Sample Size Determination for Kernel Regression Estimation Using Sequential Fixed-Width Confidence Bands

L. Sandamali Dharmasena * P. Zeephongsekul† Basil M. de Silva‡

Abstract—**We consider a random design model based on independent and identically distributed 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 nonparametric 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 procedures 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 better performance than those constructed by using Nadaraya-Watson estimator. However both estimators are shown to have asymptotically correct coverage properties.**

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

1 Introduction

Suppose that $(X_1, Y_1), ..., (X_n, Y_n)$ is a sequence of independent and identically distributed (i.i.d.) bivariate random variables having an unknown continuous probability density function $f_{XY}(x, y)$ and without loss of generality we assume that $X_i \in (0, 1)$ with a probability distribution function $f_X(x)$. 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 $\hat{m}_{NW,h_n}(x_0)$ and local linear estimator $\hat{m}_{LL,h_n}(x_0)$ (Wand and Jones (1995)) which are defined respectively by

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

and

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

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 $a < r < b$; $a, b \in \mathbb{R}$. As shown above both estimators are weighted averages of the response variable Y. Let $K(\cdot)$ satisfy $\int uK(u)du = 0$, $\int u^2K(u)du \leq \infty$, $K(u)$ and $|uK(u)|$ are bounded. Commonly used kernel functions are listed in Table 1.

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

*Date:06/07/08. Dept. Maths & Statistics, RMIT University, Melbourne, Australia Email: sandamali.dharmasena@rmit.edu.au
†Email: panlopz@rmit.edu.au
‡Email: desilva@rmit.edu.au

Table 1: Selected Kernel Functions

Kernel	K(u)	
Epanechnikov	$\frac{3}{4}(1-\frac{1}{5}u^2)/\sqrt{5}$	for $\|u\|<\sqrt{5}$
	0	otherwise
Biweight	$\frac{15}{16}(1-u^2)^2$	for $\mid u \mid<1$
	0	otherwise
Double Exponential	$\frac{1}{2}\exp(-\|u\|)$	for $\mid u \mid<1$
	0	otherwise
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-(1/2)u^2)$	$\|u\|<\infty$
Uniform	$\frac{1}{2}$	for $\mid u \mid<1$
	0	otherwise

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 performing reliably over all. Also, the Nadaraya-Watson estimator retains some optimality properties as demonstrated 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 $\hat{m}_{LL,h_N}(x)$ and $\hat{m}_{NW,h_N}(x)$ 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

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

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. According to the assumption (iii) listed above, bandwidth h_n is a sequence satisfying $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Since we take $h_n = n^{-r}$; $a < r < b\, a, b \in \mathbb{R}$, $n^{-r} \to 0$ and $n^{1-r} \to \infty$ as $n \to \infty$ which result in $0 < r < 1$. And the assumption (iv) ensures the given point x_0 which the estimation is taking place is selected in such a way x_0 is more than a bandwidth h_n away from the boundary to avoid boundary effects that is $h_n < x_0$ and $x_0 < 1 - h_n$. Combining these assumptions a range of values which r takes will be decided as shown below:

$h_n < x_0$ and $h_n < 1 - x_0 \Leftrightarrow h_n < \min\{x_0, 1 - x_0\}$ which implies

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

Since $0 < r < 1$, the above must imply $r \in (\max\{0, r_0\}, 1) = (r_{\min}, 1)$ where we let $r_{min} = \max(0, r_0)$.

A natural way of constructing a confidence band for $m(x)$ is follows. Suppose that $\hat{m}_{q,h_n}(x)$ is an estimator of $m(x)$ where $q = NW$ for (2) estimator, $q = LL$ for (3) estimator then a $100(1-\alpha)\%$ confidence band is of the form

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

There are many difficulties with finding a good solution to (8). Firstly, we must derive the asymptotic distribution of $\hat{m}_{q,h_n}(x) - m(x)$; secondly we must estimate the residual variance and distribution function of X. Also, 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 $v(x) = \int y^2 f(x, y)dy$ respec-

tively, then

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

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

In general the asymptotic bias of the $\hat{m}_{LL,h_n}(x)$ estimator is smaller than $\hat{m}_{NW,h_n}(x)$ estimator (11):

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

where

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

as $n\to\infty$ where $A=\frac{h_n^2}{2}m''(x)\mu_2(K)$ and $\mu_2(K)=\int_{\infty}^{-\infty}u^2K(u)du$.

However, both estimators have the same asymptotic variance which is

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

as $n\to\infty$ where $B=\int_{\infty}^{-\infty}K^2(u)du$.

