Kernel Type Plug-in Estimators of Functions of Multidimensional Density Functionals and Their Derivatives[1]

Kitaeva, Anna
Tomsk Polytechnic University, Institute of Engineering Entrepreneurship
9/4, Usov Street
Tomsk, 634050,Russia
E-mail: kit1157@yandex.ru

Koshkin, Gennady
Tomsk State University, Department of Applied Mathematics and Cybernetics
36, Lenin Street
Tomsk, 634050,Russia
E-mail: kgm@mail.tsu.ru

We consider a unified approach to investigation of plug-in pointwise estimators of the functions of the multidimensional density functionals and the derivatives of the functionals, including the functions of the conditional functionals and their derivatives, based on kernel estimators of the substitution elements.

Let $(X_t)_{t\geq 0}$ be a sequence of random variables satisfying the strong mixing (s.m.) condition with the coefficient $\alpha(\cdot)$ (notation: $(X_t)_{t\geq 0} \in \mathcal{S}(\alpha)$). Time series of such a kind is often used in modelling economic, financial, physical, and technical processes (see, for example, [1,2]).

The problem of identification for the stochastic systems is often connected with estimating of the functions of the following form:

$$H(A) = H\left(\{a_i(x)\},\ \{a_i^{(1j)}(x)\},\ i=\overline{1,s},\ j=\overline{1,m}\right) = H\left(a(x),\ a^{(1j)}(x)\right), \tag{1}$$

where $x \in \mathsf{R}^m$, $H(\cdot)$: $\mathsf{R}^{(m+1)s} \to \mathsf{R}^1$ is a given function, $a^{(1j)}(x) = \left(a_1^{(1j)}(x), \ldots, a_s^{(1j)}(x)\right)$, $a(x) \equiv a^{(0j)}(x) = (a_1(x), \ldots, a_s(x))$. Functionals $a_i(x)$ and their derivatives are defined as follows: $a_i(x) = \int g_i(y) f(x,y)dy$, $a_i^{(1j)}(x) = \dfrac{\partial a_i(x)}{\partial x_j}$, $i=\overline{1,s}$, $j=\overline{1,m}$, where $g_1, \ldots, g_s$ are known functions, $f(\cdot,\cdot)$ is an unknown density function of the observed random vector $Z = (X,Y) \in \mathsf{R}^{m+1}$, Y and $X = (X_1, \ldots, X_m)$ are the output and input variables respectively, $\int \equiv \int\limits_{\mathsf{R}^1}$.

Since for $g_s(y) \equiv 1$, we have $a_s(x) = \int f(x,y)dy = p\,(x)$, where $p\,(\cdot)$ is the density function of X, in the form (1) we can represent any function of the conditional functionals $b_i(x) = a_i(x)/p\,(x) = \int g_i(y) f(y|\,x)dy$ and their derivatives $b_i^{(1j)}(x) = \dfrac{\partial b_i(x)}{\partial x_j}$, $i = \overline{1, s-1}$.

Here are the well known examples of such functions:

— the conditional initial moments

$$\mu_m(x) = \int y^m f(y|x)dy,\ m \geq 1,\ H(a_1,a_2) = a_1/a_2,\ g_1(y) = y^m,\ g_2(y) = 1;$$

$\mu_1(x) = r(x)$ is the regression line;

— the conditional central moments

$$V_m(x) = \int (y - r(x))^m f(y|x)dy,\ g_1(y) = y, g_2(y) = y^2, \ldots,\ g_m(y) = y^m,\ g_{m+1}(y) = 1;$$

[1]This work was supported by the Russian Foundation for Basic Research (project no. 09-08-00595-) and FP "Scientific and scientific-pedagogical cadres of innovative Russia", project no. 102.740.11.5190.

