Deconvolving Kernel Regression Function Estimation Based On Right Censored Data

by

EROL OZKAN

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2021

To my wife and kids...

ACKNOWLEDGEMENTS

I would like to express my heartfelt thanks to my advisor, Dr.Shan Sun-Mitchell, for her advice, unwavering support, and encouragement during my graduate studies over the last 3 and half years. She has given informative conversations regarding my study, and her moral guidance is critical to the success of this work.

I would like to extend my sincere gratitude to Dr. Andrzej Korzeniowski, Dr. Suvra Pal, Dr. Li Wang and Dr. Dengdeng Yu. I learned a lot and improved my proofing abilities thanks to Dr. Andrzej Korzeniowski's courses. I am grateful to Dr. Suvra Pal for introducing me to survival analysis, which encouraged me to include this component of my study. I appreciate Dr. Li Wang's assistance with my simulation study guidance. I am thankful to Dr. Dengdeng Yu as well even though we know each other for a short time we did great summer project with him.

Thanks to Mathematics Department at UTA, which provided me the environment to pursue my research interests and made this work possible. I would like to thank to my government for providing this opportunity for me. It is an invaluable experience in my life.

To my wife Nurdan, my sons Mehmet Burak and Enes Hakan, my father Dursun Kaya, my mother Hamiyet, my brother Isa, my nephew Sare, and my cousin Salih, thank you for your love, support, and encouragement, which have enhanced my life and motivated me to continue and finish my graduate studies. Lastly, I would like to thank all of my friends at UTA and special thanks goes to my friend Hrishabh Khakurel.

July 23, 2021

ABSTRACT

Deconvolving Kernel Regression Function Estimation Based On Right Censored Data

Erol Ozkan, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Dr.Sun-Mitchell,Shan

In this study, we propose a new regression function estimator when the observation is contaminated in the convolution model with error in independent variable. We want to examine the effect of the error variables when the data is right censored. The tail behavior of the characteristic function of the error distribution is used to describe the optimum local and global rates of convergence of these kernel estimators. We show that depending on the error is either ordinary smooth or super smooth, there are two sorts of convergence rates in adjusted mean square error for the regression function estimator. It is observed that the rate of convergence is slower in super smooth model for both locally and globally, whereas it is faster in ordinary smooth model. Furthermore, it is examined that in nonparametric regression function estimation, the choice of the kernel K has very little impact on optimality (in the MSE sense), but the bandwidth h has significant impact. Simulation are drawn for different sample sizes in two different examples with 100 replications for each of the samples.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT vi

Chapter Page

1. INTRODUCTION 1

   1.1 Literature Review 3

   1.2 Model Construction 7

2. THE RATE OF CONVERGENCE 12

   2.1 Case 1: Super Smooth Distribution 14

   2.2 Case 2: Ordinary Smooth Distribution 23

3. SIMULATION 29

   3.1 Introduction 29

   3.2 Data generation 29

   3.3 Discussion & Results 32

4. CONCLUSION AND FUTURE STUDIES 37

REFERENCES 38

BIOGRAPHICAL STATEMENT 41

CHAPTER 1

INTRODUCTION

In statistics literature, many studies deal with the problem of nonparametric regression function estimation. There are many approaches to this estimation problem. Kernel functions, spline functions are some of the popular examples. Each approach has its pros and cons. Among these popular techniques, kernel regression estimators have some advantage of mathematical and clarity [1,13]. In terms of kernel-based regression function estimation, one of the approaches is called the Nadaraya-Watson estimator [20]. The Nadaraya-Watson is the very earliest and simplest estimation technique that will be given in detail in the literature review section. The other class of kernel-based regression function estimation is called local polynomial kernel estimators [23,24]. It is estimated the regression function at a specific point by locally fitting a kth degree polynomial to the given dataset using weighted least squares. When the degree equals zero, it turns out to Nadaraya-Watson estimator. The Nadaraya-Watson approach is the special case of local polynomial kernel estimators. When the degree is 1, the method is called the local linear kernel estimator. There is some resemblance between the local linear kernel estimator and other kernel-based estimation approaches, however the local linear has affirmative asymptotic features and boundary manner compared with other estimators.

In this thesis, we work on the regression function estimation for right-censored data by using the deconvolution kernel method. A massive amount of interest has been targeted on the problem of nonparametric regression function estimation. Most of the researches has been focused on the data with standard structure. The regression analysis with errors in variables is expanding fast [4]. Many studies have been directed on that problem such as Anderson (1984), Stefanski and Carroll (1987), Fuller (1987), and more. Among those studies, some of them have focused on the parametric approach in which it is assumed that the regression function follows a known probability density function. Our goal is to work on the nonparametric approach to examine the effect of errors in variables in nonparametric regression function estimation based on the right-censored data. In 1993 Fan and Young investigated the nonparametric regression function estimation with errors in variables using the complete data. In our study, as we use the right-censored data, it can be represented as incomplete observation data. The right-censored data is commonly used in health and survival analysis as well as engineering and social sciences.

The thesis is organized as follows: In Chapter 1, we will talk about the literature about parametric and nonparametric regression. Details will be given in the literature review section. Then, we introduce our proposed model under the model formulation section. In chapter 2, the rate of convergence is investigated for different error distributions such as super smooth and ordinary smooth distribution. The rate of convergence is discussed in terms of adjusted mean squared error. In chapter 3, the simulation study is represented by providing average squared error for different sample sizes as well as the illustration of regression function figure is plotted. Finally, the last chapter consists of the discussion and concluding remarks.

1.1 Literature Review

Parametric models presume strong functional linkage between the variables of interest, namely the dependent and independent variables. In the parametric models, the data are used to adjust certain parameters inside the functional structure to provide the best possible fit. For instance, the linear regression assumes that the relationship between the two variables of interest, independent variable x, and dependent variable y, is given by some version of the form as

$$\begin{aligned} y &= m(x) + \epsilon \\ &= \beta_0 + \beta_1 x + \epsilon, \end{aligned}$$

where the β_0 is the intercept, β_1 is the estimate of the regression slope and ϵ is the error term. The structure of linear regression is assumed there is no way that the model violates the structure which means that any linear model produces the fitted surface to the given data as some kind of line. The linear model can adjust its parameters namely, the slope and the intercept values to provide a better fit to the data. Another example of a parametric model is logistic regression. Different than linear regression, the dependent variable is binary in logistic regression. The logistic model takes the form as follows

$$P(y = 1) = g(\Phi(\beta_0 + \beta_1 x))$$

where $g(\cdot)$ is a functional linkage. Because the logistic regression is the functional linkage to the normal cumulative density function, it always fits the function called sigmoid. Similarly, by adjusting the parameters in the logistic model, the shape of the sigmoid function changes and tries to find the best fit possible.

Non-parametric statistics is a subfield of statistics that isn't entirely rely on parametrized probability distribution families. The nonparametric statistics is predicated on being either having a predefined distribution with undetermined parameters or distribution-free. Both descriptive statistics and statistical inference are included in non-parametric statistics. When the assumptions of parametric tests are violated, non-parametric tests are frequently applied [13,14]. In the above models, as mentioned the regression function requires to take some kind of parametric family function which might be restrictive for adequate estimation of the true regression function. There is an obvious risk of getting inaccurate findings in the regression analysis if a parametric family is not of appropriate form. Thus, by eliminating the requirement that the regression function be a member of a parametric family, the rigidity of parametric regression may be solved. This method produces what is known as a non-parametric regression. A non-parametric regression estimate can be obtained using a variety of approaches. Some are based on very simple concepts, while others are more mathematically complex. Kernel regression, local regression, neural networks, support vector regression and smoothing splines can be given as an example of the non-parametric regression techniques.

One of the simplest approaches of kernel regression function estimation is called Nadaraya-Watson estimator. The Nadaraya-Watson estimator is effectively a weighted average of the data points inside the appropriate bandwidth window. The formal function of the Nadaraya-Watson estimator is

$$\hat{m}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) y_i}{\frac{1}{n}\sum_{j=1}^{n} K_h(x - x_j)}$$

where $\hat{m}(x)$ is the estimate of the regression function, K_h is the kernel function with associated bandwidth h. Note that, the reason of the index is being j in the denominator is to make sure that it is the appropriate weighted average inside the bandwidth window. The Nadaraya-Watson estimator is known as an optimal non-parametric regression function estimator [11,16].

Another technique of kernel-type regression estimation is called local polynomial kernel estimators [23, 25]. These estimate the regression function at a specific location by using weighted least squares to locally fit the kth degree polynomial to the data. The Nadaraya-Watson estimator is the special case of the local polynomial kernel estimator when the degree k equals 0 that is, local constants. The local linear kernel estimators, which corresponds to $k = 1$, are of particular importance and simplicity. The local linear kernel estimators have some similarities to the standard kernel regression estimators, but it has better asymptotic characteristics and boundary behavior than those [11]. The average mean squared error characteristics of the local linear kernel regression estimator's are similar to those of the kernel density estimators. This clearly indicates that most of the concepts developed in the domain of density estimation may be simply transferred to the context of regression. The local kernel regression turns out to be a polynomial regression when the degree k increases.

The non-parametric estimation of the regression function under right censored data has been discussed by Guessoum and Ould-Said in 2008. By using the incomplete data (right censored), they investigated the behavior of a kernel estimator for the regression function. They also discussed the pointwise and uniform strong consistency over a compact set and calculate the estimator's rate of convergence. The asymptotic normality of the estimate is also established. For various scenarios, simulations are constructed to demonstrate both convergence and asymptotic normality. Recently, Aydin and Yilmaz (2016) also give some improvement in terms of bandwidth selection for the non-parametric regression by using the right censored data. To account for censoring, they utilized Kaplan-Meier estimator which proposed by Stute in 1993. In this study, they use three different approaches to select the optimal bandwidth, namely improved Akaike information criterion (AICc), Risk estimation using classsical pilots (RECP), and Generalized cross-validation(GCV). A Monte-Carlo simulation is conducted for this aim to show which criterion provides the best estimates for various sample sizes and censoring levels. As a conclusion, they investigated for all sample sizes and censoring levels, the RECP criterion shows superior performance over the other criteria. In general, the upgraded versin of the AIC and the GCV criteria perform similarly, although GCV outperforms than the AICc.

