УДК 519.22 MSC 62G05, 62G08
DOI 10.17223/19988621/57/3

Е.А. Пчелинцев, С.С. Перелевский

АДАПТИВНОЕ ОЦЕНИВАНИЕ В ГЕТЕРОСКЕДАСТИЧНОЙ НЕПАРАМЕТРИЧЕСКОЙ РЕГРЕССИИ[1]

Рассматривается задача оценивания неизвестной функции гетероскедастичной регрессии. Предлагается адаптивная процедура выбора модели, основанная на улучшенных взвешенных оценках наименьших квадратов со специально подобранными весовыми коэффициентами. Устанавливается, что процедура имеет более высокую среднеквадратическую точность по сравнению с процедурой, основанной на классических взвешенных оценках наименьших квадратов. Для среднеквадратического риска предлагаемой процедуры доказывается неасимптотическое оракульное неравенство, определяющее для него точную верхнюю границу по всевозможным оценкам. Приводятся результаты численного моделирования.

Ключевые слова: *гетероскедастичная регрессия, улучшенное непараметрическое оценивание, процедура выбора модели, оракульное неравенство.*

1. Введение

Предположим, что наблюдения $(y_j)_{1\le j\le n}$ описываются уравнением гетероскедастичной регрессии

$$y_j = S(x_j) + \sigma_j \xi_j,\ 1 \le j \le n, \tag{1}$$

где $x_j = j/n$; $S(\cdot) \in L_2[0,1]$ – неизвестная 1-периодическая $\mathbb{R} \to \mathbb{R}$-функция, которую требуется оценить; $(\xi_j)_{1\le j\le n}$ – последовательность независимых одинаково распределённых случайных величин, таких, что $\mathbf{E}(\xi_j) = 0$, $\mathbf{E}(\xi_j^2) = 1$, $\mathbf{E}(\xi_j^4) = \xi^* < \infty$; $(\sigma_j)_{1\le j\le n}$ – неизвестные коэффициенты волатильности, которые в общем случае зависят от x_j и функции S; n – число наблюдений. Предполагается, что коэффициенты σ_j в (1) удовлетворяют следующим неравенствам:

$$\max_{1\le j\le n} \sigma_j^2 \le \overline{\sigma},\ \min_{1\le j\le n} \sigma_j^2 \ge \sigma_0, \tag{2}$$

где граница $\overline{\sigma}$ является функцией от n, т.е. $\overline{\sigma} = \overline{\sigma}_n$ такая, что для любого $\delta > 0$

$$\lim_{n\to\infty} \frac{\overline{\sigma}_n}{n^\delta} = 0,$$

σ_0 – некоторая положительная величина.

[1] Работа выполнена в рамках государственного задания Минобрнауки России (проект № 2.3208.2017/4.6) и при финансовой поддержке РФФИ (грант № 16-01-00121 А).

Модели вида (1) с $\sigma = \sigma(x_j)$ были введены как обобщение непараметрической модели ANCOVA [1]. Следует отметить, что гетероскедастичные регрессии с таким коэффициентом волатильности встречаются в эконометрических исследованиях, а именно, в анализе инвестиционного поведения фирм [2, 3]. Например, для задачи потребления используют некоторые параметрические аналоги модели (1) с коэффициентами волатильности

$$\sigma_j = c_0 + c_1 x_j + c_2 S^2(x_j^2),$$

где c_0, c_1, c_2 – некоторые неотрицательные неизвестные постоянные. Отметим, что регрессионная модель (1) является важной в стохастических дифференциальных уравнениях для аппроксимации диффузионных процессов с непрерывным временем путем использования последовательных ядерных оценок, имеющих асимптотически минимальные отклонения (см., например, [4–6]).

Цель работы — построить адаптивную процедуру выбора модели для оценивания функции S в модели (1), которая имеет более высокую среднеквадратическую точность по сравнению с оценками МНК для конечного объема наблюдений, и получить оракульное неравенство для среднеквадратического риска построенной процедуры выбора модели.

Напомним, что метод выбора моделей появился в пионерских работах Акайке [7] и Мэллоуза [8], которые предложили ввести пенализационное слагаемое в критерий максимального правдоподобия. Далее, Барон, Бирже и Массар [9], Массар [10] и Кнайп [11] развили этот метод для получения неасимптотических оракульных неравенств в непареметрических регрессионных моделях с гауссовскими шумами в дискретном времени. К сожалению, этот метод не может быть применен к регрессиям с негауссовскими шумами, поскольку для таких моделей оценки коэффициентов Фурье являются, вообще говоря, зависимыми случайными величинами и, как правило, с неизвестными корреляционными коэффициентами. Этот факт делает невозможным применение метода Барона–Бирже–Массара–Кнайпа для получения оракульных неравенств для таких моделей. По этой причине в данной работе для оценивания функции в модели (1) применяется метод, развитый в [12].

В отличие от упомянутых работ, в статье разрабатывается процедура выбора моделей для адаптивного оценивания функции, основанная на взвешенных улучшенных оценках наименьших квадратов со специально подобранными коэффициентами. Применение метода улучшенного оценивания, известного как феномен Стейна, для параметрических моделей регрессии в непрерывном времени стало возможно благодаря работам [13, 14], а для непараметрических моделей – работам [15, 16], в которых предложены соответствующие модификации известной процедуры Джеймса – Стейна для регрессионных моделей с шумами, содержащими импульсные компоненты. В данной работе этот метод применен к процедуре выбора моделей с заменой классических оценок наименьших квадратов на их улучшенные версии. Это позволяет улучшить среднеквадратическую точность оценивания. Процедура выбора моделей дает адаптивное правило выбора лучшей оценки (в смысле точного неасимптотического оракульного неравенства) в определенном семействе проекционных оценок. Точное оракульное неравенство означает оценивание сверху неасимптотического риска минимальным риском по выбранному семейству оценок, умноженным на коэффициент, стремящийся к единице с ростом отношения сигнал/шум. Такие неравенства позволяют доказывать эффективность процедур без знания регулярности оцениваемой функции, т.е. в адаптивной постановке [17].

Статья состоит из следующих разделов. В разделе 2 предлагаются улучшенные взвешенные оценки наименьших квадратов. Доказывается, что такие оценки превосходят классические оценки МНК по среднеквадратической точности. В разделе 3 строится адаптивная процедура выбора моделей для оценивания функции, основанная на взвешенных улучшенных оценках наименьших квадратов. Получено точное неасимптотическое оракульное неравенство для риска процедуры. Раздел 4 содержит результаты численного сравнения эмпирических рисков данной адаптивной улучшенной процедуры с процедурой, предложенной в [12]. Доказательства основных результатов приводятся в приложении.

