

# Nonparametric Regression Estimation with Mixed Measurement Errors

Zanhua Yin, Fang Liu, Yuanfu Xie

College of Mathematics and Computer Science, Gannan Normal University, Ganzhou, China

Email: yinzh226@163.com

**How to cite this paper:** Yin, Z.H., Liu, F. and Xie, Y.F. (2016) Nonparametric Regression Estimation with Mixed Measurement Errors. *Applied Mathematics*, 7, 2269-2284. <http://dx.doi.org/10.4236/am.2016.717179>

**Received:** May 20, 2016

**Accepted:** November 27, 2016

**Published:** November 30, 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

We consider the estimation of nonparametric regression models with predictors being measured with a mixture of Berkson and classical errors. In practice, the Berkson error arises when the variable  $X$  of interest is unobservable and only a proxy of  $X$  can be measured while the inaccuracy related to the observation of the proxy causes an error of classical type. In this paper, we propose two nonparametric estimators of the regression function in the presence of either or both types of errors. We prove the asymptotic normality of our estimators and derive their rates of convergence. The finite-sample properties of the estimators are investigated through simulation studies.

## Keywords

Berkson Error, Classical Error, Deconvolution, Kernel Method, Mixed Measurement Errors

## 1. Introduction

Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  denote a sequence of independent and identically distributed random vectors. In traditional non-parametric regression model analysis, one is interested in the following model

$$Y_i = g(X_i) + e_i, \quad (1)$$

where  $g(\cdot)$  is assumed to be a smooth, continuous but unknown function; the random errors  $e_i$  are assumed to be normally and independently distributed with mean 0 and constant variance  $\sigma^2$ ; and  $i = 1, \dots, n$ . Here, the predictor  $X$  is usually assumed to be directly observable without errors. Both the direct observation and error-free assumptions are however seldom true in most epidemiologic studies. For the violation of the error-free assumption, [1] considered an environmental study which studied the relation of mean exposure to lead up to age 10 (denoted as  $X$ ) with

intelligence quotient (IQ) among 10-year-old children (denoted as  $Y$ ) living in the neighborhood of a lead smelter. Each child had one measurement made of blood lead (denoted as  $W$ ), at a random time during their life. The blood lead measurement (*i.e.*,  $W$ ) became an approximate measure of mean blood lead over life ( $X$ ). However, if we were able to make many replicate measurements (at different random time points), the mean would be a good indicator of lifetime exposure. In other words, the measurements of  $X$  are subject to errors and  $W$  is a perturbation of  $X$ . In the measurement error literature, this is known as the *classical error model* and Model (1) becomes

$$Y_i = g(X_i) + e_i, \quad W_i = X_i + \varepsilon_i, \quad (2)$$

where  $(e_i, X_i, \varepsilon_i), i = 1, \dots, n$ , are mutually independent and  $\varepsilon$  represents the classical measurement error variable. Various methods and approaches for analyzing Model (2) such as deconvolution kernel approaches (e.g., [2] [3] [4]), design-adaptive local polynomial estimation method (e.g., [5]), methods based on simulation and extrapolation (SIMEX) arguments (e.g., [6] [7] [8] [9]), and Bayesian approach (e.g., [10]) have been extensively studied in the literature.

In many studies, it is however too costly or impossible to measure the predictor  $X$  exactly or directly. Instead, a proxy  $W$  of  $X$  is measured. For the violation of the direct observation assumption, [1] modified the aforementioned environmental study in which the children's place of residence at age 10 (assumed known exactly) were classified into three groups by proximity to the smelter—close, medium, far. Random blood lead samples, collected as describe in the aforementioned design, were averaged for each group (denoted as  $W$ ), and this group mean used as a proxy for lifetime exposure for each child in the group. Here, the same approximate exposure (proxy) is used for all subjects in the same group, and true exposures, although unknown, may be assumed to vary randomly about the proxy. This is the well-known *Berkson error model*. In other words, the predictor  $X$  are not directly observable and measurements on its surrogates  $W$  are available instead. The true predictor  $X$  is then a perturbation of  $W$ . The model of interest now becomes

$$Y_i = g(X_i) + e_i, \quad X_i = W_i + \delta_i, \quad (3)$$

where  $(e_i, W_i, \delta_i), i = 1, 2, \dots, n$ , are mutually independent. Model (3) was first considered by [11] and the estimation of the linear Berkson measurement error models was discussed in [12]. Methods based on least squares estimation ([13]), minimum distance estimation ([14] [15]), regression calibration ([16]) and trigonometric functions ([17]) have been studied.

The stochastic structure of Model (3) is fundamentally different from Model (2). Here, the measurement error of Model (2) is independent of  $X$ , but dependent on  $W$ . This distinctive feature leads to completely different procedures in estimation and inference for the models. In particular, nonparametric estimators that are consistent in Model (2) are no longer valid in Model (3), and vice versa. In most of the existing literature, the measurement error is supposed to be only one of the two types. In the Berkson model (3), it is usually assumed that the observable variable  $W$  is measured

with perfect accuracy. However, this may not be true in some situations. In such cases,  $W$  is observed through  $V = W + \varepsilon$ , where  $\varepsilon$  is a classical measurement error. [18] presented a good discussion of the origins of mixed Berkson and classical errors in the context of radiation dosimetry. Under this mixture of measurement errors, we observe a random sample of independent pairs  $(V_i, Y_i)$ , for  $i = 1, \dots, n$ , generated by

$$Y_i = g(X_i) + e_i, \quad X_i = W_i + \delta_i, \quad V_i = W_i + \varepsilon_i, \quad (4)$$

where  $W_i \sim f_w$ ,  $\delta_i \sim f_\delta$ ,  $\varepsilon_i \sim f_\varepsilon$  and  $e_i$  are mutually independent, and the respective error densities  $f_\delta$  and  $f_\varepsilon$  are assumed to be known. Due to its potentially wide applications, statistical procedures for analyzing Model (4) has received more attention recently. For instance, a regression calibration approach was proposed by [19] and [20] in a parametric context of random exposure. [21] considered a bayesian approach for a semi-parametric regression function. [22] developed a nonparametric density estimation approach for contaminated data with a mixture of Berkson and classical errors but without further extending to estimate the regression function. [23] proposed a two-step nonparametric kernel method for estimating the regression function but its calculation is complicated. In this paper, we propose two non-parametric estimators for the regression function curve  $g(\cdot)$  with the predictor  $X$  being measured with either classical error, Berkson error, or a combination of both. The difficulty primarily depends on the relative smoothness of the error densities  $f_\delta$  and  $f_\varepsilon$ . When  $f_\delta$  is smooth enough (relative to  $f_\varepsilon$ ), we are able to construct a nonparametric estimator that converges to the target curve at the parametric  $\sqrt{n}$  rate. For less smooth density  $f_\delta$ , we propose a kernel estimator that converges at rates ranging from  $\sqrt{n}$  to rates that are close to the deconvolution rates.

This paper is organised as follows. In Section 2, we propose estimators for the regression function curve  $g(\cdot)$ . We then derive the asymptotic normality of our estimators under some regularity conditions and give the rates of convergence in Section 3. Section 4 presents some numerical results from simulation studies. A brief discussion will be given in Section 5. All technical results and proofs are deferred to the **Appendix**.