Hence in the univariate case of Theorem 1 we have

$$\sqrt{nh_n}\left[\hat{m}_{q,h_n}(x)-m(x)\right]\sim N\left(0,\frac{B\sigma^2}{f(x)}\right)$$

which leads to

$$\frac{\hat{m}_{q,h_n}(x)-m(x)}{\sigma\sqrt{\frac{B}{f(x)nh_n}}}\sim N(0,1)\qquad \text{as } n\to\infty. \tag{12}$$

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

$$\mathcal{P}\left\{m(x)\in I_n=[\hat{m}_{q,h_n}(x)\pm d]\right\}\approx 1-\alpha \tag{13}$$

which we can rewrite as

$$\mathcal{P}\left(\left|\frac{\hat{m}_{q,h_n}(x)-m(x)}{\sqrt{\mathbf{Var}\left[\hat{m}_{q,h_n,n}(x)\right]}}\right|<\frac{d}{\sqrt{\mathbf{Var}\left[\hat{m}_{q,h_n}(x)\right]}}\right)\geq 1-\alpha \tag{14}$$

for large n where x is fixed.

Using (14) one can see that the probability requirement (13) leads to the implicit solution-equation

$$n\geq n_{opt}=\left\{\frac{BZ_{\alpha/2}^2\sigma^2}{d^2f(x)}\right\}^{\frac{1}{1-r}} \tag{15}$$

where Z_α is the $100\%\alpha$ upper percentile of the standard normal distribution.

3.1 Purely Sequential Procedure

In general σ^2 in (15) is unknown and purely sequential procedure suggest to substitute the variance parameter σ^2 by a estimator $\hat{\sigma}^2_{n_0}$ based on a sample $(X_1,Y_1),...,(X_{n_0},Y_{n_0})$ of size $n_0<n_{opt}$. Here we use the residual variance estimate of σ^2 proposed by Ursula et.al. (2003) based on covariate matched U-statistics:

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

where

$$\hat{g}_i=\frac{1}{n-1}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h_n}\right) \tag{16}$$

and

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2+o_p(n^{-\frac{1}{2}}) \tag{17}$$

Hence from the optimal sample size n_{opt} given in (15) we continue sampling until

$$n\geq\left\{\frac{Z_{\alpha/2}^2B\hat{\sigma}_n^2}{d^2f(x)}\right\}^{\frac{1}{1-r}} \tag{18}$$

where $\hat{\sigma}_n^2$ is residual variance estimator $\hat{\sigma}^2$ based on sample size n. By taking $n=n_0$ we propose the following stopping rule for purely sequential procedure which is given by

$$N=\max\left\{n,\left\lfloor\left\{\frac{Z_{\alpha/2}^2B\hat{\sigma}_n^2}{d^2f(x)}\right\}^{\frac{1}{1-r}}\right\rfloor+1\right\} \tag{19}$$

where $\lfloor n\rfloor$ refers to the floor function.

In purely sequential procedure we take one observation at a time until the condition given in (18) is satisfied and steps involved in this procedure are as follows:
Step 1: Take an initial sample of size n_0, that is select $\{(X_1,Y_1),...,(X_{n_0},Y_{n_0})\}$ where Y_i is the observed value of $m(X_i)$ at X_i for $i=1,...,n_0$.
Step 2: Now let $n=n_0$ and calculate

$$\left\{\frac{Z_{\alpha/2}^2B\hat{\sigma}_n^2}{d^2f(x)}\right\}^{\frac{1}{1-r}}.$$

Step 3: Compare n with $\left\{\frac{Z^2_{\alpha/2}B\hat{\sigma}^2_n}{d^2 f(x)}\right\}^{\frac{1}{1-r}}$. If

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

then final sample size N equals to n i.e. $N = n$ and no further observations are required and hence the process terminates. Go to step 5. Otherwise go to step 4.
Step 4: If (18) is not satisfied increase sample size by one that is new sample size is $n_0 + 1$ and let $n = n_0 + 1$. Go to step 3.
Step 5: Use the sample $\{(X_1, Y_1), ..., (X_N, Y_N)\}$ to compute $\hat{m}_{NW,h_N}(x_0)$ and $\hat{m}_{LL,h_N}(x_0)$ estimates for $m(x_0)$ and hence construct the confidence band given by (8).

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

$$\frac{\sqrt{nh_n}\left\{\hat{m}_{q,h_n}(x) - m(x)\right\}}{\sigma\sqrt{B(f(x))^{-1}}} \to N(0,1) \tag{20}$$

From (17) for a random sample of normally distributed residuals $\{\varepsilon_i\}_{i=1}^n$ with mean 0 and variance σ^2

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_n \tag{21}$$

where χ^2_n is the chi-squared distribution with n degrees of freedom. Hence we combine (20) and (21) to claim that

$$\frac{\frac{\sqrt{nh_n}\{\hat{m}_{q,h_n}(x) - m(x)\}}{\sigma\sqrt{B(f(x))^{-1}}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} \sim t_n. \tag{22}$$