2. Улучшенные взвешенные оценки МНК

Для оценивания неизвестной функции S в модели (1) воспользуемся ее разложением в ряд Фурье по тригонометрическому базису $\phi_j(x)_{j\ge1}$ в пространстве $L_2[0,1]$:

$$S(x) = \sum_{j\ge1} \theta_j \phi_j(x), \tag{3}$$

где $\phi_1 = 1, \quad \phi_j(x) = \begin{cases} \sqrt{2}\cos(2\pi[j/2]x), & \text{для четных } j, \\ \sqrt{2}\sin(2\pi[j/2]x), & \text{для нечетных } j, \end{cases}$

$[a]$ – целая часть числа a и коэффициенты Фурье

$$\theta_j = (S, \phi_j)_n = \frac{1}{n}\sum_{l=1}^{n} S(x_l)\phi_j(x_l)$$

– эмпирическое скалярное произведение функций S и ϕ_j на сетке $(x_j)_{1\le j\le n}$. Поскольку они зависят от неизвестной функции S, то также являются неизвестными и подлежат оцениванию. Если n является нечетным, то $\phi_j(x)_{j\ge1}$ – ортонормированный базис относительно введенного эмпирического скалярного произведения, т.е. для всех $1 \le j \le n$

$$(\phi_i, \phi_j)_n = \frac{1}{n}\sum_{l=1}^{n} \phi_i(x_l)\phi_j(x_l) = \delta_{ij},$$

где δ_{ij} – символы Кронекера. Используя этот факт, получаем, что оценки МНК коэффициентов Фурье определяются формулой

$$\hat{\theta}_{j,n} = (Y, \phi_j)_n.$$

Здесь $Y=(y_1,...,y_n)'$ и штрих обозначает транспонирование. Из (1) следует, что оценки МНК удовлетворяют следующему уравнению:

$$\hat{\theta}_{j,n}=\theta_j+\frac{1}{n}\xi_{j,n},\ \xi_{j,n}=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\sigma_l\xi_l\phi_j(x_l). \tag{4}$$

Теперь, как и в [12], определим класс взвешенных оценок МНК функции S следующим образом:

$$\hat{S}_\lambda(x)=\sum_{j=1}^{n}\lambda(j)\hat{\theta}_{j,n}\phi_j(x), \tag{5}$$

где $x\in[0,1]$; вектор весовых коэффициентов $\lambda=(\lambda_1,...,\lambda_n)$ принадлежит некоторому конечному множеству $\Lambda\subset[0,1]^n$ для $n\geq 3$. Пусть $\nu=|\Lambda|$ – мощность множества Λ,

$$\upsilon_n=\max_{\lambda\in\Lambda}\sum_{j=1}^{n}\lambda(j),\ \upsilon_{1,n}=\max_{\lambda\in\Lambda}\sup_{1\leq l\leq n}\left|\sum_{j=1}^{n}\lambda(j)(\phi_j^2(x_l)-1)\right|.$$

Предположим, что существует целое число $d=d(n)\leq n$, такое, что $\lambda(j)=1$ для $j=\overline{1,d}$. Учитывая такой вид весовых коэффициентов, для оценивания неизвестной функции S в (1) вместо взвешенных оценок МНК предлагается использовать оценки вида

$$S_\lambda^*(x)=\sum_{j=1}^{n}\lambda(j)\theta_{j,n}^*\phi_j(x),\ \theta_{j,n}^*=(1-g(j))\hat{\theta}_{j,n}, \tag{6}$$

где
$$g(j)=\frac{c_n}{\|\tilde{\theta}_n\|_d}I_{\{1\leq j\leq d\}},\quad \|\tilde{\theta}_n\|=\sum_{j=1}^{d}\hat{\theta}_{j,n}^2,$$

I_A – индикатор множества A и

$$c_n=\frac{(d-1)\sigma_0^2}{n(r_n+\sqrt{d\overline{\sigma}^2/n})}.$$

Здесь r_n – положительный параметр, такой, что $\lim_{n\to\infty}r_n=\infty$ и для любого $\delta>0$

$$\lim_{n\to\infty}\frac{r_n}{n^\delta}=0.$$

Далее обозначим разность эмпирических среднеквадратических рисков предложенной оценки (6) и оценки МНК (5) как

$$\Delta_n(S)=\mathbf{E}_S\|S_\lambda^*-S\|_n^2-\mathbf{E}_S\|\hat{S}_\lambda-S\|_n^2, \tag{7}$$

где E_S – математическое ожидание относительно распределения наблюдений $Y=(y_1,...,y_n)'$ при фиксированной функции S и

$$\|S\|_n^2=\frac{1}{n}\sum_{l=1}^{n}S^2(x_l).$$

Предложенная оценка (6) обладает следующим свойством.

Теорема 1. *Пусть в модели* (1) *коэффициенты волатильности* $(\sigma_j)_{1\le j\le n}$ *удовлетворяют неравенствам* (2). *Тогда для всех* $\lambda\in\Lambda$ *оценка* (6) *превосходит по среднеквадратической точности взвешенную оценку МНК* (5). *При этом разность среднеквадратических рисков* (7) *удовлетворяет следующему неравенству:*

$$\Delta_n(S)\le -c_n^2. \tag{8}$$

Замечание 1. *Теорема 1 утверждает, что оценка* (6) *обладает более высокой среднеквадратической точностью по сравнению с оценкой МНК* (5), *причем минимальный выигрыш в точности равен* c_n^2.