## 2. Proposed Estimators

Let  $(Y_1, V_1), \dots, (Y_n, V_n)$  be a random sample from Models (4), and  $\phi_X(t)$ ,  $\phi_W(t)$ ,  $\phi_\delta(t)$ ,  $\phi_V(t)$  and  $\phi_\varepsilon(t)$  be the characteristic functions of  $X_i$ ,  $W_i$ ,  $\delta_i$ ,  $V_i$  and  $\varepsilon_i$ , respectively. We have the following relationships:

$$\phi_X(t) = \phi_W(t)\phi_\delta(t), \quad \phi_V(t) = \phi_W(t)\phi_\varepsilon(t).$$

Hence, if  $\phi_\varepsilon(t)$  does not vanish,

$$\phi_X(t) = \phi_V(t)\phi_\delta(t)/\phi_\varepsilon(t).$$

Since  $\phi_\delta(t)$  and  $\phi_\varepsilon(t)$  are assumed to be known, an estimate of  $\phi_X(t)$  can be computed as

$$\hat{\phi}_X(t) = \hat{\phi}_V(t) \phi_\delta(t) / \phi_\varepsilon(t), \quad \text{with } \hat{\phi}_V(t) = \frac{1}{n} \sum_{j=1}^n \exp(itV_j).$$

Noticing that, if  $\phi(t)$  is absolutely integrable, the characteristic function  $\phi(t)$  and its density function  $f(x)$  have the following relation

$$f(x) = \frac{1}{2\pi} \int \exp(-itx) \phi(t) dt,$$

under the condition that  $\phi_\delta(t) / \phi_\varepsilon(t) \in L_1(\mathbb{R})$ , the density estimator of  $X_i$  is then given by

$$\hat{f}_X(x) = \frac{1}{2\pi} \int \exp(-itx) \hat{\phi}_X(t) dt = \frac{1}{n} \sum_{i=1}^n G_\varepsilon^\delta(x - V_i), \tag{5}$$

where

$$G_\varepsilon^\delta(x) = \frac{1}{2\pi} \int e^{-itx} \phi_\delta(t) / \phi_\varepsilon(t) dt.$$

As a result, we propose the following estimator for  $g(x)$

$$\hat{g}(x) = \frac{1/n \sum_{i=1}^n G_\varepsilon^\delta(x - V_i) Y_i}{1/n \sum_{i=1}^n G_\varepsilon^\delta(x - V_i)}. \tag{6}$$

**Example 1** Let the error densities  $f_\delta$  and  $f_\varepsilon$  in Model (4) be normal densities with mean zero and variances  $\sigma_\delta^2$  and  $\sigma_\varepsilon^2$ , respectively. It follows that

$\phi_\delta(t) / \phi_\varepsilon(t) = \exp\left(-\frac{1}{2} \sigma_1^2 t^2\right)$  with  $\sigma_1^2 = \sigma_\delta^2 - \sigma_\varepsilon^2$ . If we assume  $\sigma_\delta^2 > \sigma_\varepsilon^2$ , then the ratio  $\phi_\delta(t) / \phi_\varepsilon(t)$  is the characteristic function of another normal random variable. By (6), the estimator of  $g(x)$  can be written as

$$\hat{g}(x) = \frac{1/n \sum_{i=1}^n \psi_{(0,\sigma_1)}(x - V_i) Y_i}{1/n \sum_{i=1}^n \psi_{(0,\sigma_1)}(x - V_i)}$$

where  $\psi_{(0,\sigma_1)}(x)$  is the density of the Normal(0,  $\sigma_1^2$ ) variable. If  $\sigma_\delta^2 < \sigma_\varepsilon^2$ , the ratio  $\phi_\delta(t) / \phi_\varepsilon(t)$  is not integrable, and the estimators (5) and (6) can not be calculated. To overcome this issue, we propose an alternative approach for estimating  $g(x)$ .

Using a kernel function  $K(x)$  with a bandwidth  $h$ , we consider the following kernel estimator for  $f_V(x)$

$$\hat{f}_V(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - V_i}{h}\right),$$

and an estimator for  $\phi_V(t)$  is then given by

$$\hat{\phi}_V(t) = \frac{1}{n} \sum_{i=1}^n \exp(itV_i) \phi_K(th),$$

where  $\phi_K(t)$  is the characteristic function of the kernel function  $K(x)$ .

Proceeding as above, we get an alternative estimator of  $f_X(x)$  by

$$\tilde{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K_\varepsilon^\delta\left(\frac{x - V_i}{h}\right), \tag{7}$$

where

$$K_\varepsilon^\delta(x) = \frac{1}{2\pi} \int e^{-itx} \phi_K(t) \frac{\phi_\delta(t/h)}{\phi_\varepsilon(t/h)} dt. \tag{8}$$

Therefore, when (6) is no longer valid, we propose the following estimator for  $g(x)$

$$\tilde{g}(x) = \frac{1/nh \sum_{i=1}^n K_\varepsilon^\delta\left(\frac{x-V_i}{h}\right) Y_i}{1/nh \sum_{i=1}^n K_\varepsilon^\delta\left(\frac{x-V_i}{h}\right)}. \tag{9}$$

**Remark 1** To ensure that the proposed estimator (9) is well-behaved, we need to make the following assumption.

**Condition A:**

1.  $\phi_\varepsilon(t) \neq 0$  for all  $t$ ; and
2.  $\sup_t |\phi_K(t) \phi_\delta(t/h) / \phi_\varepsilon(t/h)| < \infty$  and  $\int |\phi_K(t) \phi_\delta(t/h) / \phi_\varepsilon(t/h)| dt < \infty$ .

**Example 2** We use the same model as in Example 1 with  $\sigma_\delta^2 < \sigma_\varepsilon^2$ . In this case, to ensure (A2) to be valid, it is rather common to choose kernels that have a compactly supported characteristic function  $\phi_K(t)$ . For example, we choose the sinc kernel  $K(x) = \sin x / (\pi x)$ , which has characteristic function  $\phi_K(t) = 1_{[-1,1]}(t)$ , the indicator function of the interval  $[-1,1]$ . From (8), we have

$$K_\varepsilon^\delta(x) = \frac{1}{\pi} \int_0^1 \cos(tx) \exp\left\{ \frac{(\sigma_\varepsilon^2 - \sigma_\delta^2)t^2}{2h^2} \right\} dt.$$

**Remark 2**

1. The above two nonparametric estimators of  $f_X(x)$  were given by [22];
2. When the variance of  $\varepsilon$  in Models (4) is equal to 0, which is the Berkson error model, the estimator (6) becomes

$$\hat{g}(x) = \frac{1/n \sum_{i=1}^n f_\delta(x-V_i) Y_i}{1/n \sum_{i=1}^n f_\delta(x-V_i)}, \tag{10}$$

where  $f_\delta(x)$  is the density function of  $\delta$ ; and;

3. When the variance of  $\delta$  in Models (4) is equal to 0, which is the classical error model,  $\tilde{g}(x)$  given in (9) reduces to the estimator of [2].

