International Journal of Theoretical and Applied Finance
Vol. 10, No. 3 (2007) 505–516
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

KERNEL-BASED SEMI-LOG-OPTIMAL EMPIRICAL PORTFOLIO SELECTION STRATEGIES

LÁSZLÓ GYÖRFI*, ANDRÁS URBÁN† and ISTVÁN VAJDA‡

Department of Computer Science and Information Theory
Budapest University of Technology and Economics
1521 Stoczek u. 2, Budapest, Hungary
**gyorfi@szit.bme.hu*
†*urbi@szit.bme.hu*
‡*vajda@szit.bme.hu*

Received 11 April 2006
Accepted 11 September 2006

The purpose of this paper is to introduce an approximation of the kernel-based log-optimal investment strategy that guarantees an almost optimal rate of growth of the capital under minimal assumptions on the behavior of the market. The new strategy uses much less knowledge on the distribution of the market process. It is analyzed both theoretically and empirically. The theoretical results show that the asymptotic rate of growth well approximates the optimal one that one could achieve with a full knowledge of the statistical properties of the underlying process generating the market, under the only assumption that the market is stationary and ergodic. The empirical results show that the proposed semi-log-optimal and the log-optimal strategies have essentially the same performance measured on past NYSE data.

Keywords: Sequential investment; semi-log-optimal portfolios; kernel-based empirical portfolio selections.

1. Introduction

The purpose of this paper is to investigate sequential investment strategies for financial markets. 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. The only assumption we use in our mathematical analysis is that the daily price relatives form a stationary and ergodic process. Under this assumption the asymptotic rate of growth has a well-defined maximum which can be achieved in full knowledge of the distribution of the entire process, see Algoet and Cover [2].

Universally consistent procedures achieving the same asymptotic growth rate without any previous knowledge have been known to exist, see Algoet [1], Györfi

and Schäfer [8], Györfi *et al.*, [7]. In this paper, new strategy, called semi-log-optimal strategy, is proposed which guarantees an almost optimal asymptotic growth rate of capital for all stationary and ergodic markets, and is of small computational complexity. The procedure is an approximation of the kernel-based strategy introduced by Györfi *et al.*, [7] such that it uses only the first and second moments of the market vector. We perform an experimental study in which we compare the performance of the proposed method and the method in [7] for the data sets of New York Stock Exchange (NYSE) spanning a twenty-two-year period with thirty-six stocks included.

The rest of the paper is organized as follows. In Sec. 2 the mathematical model is described, and related results are surveyed briefly. In Sec. 4 the new kernel-based nonparametric sequential investment strategy is introduced and its main consistency properties are stated. Numerical results based on various data sets are described in Sec. 5. The proof of the main theoretical result (Theorem 4.1) is given in Sec. 6.

2. Setup, the Log-Optimal Strategy

The model of stock market investigated in this paper is the one considered, among others, by Breiman [4], Algoet and Cover [2]. Consider a market of d assets. A *market vector* $\mathbf{x} = (x^{(1)}, \ldots, x^{(d)})^T \in \mathbb{R}_+^d$ is a vector of d nonnegative numbers representing price relatives for a given trading period. That is, the jth component $x^{(j)} \geq 0$ of $\mathbf{x}$ expresses the ratio of the closing and opening prices of asset j. In other words, $x^{(j)}$ is the factor by which capital invested in the jth asset grows during the trading period.

The investor is allowed to diversify his capital at the beginning of each trading period according to a portfolio vector $\mathbf{b} = (b^{(1)}, \ldots, b^{(d)})^T$. The jth 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. Let S_0 denote the investor's initial capital. Then at the end of the trading period the investor's wealth becomes

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

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

The evolution of the market in time is represented by a sequence of market vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \mathbb{R}_+^d$, where the jth component $x_i^{(j)}$ of $\mathbf{x}_i$ denotes the amount obtained after investing a unit capital in the jth asset on the ith trading period. For $j \leq i$ we abbreviate by $\mathbf{x}_j^i$ the array of market vectors $(\mathbf{x}_j, \ldots, \mathbf{x}_i)$ and denote

by Δ_d the simplex of all vectors $\mathbf{b} \in \mathbb{R}_+^d$ with nonnegative components summing up to one. An *investment strategy* is a sequence $\mathbf{B}$ of functions

$$\mathbf{b}_i : \left(\mathbb{R}_+^d\right)^{i-1} \to \Delta_d \ , \quad i = 1, 2, \ldots$$

so that $\mathbf{b}_i(\mathbf{x}_1^{i-1})$ denotes the portfolio vector chosen by the investor on the ith trading period, upon observing the past behavior of the market. We write $\mathbf{b}(\mathbf{x}_1^{i-1}) = \mathbf{b}_i(\mathbf{x}_1^{i-1})$ to ease the notation.

