Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol I
IMECS 2008, 19-21 March, 2008, Hong Kong

Sequential Fixed-width Confidence Bands for Kernel Regression Estimation

L. S. Dharmasena * B. M. de Silva† P. Zeephongsekul‡

Abstract—**We consider a random design model based on independent and identically distributed (iid) pairs of observations (X_i, Y_i), where the regression function $m(x)$ is given by $m(x) = E(Y_i|X_i = x)$ with one independent variable. In a nonparametric setting the aim is to produce a reasonable approximation to the unknown function $m(x)$ when we have no precise information about the form of the true density, $f(x)$ of X. We describe an estimation procedure of non-parametric regression model at a given point by some appropriately constructed fixed-width $(2d)$ confidence interval with the confidence coefficient of at least $1-\alpha$. Here, $d(>0)$ and $\alpha \in (0,1)$ are two preassigned values. Fixed-width confidence intervals are developed using both Nadaraya-Watson and local linear kernel estimators of nonparametric regression with data-driven bandwidths. The sample size was optimized using the purely and two-stage sequential procedure together with asymptotic properties of the Nadaraya-Watson and local linear estimators. A large scale simulation study was performed to compare their coverage accuracy. The numerical results indicate that the confidence bands based on the local linear estimator have the best performance than those constructed by using Nadaraya-Watson estimator. However both estimators are shown to have asymptotically correct coverage properties.**

Keywords: ***Nonparametric regression, Nadaraya-Watson estimator, Local linear estimator, fixed-width confidence interval, random design, purely sequential procedure, two-stage sequential procedure***

1 Introduction

Suppose that $(X_1, Y_1), ..., (X_n, Y_n)$ is a sequence of iid distributed (i.i.d.) bivariate random variables having an unknown continuous pdf $f_{XY}(x, y)$ and for simplicity we assume that $X_i \in (0, 1)$ with an unknown pdf $f_X(x)$. Consider the nonparametric regression model

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, ..., n \tag{1}$$

where ε_i is a sequence of iid random variables with $\mathbf{E}[\varepsilon_i] = 0$, $\mathbf{E}[\varepsilon_i^2] = \sigma^2$ and $m(\cdot)$ is an unknown function.

The present article attempts to estimate fixed width confidence bands for the unknown function $m(x)$ at a given point $x = x_0$. Estimation is based on kernel type estimators and consider two most popular kernel estimators namely, Nadaraya-Watson estimator $\hat{m}_{h_n,NW}(x_0)$ and local linear estimator $\hat{m}_{h_n,LL}(x_0)$ (Wand and Jones (1995)) which are defined respectively by

$$\hat{m}_{h_n,NW}(x_0) = \frac{\sum_{i=1}^n y_i K(\frac{x_0-x_i}{h_n})}{\sum_{j=1}^n K(\frac{x_0-x_j}{h_n})} \tag{2}$$

and

$$\hat{m}_{h_n,LL}(x_0) = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i} \tag{3}$$

where

$$w_i = K\left(\frac{x_0 - x_i}{h_n}\right)(s_{n,2} - (x_0 - x_i)s_{n,1}) \tag{4}$$

with

$$s_{n,l} = \sum_{i=1}^n K\left(\frac{x_0 - x_i}{h_n}\right)(x_0 - x_i)^l, \quad l = 1, 2 \tag{5}$$

here $K(\cdot)$ is the kernel function and h_n is the bandwidth. In this paper, as in Isogai (1987), we take $h_n = n^{-r}$ for $0.2 < r < 1$. Let $K(\cdot)$ satisfy $\int uK(u)du = 0$, $\int u^2K(u)du \leq \infty$, $K(u)$ and $|uK(u)|$ are bounded, $\lim nh_n^3 = \infty$ and $\lim nh_n^5 = 0$.

