

УДК 512.234

В.А. Симахин

**АДАПТИВНЫЕ РОБАСТНЫЕ НЕПАРАМЕТРИЧЕСКИЕ  
АЛГОРИТМЫ ПРОГНОЗА**

В работе на основе взвешенного метода максимального правдоподобия синтезированы робастные непараметрические алгоритмы прогноза стационарных временных рядов. Данные алгоритмы включают как частный случай классические непараметрические оценки прогноза.

**Ключевые слова:** *временной ряд, робастность, непараметрический, прогноз.*

Статистическому анализу временных рядов посвящено большое количество исследовательских работ (см., например, [1 – 3, 5]). В качестве математической модели случайного процесса используются общие гетероскедастические модели нелинейной авторегрессии

$$X_i = m(\bar{X}_{i-m-1}) + g(\bar{X}_{i-m-1})\varepsilon_i, \quad (1)$$

где  $\bar{X}_{i-m-1} = (X_{i-1}, \dots, X_{i-m})^T$ ,  $m(\bar{X}_{i-m-1})$  – функция авторегрессии,  $g(\bar{X}_{i-m-1})$  – функция волатильности,  $\varepsilon_i$  – последовательность независимых и одинаково распределенных случайных величин. При непараметрической постановке задачи структура функций  $m(\bar{X}_{i-m-1})$ ,  $g(\bar{X}_{i-m-1})$  и распределение  $\varepsilon_i$  считаются неизвестными, удовлетворяющими достаточно общим предположениям [2, 3]. Необходимо отметить, что работы на основе непараметрического подхода занимают достаточно скромное место в исследованиях временных рядов, тому свидетельство обзоры [2, 3]. Большая часть этих работ связана с анализом условных функционалов в условиях зависимой выборки [2, 3, 5]. Начиная с 60-х гг. прошлого столетия возрос интерес к исследованиям, связанным с созданием робастных алгоритмов прогноза временных рядов [4]. Для задачи прогноза в непараметрической постановке задачи предложено три непараметрических алгоритма – регрессионного, медианного и модального типов [2, 3, 5 – 7]. Данные алгоритмы были получены по аналогу регрессионных задач путем переноса их на авторегрессионные задачи. Алгоритм регрессионного типа [3, 5, 6] не является робастным и, очевидно, считалось, что алгоритмы медианного и модального типов решают задачу робастного непараметрического прогноза. Причем робастные свойства алгоритма модального типа мало изучены. Анализ показывает, что алгоритмы прогноза, наследуя свойства алгоритмов регрессионных задач, в то же время не в полной мере учитывают специфику задачи прогноза. Прогноз проводится по последним членам временного ряда, и любой выброс в этих данных приводит к катастрофическим результатам. Меры борьбы с выбросами в независимых переменных, применяемые в регрессионном анализе, в данном случае неприменимы. Необходимо принимать дополнительные меры, чтобы получить робастные алгоритмы прогноза. В данной работе на основе взвешенного метода максимального правдоподобия [9] синтезирован класс робастных (устойчивых) непараметрических алгоритмов для задачи прогноза. Данные алгоритмы включают как частный случай непара-

метрические алгоритмы регрессионного, медианного и модального типов, введенные ранее [2, 3, 5–7].

## 1. Постановка задачи и анализ выбросов

### 1.1. Постановка задачи

Рассмотрим задачу прогнозирования случайного стационарного в узком смысле процесса  $X(t)$  вида (1), удовлетворяющего условию слабой зависимости (сильного перемешивания (СП) или равномерно сильного перемешивания (РСП)) [5] с неизвестной функцией распределения  $F(\bar{x}, t) \in \Omega$ , где  $\Omega$  – непараметрический класс абсолютно непрерывных функций распределения.

Пусть  $x_1, \dots, x_N$  – выборка из случайного процесса  $X(t)$ , наблюдаемого на интервале  $[0, T]$  ( $x_i = X(t_i)$ ,  $t_{i+1} - t_i = \Delta t$ ,  $x_N = X(T)$ ). Требуется по  $m$  последним значениям временного ряда  $\bar{x}_{N-m} = (x_N, \dots, x_{N-m+1})^T$  найти прогноз  $X_{N+k}$  на  $k$  тактов вперед.