### 3. Theoretical Properties

In this section, we study asymptotic properties of the estimators proposed in Section 2. In particular, the properties of the estimator  $\hat{g}(x)$  at (6) are clear. It is easy to check that the numerator and the denominator are both unbiased estimators of  $g(x)f_X(x)$  and  $f_X(x)$ , respectively and that,  $\hat{g}(x)$  converges at the fast parametric  $\sqrt{n}$  rate. Properties of the estimator  $\tilde{g}(x)$  at (9) need to further explore and, in what follows, we derive them.

### 3.1. Asymptotic Results for $\tilde{g}$

In this section, we investigate the large-sample properties of the estimator  $\tilde{g}(x)$  at (9). For this purpose, we present the following regular conditions which are mild and can be found in [2].

**Condition B:**

1.  $e_1, \dots, e_n$  have zero means and uniformly bounded variances;
2.  $f_W, f_X$  and  $f_\varepsilon$  are bounded, and  $f_X$  and  $g$  have bounded  $k$ th derivatives;
3.  $K(x)$  is a real and symmetric kernel and has finite moment of order  $k$ . Namely,  $\mu_{K,j} = \int u^j K(u) du = 1_{\{j=0\}}$  for  $j = 1, \dots, k-1$  and  $\mu_{K,k} \neq 0$ ; and
4. The conditional moment  $E(Y^{2+\eta} | W = u)$  is bounded for all  $u$  and some  $\eta > 0$ .

Let  $K_h(x) = h^{-1}K(x/h)$ . The mean squared error (MSE) of the estimator  $\tilde{g}(x)$  is described in the next Theorem.

**Theorem 1 (MSCE)** Suppose that Conditions A and B hold. Then, for each  $x$  such that  $f_X(x) > 0$ ,

$$\begin{aligned}
 & E \left| \left\{ \tilde{g}(x) - g(x) \right\} \frac{\tilde{f}_X(x)}{f_X(x)} \right|^2 \\
 &= \left[ \frac{1}{f_X(x)} \int K_h(x-v) \{g(v) - g(x)\} f_X(v) dv \right]^2 \\
 &+ \frac{1}{nh^2 f_X^2(x)} \int \{K_\varepsilon^\delta(u)\}^2 (\tau^2 f_W) * f_\varepsilon(x-uh) du + o\left(\frac{1}{n}\right),
 \end{aligned} \tag{11}$$

where  $\tau^2(\cdot) = E\left[\{Y - g(x)\}^2 | W = \cdot\right]$ .

Explicit rates of convergence of the estimator  $\tilde{g}(x)$  can be found by examination of the asymptotic behaviour of the MSE. For the bias, using the Taylor expansion of the first term on the right-hand side of Equation (11), we have

$$\text{Bias}^2(\tilde{g}(x)) = b_k^2(x) h^{2k} + o(h^{2k}),$$

where  $b_k(x) = (-1)^k f_X^{-1}(x) \mu_{K,k} \left[ \{g(x) f_X(x)\}^{(k)} / (k!) - g(x) f_X^{(k)}(x) / (k!) \right]$ .

The second term on the right-hand side of Equation (11) describes the variance of  $\tilde{g}(x)$ . The asymptotic behaviour of this term is more difficult to evaluate since it depends on the tail behaviour of the ratio  $|\phi_\delta(t)| / |\phi_\varepsilon(t)|$ , as [14] discussed, which can be classified into the following:

1. An exponential ratio of order  $\beta$  is

$$d_0 |t|^{\beta_0} \exp\left(|t|^\beta / \gamma\right) \leq |\phi_\delta(t)| / |\phi_\varepsilon(t)| \leq d_1 |t|^{\beta_1} \exp\left(|t|^\beta / \gamma\right), \text{ as } |t| \rightarrow \infty, \tag{12}$$

with  $d_0, d_1, \gamma, \beta > 0$  and  $\beta_0, \beta_1 \in \mathbb{R}$ .

2. A polynomial ratio of order  $\beta$  is

$$d_0 |t|^{-\beta} \leq |\phi_\delta(t)| / |\phi_\varepsilon(t)| \leq d_1 |t|^{-\beta} \text{ as } |t| \rightarrow \infty, \tag{13}$$

with  $d_0, d_1 > 0$  and  $\beta \leq 1$ .

### 3.1.1. Asymptotic Mean Squared Error (AMSE)

In this section, we study the asymptotic behaviour of the MSE where  $|\phi_\delta(t)|/|\phi_\varepsilon(t)|$  behaves like an exponential or a polynomial.

**Theorem 2** *Suppose that Conditions A and B hold and that the first half inequality of (12) is satisfied. Assume that  $\phi_\kappa(t)$  is supported on  $[-1, 1]$ . Then, for each  $x$  such that  $f_x(x) > 0$ , we have*

$$\text{AMSE}(\tilde{g}(x)) = b_k^2(x)h^{2k} + \kappa_1 n^{-1} h^{-2\beta_2 - 1} \exp\left(2|1/h|^\beta / \gamma\right),$$

with  $\kappa_1$  being some positive constant and  $\beta_2 = \beta_1 1_{\{\beta_1 > \frac{1}{2}\}}$ .

When  $|\phi_\varepsilon(t)|$  is exponentially smoother than  $|\phi_\delta(t)|$ , we obtain a slower logarithmic rate which is similar to the deconvolution rate for supersmooth error given in [2]. More precisely, the optimal bandwidth is of order  $h = d(2/\gamma)^{1/\beta} (\ln n)^{-1/\beta}$  with  $d \geq 1$ , and the estimator  $\tilde{g}(x)$  then converges at the rate of  $(\log n)^{k/\beta}$ .

**Theorem 3** *Suppose Conditions A and B hold, and that  $\int |t|^{-2\beta} |\phi_\kappa(t)|^2 dt < \infty$ . Then, under the polynomial ratio (13), for each  $x$  such that  $f_x(x) > 0$ , we have*

$$\text{AMSE}(\tilde{g}(x)) = b_k^2(x)h^{2k} + \kappa_2 \lambda(h)/n,$$

with  $\kappa_2$  being some positive constant, and  $\lambda(h) = h^{2\beta-1} 1_{\{\beta < 1/2\}} + 1_{\{1/2 \leq \beta \leq 1\}}$ .

We obtain that, when  $|\phi_\delta(t)|/|\phi_\varepsilon(t)|$  behaves like a polynomial ratio of order  $\beta$  in the tail, the convergence rates range from  $\sqrt{n}$  to deconvolution rate of ordinary smooth error of [2]. More precisely, the optimal bandwidth is of order  $h \sim \text{const} \cdot n^{-1/(2k)}$  when  $1/2 \leq \beta \leq 1$ , and the estimator  $\tilde{g}(x)$  then converges at the rate of  $\sqrt{n}$ . When  $\beta < 1/2$ , the optimal bandwidth is of order  $h \sim \text{const} \cdot n^{-1/(2k-2\beta+1)}$  and the estimator  $\tilde{g}(x)$  converges at the rate of  $n^{k/(2k-2\beta+1)}$ .