3. Выбор модели. Оракульное неравенство

В данном разделе предлагается правило выбора из введенного класса оценок (6) наилучшей оценки в смысле точного оракульного неравенства. Такой метод позволяет решить задачу в адаптивной постановке, т.е. в случае, когда неизвестна степень гладкости функции S в (1). Для построения адаптивной процедуры выбора модели определим множество Λ. С этой целью зададим сетку

$$A=\{1,...k\}\times\{t_1,...,t_m\},$$

где $t_i=i\varepsilon$ и $m=\left[\dfrac{1}{\varepsilon^2}\right]$. Считаем, что параметры $k\ge 1$ и $0<\varepsilon<1$ – функции от n, т.е. $k=k_n$ и $\varepsilon=\varepsilon_n$, такие, что

$$\begin{cases}\lim\limits_{n\to\infty}k_n=+\infty, & \lim\limits_{n\to\infty}\dfrac{k_n}{\ln n}=0,\\ \lim\limits_{n\to\infty}\varepsilon_n=0, & \lim\limits_{n\to\infty}n^{\delta}\varepsilon_n=+\infty,\end{cases} \tag{9}$$

для любого $\delta>0$. Например, можно взять для $n\ge 3$

$$\varepsilon_n=\frac{1}{\ln n},\ k_n=\overline{k}+\sqrt{\ln n},$$

где $\overline{k}$ – некоторая неотрицательная постоянная. Для любого $\alpha=(\beta,t)\in A$ положим, что весовой вектор $\lambda_\alpha=(\lambda_\alpha(1),...,\lambda_\alpha(n))'$ с компонентами

$$\lambda_\alpha(j)=1_{\{1\le j\le j_o\}}+(1-(j/\omega_\alpha)^\beta)1_{\{1\le j\le\omega_\alpha\}}. \tag{10}$$

Здесь $d=d(n)=[\omega_\alpha/\ln n]$,

$$\omega_\alpha=\varpi+(A_\beta tn)^{1/(2\beta+1)},\ A_\beta=(\beta+1)(2\beta+1)/(\pi^{2\beta}\beta)$$

и ϖ – некоторая неотрицательная постоянная. Отметим, что в этом случае $\nu=mk$.

Замечание 2. *Весовые коэффициенты вида* (10) *были введены Пинскером* [18] *и Нюссбаумом* [19] *для доказательства асимптотической эффективности оценок гауссовских сигналов. Для негауссовской гетероскедастичной модели весовые коэффициенты* (10) *использовались в работах* [12, 17].

Для того чтобы выбрать вектор весовых коэффициентов $\lambda \in \Lambda$ в (6), необходимо минимизировать эмпирическую квадратическую ошибку вида

$$Err_n(\lambda) = \left\| S_\lambda^* - S \right\|_n^2 = \frac{1}{n}\sum_{l=1}^{n}(S_\lambda^*(x_l) - S(x_l))^2,$$

которую в силу равенств (6) и (3) можно переписать, как

$$Err_n(\lambda) = \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{l=1}^{n}(\lambda(j)\theta_{j,n}^* - \theta_j)\phi_j(x_l)\right)^2. \tag{11}$$

Затем, учитывая, что $n^{-1}\sum_{l=1}^{n}\phi_j^2(x_l) = 1$, перепишем (11) следующим образом:

$$Err_n(\lambda) = \sum_{j=1}^{n}\lambda^2(j)\theta_{j,n}^{*2} - 2\sum_{j=1}^{n}\lambda(j)\theta_{j,n}^*\theta_j + \|S\|^2,$$

Так как в этом равенстве коэффициенты θ_j неизвестны, необходимо величины $\theta_{j,n}^*\theta_j$ заменить некоторыми оценками. В качестве таких оценок предлагаются

$$\overline{\theta}_{j,n} = \theta_{j,n}^*\hat{\theta}_{j,n} - \frac{1}{n}\hat{\varsigma}_n,$$

где $\hat{\varsigma}_n$ – некоторая оценка интегрированной дисперсии шума

$$\varsigma_n = \frac{1}{n}\sum_{l=1}^{n}\sigma_l^2.$$

При осуществлении этой замены в эмпирической ошибке нужно заплатить «штраф». Определим платежную функцию как

$$J_n(\lambda) = \sum_{j=1}^{n}\lambda^2(j)\theta_{j,n}^{*2} - 2\sum_{j=1}^{n}\lambda(j)\overline{\theta}_{j,n} + \rho\hat{P}_n(\lambda), \tag{12}$$

где $0 < \rho < 1$ – некоторый положительный коэффициент, который будет определен ниже. Пенализационное слагаемое

$$\hat{P}_n(\lambda) = \frac{|\lambda|^2\,\hat{\varsigma}_n}{n},\ \ |\lambda|^2 = \sum_{j=1}^{n}\lambda^2(j). \tag{13}$$

Отметим, что если последовательность $(\sigma_j)_{1\le j\le n}$ известна, то

$$P_n(\lambda) = \frac{|\lambda|^2\,\varsigma_n}{n}.$$

Теперь, полагая

$$\lambda^* = \operatorname*{arg\,min}_{\lambda\in\Lambda} J_n(\lambda),$$

определим процедуру выбора модели равенством

$$S^* = S_{\lambda^*}^*. \tag{14}$$

Заметим, что λ^* существует, поскольку множество Λ конечно. В случае, когда λ^* не единственное, берем любое из них. Далее получим неасимптотическую верхнюю границу для среднеквадратического риска оценки (14).

Теорема 2. *Пусть наблюдения описываются уравнением* (1). *Тогда для любого* $0 < \rho < 1/3$ *среднеквадратический риск предложенной процедуры выбора модели* (14) *для функции S удовлетворяет следующему оракульному неравенству*:

$$R(S^*, S) \le \frac{1-\rho}{1-2\rho} \min_{\lambda \in \Lambda} R(S^*_\lambda, S) + \frac{1}{n\rho} \Psi_n, \tag{15}$$

где
$$R(S^*_\lambda, S) = \mathbf{E}_S \left\| S^*_\lambda - S \right\|_n^2$$

и
$$\Psi_n = 3\left[(8c_n + 64\bar{\sigma} + \nu\bar{\sigma})\left(1 + 2n^{-1} + \bar{\sigma}(\upsilon_{1,n} + \upsilon_n)\sqrt{\xi^*/n}\right) + \upsilon_n \mathbf{E}_S \left| \varsigma_n - \hat{\varsigma}_n \right| \right].$$

4. Численное моделирование

В этом разделе проиллюстрируем теоретически установленный результат в теореме 1 с помощью численного моделирования в среде Matlab. В качестве функции S в модели (1) выберем

$$S(x) = x\sin(2\pi x) + x^2(1-x)\cos(4\pi x),\ \sigma_j^2 = 2 + S(x_j),$$

$(\xi_j)_{1 \le j \le n}$ – последовательность н.о.р. $N(0,1)$ с.в. Для вычисления весовых коэффициентов $(\lambda_\alpha)_{\alpha \in A}$ положим в (10)

$$k^* = 100 + \sqrt{\ln n},\ \varepsilon = \frac{1}{\ln n},\ m = [\ln^2 n],\ \rho = \frac{1}{3 + \ln^2 n},\ \omega_\alpha = 100 + (A_\beta t n)^{\frac{1}{2\beta+1}}.$$

В таблице приведены результаты моделирования эмпирических среднеквадратических рисков по $N = 1000$ реализациям процедур выбора модели (14), построенным на основе предложенных улучшенных оценок (6) и на основе оценок МНК (5):

$$\tilde{R}(S^*, S) = \frac{1}{N} \sum_{m=1}^{N} \| S^*_m - S \|_n^2,$$

где S^*_m – оценка, полученная по m-й реализации выборки.