In general, local polynomial estimator (Fan and Gijbels, 1996) are superior to Nadaraya-Watson estimator in some respects (Fan, 1993), but recent contributions by Boularan et al. (1995), Einmahl and Mason (2000) as well as Quian and Mammitzsch (2000), among others, have given evidence of continuing interest in the Nadaraya-Watson estimator. One of the strengths of this estimator certainly consists in its automatic adaptation to designs where the local polynomial estimator may not be

*Date:16/11/07. Dept. Maths & Statistics, RMIT University, Melbourne, Australia Email: sandamali.dharmasena@rmit.edu.au
†Email: desilva@rmit.edu.au
‡Email: desilva@rmit.edu.au

performing reliably over all since its variance can fail to exist in random designs. Also, the Nadaraya-Watson estimator retains some optimality properties as exposed in Hardle and Marron (1985). Methods for obtaining confidence bands for $m(x)$ can be found in Hall and Titterington (1988), Eubank and Speckman (1993) and Diebolt (1995). The most widely used confidence band for $m(x)$ is based on the theorem of Bickel and Rosenblatt (1973) for kernel estimation of a density function. Bias-corrected confidence bands for general nonparametric regression models are considered by Xia (1998). In principle, confidence intervals can be obtained from asymptotic normality results for $\hat{m}(x)$. However, the limiting bias and variance depend on unknown quantities which have to be estimated consistently in order to construct asymptotic confidence intervals.

Sequential analysis, in general, comes in handy when the experimenter's objective is to control the error of estimation at some preassigned level. Whether one wants to estimate $m(x)$ at one single point x_0 or for all $x \in \mathbb{R}$, depending on the specific goal and error criterion, one would like to determine the sample size n in an optimal fashion. That is in order to have the error controlled at a preassigned level, sample size has to be adaptively estimated in the process by a positive integer valued random variable N where the event $[N = n]$ will depend only on $(X_1, Y_1), ..., (X_n, Y_n)$ for all $n \geq 1$. Finally $m(x)$ is estimated by $\hat{m_{h_N}}(x)$ constructed from $(X_1, Y_1), ..., (X_N, Y_N)$.

2 Nonparametric Kernel Regression

Throughout the present work, we will consider the following regression model with a random design. Let

$$m(x) = \mathbf{E}[Y|X = x] \qquad (6)$$

be the unknown regression function which describes the dependance of the so-called response variable Y on the value of X. The following assumptions are used in this study (Wand and Jones (1995)):

(i) $m''(x)$ is continuous for all $x \in [0, 1]$.
(ii) $K(x)$ is symmetric about $x = 0$ and supported on $[-1, 1]$.
(iii) $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.
(iv) The given point $x = x_0$ must satisfy $h_n < x_0 < 1-h_n$ for all $n \geq n_0$ where n_0 is a fixed number.

The obvious problem that occurs when using (2) and (3) is the choice of bandwidth, h_n. Since $h_n = n^{-r}$ for $0.2 < r < 1$ using the property $h_n < x_0 < 1 - h_n$ one can prove that $r \in (r_{\min}, 1)$ where $r_{\min} = \max(0.20, r_0)$ and

$$r_0 = \left\{ \frac{-\ln [\min(x_0, 1 - x_0)]}{\ln(n)} \right\} \qquad (7)$$

A natural way of constructing a confidence band for $m(x)$ is follows. Suppose that $\hat{m_{h_n}}(x)$ is an estimator of $m(x)$ then $100(1 - \alpha)\%$ confidence band is of the form

$$Pr\left\{|\hat{m}_{h_n}(x) - m(x)| \leq d\right\} \geq 1 - \alpha \quad \forall x \in [0, 1] \qquad (8)$$

There are many difficulties with finding a good solution to (8). Firstly, we must derive the asymptotic distribution of $\hat{m}(x)-m(x)$; secondly the estimation of residual variance and distribution function of X. Consequently, a good estimator of bandwidth h_n is needed.

The kernel estimators are asymptotically normal, as was first shown in Schuster (1972).