### 3.1.2. Asymptotic Normality

The theorem below establishes asymptotic normality in the exponential ratio case.

**Theorem 4** *Under the conditions of Theorem (2), and for bandwidth  $h = d(2/\gamma)^{1/\beta} (\ln n)^{-1/\beta}$  with  $d \geq 1$ ,*

$$\frac{\tilde{g}(x) - g(x) - \text{Bias}(\tilde{g}(x))}{\sqrt{\text{var}(\tilde{g}(x))}} \xrightarrow{d} N(0, 1),$$

where  $\text{Bias}(\tilde{g}(x)) = b_k(x)h^k + o(h^k)$  and  $\text{var}(\tilde{g}(x)) = o\left([\text{Bias}(\tilde{g}(x))]^2\right)$ .

The next theorem establishes asymptotic normality in the polynomial ratio case.

**Theorem 5** *Suppose that Conditions A and B hold and that the inequality of (13) is satisfied. Assume that  $\int |t|^{-\beta} |\phi_\kappa(t)| dt < \infty$  and  $\int |t|^{-2\beta} |\phi_\kappa(t)|^2 dt < \infty$ . Then, under  $\lambda(h)/n = O(h^{2k})$  and  $h \rightarrow 0$  as  $n \rightarrow \infty$ , for each  $x$  such that  $f_x(x) > 0$ , we have*

$$\frac{\tilde{g}(x) - g(x) - \text{Bias}(\tilde{g}(x))}{\sqrt{\text{var}(\tilde{g}(x))}} \xrightarrow{d} N(0, 1),$$

where  $\text{Bias}(\tilde{g}(x))$  is the same as given in Theorem (4) and  $\text{var}(\tilde{g}(x))$  is equal to

the second term on the right-hand side of Equation (11).

The proofs of all theorems are postponed to the **Appendix**.

### 3.2. Unknown Measurement Error Distribution

When the error densities are unknown, they can be readily estimated from additional observations (e.g., a sample from the error densities, replicated data or external data) and these estimates can be substituted into (6) and (9) to produce the estimate of  $g(x)$ . For sufficiently large sample size, the rates of convergence of the estimates remain unchanged when  $\phi_\delta$  and  $\phi_\varepsilon$  are replaced by their consistent estimators (e.g., [4] [17] [24]).

## 4. Simulation Studies

We study numerical properties of the estimators proposed in Section 2. Note that we have defined two estimators, at (6) and (9). The first exists when  $\phi_\delta(t)/\phi_\varepsilon(t)$  is integrable, and the estimator (9) otherwise. We use the notations  $\hat{g}$  and  $\tilde{g}$  for the estimators (6) and (9) respectively. We use the notation  $\hat{g}_I$  for the estimator that ignores the errors, that is, the estimator is the classical Nadaraya-Watson estimator of  $g$  based on direct data from  $(Y_i, V_i)$ ,  $i = 1, \dots, n$ . Note that  $\hat{g}_I$  is exactly equal to  $\tilde{g}$  when  $\phi_\delta(t) = \phi_\varepsilon(t)$ . In addition, we use  $\hat{g}_C$  for the estimator of [23].

We apply the various estimators introduced above to some simulated examples (see, [23]):

1.  $g(x) = 5 \sin(2x) \exp(-16x^2/50)$  (sinusoidal),
2.  $g(x) = (50x^2 + 10x + 25)^{-1}$  (sharp unimodal), and
3.  $g(x) = \phi_{0,1.5}(4x) + \phi_{1,2}(4x) + \phi_{2,5}(4x)$  (asymmetric);

where  $\phi_{\mu,\sigma}$  is the density of an  $\text{Normal}(\mu, \sigma^2)$  variable. For each of the above regression functions, we generate 200 data sets of  $n$  randomly sampled vectors  $(V_1, Y_1), \dots, (V_n, Y_n)$ , as follows. We generate a random sample  $\delta_1, \dots, \delta_n$  from  $f_\delta$ , a random sample  $\varepsilon_1, \dots, \varepsilon_n$  from  $f_\varepsilon$  and a random sample  $W_1, \dots, W_n$  from  $f_W$ , and put  $X_i = W_i + \delta_i$  and  $V_i = W_i + \varepsilon_i$ ,  $i = 1, \dots, n$ , where  $f_W$  is the density of an  $\text{Normal}(0, \sigma_W^2)$  variable, and we take  $f_\delta$  and  $f_\varepsilon$  to be either normal or Laplace with zero mean. Then we generate a random sample  $Y_1, \dots, Y_n$  as  $Y_i = g(X_i) + e_i$ , where the errors  $e_i$  are normally distributed with zero mean and variance  $\sigma_e^2$ , where  $\sigma_e^2 = 0.1 \times \sigma^2(g)$  with  $\sigma^2(g)$  denoting the mean-squared deviation of  $g$  from its average value. We simply denoted  $\delta \sim \text{Normal}$  and  $\varepsilon \sim \text{Laplace}$  by  $(\delta, \varepsilon) \sim (\text{N}, \text{L})$ , and other similar.

In our simulations we consider sample sizes  $n = 50, 100, 250$ , and in each case we generate 200 samples from the distribution of the random vector  $(V, Y)$ . Except if stated otherwise, we adopt the second order kernel  $K$  corresponding to

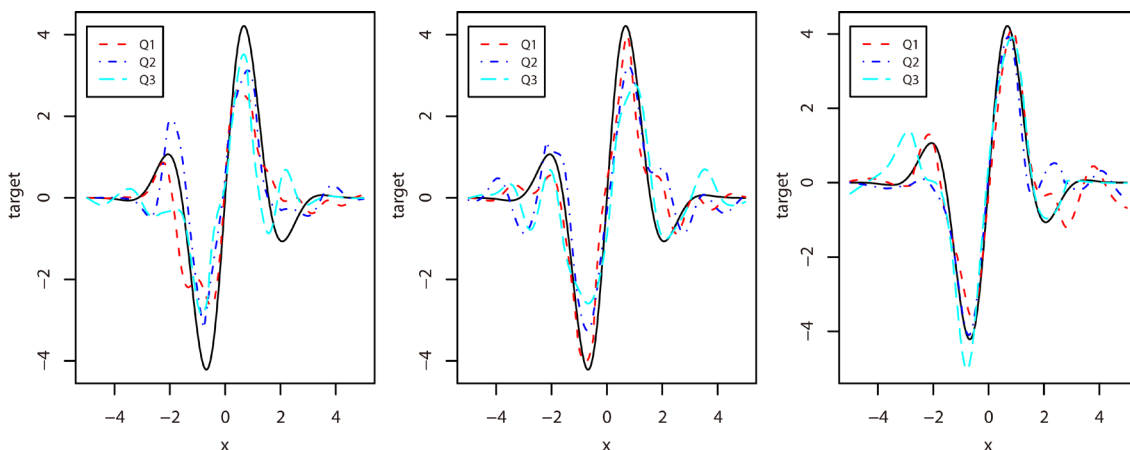
$\phi_K(t) = (1-t^2)^3 \cdot 1_{[-1,1]}(t)$ , which is necessary to calculate  $\tilde{g}$  and  $\hat{g}_C$ . For the bandwidth  $h$ , it is necessary to calculate  $\tilde{g}$ ,  $\hat{g}_I$  and  $\hat{g}_C$ , we select the value  $h$  that minimises the cross-validation (CV) criterion,  $CV = \sum_{j=1}^n \{Y_j - g_{-j}^{\text{est}}(V_j)\}^2$ , where the subscript  $-j$  meant that the estimator was constructed without using the  $j$ th observation.



