Nonparametric Identification of Static and Dynamic Production Functions

Gennady Koshkin, and Anna Kitayeva

***Abstract*—A class of semi-recursive kernel plug-in estimators of functions depending on multivariate density functionals and their derivatives is considered. The approach enables to estimate the static production function, marginal productivity and marginal rate of technical substitution of inputs. The piecewise smoothed approximations of these estimators are proposed. The main parts of the asymptotic mean square errors (AMSEs) of the estimators are found. The results are generalized to the dynamic production functions with the lagged values of the inputs and output.**

***Index Terms*—Kernel recursive estimator, mean square convergence, piecewise smooth approximation, production function.**

I. Introduction

THIS paper is based on the results published in the Proceedings of the World Congress on Engineering 2011, that was held in Imperial College London, London, U.K., July 6-8, 2011 [1].

Numerous statistical problems (such as identification, classification, filtering, prediction, etc.) is connected to estimation of certain characteristics of the following expressions:

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

$$= H\left(a(x),\ a^{(1j)}(x)\right). \qquad (1)$$

Here $x \in \mathsf{R}^m$, $H(\cdot) : \mathsf{R}^{(m+1)s} \to \mathsf{R}^1$ is a given function,

$$a^{(0j)}(x) = a(x) = (a_1(x), \ldots, a_s(x)),$$

$$a^{(1j)}(x) = \left(a_1^{(1j)}(x), \ldots, a_s^{(1j)}(x)\right),$$

$$a_i(x) = \int g_i(y) f(x,y)dy, \quad i=\overline{1,s},$$

$$a_i^{(1j)}(x) = \frac{\partial a_i(x)}{\partial x_j}, \quad i=\overline{1,s}, \quad j=\overline{1,m},$$

where $g_1, \ldots, g_s$ are the known Borel functions, $\int \equiv \int_{\mathsf{R}^1}$, $f(\cdot,\cdot)$ is an unknown probability density function (p.d.f.) for the observed random vector $Z = (X,Y) \in \mathsf{R}^{m+1}$.

Remark. Note that in (1) some variables of function $H(\cdot)$ may be omitted (for example, all derivatives).

Manuscript received August 20, 2011. 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.

A. Kitayeva is with International Management Institute, Tomsk Polytechnic University, Russia, (email: olz@tomsknet.ru.),

G. Koshkin is with Tomsk State University, Department of Applied Mathematics and Cybernetics, 36, Lenin str., Tomsk, Russia and with Department of Informatization Problem, Tomsk Scientific Center SB RAS, (email: kgm@mail.tsu.ru.)

If $g_i(y) \equiv 1$, then

$$a_i(x) = \int f(x,y)dy = p(x),$$

where $p(\cdot)$ is the marginal p.d.f. of the random variable X, and $f(y|\,x) = f(x,y)/p(x)$ is the conditional p.d.f.

Here are the well known examples of such functions:

— the conditional initial moments

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

$$g_1(y) = y^m, \quad g_2(y) = 1;$$

$\mu_1(x) = r(x)$ is the regression function of the output y of a stochastic object to relative to the inputs x ($r(x)$ minimizes the mean square error (MSE) of the outputs of an object and a model);

— the conditional central moments

$$V_m(x) = \int (y - r(x))^m f(y|x)dy, \ g_1(y) = y,$$

$$g_2(y) = y^2, \ldots,\ g_m(y) = y^m,\ g_{m+1}(y) = 1;$$

$V_2(x) = D(x)$ is the conditional variance or the scedastic curve [2], $\sigma(x) = \sqrt{D(x)}$ is the conditional standard deviation, and $D(x), \sigma(x)$ specify errors of the regressive model $r(x)$;

— the conditional coefficient of skewness or the clitic curve [2]

$$\beta_1(x) = \frac{\mathsf{E}((Y - r(x))|x)^3}{[\mathsf{D}(Y|x)]^{3/2}},\ b_i = a_i/a_1,\ g_i(y) = y^{i-1},$$

$$H(a_1,a_2,a_3,a_4) = (b_4 - 3b_3b_2 + 2b_2^3)/(b_3 - b_2^2)^{3/2};$$

— the curtic curve [2]

$$\beta_2(x) = \frac{\mathsf{E}((Y - r(x))^4|x)}{[\mathsf{D}(Y|x)]^2};$$

— the sensitivity functions

$$T_j(x) = \frac{\partial r(x)}{\partial x_j},\ g_1(y) = 1,\ g_2(y) = y,$$

$$H\left(a_1, a_2, a_1^{(1j)}, a_2^{(1j)}\right) = \frac{a_1^{(1j)}}{a_2} - \frac{a_1 a_2^{(1j)}}{a_2^2} = b_2^{(1j)};$$

$T_j(x)$ defines the degree of the relation between changes of the input x_j and output y of an object model. We note that in [2] the useful possibility is pointed out for the application of conditional cumulants or semi-invariants κ_j, $j \geq 1$, instead of conditional moments; thus, for example, for the clitic curve, the numerator of the ratio H is the conditional semi-invariant κ_3 [3].

(Advance online publication: 24 August 2011)

II. Problem statement

In the solution of various problems of identification, recurrent procedures find wide application, which has a number of advantage in comparison with common procedures. As a rule, they are easily realized on computers, thus economizing their memory, afford the result at each step of working of the algorithm, and the coming of new measurements does not lead to cumbersome recomputations, so this provides the data processing in the real-time mode.

