Asymptotically Optimal Algorithm for Short-Term Trading Based on the Method of Calibration

Vladimir V'yugin and Vladimir Trunov

Institute for Information Transmission Problems, Russian Academy of Sciences,
Bol'shoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia
email: vyugin@iitp.ru

Abstract. A trading strategy based on a natural learning process, which asymptotically outperforms any trading strategy from RKHS (Reproduced Kernel Hilbert Space), is presented. In this process, the trader rationally chooses his gambles using predictions made by a randomized well calibrated algorithm. Our strategy is based on Dawid's notion of calibration with more general changing checking rules and on some modification of Kakade and Foster's randomized algorithm for computing calibrated forecasts. We use also Vovk's method of defensive forecasting in RKHS.

1 Introduction

Predicting sequences is the key problem of machine learning and statistics. The learning process proceeds as follows: observing a finite-state sequence given online a forecaster assigns a subjective estimate to future states. A minimal requirement for testing any prediction algorithm is that it should be calibrated (see Dawid [3]). Dawid gave an informal explanation of calibration for binary outcomes as follows. Let a sequence $\omega_1, \omega_2, \dots, \omega_{n-1}$ of binary outcomes be observed by a forecaster whose task is to give a probability p_n of a future event $\omega_n = 1$. In a typical example, p_n is interpreted as a probability that it will rain. Forecaster is said to be well calibrated if it rains as often as he leads us to expect. It should rain about 80% of the days for which $p_n = 0.8$, and so on.

A more precise definition is as follows. Let $I(p)$ denote the characteristic function of a subinterval $I \subseteq [0,1]$, i.e., $I(p) = 1$ if $p \in I$, and $I(p) = 0$, otherwise. An infinite sequence of forecasts $p_1, p_2, \dots$ is calibrated for an infinite binary sequence of outcomes $\omega_1\omega_2\dots$ if for characteristic function $I(p)$ of any subinterval of $[0,1]$ the calibration error tends to zero, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} I(p_i)(\omega_i - p_i) \to 0$$

as $n \to \infty$. The indicator function $I(p_i)$ determines some "checking rule" which selects indices i where we compute the deviation between forecasts p_i and outcomes ω_i.

If the weather acts adversatively, then, as shown by Oakes [9] and Dawid [4], any deterministic forecasting algorithm will not always be calibrated.

Foster and Vohra [6] show that calibration is almost surely guaranteed with a randomizing forecasting rule, i.e., where the forecasts p_i are chosen using internal randomization and the forecasts are hidden from the weather until weather makes its decision whether to rain or not.

The origin of the calibration algorithms is the Blackwell's [1] approachability theorem but, as its drawback, the forecaster has to use linear programming to compute the forecasts. We modify and generalize a more computationally efficient method from Kakade and Foster [8], where "an almost deterministic" randomized rounding universal forecasting algorithm is presented. For any sequence of outcomes $\omega_1\omega_2\dots$ and for any precision of rounding $\Delta > 0$, an observer can simply randomly round the deterministic forecast p_i up to Δ to a random forecast $\tilde{p}_i$ in order to calibrate for this sequence with probability one :

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} I(\tilde{p}_i)(\omega_i-\tilde{p}_i)\right| \le \Delta, \qquad (1)$$

where $I(p)$ is the characteristic function of any subinterval of $[0,1]$. This algorithm can be easily generalized such that the calibration error tends to zero as $n\to\infty$.

We consider real valued outcomes ω_i (for example, prices of a stock). In this case, predictions could be interpreted as mean values of future outcomes under some unknown for us probability distributions. We need not the precise form of such distributions – we should predict only future means.

The well known applications of the method of calibration are the following. Kakade and Foster proved that empirical frequencies of play in any normal-form game with finite strategy sets converges to a set of correlated equilibrium if each player chooses his gamble as the best response to the well calibrated forecasts of the gambles of other players. In series of papers: [13], [14], [15], [16], [17], Vovk developed the method of calibration for the case of more general RKHS (Reproduced Kernel Hilbert Space) and Banach spaces. Vovk called his method defensive forecasting (DF). He also applied his method for recovering unknown functional dependencies presented by arbitrary functions from RKHS and Banach spaces. Chernov et al. [5] show that well calibrated forecasts can be used to compute predictions for the Vovk's [12] aggregating algorithm.

In this paper we present a new application of the method of calibration. We show that, using the well calibrated forecasts, it is possible to construct an asymptotically optimal trading strategy in the Stock Market which outperforms any trading strategy presented by a function from a given RKHS. The learning process is the most traditional one. At each step *Forecaster* makes a prediction of a future price of the stock and *Speculator* takes the best response to this prediction. He chooses a strategy: dealing for a rise or for a fall, or passes the step. *Forecaster* uses some randomized algorithm for computing calibrated forecasts. Our main result, Theorem 1 (Section 3), says that this trading strategy is optimal – it outperforms any trading strategy presented by a function from a given

RKHS. To achieve this goal we extend in Theorem 2 (Section 4) Kakade and Foster's forecasting algorithm to arbitrary real valued outcomes and to a more general notion of calibration with changing parameterized checking rules. We combine it with Vovk's [13] defensive forecasting method in RKHS (see also [14]). In Section 6 results of numerical experiments are presented.

2 Preliminaries

A Hilbert space $\mathcal{F}$ of real-valued functions on a compact set X is called RKHS (Reproducing Kernel Hilbert Space) on X if the evaluation functional $f \to f(x)$ is continuous for each $x \in X$. Let $\|\cdot\|$ be a norm on $\mathcal{F}$ and $c_{\mathcal{F}}(x) = \sup\limits_{\|f\|\le 1} |f(x)|$. The embedding constant of $\mathcal{F}$ is defined: $c_{\mathcal{F}} = \sup\limits_{x} c_{\mathcal{F}}(x)$. We consider RKHS $\mathcal{F}$ with $c_{\mathcal{F}} < \infty$.

