Team Members: **Zhang Ketian, Liang Hao**

School: **Affiliated High School of South China Normal**

Province: **Guangdong**

Tutor: **Huang Yiwen, Ye Qiaoka**

Title of Dissertation: **The Model of Compound Expectation in Financial Investment**

Abstract: It's unwise to apply mathematical expectation to the area of financial investment, though it is widely used in statistics research. It is compound interest that counts mostly in financial investment. In order to work out the investing proceeds and to evaluate the pros and cons of an opportunity, we put forward the model of compound expectation. The Yield rate of compound expectation is defined as "a sustainable and theoretical rate of return on the premise of long-term investment." With the aid of this model, an investor can judge an investing project more easily, integrating its proceeds with risks, and directly appraise the investment opportunity with its final yield rate of compound interest.

When it comes to the quantitative study in the model of compound expectation, we put the major part under the assumption that the distribution of expected yield rate corresponds with normal distribution, trying to develop the optimum strategy of capital input and to calculate the maximum rate of return of compound expectation. We have also strictly proved that the theoretical optimization strategy of funds allocation is *unique*, with its pattern fixed.

Further significance and application of the model of compound expectation is shown afterwards. Whatever the certain distribution of expected yield rate is, such valuable conclusions as ①"Mathematical expectation is always bigger than that of compound interest" ②"The increase in σ of the expected proceeds rate will theoretically only exert negative impact on the final income" ③There will be no more than one maximum point in the function of compound expectation and proportion of investment ④"Marginal relative yield rate in the model of compound interest is bound to diminish as the investment input increases" and are discovered and demonstrated, which may be of great importance as well as potency to solve practical investing problems in the capital market.

It has been generally believed that diversification of investment aims at spreading the risks of portfolio. But in the last chapter, by applying the model of compound expectation to it, we draw a distinctive conclusion that the goal of a shrewd diversification is mainly to allocate the assets in a more efficient way, leading the investor to a better outcome. Given several investing projects at the same time, in each of which an investor allocate a proportion of funds, the optimization strategy can be formulated with the help of our computer programming and an important indicator called "marginal return rate". In this way can our model facilitates investors to get a higher long-term return in the capital market.

Key words:

Mathematical Expectation | **Model of Compound Expectation** | **Normal Distribution** | **Allocation of Funds** | **Optimization Strategy** | **Proceeds and Risks** | **Diversification** | **Capital Efficiency**

Common symbols in the text:

1. $F_X(x)$: The distribution function of the continuous random variable *X*.

2. $p_X(x)$: The density function of the continuous random variable *X*.

3. *C*: $C = P(X > \mu - 3\sigma)$, where $C \sim N(\mu, \sigma^2)$.

Important concepts in the text:

Yield rate of compound expectation – A sustainable and **theoretical** compounded rate of return on the premise of long-term investment

Compound expectation – A theoretical ratio of capital input and interest to capital input, equal to (**Yield rate of compound expectation+1**)

Ratio of capital input and interest to capital input – The **actual** ratio of capital input and interest to the capital input in an investment project, equal to (Yield rate + 1)

Zero-removal normal distribution – The curve obtained through removing the image on the left side of the y-axis of normal distribution

Constant-positive normal distribution – The curve obtained through removing the image of $x < \mu - 3\sigma$ of normal distribution

Even-removal normal distribution – The curve obtained through removing the images of $x < \mu - 3\sigma$ and $x > \mu + 3\sigma$ of normal distribution

Relative yield rate – The ratio of the yield rate of compound expectation on a certain capital input to the capital input

Marginal relative yield rate – The relative yield rate for an increased unit of capital input

Meaning of important symbols in the text:

m – Yield rate of compound expectation
X – Ratio of capital input and interest to capital input for all-capital input
Y – Ratio of capital input and interest to capital input for partial input
t – Proportion of capital input to all principal for an investment project
x_i – Ratio of capital input and interest to capital input obtained under a certain probability
p_i – Probability of occurrence corresponding with x_i
G (X) – The theoretical compound expectation of a project under the all-capital input
G (Y) – The theoretical compound expectation of a project under the proportion *t*
R – Relative yield rate on investment
MR – Marginal relative yield rate

Table of Contents

I. Overview of Compound Expectation

1.1 Leading question: Is the mathematical expectation always effective for the expected profitability of financial investment?

We found a question on a simulated mathematics test paper of college entrance examination: A person engages in one long-term investment project, and the principal is RMB 10,000. There is a 50% probability of principal doubling every time, and also 50% probability of principal loss by 50%, find the expected yield rate for each investment.

"Standard answer": $m = 50\% \times 100\% - 50\% \times 50\% = 25\%$

Through simply conducting the calculation from the viewpoint of mathematical expectation, it obtained the expected yield rate for each investment, i.e. 25%. On the analogy of this, the amount will exceed RMB 100,000 for the principal of RMB 10,000 after 11 years in theory (calculated by compound interest).

In fact, such thinking is very absurd.

Because the "standard answer" actually contains such a requirement: Under the conditions of the question, the principal must be divided into innumerable small portions, and each small portion is put into the same investment projects as mentioned above, and the results of the innumerable investments are independent of each other. However, this is **self-contradictory!** There is only one investment project at the same time, so that only one investment result exists (capital input doubled or reduced by half) and it is impossible to obtain different investment results independent of each other.

Why does this paradox exist? In fact, there are essential differences between solving the issues about the rate of return on financial investment and solving the issues about the absolute returns. The transaction results for each time have an impact on the subsequent absolute returns. It is inappropriate to solve such issues using the thinking mode of mathematical expectation.

Let's have a look at the correct analysis:

As the yield rate *m* has the rolling characteristics of compound interest, that is, the returns of the previous investment will directly affect the absolute returns of the current investment. According to the theoretical analysis, if someone puts the principal of RMB 10,000 into the above-mentioned investments **for multiple times on a long-term basis**, **there should be theoretically n times of principal doubling as well as n times of principal halving after 2n (*n* is big enough) times**. Therefore, in the long run, such investments have no meaning, and there is no loss or profit for the investor.

Under the conditions of the question, the expression of correct thinking is:

$$m = (1+100\%)^{50\%} \times (1-50\%)^{50\%} - 1 = 0$$

Through the comparison between the correct solution and false solution to the above question, we can find that the mathematical expectation is incapable of solving the expectation issues about the rate of return on financial investment, and it is reasonable to solve the issues using the compound expectation. Therefore, we need to construct the model of compounded expectation systematically.

1.2 Development of the Model of Compound Expectation

The investment community generally believes a saying: “The compound interest is the eighth wonder of the world”. In the investment process like rolling a “snowball”, the sustainable compounded rate of return is the only indicator that dominates the final yield rate.

It is found: Mathematical expectation can not scientifically solve the issues about compound interest in the investment in many cases. Based on this, we put forward the concept of **Yield rate of compound expectation**, that is, a sustainable and theoretical compounded rate of return on the premise of long-term investment. **The compound expectation** is the (**Yield rate of compound expectation +1**).

The basic assumptions of the model of compound expectation:

(1) The investor engages in long-term multiple transactions;

(2) The investors know about the distribution of investment yield rate in advance, and the distribution of the expected yield rate is not subject to the influence of the capital input;

(3) The investment gains and losses are calculated by the income ratio rather than the absolute amount of income;

(4) The taxes and transaction costs are negligible.

(5) The actual results of investments are independent of each other, causing no impact on the future investment results or the investor’s state of mind.

Let the investment principal be “1”, if the scattered distribution of *X*, i.e. the ratio of capital input and interest to capital input, obtained under a certain investment opportunity is $x_1, x_2, \dots x_n$, and the corresponding probabilities of occurrence are $p_1, p_2, \dots p_n$ respectively.

Then the definition of *m*, i.e. the expected value of the compounded rate of return, is

$m = \prod_{i=1}^{n} x_i^{p_i}$ **-1 (Formula 1)**

It is clear that: As the debt leverage transaction is not introduced, the principal will not become negative due to loss, therefore, there must be $x_i \geq 0, i = 1,2,\dots\dots n$.

However, the above situations only consider the case of putting all principal into the investment projects. In fact, all-capital input is not a good strategy. For example, if only a portion of the principal is put into the operation, the overall investment rate of return may be greater. Let’s take the conditions of the leading question as an example: Suppose that a fixed proportion extracted from the principal, i.e. *t*, is put into the investment project ($0 \leq t \leq 1$), it is thus obtained according to **Formula 1**:

$$m' = (1+t)^{50\%} \times (1-0.5t)^{50\%} - 1$$

$$= \sqrt{(1+t)(1-0.5t)} - 1$$

Through simplification, it is obtained: When t = 0.5, m' obtains the maximum value, i.e. 6.07%, which is significantly better than the zero returns obtained at the time of all-capital input. Therefore, we find that under the model of compound expectation, the theoretical compound yield is subject to the influence of the proportion of the invested funds, and the optimum investment strategy exists.

II. Preliminary Quantitative Study of the Model of Compound Expectation

Lemma: If the function $f(x)$ is continuous and has the only zero point *c* in the interval (*a*, *b*), then $f(x)$ has the constant same sign in the intervals (*a*, *c*) and (*c*, *b*) respectively.

Proof: Suppose that $\exists x_1, x_2 \in (a,c)$ (let $x_1 < x_2$), $f(x_1)$ and $f(x_2)$ have different sign. Also, $f(x)$ is continuous on $[x_1, x_2]$, $\exists \xi \in (x_1, x_2)$ by the zero point theorem, so that $f(\xi) = 0$. Obviously, $\xi \neq c$, that is to say, c is not the only zero point of $f(x)$ in the interval (*a*, *b*), so that a contradiction exists. Thus, $f(x)$ has the same sign in the interval (*a*, *c*)

Similarly, $f(x)$ has the constant same sign in the interval (*c*, *b*). This completes the proof.

2.1 Two-point random distribution and optimization of the ratio of capital input and interest to capital input

Next, we will explore the optimum investment proportion when the expected rate of return is consistent with when the two-point distribution, and the obtained maximum value of compound expectation.

Distribution column of *X*, i.e. the expected ratio of capital input and interest to capital input

X	x_1	x_2

P	p_1	p_2

(As the debt leverage transaction is not introduced, the principal will not become negative, and let $x_2 > x_1 \geq 0$)

Suppose that a fixed proportion extracted from the principal, i.e. t, is put into the investment project, the final ratio of the principal and interest to the principal is Y, thus:

$Y = tX + (1-t)$

Let $t = t_{\max}$, Y obtains the maximum value.

