

ОБРАБОТКА ИНФОРМАЦИИ

УДК 621.391.1:519.25

Г.М. Кошкин, В.С. Лаходынов

ПОЛУРЕКУРРЕНТНАЯ НЕПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ УСЛОВНЫХ ФУНКЦИОНАЛОВ СЛАБОЗАВИСИМЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Исследуется сходимость в среднеквадратическом полурекуррентной оценки условного функционала, построенной по слабозависимым последовательностям сильного перемешивания. Полурекуррентные оценки используются при идентификации нелинейной авторегрессии произвольного порядка. Решается задача прогноза стоимости пая инвестиционного фонда по реальным данным с использованием полурекуррентных алгоритмов.

Ключевые слова: *полурекуррентные алгоритмы, условные функционалы, непараметрическая идентификация.*

При решении многих задач идентификации экономических объектов, как правило, используются условные математические ожидания или функции регрессии. Например, в качестве модели производственной функции фирмы, т.е. зависимости объема выпуска Y от объема используемых для производства ресурсов X , можно взять условное математическое ожидание объема выпуска при условии, что $X = x$. В данной работе функция регрессии применяется при идентификации нелинейных процессов авторегрессии, что позволяет строить прогнозы цен акций, облигаций, опционов и т.п.

Для идентификации функции регрессии воспользуемся непараметрическими полурекуррентными оценками подстановки, так как реальные объекты обычно имеют сложную структуру, и априорной информации часто недостаточно для построения параметрических оценок. Проводится статистическое моделирование нелинейной авторегрессии 1-го порядка с аддитивными шумами, имеющими различную дисперсию. Результаты моделирования подтверждают возможность использования полурекуррентных непараметрических оценок для идентификации нелинейных процессов авторегрессии.

Проверка работоспособности предложенных алгоритмов проводится на данных изменения цены пая открытого паевого инвестиционного фонда «Альфа-Капитал Акции».

1. Постановка задачи

Рассмотрим стохастический объект, на l входах и выходе которого наблюдаются соответственно случайные величины $X = (X_1, \dots, X_l)$ и Y , $X \in R^l$, $Y \in R^1$, где R^l – l -мерное евклидово пространство.

Значение выходной переменной Y определяется неизвестным функциональным преобразованием $F(X)$ входных переменных X , т.е. $Y=F(X)$. В качестве приближения неизвестной функции F возьмем функцию регрессии. Функция регрессии, или условное математическое ожидание выхода стохастического объекта относительно входов, является моделью, минимизирующей среднеквадратическое отклонение (СКО) истинного выхода объекта и модели:

$$r(x) = \int_{R^1} yf(y|x)dy = \frac{\int_{R^1} yf(x,y)dy}{p(x)} = \frac{a(x)}{p(x)},$$

где $f(x,y)$ – неизвестная плотность распределения наблюдаемой $(l+1)$ -мерной случайной величины $Z=(X,Y) \in R^{l+1}$, $f(y|x) = \frac{f(x,y)}{p(x)}$ – условная плотность распределения случайной величины $Y \in R^1$ при условии, что $X=x$, $p(x)$ – маргинальная плотность распределения величины X , $a(x) = \int_{R^1} yf(x,y)dy$.

При построении модели учитываются не все входы реального объекта, поэтому возникает погрешность, связанная с упрощением регрессионной модели относительно истинной структуры объекта. Эту погрешность можно измерять остаточной (условной) дисперсией:

$$D(Y|x) = \frac{\int_{R^1} y^2 f(x,y)dy}{p(x)} - r^2(x).$$

Для характеристики уровня сложности объекта применяются такие его статистические характеристики, как условные коэффициенты асимметрии и эксцесса. В этом случае $g(y) = y^k$, $k = 1, 2, 3, 4$.

Удобно представить указанные характеристики стохастических объектов в виде функций от условных функционалов

$$J(x) = \int_{R^1} g(y)f(y|x)dy = \frac{\int_{R^1} g(y)f(x,y)dy}{p(x)} = \frac{G(x)}{p(x)}, \quad (1.1)$$

где $g(\cdot)$ – известная измеримая по Борелю скалярная функция. Например, для остаточной дисперсии $D(Y|x) = J_1(x) - J_2^2(x)$, где $g_1(y) = y^2$, $g_2(y) = y$.

