ESKİŞEHİR TEKNİK ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
B- TEORİK BİLİMLER

Eskişehir Technical University Journal of Science and Technology B- Theoretical Sciences

Internatıonal Conference on Natural Sciences and Technologies
Iconat Special Issue 2021

2021, Vol:9, pp. 94-102, DOI: 10.20290/estubtdb.1015940

NONPARAMETRIC REGRESSION WITH ERROR-IN-VARIABLES MODEL BASED ON DIFFERENT KERNEL FUNCTIONS

Dursun AYDIN [1], , Ersin YILMAZ [1] *

[1] Department of Statistics, Faculty of Science, Mugla Sıtkı Koçman University, Mugla, Turkey

ABSTRACT

Estimation of error-invariable models is a specific problem in different fields such as medicine, economics, industry, and biostatistics. The main different between classical regression and error-in-variable models is that explanatory variables involve random error terms. Therefore, classical estimation methods that do not include the necessary adjustments for the contaminated explanatory variables give biased results. Regarding the error-in variables, there are important studied in the literature such as [1], [2], [3], [4], [5] and [6]. In this paper, nonparametric regression with measurement error is considered and estimated by kernel smoothing estimator which is studied detailed by [6]. This paper differs from their study with the idea of using two different kernel functions to compared them on quality of estimations. These functions are suitable for different error behaviors (see [7]). The goal of the paper is encouraged by a Monte Carlo simulation study and results are presented.

Keywords: Error in variables, kernel smoothing, nonparametric regression, kernel functions

1. INTRODUCTION

The regression models are used in a number of science fields such as economics, medicine, biostatistics and so on. Usually, in classical regression models, explanatory variables are assumed as measured without errors. However, in in the real world, this assumption is not always ensured since variables are contaminated by error. In this paper, two different approaches are introduced to estimate a nonparametric regression model with errors-in-variables (EIV) based on kernel deconvolution method. There are different kinds of kernel functions that used in the deconvolution process. This study aims to show the effects of two common kernel functions to the estimation of the nonparametric model. Suppose that we are interested in relationship between a response variable $Y \in R^+$ and a covariate $X \in R$. In the classical nonparametric regression problem, the purpose is to estimate a regression function $m(.)$, using a sample $\{Y_i, X_i\}_{i=1}^n$:

$$Y_i = m(X_i) + \varepsilon_i, \;\; a = X_1 < \cdots < X_n = b \tag{1.1}$$

where the $\varepsilon_i \sim N(0, \sigma^2)$ and variance σ_ε^2, and m(.) is an unknown smooth regression function to be estimated. Therefore, we observe a new updated dataset $(Y_1, W_1), \ldots, (Y_n, W_n)$, where

$$\begin{aligned} W_i &= X_i + U_i \\ Y_i &= m(X_i) + \varepsilon_i, \;\; E(\varepsilon_i | X_i) = 0. \end{aligned} \tag{1.2}$$

Here X_i's have measurement error, Y_i and W_i's are the observed variables, and ε_i's are the model or regression errors, while U_i's are the so called-measurement errors.

*Corresponding Author:yilmazersin13@hotmail.com
Received: 28.10.2021 Published: 24.12.2021

The key idea of this study is to estimate $m(.)$ in a nonparametric regression model with when the covariate is subject to measurement error. Figure 1 shows the data with measurement error. Black dots denote the covariate with errors and red dots Show the covariate without errors. The main problem can be said as the presence of measurement errors leads to biased and inconsistent parameter estimates. This paper aims to introduce two modified estimators for estimation of the nonparametric regression that solve the EIV problem.