Let $X = [0,1]^m$ for $m \ge 1$. An example of RKHS is the Sobolev space $\mathcal{F} = H^1([0,1])$, which consists of absolutely continuous functions $f : [0,1] \to \mathcal{R}$ with $\|f\| \le 1$, where $\|f\| = \sqrt{\int_0^1 (f(t))^2 dt + \int_0^1 (f'(t))^2 dt}$. For this space, $c_{\mathcal{F}} = \sqrt{\coth 1}$ (see [14]). Other examples and details of the kernel theory see in Smola and Scholkopf [10].

Let $\mathcal{F}$ be an RKHS on X with the dot product $(f \cdot g)$ for $f, g \in \mathcal{F}$. By Riesz–Fisher theorem, for each $x \in X$ there exists $k_x \in \mathcal{F}$ such that $f(x) = (k_x \cdot f)$. The reproduced kernel is defined $K(x,y) = (k_x \cdot k_y)$. Main properties of the kernel: 1) $K(x,y) \ge 0$ for all $x, y \in X$; 2) $K(x,y) = K(y,x) \ge 0$ for all $x, y \in X$; 3) $\sum\limits_{i,j=1}^{k} \alpha_i \alpha_j K(x_i, x_j) \ge 0$ for all k, for all $x_i \in X$, and for all real numbers α_i, where $i = 1, \dots, k$. Mercer theorem says that 1)–3) define a kernel $K(x,y)$ and a mapping Φ to some RKHS $\mathcal{F}$ such that $K(x,y) = (\Phi(x) \cdot \Phi(y))$. Also, $c_{\mathcal{F}}(x) = \|k_x\| = \|\Phi(x)\|$. For Sobolev space $H^1([0,1])$, the reproducing kernel is $K(t,t') = (\cosh \min(t,t') \cosh \min(1-t, 1-t'))/\sinh 1$ (see [14]).

Some special kernel corresponds to the method of randomization defined below. A random variable $\tilde{y}$ is called randomization of a real number $y \in [0,1]$ if $E(\tilde{y}) = y$, where E is the symbol of mathematical expectation with respect to the corresponding to $\tilde{y}$ probability distribution.

We use a specific method of randomization of real numbers from unit interval proposed by Kakade and Foster [8]. Given positive integer number K divide the interval $[0,1]$ on subintervals of length $\Delta = 1/K$ with rational endpoints $v_i = i\Delta$, where $i = 0, 1, \dots, K$. Let V denotes the set of these points. Any number $p \in [0,1]$ can be represented as a linear combination of two neighboring endpoints of V defining subinterval containing p :

$$p = \sum_{v \in V} w_v(p) v = w_{v_{i-1}}(p) v_{i-1} + w_{v_i}(p) v_i, \quad (2)$$

where $p \in [v_{i-1}, v_i]$, $i = \lfloor p^1/\Delta + 1 \rfloor$, $w_{v_{i-1}}(p) = 1 - (p - v_{i-1})/\Delta$, and $w_{v_i}(p) = 1 - (v_i - p)/\Delta$. Define $w_v(p) = 0$ for all other $v \in V$. Define a random variable

$$\tilde{p} = \begin{cases} v_{i-1} \text{ with probability } w_{v_{i-1}}(p) \\ v_i \text{ with probability } w_{v_i}(p) \end{cases}$$

Let $\bar{w}(p) = (w_v(p) : v \in V)$ be a vector of probabilities of rounding.

For any k-dimensional vector $\bar{x} = (x_1, \dots, x_m) \in [0,1]^m$ and $k \le m$, we round each coordinate x_s, $s = 1, \dots k$ to v_{j_s-1} with probability $w_{v_{j_s-1}}(x_s)$ and to v_{j_s} with probability $w_{v_{j_s}}(x_s)$, where $x_s \in [v_{j_s-1}, v_{j_s}]$.

Let $v = (v^1, \dots, v^k) \in V^k$ and $W_v(\bar{x}) = \prod_{s=1}^{k} w_{v^s}(x_s)$. For any $\bar{x}$, let $\bar{W}(\bar{x}) = (W_v(\bar{x}) : v \in V^k)$ be a vector of probability distribution in $V^k : \sum\limits_{v \in V^k} W_v(\bar{x}) = 1$.

For $\bar{x}, \bar{y} \in [0,1]^k$, the dot product $K(\bar{x}, \bar{x}') = (\bar{W}(\bar{x}) \cdot \bar{W}(\bar{x}'))$ satisfies properties 1)–3) of Mercer theorem. Hence it is a kernel function.

3 Main result: an optimal trading strategy

Consider a game with players: *Speculator* and *Stock Market*. We suppose that the prices $S_1, S_2, \dots$ of a stock are bounded and rescaled such that $0 \le S_i \le 1$ for all t.

We present the process of trading in Stock Market in form of a game regulated by the following protocol.

FOR $i = 1, 2 \dots$

Stock Market announces a signal $\bar{x}_i = (x_{i,1} \dots, x_{i,m}) \in [0,1]^m$,where $m \ge 1$. In what follows we suppose that $x_{i,1} = S_{i-1}$.

Speculator bets by buying or selling a number M_i of shares of the stock by S_{i-1} each. [1]

Stock Market reveals a price S_i of the stock.

Speculator receives his total gain (or suffer loss) at the end of step i :

$\mathcal{K}_i = \mathcal{K}_{i-1} + M_i(S_i - S_{i-1})$. We get $\mathcal{K}_0 = 0$.

ENDFOR

We will define an optimal trading strategy as a random variable $\tilde{M}_i$. To construct such a strategy, at each step i we will compute a forecast p_i of a future price S_i and randomize it to $\tilde{p}_i$. We also randomize the past price S_{i-1} of the stock to $\tilde{S}_{i-1}$. Details of this computation and randomization will be given in Section 4. Define

$$\tilde{M}_i = \begin{cases} c_{\mathcal{F}} l, \text{ if } \tilde{p}_i > \tilde{S}_{i-1}, \\ 0, \text{ otherwise,} \end{cases}$$

where l is a parameter. We suppose that some RKHS $\mathcal{F}$ on $[0,1]^m$ with a kernel $M(\bar{x}, \bar{x}')$ and a finite embedding constant $c_{\mathcal{F}}$ is given. We emphasize that we consider only playing for a rise: $\tilde{M}_i \ge 0$ for all i. [2]

[1] We call M_i a trading strategy. In case $M_i > 0$ *Speculator* playing for a rise, in case $M_i < 0$ *Speculator* playing for a fall, *Speculator* passes the step if $M_i = 0$. We suppose that *Speculator* can borrow money for buying M_i shares and can incur debt.