$V_2(x)$ is the conditional variance;

— the conditional coefficient of skewness

$$\beta_1(x) = \frac{\mathsf{E}[(Y-r(x))|x]^3}{V_2(x)^{3/2}},\ b_i = a_i/a_1,\ g_i(y) = y^{i-1},\ H(a_1,a_2,a_3,a_4) = (b_4 - 3b_3b_2 + 2b_2^3)/(b_3 - b_2^2)^{3/2};$$

— the sensitivity functions

$$T_j(x) = \frac{\partial r(x)}{\partial x_j},\ g_1(y) = 1,\ g_2(y) = y,\ H\left(a_1, a_2, a_1^{(1j)}, a_2^{(1j)}\right) = \frac{a_1^{(1j)}}{a_2} - \frac{a_1 a_2^{(1j)}}{a_2^2} = b_2^{(1j)}.$$

Other examples of using expressions of the form (1) can be found in [3].

As the estimators for the functionals $a^{(0j)}(x) \equiv a(x)$ and their derivatives $a^{(1j)}(x)$ at x we take the kernel type statistics

$$a_{n0}^{(rj)}(x) = \frac{1}{nh_n^{m+r}} \sum_{l=1}^{n} g(Y_l)\mathbf{K}^{(rj)}\left(\frac{x - X_l}{h_n}\right),\quad r = 0, 1,$$

where $r = 0, 1$, $Z_l = (X_l, Y_l)$, $l = \overline{1, n}$, is $(m+1)$-dimensional sample characterized by the density $f(\cdot,\cdot)$, $(Z_j)_{j\ge1} \in \mathcal{S}(\alpha)$; $\mathbf{K}^{(0j)}(u) \equiv \mathbf{K}(u) = \prod_{i=1}^{m} K(u_i)$ is an m-dimensional product-form kernel; $\mathbf{K}^{(1j)}(u) = \dfrac{\partial \mathbf{K}(u)}{\partial u_j} = K(u_1)\cdots K(u_{j-1})K^{(1)}(u_j)K(u_{j+1})\cdots K(u_m)$, $K^{(1)}(u_j) = \dfrac{dK(u_j)}{du_j}$; $(h_n) \downarrow 0$ is the number sequence; $g(y) = (g_1(y), \dots, g_s(y))$; $a_{n0}^{(rj)}(x) = \left(a_{1n0}^{(rj)}(x), \dots, a_{sn0}^{(rj)}(x)\right)$, and $a_{n0}^{(0j)}(x) \equiv a_{n0}(x)$.

The recursive estimators are

$$a_n^{(rj)}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{g(Y_i)}{h_i^{m+r}} \mathbf{K}^{(rj)}\left(\frac{x - X_i}{h_i}\right) = a_{n-1}^{(rj)}(x) - \frac{1}{n}\left[a_{n-1}^{(rj)}(x) - \frac{g(Y_n)}{h_n^{m+r}}\mathbf{K}^{(rj)}\left(\frac{x - X_n}{h_n}\right)\right].$$

Recursive procedures have a number of advantages: as a rule, they are easily computer implemented, they are memory-saving, provide a finished result at every step of the algorithm, newly obtained measurements do not lead to cumbersome re-computations. Thus, real-time data processing is possible.

The recursive kernel estimators were first proposed and studied in [4, 5] for a one-dimensional density ($m = 1$, $s = 1$, $g(y) = 1$, and $H(a_1) = a_1$).

Kernel plug-in estimators for the conditional functionals $b(x) = (b_1(x), \dots, b_{s-1}(x))$ at x are of the form

$$b_{n0}(x) = \sum_{i=1}^{n} g(Y_i)\mathbf{K}(x - X_i) \Big/ \sum_{i=1}^{n} \mathbf{K}(x - X_i) = \frac{a_{n0}^{(0j)}(x)}{a_{sn0}^{(0j)}(x)}. \tag{2}$$

and the semi-recursive estimators are

$$b_n(x) = \sum_{i=1}^{n} \frac{g(Y_i)}{h_i^m}\mathbf{K}\left(\frac{x - X_i}{h_i}\right) \Big/ \sum_{i=1}^{n} \frac{1}{h_i^m}\mathbf{K}\left(\frac{x - X_i}{h_i}\right) = \frac{a_n^{(0j)}(x)}{a_{sn}^{(0j)}(x)},\ g(y) = (g_1(y), \dots, g_{s-1}(y)). \tag{3}$$

For $g_1(y) = y$, $g_2(y) = 1$ $(s = 2)$, $m = 1$ we obtain respectively the Nadaraya–Watson kernel estimator [6, 7] for the one-dimensional regression function and the semi-recursive analog of the estimator, which in the case of independent observations was considered in [8—10]. Estimators of this type are said to be semi-recursive, since only the numerator and denominator are computed recursively.