Starting with an initial wealth S_0, after n trading periods, the investment strategy $\mathbf{B}$ achieves the wealth

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

where $W_n(\mathbf{B})$ denotes the *average growth rate*

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

Obviously, maximization of $S_n = S_n(\mathbf{B})$ and maximization of $W_n(\mathbf{B})$ are equivalent.

To make the analysis feasible, some simplifying assumptions are used that need to be taken into account in the usual model of log-optimal portfolio theory. Assume

- the assets are arbitrarily divisible,
- the assets are available in unbounded quantities at the current price at any given trading period,
- there are no transaction costs,
- the behavior of the market is not affected by the actions of the investor using the strategy under investigation.

In this paper we assume that the market vectors are realizations of a random process, and describe a statistical model. Our view is completely nonparametric in that the only assumption we use is that the market is stationary and ergodic, allowing arbitrarily complex distributions. More precisely, assume that $\mathbf{x}_1, \mathbf{x}_2, \ldots$ are realizations of the random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots$ drawn from the vector-valued stationary and ergodic process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$. The sequential investment problem, under these conditions, have been considered by, e.g., Breiman [4] and Algoet and Cover [2]. The fundamental limits, determined in [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

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

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

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 \quad \text{almost surely}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log S_n^* = W^* \quad \text{almost surely,}$$

where

$$W^* = \mathbb{E}\left\{\log \left\langle \mathbf{b}^*(\mathbf{X}_{-\infty}^{-1}) \, , \, \mathbf{X}_0 \right\rangle\right\}$$

is the maximal possible growth rate of any investment strategy.

3. The Semi-Log-Optimal Strategy

Thus (almost surely), no investment strategy can have a faster rate of growth than W^*. Of course, to determine a log-optimal portfolio, full knowledge of the (infinite-dimensional) distribution of the process is required. Strategies achieving the same rate of growth without knowing the distribution are called *universally consistent*, i.e., an investment strategy $\mathbf{B}$ is called universally consistent with respect to a class of stationary and ergodic processes $\{\mathbf{X}_n\}_{-\infty}^{\infty}$, if for each process in the class,

$$\lim_{n\to\infty} \frac{1}{n} \log S_n(\mathbf{B}) = W^* \quad \text{almost surely.}$$

In order to construct $\mathbf{b}^*$, one has to know the conditional distribution of $\mathbf{X}_n$ given $\mathbf{X}_1^{n-1}$. The classical Markowitz mean-variance approach to portfolio optimization for single period investment selects portfolio $\mathbf{b}$ by performance $\mathbb{E}\{\langle \mathbf{b} \, , \, \mathbf{X}_n \rangle\}$ and risk $\mathbf{Var}\{\langle \mathbf{b} \, , \, \mathbf{X}_n \rangle\}$ such that only the first and second moments of $\mathbf{X}_n$ are used in the calculations (cf. Francis [6]). Similarly, if the process $\{\mathbf{X}_n\}$ is log-normally distributed, then again only the first and second moments are needed in the derivations (cf. Schäfer [10]).

Next, for portfolio selection, we introduce a new principle, which supposes only the knowledge of the conditional first and second moments, and has almost optimal performance.

Put

$$h(z) = z - 1 - \frac{1}{2}(z-1)^2,$$

which is the second order Taylor expansion of the function $\log z$ at $z = 1$. Then, the semi-log-optimal portfolio selection is defined by

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

For $\bar{S}_n = S_n(\bar{\mathbf{B}})$, Vajda [12] proved that under the condition (4.4)

$$\liminf_{n\to\infty} \frac{1}{n} \log \bar{S}_n \geq W^* - \frac{5}{6}\mathbb{E}\Big\{ \max_m |X^{(m)} - 1|^3 \Big\} \quad \text{almost surely.}$$

4. Kernel-Based Semi-Log-Optimal Strategy

The surprising fact is that there exists empirical strategies, universally consistent with respect to the class of all stationary and ergodic processes with $\mathbb{E}|\log X^{(j)}| < \infty$

for all $j = 1, \ldots, d$ (cf. Algoet [1] and Györfi and Schäfer [8]). Györfi *et al.*, [7] introduced kernel-based strategies; here we describe only the simplest "moving-window" version, corresponding to a uniform kernel function.

