# SEQUENTIAL KERNEL ESTIMATION OF A MULTIVARIATE REGRESSION FUNCTION

D.N. Politis
*University of California, San Diego*
USA, San Diego, La Jolla, CA 92093-0112
E-mail: dpolitis@ucsd.edu

V.A. Vasiliev[1]
*Tomsk State University*
Russia, 634050, Tomsk, Lenin Ave., 36
E-mail: vas@mail.tsu.ru

**Key words:** Kernel regression estimation; dependent observations; sequential approach; guaranteed mean square accuracy; finite sample size

This paper presents a sequential estimation procedure for an unknown multivariate regression function. Observed regressors and noises of the model are supposed to be dependent and form sequences of dependent vectors and numbers respectively. Two types of estimators are considered. Both estimators are constructed on the basis of Nadaraya–Watson kernel estimators.
First, sequential estimators with given bias and mean square error are defined. According to the sequential approach the duration of observations is a special stopping time. Then on the basis of these estimators, truncated sequential estimators of a regression function are constructed on a time interval of a fixed length. At the same time the variance of these estimators is also bounded by a non-asymptotic bound.
Together with finite-sample, asymptotic properties of the presented estimators are investigated. It is shown, in particular, that by the appropriately chosen bandwidths both estimators have optimal (as compared to the case of independent data) rates of convergence.

## 1. Introduction

In many problems of identification the solution is reduced to finding some statistics in the form of a ratio of some random functions as estimators (e.g., ratio estimators). Such situation arises in the regression estimation problem by use of the Nadaraya–Watson estimators. There are a lot of results and publications dedicated this problem for independent and dependent observations, in asymptotic and non-asymptotic problem statements (see, e.g, [1], [3]-[6], [8], [9], [16], [20]-[27], [30] etc.) It should be noted, that often regressors $X$ (see (1) below) are supposed to be non-random or bounded and noises $\Delta$ of the model are independent [1, 4, 8, 17], [22]-[26] etc.

If, in general, the denominator of such ratio may be small enough with positive probability, then the functional is no longer finite and one has, e.g., to use some

[1]Research was supported by RFBR Grant N 09-01-00172

variants of Cramer's theorem [3] to find even the first moment of such estimator. Furthermore it is not trivial to find dominant sequences for moments of ratio estimators (see, e.g., [30]).

The functionals of this type were unified in [30] (see also references therein) into the class of functionals with singularities and special estimation procedures for them, called piecewise smooth approximation, were considered. The method proposed in [30] gives a possibility to get estimators based on dependent observations with known principal term of the mean square error (MSE) and known asymptotic properties (when the sample size unboundedly increases).

However in practice the observation time of a system is always finite which does not allow us to judge the quality of such estimators. One of the possibilities for finding estimators with a guaranteed quality of inference is provided by the stand-point of sequential analysis; this approach presumes a special choice of the time for stopping observations with the help of some functional of the process under observation. The principle of sequential analysis has been primarily proposed by A.Wald for a scheme of independent observations [33]; later, the sequential approach was applied to parameter estimation problem of one-parameter stochastic differential equations in [15, 18] and for multiparameter continuous and discrete-time dynamic systems in many papers and books (see [2], [7], [10]-[14], [28]-[32] among others). Sequential approach has also been applied to non-parametric regression and density function estimation problems as well (see, e.g., [4]-[6], [9], [23, 28, 30, 32] and references in [6]). Peculiarities and difficulties of sequential approach to nonparametric problems are discussed in detail by Efroimovich in [6].

In this paper (see Section 3) we solve the problem of estimation of a regression function with a prescribed statistical quality using a sequential approach, which supposes unboundedness of a sample size. In Section 4 we consider the most realistic problem statement, where the observation time of a system is not only finite but bounded. One of the possibilities for finding estimators with the guaranteed quality of inference from a sample of fixed size is provided by the approach of truncated sequential estimation.

Truncated sequential estimation method was developed by Konev and Pergamenchtchikov [11]-[13], as well as in [7] for parameter estimation problems in discrete and continuous time dynamic models. Using sequential approach, they have constructed estimators of dynamic systems parameters with known variance from samples of fixed size.

Non-parametric truncated sequential estimators of a regression function presented in this paper constitute Nadaraya–Watson estimators calculated at a special stopping time. These estimators have known mean square errors as well. The duration of observations is also random but bounded from above by a non-random fixed number.

Non-asymptotic and asymptotic properties of estimators of both types are investigated. It is shown, that truncated sequential estimators coincide with the afore-mentioned sequential estimators (i.e. with the Nadaraya–Watson estimators calculated at a special stopping time) for sufficiently large sample size.

The assumption on the independence of observations is often not satisfied in practice and is an essential restriction. In this paper the sequential analysis approach is employed in the procedures of non-parametric estimation of a regression function from dependent observations. In particular, model inputs and noises can be dependent and unbounded with a positive probability. The example of estimation of a nonlinear autoregressive function is considered.

## 2. Problem statement

This paper considers the problem of non-parametric kernel estimation of a regression function $f(x)$ with multivariate argument $x \in \mathbb{R}^m$, $m \geq 1$ at a point $x = x^0$ satisfying the equation

$$(1) \qquad Y_i = f(x_i) + \delta_i, \quad i \geq 1,$$

in the case of dependent regressors $X = (x_i)_{i \geq 1}$ and noises $\Delta = (\delta_i)_{i \geq 1}$. It is supposed, that pairs $(Y_i, x_i)$, $i = 1, 2, \ldots$ are under observations but the errors $\delta_i$ are unobservable. We allow for general dependence conditions on random variables $(x_i)$ and $(\delta_i)$ supposing mutual dependence of the processes $X$ and $\Delta$. In particular, model inputs can be unbounded with a positive probability.