Usual and semi-recursive (the functionals $a_i(x)$ and their derivatives are computed recursively) plug-in estimators for (1) are of the form

(4) $H(A_{n0}) = H\left(\left\{a_{n0}^{(rj)}(x)\right\},\ j=\overline{1,m},\ \ r=0,1\right), H(A_n) = H\left(\left\{a_{n}^{(rj)}(x)\right\},\ j=\overline{1,m},\ \ r=0,1\right).$

Problems of instability of estimators (2)—(3) and, possibly, (4) (depending on the form of a function $H(\cdot)$) will be solved with the help of the piecewise smooth approximations (PWSA) $\widetilde{H}(A_n,\delta_n)$ of estimators (4) [11], similar to Tikhonov's regularization procedure:

$$\widetilde{H}(A_n,\delta_n) = \frac{H(A_n)}{(1+\delta_n|H(A_n)|^{\tau})^{\rho}},$$

where $\tau>0,\ \ \rho>0,\ \ \rho\tau\geq 1,\ \ (\delta_n)\downarrow 0$ as $n\to\infty$.

The PWSA 1) solves the problem of instability of the estimators related to a possible closeness of the denominator to zero, for example, when alternating-sign kernels [12] are used for conditional functionals' estimation (the kernels improve the rate of mean square convergence); and 2) removes obstacles to the main parts of asymptotic mean-square errors (MPAMSE) calculation.

Dependence of the observations makes analysis of properties of the estimators much more complicated: for example, the MPAMSE of the Nadaraya–Watson estimator for s. m. sequences was found only in 1999 [13]; in the same paper convergence of this estimator with probability 1 was also proved.

We find the MPAMSE and the mean-square rate of convergence (improved by a choice of a kernel) of estimators (4) and their PWSA to H(A).

The results for the semi-recursive estimators are given below. Let us introduce necessary definitions and notations.

Denote: $\sup\limits_{x} = \sup\limits_{x\in \mathsf{R}^m},\ \ T_j = \int u^j K(u)du,\ j=1,2,\ldots.$

Definition 1 is related to smoothness conditions for the estimated function $H(\cdot)$, and the next two definitions are related to the estimation procedure.

Definition 1. A function $H(\cdot):\mathsf{R}^s\to\mathsf{R}^1$ belongs to the class $\mathcal{N}_\nu(t)$ $(H(\cdot)\in\mathcal{N}_\nu(t))$ if it and all of its partial derivatives (up to the ν-th inclusive) are continuous at the point $t\in\mathsf{R}^s$. A function $H(\cdot)\in\mathcal{N}_\nu(\mathsf{R})$ if these properties of $H(\cdot)$ are satisfied for all $z\in\mathsf{R}^s$.

Definition 2. A Borel function $K(\cdot)\in\mathcal{A}^{(r)}$, if $\int |K^{(r)}(u)|\,du<\infty$, and $\int K(u)\,du=1$.

Definition 3. A Borel function $K(\cdot)\in\mathcal{A}_\nu^{(r)}$, if $K(\cdot)\in\mathcal{A}^{(r)}$, $T_j=0$, $j=1,\ldots,\nu-1$, $T_\nu\neq 0$, $\int |u^\nu K(u)|du<\infty$, and $K(u)=K(-u)$.

The parameter ν in Definition 3 determines the mean-square convergence rate for estimators (4). For simplicity, we use the notation $\mathcal{A}^{(0)}=\mathcal{A}$, and $\mathcal{A}_\nu^{(0)}=A_\nu$.

Definition 4. A sequence $(h_n)\in\mathcal{H}(\lambda)$, if $\dfrac{1}{n}\sum\limits_{i=1}^{n} h_i^\lambda = S_\lambda h_n^\lambda + o(h_n^\lambda)$, where λ is a real number, S_λ is a constant independent on n.