In 1993, Fan and Young has pointed out about the non-parametric regression with errors in variables. They have investigated the effect of the errors in variables in non-parametric regression function estimation. Deconvolution is used in the implementation of a new class of kernel estimators to account for errors in covariates. They have showed that the tail behavior of the characteristic function of the error distribution may be used to determine the optimal local and global rates of convergence of these kernel estimators. In reality, depending on whether the error is ordinary smooth or super smooth, there are two sorts of convergence rates. It is further demonstrated that these conclusions hold consistently throughout a class of combined response and covariate distributions, which is good enough for practical applications. Furthermore, they demonstrate that the kernel estimators have a lower bound on the convergence rates of all feasible estimators in order to acquire optimality. Another study has been done by Ioannides and Alevizos in 1996 about non-parametric regression with errors in variables and applications. The nonparametric estimator's uniform consistency with sharp rates is demonstrated using the Pollard empirical procedure. The Engel curve analysis and its applications are addressed.

1.2 Model Construction

Consider the model

$$Z = m(X) + \varepsilon \tag{1.1}$$

where $(X_1, ..., X_n)$ is unobservable and right censored data which is contaminated in the following

$$Y = X + E \tag{1.2}$$

where $(Y_1, ..., Y_n)$ is also right censored data from unknown common distribution function $F_Y(\cdot)$ and censored from the right by the censoring time $C_i \sim G(\cdot)$. Assume Y_i and C_i are independent and both of them are non-negative random variables.

In this case, we are only able to observe $W_i = min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$, δ_i is the indicator function that identify if W_i is censored or not. Let J is a distribution function of W_i which implies $J = 1 - (1 - G)(1 - F_Y)$. The censoring indicators are assumed to follow a Bernoulli distribution such as:

$$P(\delta_i = 1) = p \quad and \quad P(\delta_i = 0) = 1 - p \quad for \quad i = 1, ..., n \quad , \tag{1.3}$$

where p lies between 0 and 1. Thus, the joint distribution of (W_i, Z_i, δ_i) can be written as

$$\begin{aligned} g_{W,Z,\delta}(w, z, \delta) &= g_{(W,Z)|\delta=1}(w, z)P(\delta = 1) + g_{(W,Z)|\delta=0}(w, z)P(\delta = 0) \\ &= g_{(Y,Z)|\delta=1}(y, z)p + g_{(C,Z)|\delta=0}(c, z)(1 - p) \end{aligned} \tag{1.4}$$

X_i is contaminated in the model $Y_i = X_i + E_i$, where E_i is the error term and Y_i is the survival time, which is observable, along with the unknown probability density function f_Y in $W_i = min(Y_i, C_i)$. We assume that E_i is a random variable which is independent of X_i and δ_i, $\{i = 1, ..., n\}$. In order to estimate the regression function $m(x) = E(Z|X = x)$, we need to first estimate the unknown distribution of random variable X in (1.2). Hence our first objective is to estimate the unknown density of f_X based on the observations of Y_i.

To start with estimation of f_X, we need to consider the convolution namely $f_Y = f_X * f_E$, and

$$\varphi_{f_X(t)} = \frac{\varphi_{f_Y}(t)}{\varphi_{f_E}(t)} \tag{1.5}$$

where $\varphi_f(\cdot)$ is the characteristic function of f. By using the Fourier inversion theorem, the estimator of f_X, $\hat{f}_X$ can be expressed as

$$\hat{f}_X(x) = \frac{1}{2\pi}\int e^{-itx}\frac{\varphi_{\hat{f}_Y}(t)}{\varphi_{f_E}(t)}dt \tag{1.6}$$

where $\varphi_{\hat{f}_Y}(t) = \int exp(ity)\hat{f}_Y(y)dy$, and $\hat{f}_Y(y)$ can be obtained using two methods.

A common kernel density estimator of f_Y can be obtained by using Kaplan-Meier estimator $\hat{F}_{KM}$ of F_Y that is [27]

$$\hat{F}_{KM}(y) = 1 - \hat{S}_{KM}(y) = \begin{cases} 0 & \text{if } 0 \leq y \leq W_{(1)} \\ 1 - \prod_{i=1}^{j-1}\left(\frac{n-i}{n-i+1}\right)^{\delta_{[i]}} & \text{if } W_{(j-1)} < y \leq W_{(j)} \quad \text{j=2,3,...,n} \\ 1 & \text{if } y > W_{(n)}, \end{cases}$$

where $(W_i, \delta_{[i]})$, $i = 1, ...n$ represents the (W_i, δ_i) ordered in terms of W_i's. By using the Kaplan-Meier method, the kernel density estimator of f_Y can be written

$$\begin{aligned} \hat{f}_Y^{KM} &= \frac{1}{h}\int K\left(\frac{y-x}{h}\right)d\hat{F}_{KM}(y) \\ &= \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{y-W_{(j)}}{h}\right)s_j, \end{aligned} \tag{1.7}$$

where s_j denotes the size of jump of $\hat{F}_{KM}$ at $W_{(j)}$, h is the bandwidth which is assumed to be positive, and K is the regular kernel function which holding

$$\int K(t)dt = 1 \quad , \quad \int tK(t)dt = 0 \quad and \quad \int t^2K(t)dt < \infty.$$

Therefore, Chakrabarty [2] showed that the $\hat{f}_X$ is defined by combining the (1.6) and (1.7) as following

$$\hat{f}_X^{KM}(x) = \frac{1}{h}\sum_{i=1}^{n} K^z\left(\frac{x-W_{(j)}}{h}\right)s_j, \tag{1.8}$$

where $K^{(z)}\left(\frac{x-W_{(j)}}{h}\right)=\frac{1}{2\pi}\int exp\left(\frac{-it(x-W_{(j)})}{h}\right)\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}dt.$

In this study, we introduce an another way of estimating the unknown density of f_X. The reason that we do not want to use the Kaplan-Meier is that the $W_{(i)}$'s are order statistic in the (1.8) which are not independent. Therefore it is extremely hard to evaluate the properties of this estimator (1.8). Satten and Datta[9] stated that the Kaplan-Meier estimator $\hat{F}_{KM}$ is parallel to the following estimator using inverse probability of censoring weighted idea as

$$\hat{F}_{IP}(y)=\frac{1}{n}\sum_{i=1}^{n}\frac{I(W_i\leq y)\delta_i}{1-G(W_i)}. \tag{1.9}$$

Hence, the application of the inverse probability of weighted censoring idea f_Y produces:

$$\begin{aligned}\hat{f}_Y^{IP}(y)&=\frac{1}{h}\int K\left(\frac{y-x}{h}\right)d\hat{F}_{IP}(x)\\&=\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{W_j-y}{h}\right)\frac{\delta_j}{1-G(W_j)}.\end{aligned} \tag{1.10}$$

Now, we need to find the characteristic function of $\hat{f}_Y^{IP}$ which is denoted by $\varphi_{\hat{f}_Y^{IP}}(t)$. In order to formulate $\varphi_{\hat{f}_Y^{IP}}(t)$, we use the symmetric property of K as follows [21,22]

$$
\begin{aligned}
\varphi_{\hat{f}_Y^{IP}}(t) &= \int e^{ity} \hat{f}_Y^{IP}(y) dy \\
&= \int e^{ity} \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{W_j - y}{h}\right) \frac{\delta_j}{1 - G(W_j)} dy \\
&= \frac{1}{nh} \sum_{j=1}^{n} \int e^{ity} K\left(\frac{W_j - y}{h}\right) \frac{\delta_j}{1 - G(W_j)} dy \\
&= \frac{1}{nh} \sum_{j=1}^{n} \frac{\delta_j}{1 - G(W_j)} \int e^{ity} K\left(\frac{W_j - y}{h}\right) dy \\
&= \frac{1}{n} \sum_{j=1}^{n} \frac{\delta_j}{1 - G(W_j)} \int e^{it(hu + W_j)} K(u) du \\
&= \frac{1}{n} \sum_{j=1}^{n} \frac{\delta_j}{1 - G(W_j)} e^{itW_j} \varphi_K(ht)
\end{aligned} \tag{1.11}
$$

By using (1.6) and (1.11), we obtain the deconvolving kernel density estimation of $f_X(x)$ as

$$
\begin{aligned}
\hat{f}_X^{IP}(x) &= \frac{1}{2\pi} \int e^{-itx} \frac{\varphi_{\hat{f}_Y}(t)}{\varphi_{f_E}(t)} dt \\
&= \frac{1}{2\pi} \int e^{-itx} \frac{\frac{1}{n} \sum_{j=1}^{n} \frac{\delta_j}{1 - G(W_j)} (e^{itW_j}) \varphi_K(ht)}{\varphi_{f_E}(t)} dt \\
&= \frac{1}{2\pi n} \sum_{j=1}^{n} \int \frac{e^{it(W_j - x)} \varphi_K(ht)}{\varphi_{f_E}(t)} dt \frac{\delta_j}{1 - G(W_j)} \\
&= \frac{1}{nh} \sum_{j=1}^{n} \frac{1}{2\pi} \int \frac{e^{iy \frac{(x - W_j)}{h}} \varphi_K(y)}{\varphi_{f_E}(\frac{y}{h})} dy \frac{\delta_j}{1 - G(W_j)} \\
&= \frac{1}{nh} \sum_{j=1}^{n} K^Z\left(\frac{x - W_j}{h}\right) \frac{\delta_j}{1 - G(W_j)},
\end{aligned} \tag{1.12}
$$

where

$$
K^Z\left(\frac{x - W_j}{h}\right) = \frac{1}{2\pi} \int e^{\left(-it\left(\frac{x - W_j}{h}\right)\right)} \frac{\varphi_K(t)}{\varphi_{f_E}\left(\frac{t}{h}\right)} dt \tag{1.13}
$$

We propose the following kernel regression function estimator involving errors in variables:

$$\begin{aligned}\hat{m}(x) &= \frac{\sum_{j=1}^{n} K^Z\left(\frac{x-W_j}{h}\right) Z_j}{\sum_{j=1}^{n} K^Z\left(\frac{x-W_j}{h}\right) \frac{\delta_j}{1-G(W_j)}} \\ &= \frac{1}{nh} \frac{\sum_{j=1}^{n} K^Z\left(\frac{x-W_j}{h}\right) Z_j}{\hat{f}_X^{IP}(x)}\end{aligned} \tag{1.14}$$

where $\hat{f}_X^{IP}(x)$ and $K^Z(\cdot)$ are given in (1.12) and (1.13), respectively.

