DE GRUYTER Depend. Model. 2020; 8:221–238

Research Article Open Access

Yousri Slaoui* and Salah Khardani

Nonparametric relative recursive regression

https://doi.org/10.1515/demo-2020-0013
Received May 15, 2020; accepted August 29, 2020

Abstract: In this paper, we propose the problem of estimating a regression function recursively based on the minimization of the Mean Squared Relative Error (*MSRE*), where outlier data are present and the response variable of the model is positive. We construct an alternative estimation of the regression function using a stochastic approximation method. The Bias, variance, and Mean Integrated Squared Error (*MISE*) are computed explicitly. The asymptotic normality of the proposed estimator is also proved. Moreover, we conduct a simulation to compare the performance of our proposed estimators with that of the two classical kernel regression estimators and then through a real Malaria dataset.

Keywords: nonparametric regression, stochastic approximation algorithm, smoothing, curve fitting, relative regression

MSC: 62G08, 62L20, 65D10

1 Introduction

Nonparametric regression provides a useful diagnostic tool for data analysis. A useful mathematical model is to estimate the link between a Borelian function $m(T)$ and X, by means of a function $r(x)$ which achieves the minimum of the mean squared error (MSE)

$$\mathbb{E}\Big(r(X)-m(T)\Big)^2=\min_{\eta}\mathbb{E}\Big(\eta(X)-m(T)\Big)^2, \tag{1}$$

based on a random sample of data $(X_1,Y_1),\dots,(X_n,Y_n)$ from a unknown joint density $f(\cdot,\cdot)$, when the covariates (X_i) for $i\in\{1,\dots,n\}$ take values in finite dimensional $\mathbb{R}^d$. Nonparametric regression methods have attracted much attention among statisticians in the last several decades, and a large literature now exists. When it comes to the situation with multiple covariates, multivariate nonparametric regression has been proved to be very useful in practice. [45, 46] have shown that the local regression estimators having optimal rates of convergence, and Cleveland and Devlin [6] have proved that they are very useful in modeling data. [34] derived the asymptotic properties of the multivariate local linear and local quadratic estimators. [49] studied multivariate plug-in bandwidth selection and Herman et al. [14] proposed plug-in approaches for bivariate convolution kernel estimator. Eubank [10], [48], and [12] described thin plate smoothing splines.

Often, in nonparametric estimation, we use the least squares and the least absolute deviation as criteria to construct the predictors. However, for many practical situations the MSRE is more appropriate as measure of performance than the two previous criteria, see, [18] for some models in software engineering, [4] for some examples in medicine or [5] for some financial applications. Let us underline that, the classical procedure estimation (MSE) is based on some restrictive condition that is the homoscedasticity. This consideration gives the same weight for all observations, which is inadequate when the data contains some outliers.

Although relative error is not widely studied in the statistical literature there are methods designed with rel-

***Corresponding Author: Yousri Slaoui:** Univ. de Poitiers, Lab. de Mathématiques et App., Futuroscope Chasseneuil, France, E-mail: Yousri.Slaoui@math.univ-poitiers.fr
Salah Khardani: Laboratoire des Réseaux Intelligents et Nanotechnologie, Ecole Nationale des Sciences et Technologies Avancées à Borj-Cédria, Tunisia

ative error performance. We can list the work of [28], which consist on the study of an estimation method for minimizing the sum of absolute relative residuals. However, [11] developed an estimation method designed to reduce absolute relative-error. Moreover, [18] studied the asymptotic properties of the estimators by minimizing the sum of the squared relative errors. [15] introduced and studied local constant and local linear nonparametric regression estimators when it is appropriate to assess performance in terms of mean squared relative error of prediction. [51] established the connection between relative error estimators and the M-estimation in the linear model. [2] considered the case of spatial data. [7] considered the case where the explanatory variable are of functional type of data, [1] investigate the functional nonparametric regression estimation in the case when the response is subject to left-truncation by an other random variable, [43] considered the case of recursive estimation of the regression estimation in the case of the functional data, while [8] study the M-estimation of the functional nonparametric regression when the response variable is subject to left-truncation by an other random variable. Relative error is sometimes a more meaningful measure of performance of a predictor than the absolute error. Generally, this occurs when the range of predicted values is large.
In this paper, we construct an alternative kernel estimate regression function using a recursive methods by considering the problem of estimating the regression function based on the minimization of the MSRE.

We address recursive kernel estimators for which recursive means that the estimator calculated from the first n observations, say f_n, is a function of only f_{n-1} and the $(n)^{-th}$ observation.
The Robbins-Monro algorithm was originally proposed by [32] and further developed and investigated as well as applied in many different situations (see, among many others, [9, 19–21, 25, 26, 30, 31, 33, 37, 39–42, 47]).

As is well known, such a recursive property works well within the framework a data streams. Streaming data are massive data arriving in streams, and if they are not processed immediately or stored, then they are lost forever. The sample data are obtained by means of an observational mechanism that allows for a rapid increase in the sample size over time. In recent years, data streams have become an increasingly important area of research. Common data streams include Twitter activity, the Facebook news stream, Internet packet data, stock market activity, credit card transactions and Internet and phone usage. In those situations, the data arrive so rapidly that it is impossible for the user to store them all in disk (as a traditional database), and then interact with them at the time of our choosing. Consequently, to deal with such big data, the traditional nonparametric techniques rapidly require a lot of time to be computed and therefore become useless in practice. Therefore, the development of methods of processing and analyzing these data streams effectively and efficiently has become a challenging problem in statistics and computational science. This is why we consider the regression estimation problem in the context of data streams in this paper. This recursive estimator shows good theoretical properties, from the point of view of relative mean square error.

The general idea of the proposed recursive methods is described in Section 2. Asymptotic MSRE properties of the recursive regression estimator are given and discussed in Section 3. A simulation study is presented in Section 4. In Section 5, we consider a real Malaria dataset. We conclude the paper in Section 6, whereas the technical details are deferred to Section 7.

2 Presentation of estimates

Given identically distributed (i.i.d.) observations $(X_1, Y_1), \dots (X_n, Y_n)$ with joint density function $f(x, y)$, and f denote the probability density of X. In regression analysis, our interest is the estimation of Y given X, which consist on finding a function η(X) which satisfies the problem (1).

In order to construct a stochastic algorithm for the estimation of the regression function $r(x)$ at a point x, we define an algorithm that calculates the zero of the function $h : y \to r(x) - y$. Because x is a fixed point, then the value $r(x)$ is the unique solution of the equation $h(y) = 0$ with unknown y.

A typical kernel based estimator of the Robbins-Monro's procedure (see [32]) is

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

However, the use of previous loss function as a measure of prediction performance may be not suitable in some situation. In particular, in the case when the presence of outliers can lead to unreasonable results since all variables have the same weight. Now, to overcome this limitation we propose to estimate the function r by an alternative loss function.

In the relative regression analysis $r(x)$ is obtained by minimizing the mean squared relative error (MSRE) ie: $r(x)$ is the solution of the optimisation problem:

$$r(x) = \arg\min_{\theta}\left(\mathbb{E}\left[\left(\frac{Y-\theta(x)}{Y}\right)^2 |X=x\right]\right), \quad \text{for} \quad Y>0.$$

It is clear that this criterion is a more meaningful measure of prediction performance than the least squares error, in particular, when the range of predicted values is large.
Moreover, the solution of this problem can be expressed by the ratio of first two conditional inverse moments of Y given X. As proposed by [29], $r(x) = \frac{\mathbb{E}[Y^{-1}|X=x]}{\mathbb{E}[Y^{-2}|X=x]}$ is the best MSRE predictor of Y given X. Thus, we can estimate $r(x)$ by

$$\widetilde{r}_n(x) = \frac{\phi_n(x)}{\psi_n(x)},$$

where $\phi_n(x)$ is an estimator of $\mathbb{E}\left[Y^{-1}|X=x\right]f(x)$ and $\psi_n(x)$ is an estimator of $\mathbb{E}\left[Y^{-2}|X=x\right]f(x)$. In order to construct a stochastic algorithm for the estimation of the regression function $\phi : x \to \mathbb{E}\left[Y^{-1}|X=x\right]f(x)$ at a point x, we define an algorithm of search of the zero of the function $h : y \to \phi(x) - y$. Following Robbins-Monro's procedure (see [32]), this algorithm defined by setting $\phi_0(x) \in \mathbb{R}$, and, for all $n \geqslant 1$,

$$\phi_n(x) = \phi_{n-1}(x) + \gamma_n W_n(x),$$

where $W_n(x)$ is an observation of the function h at the point $\phi_{n-1}(x)$, and the stepsize (γ_n) is a sequence of positive real numbers that goes to zero. Taking $W_n(x) = h_n^{-d} Y_n^{-1} K\left(\frac{x-X_k}{h_k}\right) - \phi_{n-1}(x)$, then, the estimator ϕ_n to recursively estimate the function ϕ at the point x can be written as

$$\phi_n(x) = (1-\gamma_n)\phi_{n-1}(x) + \gamma_n h_n^{-d} Y_n^{-1} K\left(\frac{x-X_k}{h_k}\right). \tag{2}$$