Эмпирические квадратические риски

n	101	401	1001
$R(S^*,S)$	0.2779	0.0596	0.0196
$R(\hat{S},S)$	0.9276	0.1297	0.1011
$\Delta_n(S)$	-0.6497	-0.0701	-0.0815

Из таблицы видно, что эмпирический риск предложенной процедуры (6) меньше, чем для процедуры выбора модели $\hat{S}$, построенной на основе оценок МНК (5).

Далее на рис. 1 представлены графики истинной функции S (сплошная линия) и ее оценок S^* (штрихованная линия) и $\hat{S}$ (пунктирная линия) при $n = 401$.

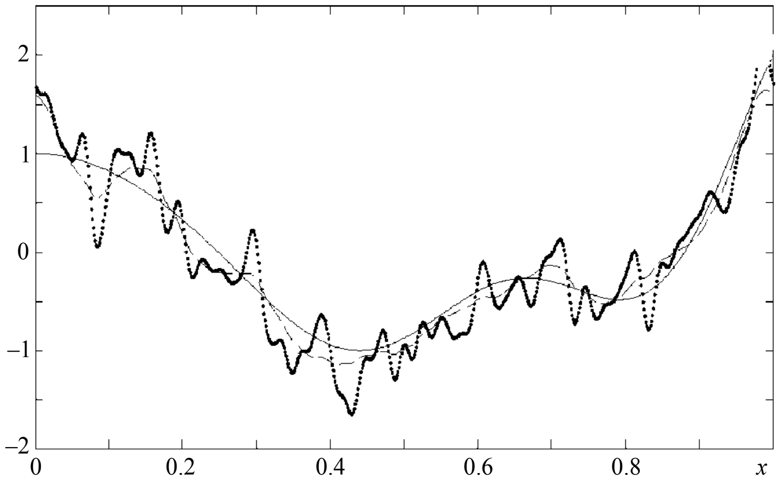

Рис. 1. Графики функции регрессии и ее оценок при $n = 401$
Fig. 1. Plots of the regression function and its estimates at $n = 401$

5. Приложение

Доказательство теоремы 1. Рассмотрим среднеквадратический риск оценки (6)

$$\mathbf{E}\left\|S^*_\lambda - S\right\|_n^2 = \frac{1}{n}\mathbf{E}\sum_{l=1}^{n}(S^*_\lambda(x_l) - S(x_l))^2 =$$

$$= \frac{1}{n}\mathbf{E}\sum_{l=1}^{n}\left(\sum_{j=1}^{d}\theta^*_{j,n}\varphi_j(x_l) + \sum_{j=d+1}^{n}\lambda(j)\hat{\theta}_{j,n}\varphi_j(x_l) - S(x_l)\right)^2,$$

где $g(j) = c_n / \left\|\tilde{\theta}_n\right\|$. Так как

$$\hat{S}(x_l) - S(x_l) = \sum_{k=1}^{n}(\lambda_k\hat{\theta}_{k,n} - \theta_k)\varphi_k(x_l)$$

и $(\phi_k)_{k\geq 1}$ – ортонормированная система функций, то

$$2\mathbf{E}\left(\hat{S}_\lambda - S,\ \sum_{j=1}^{d}g(j)\hat{\theta}_{j,n}\phi_j\right)_n = \frac{2}{n}\mathbf{E}\sum_{l=1}^{n}\left(\sum_{k=1}^{n}(\lambda_k\hat{\theta}_{k,n} - \theta_k)\phi_k(x_l)\right) = 2c_n\mathbf{E}\sum_{j=1}^{d}(\hat{\theta}_{j,n} - \theta_j)\frac{\hat{\theta}_{j,n}}{\left\|\tilde{\theta}_n\right\|}$$

и

$$\mathbf{E}\left\|\sum_{j=1}^{d}g(j)\hat{\theta}_{j,n}\phi_j\right\|_n^2 = c_n^2.$$

Отсюда

$$\Delta_n(n) = c_n^2 - 2c_n\mathbf{E}\sum_{j=1}^{d}(\hat{\theta}_{j,n} - \theta_j)\frac{\hat{\theta}_{j,n}}{\left\|\tilde{\theta}_n\right\|}.$$

Далее, вычисляя здесь математическое ожидание, как показано в доказательстве теоремы 2.1 в [13], получаем неравенство

$$\Delta_n(n) \le c_n^2 - 2c_n \frac{(d-1)\sigma_0^2}{n(r_n + \sqrt{\bar{\sigma}^2 / n})}.$$

Минимизируя правую часть относительно c_n, имеем

$$\Delta_n(n) \le -c_n^2.$$

Теорема 1 доказана.

Доказательство теоремы 2. Представим эмпирическую квадратическую ошибку $Err_n(\lambda)$ следующим образом:

$$Err_n(\lambda) = J_n(\lambda) + 2\sum_{j=1}^{n} \lambda(j)(\bar{\theta}_{j,n} - \theta^*_{j,n}\theta_j) + \|S\|_n^2 - \rho\hat{P}_n(\lambda), \qquad (16)$$

где $J_n(\lambda)$ определена в (12). Используя (4) и (6), перепишем (16) в виде

$$Err_n(\lambda) = J_n(\lambda) + 2\sum_{j=1}^{n} \lambda(j)\hat{\theta}_{j,n}\xi_{j,n} - \frac{2}{\sqrt{n}}\sum_{j=1}^{n} \lambda(j)g(j)\hat{\theta}_{j,n}\xi_{j,n} + \|S\|_n^2 - \rho\hat{P}_n(\lambda) + \frac{2}{n}\sum_{j=1}^{n} \lambda(j)\hat{\varsigma}_n.$$

Положим

$$\varsigma_{j,n} = \frac{1}{n}\sum_{l=1}^{n} \sigma_l^2 \phi_j^2(x_l), \quad \mu_{j,n} = \xi_{j,n}^2 - \varsigma_{j,n} = \mu'_{j,n} + \mu''_{j,n},$$

