2021

Parameter Estimation for the Logistic Regression Model with Errors in Covariate

Huyen Dieu Nguyen
Illinois State University, dieuhuyen0708@gmail.com

PARAMETER ESTIMATION FOR THE LOGISTIC REGRESSION MODEL
WITH ERRORS IN COVARIATE

HUYEN D. NGUYEN

54 pages

In a logistic regression model, when the covariate is measured with error, the estimators of the regression coefficient parameters can be biased. We propose a method for estimating parameters of a logistic regression with case-control data, when the covariate is subject to measurement error. The density of the covariate is estimated by using the deconvolution kernel density estimation. The parameters of the regression are estimated by the integrated squared distance based on the log ratio of the estimated density. We show the consistency and the asymptotic normality of the proposed estimators. Simulation study shows the superiority of the proposed method in different sample sizes and measurement error magnitudes scenario compared to method in which the measurement error was ignored. The methodology is applied to estimating the relationship of systolic blood pressure and the presence of coronary heart disease.

PARAMETER ESTIMATION FOR THE LOGISTIC REGRESSION MODEL

WITH ERRORS IN COVARIATE

HUYEN D. NGUYEN

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Mathematics

ILLINOIS STATE UNIVERSITY

2021

PARAMETER ESTIMATION FOR THE LOGISTIC REGRESSION MODEL

WITH ERRORS IN COVARIATE

HUYEN D. NGUYEN

COMMITTEE MEMBERS:

Pei Geng, Chair

Maochao Xu

ACKNOWLEDGMENTS

CONTENTS

TABLES

FIGURES

CHAPTER I: INTRODUCTION

Logistic regression is one of the most commonly applied methods for binary repsonse. In the model, the binary response Y relates to covariate X and the conditional mean function of Y given X is

$$E(Y|X = x) = \frac{\exp(\alpha^* + \beta x)}{1 + \exp(\alpha^* + \beta x)}. \tag{1}$$

The maximum likelihood estimator is often used in practice to estimate parameters α^* and β. It is asymptotically normal and consistent subject to regularity condition. However, the quality of the estimation can be degraded when the covariate X may not be accurately measured. Stefanski and Carroll (1985) showed that ignoring measurement errors may result in biased estimators. In this paper, we discuss a method to estimate logistic regression parameters when covariate is contaminated.

The motivation to study this problem stems from medical research and health science. An example of this problem is studying the presence of a disease and its relationship with a risk factor, such as blood pressure or dietary intake, that is difficult to measure and subject to measurement error. The error could be caused by both measuring instrument variability and biological fluctuation of human body. In fact, pioneer papers in addressing this problem such as Carroll et al. (1984) and Stefanski and Carroll (1985) were both motivated by Framingham Heart Study, which investigated the development of cardiovascular disease. The study was based on logistic regression with response Y being the presence pr absence of heart disease and covariate X being risk factors such as systolic blood pressure, serum cholesterol,etc. Systolic blood pressure measure, for example, is subject to substantial error. While a person's true blood pressure is defined as long-term average, the readings can be

affected by the blood pressure cuff placement, the person's stress level, the amount of time after eating at the time of the readings. Yi et al. (2018) studied the calcium intake in relation to distal colon cancer. Dietary intake is also a variable that is difficult to measure accurately due to the variability of the food source and the amount of food consumed. Therefore, it is crucial to develop methods that can account for the error.

The method for logistic regression parameter estimation that we discuss in this paper is under the case-control framework. This work is based on the paper by Geng and Sakhanenko (2016) in which the authors proposed the method for when the covariate sample is free of error. Let $X_{0i}, i = 1, ..., n_0$ be the true measure of the covariate X in the control group $(Y_i = 0)$ and $X_{1j}, j = 1, ..., n_1$ be the true measure of the covariate X in the case group $(Y_i = 1)$. Let n be the total sample size of the data such that $n = n_1 + n_0$. Then,

$Y = 0 : X_{01}, X_{02}, ..., X_{0n_0}$ is a random sample from a random variable with density $f_0(x)$;

$Y = 1 : X_{11}, X_{12}, ..., X_{1n_1}$ is a random sample from a random variable with density

$$f_1(x) = \exp(\alpha + \beta x) f_0(x) \text{ where } \alpha = \alpha^* + \ln \frac{P(Y = 0)}{P(Y = 1)}.$$

$$(2)$$

Under this framework, the logarithm ratio of the true densities f_0, and f_1 is the straight line $\alpha + \beta x$. The estimated logarithm ratio of the densities $\ln\{\hat{f}_1(x)/\hat{f}_0(x)\}$ is approximately a straight line and the idea of the method is to have the distance between this line and the true $\alpha + \beta x$ line minimized. Geng and Sakhanenko (2016) estimated the densities f_0 and f_1 using the kernel density estimation method. The authors derived the logistic regression parameter estimators by minimizing the integrated square distance of the estimated logarithm

2

ratio of the densities and the true one. They established the consistency and the asymptotic normality of the estimators and investigated its performance compared with some other existing methods. The advantages of this method is that it does not require assumptions on the distribution of the covariate and the estimator has closed form expression. The simulation study showed the method outperforms when the sample sizes are small and severely unbalanced.

In this paper, we consider the case in which instead of observing the true measures of the covariate X, we observe the data contaminated with error $U_{0i}, i = 1, ..., n_0$ and $U_{1j}, j = 1, ..., n_1$ such that $U_{0i} = X_{0i} + \varepsilon_{0i}$ and $U_{1j} = X_{1j} + \varepsilon_{1j}$ with $\{\varepsilon_{01}, \varepsilon_{02}, ..., \varepsilon_{0n_0}\}$ and $\{\varepsilon_{11}, \varepsilon_{12}, ..., \varepsilon_{1n_1}\}$ be random samples of known distributions.

There is a wealth of literature on the covariate measurement error problem in logistic regression. As mentioned, pioneer works include Carroll et al. (1984) and Stefanski and Carroll (1985). Carroll et al. (1984) proposed a structural model based on the maximum likelihood of the response Y_i conditioned on the observed covariate. Both covariate and the measurement error were assumed to be independent normally distributed with known variance. The author considered a particular case of the structural model in which the regression is probit due to the difficulty of the calculation of the likelihood for logistic function, and suggested that both logistic and probit regression often result in similar estimates. Schafer (1987) followed this suggestion and proposed the Expectation-Maximization (EM) algorithm to compute the estimates of the logistic regression. In the algorithm, the maximum likelihood estimation of β is estimated by iterative reweighted least squares. The nuisance parameters of conditional density of true covariate given the errors are estimated using the validation data, then used to calculate the likelihood function and the estimated β is obtained by

maximizing the likelihood function over the primary data. Stefanski and Carroll (1985) proposed a functional maximum likelihood estimator. The errors are normally distributed with covariance matrix $\sigma^2 \Sigma$ and the log-likelihood for estimating β is

$$\sum_1^n \{Y_i \log(F(X_i^T \beta)) + (1 - Y_i \log(F(-X_i^T \beta)) - (2\sigma^2)^{-1}(U_i - X_i)^T \Sigma^{-1}(U_i - X_i)\}, \quad (3)$$

in which F is the logistic function defined by $F(t) = (1 + \exp(-t))^{-1}$. The estimator is obtained by maximizing the log-likelihood function. Amstrong et al. (1989) proposed a method based on the normal discriminant analysis model for true covariate values in cases and control. Similar to other methods, it requires the assumptions that both the true covariate and the error are from normal distributions.

