# A Robust Log-Optimal Strategy and its Application in NYSE

Guangdong Experimental High School

YifengGuo

Chenguang Huang

# 摘要

在赌博与资产配置问题中，对数最优策略最大化的并不是整个资产，而是对数收益率的期望。相对于其他策略，这个策略在理论上具有长期的优越性。它另外一个显著的优点是能够实时反映市场的动态信息，从而使对数收益率得到增长。我们采用简单的方式证明了这些性质。此外，我们考虑如何把对数最优策略应用到资产配置的问题中。因为收益的分布函数在实际问题中一般是未知的，因此需要通过历史数据去估计。估计误差会影响到资产配置的效果，从而犹如“蝴蝶效应”一样对资产配置造成毁灭性的灾难。因此，如何在对数最优策略中克服“蝴蝶效应”是一个很重要的问题。我们根据收益效用得到了一个新的对数最优策略，这个策略是对数收益率期望的二阶近似。我们证明了收益效用的偏差上界能够被投资组合的$l_1$范数和协方差矩阵估计偏差的$l_\infty$范数所控制。这说明了我们提出的策略是稳健的。我们将新策略应用到 NYSE 数据中，模拟结果表明了新策略使收益有很大的提高。

关键词：资产配置效用值、资产增长对数速率、对数最优策略、稳健、资产配置

## Abstract

The log-optimal strategy in gambling and asset allocation problem attempts to maximize the expectation of the logarithmic rate of return but not the gross wealth itself. This strategy has been shown to have long termsuperiority overother strategies in theory. Another remarkable virtue is it allows updating the real-time market information quickly which could increase of the logarithmic rate of return. We will prove these properties ourselves in a simple way. We further consider how to do asset allocation by applying the log-optimal strategy in practice. The distribution function of the returns is usually unknown and need to be estimated from the history in a real world. The estimation error will affect the asset allocation directly, but such effect mayresult in a "butterfly effect" which could bring an investment disaster. Therefore, it is an important issue to answer whether there is log-optimal strategy resisted "butterfly effect". We develop a new log-optimal strategy based on the allocation utility which is a quadratic approximation to the expectation of the logarithmic rate of return. We show that the upper bound of the allocation utility deviation can be controlled by the $l_1$-norm of portfolio and $l_\infty$-norm of the bias of variance matrix estimator. This turns out that our proposed strategy is robust. We apply our strategy for NYSE data. The results showed that our strategy has high performance in return.

Keyword: allocation utility, logarithmic rate of return, log-optimal strategy, robustness, asset allocation.

## 1 Introduction

In investment portfolio, asset allocation is a primary strategythat aims to balance risk and reward by apportioning a portfolio's assets according to an individual's goals, risk tolerance and investment horizon [6]. The log-optimal strategy as one of well-known asset allocation methods attemptsto maximizethe expectation of the logarithmic rate of return but not the capital itself. This strategy has been shown to have long-termsuperiority overother strategies in theory and allows updatingthe real-time market information quickly which could increase the logarithmic rate of return. In practice, the distribution function of returnsis usually unknown and need to be estimated from the history. The attendant estimation error will affect the asset allocation directly, but we expect that the reasonably small error could not affect the allocation tempestuously, in other words,we want to avoid "butterfly effect" which could bring an investment disaster. To take this concern into consideration,we will develop a proper robustlog-optimal strategywhich is not only gain reasonable return but also resist"butterfly effect".

Due to the common ground between investment and gamble:the randomness of return, we introduce fundamental concepts and properties of the log-optimal strategy by beginning with classic gambling capital allocation problem.

### 1.1 Classic gambling capital allocation problem

Suppose a gambler uses a part of his capital ineach game. His returnwill doublewhen he wins or vanishes when he loses.If the games are consecutive and the probability of winning keeps the same in each game, then denote this probability of winning by $p$, and let$p \in [0.5,1)$. We are looking fora strategy to maximize the capital after severalconsecutivegames. Throughout the paper,the return is defined as the ratio of the final capital (price) to the initialcapital (price)over the overall period, which is different from the usualdefinition of the return.

**Analysis:** In an ideal condition, there is no floor limits on bets. The gambler cannot changethe gambling rule but can distribute different bets in different games.Suppose the initial capital of the gambler is$X_0$. We consider two extreme cases:

1) If the gambler uses up all of the capitalin each game, then the capital will become $p^n X_0$ after $n$ games. Note thatthe limit of the rate of return: $lim_{n\to\infty} p^n = 0$for $p \in [0.5,1)$, the capital

goes to zero after infinite games.

2) If the gambler bets nothing in each game,the return is always 1, there will be no space to increase the capital.

Hence, in order to get the maximallogarithmic rate of return and avoid the bankrupt,the optimal betproportion of the capital in each game must be in $[0,1)$and the same if the probability of winning is the same in each game. Wedenote this proportion by$b$.

Supposethe gambler wins$S$ times and loses$n - S$times in an $n$-game gambling, then the capital after $n$ gamesis$X_n = X_0(1+b)^S(1-b)^{n-S}$.

Let $r_n(b)$ to bethe logarithmic rate of return, the average of the logarithmic returns,

$$r_n(b) = \frac{1}{n}\left[log\left(\frac{X_1}{X_0}\right) + \cdots + log\left(\frac{X_n}{X_{n-1}}\right)\right]$$

$$= log\left(\frac{X_n}{X_0}\right)^{\frac{1}{n}} = \frac{s}{n}log(1+b) + \frac{n-s}{n}log(1-b).$$

It actually assesses the capital exponent growth rate and its expectation is

$$r(b) = E\left[log\left(\frac{X_n}{X_0}\right)^{\frac{1}{n}}\right] = E\left[\frac{s}{n}log(1+b) + \frac{n-s}{n}log(1-b)\right]$$

$$= p\,log(1+b) + (1-p)\,log(1-b),$$

which does not depend on $n$, the game number. To maximize the expectationwith respect to$b$,

$$r'(b) = \frac{p}{1+b} - \frac{1-p}{1-b} = \frac{2p-1-b}{(1+b)(1-b)} = 0.$$

We get the solution $b = 2p - 1 \in [0,1)$. Because $r''(b) = \frac{-p}{(1+b)^2} - \frac{1-p}{(1-b)^2} < 0$, $r(2p-1) = log2 + plogp + (1-p)log\,(1-p)$ is the maximumof the expectation in [0, 1).When$p = 1$, it corresponds to a bet which the gamblerwon't fail. Thus, betting all the money in each game is the best strategy.

In addition, if the condition that$p \in [0.5,1)$ doesn't hold, the optimal b is negative. It means that to sell short the $b$ proportion of the capital can increase thecapital if it allows. A gambler usually cannot oversell in aclassicgame, but an investor can do it in many financial markets. So we can only reach the optimal growth ratefor$b = 0$ here.

Through this simple example, we can find that only one fixed optimal bet proportion maximize the expected value of the logarithmic rate of return in a classic gambling with unchangeableprobabilityof winning.

## 1.2 Extension to multivariate assets

We will consider the possible extension of the concepts to multivariate assets from a single capital in last subsection. We meet many multivariate assets in real life. For instance, one may hold several stocks in a stock market, the values ofthose stocksarea kind of multivariate assets.Asset allocation is one of common investment managementsfor multivariate assets by creating an asset mix that will optimize a certain object function. The object function is often the trade-off between expected risk and return for a long-term investment horizon, but ours is different in that this function is the expectation of the logarithmic rate of return as we illustrated in the classic gambling capital allocation problem. Specially, supposethe $d$-dim vector $\boldsymbol{X} = (X_1, X_2, \dots, X_d)^T$ represents theasset returns in a single period, its distributionfunctionis $F$, and denote the asset allocation vector or portfolioby $\boldsymbol{b} = (b_1, b_2, \dots, b_d)^T$. The expectation ofthe logarithmic rate of returnis defined as

$$r_{\boldsymbol{X}}(\boldsymbol{b}) = Elog(\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{X}) = \int log(\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{x})\, dF(\boldsymbol{x}).$$

The optimal portfolio $\boldsymbol{b}_X^*$ in some feasible region $B$ which constrains $\boldsymbol{b}$ is

$$\boldsymbol{b}_X^* = \underset{\boldsymbol{b}\in B}{argmax}\, r_{\boldsymbol{X}}(\boldsymbol{b}).$$

If the distribution of returnsis i.i.d. (independent identically distributed) over periods, then optimal portfoliowhich maximizes the expectationofthe logarithmic rate of returnis fixed in each period. We thus called this approach to asset allocationthe log-optimal strategy for convenience.