Theorem 1. *Let $K(\cdot)$ satisfy $\int uK(u)du = 0$, $\int u^2K(u)du \leq \infty$, $K(u)$ and $|uK(u)|$ are bounded, h_n is such that $\lim nh_n^3 = \infty$ and $\lim nh_n^5 = 0$. Suppose $x_1, ..., x_k$ are distinct points and $g(x_i) > 0$ for $i = 1, 2, ..., k$. If $E[Y^3]$ is finite and if g', w', v', g'' and w'' exist and bounded where $g(x) = \int f(x, y)dy$, $w(x) = \int yf(x, y)dy$ and $v(x) = \int y^2f(x, y)dy$ respectively, then*

$$\sqrt{nh_n}\left(m_{h_n}(x_1) - m(x_1), ..., m_{h_n}(x_k) - m(x_k)\right) \xrightarrow{d} Z^* \qquad (9)$$

where Z^* is multivariate normal with mean vector $\mathbf{0}$ and diagonal covariance matrix $\mathbf{C} = [\mathbf{C}_{ii}]$ where $C_{ii} = V[Y|X = x_i] \int K^2(u)du/g(x_i)$ $(i = 1, ..., k)$.

In general the bias of the $\hat{m}_{q,ll}(x)$ estimator is smaller than $\hat{m}_{q,nw}(x)$ estimator (11).

$$\mathbf{E}[\hat{m}_{q,h_n}(x_0)] = m(x_0) + Bias_q \qquad (10)$$

and

$$\mathbf{Var}[\hat{m}_{q,h_n}(x_0)] = \frac{B\sigma^2}{nh_nf(x)} + o\left\{(nh_n)^{-1}\right\}$$

where $q = nw$ for (2) estimator, $q = ll$ for (3) estimator,

$$Bias_q = \begin{cases} A + \frac{h_n^2\mu_2(K)m'(x)f'(x)}{f(x)} + o(h_n^2) & \text{if} \quad \text{q=nw} \\ A + o(h_n^2) & \text{if} \quad \text{q=ll} \end{cases} \qquad (11)$$

$A = \frac{h_n^2}{2}m''(x)\mu_2(K)$, $\mu_2(K) = \int_\infty^\infty u^2K(u)du$ and $B = \int_\infty^\infty K^2(u)du$.

3 Sequential Fixed-Width Confidence Interval

Given $d(> 0)$ and $\alpha \in (0, 1)$ with $h_n = n^{-r}$ for $r \in (r_{\min}, 1)$, suppose we wish to cliam

$$Pr\left\{m(x) \in I_n = [\hat{m}_{q,h_n}(x) \pm d]\right\} \approx 1 - \alpha \qquad (12)$$

for large n where x is fixed. Using Theorem 1 one can see that the probability requirement (12) leads to the implicit solution-equation

$$n \geq n_{opt} = \left\{ \frac{B Z^2_{\alpha/2} \sigma^2}{d^2 f(x)} \right\}^{\frac{1}{1-r}} \qquad (13)$$

where $Z_{\alpha/2}$ is the 50%α upper percentile of the standard normal distribution.

3.1 Purely Sequential Procedure

In general σ^2 in (13) is unknown and purely sequential procedure suggest to substitute the variance parameter by a estimator $\hat{\sigma}^2_{n_0}$ based on a small sample $(X_1, Y_1), ..., (X_{n_0}, Y_{n_0})$ of size $n_0 < n_{opt}$. Here we use the estimate of σ^2 proposed by Ursula etal. (2003) based on covariate matched U-statistics, that is

$$\hat{\sigma}^2 = \frac{\sum\sum_{i \neq j} \frac{1}{2}(Y_i - Y_j)^2 \frac{1}{2}\left(\frac{1}{\hat{g}_i - \hat{g}_j}\right) K\left(\frac{X_i - X_j}{h_n}\right)}{n(n-1)} \qquad (14)$$

where

$$\hat{g}_i = \frac{1}{n-1} \sum_{i \neq j} K\left(\frac{X_i - X_j}{h_n}\right) \qquad (15)$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 + o_p(n^{-\frac{1}{2}}) \qquad (16)$$