$\mathrm{G}(Y) = [tx_1 + (1-t)]^{p_1}[tx_2 + (1-t)]^{p_2}$

(1) When $0 < x_1 < x_2 \leq 1$, the capital input will not produce any profit, thus the investment shall not be made: $t_{\max} = 0$.

(2) When $1 \leq x_1 < x_2$, the capital input will surely produce profits or keep break-even at least, thus all-capital input shall be made: $t_{\max} = 1$.

(3) When $0 \leq x_1 < 1 < x_2$,

Let $g(x) = [xx_1 + (1-x)]^{p_1}[xx_2 + (1-x)]^{p_2}$, thus

$$g^{'}(x) = p_1(x_1-1)(\frac{x_2x+1-x}{x_1x+1-x})^{p_2} + p_2(x_2-1)(\frac{x_1x+1-x}{x_2x+1-x})^{p_1}.$$

Obviously, $g(x)$ is continuous in $(\frac{1}{1-x_2}, \frac{1}{1-x_1})$, where $\frac{1}{1-x_2} < 0$, $\frac{1}{1-x_1} > 1$.

$g^{'}(x)$ has the only zero point.

$$x_0 = \frac{\mathrm{E}(X)-1}{(1-x_1)(x_2-1)}$$

According to $x_1 < E(X) < x_2$, $\frac{1}{1-x_2} < x_0 < \frac{1}{1-x_1}$ is derived. Also, $\lim\limits_{x \to \frac{1}{1-x_2}} g^{'}(x) = +\infty$,

$\lim\limits_{x \to \frac{1}{1-x_1}} g'(x) = -\infty$, $g'(x)$ is constant positive in $(\frac{1}{1-x_2}, x_0)$ and constant negative in $(x_0, \frac{1}{1-x_1})$ according to the lemma. Thus $g(x)$ is monotonically increasing in $(\frac{1}{1-x_2}, x_0)$ and monotonically decreasing in $(x_0, \frac{1}{1-x_1})$.

Therefore, if $0 \le \frac{\mathrm{E}(X)-1}{(1-x_1)(x_2-1)} \le 1$, then $t_{\max} = \frac{\mathrm{E}(X)-1}{(1-x_1)(x_2-1)}$; if $\frac{\mathrm{E}(X)-1}{(1-x_1)(x_2-1)} < 0$, then $t_{\max} = 0$; if $\frac{\mathrm{E}(X)-1}{(1-x_1)(x_2-1)} > 1$, then $t_{\max} = 1$.

In summary, the optimum solution in the two-point yield distribution is as follows:

When $0 \le x_1 < x_2 \le 1$ or $0 \le x_1 < 1 < x_2$ and $\frac{\mathrm{E}(X)-1}{(x_1-1)(1-x_2)} < 0$, $t_{\max} = 0$, G(Y) = 1;

When $1 \le x_1 < x_2$ or $0 \le x_1 < 1 < x_2$ and $\frac{\mathrm{E}(X)-1}{(x_1-1)(1-x_2)} > 1$, $t_{\max} = 1$, $\mathrm{G}(Y) = x_1^{p_1} x_2^{p_2}$;

When $0 \le x_1 < 1 < x_2$ and $0 \le \frac{\mathrm{E}(X)-1}{(x_1-1)(1-x_2)} \le 1$, $t_{\max} = \frac{\mathrm{E}(X)-1}{(x_1-1)(1-x_2)}$, thus the optimum compound expectation value is G(Y) $= \frac{p_1^{p_1} p_2^{p_2} (x_2 - x_1)}{(1-x_1)^{p_2} (x_2-1)^{p_1}}$.

2.2 Multi-point random distribution and optimization of yield rate

Distribution column of *X*, i.e. the expected ratio of capital input and interest to capital input

X	x_1	x_2	x_3	$\cdots$	x_n
P	p_1	p_2	p_3	$\cdots$	p_n

(As the debt leverage transaction is not introduced, the principal will not become negative, thus $x_i \geq 0$, i = 1, 2…n)

It is obtained from the definition of the model of compound expectation:

$$G(X) = \prod_{i=1}^{n} x_i^{p_i}$$ (Input of all principal each time)

$$G(Y) = f(t) = \prod_{i=1}^{n} (1 - t + tx_i)^{p_i}$$ (Input of the proportion t of the principal each time)

When t_0 exists, which makes $f'(t_0) = 0$, the t_0 value may make G(Y) get the maximum value under the distribution of the ratio of capital input and interest to capital input. However, as it is impossible for the general algebraic equation of quartic or higher degree to get the general form of algebraic solution, when n is more than 5, there is no general solution for $f'(t\ ) = 0$, so concrete analysis shall be conducted when faced with specific issues. Therefore, the condition of "**The expected rate of return is consistent with the normal distribution**" will be considered as the study focus of the model when carrying out in-depth quantitative study in the fourth section of this dissertation.

III. Practical Significance of Compound Expectation in Financial Market (Preliminary)

In previous studies, the risks and proceeds are often divided. However, as the financial investment belongs to long-term transactions, in the long run, the result of each investment has an impact on the absolute yield on the next investment, and the ultimate "high yield" is the only criterion for evaluating the success or failure of investment. "Risk" (defined as the possibility and severity of loss for each investment) only has an influence on the long-term returns. **The development and application of compound expectation directly unifies the above risks and proceeds, and the compound expectation value is taken as the only judgment indicator of investment projects.**

"High risk, high return" is the basic rule embraced by the market, and even the risk is considered to be a prerequisite for getting returns, such as CITIC Securities' questionnaire on risk preference of investors:

Suppose you invest RMB 100,000 for purchase of a certain securities and hold for one year, what kind of securities will you choose?

A. 100% earning of RMB 3,200

B. 80% possibility for earning RMB 29,000, and 20% possibility for losing RMB 100,000

C. 50% possibility for earning RMB 100,000, and 50% possibility for losing RMB 100,000

D. 50% possibility for earning RMB 10,000, and 50% possibility for losing RMB 10,000

E. 20% possibility for earning RMB 100,000, and 80% possibility for losing RMB 21,000

On the surface, the mathematical expectation for A, B and E is consistent, and the difference between C and D is also insignificant. The risk appetite is different for them (It seems that the earnings are more attractive for greater risk of loss). However, as the investors will continue to make long-term transactions under different opportunities in the stock market, the model of compound expectation is applicable. Suppose that the investor makes all-capital input (according to the meaning of the question), we found that the "risk averse" option A is the best choice after calculating the yield rate of the compound expectation (according to Formula 1).

$$m_A = 3.2\% \quad m_B = -100\% \quad m_C = -100\% \quad m_D = 2.18\% \quad m_E = -4.87\%$$

It can be seen that for the different projects under the same mathematical expectation value, high risk can not brings high returns (such as Option E), and low risk can make the investors become the ultimate winners. The study inference of the model of compound expectation is essentially consistent with the risk premium principle and the risk aversion principle in economics.

The following more specific quantitative study on the model of compound expectation on the assumption of normal distribution of the expected rate of return on investment will also give the readers more in-depth understanding of the significance of the model.

IV. Quantitative Study on the Model of Compound Expectation When the Expected Rate of Return Is under Normal Distribution

4.1 Basic theorem proof and background overview

Before conducting more in-depth quantitative study, we firstly prove an important theorem relating to the definition of compound expectation value.

Theorem I: For the constant-positive random variable *X*, if E (lnX) exists, then the compound expectation of *X* under all-capital input

$$\mathrm{G}(X) = e^{\mathrm{E}(\ln X)}.$$

Proof: (1) If *X* is a discrete random variable, then the definition of the compound expectation:

$$\mathrm{G}(X) = \prod_{i=1}^{n} x_i^{\,p_i}$$

where $p_i = P(X = x_i)$**.**

$$\ln \prod_{i=1}^{n} x_i^{\,p_i} = \sum_{i=1}^{n} \ln x_i \cdot p_i = \mathrm{E}(\ln X), \text{ thus } \prod_{i=1}^{n} x_i^{\,p_i} = e^{\mathrm{E}(\ln X)} = \mathrm{G}(X).$$

(2) If *X* is a continuous random variable, then the definition of the compound expectation under discrete distribution $\mathrm{G}(X) = \prod_{i=1}^{n} x_i^{\,p_i}$

Under continuous distribution, it can be converted

to $\mathrm{G}(X)=\lim_{\lambda\to 0}\prod_i \xi_i^{\Delta F_i}$ **,**

where $x_i \le \xi_i \le x_{i+1}$**,** $\Delta F_i = F_X(x_{i+1}) - F_X(x_i) = P(x_i \le X < x_{i+1})$**,**

$\lambda = \max\{\Delta F_1, \Delta F_{2,} \ldots\ldots\}$**.**

$$\ln\lim_{\lambda\to 0}\prod_i \xi_i^{\Delta F_i} = \lim_{\lambda\to 0}\ln\prod_i \xi_i^{\Delta F_i} \text{ (ln}x\text{ is continuous in the definition domain)}$$

$$= \lim_{\lambda\to 0}\sum_i \ln\xi_i \cdot \Delta F_i$$

$$= \int_0^{+\infty} \ln x dF_X(x) \text{ (Definition of definite integral)}$$

$$= \int_0^{+\infty} \ln x \cdot p_X(x)dx \; (\frac{dF_X(x)}{dx} = p_X(x))$$

$$= \mathrm{E}(\ln X) \text{【1】},$$

thus $\lim_{\lambda\to 0}\prod_i x_i^{\Delta F_i} = e^{\mathrm{E}(\ln X)} = \mathrm{G}(X)$.

In summary, $\mathrm{G}(X)=e^{\mathrm{E}(\ln X)}$ **is actually the compound expectation of the constant-positive random variable *X*, and the proof of Theorem I is now completed.**

Currently, it is generally recognized in academic circle of financial investment: On the assumption of “geometric Brownian motion of asset prices”, the distribution of the expected rate of return on assets roughly corresponds with normal distribution (which is also clarified in the book named “Financial Market Econometrics”). The empirical studies based on stock returns are also mostly based on the condition that the yield rate is subject to normal distribution (such as CAPM and Black-Scholes’ pricing formulas). The VaR RiskMetrics launched by JPMorgan Bank in 1994 is also based on the assumption that the yield rate of securities is subject to normal distribution.