Возьмем в качестве непараметрической полурекуррентной ядерной оценки подстановки условного функционала (1.1) в точке x статистику вида

$$J_{[n]}(x) = \frac{G_{[n]}(x)}{p_{[n]}(x)}, \quad (1.2)$$

$$\text{где} \quad G_{[n]}(x) = G_{[n-1]}(x) - \frac{1}{n} \left[G_{[n-1]}(x) - \frac{g(Y_n)}{\prod_{k=1}^l h_{[n]k}} K\left(\frac{x - X_n}{h_{[n]}}\right) \right] \quad (1.3)$$

$$\text{и} \quad p_{[n]}(x) = p_{[n-1]}(x) - \frac{1}{n} \left[p_{[n-1]}(x) - \frac{1}{\prod_{k=1}^l h_{[n]k}} K\left(\frac{x - X_n}{h_{[n]}}\right) \right]$$

– соответственно рекуррентные оценки функционала $G(x)$ и плотности $p(x)$; $Z_i = (X_i, Y_i)$, $X_i = (X_{i,1}, \dots, X_{i,l})$, $i = \overline{1, n}$, – $(l+1)$ -мерная выборка, характеризующая плотностью $f(z) = f(x, y)$; $K(u)$ – ядерная функция; $h_{[n]k} > 0$ – последовательность чисел, сходящаяся к нулю для каждого $k = \overline{1, l}$.

Данные оценки являются полурекуррентными [1], так как и числитель, и знаменатель в формуле (1.2) определяются рекуррентными соотношениями, но при этом сама оценка (1.2) не является рекуррентной.

Отметим, что последовательные процедуры обладают рядом преимуществ перед обычными: они, как правило, легко реализуются на компьютерах, экономя при этом машинную память, на каждом такте работы алгоритма дают готовый результат, и поступление новых измерений не приводит к громоздким перерасчетам, что позволяет обрабатывать информацию в режиме реального времени.

2. Ядерные рекуррентные оценки базовых функционалов по независимой выборке

Введем следующие обозначения:

$$a_r(x) = \int_{R^1} g^r(y) f(x, y) dy, \quad a_r^+(x) = \int_{R^1} |g^r(y)| f(x, y) dy,$$

где $r = 0, 1$. В качестве оценки плотности $p(x) = a_0(x)$ и базового функционала $a_1(x)$ будем использовать статистики

$$a_{[n]r}(x) = a_{[n-1]r}(x) - \frac{1}{n} \left[a_{[n-1]r}(x) - \frac{1}{\prod_{k=1}^l h_{[n]k}} g^r(Y_n) K\left(\frac{x - X_n}{h_{[n]}}\right) \right], \quad (2.1)$$

где $K\left(\frac{x - X_i}{h_{[n]}}\right) = K\left(\frac{x_1 - X_{i,1}}{h_{[i]1}}, \dots, \frac{x_l - X_{i,l}}{h_{[i]l}}\right)$, $r = 0, 1$.

Преобразуем выражение (2.1) к удобному для исследований виду. Для этого умножим обе его части на n , заменим n на i и просуммируем по i от 1 до n . В результате получим

$$a_{[n]r}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\prod_{k=1}^l h_{[i]k}} g^r(Y_i) K\left(\frac{x - X_i}{h_{[i]}}\right). \quad (2.2)$$

Сформулируем результаты об асимптотической несмещенности оценки (2.2) (лемма 1), скорости сходимости ее смещения (лемма 2), главных частей асимптотической дисперсии (следствие 1) и ковариации оценок $a_{[n]r}(x)$ и $a_{[n]s}(x)$ (лемма 3).

Лемма 1. Если функция $a_r(x)$ непрерывна, ядро удовлетворяет условиям $\int_{R^l} K(u)du = 1$, $K(u) \geq 0$, $\sup_{x \in R^l} a_r^+(x) < \infty$, последовательность вещественных чисел $h_{[n]k} \downarrow 0$ для каждого $k = \overline{1, l}$, то

$$\lim_{n \rightarrow \infty} Ea_{[n]r}(x) - a_r(x) = 0. \quad (2.3)$$

Предположим, что последовательность чисел $h_{[n]}$ удовлетворяет следующим условиям:

$$\frac{1}{n} \sum_{i=1}^n h_{[i]k}^q = S_{qk} h_{[n]k}^q + o(h_{[n]k}^q); \quad (2.4)$$

$$\prod_{k=1}^l h_{[i]k}^q = C_{qs} h_{[i]s}^{ql} + o(h_{[i]s}^{ql}), \quad s = \overline{1, l}, \quad (2.5)$$

где q – вещественное число; S_{qk} и C_{qs} – некоторые постоянные, не зависящие от n .