The following statement (23) is obviously equivalent to (13)

$$\begin{aligned} Pr\left\{m(x) \in I_n\right\} &\approx t\left(\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))^{-1}}\hat{\sigma}}\right) - t\left(-\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))^{-1}}\hat{\sigma}}\right) \\ &= 2t\left(\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))}}\right) - 1 \end{aligned} \tag{23}$$

where $t(\cdot)$ 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

$$2t\left(\frac{\sqrt{nh_n}d}{\sqrt{B(f(x))^{-1}}\hat{\sigma}}\right) - 1 \geq 1 - \alpha \tag{24}$$

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

$$n \geq \left(\frac{t^2_{n,\alpha/2}B\hat{\sigma}^2}{d^2 f(x)}\right)^{\frac{1}{1-r}} \tag{25}$$

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

Two-stage sampling procedure is initiated by taking a pilot bi-variate sample of size n_0 i.e. $\{X_i, Y_i\}_{i=1}^{n_0}$ and then estimate the required final sample size N. Now using (25) we propose the following stopping rule for a two-stage procedure

$$N \equiv N(d) = \max\left\{n_0, \left\lfloor\left(\frac{t^2_{n_0,\alpha/2}B\hat{\sigma}^2_{n_0}}{d^2 f(x)}\right)^{\frac{1}{1-r_1}}\right\rfloor + 1\right\} \tag{26}$$

where $r_1 \in (r_{\min}, 1)$ and from (7) $r_{\min} = \max(0, -\ln[\min(x_0, 1-x_0)]/\ln[n_0])$

If $N = n_0$ then we need no more observations in the second stage. However, 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 (13). In an application of the stopping rule (26), it is important to select the best available values for the design constants r_1 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

$$B = \int_{-\infty}^{\infty} K^2(u)du = \int_{-\infty}^{\infty} \frac{1}{2\pi}e^{-u^2}du = (2\sqrt{\pi})^{-1}.$$

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 $\overline{\hat{m}_{LL,h_N}}$, average Nadaraya-Watson $\overline{\hat{m}_{NW,h_N}}$ 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 2 and 3 for $\alpha = 0.05$. In Tables 2 and 3, figures enclosed in brackets under estimated values refer to their standard errors of the corresponding estimated value.

The following are further definitions of the statistics which head some of the columns in the tables:

- $\bar{n} = \frac{\sum_{j=1}^{n_{sim}} (N)_j}{n_{sim}}$

where $(N)_j$ is final sample size given in either (26) or (19) depending on the sequential procedure being used, calculated from j^{th} simulated sample and n_{sim} (= 15000) is number of simulated samples.

- $SE_{\bar{n}} = \sqrt{\frac{\sum_{j=1}^{n_{sim}} ((N)_j - \bar{n})^2}{(n_{sim}-1)n_{sim}}}$

- $\%Over = ((\bar{n} - n_{opt})/n_{opt})\,100\%$

- $\overline{\hat{m}_{LL,h_N}(x_0)} = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} (\hat{m}_{LL,h_N}(x_0))_j$

- $SE_{\overline{\hat{m}_{LL,h_N}(x_0)}} = \sqrt{\frac{\sum_{j=1}^{n_{sim}} \left(\hat{m}_{LL,h_N}(x_0))_j - \overline{\hat{m}_{LL,h_N}(x_0)}\right)^2}{n_{sim}(n_{sim}-1)}}$

where $(\hat{m}_{LL,h_N}(x_0))_j$ is the estimated value of local linear estimator for j^{th} simulated sample.

- $\overline{\hat{m}_{NW,h_N}(x_0)} = \frac{1}{n_sim} \sum_{j=1}^{n_{sim}} (\hat{m}_{NW,h_N}(x_0))_j$

- $SE_{\overline{\hat{m}_{NW,h_N}(x_0)}} = \sqrt{\frac{\sum_{j=1}^{n_{sim}} \left(\hat{m}_{NW,h_N}(x_0))_j - \overline{\hat{m}_{NW,h_N}(x_0)}\right)^2}{n_{sim}(n_{sim}-1)}}$

where $(\hat{m}_{NW,h_N}(x_0))_j$ is the estimated value of Nadaraya-Watson estimator for j^{th} simulated sample.

- $\overline{\hat{\sigma}^2} = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} (\hat{\sigma}^2)_j$

- $SE_{\overline{\hat{\sigma}^2}} = \left\{ \frac{1}{n_{sim}-1} \sum_{j=1}^{n_{sim}} \left((\hat{\sigma}^2)_j - \overline{\hat{\sigma}^2} \right)^2 \right\}^{1/2}$

where $(\hat{\sigma}^2)_j$ is the estimated value of local linear estimator for j^{th} simulated sample.