In purely sequential procedure we take one observation at time until the condition given in (17) is satisfied. Since we already have a sample of size n_0, the stopping rule for purely sequential procedure is given by

$$N = \max\left\{ n_0, \left\lfloor \left\{ \frac{Z^2 \alpha/2}{B} \hat{\sigma}^2_{n_0} d^2 f(x) \right\}^{\frac{1}{1-r_1}} \right\rfloor + 1 \right\} \qquad (17)$$

where $\lfloor n \rfloor$ refers to the floor function, $r_1 \in (r_{\min}, 1)$ and from (7) $r_{\min} = \max\left(0.20, -\ln\left[\min(x_0, 1 - x_0)\right] / \ln n_0\right)$.

3.2 Two-stage Sequential Procedure

The above purely sequential procedure involves a lot of computational effort. Stein (1945) introduced a sampling procedure which requires only two sampling operations. However, it turned out that this two-stage procedure is less efficient than the purely sequential procedure. Using the asymptotic normality results in the Theorem 1 for univariate random design case we can write

$$\frac{\sqrt{nh_n}\left\{\hat{m}_{q,h_n}(x) - m(x)\right\}}{\sigma\sqrt{B(f(x))^{-1}}} \to N(0,1) \qquad (18)$$

From (24) for a random sample of normally distributed residuals $\{\varepsilon_i\}_{i=1}^n$ with mean 0 and variance σ^2

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n)} \qquad (19)$$

where $\chi^2_{(n)}$ is the chi-squared distribution with n degrees of freedom. Hence we combine (18) and (19) to claim that

$$\frac{\frac{\sqrt{nh_n}\{\hat{m}_{q,h_n}(x) - m(x)\}}{\hat{\sigma}\sqrt{B(f(x))^{-1}}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} \sim t_n \qquad (20)$$

The following statement (21) is obviously equivalent to (12)

$$\begin{aligned} Pr\left\{m(x) \in I_n\right\} &\approx t\left(\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))^{-1}}\hat{\sigma}}\right) - t\left(-\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))^{-1}}\hat{\sigma}}\right) \\ &= 2t\left(\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))}}\right) - 1 \qquad (21) \end{aligned}$$

where $t()$ is the cumulative student-t distribution and an approximate solution to the problem is provided by taking the smallest integer $n \geq 1$ such that

$$2t\left(\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))^{-1}}\hat{\sigma}}\right) - 1 \geq 1 - \alpha \qquad (22)$$

and since $h_n = n^{-r}$

$$n \geq \left(\frac{t^2_{\alpha/2,n} B \hat{\sigma}^2}{d^2 f(x)}\right)^{\frac{1}{1-r}} \qquad (23)$$

where $t_{\alpha/2,n} = t^{-1}(1 - \alpha/2)$ the $(1 - \alpha/2)^{th}$ quantile of the student-t distribution function $t(\cdot)$.

Two-stage sampling procedure is started by taking a pilot bi-variate sample $\{X_i, Y_i\}_{i=1}^{n_0}$ and then estimate the required final sample size by N. Now using (23) we propose the following stopping rule for a two-stage procedure

$$N \equiv N(d) = \max\left\{ n_0, \left\lfloor \left(\frac{t^2_{n_0,\alpha/2} B \hat{\sigma}^2_{n_0}}{d^2 f(x)}\right)^{\frac{1}{1-r_1}} \right\rfloor + 1 \right\} \qquad (24)$$

If $N = n_0$ then we need no more observations in the second stage. But if $N > n_0$ then we take additional bivariate sample $\{X_i, Y_i\}_{i=n0+1}^{N}$ of size $N - n_0$ in the second stage. Finally we use the sample $\{(X_1, Y_1), ..., (X_N, Y_N)\}$ to compute Nadaraya-Watson (2) and local linear (3) estimates for $m(x_0)$ and construct the confidence band given in (12). In an application of the stopping rule (24), it is important to select the best available values for the design constants r and n_0 for fixed predesigned values of d and α.