The main goal of the paper is twofold. We will construct sequential and truncated sequential estimators. Estimators of both types have known upper bounds of mean square errors. Non-parametric truncated sequential estimators of a regression function presented in this paper are constitute the Nadaraya–Watson estimators calculated at a special stopping time. These estimators have known upper bound of mean square errors as well. The duration of observations is also random but bounded from above to a non-random fixed number.

Now we give some needed notation and definitions. The notation $\overline{1, m}$ means the set of integers $1, 2, \ldots, m$, and $\chi(A)$ is the indicator function of set $A$.

For a fixed vector of nonnegative integers $a = (\alpha_1, \ldots, \alpha_m)$, we consider the partial derivative

$$f_a^{(\alpha)}(x) = \frac{\partial^\alpha f(x)}{\partial x_1^{\alpha_1} \ldots \partial x_m^{\alpha_m}}, \qquad f_0^{(0)}(x) = f(x)$$

of a function $f(x)$ at a given point $x \in \mathcal{R}^m$, where $\alpha_1 + \alpha_2 + \ldots + \alpha_m = \alpha$.

Denote by $\beta(k)$ the set of all vectors $b = (\beta_1, \ldots, \beta_m)$ with nonnegative integer-valued components $\beta_1, \ldots, \beta_m$ such that $\beta_1 + \ldots + \beta_m = k$. Omitting the subscript $b = (\beta_1, \ldots, \beta_m)$ of partial derivatives $f_b^{(k)}(x)$ will mean that the set of indices $\beta_1, \ldots, \beta_m$ is not specified.

In the sequel, we will use the following notation: $\sum\limits_{\beta(k)}$ means the summation over all $b \in \beta(k)$; we write $x^k = x_1^{\beta_1} \cdot \ldots \cdot x_m^{\beta_m}$ for vector $x = (x_1, \ldots, x_m)'$ and $|k|! = \beta_1! \cdot \ldots \cdot \beta_m!$, where $k = \beta_1 + \ldots + \beta_m$.

We suppose that the function to be estimated satisfies the following assumption

**Assumption 1.** *The function $f(x)$ is $p$-th differentiable at a point $x^0$, $p \geq 0$ and for some positive numbers $C_{f,i}$, $i = \overline{0, p}$ $(p \geq 0)$, $L_f$, $\gamma_f \in (0, 1]$, such that*

$$|f^{(i)}(x^0)| \leq C_{f,i}, \; i = \overline{0, p}$$

*and for every* $x, y \in \mathbb{R}^m$,

$$|f^{(p)}(x) - f^{(p)}(y)| \le L_f ||x - y||^{\gamma_f}, \quad \|x\|^2 = \sum_{j=1}^{m} x_j^2.$$

For estimation of the function $f(\cdot)$ at a point $x^0$, $f = f(x^0)$ we use so-called sequential estimators, constructed on the basis of kernel Nadaraya–Watson estimators of the form

$$\hat{f}_N = \left( \sum_{i=1}^{N} K\left(\frac{x^0 - x_i}{h}\right)\right)^{+} \cdot \sum_{i=1}^{N} Y_i K\left(\frac{x^0 - x_i}{h}\right), \quad N \ge 1, \tag{2}$$

where $a^+ = a^{-1}$ for $a \ne 0$; $a^+ = 0$ when $a = 0$; $K(\cdot)$ is a kernel function and $h$ is a bandwidth, satisfying the Assumptions 4 and 5 below respectively.

Define for every $i \ge 1$ density functions $f_{i-1}(t)$ of conditional distribution functions $F_{i-1}(t) = P(x_i \le t | \mathcal{F}_{i-1}^x)$, $\mathcal{F}_i^x = \sigma\{x_1, \ldots, x_i\}$.

As regards to the regressors $X$ and $\Delta$ we suppose

**Assumption 2.** *Assume the conditional density functions* $f_{i-1}(t)$, $i \ge 1$ *are* $q$*-th differentiable at a point* $x^0$, $q \ge 0$ *and for some positive numbers* $c_x^0$, $C_{x,k}$, $k = \overline{0, q}$ $(q \ge 0)$, $L_x$ *and* $\gamma_x \in (0, 1]$ *the following relations*

$$\inf_{i \ge 1} f_{i-1}(x^0) \ge c_x^0,$$

$$\sup_{i \ge 1} |f_{i-1}^{(k)}(x^0)| \le C_{x,k}, \ \ k = \overline{0, q}$$

*and for every* $x, y \in \mathbb{R}^m$,

$$\sup_{i \ge 1} |f_{i-1}^{(q)}(x) - f_{i-1}^{(q)}(y)| \le L_x ||x - y||^{\gamma_x}$$

*are fulfilled.*

**Assumption 3.** *Let the noises* $\Delta$ *be conditionally zero mean* $E(\delta_i | \mathcal{F}_i^x) = 0$, $i \ge 1$ *and for some monotonically non-increasing as* $k \to \infty$ *non-negative function* $\varphi(k)$, $k \ge 1$ *the following "general mixing condition" for the sequence* $\Delta$ *holds true*

$$|E(\delta_i \delta_j | \mathcal{F}_{i \vee j}^x)| \le \varphi(|i - j|), \quad i, j \ge 1.$$

We now define admissible sets of kernels $K(\cdot)$ and bandwidths $h$ as follows.

Let

$$T_j = \int_{\mathcal{R}^m} u_1^{\alpha_1} \ldots u_m^{\alpha_m} K(u)\, du, \qquad \sum_{i=1}^{m} \alpha_i = j,$$

where $\alpha_1, \ldots, \alpha_m$ are nonnegative integers.