## 2 Information'sbenefit

Some market information may help to predict the future profits and also affect the asset allocationdirectly. We will investigate the effect of informationon the log-optimal strategy in this section. When new information comes in each period, thei.i.d. assumption of the distribution will be broken up, but it also brings benefit: it will increase the expectation ofthe optimal logarithmic rate of return in general.

Denotethe information by$Y$and the conditional distribution of $\boldsymbol{X}$given $Y = y$by $F(\boldsymbol{X}|Y = y)$at certain time point.Let $\boldsymbol{b}^{*T}_{\boldsymbol{X}|Y}$ to be the optimal portfoliosuch that

$$\boldsymbol{b}^{*T}_{\boldsymbol{X}|Y} = \underset{\boldsymbol{b}\in B}{argmax}\, r_{\boldsymbol{X}|Y}(\boldsymbol{b}) = \underset{\boldsymbol{b}\in B}{argmax} \int log(\boldsymbol{b}^T\boldsymbol{x})dF(\boldsymbol{x}|Y = y).$$

The increment of the expectation of the growth rate is defined as

$$\Delta V_Y = r_{\boldsymbol{X}|Y}\left(\boldsymbol{b}^{*T}_{\boldsymbol{X}|Y}\boldsymbol{x}\right) - r_{\boldsymbol{X}|Y}(\boldsymbol{b}^{*T}_{\boldsymbol{X}}\boldsymbol{x}).$$

According to the definition of $\boldsymbol{b}^{*T}_{\boldsymbol{X}|Y}$, we know$\Delta V_Y \geq 0$, which explains that theoretically the information Y will not decrease theexpectation of theoptimal logarithmic rate of return. The result is summarized in the following theorem.

**Theorem 2.1**$\Delta V_Y$ has an upper bound.

We prove two lemmas first.

**Lemma2.1.1**$E(log\,(\varphi(X))) \leq log\left(E\big(\varphi(X)\big)\right)$,$\forall$ r. v.$X \geq 0, \varphi(X) > 0$.

**Proof:**Sincethe logarithm function satisfies

$log(\lambda x_1 + (1-\lambda)x_2) \geq \lambda log(x_1) + (1-\lambda)\, log(x_2)\, , \lambda \in [0,1]$,

it is a concave function.

Let$\lambda = \frac{x_2 - x}{x_2 - x_1}$, when $x \in [x_1, x_2]$

We have $log(x) \geq \frac{x_2 - x}{x_2 - x_1} log(x_1) + \frac{x - x_1}{x_2 - x_1} log(x_2)$

The inequality equivalences

$$\frac{1}{x - x_1}[log(x) - log(x_1)] \geq \frac{1}{x_2 - x_1}[log(x_2) - log(x_1)].$$

Let $x \to x_1$, we have

$(x_2 - x_1)log'(x_1) \ \geq [log(x_2) - log(x_1)].$

Let

$x_0 = \sum_{i=1}^{m} \lambda_i \, x_i$, when$\sum_{i=1}^{m} \lambda_i = 1$， $\lambda_i > 0$.

For each$i$, we have

$$\lambda_i(x_i - x_0)log'(x_0) \geq \lambda_i[log(x_i) - log(x_0)].$$

Thus

$$\sum_{i=1}^{m} \lambda_i(x_i - x_0)log'(x_0) \ \geq \sum_{i=1}^{m} \lambda_i[log(x_i) - log(x_0)],$$

that is

$$log\left(\sum_{i=1}^{m} \lambda_i \, x_i\right) \geq \sum_{i=1}^{m} \lambda_i \, log(x_i).$$

Since

$$E\big(log\big(\varphi(X)\big)\big) = \int log\big(\varphi(x)\big) \, dF(x)$$

$$= lim_{n\to\infty} \sum_{k=1}^{\infty} log\big(\varphi(x_k)\big)(F\left(\frac{k}{n}\right) - F(\frac{k-1}{n})),$$

$$\log\Big(\mathrm{E}\big(\varphi(\mathrm{X})\big)\Big) = \log\left(\int \varphi(x) \, \mathrm{dF}(x)\right)$$

$$= log \ (lim_{n\to\infty} \sum_{k=1}^{\infty} \varphi(x_k) \left(F\left(\frac{k}{n}\right) - F\left(\frac{k-1}{n}\right)\right),$$

and the logarithm function is continuous. Thus, let

$$\lambda_k = F\left(\frac{k}{n}\right) - F(\frac{k-1}{n}), \ \ m \to \infty,$$

weget

$$\int log\big(\varphi(x)\big) \, dF(x) \leq log \ (\int \varphi(x) \, dF(x)),$$

that is

$$E\big(lo \ g\big(\varphi(X)\big)\big) \leq lo \ g \left(E\big(\varphi(X)\big)\right).$$

**Lemma 2.1.2** If$b^*$ is the optimalportfolioand $E \frac{b^T X}{b^{*T} X}$exists, we have $E \frac{b^T X}{b^{*T} X} \leq 1$ ,for any other portfolio$b$.

**Proof:** Let

$$W(b_\lambda, F) = \int \log(b_\lambda^{T} \boldsymbol{x})\, dF(\boldsymbol{x}),\ b_\lambda = \lambda b + (1-\lambda) b^*$$

where $b$ is other portfolio. When $\lambda$=0, we have $b_0 = b^*$. According to the definition we have the greatest value

$$W(b_0, F) = W(b^*, F) = \max_{\boldsymbol{b} \in A} \int log(\boldsymbol{b}^{\boldsymbol{T}} \boldsymbol{x})\, dF(\boldsymbol{x}).$$

That is $W(b_0, F) \geq W(b_k, F), k \in [0,1]$. By the definition of derivative, we know when $\lambda \to 0_+$, there is $\frac{dW(b_\lambda, F)}{d\lambda} \leq 0$

That is to say,

$$\begin{aligned}
\lim_{\lambda \to 0_+} \frac{dW(b_\lambda, F)}{d\lambda} &= \lim_{\lambda \to 0_+} \frac{1}{\lambda}[W(b_\lambda, F) - W(b_0, F)] \\
&= \lim_{\lambda \to 0_+} \frac{1}{\lambda}[E(log(\lambda b^T \boldsymbol{X} + (1-\lambda) b^{*T} \boldsymbol{X})) - E(log(b^{*T} \boldsymbol{X}))] \\
&= E\left(\lim_{\lambda \to 0_+} \frac{1}{\lambda} log\left(\lambda \frac{b^T \boldsymbol{X}}{b^{*T} \boldsymbol{X}} + 1 - \lambda\right)\right) \qquad (*) \\
&= E\left(\lim_{\lambda \to 0_+} \frac{1}{\lambda} log\left(1 + \lambda(\frac{b^T \boldsymbol{X}}{b^{*T} \boldsymbol{X}} - 1)\right)\right) \\
&= E\left(\frac{b^T \boldsymbol{X}}{b^{*T} \boldsymbol{X}} - 1\right) (**) \\
&\leq 0.
\end{aligned}$$

The equality($*$) can be referred to dominated convergence theorem in [5] and the equality($**$) is due to the L'Hospital's rule

$$\lim_{x \to 0} log(1 + c * x)/x = c,$$

where $c = (\frac{b^T \boldsymbol{X}}{b^{*T} \boldsymbol{X}} - 1)$.