Define an infinite array of experts $\mathbf{H}^{(k,\ell)} = \{\mathbf{h}^{(k,\ell)}(\cdot)\}$, where k, ℓ are positive integers. For fixed positive integers k, ℓ, choose the radius $r_{k,\ell} > 0$ such that for any fixed k,

$$\lim_{\ell \to \infty} r_{k,\ell} = 0.$$

Then, for $n > k + 1$, define the expert $\mathbf{h}^{(k,\ell)}$ as follows. Let J_n be the locations of matches:

$$J_n = \left\{ k < i < n : \|\mathbf{x}_{i-k}^{i-1} - \mathbf{x}_{n-k}^{n-1}\| \leq r_{k,\ell} \right\},$$

where $\|\cdot\|$ denotes the Euclidean norm. Put

$$\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b} \in \Delta_d} \prod_{\{i \in J_n\}} \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle\,, \tag{4.1}$$

if the product is non-void, and $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise.

These experts are mixed as follows: let $\{q_{k,\ell}\}$ be a probability distribution over the set of all pairs (k, ℓ) of positive integers such that for all k, ℓ, $q_{k,\ell} > 0$. If $S_n(\mathbf{H}^{(k,\ell)})$ is the capital accumulated by the elementary strategy $\mathbf{H}^{(k,\ell)}$ after n periods when starting with an initial capital $S_0 = 1$, then, after period n, the investor's capital becomes

$$S_n(\mathbf{B}) = \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{H}^{(k,\ell)}). \tag{4.2}$$

Györfi *et al.*, [7] proved that the kernel-based portfolio scheme $\mathbf{B}$ is universally consistent with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots, d$.

Next we introduce a modification of the previously defined strategy. Equation (4.1) can be formulated in an equivalent form:

$$\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b} \in \Delta_d} \sum_{\{i \in J_n\}} \log \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle\,.$$

The semi-log-optimal kernel-based experts $\bar{\mathbf{H}}^{(k,\ell)} = \{\bar{\mathbf{h}}^{(k,\ell)}(\cdot)\}$ are as follows:

$$\bar{\mathbf{h}}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b} \in \Delta_d} \sum_{\{i \in J_n\}} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle). \tag{4.3}$$

The semi-log-optimal kernel-based strategy $\bar{\mathbf{B}}$ is the mixture (4.2) of the experts $\{\bar{\mathbf{H}}^{(k,\ell)}\}$.

In order to compute $\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1})$, one has to make an optimization over $\mathbf{b}$. In each optimization step the computational complexity is proportional to the number of matches ($|J_n|$). For $\bar{\mathbf{h}}^{(k,\ell)}(\mathbf{x}_1^{n-1})$ this complexity can be reduced. We have that

$$\sum_{\{i \in J_n\}} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle) = \sum_{\{i \in J_n\}} (\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle - 1) - \frac{1}{2} \sum_{\{i \in J_n\}} (\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle - 1)^2.$$

If $\mathbf{1}$ denotes the all 1 vector, then

$$\sum_{\{i\in J_n\}} h(\langle \mathbf{b}\,,\,\mathbf{x}_i\rangle) = \langle \mathbf{b}\,,\,\mathbf{m}\rangle - \langle \mathbf{b}\,,\,\mathbf{Cb}\rangle,$$

where

$$\mathbf{m} = \sum_{\{i\in J_n\}} (\mathbf{x}_i - \mathbf{1})$$

and

$$\mathbf{C} = \frac{1}{2}\sum_{\{i\in J_n\}} (\mathbf{x}_i - \mathbf{1})(\mathbf{x}_i - \mathbf{1})^T.$$

If we calculate the vector $\mathbf{m}$ and $\mathbf{C}$ beforehand then in each optimization step the complexity does not depend on the number of matches, so the running time for calculating $\bar{\mathbf{h}}^{(k,\ell)}(\mathbf{x}_1^{n-1})$ is much smaller than that for $\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1})$.