Лемма 2. Если функция $a_r(x)$ дважды непрерывно дифференцируема на R^l , $\int_{R^l} K(u)du = 1$, $\int_{R^l} u_k^2 K(u)du = 1$, $k = 1, 2, \dots, l$, $K(u) = K(-u)$, $\sup_{x \in R^l} \left| \sum_{k=1}^l \frac{\partial^2 a_r(x)}{\partial x_k^2} \right| < \infty$, то при $n \rightarrow \infty$

$$Ea_{[n]r}(x) = a_r(x) + \sum_{k=1}^l \frac{S_{2k}}{2} \frac{\partial^2 a_r(x)}{\partial x_k^2} h_{[n]k}^2 + o\left(\sum_{j=1}^l h_{[n]j}^2\right). \quad (2.6)$$

Лемма 3. Если $\int_{R^l} K(u)du = 1$, $\int_{R^l} u_k^2 K(u)du = 1$, $k = 1, 2, \dots, l$, $K(u) = K(-u)$, $a_r(x)$ непрерывна на R^l , $\sup_{x \in R^l} |a_r(x)| < \infty$, $\int_{R^l} K^2(u)du < \infty$, то при $n \rightarrow \infty$

$$\begin{aligned} & \text{cov}(a_{[n]r}(x), a_{[n]s}(x)) = \\ & = \frac{a_{r+s}(x) S_{-1j}}{C_{1j} n h_{[n]j}^l} \int_{R^l} K^2(u)du + o\left(\frac{1}{n h_{[n]j}^l}\right), \quad j = \overline{1, l}. \end{aligned} \quad (2.7)$$

Следствие 1. Если выполнены условия леммы 3, то при $r = s$ и $n \rightarrow \infty$ справедливо соотношение

$$Da_{[n]r} \approx \frac{a_{2r}(x) S_{-1j}}{C_{1j} n h_{[n]j}^l} \int_{R^l} K^2(u)du, \quad j = \overline{1, l}.$$

Обозначим среднеквадратическую ошибку (СКО) оценки $a_{[n]r}(x)$ через

$$u^2(a_{[n]r}(x)) = E(a_{[n]r}(x) - a_r(x))^2.$$

Теорема 1. Если выполнены условия лемм 2 и 3, то при $n \rightarrow \infty$ асимптотически оптимальная СКО оценки $a_{[n]r}(x)$ задается формулой

$$u_*^2(a_{[n]r}(x)) \approx A \left(a_{2r}(x) \sqrt{\prod_{s=1}^l \frac{\partial^2 a_r(x)}{\partial x_s^2}} \int_{R^l} K^2(u) du \right)^{\frac{4}{4+l}} n^{-\frac{4}{4+l}}, \quad (2.8)$$

где

$$A = (4+2l)^{\frac{4+2l}{4+l}} l^{\frac{1}{4+l}} (4+l)^{\frac{4+3l}{4+l}} + \frac{l^{\frac{4+2l}{4+l}} (4+l)^{\frac{2l}{4+l}}}{(4+2l)^{\frac{4+2l}{4+l}}},$$

а для оптимальных параметров размытости выполняются соотношения

$$h_{[i]k} = \sqrt{\frac{\frac{\partial^2 a_r(x)}{\partial x_i^2}}{\frac{\partial^2 a_r(x)}{\partial x_k^2}}} h_{[i]l}, \quad i = \overline{1, n}, \quad k = \overline{1, l}; \quad (2.9)$$

$$h_{[n]k}^o = \left(\frac{2(2+l)a_{2r}(x) \sqrt{\prod_{s=1}^l \frac{\partial^2 a_r(x)}{\partial x_s^2}}}{l(4+l)^2} \int_{R^l} K^2(u) du \right)^{\frac{1}{4+l}} \frac{1}{\sqrt{\frac{\partial^2 a_r(x)}{\partial x_k^2}}} n^{-\frac{1}{4+l}},$$

для всех $k = \overline{1, l}$. (2.10)

Доказательства лемм 1 – 3 и теоремы 1 раздела 2 не приводятся, так как ключевые моменты и идеология этих доказательств имеются в [3].

3. Ядерные рекуррентные оценки базовых функционалов последовательностей сильного перемешивания

Пусть $X_t \in R^l, t=0, 1, 2, \dots$, – последовательность случайных величин, определенных на вероятностном пространстве (Ω, F^x, P) , и пусть $F_{s,t}^x = \sigma\{X_u, s \leq u \leq t\}$ – σ -алгебра событий, генерируемых случайными величинами $X_u, s \leq u \leq t$. Говорят, что последовательность случайных величин $X_1, \dots, X_n$ удовлетворяет условию сильного перемешивания с коэффициентом сильного перемешивания $\alpha(\tau)$, если

$$\alpha(\tau) = \sup_t \sup_{A \in F_{0,t}; B \in F_{t+\tau, \infty}} |P(AB) - P(A)P(B)| \downarrow 0, \quad \tau > 0.$$

Для краткости в этом случае будем использовать обозначение $(X_i)_{i \geq 1} \in S(\alpha)$.