“The distribution of the expected ratio of capital input and interest to capital input (or expected rate of return) corresponds with normal distribution” is also the basic premise for our study in this section. As the normal distribution of the rate of return on investment is relatively reasonable and in line with the market realities (See **Mean-variance Theory in Portfolio Management put forward by Markowitz in** *Mean-Variance Analysis in Portfolio Choice and Capital Markets*), quantitative study will be conducted for the model of compound expectation based on this condition. We define μ as the expected median ratio of capital input and interest to capital input, and use σ to measure the possible deviation between the actual ratio of capital input and interest to capital input and the expected median.

It is declared that we do not simply use the mathematical expectation in the normal distribution curve of the rate of return on investment to measure the income and use the variance to measure the risk. Under this condition, we calculate the yield rate of compound expectation in the normal distribution curve of a given rate of return on investment, and use this indicator to measure the final expected yield of the investment project.

Basis assumptions of the model:

(1) The investment model is the model of compound expectation;

(2) The expected ratio of capital input and interest to capital input corresponds with normal distribution, and its distribution is not subject to the impact of capital input;

(3) The investors know about the probability distribution of the investment yield rate in advance (i.e. the values μ and σ are known);

(4) The theoretical compound expectation value is the only factor influencing the decisions of the investors;

(5) The taxes and transaction costs are negligible;

(6) As the debt leverage transaction is not introduced, the principal will not become negative, and the case of less than zero does not exist in the distribution of *X*.

4.2 Distribution Study and Standardization of the Ratio of Capital Input and Interest to Capital Input

Firstly, when $\mu \le 1$ in the normal distribution of the expected ratio of capital input and interest to capital input of an investment opportunity, no matter how the investors develop the investment schemes, they can not obtain a positive theoretical rate of return in the model of compound expectation (Regarding this point, a comprehensive and rigorous proof will be given in the sixth section of this dissertation). Therefore, this type of investment opportunities will be regarded by us as **invalid investment opportunities**, and the proportion of funds that should be invested by the investors in the most optimum case is $t = 0$. This is the risk aversion choice for reasonable investors in the absence of positive yield expectation.

If the mathematical expectation is greater than 1 in the distribution of the expected ratio of capital input and interest to capital input of an investment opportunity, then we call it **"valid distribution"**. Under the additional condition that the investment opportunities are of valid distribution, the theoretical rate of return and the optimization issues of the model of compound expectation will be studied below.

According to the basic assumption (6), there shall be no negative values in the distribution of *X*, so we need to remove the part of $x \le 0$ in the normal distribution image of the ratio of capital input and interest to capital input, and the new curve is called zero-removal normal distribution (See Figure 1).

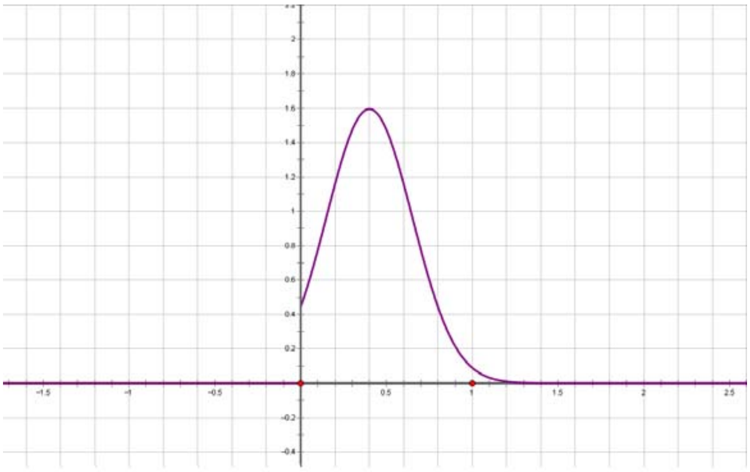


Figure 1

Definition 1 Denote the distribution function of the normal distribution $\mathrm{N}(\mu,\sigma^2)$ **as** $F(x)$**, then**

$$F_0(x)=\begin{cases}\dfrac{F(x)-F(0)}{1-F(0)} & (x>0)\\ 0 \quad (x\le 0)\end{cases}$$

is taken as the distribution of the distribution function, called zero-removal normal distribution and denoted by $\mathrm{N}_0(\mu,\sigma^2)$**.**

However, as this function is relatively complex, we can not effectively simplify and draw a conclusion in the later calculations. On the other hand, direct removal of *x <0* part is equivalent to reduction of the overall effect of principal loss, which is likely to cause relatively big errors.

For the purpose of simplification and difficulty reduction, we consider converting the original normal distribution into the constant-positive normal distribution, that is, the new curve obtained through removing the image of $x<\mu-3\sigma$ of the normal distribution curve, so as to replace **zero-removal normal distribution**. As the probability that *X* falls in the interval $(-\infty,\ \mu-3\sigma)$ is less than only 0.13%, the error generated from image change is within the acceptable range (See Section 4.4 for the quantitative analysis on the error).

Definition 2 Denote the distribution function of the normal distribution $\mathrm{N}(\mu,\sigma^2)$ **as** $F(x)$**, if** $\mu>3\sigma$**, then**

$$F_+(x)=\begin{cases}\dfrac{F(x)-F(\mu-3\sigma)}{F(\mu+3\sigma)} & (x>\mu-3\sigma)\\ 0 \quad (x\le \mu-3\sigma)\end{cases}$$

is taken as the distribution of the distribution function, called constant-positive normal distribution and denoted by $N_+(\mu, \sigma^2)$ (See Figure 2).

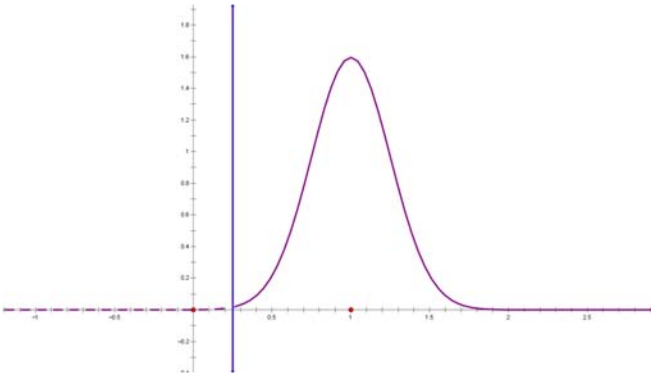
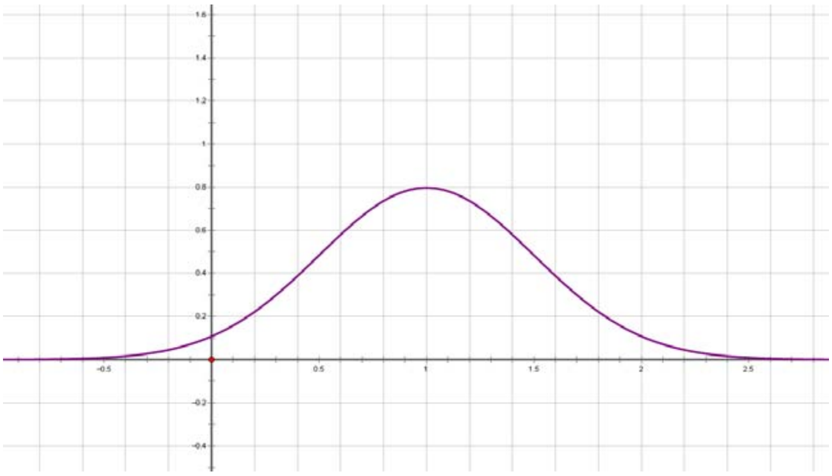


Figure 2

Important condition for the study on the model of constant-positive normal distribution: $\mu > 3\sigma$

Through referring to the empirical studies on the rate of return in some stock markets, such as *Options, Futures and Other Derivatives* written by John Hull (Canada), *The Econometrics of Financial Markets* and other books, combining with the long-standing risk-free annualized rate of return (about 3%) in the markets and collecting the statistics of the income results of actual investment projects, we found that: in the normal distribution of the expected ratio of capital input and interest to capital input, more than 95% meets the condition of $\mu > 3\sigma$.

In fact, there are reasons why the investment opportunities are rare in the yield distribution of $\mu \le 3\sigma$. In the figure below, it is the normal distribution curve of the rate of return when $\mu = 1$ and $\sigma = 0.5$, i.e. $\mu = 2\sigma$. It is clear that the expected rate of return is scattered in the figure, and there is a big probability of overly deviation between the actual rate of return and the expected median rate of return. In a sense, the yield expectation of such investments is invalid, as there is a big uncertainty for the expectation itself.



It has been shown through the empirical analysis of the stock market: the rate of return for most of the investment opportunities in the market corresponds with $\mu > 3\sigma$, and the normal distribution of the rate of return is relatively valid only in the event of $\mu > 3\sigma$, because the distribution of the expected rate of return is relatively concentrated, and the investors can know about the expected rate of return on investment accurately.

Therefore, we call the normal distribution of the rate of return corresponding with $\mu > 3\sigma$ and $\mu > 1$ "**Valid yield rate expectation**", and **take $\mu > 3\sigma$ as one of the basic conditions for the calculation of compound interest in this section**. In this section, the study will be conducted only in this domain.

4.3 Calculation of compound expectation under all-capital input

When the expected ratio of capital input and interest to capital input corresponds with normal distribution, i.e. $X \sim N_{+}(\mu, \sigma^2)$ and all principal is put into the investment project, the compound expectation $\mathrm{G}(X)$ is calculated as follows:

Denoted by $C = F(\mu + 3\sigma) \approx 0.9987$.

$$p(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}C\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & (x > \mu - 3\sigma) \\ 0 & (x \le \mu - 3\sigma) \end{cases},$$

$$\mathrm{E}(\ln x) = \int_{-\infty}^{+\infty} \ln x \cdot p(x) dx$$

$$= \frac{1}{\sqrt{2\pi}C\sigma} \int_{\mu-3\sigma}^{+\infty} \ln x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$=\frac{1}{\sqrt{2\pi}C\sigma}\int_{k-3}^{+\infty}\ln(\sigma v)\cdot e^{-\frac{(v-k)^2}{2}}\sigma dv\ (v=\frac{x}{\sigma},\ \ k=\frac{\mu}{\sigma})$$

$$=\frac{1}{\sqrt{2\pi}C}\int_{k-3}^{+\infty}\ln v\cdot e^{-\frac{(v-k)^2}{2}}dt+\frac{\ln\sigma}{C}\int_{k-3\times1}^{+\infty}\frac{1}{\sqrt{2\pi}\times1}e^{-\frac{(v-k)^2}{2\times1^2}}dv$$

$$=\frac{1}{\sqrt{2\pi}C}\int_{k-3}^{+\infty}\ln v\cdot e^{-\frac{(v-k)^2}{2}}dv+\ln\sigma\,.$$

Denoted by $f(k)=\int_{k-3}^{+\infty}\ln v\cdot e^{-\frac{(v-k)^2}{2}}dv$, then $\mathrm{G}(X)=\sigma e^{\frac{1}{\sqrt{2\pi}C}f(k)}$.

