

Empirical portfolio selection with transaction costs for discrete market

Abstract

Discrete time growth optimal investment in stock markets with proportional transactions costs is considered. The market process is modeled by a discrete valued stationary and ergodic process. Without the knowledge of the underlying distribution of the process, we show an empirical investment strategy such that, in the long run, the growth rate on trajectories achieves the maximum with probability 1.

1. Introduction

The purpose of this paper is to investigate sequential investment strategies for financial markets in which at the beginning of each trading period a portfolio selection has to be chosen for investing the current capital in the available assets. After buying and selling transaction cost has to be paid proportionally to the amount of money moved. The sequence of price relatives is supposed to form a stationary and ergodic process with discrete values. The strategies are allowed to use the data collected from the past, however, the underlying distribution of the market process is unknown. The investor's aim is to find a strategy selecting scheme such that his wealth grows on average as fast as for any other strategy with probability 1. In particular, such optimal strategy will perform at least as well as any strategy that is even allowed to use knowledge of the distribution of the market vectors.

In case no transaction cost is involved, a number of empirical optimal strategies are known, for a comprehensive survey see Györfi et al. [4]. The problem of growth optimal investment with proportional transaction costs was studied by Cover and Iyengar [7] in horse markets, where in every market period one of the assets has positive payoff and all the others pay nothing. Iyengar [6] investigated growth optimal investment with several assets assuming independent and identically distributed sequence of market returns. The most far reaching study was Schäfer [8] who considered the maximization of expected growth rate when the asset returns are Markovian. Györfi and Vajda [5] extended it to almost sure optimality. These results, however, are based on the knowledge of the underlying distribution.

The rest of the paper is organized as follows. In Section 2, we describe the market model with transaction costs. In Section 3, we formulate the problem of short-term optimization and show its equivalence with the average growth rate maximization. The portfolio strategies are introduced and their optimality is stated. In Section 4, some technical results are derived. These show that, in a sense, short-term optimization is not very different from the one-step optimization, which corresponds to the well-examined problem of log-optimal portfolios. In Section 5, the optimality of the introduced strategies is proven and Section 6 concludes the paper.

2. Mathematical setup

Consider a market consisting of d assets. The evolution of the market in time is represented by a sequence of market vectors $s_1, s_2, \dots \in \mathbb{R}_+^d$, where $s_i = (s_{i,1}, \dots, s_{i,d})$ such that the j -th component $s_{i,j}$ of s_i denotes the price of the j -th asset at the end of the i -th period.

In order to apply the usual prediction techniques for time series analysis one has to transform the sequence $\{s_i\}$ into a more or less stationary sequence of return vectors $\{x_i\}$ as follows: $x_i = (x_{i,1}, \dots, x_{i,d})$ such that $x_{i,j} = \frac{s_{i,j}}{s_{i-1,j}}$, where $s_{0,j} = 1$ by definition. Thus, the j -th component $x_{i,j}$ of x_i denotes the amount obtained after investing a unit capital in the j -th asset on the i -th trading period.

The investor is allowed to diversify his capital at the beginning of each trading period according to a portfolio vector $b = (b^1, \dots, b^d)$. The j -th component b^j of b denotes the proportion of the investor's capital invested in asset j . Throughout the paper we assume that the portfolio vector 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 the capital is excluded. The nonnegativity of the components means that short selling and buying stocks on margin are excluded. To make the analysis feasible, some simplifying assumptions are made. We assume that assets are arbitrarily divisible and all assets are available in unbounded quantities at the current price at any given trading period. We also assume that the behavior of the market is not affected by the actions of the investor using the strategies under investigation.

For $j \leq i$ we abbreviate by x_j^i the array of return vectors $(x_j, \dots, x_i)$. Denote by Δ_d the simplex of all vectors $b \in \mathbb{R}^d$ with nonnegative components summing up to 1. An investment strategy is a sequence $\mathbf{B}$ of functions

$$b_i : (\mathbb{R}_+^d)^{i-1} \rightarrow \Delta_d, \quad i = 1, 2, \dots$$

so that $b_i(x_1^{i-1})$ denotes the portfolio vector chosen by the investor on the i -th trading period, upon observing the past behavior of the market. We write $b(x_1^{i-1}) = b_i(x_1^{i-1})$ to ease the notation.