В общем случае оптимальный прогноз в зависимости от критерия качества есть функционал от условной функции распределения  $F(t / \bar{x}_{N-m})$

$$X_{N+k} = J(F(t / \bar{x}_{N-m})). \quad (2)$$

Вид функционала  $J$  зависит от используемого критерия качества. В частности, для квадратичной функции потерь оптимальным прогнозом является условное среднее. В дальнейшем были введены алгоритмы на условной медиане  $F(X_{N+k} / \bar{x}_{N-m}) = 0,5$  и условной моде  $X_{N+k} = \arg \max_t F'(t / \bar{x}_{N-m})$  [2, 3, 7]. Данные алгоритмы рекомендовались к применению в случае, если выборка «засорена» (имеются, например, единичные дельтообразные выбросы).

В непараметрическом случае, когда вид  $F(\bar{x}, t)$  неизвестен, в качестве алгоритмов прогноза берутся оценки

$$\hat{X}_{N+k} = J(F_N(t / \bar{x}_{N-m})), \quad (3)$$

где  $F_N(t / \bar{x}_{N-m})$  – непараметрическая оценка условной функции распределения.

В качестве  $F_N(t / \bar{x}_{N-m})$  будем использовать две классические оценки из класса дискретных и непрерывных оценок, введенных в [10] и [11] соответственно:

$$F_N(t / \bar{x}_{N-m}) = \frac{1}{N-m-k} \sum_{i=1}^{N-m-k} C(t - x_{i+m+k-1}) W(\bar{x}_i / \bar{x}_{N-m}); \quad (4)$$

$$\tilde{F}_N(t / \bar{x}_{N-m}) = \int G\left(\frac{t-z}{h_N}\right) dF_N(z / \bar{x}_{N-m}), \quad (5)$$

где 
$$W(\bar{x}_i / \bar{x}_{N-m}) = \frac{1}{h_N^m} \prod_{j=1}^m K\left(\frac{x_{N+j-1} - x_{i+j-1}}{h_N}\right) \cdot f_N^{-1}(\bar{x}_{N-m}), \quad (6)$$

$$f_N(\bar{x}_{N-m}) = \frac{1}{N-m} \sum_{i=1}^{N-m} \frac{1}{h_N^m} \prod_{j=1}^m K\left(\frac{x_{N+j-1} - x_{i+j-1}}{h_N}\right) \quad (7)$$

– оценка плотности Розенблатта – Парзена;

$$\begin{aligned} \bar{x}_i &= (x_i, \dots, x_{i+m-1})^T, \quad 0 \leq G(u) \leq 1, \quad G(-\infty) = 0, \quad G(\infty) = 1, \quad G(-u) = 1 - G(u), \\ \int u^2 dG(u) &< \infty, \quad G'(u) = K(u), \quad \int K(u) du = 1, \quad \int uK(u) du = 0, \quad \int u^2 K(u) du < \infty, \\ \int K^2(u) du &= Q < \infty, \quad h_N \rightarrow 0, \quad N \cdot h_N^m \rightarrow \infty, \quad \text{при } N \rightarrow \infty, \end{aligned}$$

$C(u) = [1, u \geq 0; 0, u < 0]$  – функция Хевисайда.

**Замечание 1.** Выше приведены классические ограничения на ядерные функции  $K(u)$ ,  $G(u)$  и коэффициент размытости  $h_N$ . На коэффициенты перемешивания процесса обычно накладываются условия, обеспечивающие сходимость оценок, например для РСР  $\sum_{\tau=1}^{\infty} \sqrt{\beta(\tau)} < \infty$  [5, разд. 3.1]. В дальнейшем будем считать, что данные условия регулярности выполняются [2, 3, 5 – 7].

**Замечание 2.** К настоящему времени предложены различные алгоритмы на основе активных ядер (см., например, [2, 3, 5]), позволяющие повысить скорость сходимости оценок. Применение методов получения активных ядер (например, метод локальной аппроксимации) усложняет математику, но не меняет сути, в связи с этим в данной статье на них останавливаться не будем.