Результаты лемм 1 и 2 остаются справедливыми и в случае, когда выборочные наблюдения $Z_i, i = \overline{1, n}$, зависимы между собой и, в частности, если они удовлетворяют условию сильного перемешивания.

В данном разделе показано, что для ковариации, дисперсии и СКО оценки представление главной части при $n \rightarrow \infty$ совпадает с соответствующим выражением, когда выборочные наблюдения $Z_1, \dots, Z_n$ являются независимыми.

Лемма 4. Пусть $(Z_i)_{i \geq 1} \in S(\alpha)$, причем $\int_0^\infty [\alpha(m)]^q dm < \infty$, $0 < q < \frac{1}{2}$, $\int_{R^l} K(u) du = 1$, $\int_{R^l} u_k^2 K(u) du = 1$, $k = 1, 2, \dots, l$, $\int_{R^l} K^2(u) du < \infty$, $K(u) = K(-u)$, $a_r(x)$ непрерывна на R^l , $\sup_{x \in R^l} |a_r(x)| < \infty$, $\sup_{x \in R^l} a_{(2+\delta)r}^+(x) < \infty$, $\delta > 0$, $\sup_{x \in R^l} a_{1m,rs}^+(x, x) < \infty$ для любых $m > 1$. Тогда при $n \rightarrow \infty$

$$\begin{aligned} \text{cov}(a_{[n]r}(x), a_{[n]s}(x)) = \\ = \frac{a_{r+s}(x) S_{-1j}}{C_{1j} n h_{[n]j}^l} \int_{R^l} K^2(u) du + o \left(\frac{1}{n h_{[n]j}^{\frac{2(1+\delta)}{2+\delta}} l} \right), \quad j = \overline{1, l}. \end{aligned} \quad (3.1)$$

Аналогичные результаты также справедливы для дисперсии и СКО оценки (см. следствие 1 и теорему 1).

4. СКО полурекуррентной оценки условного функционала

Введем следующие обозначения:

$$\begin{aligned} f_{1i(i+j)(i+j+k)}(z, s, u, w) &= \frac{\partial^{4l+4}}{\partial z \partial s \partial u \partial w} P(Z_1 \leq z, Z_i < s, Z_{i+j} \leq u, Z_{i+j+k} \leq w) \\ - 4l + 4 \text{-мерная плотность распределения случайных величин} \\ (Z_1, Z_i, Z_{i+j}, Z_{i+j+k}), \quad a_{1i(i+j)(i+j+k),r}^+(x, y, x', y') &= \int_{R^l} \int_{R^l} |g^r(v) g^r(s) g^r(v') g^r(s') \times \\ &\times g^r(s')|^{2+\delta} f_{1i(i+j)(i+j+k)}(x, v, y, s, x', v', y', s') dv ds dv' ds', \\ a_{1(1+j)(1+j+k), (2+\delta)r}^+(x, y, x') &= \int_{R^l} |g^r(v) g^r(s) g^r(v')|^{2+\delta} \times \\ &\times f_{1(1+j)(1+j+k)}(x, v, y, s, x', v') dv ds dv'. \end{aligned}$$

Выражение для СКО оценки $J_{[n]}(x)$ определяется теоремой 3.

Теорема 2. Пусть выполнены условия:

$$\begin{aligned} 1) (Z_i)_{i \geq 1} \in S(\alpha), \quad \text{причем} \quad \int_0^\infty [\alpha(m)]^q dm < \infty, \quad 0 < q < \frac{1}{2}, \quad \int_{R^l} K(u) du = 1, \\ \int_{R^l} u_k^2 K(u) du = 1, \quad k = 1, 2, \dots, l, \quad \int_{R^l} K^2(u) du < \infty, \quad K(u) = K(-u), \quad a_r(x) \text{ непрерыв-} \\ \text{на на } R^l, \quad \sup_{x \in R^l} |a_r(x)| < \infty, \quad \sup_{x \in R^l} a_{1i(i+j)(i+j+k),r}^+(x, x, x, x) < \infty, \quad i, j, k \geq 1, \quad i+j+k \leq n, \\ \sup_{x \in R^l} a_{1(1+j)(1+j+k), (2+\delta)r}^+(x, x, x) < \infty, \quad j, k \geq 1, \quad 1+j+k \leq n, \quad \sup_{x \in R^l} a_{1i(i+j), (2+\delta)r}^+(x, x, x) < \infty, \\ i \geq 2, \quad j \geq 1, \quad i+j \leq n, \quad \sup_{x \in R^l} a_{1i, (2+\delta)r}^+(x, x) < \infty, \quad 2 \leq i \leq n, \quad \sup_{x \in R^l} a_{(2+\delta)r}^+(x) < \infty; \end{aligned}$$

2) последовательность $\left(\left| \frac{a_{[n]l}(x)}{a_{[n]0}(x)} \right| \right)$ мажорируется числовой последовательностью $(Cd_{[n]}^\gamma)$, где $d_{[n]} \rightarrow \infty$ при $n \rightarrow \infty$, $0 \leq \gamma < \infty$, C – некоторая постоянная.