The condition in Definition 4 is related to the recursive structure of the estimators. Let $f_{1(i+1)(i+j+1)(i+j+k+1)}$ be the density function of the sample variables $(Z_1, Z_{i+1}, Z_{i+j+1}, Z_{i+j+k+1})$,

$$a^{+}_{1(i+1)(i+j+1)(i+j+k+1),t}(x,\,y,x',y') = \int_{\mathsf{R}^4} |g_t(v)g_t(s)g_t(v')g_t(s')|f_{1(i+1)(i+j+1)(i+j+k+1)}(x,v,y,s,x',v',y',s')$$

$dvdsdv'ds',\ \ 1\leq i,j,k<n,\ \ i{+}j{+}k\leq n{-}1;\ \ a^{(2+\delta)+}_{1(i+1),\,t}(x,x') = \int_{\mathsf{R}^2}|g_t(v)g_t(s)|^{2+\delta}f_{1(1+j)}(x,v,s,x')dvds;$

$$a^{(2+\delta)+}_{1(1+j)(1+j+k),\,t}(x,\,y,x') = \int_{\mathsf{R}^3}|g_t(v)g_t(s)g_t(v')|^{2+\delta}f_{1(1+j)(1+j+k)}(x,v,y,s,x',v')dvdsdv'.$$

Introduce also the following notation $r=0,1,\ j=\overline{1,m}\ \ i=\overline{1,s}$:

$$A = A(x) = \left\{a^{(rj)}(x)\right\}; \quad H_{tjr} = \partial H(A)/\partial a_t^{(rj)}; \quad H\left(\left\{a_n^{(rj)}(x)\right\}\right) = H(A_n);$$

$$a^{s+}(x) = \int |g^s(y)| f(x,y)dy; \quad a_{t,p}(x) = \int g_t(y) g_p(y) f(x,y)\,dy; \quad a_{t,p}^{1+}(x) = \int |g_t(y) g_p(y)| f(x,y)\,dy;$$

$$L^{(r,q)} = \int K^{(r)}(u) K^{(q)}(u)\,du; \quad \mathcal{B}_{t,p}^{(r,q)} = L^{(r,q)}\left(L^{(0,0)}\right)^{m-1} a_{t,p}(x); \quad \omega_{i\nu}^{(rj)}(x) = \frac{T_\nu}{\nu!}\sum_{l=1}^{m}\frac{\partial^\nu a_i^{(rj)}(x)}{\partial x_l^\nu};$$

$$\text{the set } Q = \begin{cases} \{0\}, & \forall j \;\; r = 0, \\ \{1\}, & \forall j \;\; r = 1, \\ \{0,1\}, & \exists j \;\; r = 0 \bigwedge r = 1; \end{cases} \qquad \max(r) = \max_{r \in Q}(r); \quad d_n = \left[\frac{1}{nh_n^{m+2\max(r)}} + h_n^{2\nu}\right]^{-1}.$$

Theorem 2 (MSE of the estimator $H(A_n)$). *Assume that for* $t,p = \overline{1,s}, j = \overline{1,m}$ *and* $r \in Q$ *:*

1) $(Z_i)_{i \geq 1} \in \mathcal{S}(\alpha)$, *and* $\int_0^\infty \tau^2 [\alpha(\tau)]^{\frac{\delta}{2+\delta}} d\tau < \infty$, *for some* $0 < \delta < 2$;

2) $a_{t,p}(\cdot) \in \mathcal{N}_0(\mathsf{R})$, $a_t^{(2+\delta)+}(\cdot) \in \mathcal{N}_0(x)$; $\sup_x a_{t,p}^{+}(x) < \infty$, $\sup_x a_t^{\beta+}(x) < \infty$, $\beta = 0, 4$;

3) $K(\cdot) \in \mathcal{A}_\nu^{(r)}$, $\sup_{u \in \mathsf{R}^1} |K^{(r)}(u)| < \infty$; *and if* $1 \in Q$, *then* $\lim_{|u| \to \infty} K(u) = 0$, *for* $m > 1$ $\sup_{u \in \mathsf{R}^1} |K(u)| < \infty$; 4) $a_t^{(rj)}(\cdot) \in \mathcal{N}_\nu(\mathsf{R})$, $\sup_x |a_t^{(rj)}(x)| < \infty$; $\sup_x \left|\frac{\partial^\nu a_t^{(rj)}(x)}{\partial x_l \ldots \partial x_q}\right| < \infty$, $l, \ldots, q = \overline{1,m}$;