We let $\phi_n(0) = 0$, then, we can estimate ϕ recursively at the point x by

$$\phi_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-d} Y_k^{-1} K\left(\frac{x-X_k}{h_k}\right),$$

where $\Pi_n = \prod_{i=1}^{n}(1-\gamma_i)$, following similar steps, we can estimate recursively the function ψ at the point x by

$$\psi_n(x) = (1-\gamma_n)\psi_{n-1}(x) + \gamma_n h_n^{-d} Y_n^{-2} K\left(\frac{x-X_k}{h_k}\right), \tag{3}$$

moreover, we let $\psi_n(0) = 0$, then, we can estimate ψ recursively at the point x by

$$\psi_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-d} Y_k^{-2} K\left(\frac{x-X_k}{h_k}\right).$$

Then, our proposal in this paper is the following estimator:

$$r_n(x) = \frac{\sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-d} Y_k^{-1} K\left(\frac{x-X_k}{h_k}\right)}{\sum_{j=1}^{n} \Pi_j^{-1} \gamma_j h_j^{-d} Y_j^{-2} K\left(\frac{x-X_j}{h_j}\right)}. \tag{4}$$

The purpose of this paper is the study of the properties of the proposed relative recursive regression estimators (4), and its comparison with the direct analogue of the well-known Nadaraya-Watson estimator introduced separately by [27] and [50], and defined as

$$\widetilde{r}_n(x) = \frac{\sum_{k=1}^{n} Y_k^{-1} K\left(\frac{x-X_k}{h_n}\right)}{\sum_{j=1}^{n} Y_j^{-2} K\left(\frac{x-X_j}{h_n}\right)}. \tag{5}$$

This estimator was proposed in[15]. However, the strong consistency and the asymptotic normality of this estimator under weak dependence conditions is given in [22], while the case of censored data was considered in [16].

3 Assumptions and main results

We define the following class of regularly varying sequences.

Definition 1. Let $\gamma \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if

$$\lim_{n\to+\infty} n\left[1-\frac{v_{n-1}}{v_n}\right]=\gamma. \tag{6}$$

Condition (6) was introduced by [13] to define regularly varying sequences (see also [3]) and by [24] in the context of stochastic approximation algorithms. Noting that the acronym $\mathcal{GS}$ stand for (Galambos and Seneta). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b\in\mathbb{R}$, $n^{\gamma}(\log n)^b$, $n^{\gamma}(\log\log n)^b$, and so on.
In this section, we investigate the asymptotic properties of our proposed estimators (4). The assumptions to which we shall refer are the following

(A1) $K:\mathbb{R}^d\to\mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}^d}K(z)\,dz=1$, and, for all $j\in\{1,\ldots,d\}$, $\int_{\mathbb{R}} z_jK(z)\,dz_j=0$ and $\int_{\mathbb{R}^d} z_j^2\|K(z)\|dz<\infty$.

(A2) *i*) $(\gamma_n)\in\mathcal{GS}(-\alpha)$ with $\alpha\in\left(1/2,1\right]$.
ii) $(h_n)\in\mathcal{GS}(-a)$ with $a\in\left(0,\alpha/d\right)$.
iii) $\lim_{n\to\infty}(n\gamma_n)\in\left(\min\{2a,(\alpha-ad)/2\},\infty\right]$.

(A3) *i*) $f(s,t)$ is twice continuously differentiable with respect to s.
ii) For $q\in\{-4,-3,-2,-1,0\}$, $s\mapsto\int_{\mathbb{R}}t^qf(s,t)\,dt$ is a bounded function continuous at $s=x$.
For $q\in[-3,-2]$, $s\mapsto\int_{\mathbb{R}}|t|^qf(s,t)\,dt$ is a bounded function and for $q'\in[-5,-4]$, $s\mapsto\int_{\mathbb{R}}|t|^{q'}f(s,t)\,dt$ is a bounded function..
iii) For $q\in\{-2,-1,0\}$, $\int_{\mathbb{R}}|t|^q\left|\frac{\partial f}{\partial x}(x,t)\right|dt<\infty$, and $s\mapsto\int_{\mathbb{R}}t^q\frac{\partial^2 f}{\partial s^2}(s,t)\,dt$ is a bounded function continuous at $s=x$.
iv) The function $\psi(x)>0$ and the inverse moments of the response variable $\forall m\geq 1$, $\mathbb{E}\left[Y^{-m}|X=x\right]<C<\infty$.

Discussion of the assumptions

It is interesting to underline that the intuition behind the use of such bandwidth (h_n) belonging to $\mathcal{GS}(-a)$ is that the ratio h_{n-1}/h_n is equal to $1+a/n+o\left(1/n\right)$, then using such bandwidth and using the assumption $(A2)$ on the bandwidth and on the stepsize, Lemma 2 ensures that the bias and the variance will depend only on h_n and not on $h_1,\ldots,h_n$, then the *MISE* will depend also only on h_n, which will be helpful to deduce an optimal bandwidth. Moreover, in order to help the readers to follow the main results obtained in this paper, we underline that under the assumption $(A2)$, we have $\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_k=1+o(1)$, $\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_kh_k^2=O\left(h_n^2\right)$ and $\Pi_n^2\sum_{k=1}^n\Pi_k^{-2}\gamma_k^2h_k^{-d}=O\left(\gamma_nh_h^{-d}\right)$. Assumptions $(A1)$ and $(A3)$ are regularity conditions which permit us to evaluate the bias term, the variance term of the estimator (4). Moreover, $(A3)$ include some technical condition to attain brevity of proofs and to obtain a convergence rate. Somme work in progress plain to consider less restrictive conditions. Assumption $(A2)(iii)$ is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\gamma_n]^{-1}\right)$ is finite. For simplicity, we introduce the following notations:

$$\xi = \lim_{n\to\infty}(n\gamma_n)^{-1}, \tag{7}$$

$$R(K) = \int_{\mathbb{R}^d}K^2(z)\,dz, \qquad \mu_j(K)=\int_{\mathbb{R}}z^jK(z)\,dz,$$

$$\phi_{j,j}^{(2)}(x) = \frac{\partial^2\phi}{\partial x_i \partial x_j}(x), \qquad \psi_{j,j}^{(2)}(x) = \frac{\partial^2\psi}{\partial x_i \partial x_j}(x),$$

$$\mathcal{B}(x) = \frac{1}{\psi(x)}\sum_{j=1}^{d}\left(\mu_j^2\left[\phi_{j,j}^{(2)}(x) - r(x)\psi_{j,j}^{(2)}(x)\right]\right),$$

$$\mathcal{V}(x) = \frac{\left\{\mathbb{E}\left[Y^{-2}|X=x\right] + r^2(x)\mathbb{E}\left[Y^{-4}|X=x\right] - 2r(x)\mathbb{E}\left[Y^{-3}|X=x\right]\right\}f(x)}{\psi^2(x)}.$$

3.1 Results on the relative recursive regression estimators 4

In this section, we explicit the choice of the bandwidth (h_n) through a plug-in method, which consist on considering an asymptotic unbiased estimator of the unknown quantities which can be appeared in the expression of the theoretical bandwidth and then in the expression the corresponding *MISE*. Our first result is the following proposition, which gives the bias and the variance of r_n.

Theorem 1. [Bias and variance of r_n]. Let Assumptions (A1)–(A3) hold, and assume that, for all $i, j \in \{1, \dots, d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x.
If $a \in [0, \alpha/(d+4)]$, then

$$\mathbb{E}\left[r_n(x)\right] - r(x) = \frac{h_n^2}{2(\alpha - 2a\xi)}\mathcal{B}(x) + o\left(h_n^2\right). \tag{8}$$

If $a \in (\alpha/(d+4), 1)$, then

$$\mathbb{E}\left[r_n(x)\right] - r(x) = o\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{9}$$

If $a \in [\alpha/(d+4), 1)$, then

$$Var\left[r_n(x)\right] = \frac{\gamma_n h_n^{-d}}{(2-(\alpha - ad)\xi)}\mathcal{V}(x)R(K)\frac{\gamma_n}{h_n^d} + o\left(\gamma_n h_n^{-d}\right). \tag{10}$$

If $a \in [0, \alpha/(d+4))$, then

$$Var\left[r_n(x)\right] = o\left(h_n^4\right). \tag{11}$$

The bias and the variance of the estimator r_n defined by the stochastic approximation algorithm (4) then heavily depend on the choice of the stepsize (γ_n).
We propose now to state the following theorem, which gives the weak convergence rate of the estimator r_n defined in (4).

Theorem 2 (Weak pointwise convergence rate). Let Assumptions $(A1)-(A3)$ hold, and assume that, for all $i, j \in \{1, \dots, d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x.

