

## Адаптивное прогнозирование и его применение для технического анализа финансовых инструментов<sup>1</sup>

В. В. Вьюгин, В. Г. Трунов

*Институт проблем передачи информации им. А.А. Харкевича,  
Российская академия наук, Москва, Россия  
vugin@iitp.ru*

Поступила в редколлегию 08.02.2011

**Аннотация**—Изучается понятие калибруемости последовательности предсказаний на последовательности исходов. Приводится модифицированный вариант алгоритма универсального прогнозирования Фостера, Вохра и Какаде, доказана асимптотическая состоятельность этого алгоритма. Разработаны стратегии игры на фондовом рынке, построенные на техническом анализе, который использует прогнозы этого алгоритма. Приведены результаты экспериментов на реальных данных.

### 1. ВВЕДЕНИЕ

В последнее десятилетие были разработаны методы прогнозирования, отличные от традиционных подходов математической статистики. В частности, можно «успешно» прогнозировать (с точки зрения некоторого критерия) любую последовательность исходов, не используя при этом никаких гипотез об источнике их порождения. Один из таких методов в модифицированной форме приводится в данной работе.

Прикладная часть работы посвящена применениям универсальных методов прогнозирования для технического анализа в финансовой сфере. Технический анализ состоит из методов анализа временных рядов цен финансовых инструментов и не использует информацию о работе соответствующих им предприятий и организаций. Как правило, конечной целью технического анализа является построение прогнозов будущего поведения цен финансовых инструментов. В рамках технического анализа, в данной работе будут строиться стратегии игры на финансовом рынке, основанные на универсальном прогнозировании цен различных финансовых инструментов (акций, фьючерсов и т.д.).

Рассматривается следующая задача прогнозирования: предсказатель получает в режиме онлайн некоторую числовую последовательность исходов  $\omega_1, \omega_2, \dots, \omega_{n-1}, \dots$ , при этом его задачей является вычисление числовой оценки  $p_n$  для будущего события  $\omega_n$  по известным  $n - 1$  исходам  $\omega_1, \omega_2, \dots, \omega_{n-1}$ .

Число  $p_n$  может рассматриваться как прогноз среднего значения будущего исхода  $\omega_n$ . В случае, когда  $\omega_i$  принимают значения 0 или 1, это среднее равно вероятности события  $\omega_n = 1$ .

Типичным примером такого прогнозирования является задача вероятностного предсказания погоды на завтра. Событие  $\omega_n = 1$  может интерпретироваться как дождь в  $n$ -ый день, а число  $p_n$  — как его вероятность, вычисленная на основе наблюдений погоды  $\omega_1, \omega_2, \dots, \omega_{n-1}$  за предыдущие  $n - 1$  дней. Отсутствие дождя соответствует событию  $\omega_n = 0$ , его вероятность равна  $1 - p_n$ .

<sup>1</sup> Работа частично поддержана грантом РФФИ 09-07-00180-а.

Предсказатель погоды считается калибруемым, если дождь случается так же часто, как он прогнозируется предсказателем. Например, если дождь случается в 80% всех дней, для которых предсказатель давал прогноз  $p_n = 0.8$  и т.д. В этом примере  $p_n \in [0, 1]$ .

В данной работе мы будем рассматривать случай  $\omega_n \in [0, 1]$  и  $p_n \in [0, 1]$  для всех  $n$ . Например, действительные числа  $\omega_n$  могут быть ценами некоторого финансового инструмента в моменты времени  $n = 1, 2, \dots$ ; предполагаем, что цена нормирована, так что  $0 \leq \omega_n \leq 1$  для всех  $n$ . Число  $p_n$  может рассматриваться как прогноз значения этой цены на момент времени  $n$ , так что  $0 \leq p_n \leq 1$  для всех  $n$ . В следующих разделах мы рассмотрим задачу прогнозирования цены финансового инструмента и построим игровые стратегии, использующие прогнозы этой цены.

При вероятностном подходе предполагается, что источник, генерирующий последовательность  $\omega_1, \omega_2, \dots$ , описывается некоторой вероятностной мерой  $P$ . В этом случае,  $p_n = E(\omega_n | \omega_1, \omega_2, \dots, \omega_{n-1})$ , где  $E$  - условное математическое ожидание случайной величины  $\omega_n$  относительно  $\omega_1, \omega_2, \dots, \omega_{n-1}$  по этой мере.

На практике мы часто имеем дело с единственной исторической последовательностью исходов  $\omega_1, \omega_2, \dots, \omega_{n-1}$ , по которой довольно трудно оценить меру, описывающую будущие значения величины  $\omega_n$ . В данной работе предположение о наличии такой меры не используется. Тем не менее, можно указать метод прогнозирования произвольной последовательности  $\omega_1, \omega_2, \dots, \omega_{n-1}$ , удовлетворяющий требованию *калибруемости*.

Точные определения и алгоритм построения калибруемых прогнозов будут приведены в разделе 2.

Данная работа продолжает работу [1]. В частности, в теореме 1 получен новый результат об асимптотической состоятельности алгоритма построения калибруемых прогнозов. Также получена новая дискретизация прогнозов и новые, более точные, нижние оценки длины подпоследовательности входов в рынок для торговой стратегии. Проведено более полное сравнительное моделирование игр на временных рядах цен финансовых инструментов.

## 2. КАЛИБРУЕМЫЕ ПРОГНОЗЫ

Приведем точное определение калибруемости, предложенное Дейвидом [4]. Мы рассмотрим произвольные подынтервалы  $I = [a, b]$ ,  $(a, b]$ ,  $[a, b)$ ,  $(a, b)$  интервала  $[0, 1]$  и их характеристические функции

$$I(p) = \begin{cases} 1, & \text{если } p \in I, \\ 0, & \text{в противном случае.} \end{cases}$$

Последовательность прогнозов  $p_1, p_2, \dots$  называется *калибруемой* на бесконечной последовательности  $\omega_1, \omega_2, \dots$ , если для характеристической функции  $I(p)$  любого подынтервала  $I \subseteq [0, 1]$  *калибровочная ошибка* стремится к нулю, т.е.

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

при условии, что знаменатель отношения (1) стремится к бесконечности при  $n \rightarrow \infty$ . Характеристическая функция  $I(p_i)$  определяет некоторое правило выбора, которое определяет те номера исходов  $i$ , для которых мы вычисляем отклонение прогноза  $p_i$  от соответствующего исхода  $\omega_i$ .

Под *детерминированной прогнозирующей системой* понимается произвольная, возможно частичная, функция  $f : \cup_{n=1}^{\infty} [0, 1]^n \rightarrow [0, 1]$ , отображающая конечные последовательности действительных чисел из отрезка  $[0, 1]$  в числа из этого отрезка.

Простые соображения, впервые приведенные в [7, 9], показывают, что никакая детерминированная прогнозирующая система не может калиброваться на любой бесконечной последовательности исходов. А именно, для произвольной прогнозирующей системы  $f$  можно определить последовательность  $\omega = \omega_1, \omega_2, \dots$  так, что

$$\omega_i = \begin{cases} 1, & \text{если } p_i < \frac{1}{2}, \\ 0, & \text{в противном случае,} \end{cases}$$

где  $p_i = f(\omega_1, \dots, \omega_{i-1})$ ,  $i = 1, 2, \dots$ . Легко видеть, что для  $I = [0, \frac{1}{2}]$  или для  $I = [\frac{1}{2}, 1]$  условие (1) нарушается при  $n \rightarrow \infty$ .

Мы увидим далее, что подобные трудности можно преодолеть, если прогнозы  $p_n$  «слегка» рандомизировать. Пусть  $\mathcal{P}[0, 1]$  – множество всех вероятностных мер на отрезке  $[0, 1]$ .<sup>1</sup>

Согласно [8], под *рандомизированной прогнозирующей системой* понимается произвольная функция  $f : \cup_{n=1}^{\infty} ([0, 1]^{n-1} \times [0, 1]^{n-1}) \rightarrow \mathcal{P}[0, 1]$ . Таким образом, значениями рандомизированной прогнозирующей системы являются вероятностные меры на единичном отрезке  $[0, 1]$ .

Легко видеть, что для произвольной фиксированной последовательности исходов  $\omega = \omega_1, \omega_1, \dots$  существует вероятностная мера  $Pr$  на бесконечных траекториях прогнозов  $p_1, p_2, \dots \in [0, 1]^\infty$ , для которой выполнено

$$f(\omega^{n-1}, p^{n-1}) = P_{\omega^{n-1}}(\cdot | p^{n-1}) \quad (2)$$

для всех  $n$ , где  $p^{n-1} = p_1, \dots, p^{n-1}$  и  $\omega^{n-1} = \omega_1, \dots, \omega_{n-1}$ . По теореме Ионеску–Тульчи (см. [2]) о продолжении меры, такая мера существует также и в случае, когда для произвольного  $n$  исход  $\omega_n$  зависит от предыдущих исходов и прогнозов  $\omega^{n-1}, p^n$ , т.е. исходы  $\omega_n$  являются измеримыми функциями от  $\omega^{n-1}, p^n$ . Это есть так называемый случай “адаптивно враждебной Природы” – подобный способ порождения последовательности исходов рассматривался в приведенном выше примере (в этом примере исход  $\omega_i$  зависит от прогноза  $p_i$ ).

Результаты этой работы будут относиться к частному случаю этого определения. Мы будем рассматривать рандомизированные прогнозирующие системы типа  $f : \cup_n \{0, 1\}^{n-1} \rightarrow \mathcal{P}[0, 1]$ . Данные прогнозирующие системы не зависят от прошлых прогнозов. Мы также введем обозначение  $f(\alpha) = Pr_\alpha(\cdot)$  для вероятностного распределения на прогнозах, где  $\alpha$  – произвольная конечная последовательность исходов, принадлежащих  $[0, 1]$ .

При использовании прогнозирующих систем данного типа мера на траекториях прогнозов устроена проще: для каждой бесконечной последовательности  $\omega$  вероятностные распределения  $Pr_{\omega^{i-1}}(\cdot)$  порождают распределение вероятностей  $Pr = \prod_{i=1}^{\infty} Pr_{\omega^{i-1}}$  на множестве всех бесконечных траекторий прогнозов  $p_1, p_2, \dots$ , где  $p_i \in [0, 1]$ ,  $i = 1, 2, \dots$ . Мера  $Pr$  зависит от бесконечной последовательности  $\omega$ .

Фостер и Вохра [5], а также Какаде и Фостер [6] предложили вероятностные алгоритмы для построения калиброванных предсказаний: для каждого числа  $\Delta > 0$  можно построить рандомизированную прогнозирующую систему  $f(\alpha) = Pr_\alpha(\cdot)$  такую, что для произвольной бесконечной последовательности  $\omega = \omega_1, \omega_2, \dots$  вероятность  $Pr$  события

$$\limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n I(p_i)(\omega_i - p_i) \right| \leq \Delta$$

равна 1, где траектория  $p_1, p_2, \dots$  распределена по вероятностной мере  $Pr$ ,  $I(p)$  – характеристическая функция произвольного подынтервала  $[0, 1]$ . Такая функция называется правилом выбора, определенным прогнозом.

<sup>1</sup> При этом также рассматривается сигма-алгебра борелевских множеств на  $[0, 1]$ .

Мы будем использовать правила выбора, которые будут определяться на основе подынтервалов  $I$  единичного интервала  $[0, 1]$ , Пусть

$$I(p) = \begin{cases} 1, & \text{если } p \in I, \\ 0, & \text{в противном случае} \end{cases}$$

– характеристическая функция подынтервала  $I$ .

Мы приводим некоторый вариант основного результата работы [6] о существовании универсальной прогнозирующей системы.

**Теорема 1.** *Можно построить рандомизированную прогнозирующую систему  $Pr_\alpha(\cdot)$  такую, что для произвольной последовательности исходов  $\omega_1, \omega_2, \dots$  и для характеристической функции  $I(p)$  произвольного подынтервала  $I$  единичного интервала  $[0, 1]$  событие*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n I(p_i)(\omega_i - p_i) = 0 \quad (3)$$

*имеет  $Pr$ -вероятность 1, где  $Pr$  есть вероятностное распределение на  $[0, 1]^\infty$ , порожденное последовательностью распределений  $Pr_{\omega_{n-1}}(\cdot)$ ,  $n = 1, 2, \dots$ , а траектория прогнозов  $p_1, p_2, \dots$  распределена по мере  $Pr$ .*

Доказательство этой теоремы приведено в разделе 2.1.