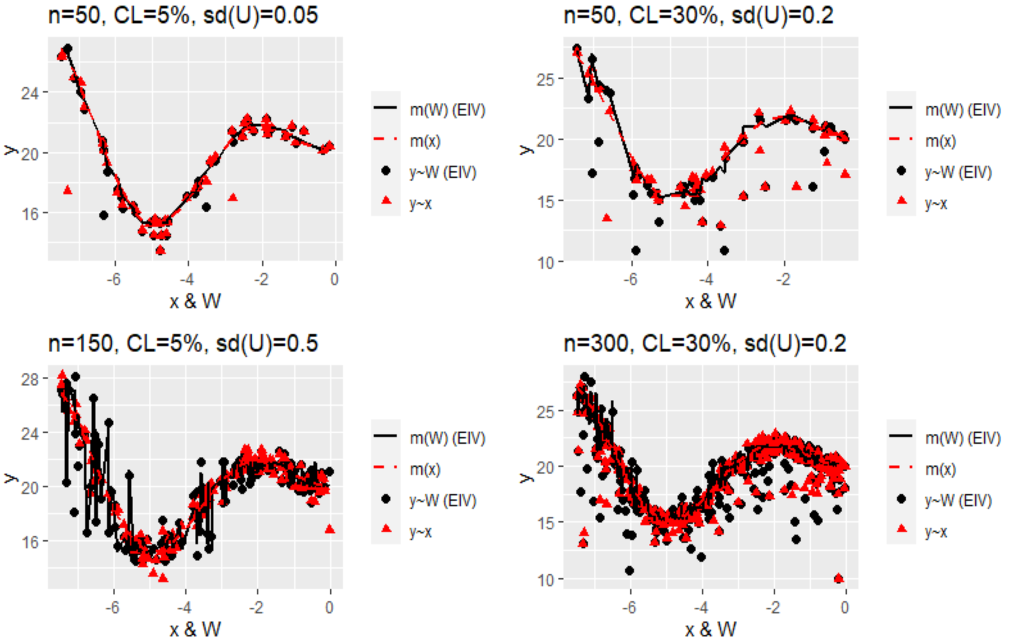

Figure 1. Generated dataset with measurement error

This estimation problem based on model (1.2), referred to as an errors-in-variables problem, has attracted great attention in the literature. Examples of these works include [1], [2], [3], [8], [4], [9], [5], [6].

Remain of the paper is organized as follows: In Section 2, derivation of the estimators is introduced. Section 3 presents some statistical properties of the method. Section 4 involves the simulation study and conclusions are given in Section 5.

2. DERIVATION of THE ESTIMATORS for EIV MODEL

In the regression analysis with measurement error, one needs first to make some identification assumptions on variables and their dependence relationships. The following assumptions to assure that the model is identifiable.

Assumptions:
A1. The random error vector U has a known distribution, and it is mutually independent of X, and Y.
A2.The measurement error U, the model error ε, and the covariate X are independent mutually and satisfy $E(U) = 0, Var(U) = \sigma_u^2, E(\varepsilon) = 0, Var(\varepsilon) = \sigma_\varepsilon^2$.
A3. The random error vector U has a known distribution, and it is mutually independent of X and Y.

Let $W_i = X_i + U_i, i = 1,2, \dots, n$ be a random sample from the distribution of $W_i = X_i + U_i$. One of the main problems in errors in variable regression is to estimate the density function $f_X(.)$ of a random variable X. From A3 we see that the distribution of U is known, and that X and U are independent. We assume that the unobservable X_i's have probability distribution function F_X and probability density function f_X, and the U_i's have a known density function f_U. Then the observations from random variable $W_i = (X_i + U_i)$ have the probability density function, given by

$$f_W(w) = \int f_X(w-x) f_U(x) dx = f_X * f_U \tag{2.1}$$

where "$*$" indicates convolution of the two density functions. Supposing f_U is known we consider estimating f_X from a set of independent data points $\{W_i, i = 1, \ldots, n\}$ with ordinary probability density function f_W. Since the variables $X_1, \ldots, X_n$ are not directly observable due to the measurement error, $f_X(.)$ will be estimated from the data $W_i = X_i + U_i$. This procedure, the problem of estimating the density f_X from the values W_i by using the convolution density f_W (2.1), is often referred to as a non-parametric deconvolution problem.