**Assumption 4.** *Let the function $K(\cdot)$ is uniformly bounded from above:*

$$\sup_z |K(z)| \le \overline{K} < \infty$$

*and such that*

$$\int K(z)dz = 1, \quad \int |z^s K(z)|dz < \infty, \quad 0 \le s \le q + 2p + 3.$$

*Moreover, we suppose that the integrals $T_j = 0$, $j = \overline{1, p+q}$.*

**Assumption 5.** *Let $(h_n)_{n\ge1}$ be a sequence of monotonically decreasing to zero positive numbers such that $nh_n^m \to \infty$ as $n \to \infty$ and $h_1 \le 1$.*

It should be noted that, according to Assumption 4, the kernel $K(\cdot)$ does not have to be nonnegative. For example, the infinite-order flat-top kernels of Politis [19] are allowed, including the infinitely differentiable flat-top kernel of McMurry and Politis [16].

## 3. Sequential estimation of the regression function f

In this section we consider sequential estimators of the regression function $f$ at a point $x^0$ with a given variance, calculated at a special stopping time. These estimators are constructed on the basis of Nadaraya-Watson kernel estimators. Bandwidth giving an optimal convergence rate of estimators (which coincides with the rate of Nadaraya-Watson estimators constructed by independent observations) is found.

According to sequential approach we introduce the following definition.

**Definition 1.** *Let $(H_n)_{n\ge1}$ be an unboundedly monotonically increasing given sequence of positive numbers and $(h_n)_{n\ge1}$ satisfies Assumption 5.*

*The sequential plans $(\tau_n,\ f_n^*)$, $n \ge 1$ of estimation of the function $f = f(x^0)$ will be defined by the formulae*

$$\tau_n = \inf\{j \ge 1 : \ \sum_{i=1}^{j} K\left(\frac{x^0 - x_i}{h_n}\right) \ge H_n\},$$

$$f_n^* = \frac{1}{Q_n}\sum_{i=1}^{\tau_n} Y_i K\left(\frac{x^0 - x_i}{h_n}\right), \quad Q_n = \sum_{i=1}^{\tau_n} K\left(\frac{x^0 - x_i}{h_n}\right), \tag{3}$$

*where $\tau_n$ is the duration of estimation, and $f_n^*$ is the estimator of $f$ with given accuracy in the mean square sense.*

We have introduced in Definition 1 the sequence of sequential estimation plans. At the same time Theorem 1 below gives non-asymptotic properties of these plans for every $n \ge 1$, in particular, upper bound for the MSE of the estimator $f_n^*$. It gives a possibility to investigate asymptotic properties of defined plans as well (see Section 3.1.).

To formulate this theorem we need the following notation.

Assume $(l_n)_{n\geq 1}$ to be an arbitrary non-decreasing sequence of positive numbers.

Define the function $\gamma_{x,\delta} = 0$ if processes $X$ and $\Delta$ are independent and 1 otherwise; $\gamma_\delta = 0$ if $\varphi(k) = 0$, $k > 0$ and 1 otherwise; the numbers

$$\varphi_n = \varphi(l_n), \quad K_\alpha(\gamma) = \int |z^\alpha| \cdot ||z||^\gamma \cdot |K(z)| dz, \quad \alpha \geq 0, \quad \gamma \geq 0, \quad K_0 = K_0(0),$$

$$\kappa = \frac{L_x}{|q|!} \sum_{\beta(q)} K_q(\gamma_x), \quad C_1 = \sum_{j=1}^{p} \sum_{\beta(q+j)} \frac{1}{|q+j|!} C_{f,j} K_{q+j}(\gamma_x) \chi(p>0),$$

$$C_2 = \frac{1}{|p|!} C_{x,0} \sum_{\beta(p)} K_p(\gamma_f), \quad C_3 = C_{x,0} \int K^2(y)\, dy,$$

$$C_4 = 4\overline{K}(p+2) \sum_{j=0}^{q+1} \sum_{k=0}^{p+1} \sum_{\beta(j+2k)} \frac{1}{|j|!(|k|!)^2} C_{x,j} C_{f,k}^2 K_{\nu_{j,k}}(\mu_{j,k}),$$

$$C_{x,q+1} = \frac{1}{|q|!} L_x, \quad C_{f,p+1} = \frac{1}{|p|!} L_f,$$

$$\nu_{j,k} = j+2k-\chi(j=q+1)-\chi(k=p+1), \quad \mu_{j,k} = \gamma_x \chi(j=q+1) + 2\gamma_f \chi(k=p+1),$$

as well as

$$n_0 = \inf\{n \geq 1 : \ c_x^0 - \kappa h_n^{q+\gamma_x} > 0\},$$

$$t_n = \frac{(C_{x,0} + \kappa h_n^{q+\gamma_x}) \cdot \sqrt{C_3(H_n + \overline{K})}}{(c_x^0 - \kappa h_n^{q+\gamma_x})^{5/2} h_n^m} +$$

$$+ \frac{\sqrt{[C_3(C_{x,0} + \kappa h_n^{q+\gamma_x})^2 \cdot (c_x^0 - \kappa h_n^{q+\gamma_x})^{-3} + 2\overline{K}^2 (c_x^0 - \kappa h_n^{q+\gamma_x})^{-1}](H_n + \overline{K}) + H_n^2}}{(c_x^0 - \kappa h_n^{q+\gamma_x}) h_n^m},$$