Thus $E \frac{b^T \boldsymbol{X}}{b^{*T} \boldsymbol{X}} \leq 1$.

**The proof of Theorem 2.1:**

$$\Delta V_Y = r_{\boldsymbol{X}|Y}\left(\boldsymbol{b}_{\boldsymbol{X}|Y}^{*T} \boldsymbol{x}\right) - r_{\boldsymbol{X}|Y}(\boldsymbol{b}_{\boldsymbol{X}}^{*T} \boldsymbol{x})$$

$$= \int log(\boldsymbol{b}_{\boldsymbol{X}|Y}^{*T} \boldsymbol{x}) dF(\boldsymbol{x}|Y=y) - \int log(\boldsymbol{b}_{\boldsymbol{X}}^{*T} \boldsymbol{x}) dF(\boldsymbol{x}|Y=y)$$

$$= \int log \frac{\boldsymbol{b}_{X|Y}^{*T}\boldsymbol{x}}{\boldsymbol{b}_{X}^{*T}\boldsymbol{x}} dF(\boldsymbol{x}|Y=y)$$

$$= \int log \left(\frac{\boldsymbol{b}_{X|Y}^{*T}\boldsymbol{x}}{\boldsymbol{b}_{X}^{*T}\boldsymbol{x}} \frac{f(\boldsymbol{x})}{f_{X|Y=y}(\boldsymbol{x})}\right) dF(\boldsymbol{x}|Y=y) + \int log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} dF(\boldsymbol{x}|Y=y)$$

$$= \int log \left(\frac{\boldsymbol{b}_{X|Y}^{*T}\boldsymbol{x}}{\boldsymbol{b}_{X}^{*T}\boldsymbol{x}} \frac{f(\boldsymbol{x})}{f_{X|Y=y}(\boldsymbol{x})}\right) dF(\boldsymbol{x}|Y=y) + \int f_{X|Y=y}(\boldsymbol{x}) log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} d\boldsymbol{x}$$

$$\leq log \int \frac{\boldsymbol{b}_{X|Y}^{*T}\boldsymbol{x}}{\boldsymbol{b}_{X}^{*T}\boldsymbol{x}} \frac{f(\boldsymbol{x})}{f_{X|Y=y}(\boldsymbol{x})} dF(\boldsymbol{x}|Y=y) + \int f_{X|Y=y}(\boldsymbol{x}) log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} d\boldsymbol{x}$$

$$(lemma\ 2.1.1)$$

$$= log \int \mathrm{f}_{X|\mathrm{Y=y}}(\boldsymbol{x}) \frac{\boldsymbol{b}_{X|Y}^{*T}\boldsymbol{x}}{\boldsymbol{b}_{X}^{*T}\boldsymbol{x}} \frac{\mathrm{f}(\boldsymbol{x})}{\mathrm{f}_{X|\mathrm{Y=y}}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} + \int \mathrm{f}_{X|\mathrm{Y=y}}(\boldsymbol{x}) \mathrm{log} \frac{\mathrm{f}_{X|\mathrm{Y=y}}(\boldsymbol{x})}{\mathrm{f}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}$$

$$= log \int \frac{\boldsymbol{b}_{X|Y}^{*T}\boldsymbol{x}}{\boldsymbol{b}_{X}^{*T}\boldsymbol{x}} dF(\boldsymbol{x}) + \int f_{X|Y=y}(\boldsymbol{x}) log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} d\boldsymbol{x}$$

$$\leq log\, 1 + \int f_{X|Y=y}(\boldsymbol{x}) log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} d\boldsymbol{x} (lemma\ 2.1.2)$$

$$= \int f_{X|Y=y}(\boldsymbol{x}) log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} d\boldsymbol{x}.$$

Furthermore, we define $\Delta V = E(\Delta V_Y)$, the expectation of the increment $\Delta V_Y$ with respect to $Y$, then $\Delta V$also has an upper bound. Denote by$G(H)$ the cumulative distribution function of $Y$ $((X,Y))$, and $g(h)$the density function of $Y((X,Y))$, we verify that

$$\Delta V = \int \Delta V_{Y=y} dG(y)$$

$$\leq \int\int f_{X|Y=y}(\boldsymbol{x}) log \frac{f_{X|Y=y}(\boldsymbol{x})}{f(\boldsymbol{x})} d\boldsymbol{x} dG(y)$$

$$= \int\int f_{X|Y=y}(\boldsymbol{x}) \cdot g(y) \cdot log \frac{f_{X|Y=y}(\boldsymbol{x}) g(y)}{f(\boldsymbol{x}) g(y)} d\boldsymbol{x}\, dy$$

$$= \int\int h(\boldsymbol{x}, y)\, log \frac{h(\boldsymbol{x}, y)}{f(\boldsymbol{x}) g(y)} dx\, dy.$$

$\Delta V$ hasan upper bound $\int\int h(\boldsymbol{x}, y)\, log \frac{h(x,y)}{f(x)g(y)} d\boldsymbol{x}\, dy$ ,that is the mutual information of $\boldsymbol{X}$ and $Y$ [1].When $\boldsymbol{X}$ and $Y$are independent, $\Delta V = 0$ means no increment because the information $Y$ does not affect$\boldsymbol{X}$ in any way; When $\boldsymbol{X}$is completely determined by $Y$, this upper bound is exactly the entropy of the information $Y$. Further exploration may be interesting but is out of the scope of this paper.

## 3 Greed characteristic and optimal property

When thesequence of the returns $\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots$arei.i.d., the best strategy maintains a certain fixed portfolio. Owing to the information of investment which we discussed in section 2, the i.i.d.assumption becomes unrealistic. Fortunately, this assumption is not necessary, the log-optimal could vary the optimal portfolios over the periods but the strategy is still superior to other strategies asymptotically in view of gross wealth.Recall that the logarithmic return can be presented as $log\left(\prod_i \boldsymbol{b_i^T X_i}\right) = \sum_i log\ (\boldsymbol{b_i^T X_i})$ . To optimize $log\left(\prod_i \boldsymbol{b_i^T X_i}\right)$ is therefore equivalent to optimize$log\ (\boldsymbol{b_i^T X_i})$in the$i^{th}$ period. This means that the local optimaldynamic portfolios together can make a global optimal strategy, in other words, the log-optimal strategy hasthe greed characteristic.

Denote the gross wealth at the end of the $n^{th}$period with an sequence of portfolios$\{\boldsymbol{b_i}\}$by$S_n = S_0 \prod_{i=1}^{n} \boldsymbol{b_i^T X_i}$and thegross wealth using the log-optimal strategy$S_n^* = S_0 \prod_{i=1}^{n} \boldsymbol{b_i^{*T} X_i}$. Nexttheorem indicates the optimal property of thelog-optimal strategy.

