Nonparametric kernel-based sequential investment strategies *

László Györfi
Department of Computer Science and Information Theory
Technical University of Budapest
1521 Stoczek u. 2,
Budapest, Hungary
gyorfi@szit.bme.hu

Gábor Lugosi and Frederic Udina
Department of Economics,
Pompeu Fabra University
Ramon Trias Fargas 25-27,
08005 Barcelona, Spain,
lugosi@upf.es and udina@upf.es

May 24, 2005

Abstract

The purpose of this paper is to introduce sequential investment strategies that guarantee an optimal rate of growth of the capital under minimal assumptions on the behavior of the market. The new

*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. The work of the second and third authors was supported by the Spanish Ministry of Science and Technology and FEDER, grant BMF2003-03324. The second author acknowledges support of the PASCAL Network of Excellence under EC grant no. 506778.

strategies are analyzed both theoretically and empirically. The theoretical results show that the asymptotic rate of growth matches 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 performance of the proposed investment strategies measured on past NYSE and currency exchange data is solid, and sometimes even spectacular.

Key words and phrases: sequential investment, universal portfolios, kernel estimation.

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 on the long run without knowing the underlying distribution generating the stock prices. Since accurate statistical modeling of stock market behavior has been known as a notoriously difficult problem, we take an extreme point of view and work with minimal assumptions on the distribution of the time series. In fact, 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 [3].

Universal procedures achieving the same asymptotic growth rate without any previous knowledge have been known to exist, see Algoet [1], Györfi and Schäfer [17]. In this paper new universal strategies are proposed which not only guarantee an optimal asymptotic growth rate of capital for all stationary and ergodic markets, but also have a good finite-horizon performance in practice. This is demonstrated on an experimental study in which the performance of the proposed methods is measured on different data sets, including past New York Stock Exchange (NYSE) data spanning a twenty-two-year period with thirty six stocks included, and currency exchange values of eight currencies over a fifteen-year period.

The experimental results demonstrate that the proposed methods are able to find, and effectively exploit, hidden complicated dependences of asset prices on the past evolution of the market.

To make the analysis feasible, some simplifying assumptions are used that need to be taken into account. First of all, 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 ignore transaction costs in the mathematical analysis, though some of the experimental results are also presented with transaction costs present. Another key assumption is that the behavior of the market is not affected by the actions of the investor using the strategy under investigation. This assumption is realistic when the investor handles small amounts of capital compared to the total trading volume on the market. Under this hypothesis, testing the methods on past stock-market data is meaningful. On the other hand, the spectacular growth of capital demonstrated by some of the proposed methods (e.g., by a factor of more than 10^8 during twenty two years on the New York Stock Exchange) should be interpreted with care, since such an explosive growth in real markets would inevitably be accompanied by some reaction of the market whose effect is not taken into account neither in the theoretical results, nor in the experimental figures based on past price fluctuations. In spite of these simplifications, we feel that our numerical results provide a strong empirical evidence for the inefficiency of the stock markets. This may partially be explained by the fact that the dependence structures of the markets revealed by the proposed investment strategies are quite complex and even though all information we use is publicly available, the way this information can be exploited remains hidden from most traders.

The rest of the paper is organized as follows. In Section 2 the mathematical model is described, and related results are surveyed briefly. In Section 3 a family of kernel-based nonparametric sequential investment strategies is introduced and its main consistency properties are stated. Numerical results based on various data sets are described in Section 4. The proof of the main theoretical result (Theorems 1 and 2) is given in Section 5.

2 Setup, mathematical model

The model of stock market investigated in this paper is the one considered, among others, by Breiman [8], Algoet and Cover [3], Cover [10]. Consider a market of d assets. A *market vector* $\mathbf{x} = (x^{(1)}, \dots x^{(d)}) \in \mathbb{R}_+^d$ is a vector of d nonnegative numbers representing price relatives for a given trading period. That is, the j-th 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 j-th 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 $\mathbf{b} = (b^{(1)}, \dots b^{(d)})$. 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. 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 $\mathbf{x}_1, \mathbf{x}_2, \dots \in \mathbb{R}_+^d$, where the j-th component $x_i^{(j)}$ of $\mathbf{x}_i$ denotes the amount obtained after investing a unit capital in the j-th asset on the i-th trading period. For $j \le i$ we abbreviate by $\mathbf{x}_j^i$ the array of market vectors $(\mathbf{x}_j, \dots, \mathbf{x}_i)$ 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

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

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

In modeling the behavior of the evolution of the market, two main approaches have been considered in the theory of sequential investment. One of them allows the market sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$ to take completely arbitrary values and no stochastic model is imposed on the mechanism generating the price relatives, see, for example, Cover [10], Cover and Ordentlich [11], Singer [22], Hembold, Schapire, Singer, and Warmuth [18], Ordentlich and Cover [21], Vovk and Watkins [26], Blum and Kalai [5], Borodin, El-Yaniv, and Gogan [6], Cesa-Bianchi and Lugosi [9], Cross and Barron [12], Stoltz and Lugosi [24]. In this approach the achieved wealth is compared with that of the best in a class of reference strategies. For example, Cover [10] considers the class of all *constantly rebalanced portfolios* (CRP) defined by strategies $\mathbf{B}$ for which $\mathbf{b}_i(\mathbf{x}_1^{i-1})$ equals a fixed portfolio vector independently of i and the past $\mathbf{x}_1^{i-1}$. Cover showed that there exist investment strategies $\mathbf{B}$ (so-called *universal portfolios*) which perform almost as well as the best constantly rebalanced portfolio in the sense that

$$W_n(\mathbf{B}) \geq \max_{\mathbf{C} \in \mathcal{C}} W_n(\mathbf{C}) - \left(\frac{d-1}{2n} \log n + O\left(\frac{1}{n}\right) \right)$$

for all possible market sequences $\mathbf{x}_1^n$, where $\mathcal{C}$ denotes the class of all constantly rebalanced portfolios. This result has been extended in various ways in the above-mentioned references.

The advantage of this "worst-case" approach is that it avoids imposing statistical models on the stock market and the results hold for all possible sequences $\mathbf{x}_1^n$. In this sense this approach is extremely robust. However, it is difficult to control the behavior of the best strategy in the reference class. For example, constantly rebalanced portfolios are known to be asymptotically optimal if the market vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are realizations of an independent, identically distributed sequence of random vectors (see below) but are insufficient if the market vectors of different trading periods have a statistical dependence, which seems to be the case in real-world markets. For this reason, larger reference classes have also been considered (see, e.g., the side-information model of Cover and Ordentlich [11], the switching portfolios of Singer [22] and also Cross and Barron [12]) but similar limitations still hold.

Another possibility is to assume that the market vectors are realizations of a random process, and describe a statistical model. The advantage of this view more classical view is that, for each process, an optimal strategy may be determined (in a sense specified below), which depends on the unknown distribution of the process, and use the past market sequence to estimate the statistical features necessary to approximate the optimal strategy. However, one has to proceed with care since complicated dependences in time and across stocks make statistical modeling extremely difficult.

In this paper we adopt a compromise between the worst-case and the statistical approaches. Even though we assume that the market sequence is a realization of a random process, we do not assume any parametric structure on the distribution or on the time dependences. Our view is completely nonparametric in that the only assumption we use is that the market is stationary and ergodic, allowing arbitrarily complex distributions. The main message of this paper is that there exist completely nonparametric investment strategies that effectively find these hidden complex dependences in the past data and are able to use this information to produce a rapid growth of the capital.