$$w_n = \sqrt{\begin{array}{l}\varphi(0)\overline{K}^2 \gamma_\delta \dfrac{l_n}{H_n^2}(1+3l_n) + C_3 \varphi(0)(c_x^0 - \kappa h_n^{q+\gamma_x})^{-1} \dfrac{1}{H_n}(1 + \dfrac{\overline{K}}{H_n}) + \\ +2\gamma_\delta \varphi(0) K_0 \overline{K} (c_x^0 - \kappa h_n^{q+\gamma_x})^{-1} \dfrac{l_n}{H_n}(1 + \dfrac{\overline{K}}{H_n}) + 2\gamma_\delta K_0 \overline{K} \varphi_n \dfrac{h_n^m t_n^2}{H_n^2}\end{array}},$$

$$b_n = (c_x^0 - \kappa h_n^{q+\gamma_x})^{-1} [C_1 h_n^{q+1+\gamma_x} + C_2 h_n^{p+\gamma_f}] \left(1 + \frac{\overline{K}}{H_n}\right) + w_n$$

and

$$V_n = 2\left\{(2C_3 C_{f,0}^2 + (1+\gamma_{x,\delta}) C_4)(c_x^0 - \kappa h_n^{q+\gamma_x})^{-1} \frac{1}{H_n}(1 + \frac{\overline{K}}{H_n}) + \right.$$
$$\left. +(1+\gamma_{x,\delta}) w_n^2 + \frac{2(C_1 h_n^{q+1+\gamma_x} + C_2 h_n^{p+\gamma_f})^2 h_n^{2m} t_n^2}{H_n^2}\right\}.$$

**Theorem 1.** *Under model (1), let the Assumptions 1-5 be fulfilled and* $n \geq n_0$*. Then the sequential estimation plans* $(\tau_n, f_n^*)$ *are closed, i.e.* $\tau_n < \infty$ *a.s. and have the following non-asymptotic properties:*

*1. for the expected duration of observations*

*a)*

$$(C_{x,0} + \kappa h_n^{q+\gamma_x})^{-1} h_n^{-m} H_n \leq E\tau_n \leq (c_x^0 - \kappa h_n^{q+\gamma_x})^{-1} h_n^{-m}(H_n + \overline{K}), \tag{4}$$

*b)*

$$E\tau_n^2 \leq t_n^2; \tag{5}$$

*2. for the bias*

$$|Ef_n^* - f| \leq b_n;$$

*3. for the MSE*

$$E(f_n^* - f)^2 \leq V_n.$$

**Remark 1.** *From Definition 1 it follows that the presented sequential estimators coincide with the usual Nadaraya–Watson estimators (2) calculated at a special stopping time. It follows that, at least in the case of independent inputs of the model, these estimators have the same asymptotic properties as Nadaraya–Watson estimators (see Section 3.1.). However, as shown in Theorem 1, sequential estimators have the above exact, non-asymptotic properties, that may be important for practitioners.*

### 3.1. Bandwidth choice for sequential estimators

From Theorem 1 follows, that it is natural to define $H_n = nh_n^m$. In this case and under the condition on the bandwidth $nh_n^m \to \infty$ as $n \to \infty$ from Assumption 5 the mean of the duration of observations is proportional, according to (4) to $n$, i.e.,

$$C_{x,0}^{-1} n - \frac{C_{x,0}^{-1}\kappa h_n^{q+\gamma_x} n}{C_{x,0} + \kappa h_n^{q+\gamma_x}} \leq E\tau_n \leq$$

$$\leq (c_x^0)^{-1} n + \frac{(c_x^0)^{-1}\kappa h_n^{q+\gamma_x} n + \overline{K} h_n^{-m}}{c_x^0 - \kappa h_n^{q+\gamma_x}}, \quad n \geq n_0,$$

as well as, from (5) it follows,

$$E\tau_n^2 \leq (c_x^0)^{-2} n^2 + o(n^2) \quad \text{as} \quad n \to \infty$$

and

$$|Ef_n^* - f| = O\left(h_n^{\varrho} + \sqrt{\varphi_n h_n^{-m}} + \frac{1 + \gamma_\delta \sqrt{l_n}}{\sqrt{nh_n^m}}\right), \quad \varrho = (q + 1 + \gamma_x) \wedge (p + \gamma_f),$$

$$E(f_n^* - f)^2 = O\left(h_n^{2\varrho} + \varphi_n h_n^{-m} + \frac{1 + \gamma_\delta l_n}{nh_n^m}\right) \quad \text{as } n \to \infty.$$

From these formulae it follows, that for the asymptotic unbiasnessy and $L_2$-convergency of $f_n^*$ we have to use the sequences $(l_n)$ and $(h_n)$ satisfying the condition

$$\varphi_n h_n^{-m} + \frac{l_n}{nh_n} = o(1) \quad \text{as } n \to \infty. \tag{6}$$

**Remark 2.** *If the regressors $x_i$ form a sequence of i.i.d.r.v's, then $c_x^0 = C_{x,0}$. Moreover if the number $c_x^0 = f_i(x^0)$ is known, we can put $H_n = c_x^0 n h_n^m$. In this case*

$$n - \frac{\kappa h_n^{q+\gamma_x} n}{c_x^0 + \kappa h_n^{q+\gamma_x}} \leq E\tau_n \leq n + \frac{\kappa h_n^{q+\gamma_x} n + \overline{K} h_n^{-m}}{c_x^0 - \kappa h_n^{q+\gamma_x}}$$

*and*

$$\overline{\lim_{n\to\infty}}\, [(h_n^{q+\gamma_x} n)^{-1} \wedge h_n^m] |E\tau_n - n| < \infty,$$

