```

A.2 Univariate stock

Solve for maximum of the target function:

```
GStock[f_?NumericQ, mu_?NumericQ, sigma_?NumericQ] :=
 NIntegrate[
  Log[1 + f (Exp[u /252 ] -
        1)] Exp[-(u - (mu - sigma^2/2))^2 /(2 sigma^2 252) ]/
    Sqrt[2 Pi sigma^2 252 ], {u, -Infinity, Infinity},
  AccuracyGoal -> 5, PrecisionGoal -> 4, MaxRecursion -> 0]
f /. Last@NMaximize[GStock[f, 0.05, 0.3], f]
```

Generate stock data us Geometric Brownian motion:

```
data = Part[
 RandomFunction[
  GeometricBrownianMotionProcess[0.05, 0.3, 1], {0, 100 - 1/252,
   1/252}], 2, 1, 1];
```

Daily return:

```
r = (Drop[#, 1] - Drop[#, -1])/Drop[#, -1] &@data;
```

Simulation:

```
First@Transpose@
NestList[{#[[1]] (1 + 0.555562 Part[r, #[[2]]]), #[[2]] + 1} &, {1,
  1}, 4 252 - 1]
```

A.3 Uncorrelated multivariate stocks

Solve for maximum of the target function:

```
G3Stock[f1_?NumericQ, mu1_?NumericQ, sigma1_?NumericQ, f2_?NumericQ,
mu2_?NumericQ, sigma2_?NumericQ, f3_?NumericQ, mu3_?NumericQ,
sigma3_?NumericQ] :=
NIntegrate[
Log[1 + f1 (Exp[u1 /252 ] - 1) + f2 (Exp[u2 /252 ] - 1) +
   f3 (Exp[u3 /252 ] -
      1)] Exp[-(u1 - (mu1 - sigma1^2/2))^2 /(2 sigma1^2 252) ]/
  Sqrt[2 Pi sigma1^2 252 ] Exp[-(u2 - (mu2 -
         sigma2^2/2))^2 /(2 sigma2^2 252) ]/
  Sqrt[2 Pi sigma2^2 252 ] Exp[-(u3 - (mu3 -
         sigma3^2/2))^2 /(2 sigma3^2 252) ]/
  Sqrt[2 Pi sigma3^2 252 ], {u1, -Infinity,
 Infinity}, {u2, -Infinity, Infinity}, {u3, -Infinity, Infinity},
AccuracyGoal -> 5, PrecisionGoal -> 4, MaxRecursion -> 0]
NMaximize[{G3Stock[f1, 0.1, 0.4, f2, 0.07, 0.25, f3, 0.04, 0.1],
f1 >= 0 && f2 >= 0 && f3 >= 0 && f1 + f2 + f3 <= 1}, {f1, f2, f3}]
```

Generate stock data us Geometric Brownian motion:

```
data1 = RandomFunction[
 GeometricBrownianMotionProcess[0.10, 0.4, 1], {0, 100 - 1/252,
  1/252}];
data2 = RandomFunction[
 GeometricBrownianMotionProcess[0.07, 0.25, 1], {0, 100 - 1/252,
  1/252}];
data3 = RandomFunction[
 GeometricBrownianMotionProcess[0.04, 0.10, 1], {0, 100 - 1/252,
  1/252}];
```

Daily return:

```
r1 = (Drop[#, 1] - Drop[#, -1])/Drop[#, -1] &@data1;
r2 = (Drop[#, 1] - Drop[#, -1])/Drop[#, -1] &@data2;
r3 = (Drop[#, 1] - Drop[#, -1])/Drop[#, -1] &@data3;
```

Simulation:

```
First@Transpose@
NestList[{#[[
     1]] (1 + 0.13 Part[r1, #[[2]]] + 0.607 Part[r2, #[[2]]] +
      0.263 Part[r3, #[[2]]]), #[[2]] + 1} &, {1, 1}, 4 252 - 1]
```

Sharpe's ratio:

```
r = Table[{0.13, 0.607, 0.263}.Part[Transpose@{r1, r2, r3}, i], {i, 1,
   252 100 - 1}];
ListLinePlot[{Table[
```