## 1.2. Анализ выбросов

Без потери общности будем считать  $\kappa=1$ . Основные алгоритмы прогноза относятся к «гусеничным алгоритмам». По временному ряду  $x_1, \dots, x_N$  прокатывается гусеница, которая последовательно нарезает блоки  $\bar{x}_i = (x_i, \dots, x_{i+m-1})^T$ ,  $y_i = x_{i+m}$ ,  $i = 1, \dots, N - m$ . В результате получаем выборку  $\{\bar{x}_i, y_i\}$ ,  $i = 1, \dots, N - m$ , где наблюдение  $y_i$  играет роль зависимой переменной, а вектор  $\bar{x}_i$  выступает в виде вектора независимых переменных. Таким образом, задача прогноза сводится к задаче многомерной регрессии при зависимых наблюдениях. Нетрудно заметить, что в «гусеничных алгоритмах» соседние блоки частично перекрываются, а это приводит к сильной зависимости между ними. Отметим ряд моментов, полученных при анализе условных функционалов для процессов, удовлетворяющих условию слабой зависимости [5]. Условия СП и РСР влекут за собой эргодичность процесса. Свойства непараметрических оценок условных функционалов в случае независимых и слабовзависимых наблюдений совпадают – «принцип локальности» («the whitening by windowing principle» Hart (1996)) [5]. «Принцип локальности» имеет место и для случая непараметрических процедур обработки с частично перекрывающимися блоками. В соседних блоках зависимые переменные  $y_i$  и независимые  $x_i$  меняются местами. В связи с этим появление даже одного выброса во временном ряду приводит к выбросу как по зависимой переменной, так и к  $(m-1)$  выбросам в независимых переменных  $\bar{x}_i$ . Следовательно, необходимы робастные алгоритмы регрессии, устойчивые как по оси  $X$ , так и по оси  $Y$ . Такие непараметрические алгоритмы имеются [8]. Особую роль для задачи прогноза играет вектор  $\bar{x}_{N-m}$ , который единственен, и только на его основе строится прогноз. Появление выброса в векторе  $\bar{x}_{N-m}$  приводит к катастрофическим результатам для прогноза. Действительно, весовая функция  $W(\bar{x}_i / \bar{x}_{N-m})$  играет роль фильтра, вырезающего из временного ряда  $x_1, \dots, x_{N-m}$  однородные с  $\bar{x}_{N-m}$

блоки  $\bar{x}_i$  (кластеры). Задача прогноза в этом случае сводится к задаче оценки параметра положения фильтрованных данных  $y_i$ . Параметр  $m$  играет роль размера (интервала зависимости) кластера, а параметр  $h_N$  определяет ширину пропуска фильтра. Область применения таких алгоритмов связана с некоторыми особенностями исходного процесса  $X(t)$ , в частности со стационарностью по кластерам. Причем исходный процесс  $X(t)$  может состоять из совокупности кластеров с разной волатильностью. Отметим, что с «кластерностью» при изучении случайных процессов столкнулись и при параметрическом подходе при изучении моделей случайных процессов с волатильностью (ARCH, GARCH, ...) [1].

## 2. Устойчивые непараметрические алгоритмы прогноза

В работе [12] предложены методы получения робастных оценок на основе критерия устойчивости. Оказалось, что класс устойчивых оценок может быть синтезирован на основе взвешенного метода максимального правдоподобия (ВММП) [8, 9, 14].

Обозначим через  $\zeta_i = x_{i+m} - X_{N+k} = y_i - X_{N+k}$  – невязки прогноза и через  $p(z, \theta / \bar{X}_{N-m})$ ,  $P(z, \theta / \bar{X}_{N-m})$  – плотность и функцию распределения невязок в зависимости от реализации вектора  $\bar{X}_{N-m}$ . В данном случае нас интересует неизвестный параметр распределения  $\theta = (X_{N+k}, S)^T$ . Прогноз  $X_{N+k}$  выступает как параметр сдвига, а оценка параметра масштаба  $S$  позволяет оценить вариацию прогноза, то есть

$$p(x, \theta / \bar{X}_{N-m}) = \frac{1}{S} h(u), \quad (8)$$

где  $h(u)$  – стандартная плотность распределения семейства  $P(x)$  и

$$u = [x - X_{N+k}] \cdot S^{-1}.$$

Условную М-оценку  $\hat{\theta}$  неизвестного параметра  $\theta$  можно определить на основе решения системы эмпирических уравнений вида

$$\int \varphi_i(t, \hat{\theta} / \bar{X}_{N-m}) dF_N(t / \bar{X}_{N-m}) = 0, \quad i = 1, 2, \quad (9)$$

где  $\varphi_i(t, \theta / \bar{X}_{N-m})$  – оценочная функция.