In the deconvolution problem, the first operation is to estimate the density f_W of the values W_i. Therefore, we consider a problem of density estimation that can be solved by different methods. But we see that the focus at this stage is on the Fourier transform and kernel smoothing. Let $\phi_X(.) = E(e^{itx}),\ \phi_W(.) = E(e^{itw}),\ and\ \phi_U(.) = E(e^{itu})$ indicate the characteristic functions of X, W and U, respectively. Suppose that $\phi_U(t) \neq 0$ and for all $t \in R$. An inverse Fourier transform provides that the density function f_X of random variable X which is defined as follows

$$f_X(x) = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{-itx} \phi_X(t) dt = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{-itx} \frac{\phi_W(t)}{\phi_U(t)} dt \tag{2.2}$$

Since that X and U are assumed to be independent, $\phi_X(t) = \frac{\phi_W(t)}{\phi_U(t)}$. Note that a naive estimator of the density f_X can be computed by using the equation (2.2). However, in practice, this estimator is unstable because the sample characteristic function has large variations at its tails. To avoid this problem, the estimation of f_X can be expressed by its kernel estimator:

$$\hat{\phi}_W(t) = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{itw} \hat{f}_W(w)\, dw = (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{itw} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{w - W_i}{h}\right) dw \tag{2.3}$$

where $\hat{f}_W(w)$ is the ordinary kernel probability density estimator of the density function $f_W, h > 0$ is a bandwidth or smoothing parameter.

$K(.)$ is a measurable function (or kernel function) that satisfies the following properties:
$\int K(x)dx = 1, \int xK(x)dx = 0, \mu_2 = \int x^2 K(x) dx < \infty, K(x) \geq 0\ for\ all\ x, and\ \ K(x) = K(-x)$

These properties denote that a kernel function needs to be symmetric, and continuous probability density function with mean zero and constant variance. It is also noted that the smoothing parameter h should be chosen optimally in kernel estimation. A large $h > 0$ provides an extremely smooth estimate, while a small $h > 0$ produces a wiggly function curve. In this context, the generalized cross validation (GCV) method is used to determine the parameter h that gives the required amount of smoothness. If $K(x)$ is a kernel function, then $K(x) = K(x/h)$ is also a kernel function based on h, a positive bandwidth parameter. Note that the resulting deconvolution kernel estimator for the density f_X based on $\hat{\phi}_W(t)$ in (2.3), discussed by [10] is defined by

$$\hat{f}_X(x) = (nh)^{-1} \sum_{i=1}^{n} K_U\left(\frac{x - W_i}{h}\right) \tag{2.4}$$

The main difference of the study from the others is to use two types of $K_U(.)$ that are introduced in [6] obviously;

$$K_U^1(t) = \frac{1}{\sqrt{2\pi\left(1-\frac{\sigma^2}{h^2}\right)}} e^{-\frac{t^2}{2\left(1-\frac{\sigma^2}{h^2}\right)}}, \quad K_U^2(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \left(1 + \left(\frac{\sigma}{h}\right)^2 (1-t^2)\right) \tag{2.5}$$

Note that $K_U^1(t)$ provides satisfying estimates when the magnitude of the error variance is small. $K_U^2(t)$ works under general cases. This study aims to put forward their behaviours under EIV model estimation in nonparametric setting.

2.1. Nonparametric Kernel Estimator with EIV

After the sloution of the deconvolution problem, nonparametric regression model is estimated by using any smoothing method in the literature such as Local Polynomial smoothing, Kernel smoothing, smoothing splines etc…In this study, Kernel smoothing is determined to estimate the model. Thus, difference between kernel functions used in deconvolution process can be inspected more easily. Because other smoothing techniques have additional parameters to be selected. Let $(X_1, Z_1), \ldots, (X_n, Z_n)$ show a random sample from a pair of random variables (X, Z) and let $K(.)$ present a kernel function. To describe the deconvolution kernel approach, we first consider a classical nonparametric regression model in which X_i's are not contaminated.

The kernel estimator of the regression function in model (1.1) is obtained as follows

$$\begin{aligned} \widehat{m}(x) &= \textstyle\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Z_i \Big/ \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \\ &= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Z_i \Big/ \hat{f}_X(x) = \frac{\hat{g}(x)}{\hat{f}_X(x)} \end{aligned} \tag{2.6}$$

where $\hat{f}_X(x)$ denotes the kernel estimator of the marginal probability density function f_X of covariate X, whereas $\hat{g}(x)$ shows the kernel estimator of the joint density function $g(X,Z)$ of (X,Z). Note that the estimator in (2.6) expressed as the ratio of two density functions is known as the Nadaraya-Watson estimator proposed by [11-12]. Now we consider the right censored nonparametric regression model with measurement errors.

