Journal of Investment Strategies Volume 2/Number 1, Winter 2012/13 (63–88)

Universal algorithmic trading

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

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

The problem of universal sequential investment in stock markets is considered. We construct an algorithmic trading strategy that is asymptotically at least as good as any trading strategy that is not excessively complex and that computes the investment at each step using a fixed continuous function of the side information. This strategy uses predictions of stock prices computed using the theory of well-calibrated forecasting. Unlike in statistical theory, no stochastic assumptions are made about stock prices. The empirical results obtained on historical markets provide strong evidence that this type of technical trading can "beat" some generally accepted trading strategies if transaction costs are ignored.

1 INTRODUCTION

We study the problem of universal sequential investment in a stock market with side information. We consider a trading method called algorithmic trading, or systematic quantitative trading. This method is used in financial industrial applications and involves rule-based automatic trading strategies, usually implemented using computer-based trading systems.

The problem of algorithmic trading is considered in a machine learning framework, where algorithms that are adaptive to input data are designed and their performance is evaluated.

There are three common types of analysis for adaptive algorithms: average case analysis, which requires a statistical model of input data; worst-case analysis, which is noninformative because, for any trading algorithm, we can present a sequence of stock prices moving in the opposite direction to the trader's decisions; and competitive analysis, which is popular in the "prediction with expert advice" framework.

A nontraditional objective (in computational finance) is to develop algorithmic trading strategies that are, in some sense, always guaranteed to perform well. In competitive analysis, the performance of an algorithm is measured against any trading algorithm from a broad class. We only ask that an algorithm performs well in the worst case, ie, relative to the input data of any type. Given a particular performance measure, an adaptive algorithm is strongly competitive against a class of trading algorithms if it achieves the maximum possible regret over all input sequences. Unlike in statistical theory, no stochastic assumptions are made about the stock prices.

This line of research in finance was pioneered by Cover (see Cover and Gluss (1986); Cover (1991); and Cover and Ordentlich (1996)), who designed universal portfolio selection algorithms that are proven to perform well (in terms of their total return) with respect to some adaptive online or offline benchmark algorithms. Such algorithms are called universal.

In this framework we consider a universal trading for one stock. We construct a "universal" strategy for algorithmic trading in the stock market that performs at least as well as any trading strategy that is not excessively complex. By "performance" we mean the return per unit of currency on an investment.

The process proceeds as follows. Observing a sequence of past prices of a stock and the side information, a forecaster assigns a subjective estimate to a future price. Then, using this forecast, a trader makes a decision on whether to buy or sell shares of the stock.

The forecasting method is based on Dawid's notion of calibration with more general checking rules and on a modification of Kakade and Foster's randomized algorithm for computing well-calibrated forecasts (Dawid (1982); Foster and Vohra (1998)).

Asymptotic calibration is an area of intensive research where several algorithms for computing well-calibrated forecasts have been developed. Several applications of well-calibrated forecasting have been proposed, including convergence to correlated equilibrium (Kakade and Foster (2008)), recovering unknown functional dependencies (Vovk (2006)) and predictions with expert advice (Chernov *et al* (2010)).

We present a new application: a universal trading strategy based on well-calibrated forecasts.

A minimal requirement for testing any prediction algorithm is that it should be calibrated. Dawid (1982) gave the following informal explanation of calibration for binary outcomes. 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 the probability that it will rain. The forecast is said to be well-calibrated if it rains as often as the forecaster 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 $I \subseteq [0, 1]$, ie, $I(p) = 1$ if $p \in I$, and $I(p) = 0$ otherwise. An infinite sequence of forecasts $p_1, p_2, \dots$ is calibrated for an infinite sequence of outcomes $\omega_1, \omega_2, \dots$ (binary or real) if, for a characteristic function $I(p)$ of any subinterval of

[0, 1], the calibration error tends to zero, ie:

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

The indicator function $I(p_i)$ determines some "checking rule" that 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 (1985) and Dawid (1985), any deterministic forecasting algorithm will not always be calibrated.

Foster and Vohra (1998) show that calibration is almost surely guaranteed with a randomizing forecasting rule, ie, where the forecasts p_i are chosen using internal randomization and the forecasts are hidden from the weather until it makes its "decision" whether to rain or not.

Unlike elsewhere in the literature, we consider real-valued outcomes ω_i (for example, scaled prices of a stock). In this case, predictions can be interpreted as mean values of future outcomes under some probability distributions that are unknown to us. We do not need to know the precise form of such distributions – we should only predict future means.

We use the forecasting method of Kakade and Foster (2008), where an "almost deterministic" randomized rounding universal forecasting algorithm is presented. For any sequence of real outcomes $\omega_1, \omega_2, \dots$, an observer can simply randomly round the deterministic forecast p_i up to some precision bound to a random forecast $\tilde{p}_i$ in order to calibrate for this sequence, with probability 1:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I(\tilde{p}_i)(\omega_i - \tilde{p}_i) = 0 \tag{1.1}$$

where $I(p)$ is the characteristic function of any subinterval of [0, 1].

In this paper a forecast is a single number from the unit interval [0, 1] (which can be the output of a random variable), whereas Kakade and Foster considered a finite outcome space and a probability distribution on this space as a forecast.

We show that, using well-calibrated forecasts, it is possible to construct a universal trading strategy for the stock market that performs asymptotically at least as well as any trading strategy presented by a continuous function. The learning process is as follows. At each step, a forecaster makes a prediction of a future price of the stock. A trader randomizes this prediction and the past price of the stock and takes the best response in terms of these randomized values. He chooses a strategy: going long or going short, or skips the step. In finance, a long position in a security, such as a stock or a bond, or equivalently to be long in a security, means the holder of the position owns the security and will profit if the price of the security goes up. Short selling (also known as shorting or going short) is the practice of selling securities or other financial instruments, with the intention of subsequently repurchasing them ("covering") at a lower price price. The forecaster uses a randomized algorithm for computing the well-calibrated forecasts. Our main result, Theorem 2.1, says that this trading strategy is universal – it performs asymptotically at least as well as any trading strategy presented by any continuous function. To achieve this result, in Theorem 4.1 we extend Kakade and Foster's forecasting algorithm for arbitrary real-valued outcomes to a more general notion of calibration with changing checking rules. We combine it with the Vovk *et al* (2005) defensive forecasting method in a reproducing kernel Hilbert space (RKHS) (see Vovk (2006)).

In Section 6 we present the results of our numerical experiments. Our empirical results on historical markets provide strong evidence that this type of technical trading can "beat the market" if transaction costs are ignored.

2 UNIVERSAL TRADING STRATEGY

We consider a trading process in a stock market with one stock. Suppose that prices $S_1, S_2, \dots$ of the stock are bounded and rescaled such that $0 \leqslant S_i \leqslant 1$ for all i. We present the process of trading in a stock market in the form of the perfect information protocol of a game with two traders. Trader M uses the randomized strategy $\tilde{M}_i$, Trader D uses an arbitrary stationary trading strategy D that is a real function defined on $[0, 1]$.

In this paper, a forecast is a single number from the unit interval $[0, 1]$ (which can be the output of a random variable). We could interpret the forecast p_i as the mean value of a future price S_i under some probability distribution in $[0, 1]$ that is unknown to us.

In the perfect-information protocol, every player can see the other players' moves so far. At the beginning of each step i, traders are given some data x_i that is relevant to predicting a future price S_i of the stock. We call x_i a signal or a piece of side information. The real number x_i belongs to $[0, 1]$ and can encode any information. For example, it could even be the future price S_i.

Past prices, signals and predictions are also known to traders in the perfect-information protocol. For example, at any step i, past prices and predictions can be encoded in the signal x_i and used by Trader D.

Define a universal trading strategy as a random variable $\tilde{M}_i$. To construct such a strategy, at each step i, 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 sequential randomization will be given in Section 4. Define:

$$\tilde{M}_i = \begin{cases} 1 & \text{if } \tilde{p}_i > \tilde{S}_{i-1} \\ -1 & \text{otherwise} \end{cases}$$

TABLE 1 Protocol of the trading game.

FOR $i = 1, 2 \dots$

Stock market announces a signal $x_i \in [0, 1]$.