[2] The case of playing for a fall is considered analogously.

We will prove that if the forecasts $\tilde{p}_i$ are well calibrated on the sequence of prices S_i, $i = 1, 2, \ldots$, this strategy outperforms any trading strategy $D(\bar{x}_i) \in \mathcal{F}$ with bounded norm: $\|D\| \le l$. The main result of this paper is presented in the following theorem.

Let $S_1, S_2, \ldots \in [0,1]$ and $\bar{x}_1, \bar{x}_2, \ldots \in [0,1]^m$ be given online according to the protocol of our game.

Theorem 1. *An algorithm for computing forecasts and a sequential method of randomization can be constructed such that given $l > 0$ for any nonegative trading strategy $D(\bar{x}_i) \in \mathcal{F}$ such that $\|D\| \le l$*

$$\liminf_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{M}_i(S_i - S_{i-1}) - \frac{1}{n}\sum_{i=1}^{n} D(\bar{x}_i)(S_i - S_{i-1}) \right) \ge 0 \tag{3}$$

holds almost surely with respect to a probability distribution generated by the corresponding sequential randomization.

Moreover, for any $\epsilon > 0$ this trading strategy $\tilde{M}_i$ can be tuned such that for any $\delta > 0$, with probability at least $1 - \delta$, for all n

$$\sum_{i=1}^{n} \tilde{M}_i(S_i - S_{i-1}) \ge \sum_{i=1}^{n} D(\bar{x}_i)(S_i - S_{i-1}) -$$
$$-c_{\mathcal{F}} l \left(\frac{4}{9}(7e-3)(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{\frac{3}{4}+\epsilon} + \sqrt{(c_{\mathcal{F}}^2+1)n} + 3\sqrt{\frac{n}{2}\ln\frac{6}{\delta}} \right), \tag{4}$$

where e is the base of the natural logarithm.

The proof of this theorem is given in Section 5, where we construct the corresponding optimal trading strategy based on the well calibrated forecasts defined in Section 4.

4 Computing the well calibrated forecasts

We consider checking rules of general type. Let $\bar{x} \in [0,1]^m$ be a signal and $k \le m$. For any measurable subset $\mathcal{S} \subseteq [0,1]^{k+1}$ define

$$I_S(p, \bar{x}) = \begin{cases} 1, & (p, \bar{x}) \in \mathcal{S}, \\ 0, & \text{otherwise,} \end{cases}$$

In Section 5 we get $k = 1$. In what follows, given a sequence $\Delta_1 \ge \Delta_2 \ge \ldots \to 0$ of rational numbers we will define the corresponding randomization $\tilde{p}_1, \tilde{p}_2, \ldots$ of any sequence $p_1, p_2, \ldots$ of reals from $[0,1]$ such that $E_i(\tilde{p}_i) = p_i$ and $Var(\tilde{p}_i) = E_i(\tilde{p}_i - p_i)^2 \le \Delta_i$ for all i, where E_i is the symbol of the mathematical expectation with respect to the random rounding corresponding to Δ_i.

Analogously, a sequence of of k-dimensional random variables $\tilde{x}_1, \tilde{x}_2, \ldots$ will be defined which are randomization of of the first k coordinates of a sequence of

m-dimensional signals $\bar{x}_1, \bar{x}_2, \ldots$. We call this the sequential method of randomization.

Let Pr be an overall probability distribution generated by a sequential method of randomization.

The following theorem is the main tool for analysis presented in Section 5.

Let $S_1, S_2, \ldots \in [0,1]$ be a sequence of real numbers and $\bar{x}_1, \bar{x}_2, \ldots \in [0,1]^m$ be a sequence of m-dimensional signals given online according to the protocol of our game. Let also, $\mathcal{F}$ be RKHS on $[0,1]^m$ with a kernel $M(\bar{x}, \bar{x}')$ and a finite embedding constant $c_{\mathcal{F}}$.

Theorem 2. *Given $\epsilon > 0$ and k an algorithm for computing forecasts $p_1, p_2, \ldots$ and a sequential method of randomization of the forecasts and signals can be constructed such that two conditions hold:*

- *(i) for the characteristic function I_S of any subset $S \subseteq [0.1]^{k+1}$ and for any $\delta > 0$, with probability at least $1 - \delta$,*

$$\left|\sum_{i=1}^{n} I_S(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)\right| \le 8e\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}} (c_{\mathcal{F}}^2 + 1)^{\frac{1}{k+3}} n^{1-\frac{1}{k+3}+\epsilon} + \\ + \sqrt{(c_{\mathcal{F}}^2+1)n} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}} \quad (5)$$

  for all n, where $\tilde{p}_1, \tilde{p}_2, \ldots$ are corresponding randomization of $p_1, p_2, \ldots$ and $\tilde{x}_1, \tilde{x}_2, \ldots$ are corresponding randomization of k-dimensional initial fragments of the signals $\bar{x}_1, \bar{x}_2, \ldots$;
- *(ii) for any $D \in \mathcal{F}$*

$$\sum_{i=1}^{n} D(\tilde{x}_i)(S_i - p_i) \le \|D\| \sqrt{(c_{\mathcal{F}}^2+1)n} \quad (6)$$

  for all n.

Proof. At first, in Proposition 1, given $\Delta > 0$, we modify a randomized rounding algorithm of Kakade and Foster [8] to construct some Δ-calibrated forecasting algorithm, and combine it with Vovk's [14] defensive forecasting algorithm. After that, we revise it by applying some doubling trick argument such that (5) will hold.