In this paper, we propose a modification of the minimum integrated square error method for bias correction when the covariate is measured with error by utilizing the deconvolution kernel density method to estimate the densities f_1 and f_0 of covariate of case and control groups. Deconvolution kernel density estimation is a well-studied nonparametric density estimation method when the random variable is observed with independent additive error of a known distribution. The method was proposed by Carroll and Hall (1988) and Stefanski and Carroll (1990). Theoretical properties of the estimator has been extensively researched. Devroye (1989), Carroll and Hall (1989), Stefanski (1990), Fan (1991a), and Fan (1992) studied the consistency and rate of convergence of the estimator, and Fan (1991b) studied its asymptotic normality. Many of these properties depend on the smoothness of the error distribution, which is characterized by the rate of decay of its characteristic function in the tails. The distribution is called ordinary smooth if the rate of decay is geometric, and

super smooth if the rate is exponential. In this paper, we focus on the the case in which the error distribution is ordinary smooth.

The paper is organized as follows. In chapter 2, we introduce the background information on the deconvolution kernel density estimation and the integrated square distance estimators for logistic regression parameters. In chapter 3, we state the needed assumptions and the result on consistency and asymptotic normality of the estimator. Chapter 4 shows simulation study and real data application. In chapter 5, we discuss some relevant issues of the estimator. Detailed proofs of the theoretical results are presented in the Appendix.

CHAPTER II: BACKGROUND

II.1 Kernel Density Estimator

In this section, we briefly introduce the measurement error model and the deconvolution kernel density estimator. We begin with the overview of the ordinary kernel density estimation when the sample is free of error. Suppose we observe the error-free data $X_{11}, X_{12}, ..., X_{1n_1}$ from the case group and $X_{01}, X_{02}, ..., X_{0n_0}$ from the control group. X_{1i}'s are independent and identically-distributed random variables from the distribution with density f_1 and X_{0i}'s are independent and identically-distributed random variables from the distribution with density f_0 . The naive kernel density estimator is defined as

$$\tilde{f}_1(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{h_1} K\left(\frac{x - X_{1i}}{h_1}\right), \ \ \tilde{f}_0(x) = \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{1}{h_0} K\left(\frac{x - X_{0i}}{h_0}\right), \tag{4}$$

in which $K(x)$ is a kernel function that is generally smooth and satisfying

$$K(x) = K(-x); \ \int K(x)dx = 1; \ \lim_{x \to -\infty} K(x) = \lim_{x \to \infty} K(x) = 0; \tag{5}$$

and $h_i > 0$, $i = 0, 1$ is the smoothing parameter or bandwidth. The idea of the kernel density estimation is to smooth each data point into small density "bumps" then add these bumps to obtain the final density estimate. Figure (1) illustrates the idea of the kernel density estimation with 10 data points drawn from two normal distributions, 5 data points from the normal distribution with mean 0 and standard deviation 0.5 and 5 data points from the normal distribution with mean 0 and standard deviation 0.5, and using Gaussian kernel and bandwidth 0.3. The bumps are center at each data point; their shape is determined

by the kernel function and their width is determined by the bandwidth. Figure (2) displays the shapes of three commonly used kernels in the naive kernel density estimator: Gaussian (standard normal), Epanechnikov, and Uniform. Figure (3) illustrates an example of how bandwidth choice affects the quality of the density estimator. A small bandwidth results in an undersmoothed density estimate, which contained too many bumps, whereas a large bandwidth results in an oversmoothed density estimate, which mask underlying features of the true density.
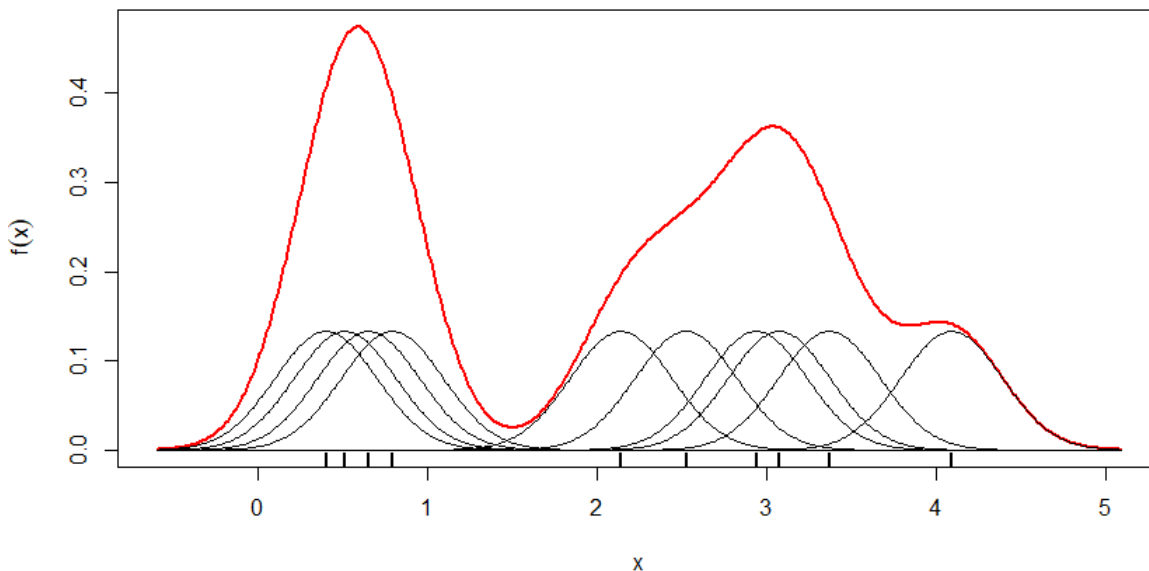
Figure 1: An example of kernel density estimation