Theorem 4.1. *Assume the market process is ergodic such that*

$$0.6 \le X^{(j)} \text{ and } \mathbb{E}\{|X^{(j)} - 1|^3\} < \infty, \tag{4.4}$$

for $j = 1, 2, \ldots, d$. Then, for $\bar{S}_n = S_n(\bar{\mathbf{B}})$,

$$\liminf_{n\to\infty} \frac{1}{n}\log \bar{S}_n \ge W^* - \frac{5}{6}\mathbb{E}\Big\{\max_m |X^{(m)} - 1|^3\Big\} \quad \textit{almost surely.}$$

In the next section we have some experiments for NYSE data, where the bound $\mathbb{E}\{\max_m |X^{(m)} - 1|^3\}$ in the theorem is of order 10^{-6}–10^{-4}.

5. Empirical Results

In this section we present some numerical results obtained by applying the described algorithms to some financial data consisting of the prices for 36 NYSE stocks along 22 years. The dataset we use is a standard set of NYSE data used by Cover [5], Singer [11], Hembold *et al.*, [9], and others. It includes daily prices of 36 assets along a 22-year period (5651 trading days) ending in 1985.

All the proposed algorithms use an infinite array of experts. In practice we take a finite array of size $K \times L$. In all cases select $K = 5$ and $L = 10$. Choose the uniform distribution $\{q_{k,\ell}\} = 1/(KL)$ over the experts in use, and the radius

$$r_{k,l}^2 = 0.0001 \cdot d \cdot k \cdot \ell,$$

($k = 1, \ldots, K$ and $\ell = 1, \ldots, L$).

Table 1 summarizes the wealth achieved by several portfolio. In the second column we show the wealth achieved by the best stock of the two involved, by the best constantly rebalanced portfolio (BCRP), by an oracle (defined as the best possible "anticipating" strategy which invests all the capital in the best stock each day), and the results reported in the literature for Cover's [5] universal portfolio

Table 1. Wealth achieved by different strategies by investing in the pairs of NYSE stocks.

Stocks					Best Exp. $[k, \ell]$
Iroquois	best asset	8.92	$\mathbf{B}$	2.6e+10	3.6e+11 [2,10]
Kin Ark	BCRP	73.70	$\bar{\mathbf{B}}$	2.6e+10	3.6e+11 [2,10]
	oracle	6.85e+53			
	Cover UP	39.97			
	Singer SAP	143.7			
Com. Met.	best asset	52.02	$\mathbf{B}$	1224	4765 [3,10]
Mei. Corp	BCRP	103.0	$\bar{\mathbf{B}}$	1219	4685 [3,10]
	oracle	2.12e+35			
	Cover UP	74.08			
	Singer SAP	107.7			
Com. Met.	best asset	52.02	$\mathbf{B}$	1.5e+11	1.9e+12 [2,8]
Kin Ark	BCRP	144.0	$\bar{\mathbf{B}}$	1.5e+11	1.9e+12 [2,8]
	oracle	1.84e+49			
	Cover UP	80.54			
	Singer SAP	206.7			
IBM	best asset	13.36	$\mathbf{B}$	52.3	182.4 [1,1]
Coca-Cola	BCRP	15.02	$\bar{\mathbf{B}}$	52.2	182.6 [1,1]
	oracle	1.08e+15			
	Cover UP	14.24			
	Singer SAP	15.05			

(UP) and Singer's [11] switching adaptive portfolio (SAP). (Note that the "anticipating" portfolio BRCP does not correspond to any valid investment strategy since it can only be determined in hindsight.) The third column lists our results for the kernel ($\mathbf{B}$) and the semi-log-optimal version of the kernel ($\bar{\mathbf{B}}$) portfolios. The last column lists the wealth and the index of the best expert among the KL competing experts.

Table 2 summarizes the wealth achieved by each expert at the last period when investing one unit in the Iroquois/Kin-Ark NYSE stock pair. Upper part is for the kernel-based log-optimal portfolio $\mathbf{B}$, while the lower part is for the kernel-based semi-log-optimal portfolio $\bar{\mathbf{B}}$. Experts are indexed by $k = 1, \ldots, 5$ in columns and $\ell = 1, \ldots, 10$ in rows.

From both tables we can conclude that the log-optimal and the semi-log-optimal portfolios have the same performance.

6. Proofs

The proof of Theorem 4.1 uses the following three auxiliary results. The first is known as Breiman's generalized ergodic theorem [3].

Lemma 6.1. (Breiman [3]). *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 $\{\ldots, z_{-1}, z_0, z_1, \ldots\}$ by i digits to the left. Let $f_1, f_2, \ldots$ be a sequence of real-valued*

Table 2. Wealth achieved for the Iroquois/Kin-Ark NYSE stock pair.