Since the variables $X_1, \ldots, X_n$ are not directly observable, the kernel estimator $\hat{f}_X(x)$ given in (2.4) is constructed by the deconvolution method (2.1). Based on this idea, we replace the estimators of denominator of the equation (2.6) with the deconvolution kernel estimator (2.4). Note also that the estimator of the numerator in the same equation should be constructed from the contaminated observations $W_i's$. In keeping with the spirit of deconvolution kernel density estimator, we propose to replace $\hat{g}(x)$ in (2.6) with

$$\hat{g}(x) = (nh)^{-1} \sum_{i=1}^n K_U\left(\frac{x-W_i}{h}\right) Y_i \tag{2.7}$$

Accordingly, the estimate of $m(.)$ at fixed x (expressed by $\widehat{m}_{DKE}(x;h)$) can be computed as

$$\widehat{m}_{DKE}(x;h) = \frac{\hat{g}(x)}{\hat{f}_X(x)} \sum_{i=1}^n K_U\left(\frac{x-W_i}{h}\right) Y_i \Big/ \sum_{i=1}^n K_U\left(\frac{x-W_i}{h}\right) = \sum_{i=1}^n W_i(x, W_i) Y_i = \hat{Y}_i \tag{2.8}$$

where $\hat{f}_X(x)$ and K_U's are defined by (2.5). Note that the deconvoluted kernel regression estimator (denoted by $\widehat{\boldsymbol{m}}_h^{DKE}$) can be rewritten in the following matrix and vector form

$$\widehat{m}_h(x) = \left(\widehat{m}_h(X_1), \ldots, \widehat{m}_h(X_n)\right)' = \widehat{\boldsymbol{m}}_h^{DKE} = \boldsymbol{S}_h^{DK}\boldsymbol{Y} = \widehat{\boldsymbol{Y}}_{DKE} \tag{2.9}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ is the vector of the censored response observations and $\boldsymbol{S}_h^{DK}$ is a deconvoluted kernel (DK) smoother matrix defined by kernel weights multiplied with $(nh)^{-1}$:

$$W_i(x, W_i) = K_U\left(\frac{x - W_i}{h}\right) \Big/ \sum_{i=1}^{n} K_U\left(\frac{x - W_i}{h}\right) = K_U \Big/ \sum K_U \tag{2.10}$$

with a smoothing parameter h, which is a nonnegative scaler determining the degree smoothness of curve obtained by $\widehat{\boldsymbol{m}}_h^{DKE}$.

Based on the ideas of [7], the smoothness of distribution of a random variable X is described in terms of the function $\phi_X(t) as\ t \to \infty$. Note that the optimal rates of convergence for the deconvoluted estimator given in this study depends on the degree (b) of smoothness of the error distributions. The optimal rates of convergence for these estimators also depend on the smoothness of function $m(x)$ and the regularity conditions on the marginal density function. Essentially, these conditions are based on the following assumptions:

Assumptions:
B1. The marginal density $f_X(.)$ of the unobserved variable X is bounded away from on 0 on $[a, b]$, where $a < b$, and has a bounded k^{th} derivative, where k is a positive scaler.
B2. The characteristics function of error distribution $\phi_U(.)$ does not vanish
B3. The nonparametric regression function $m(x)$ has a continuous k^{th} derivative on $[a, b]$
B4. The second moment $E(Y^2|X = x)$ is continuous on $[a, b]$ and also, $E(Y^2) < \infty$.
B5. The distribution of the measurement error variable U is ordinary smooth or super smooth

3. STATISTICAL PROPERTIES OF ESTIMATOR

In this section, some statistical properties are introduced with Definition 2.1-2.2.

Definition 5.1: The distribution of random variable U is stated to be super smooth of degree b: if the characteristic function of the error distribution $\phi_U(.)$ satisfies

$$d_0|t|^{b_0} exp\left(-\frac{|t|^b}{\gamma}\right) \leq |\phi_U(t)| \leq d_1|t|^{b_1} exp\left(-\frac{|t|^b}{\gamma}\right) as\ t \to \infty \tag{3.1}$$

where d_0, d_1, b and γ are nonnegative constants and b_0, b_1 are constants.