Trader M bets by buying or selling the random number $\tilde{M}_i$ of shares of the stock by S_{i-1} each.

Trader D bets by buying or selling a number $D(x_i)$ of shares of the stock by S_{i-1} each.

Stock market reveals a price S_i of the stock.

Trader M updates his total gain (loss):

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

Trader D updates his total gain (loss):

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

ENDFOR

In case $\tilde{M}_i > 0$ Trader M goes long, and goes short otherwise. The same holds for Trader D. We suppose that traders can borrow money for buying shares and can incur debt. The core of the universal strategy is the algorithm for computing the well-calibrated forecasts $\tilde{p}_i$. This algorithm is presented in Section 4.

Recall the norm $\|D\|_\infty = \sup_{0 \leqslant x \leqslant 1} |D(x)|$, where D is a continuous function. Assume that $S_1, S_2, \dots \in [0, 1]$ and $x_1, x_2, \dots \in [0, 1]$ are given sequentially according to the protocol presented in Table 1.

The main theoretical result of this paper is presented in the following theorem. It says that, with probability one, the average gain of the universal trading strategy is asymptotically not less than the average gain of stationary trading strategy from one share of the stock.

THEOREM 2.1 *An algorithm for computing forecasts and a sequential method of randomization can be constructed such that, for any continuous nonzero function D:*

$$\liminf_{n\to\infty} \frac{1}{n}(\mathcal{K}_n^{\mathrm{M}} - \|D\|_\infty^{-1}\mathcal{K}_n^{\mathrm{D}}) \geqslant 0 \tag{2.1}$$

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

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.

3 REPRODUCING KERNEL HILBERT SPACES

First, we will prove in Section 5 that the trading strategy $\tilde{M}_i$ defined in Section 2 is universal with respect to all stationary trading strategies from any benchmark class called an RKHS. After that, using a universal RKHS, we extend the universality property to the class of all continuous stationary trading strategies.

A Hilbert space $\mathcal{F}$ of real-valued functions on a compact set X is called an RKHS on X if the evaluation functional $f \to f(x)$ is continuous for each $x \in X$. Let $\|\cdot\|_{\mathcal{F}}$ be a norm on $\mathcal{F}$ and let:

$$c_{\mathcal{F}}(x) = \sup_{\|f\|_{\mathcal{F}} \leqslant 1} |f(x)|$$

The embedding constant of $\mathcal{F}$ is defined as $c_{\mathcal{F}} = \sup_x c_{\mathcal{F}}(x)$.

We consider an RKHS $\mathcal{F}$ on $X = [0, 1]$ with $c_{\mathcal{F}} < \infty$. An example of such an RKHS is the Sobolev space $\mathcal{F} = H^1([0, 1])$, which consists of absolutely continuous functions $f\colon [0, 1] \to \mathcal{R}$ with $\|f\|_{\mathcal{F}} < \infty$, where:

$$\|f\|_{\mathcal{F}} = \sqrt{\int_0^1 (f(t))^2 \, \mathrm{d}t + \int_0^1 (f'(t))^2 \, \mathrm{d}t}$$

For this space, $c_{\mathcal{F}} = \sqrt{\coth 1}$ (see Vovk (2006)).

Let $\mathcal{F}$ be an RKHS on X with the dot product $(f \cdot g)$ for $f, g \in \mathcal{F}$. By the 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 as $K(x, y) = (k_x \cdot k_y)$.

Conversely, any kernel $K(x, y)$ defines some canonical RKHS $\mathcal{F}$ and a mapping $\Phi\colon X \to \mathcal{F}$ such that $K(x, y) = (\Phi(x) \cdot \Phi(y))$. For other examples and details of kernel theory, see Scholkopf and Smola (2002).

4 WELL-CALIBRATED FORECASTING WITH SIDE INFORMATION

In this section we define an algorithm for computing well-calibrated forecasts, which is a core of the investment strategy $\tilde{M}_i$ defined in Section 2.

We consider checking rules of a more general type than those used by Kakade and Foster (2008). For any measurable subset $\mathcal{S} \subseteq [0, 1] \times [0, 1]$, define:

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

where $p, x \in [0, 1]$. In Section 5 we obtain $S = \{(p, x)\colon p > x\}$.

Some special kernel corresponds to the method of randomization defined below. A random variable $\tilde{y}$ is called a 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 probability distribution corresponding to $\tilde{y}$.

We use a specific method of randomization of real numbers from the unit interval proposed by Kakade and Foster (2008). Given a positive integer number K, divide the interval [0, 1] into subintervals of length $\Delta = 1/K$ with rational endpoints $v_i = i\Delta$, where $i = 0, 1, \dots, K$. Let V denote 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 the subinterval containing p:

$$p = \sum_{v \in V} w_v(p)v = w_{v_{i-1}}(p)v_{i-1} + w_{v_i}(p)v_i \tag{4.1}$$

where:

$$p \in [v_{i-1}, v_i], \quad i = \lfloor p^1/\Delta + 1 \rfloor,$$
$$w_{v_{i-1}}(p) = 1 - (p - v_{i-1})/\Delta, \qquad w_{v_i}(p) = 1 - (v_i - p)/\Delta$$

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 $\bar{w}(p) = (w_v(p) \colon v \in V)$ be a vector of probabilities of rounding. For $z, z' \in [0, 1]$, define the dot product $K(z, z') = (\bar{w}(z) \cdot \bar{w}(z'))$, which is a kernel function.[1]

In what follows, given a sequence $\Delta_1 \geqslant \Delta_2 \geqslant \cdots \to 0$ of rational numbers, we will define the corresponding randomizations $\tilde{p}_1, \tilde{p}_2, \dots$ of reals $p_1, p_2, \dots \in [0, 1]$ such that $E_i(\tilde{p}_i) = p_i$ and the standard deviation $\sigma_i = \sqrt{E_i(\tilde{p}_i - p_i)^2} \leqslant \Delta_i$ for all i, where E_i is the symbol of the mathematical expectation with respect to the random rounding up to Δ_i. Similarly, a sequence of random variables $\tilde{z}_1, \tilde{z}_2, \dots$ will be defined as randomizations of real numbers $z_1, z_2, \dots$, where $z_i = S_{i-1}$ is the past outcomes, $i = 1, 2, \dots$ (setting $S_0 = 0$). 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 used in the analysis presented in Section 5.

Let $\mathcal{F}$ be an RKHS on [0, 1] with a finite embedding constant $c_{\mathcal{F}}$, let $\|\cdot\|_{\mathcal{F}}$ be the norm and let $M(x, x')$ be the kernel on [0, 1].

Suppose that $S_1, S_2, \dots \in [0, 1]$ is a sequence of real numbers and that $x_1, x_2, \dots \in [0, 1]$ is a sequence of signals given sequentially according to the protocol presented in Table 1 on page 67.

THEOREM 4.1 *Given $\epsilon > 0$, an algorithm for computing forecasts $p_1, p_2, \dots$ and a sequential method of randomization can be constructed such that the following two conditions hold.*

---

[1] Many other methods of randomization also work.

(i) *For the characteristic function* I_S *of any subset* $S \subseteq [0.1]^2$ *and for any* $\delta > 0$, *with* Pr *at least* $1 - \delta$*:*

$$\left|\sum_{i=1}^{n} I_S(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i)\right| = O\left(n^{(3/4)+\epsilon}(c_{\mathcal{F}}^2 + 1)^{1/4} + \left(n \ln\left(\frac{1}{\delta}\right)\right)^{1/2}\right) \quad (4.2)$$

as $n \to \infty$, *where* $\tilde{p}_1, \tilde{p}_2, \dots$ *are the corresponding randomizations of* $p_1, p_2, \dots$ *and* $\tilde{z}_1, \tilde{z}_2, \dots$ *are randomizations of reals* $z_1, z_2, \dots$, *where* $z_i = S_{i-1}$, $i = 1, 2, \dots$.

(ii) *For any* $D \in \mathcal{F}$*:*

$$\left|\sum_{i=1}^{n} D(x_i)(S_i - p_i)\right| \leqslant \|D\|_{\mathcal{F}} \sqrt{(c_{\mathcal{F}}^2 + 1)n} \qquad (4.3)$$

for all n.

