Experiments on universal portfolio selection using data from real markets

László Györfi, Frederic Udina, Harro Walk

January 17, 2008

Abstract

In recent years optimal portfolio selection strategies for sequential investment have been shown to exist. Although their asymptotical optimality is well established, finite sample properties do need to be tested in practice. In this paper we present some experiments based on real data form financial markets. We show that these methods show very good performance that reflect his log-optimal asymptotic properties. In our comparisons, nearest-neighbor based methods appear to be much more robust than others. They are quite insensitive to changes in scale or dimensionality that force other method to adjust smoothing parameters.

1 Introduction

In a financial market, on the basis of the past market data, without knowledge of the underlying statistical distribution, a portfolio selection has to be chosen for investment of the current capital in the available assets at the beginning of the new market period. The goal is to find a portfolio selection scheme such that the investor's wealth grows on the average as fast as by the optimum strategy based on the full knowledge of the underlying distribution. Nonparametric statistical methods allow to construct asymptotically optimal strategies for sequential investment in financial markets.

Throughout the paper it is assumed that the vectors of daily price relatives (return vectors) form a stationary and ergodic process. Then a log-optimal rate of growth exists and is achieved with probability one by a strategy based on the knowledge of the underlying distribution (Algoet and Cover [2]). Even in the

AMS 2000 subject classifications. Primary: 62G10 secondary: 62G05, 62L12, 62M20, 62P05, 91B28

Key words and phrases: sequential investment, universally consistent portfolios, nearest neighbor estimation, portfolio selection

more realistic case that only the past data are available, with no knowledge of the underlying distribution, selection schemes with log-optimal growth rate exist (Algoet [1]). Such investment schemes are called *universally consistent*. Györfi and Schäfer [10] constructed universally consistent schemes using histograms from nonparametric statistics, and Györfi, Lugosi, and Udina [9] using kernel estimates. In this paper a new universal strategy, called nearest neighbor strategy, is proposed which not only guarantees a log-optimal growth rate of capital for all stationary and ergodic markets, but also have a good finite-horizon performance in practice, and, as main novelty, is very robust in the sense that no design parameter tuning is needed to guarantee this good finite-horizon performance. The reason may be that nearest neighbor methods can be interpreted as well tractable kernel methods with data-based local choice of bandwidths. In [11], we present a numerical comparison of some empirical portfolio strategies for NYSE and currency exchange data, according to which the nearest neighbor based portfolio selection outperform the histogram and the kernel strategy.

The rest of the paper is organized as follows. In Section 2 the mathematical model is described. In Section 3 a nearest neighbor (NN) based nonparametric sequential investment strategy is introduced and its universal consistency is stated. The proof of this theoretical result (Theorem 3.1) is given in Section 4.

2 Mathematical model

The following stock market model has been investigated, among others, by Algoet and Cover [2]. Further references can be found in Györfi, Lugosi, and Udina [9]. Also the monographs of Cover and Thomas [7], and Luenberger [13] deal with the concept of log-optimality below.

Consider a market of d assets. The evolution of the market in time is represented by a sequence of return vectors $\mathbf{x}_1, \mathbf{x}_2, \dots$ with values in $\mathbb{R}_+^d$, where the j-th component $x_n^{(j)}$ of the return vector $\mathbf{x}_n$ denotes the amount obtained after investing a unit capital in the j-th asset on the n-th trading period. That is, the j-th component $x_n^{(j)} \geq 0$ of $\mathbf{x}_n$ expresses the ratio of the closing and opening prices of asset j during the n-th 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 $\sum_{j=1}^d b^{(j)} = 1$. It means that the investor neither consumes money nor deposits new money and that no transaction costs appear. The non-negativity of the components of $\mathbf{b}$ means that short selling and buying stocks on margin are not permitted. Denote by Δ_d the simplex of all vectors $\mathbf{b} \in \mathbb{R}_+^d$ with nonnegative components summing up to one.