4 Simulation Results

We use the following two models to assess the performance of the confidence bands developed in Section 3:
Model I : $Y = \sqrt{4x+3} + \epsilon$
Model II: $Y = 2\exp\{\frac{-x^2}{0.18}\} + 3\exp\{-\frac{(x-1)^2}{0.98}\} + \epsilon$
where $\epsilon \sim N(0, \sigma^2)$.
Widths of the interval $d = 0.05, 0.07, 0.09, 0.11, 0.13$ were used. The initial sample size n_0 and σ were chosen to be 25 and 0.5 respectively. The confidence bands were investigated for $\alpha = 0.05$. For all the data analysed, we used standard normal kernel $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and hence $B = 2\sqrt{\pi}$. In both models 15000 replicate samples for each experimental setting were carried out to obtain the final sample sizes required to estimate $m(x)$ at $x_0 = 0.306$ given fixed-width, $2d$.

We obtained 15000 random samples of $\{X_i\}_{i=1}^{25}$ from uniform distribution and then calculate corresponding y_i for each stated relation (Models I and II). Random errors ϵ were generated from $N(0, .5^2)$ distribution and added to the above y_i to obtained Y_i. First we consider two-stage sequential procedure for $\alpha = 0.05$ and then purely sequential procedure. The average final sample size $\bar{n}$, average residual variance estimate $\overline{\hat{\sigma}^2}$, average local linear $\overline{\hat{m}_{LL}}$, average Nadaraya-Watson $\overline{\hat{m}_{NW}}$ estimates and coverage probability $\tilde{p}$ which is the proportion of the confidence intervals that contains the theoretical value, $m(x_0)$ estimated at the point $x_0 = 0.306$ are reported in Tables 1 and 2 for $\alpha = 0.05$. Here (.) in the tables gives the standard error of the estimated value.

Coverage probabilities of both Nadaraya-Watson ($\tilde{p}_{NW}$) and local linear estimators ($\tilde{p}_{LL}$) have achieved preset confidence coefficient 95% at $x_0 = .306$ in Model II except when $d = .13$. But the coverage probabilities for Model I shows a different picture as Nadaraya-watson estimator fails to achieve required coverage probabilities except when $d = .05$ where as local linear method does. This noticeable difference is mainly due to structural differences in the selected models. And also due to the bias terms which heavily depend on derivatives of the unknown function $m(\cdot)$ associated with each estimator. ($\tilde{p}_{NW}$) for Model I is increasing with decreasing d due to large sample sizes resulted in increase in sample sizes. This is consistent with both procedures

Table 1: Empirical coverage of LL and NW for Model I
$\alpha = .05;\ m(x_0) = 2.055$

d	n_{opt}	$\bar{n}$	$\tilde{p}_{LL}$	$\tilde{p}_{NW}$	$\overline{\hat{m}_{LL}}$	$\overline{\hat{m}_{NW}}$	$\overline{\hat{\sigma}^2}$
			Two – stage Procedure				
.13	81.8	109.3	.947	.902	2.046	2.108	.265
		(.40)	(.00)	(.00)	(.001)	(.001)	(.001)
.11	139.0	185.9	.965	.912	2.048	2.105	.262
		(.69)	(.00)	(.00)	(.000)	(.000)	(.001)
.09	262.8	340.0	.978	.921	2.048	2.099	.260
		(1.28)	(.00)	(.00)	(.000)	(.000)	(.000)
.07	583.6	776.7	.989	.932	2.047	2.091	.265
		(2.83)	(.00)	(.00)	(.000)	(.000)	(.000)
.05	1698.2	2259.7	.996	.958	2.048	2.076	.265
		(8.34)	(.00)	(.00)	(.000)	(.000)	(.000)
			Purely Sequential Procedure				
.13	81.8	80.1	.918	.869	2.046	2.219	.242
		(.00)	(.00)	(.00)	(.001)	(.001)	(.001)
.11	139.0	137.6	.954	.901	2.046	2.189	.246
		(.00)	(.00)	(.00)	(.001)	(.001)	(.001)
.09	262.8	261.1	.980	.914	2.047	2.109	.248
		(.00)	(.00)	(.00)	(.000)	(.000)	(.000)
.07	583.6	581.7	.991	.926	2.047	2.097	.249
		(.00)	(.00)	(.00)	(.000)	(.000)	(.000)
.05	1698.2	1695.6	.998	.947	2.051	2.081	.250
		(.00)	(.00)	(.00)	(.000)	(.000)	(.000)