Let S_n denote the gross wealth at the end of trading period n , $n = 0, 1, 2, \dots$, where without loss of generality let the investor's initial capital S_0 be 1 dollar, while N_n stands for the net wealth at the end of trading period n . Using the notations above, for the trading period n , the net wealth N_{n-1} can be invested according to portfolio b_n , therefore the gross wealth S_n at the end of trading period n is

$$S_n = N_{n-1} \sum_{j=1}^d b_{n,j} x_{n,j} = N_{n-1} \langle b_n, x_n \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product.

At the beginning of a new market day $n+1$, the investor sets up his new portfolio, i.e. buys/sells stocks according to the actual portfolio vector b_{n+1} . During this rearrangement, he has to pay transaction cost, therefore at the beginning of a new market day $n+1$, the net wealth N_n in the portfolio vector b_{n+1} is less than S_n . The

rate of proportional transaction costs levied on one asset are denoted by $0 < c_s < 1$ and $0 < c_p < 1$, i.e the sale of 1 dollar worth of asset nets only $1 - c_s$ dollars and the purchase of 1 dollar worth of asset costs an extra c_p dollars. We consider the special case when these rates of costs are constants over the assets. Let's calculate the transaction cost to be paid after selecting the portfolio b_{n+1} . Before rearranging the capitals, at the j -th asset there are $b_{n,j}x_{n,j}N_{n-1}$ dollars, while after rearranging we need $b_{n+1,j}N_n$ dollars. If $b_{n,j}x_{n,j}N_{n-1} > b_{n+1,j}N_n$, then we have to sell and the transaction cost is $c_s (b_{n,j}x_{n,j}N_{n-1} - b_{n+1,j}N_n)$, otherwise we have to buy and the transaction cost is $c_p (b_{n+1,j}N_n - b_{n,j}x_{n,j}N_{n-1})$.

Let x^+ denote the positive part of x . Thus, the gross wealth S_n decomposes to the sum of net wealth and cost in the following way

$$S_n = N_n + \sum_{j=1}^d c_s (b_{n,j}x_{n,j}N_{n-1} - b_{n+1,j}N_n)^+ + \sum_{j=1}^d c_p (b_{n+1,j}N_n - b_{n,j}x_{n,j}N_{n-1})^+.$$

Dividing both sides by S_n and introducing the ratio $w_n = \frac{N_n}{S_n}$, we get

$$1 = w_n + c_s \sum_{j=1}^d \left(\frac{b_{n,j}x_{n,j}}{\langle b_n, x_n \rangle} - b_{n+1,j}w_n \right)^+ + c_p \sum_{j=1}^d \left(b_{n+1,j}w_n - \frac{b_{n,j}x_{n,j}}{\langle b_n, x_n \rangle} \right)^+.$$

The equation above always has a unique solution $w_n \in (0, 1]$. This ratio only depends on b_n, b_{n+1} , and x_n , i.e. $w_n = w(b_n, b_{n+1}, x_n)$. This function w is continuous, and has the bounds $\frac{1-c_s}{1+c_p} \leq w \leq 1$.

In the sequel x_i will be a random variable and is denoted by X_i . Assume that the sequence $X_1, X_2, \dots$ form a stationary and ergodic process. Furthermore, we assume that X_i takes values in a countable set $\mathcal{S} \subset (\delta, K)^d$ for some $0 < \delta < K < \infty$.

Define the asymptotic growth rate of a strategy $\mathbf{B}$ by

$$W(\mathbf{B}) = \liminf_{n \rightarrow \infty} \frac{1}{n} \ln S_n(\mathbf{B}),$$

i.e. the number for which $S_n \approx e^{nW}$ for large n . Our aim is to maximize this rate without the knowledge of the distribution of X , that is, to find a strategy whose average growth rate is almost surely (a.s.) as large as the average growth rate of any other investment strategy.