**Theorem3**: $S_n^*$is asymptotically superior to$S_n$.

**Proof:**According to lemma 2.1.2, we have $E\frac{S_n}{S_n^*} \le 1$, and

$$Pr(S_n > n^2 \cdot S_n^*) = Pr\left(\frac{S_n}{S_n^*} > n^2\right)$$

$$= \int_{n^2}^{+\infty} dF(\frac{S_n}{S_n^*})$$

$$\le \frac{1}{n^2}\int_{n^2}^{+\infty} \frac{S_n}{S_n^*} dF(\frac{S_n}{S_n^*})$$

$$\le \frac{1}{n^2}\int_{0}^{+\infty} \frac{S_n}{S_n^*} dF(\frac{S_n}{S_n^*})$$

$$\le \frac{1}{n^2} E\frac{S_n}{S_n^*} \le \frac{1}{n^2},$$

that is

$$Pr\left(\frac{1}{n} log\frac{S_n}{S_n^*} > \frac{1}{n} log\ n^2\right) \le \frac{1}{n^2}, \quad \sum_{n=1}^{\infty} \Pr\left(\frac{1}{n} log\frac{S_n}{S_n^*} > \frac{2logn}{n}\right) \le \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

and

$$Pr\left(\overline{\lim_{n\to\infty}}\left\{\frac{1}{n}log\frac{S_n}{S_n^*} > \frac{2\,log\ n}{n}\right\}\right) = \lim_{k\to\infty} Pr\left(\bigcup_{n=k}^{\infty}\left\{\frac{1}{n}log\frac{S_n}{S_n^*} > \frac{2\,log\ n}{n}\right\}\right)$$

$$\leq \lim_{k\to\infty}\sum_{n=k} Pr\left(\left\{\frac{1}{n}log\frac{S_n}{S_n^*} > \frac{2\,log\ n}{n}\right\}\right)$$

$$= 0.$$

This implies,$\exists N > 0$, for $\forall n > N$, we have

$$\frac{1}{n}log\frac{S_n}{S_n^*} \leq \frac{2\,log\,n}{n}.$$

Thus

$$\overline{\lim}_{n\to\infty}\frac{1}{n}log\frac{S_n}{S_n^*} \leq 0, \text{ with probability 1.}$$

Hence, we have the conclusion that$S_n^*$is asymptotically superior to $S_n$[1].

## 4 From theory to practice: thetrueprobability distribution is unknown

We assume $F(\boldsymbol{X})$is knownin section 2. But in practice, $F(\boldsymbol{X})$is unknown and need to be estimated from the history with certain assumptions. We shall be alert for the deviation of the estimation and these assumptions, especially for possible "butterfly effects". In this section, wedon't estimate$F(\boldsymbol{X})$directly, but introduce a new object function called the allocation utility first. This function is a quadratic approximation to the expectation of the logarithmic rate of return, and it is simple enough to involving only the expectation and covariance of $\boldsymbol{X}$.Indeed, it can be considered as the balance between the logarithmic returnand its squared coefficient ofvariation. We optimize the allocation utility with the estimations of the expectation and covariance and find the optimal portfolio, but not the expectation of the logarithmic rate of return with the empirical cumulative distribution function of $\boldsymbol{X}$. The latter is usually so complicate specially for the high dimension of $\boldsymbol{X}$. Our strategy is computationallyeffective, and what's more, it is robust in the sense that the upper bound of the allocation utility deviation can be controlled by the $l_1$-norm of portfolioand$l_\infty$-norm of the bias of variance matrix estimator.We will concern the constraint on portfolio because of limited knowledge on the estimationof variance matrix.

**Allocation utility**

Suppose $\boldsymbol{X}$has itsexpectation $\boldsymbol{\mu}$and the covariance matrix$\boldsymbol{\Sigma}$.Then$E(\boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{b}^T\boldsymbol{\mu}$, $Var(\boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}$.

The accurate logarithm optimal strategy has $\boldsymbol{b} = \underset{\boldsymbol{b}\in B}{argmax} E(log(\boldsymbol{b}^T\boldsymbol{X}))$ . However, the optimization problem depends on the distribution function of $\boldsymbol{X}$and the accurate optimization needs complicate calculation. So we adopt Taylor expansion toapproximate$E(log(\boldsymbol{b}^T\boldsymbol{X}))$:

$$
\begin{aligned}
&E(log(\boldsymbol{b}^T\boldsymbol{X})) \\
&\approx E\left( log\, E(\boldsymbol{b}^T\boldsymbol{X}) + \frac{\boldsymbol{b}^T\boldsymbol{X} - E(\boldsymbol{b}^T\boldsymbol{X})}{E(\boldsymbol{b}^T\boldsymbol{X})} - \frac{\left(\boldsymbol{b}^T\boldsymbol{X} - E(\boldsymbol{b}^T\boldsymbol{X})\right)^2}{2\left(E(\boldsymbol{b}^T\boldsymbol{X})\right)^2} \right) \\
&= E(log(\boldsymbol{b}^T\boldsymbol{\mu})) + \frac{E(\boldsymbol{b}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{\mu})}{E(\boldsymbol{b}^T\boldsymbol{X})} - \frac{E((\boldsymbol{b}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{\mu})^2)}{2\left(E(\boldsymbol{b}^T\boldsymbol{X})\right)^2} \\
&= log(\boldsymbol{b}^T\boldsymbol{\mu}) + 0 - \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T]\boldsymbol{b} \\
&= log(\boldsymbol{b}^T\boldsymbol{\mu}) - \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}.
\end{aligned}
$$

Define the Allocation utilityfunction as

$$M(\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\Sigma}) = log(\boldsymbol{b}^T\boldsymbol{\mu}) - \frac{1}{2(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}.$$

So the log-optimal strategy can be approximately by the optimization of the Allocation utility: $max\, M(\boldsymbol{b},\boldsymbol{\mu},\boldsymbol{\Sigma}),\ s.t.\ \ \boldsymbol{b} \in B$

**Optimal portfolio$\boldsymbol{b}^{opt}$**

If the region $B$is simple, we can solve $\boldsymbol{b}$directly.For example, we take$B = \{\boldsymbol{b}|\boldsymbol{b}^T\boldsymbol{e} = 1，\ \boldsymbol{b}^T\ \boldsymbol{\mu} \geq c\}$ and analyze the optimal requirements. We can easily see that the first constraint is natural and the second constraint accounts for the requirements of minimum rate of return. Because there are both equality constraint and inequality constraint, we apply Karush-Kuhn-Tucker condition to solve this optimization problem.

Let

$$\boldsymbol{F}(\boldsymbol{b},\alpha,\beta) = -\boldsymbol{log}(\boldsymbol{b}^T\boldsymbol{\mu}) + \frac{\mathbf{1}}{\mathbf{2}(\boldsymbol{b}^T\boldsymbol{\mu})^2}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b} + \alpha(\boldsymbol{b}^T\boldsymbol{e} - \mathbf{1}) + \beta(c - \boldsymbol{b}^T\ \boldsymbol{\mu}),$$

where$\beta \geq \mathbf{0}$.

Take $\partial F(\boldsymbol{b},\alpha,\beta)/\partial\boldsymbol{b} = \mathbf{0}$ , and substitute $\beta(c - \boldsymbol{b}^T\ \boldsymbol{\mu}) = 0，\ \beta \geq 0，\ \boldsymbol{b}^T\boldsymbol{e} - 1 = 0$ into simultaneous equalities and inequalities，we have

$$\begin{cases} -\frac{\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{e} - \beta\boldsymbol{\mu} = \mathbf{0} & (4.1) \\ \beta(c - \boldsymbol{b}^T\ \boldsymbol{\mu}) = 0 & (4.2) \\ \beta \geq 0 & (4.3) \\ \boldsymbol{b}^T\boldsymbol{e} - 1 = 0 & (4.4) \end{cases}$$

Next we shallconsider how to solve the above equation systems.

Multiply $\boldsymbol{b}^T$ in both sides of equality (4.1), we have

$$\mathbf{0} = \boldsymbol{b}^T\left[-\frac{\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{e} - \beta\boldsymbol{\mu}\right]$$

$$= -\frac{\boldsymbol{b}^T\boldsymbol{\mu}}{\boldsymbol{b}^T\boldsymbol{\mu}} - \frac{\boldsymbol{b}^T\boldsymbol{\mu}\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^3} + \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha\boldsymbol{b}^T\boldsymbol{e} - \beta\boldsymbol{b}^T\boldsymbol{\mu}$$

$$= -\mathbf{1} - \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}{(\boldsymbol{b}^T\boldsymbol{\mu})^2} + \alpha \cdot \mathbf{1} - \beta\boldsymbol{b}^T\boldsymbol{\mu}$$

$$= -\mathbf{1} + \alpha - \beta \boldsymbol{b}^{\boldsymbol{T}} \boldsymbol{\mu}$$

(1) If $\beta = 0$, we are able to find $0 = -1 + \alpha - 0$, which refers to $\alpha = 1$.