```
    Mean[#]/StandardDeviation[#] &@Part[r1, 1 ;; n], {n, 100, 25199,
     100}], Table[
    Mean[#]/StandardDeviation[#] &@Part[r2, 1 ;; n], {n, 100, 25199,
     100}], Table[
    Mean[#]/StandardDeviation[#] &@Part[r3, 1 ;; n], {n, 100, 25199,
     100}],
   Table[Mean[#]/StandardDeviation[#] &@Part[r, 1 ;; n], {n, 100,
     25199, 100}]},
  PlotStyle -> {{Dotted, Thin}, {Dotted, Thin}, {Dotted, Thin}, Thin},
  MaxPlotPoints -> 500,
  PlotLegends ->
   Placed[{"Stock1", "Stock2", "Stock3",
     "f={13%,60.7%,26.3%}"}, {0.5, .72}], PlotRange -> Full,
  DataRange -> {0, 100}, ImageSize -> Large,
  AxesLabel -> {"Years", "Shape's ratio"}, LabelStyle -> Medium]
```

A.4 Kernel density estimation

Download share price history from 1962 to 2015:

```
ge = FinancialData["GE", {{1962, 01, 01}, {2015, 12, 31}}][[All, 2]];
dd = FinancialData["DD", {{1962, 01, 01}, {2015, 12, 31}}][[All, 2]];
ibm = FinancialData["IBM", {{1962, 01, 01}, {2015, 12, 31}}][[All,
 2]];
```

Building joint distribution using Gaussian kernel with least square cross validation:

```
dt = (2015 - 1962 + 1)/Length@ge;
{uge, udd, uibm} =
```

```
Map[Log[Drop[#, 1]/Drop[#, -1]]/dt &, {ge, dd, ibm}];
pdfmulti =
PDF[SmoothKernelDistribution[Transpose[{uge, udd, uibm}],
 "LeastSquaresCrossValidation", "Gaussian"]]
```

Solve for maximum of the target function:

```
G[f1_?NumericQ, f2_?NumericQ, f3_?NumericQ] :=
NIntegrate[
Log[1 + f1 (Exp[u1 dt] - 1) + f2 (Exp[u2 dt] - 1) +
   f3 ( Exp[u3 dt] - 1)] pdfmulti[{u1, u2, u3}], {u1, -70,
 70}, {u2, -70, 70}, {u3, -70, 70}]
NMaximize[{G[f1, f2, f3],
f1 >= 0 && f2 >= 0 && f3 >= 0 && f1 + f2 + f3 <= 1}, {f1, f2, f3},
PrecisionGoal -> 5]
```

Simulation:

```
r = (Drop[#, 1] - Drop[#, -1])/Drop[#, -1] &@Transpose@{ge, dd, ibm};
capital =
First@Transpose@
  NestList[{#[[
       1]] (1 + {0.6471, 0.1765, 0.1745}.Part[r, #[[2]]]), #[[2]] +
      1} &, {1, 1}, Length@r];
```

A.5 Copula approach

Construct joint distribution using Gaussian copula:

```
jointdist =
```

```
CopulaDistribution[{"Multinormal",
  Covariance[Transpose[{uge, udd, uibm}]]},
 Map[EstimatedDistribution[#, NormalDistribution[u, s]] &, {uge,
   udd, uibm}]];
jointpdf = PDF[jointdist];
```

Solve for maximum of the target function:

```
G[f1_?NumericQ, f2_?NumericQ, f3_?NumericQ] :=
NIntegrate[
 Log[1 + f1 (Exp[u1 dt] - 1) + f2 (Exp[u2 dt] - 1) +
    f3 ( Exp[u3 dt] - 1)] jointpdf[{u1, u2, u3}], {u1, -Infinity,
  Infinity}, {u2, -Infinity, Infinity}, {u3, -Infinity, Infinity},
 Method -> {"GlobalAdaptive", "SymbolicProcessing" -> 0,
   "SingularityHandler" -> None}];
NMaximize[{G[f1, f2, f3],
f1 >= 0 && f2 >= 0 && f3 >= 0 && f1 + f2 + f3 <= 1}, {f1, f2, f3},
PrecisionGoal -> 5]
```

Simulation