To construct an optimal strategy we will use the concept of expert mixing. Suppose that we have a finite or countably infinite array of strategies, the so-called experts, $\{\mathbf{B}_k | k \in I\}$ and a strictly positive distribution $\{q_k | k \in I\}$ on the set of indices. Consider two types of aggregations. In the first case (AG1), distribute the initial wealth $S_0 = 1$ among the experts according to the distribution $\{q_k\}$, and let them perform their actions individually. Clearly,

$$S_n(\mathbf{B}^{AG1}) = \sum_{k \in I} q_k S_n(\mathbf{B}_k).$$

However, strictly speaking $\mathbf{B}^{AG1}$ is not an investment strategy, since it may buy and

sell from the same stock at the same time. In the second case (AG2) therefore the aggregated portfolio pays for the transaction cost. The cost paid obviously can only decrease with this change, hence

$$S_n(\mathbf{B}^{AG2}) \geq \sum_{k \in I} q_k S_n(\mathbf{B}_k). \quad (1)$$

3. Short-term optimization

Temporarily suppose that after every m trading period we don't have to pay the transaction cost. Let $\tilde{S}_n$ note the wealth after nm steps in this model, and suppose we have an empirical strategy $\mathbf{H}^m$ which is optimal, i.e.

$$\liminf_{n \rightarrow \infty} \frac{1}{n} (\ln \tilde{S}_n(\mathbf{H}^m) - \ln \tilde{S}_n(\mathbf{H}')) \geq 0 \text{ a.s.} \quad (2)$$

for any other strategy $\mathbf{H}'$. Since w_n is bounded below by a nonzero constant C , we have

$$\ln \tilde{S}_{\lfloor n/m \rfloor} + (\ln S_n - \ln S_{m \lfloor n/m \rfloor}) \geq \ln S_n \geq \ln \tilde{S}_{\lfloor n/m \rfloor} + (\ln S_n - \ln S_{m \lfloor n/m \rfloor}) + \lfloor n/m \rfloor \ln C,$$

and thus

$$\liminf_{n \rightarrow \infty} \frac{1}{n} (\ln S_n(\mathbf{H}^m) - \ln S_n(\mathbf{H}')) \geq \frac{1}{m} \ln C \text{ a.s.}$$

Therefore $\mathbf{H}^m$ is almost optimal for large m , and optimal strategy can be obtained by mixing these strategies as experts for all m . From now on we focus on constructing $\mathbf{H} = \mathbf{H}^m$ for a fixed m . This can be interpreted as short-term optimization: in each period of m , we try to find a strategy that maximizes the growth in that period, but the choices of these strategies do not effect each other. The previous argument shows that to obtain a growth optimal strategy, it suffices to construct the short-term optimal strategies $\mathbf{H}^m$.

Introduce the notation $\mathcal{B} = \{\mathbf{b} = (b_1, \dots, b_m) | b_i \text{ is an } \mathcal{S}^{i-1} \rightarrow \Delta_d \text{ function}\}$, the set of the short-term strategies. At the start of every period of m days, we choose a $\mathbf{b} \in \mathcal{B}$, which determines our strategy for that period: on the first day our portfolio is b_1 , on the second day $b_2(X_1)$, on the third day $b_3(X_1, X_2)$, etc. Obviously, every investment strategy can be described with the sequence of these short-term strategies. For arbitrary $z = (x_1, \dots, x_m) \in \mathcal{S}^m$ and $\mathbf{b} = (b_1, \dots, b_m) \in \mathcal{B}$ define

$$G(\mathbf{b}, z) = \ln(\langle b_1, x_1 \rangle w(b_1, b_2(x_1), x_1) \langle b_2(x_1), x_2 \rangle \times \dots$$

$$\dots \times w(b_{m-1}(x_1, \dots, x_{m-2}), b_m(x_1, \dots, x_{m-1}), x_{m-1}) \langle b_m(x_1, \dots, x_{m-1}), x_m \rangle).$$

That is, the doubling rate of a period if our strategy is $\mathbf{b}$ and the return vectors are $(x_1, \dots, x_m) = z$. Note that there exists a universal bound $|G(\mathbf{b}, z)| \leq M$ for all $\mathbf{b} \in \mathcal{B}$ and $z \in \mathcal{S}^m$.