i.e. two-stage and purely sequential. The performance of Nadraya-Watson estimator worsens as x increases as its bias highly depends on derivatives of $m(\cdot)$. For the interior point $x_0 = .306$, the Nadraya-Watson estimator assigns symmetric weights to both sides of $x_0 = .306$. For a random design this will overweigh the points on right hand side and hence create large bias. In other words Nadaraya-Watson estimator is not design-adaptive. However local linear method assigns asymmetrical weighting scheme while maintaining the same type of smooth weighting scheme as Nadaraya-Watson estimator. Hence local linear method adapts automatically to this random design.

This simulation analysis clearly shows that the average sample sizes in two-stage procedure is much larger than corresponding values in the purely sequential procedure for both models. This evidence clearly implies that the two-stage procedure is less efficient compared to purely sequential procedure but at the same time one should note that it is also associated with the highest coverage probability which exceeds the target confidence coefficient 95%. Further note that advantage of using a two-stage procedure is reflected in computational time. The purely sequential procedure needs substantially more computations and hence during simulations it needs significantly more computational times than the two-stage procedure, particularly for small d. However purely sequential procedure at times fall somewhat short of the optimal sample size. Hence the coverage probability falls short of the target especially when half width of the interval d becomes larger as it result in small sample sizes. But achieved target coverage probability for smaller d due to larger sample sizes.

ISBN: 978-988-98671-8-8

Table 2: Empirical coverage of LL and NW for Model II
$\alpha = .05;\ m(x_0) = 3.024$

d	n_{opt}	$\bar{n}$	$\tilde{p}_{LL}$	$\tilde{p}_{NW}$	$\widehat{\hat{m}_{LL}}$	$\widehat{\hat{m}_{NW}}$	$\widehat{\hat{\sigma}^2}$
		Two − stage Procedure					
.13	81.8	105.1	.946	.956	3.038	3.011	.258
		(.40)	(.00)	(.00)	(.001)	(.001)	(.001)
.11	139.0	180.5	.959	.967	3.037	3.004	.260
		(.68)	(.00)	(.00)	(.000)	(.000)	(.001)
.09	262.8	337.0	.973	.954	3.031	2.993	.258
		(1.27)	(.00)	(.00)	(.000)	(.000)	(.000)
.07	583.6	759.8	.989	.976	3.032	3.003	.261
		(2.91)	(.00)	(.00)	(.000)	(.000)	(.000)
.05	1698.2	2149.4	.994	.954	3.027	3.001	.256
		(8.25)	(.00)	(.00)	(.000)	(.000)	(.000)
		Purely Sequential Procedure					
.13	81.8	79.6	.916	.901	3.021	2.983	.241
		(.40)	(.00)	(.00)	(.001)	(.001)	(.000)
.11	139.0	137.9	.959	.946	3.031	2.995	.246
		(.68)	(.00)	(.00)	(.001)	(.001)	(.001)
.09	262.8	261.7	.977	.964	3.033	2.999	.248
		(1.27)	(.00)	(.00)	(.000)	(.001)	(.000)
.07	583.6	581.5	.992	.956	3.029	2.997	.249
		(2.91)	(.00)	(.00)	(.000)	(.000)	(.000)
.05	1698.2	1695.6	.998	.974	3.025	3.004	.250
		(8.25)	(.00)	(.00)	(.000)	(.000)	(.000)