Take the following expression as an estimator of the functional $a(x) = a^{(0j)}(x)$ $(r = 0)$ and its derivatives $a^{(1j)}(x)$ $(r = 1)$ at a point x:

$$a_n^{(rj)}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{g(Y_i)}{h_i^{m+r}}\mathbf{K}^{(rj)}\left(\frac{x - X_i}{h_i}\right). \qquad (2)$$

Here $Z_i = (X_i, Y_i),\ i = \overline{1, n},$ is the $(m + 1)$-dimensional random sample from p.d.f. $f(\cdot,\cdot)$, (h_i) is a sequence of positive bandwidths tending to 0 as $i \to \infty$,

$$\mathbf{K}^{(0j)}(u) = \mathbf{K}(u) = \prod_{i=1}^{m} K(u_i)$$

is a m-dimensional multiplicative function which does not need to possess the properties of p.d.f.,

$$\mathbf{K}^{(1j)}(u) = \frac{\partial \mathbf{K}(u)}{\partial u_j}, g(y) = (g_1(y), \ldots, g_s(y)),$$

$$a_n^{(rj)}(x) = \left(a_{1n}^{(rj)}(x), \ldots, a_{sn}^{(rj)}(x)\right).$$

Note that (2) can be computed recursively by

$$a_n^{(rj)}(x) = a_{n-1}^{(rj)}(x) -$$
$$-\frac{1}{n}\left[a_{n-1}^{(rj)}(x) - \frac{g(Y_n)}{h_n^{m+r}}\mathbf{K}^{(rj)}\left(\frac{x - X_n}{h_n}\right)\right]. \qquad (3)$$

This property is particularly useful when the sample size is large since (3) can be easily updated with each additional observation.

In the case, when

$$m = 1,\quad s = 1,\quad g(y) = 1,\quad H(a_1) = a_1,$$

we obtain the recursive kernel estimator of $p\,(x)$ that was introduced by Wolverton and Wagner in [4] and apparently independently by Yamato [5], and has been thoroughly examined in [6].

The semi-recursive kernel estimators of conditional functionals

$$b(x) = (b_1(x), \ldots, b_{s-1}(x)),$$

$$b_i(x) = a_i(x)/p\,(x) = \int g_i(y) f(y|\,x) dy$$

at a point x are designed as $(g_s(x) = 1)$

$$b_n(x) = \frac{\displaystyle\sum_{i=1}^{n}\frac{g(Y_i)}{h_i^m}\mathbf{K}\left(\frac{x - X_i}{h_i}\right)}{\displaystyle\sum_{i=1}^{n}\frac{1}{h_i^m}\mathbf{K}\left(\frac{x - X_i}{h_i}\right)} = \frac{a_n(x)}{p_n(x)} = \frac{a_n^{(0j)}(x)}{a_{sn}^{(0j)}(x)}.$$

Such estimators are called semi-recursive because they can be updated sequentially by adding extra terms to both the numerator and denominator when new observations became available. If $g_1(y) = y$ $(s = 2)$, we obtain semi-recursive kernel estimators of the regression line [7]– [9]. Weak and strong universal consistency of such estimates was investigated in [10]– [14].

To estimate (1) we use the following semi-recursive plug-in statistic

$$J_n(x) = H\left(\left\{a_n^{(rj)}(x)\right\},\quad j = \overline{1, m},\quad r = 0, 1\right). \qquad (4)$$

Plug-in estimators $b_n(x)$ are often used for estimating of ratios. There is a possible instability of $b_n(x)$, and may be estimators (4) too, related to a proximity of the denominator to zero. Cramér considered such a problem first (see [15] for details). Here this problem has been solved by make using of the piecewise smooth approximation [16]

$$\widetilde{J}_n(x) = \frac{J_n(x)}{(1 + \delta_n |J_n(x)|^\tau)^\rho}, \qquad (5)$$

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

III. Mean square errors

Denote:

$$\sup_x = \sup_{x \in \mathsf{R}^m},\quad K^{(1)}(u) = \frac{dK(u)}{du},$$

$$T_j = \int u^j K(u) du,\ j = 1, 2, \ldots.$$

We will introduce auxiliary definitions.

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 is continuously differentiable up to the order ν at the point $t \in \mathsf{R}^s$. A function $H(\cdot) \in \mathcal{N}_\nu(\mathsf{R})$ if it is continuously differentiable up to the order ν for any $z \in \mathsf{R}^s$.

This definition is related to the required smoothness conditions for the function H in (1). The following three definitions impose conditions on the estimation procedure.

Definition 2. A Borel function $K(\cdot) \in \mathcal{A}^{(r)}$, $(\mathcal{A}^{(0)} = \mathcal{A})$ if $\int |\,K^{(r)}(u)|\,du < \infty$, and $\int K(u)\,du = 1$.

Definition 3. A Borel function $K(\cdot) \in \mathcal{A}_\nu^{(r)}, (\mathcal{A}_\nu^{(0)} = A_\nu)$ 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 specifies the rate of convergence in the mean square sense of the estimators (4) and (5).

Definition 4. A sequence $(h_n) \in \mathcal{H}(m + r + q)$ if

$$(h_n + 1/(n h_n^{m+r+q})) \downarrow 0, \qquad (6)$$

$$\frac{1}{n}\sum_{i=1}^{n} h_i^\lambda = S_\lambda h_n^\lambda + o(h_n^\lambda), \qquad (7)$$

where λ is a real number, S_λ is a constant independent on n; $r, q = 0, 1$.