Also introduce the notations $Z_i = (X_{im+1}, \dots, X_{(i+1)m})$ and $X_i^{(j)} = X_{im+j}$ for $i \geq 0$, $1 \leq j \leq m$. We can summarize all the notations regarding the sequence

$X_1, X_2, \dots$ in the following way

$$\underbrace{\underbrace{X_1 = X_0^{(1)}, \dots, X_m = X_0^{(m)}}_{Z_0}, \underbrace{X_{m+1} = X_1^{(1)}, \dots, X_{2m}}_{Z_1}, \dots, \underbrace{X_{im+1}, \dots, X_{(i+1)m}}_{Z_i}, \dots}_{Z_0^i}$$

We can now introduce the optimal strategy for a fixed m . First we define an infinite array of experts. To this end, first set for any $s \in (\mathcal{S}^m)^k$

$$\mathbf{b}^{(k)}(Z_0^n, s) = \arg \max_{\mathbf{b}=(b_1, \dots, b_m)} \sum_{k \leq i \leq n-1: Z_{i-k}^{i-1} = s} G(\mathbf{b}, Z_i).$$

$b_j: \{X_i^{(1)}, \dots, X_i^{(j-1)} | Z_{i-k}^{i-1} = s, k \leq i \leq n-1\} \rightarrow \Delta_d$

That is, we search the past for appearances of s , and design a short term strategy optimizing the return for the trading periods following each occurrence. So far this definition only defines b_j on the points of $\mathcal{S}^{j-1}$ which occurred as $(X_i^{(1)}, \dots, X_i^{(j-1)})$ after a matching $Z_{i-k}^{i-1} = s$. On the remaining points set $b_j = (1/d, \dots, 1/d)$. Also, if there is no i such that $Z_{i-k}^{i-1} = s$, set $b_j = (1/d, \dots, 1/d)$ for all j . Define the expert $\mathbf{h}^{(k)}(\cdot)$ with

$$\mathbf{h}^{(k)}(Z_0^n) = \mathbf{b}^{(k)}(Z_0^n, Z_{n-k+1}^n).$$

These experts are mixed according to (AG2) with the use of a strictly positive probability distribution $\{q_k\}$ on the positive integers, so from (1) we get

$$\tilde{S}_n(\mathbf{H}) \geq \sum_k q_k \tilde{S}_n(\mathbf{H}^{(k)}).$$

THEOREM 3.1 *The strategy $\mathbf{H} = \mathbf{H}^m$ defined above satisfies (2).*

The proof of the theorem is given in Section 5. Combined with the earlier arguments this theorem provides an empirical growth-optimal strategy.

Note that the growth rate to be maximized can be expressed as

$$\frac{1}{n} \ln \tilde{S}_n = \frac{1}{n} \sum_{i=0}^{n-1} G(\mathbf{b}_i(Z_0^{i-1}), Z_i),$$

where $\mathbf{b}_i(Z_0^{i-1}) \in \mathcal{B}$.

This leads to the following definition: let us call $\mathbf{b}^* \in \mathcal{B}$ m -logoptimal with respect to (w.r.t.) Z - or its distribution P - if

$$\mathbb{E}\{G(\mathbf{b}^*, Z)\} = \max_{\mathbf{b} \in \mathcal{B}} \mathbb{E}\{G(\mathbf{b}, Z)\}.$$

The set of m -logoptimal portfolio strategies is denoted by $B^*(P)$

Before moving on to the proof, we will need some technical results.

4. Continuity properties

The results presented in this section are about how some quantities related to $G(\mathbf{b}, Z)$ depend on the distribution P of Z . These are mainly generalizations of Section 3 and 4 of Algoet and Cover [1], which corresponds to the case $m = 1$, that is, when we ignore all transactions costs.

Introduce the notations $v(\mathbf{b}, P) = \mathbb{E}\{G(\mathbf{b}, Z)\}$ and $v^*(P) = \sup_{\mathbf{b} \in \mathcal{B}} v(\mathbf{b}, P)$. Also, let $\mathcal{P}$ denote the space of probability measures on $\mathcal{S}^m$ with the weak topology.

PROPOSITION 4.1 *The function v^* is continuous on $\mathcal{P}$.*

PROOF. Since $G(\mathbf{b}, z)$ is continuous in z , for any $\mathbf{b}^* \in B^*(P)$ and any sequence $P_n \rightarrow P$

$$\liminf v^*(P_n) \geq \liminf v(\mathbf{b}^*, P_n) = v(\mathbf{b}^*, P) = v^*(P),$$

so v^* is lower semi-continuous. On the other hand, $v(\mathbf{b}, P)$ is continuous on $\mathcal{B} \times \mathcal{P}$ as well, and since $\mathcal{B}$ is compact, it follows (Bertsekas and Shreve [3], Proposition 7.33) that v^* is upper semi-continuous on $\mathcal{P}$, too.