More precisely, assume that $x_1, x_2, \ldots$ are realizations of the random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots$ drawn from the vector-valued stationary and ergodic process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$. (Note that by Kolmogorov's theorem any stationary and ergodic process $\{\mathbf{X}_n\}_1^{\infty}$ can be extended to a bi-infinite stationary process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that ergodicity holds for both $n \to \infty$ and $n \to -\infty$). The sequential investment problem, under these conditions, have been considered by, e.g., Breiman [8], Algoet and Cover [3], Algoet [1, 2], Walk and Yakowitz [27], Györfi and Schäfer [17]. The fundamental limits, determined in [3], [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

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

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

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

is the maximal possible growth rate of any investment strategy. (Note that for i.i.d. markets $W^* = \max_{\mathbf{b}} \mathbb{E}\{\log \langle \mathbf{b} \, , \, \mathbf{X}_0 \rangle\}$ which shows that in this case the log-optimal portfolio is a constantly rebalanced portfolio, see also Breiman [8], Kelly [19], Latané [20], Finkelstein and Whitley [13], and Barron and Cover [4].)

Thus, (almost surely) no investment strategy can have a faster rate of growth than a log-optimal portfolio. 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 *universal* in this paper. More precisely, an investment strategy $\mathbf{B}$ is called universal with respect to a class of stationary and ergodic processes $\{\mathbf{X}_n\}_{-\infty}^{\infty}$, if for each process in the class,

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

The surprising fact that there exists a strategy, universal with respect to the class of all stationary and ergodic processes was proved by Algoet [1]. Algoet's construction is, however, quite complex and, despite of its theoretical importance, has little practical value. Algoet also introduced a simpler scheme and sketched the proof of its universality, which was completed by Györfi and Schäfer [17].

Next we describe Györfi and Schäfer's version of Algoet's scheme as the investment strategies defined in this paper are generalizations of this method. We call this scheme a *histogram-based* investment strategy and denote it by $\mathbf{B}^H$.

$\mathbf{B}^H$ is constructed as follows. We first define an infinite array of elementary strategies (the so-called *experts*) $\mathbf{H}^{(k,\ell)} = \{\mathbf{h}^{(k,\ell)}(\cdot)\}$, indexed by the positive integers $k, \ell = 1, 2, \ldots$. Each expert $\mathbf{H}^{(k,\ell)}$ is determined by a period length k and by a partition $\mathcal{P}_\ell = \{A_{\ell,j}\}$, $j = 1, 2, \ldots, m_\ell$ of $\mathbb{R}_+^d$ into m_ℓ disjoint sets. To determine its portfolio on the nth trading period, expert $\mathbf{H}^{(k,\ell)}$ looks at the market vectors $\mathbf{x}_{n-k}, \ldots, \mathbf{x}_{n-1}$ of the last k periods, discretizes this kd-dimensional vector by means of the partition $\mathcal{P}_\ell$, and determines the portfolio vector which is optimal for those past trading periods whose preceding k trading periods have identical discretized market vectors to the present one. Formally, let G_ℓ be the discretization function corresponding to the partition $\mathcal{P}_\ell$, that is,

$$G_\ell(\mathbf{x}) = j, \text{ if } \mathbf{x} \in A_{\ell,j} \, .$$

With some abuse of notation, for any n and $\mathbf{x}_1^n \in \mathbb{R}^{dn}$, we write $G_\ell(\mathbf{x}_1^n)$ for the sequence $G_\ell(\mathbf{x}_1), \ldots, G_\ell(\mathbf{x}_n)$. Then define the expert $\mathbf{H}^{(k,\ell)}$ by writing, for each $n > k+1$,

$$\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b} \in \Delta_d} \prod_{\left\{k<i<n: G_\ell(\mathbf{x}_{i-k}^{i-1}) = G_\ell(\mathbf{x}_{n-k}^{n-1})\right\}} \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle \ , \qquad (1)$$

if the product is non-void, and uniform $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise. That is, $\mathbf{h}_n^{(k,\ell)}$ discretizes the sequence $\mathbf{x}_1^{n-1}$ according to the partition $\mathcal{P}_\ell$, and browses through all past appearances of the last seen discretized string $G_\ell(\mathbf{x}_{n-k}^{n-1})$ of length k. Then it designs a fixed portfolio vector optimizing the return for the trading periods following each occurrence of this string.

The histogram-based strategy $\mathbf{B}^H$ forms a “mixture” of all experts $\mathbf{H}^{(k,\ell)}$ using a probability distribution $\{q_{k,\ell}\}$ on the set of all pairs (k, ℓ) of positive integers such that for all k, ℓ, $q_{k,\ell} > 0$. The strategy $\mathbf{B}^H$ simply weighs the experts $\mathbf{H}^{(k,\ell)}$ according to their past performances and $\{q_{k,\ell}\}$ such that after the nth trading period, the investor’s capital becomes

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

where $S_n(\mathbf{H}^{(k,\ell)})$ is the capital accumulated after n periods when using the portfolio strategy $\mathbf{H}^{(k,\ell)}$ with initial capital S_0. This may easily be achieved by distributing the initial capital S_0 among all experts such that expert $\mathbf{H}^{(k,\ell)}$ trades with initial capital $q_{k,\ell}S_0$. It is shown in [17] that the strategy $\mathbf{B}^H$ is universal with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, for all $j = 1, 2, \ldots, d$ under the following two conditions on the partitions used in the discretization:

(a) the sequence of partitions is nested, that is, any cell of $\mathcal{P}_{\ell+1}$ is a subset of a cell of $\mathcal{P}_\ell$, $\ell = 1, 2, \ldots$;

(b) if $\mathrm{diam}(A) = \sup_{\mathbf{x},\mathbf{y}\in A} \|\mathbf{x}-\mathbf{y}\|$ denotes the diameter of a set, then for any sphere $S \subset \mathbb{R}^d$ centered at the origin,

$$\lim_{\ell\to\infty} \max_{j:A_{\ell,j}\cap S\neq\emptyset} \mathrm{diam}(A_{\ell,j}) = 0 \ .$$

Remark. In the above-mentioned result, the only condition on the market process is that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$. However, this condition is not very restrictive for two reasons. First, most "real" markets obviously satisfy such a condition. Second, the result may be generalized so that it includes all ergodic market processes by using a slightly more complicated scheme suggested by Algoet [1]. This scheme uses a three-dimensional array $\mathbf{h}^{(k,\ell,m)}$ of experts defined by

$$\mathbf{h}^{(k,\ell,m)} = (1-\lambda_m)\mathbf{h}^{(k,\ell)} + \lambda_m \mathbf{b}_0 \ ,$$

where $\lambda_m \in (0, 1)$ is a sequence of numbers converging to zero and $\mathbf{b}_0$ is the uniform portfolio $(1/d, \ldots, 1/d)$.

3 Kernel-based investment strategies

In this section we introduce a class of kernel-based investment strategies and prove their universality. Kernel-based rules allow a more flexible way of extracting information from the history of the market. The family of methods introduced here is similar, in spirit, to the histogram-based strategy described in the previous section. The main difference is that the elementary strategies used by the strategy replace the rigid discretization of the past few market vectors by a more flexible "moving-window" rule. By appropriate weighing by a kernel function a whole rich family of strategies is obtained. The main theoretical result of this section is the universality of these strategies under general assumptions. The numerical results shown in Section 4 indicate the practical superiority of kernel-based methods.

To simplify notation we start with the simplest "moving-window" version, corresponding to a uniform kernel function, and treat the general case briefly later.

The kernel-based strategy $\mathbf{B}^K$ is constructed similarly to the histogram-based portfolio $\mathbf{B}^H$ described in the previous section. Just like before, we start by defining an infinite array of experts $\mathbf{H}^{(k,\ell)} = \{\mathbf{h}^{(k,\ell)}(\cdot)\}$, where k, ℓ are positive integers. To define $\mathbf{H}^{(k,\ell)}$, let $c > 0$ be a constant possibly depending on k and d. For fixed positive integers k, ℓ and for each vector $\mathbf{s} = \mathbf{s}_{-k}^{-1}$ of dimension kd define the portfolio vector, for $n > k+1$,

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

if the product is non-void, and $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise. If the product is non-void then we may re-write this definition as

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

Finally, we define the expert $\mathbf{h}^{(k,\ell)}$ by

$$\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}, \mathbf{x}_{n-k}^{n-1}), \qquad n = 1, 2, \ldots \tag{2}$$

That is, $\mathbf{h}_n^{(k,\ell)}$ discretizes the sequence $\mathbf{x}_1^{n-1}$, and browses through all past approximate appearances of the last seen vector $\mathbf{x}_{n-k}^{n-1}$. Then it designs a fixed portfolio vector according to the returns on the periods following these approximate appearances.