*where $a \wedge b = \min\{a, b\}$, as well as*

$$E\tau_n^2 \leq n^2 + o(n^2) \quad as \quad n \to \infty.$$

It should be noted that in the case of i.i.d. noises $(\delta_i)$ and by $q \geq p$ the MSE of the constructed sequential estimator has similar to the i.i.d. or non-random case for the regressors $(x_i)$ optimal decreasing rate

$$E(f_n^* - f)^2 = O\left(h_n^{2(p+\gamma_f)} + \frac{1}{nh_n^m}\right) \quad \text{as } n \to \infty.$$

In particular, for the case $p = 0$, $\gamma_f = 1$ (see, for comparison, [8] among others),

$$E(f_n^* - f)^2 = O\left(h_n^2 + \frac{1}{nh_n^m}\right) \quad \text{as } n \to \infty.$$

# 4. Truncated sequential estimation of the regression function f

We shall consider in this section the problem of estimation of the regression function $f$ with a known mean square accuracy based on observations of the process $(Y_i, x_i)$ for $i = 1, \ldots, N$ on the time interval $[1, N]$ for a fixed time $N$.

Such possibility gives the truncated sequential estimation method, developed by Konev and Pergamenchtchikov [11]-[13], as well as in [7] for parameter estimation problems in discrete and continuous time dynamic systems.

**Definition 2.** *Let $H$ and $h$ be positive numbers.*

*The truncated sequential plans $(\tau_N(h, H),\ f_N^*(h, H))$, $N \geq 1$ of estimation of the function $f = f(x^0)$ will be defined by the formulae*

$$\tau_N(h, H) = \inf\{j \in [1, N] :\ \sum_{i=1}^{j} K\left(\frac{x^0 - x_i}{h}\right) \geq H\}$$

*with the provision that* $\inf\{\emptyset\} = N$,

$$f_N^*(h,H) = \hat{f}_N(h,H)\cdot\chi\left(\sum_{i=1}^{N} K\left(\frac{x^0 - x_i}{h}\right) \ge H\right) +$$

$$+\frac{C_{f,0}}{2}\cdot\chi\left(\sum_{i=1}^{N} K\left(\frac{x^0 - x_i}{h}\right) < H\right),$$

$$\hat{f}_N(h,H) = \frac{1}{Q_N(h,H)}\sum_{i=1}^{\tau_N(h,H)} Y_i K\left(\frac{x^0 - x_i}{h}\right), \quad Q_N(h,H) = \sum_{i=1}^{\tau_N(h,H)} K\left(\frac{x^0 - x_i}{h}\right),$$

*where* $\tau_N(h,H)$ *is the duration of estimation, which is bounded by construction* ($\tau_N(h,H) \le N$) *and* $f_N^*(h,H)$ *is the estimator of* $f$.

The parameters $H$ and $h$ will be defined in a special way as functions of $N$ for optimization of a convergence rate of the estimator $f_N^*(h,H)$ as $N \to \infty$ in the mean square sense.

Assume $l(H)$ to be an arbitrary non-decreasing function of positive numbers. Define

$$V_N(h,H) = 2\left\{(2C_3C_{f,0}^2 + (1+\gamma_{x,\delta})C_4)\frac{h^m N}{H^2} + (1+\gamma_{x,\delta})[2\gamma_\delta K_0\overline{K}\varphi(l(H))\frac{h^m N^2}{H^2} +\right.$$

$$+C_3\varphi(0)\frac{h^m N}{H^2} + 2\gamma_\delta\varphi(0)K_0\frac{l(H)h^m N}{H^2} + \varphi(0)\overline{K}^2\gamma_\delta\frac{l(H)}{H^2}(1+3l(H))]+$$

$$\left.+\frac{2(C_1h^{q+1+\gamma_x} + C_2h^{p+\gamma_f})^2h^{2m}N^2}{H^2}\right\} + \frac{C_3C_{f,0}^2 N h^m}{4[Nh^m(c_x^0 - \kappa h^{q+\gamma_x}) - H]^2}.$$

From the proof of the following Theorem 2 it follows that the number $H$ satisfies the condition

$$Nh^m(c_x^0 - \kappa h^{q+\gamma_x}) - H > 0. \tag{7}$$

Thus it is natural to put

$$H := H_N = Nh_N^m C_\alpha^{-1}, \tag{8}$$

where

$$C_\alpha = (c_x^0 - \kappa h_{n_0}^{q+\gamma_x} - \alpha^{-1/2})^{-1}, \quad \alpha > (c_x^0 - \kappa h_{n_0}^{q+\gamma_x})^{-2}$$

and $(h_N)$ is a sequence satisfying Assumption 5. It should be noted that, according to the assumptions of the theorem below the number $C_\alpha$ and the sequence $(H_N)$ are assumed known.

Then, denoting $l_N = l(H_N)$ and $\varphi_N = \varphi(l_N)$, we have

$$V_N^0 := V_N(h_N,H_N) = 2\left\{[(2C_3C_{f,0}^2 + (1+\gamma_{x,\delta})C_4)C_\alpha^2 + \frac{C_3C_{f,0}^2\alpha}{8} + (1+\gamma_{x,\delta})\cdot\right.$$

$$\cdot(C_3C_\alpha^2\varphi(0) + 2\gamma_\delta\varphi(0)K_0C_\alpha^2 l_N)]\frac{1}{Nh_N^m} + (1+\gamma_{x,\delta})[\varphi(0)\overline{K}^2C_\alpha^2\gamma_\delta\frac{l_N}{(Nh_N^m)^2}(1+3l_N)+$$

$$\left.+2\gamma_\delta K_0\overline{K}C_\alpha^2\varphi_N h_N^{-m}] + 2C_\alpha^2(C_1h_N^{q+1+\gamma_x} + C_2h_N^{p+\gamma_f})^2\right\}.$$