CHAPTER 2

THE RATE OF CONVERGENCE

In this chapter, we are going to investigate the sampling behaviour of the kernel estimators in (1.14) proposed in the model formulation section. The type of error distribution plays an important role in terms of the rate of convergence for these estimators. Now, the definition of ordinary smooth and super smooth distribution will be given into the following:

- Definition(2.1) Super smooth distribution [15]:

  A random variable, E, is considered to possess a super smooth distribution of order β, if the characteristic function of E, represented by $\varphi_{f_E}(\cdot)$, holds:

$$d_0|t|^{\beta_0} exp\left(\frac{-|t|^\beta}{\gamma}\right) \leq |\varphi_{f_E}(t)| \leq d_1|t|^{\beta_1} exp\left(\frac{-|t|^\beta}{\gamma}\right)(t \to \infty), \tag{2.1}$$

  where d_0, d_1, β and γ are positive constants and β_0, β_1 are constants. The examples for super smooth distribution are normal,Cauchy.

- Definition(2.2) Ordinary smooth distribution [15]:

  A random variable, E, is considered to possess a ordinary smooth distribution of order β, if the characteristic function of E, represented by $\varphi_{f_E}(\cdot)$, holds:

$$d_0|t|^{-\beta} \leq |\varphi_{f_E}(t)| \leq d_1|t|^{-\beta}(t \to \infty), \tag{2.2}$$

  where d_0, d_1, β and γ are positive constants and β_0, β_1 are constants. The examples for ordinary smooth distribution are gamma, double exponential.

  The convergence rates depend upon the order of the smoothness of error distribution, β. They depend on the smoothness of the $m(x)$ as well. Furthermore, there are some conditions that are playing a crucial role in terms of the rates of convergence of $\hat{m}(x)$. Throughout this chapter, we will need the following assumptions to prove the required lemmas and theorems.

Assumption A

- **(A1)** h_n is a sequence of bandwidths that satisfies $h = h_n \to 0$ as $n \to \infty$ and $nh_n \to \infty$ as $n \to \infty$.

Assumption B

- **(B1)** The characteristic function of error distribution, $\varphi_{f_E}(\cdot)$, does not disappear.
- **(B2)** Assume $a < b$. The marginal density of the unobserved independent variable X is bounded on the interval $[a, b]$. Also, it has a bounded kth derivative.
- **(B3)** The regression function $m(\cdot)$ has continuous kth derivative on $[a, b]$.
- **(B4)** $E(Z^2|X = x)$ is continuous on $[a, b]$. Furthermore, $E(Z^2) < \infty$.

Assumption C

The kernel $K(\cdot)$ is a kth-order kernel. That is:

- **(C1)** $\int y^j K(y)dy = 0$ for $j = 1, ..., k - 1$.
- **(C2)** $\int y^k K(y)dy \neq 0$

Assumption $(B1)$ establishes that the proposed estimator $\hat{m}(x)$ clearly expressed. Assumption $(B2) - (B4)$ are similar to those necessary in the ordinary non-parametric regression. Furthermore, the rates of convergence depend upon the assumption C.

In the following sections, the rates of convergence of the proposed regression function will be discussed.

2.1 Case 1: Super Smooth Distribution

In this section, the rate of convergence will be investigated when the error distribution follows a super smooth distribution.

Theorem 2.1 (Rate for Adjusted MSE of $\hat{m}(x)$). Assume (B) and (C), and suppose that the left-hand side of inequality (2.1) is satisfied. Suppose that $\varphi_K(t)$ has a bounded support on $|t| \leq T_0$, T_0 is a non-negative constant. Then, for bandwidth $h_n = d(logn)^{-1/\beta}$ with $d > T_0 \left(\frac{2}{\gamma}\right)^{\frac{1}{\beta}}$, and k is a positive number,

$$E\left([\hat{m}(x) - m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 = (d^k a_k(x))^2 (logn)^{\frac{-2k}{\beta}}(1+o(1)) + O\left(\frac{1}{n}\right) \quad (2.3)$$

and

$$E\int_a^b \left([\hat{m}(x) - m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 dx = \int_a^b (d^k a_k(x))^2 dx (logn)^{\frac{-2k}{\beta}}(1+o(1)) + O\left(\frac{1}{n}\right), \quad (2.4)$$

where

$$a_k(x) = (-1)^k \left[\frac{[m(x)f_{X|\delta=1}(x)]^{(k)}}{k!} - \frac{m(x)[f_{X|\delta=1}^{(k)}(x)]}{k!}\right] p f_{X|\delta=1}^{-1}(x) \int u^k K(u)du.$$

Note that the multiplier $\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}$ is employed to prevent possible difficulty of having 0 in the denominator of $\hat{m}(x) - m(x)$.

In order to show (2.3) and (2.4), we need following 3 lemmas.

Lemma 2.1 Under assumption $(A1)$ and if $\varphi_K(\cdot)$ vanishes outside the interval $[-T_0, T_0]$,then

$$ED_n(x) = \frac{p}{h}\int [m(y) - m(x)]K\left(\frac{x-y}{h}\right) f_{X|\delta=1}(y)dy \quad (2.5)$$

where

$$D_n(x) = \frac{1}{nh}\sum_{j=1}^{n} K^Z\left(\frac{x - W_j}{h}\right)\frac{\delta_j}{1 - G(W_j)}[Z_j - m(x)] \quad (2.6)$$

Proof. of Lemma (2.1) First of all, we begin the proof with multiplying $D_n(x)$ by h, which gives,

$$hD_n(x) = \frac{1}{n} \sum K^Z \left(\frac{x - W_j}{h} \right) \frac{\delta_j}{1 - G(W_j)} [Z_j - m(x)]. \tag{2.7}$$

Next, take the expectation of (2.7),

$$hED_n(x) = E \left[K^Z \left(\frac{x - W}{h} \right) \frac{\delta}{1 - G(W)} [Z - m(x)] \right] \tag{2.8}$$

As it can be observed from (2.8), we have three random variables which are W, Z and δ. Note that the joint probability distribution of (W, Z, δ), which we define in the model formulation section, is given by:

$$g_{W,Z,\delta}(w, z, \delta) = g_{(Y,Z)|\delta=1}(y, z)\, p + g_{(C,Z)|\delta=0}(c, z)(1 - p) \tag{2.9}$$

Note that, when $\delta = 1$, $W = min(Y_i, C_i)$ is equal to Y_i. So the term $\frac{\delta}{1-G(W)}$ will be equal to 1 as $G(Y_i) = 0$.

Therefore, we can write (2.8) in terms of integral as

$$\begin{aligned} hED_n(x) &= \int\int K^Z \left(\frac{x - y}{h} \right) \frac{1}{1 - G(y)} [z - m(x)] g_{(Y,Z)|\delta=1}(y, z) p \, dy \, dz \\ &+ \int\int K^Z \left(\frac{x - y}{h} \right) \frac{0}{1 - G(y)} [z - m(x)] g_{(C,Z)|\delta=1}(c, z)(1 - p) dc \, dz \\ &= \int\int K^Z \left(\frac{x - y}{h} \right) \frac{1}{1 - G(y)} [z - m(x)] g_{(Y,Z)|\delta=1}(y, z) p \, dy \, dz \\ &= \frac{p}{2\pi} \int\int e^{\left(\frac{-itx}{h}\right)} e^{\left(\frac{ity}{h}\right)} \frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})} [z - m(x)] g_{(Y,Z)|\delta=1}(y, z) \, dy \, dz \end{aligned} \tag{2.10}$$

Let $g_{Y,Z|\delta=1}(\cdot)$ and $f_{X,Z|\delta=1}(\cdot)$ represent the conditional density of (Y, Z) and (X, Z), respectively. Using the independence of E and (X, Z) and $Y = X + E$,

$$g_{(Y,Z)|\delta=1}(y, z) = \int f_{(X,Z)|\delta=1}(y - x, z) dF_E(x), \tag{2.11}$$

where $F_E(\cdot)$ is the cumulative distribution function of E. Also, $f_X(x)$ represents the marginal density of X.