ℓ	k				
	1	2	3	4	5
$S_{5651}(\mathbf{B}) = 2.58e+10$					
1	1.2e+8	6.8e+3	1.7e+3	1.4e+3	2.9e+2
2	4.1e+8	3.3e+6	7.3e+4	5e+3	5.2e+2
3	2.9e+9	9.7e+7	3e+6	8.2e+4	1.4e+3
4	5.6e+9	3.7e+9	4.7e+6	1.5e+6	1.2e+5
5	9.1e+9	2.1e+10	1.8e+7	4.5e+6	3.4e+4
6	8.3e+9	4.7e+10	1.2e+8	1.6e+7	1.9e+5
7	1.2e+10	3e+11	2.6e+8	1.2e+7	1e+6
8	2.4e+10	2e+11	8.5e+8	7e+8	3e+6
9	1.4e+10	2.1e+11	1.3e+10	1.1e+9	1.4e+7
10	2.7e+10	3.6e+11	3.9e+10	5.5e+8	5e+7
$S_{5651}(\bar{\mathbf{B}}) = 2.57e+10$					
1	1.3e+8	7.2e+3	1.7e+3	1.4e+3	2.9e+2
2	4.5e+8	3.1e+6	7.2e+4	5.1e+3	5.2e+2
3	3e+9	9.9e+7	3.1e+6	8.4e+4	1.4e+3
4	5.6e+9	3.7e+9	4.8e+6	1.6e+6	1.2e+5
5	9.6e+9	2.2e+10	1.9e+7	4.5e+6	3.5e+4
6	8.8e+9	4.8e+10	1.3e+8	1.6e+7	2e+5
7	1.2e+10	2.9e+11	2.7e+8	1.2e+7	1e+6
8	2.5e+10	1.9e+11	8.8e+8	6.9e+8	2.9e+6
9	1.4e+10	2e+11	1.4e+10	1.1e+9	1.4e+7
10	2.7e+10	3.6e+11	3.8e+10	5.4e+8	5e+7

functions such that $\lim_{n\to\infty} f_n(Z) = f(Z)$ *almost surely for some function* f. *Assume that* $\mathbb{E}\sup_n |f_n(Z)| < \infty$. *Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_i(T^i Z) = \mathbb{E} f(Z) \quad \textit{almost surely.}$$

The next two lemmas are the slight modifications of the results due to Algoet and Cover [2], Theorems 3 and 4.

Lemma 6.2. *Let* $\mathbf{Q}_{n\in\mathcal{N}\cup\{\infty\}}$ *be a family of regular probability distributions over the set* $\mathbb{R}^d_+$ *of all market vectors such that* $\mathbb{E}\{|U_n^{(j)}|^2\} < \infty$ *for any coordinate of a random market vector* $\mathbf{U}_n = (U_n^{(1)}, \ldots, U_n^{(d)})$ *distributed according to* $\mathbf{Q}_n$. *In addition, let* $\mathbf{B}^*(\mathbf{Q}_n)$ *be the set of all semi-log-optimal portfolios with respect to* $\mathbf{Q}_n$, *that is, the set of all portfolios* $\mathbf{b}$ *that attain* $\max_{\mathbf{b}\in\Delta_d} \mathbb{E}\{h(\langle \mathbf{b}\,,\, \mathbf{U}_n\rangle)\}$. *Consider an arbitrary sequence* $\mathbf{b}_n \in \mathbf{B}^*(\mathbf{Q}_n)$. *If*

$$\mathbf{Q}_n \to \mathbf{Q}_\infty \quad \textit{weakly as } n \to \infty$$

then, for $\mathbf{Q}_\infty$*—almost all* $\mathbf{u}$,

$$\lim_{n\to\infty} \langle \mathbf{b}_n\,,\, \mathbf{u}\rangle = \langle \mathbf{b}^*\,,\, \mathbf{u}\rangle\,,$$

where the right-hand side is constant as $\mathbf{b}^*$ *ranges over* $\mathbf{B}^*(\mathbf{Q}_\infty)$.