```
First@Transpose@
  NestList[{#[[
       1]] (1 + {0.818854, 0.0694449, 0.111701}.Part[
          r, #[[2]]]), #[[2]] + 1} &, {1, 1}, Length@r];
```

A.6 Archimedean copula

Solve for the parameter in Density funcion of Archimedean copula:

```
copuladensity[t1_, t2_, t3_, theta_] :=
t1^(-1 - theta) t2^(-1 - theta) t3^(-1 - theta) (-2 - 1/theta) (-1 -
  1/theta) theta^2 (1 + ((-1 + t1^-theta)/theta + (-1 + t2^-theta)/
       theta + (-1 + t3^-theta)/theta) theta)^(-3 - 1/theta);
copuladensity[
  1/2 + 1/2 Sign[
         u1 - 0.0993256139801443`] (1 -
          E^(-Abs[-0.0993256139801443` + u1]/2.8445087121072263`)),
  1/2 + 1/2 Sign[
         u2 - 0.07395640112652939`] (1 -
          E^(-Abs[-0.07395640112652939` + u2]/2.853684556001697`)),
  1/2 + 1/2 Sign[
         0.1960662741617988` + u3] (1 -
          BetaRegularized[
           2.801287878066163`/(((u3 + 0.1960662741617988`)/
                 2.7078975563350096`)^2 + 2.801287878066163`),
           2.801287878066163`/2, 1/2]),
  theta
]
likelihood[{u1_, u2_, u3_}] :=
  Log[((-2 - 1/theta) (-1 -
     1/theta) theta^2 (1/2 +
      1/2 (1 - E^(-0.351555 Abs[-0.0993256 + u1])) Sign[-0.0993256 +
          u1])^(-1 - theta) (1/2 +
      1/2 (1 - E^(-0.350424 Abs[-0.0739564 + u2])) Sign[-0.0739564 +
          u2])^(-1 - theta) (1/2 +
```

```
      1/2 (1 -
         BetaRegularized[
          2.80129/(2.80129 + 0.136375 (0.196066 + u3)^2), 1.40064,
          1/2]) Sign[0.196066 + u3])^(-1 - theta) (1 +
      theta ((-1 + (1/2 +
               1/2 (1 -
                  E^(-0.351555 Abs[-0.0993256 +
                  u1])) Sign[-0.0993256 + u1])^-theta)/
          theta + (1/
            theta) (-1 + (1/2 +
               1/2 (1 -
                  E^(-0.350424 Abs[-0.0739564 +
                  u2])) Sign[-0.0739564 + u2])^-theta) +
         1/theta (-1 + (1/2 +
               1/2 (1 -
                  BetaRegularized[
                  2.80129/(2.80129 + 0.136375 (0.196066 + u3)^2),
                  1.40064, 1/2]) Sign[
                 0.196066 + u3])^-theta)))^(-3 -
      1/theta)) (0.17577727847055793` E^(-0.35155455694111587` \
      (Abs[-0.0993256139801443` +
         u1]))) (0.17521207764482244` E^(-0.3504241552896449` \
      (Abs[-0.07395640112652939` + u2]))) (0.956126 (1/(2.80129 +
        0.136375 (0.196066 + u3)^2))^1.90064)]
MemoryConstrained[
  NMaximize[{Total[Map[likelihood, Transpose[{uge, udd, uibm}], 1]],
```

```
    theta > 0}, theta, Method -> "SimulatedAnnealing"], 2 10^9];
MemoryConstrained[
  NMaximize[{Total[Map[likelihood, Transpose[{uge, udd, uibm}], 1]],
    theta > 0}, theta, Method -> "NelderMead"], 2 10^9]
```

A.7 Simulation with option

Build target function:

```
G[{f1_?NumericQ, f2_?NumericQ, f3_?NumericQ, f11_?NumericQ,
  f21_?NumericQ, f31_?NumericQ, f12_?NumericQ, f22_?NumericQ,
  f32_?NumericQ, c11_?NumericQ, c21_?NumericQ, c31_?NumericQ,
  c12_?NumericQ, c22_?NumericQ, c32_?NumericQ, m_}] :=
NIntegrate[
 Log[1 + (-1 + E^(27 u1/6797)) f1 + (-1 + E^(27 u2/6797)) f2 + (-1 +
       E^(27 u3/6797)) f3 +
    f11 (-1 + (-c11 Erfc[(-0.00572963 + 0.00572963 m -
              43.6329 Log[E^(27 u1/6797)/c11])/Sqrt[-1 + m]] +
          E^(27 u1/6797) Erfc[(0.00572963 - 0.00572963 m -
              43.6329 Log[E^(27 u1/6797)/c11])/
             Sqrt[-1 + m]])/(-c11 Erfc[(0.00572963 m -
              43.6329 Log[1/c11])/Sqrt[m]] +
          Erfc[-((43.6329 (0.000131314 m + Log[1/c11]))/
              Sqrt[m])])) +
    f12 (-1 + (E^(27 u1/6797) Erfc[(-0.00572963 + 0.00572963 m +
              43.6329 Log[E^(27 u1/6797)/c12])/Sqrt[-1 + m]] -
          c12 Erfc[(0.00572963 - 0.00572963 m +
```

```
            43.6329 Log[E^(27 u1/6797)/c12])/
          Sqrt[-1 + m]])/(-c12 Erfc[(-0.00572963 m +
            43.6329 Log[1/c12])/Sqrt[m]] +
      Erfc[(0.00572963 m + 43.6329 Log[1/c12])/Sqrt[m]])) +
f21 (-1 + (-c21 Erfc[(-0.0056456 + 0.0056456 m -
            44.2823 Log[E^(27 u2/6797)/c21])/Sqrt[-1 + m]] +
      E^(27 u2/6797) Erfc[(0.0056456 - 0.0056456 m -
            44.2823 Log[E^(27 u2/6797)/c21])/
          Sqrt[-1 + m]])/(-c21 Erfc[(0.0056456 m -
            44.2823 Log[1/c21])/Sqrt[m]] +
      Erfc[-((44.2823 (0.000127491 m + Log[1/c21]))/
          Sqrt[m])])) +
f22 (-1 + (E^(27 u2/6797) Erfc[(-0.0056456 + 0.0056456 m +
            44.2823 Log[E^(27 u2/6797)/c22])/Sqrt[-1 + m]] -
      c22 Erfc[(0.0056456 - 0.0056456 m +
            44.2823 Log[E^(27 u2/6797)/c22])/
          Sqrt[-1 + m]])/(-c22 Erfc[(-0.0056456 m +
            44.2823 Log[1/c22])/Sqrt[m]] +
      Erfc[(0.0056456 m + 44.2823 Log[1/c22])/Sqrt[m]])) +
f31 (-1 + (E^(27 u3/6797) Erfc[(0.00562847 - 0.00562847 m -
            44.417 Log[E^(27 u3/6797)/c31])/Sqrt[-1 + m]] -
      c31 Erfc[(-0.00562847 + 0.00562847 m -
            44.417 Log[E^(27 u3/6797)/c31])/
          Sqrt[-1 + m]])/(-c31 Erfc[(0.00562847 m -
            44.417 Log[1/c31])/Sqrt[m]] +
      Erfc[-((44.417 (0.000126719 m + Log[1/c31]))/Sqrt[m])])) +
```