These experts are mixed the same way as in the case of the histogram-based strategy. That is, 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, $q_{k,\ell} > 0$. The strategy $\mathbf{B}^K$ weighs the experts $\mathbf{H}^{(k,\ell)}$ according to their past performances and $\{q_{k,\ell}\}$ by

$$\mathbf{b}(\mathbf{x}_1^{n-1}) = \frac{\sum_{k,\ell} q_{kl} S_{n-1}(\mathbf{H}^{(k,\ell)}) \mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1})}{\sum_{k,\ell} q_{kl} S_{n-1}(\mathbf{H}^{(k,\ell)})},$$

where $S_n(\mathbf{H}^{(k,\ell)})$ is the capital accumulated by the elementary strategy $\mathbf{H}^{(k,\ell)}$ after n periods when starting with an initial capital S_0. Thus, after period n, the investor's capital becomes

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

The main result of this section, whose proof is given in Section 5 below, states the universality of the scheme defined above:

Theorem 1 *The portfolio scheme $\mathbf{B}^K$ is universal with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.*

Remark. The assumption of the finiteness of the $\mathbb{E}\{|\log X^{(j)}|\}$ may be weakened similarly as in the case of the histogram-based strategy described in the previous section.

Remark. (PARAMETERS.) For the universality of the method, it suffices to assume that the initial weights $q_{k,\ell}$ are strictly positive. However, in practice, for good finite-time behavior, the role of these weights is important. For good practical performance, $q_{k,\ell}$, as well as other parameters such as the constant c have to be fine tuned. Some possible choices are given in Section 4.

Remark. (TRANSACTION COSTS.) As mentioned in the introduction, a main simplifying (and unrealistic) assumption in our analysis is that transaction costs are ignored. It follows from a result of Blum and Kalai [5] that if the market process $\mathbf{X}_1, \mathbf{X}_2, \ldots$ is a sequence of independent and identically distributed vectors then there exists an investment strategy whose asymptotic rate of growth equals W^*. However, it is easy to see that, in general, if the market is stationary but not necessarily i.i.d. then achieving the optimal rate of growth in the presence of transaction costs is impossible. To see this, consider the simple markovian example in which $d = 2$ and, deterministically, when n is odd, $X_n^{(1)} = 0$ and $X_n^{(2)} = 1$ and when n is even, $X_n^{(1)} = 1$ and $X_n^{(2)} = 0$. In this case clearly $W^* = 0$, but to achieve it, the total wealth has to be moved each in period from one asset to another, and transaction costs force a rate of growth bounded away from zero (from below). However, in practice, simple modifications of the kernel-based strategies may work quite well even when transaction costs are present. Some numerical examples are offered in Section 4.

Remark. (VALIDITY OS ASSUMPTIONS.) The theoretical results assume little more than stationarity and ergodicity of the market. Obviously, there is no empirical test to decide whether a market satisfies these properties or not. The practical usefulness of these assumptions should be judged based on the numerical results the investment strategies lead to. In the next section we describe various such results based on past data. These results suggest that the market can be modeled effectively by a low-order stationary Markov process. This is evidenced by the good behavior of some experts that operate on such an assumption. We emphasize again that these numerical results ignore the effect using such a strategy may have on the market.

Remark. (VOLATILITY.) In this paper we completely ignore the issue of volatility and focus on almost sure convergence of the growth rate $(1/n) \log S_n$. Controlling the volatility of the process is obviously a relevant and nontrivial problem. Once again we refer the reader to the numerical results of the next section that suggest that the achieved wealth, in fact, has a low volatility. However, we do not have any theoretical guarantees.

Next we describe a class of general kernel-based investment strategies. These strategies are based on a sequence of *kernel* function $K_k : \mathbb{R}_+^{kd} \to \mathbb{R}_+$. The definition of a generalized kernel-based strategy parallels that of $\mathbf{B}^K$ defined above with the only difference that in the defining equation (2) of the elementary strategies $\mathbf{H}^{(k,\ell)}$, the portfolio vector $\mathbf{b}^{(k,\ell)}(x_1^{n-1}, \mathbf{s})$ is defined by

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}, \mathbf{s}) = \arg\max_{\mathbf{b}\in\Delta_d} \prod_{k<i<n} \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle^{\frac{w_i^{(k,\ell)}}{\sum_{k<j<n} w_j^{(k,\ell)}}} ,$$

where the weights $w_i^{(k,\ell)}$ are defined by

$$w_i^{(k,\ell)} = K_k\left(\ell(\mathbf{x}_{i-k}^{i-1} - \mathbf{s})\right)$$

and $0/0$ is understood as 0.

Observe that if K_k is the uniform (or moving window) kernel $K_k(\mathbf{x}) = \mathbb{I}_{\|\mathbf{x}\|\le c}$ $(\mathbf{x} \in \mathbb{R}_+^{kd})$ then we recover the definition of the strategy $\mathbf{B}^K$ introduced above. Typical nonuniform kernels assign a smaller weight to those $\mathbf{x}_i$ for which the distance of $\mathbf{x}_{i-k}^{i-1}$ from $\mathbf{s}$ is larger. Such kernels promise a better prediction of the local structure of the conditional distribution. The next result extends Theorem 1 to a class of general kernels. The proof is given in Section 5.

Theorem 2 *Assume that for each* $k = 1, 2, \ldots$ *the kernel* K_k *is such that there exists a non-increasing function* ϕ_k *defined on* $\mathbb{R}_+$ *with* $\phi_k(+0) > 0$ *and* $\lim_{t\to\infty} t^d \phi_k(t) = 0$ *such that for some constants* $c_1, c_2 > 0$, *for all* $\mathbf{x} \in \mathbb{R}_+^{kd}$,

$$c_1\phi_k(\|\mathbf{x}\|) \le K_k(\mathbf{x}) \le c_2\phi_k(\|\mathbf{x}\|) \ .$$

Then the kernel-based portfolio scheme defined above is universal with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, *for* $j = 1, 2, \ldots d$.

4 Practical implementation and numerical results

The purpose of this section is to discuss some issues of the practical implementation of the investment strategies $\mathbf{B}^H$ and $\mathbf{B}^K$ described in the previous sections and to report numerical results of the application of the algorithms to real financial data.

Both strategies have in common that they use an infinite array of experts. In practice, one chooses two positive integers K and L and replace the infinite array of elementary strategies by a finite array of KL experts $\mathbf{H}^{(k,\ell)}$, $k = 1, \ldots, K$, $\ell = 1, \ldots, L$ defined by (1) and (2), for both strategies. Recall that k is the length of the recent market history matched by data in the past and ℓ indexes the fineness of the discretization scheme in use, usually finer as ℓ increase. We also include, as an additional expert, with index $k = \ell = 0$, the strategy that uses the full history to calculate the portfolio by

$$\mathbf{h}^{(0,0)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b} \in \Delta_d} \prod_{0<i<n} \langle \mathbf{b} \,, \mathbf{x}_i \rangle \,, \qquad n > 1 \,.$$

In all cases reported below we used the uniform distribution $\{q_{k,\ell}\} = 1/(KL+1)$ over the experts in use.

The next table gives a schematic description of the algorithm implementing the strategies described in Sections 2 and 3.

Given $\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} \in \mathrm{R}^d$, to compute the portfolio for the n-th trading period,

1. for each expert (k, ℓ), $k = 1..K$, $\ell = 1..L$ do

1.1 `ComputeHistory(k,l)`: collect the data from those training periods in the past that followed a k-period string similar to $\mathbf{x}_{n-k}, \ldots, \mathbf{x}_{n-1}$ and place these periods in the history list. What "similar" means depends on whether we use a histogram-based strategy (see below) or a kernel-based strategy, where all past periods are weighted as described in the previous section.