Let S_0 denote the investor's initial capital. For the first trading period, the portfolio vector $\mathbf{b}_1$ is constant, usually $(1/d, ...1/d)$. Then at the end of the first trading period the investor's wealth becomes

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} x_1^{(j)} = S_0 \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle \, ,$$

where $\langle \cdot \, , \, \cdot \rangle$ denotes inner product. For $j \le i$ we abbreviate by $\mathbf{x}_j^i$ the array of market vectors $(\mathbf{x}_j, \dots, \mathbf{x}_i)$. Let S_{n-1} be the wealth at the end of the $n-1$-th trading period, then it is the initial capital for the n-th trading period, for which the portfolio may depend on the past return vectors: $\mathbf{b}_n = \mathbf{b}_n(\mathbf{x}_1^{n-1})$. Therefore we get by induction that

$$S_n = S_{n-1} \left\langle \mathbf{b}_n(\mathbf{x}_1^{n-1}) \, , \, \mathbf{x}_n \right\rangle = S_0 \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle = S_0 \exp \left\{ \sum_{i=1}^{n} \log \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle \right\} .$$

This may be written as $S_0 \exp\{n W_n(\mathbf{B})\}$, where $W_n(\mathbf{B})$ denotes the *average growth rate* of the investment strategy $\mathbf{B} = \{\mathbf{b}_n\}_{n=1}^{\infty}$:

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

The goal is to maximize the wealth $S_n = S_n(\mathbf{B})$ or, equivalently, maximize the average growth rate $W_n(\mathbf{B})$.

We assume that the sequence of return vectors $\mathbf{x}_1, \mathbf{x}_2, \dots$ are realizations of a random process $\mathbf{X}_1, \mathbf{X}_2, \dots$ such that $\{\mathbf{X}_n\}_{-\infty}^{\infty}$ is a stationary and ergodic process. Besides a mild moment condition on the log-returns, no other distribution assumptions are made. According to Algoet and Cover [2], for the so-called conditional log-optimum investment strategy $\mathbf{B}^* = \{\mathbf{b}_n^*\}_{n=1}^{\infty}$ defined by

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

one has

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

for each competitive strategy $\mathbf{B}$, where $S_n^* = S_n(\mathbf{B}^*)$ and $S_n = S_n(\mathbf{B})$. Furthermore

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

where

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

is the maximal possible growth rate of any investment strategy. The conditional log-optimum investment strategy $\mathbf{B}^*$ depends upon the distribution of the stationary and ergodic process $\{\mathbf{X}_n\}_{n=1}^{\infty}$. Surprisingly, according to Algoet [1], there exists investment strategy $\tilde{\mathbf{B}}$ on the basis of past return data such that

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

i.e., having the same best asymptotic growth rate as $\mathbf{B}^*$, for each stationary and ergodic processes $\{\mathbf{X}_n\}_{-\infty}^{\infty}$. Such investment strategies are called *universally consistent* with respect to a class of all stationary and ergodic processes.

The investment strategy of Györfi and Schäfer's [10] is, as Algoet's [1] strategy, histogram based. At a given time instant n one looks for correspondingly discretized k-tuples $\mathbf{x}_{n-k-j+1}^{n-j}$ of return vectors in the whole history of the market which are identical to the discretized return vectors $\mathbf{x}_{n-k}^{n-1}$. Such time instant $n-j$ is called matching time. Then design a fixed portfolio vector optimizing the return for the trading periods following each matching. For different integer $k > 0$ and histogram design parameter, mix these portfolios (see (3) below). Györfi, Lugosi, and Udina [9] modified this strategy by use of kernels ("moving-window"). In both papers, universal consistency of the strategies with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, for $j = 1, 2, \dots d$ is shown.

3 Nearest neighbor based strategy

Define an infinite array of elementary strategies (the so-called experts) $\mathbf{H}^{(k,\ell)} = \{\mathbf{h}^{(k,\ell)}(\cdot)\}$, where k, ℓ are positive integers. Just like before, k is the window length of the near past, and for each ℓ choose $p_\ell \in (0,1)$ such that

$$\lim_{\ell\to\infty} p_\ell = 0. \tag{1}$$

Put

$$\hat{\ell} = \lfloor p_\ell n \rfloor.$$

At a given time instant n, the expert searches for the $\hat{\ell}$ nearest neighbor (NN) matches in the past. For fixed positive integers k, ℓ ($n > k + \hat{\ell} + 1$) and for each vector $\mathbf{s} = \mathbf{s}_{-k}^{-1}$ of dimension kd introduce the set of the $\hat{\ell}$ nearest neighbor matches:

$$\hat{J}_{n,\mathbf{s}}^{(k,\ell)} = \{i;\ k+1 \le i \le n \text{ such that } \mathbf{x}_{i-k}^{i-1} \text{ is among the } \hat{\ell} \text{ NNs of } \mathbf{s} \text{ in } \mathbf{x}_1^k, \dots, \mathbf{x}_{n-k}^{n-1}\}.$$

Define the portfolio vector by

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}, \mathbf{s}) = \arg\max_{\mathbf{b}\in\Delta_d} \prod_{i\in\hat{J}_{n,\mathbf{s}}^{(k,\ell)}} \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle\,.$$

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, $\mathbf{h}_n^{(k,\ell)}$ is a fixed portfolio vector according to the return vectors following these nearest neighbors.

