The growth optimal investment strategy is secure, too.

László Györfi, György Ottucsák, and Harro Walk

This paper is a revisit of discrete time, multi period, sequential investment strategies for financial markets showing that the log-optimal strategies are secure, too. Using exponential inequality of large deviation type, we bound the rate of convergence of the average growth rate to the optimum growth rate both for memoryless and for Markov market processes. A kind of security indicator of an investment strategy can be the market time achieving a target wealth. We show that the log-optimal principle is optimal in this respect.

1 Introduction

This paper gives some additional features of the investment strategies in financial stock markets inspired by the results of information theory, non-parametric statistics and machine learning. Investment strategies are allowed to use information collected from the past of the market and determine, at the beginning of a trading period, a portfolio, that is, a way to distribute their current capital among the available assets. The goal of the investor is to maximize his wealth in the long run without knowing the underlying distribution generating the stock prices. Under this assumption the asymptotic

László Györfi and György Ottucsák
Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Stoczek u.2, 1521 Budapest, Hungary e-mail: gyorfi@cs.bme.hu, oti@cs.bme.hu. This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

Harro Walk
Department of Mathematics, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart e-mail: walk@mathematik.uni-stuttgart.de.

rate of growth has a well-defined maximum which can be achieved in full knowledge of the underlying distribution generated by the stock prices.

Under memoryless assumption on the underlying process generating the asset prices, the log-optimal portfolio achieves the maximal asymptotic average growth rate, that is the expected value of the logarithm of the return for the best fix portfolio vector. Using exponential inequality of large deviation type, we bound the rate of convergence of the average growth rate to the optimum growth rate. Consider a security indicator of an investment strategy, which is the market time achieving a target wealth. We show that the log-optimal principle is optimal in this respect, too.

For generalized dynamic portfolio selection, when asset prices are generated by a stationary and ergodic process, there are universal consistent (empirical) methods that achieve the maximal possible growth rate. If the market process is first order Markov process, then we extend the rate of convergence of the average growth rate to the optimal one.

Consider a market consisting of d assets. The evolution of the market in time is represented by a sequence of price vectors $\mathbf{S}_1, \mathbf{S}_2, \ldots \in \mathbb{R}_+^d$, where

$$\mathbf{S}_n = (S_n^{(1)}, \ldots, S_n^{(d)})$$

such that the j-th component $S_n^{(j)}$ of $\mathbf{S}_n$ denotes the price of the j-th asset on the n-th trading period. In order to normalize, put $S_0^{(j)} = 1$.

We transform the sequence prices $\{\mathbf{S}_n\}$ into the sequence of return (relative price) vectors $\{\mathbf{X}_n\}$ as follows:

$$\mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(d)})$$

such that

$$X_n^{(j)} = \frac{S_n^{(j)}}{S_{n-1}^{(j)}}.$$

Thus, the j-th component $X_n^{(j)}$ of the return vector $\mathbf{X}_n$ denotes the amount obtained after investing a unit capital in the j-th asset on the n-th trading period.

2 Constantly rebalanced portfolio selection

The dynamic portfolio selection is a multi-period investment strategy, where at the beginning of each trading period we rearrange the wealth among the assets. A representative example of the dynamic portfolio selection is the constantly rebalanced portfolio (CRP), was introduced and studied by Kelly

[29], Latané [30], Breiman [11], Finkelstein and Whitley [19], and Barron and Cover [8].

The investor is allowed to diversify his capital at the beginning of each trading period according to a portfolio vector $\mathbf{b} = (b^{(1)}, \dots b^{(d)})$. The j-th component $b^{(j)}$ of $\mathbf{b}$ denotes the proportion of the investor's capital invested in asset j. Throughout the paper we assume that the portfolio vector $\mathbf{b}$ has nonnegative components with $\sum_{j=1}^{d} b^{(j)} = 1$. The fact that $\sum_{j=1}^{d} b^{(j)} = 1$ means that the investment strategy is self financing and consumption of capital is excluded. The non-negativity of the components of $\mathbf{b}$ means that short selling and buying stocks on margin are not permitted. The simplex of possible portfolio vectors is denoted by Δ_d.

Let S_0 denote the investor's initial capital. Then at the beginning of the first trading period $S_0 b^{(j)}$ is invested into asset j, and it results in return $S_0 b^{(j)} x_1^{(j)}$, therefore at the end of the first trading period the investor's wealth becomes

$$S_1 = S_0 \sum_{j=1}^{d} b^{(j)} X_1^{(j)} = S_0 \langle \mathbf{b}, \mathbf{X}_1 \rangle ,$$

where $\langle \cdot , \cdot \rangle$ denotes inner product. For the second trading period, S_1 is the new initial capital

$$S_2 = S_1 \cdot \langle \mathbf{b}, \mathbf{X}_2 \rangle = S_0 \cdot \langle \mathbf{b}, \mathbf{X}_1 \rangle \cdot \langle \mathbf{b}, \mathbf{X}_2 \rangle .$$

By induction, for the trading period n the initial capital is S_{n-1}, therefore

$$S_n = S_{n-1} \langle \mathbf{b}, \mathbf{X}_n \rangle = S_0 \prod_{i=1}^{n} \langle \mathbf{b}, \mathbf{X}_i \rangle .$$

The asymptotic average growth rate of this portfolio selection is

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \ln S_n &= \lim_{n\to\infty} \left(\frac{1}{n} \ln S_0 + \frac{1}{n} \sum_{i=1}^{n} \ln \langle \mathbf{b}, \mathbf{X}_i \rangle \right) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \ln \langle \mathbf{b}, \mathbf{X}_i \rangle , \end{aligned}$$

therefore without loss of generality one can assume in the sequel that the initial capital $S_0 = 1$.

If the market process $\{\mathbf{X}_i\}$ is memoryless, i.e., it is a sequence of independent and identically distributed (i.i.d.) random return vectors then we show that the best constantly rebalanced portfolio (BCRP) is the log-optimal portfolio:

$$\mathbf{b}^* := \arg\max_{\mathbf{b}\in\Delta_d} \mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\}.$$

This optimality means that if $S_n^* = S_n(\mathbf{b}^*)$ denotes the capital after day n achieved by a log-optimum portfolio strategy $\mathbf{b}^*$, then for any portfolio strategy $\mathbf{b}$ with finite $\mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_1\rangle\}$ and with capital $S_n = S_n(\mathbf{b})$ and for any memoryless market process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

$$\lim_{n\to\infty}\frac{1}{n}\ln S_n \le \lim_{n\to\infty}\frac{1}{n}\ln S_n^* \quad \text{almost surely (a.s.)} \tag{1}$$

and maximal asymptotic average growth rate is

$$\lim_{n\to\infty}\frac{1}{n}\ln S_n^* = W^* := \mathbb{E}\{\ln\langle\mathbf{b}^*, \mathbf{X}_1\rangle\} \quad \text{a.s.}$$

The proof of the optimality is a simple consequence of the strong law of large numbers. Introduce the notation

$$W(\mathbf{b}) = \mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_1\rangle\}.$$