■

PROPOSITION 4.2 *If $\mathbf{b}^0$ and $\mathbf{b}^1$ are both m -logoptimal strategies with respect to P , then $G(\mathbf{b}^0, z) = G(\mathbf{b}^1, z)$ for P -almost every z .*

PROOF. Let $\mathbf{b}^\lambda$ denote the following strategy. Aggregate the strategies $\mathbf{b}^1$ and $\mathbf{b}^0$ according to (AG2), with the weights λ and $1 - \lambda$, respectively. From (1),

$$\exp\{G(\mathbf{b}^\lambda, z)\} \geq \lambda \exp\{G(\mathbf{b}^1, z)\} + (1 - \lambda) \exp\{G(\mathbf{b}^0, z)\},$$

and thus

$$G(\mathbf{b}^\lambda, z) \geq \lambda G(\mathbf{b}^1, z) + (1 - \lambda) G(\mathbf{b}^0, z).$$

Taking expectations gives

$$v(\mathbf{b}^\lambda, P) \geq \lambda v(\mathbf{b}^1, P) + (1 - \lambda) v(\mathbf{b}^0, P) = v^*(P),$$

where equality has the necessary condition that $G(\mathbf{b}^0, z) = G(\mathbf{b}^1, z)$ almost everywhere. ■

PROPOSITION 4.3 *The set $\{(P, \mathbf{b}) | P \in \mathcal{P}, \mathbf{b} \in B^*(P)\}$ is closed in $\mathcal{P} \times \mathcal{B}$.*

PROOF. We need to show that if $(P_n, \mathbf{b}_n) \rightarrow (P_\infty, \mathbf{b}_\infty)$ and $\mathbf{b}_n \in B^*(P_n)$, then $\mathbf{b}_\infty \in B^*(P_\infty)$. Since $v^*(P)$ is continuous in P and $v(\mathbf{b}, P)$ is continuous in both coordinates,

$$v(\mathbf{b}_\infty, P_\infty) = \lim_{n \rightarrow \infty} v(\mathbf{b}_n, P_n) = \lim_{n \rightarrow \infty} v^*(P_n) = v^*(P_\infty).$$

■

LEMMA 4.4 *If $P_n \rightarrow P_\infty$ weakly and $\mathbf{b}_i \in B^*(P_i)$ for $i = 1, \dots, \infty$, then $G(\mathbf{b}_n, z) \rightarrow G(\mathbf{b}_\infty, z)$ for P_∞ -almost every z .*

PROOF. Since $\mathcal{B}$ is sequentially compact, every subsequence of $\{\mathbf{b}_n\}$ has a convergent subsequence. The previous proposition yields that the limit of such a subsequence is m -logoptimal w.r.t. P_∞ . Therefore every subsequence of $\{G(\mathbf{b}_n, z)\}$ also has a convergent subsequence, with the limit $G(\mathbf{b}'_\infty, z)$ for some $\mathbf{b}'_\infty \in B^*(P_\infty)$. These limits coincide, therefore $\{G(\mathbf{b}_n, z)\}$ is convergent, too. ■

LEMMA 4.5 *Let Z be an $\mathcal{S}^m$ -valued random market vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{F}_k$ is an increasing sequence of σ -fields with the limiting field $\mathcal{F}_\infty \subset \mathcal{F}$, then*

$$\mathbb{E}\{\max_{\mathbf{b}} \mathbb{E}\{G(\mathbf{b}, Z)|\mathcal{F}_k\}\} \nearrow \mathbb{E}\{\max_{\mathbf{b}} \mathbb{E}\{G(\mathbf{b}, Z)|\mathcal{F}_\infty\}\} \text{ a.s.} \quad (3)$$

as $k \rightarrow \infty$ where the maximum on the left-hand side is taken over all $\mathcal{F}_k$ -measurable $\mathcal{B}$ -valued functions and the maximum on the right-hand side is taken over all $\mathcal{F}_\infty$ -measurable $\mathcal{B}$ -valued functions.