The condition (6) is a common condition for the convergence in the mean square sense of kernel estimators. The condition (7) is related to the recurrent structure of the estimators and is fulfilled, for example, $h_i = O\left(i^{-\alpha}\right)$, $0 < \alpha < 1$ (it is this form that optimal bandwidths h_n have [8]), in which case the constant S_λ can be defined according to the Euler–Macloren formula. In particular, for any $p \neq -1$ we obtain

$$\sum_{j=1}^{n} j^p = n^{p+1}/(p+1) + o(n^{p+1}).$$

Definition 5. Let $t_n,\ X_1,\dots,X_n$ are vectors, and $t_n = t_n(X_1,\dots,X_n)$. A sequence of functions $\{H(t_n)\}$ belongs to the class $\mathcal{M}(\gamma)$ if for any possible values $X_1,\dots,X_n$ the sequence $\{|H(t_n)|\}$ is dominated by a sequence of numbers $(C_0 d_n^{\gamma})$, $(d_n) \uparrow \infty$ as $n \to \infty$, $0 \le \gamma < \infty$, C_0 is a constant.

For the necessity of introduction of majorizing sequences in finding the AMSEs of unstable estimators, it was Cramér who first pointed to it in the book [15], in which he strictly formulated and proved the theorem on the mean and variance of the function of sample moments.

Put for $r,q = 0,1;\ t,p = \overline{1,s};\ j = \overline{1,m}$:

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

$$H\left(\left\{a_n^{(rj)}(x)\right\}\right) = H(A_n);\ a^{s+}(x) = \int |g^s(y)| f(x,y)dy;$$

$$a_{t,\,p}(x) = \int g_t\,(y) g_p\,(y)\, f(x,y)\, dy;$$

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

$$L^{(r,\,q)} = \int K^{(r)}(u) K^{(q)}(u)\, du;$$

$$\mathcal{B}_{t,\,p}^{(r,\,q)} = L^{(r,\,q)} \left(L^{(0,0)}\right)^{m-1} a_{t,\,p}\,(x);$$

$$\omega_{i\nu}^{(rj)} = \frac{T_\nu}{\nu!} \sum_{l=1}^{m} \frac{\partial^{\,\nu} a_i^{(rj)}(x)}{\partial x_l^{\nu}};$$

the set

$$Q = \begin{cases} \{0\} & \text{if } \forall j \ \ r = 0; \\ \{1\} & \text{if } \forall j \ \ r = 1; \\ \{0,1\} & \text{if } \exists j \ \ r = 0 \bigwedge r = 1. \end{cases}$$

Theorem 1 (AMSE of the estimator $J_{\,n}(x)$). If for $t,p = \overline{1,s},\ \ j = \overline{1,m},\ \ r \in Q$:

1) the functions $a_{t,\,p}\,(\cdot) \in \mathcal{N}_0(\mathsf{R})$, $\sup\limits_x a_{t,\,p}^{1+}\,(x) < \infty$, $\sup\limits_x a_t^{1+}(x) < \infty$, $\sup\limits_x a_t^{4+}(x) < \infty$;

2) the kernel function $K(\cdot) \in \mathcal{A}_\nu^{(\max(r))}$, $\sup\limits_x \left|K^{(r)}(x)\right| < \infty$, if $Q = \{0,1\}$ then $K^{(r)}(\cdot) \in \mathcal{N}_0(\mathsf{R})$, if $1 \in Q$ then $\lim\limits_{|u|\to\infty} K(u) = 0$;

3) $a_t^{(rj)}(\cdot) \in \mathcal{N}_\nu(\mathsf{R})$, $\sup\limits_x |\,a_t^{(rj)}(x)| < \infty$, $\sup\limits_x \left|\dfrac{\partial^{\,\nu} a_t^{(rj)}(x)}{\partial x_l \partial x_t \dots \partial x_q}\right| < \infty$, $l,t,\dots,q = \overline{1,m}$;

4) the sequence $(h_n) \in \mathcal{H}(m + 2\max(r))$;

5) $H(\cdot) \in \mathcal{N}_2(A)$;

6) $\{H(A_n)\} \in \mathcal{M}(\gamma)$, $\quad 0 \le \gamma \le 1/4$.

Then AMSE of the estimator $J_{\,n}(x)$ as $n \to \infty$

$$\mathsf{u}^2(J_{\,n}) = \sum_{t,\,p=1}^{s} \sum_{j,\,k=1}^{m} \sum_{r,\,q \in Q} H_{tjr} H_{pkq} \times$$

$$\times \left[S_{-(m+2\max(r,\,q))} \frac{\mathcal{B}_{t,\,p}^{(r,\,q)}}{n h_n^{m+r+q}} + S_\nu^2\, \omega_{t\nu}^{(rj)} \omega_{p\;\nu}^{(qk)} h_n^{2\nu} \right] +$$

$$+\, O\left(\left[\frac{1}{n h_n^{m+2\max(r)}} + h_n^{2\nu}\right]^{\frac{3}{2}}\right).$$

It is important that we do not need condition 6) of Theorem 1 when piecewise smooth approximation (5) is used.

Theorem 2 (AMSE of the piecewise smooth approximation $\widetilde{J}_{n,\,\nu}(x)$). Suppose that conditions 1)–5) of Theorem 1 hold and restriction 6) is replaced by

6*) $J(x) = H(A(x)) \neq 0$ or $\tau \ge 4$, τ is a positive integer.

Then as $n \to \infty$

$$\mathsf{u}^2(\widetilde{J}_n) \sim \mathsf{u}^2(J_{\,n}(x)).$$

The proofs are given in [17].