Then

$$\begin{aligned}
\frac{1}{n}\ln S_n &= \frac{1}{n}\sum_{i=1}^{n}\ln\langle\mathbf{b}, \mathbf{X}_i\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_i\rangle\} + \frac{1}{n}\sum_{i=1}^{n}\left(\ln\langle\mathbf{b}, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_i\rangle\}\right) \\
&= W(\mathbf{b}) + \frac{1}{n}\sum_{i=1}^{n}\left(\ln\langle\mathbf{b}, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_i\rangle\}\right).
\end{aligned}$$

The Kolmogorov strong law of large numbers implies that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\ln\langle\mathbf{b}, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_i\rangle\}\right) \to 0 \quad \text{a.s.},$$

therefore

$$\lim_{n\to\infty}\frac{1}{n}\ln S_n = W(\mathbf{b}) = \mathbb{E}\{\ln\langle\mathbf{b}, \mathbf{X}_1\rangle\} \quad \text{a.s.}$$

Similarly,

$$\lim_{n\to\infty}\frac{1}{n}\ln S_n^* = W^* := W(\mathbf{b}^*) = \max_{\mathbf{b}} W(\mathbf{b}) \quad \text{a.s.}$$

There is an obvious question here: how secure a growth optimal portfolio strategy is? The strong law of large numbers has another interpretation. Put

$$R_n := \inf_{n\le m}\frac{1}{m}\ln S_m^*,$$

then e^{nR_n} is a lower exponential envelope for S_n^*, i.e.,

$$e^{nR_n} \le S_n^*.$$

Moreover,

$$R_n \uparrow W^* \quad \text{a.s.},$$

which means that for an arbitrary $R < W^*$, we have that

$$e^{nR} \le S_n^*$$

for all n after a random time N large enough.

In the sequel we bound N, i.e., derive a rate of convergence of the strong law of large numbers. Assume that there exist $0 < a_1 < 1 < a_2 < \infty$ such that

$$a_1 \le X^{(j)} \le a_2 \tag{2}$$

for all $j = 1, \dots, d$. For the New York Stock Exchange (NYSE) daily data, this condition is satisfied with $a_1 = 0.7$ and with $a_2 = 1.2$. $a_1 = 0.7$ means that in a day three times happened 10% decrease, while $a_2 = 1.2$ corresponds to two times increase with 10%. (Cf. Fernholz [18], Horváth and Urbán [27].) Figure 1 shows the histogram of Coke's daily logarithmic relative prices such that most of the days the relative prices are in the interval $[0.95, 1.05]$. Here are some statistical data:

$$\begin{aligned} \text{minimum} &= -0.2836 \\ \text{1st qu.} &= -0.0074 \\ \text{median} &= 0.0000 \\ \text{mean} &= 0.00053 \\ \text{3rd qu.} &= 0.0083 \\ \text{maximum} &= 0.1796. \end{aligned}$$

Theorem 1. *If the market process $\{\mathbf{X}_i\}$ is memoryless and the condition (2) is satisfied, then for an arbitrary $R < W^*$, we have that*

$$\mathbb{P}\left\{e^{nR} > S_n^*\right\} \le e^{-2n\frac{(W^*-R)^2}{(\ln a_2 - \ln a_1)^2}}.$$

Proof. We have that

$$\mathbb{P}\left\{e^{nR} > S_n^*\right\} = \mathbb{P}\left\{R > \frac{1}{n}\ln S_n^*\right\}$$

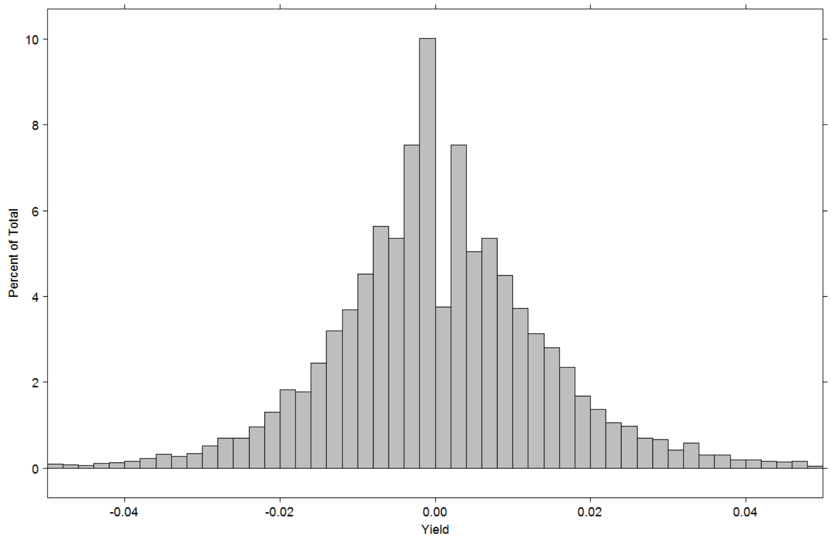

Fig. 1 The histogram of log-returns for Coke

$$= \mathbb{P}\left\{ R - W^* > \frac{1}{n}\sum_{i=1}^{n}\left(\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle\}\right) \right\}.$$

Apply the Hoeffding [26] inequality: Let $X_1,\ldots,X_n$ be independent random variables with $X_i \in [c, c+K]$ with probability one. Then, for all $\epsilon > 0$,

$$\mathbb{P}\left\{ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}\{X_i\}) < -\epsilon \right\} \le e^{-2n\frac{\epsilon^2}{K^2}}.$$

Because of the condition,

$$\ln a_1 \le \ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle \le \ln a_2,$$

therefore the theorem follows from the Hoeffding inequality for the correspondences

$$\epsilon = W^* - R$$

and

$$X_i = \ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle$$

and

$$K = \ln a_2 - \ln a_1.$$

■

Using Theorem 1, we can bound the probability that after n there is a time instant m such that $e^{mR} > S_m^*$:

Corollary 1. *If the market process $\{\mathbf{X}_i\}$ is memoryless and the condition (2) is satisfied, then for an arbitrary $R < W^*$, we have that*

$$\mathbb{P}\left\{\cup_{m=n}^{\infty}\{e^{mR} > S_m^*\}\right\} \le e^{-2n\frac{(W^*-R)^2}{(\ln a_2 - \ln a_1)^2}} \frac{e^{2\frac{(W^*-R)^2}{(\ln a_2-\ln a_1)^2}}}{e^{2\frac{(W^*-R)^2}{(\ln a_2-\ln a_1)^2}} - 1}. \tag{3}$$

Proof. From Theorem 1 we get that

$$\begin{aligned}\mathbb{P}\left\{\cup_{m=n}^{\infty}\{e^{mR} > S_m^*\}\right\} &\le \sum_{m=n}^{\infty} \mathbb{P}\left\{e^{mR} > S_m^*\right\} \\ &\le \sum_{m=n}^{\infty} e^{-2m\frac{(W^*-R)^2}{(\ln a_2-\ln a_1)^2}} \\ &= e^{-2n\frac{(W^*-R)^2}{(\ln a_2-\ln a_1)^2}} \frac{1}{1 - e^{-2\frac{(W^*-R)^2}{(\ln a_2-\ln a_1)^2}}}.\end{aligned}$$