Proposition 1. *Under the conditions of Theorem 2, an algorithm for computing forecasts and a method of randomization can be constructed such that the inequality (6) holds for all D from RKHS $\mathcal{F}$ and for all n. Also, for any $\delta > 0$ with probability at least $1 - \delta$*

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)\right| \le \Delta n + \sqrt{\frac{n}{\Delta^k}} + \sqrt{(c_{\mathcal{F}}^2+1)n} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}}$$

holds for all n, where I is the characteristic function of any measurable subset of $[0,1]^{k+1}$.

Proof. We define a deterministic forecast and after that we randomize it.

The partition $V=\{v_0,\dots,v_K\}$ and probabilities of rounding were defined above by (2). In what follows we round some deterministic forecast p_n to v_{i-1} with probability $w_{v_{i-1}}(p_n)$ and to v_i with probability $w_{v_i}(p_n)$. We also round the each coordinate $x_{n,s}$, $s=1,\dots k$, of a signal $\bar{x}_n$ to v_{j_s-1} with probability $w_{v_{j_s-1}}(x_{n,s})$ and to v_{j_s} with probability $w_{v_{j_s}}(x_{n,s})$, where $x_{n,s}\in[v_{j_s-1},v_{j_s}]$.

Let $W_v(p_n,\bar{x}_n)=w_{v^1}(p_n)w_{v^2}(\bar{x}_n)$, where $v=(v^1,v^2)$ and $v^1\in V$, $v^2=(v^2_1,\dots v^2_k)\in V^k$, $w_{v^2}(\bar{x}_n)=\prod_{s=1}^k w_{v^2_s}(x_{n,s})$, and $\bar{W}(p_n,\bar{x}_n)=(W_v(p_n,\bar{x}_n):v\in V^{k+1})$ be a vector of probability distribution in V^{k+1}. Define the corresponding kernel $K(p,\bar{x},p',\bar{x}')=(W(p,\bar{x})\cdot W(p',\bar{x}'))$.

Let the deterministic forecasts $p_1,\dots,p_{n-1}$ be already defined (put $p_1=1/2$). We want to define a deterministic forecast p_n.

By Mercer theorem the kernel $M(\bar{x},\bar{x}')$ can be represented as a dot product in some feature space: $M(\bar{x},\bar{x}')=(\Phi(\bar{x})\cdot\Phi(\bar{x}'))$. Consider

$$U_n(p,\bar{x}_n)=\sum_{i=1}^{n-1}(K(p,\bar{x}_n,p_i,\bar{x}_i)+M(\bar{x}_n,\bar{x}_i))(S_i-p_i).$$

The following lemma presents a general method for computing deterministic forecasts.

Lemma 1. *(Takemura, Vovk [13]) A sequence of forecasts $p_1,p_2,\dots$ can be computed such that $\mathcal{M}_n\le\mathcal{M}_{n-1}$ for all n, where $\mathcal{M}_0=1$ and $\mathcal{M}_n=\mathcal{M}_{n-1}+U_n(p_n,\bar{x}_n)(S_n-p_n)$ for all n.*

Proof. Indeed, if $U_n(p,\bar{x}_n)>0$ for all $p\in[0,1]$ then define $p_n=1$; if $U_n(p,\bar{x}_n)<0$ for all $p\in[0,1]$ then define $p_n=0$. Otherwise, define p_n to be some root of the equation $U_n(p,\bar{x}_n)=0$ (some root exists by the intermediate value theorem). Evidently, $M_n\le M_{n-1}$ for all n. △

Let forecasts $p_1,p_2,\dots$ be computed by the method of Lemma 1. Then for any N

$$0\ge\mathcal{M}_N-\mathcal{M}_0=\sum_{n=1}^N U_n(p_n,\bar{x}_n)(S_n-p_n)=$$

$$=\sum_{n=1}^N\sum_{i=1}^{n-1}(K(p_n,\bar{x}_n,p_i,\bar{x}_i)+M(\bar{x}_n,\bar{x}_i))(S_i-p_i)(S_n-p_n)=$$

$$=\frac{1}{2}\sum_{n=1}^N\sum_{i=1}^N K(p_n,\bar{x}_n,p_i,\bar{x}_i)(S_i-p_i)(S_n-p_n)-$$

$$-\frac{1}{2}\sum_{n=1}^N(K(p_n,\bar{x}_n,p_n,\bar{x}_n)(S_n-p_n))^2+$$

$$+\frac{1}{2}\sum_{n=1}^N\sum_{i=1}^N M(\bar{x}_n,\bar{x}_i)(S_i-p_i)(S_n-p_n)-$$

$$-\frac{1}{2}\sum_{n=1}^{N}(M(\bar{x}_n,\bar{x}_n)(S_n-p_n))^2= \quad (7)$$

$$=\frac{1}{2}\left\|\sum_{n=1}^{N}\bar{W}(p_n,\bar{x}_n)(S_n-p_n)\right\|^2-\frac{1}{2}\sum_{n=1}^{N}\|\bar{W}(p_n,\bar{x}_n)\|^2(S_n-p_n)^2+ \quad (8)$$

$$+\frac{1}{2}\left\|\sum_{n=1}^{N}\Phi(\bar{x}_n)(S_n-p_n)\right\|^2-\frac{1}{2}\sum_{n=1}^{N}\|\Phi(\bar{x}_n)\|^2(S_n-p_n)^2. \quad (9)$$

In (8), $\|\cdot\|$ is Euclidian norm, and in (9), $\|\cdot\|$ is a norm in RKHS $\mathcal{F}$.

Since $(S_n-p_n)^2\le 1$ for all n and

$$\|(\bar{W}(p_n,\bar{x}_n)\|^2=\sum_{v\in V^{k+1}}(W_v(p_n,\bar{x}_n))^2\le\sum_{v\in V^{k+1}}W_v(p_n,\bar{x}_n)=1,$$

the subtracted sum of (8) is upper bounded by N.