Proof At first, given $\Delta > 0$, we modify a randomized rounding algorithm of Kakade and Foster (2008) to construct some forecasting algorithm calibrated up to a precision Δ, and combine it with the Vovk (2006) defensive forecasting algorithm. After that, we apply a "doubling trick" argument to this algorithm so that (4.2) will hold.

Under the assumption of Theorem 4.1, an algorithm for computing forecasts and a method of randomization can be constructed such that inequality (4.3) holds for all D from the RKHS $\mathcal{F}$ and for all n. Also, for any $\delta > 0$ with probability at least $1 - \delta$:

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

holds for all n, where I is the characteristic function of any measurable subset of $[0, 1] \times [0, 1]$.

We define a deterministic forecast p_i and after that we randomize it to $\tilde{p}_i$.

The partition $V = \{v_0, \dots, v_K\}$ and probabilities of rounding were defined above by (4.1). 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 randomly round z_n to v_{s-1} with probability $w_{v_{s-1}}(z_n)$ and to v_s with probability $w_{v_s}(z_n)$, where $z_n \in [v_{s-1}, v_s]$.

Let $W_v(p_n, z_n) = w_{v^1}(p_n) w_{v^2}(z_n)$, where $v = (v^1, v^2)$ and $v^1, v^2 \in V$, and let $W(p_n, z_n) = (W_v(p_n, z_n) \colon v \in V^2)$ be a vector of probability distribution in V^2. Define the corresponding kernel $K(p, z, p', z') = (W(p, z) \cdot W(p', z'))$.

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

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

$$U_n(p) = \sum_{i=1}^{n-1} (K(p, z_n, p_i, z_i) + M(x_n, x_i))(S_i - p_i) \tag{4.4}$$

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

LEMMA 4.2 (Vovk *et al* (2005)) *A sequence of forecasts $p_1, p_2, \dots$ can be computed such that $\mathcal{M}_n \leqslant \mathcal{M}_{n-1}$ for all n, where $\mathcal{M}_0 = 1$ and $\mathcal{M}_n = \mathcal{M}_{n-1} + U_n(p_n)(S_n - p_n)$ for all n.*

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

Let forecasts $p_1, p_2, \dots$ be computed using the method in Lemma 4.2. Then, for any N:

$$\begin{aligned}
0 &\geqslant \mathcal{M}_N - \mathcal{M}_0 \\
&= \sum_{n=1}^{N} U_n(p_n)(S_n - p_n) \\
&= \sum_{n=1}^{N} \sum_{i=1}^{n-1} (K(p_n, z_n, p_i, z_i) + M(x_n, x_i))(S_i - p_i)(S_n - p_n) \\
&= \frac{1}{2} \sum_{n=1}^{N} \sum_{i=1}^{N} K(p_n, z_n, p_i, z_i)(S_i - p_i)(S_n - p_n) \\
&\quad - \frac{1}{2} \sum_{n=1}^{N} (K(p_n, z_n, p_n, z_n)(S_n - p_n))^2 \\
&\quad + \frac{1}{2} \sum_{n=1}^{N} \sum_{i=1}^{N} M(x_n, x_i)(S_i - p_i)(S_n - p_n) \\
&\quad - \frac{1}{2} \sum_{n=1}^{N} (M(x_n, x_n)(S_n - p_n))^2
\end{aligned} \tag{4.5}$$

$$= \frac{1}{2}\left\| \sum_{n=1}^{N} W(p_n, z_n)(S_n - p_n) \right\|^2$$

$$- \tfrac{1}{2}\sum_{n=1}^{N} \|W(p_n, z_n)\|^2 (S_n - p_n)^2 \tag{4.6}$$

$$+ \tfrac{1}{2}\left\| \sum_{n=1}^{N} \Phi(x_n)(S_n - p_n) \right\|_{\mathcal{F}}^2 - \frac{1}{2}\sum_{n=1}^{N} \|\Phi(x_n)\|_{\mathcal{F}}^2 (S_n - p_n)^2 \tag{4.7}$$

In (4.6), $\|\cdot\|$ is the Euclidean norm, and in (4.7), $\|\cdot\|_{\mathcal{F}}$ is a norm in the RKHS $\mathcal{F}$.

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

$$\begin{aligned} \|W(p_n, z_n)\|^2 &= \sum_{v \in V^2} (W_v(p_n, z_n))^2 \\ &\leqslant \sum_{v \in V^2} W_v(p_n, z_n) = 1 \end{aligned}$$

the subtracted sum of (4.6) is upper bounded by N.

Since $\|\Phi(x_n)\|_{\mathcal{F}} = c_{\mathcal{F}}(x_n)$ and $c_{\mathcal{F}}(x) \leqslant c_{\mathcal{F}}$ for all x, the subtracted sum of (4.7) is upper bounded by $c_{\mathcal{F}}^2 N$. As a result, we obtain:

$$\left\| \sum_{n=1}^{N} W(p_n, z_n)(S_n - p_n) \right\| \leqslant \sqrt{(c_{\mathcal{F}}^2 + 1)N} \tag{4.8}$$

$$\left\| \sum_{n=1}^{N} \Phi(x_n)(S_n - p_n) \right\|_{\mathcal{F}} \leqslant \sqrt{(c_{\mathcal{F}}^2 + 1)N} \tag{4.9}$$

for all N. Let us define:

$$\bar{\mu}_n = \sum_{i=1}^{n} W(p_i, z_i)(S_i - p_i)$$

By (4.8), $\|\bar{\mu}_n\| \leqslant \sqrt{(c_{\mathcal{F}}^2 + 1)n}$ for all n.

Let $\bar{\mu}_n = \{\mu_n(v) \colon v \in V^2\}$. By definition, for any v:

$$\mu_n(v) = \sum_{i=1}^{n} W_v(p_i, z_i)(S_i - p_i) \tag{4.10}$$

Insert the term $I(v)$ into the sum (4.10), where I is the characteristic function of an arbitrary set $\mathcal{S} \subseteq [0,1] \times [0,1]$, sum by $v \in V^2$ and exchange the order of summation. Using the Cauchy–Schwarz inequality for vectors $\bar{I} = (I(v) \colon v \in V^2)$,

$\bar{\mu}_n = (\mu_n(v) \colon v \in V^2)$ and the Euclidean norm, we obtain:

$$\begin{aligned}\left|\sum_{i=1}^{n}\sum_{v\in V^2} W_v(p_i,z_i)I(v)(S_i-p_i)\right| &= \left|\sum_{v\in V^2} I(v)\sum_{i=1}^{n} W_v(p_i,z_i)(S_i-p_i)\right| \\ &= (\bar{I}\cdot\bar{\mu}_n) \leqslant \|\bar{I}\|\cdot\|\bar{\mu}_n\| \\ &\leqslant \sqrt{|V^2|(c_{\mathcal{F}}^2+1)n} \qquad (4.11)\end{aligned}$$

for all n, where $|V^2| = (1/\Delta+1)^2 \leqslant 4/\Delta^2$ is the cardinality of the partition.

Let $\tilde{p}_i$ be a random variable taking values $v \in V$ with probabilities $w_v(p_i)$. Recall that $\tilde{z}_i$ is a random variable taking values $v \in V$ with probabilities $w_v(z_i)$.

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

$$E(I(\tilde{p}_i,\tilde{z}_i)(S_i-\tilde{p}_i)) = \sum_{v\in V^2} W_v(p_i,z_i)I(v)(S_i-v^1) \qquad (4.12)$$

where $v = (v^1, v^2)$. By the Azuma–Hoeffding inequality (see (4.20) below), for any $\delta > 0$, with probability $1-\delta$:

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i,\tilde{z}_i)(S_i-\tilde{p}_i) - \sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{z}_i)(S_i-\tilde{p}_i))\right| \leqslant \sqrt{\frac{n}{2}\ln\frac{2}{\delta}} \qquad (4.13)$$

By definition of the deterministic forecast, the sums:

$$\sum_{v\in V^2} W_v(p_i,z_i)I(v)(S_i-p_i) \quad \text{and} \quad \sum_{v\in V^2} W_v(p_i,z_i)I(v)(S_i-v^1)$$

differ at most by Δ for all i, where $v = (v^1, v^2)$. Summing (4.12) over $i = 1,\dots,n$ and using inequality (4.11), we obtain:

$$\begin{aligned}\left|\sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{z}_i)(S_i-\tilde{p}_i))\right| &= \left|\sum_{i=1}^{n}\sum_{v\in V^2} W_v(p_i,z_i)I(v)(S_i-v^1)\right| \\ &\leqslant \Delta n + 2\sqrt{\frac{(c_{\mathcal{F}}^2+1)n}{\Delta^2}} \qquad (4.14)\end{aligned}$$

for all n.