Then we substitute $\alpha = 1, \beta = 0$ into equality (4.1), we have

$$-\frac{\boldsymbol{\mu}}{\boldsymbol{b}^T \boldsymbol{\mu}} - \frac{\boldsymbol{\mu} \boldsymbol{b}^T \boldsymbol{\Sigma} \boldsymbol{b}}{(\boldsymbol{b}^T \boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma} \boldsymbol{b}}{(\boldsymbol{b}^T \boldsymbol{\mu})^2} + \boldsymbol{e} = \mathbf{0} \qquad (4.5)$$

We can get the value of optimal portfolio $\widehat{\boldsymbol{b}}^{opt}$ by solving equality(4.5).

(2) If $\beta > 0$, According to equality (4.2), we can see $\boldsymbol{b}^{\boldsymbol{T}}\ \boldsymbol{\mu} = c$.

So we have $0 = -1 + \alpha - \beta c$, that is $\alpha = 1 + \beta c$.

Thus we can get following equation system:

$$\begin{cases} -\frac{\boldsymbol{\mu}}{\boldsymbol{b}^{\boldsymbol{T}} \boldsymbol{\mu}} - \frac{\boldsymbol{\mu} \boldsymbol{b}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{b}}{(\boldsymbol{b}^{\boldsymbol{T}} \boldsymbol{\mu})^3} + \frac{\boldsymbol{\Sigma} \boldsymbol{b}}{(\boldsymbol{b}^{\boldsymbol{T}} \boldsymbol{\mu})^2} + \alpha \boldsymbol{e} - \beta \boldsymbol{\mu} = \mathbf{0} \\ \boldsymbol{b}^{\boldsymbol{T}}\ \boldsymbol{\mu} = c \\ \alpha = \mathbf{1} + \beta \boldsymbol{c} \end{cases} \qquad (4.6)$$

After solving equation system(4.6), we can get the value of optimal portfolio $\widehat{\boldsymbol{b}}^{opt}$.

From equality(4.5)and equation system(4.6), we can find that the solution of $\widehat{\boldsymbol{b}}^{opt}$ depends on the value of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ If the estimation of $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ deviates from the true value, it may lead to deviation of estimationof optimal return. So we need to control our estimation process in order to reduce deviation. Next, we will find whether the deviation of estimation could affect the utility function seriously.

**Analysis of Robustness**

Suppose that $\widehat{\boldsymbol{b}}^{\boldsymbol{opt}}$ is the optimal portfolio estimator by replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with theirestimators: $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ in the optimization.Theestimation error of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ may have a serious influence on estimated utility value of $\widehat{\boldsymbol{b}}^{opt}$ accuracy.we need to know the optimal allocation vector $\widehat{\boldsymbol{b}}^{opt}$ utility function accurately. Thus we need toinvestigatewhether the bias of the optimal utility function following theestimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be controlled and how.

We first need to make an assumption for the model.

(A1) Suppose $\boldsymbol{E}(\widehat{\boldsymbol{\mu}}) = \boldsymbol{\mu}$, which refers to $\widehat{\boldsymbol{\mu}}$ is $\boldsymbol{\mu}$ unbiased estimation.

From (A1), we can know $E\left(\widehat{\boldsymbol{b}}^{opt^T} \widehat{\boldsymbol{\mu}}\right) = \widehat{\boldsymbol{b}}^{opt^T} \boldsymbol{\mu}$. According to law of large number, we can know $\forall \varepsilon > 0$ when sample size $\mathrm{n} \to \infty$, we have $P\left(\left|\widehat{\boldsymbol{b}}^{opt^T} \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{b}}^{opt^T} \boldsymbol{\mu}\right| > \varepsilon\right) \to 0$. For $\forall \epsilon > 0$, $\exists n_0 \in$

$N_{+}$such that$\widehat{\boldsymbol{\mu}}$satisfies$\left|\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right|\leq\varepsilon$ when sample size is more than $n_0$.

In addition, we need the following lemma.

Lemma 4.1： $\forall p,q,q_1,q_2\in\mathbb{R}$, satisfy $q_1\leq q\leq q_2$,We have

$$|p-q|\leq max\{p-q_1,q_2-p\}$$

**Proof:** If $p\geq q$， $p-q_1\geq p-q\geq 0$， which refers to $p-q_1>|p-q|$.

If$p<q$， $q_2-p\geq q-\mathrm{p}>0$， which refers to $q_2-p>|p-q|$.

So we have $|p-q|\leq max\{|p-q_1|,|q_2-p|\}$

Next we will study the deviation between $M(\widehat{\boldsymbol{b}}^{opt},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}})$ and$M(\widehat{\boldsymbol{b}}^{opt},\boldsymbol{\mu},\boldsymbol{\Sigma})$.

$$\left|M(\widehat{\boldsymbol{b}}^{opt},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}})-M(\widehat{\boldsymbol{b}}^{opt},\boldsymbol{\mu},\boldsymbol{\Sigma})\right|$$

$$=\left|log(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}})-\frac{1}{2\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-log(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu})+\frac{1}{2\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$=\left|\left(log(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}})-log(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu})\right)+\left(\frac{1}{2\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{2\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}\right)\right|$$

$$\leq\left|log(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}})-log(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu})\right|+\left|\frac{1}{2\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{2\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$=\left|log\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}\right)\right|+\frac{1}{2}\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$\leq\left|log\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}\right)\right|+\frac{1}{2}\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$=\left|log\left(1+\frac{\varepsilon}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}\right)\right|+\frac{1}{2}\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|.$$

For the first element of inequality left side, due to $\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\geq c$， we have

$$\left|log\left(1+\frac{\varepsilon}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}\right)\right|\leq log\left(1+\frac{\varepsilon}{c}\right).$$

For the second element of inequality right side, according to lemma 4.1，

Since

$$\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt} < \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt} < \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt},$$

we have

$$\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}$$

$$\leq max\left\{\begin{matrix}\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}, \\ \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\end{matrix}\right\}.$$

Since $-\varepsilon \leq \widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu} \leq \varepsilon$,

on one hand,

$$\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}$$

$$= \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\left[\widehat{\boldsymbol{\Sigma}} - \left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}\right)^2\boldsymbol{\Sigma}\right]\widehat{\boldsymbol{b}}^{opt}$$

$$\leq \frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\widehat{\boldsymbol{\Sigma}} - \left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}\right)^2\boldsymbol{\Sigma}\right)\widehat{\boldsymbol{b}}^{opt}\right),$$

on the other hand，

$$\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt} - \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}$$

$$\leq \frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\left[\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon}\right)^2\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}\right]\widehat{\boldsymbol{b}}^{opt}$$

$$\leq \frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon}\right)^2\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}\right)\widehat{\boldsymbol{b}}^{opt}\right).$$

Hence

$$\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$\leq max\left\{\begin{array}{c}\frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\widehat{\boldsymbol{\Sigma}}-\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}+\varepsilon}\right)^2\boldsymbol{\Sigma}\right)\widehat{\boldsymbol{b}}^{opt}\right),\\ \frac{1}{c^2}\left(\widehat{\boldsymbol{b}}^{opt^T}\left(\left(\frac{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}-\varepsilon}\right)^2\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right)\widehat{\boldsymbol{b}}^{opt}\right)\end{array}\right\}$$

$$\approx\frac{1}{c^2}\left|\widehat{\boldsymbol{b}}^{opt^T}(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}})\widehat{\boldsymbol{b}}^{opt}\right|.$$

Let $\widehat{\boldsymbol{b}}^{opt}=\left(\hat{b}_1,\ldots\ldots,\hat{b}_n\right)^T$, the $i^{\text{th}}$ row and $j^{\text{th}}$ column element is $\sigma_{ij}$ of $\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}$.