As it is difficult to calculate $f(k)$ for its complexity, we find a suitable elementary function as an approximate substitute for $f(k)$ through using computer.

Firstly, we calculated some f (k) values with the numerical integration method.

Considering that a computer can not calculate the integral values with infinite upper bound, we conducted the following approximate conversion:

It is denoted by $h(k)=\int_{k-3}^{a}\ln v\cdot e^{-\frac{(v-k)^2}{2}}dv$.

When a is relative big, $f(k)\approx h(k)=\int_{k-3}^{a}\ln x\cdot e^{-\frac{(x-k)^2}{2}}dx$, and the absolute value of the error between $f(k)$ and $h(k)$ is $\varepsilon=\int_{a}^{+\infty}\ln x\cdot e^{-\frac{(x-k)^2}{2}}dx$.

Let $g(x)=\ln x-x\ (x>0)$, then $g'(x)=\frac{1}{x}-1$. When $x\in(0,1)$, $g'(x)>0$; when $x\in(1,+\infty)$, $g'(x)<0$. When $x=1$, $g(x)$ obtains the maximum value, i.e. $\ln x-x\le g(1)=-1$. Thus, $\ln x<x$, and $e^{-\frac{(x-k)^2}{2}}>0$, so

$$\varepsilon<\int_{a}^{+\infty}x\cdot e^{-\frac{(x-k)^2}{2}}dx$$

$$=\left[\sqrt{2}k\int_{0}^{\frac{x-k}{\sqrt{2}}}e^{-t^2}dt-e^{-\frac{(x-k)^2}{2}}\right]\Bigg|_{a}^{+\infty}$$

$$=k\sqrt{\frac{\pi}{2}}-\sqrt{2}k\int_{0}^{\frac{a-k}{\sqrt{2}}}e^{-t^2}dt-e^{-\frac{(a-k)^2}{2}}$$

$$< \sqrt{2}k(\frac{\sqrt{\pi}}{2} - \int_0^{\frac{a-k}{\sqrt{2}}} e^{-t^2} dt)$$

$$= \sqrt{2}k\int_{\frac{a-k}{\sqrt{2}}}^{+\infty} e^{-t^2} dt \; (\int_0^{+\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2} \text{【2】})$$

$$< \sqrt{2}k\int_{\frac{a-k}{\sqrt{2}}}^{+\infty} e^{-t} dt \; (\text{when } \frac{a-k}{\sqrt{2}} > 1)$$

$$= \sqrt{2}k \cdot (-e^{-t})\Big|_{\frac{a-k}{\sqrt{2}}}^{+\infty} = \sqrt{2}ke^{\frac{k-a}{\sqrt{2}}},$$

that is, when $a > k + \sqrt{2}$, $\varepsilon < \sqrt{2}ke^{\frac{k-a}{\sqrt{2}}}$. For $\forall \varepsilon_0 > 0$, as long as

$$a > \max\{k + \sqrt{2}, k - \sqrt{2}\ln\frac{\varepsilon_0}{\sqrt{2}k}\},$$

it can be ensured that $\varepsilon < \varepsilon_0$, that is, the error generated by the above approximate conversion can be controlled.

Through properly selecting *a*, the following table is obtained through doing numerical integration for $f(k) \approx \int_{k-3}^{a} \ln x \cdot e^{-\frac{(x-k)^2}{2}} dx$ with computer:

Table 1 – Approximate values of $f(k)$

k	3.5	4	4.5	5	5.5	6	6.5	7
f(*k*)	3.02373	3.38812	3.70205	3.97882	4.22674	4.45149	4.65716	4.84682

k	7.5	8	8.5	9	9.5	10	10.5	11
f(*k*)	5.02284	5.18708	5.34104	5.48596	5.62284	5.75256	5.87581	5.99323

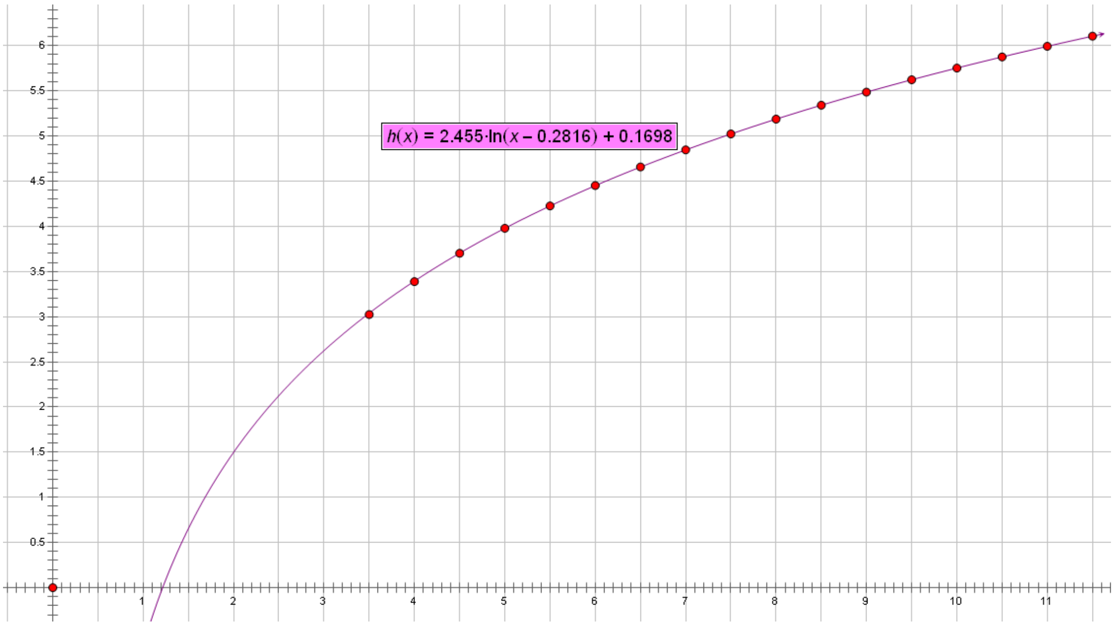
k	11.5	100	1000	10000
f(*k*)	6.10534	11.5279	17.2918	23.0557

Then, we find a primary function to fit the f (k)-k image.

According to the analysis and calculation, $f(k)$ is monotonically increasing, and its derivative is monotonically decreasing, therefore, the functions, like $y = a + b\ln(k + c)$, are used to fit, and it is calculated that $a = 0.16981$, $b = 2.4551$, $c = -0.28164$.

(See Figure 3 for the fitting effect).

Figure 3 Fitting Effect of $f(x)$



Note: Red points are on the $f(x)$, and the purple curve is the image of $h(x)$.

As there is a high degree of coincidence for the two functions, so the elementary function $y = 0.1698 + 2.455\ln(k - 0.2816)$ can be used as an approximate substitute for $f(k)$, and after the substitution, it is obtained that:

The compound expectation value under all-capital input $\mathrm{G}(X) \approx 1.0702(\frac{\mu}{\sigma} - 0.2816)^{0.9807}\sigma$

4.4 Accuracy analysis of compound expectation value in constant-positive normal distribution

Zero-removal normal distribution is obtained from the basic assumptions of the model of compound expectation, whereas the **constant-positive normal distribution** is the further amendment based on this. Therefore, we need to analyze the error between them through calculation, so as to verify whether the compound expectation value obtained under the constant-positive normal distribution is sufficiently accurate.

When the expected ratio of capital input and interest to capital input corresponds with normal distribution, i.e. $X_0 \sim N_0(\mu, \sigma^2)$ and all principal is put into the investment project, the compound expectation $\mathrm{G}(X)$ is calculated as follows:

Denoted by $C_0 = 1 - F(0) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v-k)^2}{2}} dv$ ($v = \frac{x}{\sigma}$, $k = \frac{\mu}{\sigma}$) . The density function of X_0 is

$$p_0(x) = \begin{cases} \frac{1}{\sqrt{2\pi}C_0\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & (x > 0) \\ 0 & (x \le 0) \end{cases}$$

$$\begin{aligned} \mathrm{E}(\ln X_0) &= \int_{-\infty}^{+\infty} \ln x \cdot p_0(x) dx \\ &= \frac{1}{\sqrt{2\pi}C_0\sigma} \int_0^{+\infty} \ln x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}C_0\sigma} \int_0^{+\infty} \ln(\sigma v) \cdot e^{-\frac{(v-k)^2}{2}} \sigma dv \\ &= \frac{1}{\sqrt{2\pi}C_0} \int_0^{+\infty} \ln v \cdot e^{-\frac{(v-k)^2}{2}} dt + \frac{\ln\sigma}{C_0} \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v-k)^2}{2}} dv \\ &= \frac{1}{\sqrt{2\pi}C_0} \int_0^{+\infty} \ln v \cdot e^{-\frac{(v-k)^2}{2}} dv + \ln\sigma \end{aligned}$$

Denoted by $f_0(k) = \int_0^{+\infty} \ln v \cdot e^{-\frac{(v-k)^2}{2}} dv$, then $\mathrm{G}(X_0) = \sigma e^{\frac{1}{\sqrt{2\pi}C_0} f_0(k)}$

Below is the discussion about the error of $\mathrm{G}(X)$ relative to $\mathrm{G}(X_0)$:

$$\begin{aligned} \left|\ln\frac{\mathrm{G}(X_0)}{\mathrm{G}(X)}\right| &= \frac{1}{\sqrt{2\pi}} \left|\frac{f_0(k)}{C_0} - \frac{f(k)}{C}\right| \\ &= \frac{1}{\sqrt{2\pi}} \left|\frac{f_0(k) - f(k)}{C_0} + \frac{C - C_0}{CC_0} f(k)\right| \\ &< \frac{1}{\sqrt{2\pi}} [\frac{1}{C} \left|\int_0^{k-3} \ln x \cdot e^{-\frac{(x-k)^2}{2}} dx\right| + \frac{1-C}{C^2} f(k)] \end{aligned}$$