By (4.13) and (4.14), with probability $1-\delta$:

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i,\tilde{z}_i)(S_i-\tilde{p}_i)\right| \leqslant \Delta n + 2\sqrt{\frac{(c_{\mathcal{F}}^2+1)n}{\Delta^2}} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}} \qquad (4.15)$$

By the Cauchy–Schwarz inequality:

$$\begin{aligned}\left|\sum_{n=1}^{N} D(x_n)(S_n - p_n)\right| &= \left|\sum_{n=1}^{N}(S_n - p_n)(D \cdot \Phi(x_n))\right| \\ &= \left|\left(\sum_{n=1}^{N}(S_n - p_n)\Phi(x_n) \cdot D\right)\right| \\ &\leqslant \left\|\sum_{n=1}^{N}(S_n - p_n)\Phi(x_n)\right\| \cdot \|D\|_{\mathcal{F}} \\ &\leqslant \|D\|_{\mathcal{F}}\sqrt{(c_{\mathcal{F}}^2 + 1)N}\end{aligned}$$

□

Now we complete the proof of Theorem 4.1.

The expression $\Delta n + 2\sqrt{(c_{\mathcal{F}}^2 + 1)n/\Delta^2}$ from (4.14) and (4.15) takes its minimal value for $\Delta = \sqrt{2}(c_{\mathcal{F}}^2 + 1)^{1/4}n^{-1/4}$. In this case, the right-hand side of inequality (4.14) is equal to:

$$\begin{aligned}\Delta n + 2\sqrt{\frac{n(c_{\mathcal{F}}^2 + 1)}{\Delta^2}} &= 2\Delta n \\ &= 2\sqrt{2}(c_{\mathcal{F}}^2 + 1)^{1/4}n^{3/4}\end{aligned} \tag{4.16}$$

We prove the bound (4.2) using the doubling trick argument. Let $\epsilon > 0$ and $h = \lceil 2/\epsilon \rceil$, where $\lceil r \rceil$ is the smallest integer number greater than or equal to r. Define an increasing sequence of natural numbers $n_s = (s + h)^h$, $s = 1, 2, \dots$, and $\Delta_s = \sqrt{2}(c_{\mathcal{F}}^2 + 1)^{1/4}n_s^{-1/4}$.

It can be proved by mathematical induction on s that:

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i))\right| \leqslant 4s\Delta_{s-1}n \tag{4.17}$$

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

$$\left|\sum_{i=1}^{n_s} E(I(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i))\right| \leqslant 4s\Delta_s n_s \tag{4.18}$$

also holds. Therefore, we obtain:

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i))\right| \leqslant 18(c_{\mathcal{F}}^2 + 1)^{1/4}n^{(3/4)+\epsilon} \tag{4.19}$$

for all n.[2]

[2] Here and in what follows we omit details of the calculation of constants in the bounds.

The Azuma–Hoeffding inequality states that, for any $\gamma > 0$:

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} V_i\right| > \gamma\right\} \leqslant 2\mathrm{e}^{-2n\gamma^2} \tag{4.20}$$

for all n, where V_i are martingale differences.

We define:

$$V_i = I(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i) - E(I(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i)) \quad \text{and} \quad \gamma = \sqrt{\frac{1}{2n}\ln\frac{2}{\delta}}$$

where $\delta > 0$.

Combining (4.19) with (4.20), we obtain, for any $\delta > 0$ with probability $1 - \delta$:

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i, \tilde{z}_i)(S_i - \tilde{p}_i)\right| \leqslant 18(c_{\mathcal{F}}^2 + 1)^{1/4} n^{(3/4)+\epsilon} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}}$$

for all n. Theorem 4.1 is proved. □

5 PROOF OF THEOREM 2.1

At any step i we compute the deterministic forecast p_i defined in Section 4 and its randomization to $\tilde{p}_i$ using parameters $\Delta = \Delta_s = \sqrt{2}(c_{\mathcal{F}} + 1)^{1/4}(s + h)^{-h/4}$ and $n_s = (s + h)^h$, where $n_s \leqslant i < n_{s+1}$. Also, let $\tilde{S}_{i-1}$ be a randomization of the past price S_{i-1}. In Theorem 4.1, $z_i = S_{i-1}$ and $\tilde{z}_i = \tilde{S}_{i-1}$.

The following upper bound directly follows from the method of discretization:

$$\begin{aligned}
\left|\sum_{i=1}^{n} I(\tilde{p}_i > \tilde{S}_{i-1})(\tilde{S}_{i-1} - S_{i-1})\right| &\leqslant \sum_{t=0}^{s}(n_{t+1} - n_t)\Delta_t \\
&\leqslant 4(c_{\mathcal{F}}^2 + 1)^{1/4} n_s^{(3/4)+\epsilon} \\
&\leqslant 4(c_{\mathcal{F}}^2 + 1)^{1/4} n^{(3/4)+\epsilon}
\end{aligned} \tag{5.1}$$

Let $D(x)$ be an arbitrary trading strategy from the RKHS $\mathcal{F}$. Clearly, the bound (5.1) holds if we replace $I(\tilde{p}_i > \tilde{S}_{i-1})$ on $\|D\|_\infty^{-1} D(\boldsymbol{x}_i)$.

For simplicity, we give the proof for the case of going long, where $D(x) \geqslant 0$ for all x and:

$$\tilde{M}_i = \begin{cases} 1 & \text{if } \tilde{p}_i > \tilde{S}_{i-1} \\ 0 & \text{otherwise} \end{cases}$$

We use abbreviations:

$$\nu_1(n) = 4(c_{\mathcal{F}}^2 + 1)^{1/4} n^{(3/4)+\epsilon} \tag{5.2}$$

$$\nu_2(n) = 18n^{(3/4)+\epsilon}(c_{\mathcal{F}}^2 + 1)^{1/4} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}} \tag{5.3}$$

$$\nu_3(n) = \sqrt{(c_{\mathcal{F}}^2 + 1)n} \tag{5.4}$$

The following sums are for $i = 1, \dots n$. We use the Azuma–Hoeffding inequality (4.20). For any $\delta > 0$ with probability $1 - \delta$, the following chain of equalities and inequalities is valid:

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{M}_i (S_i - S_{i-1}) & \\
&= \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (S_i - S_{i-1}) \\
&= \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (S_i - \tilde{p}_i) + \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{p}_i - \tilde{S}_{i-1}) + \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{S}_{i-1} - S_{i-1}) && (5.5) \\
&\geqslant \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{p}_i - \tilde{S}_{i-1}) - \nu_1(n) - \nu_2(n) && (5.6) \\
&\geqslant \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(\tilde{p}_i - \tilde{S}_{i-1}) - \nu_1(n) - \nu_2(n) \\
&= \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(p_i - S_{i-1}) + \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(\tilde{p}_i - p_i) \\
&\quad - \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(\tilde{S}_{i-1} - S_{i-1}) - \nu_1(n) - \nu_2(n) && (5.7) \\
&\geqslant \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(p_i - S_{i-1}) - 3\nu_1(n) - \nu_2(n) && (5.8) \\
&\geqslant \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(S_i - S_{i-1}) - \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(S_i - p_i) \\
&\quad - 4\nu_1(n) - \nu_2(n) - \|D\|_\infty^{-1} \|D\|_{\mathcal{F}} \nu_3(n) && (5.9) \\
&= \|D\|_\infty^{-1} \sum_{i=1}^{n} D(x_i)(S_i - S_{i-1}) \\
&\quad - 4\nu_1(n) - \nu_2(n) - \|D\|_\infty^{-1} \|D\|_{\mathcal{F}} \nu_3(n)
\end{aligned}$$

In transition from (5.5) to (5.6), the inequality (4.2) of Theorem 4.1 and the bound (5.1) were used, so terms (5.2) and (5.3) were subtracted. In transition from (5.7) to (5.8), the bound (5.1) was applied twice to intermediate terms, so the term (5.1) was subtracted twice. In transition from (5.8) to (5.9), the inequality (4.3) of Theorem 4.1 was used, so the term (5.4) was subtracted.