Lemma 6.3. *Let* $\mathbf{X}$ *be a random market vector defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *satisfying* $\mathbb{E}\{|X^{(j)}|^2\} < \infty$. *If* $\mathcal{F}_k$ *is an increasing sequence of sub-*σ*-fields of* $\mathcal{F}$ *with*

$$\mathcal{F}_k \nearrow \mathcal{F}_\infty \subseteq \mathcal{F} \ ,$$

then

$$\mathbb{E}\left\{\max_{\mathbf{b}} \mathbb{E}\left[h(\langle \mathbf{b}\,,\, \mathbf{X}\rangle)|\mathcal{F}_k\right]\right\} \nearrow \mathbb{E}\left\{\max_{\mathbf{b}} \mathbb{E}\left[h(\langle \mathbf{b}\,,\, \mathbf{X}\rangle)|\mathcal{F}_\infty\right]\right\}$$

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

Proof of Theorem 4.1. The proof is an easy modification of Györfi *et al.*, [7]. Without loss of generality we may assume $S_0 = 1$, so

$$\begin{aligned}
\liminf_{n\to\infty} W_n(\bar{\mathbf{B}}) &= \liminf_{n\to\infty} \frac{1}{n} \log S_n(\bar{\mathbf{B}}) \\
&= \liminf_{n\to\infty} \frac{1}{n} \log \left(\sum_{k,\ell} q_{k,\ell} S_n(\bar{\mathbf{H}}^{(k,\ell)}) \right) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log \left(\sup_{k,\ell} q_{k,\ell} S_n(\bar{\mathbf{H}}^{(k,\ell)}) \right) \\
&= \liminf_{n\to\infty} \frac{1}{n} \sup_{k,\ell} \left(\log q_{k,\ell} + \log S_n(\bar{\mathbf{H}}^{(k,\ell)}) \right) \\
&= \liminf_{n\to\infty} \sup_{k,\ell} \left(W_n(\bar{\mathbf{H}}^{(k,\ell)}) + \frac{\log q_{k,\ell}}{n} \right) \\
&\geq \sup_{k,\ell} \liminf_{n\to\infty} \left(W_n(\bar{\mathbf{H}}^{(k,\ell)}) + \frac{\log q_{k,\ell}}{n} \right) \\
&= \sup_{k,\ell} \liminf_{n\to\infty} W_n(\bar{\mathbf{H}}^{(k,\ell)}). \qquad (6.1)
\end{aligned}$$

Because of the property of the Taylor expansion, we have that $0.6 \leq z$ implies the inequalities

$$h(z) - \frac{1}{2}|z-1|^3 \leq \log z \leq h(z) + \frac{1}{3}|z-1|^3,$$

therefore

$$\begin{aligned}
W_n(\bar{\mathbf{H}}^{(k,\ell)}) &= \frac{1}{n} \sum_{i=1}^n \log \left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{X}_1^{i-1})\,,\, \mathbf{X}_i \right\rangle \\
&\geq \frac{1}{n} \sum_{i=1}^n h\left(\left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{X}_1^{i-1})\,,\, \mathbf{X}_i \right\rangle \right) - \frac{1}{2n} \sum_{i=1}^n \left| \left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{X}_1^{i-1})\,,\, \mathbf{X}_i \right\rangle - 1 \right|^3 .
\end{aligned}$$

By Jensen's inequality,

$$|\langle \mathbf{b}\,,\,\mathbf{X}_i\rangle - 1|^3 = \left|\sum_{m=1}^d b^{(m)}(X_i^{(m)}-1)\right|^3 \le \sum_{m=1}^d b^{(m)}\left|X_i^{(m)}-1\right|^3 \le \max_m \left|X_i^{(m)}-1\right|^3,$$

therefore

$$-\frac{1}{2n}\sum_{i=1}^n \left|\left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{X}_1^{i-1})\,,\,\mathbf{X}_i\right\rangle - 1\right|^3 \ge -\frac{1}{2n}\sum_{i=1}^n \max_m \left|X_i^{(m)}-1\right|^3$$
$$\to -\frac{1}{2}\mathbb{E}\left\{\max_m \left|X_0^{(m)}-1\right|^3\right\}.$$