(The absolute value inequality $|a+b| \le |a| + |b|$ is used here, and it is noted that $C \le C_0 < 1$)

$$\le \frac{1}{\sqrt{2\pi}} [\frac{\left|\int_0^{k-3} \ln x dx\right|}{Ce^{4.5}} + \frac{1-C}{C^2} f(k)]$$

where $\int_0^{k-3} \ln x dx = (x\ln x - x)\Big|_0^{k-3} = (k-3)\ln(k-3) - (k-3) - \lim_{x\to 0^+} x\ln x$, according to L Hospital Rule, $\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{\ln x}{\frac{1}{x}} = \lim_{x\to 0^+} \frac{x^{-1}}{-x^{-2}} = \lim_{x\to 0^+} -x = 0$. Thus, the error estimation is obtained as follows:

$$\left|\ln\frac{\mathrm{G}(X_0)}{\mathrm{G}(X)}\right| \le \frac{1}{\sqrt{2\pi}C}\left[\frac{|(k-3)\ln(k-3)-(k-3)|}{e^{4.5}} + (\frac{1}{C}-1)f(k)\right]$$

When $k = \frac{\mu}{\sigma} = 4$, substitute with $C \approx 0.9987$, $f(4) \approx 3.38812$, and it is obtained:

$$\left|\ln\frac{\mathrm{G}(X_0)}{\mathrm{G}(X)}\right| < 6.196\times 10^{-3}$$

$$\text{i.e. } -0.6177\% < \frac{\mathrm{G}(X_0)}{\mathrm{G}(X)} - 1 < 0.6215\%$$

When $k = 8$, $\left|\ln\frac{\mathrm{G}(X_0)}{\mathrm{G}(X)}\right| < 0.01622$, $-1.609\% < \frac{\mathrm{G}(X_0)}{\mathrm{G}(X)} - 1 < 1.636\%$, it can be seen that the error generated from the conversion of the original normal distribution into constant-positive normal distribution is within the acceptable range.

In summary, after simplification, fitting and approximation of the function, we believe that the compound expectation under the constant-positive normal distribution $\mathrm{G}(X) \approx 1.0702(\frac{\mu}{\sigma} - 0.2816)^{0.9807}\sigma$ is an acceptable result within the error range.

4.5 Exploration on the optimum investment strategy under normal distribution

Supposing that the investor extracts a fixed proportion (t) of funds from the principal ($0 \le t \le 1$) to invest in the project, explore the optimum investment strategy in this mode.

Let $Y = tX + (1-t)\,(0 \le t \le 1)$, $X \sim N_+(\mu,\sigma^2)$, it can be easily obtained that Y corresponds with constant-positive normal distribution, and the maximum value of $\mathrm{G}(Y)$ is calculated as follows:

(1). When $t = 0$, $Y \equiv 1$, obviously, $\mathrm{G}(Y) = 1$.

(2). When $0 < t \le 1$,

$$\mathrm{E}(\ln Y) = \int_{-\infty}^{+\infty} \ln[tx + (1-t)] \cdot p(x)dx$$ ($p(x)$ is the density function of *X*)

$$= \frac{1}{\sqrt{2\pi}C\sigma} \int_{\mu-3\sigma}^{+\infty} \ln[tx + (1-t)] \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi}C\sigma} \int_{k-3}^{+\infty} \ln[t\sigma v + (1-t)] \cdot e^{-\frac{(v-k)^2}{2}} \sigma dv$$ (Denoted by $v = \frac{x}{\sigma}$, $k = \frac{\mu}{\sigma}$)

$$= \frac{1}{\sqrt{2\pi}C} \int_{k-3}^{+\infty} [\ln(v + \frac{1-t}{t\sigma}) + \ln(t\sigma)] \cdot e^{-\frac{(v-k)^2}{2}} dv$$

$$= \frac{1}{\sqrt{2\pi}C} \int_{j-3}^{+\infty} \ln v \cdot e^{-\frac{(v-j)^2}{2}} dv + \frac{\ln(t\sigma)}{C} \int_{k-3}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(v-k)^2}{2}} dv \quad (j = k + \frac{1-t}{t\sigma})$$

$$= \frac{1}{\sqrt{2\pi}C} f(j) + \ln(t\sigma)$$

Thus $G(Y) = e^{\mathrm{E}(\ln Y)} = \sigma t e^{\frac{1}{\sqrt{2\pi}C} f(j)}$ ($0 < t \le 1$).

From $f(j) \approx 0.1698 + 2.455 \ln(j - 0.2816)$, $j = \frac{\mu}{\sigma} + \frac{1-t}{t\sigma}$, it is obtained that

$$\mathrm{G}(Y) \approx g_Y(t) = \begin{cases} 1.0702(\frac{1-t}{\sigma t} + \frac{\mu}{\sigma} - 0.2816)^{0.9807} \sigma t & (0 < t \le 1) \\ 1 \quad (t = 0) & \end{cases}$$

Then, what value does *t equal to for* $\mathrm{G}(Y)$ to obtain the maximum value?

I. When $t = 0$, it can be easily obtained that $g_Y(t) = 1$.

II. When $0 < t \le 1$,

$$g_Y'(t) = 1.0702[(\frac{1-t}{\sigma t} + \frac{\mu}{\sigma} - 0.2816)^{0.9807} \sigma + 0.9807(\frac{1-t}{\sigma t} + \frac{\mu}{\sigma} - 0.2816)^{-0.0193}(-\frac{1}{t^2})t]$$

$$= 1.0702(\frac{1-t}{\sigma t} + \frac{\mu}{\sigma} - 0.2816)^{-0.0193}[(\frac{1-t}{\sigma t} + \frac{\mu}{\sigma} - 0.2816)\sigma - \frac{0.9807}{t}]$$

$$= 1.0702\sigma^{0.0193}(\frac{1}{t} + \mu - 0.2816\sigma - 1)^{-0.0193}(\frac{0.0193}{t} + \mu - 0.2816\sigma - 1).$$

Denoted by $q=\mu-0.2816\sigma-1>-1$ (it is noted that $\mu>3\sigma$ and $\sigma>0$), then

$$g_Y{}'(t)=1.0702\sigma^{0.0193}(\frac{1}{t}+q)^{-0.0193}(\frac{0.0193}{t}+q).$$

(1) When $q>-0.0193$, $\because \frac{1}{t}>\frac{0.0193}{t}\geq 0.0193$, $\therefore$ $g_Y{}'(t)>0$, when $t=1$, $g_Y(t)=1.0702(\frac{\mu}{\sigma}-0.2816)^{0.9807}\sigma$, thus obtaining the maximum value. Then according to $\mu>3\sigma$, $g_Y(t)>1.0702\cdot(3-0.2816)^{0.9807}\sigma=2.854\sigma$, moreover, $q>3\sigma-0.2816\sigma-1>-0.0193$, i.e. $\sigma>0.3608$, thus $g_Y(t)>1.03>1$.

(2) When $-1<q\leq-0.0193$, $g_Y{}'(t)>0$ ($t\in(0,-\frac{0.0193}{q})$), $g_Y{}'(t)<0$ ($t\in(-\frac{0.0193}{q},1]$), thus, when $t=-\frac{0.0193}{q}$, $g_Y(t)=0.9729(-\frac{\sigma}{q})^{0.0193}$, obtaining the maximum value.

In particular, according to $\mu=0.2816\sigma+q+1>3\sigma$, it is certainly that $\sigma<0.3679(q+1)$, moreover, $0.9729(-\frac{\sigma}{q})^{0.0193}<1\Leftrightarrow\sigma<-4.152q$, $\therefore$ If $0.3679(q+1)\leq-4.152q$, i.e. $q\leq-0.0814$, then $g_Y(t)<1$.

In summary, $q=\mu-0.2816\sigma-1>-1$, let $t=t_{\max}$, and $\mathrm{G}(Y)$ obtains the maximum value, then

① When $q>-0.0193$, $t_{\max}=1$.

② When $-0.0814<q\leq-0.0193$, if $0.9729(-\frac{\sigma}{q})^{0.0193}>1$, then $t_{\max}=-\frac{0.0193}{q}$; if $0.9729(-\frac{\sigma}{q})^{0.0193}=1$, then $t_{\max}=-\frac{0.0193}{q}$ or 0; if $0.9729(-\frac{\sigma}{q})^{0.0193}<1$, then $t_{\max}=0$.

③ When $-1<q\leq-0.0814$, $t_{\max}=0$.

Note: $\mu>3\sigma$ is taken as the study range at the beginning of the fourth section, the approximate conversion of the original normal distribution of the rate of return into the constant-positive normal distribution is conducted and the approximate fitting of definite integral function is carried out, so there will be a certain deviation between the $t_{\max}$ selection strategy obtained through calculation and some other conclusions; however, the final deviation is not significant.

V. Proof of Absolute Optimization of the Scheme under the Model of Compound Expectation

In the above derivation of the model of compound expectation, we just explore the optimum scheme on the premise of "extracting a fixed proportion (*t*) of funds from the principal to invest in the project each time". Does any other set of methods exist (e.g. input of a fixed absolute amount of principal each time, or increase or decrease in the percentage of actual investment this time based on the actual rate of return last time) to get a higher rate of return on investment theoretically?