Определим оценочную функцию  $\varphi(t, \theta / \bar{X}_{N-m})$  в следующем виде [8, 9, 14]:

$$\varphi(x, \theta / \bar{X}_{N-m}) = \left[ \frac{\partial}{\partial \theta} \ln p(x, \theta / \bar{X}_{N-m}) + \beta \right] p^l(x, \theta / \bar{X}_{N-m}), \quad (10)$$

где  $\beta$  – параметр, который определяется из условия несмещенности оценки

$$M[\varphi(x, \theta / \bar{X}_{N-m})] = 0 \quad (11)$$

и  $l$  – параметр, который назовем параметром радикальности оценки.

Выражение (11) определяет ВММП с весами  $p^l(x, \theta / \bar{X}_{N-m})$ : при  $l = 0$  получаем оценки максимального правдоподобия (ОМП); при  $l = 0,5$  – радикальные оценки; при  $l = 1$  – оценки максимальной устойчивости (ОМУ) [12]. Физически роль параметра  $l$  сводится к определению степени «мягкого» усечения алгоритма,

настраивая его на вид априорного распределения. Следовательно, варьируя параметром  $l$ , можно получать эффективные оценки при локальных отклонениях распределения  $P(x, \theta / \bar{X}_{N-m})$  в классе устойчивых оценок [12].

Для параметров сдвига и масштаба в соответствии с (10) и (12) получаем оценочные функции в виде [9, 14]:

$$\begin{aligned} \varphi_1(x, \hat{\theta} / \bar{X}_{N-m}) &= h'_u(u) h^{l-1}(u), \\ \varphi_2(x, \hat{\theta} / \bar{X}_{N-m}) &= \left[ u h'_u(u) \frac{1}{p(x, \hat{\theta} / \bar{X}_{N-m})} - \frac{1}{(l+1)} \right] \cdot p^l(x, \hat{\theta} / \bar{X}_{N-m}). \end{aligned} \quad (12)$$

Система уравнений (9) и оценочные функции (12) определяют алгоритм получения устойчивых оценок параметров сдвига и масштаба в зависимости от параметра радикальности  $l$ .

Рассмотрим в качестве примеров два случая.

1.  $p(t, \theta / \bar{X}_{N-m}) = \frac{1}{\sqrt{2\pi}S} \exp\left\{-\frac{(t - \hat{X}_{N+k})^2}{2S^2}\right\}$  – нормальное распределение невязок. Из (9), (12) получаем следующую систему уравнений для оценки прогноза.

$$\sum_{i=1}^{N-m-k} u_i \exp\left\{\frac{l}{2} u_i^2\right\} W(\bar{x}_i / \bar{x}_{N-m}) = 0; \quad (13)$$

$$\sum_{i=1}^{N-m-k} \left(u_i^2 - \frac{1}{l+1}\right) \exp\left\{\frac{l}{2} u_i^2\right\} W(\bar{x}_i / \bar{x}_{N-m}) = 0, \quad (14)$$

где  $W(\bar{x}_i / \bar{x}_{N-m})$  – определено в (2.10) и  $u_i = [x_{i+m+k-1} - \hat{X}_{N+k}] \cdot \hat{S}^{-1}$ .

2.  $p(t, \theta / \bar{X}_{N-m}) = \frac{1}{2S} \exp\left\{-\frac{|t - \hat{X}_{N+k}|}{S}\right\}$  – распределение Лапласа. Из (9), (12)

получаем систему уравнений для оценки прогноза:

$$\sum_{i=1}^{N-m-k} \text{sign}(u_i) \exp\{l|u_i|\} W(\bar{x}_i / \bar{x}_{N-m}) = 0; \quad (15)$$

$$\sum_{i=1}^{N-m-k} \left(|u_i| - \frac{1}{l+1}\right) \exp\{l|u_i|\} W(\bar{x}_i / \bar{x}_{N-m}) = 0. \quad (16)$$

Можно получить аналогичные уравнения для оценки прогноза и для других распределений [9]. Сделаем некоторые выводы.