By using the equation (2.11), (2.10) can be rewritten as:

$$\frac{p}{2\pi}\int\int\int\int e^{\left(\frac{-itx}{h}\right)}e^{\left(\frac{ity}{h}\right)}\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}[z-m(x)]f_{(X,Z)|\delta=1}(y-u,z)dt\,dF_E(u)dy\,dz \quad (2.12)$$

Note that the transformation of Fourier is equal to the transformation of the product:

$$\int e^{\frac{ity}{h}}\left[\int f_{(X,Z)|\delta=1}(y-u,z)dF_E(u)\right]dy = \varphi_{f_E}\left(\frac{t}{h}\right)\int e^{\frac{ity}{h}}f_{(Y,Z)|\delta=1}(y,z)dy \quad (2.13)$$

Using (2.13), it can be acquired

$$\begin{aligned}
&\frac{p}{2\pi}\int\int\int e^{\left(\frac{-itx}{h}\right)}e^{\left(\frac{ity}{h}\right)}\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}f_{(X,Z)|\delta=1}(y-u,z)dt\,dF_E(u)dy\\
&=\frac{p}{2\pi}\int e^{\left(\frac{-itx}{h}\right)}\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\left(\int e^{\left(\frac{ity}{h}\right)}\int f(y-u,z)dF_E(u)dy\right)dt\\
&=\frac{p}{2\pi}\int e^{\frac{-itx}{h}}\varphi_K(t)\int e^{\frac{ity}{h}}f_{(Y,Z)|\delta=1}(y,z)dy\,dt\\
&=p\int K\left(\frac{x-y}{h}\right)f_{(Y,Z)|\delta=1}(y,z)dy
\end{aligned} \quad (2.14)$$

where the latter equality in (2.14) comes from the inversion of the Fourier transformation. By using the fact that the inversion of two Fourier transforms equivalent to convolution, and (2.10) and (2.14), we obtain:

$$\begin{aligned}
ED_n(x) &= \frac{p}{h}\int\int[z-m(x)]K\left(\frac{x-y}{h}\right)f(y,z)dydz\\
&= \frac{p}{h}\int[m(y)-m(x)]K\left(\frac{x-y}{h}\right)f_{X|\delta=1}(y)dy
\end{aligned} \quad (2.15)$$

The following lemma provides the upper bound of $ED_n(x)$.

Lemma 2.2 Under assumption (C) and k is a positive integer,

$$\begin{aligned}
ED_n(x) &= \frac{p}{h}\int[m(y)-m(x)]K\left(\frac{x-y}{h}\right)f_{X|\delta=1}(y)dy\\
&\leq f_{X|\delta=1}(x)a_k(x)h^k(1+o(1))
\end{aligned} \quad (2.16)$$

where $D_n(x)$ is given by (2.6) and

$$a_k(x) = (-1)^k \left[\frac{[m(x) f_{X|\delta=1}(x)]^{(k)}}{k!} - \frac{m(x)[f^{(k)}_{X|\delta=1}(x)]}{k!} \right] p f^{-1}_{X|\delta=1}(x) \int u^k K(u) du$$

Proof. By Lemma 1, we have

$$ED_n(x) = \frac{p}{h} \int [m(y) - m(x)] K\left(\frac{x-y}{h}\right) f_{X|\delta=1}(y) dy \tag{2.17}$$

Let $u = \frac{x-y}{h}$ which implies $y = x - uh$. Then, by taking derivative we obtain $dy = -hdu$ and using (A2):

$$\begin{aligned} ED_n(x) &= p \left| \int [m(x-uh) - m(x)] K(u) f_{X|\delta=1}(x-uh) du \right| \\ &\leq p \left| \int [m(x-uh) - m(x)] K(u) f_{X|\delta=1}(x-uh) du \right| \\ &= p \left| \int m(x-uh) K(u) f_{X|\delta=1}(x-uh) du - \int m(x) K(u) f_{X|\delta=1}(x-uh) du \right| \end{aligned} \tag{2.18}$$

In the first term of (2.18), let $g(x-uh) = m(x-uh) f_{X|\delta=1}(x-uh)$. For $g(x-uh)$, we use Taylor expansion:

$$\begin{aligned} g(x-uh) &= g(x) - uhg'(x) + \frac{(-1)^2 u^2 h^2 g''(x)}{2!} + ... + (-1)^{k-1} \frac{u^{k-1} h^{k-1} g^{(k-1)}(x)}{(k-1)!} \\ &\quad + (-1)^k \frac{u^k h^k g^{(k)}(x)}{k!} + o(h^k) \\ &= m(x) f_{X|\delta=1}(x) - uh[m(x) f_{X|\delta=1}(x)]' + \frac{(-1)^2 u^2 h^2 [m(x) f_{X|\delta=1}(x)]''}{2!} + ... \\ &\quad + (-1)^{k-1} \frac{u^{k-1} h^{k-1} [m(x) f_{X|\delta=1}(x)]^{(k-1)}}{(k-1)!} \\ &\quad + (-1)^k \frac{u^k h^k [m(x) f_{X|\delta=1}(x)]^{(k)}}{k!} + o(h^k) \end{aligned} \tag{2.19}$$

Multiplying (2.19) by $K(u)$ and taking the integral in terms of du, we obtain

$$\begin{aligned}\int m(x-uh)f_{X|\delta=1}(x-uh)K(u)du &= m(x)f_{X|\delta=1}(x)\int K(u)du \\ &- h[m(x)f_{X|\delta=1}(x)]\int u'K(u)du \\ &+ \frac{(-1)^2h^2[m(x)f_{X|\delta=1}(x)]''}{2!}\int u^2K(u)du + ... \\ &+ (-1)^{k-1}\frac{h^{k-1}[m(x)f_{X|\delta=1}(x)]^{(k-1)}}{(k-1)!}\int u^{k-1}K(u)du \\ &+ (-1)^k\frac{h^k[m(x)f_{X|\delta=1}(x)]^{(k)}}{k!}\int u^kK(u)du + o(h^k).\end{aligned} \tag{2.20}$$

Recall that $\int y^jK(y)dy = 0$ for $j = 1, ..., k-1$, therefore (2.20) becomes

$$\begin{aligned}&\int m(x-uh)f_{X|\delta=1}(x-uh)K(u)du \\ &\quad = m(x)f_{X|\delta=1}(x) + (-1)^kh^k\frac{[m(x)f_{X|\delta=1}(x)]^{(k)}}{k!}\int u^kK(u)du + o(h^k).\end{aligned} \tag{2.21}$$

Now, similarly apply Taylor expansion to $f_{X|\delta=1}(x-uh)$:

$$\begin{aligned}f_{X|\delta=1}(x-uh) &= f_{X|\delta=1}(x) - uhf'_{X|\delta=1}(x) + (-1)^2\frac{u^2h^2f''_{X|\delta=1}(x)}{2!} + ... \\ &+ (-1)^{k-1}\frac{u^{k-1}h^{k-1}f^{(k-1)}_{X|\delta=1}(x)}{(k-1)!} + (-1)^k\frac{u^kh^kf^{(k)}_{X|\delta=1}(x)}{k!} + o(h^k).\end{aligned} \tag{2.22}$$

Multiply (2.22) by $m(x)K(u)$ and take the integral in terms of du:

$$\begin{aligned}\int m(x)f_{X|\delta=1}(x-uh)K(u)du &= m(x)f_{X|\delta=1}(x)\int K(u)du - m(x)hf'_{X|\delta=1}(x)\int uK(u)du \\ &+ (-1)^2\frac{h^2f''_{X|\delta=1}(x)}{2!}\int u^2K(u)du + ... \\ &+ (-1)^{k-1}\frac{h^{k-1}f^{(k-1)}_{X|\delta=1}(x)}{(k-1)!}\int u^{k-1}K(u)du \\ &+ (-1)^k\frac{h^kf^{(k)}_{X|\delta=1}(x)}{k!}\int u^kK(u)du + o(h^k).\end{aligned} \tag{2.23}$$

Recall that $\int y^j K(y)dy = 0$ for $j = 1, ..., k-1$, therefore (2.23) becomes

$$\int m(x) f_{X|\delta=1}(x-uh)K(u)du$$

$$= m(x)f_{X|\delta=1}(x) + (-1)^k h^k \frac{m(x)f^{(k)}_{X|\delta=1}(x)}{k!} \int u^k K(u)du + o(h^k). \quad (2.24)$$

Next, we subtract (2.24) from (2.21),

$$\left| \int m(x-uh) f_{X|\delta=1}(x-uh)K(u)du - \int m(x) f_{X|\delta=1}(x-uh)K(u)du \right|$$

$$= (-1)^k h^k \frac{[m(x)f_{X|\delta=1}(x)]^{(k)}}{k!} \int u^k K(u)du - (-1)^k h^k \frac{m(x)f^{(k)}_{X|\delta=1}(x)}{k!} \int u^k K(u)du + o(h^k). \quad (2.25)$$

Now, plug in back the equality (2.25) into (2.18), then we obtain,

$$\begin{aligned} E(D_n(x)) &\leq p(-1)^k h^k \frac{[m(x)f_{X|\delta=1}(x)]^{(k)}}{k!} \int u^k K(u)du \\ &- (-1)^k h^k \frac{m(x)f^{(k)}_{X|\delta=1}(x)}{k!} \int u^k K(u)du + o(h^k) \\ &= f_{X|\delta=1}(x) a_k(x) h^k (1 + o(1)) \end{aligned} \quad (2.26)$$

The following lemma provides the upper bound for the norm of the deconvolving kernel $K^z(x)$.