1. If there exists $c \geq 0$ such that $\gamma_n^{-1}h_n^{d+4} \to c$, then

$$\sqrt{\gamma_n^{-1}h_n^d}\left(r_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\frac{\sqrt{c}}{2(\alpha-2a\xi)}\mathcal{B}(x), \frac{1}{(2-(\alpha-ad)\xi)}\mathcal{V}(x)R(K)\right),$$

2. If $nh_n^{d+4} \to \infty$, then

$$\frac{1}{h_n^2}\left(r_n(x) - r(x)\right) \stackrel{\mathbb{P}}{\to} \frac{1}{2\psi(x)(\alpha - 2a\xi)}\mathcal{B}(x),$$

where $\stackrel{\mathcal{D}}{\to}$ denotes the convergence in distribution, $\mathcal{N}$ the Gaussian-distribution and $\stackrel{\mathbb{P}}{\to}$ the convergence in probability.

Let us now consider the case where the bandwidth h_n is chosen so that $\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+4} = 0$ (which corresponds to under-smoothing). Thus, the proposed estimator satisfies the following central limit theorem:

$$\sqrt{\gamma_n^{-1}h_n^d}\left(r_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{1}{(2-(\alpha-ad)\xi)}\mathcal{V}(x)R(K)\right).$$

Let ϕ denote the distribution function $\mathcal{N}(0,1)$, and $t_{\alpha/2}$ be such that $\phi\left(t_{\alpha/2}\right) = 1 - t_{\alpha/2}$ (where $\alpha \in (0,1)$). Then the approximate asymptotic confidence interval of $r(x)$, with level $1-\alpha$, is given by

$$\left[r_n(x) - \phi\left(t_{\alpha/2}\right)\sqrt{\frac{1}{(2-(\alpha-ad)\xi)}}\sqrt{\frac{\widehat{\mathcal{V}}(x)}{\gamma_n^{-1}h_n^d}}, r_n(x) + \phi\left(t_{\alpha/2}\right)\sqrt{\frac{1}{(2-(\alpha-ad)\xi)}}\sqrt{\frac{\widehat{\mathcal{V}}(x)}{\gamma_n^{-1}h_n^d}}\right], \tag{12}$$

where $\widehat{\mathcal{V}}(x)$ is the empirical estimator of $\mathcal{V}(x)$, we estimate r by r_n and f by f_n, where f_n in the recursive kernel density estimator:

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^{-d} K\left(\frac{x-X_k}{h_k}\right).$$

In order to obtain a theoretical expression of the bandwidth (h_n), we state the following proposition, which gives the *MISE* of the estimator r_n.

Proposition 1. Let Assumptions $(A1)-(A3)$ hold, and assume that, for all $i,j \in \{1,\ldots,d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x.

1. If $a < \alpha/(d+4)$, then

$$MISE[r_n] = \frac{h_n^4}{4(\alpha-2a\xi)^2}\int_{\mathbb{R}^d}\mathcal{B}^2(x)\,dx + o\left(h_n^4\right).$$

2. If $a = \alpha/(d+4)$, then

$$MISE[r_n] = \frac{h_n^4}{4(\alpha-2a\xi)^2}\int_{\mathbb{R}^d}\mathcal{B}^2(x)\,dx + \frac{\gamma_n h_n^{-d}}{(2-(\alpha-ad)\xi)}R(K)\int_{\mathbb{R}^d}\mathcal{V}(x)\,dx + o\left(h_n^4 + \frac{\gamma_n}{h_n^d}\right).$$

3. If $a > \alpha/(d+4)$, then

$$MISE[r_n] = \frac{\gamma_n h_n^{-d}}{(2-(\alpha-ad)\xi)}R(K)\int_{\mathbb{R}^d}\mathcal{V}(x)\,dx + o\left(\gamma_n h_n^{-d}\right).$$

The following corollary is a direct consequence of the previous proposition,

Corollary 1. Let Assumptions $(A1)-(A3)$ hold, and assume that, for all $i,j \in \{1,\ldots,d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x and $\mathcal{B}(x) \neq 0$. To minimize the *MSE* of r_n at the point x, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and such that $(\gamma_n) = \left(\gamma_0 n^{-1}\right)$, the bandwidth (h_n) must equal

$$\left(\left[\frac{d(\alpha-2a\xi)^2}{(2-(\alpha-ad)\xi)}\frac{R(K)\int_{\mathbb{R}^d}\mathcal{V}(x)\,dx}{\int_{\mathbb{R}^d}\mathcal{B}^2(x)\,dx}\right]^{1/(d+4)}\gamma_n^{1/(d+4)}\right),$$

and then the corresponding *MISE*

$$\begin{aligned} MISE[r_n] &= \frac{d+4}{4}d^{-d/(d+4)}\gamma_0^2\left(\gamma_0 - 2/(d+4)\right)^{-(2d+4)/(d+4)} \\ &\times \left[\int_{\mathbb{R}^d}\mathcal{B}^2(x)\,dx\right]^{d/(d+4)}\left[R(K)\int_{\mathbb{R}^d}\mathcal{V}(x)\,dx\right]^{4/(d+4)} n^{-4/(d+4)}[1+o(1)]. \end{aligned}$$

We can observe that the proposed optimal bandwidth depends on the following unknown quantities: $\mathcal{V}$ and $\mathcal{B}$, in order to overcome this problem, we followed the plug-in method proposed in [36] , which leads to consider the following kernel estimators

$$\widehat{\mathcal{V}} = \frac{\Pi_n}{n}\sum_{\substack{i,k=1\\ i\neq k}}^{n}\Pi_i^{-1}\gamma_i b_i^{-d}\left\{Y_k^{-1} - \frac{\sum_{\substack{i,j=1\\ i\neq j}}^{n}\Pi_j^{-1}\gamma_j b_j^{-d}Y_j^{-1}K_b\left(\frac{X_i-X_j}{b_j}\right)}{\sum_{\substack{i,j=1\\ i\neq j}}^{n}\Pi_j^{-1}\gamma_j b_j^{-d}Y_j^{-2}K_b\left(\frac{X_i-X_j}{b_j}\right)}Y_k^{-2}\right\}^2 K_b\left(\frac{X_i-X_k}{b_k}\right),$$

and

$$\widehat{\mathcal{B}} = \frac{\Pi_n^2}{n}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n}\Pi_j^{-1}\Pi_k^{-1}\gamma_j\gamma_k b_j'^{-d-2}b_k'^{-d-2}\left\{Y_k^{-1}-\frac{\sum_{\substack{i,j=1\\ i\neq j}}^{n}\Pi_j^{-1}\gamma_j b_j'^{-d}Y_j^{-1}K_{b'}\left(\frac{X_i-X_j}{b_j'}\right)}{\sum_{\substack{i,j=1\\ i\neq j}}^{n}\Pi_j^{-1}\gamma_j b_j'^{-d}Y_j^{-2}K_{b'}\left(\frac{X_i-X_j}{b_j'}\right)}Y_k^{-2}\right\}$$
$$K_{b'}^{(2)}\left(\frac{X_i-X_j}{b_j'}\right)K_{b'}^{(2)}\left(\frac{X_i-X_k}{b_k'}\right),$$

where $K_b^{(j)}$ is the j-th derivative of a kernel K_b and b_n the associated bandwidth. We followed the approach proposed in [36, 37] and we showed that b_n and b_n' should belong to $\mathcal{GS}\left(-2/(d+4)\right)$ and $\mathcal{GS}\left(-3/(2(d+4))\right)$, respectively. In practice, we use (13) with $\beta = 2/(d+4)$ and $\beta = 3/(2(d+4))$, respectively. Where

$$b_n = n^{-\beta}\min\left\{\widehat{s}, \frac{Q_3-Q_1}{1.349}\right\}, \quad \beta\in\,]0,1[\tag{13}$$

(see [35] with $\widehat{s}$ the sample standard deviation, and Q_1, Q_3 denoting the first and third quartiles, respectively. Then, we have the following corollary.

Corollary 2. Let Assumptions $(A1)-(A3)$ hold, and assume that, for all $i,j\in\{1,\ldots,d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x and $\mathcal{B}(x)\neq 0$. To minimize the MSE of r_n at the point x, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and such that $(\gamma_n) = \left(\gamma_0 n^{-1}\right)$, the bandwidth (h_n) must equal

$$\left(\left[\frac{d\left(\alpha-2a\xi\right)^2}{\left(2-\left(\alpha-ad\right)\xi\right)}\frac{\widehat{\mathcal{V}}}{\widehat{\mathcal{B}}}\right]^{1/(d+4)}\gamma_n^{1/(d+4)}\right), \tag{14}$$

and then the corresponding $MISE$

$$MISE\left[r_n\right] = \frac{d+4}{4}d^{-d/(d+4)}\gamma_0^2\left(\gamma_0-2/\left(d+4\right)\right)^{-(2d+4)/(d+4)}\widehat{\mathcal{B}}^{d/(d+4)}\widehat{\mathcal{V}}^{4/(d+4)}n^{-4/(d+4)}\left[1+o\left(1\right)\right].$$

3.2 Results on the relative non-recursive regression estimator

Let us claimed the following Lemma which gives the bias and variance of the relative non-recursive regression estimator (5), the proof follows easily from the one of the Theorem 1.