IV. Nonparametric semi-recursive identification of the static production function and its characteristics

Apply the results to estimate the static production function and its characteristics.

A. Estimation of the three-factor production function

Let $r(x)$, $x = (x_1,x_2,x_3) \in \mathsf{R}^3$, be the regression model of the three-factor production function,

$$a(x) = (a_1(x), a_2(x)), \quad a_1(x) = \int y f(x,y)dy,$$

$$a_2(x) = \int f(x,y)dy = p\,(x).$$

Here $x_1 > 0$ is the capital input, $x_2 > 0$ is the labor input, $x_3 > 0$ is the nature input, $y > 0$ is a product, and p.d.f. $f(x,y) > 0$ only if $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, $y > 0$. Then

$$J_n(x) = r_n(x) = \frac{\sum\limits_{i=1}^{n} \dfrac{Y_i}{h_i^3} \mathbf{K}\left(\dfrac{x - X_i}{h_i}\right)}{\sum\limits_{i=1}^{n} \dfrac{1}{h_i^3} \mathbf{K}\left(\dfrac{x - X_i}{h_i}\right)} =$$

$$= \frac{a_{1n}^{(0j)}(x)}{a_{2n}^{(0j)}(x)} = \frac{a_{1n}(x)}{p_{\,n}(x)}\,. \tag{8}$$

Let

$$\mathbf{K}(u) = K(u_1)K(u_2)K(u_3), \quad K(\cdot) \in \mathcal{A}_\nu,$$

$$\sup_{u \in \mathsf{R}^1} |\,K(u)| < \infty, \quad (h_n) \in \mathcal{H}(3).$$

To find the AMSE of the estimator $r_n(x)$, we use Theorem 1. In view of 1)–4) conditions of the theorem functions $a_i(z)$, $\ i = 1,2,$ and their derivatives are continuously differentiable up to the order ν for any $z \in \mathsf{R}^3$, and the function $\int y^4 f(x,y)dy$ is bounded on R^3. If $p\,(x) > 0$, then condition 5) is fulfilled.

It seems impossible to find a majorizing sequence (d_n) (condition 6) of Theorem 1), since the denominator in the ratio (8) may be equal to zero [18], [19]. But it is shown in these papers that we can find the majorizing sequence with

$\gamma = 0$ under $\nu = 2$ according to Definition 5 if, for example, $K(u) \geq 0$, and $Y < \infty$. In this case as $n \to \infty$

$$\mathsf{u}^2(r_n) = \sum_{i,\,p=1}^{2} H_i H_p \left(S_{-3} \frac{L^{(0,0)} B_{i,\,p}}{n h_n^3} + S_2^2 \omega_{i2} \omega_{p\,2} h_n^4 \right) + + O\left(\left[\frac{1}{n h_n^3} + h_n^4 \right]^{3/2} \right),$$

where

$$H_1 = \frac{1}{p\,(x)}, \quad H_2 = -\frac{r(x)}{p\,(x)}; \quad B_{1,1} = \int y^2 f(x,y) dy,$$

$$B_{1,\,2} = B_{2,1} = \int y f(x,y) dy, \quad B_{2,\,2} = p\,(x);$$

$$\omega_{12} = \frac{T_2}{2} \left(\frac{\partial^2 a_1(x)}{\partial x_1^2} + \frac{\partial^2 a_1(x)}{\partial x_2^2} + \frac{\partial^2 a_1(x)}{\partial x_3^2} \right),$$

$$\omega_{22} = \frac{T_2}{2} \left(\frac{\partial^2 p\,(x)}{\partial x_1^2} + \frac{\partial^2 p\,(x)}{\partial x_2^2} + \frac{\partial^2 p\,(x)}{\partial x_3^2} \right).$$

For $\nu > 2$ we can use the piecewise smooth approximation $\widetilde{r}_n(x)$:

$$\widetilde{r}_n(x) = \frac{r_n(x)}{(1 + \delta_{n,\,\nu} |\, r_n(x) |^{\,\tau})^{\rho}},$$

where $\tau > 0,\ \rho > 0,\ \rho\tau \geq 1,\ \delta_{n,\,\nu} = O\left(h_n^{2\nu} + 1/(n h_n^3) \right)$, $(\delta_{n,\,\nu}) \downarrow 0$ as $n \to \infty$.

In view of condition 6*) of Theorem 2 it is enough to take even $\tau \geq 4$, and as $n \to \infty$

$$\mathsf{u}^2(\widetilde{r}_n) = \sum_{i,\,p=1}^{2} H_i H_p \left(S_{-3} \frac{L^{(0,0)} B_{i,\,p}}{n h_n^3} + S_\nu^2\, \omega_{i\nu} \omega_{p\nu} h_n^{2\nu} \right) + + O\left(\left[\frac{1}{n h_n^3} + h_n^{2\nu} \right]^{3/2} \right).$$

B. Estimation of the conditional variance

Estimate the conditional variance $D(x)$ by the following statistic:

$$D_n\,(x) = \frac{\sum_{i=1}^{n} \frac{Y_i^2}{h_i^3} \mathbf{K}\left(\frac{x - X_i}{h_i} \right)}{\sum_{i=1}^{n} \frac{1}{h_i^3} \mathbf{K}\left(\frac{x - X_i}{h_i} \right)} - r_n^2\,(x). \qquad (9)$$

The estimator (9) is a semi-recursive counterpart of a widely used volatility function estimator [20]. The piecewise smooth approximation of the estimator $D_n\,(x)$ can be written easily.