### 2.1. Алгоритм вычисления калибруемых прогнозов

В этом разделе мы приведем некоторую модификацию алгоритма Какаде и Фостера вычисляющего рандомизированные предсказания.

Предварительно приведем конструкцию алгоритма, который удовлетворяет более слабым требованиям, а именно, для произвольного  $\Delta > 0$  построим рандомизированную прогнозирующую систему  $Pr_\alpha(\cdot)$  такую, что для произвольной последовательности  $\omega_1, \omega_2, \dots$  и для характеристических функций  $I(p)$  произвольных подынтервалов единичного интервала  $Pr$ -вероятность события

$$\limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n I(p_i)(\omega_i - p_i) \right| \leq \Delta \quad (4)$$

равна 1, где  $Pr$  – распределение на  $[0, 1]^\infty$ , порожденное последовательностью распределений  $Pr_{\omega_{n-1}}(\cdot) = f(\omega_1, \dots, \omega_{n-1})$ ,  $n = 1, 2, \dots$ , последовательность прогнозов  $p_1, p_2, \dots$  распределена по мере  $Pr$ .

Для каждого  $n$  распределение вероятностей  $Pr_{\omega_{n-1}}$  будет сосредоточено только в двух точках отрезка  $[0, 1]$ , расположенных на расстоянии  $\Delta$  друг от друга. Приводимый ниже вариант универсального рандомизированного алгоритма основан на идеях из [6] и [11].

Пусть  $\omega_1, \omega_2, \dots$  – произвольная последовательность элементов  $\{0, 1\}$ , поступающая в режиме онлайн. Построим алгоритм для вычисления случайной величины, выдающей прогноз  $p_n \in [0, 1]$  будущего значения  $\omega_n \in \{0, 1\}$  по начальному фрагменту  $\omega_1, \dots, \omega_{n-1}$ . Основное требование к таким прогнозам – они должны с вероятностью близкой к единице удовлетворять условию калибруемости (4). Соответствующее распределение вероятностей является внутренним по отношению к алгоритму и строится в процессе конструкции.

Предварительно разобьем интервал  $[0, 1]$  на равные части длины  $\Delta = 1/K$  с помощью рациональных точек  $v_i = i\Delta$ ,  $i = 0, 1, \dots, K$ . Пусть  $V$  обозначает множество всех этих точек.

Любое число  $p \in [0, 1]$  представляется в виде линейной комбинации соседних точек подынтервала разбиения, содержащего  $p$ ,

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

где  $p \in [v_{i-1}, v_i]$ ,  $i = \lfloor p/\Delta + 1 \rfloor$ , и

$$w_{v_{i-1}}(p) = 1 - \frac{p - v_{i-1}}{\Delta}, \quad w_{v_i}(p) = 1 - \frac{v_i - p}{\Delta}.$$

Полагаем  $w_v(p) = 0$  для всех остальных значений  $v \in V$ .

В дальнейшем, детерминированный прогноз  $p$ , выдаваемый приведенным далее алгоритмом, будет округляться до  $v_{i-1}$  с вероятностью  $w_{v_{i-1}}(p)$  и до  $v_i$  с вероятностью  $w_{v_i}(p)$ .

Построим алгоритм, выдающий детерминированный прогноз  $p$ . Пусть прогнозы  $p_1, \dots, p_{n-1}$  уже определены этим алгоритмом (полагаем  $p_1 = 0$ ). Построим прогноз  $p_n$ .

Рассмотрим вспомогательную величину

$$\mu_{n-1}(v) = \sum_{i=1}^{n-1} w_v(p_i)(\omega_i - p_i).$$

Имеем

$$\begin{aligned} (\mu_n(v))^2 &= (\mu_{n-1}(v))^2 + \\ &+ 2w_v(p_n)\mu_{n-1}(v)(\omega_n - p_n) + \\ &+ (w_v(p_n))^2(\omega_n - p_n)^2. \end{aligned} \quad (5)$$

Суммируем (5) по  $v$ :

$$\begin{aligned} \sum_{v \in V} (\mu_n(v))^2 &= \sum_{v \in V} (\mu_{n-1}(v))^2 + \\ &+ 2(\omega_n - p_n) \sum_{v \in V} w_v(p_n)\mu_{n-1}(v) + \\ &+ \sum_{v \in V} (w_v(p_n))^2(\omega_n - p_n)^2. \end{aligned} \quad (6)$$

Изменим порядок суммирования в сумме вспомогательных величин:

$$\begin{aligned} &\sum_{v \in V} w_v(p)\mu_{n-1}(v) = \\ &= \sum_{v \in V} w_v(p) \sum_{i=1}^{n-1} w_v(p_i)(\omega_i - p_i) = \\ &= \sum_{i=1}^{n-1} \left( \sum_{v \in V} w_v(p)w_v(p_i) \right) (\omega_i - p_i) = \\ &= \sum_{i=1}^{n-1} (\bar{w}(p), \bar{w}(p_i)) (\omega_i - p_i) = \sum_{i=1}^{n-1} K(p, p_i) (\omega_i - p_i), \end{aligned}$$

где  $\bar{w}(p) = (w_1, \dots, w_{v_K}) = (0, \dots, w_{v_{i-1}}(p), w_{v_i}(p), \dots, 0)$  – вектор вероятностей округления,  $p \in [v_{i-1}, v_i]$ , и скалярное произведение

$$K(p, p_i) = (\bar{w}(p), \bar{w}(p_i)) \quad (7)$$

соответствующих векторов (ядро). По определению,  $K(p, p_i)$  – непрерывная функция (точнее, кусочно-линейная).

Второй член правой части суммы (6) при подходящем значении  $p_n$  всегда можно сделать меньшим или равным нулю. Действительно, в качестве  $p_n$  берем корень  $p_n = p$  уравнения

$$S_n(p) = \sum_{v \in V} w_v(p) \mu_{n-1}(v) = \sum_{i=1}^{n-1} K(p, p_i) (\omega_i - p) = 0, \quad (8)$$

если он существует. В противном случае, если левая часть уравнения (8) (которая является непрерывной по  $p$  функцией) больше нуля для всех значений  $p$ , то полагаем  $p_n = 1$ , если она меньше нуля, то полагаем  $p_n = 0$ . Таким образом определенное значение  $p_n$  выдаем в качестве детерминированного прогноза.

Третий член (6) ограничен числом 1. Действительно, так как  $|\omega_i - p_i| \leq 1$  для всех  $i$ , имеем

$$\sum_{v \in V} (w_v(p_n))^2 (\omega_i - p_n)^2 \leq \sum_{v \in V} w_v(p_n) = 1.$$

Отсюда и по (6), если последовательно выбирать прогнозы  $p_i$  согласно указанному правилу, получим

$$\sum_{v \in V} (\mu_n(v))^2 \leq \sum_{i=1}^n \sum_{v \in V} (w_v(p_i))^2 (\omega_i - p_i)^2 \leq n.$$

Напомним, что для каждого  $v \in V$

$$\mu_n(v) = \sum_{i=1}^n w_v(p_i) (\omega_i - p_i). \quad (9)$$

Вставляем в суммы (9)  $I(v)$ , суммируем по  $v \in V$ , меняем порядок этого суммирования и применяем неравенство Коши–Буняковского для векторов  $(I(v) : v \in V)$ ,  $(\mu_n(v) : v \in V)$  и евклидовой нормы:

$$\begin{aligned} & \left| \sum_{i=1}^n \sum_{v \in V} w_v(p_i) I(v) (\omega_i - p_i) \right| = \\ & = \left| \sum_{v \in V} I(v) \sum_{i=1}^n w_v(p_i) (\omega_i - p_i) \right| \leq \\ & \leq \sqrt{\sum_{v \in V} I(v)} \sqrt{\sum_{v \in V} (\mu_n(v))^2} \leq \sqrt{|V|n} \end{aligned} \quad (10)$$

для всех  $n$ , где  $|V| = 1/\Delta$  – число подынтервалов разбиения.

Пусть теперь  $\tilde{p}_i$  – случайная величина, принимающая значения  $v \in V$  с вероятностями  $w_v(p_i)$  (на самом деле, для каждого  $p$  ненулевыми являются только значения  $w_v(p)$  для двух соседних границ интервала разбиения, содержащего детерминированный прогноз  $p_i$ ). Пусть также  $I(p)$  – характеристическая функция произвольного подынтервала единичного интервала. Для любого  $i$  математическое ожидание случайной величины  $I(\tilde{p}_i)(\omega_i - \tilde{p}_i)$  равно

$$E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) = \sum_{v \in V} w_v(p_i) I(v) (\omega_i - v). \quad (11)$$

Согласно усиленному закону больших чисел, для некоторой величины  $\mu_n = o(n)$  (при  $n \rightarrow \infty$ )  $Pr$ -вероятность события

$$\left| \sum_{i=1}^n I(\tilde{p}_i)(\omega_i - \tilde{p}_i) - \sum_{i=1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq \mu_n \quad (12)$$

стремится к 1 при  $n \rightarrow \infty$ . Конкретная форма переменной  $\mu_n$  будет определена далее.

По определению детерминированного прогноза  $p_i$  и функции  $w_v(p)$  :

$$\left| \sum_{v \in V} w_v(p_i) I(v)(\omega_i - v) - \sum_{v \in V} w_v(p_i) I(v)(\omega_i - p_i) \right| < \Delta \quad (13)$$

для каждого  $i$ .

Суммируем (13) по  $i = 1, \dots, n$ , при этом используем неравенство (10). В результате получаем верхнюю оценку для абсолютной величины суммы математических ожиданий (11)

$$\begin{aligned} & \left| \sum_{i=1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| = \\ & = \left| \sum_{i=1}^n \sum_{v \in V} w_v(p_i) I(v)(\omega_i - v) \right| < \Delta n + \sqrt{|V|n} \end{aligned} \quad (14)$$

для всех  $n$ , где  $|V| = 1/\Delta$  - число подынтервалов разбиения.

Из (14) и (12) получаем, что  $Pr$ -вероятность события

$$\left| \sum_{i=1}^n I(\tilde{p}_i)(\omega_i - \tilde{p}_i) \right| \leq \Delta n + \mu_n + \sqrt{n/\Delta} \quad (15)$$

стремится к 1 при  $n \rightarrow \infty$ . В частности,  $Pr$ -вероятность события

$$\left| \frac{1}{n} \sum_{i=1}^n I(\tilde{p}_i)(\omega_i - \tilde{p}_i) \right| \leq \Delta \quad (16)$$

стремится к 1 при  $n \rightarrow \infty$ .

Докажем утверждение о сходимости (1) почти всюду. Для этого усложним приведенную выше конструкцию – введем переменную дискретизацию. Пусть

$$\Delta_1 > \Delta_2 > \dots$$

– последовательность рациональных чисел такая, что  $\Delta_s \rightarrow 0$  при  $s \rightarrow \infty$ .

Мы также определим возрастающую последовательность натуральных чисел

$$n_1 < n_2 < \dots$$

Для каждого  $s$  на шагах  $n_s \leq n < n_{s+1}$  используем дискретизацию с шагом  $\Delta_s$ .

Предварительно заметим, что величина  $\Delta n + \sqrt{\frac{n}{\Delta}}$  принимает наименьшее значение при  $\Delta = n^{-1/3}$ . В неравенстве (14) будет  $\Delta n + \sqrt{\frac{n}{\Delta}} \leq 2\Delta n$  при  $\Delta \geq n^{-1/3}$ .

При выборе  $n_s$  будем требовать выполнения условия  $n_s \geq 1/\Delta_s^3$  для всех  $s$ .