Lemma 1 (Bias and variance of $\widetilde{r}_n$). Let Assumptions $(A1)$, $(A2)$ *ii)* and $(A3)$ hold, and assume that $f^{(2)}$ is continuous at x.

$$\mathbb{E}\left[\widetilde{r}_n\left(x\right)\right]-r\left(x\right) = \frac{1}{2}h_n^2\mathcal{B}\left(x\right)+o\left(h_n^2\right),$$

and

$$Var\left[\widetilde{r}_n\left(x\right)\right] = \frac{1}{nh_n}\mathcal{V}\left(x\right)R\left(K\right)+o\left(\frac{1}{nh_n}\right).$$

Then, it follows from Lemma 1, that

$$MISE\left[\widetilde{r}_n\right] = \frac{1}{nh_n^d}R\left(K\right)\int_{\mathbb{R}^d}\mathcal{V}\left(x\right)dx+\frac{1}{4}h_n^4\int_{\mathbb{R}^d}\mathcal{B}_2^2\left(x\right)dx+o\left(h_n^4+\frac{1}{nh_n^d}\right).$$

Let us now consider the case where the bandwidth h_n is chosen so that $\lim_{n\to\infty}nh_n^{d+4}=0$ (which corresponds to under-smoothing). Thus, the non-recursive estimator $\widetilde{r}_n$ satisfies the following central limit theorem:

$$\sqrt{nh_n^d}\left(\widetilde{r}_n\left(x\right)-r\left(x\right)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(0,\mathcal{V}\left(x\right)R\left(K\right)\right).$$

Let ϕ denote the distribution function $\mathcal{N}(0,1)$, and $t_{\alpha/2}$ be such that $\phi\left(t_{\alpha/2}\right) = 1 - t_{\alpha/2}$ (where $\alpha \in (0,1)$). Then the approximate asymptotic confidence interval of $r(x)$, with level $1-\alpha$, is given by

$$\left[\widetilde{r}_n(x) - \phi\left(t_{\alpha/2}\right)\sqrt{\frac{\widetilde{\mathcal{V}}(x)}{nh_n^d}}, \widetilde{r}_n(x) + \phi\left(t_{\alpha/2}\right)\sqrt{\frac{\widetilde{\mathcal{V}}(x)}{nh_n^d}}\right], \tag{15}$$

where $\widetilde{\mathcal{V}}(x)$ is the empirical estimator of $\mathcal{V}(x)$, we estimate r by $\widetilde{r}_n$ and f by $\widetilde{f}_n$, where $\widetilde{f}_n$ in the non-recursive kernel density estimator:

$$\widetilde{f}_n(x) = \frac{1}{nh_n^d}\sum_{k=1}^{n} K\left(\frac{x - X_k}{h_k}\right).$$

Then, to minimize the asymptotic *MISE* of $\widetilde{r}_n$, the bandwidth (h_n) must equal to

$$\left(d^{1/(d+4)}\left\{\frac{R(K)\int_{\mathbb{R}^d}\mathcal{V}(x)\,dx}{\int_{\mathbb{R}^d}\mathcal{B}^2(x)\,dx}\right\}^{1/(d+4)} n^{-1/(d+4)}\right), \tag{16}$$

and then the corresponding *MISE*

$$MISE\left[\widetilde{r}_n\right] = \frac{d+4}{4} d^{-d/(d+4)}\left[\int_{\mathbb{R}^d}\mathcal{B}^2(x)\,dx\right]^{d/(d+4)}\left[R(K)\int_{\mathbb{R}^d}\mathcal{V}(x)\,dx\right]^{4/(d+4)} n^{-4/(d+4)}\left[1 + o(1)\right].$$

Since (16) depends on the unknown quantities: $\mathcal{V}(x)\,R(K)$ and $\mathcal{B}^2(x)$, we consider the following kernel estimators

$$\widetilde{\mathcal{V}} = \frac{1}{nb_n^d}\sum_{\substack{i,k=1\\ i\neq k}}^{n}\left\{Y_k^{-1} - \frac{\sum_{\substack{i,j=1\\ i\neq j}}^{n} Y_j^{-1}K_b\left(\frac{X_i - X_j}{b_n}\right)}{\sum_{\substack{i,j=1\\ i\neq j}}^{n} Y_j^{-2}K_b\left(\frac{X_i - X_j}{b_n}\right)}Y_k^{-2}\right\}^2 K_b\left(\frac{X_i - X_k}{b_n}\right),$$

and

$$\widetilde{\mathcal{B}} = \frac{\Pi_n^2}{nb_n'^{2(-d-2)}}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n}\left\{Y_k^{-1} - \frac{\sum_{\substack{i,j=1\\ i\neq j}}^{n} Y_j^{-1}K_{b'}\left(\frac{X_i - X_j}{b_n'}\right)}{\sum_{\substack{i,j=1\\ i\neq j}}^{n} Y_j^{-2}K_{b'}\left(\frac{X_i - X_j}{b_n'}\right)}Y_k^{-2}\right\}$$
$$K_{b'}^{(2)}\left(\frac{X_i - X_j}{b_n'}\right)K_{b'}^{(2)}\left(\frac{X_i - X_k}{b_n'}\right).$$

Following similar steps as [36, 37], we showed that b_n and b_n' should belong to $\mathcal{GS}\left(-2/(d+4)\right)$ and $\mathcal{GS}\left(-3/(2(d+4))\right)$, respectively. Then, in practice, we use (13) with $\beta = 2/(d+4)$ and $\beta = 3/(2(d+4))$, respectively. Then, we have

Corollary 3. Let Assumptions $(A1)-(A3)$ hold, and assume that, for all $i, j \in \{1,\ldots,d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x and $\mathcal{B}(x) \neq 0$. To minimize the *MSE* of $\widetilde{r}_n$ at the point x, the bandwidth (h_n) must equal

$$\left(d^{1/(d+4)}\left\{\frac{\widetilde{\mathcal{V}}}{\widetilde{\mathcal{B}}}\right\}^{1/(d+4)} n^{-1/(d+4)}\right), \tag{17}$$

and then the corresponding *MISE*

$$MISE\left[\widetilde{r}_n\right] = \frac{d+4}{4} d^{-d/(d+4)}\widetilde{\mathcal{B}}^{d/(d+4)}\widetilde{\mathcal{V}}^{4/(d+4)} n^{-4/(d+4)}\left[1 + o(1)\right].$$

4 Simulations

In order to investigate the comparison between the three estimators, we consider three sample sizes: 100, 200, and 350 and we use the following model $Z = \mu(X) + \varepsilon$ and $Y = \exp(Z)$, which ensures that the response variable is strictly positive. We consider the standard normal kernel $K(z_1, \ldots, z_d) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{d} z_i^2\right)$, and the following three models proposed in Jones et al. (2008) in the unidimensional case:

a) Model 1: $\mu(X) = 0.5 + 2x$, $X \sim \mathcal{U}(0, I_d)$ and $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I_d\right)$,

b) Model 2: $\mu(X) = \log\left(0.5 + 40(x - 0.5)^2\right)$, $X \sim \mathcal{U}(0, I_d)$ and $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I_d\right)$,

c) Model 3: $\mu(X) = \frac{5}{\sqrt{2\pi}}\left\{\exp\left(-6(x - 0.35)^2\right)\mathbb{1}_{(x<0.5)} + \exp\left(-6(x - 0.65)^2\right)\mathbb{1}_{(x\geq 0.5)}\right\}$, $X \sim \mathcal{U}(0, I_d)$ and $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I_d\right)$,

σ^2 take the values 0.2, 0.5 and 1 and $d \in \{1, 2\}$. For each considered model and sample size n, we approximate the Median *ISE* (Integrated Squared Error) of the $\int_{\mathbb{R}} \psi(x)\left\{\widehat{g}(x) - g(x)\right\}^2 dx$ using $N = 500$ trials of sample size n, where $\widehat{g}$ is one the three considered estimators.

Computational cost

In order to give some comparative elements with the direct analogue of the well-known Nadaraya-Watson estimator (5), including computational costs. We consider a 500 samples of size $n_1 = \lfloor n/2 \rfloor$ (the lower integer part of $n/2$), moreover, we suppose that we receive an additional 500 samples of size $n - n_1$.