5) *A monotone nonincreasing sequence* $(h_n) \in \mathcal{H}(\nu)$, $(h_n) \in \mathcal{H}(-m-2k)$ *for integers* $0 \leq k \leq \max(r)$, $1/(nh_n^{m+2k}) \downarrow 0$; *if* $Q = \{1\}$, *then* $k = 1$;

6) $\sup_x a_{1(i+1)(i+j+1)(i+j+k+1),t}(x,x,x,x) < \infty$, $\sup_x a_{1(i+1)(i+j+1),t}^{(2+\delta)+}(x,x,x) < \infty$, $\sup_x a_{1(i+1),t}^{(2+\delta)+}(x,x) < \infty$, $\sup_{x,y} a_{1(i+1),tp}^{+}(x,y) < \infty$ *for any* $i,j,k \geq 1$; 7) $H(\cdot) \in \mathcal{N}_2(A)$;

8) *For any values of the sample* $Z_1, \ldots, Z_n$ *the sequence* $\{|H(A_n)|\}$ *is majorized by a number sequence* $(C_0 d_n^\gamma)$, $(d_n) \uparrow \infty$, C_0 *being a constant,* $0 \leq \gamma \leq 1/4$.

Then

$$\mathsf{u}^2(H(A_n)) = \sum_{t,p,r,q,j,k} H_{tjr} H_{pkq}\left[S_{-(m+r+q))}\frac{\mathcal{B}_{t,p}^{(r,q)}(x)}{nh_n^{m+r+q}} + S_\nu^2 \omega_{t\nu}^{(rj)}(x)\,\omega_{p\nu}^{(qk)}(x) h_n^{2\nu}\right] + O\left(d_n^{-3/2}\right).$$

Theorem 3 (MSE of the estimator $\widetilde{H}(A_n, \delta_n)$). *Assume that the conditions of Theorem 3 are fulfilled, with condition (8) replaced by the condition* 8^*) $H(A) \neq 0 \quad \tau = 4, 6, \ldots$
Then as $n \to \infty$

$$\mathsf{u}^2\left(\widetilde{H}(A_n, \delta_n)\right) \sim \mathsf{u}^2(H(A_n)).$$

The proof of Theorem 2 is given in [14], based on the results of paper [11]. Theorem 3 follows from [11, Corollary 4] with $k = 2$ and $m = 4$.

Below the results are applied to estimation of the dynamic production function and its characteristics.

Suppose that a sequence $(Y_t)_{t=\ldots,-1,0,1,2,\ldots}$ is generated by a nonlinear homoscedastic ARX process of order (m, s)

$$Y_t = \Psi(Y_{t-i_1}, \ldots, Y_{t-i_m}, X_t) + \xi_t = \Psi(U_t) + \xi_t, \tag{5}$$

where $X_t = (X_{1t}, \ldots, X_{st})$ are exogenous variables, $U_t = (Y_{t-i_1}, \ldots, Y_{t-i_m}, X_t)$, $1 \leq i_1 < i_2 < \ldots < i_m$ is the known subsequence of natural numbers, (ξ_t) is a sequence of independent identically distributed (with density positive on R^1) random variables with zero mean, finite variance, zero third, and finite fourth moments, $\Psi(\cdot)$ is an unknown nonperiodic function bounded on compacts. Assume that the process is strictly stationary. Criteria for geometric ergodicity of a nonlinear heteroscedastic autoregression and ARX models which in turn imply α-mixing have been given by many authors (see, for example, [15]– [19]).