Now one forms a "mixture" of all experts using a positive probability distribution $\{q_{k,\ell}\}$ on the set of all pairs (k, ℓ) of positive integers (i. e. such that for all k, ℓ, $q_{k,\ell} > 0$). The investment strategy simply weights 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

$$S_n = \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{H}^{(k,\ell)}), \tag{3}$$

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 $S_0 = 1$. This may easily be achieved by distributing the initial capital $S_0 = 1$ among all experts such that expert $\mathbf{H}^{(k,\ell)}$ trades with initial capital $q_{k,\ell} S_0$.

We say that a tie occurs with probability zero if for any vector $\mathbf{s} = \mathbf{s}_1^k$ the random variable

$$\|\mathbf{X}_1^k - \mathbf{s}\|$$

has continuous distribution function.

Theorem 3.1 *Assume (1) and that a tie occurs with probability zero. The portfolio scheme $\mathbf{B}^{\mathrm{NN}}$ is universally consistent with respect to the class of all stationary and ergodic processes such that $\mathbb{E}\{|\log X_0^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.*

4 Empirical results

In this section we present some numerical results obtained by applying the above algorithms to two sets of financial data. The first data set, described and analyzed in Section 4.1, includes prices for 36 NYSE stocks along 22 years. In Section 4.2 we analyze currency exchange data for eight currencies during a period of more than 13 years.

Testing the algorithms with data from real financial markets is meaningful, but it needs some previous considerations about the assumptions implied in our model that are not found in real markets.

First, we assume that assets are available in the desired quantities at a given price at any trading period. The investment period is modeled as a single instant: there is a single price for the entire period, the closing price at the previous period. After we do the desired transaction, we are informed about the closing price and the period ends.

We also assume that assets are arbitrarily indivisible. As in the mathematical analysis, we ignore transaction costs to be paid when switching portfolios. Moreover, all the wealth achieved in the last period is fully invested in the next one, without any extra investment allowed (no short sales on margin). Also, the set of assets involved is fixed: no asset may disappear, no new assets are allowed to be introduced in the market.

Another implicit assumption is that prices are not affected by our actions on the market (since we use historical data, we are forced to assume that). In the NYSE data example below, this is not a realistic assumption since we are trading enormous amount of asset values, not negligible even in comparison with the full market (we multiply our initial wealth by 10^{12} after about 20 years of trading).

All the proposed universally consistent algorithms use an infinite array of experts. In practice we take a finite array of size $K \times L$ (usually 5×10). 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 we take the uniform distribution $\{q_{k,\ell}\} = 1/(KL+1)$ over the experts in use. Implementation of the histogram $\mathbf{B}^H$ and kernel $\mathbf{B}^K$ portfolios is described in detail in Györfi, Lugosi, and Udina [9]. In the results presented below, $\mathbf{B}^K(c)$ denote the kernel portfolio where the expert (k, ℓ) uses $r_{k,\ell} = c/\ell$ as radius for the kernel.

For computational complexity reasons, see below, we introduce a variant of the nearest neighbor rule based on the one described previously. Just like before, k is the window length of the near past, however, s_ℓ is a segment length. At a given time instant n, the expert searches for a single nearest neighbor match within each segment of length s_ℓ. To define $\mathbf{H}^{(k,\ell)}$, for fixed positive integers k, ℓ introduce first the nearest neighbors within each segment. For $1 \le i \le \lfloor (n-1)/s_\ell \rfloor$, let $N_i - 1$ be the instant of nearest neighbor match of $\mathbf{x}_{n-k}^{n-1}$ within the ith segment:

$$N_i = \arg\min_{(i-1)s_\ell + k \le j < i s_\ell} \|\mathbf{x}_{j-k}^{j-1} - \mathbf{x}_{n-k}^{n-1}\|.$$

We define the expert $\mathbf{h}^{(k,\ell)}$, for $n > s_\ell + 1$, by

$$\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b} \in \Delta_d} \prod_{\{1 \le i \le \lfloor (n-1)/s_\ell \rfloor\}} \langle \mathbf{b} \, , \, \mathbf{x}_{N_i} \rangle \ .$$

Then we combine experts in the usual way and this way we obtain the first-neighbor that we shall denote by $\mathbf{B}^{fN}$.