■

Under the aspect of security we are interested in the asymptotic behavior of the relative amount of times j between 1 and n, for which S_j^* is below e^{jR} for $R(< W^*)$ near to W^*, say $R = R_n = W^* - \frac{m}{\sqrt{n}}\sigma$ for fixed $m > 0$ with $\sigma^2 = \mathbf{Var}(\ln\langle \mathbf{b}^*, \mathbf{X}_1\rangle)$ assumed to be positive and finite. For $0 \le x \le 1$ we have

$$\begin{aligned}&\mathbb{P}\left\{\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{S_j^* < e^{jR}\}} \le x\right\} \\ &= \mathbb{P}\left\{\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{\frac{1}{j}\sum_{i=1}^{j}(\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle\}) < R - W^*\}} \le x\right\} \\ &= \mathbb{P}\left\{\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{j}(\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle\}) + m\frac{j}{n} < 0\}} \le x\right\} \\ &\to \mathbb{P}\left\{\int_0^1 \mathbb{I}_{\{W(u)+mu \le 0\}} du \le x\right\}\end{aligned}$$

with standard Brownian motion W, by Donsker's functional central limit theorem (see Billingsley [9]) for the functional $f \to \int_0^1 \mathbb{I}_{\{f(u)+mu\le 0\}} du$.

By the generalized arc-sine law of Takács [37] the right hand side equals

$$F_m(x) := 2\int_0^x \left[\frac{\varphi(m\sqrt{1-u})}{\sqrt{1-u}} + m\Phi(m\sqrt{1-u})\right]\left[\frac{\varphi\left(-m\sqrt{u}\right)}{\sqrt{u}} - m\Phi\left(-m\sqrt{u}\right)\right] du$$

for $0 \le x \le 1$, where $F_m(1) = 1$, and φ and Φ are the standard normal density and distribution functions, respectively. We have a non-degenerate limit distribution. Here for $m \to \infty$ and also for the case $R = R'_n$ with $(W^* - R)\sqrt{n} \to \infty$, especially a constant $R'_n < W^*$, we have degeneration to the Dirac distribution concentrated at 0. The proof of these assertions can be as follows: For each $0 < \epsilon < 1/2$, on $[\epsilon, 1-\epsilon]$ the uniformly bounded integrand uniformly converges to 0 for $m \to \infty$, thus $F_m(1-\epsilon) - F(\epsilon) \to 0$. Further $F_m(0) = 0$ and $F_m(1) = 1$ for each m, and $F_m(x)$ is non-decreasing for each $0 \le x \le 1$. Thus, $F_m(x) \to 1$ for each $0 < x \le 1$. Finally one notices that $m < \sqrt{n}(W^* - R'_n) \to \infty$ $(n \to \infty)$ implies

$$\begin{aligned}
&\liminf_n \mathbb{P}\left\{\frac{1}{n}\sum_{j=1}^n \mathbb{I}_{\{\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^j(\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle\}) + \sqrt{n}(W^* - R'_n)\frac{j}{n} < 0\}} \le x\right\} \\
&\ge \lim_n \mathbb{P}\left\{\frac{1}{n}\sum_{j=1}^n \mathbb{I}_{\{\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^j(\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle \mathbf{b}^*, \mathbf{X}_i\rangle\}) + m\frac{j}{n} < 0\}} \le x\right\},
\end{aligned}$$

for each m. It should be mentioned that under the assumption (2) the latter of the assertions is also a consequence of Theorem 1 for $R = R'_n$.

In the literature there is a discussion on good and bad properties of log-optimal investment (see MacLean, Thorp and Ziemba [32], sections 30 and 39, with references). Beside

$$\limsup \frac{1}{n}\log(S_n/S_n^*) \le 0$$

almost surely (see (1) and (4) below, good long-run performance) one has

$$\mathbb{E}\{S_n/S_n^*\} \le 1$$

for all n (good short-term performance). Both properties were established by Algoet and Cover [3] in the much more general context of a stationary and ergodic process of daily returns $\mathbf{X}_n$ and conditionally log-optimal investment (here regarding past returns, but nothing more: myopic policy). Leaving the concept of a logarithmic utility function induced by the multiplicative structure of investment, Samuelson [34] in his critics pointed out that maximizing the expected return $\mathbb{E}\{\langle \mathbf{b}, \mathbf{X}_i\rangle\}$ instead of expected logarithmic return, with in this sense optimal portfolio choice $\mathbf{b}^{**}$ and corresponding wealth S_n^{**}, leads to $\mathbb{E}\{S_n^{**}\}/\mathbb{E}\{S_n^*\} \to \infty$, see also the comments of Markowitz [33]. But under the risk aspect of the deviation of a random variable from its expectation, use of logarithm is more advantageous. The log transform is a special case of the Box-Cox [10] transforms introduced in view of stabilization and widely used in science, e.g., in medical science. Nevertheless there is the question whether the risk aversion of log utility is big enough to save an investor with very high probability from large terminal losses for medium time horizon. Simulation studies discussed by MacLean, Thorp, Zhao and Ziemba in MacLean, Thorp and Ziemba [32], section 38, show that in a minority of scenarios such events occur. These effects depend on time horizon and distribution of the daily return, which allows a "proper use in the short and median run" provided one has a good knowledge of the distribution. Corollary 1 allows for small $\epsilon > 0$ to obtain a lower bound N for the time horizon having a probability $\geq 1 - \epsilon$ that after this time the investor's wealth is for ever at least the unit starting capital: on the right-hand side of (3) set $R = 0$ and then choose N as the lowest integer n such that the right-hand side is at most ϵ. Here as in the following, $W^* > 0$ is assumed.

Besides the growth rate of an investment strategy, one may consider the market time achieving a target wealth. We consider only strategies $\mathbf{b}$ with $\mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\} > 0$. Again, $S_n^* = S_n(\mathbf{b}^*)$ denotes the capital after day n applying log-optimum portfolio strategy $\mathbf{b}^*$, and $S_n = S_n(\mathbf{b})$ the capital using the portfolio strategy $\mathbf{b}$. For a target wealth $\bar{s}$, introduce the market times

$$\tau(\bar{s}) := \min\{m; S_m \geq \bar{s}\}$$

and similarly

$$\tau^*(\bar{s}) := \min\{m; S_m^* \geq \bar{s}\}.$$

There are some studies how to minimize the expected market time $\mathbb{E}\{\tau(\bar{s})\}$ for large $\bar{s}$ (Aucamp [5], [6], Breiman [11], Hayes [25], Kadaras and Platen [28]), where Ethier [16] established an asymptotic median log-optimality of the (mean) log-optimal investment strategy. Breiman [11] conjectured that, for large $\bar{s}$, the asymptotically best strategy is the growth optimal one such that we apply the growth optimal strategy until we reach a neighborhood of $\bar{s}$.