Definition 5.2: The distribution of random variable U is expressed to be ordinary smooth of degree b: if the tail of the characteristic function $\phi_U(.)$ satisfies $\phi_U(t) \neq 0$ for all $t \in R$, and if there are nonnegative constants d_0, d_1, b such that

$$d_0|t|^{-b} \leq |\phi_U(t)| \leq d_1|t|^{-b} as\ t \to \infty \tag{3.2}$$

The typical examples of super smooth distributions are normal $N(0,1)$ distribution with $b = 2$, Cauchy $([\pi(1 + x)]^{-1})$ distribution with $b = 1$, and their mixtures. The examples of ordinary smooth distributions are gamma distribution $\left(\frac{\alpha^p x^{p-1} exp(-\alpha x)}{\Gamma(p)}\right)$ of degree p with $b = p$ and Laplace or double exponential $\left(2^{-1}exp(-|x|)\right)$ distributions with $b = 2$, symmetric gamma and their mixtures (see, for example, [7], [1] and [13]).

3.1. Bandwidth Selection

It is useful to denote the deconvoluted estimators evaluated at the design point $\left(\hat{m}_h(X_1), \dots, \hat{m}_h(X_n)\right)$ as a linear operator:

$$\widehat{\boldsymbol{m}}_h = \boldsymbol{S}_h \boldsymbol{Y}^* = \widehat{\boldsymbol{Y}}$$

where $\boldsymbol{S}_h: R^n \rightarrow R^n$ is a smoother matrix. The considered kernel estimator in this paper need a choice of bandwidth parameter h. In this context, the aforementioned parameter h can be selected by minimizing the generalized cross validation (GCV) criterion (see [14]), defined by

$$GCV(h) = \frac{1}{n}\|(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{Y}^*\|^2 \Big/ \left[\frac{1}{n} tr(\boldsymbol{I} - \boldsymbol{S}_h)\right]^2 = \frac{n[RSS(h)]}{[EDF(h)]^2}$$

where $RSS(.)$ means the residual sum of squares while $EDF(.)$ refers to the equivalent degrees of freedom. Our task in this section is to choose the optimum value of h. The optimum h can be achieved by using the mean summed squared error (MSSE), given by

$$MSSE(\widehat{\boldsymbol{m}}_h, \boldsymbol{m}) = \|(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}\|^2 + \sigma_\varepsilon^2 tr(\boldsymbol{S}_h \boldsymbol{S'}_h) \tag{3.3}$$

Since the optimal $MSSE(\widehat{\boldsymbol{m}}_h)$ in (3.3) depends on the unknown amount of σ_ε^2, it cannot be directly applied in practice. In this case, as in classical linear regression, the variance σ_ε^2 can be estimated by the sum of residual squares;

$$\hat{\sigma}^2 = \frac{RSS(\hat{h})}{EDF(\hat{h})} = \frac{\|(\boldsymbol{I} - \boldsymbol{S}_{\hat{h}})\boldsymbol{Y}^*\|^2}{tr(\boldsymbol{I} - \boldsymbol{S}_{\hat{h}})} \tag{3.4}$$

4. Simulation Study

Data generation:
Regarding to model (1.2), $Y_i = m(X_i) + \varepsilon_i$ with $W_i = X_i + U_i$, each variable is obtained as:

$$X_i \sim Unif(0,3)\ ;\ \ \varepsilon_i \sim N(\mu_\varepsilon = 0, \sigma_\varepsilon^2 = 0.5),\ \ U_i \sim N(\mu_U = 0, \sigma_U^2);\ \ m(X_i) = -W_i sin{-}W_i)$$

Simulation Design:

- Simulation experiments are performed for three samples of size $n = 50, 150$ and 300, and all simulation combinations are repeated 1000 times.
- Also, three values of $\sigma_U (i.e., \sigma_U = 0.05, 0.2, 0.5)$ are examined to understand how the methods work at different levels of variance.

Bandwidth selection is made by GCV in simulation and it is shown in the Figure 1. The «h» value is determined as an optimal bandwidth which minimizes the GCV criterion.