We now introduce the deconvolution kernel density estimation. Assume that we observe the contaminated data $U_{11}, U_{12}, ..., U_{1n_1}$ for the case group and $U_{01}, U_{02}, ..., U_{0n_0}$ for the control group, where

$$U_{1i} = X_{1i} + \varepsilon_{1i}, \ i = 1, ...n_1, \ U_{0j} = X_{0j} + \varepsilon_{0j}, \ j = 1, ...n_0. \tag{6}$$
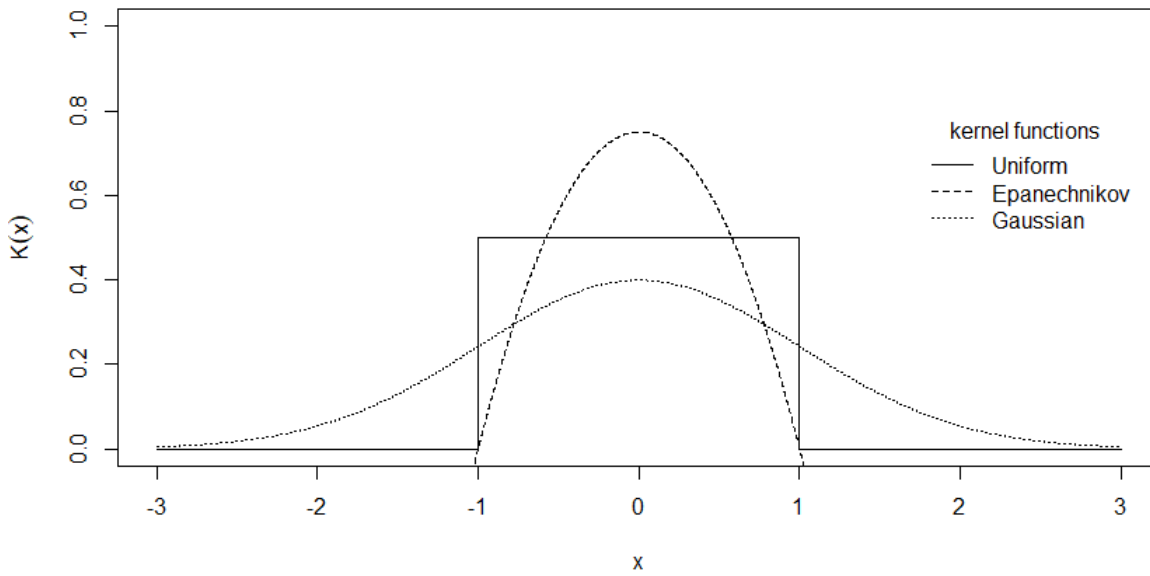
Figure 2: Commonly used kernels for kernel density estimation
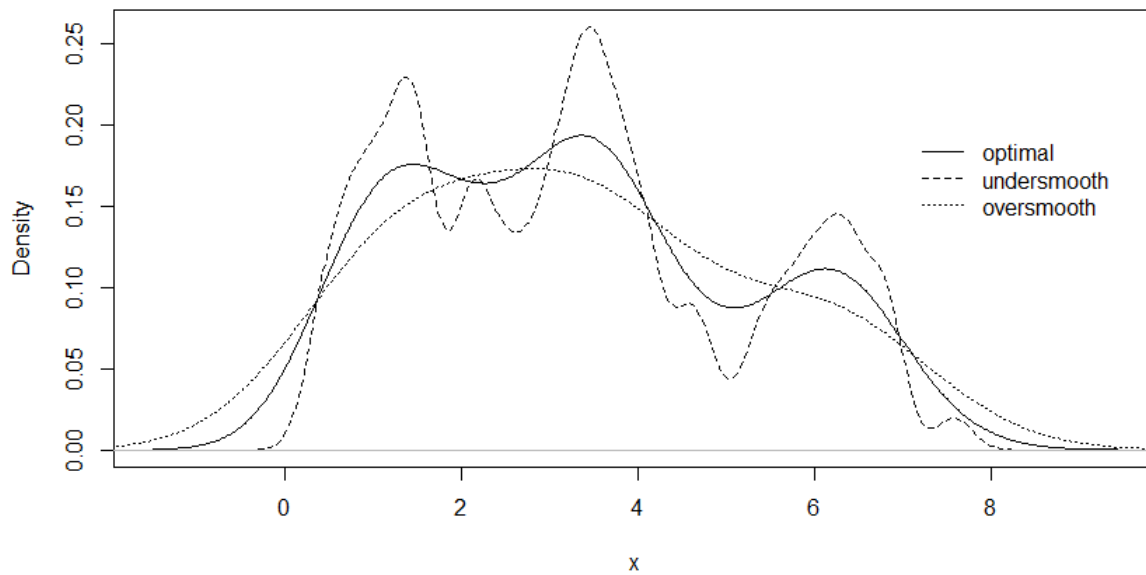
Figure 3: Comparing the kernel density estimator using different bandwidths

8

We introduce the definition of the deconvolution kernel density estimator for the case group distribution density. The deconvolution kernel density estimator of the baseline distribution density can be defined in the similar fashion. Aside from assuming that X_{1i}'s are independent and identically-distributed (i.i.d.), we further assume the errors ε_i's are i.i.d., and X_{1i}'s and ε_{1i} are mutually independent. The density function of ε_{1i}'s, denoted f_ε, is assumed known. We wish to estimate the density function f_1 of X_{1i}'s, based on the contaminated observations U_{1i}'s.

The deconvolution kernel density estimator for the unknown density f_1 of X_{1i}'s using Fourier transform was proposed by Carroll and Hall (1988) and Stefanski and Carroll (1990), is defined by

$$\hat{f}_1(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \frac{\phi_K(th_1)\hat{\phi}_n(t)}{\phi_\varepsilon(t)} dt, \tag{7}$$

where $\hat{\phi}_n(t) = n_1^{-1} \sum_{i=1}^{n} \exp(itU_{1i})$ is the empirical characteristic function of observed data $\{U_{1i}\}_{i=1}^{n_1}$, ϕ_K is the characteristic function of kernel K, and ϕ_ε is the characteristic function of the error density f_ε. Then (7) can be written as

$$\hat{f}_1(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{h_1} K^* \left(\frac{x - U_{1i}}{h_1} \right), \tag{8}$$

where

$$K^*(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \frac{\phi_K(t)}{\phi_\varepsilon(t/h)} dt. \tag{9}$$

The following conditions are needed to guarantee the estimator is well-defined

$$|\phi_\varepsilon(t)| > 0, \text{ for all real } t; \ \sup_t |\phi_K(t)/\phi_\varepsilon(t/h_1)| < \infty; \ \int |\phi_K(t)/\phi_\varepsilon(t/h_1)|dt < \infty. \tag{10}$$

This also implies that $K^*(x)$ is bounded. According to Stefanski and Carroll (1990), $K^*(x)$ possesses many properties of the naive kernel: $\int K^*(x)dx = 1$; when ϕ_K is real and even and f_ε is even, K^* is also real and even. Thus, it is called deconvolution kernel.