Полагаем  $n_0 = 0$ ,  $\Delta_0 = 1$ . Будем выбирать последовательность  $n_1, n_2, \dots$  так, чтобы было выполнено неравенство

$$\left| \sum_{i=1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq 4(s+1)\Delta_s n \quad (17)$$

для всех  $n_s \leq n \leq n_{s+1}$  и всех  $s$ . Допустим, что  $n_s$  ( $s \geq 1$ ) выбрано так, что

$$\left| \sum_{i=1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq 4s\Delta_{s-1}n \quad (18)$$

при всех  $n_{s-1} \leq n \leq n_s$ , а также выполнено

$$\left| \sum_{i=1}^{n_s} E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq 4s\Delta_s n_s. \quad (19)$$

Найдем теперь значение  $n_{s+1}$ . Для этого предварительно рассмотрим все прогнозы  $p_i$  при значении параметра  $\Delta = \Delta_{s+1}$ , вычисленные согласно указанному выше алгоритму. Первые  $n_s$  таких прогнозов мы не используем (мы будем их использовать только в оценках (20) и (21); обозначим их отдельно –  $\hat{\mathbf{p}}_1, \dots, \hat{\mathbf{p}}_{n_s}$ ), а прогнозы  $\tilde{p}_i$  при  $i > n_s$  присоединим к ранее построенным (при выборе  $n_s$ ) прогнозам. Находим значение  $n_{s+1}$ , удовлетворяющее условию

$$\begin{aligned} & \left| \sum_{i=1}^{n_{s+1}} E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq \quad (20) \\ & \leq \left| \sum_{i=1}^{n_s} E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| + \\ & + \left| \sum_{i=n_s+1}^{n_{s+1}} E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) + \sum_{i=1}^{n_s} E(I(\hat{\mathbf{p}}_i)(\omega_i - \hat{\mathbf{p}}_i)) \right| + \\ & + \left| \sum_{i=1}^{n_s} E(I(\hat{\mathbf{p}}_i)(\omega_i - \hat{\mathbf{p}}_i)) \right| \leq \\ & \leq 4(s+1)\Delta_{s+1}n_{s+1}. \end{aligned}$$

Здесь первая сумма правой части неравенства (20) ограничена величиной  $4s\Delta_s n_s$  – по предположению индукции (19). Вторая и третья суммы ограничены величинами  $2\Delta_{s+1}n_{s+1}$  и  $2\Delta_{s+1}n_s$ , соответственно, по (14) и неравенству  $\Delta_s n_s + \sqrt{\frac{n_s}{\Delta_s}} \leq 2\Delta_s n_s$ .

Предположение индукции (19) будет выполнено, если выбрать

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



Аналогично,

$$\begin{aligned}
& \left| \sum_{i=1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq \\
& \leq \left| \sum_{i=1}^{n_s} E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| + \\
& + \left| \sum_{i=n_s+1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) + \sum_{i=1}^{n_s} E(I(\hat{p}_i)(\omega_i - \hat{p}_i)) \right| + \\
& + \left| \sum_{i=1}^{n_s} E(I(\hat{p}_i)(\omega_i - \hat{p}_i)) \right| \leq \\
& \leq 4(s+1)\Delta_s n
\end{aligned} \tag{21}$$

при  $n_s < n \leq n_{s+1}$ . Здесь первая сумма правой части неравенства (20) также ограничена величиной  $4s\Delta_s n_s \leq 4s\Delta_s n$  – по предположению индукции (19). Вторая и третья суммы ограничены величинами  $2\Delta_{s+1}n \leq 2\Delta_s n$  и  $2\Delta_{s+1}n_s \leq 2\Delta_s n$ , соответственно, по (14) и по выбору  $\Delta_s$ . Предположение индукции (18) также выполнено.

Из (17) следует, что для любого  $s$

$$\left| \sum_{i=1}^n E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i)) \right| \leq 4(s+1)\Delta_s n \tag{22}$$

при всех  $n \geq n_s$ , если величины  $\Delta_s$  удовлетворяют условию  $\Delta_{s+1} \leq \Delta_s(1 - \frac{1}{s+2})$  для всех  $s$ .

Согласно усиленному закону больших чисел (26), соотношение (12) можно уточнить:

$$Pr \left\{ \sup_{n \geq n_s} \left| \frac{1}{n} \sum_{i=1}^n V_i \right| > \Delta_s \right\} \leq (\Delta_s)^{-2} e^{-2n_s \Delta_s^2} \tag{23}$$

для всех  $s$ , где

$$V_i = I(\tilde{p}_i)(\omega_i - \tilde{p}_i) - E(I(\tilde{p}_i)(\omega_i - \tilde{p}_i))$$

– последовательность мартингал-разностей.

Комбинируя неравенства (22) и (23), получим

$$Pr \left\{ \sup_{n \geq n_s} \left| \frac{1}{n} \sum_{i=1}^n I(\tilde{p}_i)(\omega_i - \tilde{p}_i) \right| \geq (4s+5)\Delta_s \right\} \leq (\Delta_s)^{-2} e^{-2n_s \Delta_s^2} \tag{24}$$

для всех  $s$ .

Ряд из полученных экспоненциальных оценок (24) сходится, если  $n_s$  удовлетворяет неравенству

$$n_s \geq \frac{\ln s + 2 \ln \ln s - 2 \ln(\Delta_s)}{2\Delta_s^2}$$

для всех  $s$ . Пусть также  $\Delta_s = o(1/s)$ . Тогда лемма Бореля–Кантелли позволит нам утверждать о сходимости (1) почти всюду.

Пишем  $A \sim B$ , если существуют такие положительные константы  $c_1$  и  $c_2$ , что  $c_1 B \leq A \leq c_2 B$  для всех значений параметров, входящих в  $A$  и  $B$ .

Можно выбрать  $n_s = s^M$  и  $\Delta_s = s^{-M/3}$ , где  $M$  – положительное целое число. Тогда  $s \sim n_s^{1/M}$  (соответствующие константы  $c_1$  и  $c_2$  зависят от  $M$ ).

Легко видеть, что последовательности  $n_s$  и  $\Delta_s$ , удовлетворяющие всем приведенным выше условиям, существуют.

Из (24) видно, что можно определить  $\mu_n = (4s + 5)\Delta_s n$ , где  $s$  такое, что  $n_s < n \leq n_{s+1}$ . При  $n_s < n \leq n_{s+1}$  будет  $n \sim n_s$ , поэтому  $\mu_n \sim n^{2/3+1/M}$ .

В этом случае для вычисления прогноза на шаге  $n$  используется дискретизация  $\Delta \sim \frac{1}{n^{1/3}}$ .

### 2.2. Неравенство больших уклонений для мартингалов

Последовательность случайных величин  $V_1, V_2, \dots$  называется мартингал-разностью относительно последовательности случайных величин  $X_1, X_2, \dots$ , если для любого  $i > 1$  величина  $V_i$  есть функция от  $X_1, \dots, X_i$  и

$$E(V_{i+1}|X_1, \dots, X_i) = 0$$

с вероятностью 1.

Следующее неравенство является следствием неравенства Хефдинга–Азумы [3].

**Лемма 1.** Пусть  $V_1, V_2, \dots$  – мартингал-разность относительно последовательности случайных величин  $X_1, X_2, \dots$ , кроме этого,  $V_i \in [A_i, A_i + 1]$  для некоторой случайной величины  $A_i$ , измеримой относительно  $X_1, \dots, X_i$ . Если  $S_n = \sum_{i=1}^n V_i$ , то для любых  $t > 0$  и натуральных чисел  $n$

$$P \left\{ \left| \frac{S_n}{n} \right| > t \right\} \leq 2e^{-2nt^2}. \quad (25)$$

Выполнен также усиленный закон больших чисел: для всех  $n$  и  $t > 0$

$$P \left\{ \sup_{k \geq n} \left| \frac{S_k}{k} \right| > t \right\} \leq t^{-2} e^{-2nt^2}. \quad (26)$$

Так как ряд экспонент в правой части неравенства (25) сходится, по лемме Бореля–Кантелли получим

**Следствие 1.** В условиях леммы 1 получаем усиленный мартингаловый закон больших чисел:

$$P \left\{ \lim_{n \rightarrow \infty} \frac{S_n}{n} = 0 \right\} = 1.$$

## 3. ИГРОВЫЕ СТРАТЕГИИ

В этом разделе мы рассмотрим финансовые приложения построенного алгоритма «универсального» прогнозирования. Будут рассмотрены игры на повышение (понижение) цены финансового инструмента. Результаты такой игры могут использоваться при так называемом техническом анализе, при котором изучаются временные ряды цен финансовых инструментов, но не используется информация о предприятиях, деятельность которых представлена этими инструментами.

Мы построим две стратегии игры на финансовом рынке, которые используют алгоритм прогнозирования, построенный в разделе 2.1. Эти стратегии используют для входа в рынок те шаги игры (моменты времени), на которых наш универсальный алгоритм прогнозирует рост цены акции на следующем шаге.

## 3.1. Игра на поиск тренда в подпоследовательности

Рассмотрим следующую игру на повышение (понижение) на фондовом рынке. Пусть  $S_1, S_2, \dots$  – последовательность цен некоторого финансового инструмента в дискретные моменты времени. Обозначим  $\Delta S_i = S_i - S_{i-1}$  – изменение цены в  $i$ -м интервале времени. Предполагаем, что  $0 < S_{i-1} \leq 1$  и  $S_1 = S_0$ .

Приведем протокол игры, который определяет последовательность действий участников игры.

**FOR**  $i = 1, 2, \dots$

В начале шага  $i$  участники рынка наблюдают цены финансового инструмента  $S_1, \dots, S_{i-1}$  за все предыдущие моменты времени.

*Предсказатель* предлагает вероятностный прогноз изменения цены финансового инструмента  $\Delta S_i \in [0, 1]$  – случайную величину  $\tilde{p}_i \in [0, 1]$ .

*Инвестор* определяет количество  $M_i$  приобретаемых или продаваемых единиц инструмента. Инвестор может пропустить данный шаг игры полагая  $M_i = 0$ .<sup>2</sup>

*Рынок* объявляет цену  $S_i$ .

*Инвестор* подсчитывает свой капитал на конец шага  $i$ :

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

**ENDFOR**

Здесь был рассмотрен случай  $M_i \geq 0$ , что соответствует так называемой длинной позиции при игре с акциями. Фондовый рынок также допускает случай  $M_i < 0$ , что соответствует игре в короткой позиции (в этом случае, дальнейшие рассуждения аналогичны).

В момент времени  $i$  наблюдаем значения  $S_1, \dots, S_{i-1}$  и вычисляем вероятностный прогноз  $p_i$  для  $\Delta S_i$  с использованием алгоритма универсального прогнозирования. Для удобства произведем нормировку и линейный сдвиг всех цен так, чтобы все рассматриваемые величины лежали в интервале  $[0, 1]$ .

Мы рассмотрим простейшую стратегию игры, которая обычно используется при игре на рынке акций. Предварительно рассмотрим игру с одной акцией.

Стратегия игры заключается в построении алгоритма для вычисления величины  $M_i$ . Для простоты рассмотрения будем покупать и продавать только одну акцию, таким образом,  $M_i = 0$  или  $M_i = 1$ . Мы оценим средний доход от игры с одной акцией. Отсюда легко перейти к случаю покупки и продажи постоянного числа акций.

Пусть число  $\epsilon > 0$  определяет условие входа в игру (правило выбора подпоследовательности для игры).

Правило входа в игру следующее:

*На шаге  $i$  входим в игру, т.е. покупаем  $M_i = 1$  акцию и продаем ее в конце этого периода, если  $p_i > \epsilon$ ; в противном случае, пропускаем этот шаг, т.е. определим  $M_i = 0$ .*

Мы будем использовать правила выбора следующего вида

$$I(p) = \begin{cases} 1, & \text{если } p > \epsilon, \\ 0, & \text{в противном случае.} \end{cases}$$

<sup>2</sup> Предполагаем, что Инвестор получает в начале шага сумму, необходимую для покупки  $M_i$  акций, а после их продажи в конце шага ее возвращает. Также  $\mathcal{K}_0 = 0$ .

В конце торгового периода получаем доход или убыток в размере  $S_i - S_{i-1}$  на одну акцию. Суммарный доход на одну акцию за  $n$  шагов игры равен

$$\mathcal{K}_n = \sum_{i=1}^n I(\tilde{p}_i > \epsilon)(S_i - S_{i-1}).$$

На шаге  $n$  используем точность округления  $\Delta_s \sim n^{-1/4}$ , где  $s$  такое, что  $n_s < n \leq n_{s+1}$ .