It is not difficult to see that Theorem 3.1 is also valid for the first neighbor variant, provided that (1) is replaced by

$$\lim_{\ell \to \infty} s_\ell = \infty. \tag{4}$$

When using real data, ties occur quite often, sometimes due to rounding. But in many cases it is also true that real data can not be assumed coming from a continuous process: for example, the NYSE data discussed below present many cases of relative price equal to one, much more than one may expect coming from rounding. This is not surprising as one may expect that a given asset does not vary its price on several trading days. Presence of ties does not result in any problem or degradation of performance of the nearest-neighbor portfolio, as can be seen in the results presented below. We deal with ties in the following way: when taking ν elements from a set that have distances d_i (assuming distances in increasing order), if $d_\nu = d_{\nu+1}$, we take also the element with index $\nu + 1$ together with all subsequent elements that show the same distance. In the first neighbor case, in the presence of ties within a segment, we take the more recent period of those tied.

In the nearest neighbor portfolio implementation, we take

$$p_l = 0.02 + 0.5\frac{\ell - 1}{L - 1},$$

so our 50 experts take from 2% to 52% of the history as matching periods. For the first neighbor portfolio we take

$$s_l = 2 + 50\frac{\ell - 1}{L - 1}.$$

With this choice the number of matches used by the experts is very similar in both variants.

It is possible that other choices of these quantities may improve the performance of the algorithms, but this would be true for some situations or data sets and not for other. We do not want to find parameter choices tailored for a particular data set. We want to stress that a single and reasonable choice of these parameters works very well in very different markets, not depending on the dimensionality or the scale of the relative prices being considered. In the examples below, the very same choice works for $d = 2$ or $d = 36$, for stock exchange data or for currency exchange data, scales in those last cases being very different one from another.

4.1 Stock market data

The first data set we use is a standard set of New York Stock Exchange data used by Cover [6], Singer [14], Hembold, Schapire, Singer, and Warmuth [12], Blum and Kalai [3], Borodin, El-Yaniv, and Gogan [4], and others. It includes daily prices of 36 assets along a 22-year period (5651 trading days) ending in 1985.

We first take the pairs used in the aforementioned papers, see Table 1. In the second and third columns of the table we show the wealth achieved by investing one US$, respectively, in the best of both assets, following the best constantly

Table 1: Wealth achieved by different strategies by investing in the pairs of NYSE stocks used in Cover [6]. In the second column we show some reference results from the literature. In the right part of the table results are shown for the histogram, kernel and nearest neighbors strategies.

Stocks					Best Exp. $[\mathsf{k},\ell]$
Iroquois	Best asset	8.92	$\mathbf{B}^{\mathsf{H}}$	2.3e+10	1.395e+11 [1,1]
Kin Ark	BCRP	73.70	$\mathbf{B}^{\mathsf{K}}$	4.038e+10	9.014e+11 [2,2]
	Oracle	6.85e+53	$\mathbf{B}^{NN}$	1.156e+12	1.439e+13 [2,8]
	Cover UP	39.97	$\mathbf{B}^{fN}$	5.094e+11	6.108e+12 [2,3]
	Singer SAP	143.7			
Com. Met.	Best asset	52.02	$\mathbf{B}^{\mathsf{H}}$	162.5	327.8 [2,1]
Mei. Corp	BCRP	103.0	$\mathbf{B}^{\mathsf{K}}$	775.1	4749. [2,5]
	Oracle	2.12e+35	$\mathbf{B}^{NN}$	3.505e+3	3.148e+4 [3,6]
	Cover UP	74.08	$\mathbf{B}^{fN}$	1.018e+4	5.63e+4 [4,2]
	Singer SAP	107.7			
Com. Met.	Best asset	52.02	$\mathbf{B}^{\mathsf{H}}$	1.331e+10	8.544e+10 [1,1]
Kin Ark	BCRP	144.0	$\mathbf{B}^{\mathsf{K}}$	1.111e+11	1.411e+12 [3,3]
	Oracle	1.84e+49	$\mathbf{B}^{NN}$	4.781e+12	8.257e+13 [3,7]
	Cover UP	80.54	$\mathbf{B}^{fN}$	9.031e+11	8.262e+12 [3,2]
	Singer SAP	206.7			
IBM	Best asset	13.36	$\mathbf{B}^{\mathsf{H}}$	63.87	112.2 [1,5]
Coca-Cola	BCRP	15.02	$\mathbf{B}^{\mathsf{K}}$	47.6	194.6 [1,6]
	Oracle	1.08e+15	$\mathbf{B}^{NN}$	74.37	296.3 [1,7]
	Cover UP	14.24	$\mathbf{B}^{fN}$	85.25	260.3 [4,9]
	Singer SAP	15.05			