The main result of this section is the following

**Theorem 2.** *Under model (1), let the Assumptions 1-4 be fulfilled, where the number $C_{f,0}$ in Assumption 1 is supposed to be known. Then:*

*1) for every positive numbers $h$, $H$ satisfying the condition (7) the truncated sequential estimator $f^*_N(h,H)$ has the MSE:*

$$E(f^*_N(h,H)-f)^2 \le V_N(h,H);$$

*2) for $H=(H_N)$ defined in (8) and $h=(h_N)$ satisfying (7) it holds*

$$E(f^*_N(h_N,H_N)-f)^2 \le V^0_N;$$

*3) assume there exists a number $r>2$, such that the bandwidth $h=(h_N)_{N\ge 1}$ from Assumption 5 satisfies the additional condition*

$$\sum_{N\ge 1} \frac{1}{h_N^{m(r-1)} N^{r/2}} < \infty \tag{9}$$

*and $H=(H_N)$ is defined in (8). Then*

$$\lim_{N\to\infty} \frac{\tau_N(h_N,H_N)}{N} < 1 \quad a.s.$$

*and the estimator $f^*_N(h_N,H_N)$ is non-degenerate as $N\to\infty$ in the following sense:*

$$P(f^*_N(h_N,H_N)=\hat f_N(h_N,H_N))=1 \quad \text{for } N \text{ large enough.}$$

**Remark 3.** *From the third assertion of Theorem 2 it follows, that similarly to sequential estimators, the truncated sequential estimators have the asymptotic properties of Nadaraya–Watson estimators (2). However, in distinction from the sequential estimators considered in Section 3., the truncated estimators have known variance for fixed sample sizes.*

**Remark 4.** *If the number $C_{f,0}$ in Assumption 1 is unknown, then the Definition 2.2 of sequential estimator is modified as follows:*

$$f^*_N(h,H)=\hat f_N(h,H)\cdot \chi\left(\sum_{i=1}^N K\left(\frac{x^0-x_i}{h}\right)\ge H\right).$$

*All assertions of Theorem 2 will be fulfilled with slightly changed (but having similar structure) functions $V_N(h,H)$ and $V^0_N$. More precisely, the number 4 in the denominator of the last summand in the definition of $V_N(h,H)$ should be omitted and the number 8 in the denominator of the third summand in the right hand side in the definition of $V^0_N$ should be changed to the number 2.*

## 4.1. Bandwidth choice for truncated sequential estimators

From Theorem 2 it follows, that by the definition of $V^0_N$ and assertion 2) of Theorem 2, the truncated sequential estimator $f^*_N(h_N,H_N)$ has similar to the sequential estimator (3) rate of convergence (as $N\to\infty$).

In the case of i.i.d. noises $(\delta_i)$ the MSE of the constructed sequential estimator has following decreasing rate

$$E(f_N^*(h_N, H_N) - f)^2 = O\left(h_N^{2\varrho} + \frac{1}{N h_N^m}\right) \quad \text{as } N \to \infty.$$

Then the bandwidth $h$ with the optimal rate is proportional to

$$h_N \sim N^{-\frac{1}{(2\varrho+m)}}$$

and the condition (9) is fulfilled for $\varrho > m/2$ and

$$r > \frac{4\varrho}{2\varrho - m}.$$

In particular, for the case $p = 0$, $\gamma_f = 1$, $m = 1$,

$$E(f_N^*(h_N, H_N) - f)^2 = O\left(h_N^2 + \frac{1}{N h_N}\right) \quad \text{and } r > 4.$$

**Remark 5.** *Theorems 1 and 2 give known upper bounds for the bias and the MSE of presented estimators as well as for the mean of observation time in Theorem 1 if the parameters of classes of functions, introduced in Assumptions 1 and 2 are known ($C_{f,k}, p, \gamma_f$, etc). In this case 'optimal' bandwidth can be found from minimization of the upper bounds for the MSE's.*

*At the same time the assertions of Theorems 1 and 2 are fulfilled even if some of these parameters are unknown. Estimators with such properties can be successfully used in various adaptive procedures as pilot estimators (see, e.g., [28]-[32]).*

# 5. Examples

In this section examples of various dependence types of regressors and noises of the model (1) are given.

## 5.1. Case of independent regressors X and Δ

Consider examples of inputs of the model in this case. We consider a scalar case $m = 1$ for simplification only. The extension to the multivariate case is immediate.

**5.1.1. Example of regressors X.** Assume that regressors $(x_i)$ satisfy the following equation

(10) $$x_i = \Psi(x_{i-1}, \ldots, x_{i-r}) + \varepsilon_i, \quad i \geq 1,$$

where $\Psi(\cdot)$ is a bounded function $|\Psi(y)| \leq \overline{\Psi} < \infty$, $y \in \mathcal{R}^r$, $r \geq 1$.

Assume that the input noises $\varepsilon_i$ in the model (10) are i.i.d. with the density function $f_\varepsilon(\cdot)$. In this case $f_{i-1}(t) = f_\varepsilon(t - \Psi(x_{i-1}, \ldots, x_{i-r})|\mathcal{F}_{i-1})$.