5 Conclusions

In this paper we have studied data-driven fixed-width confidence bands for nonparametric regression curve estimation using local linear and Nadaraya-Watson estimators. Both procedures have been produced the correct asymptotic coverage probabilities. The coverage probability of Nadaraya-Watson method was found to be generally below the preset confidence coefficients. On the other hand local linear method had near-nominal coverage probabilities in most of the cases. The performance of the purely sequential procedure is better than that of the two-stage procedure. However operationally, two-stage procedure reduces computational costs associated with the corresponding purely sequential schemes by a substantial margin. The $\hat{\sigma}^2$ appeared to be very close to it's actual value even for small sample size cases.

References

[1] Boularan, J. and Ferré, L. and Vieu P., "Location of particular points in nonparametric regression analysis," *Australian Journal of Statistics*, V37, pp. 161-168, 1995.

[2] Bickel, P. L. and Rosenblatt, M., "On some global measures of the deviations of density function estimates.," *Annals of Statistics*, V1, pp. 1071-1095, 1973.

[3] Diebolt, J., "A nonparametric test for the regression function: asymptotic theory," *Journal of Statistical Planning Inference*, V44, pp. 1-17, 1995.

[4] Einmahl, U. and Mason, D. M., "An emperical process approach to the uniform consistency of kernel-type function estimators," *Journal of Theoretical Probability*, V13, pp. 1-37, 2000.

[5] Eubank, R. L. and Speckman, P. L., "Confidence bands in nonparametric regression," *Journal of the American Statistical Association*, V88, N424, pp. 1287-1301, 1993.

[6] Fan, J., "Local linear regression smoothers and their minimax efficiency," *Annals of Statistics*, V21, pp. 196-216, 1993.

[7] Fan, J. and Gijbels, I., *Local polynomial modelling and its applications*, Chapman and Hall, London, 1996.

[8] Ghosh, M., Mukhopadhyay, N. and Sen, P.K., *Sequential Estimation*, Wiley, New York, 1997.

[9] Hall P. and Marron, J. S., "On variance estimation in nonparametric regression," *Biometrika*, V77, N2, pp. 415-419, 1990.

[10] Hall P. and Titterington, D. M., "On confidence bands in nonparametric density estimation and regression," *Journal of Multivariate Analysis*, V27, pp. 228-254, 1988.

[11] Härdle W., and Marron, J. S., "Optimal bandwidth selection in nonparametric regression function estimation," *Annals of Statistics*, V13, pp. 1465-1481, 1985.

[12] Isogai, E., "The convergence rate of fixed-width sequential confidence intervals for a probability density function," *Sequential Analysis*, V6, pp. 55-69, 1987a.

[13] Martinsek, A. T., "Fixed width confidence bands for density functions," *Sequential Analysis*, V12, pp.169-177, 1993.

[14] Quian W. and Mammitzsch V., "Strong uniform convergence for the estimator of the regression function under φ mixing conditions", *Metrika*, V52, pp.45-61, 2000.

[15] Schuster, E. F., "Joint asymptotic distribution of the estimated regression function at a finite number of distinct points", *Annals of Mathematical Statistics*, V43, pp. 84-8, 1972.

[16] Ursula U. M., Schik, A. and Wefelmeyer W., "Estimating the error variance in nonparametric regression by a covariate-matched u-statistic," *Statistics*, V37, pp. 179-188, 2003.

[17] Wand, M. P. and Jones, M.C., *Kernel Smoothing*, Chapman and Hall, London, 1995.

[18] Xia, Y., "Bia-corrected confidence bands in nonparametric regression", *Journal of the Royal Statistical Society*, Series B, V4, pp. 797-811, 1998.