C. Estimation of the marginal productivity functions

In economics, the marginal productivity functions are defined by the formulas

$$MP_j(x) = T_j(x) = \frac{\partial r(x)}{\partial x_j}, \quad j = 1, 2, 3,$$

and a dominant sequence finding difficulties force us to use the piecewise smooth approximation $\widetilde{T}_{jn}(x)$:

$$\widetilde{T}_{jn}(x) = \frac{T_{jn}(x)}{(1 + \delta_n |\, T_{jn}(x) |^{\,\tau})^{\rho}},$$

where

$$T_{jn}(x) = \left[\frac{\sum_{i=1}^{n} \frac{Y_i}{h_i^4} \mathbf{K}^{(1j)}\left(\frac{x - X_i}{h_i} \right)}{\sum_{i=1}^{n} \frac{1}{h_i^3} \mathbf{K}\left(\frac{x - X_i}{h_i} \right)} - \right. \left. - \frac{\sum_{i=1}^{n} \frac{Y_i}{h_i^3} \mathbf{K}\left(\frac{x - X_i}{h_i} \right) \sum_{i=1}^{n} \frac{Y_i}{h_i^4} \mathbf{K}^{(1j)}\left(\frac{x - X_i}{h_i} \right)}{\left[\sum_{i=1}^{n} \frac{1}{h_i^3} \mathbf{K}\left(\frac{x - X_i}{h_i} \right) \right]^2} \right], \qquad (10)$$

$$\mathbf{K}^{(11)}(u) = K^{(1)}(u_1) K(u_2) K(u_3),$$

$$\mathbf{K}^{(12)}(u) = K(u_1) K^{(1)}(u_2) K(u_3),$$

$$\mathbf{K}^{(13)}(u) = K(u_1) K(u_2) K^{(1)}(u_3).$$

Here the kernel $K(\cdot)$ satisfies such additional conditions:

$$\sup_{u \in \mathsf{R}^1} |K^{(1)}(u)| < \infty, \qquad \lim_{|u| \to \infty} K(u) = 0,$$

$$K^{(\alpha)}(\cdot) \in \mathcal{N}_0(\mathsf{R}), \quad \alpha = 1, 2;$$

functions $a_1(\cdot)$, $a_2(\cdot)$ and their derivatives up to the order $(\nu + 1)$ need to be continuous and bounded on R^3; the sequence $(h_n) \in \mathcal{H}(4)$.

The estimators (9) and (10) were used in identification of CHARN (conditional heteroscedastic autoregressive nonlinear) type model on the data of "Gazprom" stock prices for the period from January 15, 2008 till March 24, 2009 [21].

D. Estimation of the marginal rate of technical substitution and elasticity coefficients

In turns, in economics, a number of other important characteristics are defined in terms of the marginal productivity, for example, the marginal rate of technical substitution of an input x_j with an input x_i

$$MRTS_{ij}(x) = \frac{MP_i(x)}{MP_j(x)}, \ i, j = 1, 2, 3, \ i \neq j,$$

$$H(a_1, a_2, a_1^{(1j)}, a_2^{(1j)}) = \frac{b_1^{(1i)}}{b_1^{(1j)}}, \quad g_1(y) = y,$$

the elasticity coefficient of the output y by the jth variable factor

$$E_j(x) = \frac{MP_j(x)}{r(x)}\, x_j = \frac{\partial r(x)}{\partial x_j} \frac{x_j}{r(x)} = \frac{\partial \ln r(x)}{\partial x_j}\, x_j\,,$$

$$H(a_1, a_2, a_1^{(1j)}, a_2^{(1j)}) = \frac{b_1^{(1j)}}{b_1}\, x_j\,, \quad g_1(y) = y.$$

So, the plug-in estimator of the $MRTS_{ij}(x)$ takes the form

$$MRTS_{ij,\,n}(x) = \frac{T_{in}(x)}{T_{jn}(x)}.$$

The piecewise smooth approximation of the $MRTS_{ij,\,n}(x)$ can be written easily. In view of condition 5) of Theorem 1 the condition

$$r(x) \neq \frac{\partial a_1(x)}{\partial x_j} \Big/ \frac{\partial p\,(x)}{\partial x_j}$$

has to hold in addition to the previous restrictions.

The class of functions (1) enable us to describe from unique positions the system of characteristics of production function: characteristics of the zero order (production functions, average output of separate and general factors, profit); characteristics of the first order (marginal productivity of factors, elasticities of output by factors, marginal rate of technical substitution of a factor by another, elasticity of the output by the scale) [23, pp. 47–49]. As regards the characteristics of the second order, for example, elasticities of factors substitution by Allen, Mikhalevskii, Mac-Fadden (see [23, pp. 49–50]), the approaches to their investigation are similar because they can be obtained by the iterated differentiation of characteristics of the first order.

V. Simulation

The methodology of the statistical experiment (the estimation of the regression function) coincides with [22]. The two-dimensional sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is generated with the p.d.f.

$$f(x, y) = f(y|\,x)p\,(x).$$

The random variable X uniformly distributed on the interval $[0, 5]$, i.e.,

$$p\,(x) = 1/5, \quad x \in [0, 5].$$

At each X_i, the Y_i are modeled as normally distributed variables with the mean

$$r(X_i) = 10[1 + \exp(-0.5X_i)]$$

and with the variance

$$\sigma^2(X_i) = [1 + \exp(-0.5X_i)]^2,$$

i.e.,

$$f(y|\,x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left\{\frac{-[y - r(x)]^2}{2\sigma^2(x)}\right\}.$$

In the estimators, we use the Epanechnikov kernel

$$K(u) = \begin{cases} \dfrac{3}{4\sqrt{5}}\left(1 - \dfrac{u^2}{5}\right), & |u| \le \sqrt{5}, \\ 0, & |u| > \sqrt{5}, \end{cases}$$

and the bandwidths $h_n = n^{-1/5}$.