Let the integers k,ℓ, and the vector $\mathbf{s}=\mathbf{s}_{-k}^{-1}\in\mathbb{R}_+^{dk}$ be fixed. Let $\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}$ denote the (random) measure concentrated on $\{\mathbf{X}_i : 1-j+k\le i\le 0, \|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\|\le r_{k,\ell}\}$ defined by

$$\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}(A) = \frac{\sum_{i:1-j+k\le i\le 0,\|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\|\le r_{k,\ell}} \mathbb{I}_A(\mathbf{X}_i)}{|\{i : 1-j+k\le i\le 0, \|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\|\le r_{k,\ell}\}|}\,, \qquad A\subset\mathbb{R}_+^d$$

where $\mathbb{I}_A$ denotes the indicator function of the set A. If the above set of $\mathbf{X}_i$s is empty, then let $\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}=\delta_{(1,\dots,1)}$ be the probability measure concentrated on the vector $(1,\dots,1)$. Györfi *et al.*, [7] proved that for all $\mathbf{s}$, with probability one,

$$\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)} \to \mathbb{P}_{\mathbf{s}}^{*(k,\ell)} = \begin{cases} \mathbb{P}_{\mathbf{X}_0|\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}} & \text{if } \mathbb{P}(\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell})>0, \\ \delta_{(1,\dots,1)} & \text{if } \mathbb{P}(\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell})=0 \end{cases} \tag{6.2}$$

weakly as $j\to\infty$ where $\mathbb{P}_{\mathbf{X}_0|\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}}$ denotes the distribution of the vector $\mathbf{X}_0$ conditioned on the event $\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}$.

By definition, $\bar{\mathbf{b}}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{s})$ is a semi-log-optimal portfolio with respect to the probability measure $\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}$. Let $\bar{\mathbf{b}}_{k,\ell}^*(\mathbf{s})$ denote a semi-log-optimal portfolio with respect to the limit distribution $\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}$. Then, using Lemma 6.2, we infer from (6.2) that, as j tends to infinity, we have the almost sure convergence

$$\lim_{j\to\infty}\left\langle \bar{\mathbf{b}}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{s})\,,\,\mathbf{x}_0\right\rangle = \left\langle \bar{\mathbf{b}}_{k,\ell}^*(s)\,,\,\mathbf{x}_0\right\rangle$$

for $\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}$—almost all $\mathbf{x}_0$ and hence for $\mathbb{P}_{\mathbf{X}_0}$—almost all $\mathbf{x}_0$. Since $\mathbf{s}$ was arbitrary, we obtain

$$\lim_{j\to\infty}\left\langle \bar{\mathbf{b}}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{X}_{-k}^{-1})\,,\,\mathbf{x}_0\right\rangle = \left\langle \bar{\mathbf{b}}_{k,\ell}^*(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{x}_0\right\rangle \quad \text{almost surely.} \tag{6.3}$$

Next we apply Lemma 6.1 for the function

$$f_i(\mathbf{x}_{-\infty}^{\infty}) = h\left(\left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{x}_{1-i}^{-1})\,,\,\mathbf{x}_0\right\rangle\right) = h\left(\left\langle \bar{\mathbf{b}}^{(k,\ell)}(\mathbf{x}_{1-i}^{-1},\mathbf{x}_{-k}^{-1})\,,\,\mathbf{x}_0\right\rangle\right)$$

defined on $\mathbf{x}_{-\infty}^{\infty}=(\dots,\mathbf{x}_{-1},\mathbf{x}_0,\mathbf{x}_1,\dots)$. Put

$$\tilde{h}(z) = |z-1| + \frac{1}{2}(z-1)^2,$$

then for each $\mathbf{b}$, the Jensen inequality implies that

$$|h\left(\langle \mathbf{b}\,,\,\mathbf{X}_0\rangle\right)| \le \tilde{h}\left(\langle \mathbf{b}\,,\,\mathbf{X}_0\rangle\right) \le \sum_{m=1}^{d} b^{(m)}\tilde{h}(X_0^{(m)}) \le \sum_{m=1}^{d} \tilde{h}(X_0^{(m)}),$$

therefore

$$\left|f_i(\mathbf{X}_{-\infty}^{\infty})\right| = \left|h\left(\left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{X}_{1-i}^{-1})\,,\,\mathbf{X}_0\right\rangle\right)\right| \le \sum_{m=1}^{d} \tilde{h}(X_0^{(m)})\,,$$

which has finite expectation, and

$$f_i(\mathbf{X}_{-\infty}^{\infty}) \to h\left(\left\langle \bar{\mathbf{b}}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle\right) \quad \text{almost surely as } i\to\infty$$

by (6.3). As $n\to\infty$, Lemma 6.1 yields

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} f_i(T^i\mathbf{X}_{-\infty}^{\infty}) &= \frac{1}{n}\sum_{i=1}^{n} h\left(\left\langle \bar{\mathbf{h}}^{(k,\ell)}(\mathbf{X}_1^{i-1})\,,\,\mathbf{X}_i\right\rangle\right)\\ &\to \mathbb{E}\left\{h\left(\left\langle \bar{\mathbf{b}}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle\right)\right\}\\ &\stackrel{\text{def}}{=} \bar{\epsilon}_{k,\ell} \quad \text{almost surely.}\end{aligned}$$