Lemma 2.3 By using the definition of super smooth distribution in (2.1) and $\varphi_K(t)$ has a bounded support $|t| \leq T_0$,

$$\sup_x |K^Z(x)| \leq O(h) + O\left(h\, exp\left(\left| \frac{T_0}{h} \right|^\beta \right) \gamma^{-1} \right) \quad (2.27)$$

where

$$K^Z(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})} dt$$

Proof. First of all, supremum of $K^z(x)$ can be written as

$$\sup_x |K^Z(x)| \leq \int_{-T_0}^{-Th} \left| \frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})} \right| dt + \int_{-Th}^{Th} \left| \frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})} \right| dt + \int_{Th}^{T_0} \left| \frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})} \right| dt \quad (2.28)$$

By using the symmetric property of the kernel K, (2.28) can be rewritten as

$$\sup_x |K^Z(x)| \leq 2\int_0^{Th} \left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right| dt + 2\int_{Th}^{T_0} \left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right| dt \tag{2.29}$$

For the first term of (2.29) on the right hand side, as $\varphi_K(t)$ is bounded (i.e $|\varphi_K(t)| \leq 2\int K(x)dx = 2$), thus using the substitution of $t = yh$ which implies $dt = hdy$, so we get

$$\int_0^{Th} \frac{1}{\varphi_{f_E}(\frac{t}{h})} dt = \int_0^{T} \frac{h}{\varphi_{f_E}(y)} dy = O(h) \tag{2.30}$$

For the second term of (2.29) on the right hand side, we use the first half of (2.1) which states that there exists a constant T such that:

$$|\varphi_{f_E}(t)| \geq \left(\frac{d_0}{2}\right)|t|^{\beta_0} exp\left(\frac{-|t|^\beta}{\gamma}\right) \qquad , |t| > T \tag{2.31}$$

It is obvious that $t > Th$ which implies $\frac{t}{h} > T$. So, (2.31) can be rewritten as

$$\left|\varphi_{f_E}\left(\frac{t}{h}\right)\right| \geq \left(\frac{d_0}{2}\right)\left|\frac{t}{h}\right|^{\beta_0} exp\left(\frac{-|\frac{t}{h}|^\beta}{\gamma}\right) \tag{2.32}$$

Now, if we use (2.32) and plug in into the second term that mentioned above, we get

$$\begin{aligned} 2\int_{Th}^{T_0} \left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right| dt &\leq 2\left(\frac{d_0}{2}\right)^{-1} \int \left(\frac{t}{h}\right)^{-\beta_0} |\varphi_K(t)| exp\left(\left|\frac{t}{h}\right|\gamma^{-1}\right) dt \\ &\leq \frac{4}{d_0} exp\left(\left|\frac{T_0}{h}\right|^\beta \gamma^{-1}\right) h^{\beta_0} \int_{Th}^{T_0} t^{-\beta_0} dt \\ &= O\left(h\, exp\left(\left|\frac{T_0}{h}\right|^\beta \gamma^{-1}\right)\right) \end{aligned} \tag{2.33}$$

Hence, the proof of Lemma 2.3 is completed. Now, we are ready to start proving the Theorem 1.

Proof. Note that,

$$[\hat{m}(x) - m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)} = \frac{D_n(x)}{f_{X|\delta=1}(x)} \tag{2.34}$$

Then using "bias"-variance decomposition and Lemma 2.2:

$$
\begin{aligned}
E\left((\hat{m}(x)-m(x))\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 &= E\left(\frac{D_n(x)}{f_{X|\delta=1}(x)}\right)^2 \\
&= \frac{\mathrm{Var}D_n(x)}{f_{X|\delta=1}(x)^2} + \frac{E(D_n(x))^2}{f^2_{X|\delta=1}(x)} \\
&= \frac{\mathrm{Var}D_n(x)}{f_{X|\delta=1}(x)^2} + \frac{f^2_{X|\delta=1}(x)a_k^2(x)h^{2k}(1+o(1))}{f^2_{X|\delta=1}(x)} \\
&= \frac{\mathrm{Var}D_n(x)}{f_{X|\delta=1}(x)^2} + a_k^2(x)h^{2k}(1+o(1))
\end{aligned}
\tag{2.35}
$$

Next, we show that $\mathrm{Var}D_n(x) = O\left(\frac{1}{n}\right)$. To do this, we use the independence of X and E, so that

$$
\begin{aligned}
\mathrm{Var}D_n(x) &= Var\left(\frac{1}{nh}\sum_{j=1}^{n} K^Z\left(\frac{x-W_j}{h}\right)\frac{\delta_j}{1-G(W_j)}[Z_j - m(x)]\right) \\
&= Var\left(\frac{1}{nh}\sum_{j=1}^{n} K^Z\left(\frac{x-Y_j}{h}\right)[Z_j - m(x)]\right) \\
&= \frac{1}{nh^2}\mathrm{Var}\left(K^Z\left(\frac{x-Y}{h}\right)[Z-m(x)]\right) \\
&= \frac{1}{nh^2}E\left(\left|K^Z\left(\frac{x-Y}{h}\right)\right|^2[Z-m(x)]^2\right) \\
&- \frac{1}{nh^2}E\left(\left|K^Z\left(\frac{x-Y}{h}\right)\right|[Z-m(x)]\right)^2 \\
&\leq \frac{1}{nh^2}E\left(\left|K^Z\left(\frac{x-Y}{h}\right)\right|^2[Z-m(x)]^2\right) \\
&= \frac{1}{nh^2}E\left(\left|K^Z\left(\frac{x-Y}{h}\right)\right|^2[Z-m(x)]^2\right) \\
&\leq \frac{1}{nh^2}\sup_u(K^Z(u))^2E(\tau^2(X))
\end{aligned}
\tag{2.36}
$$

where $\tau^2(X) = E((Z-m(x))^2|X)$ which is assumed to be finite. Therefore, we need to show that $\sup_u(K^Z(u))^2$ is approaching to 0 with some rate. In order to show this, we consider lemma 2.3 and the following:

In the statement of the theorem 1, we have $d > T_0 \left(\frac{2}{\gamma}\right)^{\frac{1}{\beta}}$ and $h = d(logn)^{-\frac{1}{\beta}}$ which leads to:

$$exp\left(\left|\frac{T_0}{h}\right|^{\beta}\gamma^{-1}\right) \leq exp\left((logn)^{-1}\right) \tag{2.37}$$

which converges to 1 as $n \to \infty$.

Now, by using lemma 2.3

$$\sup_u |K^Z(u)| \leq O(h) + O\left(h\, exp\left(\left|\frac{T_0}{h}\right|^{\beta}\right)\gamma^{-1}\right) \tag{2.38}$$

If we take the square of both sides of (2.38), we obtain

$$\sup_u [K^Z(u)]^2 \leq O(h^2) + O(h^2 exp(logn)^{-1}) + 2\,O(h^2 exp(2logn)^{-1}) \tag{2.39}$$

Multiply (2.39) by $\frac{1}{nh^2}$, then

$$\frac{1}{nh^2}\sup_u [K^Z(u)]^2 \leq O(\frac{1}{n}) + O(\frac{1}{n}exp(logn)^{-1}) + 2\,O(\frac{1}{n}exp(2logn)^{-1}) \tag{2.40}$$

So,

$$\text{Var}D_n(x) \leq O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}exp\left((logn)^{-1}\right)\right) + O\left(\frac{1}{n}exp\left((2logn)^{-1}\right)\right) = O\left(\frac{1}{n}\right) \tag{2.41}$$

which approaches to 0 as $n \to \infty$.

Now, go back to (2.35) and plug in the result (2.41) which gives

$$E\left((\hat{m}(x) - m(x))\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 = (d^k a_k(x))^2 (logn)^{\frac{-2k}{\beta}}(1 + o(1)) + O\left(\frac{1}{n}\right) \tag{2.42}$$

Furthermore, as (2.16) and (2.41) hold uniformly in $x \in (a, b)$, the second expression in the Theorem 1 is also accurate.

$$E\int_a^b \left([\hat{m}(x) - m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 dx = \int_a^b (d^k a_k(x))^2 dx (logn)^{\frac{-2k}{\beta}}(1+o(1)) + O\left(\frac{1}{n}\right) \tag{2.43}$$

Proof of Theorem 1 is completed.

2.2 Case 2: Ordinary Smooth Distribution

In this section, the rate of convergence will be investigated when the error distribution follows an ordinary smooth distribution. We need the following assumption:

Assumption D

- **(D1)** $\varphi_{f_E}(t) \to d$ and $\varphi'_{f_E}(t)(t^{\beta+1}) \to -d\beta$ as $t \to \infty$ given $d \neq 0$ and β is non-negative constant. Furthermore, $\varphi_{f_E}(t) \neq 0$ for any given t.
- **(D2)** $\varphi_K(t)$ has $(m+2)$ bounded integrable derivatives and being symmetric function. Also, $\varphi_K(0) = 1$ and $\varphi_K(t) = 1 + O(|t|^m)$ as $t \to 0$.
- **(D3)**The unknown $f_X(\cdot)$ has m^{th} derivative and it is continuous.
- **(D4)** $\int |t^{\beta+1}|(|\varphi_K(t)| + |\varphi'_K(t)|)dt < \infty \quad , \int |t^{\beta+1}\varphi_K(t)|^2 dt < \infty$

Note that the assumption (D1) implies:

$$\left|\frac{h^\beta \varphi_K(t}{\varphi_{f_E}(t)}\right| \leq \left|\frac{max|\varphi_K(t)|}{min|\varphi_{f_E}(t)|}\right| \tag{2.44}$$

which will be used in the proof of lemma in this section.

Theorem 2.2 Assume that assumptions (B),(C) and (D4) satisfies. Then, under the ordinary smooth error distribution (2.2) and let $h = cn^{\frac{-1}{2k+2\beta+1}}$ with provided $c > 0$,

$$\begin{aligned} E\left([\hat{m}(x) - m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 &= \left[a_k^2 h^{2k} + \frac{1}{nh^{1+2\beta}}\xi(x)\right](1 + o(1)) \\ &= O(n^{\frac{-2k}{[2k+2\beta+1}}) \end{aligned} \tag{2.45}$$

and

$$\begin{aligned} E\int_a^b \left([\hat{m}(x) - m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 &= \left[a_k^2 h^{2k} + \frac{1}{nh^{1+2\beta}}\xi(x)\right](1 + o(1)) \\ &= O(n^{\frac{-2k}{[2k+2\beta+1}}), \end{aligned} \tag{2.46}$$

where $\xi(x)$ is given by

$$\xi(x) = \frac{1}{2\pi f^2_{X|\delta=1}(x)} \int \left|\frac{t^\beta}{d}\right|^2 |\varphi_K(t)|^2 dt \int \tau(x-u) f_{X|\delta=1}(x-u) dF_E(u) \qquad (2.47)$$

with $\tau^2(X) = E((Z - m(x))^2|X)$. Here k is a positive integer and $\beta > 0$.