Using the representation

$$\{S_m \geq \bar{s}\} = \left\{ \sum_{i=1}^{m} \ln \langle \mathbf{b}, \mathbf{X}_i \rangle \geq \ln \bar{s} \right\}$$

the renewal theory for extended renewal processes, i.e., random walks with drift (see, e.g., Breiman [12] and Feller [17]), yields

$$\frac{\tau(\bar{s})}{\ln \bar{s}} \to \frac{1}{\mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\}}$$

a.s.,

$$\frac{\mathbb{E}\{\tau(\bar{s})\}}{\ln \bar{s}} \to \frac{1}{\mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\}},$$

especially

$$\frac{\tau^*(\bar{s})}{\ln \bar{s}} \to \frac{1}{W^*}$$

a.s.,

$$\frac{\mathbb{E}\{\tau^*(\bar{s})\}}{\ln \bar{s}} \to \frac{1}{W^*}$$

$(\bar{s} \to \infty)$. In this sense the growth optimal strategy has another optimality property.

The result of Breiman [11] on $\mathbb{E}\{\tau^*(\bar{s})\} - \mathbb{E}\{\tau(\bar{s})\}$ can be extended to

$$\begin{aligned}
&\frac{\ln \bar{s}}{\mathbb{E}\{\ln \langle \mathbf{b}^*, \mathbf{X}_1 \rangle\}} - \frac{\ln \bar{s}}{\mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\}} + \frac{\mathbb{E}\{((\ln \langle \mathbf{b}^*, \mathbf{X}_1 \rangle)_+)^2\}}{(\mathbb{E}\{\ln \langle \mathbf{b}^*, \mathbf{X}_1 \rangle\})^2} \\
&\geq \mathbb{E}\{\tau^*(\bar{s})\} - \mathbb{E}\{\tau(\bar{s})\} \\
&\geq \frac{\ln \bar{s}}{\mathbb{E}\{\ln \langle \mathbf{b}^*, \mathbf{X}_1 \rangle\}} - \frac{\ln \bar{s}}{\mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\}} - \frac{\mathbb{E}\{((\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle)_+)^2\}}{(\mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_1 \rangle\})^2}
\end{aligned}$$

by Lorden's [31] upper bound for excess result.

Next we bound the tail distribution of $\tau^*(\bar{s})$ in case of large $\bar{s} = e^{nR}$, where $R < W^*$. We get that

$$\mathbb{P}\{\tau^*(e^{nR}) > n\} = \mathbb{P}\left\{\cap_{m=1}^n \{S_m^* < e^{nR}\}\right\} \leq \mathbb{P}\left\{S_n^* < e^{nR}\right\},$$

therefore Theorem 1 implies that

$$\mathbb{P}\{\tau^*(e^{nR}) > n\} \leq e^{-2n\frac{(W^*-R)^2}{(\ln a_2 - \ln a_1)^2}}.$$

3 Time varying portfolio selection

For a general dynamic portfolio selection, the portfolio vector may depend on the past data. As before, $\mathbf{X}_i = (X_i^{(1)}, \ldots X_i^{(d)})$ denotes the return vector on trading period i. Let $\mathbf{b} = \mathbf{b}_1$ be the portfolio vector for the first trading period. For initial capital S_0, we get that

$$S_1 = S_0 \cdot \langle \mathbf{b}_1, \mathbf{X}_1 \rangle.$$

For the second trading period, S_1 is new initial capital, the portfolio vector is $\mathbf{b}_2 = \mathbf{b}(\mathbf{X}_1)$, and

$$S_2 = S_0 \cdot \langle \mathbf{b}_1 , \mathbf{X}_1 \rangle \cdot \langle \mathbf{b}(\mathbf{X}_1) , \mathbf{X}_2 \rangle .$$

For the nth trading period, a portfolio vector is $\mathbf{b}_n = \mathbf{b}(\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) = \mathbf{b}(\mathbf{X}_1^{n-1})$ and

$$S_n = S_0 \prod_{i=1}^{n} \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}) , \mathbf{X}_i \right\rangle = S_0 e^{n W_n(\mathbf{B})}$$

with the average growth rate

$$W_n(\mathbf{B}) = \frac{1}{n} \sum_{i=1}^{n} \ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}) , \mathbf{X}_i \right\rangle .$$

The fundamental limits, determined in Algoet and Cover [3], and in Algoet [1, 2], reveal that the so-called *log-optimum portfolio* $\mathbf{B}^* = \{\mathbf{b}^*(\cdot)\}$ is the best possible choice. More precisely, on trading period n let $\mathbf{b}^*(\cdot)$ be such that

$$\mathbb{E}\left\{ \ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) , \mathbf{X}_n \right\rangle \middle| \mathbf{X}_1^{n-1} \right\} = \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) , \mathbf{X}_n \right\rangle \middle| \mathbf{X}_1^{n-1} \right\} .$$

If $S_n^* = S_n(\mathbf{B}^*)$ denotes the capital achieved by a log-optimum portfolio strategy $\mathbf{B}^*$, after n trading periods, then for any other investment strategy $\mathbf{B}$ with capital $S_n = S_n(\mathbf{B})$ and with

$$\sup_n \mathbb{E}\left\{ (\ln \left\langle \mathbf{b}_n(\mathbf{X}_1^{n-1}) , \mathbf{X}_n \right\rangle)^2 \right\} < \infty,$$

and for any stationary and ergodic process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

$$\limsup_{n \to \infty} \frac{1}{n} \ln \frac{S_n}{S_n^*} \leq 0 \quad \text{a.s.} \tag{4}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \ln S_n^* = W^* \quad \text{a.s.,} \tag{5}$$

where

$$W^* := \mathbb{E}\left\{ \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \ln \left\langle \mathbf{b}(\mathbf{X}_{-\infty}^{-1}) , \mathbf{X}_0 \right\rangle \middle| \mathbf{X}_{-\infty}^{-1} \right\} \right\}$$

is the maximal possible growth rate of any investment strategy. (Note that for memoryless markets $W^* = \max_{\mathbf{b}} \mathbb{E}\{\ln \langle \mathbf{b} , \mathbf{X}_0 \rangle\}$ which shows that in this case the log-optimal portfolio is a constantly rebalanced portfolio.)

For martingale difference sequences, there is a strong law of large numbers: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ and

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\{Z_n^2\}}{n^2} < \infty$$

then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Z_i = 0 \text{ a.s.}$$

(cf. Chow [13], see also Stout [36, Theorem 3.3.1]).

Now we can prove the optimality of the log-optimal portfolio: introduce the decomposition

$$\begin{aligned}
\frac{1}{n}\ln S_n &= \frac{1}{n}\sum_{i=1}^{n} \ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \left(\ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle - \mathbb{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \right).
\end{aligned}$$

The last average is an average of martingale differences, so it tends to zero a.s. Similarly,

$$\begin{aligned}
\frac{1}{n}\ln S_n^* &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \left(\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle - \mathbb{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \right).
\end{aligned}$$

Because of the definition of the log-optimal portfolio we have that

$$\mathbb{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \le \mathbb{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\},$$

and the proof of (4) is finished.