где

$$\mu'_{j,n} = \frac{1}{n}\sum_{l=1}^{n} \sigma_l^2 \phi_j^2(x_l)(\xi_l^2 - 1), \ \mu''_{j,n} = 2\sum_{l=2}^{n} \tau_{j,l}\xi_l, \ \tau_{j,l} = \frac{1}{n}\sigma_l \phi_j(x_l)\sum_{k=1}^{l-1} \sigma_k \phi_j(x_k)\xi_k.$$

Определим функции

$$M(\lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \lambda(j)\theta_j \xi_{j,n}, \ \Delta(\lambda) = \frac{1}{n}\sum_{j=1}^{n} \lambda(j)(\varsigma_{j,n} - \hat{\varsigma}_n), \ N'(\lambda) = \sum_{j=1}^{n} \lambda(j)\mu'_{j,n},$$

$$N''(\lambda) = \frac{1}{\sqrt{n\varsigma_n}}\sum_{j=1}^{n} \bar{\lambda}(j)\mu''_{j,n} 1_{\{\varsigma_n > 0\}}, \ B_{1,n}(\lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \lambda(j)g(j)\hat{\theta}_{j,n}\xi_{j,n},$$

где $\bar{\lambda}(j) = \lambda_j / |\lambda|$. Тогда

$$Err_n(\lambda) = J_n(\lambda) + 2M(\lambda) + 2\Delta(\lambda) + 2N'(\lambda) + 2N''(\lambda) - 2B_{1,n}(\lambda) + \|S\|^2 - \rho\hat{P}_n(\lambda). \quad (17)$$

Теперь для некоторого фиксированного $\lambda_0 \in \Lambda$ и $\lambda^* \in \Lambda$ рассмотрим разность

$$Err_n(\lambda^*) - Err_n(\lambda_0) \le 2M(x) - B_{1,n}(\lambda^*) + B_{1,n}(\lambda_0) + 2\Delta(x) +$$
$$+ \frac{2}{n}N'(x) + 4\sqrt{P_n(\lambda^*)}N''(\lambda^*) - 4\sqrt{P_n(\lambda_0)}N''(\lambda_0) - \rho\hat{P}_n(\lambda^*) + \rho\hat{P}_n(\lambda_0),$$

где $x = \lambda^* - \lambda_0 \in \Lambda_1$, $\Lambda_1 = \Lambda \setminus \{\lambda_0\}$. Далее оценим каждое слагаемое в правой части этого неравенства. Рассмотрим

$$\Delta(\lambda) = \frac{1}{n}\sum_{j=1}^{n} \lambda(j)(\varsigma_{j,n} - \varsigma_n + \varsigma_n - \hat{\varsigma}_n) = \frac{1}{n}\sum_{j=1}^{n} \lambda(j)(\varsigma_{j,n} - \varsigma_n) + \frac{1}{n}\sum_{j=1}^{n} \lambda(j)(\varsigma_n - \hat{\varsigma}_n).$$

Тогда
$$|\Delta(\lambda)| \le \frac{\overline{\sigma}\upsilon_{1,n}}{n} + \frac{\upsilon_n |\varsigma_n - \hat{\varsigma}_n|}{n} \tag{18}$$
и
$$\sup_{\lambda\in\Lambda} \mathbf{E}|N'(x)| \le \overline{\sigma}(\upsilon_{1,n} + \upsilon_n)\frac{\sqrt{\xi^*}}{\sqrt{n}}. \tag{19}$$
Используя неравенство
$$2|ab| \le \varepsilon a^2 + \frac{b^2}{\varepsilon}, \tag{20}$$
имеем для любого $\varepsilon > 0$
$$4\sqrt{P_n(\lambda)}N''(\lambda) \le \varepsilon P_n(\lambda) + \frac{4(N''(\lambda))^2}{\varepsilon}. \tag{21}$$
Заметим, что
$$\sup_{\lambda\in\mathbb{R}^n} \mathbf{E}(N''(\lambda))^2 \le \frac{2\overline{\sigma}}{n}. \tag{22}$$
Из неравенства Коши–Буняковского и неравенства (20) следует, что
$$2B_{1,n}(\lambda) = 2\frac{|\lambda|}{\sqrt{n}}\sum_{j=1}^{n}\overline{\lambda}(j)g(j)\hat{\theta}_{j,n}\xi_{j,n} \le$$
$$\le P_n(\lambda) + \frac{c_n}{\varepsilon\sqrt{\varsigma_n}}\left(\sum_{j=1}^{n}\overline{\lambda}^2\xi_{j,n}^2\right)^{1/2} \le \varepsilon P_n(\lambda) + \frac{c_n}{\varepsilon\varsigma_n}\sum_{j=1}^{n}\overline{\lambda}^2\xi_{j,n}^2.$$
Учитывая равенство
$$\sum_{j=1}^{n}\overline{\lambda}^2(j)(\xi_{j,n} - \varsigma_{j,n} - \varsigma_n) = \sum_{j=1}^{n}\overline{\lambda}^2(j)(\mu_{j,n} + \varsigma_n) = N'(\overline{\lambda}^2) + N''(\overline{\lambda}^2),$$
получаем
$$2B_{1,n}(\lambda) \le \varepsilon P_n(\lambda) + \frac{c_n}{\varepsilon\varsigma_n}N'(\overline{\lambda}^2) + N''(\overline{\lambda}^2)$$
и
$$2\mathbf{E}B_{1,n}(\lambda) \le \varepsilon\hat{P}_n(\lambda) + \frac{c_n}{\varepsilon\overline{\sigma}}\left(\frac{2\overline{\sigma}}{n} + \overline{\sigma}(\upsilon_{1,n} + \upsilon_n)\frac{\sqrt{\xi^*}}{\sqrt{n}}\right).$$
Заметим, что
$$\hat{P}_n(\lambda) = \frac{|\lambda|(\hat{\varsigma}_n - \varsigma_n + \varsigma_n)}{n} \le P_n(\lambda) + \frac{|\lambda||\hat{\varsigma}_n - \varsigma_n|}{n} \le P_n(\lambda) + \frac{\upsilon_n|\varsigma_n - \hat{\varsigma}_n|}{n}.$$
При $\varepsilon = \rho/4$ имеем
$$Err_n(\lambda^*) \le Err_n(\lambda) + \frac{(2+2\rho)\upsilon_n|\varsigma_n - \hat{\varsigma}_n|}{n} + 2M(x) +$$
$$+\frac{4c_n}{\rho\varsigma_n}\left(N'(\overline{\lambda}_0{}^2) + N''(\overline{\lambda}_0{}^2)\right) + \frac{4c_n}{\rho\varsigma_n}\left(N'(\overline{\lambda}^{*2}) + N''(\overline{\lambda}^{*2})\right) + 2N'(x) +$$
$$+\frac{16(N''(\lambda^*))^2}{\rho} + \frac{16(N''(\lambda_0))^2}{\rho} - \frac{\rho}{2}P_n(\lambda^*) + \frac{3}{2}\rho P_n(\lambda_0).$$