We report the Integrated Squared Error,  $ISE = \int \{g^{est}(x) - g(x)\}^2 dx$ , where  $g^{est}(x)$  is the estimator considered. In all graphs, to illustrate the performance of an estimator, we show the estimated curves corresponding to the first (Q1), second (Q2) and third (Q3) quartiles of the ordered ISEs. The target curve is always represented by a solid curve. In the tables we provide the average values, denoted by MISE, of the 200 calculated ISEs.

**Figure 1** and **Table 1** illustrate the way in which the estimator improves as sample size increases. We compare, for various sample sizes, the results obtained for estimating curve (a) when  $W \sim \text{Normal}(0,2)$ , and  $(\delta, \varepsilon) \sim (N, L)$  with the pair of variance ratios  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$  equals (0.1, 0.4), and for estimating curve (b) when  $W \sim \text{Normal}(0,2)$  and  $(\delta, \varepsilon) \sim (N, L), (N, N), (L, L)$  or  $(L, N)$  with  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2) = (0.1, 0.4)$ . We see clearly that, as the sample size increases, the quality of the estimators improves significantly in all cases.

For any nonparametric method for regression problem, the quality of the estimator also depends on the discrepancy of the observed sample. That is, for any given family of densities  $f_w, f_\delta$  and  $f_\varepsilon$ , and any given the noise-to-signal ratios  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$ , the performance of the estimator depends on the variances of  $W, \delta$  and  $\varepsilon$ . Here, we compare the results obtained from estimating curve (c) for different values of



**Figure 1.** Estimation of curve (a) for samples of size  $n = 50$  (left panel),  $n = 100$  (middle panel) or  $n = 250$  (right panel), when  $W \sim \text{Normal}(0,2)$ , and  $(\delta, \varepsilon) \sim (N, L)$  with  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2) = (0.1, 0.4)$ . The solid curve is the target curve.

**Table 1.** MISE ( $\times 10^{-4}$ ) for estimation of curve (b) when  $W \sim \text{Normal}(0,2)$ , and  $(\delta, \varepsilon) \sim (N, L), (N, N), (L, L)$  or  $(L, N)$  with  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2) = (0.1, 0.4)$ .

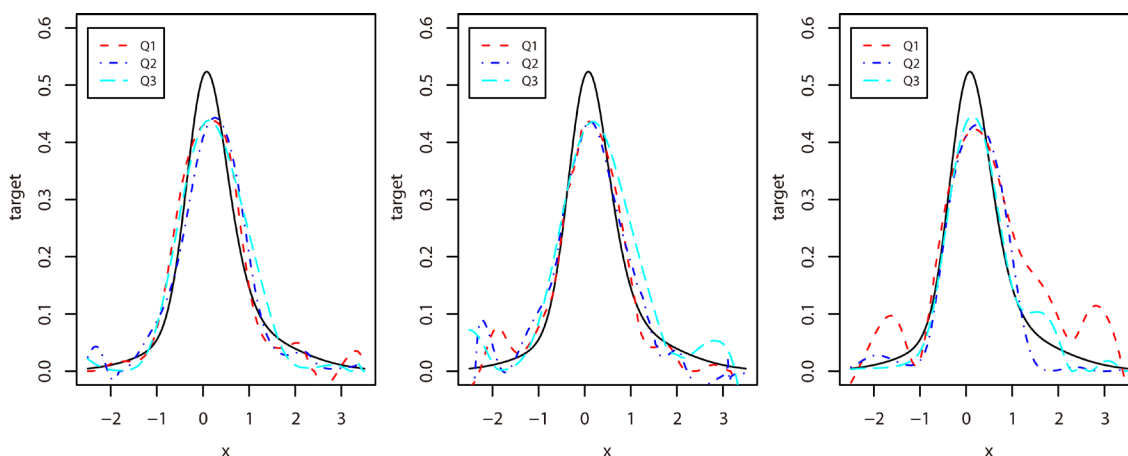
$n$	$(\delta, \varepsilon) \sim$ Method	(N, L) MISE	$(\delta, \varepsilon) \sim$ Method	(N, N) MISE	(L, L) MISE	(L, N) MISE
50	$\hat{g}$ at (6)	5.3524	$\tilde{g}$ at (9)	8.3704	21.7584	9.9570
100		3.2803		6.8685	11.1636	6.7162
250		2.7013		5.4176	6.8579	4.9409

$(\sigma_w^2, \sigma_\delta^2, \sigma_\varepsilon^2)$ . As expected, **Figure 2** shows that the best performance usually occur for smaller error variance (e.g.,  $(\sigma_w^2, \sigma_\delta^2, \sigma_\varepsilon^2) = (0.5, 0.05, 0.15)$ ). It is noteworthy that the effect of the variances on the estimator performance is obvious in model (4).