PROOF. Let P_k denote a regular conditional distribution of Z given $\mathcal{F}_k$. Lévy's martingale convergence theorem implies that

$$P_k \rightarrow P_\infty \text{ weakly a.s.}$$

There exists a selector $\mathbf{b}^*$ of m -logoptimal strategies such that $\mathbf{b}^*(P)$ is measurable in P by the measurable selection theorem of Kuratowski and Ryll-Nardzewski. With such a selector, $\mathbf{b}^*(P_k)$ is $\mathcal{F}_k$ -measurable and therefore the expression inside the expectation on the left-hand side of (3) is $v^*(P_k)$. For $l < k$, $\mathbf{b}^*(P_l)$ is $\mathcal{F}_k$ -measurable and thus

$$\mathbb{E}\{G(\mathbf{b}^*(P_l), Z)|\mathcal{F}_k\} \leq \max_{\mathbf{b}} \mathbb{E}\{G(\mathbf{b}, Z)|\mathcal{F}_k\} = v^*(P_k).$$

Taking $\mathcal{F}_l$ -conditional expectations yields that

$$v^*(P_l) = \mathbb{E}\{G(\mathbf{b}^*(P_l), Z)|\mathcal{F}_l\} \leq \mathbb{E}\{v^*(P_k)|\mathcal{F}_l\}$$

and hence $\{v^*(P_k), \mathcal{F}_k\}$ is a submartingale. Therefore, the sequence $\mathbb{E}\{v^*(P_k)\}$ is monotonically increasing. On the other hand, from the continuity and boundedness and of $v^*(\cdot)$ we get that $v^*(P_k) \rightarrow v^*(P_\infty)$ a.s. and in L_1 , too. Thus,

$$\mathbb{E}\{v^*(P_k)\} \nearrow \mathbb{E}\{v^*(P_\infty)\}.$$

■

REMARK. The countability of $\mathcal{S}$ provided the first countability and the compactness of $\mathcal{B}$ which latter turned out to be the key properties for this section. If one wants to drop the countability condition, these properties of $\mathcal{B}$ have to be ensured some other way, for example by considering only Lipschitz-continuous coordinate functions b_i with a fixed Lipschitz-constant.

5. Proof of Theorem 3.1

Besides the results of Section 4, the proof uses the following result due to Breiman [2].

LEMMA 5.1 *Let $Z = \{Z_i\}_{-\infty}^{\infty}$ be a stationary and ergodic process. For each positive integer i , let T^i denote the operator that shifts any sequence $\{\dots, z_{-1}, z_0, z_1, \dots\}$ by i digits to the left. Let $f_0, f_1, f_2, \dots$ be a sequence of real-valued functions such that $\lim_{n \rightarrow \infty} f_n(Z) = f(Z)$ almost surely for some function f . Assume that $\mathbb{E}\{\sup_n |f_n(Z)|\} < \infty$. Then almost surely*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_i(T^i Z) = \mathbb{E}f(Z).$$

An upper bound for the average growth rate can be easily calculated. Let $\mathbf{b}^*(Z_{-\infty}^{i-1})$ be m -logoptimal w.r.t. the conditional distribution $P_{Z_i|Z_{-\infty}^{i-1}}$, i.e.

$$\mathbf{b}^*(Z_{-\infty}^{i-1}) = \arg \max_{\mathbf{b}(\cdot)} \mathbb{E}\{G(\mathbf{b}(Z_{-\infty}^{i-1}), Z_i) | Z_{-\infty}^{i-1}\}.$$

For any strategy $\mathbf{H}' = \{\mathbf{b}'_i(Z_0^{i-1})\}_0^\infty$ the sum

$$\frac{1}{n} \sum_{i=0}^{n-1} G(\mathbf{b}'_i(Z_0^{i-1}), Z_i) - G(\mathbf{b}^*(Z_{-\infty}^{i-1}), Z_i)$$

is an average of supermartingale differences w.r.t. the sequence of sigma algebras $\sigma(Z_{-\infty}^{i-1})$. Indeed,

$$\mathbb{E}\{G(\mathbf{b}'_i(Z_0^{i-1}), Z_i) - G(\mathbf{b}^*(Z_{-\infty}^{i-1}), Z_i) | Z_{-\infty}^{i-1}\} \leq 0 \text{ a.s.}$$

by the definition of $\mathbf{b}^*$. The limit of the average of bounded supermartingale differences is non-positive, therefore by the ergodic theorem

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \tilde{S}_n(\mathbf{H}') &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} G(\mathbf{b}^*(Z_{-\infty}^{i-1}), Z_i) = \\ &= \mathbb{E}\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\{G(\mathbf{b}(Z_{-\infty}^{-1}), Z_0) | Z_{-\infty}^{-1}\}\} \stackrel{\text{def}}{=} W^* \text{ a.s.} \end{aligned} \quad (4)$$

Hence, if a strategy achieves the growth rate W^* , it is necessarily optimal in the sense of (2).