The empirical MSEs of estimates are: for common estimates

$$\mathsf{u}^2_{n,com} = \frac{1}{m}\sum_{l=1}^{m}[r_{n,com}(x_l) - r(x_l)]^2,$$

for recursive estimates

$$\mathsf{u}^2_{n,rec} = \frac{1}{m}\sum_{l=1}^{m}[r_{n,rec}(x_l) - r(x_l)]^2,$$

where $x_l = l \times 0.5,\ m = 10,$

$$r_{n,com}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)},$$

$$r_{n,rec}(x) = \frac{\sum_{i=1}^{n} \frac{Y_i}{h_i} K\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - X_i}{h_i}\right)}.$$

The results are given in Table 1.

TABLE 1

Empirical MSEs

n	40	60	200	1000
$\mathsf{u}^2_{n,com}$	0.264	0.236	0.128	0.083
$\mathsf{u}^2_{n,rec}$	0.243	0.207	0.065	0.038

VI. Nonparametric semi-recursive identification of the dynamic production function

Generalize the above results, given for independent observations (random samples), to time series. In [21] an autoregressive heteroscedastic model satisfying geometric ergodicity conditions is considered. The approach allows us to estimate dynamic production functions with lagged values of the inputs and output.

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)

$$Y_t = \Psi(Y_{t-i_1}, \ldots, Y_{t-i_m}, X_t) + \xi_t = \Psi(U_t) + \xi_t, \quad (11)$$

where $X_t = (X_{1t}, \ldots, X_{st})$ are exogenous variables,

$$U_t = (Y_{t-i_1}, \ldots, Y_{t-i_m}, X_t),\ 1 \le i_1 < i_2 < \ldots < i_m,$$

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 non-periodic 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, [24]– [28]).

Let $Y_1, \ldots, Y_n$ be observations generated by process (8). The conditional expectation

$$\Psi(x, z) = \Psi(u) = \mathsf{E}(Y_t|U_t = u) = \mathsf{E}(Y_t|u),$$

$(x, z) = u \in \mathsf{R}^{m+s}$, we estimate by the statistic, which is a semi-recursive counterpart of the Nadaraya–Watson estimator [18], [29] (similarly to (8)):

$$\Psi_{n,\,m+s}\,(u) = \frac{\sum_{t=2}^{n} \frac{Y_t}{h_t^{m+s}} \mathbf{K}\left(\frac{u - U_t}{h_t}\right)}{\sum_{t=2}^{n} \frac{1}{h_t^{m+s}} \mathbf{K}\left(\frac{u - U_t}{h_t}\right)}. \quad (12)$$

Since the observations are dependent, investigation of the estimators properties becomes much harder. For example, the main part of the Nadaraya-Watson estimator's AMSE for strongly mixing (s.m.) sequences was found only in 1999 [30].

According to [21], if the observed sequence satisfies the s.m. condition with an s.m. coefficient $\alpha(\tau)$ such that

$$\int_0^\infty \tau^2 [\alpha(\tau)]^{\frac{\delta}{2+\delta}} d\tau < \infty \tag{13}$$

for some $0 < \delta < 2$, then Theorem 1 holds. Note that an s.m. coefficient with the geometric rate satisfies condition (13).

We will examine the dependence of Russian Federation's Industrial Production Index (IPI) Y on the dollar exchange rate X_1, import X_2, and direct investment 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/. Apply (12) under

$$U_t = (Y_{t-1}, X_{1t}, X_{2t}, X_{3t}, X_{3(t-1)}). \tag{14}$$

The structure of data (14) provides the following estimator for Y_n :

$$\hat{Y}_n = \Psi_{n,5}\left(Y_{n-1}, X_{1n}, X_{2n}, X_{3n}, X_{3(n-1)}\right) = $$
$$= \sum_{t=2}^{n-1} Y_t \frac{K_t}{H_t} \Big/ \sum_{t=2}^{n-1} \frac{K_t}{H_t}, \tag{15}$$

where $H_t = \prod_{j=1}^{5} h_{jt}$, and the five-dimensional kernel K_t is defined by the formula

$$K_t = K\left(\frac{Y_{n-1} - Y_{t-1}}{h_{1t}}\right) \prod_{j=1}^{3} K\left(\frac{X_{jn} - X_{jt}}{h_{(j+1)t}}\right) \times$$
$$\times\, K\left(\frac{X_{3(n-1)} - X_{3(t-1)}}{h_{5t}}\right).$$

The kernel used is the Gaussian kernel and the bandwidths

$$h_{jt} = 0.17 \hat{\sigma}_j t^{-1/9},$$

where $\hat{\sigma}_j,\ j = 1,2,3,4,5$ are the corresponding sample standard deviations, the constant 0.17 is chosen subjectively. To compare nonparametric algorithm (NPA) (15) and the least-squares estimator (LSE), we has calculated the relative average error (RAE) A and relative average annual errors (RAAEs) $A(t), t = 1995, \ldots, 2003$, for both the approaches:

$$A = \frac{1}{113} \sum_{i=1}^{113} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|, \quad A(t) = \frac{1}{12} \sum_{i=1}^{12} \left| \frac{Y_i(t) - \hat{Y}_i(t)}{Y_i(t)} \right|,$$

$$A(2004) = \frac{1}{3} \sum_{i=1}^{3} \left| \frac{Y_i(2004) - \hat{Y}_i(2004)}{Y_i(2004)} \right|,$$

where Y_i is the IPI true value, and $\hat{Y}_i$ is its estimate. The results are given in Tables 2 and 3.