Представим величину дохода от одной акции за  $n$  шагов игры в виде

$$\begin{aligned} \mathcal{K}_n &= \sum_{\tilde{p}_i > \epsilon} \Delta S_i = \\ &= \sum_{i=1}^n I(\tilde{p}_i > \epsilon)(\Delta S_i - \tilde{p}_i) + \sum_{i=1}^n I(\tilde{p}_i > \epsilon)\tilde{p}_i. \end{aligned} \quad (27)$$

По определению второе слагаемое суммы (27) больше или равно

$$\epsilon \sum_{i=1}^n I(\tilde{p}_i > \epsilon).$$

Из (15) следует, что вероятность того, что первое слагаемое суммы (27) больше или равно

$$-(\Delta n + \mu_n + 2\sqrt{n/\Delta})$$

стремится к 1, при  $n \rightarrow \infty$ . Отсюда получаем, что вероятность события

$$\mathcal{K}_n \geq \epsilon \sum_{i=1}^n I(\tilde{p}_i > \epsilon) - (\Delta n + \mu_n + 2\sqrt{n/\Delta}) \quad (28)$$

стремится к 1, при  $n \rightarrow \infty$ . Здесь и далее  $I(\tilde{p}_i > \epsilon)$  есть конкретное правило выбора, которое в предыдущем разделе обозначалось  $I(\tilde{p}_i)$ .

Отсюда, вероятность того, что средний доход от одной акции (на один вход на рынок) удовлетворяет неравенству

$$\begin{aligned} k_n = \frac{\mathcal{K}_n}{\sum_{i=1}^n I(\tilde{p}_i > \epsilon)} &\geq \left( \epsilon - \frac{\Delta n + \mu_n + 2\sqrt{n/\Delta}}{\sum_{i=1}^n I(\tilde{p}_i > \epsilon)} \right) \sim \\ &\sim \left( \epsilon - \frac{2n^{2/3} + n^{2/3+1/M}}{\sum_{i=1}^n I(\tilde{p}_i > \epsilon)} \right) \end{aligned} \quad (29)$$

стремится к 1, при  $n \rightarrow \infty$ . В частности, для произвольного  $\gamma$ ,  $0 < \gamma < 1$ , вероятность того, что событие

$$\sum_{i=1}^n I(\tilde{p}_i > \epsilon) \geq \frac{\Delta n + \mu_n + 2\sqrt{n/\Delta}}{\gamma \epsilon} \quad (30)$$

влечет событие  $k_n \geq (1 - \gamma)\epsilon$ , стремится к 1 при  $n \rightarrow \infty$ . Условие (30) можно заменить на

$$\sum_{i=1}^n I(\tilde{p}_i > \epsilon) \geq \frac{cn^{2/3+\nu}}{\gamma \epsilon}, \quad (31)$$

где  $c$  - некоторая константа и  $\nu = 1/M$ .

Число  $\epsilon$  определяет условие выбора подпоследовательности входов в игру. В численных экспериментах, описанных далее в разделе 4, наилучшие результаты были получены для значений  $\epsilon = \epsilon' \sigma$ , где  $\sigma$  - среднее квадратичное отклонение начального отрезка последовательности исходов (приращений нормированных цен), а  $\epsilon'$  - число между 0 и 1.

Найдется ли достаточно длинная подпоследовательность входов, удовлетворяющая (30) или (31) зависит от величины отклонения исходной последовательности цен  $S_1, S_2, \dots$  от модели броуновского движения.

### 3.2. Игровая стратегия с ограниченным снизу риском

Важнейшим требованием к стратегии игры на финансовом рынке является требование к ее безопасности: должна существовать нижняя граница возможных потерь инвестора при любом неблагоприятном изменении цен акции.

Стратегия игры с одной акцией может быть легко модифицирована в стратегию с ограниченным снизу риском, если выбирать число акций  $M_i$  зависящим от текущего капитала (см. [10]). Пусть начальный капитал инвестора составлял  $\mathcal{K}_0$  единиц. Пусть также  $0 \leq \delta \leq 1$ . Пусть теперь при входе в игру приобретается не одна акция, как прежде, а некоторое их число, пропорциональное величине текущего капитала инвестора.

Определим число приобретаемых акций при входе в игру равным  $M_i = \delta \mathcal{K}_{i-1} / S_{i-1}$ , где  $\mathcal{K}_{i-1}$  - капитал инвестора перед  $i$ -м входом в игру. Предполагаем, что  $0 < S_{i-1} \leq 1$  и  $S_1 = S_0$ .

Тогда изменение капитала на  $i$ -входе в игру определяется соотношением

$$\mathcal{K}_i = \mathcal{K}_{i-1} + (\delta \mathcal{K}_{i-1} / S_{i-1})(S_i - S_{i-1}). \quad (32)$$

Логарифм текущего капитала инвестора на шаге  $n$  удовлетворяет неравенству

$$\begin{aligned} \ln \mathcal{K}_n &= \ln \mathcal{K}_0 + \sum_{i=1}^n I(\tilde{p}_i > \epsilon) \ln(1 + \delta(S_i - S_{i-1})/S_{i-1}) \geq \\ &\geq \ln \mathcal{K}_0 + \sum_{i=1}^n I(\tilde{p}_i > \epsilon) \ln(1 + \delta(S_i - S_{i-1})) \geq \\ &\geq \ln \mathcal{K}_0 + \delta \sum_{i=1}^n I(p_i > \epsilon)(S_i - S_{i-1}) - \\ &\quad - \delta^2 \sum_{i=1}^n I(p_i > \epsilon)(S_i - S_{i-1})^2. \end{aligned} \quad (33)$$

Здесь мы использовали неравенства  $|\Delta S_{i-1}| \leq 1$  и  $\ln(1+x) \geq x - x^2$  при  $|x| \leq 1$ .

Из  $|\Delta S_{i-1}| \leq 1$  имеем

$$\sum_{i=1}^n I(p_i > \epsilon)(S_i - S_{i-1})^2 \leq \sum_{i=1}^n I(p_i > \epsilon).$$

Из (28) следует, что вероятность события

$$\begin{aligned} &\sum_{i=1}^n I(p_i > \epsilon)(S_i - S_{i-1}) \geq \\ &\geq \epsilon \sum_{i=1}^n I(\tilde{p}_i > \epsilon) - cn^{2/3+1/M} \end{aligned} \quad (34)$$

стремится к 1 при  $n \rightarrow \infty$ .

Из (33) и (34) получаем, что вероятность события

$$\ln \mathcal{K}_n \geq \ln \mathcal{K}_0 + \delta \left( \epsilon \sum_{i=1}^n I(\tilde{p}_i > \epsilon) - cn^{2/3+1/M} \right) - \delta^2 \sum_{i=1}^n I(\tilde{p}_i > \epsilon) \quad (35)$$

стремится к 1 при  $n \rightarrow \infty$ .

Отсюда получаем, что вероятность того, что суммарный доход при такой игре удовлетворяет неравенству

$$\mathcal{K}_n \geq \mathcal{K}_0 e^{\delta \sum_{i=1}^n I(\tilde{p}_i > \epsilon) \left( (\epsilon - \delta) - \frac{cn^{2/3+1/M}}{\sum_{i=1}^n I(\tilde{p}_i > \epsilon)} \right)} \quad (36)$$

стремится к 1. Отсюда получаем, что для произвольного  $\gamma$ , такого, что  $0 < \gamma < 1$ , вероятность того, что событие

$$\sum_{i=1}^n I(\tilde{p}_i > \epsilon) \geq \frac{cn^{2/3+\nu}}{\gamma(\epsilon - \delta)}$$

влечет событие

$$\mathcal{K}_n \geq \mathcal{K}_0 e^{\delta(\epsilon - \delta)(1 - \gamma) \sum_{i=1}^n I(\tilde{p}_i > \epsilon)} \geq \mathcal{K}_0 e^{\delta(\frac{1}{\gamma} - 1)cn^{2/3+\nu}} \quad (37)$$

стремится к 1 при  $n \rightarrow \infty$ , где  $\nu = 1/M$ . В этих оценках мы использовали условие  $\epsilon > \delta$ .

Правая часть (37) всегда положительна, поэтому в игре можно потерять только начальный капитал  $\mathcal{K}_0$ , т.е.  $\mathcal{K}_n > 0$  для всех  $n$ .

## 4. ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ

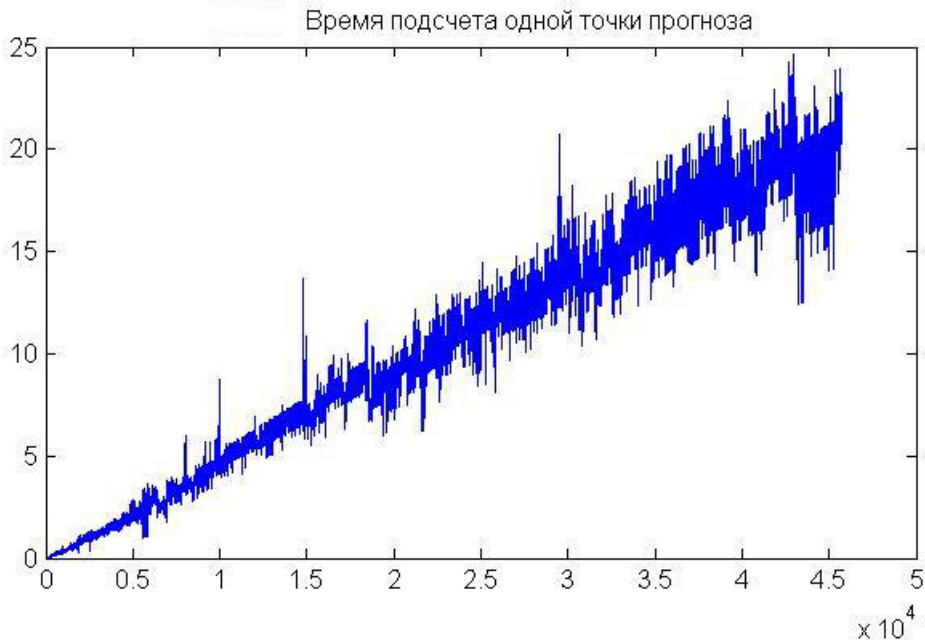
### 4.1. Методика расчетов

В соответствии с теорией универсального прогнозирования, изложенной в разделе 1, для получения детерминированного прогноза ряда на шаге  $n$  необходимо найти корень выражения (8), представляющего взвешенную сумму невязок на предыдущих шагах работы алгоритма. В качестве ядерной функции использовали гладкое приближение к ядру (7) – функцию  $K(p, p_n) = \cos((\pi(p - p_n)/2))$ .

Результаты численных экспериментов показывают, что наиболее эффективно прогнозировать не приращения цен  $\Delta S_i$ , а сами цены  $S_i$ . Кроме этого, цены  $S_i$  были нормированы и путем прибавления константы сдвинуты к центру интервала  $[0, 1]$ . Шкала нормирования цены подбиралась экспериментальным путем на обучающем сегменте последовательности (первые 2000 точек). Отбор подпоследовательности для игры производился по условию  $p_i > S_{i-1} + \epsilon$ , где  $p_i$  – прогноз цены, т.е. с помощью интервала  $I = (S_{i-1} + \epsilon, 1]$ .

Очевидно, что наибольшие временные затраты связаны с вычислением взвешенных сумм в (8) и поиском корней этого уравнения. На рисунке 1 приведены графики времени численного решения уравнения (8) в среде MATLAB в зависимости от длины предыстории.