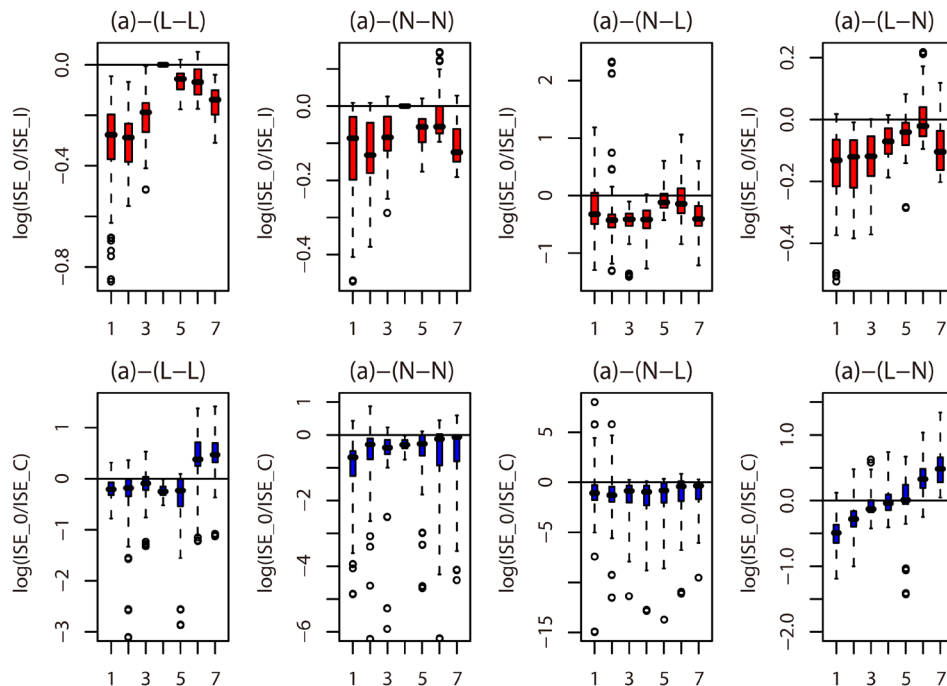
Finally, we compare  $\hat{g}$  (or  $\tilde{g}$ ),  $\hat{g}_I$  and  $\hat{g}_C$ . **Figure 3** shows the boxplots of the quantities of  $\log(\text{ISE}_o/\text{ISE}_I)$  and  $\log(\text{ISE}_o/\text{ISE}_C)$  for estimating curve (a) when  $n = 250$  and  $W \sim \text{Normal}(0, 2)$ , where  $\text{ISE}_o$  is the ISE of our proposed estimator,  $\text{ISE}_I$  is the ISE of the estimator that ignores the errors, and  $\text{ISE}_C$  is the ISE of the estimator of [23]. Here, each boxplot is constructed from 200 samples. Here, in panel (a)-(L-L) (or (a)-(N-N)), the mixed errors are both Laplace (or both normal). Here, and also in panel (a)-(N-L) (or (a)-(L-N)), the errors are  $\delta \sim \text{Normal}(0, \sigma_\delta^2)$  and  $\varepsilon \sim \text{Laplace}$  (or  $\delta \sim \text{Laplace}$  and  $\varepsilon \sim \text{Normal}(0, \sigma_\varepsilon^2)$ ). In each panel, for X-axis = 1 to 7,  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2) = (0.1, 0.4), (0.1, 0.3), (0.2, 0.3), (0.2, 0.2), (0.3, 0.2), (0.3, 0.1)$  or  $(0.4, 0.1)$ . The more a boxplot is located below the zero horizontal line, the better our method compared with the other two estimators. In the same situation, **Table 2** and **Table 3** report the average integrated square error (MISE) for estimating curves (b) and (c) respectively. As expected, our proposed estimator substantially outperformed the estimator that completely ignores any measurement errors. Our results show that our proposed estimator usually works better than the estimator proposed by [23] for estimating curves (a) and (b). It is noteworthy that the estimator proposed by [23] may perform better than our proposed estimator when curve (c) with  $\sigma_\delta^2 > \sigma_\varepsilon^2$  is estimated.

### 5. Discussion

In this paper, we propose a new method for estimating non-parametric regression models with the predictors being measured with a mixture of Berkson and classical errors. The method is based on the relative smoothness of  $\phi_\delta$  and  $\phi_\varepsilon$ . When  $\phi_\delta$  is



**Figure 2.** Estimation of function (c) for samples of size  $n = 250$ , when  $W \sim \text{Normal}(0, \sigma_w^2)$ ,  $\delta \sim \text{Laplace}$  and  $\varepsilon \sim \text{Normal}(0, \sigma_\varepsilon^2)$  with  $(\sigma_w^2, \sigma_\delta^2, \sigma_\varepsilon^2)$  being  $(0.5, 0.05, 0.15)$ ,  $(1.0, 0.1, 0.3)$ , and  $(2.0, 0.15, 0.45)$  (from left to right). The solid curve is the target curve.



**Figure 3.** Boxplots of the quantities of  $\log(\text{ISE}_0/\text{ISE}_I)$  (row 1) and  $\log(\text{ISE}_0/\text{ISE}_C)$  (row 2) for estimating regression curve (a) when  $W \sim \text{Normal}(0,2)$  and  $n=250$ , for various error densities  $f_\delta$  and  $f_\varepsilon$  and various values of  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$ .

**Table 2.**  $\text{MISE} (\times 10^{-4})$  for estimation of curve (b) when  $W \sim \text{Normal}(0,2)$  and  $n=250$ , for various error densities  $f_\delta$  and  $f_\varepsilon$  and various values of  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$ .

		$(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$						
$(\delta, \varepsilon)$	Method	(0.1,0.4)	(0.1,0.3)	(0.2,0.3)	(0.2,0.2)	(0.3,0.2)	(0.3,0.1)	(0.4,0.1)
		MISE	MISE	MISE	MISE	MISE	MISE	MISE
(N, L)	$\hat{g}$	2.7013	3.2803	3.2877	3.0648	3.0751	3.1708	3.2467
	$\hat{g}_I$	4.7107	4.3962	4.2197	4.0074	3.9953	4.2278	4.1772
	$\hat{g}_C$	4.2953	3.8815	3.5265	3.4723	3.2630	3.1153	2.8465
(N, N)	$\hat{g}^*$	5.4176	3.8075	4.0953	3.8031	3.8860	4.2107	5.1018
	$\hat{g}_I$	5.7508	4.4523	4.3278	3.8031	4.5206	4.6225	5.5277
	$\hat{g}_C$	5.8240	4.1611	4.2753	3.5777	4.0363	4.3566	4.2559
(L, L)	$\hat{g}$	6.8579	5.6354	4.2114	3.3682	4.3915	3.9042	4.3463
	$\hat{g}_I$	8.2793	5.5021	4.2403	3.3682	4.2050	4.2479	4.7129
	$\hat{g}_C$	7.7004	7.8699	5.8145	3.3493	4.3965	3.2047	3.8581
(L, N)	$\hat{g}$	4.9409	4.3785	4.3101	3.6858	3.7947	4.5531	4.1757
	$\hat{g}_I$	5.3184	4.8508	5.3981	4.6511	4.3452	4.7562	4.8375
	$\hat{g}_C$	5.0408	4.4118	4.5309	3.9896	3.5704	3.3006	3.3726

**Table 3.** MISE for estimation of curve (c) when  $W \sim \text{Normal}(0, 2)$  and  $n = 250$ , for various error densities  $f_\delta$  and  $f_\varepsilon$  and various values of  $(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$ .

$(\delta, \varepsilon)$	Method	$(\sigma_\delta^2/\sigma_w^2, \sigma_\varepsilon^2/\sigma_w^2)$						
		(0.1,0.4) MISE	(0.1,0.3) MISE	(0.2,0.3) MISE	(0.2,0.2) MISE	(0.3,0.2) MISE	(0.3,0.1) MISE	(0.4,0.1) MISE
(N, L)	$\hat{g}$	0.04895	0.04547	0.05037	0.05615	0.07006	0.06539	0.07410
	$\hat{g}_t$	0.07716	0.06840	0.06457	0.06395	0.07383	0.07897	0.07455
	$\hat{g}_c$	0.05446	0.05133	0.05042	0.05125	0.06842	0.05885	0.07185
(N, N)	$\hat{g}^*$	0.06894	0.06061	0.08074	0.06868	0.07698	0.07855	0.08983
	$\hat{g}_t$	0.09306	0.07728	0.09156	0.06868	0.08166	0.08486	0.09174
	$\hat{g}_c$	0.07558	0.06368	0.08162	0.06442	0.07558	0.05729	0.08035
(L, L)	$\tilde{g}$	0.05102	0.04070	0.05352	0.05654	0.06965	0.06364	0.07761
	$\hat{g}_t$	0.07427	0.06039	0.06891	0.05654	0.06962	0.07184	0.08422
	$\hat{g}_c$	0.05678	0.05349	0.05355	0.05094	0.06400	0.04008	0.04855
(L, N)	$\tilde{g}$	0.07343	0.05983	0.07332	0.06923	0.07571	0.05997	0.06183
	$\hat{g}_t$	0.09334	0.07516	0.08357	0.07148	0.07932	0.07314	0.08148
	$\hat{g}_c$	0.07820	0.06183	0.07485	0.05864	0.06491	0.04676	0.05368

smooth enough (relative to  $\phi_\varepsilon$ ), we propose a nonparametric estimator (6) that converges to the target curve at the parametric  $\sqrt{n}$  rate. For less smooth function  $\phi_\delta$ , we propose a kernel estimator (9) that converges at rates ranging from  $\sqrt{n}$  to rates that are close to the deconvolution rates. Numerical results show that the new estimators are promising in terms of correcting the bias arising from the errors-in-variables. It generally performs better than the approach proposed by [23]. The methodology can be readily extended to the prediction problem of nonparametric errors-in-variables regression (see, e.g., [16]). Extension of our method to the problems considered in [5] is of future research interest.