In order to prove (5) we have to show that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \to W^*$$

a.s. Introduce the notations

$$\mathbf{b}^*_{-k}(\mathbf{X}_{n-k}^{n-1}) = \arg\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{\ln \left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}), \mathbf{X}_n \right\rangle \mid \mathbf{X}_{n-k}^{n-1}\right\}$$

$(1 \le k < n)$ and

$$\mathbf{b}^*_{-\infty}(\mathbf{X}_{-\infty}^{n-1}) = \arg\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{\ln\left\langle \mathbf{b}(\mathbf{X}_{-\infty}^{n-1}), \mathbf{X}_n\right\rangle \mid \mathbf{X}_{-\infty}^{n-1}\right\}.$$

Obviously,

$$\mathbb{E}\{\ln\left\langle \mathbf{b}^*_{-k}(\mathbf{X}_{i-k}^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_{i-k}^{i-1}\} \le \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}$$

$(i > k)$ and

$$\mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\} \le \mathbb{E}\{\ln\left\langle \mathbf{b}^*_{-\infty}(\mathbf{X}_{-\infty}^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_{-\infty}^{i-1}\}.$$

Thus, the ergodic theorem implies that

$$\begin{aligned} W^*_{-k} &:= \mathbb{E}\left\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{\ln\left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0\right\rangle \middle| \mathbf{X}_{-k}^{-1}\right\}\right\} \\ &= \lim_n \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*_{-k}(\mathbf{X}_{i-k}^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_{i-k}^{i-1}\} \\ &\le \liminf_n \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\} \end{aligned}$$

a.s. and

$$\begin{aligned} &\limsup_n \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\} \\ &\le \lim_n \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*_{-\infty}(\mathbf{X}_{-\infty}^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_{-\infty}^{i-1}\} = W^*. \end{aligned}$$

a.s. Using martingale argument one can check that

$$W^*_{-k} \uparrow W^*,$$

and so (5) is proved.

Put

$$\epsilon = \frac{W^* - R}{2}. \tag{6}$$

Concerning the rate of convergence we have that

Theorem 2. *If the market process $\{\mathbf{X}_i\}$ is stationary, ergodic and the condition (2) is satisfied, then for an arbitrary $R < W^*$, we have that*

$$\mathbb{P}\left\{e^{nR} > S_n^*\right\} \le e^{-n\frac{(W^*-R)^2}{2(\ln a_2 - \ln a_1)^2}} + \mathbb{P}\left\{R + \epsilon > \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\right\}.$$

Proof. Apply the previous decomposition:

$$
\begin{aligned}
&\mathbb{P}\{e^{nR} > S_n^*\}\\
&= \mathbb{P}\left\{R > \frac{1}{n}\ln S_n^*\right\}\\
&= \mathbb{P}\Big\{R+\epsilon-\epsilon > \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\\
&\quad +\frac{1}{n}\sum_{i=1}^n \Big(\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle - \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\Big)\Big\}\\
&\le \mathbb{P}\Big\{R+\epsilon > \frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\Big\}\\
&\quad +\mathbb{P}\Big\{-\epsilon > \frac{1}{n}\sum_{i=1}^n \Big(\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle - \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\Big)\Big\}
\end{aligned}
$$

For the second term of the right hand side, we apply the Hoeffding [26], Azuma [7] inequality: Let $X_1, X_2, \ldots$ be a sequence of random variables, and assume that $V_1, V_2, \ldots$ is a martingale difference sequence with respect to $X_1, X_2, \ldots$. Assume, furthermore, that there exist random variables $Z_1, Z_2, \ldots$ and nonnegative constants $c_1, c_2, \ldots$ such that for every $i > 0$, Z_i is a function of $X_1, \ldots, X_{i-1}$, and

$$Z_i \le V_i \le Z_i + c_i \quad \text{a.s.}$$

Then, for any $\epsilon > 0$ and n,

$$\mathbb{P}\left\{\sum_{i=1}^n V_i \ge \epsilon\right\} \le e^{-2\epsilon^2/\sum_{i=1}^n c_i^2}$$

and

$$\mathbb{P}\left\{\sum_{i=1}^n V_i \le -\epsilon\right\} \le e^{-2\epsilon^2/\sum_{i=1}^n c_i^2}.$$

Thus

$$
\begin{aligned}
&\mathbb{P}\Big\{-\epsilon > \frac{1}{n}\sum_{i=1}^n \Big(\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle - \mathbb{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\Big)\Big\}\\
&\le e^{-2n\frac{\epsilon^2}{(\ln a_2 - \ln a_1)^2}}\\
&= e^{-n\frac{(W^*-R)^2}{2(\ln a_2 - \ln a_1)^2}}.
\end{aligned}
$$

■

If the market process is just stationary and ergodic, then it is impossible to get rate of convergence of the term

$$\mathbb{P}\Big\{R+\epsilon > \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{\ln\Big\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\Big\rangle \mid \mathbf{X}_1^{i-1}\}\Big\}.$$

In order to find conditions, for which a rate can be derived, one possibility is that for $i > k$

$$\begin{aligned}\mathbb{E}\{\ln\Big\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\Big\rangle \mid \mathbf{X}_1^{i-1}\} &= \max_{\mathbf{b}(\cdot)}\mathbb{E}\{\ln\Big\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\Big\rangle \mid \mathbf{X}_1^{i-1}\}\\ &= \max_{\mathbf{b}}\mathbb{E}\{\ln\langle \mathbf{b}, \mathbf{X}_i\rangle \mid \mathbf{X}_1^{i-1}\}\\ &\geq \max_{\mathbf{b}}\mathbb{E}\{\ln\langle \mathbf{b}, \mathbf{X}_i\rangle \mid \mathbf{X}_{i-k}^{i-1}\},\end{aligned}$$

and so we may increase the above probability. We expected that the density of

$$\max_{\mathbf{b}}\mathbb{E}\{\ln\langle \mathbf{b}, \mathbf{X}_{k+1}\rangle \mid \mathbf{X}_1^{k}\}$$

has a small support, which moves to the right, when k increases.

We made an experiment verifying this conjecture empirically. At the web page `www.szit.bme.hu/~oti/portfolio` there are two benchmark data set from NYSE:

- The first data set consists of daily data of 36 stocks with length 22 years (5651 trading days ending in 1985). More precisely, the data set contains the daily price relatives, that was calculated from the nominal values of the *closing prices* corrected by the dividends and the splits for all trading day. This data set has been used for testing portfolio selection in Cover [15], in Singer [35], in Györfi, Lugosi, Udina [20], in Györfi, Ottucsák, Urbán [21], in Györfi, Udina, Walk [22] and in Györfi, Urbán, Vajda [23].
- The second data set contains 19 stocks and has length 44 years (11178 trading days ending in 2006) and it was generated same way as the previous data set (it was augmented by the last 22 years).

As in Györfi, Ottucsák, Urbán [21], for fixed $1 \leq \ell \leq 10$, we considered the kernel based portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$, where $k = 1,\dots 5$ and with radius

$$r_{k,l}^2 = 0.0002\cdot d\cdot k + 0.00002\cdot d\cdot k\cdot \ell.$$

Then, for $n > k+1$ and for $\ell = 7$, define the random variable $Z_{n,k}$ by

$$Z_{n,k} = \frac{\max_{\mathbf{b}\in\Delta_d}\sum_{\left\{k<i<n:\|\mathbf{X}_{i-k}^{i-1}-\mathbf{X}_{n-k}^{n-1}\|\leq r_{k,\ell}\right\}}\ln\langle \mathbf{b}, \mathbf{X}_i\rangle}{\left|\left\{k<i<n:\|\mathbf{X}_{i-k}^{i-1}-\mathbf{X}_{n-k}^{n-1}\|\leq r_{k,\ell}\right\}\right|},$$

if the sum is non-void. Then the histogram of $\{Z_{n,k}, n = k+1, \ldots N\}$ can be an approximation of the density of $\max_{\mathbf{b}} \mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_{k+1}\rangle \mid \mathbf{X}_1^k\}$.