Тогда для СКО оценки $J_{[n]}(x)$ имеем

$$\begin{aligned} u^2 \left(\frac{a_{[n]l}(x)}{a_{[n]0}(x)} \right) &= \frac{S_{-1j}}{C_{1j} n h_{[n]j}^l} \left[\frac{a_2(x)}{a_0^2(x)} - \frac{a_1^2(x)}{a_0^3(x)} \right] \int_{R^l} K^2(u) du + \\ &+ \frac{a_1^2(x)}{4a_0^4} \left[\sum_{k=1}^l S_{2k} \frac{\partial^2 a_0(x)}{\partial x_k^2} h_{[n]k}^2 \right]^2 + \frac{1}{4a_0^2} \times \\ &\times \left[\sum_{k=1}^l S_{2k} \frac{\partial^2 a_1(x)}{\partial x_k^2} h_{[n]k}^2 \right]^2 - \frac{a_1(x)}{2a_0^3} \left[\sum_{k=1}^l S_{2k} \frac{\partial^2 a_0(x)}{\partial x_k^2} h_{[n]k}^2 \right] \times \\ &\times \left[\sum_{k=1}^l S_{2k} \frac{\partial^2 a_1(x)}{\partial x_k^2} h_{[n]k}^2 \right] + o \left(\frac{1}{n h_{[n]j}^l} + \left(\sum_{j=1}^l h_{[n]j}^2 \right)^2 \right), \quad j = \overline{1, l}. \end{aligned} \quad (4.1)$$

Доказательства леммы 4 и теоремы 2 проводятся аналогично доказательствам для ковариации и СКО из [2].

5. Непараметрическая идентификация нелинейной авторегрессии порядка p

Пусть последовательность $\{X_t, t = \dots -1, 0, 1, \dots\}$ генерируется нелинейной авторегрессией порядка p :

$$X_t = \Psi(X_{t-1}, \dots, X_{t-p}, a) + \varepsilon_t, \quad (5.1)$$

где $\{\varepsilon_t\}$ – последовательность независимых и одинаково распределенных случайных величин с нулевым математическим ожиданием и ограниченной дисперсией, Ψ – неизвестная функция, такая, что $\forall x, y \in R^p$,

$$|\Psi(y+x) - \Psi(y)| \leq a_1 |x_1| + \dots + a_p |x_p|, \quad 0 < a_1 + \dots + a_p < 1, \quad \Psi(0) = 0. \quad (5.2)$$

Условие (5.2), введенное в [4], обеспечивает стационарность процесса (5.1). В [4] также показано, что если φ – плотность распределения случайной величины $\{\varepsilon_t\}$, которая удовлетворяет условию Лифшица, т.е.

$$\forall \Delta > 0 \text{ и } C < \infty \quad \int |\varphi(x+\Delta) - \varphi(x)| dx \leq C\Delta,$$

то процесс нелинейной авторегрессии X_t будет процессом сильного перемешивания с экспоненциально убывающим коэффициентом сильного перемешивания

$$\alpha(\tau) \leq C(a) e^{-\delta(a)\tau}, \quad \tau > 0 \quad C(a) < \infty, \quad \delta(a) > 0,$$

где $a = (a_1, \dots, a_p)$, числа a_i определены в (5.2).

Пусть $X_{1-p}, \dots, X_n$ – выборка, генерируемая процессом (5.1). Теперь функцию Ψ в (5.1) можно оценить статистикой

$$\Psi_{[n]p}(x) = \frac{\sum_{i=1}^n \frac{1}{\prod_{k=1}^p h_{[i]k}} X_i K\left(\frac{x_1 - X_{i-1}}{h_{[i]1}}, \dots, \frac{x_p - X_{i-p}}{h_{[i]p}}\right)}{\sum_{i=1}^n \frac{1}{\prod_{k=1}^p h_{[i]k}} K\left(\frac{x_1 - X_{i-1}}{h_{[i]1}}, \dots, \frac{x_p - X_{i-p}}{h_{[i]p}}\right)}. \quad (5.3)$$

6. Нахождение оптимальных коэффициентов размытости

Результаты теоремы 1 имеют важное теоретическое значение, однако стоит заметить, что оптимальные $h_{[n]}^o$ и $u^2(a_{[n]r}(x))$ выражаются через неизвестные нам функции $a_{2r}(x)$, $a_r(x)$ и их производные $\frac{\partial^2 a_r(x)}{\partial x_k^2}$. Это позволяет считать задачу нахождения для конкретного набора $Z_1, \dots, Z_n$ оптимального значения $h_{[n]}^o(Z_1, \dots, Z_n)$ центральной проблемой при применении на практике оценок $a_{[n]r}(x)$.