Thus Assumption 2 holds true if the density function $f_\varepsilon(\cdot)$ is $q$-th differentiable, $q \geq 0$ and for some positive numbers $c_x^0$, $C_{x,0}$, $L_x$ and $\gamma_x \in (0, 1]$ the following relations

$$\inf_{|t - x^0| \leq \overline{\Psi}} f_\varepsilon(t) \geq c_x^0,$$

$$\sup_{|t-x^0|\le\overline{\Psi}} |f_\varepsilon^{(i)}(t)| \le C_{x,i}, \ i=\overline{0,q}$$

and for every $x, y \in \mathbb{R}^1$,

$$|f_\varepsilon^{(q)}(x) - f_\varepsilon^{(q)}(y)| \le L_x |x-y|^{\gamma_x}$$

are fulfilled.

**5.1.2. Example of noises $\Delta$.** Assume that noises $(\delta_i)$ satisfy the following autoregressive equation

$$\delta_i = \lambda\delta_{i-1} + \eta_i, \quad i \ge 1, \tag{11}$$

which is supposed to be stable, $|\lambda| < 1$, and $\eta_i$ are i.i.d., $E\eta_i = 0$, $E\eta_i^2 \le \sigma_\eta^2$, $i \ge 0$; $\eta_0 = \delta_0$. Then

$$\varphi(k) = \frac{|\lambda|^k \sigma_\eta^2}{1-\lambda^2}.$$

In this case we can put (after optimization of the upper bound of the MSE)

$$l_n \sim (1+\mu)\log_{|\lambda|}\frac{1}{n}, \quad \mu > 0$$

and the summands

$$\frac{l_n}{nh_n} \sim \frac{\log n}{nh_n}, \quad \varphi_n h_n^{-1} \sim \frac{1}{n^{1+\mu}h_n} = o\left(\frac{1}{nh_n}\right).$$

Then the condition (6) (which is necessary for truncated estimators as well) is fulfilled if

$$\frac{\log n}{nh_n} = o(1) \quad \text{as } n \to \infty.$$

It is clear, that this example can be easy generalized for the stable autoregressive process (11) of an arbitrary order.

## 5.2. Case of dependent regressors X and $\Delta$

Consider the estimation problem in the nonlinear autoregressive model

$$Y_i = f(Y_{i-1}) + \delta_i, \ \ i \ge 1,$$

where $\delta_i$, $i \ge 1$ form the sequence of zero mean i.i.d.r.v's with a density function $f_\delta(\cdot)$ and finite variance $E\delta_i^2 = \sigma^2$, $i \ge 1$; $Y_0$ is a zero mean random number with finite variance and independent from $\delta_i$, $i \ge 1$.

Then $x_i = Y_{i-1}$, $i \ge 1$ and $\varphi(k) = \sigma^2\chi(k=0)$, $k \ge 0$, as well as in this case $f_i(t) = f_\delta(t - f(Y_{i-1}))|\mathcal{F}_{i-1}^Y)$, $\mathcal{F}_i^Y = \sigma\{Y_0, \delta_1, \ldots, \delta_i\}$.

Thus Assumption 2 holds true if the function $f(\cdot)$ is uniformly bounded $\sup_{t\in\mathbb{R}^1} f(t) \le C_f$, the density function $f_\delta(\cdot)$ is $q$-th differentiable, $q \ge 0$ and for some positive numbers $c_x^0$, $C_{x,0}$, $L_x$ and $\gamma_x \in (0,1]$ the following relations

$$\inf_{|t-x^0|\le C_f} f_\delta(t) \ge c_x^0,$$

$$\sup_{|t-x^0|\le C_f} |f_\delta^{(i)}(t)| \le C_{x,i}, \ i = \overline{0,q}$$

and for every $x, y \in \mathbb{R}^1$,

$$|f_\delta^{(q)}(x) - f_\delta^{(q)}(y)| \le L_x |x-y|^{\gamma_x}$$

are fulfilled.

# 6. Conclusion

The sequential approach for the problem of estimation of a multivariate regression function from dependent observations is developed. It is supposed, that the function to be estimated belongs to the Hölder class and input processes of the model can be unbounded with a positive probability.

Two types estimators are presented. Sequential estimators give the possibility to get estimators with an arbitrary mean square accuracy by finite stopping time. Truncated sequential estimators have known variance and constructed by a sample of a finite (fixed) size.

Both estimation procedures work under general dependency types of model inputs. Asymptotic rate of convergence of both estimators coincides with the optimal rate of Nadaraya–Watson estimators constructed from independent observations. At that we consider the mean $E\tau_n$ having the rate (4) as a duration of observations in sequential approach for comparison with Nadaraya–Watson estimators calculated by the sample of the size $n$.

Presented estimators can be used directly and as pilot estimators in various statistical problems.

# References


1. Er-Wei Bai, Yun Liu, Recursive Direct Weight Optimization in Nonlinear System Identification: A Minimal Probability Approach // Automatic Control, IEEE Transactions on Issue Date: July 2007 Volume: 52 Issue: 7 P. 1218-1231
2. Borisov V.Z., Konev V.V. On sequential estimation of parameters in discrete-time processes // Automat. and Remote Control. 1977. Vol. 38. No. 10 P. 58–64 (in Russian).
3. Cramér H. Mathematical methods of statistics. Princeton Univ. Press, 1948.
4. Dharmasena, S, Zeephongsekul, P and De Silva. Two stage sequential procedure for nonparametric regression estimation // Australian and New Zealand Industrial and Applied Mathematics Journal. 2008. Vol. 49. P. C699-C716.
5. Efroimovich S.Yu. Nonparametric curve estimation. Methods, theory and applications. Berlin, N. Y.: Springer-Verlag, 1999.
6. Efroimovich Sam. Sequential Design and Estimation in Heteroscedastic Nonparametric Regression // Sequential Analysis: Design Methods and Applications. 2007. Vol. 26. No. 1. P. 3-25. Author: Sam Efromovich DOI: 10.1080/07474940601109670
7. Fourdrinier, D., Konev, V. and Pergamenshchikov, S. Truncated sequential estimation of the parameter of a first order autoregressive process with dependent noises // Mathematical Methods of Statistics. 2009. Vol. 18. No. 1. P. 43-58.
8. L.Györfi, M.Kohler, A.Krzyżak, H.Walk A distribution-free theory of nonparametric regression. Berlin, N. Y.: Springer-Verlag, 2002.
9. T. Hondaa. Sequential estimation of the hgarginal density function for a strongly mixing process // Sequential Analysis: Design Methods and Applications. 1998. Vol. 17. No. 3,4. P. 239 - 251. DOI: 10.1080/07474949808836411