We must prove that: Given the distribution conditions (they can be arbitrary distributions) of the expected rate of return on investment of each investment opportunity, **there exists a fixed optimum scheme**, that is, a proportion ($t_{\max}$) of funds is predetermined and extracted from the remaining funds to invest in the project each time (however, according to the different distribution of the expected rate of return, $t_{\max}$ can be different each time), and a higher rate of return can not be obtained through other schemes. It is proved as follows:

Let $X_i\,(i = 1, 2,\cdots, n)$ be a discrete or continuous random variable, where X_i is the distribution of the expected ratio of capital input and interest to capital input for the i^{th} time of investment, t_i is the proportion of the invested funds to the current total funds for the i^{th} time of investment, a_i is the total amount of the principal prior to the i^{th} time of investment, $a_1 = 1$, $a_{i+1} = a_i[t_i X_i + (1-t_i)]\,(i = 1, 2,\cdots, n-1)$, then $G(a_n)$ is calculated as follows:

$$a_2 = a_1[t_1 X_1 + (1-t_1)],$$

$$a_3 = a_2[t_2 X_2 + (1-t_2)] = [t_1 X_1 + (1-t_1)][t_2 X_2 + (1-t_2)],$$

……

$$a_n = \prod_{i=1}^{n-1}[t_i X_i + (1-t_i)].$$

It is obtained by taking the logarithms of both sides $\ln a_n = \sum_{i=1}^{n-1}\ln[t_i X_i + (1-t_i)] = \sum_{i=1}^{n-1}\ln Y_i$, thus

$$E(\ln a_n) = \sum_{i=1}^{n-1} E(\ln Y_i),$$

$$\mathrm{G}(a_n) = e^{E(\ln a_n)} = \prod_{i=1}^{n-1} e^{E(\ln Y_i)} = \prod_{i=1}^{n-1}\mathrm{G}(Y_i)$$

According to the above proof, it is known that the result of the rate of return on each investment will affect the absolute yield value of the next transaction, but the rate of return on each investment is theoretically independent. In other words, the optimum investment proportion to be taken each time is only related to the distribution of the expected rate of return on the investment project, but independent of the previous or later several investment projects as well as their profit and loss results.

Therefore, we can draw the conclusion: **Faced with repeated long-term investment opportunities, the optimum investment scheme is to preset an investment proportion $t_{\max}$ according to the distribution of the expected rate of return on each investment from a long-term foresight, while other methods can not obtain a higher rate of return than this method theoretically.** In other words, if an investor faces a valid investment opportunity with the same distribution of the rate of return for a long term, then the optimum investment strategy is to extract a proportion ($t_{\max}$) of funds from the principal for investment each time - this is absolutely the optimum method!

VI. In-depth Significance and Important Conclusion of the Model of Compound Expectation

Prior to more in-depth study, we must make it clear: As all the returns on actual investments must be finite, there surely exists the mathematical expectation of the **valid distribution function of the expected rate of return on investment** (It will never be infinite)

6.1 Monotonicity of compound expectation value G(Y) in actual investment

In the derivation process in Section 4.5, we also proved one thing: In the study of the model of compound expectation based on the expected rate of return under normal distribution, if $0 < t_{\max} < 1$ and when $t < t_{\max}$, the function $g_Y(t)$ increases monotonically, and the theoretical rate of return on investment increases as *t* increases; when $t > t_{\max}$, the function $g_Y(t)$ decreases monotonically, and the theoretical rate of return on investment decreases as *t* increases

That is to say, "With the increase or decrease in the investment proportion *t*, there are no sharp fluctuations in the yield rate of compound expectation."

So, whether this conclusion can be extended to the precondition of the expected rate of return on investment under arbitrary distribution?

Proof: Let the function $F(t) = E(\ln Y) = \int_0^{+\infty} \ln(tx+1-t)p(x)dx$. As $E(X) = \int_0^{+\infty} xp(x)dx$ exists, then

$$F'(t) = \int_0^{+\infty} \frac{x-1}{tx+1-t} p(x)dx .$$

$\because \frac{x-1}{tx+1-t}$ decreases monotonically about t $\therefore F'(t)$ also decreases monotonically about t, thus the equation $F'(t) = 0$ has one root at most in $(0,+\infty)$. Thus, the compound expectation function $G(Y) = e^{E(\ln Y)}$ has one maximum value at most. This completes the proof.

It can be seen that for any investment opportunity, if $0 < t_{\max} < 1$ and when $t < t_{\max}$, the yield rate of compound expectation increases monotonically; when $t > t_{\max}$, the yield rate of compound expectation decreases monotonically. **There are no sharp fluctuations in the yield rate of compound expectation.**

This conclusion is very useful for actual investment. When you still have idle remaining funds, if your investment proportion for any investment opportunity has not yet reached its optimum value $t_{\max}$, no matter how its expected rate of return is distributed, a higher return can be obtained as long as the investment proportion is increased, so do not worry and increase position boldly!

6.2 Influence of σ value in Markowitz's investment theory on the optimum yield rate

The model of compound expectation put forward by us is unique, but in comparison with other investment theories and models, whether some important conclusions derived by them can be more fully proved in the model of compound expectation?

Markowitz has two classic assertions in his famous theory of portfolio: When the distribution of the expected ratio of capital input and interest to capital input corresponds with normal distribution (1) When σ is same, the investors will choose the investment opportunities with greater mathematical expectation μ for investment. (2) When μ is same, the investors will choose the investment opportunities with smaller σ. (See Markowitz's *Mean-Variance Analysis in Portfolio Choice and Capital Markets*)

The first point is undoubtedly correct, and on the premise of the same degree of dispersion in the distribution of the yield rate, the investment opportunities with high mathematical expectation will result in a higher theoretical rate of return for the investors. For the second point, our model of compound expectation does not directly judge the risks according to σ, so whether the attractiveness of the investment opportunities is relatively low with greater σ, and the optimum yield rate for the investors will be smaller theoretically as Markowitz said?

The answer is yes! The proof is as follows:

Firstly, we converted the normal distribution of the original ratio of capital input and interest to capital input into **"even-removal" normal distribution**, that is, the new curve of ratio of capital input and interest to capital input obtained through removing the images of $x < \mu - 3\sigma$ on the left side and $x > \mu + 3\sigma$ on the right side of normal distribution, on the premise that the mathematical expectation μ is kept unchanged in the new distribution of the ratio of capital input and interest to capital input.

Definition 3 Denote the distribution function of the normal distribution $\mathrm{N}(\mu,\sigma^2)$ **as** $F(x)$**, then**

$$F_A(x) = \begin{cases} \dfrac{F(x) - F(0)}{1 - 2F(0)} & (x > 0) \\ 0 & (x \le 0) \end{cases}$$

is taken as the distribution of the distribution function, called "even-removal"

normal distribution and denoted by $N_A(\mu,\sigma^2)$**.** (See Figure 4)

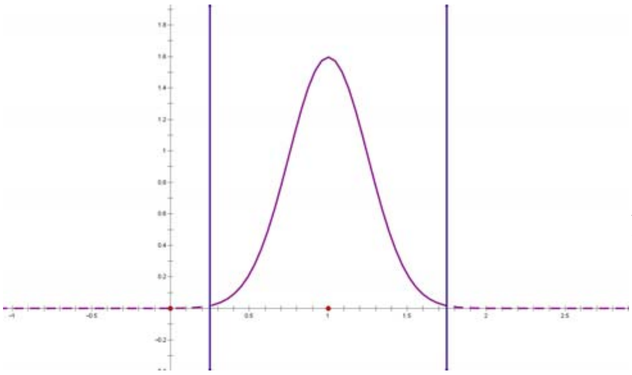


Figure 4

As the μ value of the function is kept unchanged (that is, the income risk is reduced simultaneously) and the removed parts account for a small proportion of the image (only 0.26%), the above conversion will not produce significant errors in comparison with the original normal distribution.

Let $X \sim N_A(\mu,\sigma^2)$, denote the constant as $C_A = 1-2F(\mu-3\sigma)$, take $x_{Li} = \mu - \frac{3i}{n}\sigma\,(i = 1, 2, \cdots, n)$, $x_{Ri} = \mu + \frac{3i}{n}\sigma\,(i = 1, 2, \cdots, n)$,

$\Delta F_i = F_X(x_{L(i+1)}) - F_X(x_{Li}) = F_X(x_{R(i+1)}) - F_X(x_{Ri})$ ($i = 1, 2, \cdots, n-1$),

$\Delta F_i = \frac{1}{\sqrt{2\pi}C'\sigma}\int_{\mu+\frac{3i}{n}\sigma}^{\mu+\frac{3(i+1)}{n}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \frac{1}{\sqrt{2\pi}}\int_{\frac{3i}{n}}^{\frac{3(i+1)}{n}} e^{-\frac{x^2}{2}}dx$ is not related to σ.

$$\mathrm{G}(X) = \lim_{n\to+\infty}(\mu^{\Delta F_0}\prod_{i=1}^{n-1}x_{Li}{}^{\Delta F_i}x_{Ri}{}^{\Delta F_i}) = \lim_{n\to+\infty}\prod_{i=1}^{n-1}(\mu^2 - \frac{9i^2\sigma^2}{n^2})^{\Delta F_i}$$ (

$\Delta F_0 = F_X(x_{R1}) - F_X(x_{L1})$).

The bigger the σ, the smaller the $\mu^2 - \frac{9i^2\sigma^2}{n^2}$; the smaller the $\prod_{i=1}^{n-1}(\mu^2 - \frac{9i^2\sigma^2}{n^2})^{\Delta F_i}$, the smaller the $\mathrm{G}(X)$. And

$$\mathrm{G}(X) < \lim_{n\to+\infty}\prod_{i=1}^{n-1}(\mu^2)^{\Delta F_i} = \mu\hat{}(2\lim_{n\to+\infty}\sum_{i=1}^{n-1}\Delta F_i) = \mu.$$

So we came to a conclusion: **The greater the variance** σ **in the normal distribution of the expected rate of return, the lower the theoretical rate of return obtained from all-capital input.**

After considering the optimum investment scheme, can it still get a similar conclusion?

6.3 Analysis on the characteristics and meaning of the optimum compound expectation

It aims to prove that the above conclusion still holds after considering the optimum strategy. We firstly prove this point: If the expected rate of return on investment corresponds with the even-removal normal distribution, the theoretically obtained ratio of capital input and interest to capital input still corresponds with the even-removal normal distribution after part of the principal is invested (Theorem II).