Now, before we start proving the theorem (2.2), we need the following lemma.

Lemma 2.4 Under the assumptions (B) and (C),

$$|h^\beta K^z(x)| \leq \frac{D}{1+|x|} \qquad (2.48)$$

given that D is some constant.

Proof: As stated in the definition of the ordinary smooth distribution in (2.2), there exist T and c_1 which are positive constants so that

$$|\varphi_{f_E}(t)| \geq c_1 |t|^{-\beta} \quad for \quad |t| > T.$$

As similar to in the proof of lemma 2.3, it can be written

$$\sup_x |K^Z(x)| \leq 2\int_0^{Th} \left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right| dt + 2\int_{Th}^{\infty} \left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right| dt \qquad (2.49)$$

By using the result in (2.44), the first term of (2.49) can be written

$$2\int_0^{Th} \left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right| dt \leq 2Th \left|\frac{max|\varphi_K(t)|}{min|\varphi_{f_E}(t)|}\right| \qquad (2.50)$$

Now, we work on the second term in (2.49). We use the first half of the ordinary smooth distribution definition such that

$$\varphi_{f_E}(t) \geq c_1|t|^{-\beta} \qquad , |t| > T \qquad (2.51)$$

Let $t > Th$ which implies $\frac{t}{h} > T$. So, (2.51) can be rewritten as

$$\varphi_{f_E}\left(\frac{t}{h}\right) \geq c_1 \left|\frac{t}{h}\right|^{-\beta} \qquad (2.52)$$

Thus,

$$\begin{aligned} 2\int_{Th}^{\infty}\left|\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right|dt &\leq 2\int_{Mh}^{\infty}\frac{|\varphi_K(t)|}{c_1}\left|\frac{t}{h}\right|^{\beta}dt \\ &\leq \frac{2}{h^{-\beta}c_1}\int_0^{\infty}|\varphi_K(t)|t^{\beta}dt \\ &= O(h^{-\beta}) \end{aligned} \tag{2.53}$$

So, we can write

$$\begin{aligned} |h^{\beta}K^z(x)| &\leq \frac{h^{\beta}}{2\pi}\int\frac{|\varphi_K(t)|}{|\varphi_{f_E}(\frac{t}{h})|}dt \\ &= O(1) \end{aligned} \tag{2.54}$$

After applying the integration by parts to expression $|h^{\beta}K^z(x)|$, we obtain

$$\begin{aligned} |h^{\beta}K^z(x)| &\leq \frac{h^{\beta}}{2\pi|x|}\int\left|\left(\frac{\varphi_K(t)}{\varphi_{f_E}(\frac{t}{h})}\right)'\right|dt \\ &\leq \frac{C}{|x|} \end{aligned} \tag{2.55}$$

where $C > 0$. By combining (2.54) and (2.55), we get

$$|h^{\beta}K^z(x)| \leq \frac{D}{1+|x|} \tag{2.56}$$

Now, we are ready to start proving Theorem (2.2).

Proof: Using (2.35), it is enough to evaluate the "adjusted bias" and the variance of $D_n(x)$ given by (2.6). Recall in lemma 2.2, we have

$$ED_n(x) \leq f_{X|\delta=1}(x)a_k(x)h^k(1+o(1))$$

Now, remaining part is the computation of the variance of $D_n(x)$. We start with writing the expression of variance of $D_n(x)$ as

$$
\begin{aligned}
VarD_n(x) &= \frac{1}{nh^2}Var\left(K^z\left(\frac{x-Y}{h}\right)[Z-m(x)]\right) \\
&= \frac{1}{nh^2}E\left(K^z\left(\frac{x-Y}{h}\right)^2[Z-m(x)]^2\right) \\
&- \frac{1}{nh^2}E\left(K^z\left(\frac{x-Y}{h}\right)[Z-m(x)]\right)^2 \\
&\leq \frac{1}{nh^2}E\left(K^z\left(\frac{x-Y}{h}\right)^2[Z-m(x)]^2\right) \qquad (2.57) \\
&= \frac{1}{nh^2}E\left(K^z\left(\frac{x-Y}{h}\right)^2[Z-m(x)]^2\right) \\
&\leq \frac{1}{nh^2}\int\int K^z\left(\frac{x-u-v}{h}\right)^2\tau^2(u)f_{X|\delta=1}(u)dF_E(v)du \\
&\leq \frac{1}{nh^2}\int\int(K^z(u))^2 f_{X|\delta=1}(x-u-vh)\tau^2(x-u-vh)dF_E(v)
\end{aligned}
$$

By using the definition of ordinary smooth distribution in (2.2) and the dominated convergence theorem, we have

$$K^z(x)h^\beta \rightarrow \frac{1}{2\pi c}\int e^{-itx}\varphi_K(t)t^\beta dt \triangleq H(x) \qquad (2.58)$$

By using lemma 2.4, it can be expressed

$$\left|K^z(x)h^\beta\right| \leq \frac{D}{1+|x|} \qquad (2.59)$$

for $D > 0$.

Note that we need the following lemma from Fan(1991b):

Lemma: Assume that $K^z(\cdot)$ is a sequence of Borel functions satisfying

$$K^z(x) \rightarrow K(x) \quad and \quad sup|K^z(x)| \leq K^*(x)$$

where $K^*(x)$ satisfies

$$\int K^*(x)dx < \infty. \quad and \quad \lim_{x\rightarrow\infty}|xK^*(x)| = 0$$

If y is a continuity point of a density $f(\cdot)$, then for any sequence $h \to 0$, we have

$$\lim_{n\to\infty} \frac{1}{h} \int K^z \left(\frac{y-x}{h}\right) f(x)dx = f(y) \int K(x)dx$$

So, using (2.58) and the lemma from Fan(1991b) stated right above, the variance of $D_n(x)$ can be written as

$$VarD_n(x) = \frac{1}{h^{2\beta+1}n} \int H^2(u)du \int f_{X|\delta=1}(x-v)\tau^2(x-v)dF_E(v)[1+o(1)] \quad (2.60)$$

Then, using the Parseval's identity,

$$\int H^2(u)du = \frac{1}{|c|2\pi} \int |\varphi_K(t)|^2 |t^\beta| dt \quad (2.61)$$

Therefore,

$$\begin{aligned} VarD_n(x) &= \frac{1}{2\pi n h^{2\beta+1}} \int |\varphi_K(t)|^2 \left|\frac{t^\beta}{c}\right| \int f_{X|\delta=1}(x-v)\tau^2(x-v)dF_E(v)[1+o(1)] \\ &= \frac{1}{h^{2\beta+1}n}\xi(x)[1+o(1)] \end{aligned} \quad (2.62)$$

Note that the variance of $D_n(x)$ in (2.62) approaches to 0 as $n \to \infty$.

We now use the "bias"-variance decomposition to conclude the proof.

$$\begin{aligned} E\left((\hat{m}(x)-m(x))\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 &= E\left(\frac{D_n(x)}{f_{X|\delta=1}(x)}\right)^2 \\ &= \frac{\mathrm{Var}D_n(x)}{f_{X|\delta=1}(x)} + \frac{E(D_n(x))^2}{f^2_{X|\delta=1}(x)} \\ &= \frac{\mathrm{Var}D_n(x)}{f_{X|\delta=1}(x)} + \frac{f^2_{X|\delta=1}(x)a_k^2(x)h^{2k}(1+o(1))}{f^2_{X|\delta=1}(x)} \\ &= \frac{\mathrm{Var}D_n(x)}{f_{X|\delta=1}(x)} + a_k^2(x)h^{2k}(1+o(1)) \\ &= \frac{1}{h^{2\beta+1}n}\xi(x)[1+o(1)] + a_k^2(x)h^{2k}(1+o(1)) \end{aligned} \quad (2.63)$$

Therefore,

$$E\left((\hat{m}(x)-m(x))\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 \leq \left[a_k^2(x)h^{2k}+\frac{1}{h^{2\beta+1}n}\xi(x)\right][1+o(1)]$$
$$= O(n^{\frac{-2k}{2k+2\beta+1}}) \tag{2.64}$$

Furthermore, as (2.16) and (2.62) hold uniformly in $x \in (a,b)$, the second expression in the Theorem 2 is also accurate.

$$E\int_a^b\left([\hat{m}(x)-m(x)]\frac{\hat{f}_X^{IP}(x)}{f_{X|\delta=1}(x)}\right)^2 = \left[a_k^2h^{2k}+\frac{1}{nh^{1+2\beta}}\xi(x)\right](1+o(1))$$
$$= O(n^{\frac{-2k}{[2k+2\beta+1}}),$$

Proof of Theorem 2 is completed.