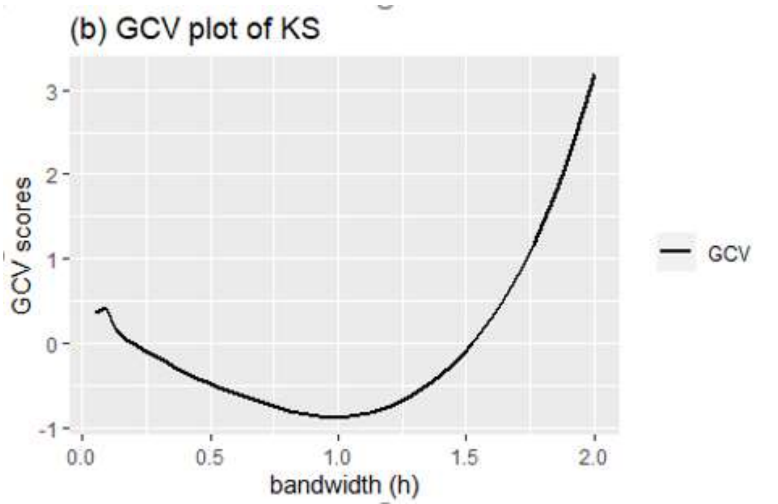

Figure 1. Bandwidth selection with GCV for KS

Then, outcomes of the estimated nonparametric models with measurement errors can be given. To measure the risk and performance of the mentioned methods, MSSE is used. Also, to evaluate risk of the methods, relative efficiency (RE) is used that is defined before. From that Table 1 involves the MSSE and RE values are given below for all simulation configurations. Results show that $K_U^1(.)$ and $K_U^2(.)$ give close scores for lower variance values. However, when the variance is getting larger, $K_U^1(.)$ gives better scores which ensures our claim.

Table 1. MSSE and RE scores

		$K_U^1(.)$		$K_U^2(.)$	
	σ_U	MSSE	RE	MSSE	RE
50	0.05	**0.306**	**0.939**	0.326	1.065
	0.20	**1.015**	**0.979**	1.037	1.022
	0.50	**1.104**	**0.745**	1.481	1.341
150	0.05	0.284	1.229	**0.231**	**0.813**
	0.20	0.789	1.467	**0.538**	**0.682**
	0.50	**0.840**	**0.815**	1.031	1.227
300	0.05	**0.121**	**0.786**	0.154	1.273
	0.20	**0.672**	**0.908**	0.740	1.101
	0.50	**0.739**	**0.624**	1.185	1.604

Figure 2 shows the fitted curves based on different scenarios. The two plots at the top denote the estimated functions based on the first kernel deconvolution function $K_U^1(.)$ and the two at the bottom represent the $K_U^2(.)$. If plots are inspected in detail, it can be seen that $K_U^1(.)$ gives better in general. However, two functions are highly close to each other when $\sigma_u = 0.05$.