1.2 `MaximizeOverHistory(k,l)`: find the portfolio $h_{k,\ell}$ that maximizes wealth for the empirical distribution of the data collected in the history list;

2. `CombinePortfolios`: weighting the experts with the wealth achieved so far, and a "prior" probability distribution $q(k, \ell)$, obtain a portfolio b_n to invest in the current period n.

Then, using the newly acquired data $\mathbf{x}_n$,

3. update the wealth for each expert and the current actual wealth;

4. store $\mathbf{x}_n$ for use in the next period, discretizing it in the histogram case.

To describe precisely the histogram-based strategy $\mathbf{B}^H$ used in our experiments, we need to define the cells of the partitions $\mathcal{P}_\ell$ determining the experts. Since typical values of the price relatives $x_n^{(j)}$ concentrate around 1, we used the following scheme. Given $\ell \in 1, \ldots, L$, in each dimension we use $M = 2 + 2\ell$ cells. For $x \in \mathbb{R}$, the index q of its cell is computed as follows. Define $a = 1/2(1 + 2\log_{10} \ell)$ and $w = a^{-1/\ell}$. Then

$$q = \begin{cases} 0 & \text{if } x \le a \\ 1 + \left\lfloor \frac{\log(x/a)}{\log w} \right\rfloor & \text{if } x \in (a, 1] \\ \ell + 1 + \left\lfloor \frac{\log(ax) + \ell \log w}{\log w} \right\rfloor & \text{if } x \in (1, 1/a] \\ 2\ell + 1 & \text{if } x > 1/a \end{cases}$$

Thus, the cell boundaries are $0, a, aw, \ldots, aw^{\ell}, 1, a^{-1}w^{-\ell}, \ldots, a^{-1}, \infty$, giving a variable-width grid that becomes finer close to 1. Then for any $\mathbf{x} \in R^d$, $G_\ell(\mathbf{x})$, is the vector of integers $\{q^{(j)}\}_{j=1,\ldots,d}$.

To implement the kernel-based investment strategy $\mathbf{B}^K$ described in Section 3, one needs to choose the kernel function K. In the experiments reported here we used the simple "moving-window" kernel $K(\mathbf{x}) = \mathbb{I}_{\|\mathbf{x}\| \le c_k}$ where $\|\cdot\|$ is the euclidean norm and tried different choices of the constants c_k, all of the form $c_k = ckd$, for different values for c. We denote the kernel-based strategy with the moving-window kernel and constants $c_k = ckd$ by $\mathbf{B}^K(c)$.

To find the portfolio that solves the maximization problem in (1) or (2) we use the routines DONLP2 of Spellucci [23].

Remark. (COMPUTATIONAL COST.) To give an idea of the computational cost of the proposed algorithms, running the experimental study using the uniform kernel on the full NYSE dataset described below took about 12 hours on a Xeon 2.00 GHz based computer. This means that to compute the portfolio of 36 assets for a single period, about 8 seconds are needed on average. Of course, real-time implementation of these investment strategies would require storage of discretized data and values of performance of the experts used, but the extra computational cost of reading these stored data is negligible.

Numerical results

We tested the investment strategies on two different sets of financial data. One of these is a standard set of New York Stock Exchange data used by Cover [10], Singer [22], Hembold, Schapire, Singer, and Warmuth [18], Blum and Kalai [5], Borodin, El-Yaniv, and Gogan [6], and others. The other is exchange rate data between US$ and eight other currencies.

The NYSE data set includes daily prices of 36 assets along a 22-year period (5651 trading days) ending in 1985. This means that $d = 36$. Because of this large dimensionality, our current implementation cannot handle the histogram strategy. Table 1 summarizes the wealth achieved by the kernel-based strategy for three different choices of the constant c. In all cases, we use $K = 5, L = 10$. For the sake of comparison, we also indicate the wealth achieved by the best constantly rebalanced portfolio (BCRP). (Note that this "anticipating" portfolio does not correspond to any valid investment strategy since the BRCP can only be determined in hindsight.) The most important feature is that after the whole 22-year period some versions of $\mathbf{B}^K$ multiply their initial wealth by a factor of more than five-hundred million. A closer inspection of the results reveal that there is a small number of elementary strategies responsible for this spectacular growth. This demonstrates how $\mathbf{B}^K$ is able to exploit effectively hidden dependences that are difficult to reveal otherwise. It is interesting to note that in the second half of the period the growth is significantly faster than in the first. This may be due to the fact that in the initial "learning" phase not enough data has been collected to discover the significant tendencies.

We also tested the discussed investment strategies on data obtained from Datastream (a commercial database) about the exchange rate to US$ of several currencies. In particular, we got daily variations, with respect to the US$, from March 25, 1988 to March 27, 2003, a total of $N = 3914$ periods, of the eight currencies listed in Table 2. The table also lists the final value of one initial US$ invested in each currency and the minimum, 25th percentile, median, 75th percentile and maximum of each series.

The achieved wealth of the histogram-, and kernel-based strategies are listed in Table 3. The numbers show the wealth achieved, in US dollars, after an initial investment of 1 US$ uniformly divided among all the currencies included in the data set (i.e., $d = 8$) and then running the strategies along the full period range. In the histogram case we use $K = 3, L = 5$ while for the kernel-based strategy we use the setting previously described. The growth of wealth for $\mathbf{B}^K(0.1)$ during the whole period is shown in Figure

After period	BCRP	$\mathbf{B}^{\mathrm{K}}(2)$	$\mathbf{B}^{\mathrm{K}}(1.00)$	$\mathbf{B}^{\mathrm{K}}(0.5)$
500	13.07	4.539	4.265	2.759
1000	7.324	3.894	5.113	4.402
1500	16.03	7.621	9.805	7.909
2000	10.21	7.052	7.535	6.901
2500	17.48	39.35	40.08	34.87
3000	18.81	321.7	853.3	505.0
3500	34.57	2876	2.231e+4	1.641e+4
4000	55.52	4.7974e+4	8.968e+5	5.531e+5
4500	106.8	2.5802e+5	5.447e+6	3.116e+6
5000	125.4	9.035e+5	4.030e+7	2.083e+7
5500	267.8	5.662e+6	4.725e+08	2.103e+8
5651	250.6	7.037e+6	5.627e+08	2.633e+8

Table 1: Wealth achieved by various versions of the kernel-based strategy $\mathbf{B}^{\mathrm{K}}$. In all cases one unit is invested in the first period uniformly in all 36 stocks included in our NYSE data set. $\mathbf{B}^{\mathrm{K}}(c)$ is the kernel strategy with constant c and BCRP is the best constantly rebalanced portfolio.

Currency	Final	Min	p_{25}	Median	p_{75}	Max
Singapoore Dollar	1.138	-0.0247	-0.0012	0.000	0.0013	0.0379
Norwegian Crown	0.856	-0.0529	-0.00345	0.000	0.0037	0.049
Swiss Franc	1.007	-0.048	-0.0044	0.000	0.0044	0.054
ECU/Euro	0.867	-0.050	-0.0035	0.000	0.0035	0.034
Israeli Shekel	0.3323	-0.1156	-0.0017	0.000	0.0015	0.1221
Indian Rupee	0.2743	-0.0869	-0.0003	0.000	0.0002	0.0667
Canadian Dollar	0.849	-0.017	-0.0018	0.000	0.0018	0.0190
British Pound	0.8504	-0.0420	-0.0031	0.000	0.0033	0.0431

Table 2: Some descriptive statistics about the exchange rate data used. Second column lists wealth in US$ achieved by investing one US$ in the first period in the corresponding currency. The rest of the columns show the minimum, 25th percentile, median, 75th percentile and maximum of each series of returns along the full range of periods ($1 \leq n \leq 3914$).

4. Even though the results here are not as spectacular as in the case of the NYSE data, after an initial learning period of about 1000 days, the kernel-based portfolio clearly outperforms the best currency, the best constantly rebalanced portfolio, and the histogram-based strategy.