CHAPTER 3

SIMULATION

3.1 Introduction

This chapter describes a simulation experiment that was used to test the behaviour of the deconvolved kernel estimate (1.14) by providing two different examples. X is a normal random variables in these examples, and it is observed through $Y = X + E$. The error E's variance σ_0^2 is selected so the reliability ratio [12] given by

$$r = \frac{Var(Y)}{\sigma_E^2 + Var(Y)} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_E^2} \approx \frac{s_X^2 - \sigma_E^2}{\sigma_X^2} = 0.7 \tag{3.1}$$

The convergence rate of estimators in the presence of super- smooth errors is slower than in the presence of ordinary smooth errors for the complete data, according to Fan [4,17]. By considering this conclusion from Fan, E is taken to be a normal and a double exponential random variable in order to investigate the effect of error distributions on the MSE of the estimator (1.14) when the right-censored data is used. In this simulation study, we consider two different regression function:

$$m_1(x) = 2x - 2 \qquad m_2(x) = sin(-4x)sin\left(\frac{11}{10}x\right) \tag{3.2}$$

3.2 Data generation

Our model is

$$Z = m(X) + \varepsilon$$

where $(X_1, ..., X_n)$ is unobservable and right censored data which is contaminated in the following

$$Y = X + E$$

Y is generated using the Weibull distribution with shape parameter 3.4 and scale parameter 1.1. E is taken as super smooth and ordinary smooth error distributions with appropriate shape and scale parameters as well. Our censoring variable C is also generated using Weibull distribution. Note that $W = min(Y, C)$. For each sample, 100 replications were performed. Table 1 and 2 represents the average of these optimal ASE's in those 100 replications. The percentage of censored data is about 45 percent for each replications.

Example 1(Truncated Double Exponential errors):

First, we suppose E has a truncated double exponential distribution, which corresponds to the ordinary smooth case, as we want E to have non-negative values.

$$f_E(u) = \frac{1}{\sigma_E} e^{\frac{-u}{\sigma_E}} \tag{3.3}$$

where u is non-negative. The characteristic function of E can be derived as following,

$$\begin{aligned}
\varphi_{f_E}(t) &= \int_0^\infty e^{itu} f_E(u) du \\
&= \frac{1}{\sigma_E} \int_0^\infty e^{itu} e^{\frac{-u}{\sigma_E}} du \\
&= \frac{1}{\sigma_E} \int_0^\infty cos(tu) e^{\frac{-u}{\sigma_E}} du + i \frac{1}{\sigma_E} \int_0^\infty sin(tu) e^{\frac{-u}{\sigma_E}} du \\
&= \frac{1}{\sigma_E} \frac{\frac{1}{\sigma_E}}{\left(\frac{1}{\sigma_E}\right)^2 + t^2} + i \frac{\frac{1}{\sigma_E}}{\left(\frac{1}{\sigma_E}\right)^2 + t^2} \\
&= \frac{1}{1 - i\sigma_E t}
\end{aligned} \tag{3.4}$$

From (1.12), the deconvolved kernel estimator can be derived as following

$$
\begin{aligned}
K^{Z}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iyx}\frac{\varphi_K(y)}{\varphi_{f_E}(\frac{y}{h})}dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iyx}\varphi_K(y)\left(1 - i\sigma_E\frac{y}{h}\right)dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iyx}\varphi_K(y) + \frac{\sigma_E}{h}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iyx}(-iy)\varphi_K(y)dy \\
&= K(x) + \frac{\sigma_E}{h}K^{'}(x)
\end{aligned}
\tag{3.5}
$$

Example 2(Normal Error):

In practice, the normal distribution $N(0, \sigma_0^2)$ is the most widely used error distribution [23]. For normal errors, the kernel functions often suggest two options. The first is the second-order kernel below, which has a compact and symmetric support for its characteristic function [1,7].

$$
K(x) = \frac{48cosx}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144sin(x)}{\pi x^5}\left(2 - \frac{5}{x^2}\right) \tag{3.6}
$$

Its characteristic function is

$$
\phi_K(t) = (1 - t^2)^3 I_{[-1,1]}(t) \tag{3.7}
$$

where $I_{[-1,1]}(t)$ is the indicator function. As a result, the deconvoluting kernel with normal error that results is

$$
K^{z}(x) = \frac{1}{\pi}\int_0^1 cos(tx)(1 - t^2)^3 exp\left(\frac{\sigma_0^2 t^2}{2h^2}\right)dt \tag{3.8}
$$

When the error variance in Gaussian deconvolution is minimal, the necessity for this support kernel can be reduced. Fan(1992) discussed the impact of the error magnitude on the deconvolution kernel methods in depth. When the standard normal density is selected as the kernel function, the corresponding deconvolving kernel becomes

$$K^Z(x) = \frac{1}{\sqrt{2\pi\left(1-\frac{\sigma_0^2}{h^2}\right)}} exp\left(-\frac{x^2}{2(1-\frac{\sigma_0^2}{h^2})}\right) \quad (3.9)$$

3.3 Discussion & Results

We use 2 different kernels for each model: (3.5) and (3.9). This is used to see how much can be gained by employing deconvolution and to evaluate the deconvolved kernel's robustness. Tables below are provided the average mean squared error for different sample sizes in 100 replications. As the sample size increases, the average mean squared error for the normal error decreases slower than for the truncated double exponential error. Furthermore, the average mean squared error for the normal error is slightly higher than the truncated double exponential error.

It is widely known that in nonparametric estimation, the choice of the kernel K has little impact on optimality (in the MSE sense), while the choice of the bandwidth h does. In this simulation, the optimal bandwidth selection is far more important than the kernel selection. In the selection of the bandwidth, we follow the similar way as in Fan(1992). In order to calculate the average squared error (ASE) at 91 grid points from 0.6 to 1.2 using a uniform sequence of 22 bandwidths ranging in [0.07,0.13] for $m_1(x)$ and [0.06,0.12] for $m_2(x)$. We have chosen the optimal bandwidth that minimizes the ASE between those 22 different bandwidth values. Note that Assumption A1 is obviously satisfied by the bandwidth selection. The average of these optimum ASE's in 100 replications is reported in Tables 1 and 2.

	Double Exponential Error			Normal Error		
Kernel	**n=250**	**n=500**	**n=1000**	**n=250**	**n=500**	**n=1000**
(3.5)	0.04895002	0.0400233	0.0233831	0.0527420	0.0425208	0.0277523
(3.9)	0.0449263	0.0383484	0.0257986	0.0448071	0.0422564	0.0280653

Table 1: The ASE for estimating the model: $m_1(x) = 2x - 2$

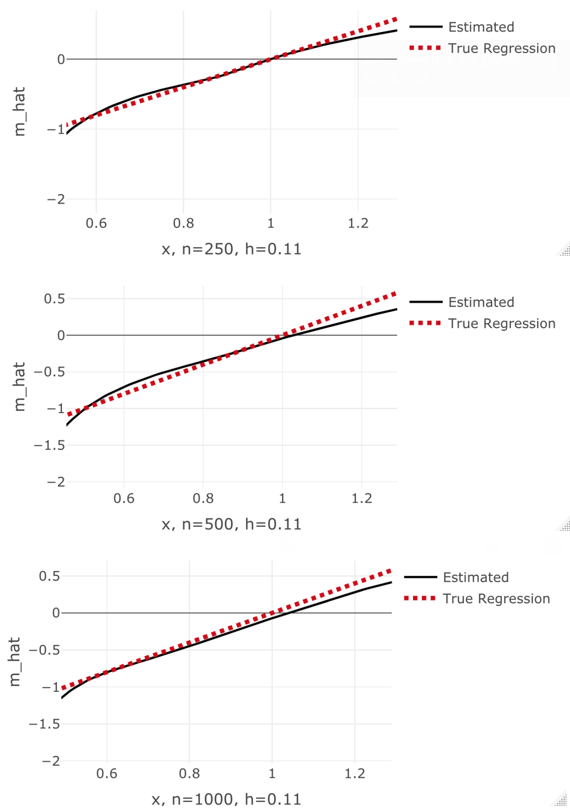

Figure 3.1. Super Smooth Case for different sample sizes ($m_1(x)$).

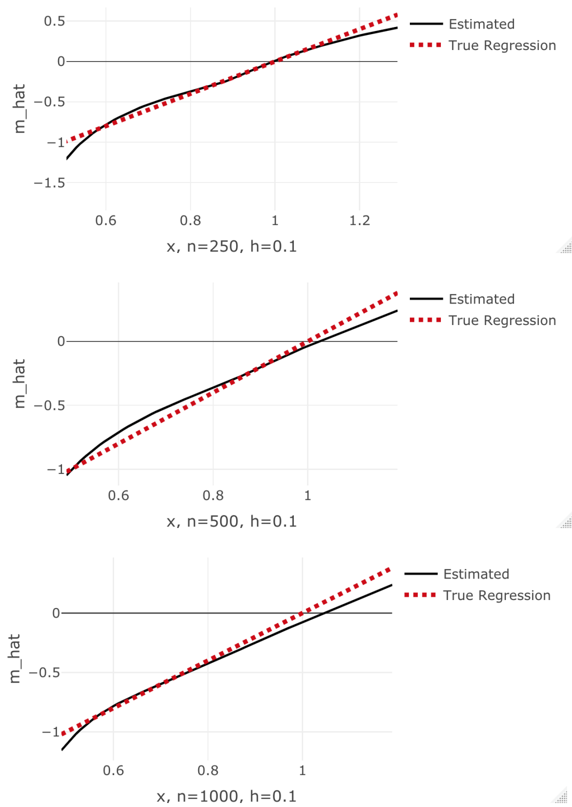

Figure 3.2. Ordinary Smooth Case for different sample sizes ($m_1(x)$).

Figure 3.1 and 3.2 show that our estimator's performance is very good with the bandwidth selection 0.11. As we increase the sample size, the quality of fit is better. Obviously, when the error distribution is ordinary smooth, we get better results but not that significant.