Воспользуемся модифицированным методом кросс-проверки при нахождении $h_{[n]}^o(Z_1, \dots, Z_n)$ для оценки $J_{[n]}(x) = \frac{a_{[n]1}(x)}{a_{[n]0}(x)}$ функции $J(x)$ из минимума критерия

$$h_{[s]}^o(Z_1, \dots, Z_n) = \arg \min_{hs > 0} \sum_{i=1}^{s-1} [g(Y_i) - \frac{\sum_{j \neq i}^{s-1} \frac{g(Y_j)}{\prod_{k=1}^l h_{[j]k}} K\left(\frac{X_i - X_j}{h_{[j]1}}\right) + \frac{g(Y_s)}{\prod_{k=1}^l hs_k} K\left(\frac{X_i - X_s}{hs}\right)}{\sum_{j \neq i}^{s-1} \frac{1}{\prod_{k=1}^l h_{[j]k}} K\left(\frac{X_i - X_j}{h_{[j]1}}\right) + \frac{1}{\prod_{k=1}^l hs_k} K\left(\frac{X_i - X_s}{hs}\right)}], \quad s = \overline{2, n}. \quad (6.1)$$

Значения $h_{[1]}^o, h_{[2]}^o, \dots, h_{[n]}^o$ находятся с помощью следующей рекуррентной процедуры:

- 1) задается начальное условие: $h_{[1]} = h1 = \frac{1}{4} \left(\frac{|X_1 - X_2|}{2} \right)$;
- 2) $h_{[1]}^o, h_{[2]}^o, \dots, h_{[n]}^o$ вычисляются из критерия (6.1) при $s = 2, 3, \dots, n$;

В результате за n шагов определяется вся последовательность $h_{[1]}^o, h_{[2]}^o, \dots, h_{[n]}^o$.

Идеология нахождения параметра размытости для оценки $\Psi_{[n]}(x)$ процесса авторегрессии порядка p основана на конкретизации процедуры (6.2). Запишем оценку $\Psi_{[n]}(x)$ в виде

$$\Psi_{[n]p}(x) = \frac{\sum_{i=1}^n \frac{X_{i+1}}{\prod_{k=1}^p h_{[i-k+1]}} K\left(\frac{x_1 - X_i}{h_{[i]}} , \dots , \frac{x_p - X_{i-p+1}}{h_{[i-p+1]}}\right)}{\sum_{i=1}^n \frac{1}{\prod_{k=1}^p h_{[i-k+1]}} K\left(\frac{x_1 - X_i}{h_{[i]}} , \dots , \frac{x_p - X_{i-p+1}}{h_{[i-p+1]}}\right)}. \quad (6.3)$$

Оптимальное значение параметра $h_{[s]}^o$ на каждом шаге процедуры (6.2) получим из минимума критерия

$$\begin{aligned} & h_{[s]}^o(X_1, \dots, X_{s+1}) = \\ & = \arg \min_{hs>0} \sum_{i=1}^{s-1} \left[X_{i+1} - \left[\sum_{j \neq i}^{s-1} \frac{X_{j+1}}{\prod_{k=1}^p h_{[j-k+1]}} K\left(\frac{X_i - X_j}{h_{[j]}} , \dots , \frac{X_{i-p+1} - X_{j-p+1}}{h_{[j-p+1]}}\right) + \right. \right. \\ & \quad \left. \left. + \frac{X_{s+1}}{hs \prod_{k=2}^p h_{[s-k+1]}} K\left(\frac{X_i - X_s}{hs} , \dots , \frac{X_{i-p+1} - X_{s-p+1}}{h_{[s-p+1]}}\right) \right] / \right. \\ & \quad \left. / \left[\sum_{j \neq i}^{s-1} \frac{1}{\prod_{k=1}^p h_{[j-k+1]}} K\left(\frac{X_i - X_j}{h_{[j]}} , \dots , \frac{X_{i-p+1} - X_{j-p+1}}{h_{[j-p+1]}}\right) + \right. \right. \\ & \quad \left. \left. + \frac{1}{hs \prod_{k=2}^p h_{[s-k+1]}} K\left(\frac{X_i - X_s}{hs} , \dots , \frac{X_{i-p+1} - X_{s-p+1}}{h_{[s-p+1]}}\right) \right] \right]^2, s = \overline{2, n}. \end{aligned}$$