Теперь оценим третье слагаемое в правой части. Заметим, что для любого $\varepsilon > 0$

$$2|M\ (x)| \le \varepsilon \|S_x\|_n^2 + \frac{Z^*}{n\varepsilon},\ Z^* = \sup_{x\in\Lambda_1} \frac{nM^2(x)}{\|S_x\|_n^2}, \tag{23}$$

где $S_x = \sum_{j=1}^{n} x(j)\theta_j \phi_j$. Поскольку

$$\mathbf{E}M^2(x) \le \frac{\overline{\sigma}}{n}\sum_{j=1}^{n} x^2(j)\theta_j^2 = \overline{\sigma}\frac{\|S_x\|_n^2}{n}, \tag{24}$$

то

$$\mathbf{E}Z^* \le \sum_{x\in\Lambda_1} \frac{nM^2(x)}{\|S_x\|_n^2} \le \nu\overline{\sigma}.$$

Рассмотрим далее

$$\left\|S_\lambda^* - S_{\lambda_0}^*\right\|_n^2 = \sum_{j=1}^{n} (x(j)+\beta(j))^2\,\hat\theta_{j,n}^2 \ge \left\|\hat S_x\right\|_n^2 + 2\sum_{j=1}^{n} x(j)\beta(j))\hat\theta_{j,n}, \tag{25}$$

где $\beta(j) = \lambda_0(j)g(j)(\lambda) - \lambda(j)g(j)(\lambda)$. Оценим разность

$$\|S_x\|_n^2 - \left\|S_\lambda^* - S_{\lambda_0}^*\right\|_n^2 \le \|S_x\|_n^2 - \left\|\hat S_x\right\|_n^2 - 2\sum_{j=1}^{n} x(j)\,\beta(j))\hat\theta_{j,n}^2 \le$$

$$\le -2M_1(x) - 2\sum_{j=1}^{n} x(j)\beta(j))\hat\theta_{j,n}\theta_j - \frac{2}{\sqrt{n}}\Phi(x).$$

Здесь

$$M_1(x) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} x^2(j)\theta_j\xi_{j,n},\ \Phi(x) = \sum_{j=1}^{n} x(j)\beta(j)\hat\theta_{j,n}\xi_{j,n}.$$

Принимая во внимание, что $|x(j)| \le 1$ для любого $x \in \Lambda_1$, получаем

$$2|M_1(x)| \le \varepsilon\|S_x\|_n^2 + \frac{Z_1^*}{n\varepsilon},\ Z_1^* = \sup_{x\in\Lambda_1}\frac{nM_1^2(x)}{\|S_x\|_n^2}.$$

Заметим, что

$$\mathbf{E}Z_1^* \le \nu\overline{\sigma}$$

и для любого $0 < \varepsilon < 1$

$$\|S_x\|_n^2 \le \varepsilon\|S_x\|_n^2 + \frac{Z_1^*}{n\varepsilon} + \left\|S_\lambda^* - S_{\lambda_0}^*\right\|_n^2 - 2\sum_{j=1}^{n} x(j)\beta(j)\hat\theta_{j,n}\theta_j - \frac{2}{\sqrt{n}}\Phi(x).$$

Так как

$$\sum_{j=1}^{n}\beta^2(j)\hat\theta_{j,n}^2 \le \frac{4c_n}{\varepsilon n} \tag{26}$$

и принимая во внимание, что $|x(j)| \le 1$ и выполняется неравенство (20), получаем для любого $\varepsilon > 0$

$$2\left|\sum_{j=1}^{n} x(j)\beta(j)\hat\theta_{j,n}\theta_j\right| \le \varepsilon\|S_x\|_n^2 + \frac{4c_n}{\varepsilon n}.$$

Используя неравенства Коши–Буняковского и (26), имеем

$$\frac{2}{\sqrt{n}}|\Phi(\lambda)| \le \varepsilon P_n(\lambda) + \frac{c_n}{\varepsilon\varsigma_n}\sum_{j=1}^{n}\overline{\lambda}^2(j)\xi_{j,n}^2 \le \varepsilon P_n(\lambda) + \frac{c_n(N'(\overline{\lambda}^2)+N''(\overline{\lambda}^2))}{n\overline{\sigma}\varepsilon}.$$

Тогда

$$\frac{2}{\sqrt{n}}|\Phi(x)| \le \varepsilon P_n(\lambda^*) + \varepsilon P_n(\lambda_0) + \frac{c_n}{n\overline{\sigma}\varepsilon}\Big(N'(\overline{\lambda}_0{}^2)+N''(\overline{\lambda}_0{}^2)+N'(\overline{\lambda}^{*2})+N''(\overline{\lambda}^{*2})\Big).$$

Применяя неравенство (23) для $M_1(x)$ в (25), получаем верхнюю границу для $\|S_x\|_n^2$, т.е.

$$\|S_x\|_n^2 \le \frac{1}{1-2\varepsilon}\left(\left\|S_\lambda^* - S_{\lambda_0}^*\right\|_n^2 + \frac{Z_1^*}{n\varepsilon} + \varepsilon P_n(\lambda^*) + \varepsilon P_n(\lambda_0) + \frac{4c_n}{\varepsilon n}\right) +$$
$$+\frac{c_n}{(1-2\varepsilon)n\overline{\sigma}\varepsilon}\Big(N'(\overline{\lambda}_0{}^2)+N''(\overline{\lambda}_0{}^2)+N'(\overline{\lambda}^{*2})+N''(\overline{\lambda}^{*2})\Big).$$

Учитывая это неравенство и $\left\|S_\lambda^* - S_{\lambda_0}^*\right\|_n^2 \le 2\left(Err_n(\lambda) - Err_n(\lambda_0)\right)$ в (23), находим

$$2|M(x)| \le \frac{\varepsilon}{1-2\varepsilon}\left(2\left((Err_n(\lambda)+Err_n(\lambda_0)\right)) + \frac{Z_1^*}{n\varepsilon} + \frac{4c_n}{\varepsilon n} + \varepsilon P_n(\lambda^*) + \varepsilon P_n(\lambda_0) + \right.$$
$$\left. + \frac{c_n}{n\overline{\sigma}\varepsilon}(N''(\overline{\lambda}_0{}^2)+N'(\overline{\lambda}^{*2})+N'(\overline{\lambda}_0{}^2)+N''(\overline{\lambda}^{*2}))\right) + \frac{Z^*}{n\varepsilon}.$$