Уравнения (13), (15) определяют:

1) при  $l = 0$  – оценки условного максимального правдоподобия (ОМП), которые являются классическими непараметрическими оценками прогноза регрессионного и медианного типов [2, 3, 5, 6, 7];

2) при  $l = 1$  – оценки условной максимальной устойчивости прогноза (ОМУ [12]) модального типа [2, 7];

3) при  $l = 0,5$  – условные радикальные оценки прогноза [12].

Оценки прогноза ОМП типа (13), (15) не являются робастными по критерию устойчивости (в том числе и оценка медианного типа, которая не справляется с асимметричными выбросами) [12]. Оценки прогноза модального типа (13), (15) (ОМУ) являются робастными и справляются как с симметричными, так и с асим-

метричными выбросами, но являются низкоэффективными [12]. В связи с этим в [12] и был предложен промежуточный вариант радикальных оценок, у которых эффективность и устойчивость совпадают (они относятся к классу MD-оценок на расстоянии Хеллингера). В ВММП радикальные оценки получаются при  $l=0,5$ . Алгоритмы вида (13) – (16) были получены в предположении, что вид распределения  $P(x, \theta / \bar{X}_{N-m})$  известен. Следовательно, их можно отнести к частично непараметрическим алгоритмам прогноза. Они являются непараметрическими только по типу априорной информации о виде функции  $m(\bar{X}_{i-m-1})$  в модели (1), но не по типу априорной информации о виде функции распределения  $F(\bar{x}, t) \in \Omega$ .

Рассмотрим случай, когда вид распределения  $P(x, \theta / \bar{X}_{N-m})$  неизвестен, и задача относится к классу непараметрических задач.

Пусть  $P(x, \theta / \bar{X}_{N-m})$  относится к непараметрическому классу абсолютно непрерывных унимодальных распределений, симметричных относительно  $X_{N+k}$ . В этом случае оценка Розенблатта – Парзена для плотности  $p(x, \theta / \bar{X}_{N-m})$  запишется в виде

$$p_N(t, \theta / \bar{X}_{N-m}) = \frac{1}{S} \int K \left( \frac{X_{N+m} - 0,5(t+z)}{S} \right) dF_N(z / \bar{X}_{N-m}). \quad (17)$$

В соответствии с ВММП для нахождения оценочных функций подставим выражение (17) в (10). В результате получаем оценочные функции

$$\begin{aligned} \varphi_1(t, \theta / \bar{X}_{N-m}) &= \int K'_u(u) dF_N(z / \bar{X}_{N-m}) \cdot \left[ p_N(t, \hat{\theta} / \bar{X}_{N-m}) \right]^{l-1}, \\ \varphi_2(t, \theta / \bar{X}_{N-m}) &= \left\{ \int u K'_u(u) dF_N(z / \bar{X}_{N-m}) \cdot \left[ p_N(t, \hat{\theta} / \bar{X}_{N-m}) \right]^{-1} - \frac{1}{(l+1)} \right\} \times \\ &\quad \times p_N^l(t, \hat{\theta} / \bar{X}_{N-m}), \end{aligned} \quad (18)$$

где  $u = [\hat{X}_{N+m} - 0,5(t+z)] \cdot \hat{S}^{-1}$ .

Подставляя (18) в (9), получаем систему оценочных уравнений для нахождения непараметрических оценок прогноза  $\hat{X}_{N+k}$  и параметра масштаба  $\hat{S}$ :

$$\int \varphi_i(t, \hat{\theta} / \bar{X}_{N-m}) dF_N(t / \bar{X}_{N-m}) = 0, \quad i = 1, 2. \quad (19)$$

Пусть

$$u_{iv} = [\hat{X}_{N+m} - z_{iv}] \cdot \hat{S}^{-1}, \quad z_{iv} = 0,5 \cdot [x_{i+m+k-1} + x_{v+m+k-1}].$$