TABLE 2

RAAEs of Identification

t	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
$A_{NPA}(t)$	0.046	0.043	0.045	0.075	0.044	0.039	0.032	0.041	0.051	0.056
$A_{LSE}(t)$	0.076	0.048	0.041	0.048	0.040	0.029	0.035	0.033	0.031	0.036

TABLE 3

RAEs of Identification

A_{NPA}	0.047
A_{LSE}	0.042

The result of 1998 can be explained by 1998 Russian financial crisis ("Ruble crisis") in August 1998.

VII. Forecasting

To predict the IPI we will apply (12) under

$$U_t = (Y_{t-1}, X_{1(t-1)}, X_{2(t-1)}, X_{3(t-1)}, X_{3(t-2)}). \tag{16}$$

The structure of data (16) provides the following forecast for Y_n :

$$\hat{Y}_n = \Psi_{n,5}\left(Y_{n-1}, X_{1(n-1)}, X_{2(n-1)}, X_{3(n-1)}, X_{3(n-2)}\right) =$$
$$= \sum_{t=3}^{n-1} Y_t \frac{K_t}{H_t} \Big/ \sum_{t=3}^{n-1} \frac{K_t}{H_t}, \tag{17}$$

where $H_t = \prod_{j=1}^{5} h_{jt}$, and the five-dimensional kernel K_t is defined by the formula

$$K_t = K\left(\frac{Y_{n-1} - Y_{t-1}}{h_{1t}}\right) \prod_{j=1}^{3} K\left(\frac{X_{j(n-1)} - X_{j(t-1)}}{h_{(j+1)t}}\right) \times$$
$$\times\, K\left(\frac{X_{3(n-2)} - X_{3(t-2)}}{h_{5t}}\right).$$

Statistic (17) may be interpreted as the predicted value based on the past information.

To find the AMSE of the estimator $\Psi_{n,5}(u)$ we use Theorem 2 [21].

Suppose that

$$K(\cdot) \in \mathcal{A}_\nu, \quad \mathbf{K}(u) = \prod_{i=1}^{5} K(u_i), \quad \sup_{u \in \mathsf{R}^1} |K(u)| < \infty,$$

the sequence $(h_n) \in \mathcal{H}(5)$, and $\lambda = -5$. Let functions $a_i(u)$, $i = 0, 1$, and their derivatives up to and including the order ν be continuous and bounded on R^5; functions $\int y^2 f(u,y)\, dy$ and $\int y^4 f(u,y)\, dy$ be bounded on R^5; and, moreover, $\int y^2 f(u,y)\, dy$ and $\int |y|^{2+\delta} f(u,y)\, dy$ be continuous at the point u. Then conditions (1)–(5) of Theorem 2 [21] hold; we also suppose that condition (6) (Theorem 2 [21]) holds. If $p(u) > 0$, then condition (7) (Theorem 2 [21]) holds too.

If the random variables Y_t are uniformly bounded, and we select a nonnegative kernel, then it is easy to show that $\Psi_{n,5}(u)$ are bounded for $\nu = 2$. By condition (8) (Theorem 2 [21]), this is equivalent to the existence of a majorizing sequence with $\gamma = 0$.

For $\nu > 2$ the piecewise smooth approximation solves the problem (see the previous section).

In Table 4 the RAAEs of forecasting for each year from 1995 till 2004 are given when we use the NPA and LSE. Similarly, in Table 5 the RAEs of forecasting for such estimators are given.

TABLE 4
RAAEs of Forecasting

t	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
$A_{NPA}(t)$	0.049	0.058	0.058	0.052	0.048	0.040	0.037	0.042	0.048	0.051
$A_{LSE}(t)$	0.113	0.045	0.039	0.049	0.027	0.030	0.029	0.020	0.020	0.023

TABLE 5

RAEs of Forecasting

A_{NPA}	0.048
A_{LSE}	0.040

The marginal productivity function and marginal rate of technical substitution are estimated in the same way on the base of (10).

VIII. Conclusion

This work presents a unifying approach to estimating both the statical and dynamical production function and its characteristics (the marginal productivity function, marginal rate of technical substitution, conditional variance). The approach is based on plug-in estimating of functions depending on functionals of the joint stationary distribution of the vector of explanatory variables $U_t = (Y_t, Y_{t-i_1}, \ldots, Y_{t-i_m}, X_{1t}, \ldots, X_{st})$, where $X_t = (X_{1t}, \ldots, X_{st})$ are exogenous variables, Y_t is an output (product), $i_2 < \ldots < i_m$ is the known subsequence of natural numbers. Note that i_m may be large, while m is small. We assume that the process Y_t is a nonlinear homoscedastic ARX process, strictly stationary and satisfies to the s. m. condition with the geometric rate. The plug-in estimators are semi-recursive, i.e., we recursively compute only the kernel estimators of functionals (3). By using the piecewise smooth approximations of the estimators, we have managed to avoid the problems concerning to the majorizing sequence's existence needed for obtaining of the main part of the estimator's AMSE.