We have that

$$\frac{1}{n} \ln \tilde{S}_n(\mathbf{H}) \geq \frac{1}{n} \ln \left(\sum_k q_k \tilde{S}_n(\mathbf{H}^{(k)}) \right) \geq \frac{1}{n} \ln \left(\sup_k q_k \tilde{S}_n(\mathbf{H}^{(k)}) \right),$$

and thus

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \tilde{S}_n(\mathbf{H}) \geq \liminf_{n \rightarrow \infty} \sup_k \frac{1}{n} \ln \left(q_k \tilde{S}_n(\mathbf{H}^{(k)}) \right) \geq$$

$$\sup_k \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \left(q_k \tilde{S}_n(\mathbf{H}^{(k)}) \right) = \sup_k \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \tilde{S}(\mathbf{H}^{(k)}).$$

This shows that the asymptotic growth rate of the strategy $\mathbf{H}$ is at least as large as the supremum of the growth rate of the elementary strategies $\mathbf{H}^{(k)}$. Thus it suffices to examine the performance of the individual experts.

Introducing the set

$$J_{j,s}^k = \{-j + k \leq i \leq 0 \mid Z_{i-k}^{-1} = s\},$$

and the probability measure

$$P_{j,s}^k(A) = \begin{cases} \frac{\sum_{i \in J_{j,s}^k} I_{\{Z_i \in A\}}}{|J_{j,s}^k|} & \text{if } J_{j,s}^k \neq \emptyset, \\ \delta_{(1, \dots, 1)} & \text{if } J_{j,s}^k = \emptyset, \end{cases}$$

where $\delta_{(1, \dots, 1)}$ is the Dirac- δ concentrated on $(1, \dots, 1)$, $\mathbf{b}^{(k)}(Z_{-j}^{-1}, s)$ is m -logoptimal w.r.t. $P_{j,s}^k$ by definition. We have the weak convergence

$$P_{j,s}^k \xrightarrow{j \rightarrow \infty} \begin{cases} P_{Z_0 \mid Z_{-k}^{-1}=s} & \text{if } P(Z_{-k}^{-1} = s) > 0, \\ \delta_{(1, \dots, 1)} & \text{if } P(Z_{-k}^{-1} = s) = 0 \end{cases}$$

a.s. Indeed, the $P(Z_{-k}^{-1} = s) = 0$ case is trivial, and if $P(Z_{-k}^{-1} = s) > 0$, the ergodic theorem implies that

$$\int f(z) dP_{j,s}^k(z) = \frac{\frac{1}{1+j-k} \sum_{i \in J_{j,s}^k} f(Z_i)}{\frac{1}{1+j-k} |J_{j,s}^k|} \xrightarrow{j \rightarrow \infty} \frac{\mathbb{E}\{f(Z_0) I_{Z_{-k}^{-1}=s}\}}{P(Z_{-k}^{-1} = s)} = \int f(z) dP_{Z_0 \mid Z_{-k}^{-1}=s}(z)$$

a.s.

Denote the limit distribution by P_s^{*k} , and let $\mathbf{b}^{*k}(s)$ be an m -logoptimal strategy w.r.t. P_s^{*k} . Lemma 4.4 yields that a.s.

$$G(\mathbf{b}^{(k)}(Z_{-j}^{-1}, s), z) \rightarrow G(\mathbf{b}^{*k}(s), z)$$

for P_s^{*k} -a.e. and hence P_{Z_0} -a.e. z . Since s was arbitrary, we obtain

$$G(\mathbf{b}^{(k)}(Z_{-j}^{-1}, Z_{-k}^{-1}), z) \rightarrow G(\mathbf{b}^{*k}(Z_{-k}^{-1}), z) \text{ a.s.}$$

Next we apply the Lemma 5.1 for the functions

$$f_i(Z_{-\infty}^\infty) = G(\mathbf{b}^{(k)}(Z_{-i}^{-1}, Z_{-k}^{-1}), Z_0) = G(\mathbf{h}^{(k)}(Z_{-i}^{-1}), Z_0).$$