rebalanced portfolio (BCRP), by an "Oracle" that knows in advance the prices for the next period, using the universal portfolio as reported in Cover [6], and the SAP portfolio as reported in Singer [14]. Note that BCRP and Oracle do not correspond to any valid investment strategy; they can only be determined in full hindsight. In the right part of the table we report our results for histogram ($\mathbf{B}^{\mathsf{H}}$), kernel ($\mathbf{B}^{\mathsf{K}}$, with constant $\mathsf{c} = 0.05$), nearest neighbor ($\mathbf{B}^{NN}$) and first neighbor ($\mathbf{B}^{fN}$) portfolios. For each pair, we start with one unit of wealth (say one US dollar), we use $\mathsf{K} = 1, \ldots, 5$ and $\mathsf{L} = 1, \ldots, 10$ for a total of 51 experts, including the expert that uses the full available past to compute the optimum portfolio. In the last column we report the wealth achieved by the best expert among the 51 competing, and its values for k, ℓ.

Investing in a fixed pair of assets involves the difficult choice of the pair, so we prefer to report results for a more *blind* strategy: simply invest in *all* assets available. Table 2 summarizes the wealth achieved by several portfolio strategies when one dollar is split between the 36 assets in the first period. Our implementation of the histogram portfolio does not allow for such a large dimensionality. For the kernel portfolio we take $\mathsf{c} = 1.00$ as a good value for this dimensionality. Nearest and first neighbor portfolios are as described before. In all cases, we use $\mathsf{K} = 5, \mathsf{L} = 10$. For the sake of reference, we also indicate the wealth achieved by BCRP.

Table 2: Wealth achieved by various strategies. In all cases one unit is invested in the first period uniformly in all 36 stocks included in our NYSE data set. $\mathbf{B}^{\mathsf{K}}$ is the kernel strategy with constant $\mathsf{c} = 1.00$, $\mathbf{B}^{NN}$ is the nearest neighbor portfolio, $\mathbf{B}^{fN}$ is the first neighbor variant of the last one, and BCRP is the best constantly rebalanced portfolio.

After period	BCRP	$\mathbf{B}^{\mathsf{K}}(1.00)$	$\mathbf{B}^{NN}$	$\mathbf{B}^{fN}$
500	13.07	3.926	8.096	11.61
1000	7.324	6.412	23.98	31.40
1500	16.03	13.35	66.44	102.2
2000	10.21	12.34	103.9	116.5
2500	17.48	77.53	1050.2	1316.
3000	18.81	1.749e+3	6.416e+5	1.326e+5
3500	34.57	4.554e+4	3.650e+6	9.578e+6
4000	55.52	1.832e+6	1.489e+8	5.054e+8
4500	106.8	8.066e+7	1.322e+9	5.364e+9
5000	125.4	8.066e+7	1.238e+10	2.822e+10
5500	267.8	9.414e+8	2.777e+11	2.837e+11
5651	250.6	1.116e+9	3.313e+11	2.302e+11

A better graphical comparison of results achieved by our algorithms can be seen in Figure 1 where the full time series is represented. It is interesting to observe that while the kernel portfolio needs about 2000 periods to start getting some wealth, the nearest neighbor is able to exploit the information very soon: even it detects around $n = 400$ that there is one particular asset that is growing fast in these periods.