	Double Exponential Error			Normal Error		
Kernel	**n=250**	**n=500**	**n=1000**	**n=250**	**n=500**	**n=1000**
(3.5)	0.04063852	0.01823147	0.0139942	0.04226132	.03843053	0.01563360
(3.9)	0.03905392	0.0224311	0.0158492	0.0428536	0.0384325	0.0141332

Table 2: The ASE for estimating the model: $m_2(x) = \sin(-4x)\sin(\frac{11}{10}x)$

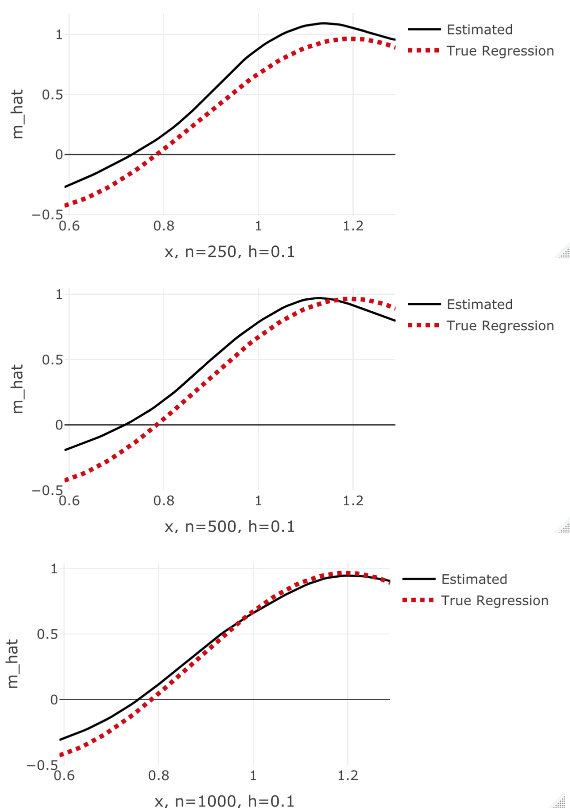

Figure 3.3. Ordinary Smooth Case for different sample sizes($m_2(x)$).

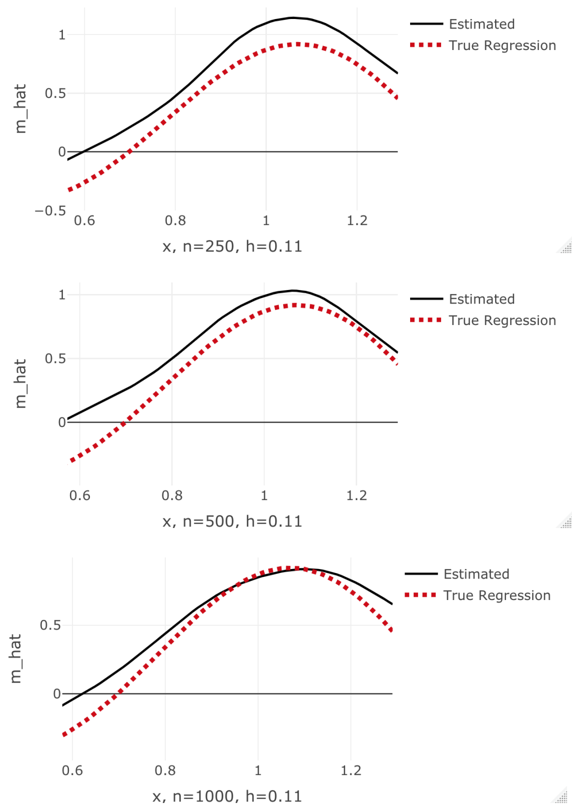

Figure 3.4. Super Smooth Case for different sample sizes $(m_2(x))$.

Figures 3.3 and 3.4 indicate that the nonlinear model's fit quality is comparable to that of the linear model. It is again better when we use ordinary smooth error distribution. As the sample size increase, the quality of fit also increase for the nonlinear regression function estimation.

CHAPTER 4

CONCLUSION AND FUTURE STUDIES

The new kernel regression function estimator is constructed by using the ordinary kernel estimator and the the idea of deconvolution in density estimation for right censored data for contaminated independent variable. We show that there are two types of convergence rates in adjusted mean square error for the regression function estimator, depending on whether the error is ordinary smooth or super smooth. The rate of convergence for our estimator is faster in terms of mean square error when we use the ordinary smooth error distribution for both locally and globally whereas it is slower in super smooth model. The choice of kernel K has very small effect on the optimality as expected; however, the bandwidth selection is very crucial. Depending on the choice of bandwidth, there might be some issues such as overestimating, underestimating etc. In order to find the optimal bandwidth, we use a proper interval and try different bandwidth to find the best one. The simulation study is done to represent our estimator's performance. We use different sample sizes to see the what happens when we use large samples and notice that the quality of fit increases with large n.

The work in this thesis could possibly be used to extend the necessary conditions for the Central Limit Theorem to hold for $\hat{m}(x)$ and the asymptotic normality of the estimator $\hat{m}(x)$. The quality of estimator might be increased by using some techniques in the selection of bandwidth. These techniques are such as cross validation, bootstrapping, Risk Estimation using Classical Pilots (RECP), Improved Akaike Criterion (AICc) etc.

REFERENCES

[1] Randall L. Eubank, Nonparametric Regression and Spline Smoothing, Second Edition, Deparment of Statistics Texas AM University, ISBN, -08247-9337-4, (1988).

[2] Chakrabarty, T. K. "Deconvolving Kernel Density estimation of right censored duration data with recall errors." *American Journal of Applied Mathematics and Statistics* 2.6 (2014): 416-422.

[3] D.A.Ioannides&P.D.Alevizos,(1996),Nonparametric regression with errors in variables and applications *Statistics Probability Letters 32 (1997) 35-43*

[4] Fan, Jianqing & Young, Troung. "Nonparametric Regression with errors in variables." *The Annals of Statistics, (1993),Vol. 21,No. 4, 1900-1925*

[5] Fan, Jianqing. "Asymptotic normality for deconvolution kernel density estimators." *Sankhyā: The Indian Journal of Statistics* Series A (1991): 97-110.

[6] Guessoum,Zohra & Ould-Said, Elias. "On nonparametric estimation of the regression function under random censorship model." *Statistics Decisions 26, 159–177 (2008) / DOI 10.1524/stnd.2008.0919*

[7] Delaigle A, Gijbels I. Bootstrap Bandwidth Selection in Kernel Density Estimation from a Contaminated Sample. Annals of the Institute of Statistical Mathematics. 2004a;56(1):19–47.

[8] Aurore Delaigle, Jianqing Fan Raymond J. Carroll (2009) A Design-Adaptive Local Polynomial Estimator for the Errors-in-Variables Problem, Journal of the American Statistical Association, 104:485, 348-359, DOI: 10.1198/jasa.2009.0114

[9] Satten, Glen A., and Somnath Datta. "The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average." *The American Statistician* 55.3 (2001): 207-210.

[10] Stefanski, Leonard A., and Raymond J. Carroll. "Deconvolving kernel density estimators." *Statistics* 21.2 (1990): 169-184.

[11] Wand, Matt P., and M. Chris Jones. *Kernel smoothing.* Chapman and Hall/CRC, (1994).

[12] Wayne A. Fuller, Measurement Error Models,ISBN 0-471-86187- 1 , ,(1987).

[13] Delaigle, A. (2021). Deconvolution kernel density estimator. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), Handbook of Measurement Error Models, Chapter x, pp. xxx–xxx. CRC.

[14] Yılmaz, Ersin Aydın, Dursun, Estimation of Right Censored Nonparametric Regression Solved by kNN Imputation: A Comparative Study. Turkiye Klinikleri Journal of Biostatistics. 11. 83-92. 10.5336/biostatic.2019-66285, (2019).

[15] Green, P. J., Silverman, B. W., Nonparametric Regression and Generalized Linear Models, Chapman Hall, London , (1994).

[16] Miller, R., Halpern, J., Regression With Censored Data, Biometrika, Vol. 69, 521-531, (1982).

[17] Zheng, Z.K., Regression Analysis with Censored Data, Ph.D Dissertation, University Of Colombia, (1984).

[18] Lee, T.C.M., Solo, V. , Bandwidth Selection for Local Linear Regression: A Simulation Study, Computational Statistics and Data Analysis, Vol. 42, 139-148, (1999).

[19] Dursun Aydin, Ersin Yilmaz. Nonparametric Regression with Randomly Right-Censored Data. International Journal of Mathematical and Computational Methods, 1, 186-189, (2016).

[20] Nadaraya, E. A. , On Estimating Regression, Theory Of Probability Its Applications, Vol. 9(1), 141-142, (1964).

[21] Wenghing, Zhu. Asymptotic Normality of the Deconvolution Kernel Density Estimators Bases on Independent as well as Strong Mixing Right Censored Data, PhD Thesis, The University of Texas at Arlington, United States, 2020

[22] Wang, Xiao-Feng, and Bin Wang. "Deconvolution Estimation in Measurement Error Models: The R Package decon." Journal of statistical software vol. 39,10 (2011): i10.

[23] Charles J. Stone. "Consistent Nonparametric Regression." Ann. Statist. 5 (4) 595 - 620, July, 1977. https://doi.org/10.1214/aos/1176343886

[24] Du, Lilun Zou, Changliang Wang, Zhaojun. (2011). Nonparametric Regression Function Estimation for Errors-in-Variables Models with Validation Data. Statistica Sinica. 21. 10.5705/ss.2009.047.

[25] Jianqing Fan and Elias Masry. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. Journal of multivariate analysis, 43(2):237-271, 1992a.

[26] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282):457-481,1958.

BIOGRAPHICAL STATEMENT

Erol Ozkan was born in Gebze, Kocaeli, Turkey in 1989. He received his B.S. degree in Mathematics from University of Ataturk in 2011, his master's degree is from the University of Nottingham, UK in 2016. Erol has been a PhD student in University of Texas at Arlington in Statistics since 2018. During his PhD education, he received the Science Dean's Excellence Scholarship, Micheal B. and Wanda G. Ray Scholarship, Summer Research Fellowship and Summer Dissertation Fellowship.