Как видно из приведенного графика, время вычисления одной точки прогноза растет линейно с ростом длины предыстории и достигает 10 секунд при длине ряда 25000 точек. Чтобы

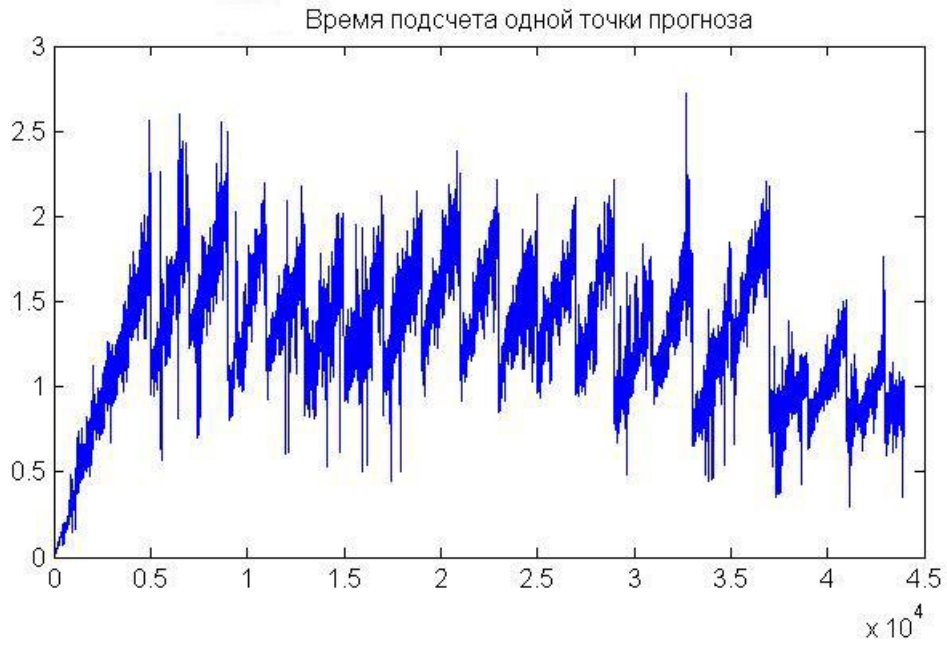


**Рис. 1.** График затрат машинного времени на прогноз очередной точки ряда (в секундах).

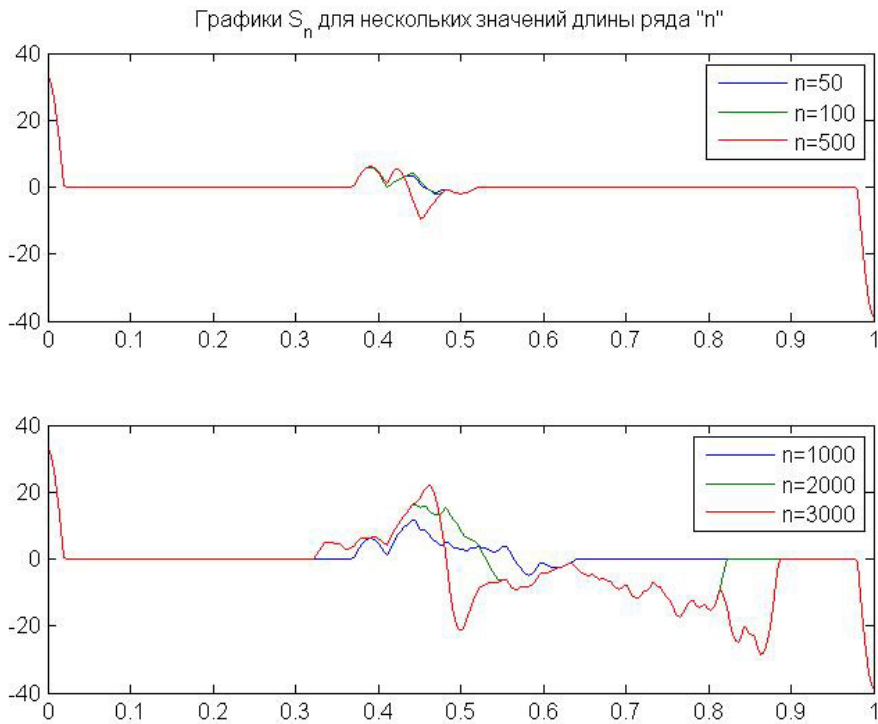
обеспечить время прогнозирования одной точки в пределах 1–3 секунд процессорного времени, необходимо ограничить длину ряда до 5000 точек. Для рядов длины, большей чем 5000 точек, применяли методику “цепного” прогнозирования, которая состоит в параллельном использовании двух процессоров, работающих на частично пересекающихся отрезках временного ряда. Пусть  $L_{max}$  - максимальная допустимая длина одного звена цепи, а  $L_{shift}$  - величина временного сдвига. При достижении первым процессом длины ряда в  $L_{shift}$  запускается второй процесс, в котором строятся прогнозы для ряда, начинающегося с точки, сдвинутой на  $L_{shift}$  относительно начала первого процесса. По достижении первым процессом длины ряда в  $L_{max}$  первый процесс останавливается и ждет сигнала от второго процесса о достижении точки  $L_{shift}$  для открывания очередного процесса прогнозирования и т.д. Результаты параллельных вычислений аккумулируются в единый суммарный прогноз. При выборе одного из значений прогноза для точек, где имеется два прогноза, предпочтение отдается процессу с большей длиной предистории. В описываемых далее экспериментах принимали  $L_{max}=5000$  точек, а  $L_{shift}=2000$  точек.

На рис. 2 приведены значения времени прогноза для цепной схемы вычислений. Как видим, процессорное время, затраченное на вычисление одной точки прогноза, не превышает 3 секунд. Это позволяет, в частности, успешно использовать данный алгоритм для прогноза временных рядов с частотой обновления 1 минута.

Рассмотрим несколько типичных примеров функций  $S_n(p)$ , корни которых являются значениями искомых прогнозов. На следующем рисунке приведены графики функций  $S_n(p)$  для разных значений длины предистории: для  $n = 10, 50, 100$  и для  $n = 1000, 2000, 3000$ . Значения  $p$  нормированы так, чтобы  $p \in [0, 1]$ . Горизонтальные участки с нулевым значением соответствуют областям, в которых еще нет экспериментальных точек и точек прогноза. Видно, как усложняется график  $S_n(p)$  по мере увеличения длины предистории.



**Рис. 2.** График затрат машинного времени на прогноз очередной точки ряда при цепной схеме вычислений (в секундах).



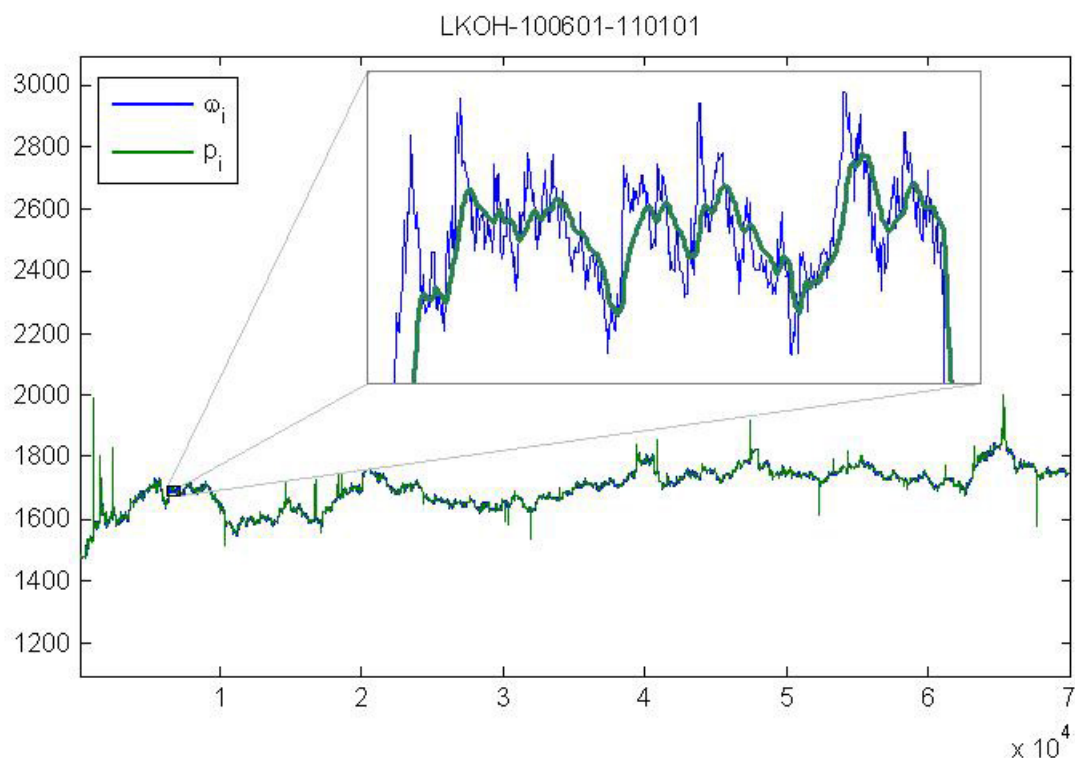
**Рис. 3.** Графики функции  $S_n(p)$  (где  $K(p; p_n) = \cos((\pi(p - p_n) = 2))$ ) для  $n = 10; 50; 100$  и для  $n = 1000; 2000; 3000$ . Значения  $p$  нормированы так, чтобы  $p \in [0; 1]$ .



## 4.2. Результаты экспериментов

Численные эксперименты по прогнозированию временных рядов и виртуальной торговле акциями проводили на материале поминутных цен акций нескольких российских компаний за период с 01.06.2010 до 01.01.2011, а также за период с 01.06.2007 до 01.01.2008, экспортированных с сайта [www.finam.ru](http://www.finam.ru). Игра велась поминутно, ее общая продолжительность около 70000 минут. На рис. 4 приведена кривая цены и прогноза цены акции ЛКОН за указанный период 2010 г.

Под транзакцией понимается покупка или продажа на финансовом рынке определенного количества единиц финансового инструмента (акции, фьючерсного контракта и др). За каждую транзакцию взимается комиссионный сбор (комиссия) в размере некоторого процента от суммы сделки. Величина комиссионного сбора зависит от статуса игрока (брокера). В данной работе использовалось минимальное значение – 0.01% от суммы сделки.



**Рис. 4.** Прогноз цены акций ЛКОН. Синим цветом показаны значения истинной цены акций, а зеленым значения прогноза.

Приведем результаты виртуальной игры на рынке ценных бумаг.

В первом столбце таблицы 1 приведены названия компаний, во втором – их тикер-символы, в третьем и последующих столбцах приведена доходность, достигнутая за 70000 минут виртуальной игры на рынке ценных бумаг с использованием нескольких стратегий. Период игры – с 01.06.2010 по 01.01.2011. Шаг игры – 1 минута.

Во всех экспериментах в качестве начального капитала  $K_0$  брали стоимость 500 акций соответствующей компании на момент, предшествующий началу игры. Доходность оценивали как отношение “заработанной суммы денег” при окончании периода игры ( $K_{\text{end}} - K_0$ ) к исходной сумме  $K_0$  в процентах:

$$D = 100 \times (K_{\text{end}} - K_0) / K_0.$$

Исследовали следующие стратегии – три варианта стратегии с ограниченным риском со значениями  $\delta = 0.5; 0.75; 1.0$  из (28), а также стратегию “buy-and-hold”, при которой в начале периода приобретается определенное число акций, которые продаются в конце игры. Порог  $\epsilon\sigma$  для входа в игру выбирали как среднее квадратичное отклонение цены соответствующей акции за первые несколько шагов временного ряда цен умноженное на число  $\epsilon$ ,  $0 < \epsilon \leq 1$ . На графиках  $\epsilon = 1$ . Рассматривались два варианта игры – (1) вариант без комиссии (отчисления за каждую транзакцию); (2) вариант с комиссией, в этом случае плата за торговую транзакцию была выбрана равной  $\tau = 0.0001$  от суммы сделки.

**Таблица 1.** Результаты игры на нескольких финансовых инструментах 2010 г. российского фондового рынка (комиссия отсутствует). В последних 4-х колонках указана доходность  $D$  в процентах.

Акция	Тикер	Частота входов в рынок	Средняя длительность	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$	Buy and hold
ОАО Газпром	GAZP	0.097	1.90	11.52	17.70	24.18	23.36
ОАО ЛУКОЙЛ	LKOH	0.099	1.85	36.28	59.02	85.48	10.84
МТС	MTSI	0.064	2.78	157.98	314.40	565.25	11.99
Роснефть	ROSN	0.099	1.90	13.37	20.64	28.31	1.47
ОАО Сбербанк	SBER	0.096	1.93	14.79	22.90	31.48	56.18
ГазпромНефть	SIBN	0.068	2.83	269.90	611.04	1264.81	11.56

**Таблица 2.** Результаты игры на нескольких финансовых инструментах 2010 г. российского фондового рынка (комиссия – 0.01%). В последних 4-х колонках указана доходность  $D$  в процентах.