To relieve the computational burden, we tested a variant of the discussed strategies which works as follows. The strategy distributes the initial wealth evenly among all $\binom{d}{2}$ pairs of assets. Then for each pair, the histogram (or kernel) based strategy is used independently. The first row of Table 4 lists the wealth achieved by this strategy using all $\binom{8}{2} = 28$ pairs of currencies of the exchange rate data by the different methods. The second row reports a version in which instead of pairs, all $\binom{8}{3} = 56$ triples of currencies are used. The third row corresponds to investing one unit divided among all $\binom{36}{2} = 630$ possible pairs of stocks in the NYSE data set. We see that, even though no theoretical guarantee can be given for the universality of these variants, the numerical performance of these simplified methods does not deteriorate significantly (it even improves in the case of the NYSE data).

In Table 5 we compare the wealth achieved by the strategies discussed

After period	BCRP	$\mathbf{B}^{\mathrm{H}}$	$\mathbf{B}^{\mathrm{K}}(0.3)$	$\mathbf{B}^{\mathrm{K}}(0.1)$	$\mathbf{B}^{\mathrm{K}}(0.05)$
500	1.078	1.112	1.060	1.145	1.085
1000	1.227	1.445	1.788	1.971	1.532
1500	1.262	3.234	5.586	8.698	5.219
2000	1.425	6.553	12.81	23.01	13.16
2500	1.273	9.009	15.21	32.99	18.53
3000	1.178	11.16	22.22	51.91	28.91
3500	1.152	20.04	43.64	111.8	60.89
3914	1.138	35.47	83.52	229.7	124.28

Table 3: Wealth achieved by various investment strategies for the currency exchange data. In all cases one US$ is invested in the first period uniformly in all 8 currencies described in Table 2. BCRP is the constant rebalanced portfolio, $\mathbf{B}^{\mathrm{H}}$ is the histogram-besed strategy, and $\mathbf{B}^{\mathrm{K}}(c)$ is the kernel strategy.

	$\mathbf{B}^{\mathrm{H}}$	$\mathbf{B}^{\mathrm{K}}(0.01)$	$\mathbf{B}^{\mathrm{K}}(0.05)$	$\mathbf{B}^{\mathrm{K}}(0.5)$
All EXCHR pairs	14.25	5.393	12.65	0.8832
All EXCHR triples	21.59	7.095	26.77	1.270
All NYSE pairs	–	–	4.31e8	1.285e10

Table 4: Wealth achieved by investing one unit divided among all the possible pairs or triples for the exchange rate (EXCHR) and NYSE data.

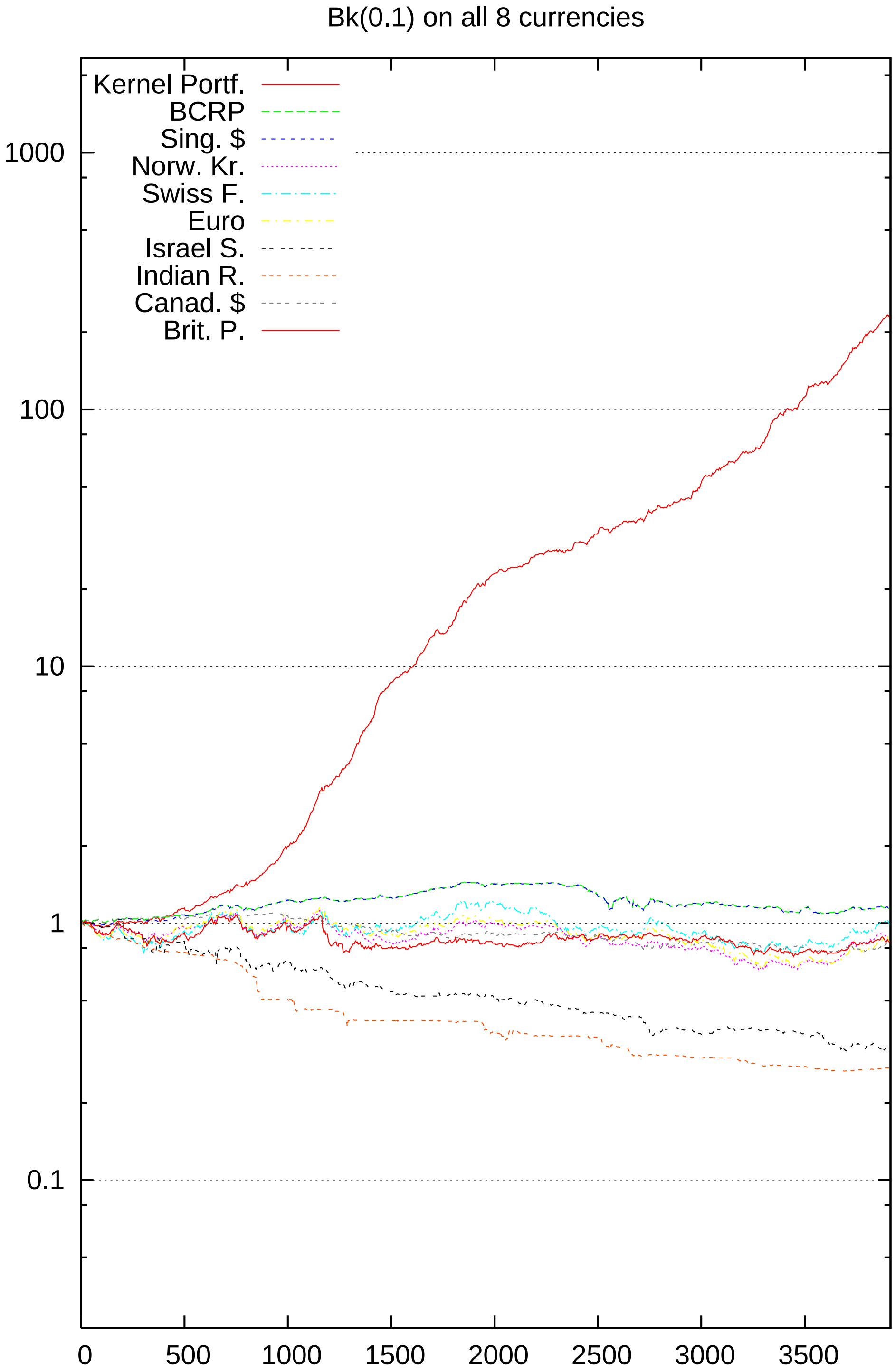

Figure 1: Wealth achieved along $N = 3925$ daily periods by investing one US$ in several currencies, and by using the kernel-based strategy computed with $d = 8$ currencies, like in column five of Table 3. Horizontal axis is time period number, vertical is the wealth achieved, in logarithmic scale.

here to other methods found in the literature. We report the wealth achieved by different strategies for the pairs of NYSE stocks used by Cover [10] (to test his universal portfolio) and by Singer [22] (for his "switching portfolios"). As a reference, we also list the wealth of some other strategies only computable with hindsight. $\mathbf{B}^H$ and $\mathbf{B}^K$ clearly outperform both Cover's universal portfolio and Singer's switching portfolios. It is also interesting to note that the presence of the stock Kin Ark makes the wealth of these strategies explode. This is interesting, since the overall growth of Kin Ark in the reported period is quite modest. The reason is that somehow the variations of the price relatives of this asset turn out to be well predictable by at least one expert and that suffices to produce this explosive growth. Indeed, the presence of this single stock is largely responsible for the wealth reported in Table 1. Removing this single stock from the portfolio, the achieved wealth of $\mathbf{B}^K(1.0)$ reduces to a much more modest value of 753.76 (which still corresponds to an annual rate of increase of about 135%).