The difficulty of deconvolution depends on the smoothness of the error distribution. Smoothness is a measure determining how many times the density function can be differentiated, which equivalent to the rate of decay of its characteristic function at its tails. In deconvolution literature, the smoothness of the error distribution is classified into two classes: ordinary smooth and super smooth (Fan 1991). A random variable ε is ordinary smooth of order τ if for some constants $0 < d_0 \leq d_1$ and $\tau > 0$,

$$d_0|t|^{-\tau} \leq |\phi_\varepsilon(t)| \leq d_1|t|^{-\tau} \quad \text{as} \quad t \to \infty. \tag{11}$$

The examples of ordinary smooth distributions are gamma and Laplace distribution. A random variable ε is super smooth of order τ if for some constants $\tau_0 \leq \tau_1$, $0 < d_0 \leq d_1$, $\tau > 0$ and $\gamma > 0$,

$$d_0|t|^{\tau_0}\exp(-|t|^\tau/\gamma) \leq |\phi_\varepsilon(t)| \leq d_1|t|^{\tau_1}\exp(-|t|^\tau/\gamma) \quad \text{as} \quad t \to \infty. \tag{12}$$

The examples of super smooth distributions are normal and Cauchy distributions. Figure (4) shows the density function of a Laplace and a normal distribution. Both distributions in the figure have the same mean 0 and the same variance 4. Figure (5) shows their corresponding characteristic function. Laplace distribution is an example of ordinary smooth distribution, whereas normal distribution is an example of super smooth distribution. At the tails, the

decay rate of the characteristic function of the normal distribution is much faster than that of the characteristic function of the Laplace distribution. The smoother the error distribution the more difficult the deconvolution is. As mentioned, in this paper, we will focus on the case in which the covariate data is contaminated with ordinary smooth error.
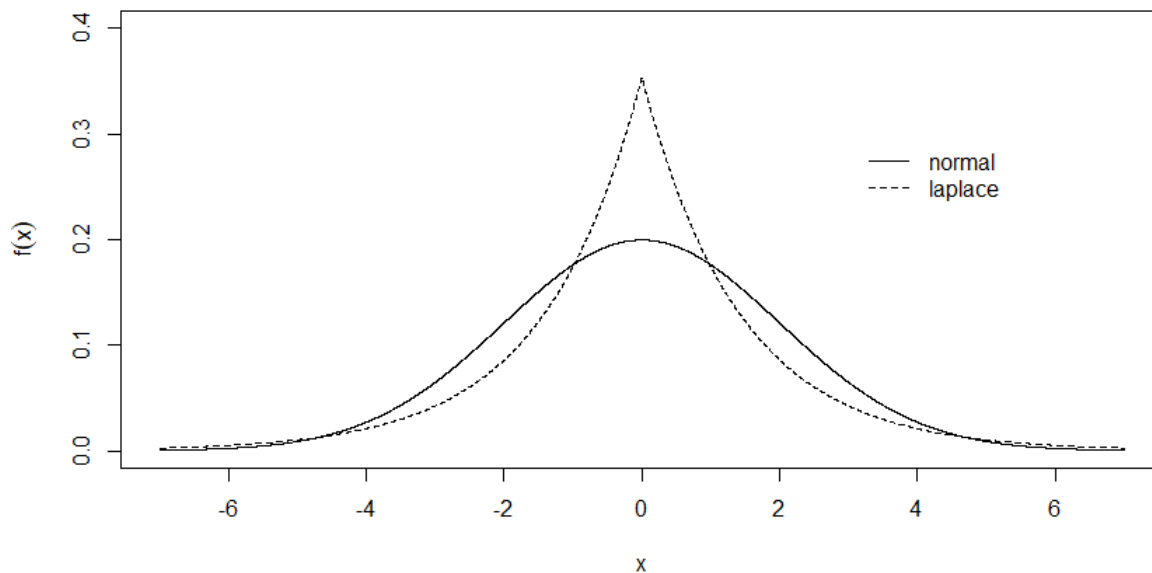
Figure 4: Probability density functions of a Laplace and a normal distribution, both with mean 0 and variance 4

II.2 Definition of the Integrated Square Distance Estimator

In this section, we introduce the minimum integrated square distance estimator of parameters in model (2). The estimator was derived by Geng and Sakhanenko (2016). Under model in (2), we have $\ln[f_1(x)/f_0(x)] = \alpha + \beta x$ The estimators of $f_0(x)$ and $f_1(x)$ are defined as follows.
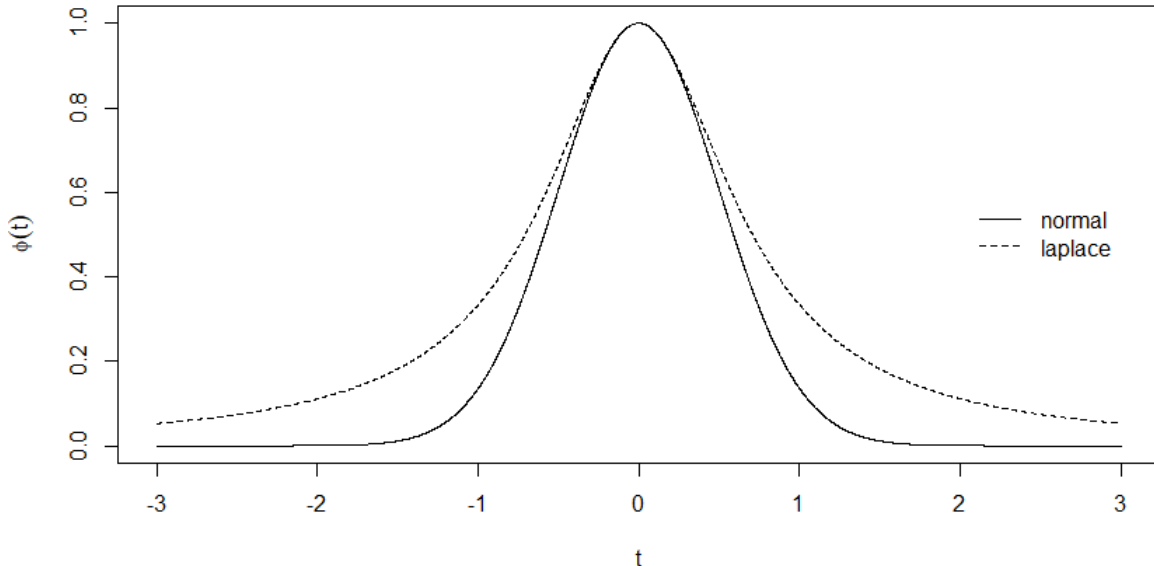
Figure 5: Characteristic functions of a Laplace and a normal distribution, both with mean 0 and variance 4

$$\hat{f}_{n_0}(x) = \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{1}{h_0} K^* \left(\frac{x - U_{0i}}{h_0} \right), \ \hat{f}_{n_1}(x) = \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{1}{h_1} K^* \left(\frac{x - U_{1j}}{h_0} \right), \quad (13)$$

where K^* is the deconvolution kernel defined as in (9). However, deconvolution kernel density estimators are not always positive, similar to the naive kernel density estimators. In fact, the deconvolution kernel lacks the nonnegativity property. Figure (6) shows an example of $K^*(x)$ that takes negative values. Thus, a small term b_{n_i} as a function of n_i is added to the density estimators so that the log ratio estimate is well-defined. The log ratio estimate becomes $\ln\{(\hat{f}_{n_1}(x) + b_{n_1})/(\hat{f}_{n_0}(x) + b_{n_0}\}$ and the integrated square distance is defined as

$$T_n(s,t) := \int_a^b \left[\ln \left\{ \frac{\hat{f}_{n_1}(x) + b_{n_1}}{\hat{f}_{n_0}(x) + b_{n_0}} \right\} - s - tx \right]^2 dx, \ s, t \in \mathbb{R}. \quad (14)$$