Since $\|\Phi(\bar{x}_n)\|=c_{\mathcal{F}}(\bar{x}_n)$ and $c_{\mathcal{F}}(\bar{x})\le c_{\mathcal{F}}$ for all $\bar{x}$, the subtracted sum of (9) is upper bounded by $c_{\mathcal{F}}^2N$. As a result we obtain

$$\left\|\sum_{n=1}^{N}\bar{W}(p_n,\bar{x}_n)(S_n-p_n)\right\|\le\sqrt{(c_{\mathcal{F}}^2+1)N} \quad (10)$$

$$\left\|\sum_{n=1}^{N}\Phi(\bar{x}_n)(S_n-p_n)\right\|\le\sqrt{(c_{\mathcal{F}}^2+1)N} \quad (11)$$

for all N. Let us denote $\bar{\mu}_n=\sum\limits_{i=1}^{n}\bar{W}(p_i,\bar{x}_i)(S_i-p_i)$. By (10), $\|\bar{\mu}_n\|\le\sqrt{(c_{\mathcal{F}}^2+1)n}$ for all n.

Let $\mu_n=(\mu_n(v):v\in V^{k+1})$. By definition for any v

$$\mu_n(v)=\sum_{i=1}^{n}W_v(p_i,\bar{x}_i)(S_i-p_i). \quad (12)$$

Insert the term $I(v)$ in the sum (12), where I is the characteristic function of an arbitrary set $\mathcal{S}\subseteq[0,1]^{k+1}$, sum by $v\in V^{k+1}$, and exchange the order of summation. Using Cauchy–Schwartz inequality for vectors $\bar{I}=(I(v):v\in V^{k+1})$, $\bar{\mu}_n=(\mu_n(v):v\in V^{k+1})$ and Euclidian norm, we obtain

$$\left|\sum_{i=1}^{n}\sum_{v\in V^{k+1}}W_v(p_i,\bar{x}_i)I(v)(S_i-p_i)\right|=$$

$$=\left|\sum_{v\in V^{k+1}}I(v)\sum_{i=1}^{n}W_v(p_i,\bar{x}_i)(S_i-p_i)\right|=$$

$$=(\bar{I}\cdot\bar{\mu}_n)\le\|\bar{I}\|\cdot\|\bar{\mu}_n\|\le\sqrt{|V^{k+1}|(c_{\mathcal{F}}^2+1)n} \quad (13)$$

for all n, where $|V^{k+1}| = 1/\Delta^{k+1}$ – is the cardinality of the partition.

Let $\tilde{p}_i$ be a random variable taking values $v \in V$ with probabilities $w_v(p_i)$ (only two of them are nonzero). Recall that $\tilde{x}_i$ is a random variable taking values $v \in V^k$ with probabilities $w_v(\bar{x}_i)$.

Let $\mathcal{S} \subseteq [0,1]^k$ and I be its indicator function. For any i the mathematical expectation of a random variable $I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)$ is equal to

$$E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) = \sum_{v \in V^{k+1}} W_v(p_i, \bar{x}_i) I(v)(S_i - v^1), \tag{14}$$

where $v = (v^1, v^2)$. By Azuma–Hoeffding inequality (see (26) below), for any $\delta > 0$, with Pr-probability $1 - \delta$

$$\left| \sum_{i=1}^{n} I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) - \sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) \right| \le \sqrt{\frac{n}{2} \ln \frac{2}{\delta}}. \tag{15}$$

By definition of the deterministic forecast

$$\left| \sum_{v \in V^{k+1}} W_v(p_i, \bar{x}_i) I(v)(S_i - p_i) - \sum_{v \in V^{k+1}} W_v(p_i, \bar{x}_i) I(v)(S_i - v^1) \right| < \Delta$$

for all i, where $v = (v^1, v^2)$. Summing (14) by $i = 1, \dots, n$ and using the inequality (13), we obtain

$$\left| \sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) \right| =$$
$$= \left| \sum_{i=1}^{n} \sum_{v \in V^{k+1}} W_v(p_i, \bar{x}_i) I(v)(S_i - v^1) \right| < \Delta n + \sqrt{(c_{\mathcal{F}}^2 + 1) n / \Delta^{k+1}} \tag{16}$$

for all n.

By (15) and (16), with Pr-probability $1 - \delta$

$$\left| \sum_{i=1}^{n} I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) \right| \le \Delta n + \sqrt{(c_{\mathcal{F}}^2 + 1) n / \Delta^{k+1}} + \sqrt{\frac{n}{2} \ln \frac{2}{\delta}}. \tag{17}$$

By Caushi–Schwartz inequality

$$\left| \sum_{n=1}^{N} D(\bar{x}_n)(S_n - p_n) \right| = \left| \sum_{n=1}^{N} (S_n - p_n)(D \cdot \Phi(\bar{x}_n)) \right| =$$
$$\left| \left(\sum_{n=1}^{N} (S_n - p_n) \Phi(\bar{x}_n) \cdot D \right) \right| \le \left\| \sum_{n=1}^{N} (S_n - p_n) \Phi(\bar{x}_n) \right\| \cdot \|D\| \le$$
$$\le \|D\| \sqrt{(c_{\mathcal{F}}^2 + 1) N}.$$

Proposition is proved.$\triangle$

Now we turn to the proof of Theorem 2.

The expression $\Delta n+\sqrt{(c^2_{\mathcal{F}}+1)n/\Delta^{k+1}}$ from (16) and (17) takes its minimal value for $\Delta=(\frac{k+1}{2})^{\frac{2}{k+3}}(c^2_{\mathcal{F}}+1)^{\frac{1}{k+3}}n^{-\frac{1}{k+3}}$. In this case, the right-hand side of the inequality (16) is equal to

$$\Delta n+\sqrt{n(c^2_{\mathcal{F}}+1)/\Delta^{k+1}}\le 2\Delta n=2\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}}(c^2_{\mathcal{F}}+1)^{\frac{1}{k+3}}n^{1-\frac{1}{k+3}}. \quad (18)$$

In what follows we use the upper bound $2\Delta n$ in (16).

To prove the bound (5) choose a monotonic sequence of rational numbers $\Delta_1>\Delta_2>\dots$ such that $\Delta_s\to 0$ as $s\to\infty$. We also define an increasing sequence of natural numbers $n_1<n_2<\dots$ For any s, we use for randomization on steps $n_s\le n<n_{s+1}$ the partition of $[0,1]$ on subintervals of length Δ_s.