Finally, we briefly present some results on the performance of these strategies in the presence of transaction costs. It is not straightforward to adapt our methods in an optimal way when transactions costs have to be paid. More precisely, assume that a fixed percentage commission $r \in (0, 1)$ has to be payed at each transaction. The results reported here are very likely improvable and correspond to the simplistic method in which each expert is weighed by the wealth achieved in presence of transaction costs, and use the resulting portfolio. Namely, let $S_n^r(\mathbf{H}^{(k,\ell)})$ be the wealth achieved by expert (k, ℓ) after period n. (This may be computed using an optimal rebalancing strategy, see Blum and Kalai [5]). Then, the portfolio is calculated by

$$\mathbf{b}(\mathbf{x}_1^{n-1}) = \frac{\sum_{k,\ell} q_{kl} S_{n-1}^r(\mathbf{H}^{(k,\ell)}) \mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1})}{\sum_{k,\ell} q_{kl} S_{n-1}^r(\mathbf{H}^{(k,\ell)})},$$

and the wealth achieved by the strategy $\mathbf{B}$ becomes

$$S_n^r(\mathbf{B}) = S_0 \prod_{i=1}^{n} \langle \mathbf{b}_i\, , \, \mathbf{x}_i \rangle \, \alpha_r(\mathbf{b}_{i-1}, \mathbf{b}_i)$$

Stocks					Best Exp. $[k, \ell]$
Iroquois	Best asset	8.92	$\mathbf{B}^{H}$	2.3e+10	1.395e+11 [1,1]
Kin Ark	BCRP	73.70	$\mathbf{B}^{K}$	2.109e+04	1.087e+06 [1,1]
	Oracle	6.85e+53		4.038e+10	9.014e+11 [2,2]
	Cover UP	39.97		2.187e+10	9.014e+11 [2,1]
	Singer SAP	143.7		7.401e+10	9.014e+11 [2,5]
Com. Met.	Best asset	52.02	$\mathbf{B}^{H}$	162.5	327.8 [2,1]
Mei. Corp	BCRP	103.0	$\mathbf{B}^{K}$	96.9	433.3 [1,2]
	Oracle	2.12e+35		775.1	4749. [2,5]
	Cover UP	74.08		373.8	4613 [4,1]
	Singer SAP	107.7		682.3	4613 [4,5]
Com. Met.	Best asset	52.02	$\mathbf{B}^{H}$	1.331e+10	8.544e+10 [1,1]
Kin Ark	BCRP	144.0	$\mathbf{B}^{K}$	1.52e+07	7.847e+08 [1,1]
	Oracle	1.84e+49		1.111e+11	1.411e+12 [3,3]
	Cover UP	80.54		5.395e+10	1.411e+12 [3,]1
	Singer SAP	206.7		2.551e+11	2.065e+12 [2,8]
IBM	Best asset	13.36	$\mathbf{B}^{H}$	63.87	112.2 [1,5]
Coca-Cola	BCRP	15.02	$\mathbf{B}^{K}$	18.92	86.1 [1,1]
	Oracle	1.08e+15		47.6	194.6 [1,6]
	Cover UP	14.24		46.46	194.6 [1,6]
	Singer SAP	15.05		18.11	60.56 [3,10]

Table 5: Wealth achieved by different strategies by investing in the pairs of NYSE stocks used in Cover [10]. In the second column we show the wealth achieved by the best stock of the two involved, by the best constantly rebalanced portfolio, by an oracle (defined as the best possible strategy which invests all the capital in the best stock each day), and the results reported in the literature for Cover's universal portfolio (UP) and Singer's switching adaptive portfolio (SAP). The third column lists our results for the histogram ($\mathbf{B}^{H}$) and kernel ($\mathbf{B}^{K}$) portfolios. In all cases we take $K = 5, L = 10$, and $c = 0.01, 0.05, 0.1, 0.5$ for $\mathbf{B}^{K}$. The last column lists the wealth and the index of the best expert among the $KL + 1$ competing experts.

where $\alpha_r(\mathbf{b}, \mathbf{b}')$ denotes the wealth loss ratio due to the transaction cost c when rebalancing the portfolio $\mathbf{b}$ to $\mathbf{b}'$.

We applied this simple approach to the NYSE data for several values of the transaction cost r. It is shown in Figure 2 that the wealth reduction is important but still gives a good result for reasonable values of the cost r.

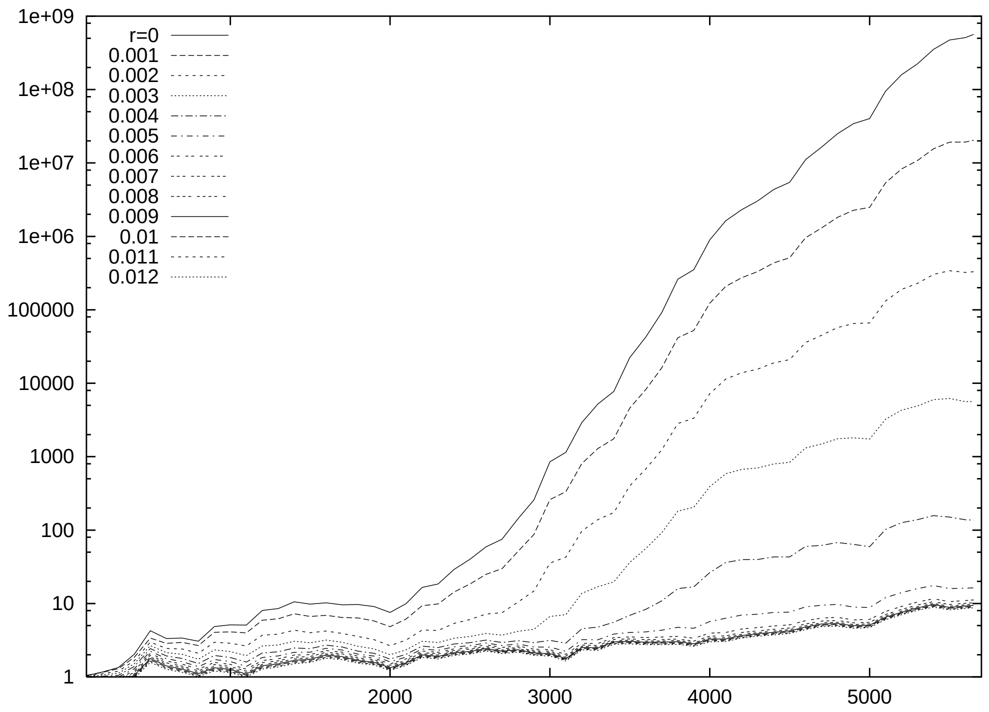

Figure 2: Wealth achieved by investing one unit uniformly in the 36 NYSE stocks and using the kernel strategy (with constant 1.0) for several values of the transaction costs $r = 0, 0.001, 0.002, \ldots, 0.012$.

5 Proofs

The proof of Theorem 1 uses the following three auxiliary results. The first is known as Breiman's generalized ergodic theorem [7], see also Algoet [2].

Lemma 1 (BREIMAN [7]). *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 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*

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

The next two lemmas are due to Algoet and Cover [3, Theorems 3 and 4].