Theorem II If $X \sim N_A(\mu,\sigma^2)$**, then** $Y = tX + (1-t) \sim N_A(t\mu + 1 - t, t^2\sigma^2)$**.**

Proof: Let the distribution functions of X and Y be $F_X(x)$ and $F_Y(x)$ respectively, $X \sim N_A(\mu,\sigma^2)$, and let the distribution function of the normal distribution $N(\mu,\sigma^2)$ be $F_N(x)$. Denote the constant as $C_A = 1 - 2F(\mu - 3\sigma)$. According to the definition,

$$F_X(x) = \begin{cases} \dfrac{F_N(x) - F_N(\mu - 3\sigma)}{C_A} & (\mu - 3\sigma < x < \mu + 3\sigma) \\ 0 & (\text{Other}) \end{cases}$$

According to $F_Y(tx + 1 - t) = F_X(x)$, it is obtained:

$$F_Y(x) = F_X(\frac{x-1}{t} + 1) = \begin{cases} \dfrac{1}{C_A}[F_N(\dfrac{x-1}{t} + 1) - F_N(\mu - 3\sigma)] & (|x - (t\mu + 1 - t)| < 3t\sigma) \\ 0 & (\text{Other}) \end{cases}$$

Let the distribution function of the normal distribution $N(t\mu + 1 - t, t^2\sigma^2)$ be $F_t(x)$, then $F_t(tx + 1 - t) = F_N(x)$, i.e. $F_N(\frac{x-1}{t} + 1) = F_t(x)$, so

$$F_Y(x) = \begin{cases} \dfrac{F_t(x) - F_N(\mu - 3\sigma)}{C_A} & (|x - (t\mu + 1 - t)| < 3t\sigma) \\ 0 & (\text{Other}) \end{cases}$$

According to the definition, Y corresponds with even-removal normal distribution. The proof of Theorem II is now completed.

The following is the proof of the subject in this section:

There are two investment projects X_1, X_2 with their expected rate of return under normal distribution, which correspond to the variances σ_1, σ_2 respectively, the optimum proportions of capital input are $t_{\max1}, t_{\max2}$ respectively to obtain the maximum theoretical rate of return.

If $\sigma_1 > \sigma_2 > 0$, the same proportion of funds, i.e. $t_{\max1}$, is invested for the above two opportunities, and the compound expectations obtained are G(Y_1) and G(Y_2) respectively.

It is known according to Theorem II and the conclusion reached in Section **6.2**: As $\sigma_1 > \sigma_2$, it is sure that G(Y_1) < G(Y_2). In the investment project X_2, when $t = t_{\max2}$, the theoretically obtained ratio of principal and interest to principal is G(Y_2'), whereas G(Y_2) is obtained when $t = t_{\max1}$ in the investment project X_2.

Obviously, G(Y_2') is the result under the optimum strategy, so G(Y_2')$\geq$G(Y_2) > G(Y_1).

In summary, it is proved: **When $\sigma_1 > \sigma_2$ for the two investment opportunities, G(Y_1) < G(Y_2') under the optimum strategy.**

It can be obtained according to the above proof: Considering the optimum input of the funds, our conclusion can be changed as: **The greater the variance σ in the normal distribution of the expected rate of return, the lower the optimum rate of return of the investment project**. Therefore, although we do not use σ to measure the investment risks, Markowitz's conclusions in his famous theory of portfolio as mentioned above are still correct in our model of compound expectation. They all reflect the same rule: **When the fluctuation in the expected rate of return increases, the theoretical rate of return finally obtained by the investors will surely decline.**

Under the model of compound expectation, for the investment opportunities where the expected ration of capital input and interest to capital input corresponds with normal distribution, the theoretical rate of return on investment obtained by the investors under the optimum investment strategy will be surely less than μ. For an opportunity of the ration of capital input and interest to capital input under normal distribution where $\mu \leq 1$, the investors shall not make any investment for this opportunity in theory.

6.4 Exploration on the magnitude relationship between compound expectation and mathematic expectation

In the first section of this dissertation, the readers can obviously find that the compound expectation calculated according to the definition is far smaller than the traditional mathematical expectation. It can also be seen from the above two sections that the compound expectation will also be surely less than the mathematical expectation under the condition that the expected ratio of capital input and interest to capital input corresponds with the normal distribution. After more in-depth study on the magnitude relationship between them, we believe that: **When the expected ratio of capital input and interest to capital input, i.e. *X*, is arbitrarily distributed (the default mathematical expectation is greater than 1), the optimum compound expectation is surely less than or equal to the mathematical expectation.**

The proof is as follows:

Let $h(x) = \ln x \qquad \because h''(x) = -\frac{1}{x^2} < 0$

$\therefore\ h(x) = \ln x$ is a convex function, and according to Jensen's inequality, it is obtained that $E(\ln x) \leq \ln[E(x)]$

$\therefore\ E(\ln Y) \leq \ln[E(Y)]$ and $Y = t(X-1)+1$

$\because 0 \leq t \leq 1 \quad E(X) > 1 \qquad \therefore E(Y) \leq E(X)$

So, the optimum compound expectation value $\mathrm{G}(Y) = e^{\mathrm{E}(\ln Y)} < E(X)$

It can be said that when $E(X) \leq 1$ in the distribution of the expected ratio of capital input and interest to capital input for an investment project, the project is totally not worthy of investment, because the optimum compound expectation is $\mathrm{G}(Y) < E(Y) \leq 1$. In actual investment, if the investment project X_i where $E(X) \leq 1$ exists among a number of investment projects faced by the investors, it can be ignored directly (i.e. $t_i = 0$), and it is not necessary to include it into the complicated calculations, which will greatly simplify the investment decision-making process about allocation of funds.

Through combining the study in **6.2 ~ 6.4** with the actual investment, we found that on the premise of keeping μ unchanged, the increase in the variance of the expected rate of return σ actually magnifies the risks and returns simultaneously, resulting in decrease in the theoretical yield rate of compound expectation, so that the long-term rate of return on investment is adversely affected. The above phenomenon actually shows that **the investment risks and returns are not equivalent**.

The public investors lose more often than they win in the stock market, which is actually related to this conclusion. In this balanced market, we can assume that **the mathematical expectation is roughly 1** (break-even) on average for all investment opportunities, but from the above argument, we can see that the compound expectation of actual investment will be surely less than the mathematical expectation, and the compound expectation will further decrease with the increase in the fluctuation of the expected rate of return. As a matter of fact, most investors pursue high profits under high risks, moreover, with the accumulated high short-term transaction fees, such investors surely lose more often than they win for the long run, and it is inevitable for the retail investors to suffer losses on the whole.

By extension, the loss risks and yield opportunities exert completely non-equivalent influence on the final rate of return in the two-point distribution, multi-point distribution and normal distribution of the rate of return. This doctrine of “High Risk, High Yield” is absurd. To achieve more excellent results in the long-term investment, it is more effective to reduce the potential risks of loss than to pursue a certain substantial yield. It also coincides with Warren Buffett’s investment motto - **Preserve the capital**.

However, the yield rate of compound expectation obtained through our calculations is always much lower than the mathematical expectation, is there any means for us to further improve the rate of return on long-term investment?

The answer is: “diversification”.

VII. Investment Diversification under the Model of Compound Expectation

7.1 Overview of diversification and decreasing principle of relative yield rate

Markowitz once specifically clarified the significance of investment diversification in his book *Portfolio Selection: Efficient Diversification of Investments*, like the saying “Do not put all your eggs in the same basket” generally embraced by the investment community, it is actually about the reduction of portfolio variance through investment diversification, so as to achieve the objective of diversification of risks.

Despite the smart diversification does improve the profitability, we can not completely agree on the viewpoint of “diversification of risks”. How does the investment diversification increase the theoretical rate of return exactly?

With regard to the investment opportunity of two-point distribution (50% probability of principal doubling, and 50% probability of principal halving) as mentioned in the leading question at the beginning of the first section, we can only get 6.07% compounded rate of return theoretically after investment optimization. However, if an investor has numerous such opportunities at the same time and averages the principal for investment, then he can get 25% compounded rate of return (equivalent to the mathematical expectation), which is much higher than the former. In other words, diversification of funds into different investment projects can improve the theoretical rate of return on investment.

In addition to reducing the fluctuation risks of portfolio returns, the strategy of investment diversification also improves the theoretical rate of return through obtaining the independent actual yield results from many opportunities, so as to achieve a higher overall profitability in the long term. When some investment projects are subject to certain risks of loss, the investment diversification becomes necessary. However, it does not just aim to balance a number of risks as many people embrace, but improve the efficiency in the use of funds to a large extent.

Prior to the exploration on the diversification strategy, we need to prove an important rule: **Faced with an investment project whose expected rate of return is of arbitrary continuous distribution, the relative rate of return on investment obtained by the newly added funds will decline gradually with gradual increase in the invested funds, that is, the decreasing principle of marginal relative yield rate.**

Firstly, we need to introduce the concept of **relative investment yield rate**: **The ratio of the yield rate of compound expectation on a certain capital input to the capital input.** (It is not related to the total principal)

Marginal relative yield rate is **the relative yield rate for an increased unit of capital input.**

In terms of quantification: The relative investment yield rate $R=\frac{G(Y)}{t}-1$.

When an increased unit of capital input tends to zero, $MR=\frac{dG(Y)}{dt}$ undoubtedly, all the capital will tend to be invested in the projects with greatest R and MR.

Theorem III The greater the proportion of the invested funds (***t***), the smaller the marginal relative yield rate (***MR***).

Proof: MR= $\frac{dG(Y)}{dt} = G(Y)\frac{d}{dt}\int_0^{+\infty}\ln(tx+1-t)\,p(x)dx = G(Y)\int_0^{+\infty}\frac{x-1}{tx+1-t}p(x)dx$

$$\frac{dMR}{dt} = \frac{d^2G(Y)}{dt^2} = G(Y)(\int_0^{+\infty}\frac{x-1}{tx+1-t}p(x)dx)^2 - G(Y)\int_0^{+\infty}\frac{(x-1)^2}{(tx+1-t)^2}p(x)dx)$$

$$= G(Y)\{[E(\frac{X-1}{Y})]^2 - E(\left(\frac{X-1}{Y}\right)^2)\}$$

$$= -G(Y)Var(\frac{X-1}{Y}) \le 0$$

(The formula $Var(X) = E(X^2) - [E(X)]^2$ will be used)

i.e. $\frac{dMR}{dt} = \frac{d^2G(Y)}{dt^2} \le 0$.

So, MR decreases monotonically in the definition domain. The proof of **Theorem III** is now completed.

This is a very special conclusion. As the financial market is considered to be a fully competitive market, the basic assumptions of the model of compound expectation also recognize that "**The capital investment will not affect the distribution of the expected rate of return of an investment opportunity**". According to the modern economic viewpoint, the marginal yield rate shall be unchanged. However, it is shown according to the above proof that the **decreasing principle of marginal yield rat**e still holds even on the premise that the capital input will not reduce the expected rate of return. Therefore, when choosing among some opportunities in the process of investment, it is only required to directly compare their marginal relative yield rates, so that the investors can directly determine where the next share of funds should be invested.