Let $\mathbf{b}_{k,\ell}^{*}(\mathbf{s})$ denote a log-optimal portfolio with respect to the limit distribution $\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}$. Then

$$\begin{aligned}\bar{\epsilon}_{k,\ell} &= \mathbb{E}\left\{h\left(\left\langle \bar{\mathbf{b}}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle\right)\right\}\\ &\ge \mathbb{E}\left\{h\left(\left\langle \mathbf{b}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle\right)\right\}\\ &\ge \mathbb{E}\left\{\log\left\langle \mathbf{b}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle\right\} - \frac{1}{3}\mathbb{E}\left\{\max_m\left|X_0^{(m)}-1\right|^3\right\}\\ &\stackrel{\text{def}}{=} \epsilon_{k,\ell} - \frac{1}{3}\mathbb{E}\left\{\max_m\left|X_0^{(m)}-1\right|^3\right\}.\end{aligned}$$

Györfi *et al.*, [7] proved that

$$\sup_{k,\ell}\epsilon_{k,\ell} = W^*,$$

therefore, by (6.1) we have

$$\begin{aligned}\liminf_{n\to\infty} W_n(\bar{\mathbf{B}}) &\ge \sup_{k,\ell}\bar{\epsilon}_{k,\ell} - \frac{1}{2}\mathbb{E}\left\{\max_m\left|X_0^{(m)}-1\right|^3\right\}\\ &\ge \sup_{k,\ell}\epsilon_{k,\ell} - \frac{5}{6}\mathbb{E}\left\{\max_m\left|X_0^{(m)}-1\right|^3\right\}\\ &= W^* - \frac{5}{6}\mathbb{E}\left\{\max_m\left|X_0^{(m)}-1\right|^3\right\} \quad \text{almost surely,}\end{aligned}$$

and the proof of the theorem is completed. □

Acknowledgments

The first author acknowledges the support of the Computer and Automation Research Institute and the Research Group of Informatics and Electronics of the Hungarian Academy of Sciences.

References

[1] P. Algoet, Universal schemes for prediction, gambling, and portfolio selection, *Annals of Probability* **20** (1992) 901–941.

[2] P. Algoet and T. Cover, Asymptotic optimality asymptotic equipartition properties of log-optimum investments, *Annals of Probability* **16** (1988) 876–898.

[3] L. Breiman, The individual ergodic theorem of information theory, *Annals of Mathematical Statistics* **28** (1957) 809–811. Correction: *Annals of Mathematical Statistics* **31** (1960) 809–810.

[4] L. Breiman, Optimal gambling systems for favorable games, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1961), pp. 65–78.

[5] T. M. Cover, Universal portfolios, *Mathematical Finance* **1** (1991) 1–29.

[6] J. C. Francis, *Investments — Analysis and Management* (McGraw-Hill, New York, 1980).

[7] L. Györfi, G. Lugosi and F. Udina. Nonparametric kernel-based sequential investment strategies, *Mathematical Finance* **16** (2006) 337–357.

[8] L. Györfi and D. Schäfer, Nonparametric prediction, in *Advances in Learning Theory: Methods, Models and Applications*, eds. J. A. K. Suykens, G. Horváth, S. Basu, C. Micchelli and J. Vandevalle, NATO Science Series (IOS Press, 2003), pp. 339–354.

[9] D. P. Helmbold, R. E. Schapire, Y. Singer and M. K. Warmuth, On-line portfolio selection using multiplicative updates, *Mathematical Finance* **8** (1998) 325–344.

[10] D. Schäfer, When should we be prepared to improve a portfolio by lacklustre stocks? — A note on log-optimal portfolio selection, *International Journal of Theoretical and Applied Finance* **6** (2003) 693–702.

[11] Y. Singer, Switching portfolios, *International Journal of Neural Systems* **8** (1997) 445–455.

[12] I. Vajda, Analysis of semi-log-optimal investment strategies, in *Prague Stochastics 2006*, eds. M. Huskova and M. Janzura (Matfyz press, Prague, 2006), pp. 719–727.