Let $Y_1, \dots, Y_n$ be observations generated by the process (5). The conditional expectation $\Psi(x, z) = \Psi(u) = \mathsf{E}(Y_t | U_t = u)$, $(x, z) = u \in \mathsf{R}^{m+s}$ we estimate by the statistic:

$$\Psi_{n,\, m+s}\,(u) = \sum_{t=i_m+1}^{n} \frac{Y_t}{h_t^{m+s}} \mathbf{K}\left(\frac{u - U_t}{h_t}\right) \Big/ \sum_{t=i_m+1}^{n} \frac{1}{h_t^{m+s}} \mathbf{K}\left(\frac{u - U_t}{h_t}\right). \tag{6}$$

This quantity may be interpreted as the predicted value based on the past information.

We proved that if the observed sequence satisfies the s. m. condition with s. m. coefficient $\alpha(\cdot)$ such that $\displaystyle\int_0^\infty \tau^2 [\alpha(\tau)]^{\frac{\delta}{2+\delta}} d\tau < \infty$ for some $0 < \delta < 2$, then Theorems 2–3 hold. Note that s. m. coefficient with the geometric rate satisfies this condition.

We apply (6) under $U_t = (Y_{t-1}, X_{1(t-1)}, X_{2(t-1)}, X_{3(t-1)})$ to investigate the dependence of Russian Federation's Industrial Production Index Y on the dollar exchange rate X_1, direct investment X_2, and export X_3 for the period from September 1994 till March 2004. The data are available from: http://www.gks.ru/, and http://sophist.hse.ru/. The estimate

$$\Psi_{n,\,4}\left(Y_{n-1}, X_{1(n-1)}, X_{2(n-1)}, X_{3(n-1)}\right) = \sum_{t=2}^{n-1} \frac{K_t}{H_t} Y_t K\left(\frac{Y_{n-1} - Y_{t-1}}{h_{1t}}\right) \Big/ \sum_{t=2}^{n-1} \frac{K_t}{H_t} K\left(\frac{Y_{n-1} - Y_{t-1}}{h_{1t}}\right),$$

where $H_t = \prod_{j=1}^{4} h_{jt}, \quad K_t = \prod_{j=1}^{3} K\left(\dfrac{X_{j(n-1)} - X_{j(t-1)}}{h_{(j+1)t}}\right)$.

To find the AMSE of the estimate $\Psi_{n,\,4}\,(u)$ we use Theorem 2.

Let $f(\cdot, \cdot)$ be the stationary distribution of the vector (U_t, Y_t). Suppose that $K(\cdot) \in \mathcal{A}_\nu$, $\mathbf{K}(u) = \prod_{i=1}^{4} K(u_i)$, $\sup_{u \in \mathsf{R}^1} |K(u)| < \infty$, the sequence $(h_n) \in \mathcal{H}(4)$, and $1/(nh_n^4) \downarrow 0$. Let functions $a_i(u)$, $i = 0, 1$, and their derivatives up to and including the order ν be continuous and bounded on R^4; functions $\displaystyle\int y^2 f(u, y)\, dy$ and $\displaystyle\int y^4 f(u, y)\, dy$ be bounded on R^4; and, moreover, $\displaystyle\int y^2 f(u, y)\, dy$ and $\displaystyle\int |y|^{2+\delta} f(u, y)\, dy$ be continuous at the point u. Then conditions (1)–(5) of Theorem 2 hold; we also suppose that condition (6) holds. If $p\,(u) > 0$, then condition (7) holds too.

If the random variables Y_t are uniformly bounded, and we select a nonnegative kernel, then it is easy to show that $\Psi_{n,\,4}\,(u)$ are bounded for $\nu = 2$. By condition (8) this is equivalent to the existence of a majorizing sequence with $\gamma = 0$. For $\nu > 2$ the PWSA solves the problem.

In Table 1 the relative errors of the forecast (REF) obtained with $\Psi_{n,\,4}\,(\cdot)$ for each year from 1995 till 2004 are given.

The result of 1998 can be explained by Russian financial crisis ("Ruble crisis") in August 1998. The kernel used is the Gaussian kernel and the bandwidths $h_{jt} = 0.17\hat{\sigma}_j t^{-1/8}$, where $\hat{\sigma}_j$, $j = 1, 2, 3, 4$ are the corresponding sample mean square deviations, the constant 0.17 is chosen subjectively.