Из неравенств (20) – (25) следует, что

$$Err_n(\lambda^*) \le Err_n(\lambda_0) + \frac{(2+2\rho)\upsilon_n|\varsigma_n - \hat{\varsigma}_n|}{n} + \frac{4c_n}{\rho\varsigma_n}\Big(N'(\overline{\lambda}^{*2})+N''(\overline{\lambda}^{*2})\Big) +$$
$$+\frac{4c_n}{\rho\varsigma_n}\left(N'(\overline{\lambda}_0{}^2)+N''(\overline{\lambda}_0{}^2) - \frac{\rho}{2}P_n(\lambda^*) + \frac{3}{2}\rho P_n(\lambda_0)\right) + 2N'(x) + \frac{16(N''(\lambda^*))^2}{\rho} +$$
$$+\frac{16(N''(\lambda_0))^2}{\rho} + \frac{\varepsilon}{1-2\varepsilon}\big(2(Err_n(\lambda)+Err_n(\lambda_0)) + \frac{Z_1^*}{n\varepsilon} + \frac{4c_n}{\varepsilon n} +$$
$$+N''(\overline{\lambda}_0{}^2)+N'(\overline{\lambda}^{*2})+N'(\overline{\lambda}_0{}^2)+N''(\overline{\lambda}^{*2})+\varepsilon P_n(\lambda^*)+\varepsilon P_n(\lambda_0)\big) + \frac{Z^*}{n\varepsilon}.$$

Используя Лемму 1 (приведенную ниже), делая необходимые преобразования при $\varepsilon = \rho/2$, получаем неравенство

$$R\left(S_{\lambda^*}^*, S\right) \le \frac{1-\rho}{1-2\rho}\min_{\lambda\in\Lambda} R\left(S_\lambda^*, S\right) + \frac{1}{n\rho}\Psi_n.$$

Теорема 2 доказана.

Лемма 1. *Для всех* $n \ge 3$, $\lambda \in \Lambda$ и $\varepsilon > 0$

$$P_n(\lambda) \le \frac{\mathbf{E}Err_n(\lambda)}{1-\varepsilon} + \frac{c_n}{\varepsilon(1-\varepsilon)}. \qquad (27)$$

Доказательство. По определению $Err_n(\lambda)$ имеем

$$\mathbf{E}Err_n(\lambda)=\sum_{j=1}^{n}(\lambda_j\theta^*_{j,n}-\theta_j)^2=\sum_{j=1}^{n}(\lambda_j(\theta^*_{j,n}-\theta_j)+(\lambda_j-1)\theta_j)^2\geq$$
$$\geq\sum_{j=1}^{n}\lambda_j^2(\theta^*_{j,n}-\theta_j)^2+\sum_{j=1}^{n}(\lambda_j(\lambda_j-1)\theta_j(\theta^*_{j,n}-\theta_j).$$

Взяв математическое ожидание и принимая во внимание, что второе слагаемое равно нулю, получаем

$$\mathbf{E}Err_n(\lambda)\geq\sum_{j=1}^{n}\lambda_j^2\mathbf{E}(\theta^*_{j,n}-\theta_j)^2=\sum_{j=1}^{n}\lambda_j^2\mathbf{E}(\hat\theta_{j,n}-\theta_j-\beta_j(\lambda)\hat\theta_{j,n})^2=$$
$$=P_n(\lambda)-2\mathbf{E}\sum_{j=1}^{n}\lambda_j^2\beta_j(\lambda)\hat\theta_{j,n}(\hat\theta_{j,n}-\theta_j)+\mathbf{E}\sum_{j=1}^{n}\lambda_j^2\beta_j^2(\lambda)\hat\theta_{j,n}^2\geq(1-\varepsilon)P_n(\lambda)-\frac{c_n}{\varepsilon}$$

Следовательно,

$$P_n(\lambda)\leq\frac{\mathbf{E}Err_n(\lambda)}{1-\varepsilon}+\frac{c_n}{\varepsilon(1-\varepsilon)}.$$

Лемма 1 доказана.

ЛИТЕРАТУРА

1. *Akritas M.G., VanKeilegom I.* ANCOVA methods for heteroscedastic nonparametric regression models // J. Amer. Statist. Assoc. 2001. No. 96. P. 220–232. DOI: https://doi.org/10.1198/016214501750332802
2. *Goldfeld S.M., Quandt R.E.* Nonlinear Methods in Econometrics. London: North–Holland, 1972.
3. *Cai T., Wang L.* Adaptive variance function estimation in heteroscedastic nonparametric regression // Annals of Statistics. 2008. V. 36. No. 5. P. 2025–2054.
4. *Galtchouk L., Pergamenshchikov S.* Nonparametric sequential estimation of the drift in diffusion processes // Mathematical Methods of Statistics. 2004. V. 13. No. 1. P. 25–49.
5. *Galtchouk L., Pergamenshchikov S.* Asymptotic efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes // Statistical Inference for Stochastic Process. 2006. V. 9. No. 1. P. 1–16.
6. *Galtchouk L., Pergamenshchikov S.* Adaptive sequential estimation for ergodic diffusion processes in quadratic metric // Journal of Nonparametric Statistics. 2011. V. 23. No. 2. P. 255–285. DOI: https://doi.org/10.1080/10485252.2010.544307.
7. *Akaike H.* A new look at the statistical model identification // IEEE Trans. on Automatic Control. 1974. V. 19. No. 7. P. 716–723.
8. *Mallows C.* Some comments on Cp // Technometrics. 1973. V. 15. P. 661–675.
9. *Barron A., Birgé L., Massart P.* Risk bounds for model selection via penalization // Probab. Theory Relat. Fields. 1999. V. 113. P. 301–415
10. *Massart P.* A non-asymptotic theory for model selection // European Congress of Mathematics. Zurich: Eur. Math. Soc. 2005.
11. *Kneip A.* Ordered linear smoothers // Annals of Statistics. 1994. V. 22. P. 835–866.
12. *Galtchouk L., Pergamenshchikov S.* Sharp non–asymptotic oracle inequalities for nonparametric heteroscedastic regression models // J. Nonparametric Statistics. 2009. V. 21. No. 1. P. 1–16. DOI: https://doi.org/10.1080/10485250802504096.
13. *Pchelintsev E.* Improved estimation in a non-Gaussian parametric regression // Statistical Inference for Stochastic Processes. 2013. V. 16. No. 1. P. 15–28. DOI: https://doi.org/10.1007/s11203-013-9075-0.