Lemma 2 (ALGOET AND COVER [3]). *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 $\mathbb{E}\{|\log U_n^{(j)}|\} < \infty$ 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 log-optimal portfolios with respect to $\mathbf{Q}_n$, that is, the set of all portfolios $\mathbf{b}$ that attain $\max_{\mathbf{b}\in\Delta_d} \mathbb{E}\{\log\langle \mathbf{b}\,,\,\mathbf{U}_n\rangle\}$. 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 $\mathbf{u}$,

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

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

Lemma 3 (ALGOET AND COVER [3]). *Let $\mathbf{X}$ be a random market vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying $\mathbb{E}\{|\log X^{(j)}|\} < \infty$. 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

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

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 1. The proof is based on techniques used in related prediction problems, see Györfi, Lugosi, and Morvai [16], Györfi and Lugosi [15], Györfi and Schäfer [17]. We need to prove that

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

Without loss of generality we may assume $S_0 = 1$, so that

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

Thus

$$\begin{aligned}
\liminf_{n\to\infty} W_n(\mathbf{B}) &\geq \liminf_{n\to\infty}\sup_{k,\ell}\left(W_n(\mathbf{H}^{(k,\ell)}) + \frac{\log q_{k,\ell}}{n}\right) \\
&\geq \sup_{k,\ell}\liminf_{n\to\infty}\left(W_n(\mathbf{H}^{(k,\ell)}) + \frac{\log q_{k,\ell}}{n}\right) \\
&= \sup_{k,\ell}\liminf_{n\to\infty} W_n(\mathbf{H}^{(k,\ell)}). \qquad (3)
\end{aligned}$$

The simple argument above shows that the asymptotic rate of growth of the strategy $\mathbf{B}$ is at least as large as the supremum of the rates of growth of all elementary strategies $\mathbf{H}^{(k,\ell)}$. Thus, to estimate $\liminf_{n\to\infty} W_n(\mathbf{B})$, it suffices to investigate the performance of expert $\mathbf{H}^{(k,\ell)}$ on the stationary and ergodic market sequence $\mathbf{X}_0, \mathbf{X}_{-1}, \mathbf{X}_{-2}, \ldots$. First 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 $\{\mathbf{X}_i : 1-j+k \leq i \leq 0, \|\mathbf{X}_{i-k}^{i-1} - \mathbf{s}\| \leq c/\ell\}$ defined by

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

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 $\mathbb{P}_{j,s}^{(k,\ell)} = \delta_{(1,\dots,1)}$ be the probability measure concentrated on the vector $(1,\dots,1)$. In other words, $\mathbb{P}_{j,s}^{(k,\ell)}(A)$ is the relative frequency of the the vectors among $\mathbf{X}_{1-j+k},\dots,\mathbf{X}_0$ which fall in the set A.

Observe that for all $\mathbf{s}$, with probability one,

$$\mathbb{P}_{j,s}^{(k,\ell)} \to \begin{cases} \mathbb{P}_{\mathbf{X}_0 | \|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell} & \text{if } \mathbb{P}(\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell) > 0, \\ \delta_{(1,\dots,1)} & \text{if } \mathbb{P}(\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell) = 0 \end{cases} \tag{4}$$

weakly as $j \to \infty$ where $\mathbb{P}_{\mathbf{X}_0 | \|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell}$ denotes the distribution of the vector $\mathbf{X}_0$ conditioned on the event $\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell$. To see this, let f be a bounded continuous function defined on $\mathbb{R}_+^d$. Then the ergodic theorem implies that

$$\begin{aligned} \int f(\mathbf{x})\mathbb{P}_{j,s}^{(k,\ell)}(d\mathbf{x}) &= \frac{\frac{1}{|1-j+k|}\sum_{i:1-j+k\le i\le 0, \|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\|\le c/\ell} f(\mathbf{X}_i)}{\frac{1}{|1-j+k|}\left|\{i : 1-j+k \le i \le 0, \|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\| \le c/\ell\}\right|} \\ &\to \frac{\mathbb{E}\{f(\mathbf{X}_0)\mathbb{I}_{\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le c/\ell\}}\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell\}} \\ &= \mathbb{E}\left\{ f(\mathbf{X}_0) \,\middle|\, \|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell \right\} \\ &= \int f(\mathbf{x})\mathbb{P}_{\mathbf{X}_0 | \|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell}(d\mathbf{x}) \quad \text{almost surely, as } j \to \infty \end{aligned}$$

if $\mathbb{P}(\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell) > 0$. On the other hand, if $\mathbb{P}(\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le c/\ell) = 0$, then with probability one $\mathbb{P}_{j,s}^{(k,\ell)}$ is concentrated on $(1,\dots,1)$ for all j, and

$$\int f(\mathbf{x})\mathbb{P}_{j,s}^{(k,\ell)}(d\mathbf{x}) = f(1,\dots,1) .$$

Denote the limit distribution of $\mathbb{P}_{j,s}^{(k,\ell)}$ by $\mathbb{P}_s^{*(k,\ell)}$.

Recall that by definition, $\mathbf{b}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{s})$ is a log-optimal portfolio with respect to the probability measure $\mathbb{P}_{j,s}^{(k,\ell)}$. Let $\mathbf{b}_{k,\ell}^*(\mathbf{s})$ denote a log-optimal portfolio with respect to the limit distribution $\mathbb{P}_s^{*(k,\ell)}$. Then, using Lemma 2, we infer from (4) that, as j tends to infinity, we have the almost sure convergence

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

for $\mathbb{P}_s^{*(k,\ell)}$-almost all x_0 and hence for $\mathbb{P}_{X_0}$-almost all x_0. Since s was arbitrary, we obtain

$$\lim_{j\to\infty} \left\langle \mathbf{b}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1}, \mathbf{X}_{-k}^{-1}) , \mathbf{x}_0 \right\rangle = \left\langle b_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1}) , \mathbf{x}_0 \right\rangle \quad \text{almost surely.} \tag{5}$$

Next we apply Lemma 1 for the function

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

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

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

which has finite expectation, and

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

by (5). As $n \to \infty$, Lemma 1 yields

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

Therefore, by (3) we have

$$\liminf_{n\to\infty} W_n(\mathbf{B}) \ge \sup_{k,\ell} \epsilon_{k,\ell} \ge \sup_{k} \liminf_{\ell} \epsilon_{k,\ell} \qquad \text{almost surely}$$

and it suffices to show that the right-hand side is at least W^*.

To this end, define, for Borel sets $A, B \subset \mathbb{R}_+^d$,

$$m_A(z) = \mathbb{P}\{\mathbf{X}_0 \in A | \mathbf{X}_{-k}^{-1} = z\}$$

and

$$\mu_k(B) = \mathbb{P}\{\mathbf{X}_{-k}^{-1} \in B\}.$$

Then for any $\mathbf{s} \in \text{support}(\mu_k)$, and for all A,

$$\begin{aligned}
\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}(A) &= \mathbb{P}\left\{\mathbf{X}_0 \in A \,\Big|\, \|\mathbf{X}_{-k}^{-1} - \mathbf{s}\| \le c/\ell\right\} \\
&= \frac{\mathbb{P}\{\mathbf{X}_0 \in A, \|\mathbf{X}_{-k}^{-1} - \mathbf{s}\| \le c/\ell\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1} - \mathbf{s}\| \le c/\ell\}} \\
&= \frac{1}{\mu_k(S_{\mathbf{s},c/\ell})} \int_{S_{\mathbf{s},c/\ell}} m_A(z)\mu_k(dz) \\
&\to m_A(\mathbf{s}) = \mathbb{P}\{\mathbf{X}_0 \in A | \mathbf{X}_{-k}^{-1} = \mathbf{s}\}
\end{aligned}$$

as $\ell \to \infty$ and for μ_k-almost all $\mathbf{s}$ by the Lebesgue density theorem (see [14, Lemma 24.5]), and therefore

$$\mathbb{P}_{\mathbf{X}_{-k}^{-1}}^{*(k,\ell)}(A) \to \mathbb{P}\{\mathbf{X}_0 \in A | \mathbf{X}_{-k}^{-1}\}$$

as $\ell \to \infty$ for all A. Thus, using Lemma 2 again, we have

$$\begin{aligned}
\liminf_{\ell} \epsilon_{k,\ell} &= \lim_{\ell} \epsilon_{k,\ell} \\
&= \mathbb{E}\left\{\log \left\langle \mathbf{b}_k^*(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0 \right\rangle\right\} \\
&\qquad \text{(where } \mathbf{b}_k^*(\cdot) \text{ is the log-optimum portfolio with respect} \\
&\qquad \text{to the conditional probability } \mathbb{P}\{X_0 \in A | \mathbf{X}_{-k}^{-1}\}) \\
&= \mathbb{E}\left\{\mathbb{E}\left\{ \log \left\langle \mathbf{b}_k^*(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0 \right\rangle \Big| \mathbf{X}_{-k}^{-1}\right\}\right\} \\
&= \mathbb{E}\left\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \log \left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0 \right\rangle \Big| \mathbf{X}_{-k}^{-1}\right\}\right\} \\
&\stackrel{\text{def}}{=} \epsilon_k^* .
\end{aligned}$$