This property can be generalized, one can check that it follows from (2) that for all $n_1 \in [0, n-1]$,

$$\begin{aligned}
\phi_n(x) &= \prod_{j=n_1+1}^{n}(1-\gamma_j)\,\phi_{n_1}(x) \\
&\quad + \sum_{k=n_1}^{n-1}\left[\prod_{j=k+1}^{n}(1-\gamma_j)\right]\frac{\gamma_k}{h_k}Y_k^{-1}K\left(\frac{x - X_k}{h_k}\right) + \frac{\gamma_n}{h_n}Y_n^{-1}K\left(\frac{x - X_n}{h_n}\right) \\
&= \alpha_1\phi_{n_1}(x) + \sum_{k=n_1}^{n-1}\beta_k\frac{\gamma_k}{h_k}Y_k^{-1}K\left(\frac{x - X_k}{h_k}\right) + \frac{\gamma_n}{h_n}Y_n^{-1}K\left(\frac{x - X_n}{h_n}\right),
\end{aligned}$$

where $\alpha_1 = \prod_{j=n_1+1}^{n}(1-\gamma_j)$ and $\beta_k = \prod_{j=k+1}^{n}(1-\gamma_j)$. Similarly, it follows from (3) that for all $n_1 \in [0, n-1]$,

$$\begin{aligned}
\psi_n(x) &= \prod_{j=n_1+1}^{n}(1-\gamma_j)\,\psi_{n_1}(x) \\
&\quad + \sum_{k=n_1}^{n-1}\left[\prod_{j=k+1}^{n}(1-\gamma_j)\right]\frac{\gamma_k}{h_k}Y_k^{-2}K\left(\frac{x - X_k}{h_k}\right) + \frac{\gamma_n}{h_n}Y_n^{-2}K\left(\frac{x - X_n}{h_n}\right) \\
&= \alpha_1\psi_{n_1}(x) + \sum_{k=n_1}^{n-1}\beta_k\frac{\gamma_k}{h_k}Y_k^{-2}K\left(\frac{x - X_k}{h_k}\right) + \frac{\gamma_n}{h_n}Y_n^{-2}K\left(\frac{x - X_n}{h_n}\right).
\end{aligned}$$

It is clear, that the use of the proposed estimator (4) can improve considerably the computational cost.

Table 1: Median ISE (approximated using $N = 500$ trials) of three models using three estimators; `non-recursive` correspond to the estimator (5) using the proposed plug-in bandwidth selection (17), `Recursive 1` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = (n^{-1})$, `Recursive 2` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = (h_n / [\sum_{k=1}^{n} h_k])$.

	Model 1 ($d = 1$)								
	$n = 100$			$n = 200$			$n = 350$		
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	0.01033	0.19377	**0.4568**	0.01757	**0.0960**	**0.4085**	0.01531	0.13881	**0.4607**
Recursive 1	0.00992	**0.1830**	0.45785	0.01806	0.10317	0.42320	0.01531	0.13736	0.48405
Recursive 2	**0.0094**	0.18377	0.45753	**0.0175**	0.10204	0.41498	**0.0150**	**0.1366**	0.48243
	Model 1 ($d = 2$)								
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	0.01542	0.21243	**0.5284**	0.03844	**0.1264**	**0.4985**	0.02438	0.13881	**0.4824**
Recursive 1	0.01284	**0.1994**	0.55563	0.03647	0.13425	0.51243	0.01531	0.14788	0.49414
Recursive 2	**0.0124**	0.19885	0.55442	**0.0356**	0.14229	0.50485	**0.0162**	**0.1444**	0.49124
	Model 2 ($d = 1$)								
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	**0.1814**	**0.4762**	0.92207	**0.1425**	**0.3503**	0.79980	**0.1166**	0.30277	0.52950
Recursive 1	0.23856	0.56041	0.92690	0.19976	0.43536	0.80378	0.15889	0.31279	0.53276
Recursive 2	0.21220	0.53840	**0.9207**	0.19241	0.40799	**0.7968**	0.15494	**0.3004**	**0.5151**
	Model 2 ($d = 2$)								
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	**0.2114**	**0.5136**	**1.1221**	**0.1754**	**0.5464**	0.98844	**0.1548**	0.46489	0.74372
Recursive 1	0.25646	0.62244	1.12692	1.31244	0.62418	0.98486	0.17648	0.53194	0.64696
Recursive 2	0.25322	0.62652	1.32292	1.42262	0.64844	**0.9794**	0.18674	**0.5242**	**0.6889**
	Model 3 ($d = 1$)								
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	**0.0266**	**0.3090**	0.92207	0.09172	0.41181	0.86676	**0.1518**	**0.1827**	**0.5890**
Recursive 1	0.02971	0.31436	0.92690	0.09241	0.41481	0.86677	0.15559	0.19767	0.60095
Recursive 2	0.02861	0.31161	**0.9207**	**0.0910**	**0.4110**	**0.8661**	0.15419	0.19272	0.59520
	Model 3 ($d = 2$)								
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	0.03414	**0.3485**	**1.1222**	0.11344	0.64114	0.92484	**0.1728**	**0.2149**	**0.7584**
Recursive 1	**0.0324**	0.35642	1.27884	1.14445	0.64585	0.93534	0.17683	0.23545	0.82147
Recursive 2	0.03568	0.35366	1.24232	**0.1132**	**0.6324**	**0.9143**	0.21435	0.21251	0.79424

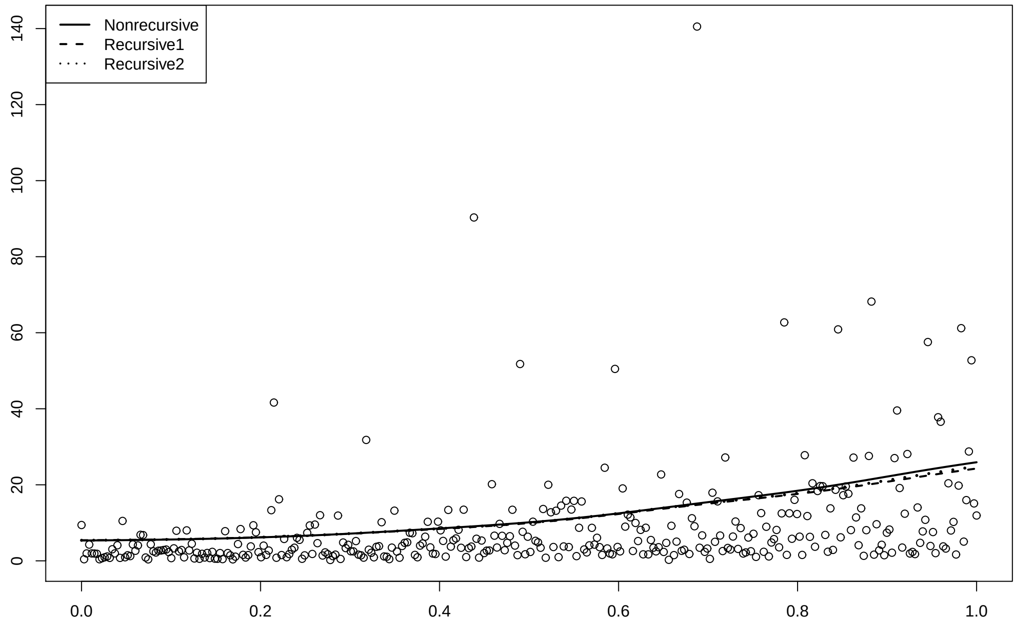

Figure 1: Qualitative comparison between three kernel relative regression estimators; `non-recursive` correspond to the estimator (5) using the proposed plug-in bandwidth selection (17), `Recursive 1` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = (n^{-1})$, `Recursive 2` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = \left(h_n / \left[\sum_{k=1}^n h_k\right]\right)$ using model 1; $\mu(X) = 0.5 + 2x$, $X \sim \mathcal{U}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,1)$.

From Figures 1, 2 and Table 1, we conclude that: 1- The proposed relative recursive regression estimators (4) and (5) are close to the true regression function. 2- The three estimators included our two relative recursive regression estimators (4) with respectively $(\gamma_n) = (n^{-1})$ and $(\gamma_n) = \left((h_n)\left[\sum_{k=1}^n h_k\right]^{-1}\right)$ and (5) performed very well, and that none of the three can be claimed to be best in all cases. 3- The estimators get closer to the true density function as sample size increases.

4.1 Feasibility in term of confidence interval

The aim of this subsection is to compare the performance of the non-recursive relative recursive regression (5) with that of the recursive estimator (4), from confidence interval point of view. We set

$$I_{i,n} = \left[\psi_n(x) - 1.96C(\psi_n)\sqrt{\frac{\mathcal{V}_{i,n}(x)}{nh_n^d}}, \psi_n(x) + 1.96C(\psi_n)\sqrt{\frac{\mathcal{V}_{i,n}(x)}{nh_n^d}}\right],$$

where, when $i = 1$, $\psi_n = \tilde{r}_n$ is the non-recursive estimator (5), and $C(\psi_n) = 1$, $\mathcal{V}_{1,n}(x) = \widetilde{\mathcal{V}}(x)$ and when $i = 2$, $\psi_n = r_n$ is the recursive estimator (4) with the choice $(\gamma_n) = (\gamma_0 n^{-1})$, $C(\psi_n) = \sqrt{\frac{\gamma_0^2}{(2\gamma_0 - (\alpha - ad))}}$ and $\mathcal{V}_{2,n}(x) = \widehat{\mathcal{V}}(x)$. It comes from (15) and (12) that both confidence intervals $I_{1,n}$ and $I_{2,n}$ have the same asymptotic level (equal to 95%), whereas $I_{2,n}$ has a smaller length than $I_{1,n}$. Table 2 give the empirical levels $(\#\left\{r(x) \in I_{i,n}\right\}/N)$ for different values of d, σ^2, the sample size n, by considering $x = 0$ (resp. $x = (0,0)$).