Акция	Тикер	Частота входов в рынок	Средняя длительность	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$	Buy and hold
ОАО Газпром	GAZP	0.097	1.90	-42.41	-56.35	-66.92	23.36
ОАО ЛУКОЙЛ	LKOH	0.099	1.85	-30.27	-41.84	-51.51	10.84
МТС	MTSI	0.064	2.78	67.18	116.13	179.24	11.99
Роснефть	ROSN	0.099	1.90	-41.89	-55.75	-66.32	1.47
ОАО Сбербанк	SBER	0.096	1.93	-39.96	-53.53	-64.08	56.18
ГазпромНефть	SIBN	0.068	2.83	132.45	254.07	438.78	11.56

В Приложении приведены графики изменения выигрыша различных стратегий в играх с несколькими финансовыми инструментами, представленными в Таблицах 1 и 2. На рис. 5–16 приведены результаты игры по поминутным временным рядам за период 01.06.2010–01.01.2011 гг, а на рис. 17–22 приведены аналогичные графики за период 01.06.2007–01.01.2007 гг. На каждом рисунке приведены значения выигрыша каждой из четырех стратегий (без начального капитала  $\mathcal{K}_0$ ):

- Стратегия “buy and hold” (“купить и держать”), при которой в начале всего периода игры покупается  $M_0$  акций по текущей цене  $S_0$ , в конце периода они продаются по цене  $S_t$ . Доход при такой игре на каждом шаге  $t$  равен  $M_0(S_t - S_0)$ . Временная кривая дохода пропорциональна временной кривой цены акции.
- Другие три кривые соответствуют изменению капитала при трех безопасных стратегиях. Каждая из этих стратегий начинает игру с суммы, равной цене 500 акций данного типа, после этого капитал игрока может изменяться только за счет выигрыша или проигрыша в течение игры.

Первая стратегия при покупке акций использует весь накопленный капитал (всегда неотрицательный), т.е.  $\delta = 1$ .

- Вторая стратегия при покупке акций использует половину накопленного капитала, т.е.  $\delta = 0.5$  (остальную часть откладывает).
- Третья стратегия при покупке акций использует четверть накопленного капитала, т.е.  $\delta = 0.25$ .

Условие покупки акции на шаге  $t$  имеет вид  $p_t > S_{t-1} + \epsilon\sigma$ , где  $\sigma$  – среднее квадратичное отклонение цены акции за предыдущие 600 точек временного ряда. Акции не продаются до тех пор, пока на некотором шаге  $t' > t$  не будет  $p_{t'} < S_{t'-1}$  или  $S_{t'-1} < S_{t'-2}$ .

Для каждой акции приведены два варианта игры – вариант без комиссии и вариант с учетом комиссии. Эти отчисления составляют  $10^{-4}$  часть от суммы стоимости приобретаемого пакета акций.

Для акций GAZP при первом варианте (рис. 5) все стратегии выигрывают, причем больше всего выигрывает наиболее рискованная стратегия, при которой весь капитал используется при покупке акций. В конце периода стратегия “купить и держать” догоняет эту стратегию. Другие две более безопасные стратегии имеют меньший доход. При втором варианте (рис. 6) все стратегии, кроме стратегии “купить и держать”, проигрывают в обратном порядке.

Аналогичные результаты получены для акции LKOH (рис. 7, 8), ROSN (рис. 11, 12) и SBER (рис. 13, 14).

Для обоих вариантов игры с акциями MTSI (рис. 9, 10) и SIBN (рис. 15, 16) все стратегии выигрывают, несмотря на то, что цены этих акций мало изменились за рассмотренный период. На это указывает стратегия “купить и держать”, выигрыш которой прямо пропорционален цене акции. Стратегии выигрывают за счет предсказуемого характера поведения цены акции, за исключением стратегии “купить и держать”, которая соответствует мало изменяющейся цене акции.

На рис. 17–20 приведены результаты поминутной игры на тех же акциях по временным рядам за период 01.06.2007–01.01.2008. В этот период происходил рост цен многих акций, поэтому все стратегии на приведенных рисунках имеют положительный выигрыш, даже при вычитании комиссионного сбора.

При одновременной работе с несколькими инструментами появляется возможность оперативного перераспределения ресурсов между игроками. В качестве иллюстрации на рисунках 21–22 приведены графики одновременной игры на нескольких инструментах. Причем на рисунке 21 показаны результаты для случая, когда игроки действуют независимо друг от друга, а результаты их деятельности просто суммируются. На следующем рисунке 22 показаны результаты коллективной игры, при которой в конце каждого торгового дня сравнивали качество работы игроков и перераспределяли активы между ними. Критерием качества работы игрока была его доходность за последние пять дней. За изъятие денег со счета и вложение взималась плата (комиссионный сбор) в виде 0.01% от соответствующей суммы. В начале игры каждый игрок получил одну и ту же условную сумму. Всего в игре участвовало 6 акций, причем с каждой из акций работало два торговых агента – один следовал стратегии активной торговли, описанной в разделе 3.2, другой использовал стратегию “купить и держать” и при этом также имел возможность в некоторые моменты времени покупать или продавать свои акции. В первом случае простого суммирования деятельности 12 условных игроков доходность за семь месяцев игры составила 40%, а при “смешивании” алгоритмов 124%. Более подробно результаты смешивания стратегий будут описаны в следующих публикациях.

## 5. ЗАКЛЮЧЕНИЕ

В данной работе предложена одна модификация алгоритма построения калибруемых прогнозов Какаде и Фостера. Этот вероятностный алгоритм строит такие прогнозы для произвольной последовательности исходов, предъявляемой этому алгоритму в режиме онлайн. При этом природа источника, генерирующего последовательность, неизвестна алгоритму. Доказана асимптотическая состоятельность алгоритма с точки зрения критерия калибруемости прогнозов.

Также рассмотрены приложения алгоритма для построения игровых стратегий, основанных на прогнозах будущих исходов временного ряда. Построены игровые стратегии для использования на финансовых рынках. Получены достаточные условия, при которых данные стратегии имеют положительный выигрыш.

Имеется ряд нерешенных задач. В частности, требуют улучшения теоретические оценки для значений параметров  $\epsilon$  и  $\delta$  и оценки для необходимого числа входов в рынок из разделов 3.1 и 3.1. Как часто случается на практике, эти оценки не соответствуют величинам этих параметров, которые использовались при численных экспериментах. В последних значения параметров подбирались экспериментальным путем. Тем не менее, оценки из разделов 3.1 и 3.1 определили ход планирования численных экспериментов.

Результаты численного моделирования приведенных игр показывают, что метод вычисления калибруемых прогнозов может применяться для построения игровых стратегий. Результаты моделирования могут применяться не только для построения финансовых роботов (компьютерных программ, играющих на финансовом рынке в режиме реального времени), но и при техническом анализе временных рядов финансовых инструментов.

Рынок называется эффективным, если никакая безопасная стратегия с ограниченным снизу риском не может иметь неограниченный выигрыш. На математическом языке, при наличии предположения о стохастическом механизме генерации цен финансового инструмента, это свойство формулируется таким образом: любой ограниченный снизу мартингал ограничен сверху с вероятностью 1.

Численные эксперименты показывают, что на всех временных рядах цен акций можно выигрывать (без комиссии). Это означает, что рынок акций не является эффективным. Наличие комиссионного сбора делает этот рынок “более” эффективным.

Результаты моделирования могут применяться при определении научно обоснованной величины комиссионного сбора, взимаемого за операции на финансовом рынке.

*ПРИЛОЖЕНИЕ*

Ниже приводятся графики с результатами численного моделирования на временных рядах акций 2010 и 2007 гг. Все эти временные ряды выгружены с сайта

[http : //www.finam.ru/analysis/export/default.asp](http://www.finam.ru/analysis/export/default.asp).

Численные эксперименты проводились в среде MATLAB.

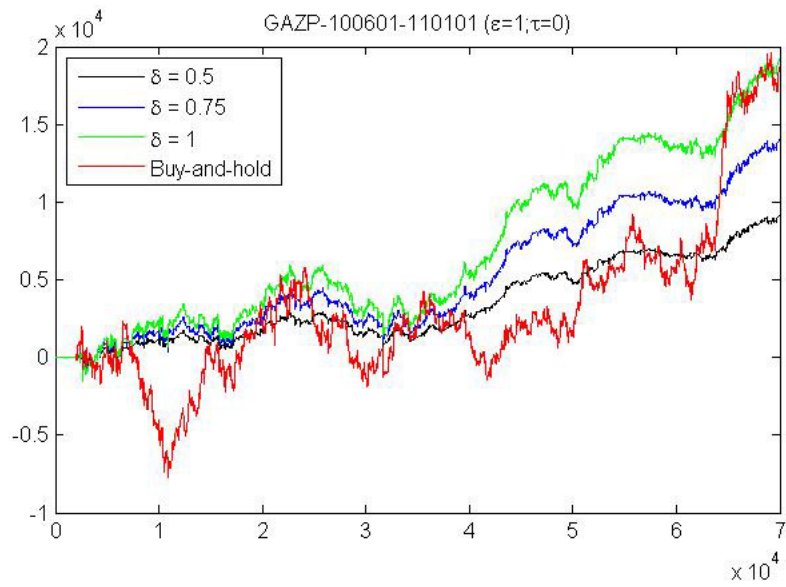


Рис. 5. Результаты игры с акциями GAZP 2010: комиссия отсутствует.

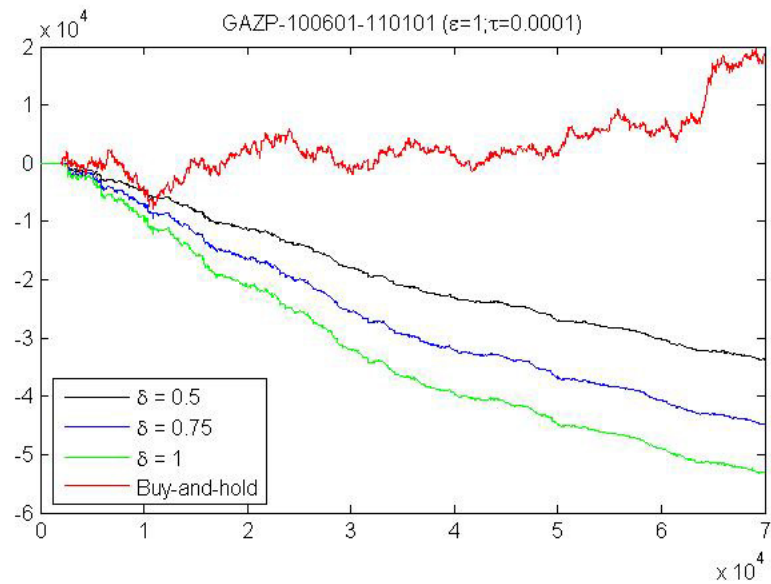


Рис. 6. Результаты игры с акциями GAZP 2010: комиссия –  $10^4$  часть от суммы сделки.

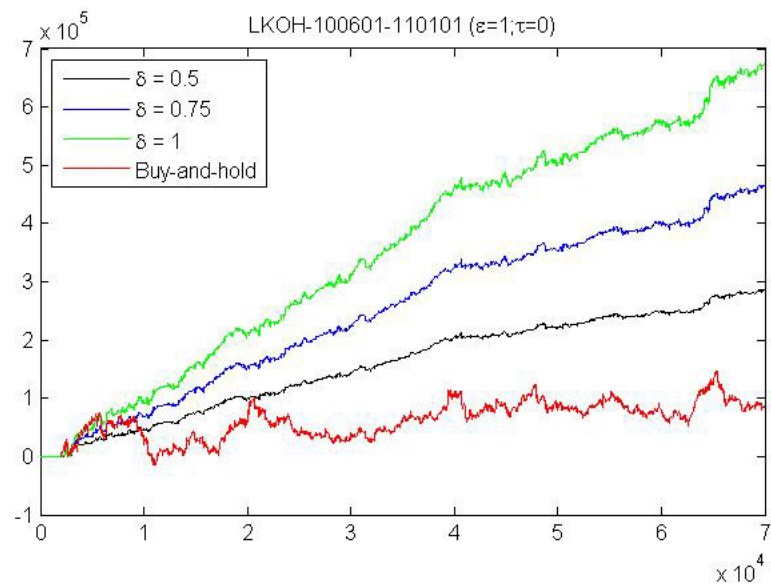


Рис. 7. Результаты игры с акциями ЛКОН 2010: комиссия отсутствует.