Оценочные уравнения (19) примут следующий вид:

$$\begin{aligned} \sum_{i=1}^{N-m-k} \sum_{v=1}^{N-m-k} K'(u_{iv}) W(\bar{x}_v / \bar{x}_{N-m}) \cdot \left[ \sum_{v=1}^{N-m-k} K(u_{iv}) W(\bar{x}_v / \bar{x}_{N-m}) \right]^{l-1} W(\bar{x}_i / \bar{x}_{N-m}) &= 0, \\ \sum_{i=1}^{N-m-k} \left\{ \sum_{v=1}^{N-m-k} u_{iv} \cdot K'(u_{iv}) W(\bar{x}_v / \bar{x}_{N-m}) \cdot \left[ \sum_{v=1}^{N-m-k} K(u_{iv}) W(\bar{x}_v / \bar{x}_{N-m}) \right]^{-1} - \frac{1}{l+1} \right\} \times \\ \times \left[ \sum_{v=1}^{N-m-k} K(u_{iv}) W(\bar{x}_v / \bar{x}_{N-m}) \right]^l W(\bar{x}_i / \bar{x}_{N-m}) &= 0. \end{aligned} \quad (20)$$

Оценочные уравнения (20) выглядят достаточно сложно, особенно если подставить выражения для ядерных функций  $K(u_{iv})$  и  $W(\bar{x}_i / \bar{x}_{N-m})$ . Однако при  $l=0$  (ОМП) и применении «идеального окна»  $K(u)$  получаем хорошо известные в непараметрической статистике алгоритмы оценки сдвига и масштаба на полусуммах Уолша.

Рассмотрим влияние одиночных выбросов на работу непараметрического алгоритма (20). Пусть выбросом является одно из наблюдений  $x_1, \dots, x_{N-m}$ . Если выброс входит в блок  $\bar{x}_i$ , то фильтр  $W(\bar{x}_i / \bar{x}_{N-m})$  приписывает этому блоку малый вес или просто не пропускает данный блок. Непараметрический алгоритм ведет себя устойчиво. Если выброс пришелся на значение  $y_i$ , то алгоритмы типа «радикального» или ОМУ решают задачу робастного оценивания прогноза, в том числе и для асимметричных распределений выбросов [8, 12]. Если выбросы отсутствуют, то при  $l=0$  алгоритм (20) сходится к оптимальному алгоритму. Таким образом, ВММП позволяет, применяя адаптивную настройку параметра радикальности  $0 \leq l \leq 1$ , получать эффективные непараметрические оценки прогноза в неопределенной ситуации с выбросами [8].

Пусть выброс присутствует в векторе  $\bar{x}_{N-m}$ . В этом случае фильтр  $W(\bar{x}_i / \bar{x}_{N-m})$  не найдет во временном ряду  $x_1, \dots, x_{N-m}$  ни одного вектора  $\bar{x}_i$ , однородного с вектором  $\bar{x}_{N-m}$ . На выходе непараметрического алгоритма прогноза получим  $\hat{X}_{N+k} = 0$ . Данная информация важна, например, в задачах определения момента перескока процесса на новый уровень. Она сигнализирует о том, что в векторе  $\bar{x}_{N-m}$  присутствует выброс. Обычные методы борьбы с выбросами, связанные с исключением данного вектора или снижением его роли с помощью специальных весовых функций, в этом случае не пригодны. В дальнейшем рассмотрим два алгоритма, позволяющие учесть данную ситуацию.

Теоретическое исследование алгоритмов вида (20) выходит за рамки данной работы и связано с исследованием непараметрических оценок параметров, заданных в неявном виде через обобщенные условные функционалы (обобщенные условные М-оценки)

$$\int \varphi(F(x, \theta/t), F'(x, \theta/t), \dots) dF(x, \theta/t) = 0.$$

### 3. Адаптивные алгоритмы

Эффективность работы алгоритмов вида (20) определяется набором параметров  $(m, h_N)$  для настройки фильтра  $W(\bar{x}_i / \bar{x}_{N-m})$  и параметром радикальности  $l$  для настройки робастности алгоритма. Для оптимизации по этим трем параметрам возможны два подхода. Первый основан на теоретических исследованиях вариации оценки прогноза и ее оптимизации по данным параметрам. Второй подход, который в основном и применяется, основан на минимизации наблюдаемой суммы квадратов невязок  $Q(m, h_N, l) = (N-m)^{-1} \sum (x_{i+m} - \hat{X}_{N+k})^2$  в режиме «скользущего экзамена».