Hence

$$\begin{aligned}\left|\widehat{\boldsymbol{b}}^{\boldsymbol{opt}^{\boldsymbol{T}}}(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}})\widehat{\boldsymbol{b}}^{\boldsymbol{opt}}\right|&=\left|\sum_{i=1}^{n}\hat{b}_i\left(\sum_{j=1}^{n}\hat{b}_j\sigma_{ij}\right)\right|\\&\leq\sum_{i=1}^{n}\left|\hat{b}_i\right|\left|\sum_{j=1}^{n}\hat{b}_j\sigma_{ij}\right|\\&\leq\sum_{i=1}^{n}\left|\hat{b}_i\right|\sum_{j=1}^{n}\left|\hat{b}_j\right|\left|\sigma_{ij}\right|\\&\leq\sum_{i=1}^{n}\left|\hat{b}_i\right|\left(\sum_{j=1}^{n}\left|\hat{b}_j\right|\sum_{i=1}^{n}\left|\sigma_{ij}\right|\right)\\&\leq\sum_{i=1}^{n}\left|\hat{b}_i\right|\max_i\left(\sum_{j=1}^{n}\left|\hat{b}_j\right|\sum_{j=1}^{n}\left|\sigma_{ij}\right|\right)\\&=\left(\sum_{i=1}^{n}\left|\hat{b}_i\right|\right)^2\max_i\sum_{j=1}^{n}\left|\sigma_{ij}\right|.\end{aligned}$$

So we have

$$\left|\mathrm{M}\left(\widehat{\boldsymbol{b}}^{opt},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}}\right)-\mathrm{M}\left(\widehat{\boldsymbol{b}}^{opt},\boldsymbol{\mu},\boldsymbol{\Sigma}\right)\right|$$

$$\leq\left|\log\left(1+\frac{\varepsilon}{\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}}\right)\right|+\frac{1}{2}\left|\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\mu}}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{b}}^{opt}-\frac{1}{\left(\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\mu}\right)^2}\widehat{\boldsymbol{b}}^{opt^T}\boldsymbol{\Sigma}\widehat{\boldsymbol{b}}^{opt}\right|$$

$$\leq \log\left(1+\frac{\varepsilon}{c}\right)+\frac{1}{2c^2}\left(\sum_{i=1}^{n}\left|\hat{b}_i\right|\right)^2 \max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|$$

$$\approx \frac{1}{2c^2}\left(\sum_{i=1}^{n}\left|\hat{b}_i\right|\right)^2 \max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|.$$

Thus if we choose a suitable constant$c_0 > 0$, and control $\sum_{i=1}^{n}\left|\hat{b}_i\right| \leq c_0$, we can sure that the estimation error is under control ($max_i \sum_{j=1}^{n}\left|\sigma_{ij}\right|$ is supposed to be small enough).

Therefore, toachieve a robust optimization effect and make the error of optimal utilitycontrolled by an upper bound, we add an extra $\sum_{i=1}^{n}|b_i| \leq c_0$ constraint in the optimizationprocedure.

# 5 Simulation studies in financial market

## 5.1 Simulation purpose, assumption and statistics introduction

**Simulation purpose**

In this section, we use NYSEdata to illustrate our strategy. We are going to develop a set of strategies which can search for the mode of stock price, select asset continuously, and achieve the growth of capital from historical stock prices.

**Simulation assumption**

We will simplify the influence of transaction cost. The simulation involves the portfolio balance and reselect on a daily basis whose operation corresponds to unknown transaction cost. So we adopt the linear transaction costsmodel [2] to simplify the unknown transaction cost involved in the simulation.

We also assume a market with a strong fluidity, which means transactions happen according to our needs.

**Data introduction**

The data comes from 36 stocks in NYSE(New York Stock Exchange), from 3rd June 1962 to 31st December 1984 and from 1st January 1985 to 31st December 2007.

## 5.2 Model

**Extracting information from markets with similar backgrounds**

We use the data of stock prices in this simulation, and we predict future stock price referencing similar price mode in the historical data. We denote$x(i,k)$as the price ratio of the $i^{th}$ stock on the $k^{th}$day. $X(k)$is the price ratios on the $k^{th}$ day.

$X(k:l) = \begin{bmatrix} x_{1k} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{pk} & \cdots & x_{pl} \end{bmatrix}$is the statistics of the price ratio of all stocks from day $k$ to day $l$, which can be regarded as $l-k+1$ day's market background environment before the $(k+1)^{th}$ day.

For trading day $t$, when we fix its thelength of market background, it is possible to find trading day with similar market background in history, whose data is valuable for the prediction of the price of the $(t+1)^{th}$ day.

**Define the similarity of market background**

$$Similar\left(X(k:\ l), X(k+m: l+m)\right) = corr^*\left(X(k: l), X(k+m: l+m)\right),$$

corr* is the Pearson correlation between the vectorization of the matrixes $X(k:\ l)$ and $X(k+m: l+m)$.

**Find the trading day with similar market background**

If we set the value of the length of the market background $n$ and the threshold value of market similarity$\rho$, we can define the index set of datescorresponding to the trading days with similar market background as:

$$C(k,n,\rho) = \left\{n < i < k | Similar\left(X(i-n:\ i-1), X(k-n: k-1)\right) > \rho\right\}$$

## 5.3 Allocate the optimization model

Before selecting asset on the $(k+1)^{th}$ trading day, we need to find the set of similar trading days. The empirical cumulative distribution function of the profit ratio on similar trading days serves as the estimation of the profit ratio on the $(k+1)^{th}$ trading day. The selection of the similar trading day is influenced by the length of the market background $n$ and the threshold value of market similarity$\rho$. The most suitable$(n,\rho)$cannot be acquired in advance, and the cross validation is not suitable for this problem due to the nature of time seriesin stocks data. In the meantime, using the whole data set as the training set may cause the problem of over-fitting.

To solve the problems above, and to make asset selection more robust, we consider several$(n,\rho)$groups as several "experts". We take all the experts'advice into consideration in the final decision. The method to select "experts" in this passage follows as below: select a $N$ as the upper bound of $n$, and consider all the integers between 2 and $N$ as the values of $n$. Select a $P$ as the upper bound of $\rho$, extract 10 values equidistantly as the values of $\rho$. The selecting $N$ and $P$ is more moderate than selecting$(n,\rho)$group when measuring the influence on the simulation result.

For a single expert$(n,\rho)$, the optimal allocation model of the$(k+1)^{th}$ day is:

$$\mathbf{b}_{k+1}(n,\rho) = \underset{\boldsymbol{b}\in A}{\operatorname{argmax}} \prod_{i\in C(k,n,\rho)} (\boldsymbol{b}\cdot\boldsymbol{X_i})$$

$A$is the region of possible portfolios. According to the robust results in section 4, we add the constraint of $L_1$ norm not great than 1 in order to control the estimation deviation's upper bound, which concords with the principle that forbids short sell.The self-adapted dynamic weight combines the experts' advice with each other through the formula below.

$$\boldsymbol{b}_{k+1} = \frac{\sum_{n,\rho} s(n,\rho)\boldsymbol{b}_{k+1}(n,\rho)}{\sum_{n,\rho} s(n,\rho)} (***),$$

where $s(n,\rho)$ is the total profit on the trading days in the set $C(k,n,\rho)$ with the allocation $\boldsymbol{b}_{k+1}(n,\rho)$.The formula$(***)$ can be referred to [3].

### 5.4 Influence of the transaction cost

Transaction cost is an important factor of assessing the strategy. We can explore the influence of transaction cost by considering simple transaction cost model in the simulation of the transaction process, but only in the real market can we calculate the accurate transaction cost. This passage adopts the linear transaction costs model. The $(\mathrm{k}+1)^{\mathrm{th}}$ day's transaction cost is:

$$\alpha \cdot \|\mathrm{b}_{\mathrm{k}+1} - \mathrm{b}_{\mathrm{k}}\|_1$$

When considering the profit, we need to minus the cost with the corresponding ratio.$\alpha$is an adjustment coefficient. If valued 0, it means that there is no transaction cost; the larger $\alpha$ is, the more transaction cost will be. This passage explores the result of the strategy with different$\alpha$.