It is easy to see that this sequence satisfies the conditions with the limiting function

$$f_i(Z_{-\infty}^\infty) \xrightarrow{i \rightarrow \infty} G(\mathbf{b}^{*k}(Z_{-k}^{-1}), Z_0).$$

We get

$$\frac{1}{n}\tilde{S}_n(\mathbf{H}^{(k)}) = \frac{1}{n} \sum_{i=0}^{n-1} G(\mathbf{h}^{(k)}(Z_0^{i-1}), Z_i) = \frac{1}{n} \sum_{i=0}^{n-1} f_i(T^i Z_{-\infty}^\infty) \rightarrow \mathbb{E}\{G(\mathbf{b}^{*k}(Z_{-k}^{-1}), Z_0)\},$$

and by our previous argument

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \tilde{S}_n(\mathbf{H}) \geq \sup_k \mathbb{E}\{G(\mathbf{b}^{*k}(Z_{-k}^{-1}), Z_0)\}. \quad (5)$$

$\mathbf{b}^{*k}(Z_{-k}^{-1})$ is m -logoptimal w.r.t. $P_{Z_0|Z_{-k}^{-1}}$, and therefore

$$\mathbb{E}\{G(\mathbf{b}^{*k}(Z_{-k}^{-1}), Z_0)\} = \mathbb{E}\{\mathbb{E}\{G(\mathbf{b}^{*k}(Z_{-k}^{-1}), Z_0)|Z_{-k}^{-1}\}\} = \mathbb{E}\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\{G(\mathbf{b}(Z_{-k}^{-1}), Z_0)|Z_{-k}^{-1}\}\}.$$

Application of Lemma 4.5 to the σ -algebras $\sigma(Z_{-k}^{-1}) \nearrow \sigma(Z_{-\infty}^{-1})$ yields that the supremum in (5) is in fact a limit, and

$$\mathbb{E}\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\{G(\mathbf{b}(Z_{-k}^{-1}), Z_0)|Z_{-k}^{-1}\}\} \nearrow \mathbb{E}\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\{G(\mathbf{b}(Z_{-\infty}^{-1}), Z_0)|Z_{-\infty}^{-1}\}\} = W^*$$

and by (4) and (5) the proof is finished.

6. Summary

We constructed an empirical strategy that is optimal under quite general conditions but it has to be noted that the result is primarily theoretical. Since we mix infinity number of experts, it cannot be directly implemented. Also, because of this generality no reasonable convergence bounds can be expected. However, experiments in the model without transaction costs - see Györfi et al. [4] - suggest that strategies modified for easier implementation also provide good results in practice, even though mathematically they would not necessarily need to.

References

- [1] P. Algoet, T. Cover, "Asymptotic optimality and asymptotic equipartition properties of log-optimum investments" *Annals of Probability*, 16, pp. 876-898, 1988.
- [2] L. Breiman, "The individual ergodic theorem of information theory" *Annals of Mathematical Statistics*, 28, pp. 809-811, 1957. Correction: *Annals of Mathematical Statistics*, 31, pp. 809-810, 1960.
- [3] D.P. Bertsekas, S.E. Shreve, "Stochastic optimal control, the discrete time case" *Academic, New York*, 1978.
- [4] L. Györfi, Gy. Ottucsák, A. Urbán, "Empirical log-optimal selections: a survey" *Machine Learning Summer School 2007, MLSS 2007*, Tuebingen, Germany, August 2007.

- [5] L. Györfi, I. Vajda, "Growth optimal portfolio selection strategies with transaction costs" in *Algorithmic Learning Theory*, Freund, Y., Györfi, L., Turán, G., Zeugmann, Th. (Eds.), Lecture Notes in Artificial Intelligence, Vol. LNAI 5254, Springer, pp. 108-122, 2008.
- [6] G. Iyengar, "Discrete time growth optimal investment with costs" *Working Paper*, 2002
<http://www.columbia.edu/~gi10/Papers/stochastic.pdf>
- [7] G. Iyengar, T. Cover, "Growth optimal investment in horse race markets with costs" *IEEE Transactions on Information Theory*, 46, pp. 2675-2683, 2000.
- [8] D. Schäfer, "Nonparametric estimation for financial investment under log-utility" PhD Dissertation, Mathematical Institute, Universität Stuttgart, Aachen: Shaker Verlag, 2002.
<http://www.szit.bme.hu/~oti/portfolio/articles/dominik.pdf>