To finish the proof we appeal to the sub-martingale convergence theorem. First note that the sequence

$$Y_k \stackrel{\text{def}}{=} \mathbb{E}\left\{ \log \left\langle \mathbf{b}_k^*(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0 \right\rangle \Big| \mathbf{X}_{-k}^{-1}\right\} = \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \log \left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0 \right\rangle \Big| \mathbf{X}_{-k}^{-1}\right\}$$

of random variables forms a sub-martingale, that is, $\mathbb{E}\left\{Y_{k+1}|X_{-k}^{-1}\right\} \geq Y_k$. To see this, note that

$$
\begin{aligned}
\mathbb{E}\left\{Y_{k+1}|X_{-k}^{-1}\right\} &= \mathbb{E}\left\{\mathbb{E}\left\{\log\left\langle \mathbf{b}_{k+1}^{*}(\mathbf{X}_{-k-1}^{-1}), \mathbf{X}_0\right\rangle \Big| \mathbf{X}_{-k-1}^{-1}\right\}\Big| \mathbf{X}_{-k}^{-1}\right\} \\
&\geq \mathbb{E}\left\{\mathbb{E}\left\{\log\left\langle \mathbf{b}_{k}^{*}(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0\right\rangle \Big| \mathbf{X}_{-k-1}^{-1}\right\}\Big| \mathbf{X}_{-k}^{-1}\right\} \\
&= \mathbb{E}\left\{\log\left\langle \mathbf{b}_{k}^{*}(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0\right\rangle \Big| \mathbf{X}_{-k-1}^{-1}\right\} \\
&= Y_k \, .
\end{aligned}
$$

This sequence is bounded by

$$
\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{\log\left\langle \mathbf{b}(\mathbf{X}_{-\infty}^{-1}), \mathbf{X}_0\right\rangle \Big| \mathbf{X}_{-\infty}^{-1}\right\}
$$

which has a finite expectation. The sub-martingale convergence theorem (see, e.g., Stout [25]) implies that this sub-martingale is convergent almost surely, and $\sup_k \epsilon_k^*$ is finite. In particular, by the submartingale property, ϵ_k^* is a bounded increasing sequence, so that

$$
\sup_k \epsilon_k^* = \lim_{k\to\infty} \epsilon_k^* \, .
$$

Applying Lemma 3 with the σ-algebras

$$
\sigma\left(\mathbf{X}_{-k}^{-1}\right) \nearrow \sigma\left(\mathbf{X}_{-\infty}^{-1}\right)
$$

yields

$$
\begin{aligned}
\sup_k \epsilon_k^* &= \lim_{k\to\infty} \mathbb{E}\left\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{\log\left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1}), \mathbf{X}_0\right\rangle \Big| \mathbf{X}_{-k}^{-1}\right\}\right\} \\
&= \mathbb{E}\left\{\max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{\log\left\langle \mathbf{b}(\mathbf{X}_{-\infty}^{-1}), \mathbf{X}_0\right\rangle \Big| \mathbf{X}_{-\infty}^{-1}\right\}\right\} \\
&= W^*
\end{aligned}
$$

and the proof of the theorem is finished. ∎

Sketch of proof of Theorem 2. The proof parallels that of Theorem 1 so we only indicate the differences.

The definition of the random measure $\mathbb{P}_{j,s}^{(k,\ell)}$ is now changed to

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

whose weak limit distribution, as $j \to \infty$, becomes

$$\begin{aligned} \mathbb{P}_{\mathbf{s}}^{*(k,\ell)}(A) &= \frac{\mathbb{E}\left\{ \mathbb{I}_{\mathbf{X}_0 \in A} K_k \left(\ell(\mathbf{X}_{-k}^{-1} - \mathbf{s}) \right) \right\}}{\mathbb{E}\left\{ K_k \left(\ell(\mathbf{X}_{-k}^{-1} - \mathbf{s}) \right) \right\}} \\ &= \frac{\int m_A(\mathbf{z}) K_k \left(\ell(\mathbf{z} - \mathbf{s}) \right) \mu_k(d\mathbf{z})}{\int K_k \left(\ell(\mathbf{z} - \mathbf{s}) \right) \mu_k(d\mathbf{z})} \end{aligned}$$

which converges, as $\ell \to \infty$, to $m_A(\mathbf{s})$ for μ_k-almost all $\mathbf{s}$ by another version of Lebesgue density theorem, see Lemma 24.8 in Györfi, Kohler, Krzyżak, and Walk [14]. The rest of the proof is identical to that of Theorem 1. ∎

Acknowledgements. We thank Yoram Singer for sending us the NYSE data set used in the experiments. We thank the reviewers and the associate editor for useful comments.

References

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

[2] P. Algoet. The strong law of large numbers for sequential decisions under uncertainity. *IEEE Transactions on Information Theory*, 40:609–634, 1994.

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

[4] A.R. Barron and T.M. Cover. A bound on the financial value of information. *IEEE Transactions on Information Theory*, 34:1097–1100, 1988.

[5] A. Blum and A. Kalai. Universal portfolios with and without transaction costs. *Machine Learning*, 35:193–205, 1999.

[6] A. Borodin, R. El-Yaniv, and V. Gogan. On the competitive theory and practice of portfolio selection (extended abstract). In *Proc. of the 4th Latin American Symposium on Theoretical Informatics (LATIN'00)*, pages 173–196, Punta del Este, Uruguay, 2000.

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

[8] L. Breiman. Optimal gambling systems for favorable games. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 65–78, Berkeley, 1961. University of California Press.

[9] N. Cesa-Bianchi and G. Lugosi. Minimax values and entropy bounds for portfolio selection problems. In *Proceedings of the First World Congress of the Game Theory Society*, 2000.

[10] T.M. Cover. Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

[11] T.M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42:348–363, 1996.

[12] J.E. Cross and A.R. Barron. Efficient universal portfolios for past-dependent target classes. *Mathematical Finance*, 13:245–276, 2003.

[13] M. Finkelstein and R. Whitley. Optimal strategies for repeated games. *Advances in Applied Probability*, 13:415–428, 1981.

[14] L. Györfi, M. Kohler, A. Krzyżak, and H Walk. *A Distribution-Free Theory of Nonparametric Regression.* Springer, New York, 2002.

[15] L. Györfi and G. Lugosi. Strategies for sequential prediction of stationary time series. In F. Szidarovszky M. Dror, P. L'Ecuyer, editor, *Modeling Uncertainty: An examination of its theory, methods, and applications*, pages 225–248. Kluwer Academic Publishers, 2001.

[16] L. Györfi, G. Lugosi, and G. Morvai. A simple randomized algorithm for consistent sequential prediction of ergodic time series. *IEEE Transactions on Information Theory*, 45:2642–2650, 1999.

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

[18] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth. Online portfolio selection using multiplicative updates. *Mathematical Finance*, 8:325–344, 1998.

[19] J.L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.

[20] H.A. Latané. Criteria for choice among risky ventures. *Journal of Political Economy*, 38:145–155, 1959.

[21] E. Ordentlich and T.M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23:960–982, 1998.

[22] Y. Singer. Switching portfolios. *International Journal of Neural Systems*, 8:445–455, 1997.

[23] P. Spellucci. DONLP2, C Version 7/99. *Netlib*, `http://www.netlib.org/`, 1997.

[24] G. Stoltz and G. Lugosi. Internal regret in on-line portfolio selection. In *Proceedings of the 16th Annual Conference on Learning Theory, COLT 2003*, pages 403–417, 2003.

[25] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 36:111–147, 1974.

[26] V.G. Vovk and C. Watkins. Universal portfolio selection. In *Proceedings of the 11th annual conference on Computational Learning Theory*, pages 12–23, 1998.

[27] H. Walk and S. Yakowitz. Iterative nonparametric estimation of a log-optimal portfolio selection function. *IEEE Transactions on Information Theory*, 48:324–333, 2002.