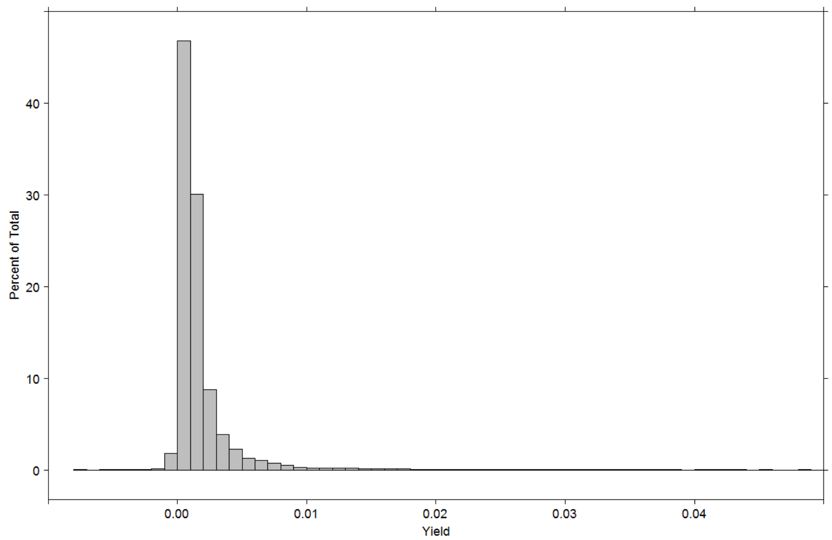

Fig. 2 The histogram of the maximum of the conditional expectations for $k = 1$

We generated these 5 histograms of the maximum of the conditional expectations. The main observation was that these histograms don't depend on k, therefore one can assume that the market process is a first order Markov process. Figure 2 shows the histogram for $k = 1$. Surprisingly, this histogram has a small support. Here are some data:

$$\begin{aligned} \text{minimum} &= -0.008 \\ \text{1st qu.} &= 0.00061 \\ \text{median} &= 0.0010 \\ \text{mean} &= 0.0019 \\ \text{3rd qu. } &= 0.0018 \\ \text{maximum} &= 0.1092. \end{aligned}$$

An important feature of this histogram is that it has a positive skewness, which means that the right hand side tail is larger than the left hand side one. The reason of this property is that $\max_{\mathbf{b}} \mathbb{E}\{\ln \langle \mathbf{b}, \mathbf{X}_{k+1}\rangle \mid \mathbf{X}_1^k\}$ is the maximum of (dependent) random variables.

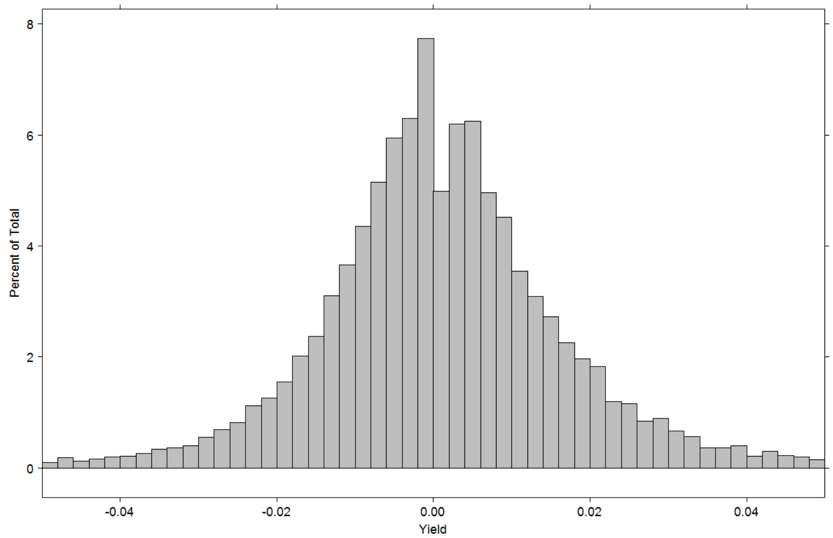

Fig. 3 The histogram of the log-returns for an empirical portfolio strategy

For the kernel based portfolio we generated the histogram of the log-return, too. The elementary portfolio is defined by

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}\in\Delta_d} \sum_{\left\{k<i<n:\|\mathbf{x}_{i-k}^{i-1}-\mathbf{x}_{n-k}^{n-1}\|\le r_{k,\ell}\right\}} \ln\langle\mathbf{b}, \mathbf{x}_i\rangle \ ,$$

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \dots, 1/d)$ otherwise. Define the random variable $Z'_{n,k}$ by

$$Z'_{n,k} = \ln\left\langle \mathbf{b}^{(k,\ell)}(\mathbf{X}_1^{n-1}), \mathbf{X}_n \right\rangle,$$

which is the daily log-return for day n. For $k = 1$ and $\ell = 7$, we generated the histogram of $\{Z'_{n,k}, n = k+1, \dots N\}$. Figure 3 shows the histogram of the log-return for the empirical portfolio strategy $\mathbf{B}^{(1,7)}$. Here are the corresponding data:

$$\begin{aligned}
\text{minimum} &= -0.1535\\
\text{1st qu.} &= -0.0077\\
\text{median} &= 0.0003\\
\text{mean} &= 0.00118\\
\text{3rd qu.} &= 0.0093\\
\text{maximum} &= 0.1522.
\end{aligned}$$

Comparing the Figures 1 and 3, one can observe that the shape and the quantiles of the histograms are almost the same. The main difference is in the mean. Since these data sets contains the relative prices for trading days only, and one year consists of 250 trading days, therefore in terms of average annual yields (AAY) the mean= 0.00118 in Figure 3 corresponds to AAY 34%, while the mean= 0.00118 for the Coke corresponds to AAY 14%.