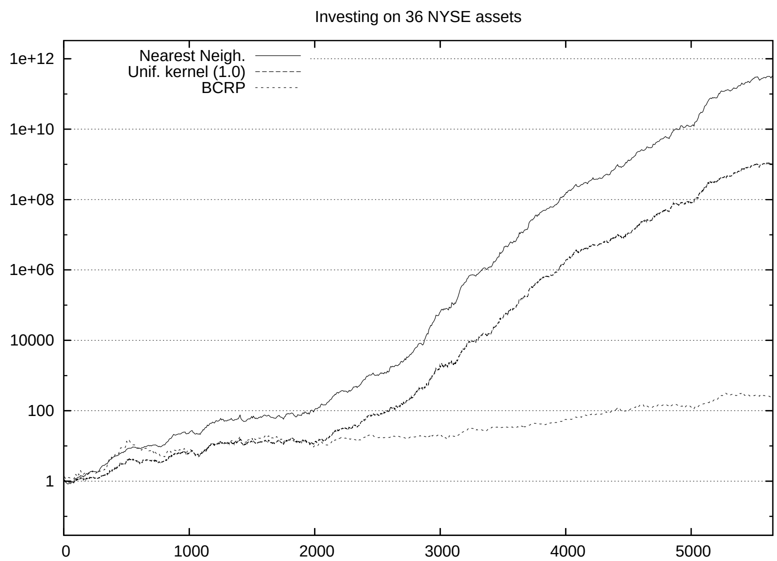

Figure 1: Wealth achieved along $N = 5651$ daily periods by investing one US\$ in 36 NYSE stocks (data set described in the text) using several asymptotically optimal strategies. Horizontal axis is time period number, vertical is wealth achieved, on logarithmic scale.

A more detailed comparison of the different performance of kernel versus nearest neighbor is made in Table 3. There we list the wealth achieved by each expert at the end of the full time period. In the lower half of the listing it can be seen that most of the experts in the nearest neighbor setting have quite good performance, all of them being within a factor of 10^5 of the overall wealth. On the contrary, in the upper half there are experts with very good performance but there are also many with very low final wealth, of the order of 10^{-8} times less than the overall final wealth. Obviously, some better choice of the parameters that define the radius of the kernel will provide better results, but these would be tailored for this particular case, dimension, and data set, while the parameters chosen for the nearest neighbor are valid without any change for all situations, for different values of k and dimensions. This is why we see (in this and the rest of the examples we studied, not all reported here) the nearest neighbor as a very robust algorithm that performs fairly well without any adjustment for different data sets, dimensionalities and scales.

Table 3: Wealth achieved by each expert at the last period when investing one unit in the 36 NYSE assets. Upper part is for the kernel ($c = 1.0$) portfolio, while the lower part is for the nearest neighbor portfolio. Experts are indexed by $k = 1..5$ in columns and $\ell = 1..10$ in rows.

$\mathbf{B}^{K}(1.0)$ on 36 NYSE assets. $S_{5651} = 1.1e+9$					
ℓ k	1	2	3	4	5
1	4.0	4.2	4.3	5.2	3.3
2	5.0	7.2	1.1e+1	4.2e+1	8.2e+1
3	2.4e+1	7.5e+2	7.6e+4	4.9e+6	1.0e+7
4	5.7e+3	4.7e+6	1.2e+7	4.9e+5	5.4e+2
5	1.7e+8	8.0e+7	7.1e+2	2.e+2	3.1e+1
6	5.2e+9	3.9e+6	1.1e+2	2.2e+1	2.7e+1
7	4.2e+10	3.5e+3	1.7e+1	2.7e+1	2.7e+1
8	8.4e+9	3.8e+1	2.4e+1	2.7e+1	2.7e+1
9	8.3e+8	4.3e+1	2.7e+1	2.7e+1	2.7e+1
10	6.5e+5	3.4e+1	2.7e+1	2.7e+1	2.7e+1
$\mathbf{B}^{NN}$ on 36 NYSE assets. $S_{5651} = 3.3e+11$					
1	5.2e+10	1.3e+8	1.2e+7	1.5e+8	3.0e+7
2	1.4e+11	6.4e+8	3.5e+7	2.2e+8	1.9e+8
3	1.9e+11	2.0e+9	1.1e+8	1.1e+9	9.5e+8
4	1.6e+11	8.1e+8	6.0e+8	1.5e+10	2.9e+9
5	3.5e+11	7.0e+8	1.3e+8	2.0e+10	7.0e+9
6	6.1e+11	2.9e+9	1.9e+9	4.3e+9	1.3e+9
7	6.8e+12	8.1e+8	1.7e+9	7.7e+9	2.0e+8
8	7.7e+12	2.4e+9	4.6e+8	1.7e+10	4.3e+7
9	2.1e+11	1.3e+9	7.3e+8	4.1e+9	1.5e+7
10	5.9e+11	9.8e+8	1.9e+7	5.5e+7	7.1e+6

4.2 Currency exchange data

Table 4: Some descriptive statistics for the exchange rate data.