Уже предварительные результаты показывают, что задача минимизации  $Q(m, h_N, l)$  хорошо сегментируется, хотя все три параметра связаны. В режиме

«скользящего экзамена» сначала настраиваются параметры  $(m, h_N)$  фильтра  $W(\bar{x}_i / \bar{x}_{N-m})$ , которые слабо зависят от выбросов внутри временного ряда  $x_1, \dots, x_{N-m}$ , а затем параметр радикальности  $l$ .

Как было отмечено в пункте 2.2, непараметрические алгоритмы прогноза не спасают от выбросов в векторе  $\bar{x}_{N-m}$ . Для отбраковки выбросов в  $\bar{x}_{N-m}$  можно предложить два метода: прореживания и прореживания с восстановлением. Предположим,  $k=1$  и выброс приходится на значение  $x_N$ .

**Прореживание.** Исключаем наблюдение  $x_N$  из вектора  $\bar{x}_{N-m}$  и будем проводить прогноз по вектору  $\bar{x}_{N-m-1} = (x_{N-1}, \dots, x_{N-m-1})^T$  на два шага вперед. Если выброс находится на  $j$ -м месте вектора  $\bar{x}_{N-m}$ , то, последовательно удаляя и возвращая в  $\bar{x}_{N-m}$  по одному наблюдению, обнаружим выброс и найдем оценку прогноза. Во временных рядах, в связи с сильной зависимостью между соседними наблюдениями, могут возникать пачки выбросов. В этом случае исключаем пары, тройки, ... соседних наблюдений в векторе  $\bar{x}_{N-m}$ .

**Прореживание с восстановлением.** Сначала с помощью прореживания выявляется выброс, например  $x_N$ . Затем наблюдение  $x_N$  с помощью непараметрического алгоритма прогноза восстанавливается  $\hat{x}_N$  и прогноз ведется по восстановленному вектору  $\bar{x}_{N-m} = (\hat{x}_N, x_{N-1}, \dots, x_{N-m})^T$ .

#### 4. Моделирование

Проведены исследования алгоритмов (13), (14) и (19) с нормальными ядрами в предположении, что выбросов в  $\bar{x}_{N-m}$  нет. В первую очередь интересовало поведение  $Q(m, h_N, l)$  в зависимости от параметра  $m$  (размер кластера) и тесно связанного с ним параметра  $h_N$  (ширина фильтра). Затем, как и в работе [8], производилась адаптация по параметру радикальности  $l$ . Модель процесса  $X_1(t) = \sin(\omega \cdot t + \varphi) + 0,15\varepsilon$ ,  $\varepsilon$  – белый шум,  $\varphi$  – равномерна на  $[0, 2\pi]$  и модель биений  $X_2(t) = \sin(\omega_1 \cdot t + \varphi) \sin(\omega_2 \cdot t) + \varepsilon$ .

На рис. 1 – 3 представлена зависимость  $Q(m, h_N, l)$  от  $m$  для  $X_1(t)$ : рис. 1 – без шума; рис. 2 – присутствует шум  $\varepsilon$ ; рис. 3 – присутствует шум  $\varepsilon$  и единичный выброс.