- $\tilde{p}_{LL} = \frac{n_{\hat{m}_{LL}(x_0)}}{n_{sim}}$

where $n_{\hat{m}_{LL}(x_0)}$ is the number of local linear fixed with confidence intervals that contain $m(x_0)$ among n_{sim} confidence intervals in other words number of confidence intervals which satisfied $|\hat{m}_{LL,N_j}(x_0) - m(x_0)| < d$ where $j = 1, ..., n_{sim}$

- $SE_{\tilde{p}_{LL}} = \sqrt{\frac{\tilde{p}_{LL}(1-\tilde{p}_{LL})}{n_{sim}}}$

- $\tilde{p}_{NW} = \frac{n_{\hat{m}_{NW}(x_0)}}{n_{sim}}$

where $n_{\hat{m}_{NW}(x_0)}$ is the number of local linear fixed with confidence intervals that contain $m(x_0)$ among n_{sim} confidence intervals in other words number of confidence intervals which satisfied $|\hat{m}_{NW,N_j}(x_0) - m(x_0)| < d$ where $j = 1, ..., n_{sim}$

- $SE_{\tilde{p}_{NW}} = \sqrt{\frac{\tilde{p}_{NW}(1-\tilde{p}_{NW})}{n_{sim}}}$

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 to the bias terms which heavily depend on derivatives of the unknown function $m(\cdot)$ associated with each estimator. $\tilde{p}_{NW}$ of Model I is increasing with decreasing d due to large sample sizes resulted in increase in sample sizes. This is consistent with both sequential procedures i.e. two-stage and purely sequential. The performance of Nadaraya-Watson estimator worsens as x increases as its bias highly depends on derivatives of $m(\cdot)$. For the interior point $x_0 = .306$, the Nadaraya-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 sequential 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 sequential procedure is reflected in computational time. The purely sequential procedure needs substantially more computations and hence during simulations

Table 2: 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)

Table 3: Empirical coverage of LL and NW for Model II
$\alpha = .05;\ m(x_0) = 3.024$

d	n_{opt}	$\bar{n}$	$\tilde{p}_{LL}$	$\tilde{p}_{NW}$	$\overline{\hat{m}_{LL}}$	$\overline{\hat{m}_{NW}}$	$\overline{\hat{\sigma}^2}$
		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)

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 the half width of the interval d becomes larger as this result in smaller sample sizes.

Figure 4.0 reflects the amount of over/under sampling ($\%Over$) of average sample sizes $\bar{n}$ from corresponding optimal sample sizes n_{opt} based on confidence intervals constructed using two-stage sequential procedure and purely sequential procedure for each half with of the interval d. Average sample size $\bar{n}$ from two-stage sampling method is over sampling whereas $\bar{n}$ computed from purely sequential procedure is slightly undersampling. In practice, the focus is on final sample size, $\bar{n}$ to be as close as possible to optimal sample size, n_{opt} with a reasonable coverage probability. Therefore, we can conclude that confidence intervals based on purely sequential procedure has fulfilled the required goal of this study.

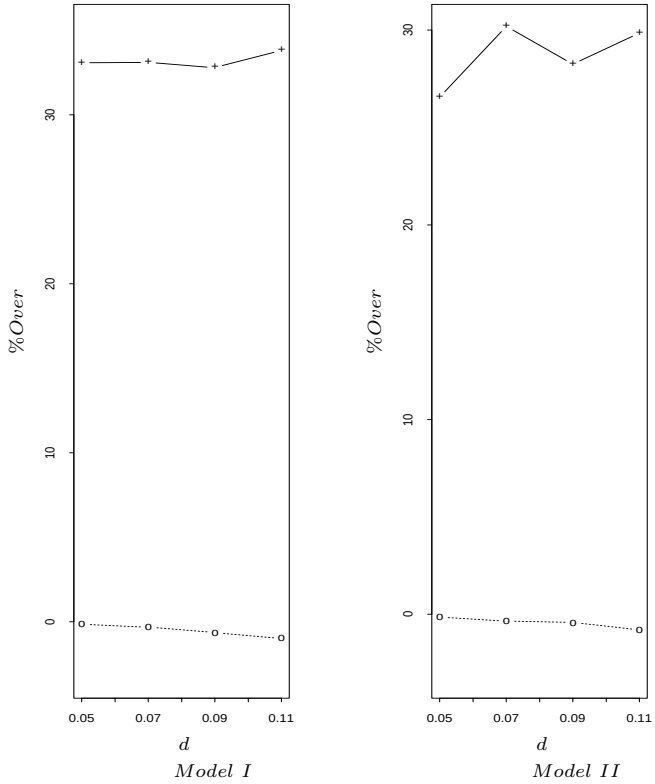

Figure 4.0 **Over % Vs half-width of the Interval, d**
(*Legend*: '+' = Two-stage, 'o' = Purely Sequential)

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 estimated residual variance estimator also appears 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.