Based on these empirical observations, in the following we assume that the market process $\{\mathbf{X}_i\}$ is a first-order stationary Markov process. In this case the log-optimum portfolio choice has the form $\mathbf{b}^*(\mathbf{X}_{n-1})$ (instead of $\mathbf{b}^*(\mathbf{X}_1^{n-1})$) maximizing $\mathbb{E}\{\ln\langle\mathbf{b},\mathbf{X}_n\rangle \mid \mathbf{X}_{n-1}\}$ such that

$$\mathbb{E}\{\ln\langle\mathbf{b}^*(\mathbf{X}_{n-1}),\mathbf{X}_n\rangle\} = W^*.$$

We assume that $\mathbf{X}_i$ has a denumerable state space $S \subset [a_1, a_2]^d$, which is realistic because the values of the components of $\mathbf{X}_i$ are quotients of integer valued prices. Further we assume that the Markov process is irreducible and aperiodic. Finally, suppose that the Markov kernel $\nu(H \mid \mathbf{x})$ defined by

$$\nu(H \mid \mathbf{x}) := \mathbb{P}\{\mathbf{X}_2 \in H \mid \mathbf{X}_1 = \mathbf{x}\}$$

($\mathbf{x} \in S$, $H \subset S$) is continuous in total variation, i.e.,

$$V(\mathbf{x},\mathbf{x}') := \sup_{H \subset S} |\nu(H \mid \mathbf{x}) - \nu(H \mid \mathbf{x}')| \to 0 \tag{7}$$

if $\mathbf{x}' \to \mathbf{x}$. Notice that by Scheffé's theorem

$$V(\mathbf{x},\mathbf{x}') := \frac{1}{2}\sum_{\mathbf{x}^* \in S} |\nu(\{\mathbf{x}^*\} \mid \mathbf{x}) - \nu(\{\mathbf{x}^*\} \mid \mathbf{x}')|.$$

The following theorem with $R < W^*$ gives exponential bounds for the probability that $e^{nR} > S_n^*$ and for the probability that after n there is a time instant m such that $e^{mR} > S_m^*$.

Theorem 3. *Let the market process $\{\mathbf{X}_i\}$ be a first-order stationary denumerable Markov chain, which is irreducible and aperiodic, satisfies (2) and (7). Then for arbitrary $R < W^*$, there exist $c, C, c^*, C^* \in (0,\infty)$ depending on $W^* - R$, $\ln a_2 - \ln a_1$ and the ergodic behavior of $\{\mathbf{X}_i\}$ such that for all n*

$$\mathbb{P}\left\{e^{nR} > S_n^*\right\} \le e^{-n\frac{(W^*-R)^2}{2(\ln a_2 - \ln a_1)^2}} + Ce^{-cn}, \tag{8}$$

and

$$\mathbb{P}\left\{\cup_{m=n}^{\infty}\{e^{mR} > S_m^*\}\right\} \le C^* e^{-c^* n}. \tag{9}$$

Proof. With the notation (6), Theorem 2 implies that

$$\mathbb{P}\left\{e^{nR} > S_n^*\right\} \le e^{-n\frac{(W^*-R)^2}{2(\ln a_2 - \ln a_1)^2}} + \mathbb{P}\left\{R+\epsilon > \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{\ln\langle \mathbf{b}^*(\mathbf{X}_{i-1}), \mathbf{X}_i\rangle \mid \mathbf{X}_{i-1}\}\right\}.$$

By stationarity, the distribution μ of $\mathbf{X}_i$ does not depend on i and satisfies

$$\int \nu(\cdot \mid \mathbf{x})\mu(d\mathbf{x}) = \mu,$$

i.e.,

$$\sum_{\mathbf{x}\in S} \nu(\{\mathbf{x}^*\} \mid \mathbf{x})\mu(\{\mathbf{x}\}) = \mu(\{\mathbf{x}^*\}). \tag{10}$$

It is well known from the theory of denumerable Markov chains (see, e.g., Feller [17]), that (10) together with irreducibility and aperiodicity of $\{\mathbf{X}_i\}$ implies that $\{\mathbf{X}_i\}$ is positive recurrent with mean recurrence time $1/\mu(\{\mathbf{x}\}) < \infty$ and weak convergence of $P_{\mathbf{X}_n|\mathbf{X}_1=\mathbf{x}}$ to μ. Thus, by Scheffé and Riesz-Vitali theorems, even

$$\begin{aligned}
&\sup_{H\subset S} |\mathbb{P}\{\mathbf{X}_n \in H \mid \mathbf{X}_1 = \mathbf{x}\} - \mu(H)| \\
&= \frac{1}{2}\sum_{\mathbf{x}^*\in S} |\mathbb{P}\{\mathbf{X}_n = \mathbf{x}^* \mid \mathbf{X}_1 = \mathbf{x}\} - \mu(\{\mathbf{x}^*\})| \\
&\to 0
\end{aligned}$$

$(n \to \infty)$ for each $\mathbf{x} \in S$. Further for each integer n

$$\begin{aligned}
&\sup_{H\subset S} |\mathbb{P}\{\mathbf{X}_n \in H \mid \mathbf{X}_1 = \mathbf{x}\} - \mathbb{P}\{\mathbf{X}_n \in H \mid \mathbf{X}_1 = \mathbf{x}'\}| \\
&= \frac{1}{2}\sum_{\mathbf{x}^*\in S} |\mathbb{P}\{\mathbf{X}_n = \mathbf{x}^* \mid \mathbf{X}_1 = \mathbf{x}\} - \mathbb{P}\{\mathbf{X}_n = \mathbf{x}^* \mid \mathbf{X}_1 = \mathbf{x}'\}| \\
&= \frac{1}{2}\sum_{\mathbf{x}^*\in S} |\sum_{\mathbf{y}\in S} \mathbb{P}\{\mathbf{X}_n = \mathbf{x}^* \mid \mathbf{X}_2 = \mathbf{y}\}(\mathbb{P}\{\mathbf{X}_2 = \mathbf{y} \mid \mathbf{X}_1 = \mathbf{x}\} - \mathbb{P}\{\mathbf{X}_2 = \mathbf{y} \mid \mathbf{X}_1 = \mathbf{x}'\})| \\
&\le \frac{1}{2}\sum_{\mathbf{x}^*\in S}\sum_{\mathbf{y}\in S} \mathbb{P}\{\mathbf{X}_n = \mathbf{x}^* \mid \mathbf{X}_2 = \mathbf{y}\}|\mathbb{P}\{\mathbf{X}_2 = \mathbf{y} \mid \mathbf{X}_1 = \mathbf{x}\} - \mathbb{P}\{\mathbf{X}_2 = \mathbf{y} \mid \mathbf{X}_1 = \mathbf{x}'\}| \\
&= \frac{1}{2}\sum_{\mathbf{y}\in S} |\mathbb{P}\{\mathbf{X}_2 = \mathbf{y} \mid \mathbf{X}_1 = \mathbf{x}\} - \mathbb{P}\{\mathbf{X}_2 = \mathbf{y} \mid \mathbf{X}_1 = \mathbf{x}'\}| \\
&= \sup_{H\subset S} |\mathbb{P}\{\mathbf{X}_2 \in H \mid \mathbf{X}_1 = \mathbf{x}\} - \mathbb{P}\{\mathbf{X}_2 \in H \mid \mathbf{X}_1 = \mathbf{x}'\}| \\
&\to 0
\end{aligned}$$

$(\mathbf{x}' \to \mathbf{x})$ by (7). Therefore even

$$\sup_{H \subset S, \mathbf{x} \in S} |\mathbb{P}\{\mathbf{X}_n \in H \mid \mathbf{X}_1 = \mathbf{x}\} - \mu(H)| \to 0.$$