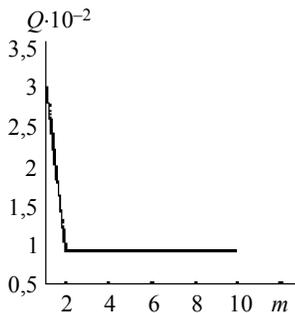


Рис. 1

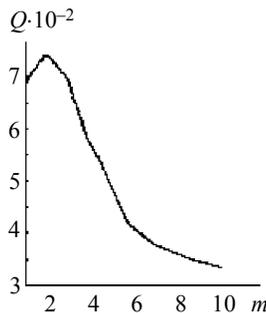


Рис. 2

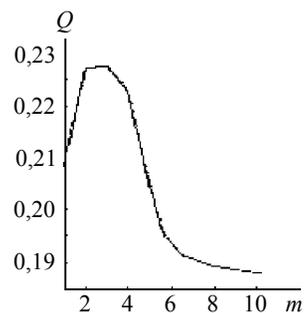


Рис. 3

Как видно из рис. 1, уже при  $m = 2$  имеет место сингулярный прогноз. Наличие белого шума (рис. 2) приводит к регулярному прогнозу и  $m$  увеличивается до 8. Наличие белого шума и выброса (рис. 3) приводит к регулярному прогнозу, причем  $m$  не увеличивается, а адаптация по  $l$  приводит к ликвидации влияния выброса. Зависимость  $Q(m, h_N, l)$  от  $l$  качественно имеет такой же вид, как и в [8], поэтому не приводится. Аналогичная картина наблюдалась и для процесса биений  $X_2(t)$ .

### Заключение

В работе на основе ВММП синтезирован ряд новых робастных непараметрических алгоритмов прогноза (алгоритмы (10), (13), (15), (20)). Данные алгоритмы включают как частный случай непараметрические алгоритмы [2, 3, 5 – 7].

Показано, что введенные ранее непараметрические алгоритмы прогноза [2, 3, 5 – 7] не являются полностью робастными и непараметрическими.

Предложены адаптивные методы настройки по параметрам  $m, h_N, l$ , которые оптимизируют работу алгоритмов как по исходному распределению, так и по распределению выбросов.

Проведено моделирование, которое показало работоспособность и эффективность алгоритмов.

В теоретическом плане алгоритмы вида (20) ставят вопрос исследования непараметрических оценок параметров, заданных в неявном виде через обобщенные условные функционалы (обобщенные условные М-оценки).

### ЛИТЕРАТУРА

1. Ширяев А.Н. Основы стохастической финансовой математики. Т. 1. М.: Фазис, 1998. 488 с.
2. Siegfried Heiler. A Survey on Nonparametric Time Series Analysis. 1999. 49 p. URL: <http://www.ub.uni-konstanz.de/kops/volltexte/1999/316/>
3. Neumann M.H., Kreiss J.P. Regression-Type inference in nonparametric autoregression // Ann. Statist. 1998. V. 26. No. 4. P. 1570–1613.
4. Мартин Р.Д. Устойчивый авторегрессионный анализ временных рядов // Устойчивые статистические методы оценки данных. М.: Машиностроение. С. 121–146.
5. Васильев В.А., Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. М.: Наука, 2004. 510 с.
6. Симахин В.А. Непараметрическое прогнозирование случайных процессов // Тез. докл. I областной научно-практической конференции по надежности научно-технических прогнозов. Новосибирск, 1978.
7. Рымар Т.Н., Симахин В.А. Непараметрическое прогнозирование стационарных случайных процессов // Тез. докл. зональной научно-технической конференции «Датчики и средства первичной обработки информации». Курган, 1990. С. 102–104.
8. Simahin V.A. Nonparametric robust regression estimate // Proc. SPIE. 2006. V. 6522. P. 130–139.
9. Симахин В.А. Непараметрическая статистика. Ч. II. Теория оценок. Курган: Изд-во КГУ, 2004. 163 с.
10. Кошкин Г.М. Об одном подходе к оцениванию переходной функции распределения и моментов для некоторых марковских процессов // Математическая статистика и ее приложения. Томск, 1976. С. 53–65.
11. Roussas G.G. Nonparametric estimation of the transition distribution function of a Markov process // Ann. Math. Statist. 1969. V. 40. No. 4. P. 1386–1400.

12. *Шурыгин А.М.* Прикладная статистика. Робастность. Оценивание. Прогноз. М.: Финансы и статистика, 2000. 223 с.
13. *Симахин В.А.* Непараметрическая статистика. Ч. I. Теория оценок. Курган: Изд-во КГУ, 2004. 216 с.
14. *Симахин В.А.* Взвешенный метод максимального правдоподобия // Материалы IX Международной науч.-технич. конф. «Кибернетика и высокие технологии XXI века». Воронеж, 2008. Т. 2. С. 661–672.

*Симахин Валерий Ананьевич*

Курганский государственный университет (г. Курган).

E-mail: sva\_full@mail.ru

Поступила в редакцию 16 сентября 2010 г.