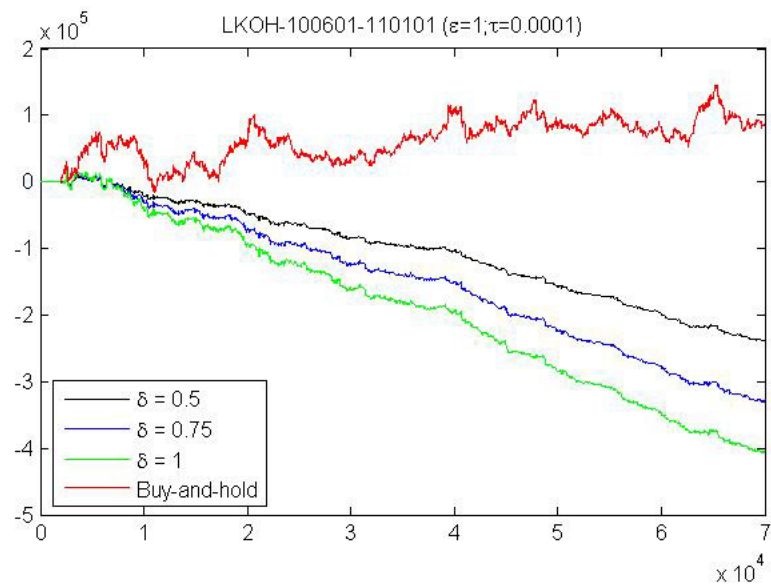


Рис. 8. Результаты игры с акциями ЛКОН 2010: комиссия –  $10^4$  часть от суммы сделки.

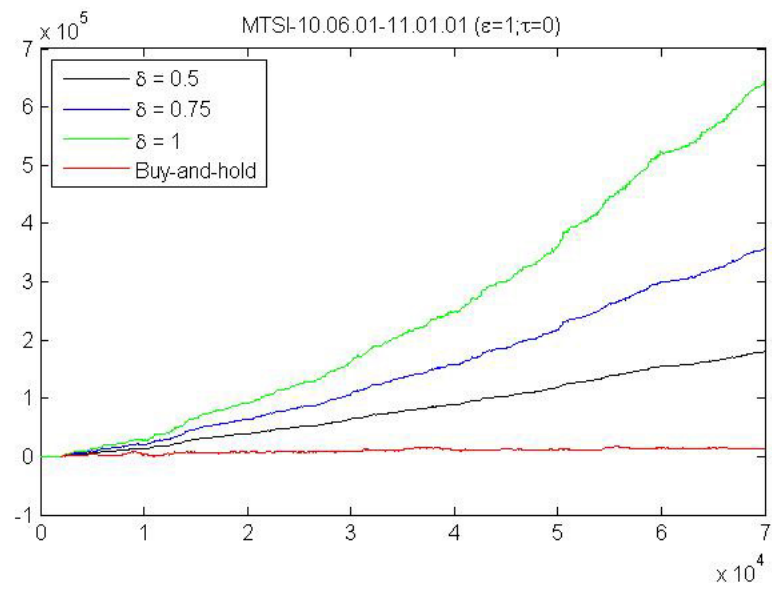


Рис. 9. Результаты игры с акциями MTSI 2010: комиссия отсутствует.

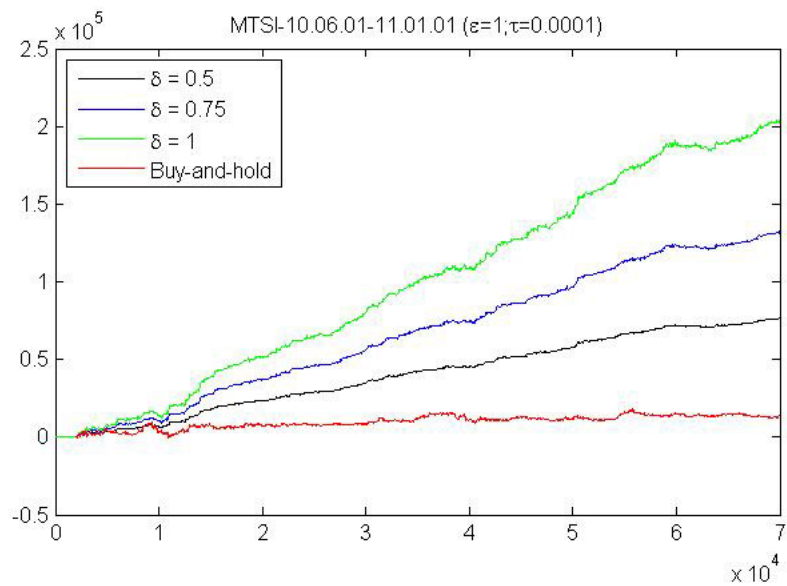


Рис. 10. Результаты игры с акциями MTSI 2010: комиссия –  $10^4$  часть от суммы сделки.

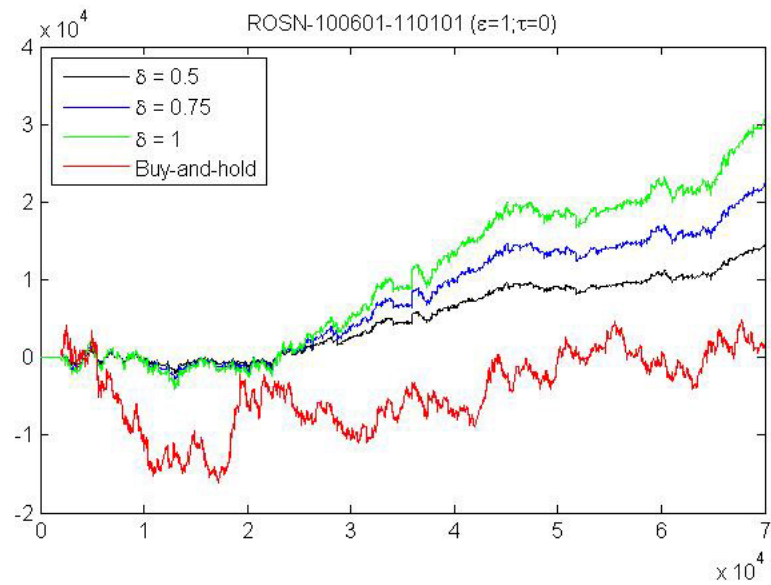


Рис. 11. Результаты игры с акциями ROSN 2010: комиссия отсутствует.

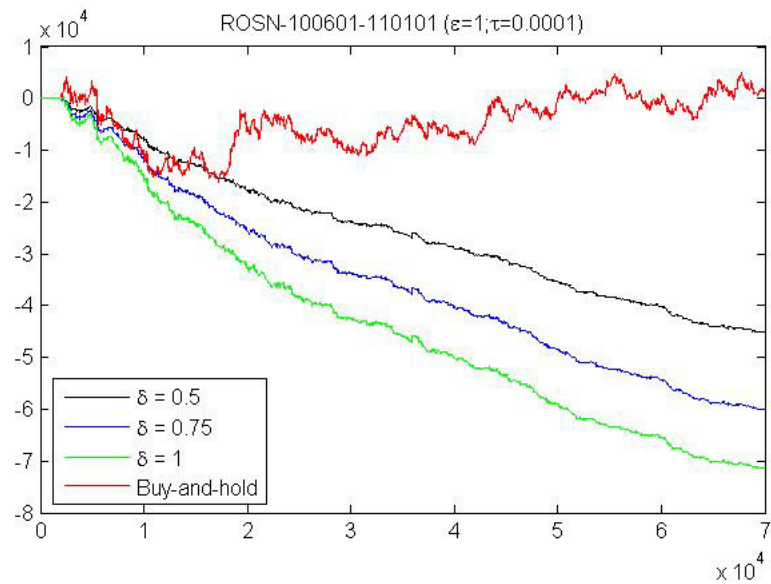


Рис. 12. Результаты игры с акциями ROSN 2010: комиссия –  $10^4$  часть от суммы сделки.



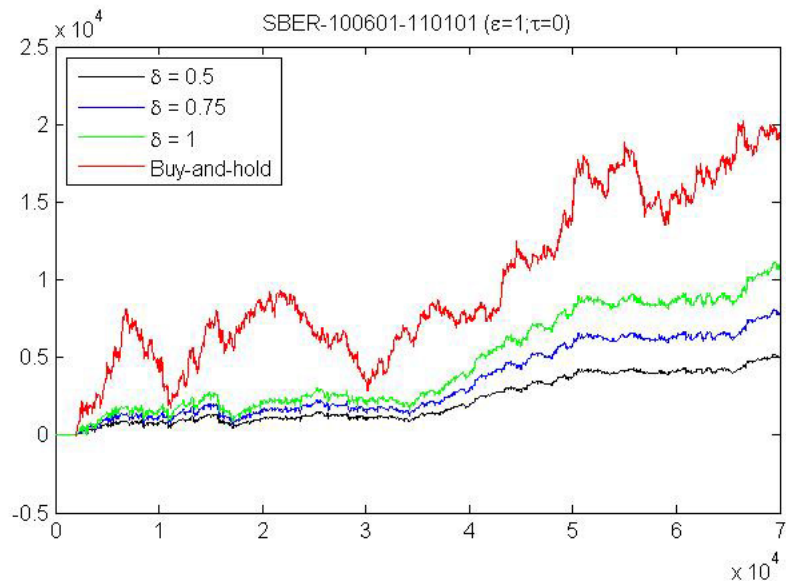


Рис. 13. Результаты игры с акциями SBER 2010: комиссия отсутствует.

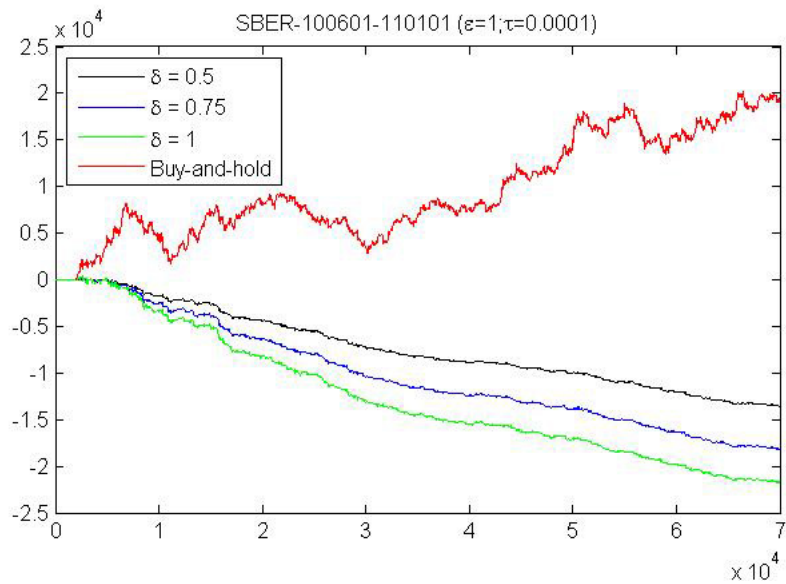


Рис. 14. Результаты игры с акциями SBER 2010: комиссия –  $10^4$  часть от суммы сделки.

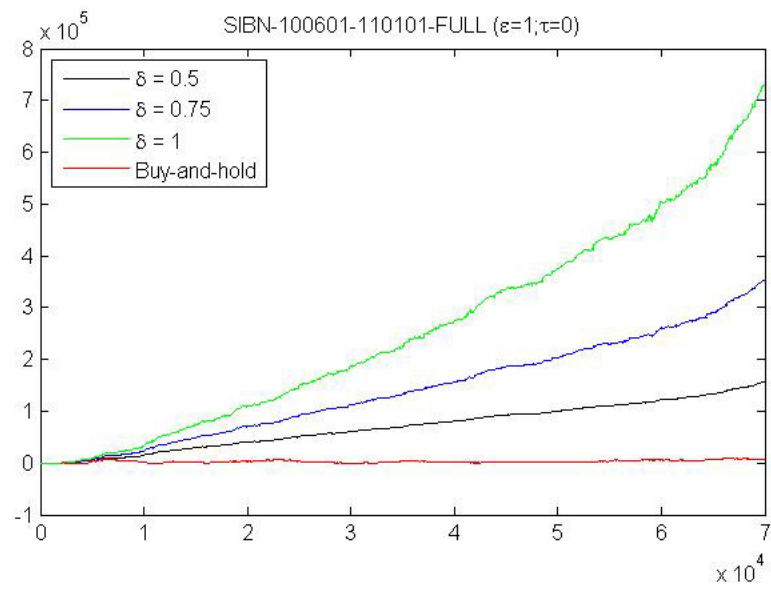


Рис. 15. Результаты игры с акциями SIBN 2010: комиссия отсутствует.