7. Моделирование нелинейной авторегрессии 1-го порядка

Пусть случайная последовательность $\{X_t, t=1, 2, \dots, 100\}$ генерируется нелинейной авторегрессией первого порядка

$$X_t = \Psi(X_{t-1}) + \varepsilon_t, \quad (7.1)$$

где $\{\varepsilon_t\}$ – последовательность независимых равномерно распределенных случайных величин с нулевым математическим ожиданием, функция $\Psi(t) = 5 + e^{-t/2}$. Заметим, что данная функция удовлетворяет условиям стационарности (5.2) процесса X_t .

В соответствии с (7.1) генерируется 6 серий для четырех различных последовательностей $\{\varepsilon_t\}$ с дисперсиями 0,0001, 0,1, 1, 10. Для каждой последовательности строятся рекуррентные оценки функции Ψ . В качестве ядра используется плотность распределения нормальной случайной величины с нулевым средним и

единичной дисперсией. Для каждой последовательности $\{\varepsilon_t\}$ вычисляются функции потерь:

$$R = \frac{1}{6} \sum_{i=1}^6 \left(\frac{1}{99} \sum_{t=2}^{100} \left| X_{i_t} - \hat{X}_{i_t} \right| \right),$$

где X_i – последовательность, сгенерированная в i -й серии, $\hat{X}_i$ – оценка, построенная по 100 испытаниям. Определим среднее значение последовательности $|X_i|$ формулой

$$X_{\text{ср}} = \frac{1}{6} \sum_{i=1}^6 \left(\frac{1}{100} \sum_{t=1}^{100} |X_{i_t}| \right),$$

а качество оценки интерполяции будем измерять коэффициентом

$$\gamma = \frac{R}{X_{\text{ср}}} 100\% .$$

Результаты статистического эксперимента отражены в табл. 1, из которой следует, что с увеличением дисперсии случайной составляющей $\{\varepsilon_t\}$ качество оценивания ухудшается.

Таблица 1

Ошибки интерполяции при различных значениях $D\{\varepsilon_t\}$

$D\{\varepsilon_t\}$	R	γ
0,0001	0,011	0,218
0,1	0,2	3,95
1	0,522	10,442
10	1,293	24,813

Далее решается задача прогноза. Ошибки прогнозирования будем вычислять по формуле

$$\Delta = \frac{1}{6} \sum_{i=1}^6 \left(\frac{1}{20} \sum_{t=81}^{100} \left| \frac{X_t - X_t^{\text{пр}}}{X_t} \right| \right) 100\% . \tag{7.1}$$

Таблица 2

Ошибки прогнозирования при различных значениях дисперсии $D\{\varepsilon_t\}$

$D\{\varepsilon_t\}$	0,0001	0,1	1	10
$\Delta, \%$	0,2	5,6	20,7	146

Из табл. 2 следует: чем меньше дисперсия помехи $D\{\varepsilon_t\}$, тем точнее прогноз. Прогнозирование направления изменения значений X_{i_t} , $t = \overline{80,100}$ показывает, что в случае, когда $D\{\varepsilon_t\}$ равна 0,0001, процент правильных прогнозов составил 85%, когда дисперсия равна 0,1 – 76,7%, когда единице – 73,3%, когда десяти – 66,7%. Таким образом, можно сделать вывод, что с увеличением дисперсии случайной составляющей количество верных прогнозов уменьшается.

8. Прогнозирование стоимости паев инвестиционного фонда «Альфа-Капитал Акции»

Полурекуррентные непараметрические алгоритмы типа (6.3) первых трех порядков применялись при обработке 100 наблюдений изменения цены пая открытого паевого инвестиционного фонда «Альфа-Капитал Акции» с 15 декабря 2006 г. по 17 мая 2007 г. В качестве ядра использовалась плотность распределения нормальной случайной величины с нулевым вектором средних и единичной дисперсионной матрицей. Была выбрана следующая стратегия инвестирования:

- 1) по первым 20 значениям строилась оценка авторегрессии по формуле (6.3);
- 2) строился прогноз на 21-е значение цены пая;
- 3) если по результатам прогноза цена пая на следующий день увеличивается, то паи покупаются, если цена уменьшается, то все паи погашаются;
- 4) после этого строилась оценка авторегрессии по 21-му наблюдению;
- 5) процедура повторялась до тех пор, пока не был построен прогноз на 17 мая.