12

Geng and Sakhanenko (2016) derived $\hat{\alpha}$ and $\hat{\beta}$ by minimizing the integrated square distance and obtained

$$
\begin{aligned}
\hat{\beta} &= \frac{12}{(b-a)^3} \int_a^b \left[\ln \left\{ \frac{\hat{f}_{n_1}(x) + b_{n_1}}{\hat{f}_{n_0}(x) + b_{n_0}} \right\} \right] \left\{ x - \frac{a+b}{2} \right\} dx, \\
\hat{\alpha} &= \frac{1}{b-a} \int_a^b \left[\ln \left\{ \frac{\hat{f}_{n_1}(x) + b_{n_1}}{\hat{f}_{n_0}(x) + b_{n_0}} \right\} - \hat{\beta}x \right] dx.
\end{aligned}
\tag{15}
$$

Similarly, define

$$
T(s,t) := \int_a^b \left[\ln \left\{ \frac{f_1(x)}{f_0(x)} \right\} - s - tx \right]^2 dx, \quad s, t \in \mathbb{R}.
\tag{16}
$$

Since $\ln[f_1(x)/f_0(x)] = \alpha + \beta x$, $T(s,t) = 0$ by definition and α and β defined as

$$
\begin{aligned}
\beta &= \frac{12}{(b-a)^3} \int_a^b \left[\ln \left\{ \frac{f_{n_1}(x)}{f_{n_0}(x)} \right\} \right] \left\{ x - \frac{a+b}{2} \right\} dx, \\
\alpha &= \frac{1}{b-a} \int_a^b \left[\ln \left\{ \frac{f_{n_1}(x)}{f_{n_0}(x)} \right\} - \beta x \right] dx,
\end{aligned}
\tag{17}
$$

which minimize $T(s,t)$. Then, the error of the estimators are

$$
\begin{aligned}
\hat{\alpha} - \alpha &= c_1 \left[\int_a^b \left\{ \ln(\hat{f}_{n_1}(x) + b_{n_1}) - \ln f_1(x) \right\} \left\{ x - \frac{2(a^2 + ab + b^2)}{3(a+b)} \right\} dx \right. \\
&\quad \left. - \int_a^b \left\{ \ln(\hat{f}_{n_0}(x) + b_{n_0}) - \ln f_0(x) \right\} \left\{ x - \frac{2(a^2 + ab + b^2)}{3(a+b)} \right\} dx \right], \\
\hat{\beta} - \beta &= c_2 \left[\int_a^b \left\{ \ln(\hat{f}_{n_1}(x) + b_{n_1}) - \ln f_1(x) \right\} \left\{ x - \frac{a+b}{2} \right\} dx \right. \\
&\quad \left. - \int_a^b \left\{ \ln(\hat{f}_{n_0}(x) + b_{n_0}) - \ln f_0(x) \right\} \left\{ x - \frac{a+b}{2} \right\} dx \right],
\end{aligned}
\tag{18}
$$

in which $c_1 = -6(a+b)/(b-a)^3$ and $c_2 = 12/(b-a)^3$.

Geng and Sakhanenko (2016) discussed the choice of the finite interval $[a, b]$. The

authors recommended to choose the a and b with a lot of different responses and both f_0 and f_1 are both relatively high for more accurate classifications. They also investigated practical choices of (a, b): group sample means, group sample medians, and the boundaries of the overlap interval of the two group observation; and compared the performance of the these choices with the true group mean, which was not feasible in practice. Their simulation study showed that both group sample means and sample medians were good options in practice, where as the boundaries of the overlap interval of the two group observation were not reliable in different sample size combinations and underlying distributions of the covariate.

CHAPTER III: MAIN RESULTS

In this section, we state the assumptions and the main results of consistency and asymptotic normality of the integrated square distance estimators with deconvolution.

Assumptions

1. f_0 and f_1 have m continuous derivatives.

2. ϕ_K is a symmetric function having $m + 2$ bounded integrable derivatives; $\phi_K(0) = 1$;

   $\phi_K(t) = 1 + O(t^m)$ as $t \to 0$.

3. $|\phi_\varepsilon(t)| > 0$ for all real t; $d_0|t|^{-\tau} \le |\phi_\varepsilon(t)| \le d_1|t|^{-\tau}$ as $t \to \infty$ for some constants

   $0 < d_0 \le d_1$ and $\tau > 0$.

4. $\int_{-\infty}^{\infty}[|\phi_K(t)| + |\phi_K'(t)|]|t|^\tau dt < \infty$.

5. $h_i \to 0$ and $n_i h_i \to \infty$ for $i = 0, 1$.

6. There exists δ_i such that $n_i^{\delta_i/2} h_i^{\delta_i + \tau(2+\delta)} \to \infty$ for $i = 0, 1$

7. $n_i h_i^2 \to \infty$ and $n_i h_i^{2m} \to 0$ for $i = 0, 1$.

8. $n_i h_i / \log n_i \to \infty$ and $|\log h_i| / \log(\log n_i) \to \infty$ for $i = 0, 1$.

9. $n_i^{1/2} b_i \to 0$ for $i = 0, 1$.

10. $n_1/n \to \rho$, $0 < \rho < 1$, $n = n_1 + n_0$

Let f_{1U} denote the density of $U_{1i} = X_{1i} + \varepsilon_{1i}$ and f_{0U} denote the density of $U_{0i} = X_{0i} + \varepsilon_{0i}$.

We also define

$$g_{1i}(x) = \frac{c_1\{x - 2(a^2 + ab + b^2)/(3a + 3b)\}}{f_i(x)}; \; g_{2i}(x) = \frac{c_2\{x - (a + b)/2\}}{f_i(x)}; \; i = 0, 1. \qquad (19)$$

Theorem 1. $\hat{\alpha}$ and $\hat{\beta}$ is consistent estimators of α and β, respectively, and

$$\sqrt{n}\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} \to N(0, \Sigma), \qquad (20)$$

in which

$$\Sigma = \rho^{-1}\Sigma_1 + (1 - \rho)^{-1}\Sigma_0. \qquad (21)$$

The kl-th entries $(k, l = 1, 2)$ of Σ_1 and Σ_0 are

$$\begin{aligned}
\Sigma_1^{(kl)} &= \int_a^b g_{k1}(x)g_{l1}(x)f_{1U}(x)dx - \int_a^b f_1(x)g_{k1}(x)dx \int_a^b f_1(x)g_{l1}(x)dx, \\
\Sigma_0^{(kl)} &= \int_a^b g_{k0}(x)g_{l0}(x)f_{0U}(x)dx - \int_a^b f_0(x)g_{k0}(x)dx \int_a^b f_0(x)g_{l0}(x)dx.
\end{aligned} \qquad (22)$$

Remarks on theorem 1

1. The role of sample size ratio.

   The covariance matrix Σ is the sum of Σ_1 and Σ_0 with the coefficient ρ^{-1} and $(1 - \rho)^{-1}$ determined by the sample size ratio of case and control groups. In a severely unbalanced sample size ratio situation, the covariance matrix Σ could get very large since the one of the coefficients, either ρ^{-1} or $(1 - \rho)^{-1}$ could get very large. This could result in a large mean square error of the estimators. For example, when the sample size of the

case group is much larger than of the control group, $\rho^{-1} \to 1$ and $(1 - \rho)^{-1} \to \infty$. Then, Σ gets very large because of $(1 - \rho)^{-1}\Sigma_0$ being very large. Vice versa, when the sample size of of the case group is much smaller than of the control group, $\rho^{-1} \to \infty$ and $(1 - \rho)^{-1} \to 1$. Then, Σ gets very large because of $\rho^{-1}\Sigma_0$ being very large. Though it is ideal to have balanced sample size, in practice we often encounter the situation in which the sample size of the control group is larger than of case group. In our real data application in chapter IV section 2, the control-case sample size ratio is approximately $8 : 1$.