### Acknowledgements

This work was supported by Natural Science Foundation of Jiangxi Province of China under grant number 20142BAB211018.

### References

- [1] Armstrong, B.G. (1998) Effect of Measurement Error on Epidemiological Studies of Environmental and Occupational Exposures. *Occupational and Environmental Medicine*, **55**, 651-656. <https://doi.org/10.1136/oem.55.10.651>
- [2] Fan, J. and Truong, Y.K. (1993) Nonparametric Regression with Errors in Variables. *Annals of Statistics*, **21**, 1900-1925. <https://doi.org/10.1214/aos/1176349402>
- [3] Delaigle, A. and Meister, A. (2007) Nonparametric Regression Estimation in the Heteroscedastic Errors-in-Variables Problem. *Journal of the American Statistical Association*, **102**, 1416-1426. <https://doi.org/10.1198/016214507000000987>

- [4] Delaigle A., Hall, P. and Meister A. (2008) On Deconvolution with Repeated Measurements. *Annals of Statistics*, **36**, 665-685. <https://doi.org/10.1214/009053607000000884>
- [5] Delaigle, A., Fan, J. and Carroll, R.J. (2009) A Design-Adaptive Local Polynomial Estimator for the Errors-in-Variables Problem. *Journal of the American Statistical Association* **104**, 348-359. <https://doi.org/10.1198/jasa.2009.0114>
- [6] Cook, J.R. and Stefanski, L.A. (1994) Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association*, **89**, 1314-1328. <https://doi.org/10.1080/01621459.1994.10476871>
- [7] Stefanski, L.A. and Cook, J.R. (1995) Simulation-Extrapolation: The Measurement Error Jackknife. *Journal of the American Statistical Association*, **90**, 1247-1256. <https://doi.org/10.1080/01621459.1995.10476629>
- [8] Carroll, R.J., Maca, J.D. and Ruppert, D. (1999) Nonparametric Regression in the Presence of Measurement Error. *Biometrika*, **86**, 541-554. <https://doi.org/10.1093/biomet/86.3.541>
- [9] Staudenmayer, J. and Ruppert, D. (2004) Local Polynomial Regression and Simulation-Extrapolation. *Journal of the Royal Statistical Society: Series B*, **66**, 17-30. <https://doi.org/10.1046/j.1369-7412.2003.05282.x>
- [10] Berry, S.M., Carroll, R.J. and Ruppert, D. (2002) Bayesian Smoothing and Regression Splines for Measurement Error Problems. *Journal of the American Statistical Association*, **97**, 160-169. <https://doi.org/10.1198/016214502753479301>
- [11] Berkson, J. (1950) Are There Two Regression Problems? *Journal of the American Statistical Association*, **45**, 164-180. <https://doi.org/10.1080/01621459.1950.10483349>
- [12] Fuller, W. (1987) Measurement Error Models. Wiley, New York. <https://doi.org/10.1002/9780470316665>
- [13] Huwang, L. and Huang, H.Y.S. (2000) On Errors-in-Variables in Polynomial Regression—Berkson Case. *Statistica Sinica*, **10**, 923-936.
- [14] Wang, L. (2003) Estimation of Nonlinear Berkson-Type Measurement Error Models. *Statistica Sinica*, **13**, 1201-1210.
- [15] Wang, L. (2004) Estimation of Nonlinear Models with Berkson Measurement Errors. *Annals of Statistics*, **32**, 2559-2579. <https://doi.org/10.1214/009053604000000670>
- [16] Carroll, R.J., Delaigle, A. and Hall, P. (2009) Nonparametric Prediction in Measurement Error Models. *Journal of the American Statistical Association*, **104**, 993-1003. <https://doi.org/10.1198/jasa.2009.tm07543>
- [17] Delaigle, A., Hall, P. and Qiu, P. (2006) Nonparametric Methods for Solving the Berkson Errors-in-Variables Problem. *Journal of the Royal Statistical Society: Series B*, **68**, 201-220. <https://doi.org/10.1111/j.1467-9868.2006.00540.x>
- [18] Schafer, D.W. and Gilbert, E.S. (2006) Some Statistical Implications of Dose Uncertainty in Radiation Dose-Response Analyses. *Radiation Research*, **166**, 303-312. <https://doi.org/10.1667/RR3358.1>
- [19] Reeves, G.K., Cox, D.R., Darby, S.C. and Whitley, E. (1998) Some Aspects of Measurement Error in Explanatory Variables for Continuous and Binary Regression Models. *Statistics in Medicine*, **17**, 2157-2177. [https://doi.org/10.1002/\(SICI\)1097-0258\(19981015\)17:19<2157::AID-SIM916>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-0258(19981015)17:19<2157::AID-SIM916>3.0.CO;2-F)
- [20] Schafer, M., Mullhaupt, B. and Clavien, P.A. (2002) Evidence-Based Pancreatic Head Resection for Pancreatic Cancer and Chronic Pancreatitis. *Annals of Surgery*, **236**, 137-148. <https://doi.org/10.1097/00000658-200208000-00001>
- [21] Mallick, B., Hoffman, F.O. and Carroll, R.J. (2002) Semiparametric Regression Modeling

- with Mixtures of Berkson and Classical Error, with Application to Fallout from the Nevada Test Site. *Biometrics*, **58**, 13-20. <https://doi.org/10.1111/j.0006-341X.2002.00013.x>
- [22] Delaigle, A. (2007) Nonparametric Density Estimation from Data with a Mixture of Berkson and Classical Errors. *Canadian Journal of Statistics*, **35**, 89-104. <https://doi.org/10.1002/cjs.5550350109>
- [23] Carroll, R.J., Delaigle, A. and Hall, P. (2007) Nonparametric Regression Estimation from Data Contaminated by a Mixture of Berkson and Classical Errors. *Journal of the Royal Statistical Society: Series B*, **69**, 859-878. <https://doi.org/10.1111/j.1467-9868.2007.00614.x>
- [24] Hu, Y. and Schennach, S.M. (2008) Identification and Estimation of Nonclassical Nonlinear Errors-in-Variables Models with Continuous Distributions. *Econometrica*, **76**, 195-216. <https://doi.org/10.1111/j.0012-9682.2008.00823.x>
- [25] Fan, J. (1991) Asymptotic Normality for Deconvolution Kernel Density Estimators. *Sankhya A*, **53**, 97-110.