### 5.5 Simulation results

#### Cumulative capital

We set$N = 10, \rho = 0, \alpha = 0$. After the implementation of the strategy for the NYSE data, the condition of the cumulative growth of capital is showed in Figure 1 and Figure 2.

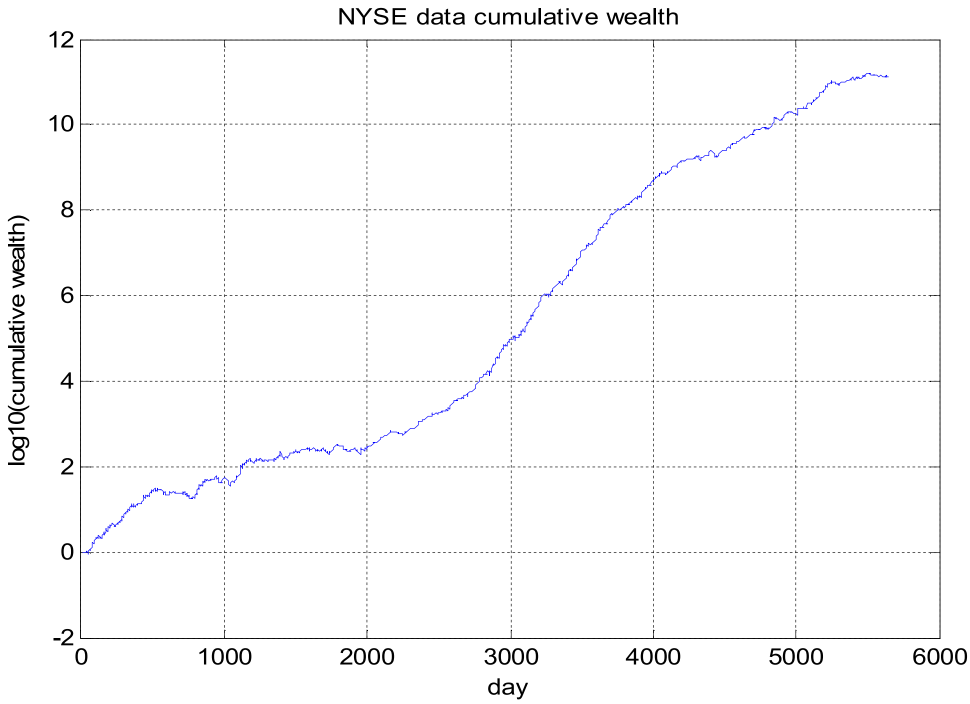


**Fig. 1.NYSE data accumulated asset（3rd June 1962 to 31st December 1984）**

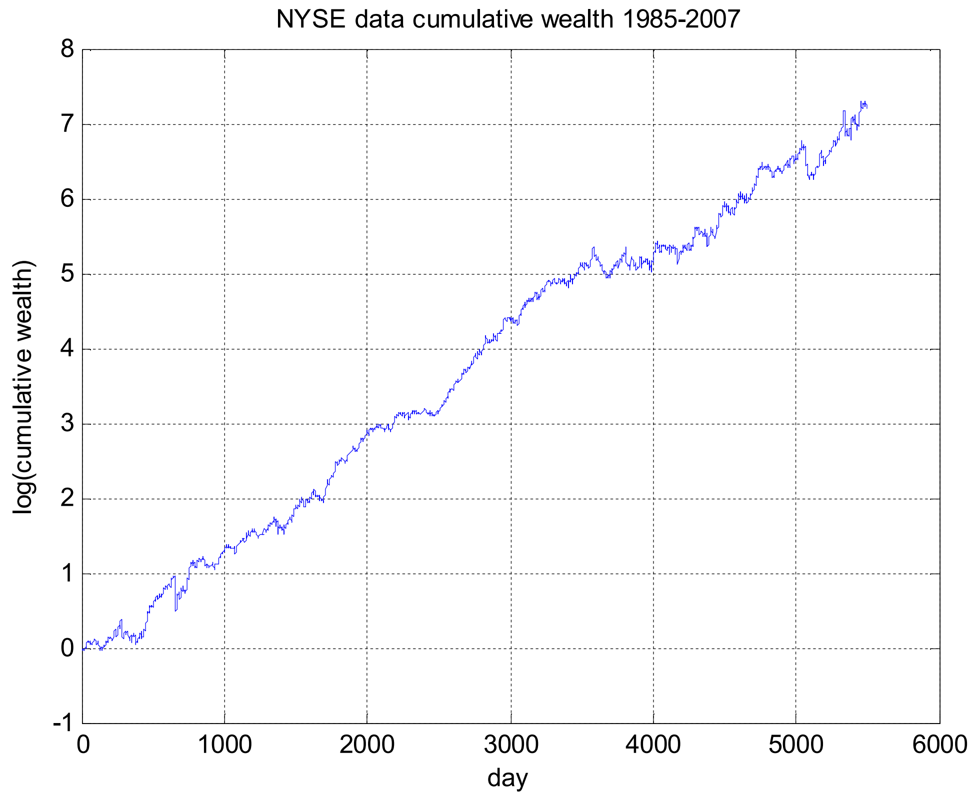


**Fig. 2.NYSE data accumulated asset（1st January 1985 to 31st December 2007）**

Figure 1 and Figure 2 indicate that if we don't consider transaction costs, our investment strategy makes asset increase robustly at each stage in 2 period of more than 5000 trading days. What's more, its order of magnitude has never decreased sharply. It shows robustness of our strategy. The reason why it has such a satisfactory performance is that the law of large number ensures the stability of the logarithmicrateof return. According to similarity of market background, the prediction accords to market principle. That's why the expectation of the gross wealth can increase.

**Transaction cost influence**

In practice, we need to analyze how transaction costs have influence on our cumulative wealth. From linear cost model in section 5.2, we consider deducting ratio of different $\alpha$ position capital after adjusting asset portfolio. We take the data from 3rd June 1962 to 31st December 1984as an example here. $\alpha$takes the degree of deduction of capital from 0 to 0.01. It corresponds to cumulative wealth inFigure 3.

As Figure 3 shows, we can find that the logarithm of cumulative wealth and $\alpha$ subject a linear relationship under the linear costs model. It's easy to see that when$\alpha = 0.0045$, the logarithm of wealth is 0,which means our asset in balance. So when $\alpha$ isless than 0.0045, cumulative wealth is can increase exponentially as time flies. In American stock market, transaction costs are usually less than 0.001. Thus the robust log-optimal strategy is significance in practice.

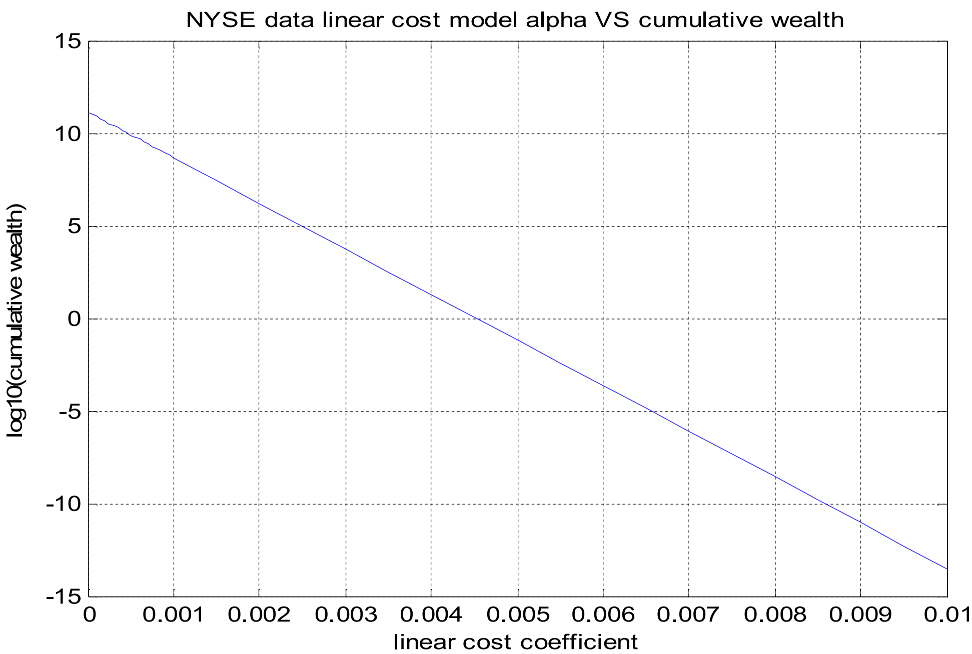


**Fig. 3. Transaction cost effect analysis**

**Parameter sensitivity analysis (date from 3rd June 1962 to 31st December 1984)**

We set largest length of market background as 10 in this paper. To analyze if parameter setting has a sensitive influence on result, we adopt integers range from 0 to 25 and get corresponding ultimately accumulated asset situation in figure 4.