Currency	Final	Mean	St. Dev.	Min.	p_{25}	p_{75}	Max.
Jap. Yen	1.2456	0.000089	0.0071	-0.0405	-0.0037	0.0035	0.0799
Swiss Fr.	1.1891	0.000077	0.0073	-0.0477	-0.0044	0.0043	0.0545
Norw. Kr.	0.9774	0.000016	0.0068	-0.0529	-0.0036	0.0038	0.0491
Ecu/Euro	1.0097	0.000022	0.0062	-0.0498	-0.0034	0.0034	0.0338
Brit. Pnd.	1.0404	0.000028	0.0058	-0.0420	-0.0029	0.0031	0.0431
Can. Dol.	0.9184	-0.000018	0.0037	-0.0172	-0.0020	0.0019	0.0168
Hung. Fnt.	0.4098	-0.000233	0.0073	-0.0866	-0.0031	0.0028	0.0370

We applied our algorithms to a data set of currency exchange rates to check that the nearest neighbor is very robust against changes of scale or dimensionality. Data were obtained from Datastream Advance, a commercial database, and include exchange rates to US$ for seven currencies (see Table 4) from December 6, 1991 to January 27, 2005, a total of 3429 daily periods in just over 13 years. The currencies are Japanese Yen, Swiss Franc, Norwegian Crown, European Currency Unit followed by the Euro when it was introduced, British Pound, Canadian Dollar and Hungarian Forint. In Table 4 we report, for each currency, the final wealth of one US$ invested in that currency from the first period, and also some statistics of the series of daily returns (ratio of price to previous price, minus one) for all 3429 periods. Statistics include the mean, standard deviation, minimum, 25th and 75th percentiles, and maximum. We do not include the median because it is identical to zero in all cases.

Table 5 and Figure 2 show the results of investing one US$ split into the seven currencies and cash and then selecting the portfolio for each period according the histogram, kernel, nearest neighbors and first neighbor. We use $K = 5$ in all cases, and $L = 10$ in all cases except for the histogram that uses $L = 6$. The kernel algorithm is used here with $c = 0.1$ to adjust it to the dimensionality and scale of this case. The nearest neighbor algorithms are used as described before, without any parameter tuning. The second column of Table 5 reports results for the best constant rebalanced portfolio. Note that we introduce cash as a possible asset to invest in along with the seven currencies. We consider cash not having any variation, i. e. no interest rate at all.

Results here are fairly good, note that a wealth factor of 393.0 along 3429 periods is roughly equivalent to a yearly increase rate of 57%, so even in this case our algorithms are able to efficiently exploit the existing inefficiency of this market. What should be remarked again is that the nearest neighbor algorithms perform better than the rest without needing any parameter adjustment.

Table 5: Wealth achieved by various strategies. In all cases one unit is invested in the first period uniformly in all currencies included in our data set, cash is also included. BCRP is the best constantly rebalanced portfolio, $\mathbf{B}^{\mathsf{H}}$ is the histogram strategy, $\mathbf{B}^{\mathsf{K}}(0.1)$ is the kernel strategy, $\mathbf{B}^{NN}$ is the nearest neighbor portfolio, and $\mathbf{B}^{fN}$ is the first neighbor variant of the last one.

After period	BCRP	$\mathbf{B}^{\mathsf{H}}$	$\mathbf{B}^{\mathsf{K}}(0.1)$	$\mathbf{B}^{NN}$	$\mathbf{B}^{fN}$
500	1.184	1.129	1.419	2.734	2.606
1000	1.276	1.174	2.904	9.095	7.996
1500	1.058	1.898	3.775	13.78	11.26
2000	1.115	2.551	5.605	23.33	18.01
2500	1.028	3.407	10.14	48.84	35.98
3000	1.118	8.215	26.71	142.3	100.4
3429	1.273	22.22	70.11	393.0	280.1

4.3 Computational cost

The methods introduced here are quite costly to compute. There are essentially two steps that are very computing intensive: the selection of the matches in the past and the optimization step. The computing cost of the optimization step is related to the number of matches found. To give an idea of the importance of this, note that the kernel algorithm, when used for our NYSE data with $c = 0.5$, took half the time (12 hours) than when used with $c = 1.0$. All times reported here are approximate and refer to a run over the full period of the data set using a PC with a Pentium Xeon CPU at 2.0Ghz.