2. The connection with the error-free case.

   In each kl-th entry of the covariance matrix Σ_1, the integrand of the first integration term includes the density f_{1U} of the observed contaminated covariate data in the case group $U_{1i} = X_{1i} + \varepsilon_{1i}$, $i = 1, ... n_1$. When the data is free of error, the density $f_{1U}(x)$ becomes the density $f_1(x)$ and the kl-th entry of Σ_1 becomes

$$\Sigma_1^{(kl)} = \int_a^b g_{k1}(x)g_{l1}(x)f_1(x)dx - \int_a^b f_1(x)g_{k1}(x)dx \int_a^b f_1(x)g_{l1}(x)dx. \qquad (23)$$

Similar result applies to the kl-th entry of Σ_0 when the data is free of error:

$$\Sigma_0^{(kl)} = \int_a^b g_{k0}(x)g_{l0}(x)f_0(x)dx - \int_a^b f_0(x)g_{k0}(x)dx \int_a^b f_0(x)g_{l0}(x)dx. \qquad (24)$$

Thus, when there is no measurement error, our covariance structure is the same as the one obtained by Geng and Sakhanenko (2016).

Outline of the proof of theorem 1

We state four lemmas needed to prove theorem 1 and outline the main ideas of the proof.

For detailed proof of theorem 1 and proofs of the four lemmas, see Appendix.

Lemma 1. *Let $g(x)$ be a continuous function over $[a, b]$.*

$$E\left[\frac{1}{h_i}K^*\left(\frac{x-U}{h_i}\right)\right] = f_i(x) + O(h_i^m) \quad for \ i = 0, 1. \tag{25}$$

and

$$\int_a^b E\left[\frac{1}{h_i}K^*\left(\frac{x-U}{h_i}\right)\right]g(x)dx = \int_a^b f_i(x)g(x)dx + O(h_i^m) \quad for \ i = 0, 1. \tag{26}$$

Lemma 2.

$$n^{1/2}\left[\hat{\alpha} - \alpha - \int_a^b\{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx + \int_a^b\{\hat{f}_{n_0}(x) - f_0(x)\}g_{10}(x)dx\right] = o_p(1);$$

$$n^{1/2}\left[\hat{\beta} - \beta - \int_a^b\{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx + \int_a^b\{\hat{f}_{n_0}(x) - f_0(x)\}g_{20}(x)dx\right] = o_p(1). \tag{27}$$

Lemma 3. *Let $g_1(x)$ and $g_2(x)$ be continuous functions over $[a, b]$. Then,*

$$E\left[\int_a^b\left\{\frac{1}{h_1}K^*\left(\frac{x-U}{h_1}\right) - f_1(x)\right\}g_1(x)dx\int_a^b\left\{\frac{1}{h_1}K^*\left(\frac{x-U}{h_1}\right) - f_1(x)\right\}g_2(x)dx\right]$$

$$\rightarrow \int_a^b g_1(x)g_2(x)f_{1U}(x)dx - \int_a^b g_1(x)f_1(x)dx\int_a^b g_2(x)f_1(x)dx \quad as \ h_1 \rightarrow 0 \tag{28}$$

and

$$E\left[\int_a^b\left\{\frac{1}{h_0}K^*\left(\frac{x-U}{h_0}\right) - f_0(x)\right\}g_1(x)dx\int_a^b\left\{\frac{1}{h_0}K^*\left(\frac{x-U}{h_0}\right) - f_0(x)\right\}g_2(x)dx\right]$$

$$\rightarrow \int_a^b g_1(x)g_2(x)f_{0U}(x)dx - \int_a^b g_1(x)f_0(x)dx\int_a^b g_2(x)f_0(x)dx \quad as \ h_0 \rightarrow 0. \tag{29}$$

Lemma 4. *Let $g(x)$ be a continuous function over $[a, b]$.*

$$\sqrt{n_1} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g(x)dx \to N(0, s_1^2), \quad \sqrt{n_0} \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g(x)dx \to N(0, s_0^2) \quad (30)$$

in which

$$
\begin{aligned}
s_1^2 &= \int_a^b g^2(x) f_{1U}(x)dx - \left\{ \int_a^b f_1(x)g(x)dx \right\}^2, \\
s_0^2 &= \int_a^b g^2(x) f_{0U}(x)dx - \left\{ \int_a^b f_0(x)g(x)dx \right\}^2.
\end{aligned}
\quad (31)
$$

.

Lemma 1 results imply the bias of the deconvolution kernel density estimators $\hat{f}_{n_i}(x)$ and of $\int_a^b \hat{f}_{n_i}(x)g(x)dx$ is $O(h_i^m)$ given $g(x)$ is a continuous function over $[a, b]$. Lemma 2 implies the main terms of the error of estimators are

$$
\begin{aligned}
&\sqrt{n} \begin{pmatrix} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx \\ \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx \end{pmatrix} - \sqrt{n} \begin{pmatrix} \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g_{10}(x)dx \\ \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g_{20}(x)dx \end{pmatrix} + o_p(1) \\
&= \sqrt{\frac{n}{n_1}} \sqrt{n_1} \begin{pmatrix} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx \\ \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx \end{pmatrix} - \sqrt{\frac{n}{n_0}} \sqrt{n_0} \begin{pmatrix} \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g_{10}(x)dx \\ \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g_{20}(x)dx \end{pmatrix} + o_p(1)
\end{aligned}
$$

$$(32)$$

To show $\hat{\beta}$ is consistent, we show the following expectation converges to 0.

$$E\left[\int_a^b \{\hat{f}_{n_1} - f_1(x)\}g_{21}(x)dx - \int_a^b \{\hat{f}_{n_0} - f_0(x)\}g_{20}(x)dx\right]^2$$

$$= E\left[\int_a^b \{\hat{f}_{n_1} - f_1(x)\}g_{21}(x)dx\right]^2 + E\left[\int_a^b \{\hat{f}_{n_0} - f_0(x)\}g_{20}(x)dx\right]^2 \quad (33)$$

$$- 2E\left[\int_a^b \{\hat{f}_{n_1} - f_1(x)\}g_{21}(x)dx \int_a^b \{\hat{f}_{n_0} - f_0(x)\}g_{20}(x)dx\right].$$

By lemma 1 and 3, we can show each term of the above expression converges to 0. Using the similar argument, we can show $\hat{\alpha}$ is a consistent estimator of α.