Therefore, with probability $1 - \delta$:

$$\mathcal{K}_n^{\mathrm{M}} \geqslant \|D\|^{-1} \mathcal{K}_n^{\mathrm{D}} - O\left(n^{(3/4)+\epsilon} + \left(n \ln\left(\frac{1}{\delta}\right)\right)^{1/2}\right) \tag{5.10}$$

The inequality (2.1) follows from (5.10). Theorem 2.1 is proved for any $D \in \mathcal{F}$.

Using a universal kernel and the corresponding canonical universal RKHS, we can extend our asymptotic results for all continuous stationary trading strategies D.

An RKHS $\mathcal{F}$ on X is universal if X is a compact metric space and every continuous function f on X can be arbitrarily well approximated in the metric $\|\cdot\|_\infty$ by a function from $\mathcal{F}$. For any $\epsilon > 0$, there exists $D \in \mathcal{F}$ such that (see Steinwart (2001, Definition 4)):

$$\sup_{x \in X} |f(x) - D(x)| \leqslant \epsilon$$

We use $X = [0, 1]$. The Sobolev space $\mathcal{F} = H^1([0, 1])$ is the universal RKHS (see Steinwart (2001) and Vovk (2006)).

The existence of the universal RKHS on $[0, 1]$ implies the full version of Theorem 2.1.

An algorithm for computing forecasts p_i and a sequential method of randomization can be constructed such that the randomized trading strategy $\tilde{M}_i$ defined in Section 2 performs at least as well as any nontrivial continuous trading strategy f:

$$\liminf_{n \to \infty} \frac{1}{n}(\mathcal{K}_n^{\mathrm{M}} - \|f\|_\infty^{-1} \mathcal{K}_n^{f}) \geqslant 0 \tag{5.11}$$

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

This result directly follows from the inequality (5.10) and the possibility of approximating arbitrarily closely any continuous function f on $[0, 1]$ by a function D from the universal RKHS $\mathcal{F}$.

Any trading strategy $\tilde{M}_i$ satisfying (5.11) is called universally consistent. The notion of the universal trading strategy for the minimax case was first studied by Cover (1991).

The property of universal consistency (5.11) is strictly asymptotic and does not tell us anything about finite data sequences. We have obtained the convergence bound (5.10) for more narrow classes of functions like the RKHS.

6 NUMERICAL EXPERIMENTS

6.1 Computer technology

In the numerical experiments we used historical data in the form of per-minute time series of prices of arbitrarily chosen stocks.

The computation scheme defined in the proof of Theorem 4.1 was essentially simplified. Since the kernels $K(p, z_n, p_i, z_i)$ and $M(x_n, x_i)$ from the proof of Theorem 4.1 have a similar form in most practical applications, we approximate them with one smooth kernel function $\mathcal{K}(p, p_n) = \cos(\pi(p - p_n)/2)$ also used by Vovk (2007). We do not include signals in the kernel. Indeed, the impact of signals is taken into account when we use the decision rules $\tilde{p}_i > \tilde{S}_{i-1}$.

The time characteristics are crucial in any short-term trading algorithm. The greatest time cost is associated with the calculation of sums (4.4) and finding roots of the corresponding equation. The experiments carried out show that the computation time of calculating any forecast increases in a linear way with the length of the history. To optimize this time, the computation has been parallelized on overlapping processes performed for a time series of fixed length L_1. The first $L_2 < L_1$ time points of any series are only used for scaling the stock prices, defining the size of randomization grid and initial learning of the forecasting algorithm. Trading is not performed at the first L_2 time points of such series. When a previous process terminates, we switch to the next process. The results of this parallel computing are accumulated into a single overall forecasting series.

The prices of a stock are scaled such that $S_i \in [0, 1]$ for all i. The scaling is performed for the time series of each process separately. The first L_2 time points of any process are used for computing a scaling constant. The forecasting algorithm is performed for the scaled prices S_i.

We implement this computer technology for two forecasting algorithms: the universal strategy constructed in Section 2 (UN model) and the autoregressive moving average (ARMA) algorithm model (see Peng and Aston (2011)).[3] We also use the buy-and-hold trading strategy. With this strategy, we buy a holding of shares at the starting point of trading and sell them at the end of the trading period.

6.2 Results of numerical experiments

In the numerical experiments we used historical data in the form of per-minute time series of prices of seventeen arbitrarily chosen stocks (eleven US stocks and six Russian stocks) and of one simulated stock, TEST. The data was downloaded from

[3] See also the state space models toolbox for MATLAB. URL: http://sourceforge.net/projects/ssmodels/.

the FINAM website.[4] The number n of trading points in each time series is from 88 000 to 116 000. The period of the game is from March 26, 2010 to March 25, 2011 and is divided on one-minute intervals.

The artificial stock TEST is simulated as $S_i = S_{i-1} + \xi_i$, $i = 1, 2, \dots, N$, where ξ_i is the Gaussian random variable with mean 0 and a variance equal to the variance of the scaled GAZP stock.

We implement the trading strategy defined in Section 2. Two series of numerical experiments are performed.

In the first series, at each step, starting from initial capital $\mathcal{K}_0^R = \mathcal{K}_0^F = \mathcal{K}_0 = KS_0$, where S_0 is the price of a stock at the first time point, the strategy $\tilde{M}_i$ performs by going long or going short with K shares of the stock.

At each step i, in the case of going long, we buy a holding of K shares at the beginning and sell them at the end of this step. In the case of going short, we sell a holding of K shares at the beginning and buy them at the end of this step. We take $K = 5$ in our experiments.

We calculate the capital separately for each trading style. In the case of going long, the capital changes at any step i as $\mathcal{K}_i^R = \mathcal{K}_{i-1}^R + K(S_i - S_{i-1})$ if $\tilde{p}_i > \tilde{S}_{i-1}$, and $\mathcal{K}_i^R = \mathcal{K}_{i-1}^R$ otherwise.[5]

In the case of going short, $\mathcal{K}_i^F = \mathcal{K}_{i-1}^F - K(S_i - S_{i-1})$ if $\tilde{p}_i \leqslant \tilde{S}_{i-1}$, and $\mathcal{K}_i^F = \mathcal{K}_{i-1}^F$ otherwise, where $i = 1, 2, \dots, N$.

The results of the numerical experiments are shown in Table 2 on the next page. In the first column, the ticker symbols of the stocks are shown. The second column shows the profit of the buy-and-hold trading strategy. With this strategy, we buy a holding of shares using capital $\mathcal{K}_0$ and sell them for $\mathcal{K}_N$ at the end of the trading period.

In the third and fourth columns, the results of one run of trading based on the universal (UN) randomized forecasting strategy are shown. In the third column, a relative return, percentage-wise, to the initial capital $(\mathcal{K}_N - \mathcal{K}_0)/\mathcal{K}_0 100\%$ is shown for the going long case. In the fourth column, the same relative return is shown for the going short case. In the fifth and sixth columns, the same results are shown for trading using ARMA forecasts.

According to the results for the trades using the UN strategy, all stocks can be divided into two groups. For the first group of stocks, both UN strategies (going long and going short) essentially outperform the buy-and-hold strategy and receive a positive relative return during the whole period of trading. The evolution of the capital of typical stocks from this group (GOOG and KOCO) is shown in Figure 1 on page 81

[4] See www.finam.ru/analysis/profile041CA00007/default.asp.

[5] Because of the discretization there is a difference between rules $\tilde{p}_i > \tilde{S}_{i-1}$ and $\tilde{p}_i \geqslant \tilde{S}_{i-1}$. This difference may be the subject of further study.