Table 2 shows that the recursive estimator with the choice $(\gamma_n) = \left(h_n / \left[\sum_{k=1}^n h_k\right]\right)$ outperforms the non-recursive estimator and the recursive one with the choice $(\gamma_n) = (n^{-1})$: the empirical levels of the intervals $I_{2,n}$ are greater than those of $I_{1,n}$.

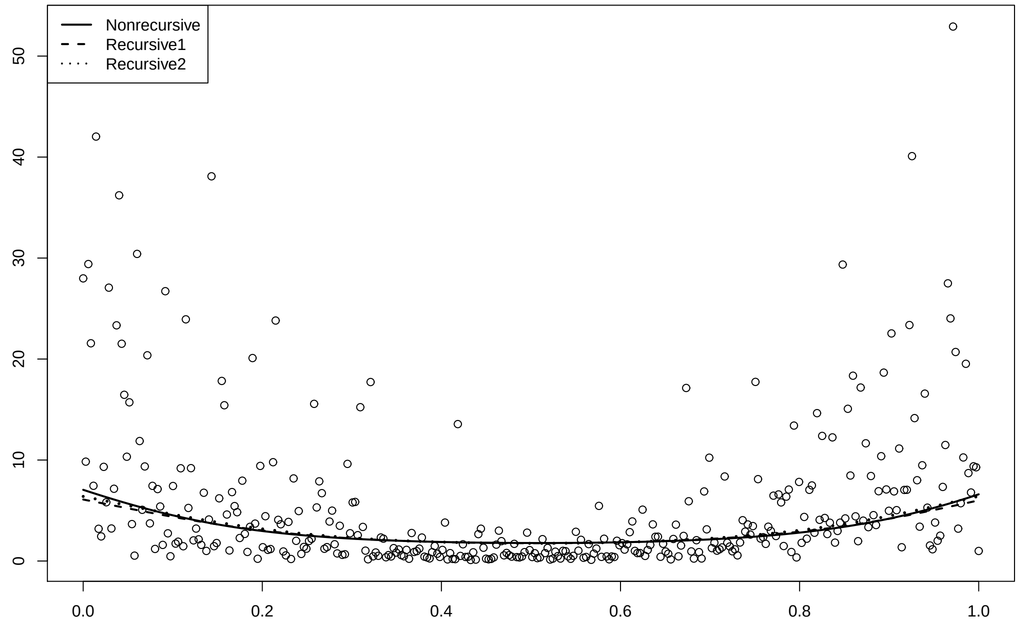

Figure 2: Qualitative comparison between three kernel relative regression estimators; `non-recursive` correspond to the estimator (5) using the proposed plug-in bandwidth selection (17), `Recursive 1` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = (n^{-1})$, `Recursive 2` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = \left(h_n / \left[\sum_{k=1}^{n} h_k\right]\right)$ using model 2; $\mu(X) = \log\left(0.5 + 40(x-0.5)^2\right)$, $X \sim \mathcal{U}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,1)$.

5 Real dataset

We considered a dataset of 176 families in Senegal, totalizing 505 children between 2 and 19 years old, living in two villages of Niakhar (Diohine and Toucar). The number of observations was 6986. We measured `Plasmodium falciparum` Parasite Load (`PL`) from thick blood smears obtained by finger-prick during two different seasons and regularly over a three-year observation period (2001-2003), the number of measurements per child ranged from 1 to 15, for more details see ([23]), this data was used also in [38] in a parametric context.

We had the following variables: 1- `Family identification` : A factor with 176 levels; 2- `Child identification` : A factor with 505 levels; 3- `PL` : Parasite Load (is strictly positive since the 505 children have a positive PL); 4- `infection` : A factor with two levels (infected: 1 or not infected: 0); 5- `year` : A factor with three levels (0 for 2001, 1 for 2002 and 2 for 2003); 6- `number of measurements per child` : A factor with 15 levels; 7- `age` : Age of the child in years between 2 and 19; 8- `season` : A factor with two levels (July-October and October-March); 9- `village` : A factor with two levels (Diohine and Toucar).

Figure 3 show that the parasite load density can be higher in some specific age classes. Moreover, one can observe that the three considered estimators can give quite similar classes of age, the different between the estimators aren't significative.

6 Conclusion

In this paper we propose a recursive relative regression estimators given in (4). The proposed estimators asymptotically follows normal distribution. Moreover, our proposed estimators attained the asymptotic con-

Table 2: The empirical levels ($\# \{r(x) \in I_{i,n}\} / N$, with $N = 500$) of three models using three estimators; `non-recursive` correspond to the estimator (5) using the proposed plug-in bandwidth selection (17), `Recursive 1` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = (n^{-1})$, `Recursive 2` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = \left(h_n / \left[\sum_{k=1}^{n} h_k\right]\right)$.

Model 1 (d = 1), x = 0									
	n = 100			n = 200			n = 350		
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	96.26%	95.83%	95.67%	96.34%	95.96%	95.86%	96.14%	95.88%	95.84%
Recursive 1	97.56%	97.24%	97.12%	97.44%	97.12%	97.04%	97.26%	97.14%	97.08%
Recursive 2	**98.14%**	**97.86%**	**97.82%**	**97.95%**	**97.86%**	**97.57%**	**97.82%**	**97.66%**	**97.36%**
Model 1 (d = 2), x = (0, 0)									
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	95.82%	95.53%	95.43%	95.84%	95.62%	95.46%	95.74%	95.68%	95.44%
Recursive 1	97.16%	97.04%	96.82%	97.04%	96.62%	96.44%	96.86%	96.54%	96.38%
Recursive 2	**97.64%**	**97.46%**	**97.32%**	**97.55%**	**97.46%**	**97.27%**	**97.42%**	**97.26%**	**97.16%**
Model 2 (d = 1), x = 0									
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	95.87%	95.51%	95.32%	95.84%	95.56%	95.24%	95.37%	95.18%	95.02%
Recursive 1	96.23%	96.13%	96.42%	96.14%	96.52%	96.44%	96.76%	96.54%	96.48%
Recursive 2	**97.77%**	**97.63%**	**97.59%**	**97.47%**	**97.38%**	**97.33%**	**97.22%**	**97.18%**	**97.06%**
Model 2 (d = 2), x = (0, 0)									
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	95.13%	95.07%	94.97%	95.04%	94.89%	94.83%	94.94%	94.88%	94.84%
Recursive 1	96.96%	96.82%	96.79%	96.84%	96.82%	96.64%	96.46%	96.34%	96.29%
Recursive 2	**97.54%**	**97.39%**	**97.33%**	**97.35%**	**97.28%**	**97.19%**	**97.22%**	**97.17%**	**97.15%**
Model 3 (d = 1), x = 0									
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	96.12%	95.98%	95.77%	96.13%	95.29%	95.13%	95.84%	95.18%	95.14%
Recursive 1	97.15%	97.17%	97.22%	97.33%	97.21%	97.12%	97.06%	96.82%	96.68%
Recursive 2	**97.88%**	**97.62%**	**97.53%**	**97.43%**	**97.32%**	**97.23%**	**97.11%**	**97.03%**	**96.96%**
Model 3 (d = 2), x = (0, 0)									
	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 0.2$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
non-recursive	95.42%	95.32%	95.27%	95.14%	95.12%	95.02%	94.82%	94.78%	94.54%
Recursive 1	96.82%	96.54%	96.42%	96.24%	96.12%	96.04%	96.56%	96.34%	96.24%
Recursive 2	**97.28%**	**97.15%**	**97.11%**	**97.23%**	**97.16%**	**97.05%**	**97.24%**	**97.19%**	**97.13%**

vergence rate $O\left(n^{-4/(d+4)}\right)$. The recursive estimators using the plug-in bandwidth selection developed in the subsection 3.1 (see, (14)) are then compared with the non-recursive one proposed by [15] using the plug-in bandwidth selection developed in the subsection 3.2 (see, (17)). We showed that, using some particularly choice, the proposed estimators can give in some situation a better results compared to the non-recursive approach in terms of estimation error. The simulation study confirms the nice feature of our proposed recursive estimators.

In conclusion, the proposed estimators allowed us to obtain a good results. A future research direction would be to extend our findings to the α-mixing framework see [17]. Another direction is to investigate the relative regression estimation based on the transformation of the data, in the case when the response variable are not positive see [44].