To show the asymptotic normality, we show that $\sqrt{n}\begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix}$ is the sum of two independent normal variables. By lemma 4, we can show for all $a_{11}, a_{21} \in \mathbb{R}$,

$$a_{11}\sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx + a_{21}\sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx \quad (34)$$

is normally distributed by letting $g(x) = a_{11}g_{11}(x) + a_{21}g_{21}(x)$. Then,

$$\sqrt{n_1}\begin{pmatrix} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx \\ \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx \end{pmatrix} \quad (35)$$

is a bivariate normal random variable. By lemma 1 and 3, we obtain the covariance matrix Σ_1

$$Var\left(\sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx\right) \to \int_a^b g_{11}^2(x)f_{1U}(x)dx - \left(\int_a^b f_1(x)g_{11}(x)dx\right)^2$$
$$Var\left(\sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx\right) \to \int_a^b g_{21}^2(x)f_{1U}(x)dx - \left(\int_a^b f_1(x)g_{21}(x)dx\right)^2 \quad (36)$$

and

$$Cov\left(\sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx, \sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx\right)$$

$$\to \int_a^b g_{11}(x)g_{21}(x)f_{1U}(x)dx - \int_a^b f_1(x)g_{11}(x)dx \int_a^b f_1(x)g_{21}(x)dx. \tag{37}$$

Then the kl-th entry of Σ_1 for $k, l = 1, 2$ is

$$\Sigma_1^{(kl)} = \int_a^b g_{k1}(x)g_{l1}(x)f_{1U}(x)dx - \int_a^b f_1(x)g_{k1}(x)dx \int_a^b f_1(x)g_{l1}(x)dx. \tag{38}$$

By similar argument,

$$\sqrt{n_1}\begin{pmatrix}\int_a^b\{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx \\ \int_a^b\{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx\end{pmatrix} \to N(0, \Sigma_0), \tag{39}$$

in which the kl-th entry of Σ_0 for $k, l = 1, 2$

$$\Sigma_0^{(kl)} = \int_a^b g_{k0}(x)g_{l0}(x)f_{0U}(x)dx - \int_a^b f_0(x)g_{k0}(x)dx \int_a^b f_0(x)g_{l0}(x)dx. \tag{40}$$

Then, by independence

$$\sqrt{n}\begin{pmatrix}\hat{\alpha} - \alpha \\ \hat{\beta} - \beta\end{pmatrix} \to N(0, \Sigma) \tag{41}$$

in which

$$\Sigma = \rho^{-1}\Sigma_1 + (1 - \rho)^{-1}\Sigma_0. \tag{42}$$

Remarks on assumptions

Assumption (2) implies that the kernel $K(x)$ is symmetric and satisfying the m-th order constraint

$$\int x^j K(x) dx = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \leq j \leq m - 1 \\ B & \text{if } j = m, \end{cases}$$

in which B is a constant. Note that since $K(x)$ is symmetric, all the odd moments are 0. The assumption of the higher-order kernel functions is used to ensure the bias of the deconvolution kernel density estimators converges to 0 sufficiently fast (see Lemma 1). The higher the order of the kernel, the faster the rate of convergence. In our simulation study in chapter IV section 1, we use the Gaussian kernel of order 2. Assumption (3) is the assumption on the tail of the characteristic function ϕ_ε when the error density is ordinary smooth. Assumption (4) on the characteristic function ϕ_K of kernel K, along with the bandwidth assumption (6), is used to verify the Lyapunov's Central Limit Theorem conditions for the proposed estimator. Assumption (5) is the assumption used to establish the consistency of the kernel density estimators. Assumption (7) is used to eliminate the bias of the deconvolution kernel density estimators. Assumption (8) is required to obtained the upper bound of $\sup_x |\hat{f}_{n_i}(x) - E\hat{f}_{n_i}(x)|$ established by Gine and Guillou (2002).

CHAPTER IV: DATA ANALYSIS

IV.1 Simulation

We investigate the performance of the deconvolution integrated square (ISD) method and compare it with the naive ISD method via simulation study. In particular, we are interested in investigating if the estimator is superior to the one obtained by ignoring the measurement error. We study the case in which the densities of the true covariate are two normal distributions with common variance. In particular, the baseline distribution is $N(0,1)$, and the control group distribution is $N(\beta, 1)$ with $\beta = 1$. We generate 500 samples of various combinations of sample sizes of case and control groups: (i) when the sample sizes of both groups are equal, (ii) when the sample size ratio of control to case group is $3:1$, and (iii) when the ratio is $5:1$. We choose the sample size of the control group to be greater than of the case group since in practice, there are usually more instances in which the disease is absent than ones in which it is present. The sample is contaminated by Laplace error density with variance σ_ε^2

$$f_\varepsilon(x) = \frac{1}{\sqrt{2}\sigma_\varepsilon} \exp\left(\frac{-\sqrt{2}x}{\sigma_\varepsilon}\right), \tag{43}$$

with the characteristic function

$$\phi_\varepsilon(t) = \frac{1}{1 + (1/2)\sigma_\varepsilon^2 t^2}. \tag{44}$$

In our simulation study, we consider two cases: $\sigma_\varepsilon = 0.3$ and $\sigma_\varepsilon = 0.6$. We choose $K(x)$ to be the Gaussian (standard normal) kernel with the characteristic function

$$\phi_K(t) = \exp\left(-\frac{1}{2}t^2\right). \tag{45}$$

Then, our deconvolution kernel $K^*(x)$ has the form

$$K^*(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right)\left[1 + \left(\frac{\sigma_\varepsilon}{\sqrt{2}h}\right)^2(1 - x^2)\right]. \tag{46}$$

Note that the Gaussian kernel $K(x)$ has order of 2. In order for the bias of $\int_a^b \hat{f}_{n_i}(x)g(x)dx$ to go to 0 fast enough and the asymptotic normality to hold, when the covariate is contaminated with Laplace error, we need kernel $K(x)$ to be of order 4. However, higher order kernels often do not have a close form for $K^*(x)$ and are difficult to compute. Therefore, in this simulation study, we use the standard normal kernel for the simplicity of the computation. As figure (6) displays, the range of $K^*(x)$ does have negative values. For the log ratio of the density to be defined, we use $b_{n_i} = n_i^{-2}$.