TABLE 2 Universal trading.

Ticker	Buy and hold (profit %)	UN going long (profit %)	UN going short (profit %)	ARMA going long (profit %)	ARMA going short (profit %)
TEST	6.85	–1.39	–8.19	9.88	3.08
AT-T	7.71	137.40	129.70	30.73	23.02
CTGR	15.04	1594.34	1579.34	1167.22	1152.53
KOCO	16.55	62.66	46.15	2.90	–13.61
GOOG	10.25	114.85	104.62	12.85	2.62
InBM	24.28	85.38	61.09	29.31	5.02
INTL	4.29	111.70	107.50	25.86	21.66
MSD	10.71	58.32	47.60	18.66	7.95
US1.AMT	22.01	22.74	0.77	28.46	6.49
US1.IP	2.40	19.83	17.47	9.36	7.00
US2.BRCM	25.30	53.62	28.28	20.06	–5.27
US2.FSLR	40.15	143.92	103.61	–9.86	–50.16
SIBN	–6.54	732.87	739.33	357.74	364.20
GAZP	22.75	101.20	78.45	31.75	9.00
LKOH	19.39	261.84	242.45	87.08	67.68
MTSI	–1.61	669.16	670.68	326.12	327.64
ROSN	9.69	188.89	179.12	34.40	24.63
SBER	14.21	108.97	94.90	37.53	23.46

and Figure 2 on the facing page. We note an unusually high relative return when trading with the CTGR stock.

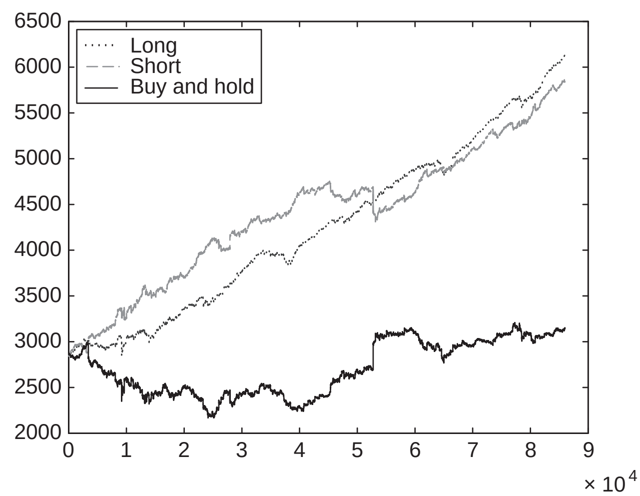
For the second group of stocks (TEST, US1.AMT, US1.IP, US2.BRCM and MCD), UN going long and going short does not outperform the buy-and-hold strategy at most points during the trading period (through it does outperform it at the end of this period). Also, UN trading strategies receive a negative relative return at some intermediate points of the trading period. The evolution of the capital of typical stocks from this group (TEST, US1.IP and US2.BRCM) is shown on Figure 3 on page 82, Figure 4 on page 82 and Figure 5 on page 83.

It was found that $\mathcal{K}_i > 0$ for $i = 1, 2, \ldots, N$, ie, we never lose all initial capital $\mathcal{K}_0$.

The results presented in Table 3 on page 84 show that trading based on the UN model using universal forecasting performs at least as well as the trading based on the ARMA forecasting model and essentially outperforms it for some stocks.

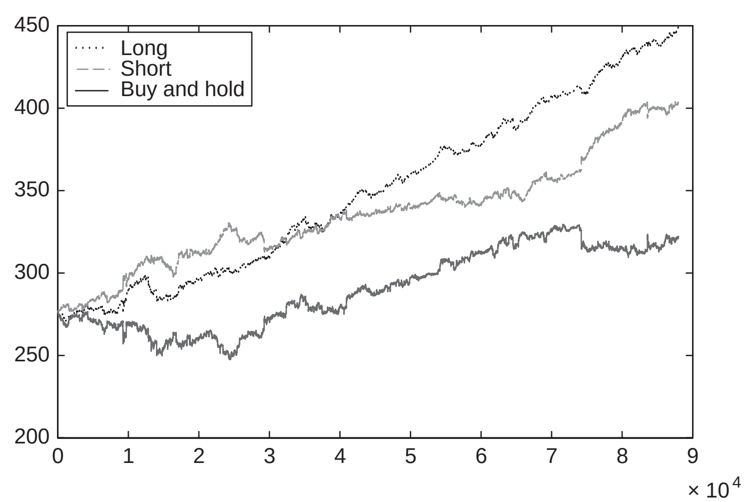
For each stock, we perform ten runs of UN- and ARMA-based algorithms. The evolution of capital for ten runs of the UN strategy using well-calibrated forecasts of the US1.IP stock for a grid of size 0.2σ is shown on Figure 6 on page 85. The results

FIGURE 1 Evolution of capital of the three trading strategies (GOOG stock).



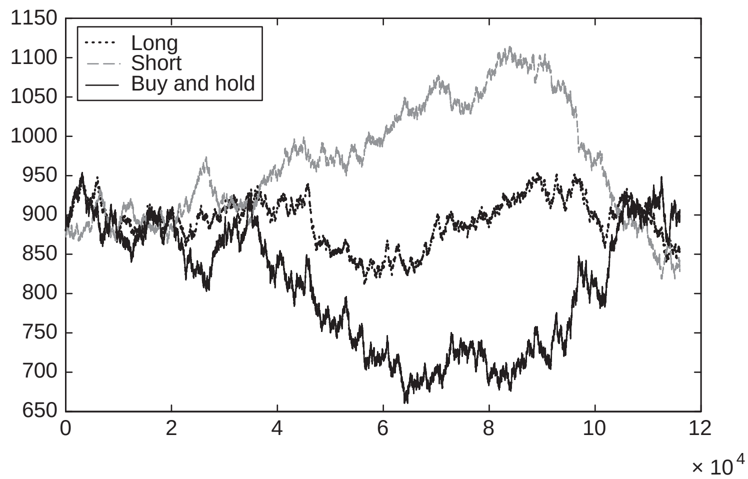
Solid black line: buy and hold. Dotted black line: going long. Dashed gray line: going short. One run of trading is performed with the GOOG stock (see Table 2 on the facing page).

FIGURE 2 Evolution of capital of the three trading strategies (KOCO stock).



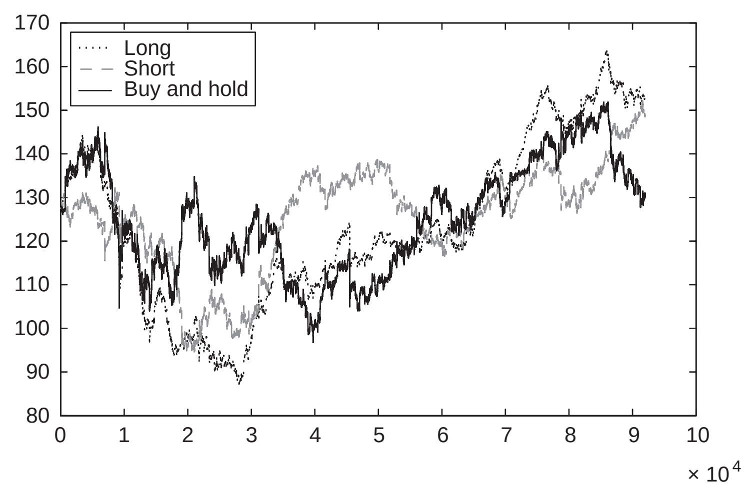
Solid black line: buy and hold. Dotted black line: going long. Dashed gray line: going short. One run of trading is performed with the KOCO stock (see Table 2 on the facing page).

FIGURE 3 Evolution of capital of the three trading strategies (TEST stock).