10. Konev, V.V. Sequential parameter estimation of stochastic dynamical systems. Tomsk: Tomsk Univ. Press, 1985 (in Russian).
11. Konev, V.V. and Pergamenshchikov, S.M. Truncated sequential estimation of the parameters in random regression // Sequential analysis. 1990. Vol. 9. No. 1. P. 19-41.
12. Konev, V.V. and Pergamenshchikov, S.M. On truncated sequential estimation of the drifting parameter mean in the first order autoregressive model // Sequential analysis. 1990. Vol. 9. No. 2. P. 193-216.
13. Konev, V. V. and Pergamenshchikov, S. M. On Truncated Sequential Estimation of the Parameters of Diffusion Processes // Methods of Economical Analysis. Central Economical and Mathematical Institute of Russian Academy of Science, Moscow. 1992. P. 3-31.
14. U. Küchler and Vasiliev, V. On guaranteed parameter estimation of a multiparameter linear regression process // Automatica, Journal of IFAC. Elsevier. 2010. No. 46. P. 637-646.
15. Liptser R.Sh., Shiryaev A.N. Statistics of random processes. I: General theory. N. Y.: Springer-Verlag, 1977. 2: Applications. N. Y.: Springer-Verlag, 1978.
16. T. McMurry and D. N. Politis. Nonparametric regression with infinite order flat-top kernels // J. Nonparam. Statist. 2004. Vol. 16. No. 3-4. P. 549–562.
17. A.V. Nazin and L. Ljung. Asymptotically Optimal SMoothing of Averaged LMS for Regression // Proceedengs of the 15th IFAC World Congress, Barcelona, July, 2002.
18. Novikov A.A. Sequential estimation of the parameters of the diffusion processes // Theory Probab. Appl. 1971. Vol. 16. No. 2. P. 394–396 (in Russian).
19. D.N.Politis. On nonparametric function estimation with infinite-order flat-top kernels // Probability and Statistical Models with applications, Ch. Charalambides et al. (Eds.), Chapman and Hall/CRC, Boca Raton. 2001. P. 469-483.
20. D.N.Politis, J.P.Romano. On a Family of Smoothing Kernels of Infinite Order in Computing Science and Statistics // Proceedings of the 25th Symposium on the Interface, San Diego, California, April 14-17, 1993. M. Tarter and M. Lock, eds., The Interface Foundation of North America, P. 141-145.
21. Dimitris N. Politis and Joseph P. Romano, Multivariate density estimation with general flat-top kernels of infinite order // Journal of Multivariate Analysis. 1999. Vol. 68. P. 1-25.
22. D. Politis, J.Romano and M. Wolf. Subsumpling. Springer Verlag, Berlin, Heidelberg, New York, 1999.
23. Prakasa Rao, B.L.S. Nonparametric Functional Estimation. New York: Academic Press, 1983.
24. J. Roll, A. Nazin, and L. Ljung. A Non-Asymptotic Approach to Local Modelling // The 41st IEEE CDC, Las Vegas, Nevada, USA, 10-13 Dec. 2002. (Regular paper.) Predocumentation is available at http://www.control.isy.liu.se/research/reports/2003/2482.pdf
25. J. Roll, A. Nazin, and L. Ljung. Non-linear System Identification Via DirectWeight Optimization // Automatica, Special Issue on "Data-Based Modelling and System Identification", V.41, No.3, 2005, pp.475–490. IFAC Journal. Predocumentation is available at http://www.control.isy.liu.se/research/reports/2005/2696.pdf
26. M. Rosenblatt, Stochastic curve estimation // NSF-CBMS Regional Conference Series, 1991. Institute of Mathematical Statistics, Hayward, 1991, Vol. 3.
27. D. W. Scott, Multivariate density estimation: theory, practice and visualization. Wiley, New York, 1992.
28. Vasiliev V.A. On Identification of Dynamic Systems of Autoregressive Type // Automat. and Remote Control. 1997. Vol. 58. No 12, P. 106-118.
29. Vasiliev V.A. On adaptive control problems of discrete-time stochastic systems // Proceedings of the 18th World Congress The Int. Fed. Autom. Contr., Milan, Italy, 2011. August 27 - September 2. 6 pages (to be published).
30. V.A.Vasiliev, A.V.Dobrovidov and G.M.Koshkin. Nonparametric estimation of functionals of stationary sequences distributions. Moscow.: Nauka, 2004, 508 p. (in Russian)
31. Vasiliev V.A., Konev V.V. On Optimal Adaptive Control of a Linear Stochastic Process // Proc.15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics, August 24-29, 1997, Berlin, Germany. Vol. 5. P. 87-91.
32. Vasiliev V.A., Koshkin G.M. Nonparametric Identification of Autoregression // Proba-

bility Theory and Its Applications. 1998. Vol. 43. No. 3. P. 577-588.
33. Wald A. Sequential analysis. N. Y.: Wiley, 1947.