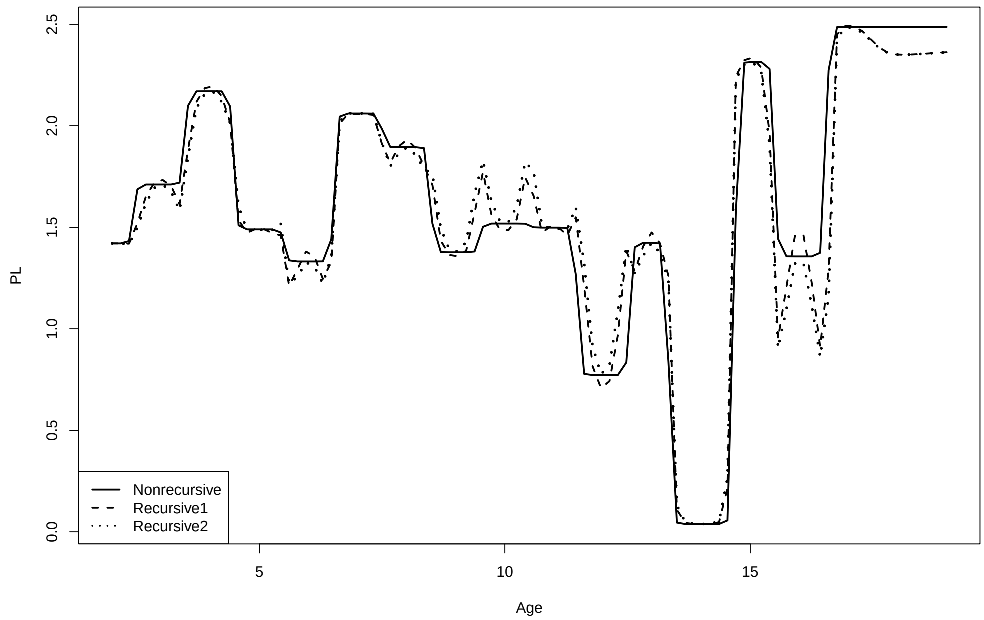

Figure 3: The `Plasmodium falciparum` Parasite Load density with automatically bandwidth selection using the `Non-recursive` correspond to the estimator (5) using the proposed plug-in bandwidth selection (17), `Recursive 1` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = (n^{-1})$, `Recursive 2` correspond to the estimator (4) using the proposed plug-in bandwidth selection (14) and the stepsize $(\gamma_n) = \left(h_n/\left[\sum_{k=1}^n h_k\right]\right)$.

7 Proofs

Throughout this section we use the following notation:

$$\Pi_n = \prod_{j=1}^n \left(1-\gamma_j\right), \quad W_n(x) = h_n^{-d} Y_n^{-1} K\left(\frac{x-X_n}{h_n}\right), \quad Z_n(x) = h_n^{-d} Y_n^{-2} K\left(\frac{x-X_n}{h_n}\right).$$

Let us first state the following technical lemma.

Lemma 2. Let $(v_n) \in \mathcal{GS}\left(v^*\right)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, and $m > 0$ such that $m - v^*\xi > 0$ where ξ is defined in (7). We have

$$\lim_{n\to+\infty} v_n \Pi_n^m \sum_{k=1}^n \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\xi}.$$

Moreover, for all positive sequence (α_n) such that $\lim_{n\to+\infty} \alpha_n = 0$, and all $\delta \in \mathbb{R}$,

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^n \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + \delta\right] = 0.$$

Lemma 2 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption $(A2)(iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity.

7.1 Proof of Theorem 1

Let us first note that, for x such that $\psi_n(x) \neq 0$, we have

$$r_n(x) - r(x) = B_n(x)\frac{\psi(x)}{\psi_n(x)}, \tag{18}$$

with

$$B_n(x) = \frac{1}{\psi(x)}(\phi_n(x) - \phi(x)) - \frac{r(x)}{\psi(x)}(\psi_n(x) - \psi(x)). \tag{19}$$

It follows from (18), that the asymptotic behavior of $r_n(x) - r(x)$ can be deduced from the one of $B_n(x)$. Moreover, the following Lemma follows from the Proposition 1 of [25].

Lemma 3. (Bias and variance of ϕ_n and ψ_n) Under Assumptions $(A1)$-$(A3)$, and assume that, for all $i, j \in \{1, \ldots, d\}$ $\phi_{ij}^{(2)}$ and $\psi_{ij}^{(2)}$ are continuous at x.
If $a \in [0, \alpha/(d+4)]$, then

$$\mathbb{E}[\phi_n(x)] - \phi(x) = \frac{1}{2(\alpha - 2a\xi)}\sum_{j=1}^{d}\left(\mu_j^2\phi_{j,j}^{(2)}(x)\right)h_n^2 + o\left(h_n^2\right). \tag{20}$$

$$\mathbb{E}[\psi_n(x)] - \psi(x) = \frac{1}{2(\alpha - 2a\xi)}\sum_{j=1}^{d}\left(\mu_j^2\psi_{j,j}^{(2)}(x)\right)h_n^2 + o\left(h_n^2\right). \tag{21}$$

If $a \in (\alpha/(d+4), 1)$, then

$$\mathbb{E}[\phi_n(x)] - \phi(x) = o\left(\sqrt{\gamma_n h_n^{-d}}\right), \quad \mathbb{E}[\psi_n(x)] - \psi(x) = o\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{22}$$

If $a \in [\alpha/(d+4), 1)$, then

$$Var[\phi_n(x)] = \frac{\mathbb{E}\left[Y^{-2}|X=x\right]f(x)}{(2-(\alpha - ad)\xi)}\frac{\gamma_n}{h_n^d}R(K) + o\left(\frac{\gamma_n}{h_n^d}\right). \tag{23}$$

$$Var[\psi_n(x)] = \frac{\mathbb{E}\left[Y^{-4}|X=x\right]f(x)}{(2-(\alpha - ad)\xi)}\frac{\gamma_n}{h_n^d}R(K) + o(\frac{\gamma_n}{h_n^d}). \tag{24}$$

If $a \in [0, \alpha/(d+4))$, then

$$Var[\phi_n(x)] = o(h_n^4), \quad Var[\psi_n(x)] = o(h_n^4). \tag{25}$$

Then, (8) follows from (20), (21) and (18) and (9) follows from (22) and (18).
Now, it follows from (19) that

$$Var[B_n(x)] = \frac{1}{\psi^2(x)}\left\{Var[\phi_n(x)] + r^2(x)\,Var[\psi_n(x)] - 2r(x)\,Cov(\phi_n(x), \psi_n(x))\right\}. \tag{26}$$

Using Lemma 2, and classical computations, we obtain that

$$Cov(\phi_n(x), \psi_n(x)) = \frac{\mathbb{E}\left[Y^{-3}|X=x\right]f(x)}{(2-(\alpha - ad)\xi)}\frac{\gamma_n}{h_n^d}R(K) + o\left(\frac{\gamma_n}{h_n^d}\right). \tag{27}$$

Then, the combination of (18), (26), (23), (24) and (27), gives (10), and the combination of (18), (26), (25) and (27), gives (11).

7.2 Proof of Theorem 2

Let us at first assume that, if $a \geq \alpha/(d+4)$, then

$$\sqrt{\gamma_n^{-1} h_n^d}\,(r_n(x) - \mathbb{E}[r_n(x)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{V}(x) R(K)). \tag{28}$$

In the case when $a > \alpha/(d+4)$, Part 1 of Theorem 2 follows from the combination of (9) and (28). In the case when $a = \alpha/(d+4)$, Parts 1 and 2 of Theorem 2 follow from the combination of (8) and (28). In the case $a < \alpha/(d+4)$, (11) implies that

$$h_n^{-2}(r_n(x) - \mathbb{E}(r_n(x))) \xrightarrow{\mathbb{P}} 0,$$

and the application of (8) gives Part 2 of Theorem 2.