We start our sequences from $n_1=1$ and $\Delta_1=1$. Also, define the numbers $n_2,n_3,\dots$ such that the inequality

$$\left|\sum\limits_{i=1}^{n}E(I(\tilde p_i,\tilde x_i)(S_i-\tilde p_i))\right|\le 4(s+1)\Delta_s n \quad (19)$$

holds for all $n_s\le n\le n_{s+1}$ and for all $s\ge 1$.

We define this sequence by mathematical induction on s. Suppose that n_s $(s\ge 1)$ is defined such that the inequality

$$\left|\sum\limits_{i=1}^{n}E(I(\tilde p_i,\tilde x_i)(S_i-\tilde p_i))\right|\le 4s\Delta_{s-1}n \quad (20)$$

holds for all $n_{s-1}\le n\le n_s$, and the inequality

$$\left|\sum\limits_{i=1}^{n_s}E(I(\tilde p_i,\tilde x_i)(S_i-\tilde p_i))\right|\le 4s\Delta_s n_s \quad (21)$$

also holds. Let us define n_{s+1}. Consider all forecasts $\tilde p_i$ defined by the algorithm given above for discretization $\Delta=\Delta_{s+1}$. We do not use first n_s of these forecasts (more correctly we will use them only in bounds (22) and (23); denote these forecasts $\mathbf{\hat p_1},\dots,\mathbf{\hat p_{n_s}}$). We add the forecasts $\tilde p_i$ for $i>n_s$ to the forecasts defined before this step of induction (for n_s). Let n_{s+1} be such that the inequality

$$\left|\sum\limits_{i=1}^{n_{s+1}}E(I(\tilde p_i,\tilde x_i)(S_i-\tilde p_i))\right|\le\left|\sum\limits_{i=1}^{n_s}E(I(\tilde p_i,\tilde x_i)(S_i-\tilde p_i))\right|+$$
$$+\left|\sum\limits_{i=n_s+1}^{n_{s+1}}E(I(\tilde p_i,\tilde x_i)(S_i-\tilde p_i))+\sum\limits_{i=1}^{n_s}E(I(\mathbf{\hat p_i},\tilde x_i)(S_i-\mathbf{\hat p_i}))\right|+$$
$$+\left|\sum\limits_{i=1}^{n_s}E(I(\mathbf{\hat p_i},\tilde x_i)(S_i-\mathbf{\hat p_i}))\right|\le 4(s+1)\Delta_{s+1}n_{s+1} \quad (22)$$

holds. Here the first sum of the right-hand side of the inequality (22) is bounded by $4s\Delta_s n_s$ – by the induction hypothesis (21). The second and third sums are bounded by $2\Delta_{s+1}n_{s+1}$ and by $2\Delta_{s+1}n_s$, respectively, where $\Delta = \Delta_{s+1}$ is defined such that (18) holds. This follows from (16) and by choice of n_s.

The induction hypothesis (21) is valid for

$$n_{s+1} \geq \frac{2s\Delta_s + \Delta_{s+1}}{\Delta_{s+1}(2s+1)} n_s.$$

Analogously,

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \leq \left|\sum_{i=1}^{n_s} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| +$$
$$+\left|\sum_{i=n_s+1}^{n} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i)) + \sum_{i=1}^{n_s} E(I(\hat{\mathbf{p}}_{\mathbf{i}},\tilde{x}_i)(S_i-\hat{\mathbf{p}}_{\mathbf{i}}))\right| +$$
$$+\left|\sum_{i=1}^{n_s} E(I(\hat{p}_i,\tilde{x}_i)(S_i-\hat{\mathbf{p}}_{\mathbf{i}}))\right| \leq 4(s+1)\Delta_s n \tag{23}$$

for $n_s < n \leq n_{s+1}$. Here the first sum of the right-hand inequality (22) is also bounded by $4s\Delta_s n_s \leq 4s\Delta_s n$ – by the induction hypothesis (21). The second and the third sums are bounded by $2\Delta_{s+1}n \leq 2\Delta_s n$ and by $2\Delta_{s+1}n_s \leq 2\Delta_s n$, respectively. This follows from (16) and from choice of Δ_s. The induction hypothesis (20) is valid.

By (19) for any s

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \leq 4(s+1)\Delta_s n \tag{24}$$

for all $n \geq n_s$ if Δ_s satisfies the condition $\Delta_{s+1} \leq \Delta_s(1-\frac{1}{s+2})$ for all s.

We show now that a sequences n_s and Δ_s satisfying all the conditions above exist.

Let $\epsilon > 0$ and $M = \lceil 2/\epsilon \rceil$, where $\lceil r \rceil$ is the least integer number $m \geq r$. Define $n_s = (s+M)^M$ and $\Delta_s = \left(\frac{k+1}{2}\right)^{\frac{2}{k+3}} (c_{\mathcal{F}}^2+1)^{\frac{1}{k+3}} n_s^{-\frac{1}{k+3}}$. Easy to verify that all requirements for n_s and Δ_s given above are satisfied.

We have in (24) for all $n_s \leq n < n_{s+1}$

$$4(s+1)\Delta_s n \leq 8s\Delta_s n_{s+1} =$$
$$= 8\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}} (c_{\mathcal{F}}^2+1)^{\frac{1}{k+3}} s(s+M+1)^M (s+M)^{-\frac{M}{k+3}} \leq$$
$$\leq 8e\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}} (c_{\mathcal{F}}^2+1)^{\frac{1}{k+3}} n_s^{1-\frac{1}{k+3}+2/M} \leq$$
$$\leq 8e\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}} (c_{\mathcal{F}}^2+1)^{\frac{1}{k+3}} n^{1-\frac{1}{k+3}+\epsilon},$$

where e is the base of the natural logarithm. Therefore, we obtain

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \le 8e\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}}(c_{\mathcal{F}}^2+1)^{\frac{1}{k+3}}n^{1-\frac{1}{k+3}+\epsilon} \quad (25)$$

for all n. Azuma–Hoeffding inequality says that for any $\gamma > 0$

$$Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} V_i\right| > \gamma\right\} \le 2e^{-2n\gamma^2} \quad (26)$$

for all n, where V_i are martingale–differences.