```
f32 (-1 + (-c32 Erfc[(0.00562847 - 0.00562847 m +
          44.417 Log[E^(27 u3/6797)/c32])/Sqrt[-1 + m]] +
     E^(27 u3/6797) Erfc[(-0.00562847 + 0.00562847 m +
          44.417 Log[E^(27 u3/6797)/c32])/
        Sqrt[-1 + m]])/(Erfc[(0.00562847 m +
         44.417 Log[1/c32])/Sqrt[m]] -
     c32 Erfc[(-0.00562847 m + 44.417 Log[1/c32])/
        Sqrt[m]]))] ((0.111873 E^(-0.351555 Abs[-0.0993256 +
       u1] - 0.350424 Abs[-0.0739564 + u2]) (1/(2.80129 +
      0.136375 (0.196066 + u3)^2))^1.90064)/((1/2 +
    1/2 (1 -
       E^(-0.351555 Abs[-0.0993256 + u1])) Sign[-0.0993256 +
       u1])^1.65073 (1/2 +
    1/2 (1 -
       E^(-0.350424 Abs[-0.0739564 + u2])) Sign[-0.0739564 +
       u2])^1.65073 (1/2 +
    1/2 (1 -
       BetaRegularized[
        2.80129/(2.80129 + 0.136375 (0.196066 + u3)^2),
        1.40064, 1/2]) Sign[0.196066 + u3])^1.65073 (1 +
    0.650734 (1.53673 (-1 +
          1/(1/2 +
             1/2 (1 -
             E^(-0.351555 Abs[-0.0993256 +
             u1])) Sign[-0.0993256 + u1])^0.650734) +
       1.53673 (-1 +
```

```
                1/(1/2 +
                   1/2 (1 -
                   E^(-0.350424 Abs[-0.0739564 +
                   u2])) Sign[-0.0739564 + u2])^0.650734) +


              1.53673 (-1 +
                 1/(1/2 +
                    1/2 (1 -
                    BetaRegularized[
                    2.80129/(2.80129 + 0.136375 (0.196066 + u3)^2),
                    1.40064, 1/2]) Sign[
                    0.196066 +
                    u3])^0.650734)))^4.53673)), {u1, -Infinity,
  Infinity}, {u2, -Infinity, Infinity}, {u3, -Infinity, Infinity},
 MinRecursion -> 2, MaxRecursion -> 12, PrecisionGoal -> 3,
 AccuracyGoal -> 4,
 Method -> {"GlobalAdaptive", "SymbolicProcessing" -> 0,
   "SingularityHandler" -> None}];
```

Set domain:

```
Table[RegionProduct[
 RegionIntersection[
  Simplex[{{0., 0., 0., 0., 0., 0., 0., 0., 0.}, {0., 0., 0., 0.,
     0., 0., 0., 0., 1.}, {0., 0., 0., 0., 0., 0., 0., 1., 0.}, {0.,
      0., 0., 0., 0., 0., 1., 0., 0.}, {0., 0., 0., 0., 0., 1., 0.,
     0., 0.}, {0., 0., 0., 0., 1., 0., 0., 0., 0.}, {0., 0., 0., 1.,
```

```
      0., 0., 0., 0., 0.}, {0., 0., 1., 0., 0., 0., 0., 0.,
     0.}, {0., 1., 0., 0., 0., 0., 0., 0., 0.}, {1., 0., 0., 0., 0.,
      0., 0., 0., 0.}}],
   RegionProduct[
    Simplex[{{0., 0., 0.}, {0., 0., 1.}, {0., 1., 0.}, {1., 0.,
       0.}}], Simplex[{{0., 0., 0., 0., 0., 0.}, {0., 0., 0., 0., 0.,
        0.02}, {0., 0., 0., 0., 0.02, 0.}, {0., 0., 0., 0.02, 0.,
       0.}, {0., 0., 0.02, 0., 0., 0.}, {0., 0.02, 0., 0., 0.,
       0.}, {0.02, 0., 0., 0., 0., 0.}}]]],
  Cuboid[{0.75, 0.75, 0.75, 0.75, 0.75, 0.75}, {1.25, 1.25, 1.25,
    1.25, 1.25, 1.25}], Point[{m}]], {m, 3., 17., 1.}];
```

Solve maximum using Bayesian method:

```
bo3 = BayesianMaximization[G, reg3, AssumeDeterministic -> True,
MaxIterations -> 5]
```

Get all the Fridays among trading days:

```
days = FinancialData["GE", {{1962, 01, 01}, {2016, 01, 31}}][[All,
 1]];
fridays =
 Flatten@Position[Map[DateValue[#, "DayName"] &, days], Friday];
```

Calculate the maturity:

```
maturity =
Drop[Flatten[
   Drop[Map[ConstantArray[1, #] &,
      Prepend[Differences[fridays],
```

```
        First[fridays]]], -1] Differences[fridays]], {13595, -1}] -
  Drop[Range[Length[days]] -
    Flatten[Map[ConstantArray[1, #] &,
       Prepend[Differences[fridays],
        First[fridays]]] fridays], {13595, -1}];
```

simulation:

```
return[i_] := Module[
{m = maturity[[i]]}, {f1, f2, f3, f11, f21, f31, f12, f22, f32, c11,
   c21, c31, c12, c22, c32} = allocation[[m - 2]];
f1 ge[[i + 1]]/ge[[i]] + f2 dd[[i + 1]]/dd[[i]] +
 f3 ibm[[i + 1]]/ibm[[i]] +
 f11 call[ge[[i + 1]], c11 ge[[i]], vge, (m - 1) dt, 0]/
   call[ge[[i]], c11 ge[[i]], vge, m dt, 0] +
 f21 call[dd[[i + 1]], c21 dd[[i]], vdd, (m - 1) dt, 0]/
   call[dd[[i]], c21 dd[[i]], vdd, m dt, 0] +
 f31 call[ibm[[i + 1]], c31 ibm[[i]], vibm, (m - 1) dt, 0]/
   call[ibm[[i]], c31 ibm[[i]], vibm, m dt, 0] +
 f12 put[ge[[i + 1]], c12 ge[[i]], vge, (m - 1) dt, 0]/
   put[ge[[i]], c12 ge[[i]], vge, m dt, 0] +
 f22 put[dd[[i + 1]], c22 dd[[i]], vdd, (m - 1) dt, 0]/
   put[dd[[i]], c22 dd[[i]], vdd, m dt, 0] +
 f32 put[ibm[[i + 1]], c32 ibm[[i]], vibm, (m - 1) dt, 0]/
   put[ibm[[i]], c32 ibm[[i]], vibm, m dt, 0] + 1 - f1 - f2 - f3 -
 f11 - f21 - f31 - f12 - f22 - f32
]
```

```
capital =
First@Transpose@
  NestList[{#[[1]]  Part[r, #[[2]]], #[[2]] + 1} &, {1, 1},
   Length@r];
```

References

[1] Richard Bellman and Robert Kalaba. On the role of dynamic programming in statistical communication theory. *Information Theory, IRE Transactions on*, 3(3):197–203, 1957.

[2] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.

[3] L Breiman. Optimal gambling systems for favorable games. *The Kelly Capital Growth Investment Criterion*, page 47, 1961.

[4] Bruno Dupire et al. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[5] David Edelman. Supremum of mean–median differences for the binomial and poisson distributions: ln 2. Technical report, Tech. Rep., Department of Mathematical Statistics, Columbia University, 1979.

[6] Stewart N Ethier. The kelly system maximizes median fortune. *Journal of Applied Probability*, 41(4):1230–1236, 2004.

[7] Kais Hamza. The smallest uniform upper bound on the distance between the mean and the median of the binomial and poisson distributions. *Statistics & Probability Letters*, 23(1):21–25, 1995.

[8] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[9] John C Hull. *Options, futures, and other derivatives*. Pearson Education India, 2006.

[10] Harry Joe and James Jianmeng Xu. The estimation method of inference functions for margins for multivariate models. 2016.

[11] John L Kelly Jr. A new interpretation of information rate. *Information Theory, IRE Transactions on*, 2(3):185–189, 1956.

[12] Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice*. Princeton University Press, 2007.

[13] Robert C Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[14] Robert C Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8(4):323–361, 1980.

[15] Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.

[16] Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.

[17] M Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[18] Paul Wilmott. *Paul Wilmott introduces quantitative finance*. John Wiley & Sons, 2007.

[19] Dennis Yang and Qiang Zhang. Drift-independent volatility estimation based on high, low, open, and close prices*. *The Journal of Business*, 73(3):477–492, 2000.