Thus, the process $\{\mathbf{X}_i\}$ is φ-mixing. Also the sequence

$$\{\mathbb{E}\{\ln \langle \mathbf{b}^*(\mathbf{X}_{i-1}), \mathbf{X}_i \rangle \mid \mathbf{X}_{i-1}\}\}$$

is φ-mixing with mixing coefficients $\varphi_m \to 0$. Now we can apply Collomb's exponential inequality (p. 449 in [14]) with $d = \delta = \sqrt{D} = \frac{1}{n}(\ln a_2 - \ln a_1)$. For $m \in \{1, \dots, n\}$ we obtain

$$\begin{aligned} &\mathbb{P}\left\{ R + \epsilon > \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{\ln \langle \mathbf{b}^*(\mathbf{X}_{i-1}), \mathbf{X}_i \rangle \mid \mathbf{X}_{i-1}\} \right\} \\ &\leq \exp\left\{ \frac{n}{m} \left(3\sqrt{e}\varphi_m + \frac{3}{8} \frac{1 + 4\sum_{i=1}^{m} \varphi_i}{m} - \frac{\epsilon}{4(\ln a_2 - \ln a_1)} \right) \right\}. \end{aligned}$$

Suitable choice of $m = M(\epsilon)$ with $n \geq N(\epsilon)$ leads to the second term on the right hand side of (8) as a bound for all n. Finally, from (8) we obtain (9) as in the proof of Corollary 1. ∎

Remark. Theorem 3 can be extended to the case of a Harris-recurrent, strongly aperiodic Markov chain, not necessarily being stationary or having denumerable state space; compare in a somewhat other context Theorem 2 in Györfi and Walk [24], where Theorem 4.1 (i) of Athreya and Ney [4] and Collomb's inequality are used.

References

1. Algoet, P.: Universal schemes for prediction, gambling, and portfolio selection. *Annals of Probability*, **20**, 901–941 (1992)
2. Algoet, P.: The strong law of large numbers for sequential decisions under uncertainty. *IEEE Transactions on Information Theory*, **40**, 609–634 (1994)
3. Algoet, P. and Cover, T. M.: Asymptotic optimality asymptotic equipartition properties of log-optimum investments. *Annals of Probability*, **16**, 876–898 (1988)
4. Athreya, K. B. and Ney, P.: A new approach to the limit theory of recurrent Markov chains, *Transactions of the American Mathematical Society*, **245**, 493–501 (1978)
5. Aucamp, D. C.: An investment strategy with overshoot rebates which minimizes the time to attain a specified goal. *Management Science*, **23**, 1234–1241 (1977)
6. Aucamp, D. C.: On the extensive number of plays to achieve superior performance with the geometric mean strategy. *Management Science*, **39**, 1163–1172 (1993)
7. Azuma, K.: Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, **68**, 357–367 (1967)
8. Barron, A. R. and Cover, T. M.: A bound on the financial value of information. *IEEE Transactions on Information Theory*, **34**, 1097–1100 (1988)
9. Billingsley, R.: *Convergence of Probability Measures (2nd ed.),* New York: Wiley (1999)

10. Box, G. E. P. and Cox, D. R.: An analysis of transformations. *Journal of the Royal Statistical Society Series B*, **26**, 211–252 (1964)
11. Breiman, L.: Optimal gambling systems for favorable games. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, **1**, 65–78, Univ. California Press, Berkeley (1961)
12. Breiman, L.: *Probability,* Reading, Ma: Addison-Wesley (1968)
13. Chow, Y. S.: Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics*, **36**, 552–558 (1965)
14. Collomb, G.: Propriétés de convergence presque complète du prédicteur à noyau. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **66**, 441–460 (1984)
15. Cover, T.: Universal portfolios. *Mathematical Finance*, **1**, 1–29 (1991)
16. Ethier, S. N.: The Kelly system maximizes median fortune. *J. Appl. Probab.* **41**, 1230–1236 (2004)
17. Feller, W.: *An Introduction to Probability Theory and its Applications, vol. 1 (3rd ed.), vol. 2 (2nd ed.)* New York: Wiley (1968, 1971)
18. Fernholz, E. R.: *Stochastic Portfolio Theory,* New York: Springer (2000)
19. Finkelstein, M. and Whitley, R.: Optimal strategies for repeated games. *Advances in Applied Probability*, **13**, 415–428 (1981)
20. Györfi, L., Lugosi, G. and Udina, F.: Nonparametric kernel based sequential investment strategies. *Mathematical Finance*, **16**, 337–357 (2006)
21. Györfi, L., Ottucsák, G. and Urbán, A.: Empirical log-optimal portfolio selections: a survey, in *Machine Learning for Financial Engineering*, eds. L. Györfi, G. Ottucsák, H. Walk, pp. 81–118, Imperial College Press, London (2012)
22. Györfi, L., Udina, F. and Walk, H.: Nonparametric nearest-neighbor-based sequential investment strategies, *Statistics and Decisions*, **26**, 145–157 (2008)
23. Györfi, L., Urbán, A. and Vajda, I.: Kernel-based semi-log-optimal empirical portfolio selection strategies. *Interantional Journal of Theoretical and Applied Finance*, **10**, 505–516 (2007)
24. Györfi, L. and Walk, H.: Empirical portfolio selection strategies with proportional transaction cost, *IEEE Trans. Information Theory*, **58**, 6320–6331 (2012)
25. Hayes, T. P.: How Not to Win a Million Dollars: A Counterexample to a Conjecture of L. Breiman. 2011.
http://arxiv.org/pdf/1112.0829v1.pdf
26. Hoeffding, W.: Probability inequalities for sums of bounded random variables. *Journal of American Statistical Association*, **58**, 13–30 (1963)
27. Horváth, M. and Urbán, A.: Growth optimal portfolio selection with short selling and leverage, in *Machine Learning for Financial Engineering*, eds. L. Györfi, G. Ottucsák, H. Walk, pp. 153–178, Imperial College Press, London (2012)
28. Kardaras, C. and Platen, E.: Minimizing the expected market time to reach a certain wealth level. *SIAM Journal of Financial Mathematics*, **1**, 16–29 (2010)
29. Kelly, J. L.: A new interpretation of information rate. *Bell System Technical Journal*, **35**, 917–926 (1956)
30. Latané, H. A.: Criteria for choice among risky ventures. *Journal of Political Economy*, **67**, 145–155 (1959)
31. Lorden, G.: On excess over the boundary. *Annals of Mathematical Statistics*, **41**, 520–527 (1970)
32. MacLean, L. C., Thorp, E. O. and Ziemba, W. T. (Eds.): *The Kelly Capital Growth Investment Criterion: Theory and Practice.* World Scientific, New Jersey (2011)
33. Markowitz, H. M.: Investment for the long run: new evidence for an old rule. *Journal of Finance*, **31**, 1273–1286 (1976)
34. Samuelson, P. A.: The "fallacy" of maximizing the geometric mean in long sequences of investment or gambling. *Proceedings of the National Academy of Sciences U.S. A.*, **68**, 2493–2496 (1971)
35. Singer, Y.: Switching portfolios. *International Journal of Neural Systems*, **8**, 445–455 (1997)

36. Stout. W. F.: *Almost sure convergence.* Academic Press, New York (1974)
37. Takács, L.: On a generalization of the arc-sine law. *Annals of Applied Probability*, **6**, 1035–1040 (1996)