Оказалось, что доходность за период с 18 января 2007 г. по 17 мая 2007 г. при выборе модели авторегрессии первого порядка составила 12,7%, второго порядка – 1,3%, третьего порядка – 6,4%. Цена пая за весь период увеличилась на 8,3, что на 50% меньше, чем доходность, полученная с использованием оценки авторегрессии 1-го порядка.

Можно считать, что в этом случае изменение цены пая более адекватно описывается процессом авторегрессии 1-го порядка, чем процессами авторегрессии 2-го и 3-го порядков.

Качество прогноза в зависимости от объема выборки представлено в табл. 3, в которой X_t – стоимость пая в момент времени t , X_t^{np} – прогнозное значение стоимости пая в момент времени t , построенное по $(t-1)$ -му значению,

$$\left| \frac{X_t - X_t^{\text{np}}}{X_t} \right| - \text{ошибка прогнозирования цены пая в момент времени } t.$$

Таблица 3

Сравнение качества прогноза в зависимости от объема выборки

T	X_t , руб.	X_t^{np} , руб.	$\left \frac{X_t - X_t^{\text{np}}}{X_t} \right $
21	3705,1	3631	0,02
30	3834,53	3796,31	0,01
40	3852,67	3839,77	0,003
50	3398,42	3631,84	0,069
60	3771,8	3740,17	0,008
70	3976,24	3919,44	0,014
80	4125,75	4127,9	0,001
90	3955,54	3972,77	0,004
100	3930,47	3919,52	0,003

Выбранная стратегия инвестирования предполагает, что на каждом этапе мы либо покупаем паи на всю имеющуюся сумму, либо погашаем все паи. Возможен вариант, когда мы тратим не все имеющиеся средства и погашаем не все паи. Зависимость доходности от доли используемых средств показана на рис. 1.

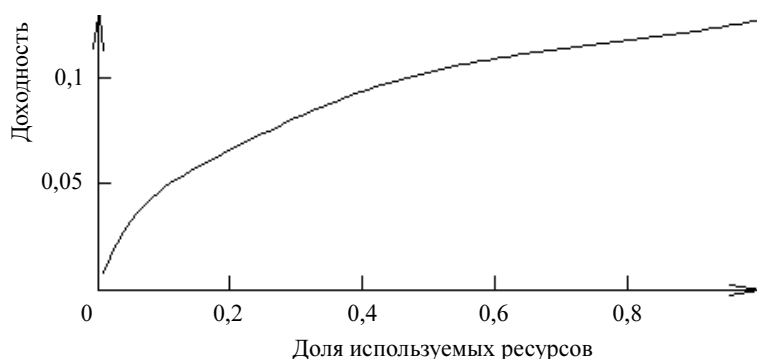


Рис. 1. Зависимость доходности от доли используемых средств

Монотонный характер зависимости доходности от доли используемых средств указывает на то, что максимальная доходность достигается в случае, когда мы располагаемся всеми имеющимися ресурсами.

Заключение

Выделим основные результаты работы:

- 1) Получено выражение для главной части СКО полурекуррентной оценки условного функционала, построенного по последовательностям сильного перемешивания. Установлено, что главные части совпадают с соответствующими главными частями СКО оценок, построенных по независимым наблюдениям.
- 2) Найдены теоретические значения оптимальных параметров размытости для каждой компоненты вектора наблюдений, которые минимизируют главную часть асимптотической СКО оценки; при этом оптимальные параметры размытости компонент выражаются друг через друга.
- 3) Построен рекуррентный алгоритм нахождения оптимальных параметров размытости.
- 4) Применение разработанных алгоритмов при обработке данных изменения цены пая открытого паевого инвестиционного фонда «Альфа-Капитал Акции» позволило увеличить доходность инвестирования на 50%.

ЛИТЕРАТУРА

1. Györfi L., Kohler M., and Walk H. Weak and strong universal consistency of semi-recursive kernel and partitioning regression estimates // *Statist. Decisions*. 1998. V. 16. P. 1 – 18.
2. Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание сигналов. М.: Наука, Физматлит, 1997. 336 с.
3. Кошкин Г.М., Пивен И.Г. Непараметрическое оценивание функционалов от условных распределений последовательностей сильного перемешивания // *Вестник ТГУ*. 2003. № 289. С. 187 – 200.
4. Балтрунас Й.Й., Рудзкене В.Ю. Регулярность процесса нелинейной авторегрессии // *Тр. АН ЛитССР*. Сер. Б. 1986. Т. 2(153). С. 118 – 122.
5. Васильев В.А., Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. М.: Наука, 2004. 508 с.

Статья представлена кафедрой теоретической кибернетики факультета прикладной математики и кибернетики Томского государственного университета, поступила в научную редакцию 17 сентября 2007 г.