We get $V_i = I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i) - E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))$ and $\gamma = \sqrt{\frac{1}{2n}\ln\frac{2}{\delta}}$, where $\delta > 0$. Denote $\nu(n) = 8e\left(\frac{k+1}{2}\right)^{\frac{2}{k+3}}(c_{\mathcal{F}}^2+1)^{\frac{1}{k+3}}n^{1-\frac{1}{k+3}+\epsilon}$.

Combining (25) with (26), we obtain that for any $\delta > 0$ with probability $1-\delta$

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i)\right| \le \nu(n) + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}}$$

for all n. Theorem 2 is proved. $\triangle$

5 Proof of Theorem 1

Recall that $\epsilon > 0$ and $M = \lceil 2/\epsilon \rceil$. At any step i we compute the deterministic forecast p_i defined in Section 4 and its randomization to $\tilde{p}_i$ using parameters $\Delta = \Delta_s = (s+M)^{-\frac{M}{4}}$ and $n_s = (s+M)^M$, where $n_s \le i < n_{s+1}$. Let also, $\tilde{S}_{i-1}$ be a randomization of the past price S_{i-1}. At first, we bound

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i > \tilde{S}_{i-1})(\tilde{S}_{i-1}-S_{i-1})\right| \le \sum_{t=0}^{s}(n_{t+1}-n_t)\Delta_t \le$$
$$\le \frac{4}{3}(e-1)(c_{\mathcal{F}}^2+1)^{\frac{1}{4}}n_s^{\frac{3}{4}+\epsilon} \le \frac{4}{3}(e-1)(c_{\mathcal{F}}^2+1)^{\frac{1}{4}}n^{\frac{3}{4}+\epsilon}. \quad (27)$$

Let $D(\bar{x})$ be an arbitrary nonnegative traiding strategy from RKHS $\mathcal{F}$ such that $\|D\| \le l$. At any step n we use the randomized trading strategy $\tilde{M}_n$ defined in Section 3. We use abbreviations:

$$\nu_0(n) = lc_{\mathcal{F}}(c_{\mathcal{F}}^2+1)^{\frac{1}{4}}(e-1)\frac{4}{3}n^{\frac{3}{4}+\epsilon}, \quad (28)$$

$$\nu_1(n) = lc_{\mathcal{F}}\sqrt{\frac{n}{2}\ln\frac{6}{\delta}}, \quad (29)$$

$$\nu_2(n) = lc_{\mathcal{F}}\left(8en^{\frac{3}{4}+\epsilon}(c_{\mathcal{F}}^2+1)^{\frac{1}{4}} + \sqrt{\frac{n}{2}\ln\frac{8}{\delta}}\right), \quad (30)$$

$$\nu_3(n) = lc_{\mathcal{F}}\sqrt{(c_{\mathcal{F}}^2+1)n} \quad (31)$$

All sums below are for $i = 1, \dots n$. By definition $0 \le D(\bar{x}_i) \le c_{\mathcal{F}} l$. We use below the Azuma–Hoeffding inequality (26).

For any $\delta > 0$, with probability $1 - \delta$ the following chain of equalities and inequalities is valid:

$$\sum_{\tilde{p}_i > \tilde{S}_{i-1}} c_{\mathcal{F}} l (S_i - S_{i-1}) = c_{\mathcal{F}} l \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (S_i - \tilde{p}_i) +$$

$$+ c_{\mathcal{F}} l \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{p}_i - \tilde{S}_{i-1}) + c_{\mathcal{F}} l \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{S}_{i-1} - S_{i-1}) \ge \tag{32}$$

$$\ge c_{\mathcal{F}} l \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{p}_i - \tilde{S}_{i-1}) - \nu_0(n) - \nu_2(n) \ge \tag{33}$$

$$\ge \sum_{i=1}^{n} D(\bar{x}_i)(\tilde{p}_i - \tilde{S}_{i-1}) - \nu_0(n) - \nu_2(n) =$$

$$= \sum_{i=1}^{n} D(\bar{x}_i)(p_i - S_{i-1}) + \sum_{i=1}^{n} D(\bar{x}_i)(\tilde{p}_i - p_i) -$$

$$- \sum_{i=1}^{n} D(\bar{x}_i)(\tilde{S}_{i-1} - S_{i-1}) - \nu_0(n) - \nu_2(n) \ge \tag{34}$$

$$\ge \sum_{i=1}^{n} D(\bar{x}_i)(p_i - S_{i-1}) - \nu_0(n) - 2\nu_1(n) - \nu_2(n) \ge \tag{35}$$

$$\ge \sum_{i=1}^{n} D(\bar{x}_i)(p_i - S_{i-1}) + \sum_{i=1}^{n} D(\bar{x}_i)(S_i - p_i) -$$

$$- \nu_0(n) - 2\nu_1(n) - \nu_2(n) - \nu_3(n) = \tag{36}$$

$$= \sum_{i=1}^{n} D(\bar{x}_i)(S_i - S_{i-1}) - \nu_0(n) - 2\nu_1(n) - \nu_2(n) - \nu_3(n).$$

In change from (32) to (33) the inequality (5) of Theorem 2 and the bound (27) were used, and so, terms (28) and (30) were subtracted. In change from (34) to (35) Azuma–Hoeffding inequality (26) (where $\gamma = \delta/3$) was applied twice to intermediate terms, and so, term (29) was subtracted twice. In change from (35) to (36) the inequality (6) of Theorem 2 was used, and so, term (31) was subtracted. Therefore, we have (4).

The inequality (3) follows from (4). Theorem 1 is proved. △

6 Numerical experiments

In the numerical experiments, we have used historical data in form of per minute time series of prices of “randomly” chosen ten stocks (four of them are from Russian Stock Market). Data has been downloaded from FINAM site: *www.finam.ru*. Number of trading points in each game is 88000–116000 min. (From March 26

2010 to March 25 2011) In our experiments, we dealing only for a rise starting with the same initial capital $\mathcal{K}_0$.