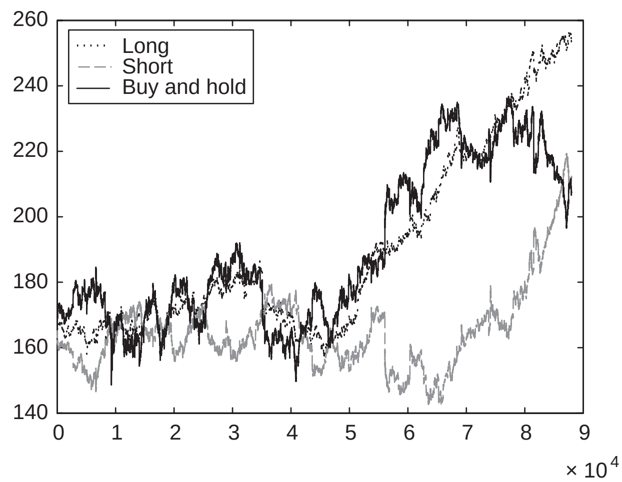
Solid black line: buy and hold. Dotted black line: going long. Dashed gray line: going short. One run of trading is performed with the simulated stock TEST (see Table 2 on page 80). Horizontal axis: rounds of trading. Vertical axis: return in the currency.

FIGURE 4 Evolution of capital of the three trading strategies (US1.IP stock).



Solid black line: buy and hold. Dotted black line: going long. Dashed gray line: going short. One run of trading is performed with the US1.IP stock (see Table 2 on page 80).

FIGURE 5 Evolution of capital of the three trading strategies (US2.BRCM stock).



Solid black line: buy and hold. Dotted black line: going long. Dashed gray line: going short. One run of trading is performed with the US2.BRCM stock (see Table 2 on page 80).

for the trading strategy using ARMA forecasts are shown in Figure 7 on page 85. Here, σ is the empirical standard deviation of stock prices computed on the basis of the initial subseries L_2 of each series.

These results show that the UN strategy is stable with respect to randomization. The ARMA-based strategy is much more unstable with respect to randomization.

We also study the dependence of results of trading on the precision of random rounding of past prices and forecasts. Ten runs were performed for each stock and for each size of grid: 0.5σ, 0.2σ, 0.1σ, 0.05σ, 0.02σ and 0.001σ. The means of the capital curves of UN trading with the US1.IP stock for these randomization grids are shown in Figure 8 on page 86. The result corresponding to 0.1σ precision is the best one.

The second series of experiments is closer to a real short-term trading strategy. The trading strategy has a defense guarantee. Starting with the same initial capital $\mathcal{K}_0 = KS_0$, where S_0 is the initial price of a stock, and $K = 5$, we only carry out the going long case using a "defensive" trading strategy. At any step i, our working capital is $\mathcal{L}_{i-1} = \min\{\mathcal{K}_0, \mathcal{K}_{i-1}\}$. In this way we can set aside the extra income.

If our current portfolio is empty at the beginning of any step i, then, using this capital, we buy $M_i = \mathcal{L}_{i-1}/S_{i-1}$ shares of the stock if $\tilde{p}_i > \tilde{S}_{i-1}$ and $\mathcal{L}_{i-1} > 0$. We stop trading as soon as $\mathcal{L}_{i-1} \leqslant 0$. If our portfolio is nonempty at the beginning of step i and $\tilde{p}_i > \tilde{S}_{i-1}$, we skip the step: $M_i = 0$. We sell all shares

TABLE 3 Defensive trading.

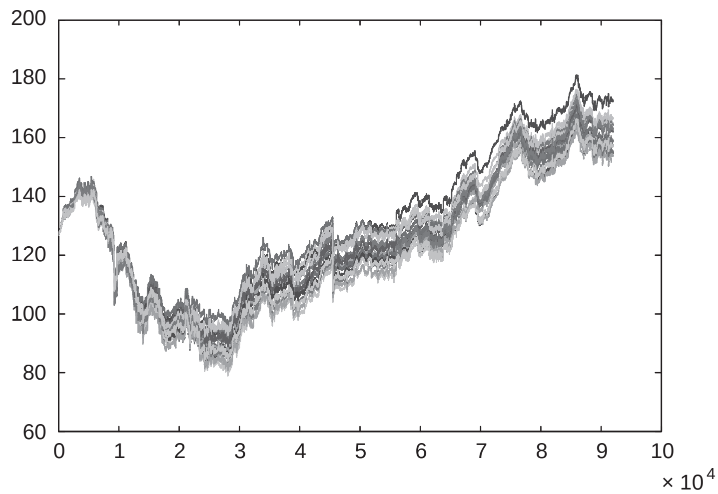
Ticker	Buy and hold (%)	UN profit (%)	UN profit (–0.01%)	ARMA profit (%)	ARMA profit (–0.01%)	UN (in)	ARMA (in)	UN (D)	ARMA (D)
TEST	6.85	3.58	–80.93	3.58	–80.90	0.232	0.163	1.453	1.890
AT-T	7.71	69.01	–79.19	29.86	–79.19	0.218	0.205	1.611	1.576
CTGR	15.04	1030.12	658.13	937.46	540.18	0.238	0.253	1.654	1.479
KOCO	16.55	36.47	–78.62	15.69	–78.55	0.216	0.198	1.604	1.502
GOOG	10.25	46.54	–80.57	3.53	–82.68	0.231	0.211	1.462	1.474
InBM	24.28	54.79	–78.53	34.66	–78.10	0.219	0.187	1.514	1.517
INTL	4.29	43.06	–76.60	5.63	–76.28	0.220	0.179	1.630	1.585
MCD	10.71	34.22	–78.56	19.21	–78.41	0.222	0.190	1.571	1.876
AMT	22.01	16.47	–77.01	24.04	–77.09	0.212	0.183	1.654	1.758
IP	2.40	4.45	–82.78	–14.79	–81.06	0.213	0.181	1.657	1.760
BRCM	25.30	11.40	–80.47	23.98	–76.10	0.216	0.172	1.585	1.876
FLSR	40.15	21.02	–80.04	–27.50	–80.03	0.227	0.196	1.499	1.506
SIBN	–6.54	600.62	249.87	287.48	–58.55	0.169	0.179	2.460	2.292
GAZP	22.75	51.29	–82.04	4.34	–82.16	0.224	0.210	1.539	1.526
LKOH	19.39	149.03	–79.91	46.44	–80.62	0.230	0.244	1.527	1.501
MTSI	–1.61	482.83	79.23	275.13	–69.36	0.188	0.195	2.174	1.959
ROSN	9.69	101.15	–83.14	–0.53	–83.54	0.228	0.240	1.549	1.499
SBER	14.21	51.56	–82.52	–14.47	–82.73	0.225	0.196	1.559	1.674

from the portfolio if $\tilde{p}_i \leqslant \tilde{S}_{i-1}$. At each step i we update the cumulative capital, $\mathcal{K}_i = \mathcal{K}_{i-1} + M_i(S_i - S_{i-1})$. At the end of the trading period we sell all shares from the portfolio.

The results of the second series of numerical experiments are shown in Table 3. In the first column, the ticker symbols of the stocks are shown. The second column contains the relative return of the buy-and-hold trading strategy. In the set of columns called "UN" the relative returns of one run of randomized trading, percentage-wise, for the initial capital are presented for the case with no transaction costs and for the case where a transaction cost at the rate 0.01% is subtracted. We compute the forecast of a future stock price using the method of calibration and universal forecasting presented in Theorem 4.1.

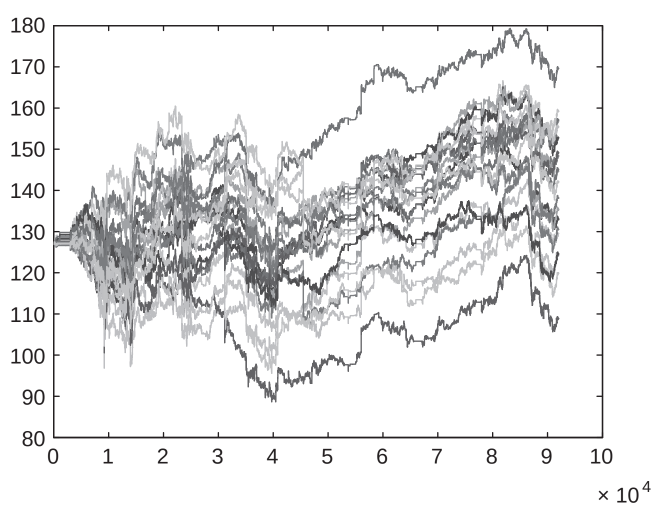
The two columns with the "ARMA" heading show the same results when using the ARMA forecasting model for computing forecasts. The frequencies of market entry steps i, where $\tilde{p}_i > \tilde{S}_{i-1}$, are given in the next two columns marked "in" (for UN and ARMA). The average time spent in the market is shown in the two columns marked "D" (for UN and ARMA).