The marginal productivity function and marginal rate of technical substitution are estimated in the same way.

TABLE 1. Errors of Forecasts

1995	1996	1997	1998	1999
0.189	0.057	0.04	0.133	0.05

2000	2001	2002	2003	2004
0.055	0.047	0.042	0.057	0.043

REFERENCES

1. Härdle, W., Tsybakov, A., and Yang, L., Nonparametric vector autoregression, Journal of Statistical Planning and Inference, 1998, vol. 68, pp. 221—245.

2. Chen, G., Choi, Y.K., and Zhou, Y., Nonparametric estimation of structural change points in volatility models for time series, Journal of Econometrics, 2005, vol. 126, pp. 79—114.

3. Kitaeva A. V., and Koshkin G. M., 'Recurrent Nonparametric Estimation of Functions from Functionals of Multidimensional Density and Their Derivatives, Autom. Remote Control, 2009, no. 3, pp. 389–407.

4. Wolverton, C. T., and Wagner, T. J., Recursive estimates of probability densities, IEEE Trans. Syst. Sci. and Cybernet., 1969, vol. 5, no. 3, pp. 246—247.

5. Banon, G., Sur un estimateur non parametrique de la densite de probabilite, Rev. Statist. Appl., 1976, vol. 24, no. 4, pp. 61–73.

6. Nadaraya, E.A., On Estimation of Regression, Teor. Veroyatnost. i Primenen., 1964, vol. 9, no. 1, pp. 157–159.

7. Watson, G.S., Smooth Regression Analysis, Sankhya, Ser. A, 1964, vol. 26, pp. 359–372.

8. Ahmad, J.A., and Lin, P.E., Nonparametric Sequential Estimation of a Multiple Regression Function, Bull. Math. Statist., 1976, vol. 17, no. 12, pp. 63–75.

9. Buldakov, V.M., and Koshkin, G.M., On Recursive Estimates of Probability Density and Regression Line, Probl. Peredachi Inf., 1977, vol. 13, no. 1, pp. 58–66 [Probl. Inf. Trans. (Engl. Transl.), 1977, vol. 13, no. 1, pp. 41–48].

10. Devroye, L.P., and Wagner, T.J., On the L1 Convergence of Kernel Estimator of Regression Functions with Applications in Discrimination, Z. Wahrsch. Verw. Gebiete, 1980, vol. 51, no. 1, pp. 15–25.

11. Koshkin, G.M., Deviation Moments of the Substitution Estimator and of Its Piecewise-Smooth Approximations, Sibirsk. Mat. Zh., 1999, vol. 40, no. 3, pp. 605–618 [Siberian Math. J. (Engl. Transl.), 1999, vol. 40, no. 3, pp. 515–527].

12. Gasser, T., and Muller, H.-G., Kernel Estimation of Regression Functions, Lecture Notes Math., 1979, vol. 757, pp. 23—68.

13. Bosq, D., and Cheze-Payaud, N., Optimal Asymptotic Quadratic Error of Nonparametric Regression Function Estimates for a Continuous-Time Process from Sampled-Data, Statistics, 1999, vol. 32, no. 3, pp. 229–247.

14. Kitaeva, A.V., and Koshkin G.M., Nonparametric Semirecursive Identification in a Wide Sense of Strong Mixing Processes, Problems of Information Transmission, 2010, vol. 46, no. 1, pp. 22—37.

15. Masry, E., and Tjøstheim, D., Nonparametric estimation and identification of nonlinear ARCH time series, Econom. Theory, 1995, vol. 11, pp. 258—289.

16. Lu, Z. D., On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term, Statist. Sinica, 1998, vol. 8, pp. 1205–1217.

17. Lu, Z. D., and Jiang, Z., L1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term, Statist. Probab. Lett., 2001, vol. 51, pp. 121--130.

18. Doukhan, P., and Tsybakov, A., Nonparametric robust estimation in nonlinear ARX models, Problems of Information Transmission, 1993, vol. 29, no. 4, pp. 318--327.

19. Doukhan, P., Mixing: properties and examples. Lecture Notes in Statistics, 1994, vol. 85, Springer-Verlag.