Thus we can draw a conclusion: Faced with the same investment opportunity whose expected rate of return is of normal distribution, when the proportion of the invested principal increases, the use efficiency of funds will decline continuously with the increase in the proportion of funds input. We can say with certainty that the investors should diversify the funds into some similar opportunities rather than invest a considerable amount of funds in an investment opportunity with risks. Therefore, if the investors face multiple same opportunities, the **investment diversification** can **improve the overall efficiency** of funds, resulting in higher yield rate of compound expectation.

7.2 Study on the optimum strategy under multiple investment opportunities

In the study of the previous six sections, we just discussed the optimum selection for fund input when faced with an investment project in a certain period of time; therefore, when a proportion (*t*) of the funds is used for investment, the other funds are laid idle obviously. There always exist risk-free interest rates in the markets as well as many different investment options at the same time, which requires us to use the model of compound expectation to **construct the optimum portfolio** under the given **risk-free interest rates and multiple investment projects**, so as to determine how to allocate a proportion of funds into each project, improve the efficiency in the use of funds and get the highest theoretical rate of return on the whole.

Suppose that the valid investment opportunities faced by the investors at the same time are A_0, $A_1, A_2, A_3 \ldots, A_n$ ($\mu \geq 1$ for each investment opportunity), whose corresponding ratios of capital input and interest to capital input are (1 + Q), $X_1, X_2, X_3 \ldots, X_n$ (where Q is the fixed rate of return for the risk-free investment opportunity A_0, Q> 0 by default. The remaining X_i is subject to arbitrary continuous distribution, where i = 1, 2, ..., n) , the proportion of the funds to the principal invested by the investor for each investment opportunity is t_0, $t_1, t_2, t_3 \ldots, t_n$ ($0 \leq t_i \leq 1$, i = 0, 1, 2, …, n) respectively, and the theoretical ratio of the capital input and interest to the total principal obtained by the investor for the corresponding investment opportunity under the corresponding investment proportion is Y_0, $Y_1, Y_2, Y_3 \ldots, Y_n$ respectively. The above investment opportunities have the same investment cycle, and the final investment yield result of each investment opportunity is independent.

We require to solve t_0, $t_1, t_2, t_3 \ldots, t_n$, so as to make the original principal obtain the maximum rate of return in theory. (As a risk-free rate of return exists, there will be no idle funds, so $\sum_{i=0}^{n} t_i = 1$)

According to the definitions of compound expectation and diversification, the specific method for directly solving t_0, $t_1, t_2, t_3 \ldots, t_n$ under the optimum strategy is as follows:

Compound expectation value of the total principal input

$$G(Y) = 1 + \sum_{i=0}^{n} [G(Y_i) - 1]$$

$$G(Y_i) = \begin{cases} 1 + Qt_0 & (i = 0) \\ t_i X_i + 1 - t & (i \geq 1) \end{cases}$$

And $t_0 + t_1 + t_2 + \ldots + t_n = 1$

Combine the above equations to solve t_0, $t_1, t_2, t_3 \ldots, t_n$, making G(Y) obtain the maximum value.

As the equations are very complex, which contain many undetermined parameters, and many classified discussions are involved, it is difficult to obtain the best funds allocation through direct solving.

We found that it is also difficult to find the solution based on the fundamental theorem of economics – "When the marginal yield rate is equal for all investment projects to which the funds have been allocated, the optimization strategy of funds allocation is obtained".

In order to better solve the practical issue about **efficient allocation of the funds for multiple different investment opportunities at the same time**, we used the computer programming method. The basic principle is:

(1) In any case, choose the opportunities with the theoretical highest relative yield rate for the funds.

(2) According to the decreasing principle of marginal yield rate, the yield rate obtained from the added capital input is surely less than the marginal yield rate of the previous investment.

The specific steps for solving t_0, $t_1, t_2, t_3 \ldots, t_n$ with programming method are as follows:

(1) Initialization: Divide the total capital input into equal small portions of enough quantity, i.e. *P* (a larger positive integer can be taken for *P*, such as 100,000), and the proportion of each small portion to the total capital is $t = \frac{1}{P}$. The number of the shares of funds which has been invested in each project is assigned to 0.

(2) Make a judgment for the first share of invested funds: Put it into X_i and the corresponding compound expectation value obtained is G($Y_{i,1}$), the relative investment yield rate is $R_{i,1} = \frac{G(Y_{i,1})}{t} - 1$, select the maximum $R_{i,1}$. If $R_{i,1} > Q$, put the first share of funds into the project with maximum $R_{i,1}$, assign the corresponding value for $S_i := S_i + 1$, go to the third (3) step; if $R_{i,1} \leq Q$, go to the

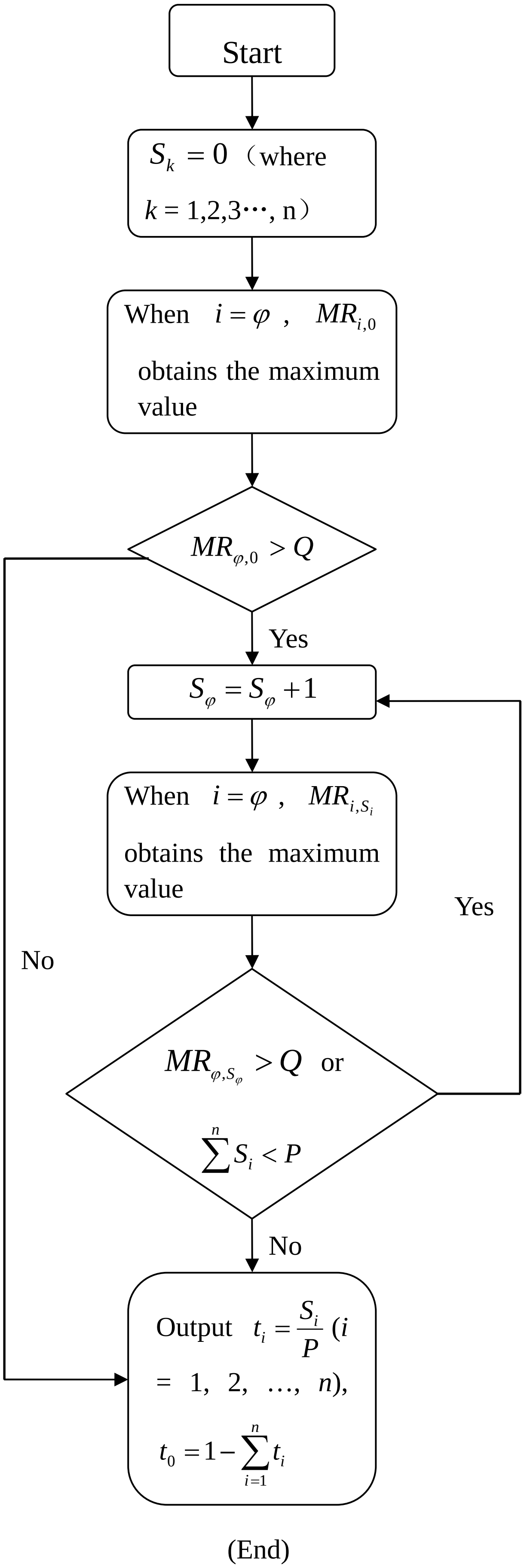
fifth (5) step.

(3) If the total capital input is less than P at the time (that is, there are still remaining funds), increase of 1 share may be considered: Currently, S_i of funds were invested in the project X_i, if 1 more share is increased, then the marginal relative yield rate is $MR_{i,S_i} = \frac{G(Y_{i,S_i+1}) - G(Y_{i,S_i})}{t} - 1$. After comparison one by one, select the maximum MR_{i,S_i}. If the funds have run out at this time, go to the fifth (5) step.

(4) Judge whether to increase the funds: If the marginal relative yield rate $MR_{i,S_i} \leq Q$, that is, this newly added share of funds can not obtain the rate of return higher than the risk-free yield rate in all investment projects, increase in funds shall be stopped, and go to the fifth (5) step; conversely, if $\Delta R_{i,S_i} > Q$, then put this newly added share of funds into the investment project with the maximum MR_{i,S_i}, and assign the corresponding value for $S_i := S_i + 1$.

(5) Output: Output $t_i = \frac{S_i}{P}$ (i = 1, 2, $\cdots$, n), $t_0 = 1 - \sum_{i=1}^{n} t_i$.

The block diagram for the above solution is as follows: (part of the calculation process is skipped)


Start
$S_k = 0$（where $k = 1,2,3\cdots$, n）
When $i = \varphi$, $MR_{i,0}$ obtains the maximum value
$MR_{\varphi,0} > Q$
Yes
$S_\varphi = S_\varphi + 1$
When $i = \varphi$, MR_{i,S_i} obtains the maximum value
Yes
No
$MR_{\varphi,S_\varphi} > Q$ or $\sum^n S_i < P$
No
Output $t_i = \frac{S_i}{P}$ (i = 1, 2, …, n), $t_0 = 1 - \sum_{i=1}^{n} t_i$
(End)

From the model of compound expectation to diversified investment strategy, their essence is to assess the risks and opportunities, so as to satisfy the investors' pursuit of higher theoretical yield rate in the long term. In fact, we can throw off the unscientific traditional investment concepts under this model, and make decisions boldly for the final goal. For example, if there exists a rather good investment project, and its theoretical rate of return is much higher than that of other opportunities according to the calculations, please do not hesitate and put your funds into it!

Bibliography

【1】*A Course in Probability and Statistics*, written by Mao Shisong, Cheng Yiming and Pu Xiaolong, Higher Education Press

【2】http://www.wolframalpha.com/input/?i=integral+exp%28-x%5E2%29

【3】*The Econometrics of Financial Market*, written by J.Y.Campbell (America) et al. //translated by Zhu Pingfang et al., Shanghai University of Finance and Economics Press

【4】*Mean-Variance Analysis in Portfolio Choice and Capital Markets*, written by Markowitz (America)

【5】*Mathematical Analysis*, written by Xu Senlin, Tsinghua University Press

【6】*Portfolio Selection: Efficient Diversification of Investments*, written by Markowitz (America)

【7】*Options, Futures and Other Derivatives*, written by John C. Hull (Canada), translated by Wang Yong (Canada) et al.

Acknowledgements:

I sincerely express my appreciation for my classmates Zhuang Ziquan and He Lvning as well as my tutor Li Ruqing, who helped me with this work.