14. *Конев В.В., Пергаменщиков С.М., Пчелинцев Е.А.* Оценивание регрессии с шумами импульсного типа по дискретным наблюдениям // Теория вероятностей и ее применения. 2013. Т. 58. № 3. С. 454–471. DOI: https://doi.org/10.4213/tvp4520.
15. *Pchelintsev E., Pchelintsev V., Pergamenshchikov S.* Non asymptotic sharp oracle inequality for the improved model selection procedures for the adaptive nonparametric signal estimation problem // Communications – Scientific Letters of the University of Zilina. 2018. V. 20. No. 1. P. 72–76.
16. *Pchelintsev E., Pergamenshchikov S.* Oracle inequalities for the stochastic differential equations // Statistical Inference for Stochastic Processes. 2018. V. 21. No. 2. P. 469–483. DOI: https://doi.org/10.1007/s11203-018-9180-1.
17. *Galtchouk L., Pergamenshchikov S.* Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression // J. Korean Statistical Society. 2009. V. 38. No. 4. P. 305–322.
18. *Pinsker M.S.* Optimal filtration of square integrable signals in Gaussian white noise // Problems Transimis. information. 1981. No. 17. P. 120–133.
19. *Nussbaum M.* Spline smoothing in regression models and asymptotic efficiency in L2 // Ann. Statist. 1985. No. 13. P. 984–997.

Статья поступила 20.10.2018 г.

Pchelintsev E.A., Perelevskiy S.S. (2019) ADAPTIVE ESTIMATION IN A HETEROSCEDASTIC NONPARAMETRIC REGRESSION. *Vestnik Tomskogo gosudarstvennogo universiteta. Matematika i mekhanika* [Tomsk State University Journal of Mathematics and Mechanics]. 57. pp. 39–53

DOI 10.17223/19988621/57/3

Keywords: heteroscedastic regression, improved nonparametric estimation, model selection procedure, oracle inequality.

The paper considers the problem of estimating the unknown function of heteroscedastic regression. An adaptive model selection procedure based on improved weighted estimates of least squares with specially selected weight coefficients is proposed. It is established that the procedure has a higher mean-square accuracy than the procedure based on classical weighted least–squares estimates. For the mean square risk of the proposed procedure, a non-asymptotic oracle inequality is proved that determines the exact upper bound for it in all possible estimates. The results of numerical simulation are given.

AMS Mathematical Subject Classification: 62G05, 62G08

Financial support. This work was done under the Ministry of Science and Higher Education of the Russian Federation in theframework of the research Project No. 2.3208.2017/4.6 and by RFBR Grant No.16-01-00121.

PCHELINTSEV Evgenii Anatol'evich (Tomsk State University, Tomsk, Russian Federation). E-mail: evgen-pch@yandex.ru

PERELEVSKY Svyatoslav Sergeevich (Tomsk State University, Tomsk, Russian Federation). E-mail: slavaperelevskiy@mail.ru

REFERENCES

1. Akritas M.G., VanKeilegom I. (2001) ANCOVA methods for heteroscedastic nonparametric regression models. *J. Amer. Statist. Assoc.* 96. pp. 220–232. DOI: https://doi.org/10.1198/016214501750332802.
2. Goldfeld S.M., Quandt R.E. (1972) *Nonlinear Methods in Econometrics.* London: North–Holland.
3. Cai T., Wang L. (2008) Adaptive variance function estimation in heteroscedastic nonparametric regression. *Annals of Statistics*. 36(5). pp. 2025–2054.

4. Galtchouk L., Pergamenshchikov S. (2004) Nonparametric sequential estimation of the drift in diffusion processes. *Mathematical Methods of Statistics*. 13(1). pp. 25–49.
5. Galtchouk L., Pergamenshchikov S. (2006) Asymptotic efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Process*. 9(1). pp. 1–16.
6. Galtchouk L., Pergamenshchikov S. (2011) Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *Journal of Nonparametric Statistics*. 23(2). pp. 255–285. DOI: https://doi.org/10.1080/10485252.2010.544307
7. Akaike H. (1974) A new look at the statistical model identification. *IEEE Trans. on Automatic Control*. 19(7). pp. 716–723.
8. *Mallows C.* (1973) Some comments on Cp. *Technometrics*. 15. pp. 661–675.
9. Barron A., Birgé L., Massart P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields.* 113. pp. 301–415
10. Massart P. (2005) A non asymptotic theory for model selection. *Proceedings of the 4th European Congress of Mathematicians* (Ed. Ari Laptev). European Mathematical Society. pp. 309–323.
11. Kneip A. (1994) Ordered linear smoothers. *Annals of Statistics*. 22. pp. 835–866.
12. Galtchouk L., Pergamenshchikov S. (2009) Sharp non–asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*. 21(1). pp. 1–16. DOI: https://doi.org/10.1080/10485250802504096.
13. Pchelintsev E. (2013) Improved estimation in a non–Gaussian parametric regression // *Statistical Inference for Stochastic Processes*. 16(1). pp. 15–28. DOI: https://doi.org/10.1007/s11203-013-9075-0
14. Konev V.V., Pergamenshchikov S.M., Pchelintsev E.A. Estimation of a Regression with the Pulse Type Noise from Discrete Data. *Theory Probab. Appl.* 58(3). pp. 442–457. DOI: https://doi.org/10.1137/S0040585X9798662X.
15. Pchelintsev E., Pchelintsev V., Pergamenshchikov S. (2018) Non asymptotic sharp oracle inequality for the improved model selection procedures for the adaptive nonparametric signal estimation problem. *Communications - Scientific Letters of the University of Zilina*. 20(1). pp. 72–76.
16. Pchelintsev E., Pergamenshchikov S. (2018) Oracle inequalities for the stochastic differential equations. *Statistical Inference for Stochastic Processes.* 21(2). pp. 469–483. DOI: https://doi.org/10.1007/s11203-018-9180-1
17. Galtchouk L., Pergamenshchikov S. (2009) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *Journal of the Korean Statistical Society*. 38(4). pp. 305–322.
18. Pinsker M.S. (1981) Optimal filtration of square integrable signals in Gaussian white noise. *Problems Transimis. information*. 17. pp. 120–133.
19. Nussbaum M. (1985) Spline smoothing in regression models and asymptotic efficiency in L2. *Ann. Statist*. 13. pp. 984–997.

Received: October 20, 2018