### Appendix

#### Proof of Theorem 1

Let  $A_n(x) = \{\tilde{g}(x) - g(x)\}$ ,  $\tilde{f}_X(x) = n^{-1} \sum_{j=1}^n U_{n,j}$ , where

$$U_{n,j} = h^{-1} K_\varepsilon^\delta \left( \frac{x - V_j}{h} \right) \{Y_j - g(x)\}, \text{ we have}$$

$$\begin{aligned} E(U_{n,j}) &= h^{-1} E \left[ E \left[ \{Y_j - g(x)\} | W \right] E \left[ K_\varepsilon^\delta \left( \frac{x - V_j}{h} \right) | W \right] \right] \\ &= h^{-1} \int K \left( \frac{x - v}{h} \right) \{g(v) - g(x)\} f_X(v) dv, \end{aligned} \tag{14}$$

and

$$\begin{aligned} E(U_{n,j}^2) &= h^{-2} \int E \left[ \{Y_j - g(x)\}^2 | W = u \right] E \left[ \left\{ K_\varepsilon^\delta \left( \frac{x - V_j}{h} \right) \right\}^2 | W = u \right] f_W(u) du \\ &= h^{-1} \int \{K_\varepsilon^\delta(u)\}^2 (\tau^2 \cdot f_W) * f_\varepsilon(x - hu) du \end{aligned} \tag{15}$$

where  $\tau^2(\cdot) = E \left[ \{Y - g(x)\}^2 | W = \cdot \right]$ . The result follows immediately from 14 and 15.

#### Proofs of the Results of Section 3.1.1.

**Lemma 1** Suppose that  $\phi_K(t)$  is supported on  $[-1,1]$ , and  $\phi_\varepsilon \neq 0$  for all  $t$ . Then, for  $\beta_2 = \beta_1 1_{\{\beta_1 > -\frac{1}{2}\}}$ , we have

$$\|K_\varepsilon^\delta(x)\|_\infty \leq Ch^{-\beta_2} \exp(h^{-\beta}/\gamma) \text{ and } \int \{K_\varepsilon^\delta(x)\}^2 dx \leq Ch^{-2\beta_2} \exp(h^{-2\beta}/\gamma),$$

where, here, and below,  $C$  denotes a generic positive and finite constant.

*Proof.* It follows from (A2) of Condition A that  $d_\varepsilon \equiv \min_{|t| \leq M} |\phi_\varepsilon(t)| > 0$  for some large enough constant  $M > 0$ . Since  $|\phi_\delta(t)| \leq 1$ , we have

$$\begin{aligned} |K_\varepsilon^\delta(x)| &\leq (2\pi)^{-1} \int_{-\infty}^{+\infty} |\phi_K(t)| |\phi_\delta(t/h)/\phi_\varepsilon(t/h)| dt \\ &\leq d_\varepsilon^{-1} \int_{|t| \leq Mh} |\phi_K(t)| dt + (2\pi)^{-1} \int_{Mh < |t| \leq 1} |\phi_K(t)| d_1 |t/h|^{\beta_1} \exp(|t/h|^\beta/\gamma) dt \\ &\leq C \int_{|t| \leq Mh} |\phi_K(t)| dt + Ch^{-\beta_1} \exp(h^{-\beta}/\gamma) \int_{Mh < |t| \leq 1} |t|^{\beta_1} dt \end{aligned}$$

The conclusion follows from

$$\int_{Mh < |t| \leq 1} |t|^{\beta_1} dt = \begin{cases} C [t^{\beta_1+1}]_{Mh}^1 = O(1) & \beta_1 > -1, \\ C [\ln t]_{Mh}^1 = O(\ln h) & \beta_1 = -1, \\ C [t^{\beta_1+1}]_{Mh}^1 = O(h^{\beta_1+1}) & \beta_1 < -1. \end{cases}$$

The proof for the other result is similar and requires Parseval's Theorem.

From (14) and Lemma 1, we have

$$E(U_{n,j}^2) = O\left(h^{-2\beta_2-1} \exp\left(2|1/h|^\beta/\gamma\right)\right).$$

The proof of Theorem 2 follows from the expressions of  $E(U_{n,j})$  and  $E(U_{n,j}^2)$ .

The proof of Theorem 3 is the same as the proof of Theorem 2, but in this case we need the following lemma.

**Lemma 2** Suppose that  $\phi_\varepsilon \neq 0$  for all  $t$ ,  $\int |t|^{-\beta} |\phi_K(t)| dt < \infty$  and  $\int |t|^{-2\beta} |\phi_K(t)|^2 dt < \infty$ . Then, we have

$$\|K_\varepsilon^\delta(x)\|_\infty \leq Ch^\beta \text{ and } \int \{K_\varepsilon^\delta(x)\}^2 dx \leq Ch\lambda(h)$$

with  $\lambda(h) = h^{2\beta-1} 1_{\{\beta < 1/2\}} + 1_{\{1/2 \leq \beta \leq 1\}}$ .

The proof of Lemma 2 is similar to the proof of Lemma 1 and is omitted.

### Proofs of the Results of Section 3.1.2.

A standard decomposition gives  $\tilde{g}(x) - g(x) = \tilde{f}_X^{-1}(x)A_n(x)$ ,  $\tilde{f}_X(x)$  goes in probability to  $f_X(x)$  and thus we only need to prove the asymptotic normality for  $A_n(x)$ . As given in [25], a sufficient condition for the following asymptotical normality

$$\frac{\sum_{j=1}^n U_{n,j} - nE(U_{n,j})}{\sqrt{n\text{var}(U_{n,j})}} \rightarrow N(0,1)$$

is that the Lyapounov's condition holds, i.e., for some  $\eta > 0$ ,

$$\frac{E|U_{n,j}|^{2+\eta}}{n^{\eta/2} \{E(U_{n,j})^2\}^{1+\eta/2}} = 0.$$

Letting  $J(u) = E\{|Y + g(x)|^{2+\eta} | W = u\}$ , we have

$$\begin{aligned} E|U_{n,j}|^{2+\eta} &= \iint J(u) |h^{-1}K_\varepsilon^\delta\{(x-u-v)/h\}|^{2+\eta} f_w(u) f_\varepsilon(v) dudv \\ &\leq Ch^{-1-\eta} \max \|K_\varepsilon^\delta(u)\|_\infty^\eta \int \{K_\varepsilon^\delta(u)\}^2 du. \end{aligned}$$

Under the conditions given in the theorem 4, we can prove that

$$E|U_{n,j}|^{2+\eta} = O\left(h^{-\beta_2(2+\eta)-1-\eta} \exp\{(2+\eta)h^{-\beta}/\gamma\}\right).$$

Under the conditions given in the theorem 5, we can prove that

$$E|U_{n,j}|^{2+\eta} = O\left(\lambda(h)h^{(\beta-1)\eta}\right).$$

The rest is standard and is omitted.





**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [am@scirp.org](mailto:am@scirp.org)