References

[1] G. Koshkin and A. Kitayeva, “Nonparametric identification of the production functions, ”*Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2011, WCE 2011, 6-8 July, 2011, London, U.K.*, vol. 1, pp. 276-280.

[2] M. G. Kendall and A. Stuart, *Inference and Relationship*, 3rd edition. London: Griffin, 1973.

[3] S. R. Rao, *Linear Statistical Methods and Its Applications*. New York: Willey, 1965.

[4] C. T. Wolverton and T. J. Wagner, “Recursive estimates of probability densities,” *IEEE Trans. Syst. Sci. and Cybernet.*, vol. 5, no. 3, pp. 246–247, 1969.

[5] H. Yamato, “Sequential estimation of a continuous probability density function and mode,” *Bulletin of Mathematical Statistics*, vol. 14, pp. 1–12, 1971.

[6] E. J. Wegman and H. I. Davies, “Remarks on some recursive estimates of a probability density function,” *Ann. Statist.*, vol. 7, no. 2, pp. 316-327, 1979.

[7] J. A. Ahmad and P. E. Lin, “Nonparametric sequential estimation of a multiple regression function,” *Bull. Math. Statist.*, vol. 17, no. 1–2, pp. 63–75, 1976.

[8] V. M. Buldakov and G. M. Koshkin, “On the recursive estimates of a probability and a regression line,” *Problems Inform. Trans.*, vol. 13, pp. 41–48, 1977.

[9] L. Devroye and T. J. Wagner, “On the L_1 convergence of kernel estimates of regression functions with applications in discrimination,” *Z. Wahrsch. Verw. Gebiete*, vol. 51, pp. 15–25, 1980.

[10] A. Krzyźak and M. Pawlak, “Almost everywhere convergence of a recursive regression function estimate and classification,” *IEEE Trans. Inform. Theory*, vol. IT–30, pp. 91–93, 1984.

[11] W. Greblicki and M. Pawlak, “Necessary and sufficient consistency conditions for recursive kernel regression estimate,” *J. Multivariate Anal.*, vol. 23, pp. 67–76, 1987.

[12] A. Krzyźak, “Global convergence of the recursive kernel estimates with applications in classification and nonlinear system estimation,” *IEEE Trans. Inform. Theory*, vol. IT–38, pp. 1323–1338, 1992.

[13] L. Györfi, M. Kohler and H. Walk, “Weak and strong universal consistency of semi-recursive kernel and partitioning regression estimates,” *Statist. Decisions*, vol. 16, pp. 1–18, 1998.

[14] H. Walk, “Strong universal pointwise consistency of recursive kernel regression estimates,” *Ann. Inst. Statist. Math.*, vol. 53, no. 4, pp. 691–707, 2001.

[15] H. Cramér, *Mathematical Methods of Statistics*. Princeton, 1948.

[16] G. M. Koshkin, “Deviation moments of the substitution estimate and its piecewise smooth approximations,” *Siberian Math. J.*, vol. 40, no. 3, pp. 515–527, 1999.

[17] A. V. Kitaeva and G. M. Koshkin, “Recurrent nonparametric estimation of functions from functionals of multidimensional density and their derivatives”, *Autom. Remote Control*, no. 3, pp. 389-407, 2009.

[18] E. A. Nadaraya, “On regression estimates”, *Theor. Prob. App.*, vol. 19, no. 1, pp. 147-149, 1964.

[19] G. Collomb, “Estimation non parametrique de la regression par la methode du noyau,” Thèse Docteur Ingénieur, Univ. Paul-Sabatier, Toulouse, 1976.

[20] D. Tjøstheim and B. H. Auestad, “Nonparametric identification of nonlinear time series: projections,” *J. Amer. Statist. Associat.*, vol. 89, no. 428, pp. 1398–1409, 1994.

[21] A. V. Kitaeva and G. M. Koshkin, “Semi-recursive nonparametric identification in the general sense of a nonlinear heteroscedastic autoregression,” *Autom. Remote Control*, vol. 71, no. 2, pp. 257–274, 2010. Available: DOI: 10.1134/S0005117910020086

[22] N. N. Aprausheva and V. D. Konakov, “Use of nonparametric estimates in regressive analysis,” *Zavod. Lab.*, no. 5, pp. 556–569, 1976 (in Russian).

[23] G. B. Klejner, *Production Functions: Theory, Methods, Application*. Moscow: Finances and Statistics, 1986 (in Russian).

[24] E. Masry and D. Tjøstheim, “Nonparametric estimation and identification of nonlinear ARCH time series”, *Econom. Theory*, vol. 11, pp. 258–289, 1995.

[25] Z. D. Lu, “On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term”, *Statist. Sinica*, vol. 8, pp. 1205-1217, 1998.

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

[27] P. Doukhan and A. Tsybakov, “Nonparametric robust estimation in nonlinear ARX models”, *Problems of Information Transmission*, vol. 29, no. 4, pp. 318-327, 1993.

[28] P. Doukhan, *Mixing: properties and examples*. *Lecture Notes in Statistics*, vol. 85, Springer-Verlag, 1994.

[29] G. S. Watson, “Smooth regression analysis”, *Sankhya. Indian J. Statist.*, vol. A26, pp. 359–372, 1964.

[30] D. Bosq, N. Cheze-Payaud, “Optimal asymptotic quadratic error of nonparametric regression function estimates for a continuous-time process from sampled-data”, *Statistics*, vol. 32. pp. 229–247, 1999.