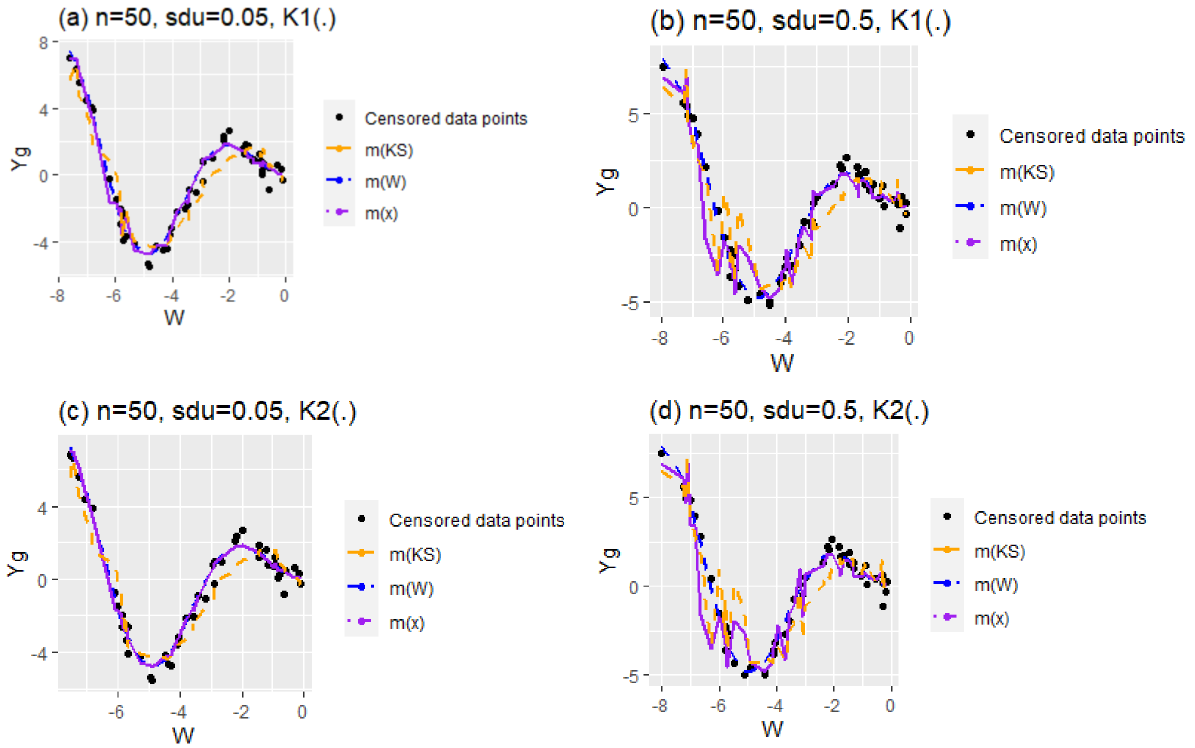

Figure 2. Fitted functions with KS under EIV

5. CONCLUSIONS

This paper focused on the estimation of the data with measurement error by using two different kernel deconvolution methods that are notated *as* $K_U^1(.)$ and $K_U^2(.)$. Note that $K_U^2(.)$ Works well for lower variance of error of nonparametric covariate U, σ_u and $K_U^1(.)$ is used in general cases. Accorndingly, two nonparametric estimators are introduced based on kernel smoothing approach. To show estimation performances of the mentioned methods practically, a simulation study is carried out. Based on the results, as we claimed, $K_U^1(.)$ gives better scores in terms of MSSE and RE measures. However, K_U^2 (.) shows itself when variance is low. Also, the comparative results from numerical studies and Figure 2 show that Kernel smoothing gives satisfying fitted curves that can be recommended to estimate nonparametric model with measurement error.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] Fan J & Truong YK. Nonparametric regression with errors in variables. The Annals of Statistics, 1993; 1900-1925.

[2] Stefanski LA & Cook JR. Simulation-extrapolation: the measurement error jackknife. Journal of the American Statistical Association, 1995; 90(432): 1247-1256.

[3] Carroll R J Maca, JD & Ruppert D. Nonparametric regression in the presence of measurement error. Biometrika, 1999; 86(3): 541-554.

[4] Carroll RJ & Hall P. Low order approximations in deconvolution and regression with errors in variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2004; 66(1): 31-46.

[5] Delaigle A & Meister A. Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. Journal of the American Statistical Association, 2007; 102(480): 1416-1426.

[6] Wang XF & Wang B. Deconvolution estimation in measurement error models: the R package decon. Journal of statistical software, 2011; 39(10): i10.

[7] Fan J. On the optimal rates of convergence for nonparametric deconvolution problems. The Annals of Statistics, 1991; 1257-1272.

[8] Berry SM Carroll RJ & Ruppert D. Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association, 2002; 97(457): 160-169.

[9] Liang H & Wang N. Partially linear single-index measurement error models. Statistica Sinica, 2005; 99-116.

[10] Stefanski LA & Carroll RJ. Deconvolving kernel density estimators. Statistics, 1990; 21(2): 169-184.

[11] Nadaraya EA. On estimating regression. Theory of Probability & Its Applications, 1964; 9(1): 141-142.

[12] Watson GS. Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A, 1964; 26(4): 359-372.

[13] Li T & Vuong Q. Nonparametric estimation of the measurement error model using multiple indicators. Journal of Multivariate Analysis, 1998; 65(2): 139-165.

[14] Craven P & Wahba G. Smoothing noisy data with spline functions. Numerische Mathematik, 1979; 31(4): 377-403.