We now prove (28). In view of (19), we have

$$B_n(x) - \mathbb{E}[B_n(x)] = \frac{1}{\psi(x)} \Pi_n \sum_{k=1}^{n} (T_k(x) - \mathbb{E}[T_k(x)]), \text{ with } T_k(x) = \Pi_k^{-1}\gamma_k (W_k(x) - r(x) Z_k(x)).$$

Now, we let $Y_k(x) = T_k(x) - \mathbb{E}(T_k(x))$.
Moreover, we have

$$\begin{aligned} v_n^2 &= \sum_{k=1}^{n} Var(Y_k(x)) \\ &= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left\{ Var(W_k(x)) + r^2(x) Var(Z_k(x)) - 2r(x) Cov(W_k(x), Z_k(x)) \right\}. \end{aligned}$$

Moreover, in view of $(A3)$, classical computations give

$$\begin{aligned} Var(W_k(x)) &= \frac{1}{h_k^d}\left[\mathbb{E}\left[Y^{-2}|X=x\right] f(x) R(K) + o(1)\right], \\ Var(W_k(x)) &= \frac{1}{h_k^d}\left[\mathbb{E}\left[Y^{-4}|X=x\right] f(x) R(K) + o(1)\right], \\ Cov(Z_k(x), W_k(x)) &= \frac{1}{h_k^d}\left[\mathbb{E}\left[Y^{-3}|X=x\right] f(x) R(K) + o(1)\right]. \end{aligned}$$

The application of Lemma 2 ensures that

$$v_n^2 = \psi^2(x) \sum_{k=1}^{n} \frac{\Pi_k^{-2}\gamma_k^2}{h_k^d} [\mathcal{V}(x) R(K) + o(1)] = \frac{\psi^2(x)}{\Pi_n^2} \frac{\gamma_n}{h_n^d} [\mathcal{V}(x) R(K) + o(1)].$$

On the other hand, we have, for all $p > 0$, $\mathbb{E}\left[|T_k(x)|^{2+p}\right] = O\left(\frac{1}{h_k^{1+p}}\right)$,

and, since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - ad)/2$, there exists $p > 0$ such that $\lim_{n\to\infty}(n\gamma_n) > \frac{1+p}{2+p}(\alpha - ad)$. Applying Lemma 2, we get

$$\sum_{k=1}^{n} \mathbb{E}\left[|Y_k(x)|^{2+p}\right] = O\left(\sum_{k=1}^{n} \Pi_k^{-2-p}\gamma_k^{2+p} \mathbb{E}\left[|T_k(x)|^{2+p}\right]\right) = O\left(\sum_{k=1}^{n} \frac{\Pi_k^{-2-p}\gamma_k^{2+p}}{h_k^{1+p}}\right) = O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p} h_n^{1+p}}\right),$$

and we thus obtain

$$\frac{1}{v_n^{2+p}} \sum_{k=1}^{n} \mathbb{E}\left[|Y_k(x)|^{2+p}\right] = O\left(\left[\gamma_n h_n^{-d}\right]^{p/2}\right) = o(1).$$

The convergence in (28) then follows from the application of Lyapounov's Theorem.

7.3 Proof of Proposition 1

Following similar steps as the proof of the Proposition 2 of [25], we proof Propostion 1.

Acknowledgments: The authors would like to thank the Editor of Dependence Modeling and the two reviewers for their helpful comments, which helped us to focus on improving the original version of the paper.

Funding: This work benefited from the financial support of the GDR 3477 GeoSto.

References

[1] Altendji, B., J. Demongeot, A. Laksaci, and M. Rachdi (2018). Functional data analysis: estimation of the relative error in functional regression under random left-truncation model. *J. Nonparametr. Stat. 30*(2), 472–490.
[2] Attouch, M., A. Laksaci, and N. Messabihi (2017). Nonparametric relative error regression for spatial random variables. *Statist. Papers 58*(4), 987–1008
[3] Bojanic, R. and E. Seneta (1973). A unified theory of regularly varying sequences. *Math. Z. 134*, 91–106.
[4] Chatfield, C (2007). The joys of consulting. *Significance 4*(1), 33–36.
[5] Chen, K., S. Guo, Y. Lin, and Z. Ying (2010). Least absolute relative error estimation. *J. Amer. Stat. Assoc. 105*(491), 1104–1112.
[6] Cleveland, W. S. and S. J. Devlin (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc. 83*(403), 596–610.
[7] Demongeot, J., A. Hamie, A. Laksaci, and M. Rachdi. (2016), Relative-error prediction in nonparametric functional statistics: theory and practice. *J. Multivariate Anal. 146*, 261–268.
[8] Derrar, S., A. Laksaci, and E. Ould Saïd (2020). M-estimation of the regression function under random left truncation and functional time series model. *Statist. Papers 61*(3), 1181–1202.
[9] Duflo, M. (1997) *Random Iterative Models*. Springer, Berlin.
[10] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.
[11] Farnum, N. R. (1990). Improving the relative error of estimation. *Amer. Stat.*, *44*(4), 288–289.
[12] Green, P. J. and B. W. Silverman (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, London.
[13] Galambos, J., and E. Seneta (1973). Regularly varying sequences. *Proc. Amer. Math. Soc. 41*(1), 110–116.
[14] Herrmann, E., M. Wand, J. Engel and T. Gasser (1995). A bandwidth selector for bivariate kernel regression. *J. R. Stat. Soc. Ser. B Stat. Methodol. 57*(1), 171–180.
[15] Jones, M. C., H. Park, K. I. Shin, S. K. Vines and S. H. Jeong (2008). Relative error prediction via kernel regression smoothers. *J. Statist. Plann. Inference 138*(10), 2887–2898.
[16] Khardani, S. and Y. Slaoui (2019a). Nonparametric relative regression under random censorship model. *Statist. Probab. Lett. 151*, 116–122.
[17] Khardani, S. and Y. Slaoui (2019b). Recursive kernel density estimation and optimal bandwidth selection under α: Mixing Data. *J. Stat. Theory Pract. 13*(2), 1–21
[18] Khoshgoftaar, T. M., B. B. Bhattacharyya, G. D. Richardson (1992). Predicting software errors, during development, using nonlinear regression models: a comparative study. *IEEE Trans. Reliab. 41*(3), 390–395.
[19] Kushner, H. J. (1977). General convergence results for stochastic approximations via week convergence theory. *J. Math. Anal. Appl. 61*(2), 490–503.
[20] Kushner, H. J. and G. G. Yin (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Second edition. Springer, New York.
[21] Ljung, L. (1978). Strong convergence of a stochastic approximation algorithm. *Ann. Statist. 6*(3), 680–696.
[22] Mechab, W. and A. Laksaci (2016). Nonparametric relative regression for associated random variables. *Metron 74*(1), 75–97.
[23] Milet, J., G. Nuel, L. Watier, D. Courtin, Y. Slaoui, P. Senghor, F. Migot-Nabias, O. Gaye and A. Garcia (2010). Genome wide linkage study, using a 250K SNP map, of Plasmodium falciparum infection and mild malaria attack in a Senegalese population. *PLoS ONE 5*(7), Article ID e11616, 11 pages.
[24] Mokkadem, A., and M. Pelletier (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist. 35*(4), 1749–1772.
[25] Mokkadem, A., M. Pelletier, and Y. Slaoui (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference 139*(7), 2459–2478.
[26] Mokkadem, A., M. Pelletier, and Y. Slaoui (2009b). Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA. Lat. Am. J. Probab. Math. Stat. 6*, 63–114.
[27] Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl. 9*(1), 141–142.

[28] Narula, S.C. and J.F. Wellington (1977). Prediction, linear regression and the minimum sum of relative errors. *Technometrics 19*(2), 185–190.

[29] Park, H and L. Stefanski (1998). Relative-error prediction. *Statist. Probab. Lett. 40*(3), 227–236.

[30] Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hungar. 8*, 391–398.

[31] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist. Ser. Statist. 8*(1), 119–126.

[32] Robbins, H. and S. Monro (1951). A Stochastic Approximation Method. *Ann. Math. Statist. 22*(3), 400–407.

[33] Ruppert, D. (1982). Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. Probab. 10*(1), 178–187.

[34] Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *Ann. Statist. 22*(3), 1346–1370.

[35] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

[36] Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *J. Probab. Stat. 2014*, Article ID 739640, 11 pages.

[37] Slaoui, Y. (2014b). The stochastic approximation method for estimation of a distribution function. *Math. Methods Statist. 23*(4), 306–325.

[38] Slaoui, Y. and G. Nuel (2014c). Parameter estimation in a hierarchical random intercept model with censored response: an approach using a SEM algorithm and Gibbs sampling. *Sankhya B 76*(2), 210–233.

[39] Slaoui, Y. (2015). Plug-in Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl. 69*(4), 483–509.

[40] Slaoui, Y. (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface. 9*(3), 375–388.

[41] Slaoui, Y. (2017). Recursive kernel density estimators under missing data. *Comm. Statist. Theory Methods 46*(18), 9101–9125.

[42] Slaoui, Y. (2018). Bias reduction in kernel density estimation. *J. Nonparametr. Stat. 30*(2), 505–522.

[43] Slaoui, Y. (2019). Wild Bootstrap Bandwidth Selection of Recursive Nonparametric Relative Regression for Independent Functional Data. *J. Multivariate Anal. 173*, 494–511.

[44] Slaoui, Y. (2020). Two new nonparametric kernel distribution estimators based on a transformation of the data. *J. Appl. Stat.*, to appear. Available at https://doi.org/10.1080/02664763.2020.1786675.

[45] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist. 8*(6), 1348–1360.

[46] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist. 10*(4), 1040–1053.

[47] Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Probl. Inf. Transm. 26*(1), 31–37.

[48] Wahba, G. (1990). *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics, Philadelphia.

[49] Wand, M. P. and M. C. Jones. (1994). Multivariate plug-in bandwidth selection. *Comput. Stat. 9*(2), 97–116.

[50] Watson, G. S. (1964). Smooth regression analysis. *Sankhya A 26*(4), 359–372.

[51] Yang, Y. and F. Ye. (2013). General relative error criterion and M-estimation. *Front. Math. China 8*(3), 695–715.