We implement the trading strategy defined in Section 3. In this strategy we used two forecasting algorithms: calibration algorithm constructed in Section 4 (DF–model) and Autoregressive Moving Average algorithm (ARMA–model) of Jyh-Ying Peng and Aston [7].

Results of numerical experiments are shown in Table 1. In the first column, stocks ticker symbols are shown. The second column contains the gain of Buy-and-Hold trading strategy. By this strategy, we buy a holdings of shares using capital $\mathcal{K}_0$ and sell them at the end of the trading period. In the next pair of columns marked DF the relative returns (in percentage wise on initial capital) are presented for the case where no transaction cost is subtracted and for the case where transaction cost at the rate 0.01% is subtracted. The method of calibration (defensive forecasting – DF) for computing forecasts of future stock price was used for these columns. We buy a holdings of shares using the cumulative capital but no more than $\mathcal{K}_0$ at steps where the forecast show a rise and sell it at steps where the forecast show a fall. Thereby, we set aside the extra capital. The next two columns marked by ARMA are analogous but ARMA forecasting model was used for computing forecasts. The frequencies of market entry steps are given in the next two columns marked In (for DF and ARMA). The average duration of a gamble is shown in the rest two columns marked Delay (for DF and ARMA).

These results show that trading based on DF model of forecasting essentially outperforms trading based on ARMA model.

Table 1.

Ticker	Buy& hold Gain %	DF Gain %	DF -0.01% Gain %	ARMA Gain %	ARMA -0.01% Gain %	DF In	ARMA In	DF Delay	ARMA Delay
AT-T	7.73	42.53	-60.20	0.45	-7.06	0.0913	0.0050	1.97	2.13
CTGR	14.87	302.79	202.38	38.08	31.10	0.0638	0.0045	2.16	1.89
KOCO	16.55	27.32	-64.83	-1.04	-9.06	0.0942	0.0050	1.93	1.30
GOOG	10.27	35.06	-61.47	-0.19	-1.62	0.0906	0.0010	1.79	1.77
InBM	24.25	52.05	-54.71	5.53	-12.20	0.0895	0.0118	1.84	1.89
INTL	4.29	20.96	-45.56	3.09	2.74	0.0879	0.0004	2.00	1.45
MTSI	-1.51	273.05	130.65	29.58	21.47	0.0669	0.0038	2.54	2.10
SBER	14.70	36.20	-76.34	1.11	0.91	0.0955	0.0001	1.92	1.82
SIBN	-6.35	409.31	263.76	55.17	48.03	0.0684	0.0034	2.82	2.65
GAZP	22.97	38.72	-77.28	0.55	0.15	0.1000	0.0002	1.88	1.50

7 Conclusion

Calibration is an area of intensive research where several algorithms for computing calibrated forecasts have been developed. Several applications of well calibrated forecasting have been proposed (convergence to correlated equilibrium, recovering unknown functional dependencies, predictions with expert advice). We present a new application of the calibration method.

We show that an asymptotically optimal trading strategy can be constructed using the well calibrated forecasts. We prove that this strategy outperforms any trading strategy presented by a rule from any RKHS. To construct optimal trading strategy, we generalize Kakade and Foster's algorithm and combine it with Vovk's DF–model for arbitrary RKHS. Using Vovk's [15] theory of defensive forecasting in Banach spaces, these results can be generalized to these spaces.

Numerical experiments show a positive return for all chosen stocks, and for three of them we receive a gain even when transaction costs are subtracted. Results of this type can be useful for technical analysis in finance.

References

1. Blackwell, D.: An analog of the minimax theorem for vector payoffs. Pacific Journal of Mathematics, 6, 1956, 1–8
2. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning and Games. Cambridge University Press, 2006
3. Dawid, A.P.: The well-calibrated Bayesian [with discussion]. J. Am. Statist. Assoc. 77, 1982, 605–613
4. Dawid, A.P.: Calibration-based empirical probability [with discussion]. Ann. Statist., 13, 1985, 1251–1285
5. Chernov, A., Kalnishkan, Y., Zhdanov, F., Vovk, V.: Supermartingales in Prediction with Expert Advice. Lecture Notes in Computer Science, 5254, 2008, 199–213
6. Foster, D.P., Vohra, R.: Asymptotic calibration. Biometrika, 85, 1998, 379–390
7. Peng, Jyh-Ying, Aston, J.A.D.: The State Space Models Toolbox for MATLAB. Journal of Statistical Software, 41, Issue 6, 2011
8. Kakade, S.M., Foster, D.P.: Deterministic calibration and Nash equilibrium. LNCS 3120 (John Shawe Taylor and Yoram Singer. ed), 2004, 33–48
9. Oakes, D.: Self-Calibrating Priors Do not Exist [with discussion]. J. Am. Statist. Assoc. 80, 1985, 339–342
10. Scholkopf, B., Smola, A.: Learning with Kernels. MIT Press, Cambridge, MA, 2002
11. Shafer, G., Vovk, V.: Probability and Finance. It's Only a Game! New York: Wiley, 2001
12. Vovk, V.: A game of prediction with expert advice. Journal of Computer and System Sciences, 56, Issue 2, 1998, 153–173
13. Vovk, V., Takemura, A., Shafer, G.: Defensive forecasting. Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics (ed. by R. G. Cowell and Z. Ghahramani) – Cambridge UK: Society for Artificial Intelligence and Statistics, 2005, 365–372
14. Vovk, V.: On-line regression competitive with reproducing kernel Hilbert spaces (extended abstract). TAMS Lecture Notes in Computer Science – Berlin: Springer, 3959, 2006, 452–463
15. Vovk, V.: Competing with wild prediction rules. Machine Learning, 69, Issue 2-3, 2007, 193–212
16. Vovk, V.: Predictions as statements and decisions. Theoretical Computer Science, 405, Issue 3, 2008, 285–296
17. Vovk, V.: Defensive Forecasting for Optimal Prediction with Expert Advice. arXiv:0708.1503v1 2007