Figure 4 shows that if the length of market window parameter is from 0 to 5, the larger the parameter, the better the result. If it is from 5 to 25, the result remains stable. In practice, we can choose optimal length of market background by the training set. Because of its stability, it can be utilized in the other fields of data mining apart from the stock market.We analyze the sensitivity of parameter $\rho$ as well. Let $n$=10 and $\rho$={-0.15，-0.05，0，0.05，0.1，0.15}, and get ultimately different accumulated asset as shown in Figure 5. if$\rho$=0, we can get the best effect. If $\rho$<0, similar trading days which we extract will have negative effect. So it's normal not to get an optimal result. If $\rho$ is so large that there is not enough trading days which meet the requirements,we cannot get an optimal result.

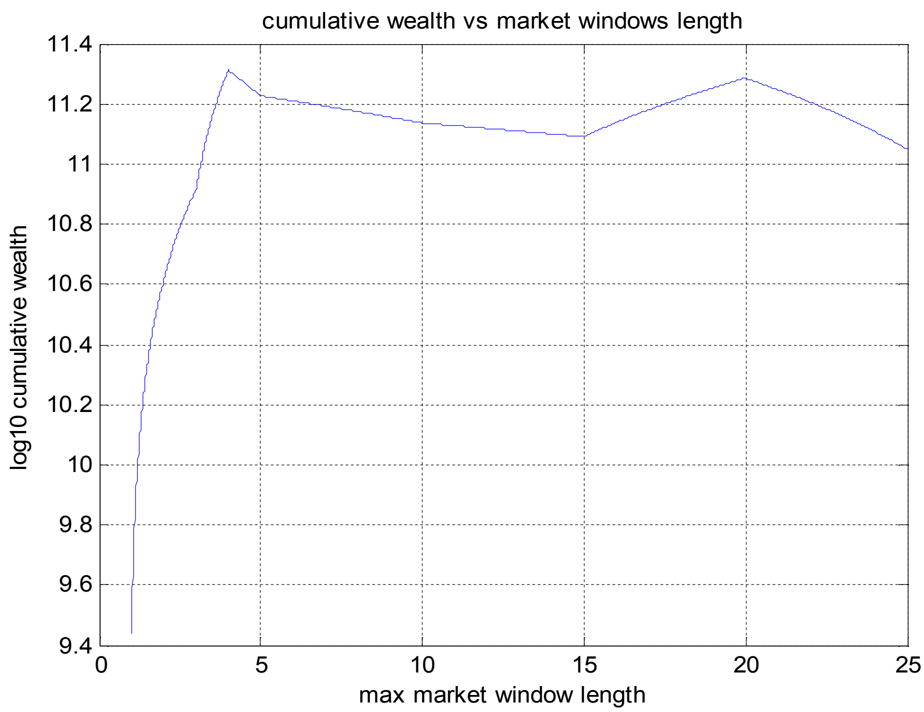


**Fig.4. Largest market window length sensitivity analysis**

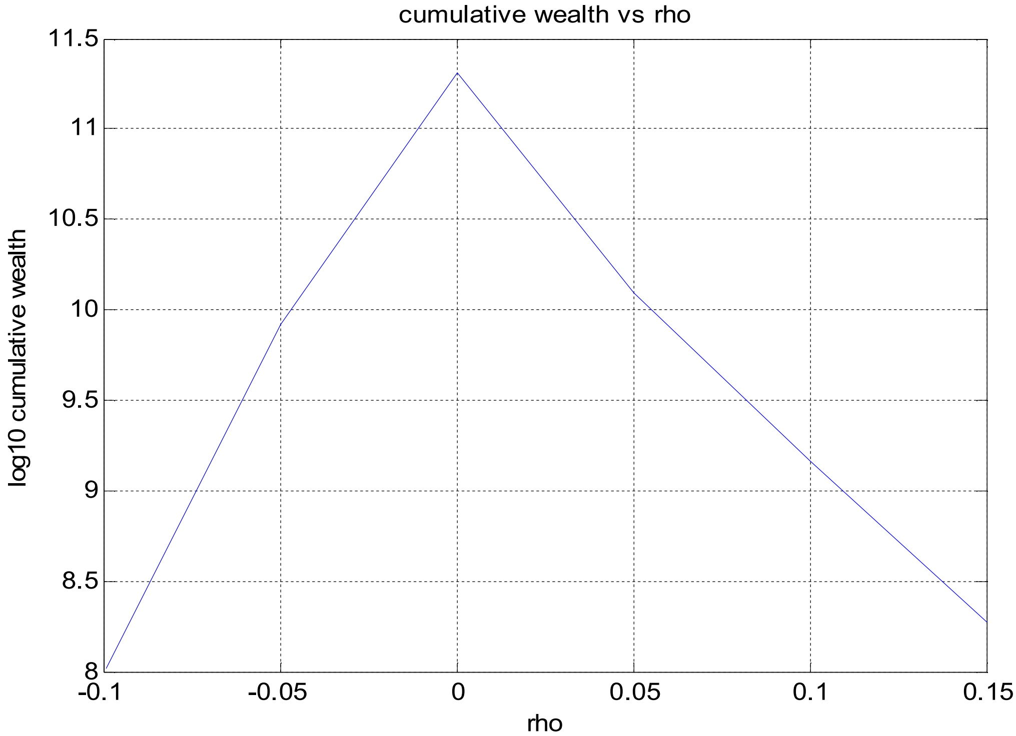


**Fig. 5. Correlated threshold sensitivity analysis**

## 6. Conclusion

We develops a new robust log-optimal strategy, many virtues meet in it:

1) an extremely explanation in theory: the balance between the logarithmic returnand its squared coefficient ofvariation;
2) computationaleffectiveness;
3) resistanceto"butterfly effect": the upper bound of the allocation utility deviation can be controlled by the $l_1$-norm of asset allocation vector and $l_\infty$-norm of the bias of variance matrix estimator;
4) highperformance in return: application on NYSE data.

This log-optimal strategy is for general cases, we still have many problems under consideration for further exploration. For instance, How to adjust our strategy to special asset class such as bonds, Insurance products,derivatives, foreign currency and etc.? How to select assets in our strategy? How to use information effectively? How to improve the accuracy of the estimations?

## References

[1] Cover T M, Thomas J A. Elements of information theory [M]. John Wiley & Sons, 2012.

[2] Blum A, Kalai A. Universal portfolios with and without transaction costs [J]. Machine Learning, 1999, 35(3): 193-205.

[3] Cover T M. Universal portfolios [J]. Mathematical finance, 1991, 1(1): 1-29.

[4] Singer Y. Switching portfolios [J]. International Journal of Neural Systems, 1997, 8(04): 445-455.

[5] R. Bell and T. M. Cover.Game-theoretic optimal portfolios. Manage. Sci.,34(6):724–733, 1988.

[6]"Asset Allocation Definition". Investopedia.Retrieved June 2011.