In the nearest neighbor case, running all $n = 5651$ periods for the $d = 36$ assets took 36 hours, while using the first neighbor variant this time reduced to 12 hours. The computing load due to the optimization is the same in both cases, but searching for the matches in the first case is much more costly because it requires sorting all previous data sequences according to the distance to the current one.

To show a more complete comparison of the discussed algorithms, we may mention that computation of the optimal portfolio for the Iroquois/Kin Ark pair mentioned in Section 4.1 took 46 minutes for the nearest neighbor portfolio, 37 minutes for the first neighbor, and 10 minutes for the kernel portfolio with $c = 0.05$.

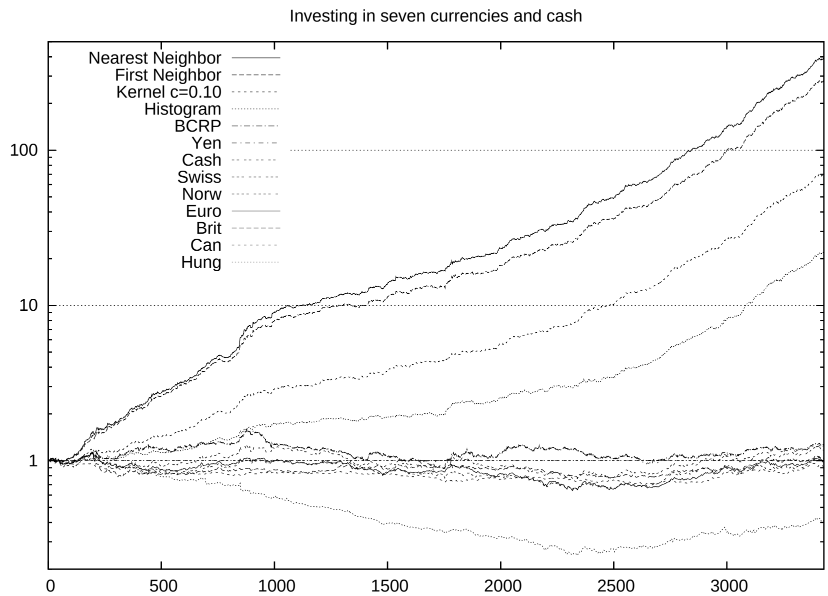

Figure 2: Wealth achieved along $N = 3429$ daily periods by investing one US$ in seven foreign currencies and cash. Labels in the legend are in the same top down order as the lines in the graph. Horizontal axis is time period number, vertical is the wealth achieved, in logarithmic scale.

5 Conclusion

We have shown how to construct and implement an algorithm for sequential investment. The algorithm is based on nearest neighbor estimation and we prove that it has asymptotical optimality for the very general class of stationary and ergodic processes. The more interesting property of the new algorithm is its robustness: it may be applied to different situations without needing any adjustment to the scale or dimensionality. Empirical results on real financial data show very good finite-horizon performance and are even spectacular in some cases.

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. The work of the second author was supported by the Spanish Ministry of Science and Technology and FEDER, grant MTM2006-05650 We thank Michael Greenacre for his careful and annotated reading.

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

[4] 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).

[5] 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.

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

[7] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York (1991).

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

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

[10] 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).

[11] L. Györfi, F. Udina, and H. Walk. Experiments on empirical portfolio selection strategies. (submitted). www.econ.upf.es/ udina/NNexp.pdf.

[12] 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.

[13] D. G. Luenberger *Investment Science*. Oxford University Press, (1998).

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

[15] W. F. Stout. *Almost Sure Convergence.* Academic Press, New York (1974).

László Györfi
Department of Computer Science and Information Theory,
Budapest University of Technology and Economics,
1521 Stoczek u. 2, Budapest, Hungary.
e-mail: gyorfi@szit.bme.hu

Frederic Udina
Department of Economics and Business,
Universitat Pompeu Fabra,
Ramon Trias Fargas 25-27, 08005 Barcelona, Spain.
e-mail: udina@upf.es

Harro Walk
Institute of Stochastics and Applications,
Universität Stuttgart,
Pfaffenwaldring 57, D-70569 Stuttgart, Germany.
e-mail: walk@mathematik.uni-stuttgart.de