FIGURE 6 Evolution of capital for ten runs of UN strategy (going long) using well-calibrated forecasts of the US1.IP stock for the period March 26, 2010 to March 25, 2011.



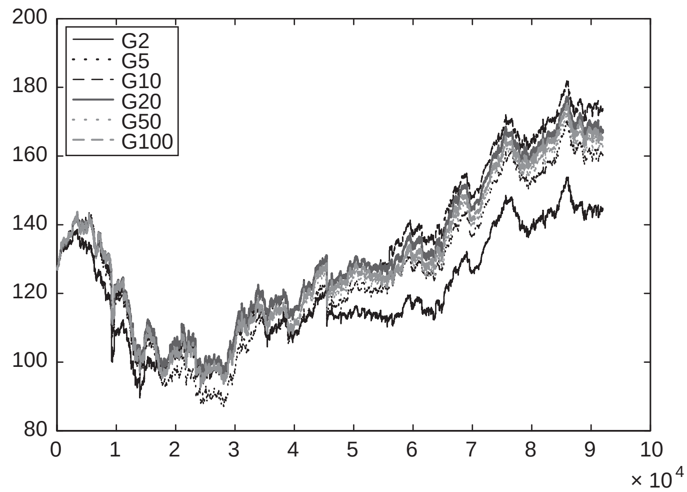
The size of the grid for randomization is 0.2σ.

FIGURE 7 Evolution of capital for ten runs of trading strategy (going long) using ARMA forecasts of the US1.IP stock for the period March 26, 2010 to March 25, 2011.



The size of the grid for randomization is 0.2σ.

FIGURE 8 Means of capital curves of UN trading (going long) with stock US1.IP for randomization grid 0.5σ (G2), 0.2σ (G5), 0.1σ (G10), 0.05σ (G20), 0.02σ (G50), 0.01σ (G100).



7 CONCLUSION

We have presented a universal algorithmic strategy that is shown to perform at least as well as any stationary trading strategy that computes the investment at each step using a fixed continuous function of the side information.

The idea of universality in financial mathematics was proposed by Cover (1991) for the problem of universal portfolio selection. Unlike in statistical theory, no stochastic assumptions are made about stock prices in Cover's approach.

The core of our universal investment strategy is the method of universal prediction – the randomized algorithm for computing well-calibrated forecasts.

Asymptotic calibration is an area of intensive research where several algorithms for computing well-calibrated forecasts have been developed. Several applications of well-calibrated forecasting have been proposed (for example, convergence to correlated equilibrium in game theory by Kakade and Foster (2008), universal methods of recovering unknown functional dependencies by Vovk (2006) and predictions with expert advice by Chernov *et al* (2010)). We present a new application of the calibration method.

Our strategy is based on the randomized forecasting algorithm of Kakade and Foster (2008) and the Vovk (2007) method of defensive forecasting in an RKHS. We show that the universal trading strategy can be constructed using well-calibrated forecasts. We prove that this strategy is at least as good as any stationary trading strategy presented using a rule from any RKHS with regret $O(n^{(3/4)+\epsilon})$. Using the universal kernel, we prove that this strategy asymptotically performs at least as well as any stationary continuous trading strategy.

We considered trading with one stock. It would be helpful to extend these results to the problem of universal portfolio selection.

The theoretical results are supplemented by the corresponding numerical experiments. Our empirical results on historical markets provide strong evidence that this type of technical trading can "beat the market" if transaction costs are ignored. Numerical experiments show a positive return for all chosen stocks, and for some of them we receive a positive return even when transaction costs are subtracted. Results of this type can be useful for technical analysis in finance.

We consider our experimental results as a preliminary verification of the theory. In order to use them in practice it is necessary to refine strategies and to test them. It is also necessary to clarify a number of technical details.

In the numerical experiments we used historical data about previous transactions of purchase and sale of shares, ignoring the process of market order execution. In high-frequency trading, the bid–ask spread is essential. The size of the spread from one asset to another will differ mainly because of the difference in liquidity of each asset. Accounting for this difference would reduce the income of our strategies. A more profound study of the influence of stock liquidity is possible using actual financial experiments. Another, more simple, way to integrate liquidity is to add additional value to transaction costs.

The obvious drawback of a universal strategy is that it uses high-frequency trading, which prevents it from practical application when transaction costs apply. The frequency of market entry can be reduced by using rules of the type $\tilde{p}_i > \tilde{S}_{i-1} + \epsilon$, where $\epsilon > 0$.

Several unresolved issues remain. The role of randomization is unclear. Randomization is necessary to obtain the theoretical results. However, we have not obtained convincing examples confirming the necessity of randomization in the experimental calculations. Randomization seems to be necessary for cases where the behavior of prices is similar to that of an "adaptive adversary". The behavior of the time series exhibiting high volatility is most similar to this. In this case, randomization will protect us against big losses.

One possible improvement in our algorithm would be preprocessing of input data. Uniformly splitting the time series into one-minute intervals does not reflect the movement of prices in the best way.

The idea of a universal strategy presented in this paper could serve as a basis for further attempts to find universal methods of trading and for developing more effective practical trading strategies for use in financial markets.

REFERENCES

Chernov, A., Kalnishkan, Y., Zhdanov, F., and Vovk, V. (2008). In *Supermartingales in Prediction with Expert Advice*. Lecture Notes in Computer Science, Volume 5254, pp. 199–213. Springer.

Cover, T. (1991). Universal portfolios. *Mathematical Finance* **1**, 1–29.

Cover, T., and Gluss, D. (1986). Empirical Bayes stock market portfolios. *Advances in Applied Mathematics* **7**, 170–181.

Cover, T., and Ordentlich, E. (1996). Universal portfolio with side information. *IEEE Transactions on Information Theory* **42**, 348–363.

Dawid, A. P. (1982). The well-calibrated Bayesian (with discussion). *Journal of the American Statistical Association* **77**, 605–613.

Dawid, A. P. (1985). Calibration-based empirical probability (with discussion). *Annals of Statistics* **13**, 1251–1285.

Foster, D. P., and Vohra, R. (1998). Asymptotic calibration. *Biometrika* **85**, 379–390.

Kakade, S. M., and Foster, D. P. (2008). Deterministic calibration and Nash equilibrium. *Journal of Computer and System Sciences* **74**(1), 115–130.

Oakes, D. (1985). Self-calibrating priors do not exist (with discussion). *Journal of the American Statistical Association* **80**, 339–342.

Peng, J.-Y., and Aston, J. A. D. (2011). The state space models toolbox for Matlab. *Journal of Statistical Software* **41**, 1–26.

Scholkopf, B., and Smola, A. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA.

Steinwart, I. (2001). On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research* **2**, 67–93.

Vovk, V. (2006). On-line regression competitive with reproducing kernel Hilbert spaces. In *Theory and Applications of Models of Computation: Proceedings of the Third Annual Conference on Computation and Logic*, Cai, J.-Y., Cooper, S., and Li, A. (eds), pp. 452–463. Springer.

Vovk, V. (2007). Defensive forecasting for optimal prediction with expert advice. Working Paper (arXiv:0708.1503v1).

Vovk, V. (2008). Predictions as statements and decisions. *Theoretical Computer Science* **405**(3), 285–296.

Vovk, V., Takemura, A., and Shafer, G. (2005). Defensive forecasting. In *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*, Cowell, R. G., and Ghahramani, Z. (eds), pp. 365–372. Society for Artificial Intelligence and Statistics, Cambridge, UK.

V'yugin, V. V., and Trunov, V. G. (2012). Universal algorithm for online trading based on the method of calibration. Working Paper (arXiv:1205.3767v2).