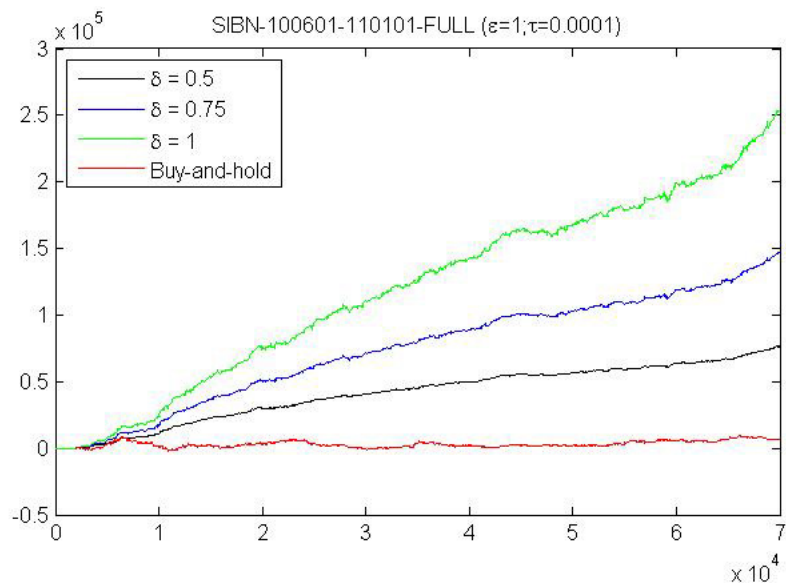


Рис. 16. Результаты игры с акциями SIBN 2010: комиссия –  $10^4$  часть от суммы сделки.

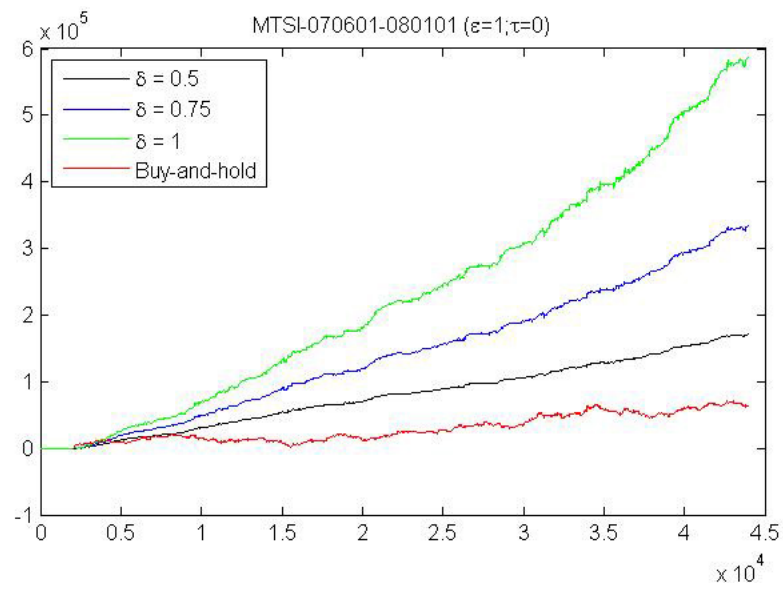


Рис. 17. Результаты игры с акциями MTSI 2007: комиссия отсутствует.

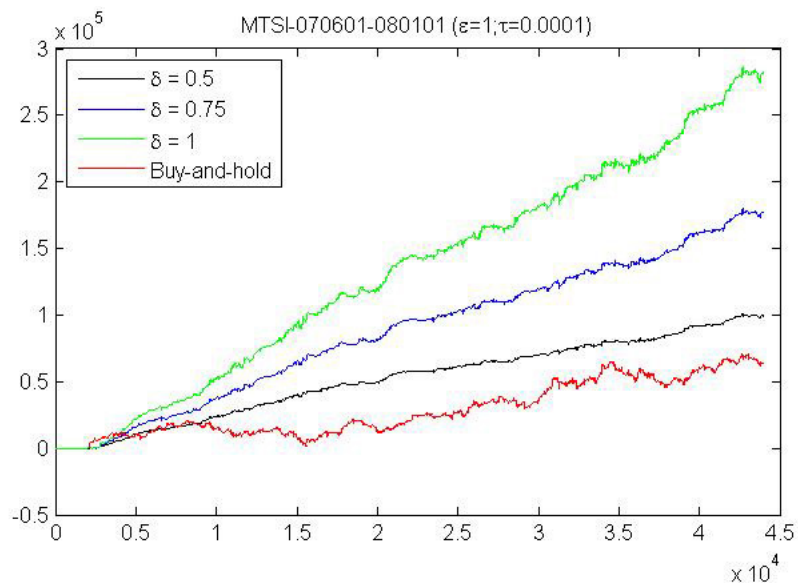


Рис. 18. Результаты игры с акциями MTSI 2007: комиссия –  $10^4$  часть от суммы сделки.

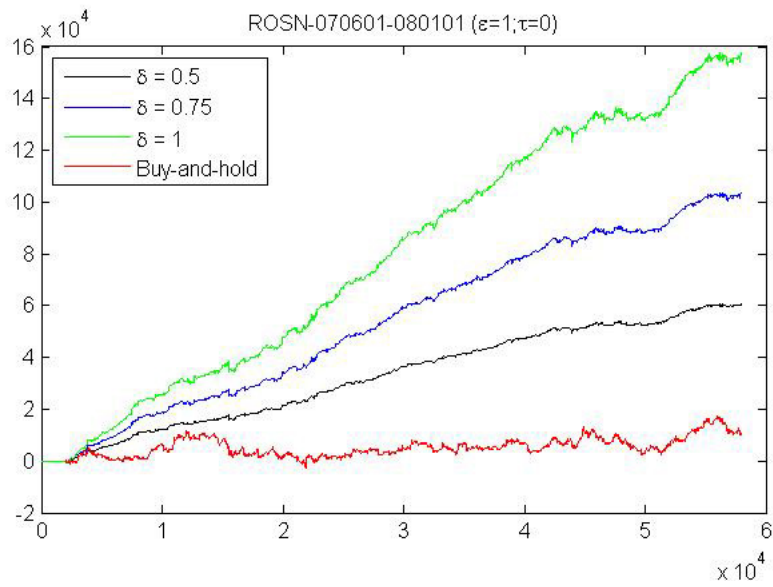


Рис. 19. Результаты игры с акциями ROSN 2007: комиссия отсутствует.

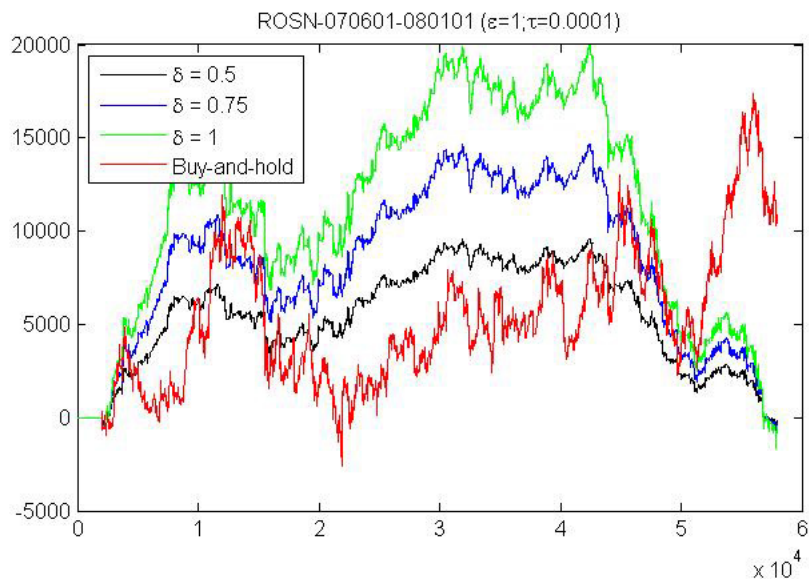


Рис. 20. Результаты игры с акциями ROSN 2007: комиссия –  $10^4$  часть от суммы сделки.

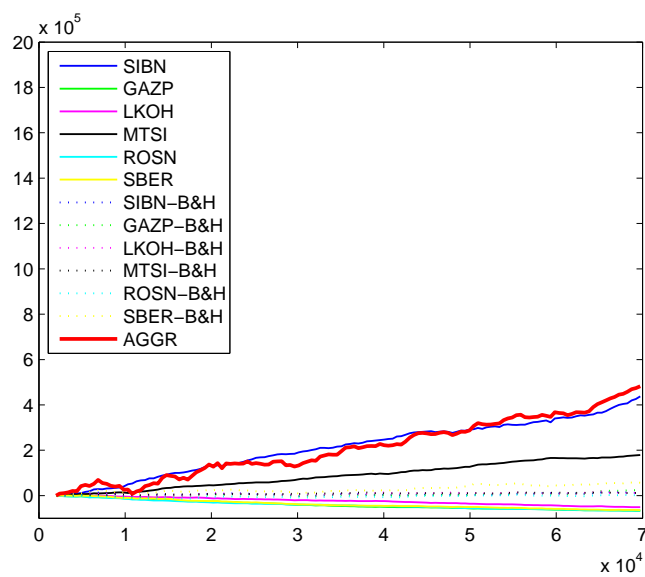


Рис. 21. Результаты суммирования работы независимых торговых агентов.

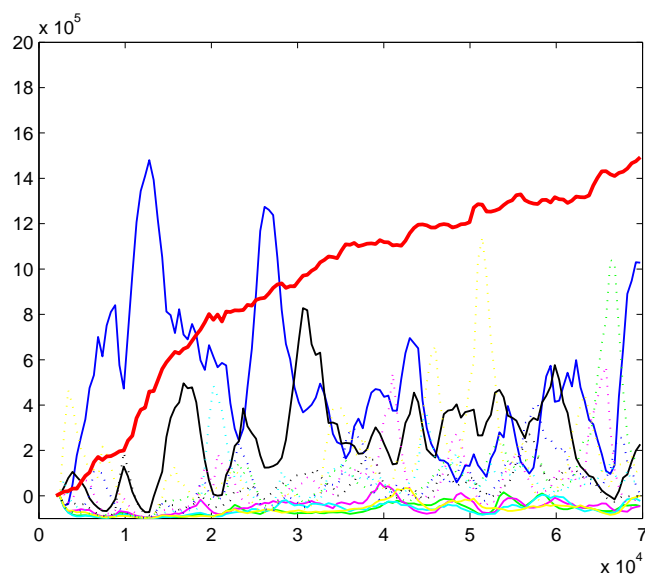


Рис. 22. Результаты совместной работы торговых агентов.

## СПИСОК ЛИТЕРАТУРЫ

1. В.В.Вьюгин, Е.Л.Румянцев, Ф.М.Цфасман. Игровые стратегии и адаптивное прогнозирование // Сборник трудов конференции ИТиС'08, М.: ИППИ РАН, С.391–396, 2008.
2. Ширяев А.Н. Вероятность. М.: Наука, 1980.
3. N.Cesa-Bianchi, G.Lugosi. Prediction, Learning and Games. Cambridge University Press, 2006.
4. A.P.Dawid. Calibration-based empirical probability [with discussion] // Ann. Statist. 13, 1251-1285, 1985.
5. D.P.Foster, R. Vohra. Asymptotic calibration // Biometrika. 85, 379–390, 1998.
6. S.M.Kakade, D.P.Foster. Deterministic calibration and Nash equilibrium // LNCS 3120 (John Shawe Taylor and Yoram Singer. ed). P.33–48, 2004.
7. D.Oakes. Self-Calibrating Priors Do not Exist [with discussion] // J. Am. Statist. Assoc. V.80. P.339–342, 1985.
8. A.Sandroni, R.Smorodinsky, R.Vohra. Calibration with Many Checking Rules // Mathematics of Operations Research. V.28. P.141–153, 2003.
9. V.Schervish, Comment [to Oakes, 1985] // J. Am. Statist. Assoc. V.80. P.341–342, 1985.
10. V.Vovk, A.Takemura, G.Shafer. Defensive Forecasting // Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics. 365–372, (<http://arxiv.org/abs/cs/0505083>), 2005.
11. V.Vovk. Defensive Forecasting for Optimal Prediction with Expert Advice // arXiv:0708.1503v1. 2007.