We use bootstrap method proposed by Delaigle and Gijbels (2004a) to select bandwidth h_i. The bootstrap bandwidth selection method is described as follows. Suppose we wish to obtain the optimal bandwidth h. The method first requires obtaining a pilot bandwidth h_{pilot}. We use the rule of thumb method to calculate the pilot bandwidth $h_{\text{pilot}} = O(n^{-1/9})$ for Laplace error in our simulation study. This bandwidth and the contaminated observed data are then used to obtain the pseudo deconvolution kernel density estimator $\hat{f}_X(x; h_{\text{pilot}})$. A bootstrap sample $X_1^*, X_2^*, ..., X_n^*$ is drawn from $\hat{f}_X(x; h_{\text{pilot}})$ and
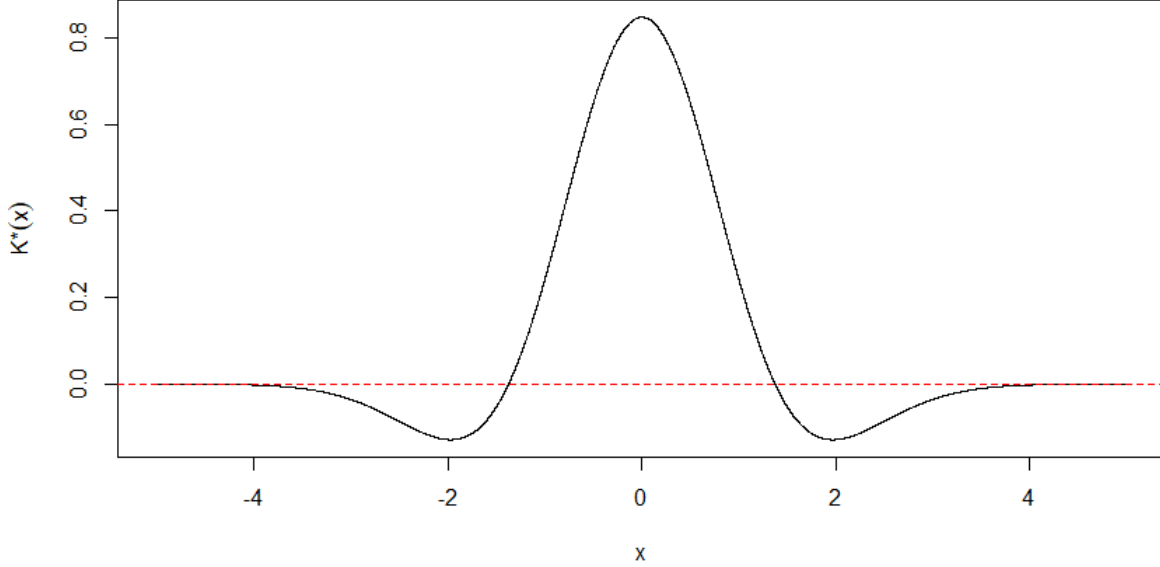
Figure 6: The deconvolution kernel $K^*(x)$ with Gaussian kernel $K(x)$ and the Laplace error density

the error ε is added to the sample. The contaminated bootstrap sample is then used to construct the deconvolution kernel density estimator $\hat{f}_X^*(x; h)$. The optimal bandwidth is obtained by minimizing the bootstrap estimator of mean integrated square error.

$$MISE^*(h) = E^* \int \{\hat{f}_X^*(x; h) - \hat{f}_X(x; h_{\text{pilot}})\}^2 dx$$
$$= \int [Bias^*\{\hat{f}_X^*(x; h)\}]^2 dx + \int Var^*\{\hat{f}_X^*(x; h)\} dx. \tag{47}$$

The bootstrap bandwidth selection method is proven to be consistent Delaigle and Gijbels (2004a) and its performance in deconvolution density estimation is shown to be superior to other commonly used methods such as cross-validation via simulation study by Delaigle and Gijbels (2004b).

For the naive ISD estimator, we choose the bandwidth $h_i = n_i^{-1/3}$ the Gaussian kernel to construct the naive kernel density estimator. We use these settings to be consistent

with the theoretical result and the simulation study of Geng and Sakhanenko (2016). Table 1 through 3 show bias and root mean square error of the simulations when the covariate sample is contaminated with Laplace error with parameter $\sigma_\varepsilon = 0.3$. Table 4 through 6 show the results when the sample is contaminated with Laplace error with parameter $\sigma_\varepsilon = 0.6$. We use the true mean of the covariate groups for the values of a and b. In all cases, the deconvolution ISD estimator obtains a smaller bias compared to one obtained by the naive ISD estimator. Especially in the case in which the variance is large ($\sigma_\varepsilon = 0.6$), the bias of the naive ISD estimator increases by a much larger amount compared to the deconvolution ISD estimator. The increase in bias of the deconvolution ISD estimator can be explained by the fact that it is more difficult to deconvolute when the error variance is large. However, the deconvolution ISD estimator is less efficient than the naive ISD estimator, especially when the error variance is large and the sample size is small.

	$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
Deconvolution Estimator	0.0697 (0.7084)	0.0635 (0.4550)	0.0527 (0.3659)	0.0286 (0.2854)
Naive Estimator	0.1073 (0.5643)	0.0994 (0.3809)	0.0835 (0.3009)	0.0693 (0.2401)

Table 1: Comparison of deconvolution integrated square distance (ISD) estimator and naive ISD estimator for $\sigma_\varepsilon = 0.3$, equal sample size, and $(a, b) = (0, 1)$

	$n_0 = 300$ $n_1 = 100$	$n_0 = 600$ $n_1 = 200$	$n_0 = 900$ $n_1 = 300$	$n_0 = 1200$ $n_1 = 400$
Deconvolution Estimator	0.0612 (0.5728)	0.0522 (0.4218)	0.0400 (0.3827)	0.0355 (0.3501)
Naive Estimator	0.0887 (0.4724)	0.0849 (0.3607)	0.0747 (0.3186)	0.0708 (0.2895)

Table 2: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.3$, sample size ratio $3 : 1$, and $(a, b) = (0, 1)$

	$n_0 = 250$ $n_1 = 50$	$n_0 = 500$ $n_1 = 100$	$n_0 = 1000$ $n_1 = 200$	$n_0 = 2000$ $n_1 = 400$
Deconvolution Estimator	0.0682 (0.7322)	0.0657 (0.5375)	0.0553 (0.4087)	0.0346 (0.3179)
Naive Estimator	0.0938 (0.5510)	0.0918 (0.4329)	0.0994 (0.3445)	0.0719 (0.2646)

Table 3: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.3$, sample size ratio $5:1$, and $(a, b) = (0, 1)$

	$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
Deconvolution Estimator	0.0858 (1.2249)	0.0771 (1.0421)	0.0521 (0.6493)	0.0570 (0.4978)
Naive Estimator	0.2294 (0.6246)	0.2144 (0.4631)	0.2125 (0.3806)	0.2053 (0.3143)

Table 4: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.6$, equal sample size, and $(a, b) = (0, 1)$

	$n_0 = 300$ $n_1 = 100$	$n_0 = 600$ $n_1 = 200$	$n_0 = 900$ $n_1 = 300$	$n_0 = 1200$ $n_1 = 400$
Deconvolution Estimator	0.1547 (1.0014)	0.0722 (0.8107)	0.0531 (0.8085)	0.0263 (0.6254)
Naive Estimator	0.2634 (0.5617)	0.2182 (0.4152)	0.2212 (0.3914)	0.1693 (0.3388)

Table 5: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.6$, sample size ratio $3:1$, and $(a, b) = (0, 1)$

	$n_0 = 250$ $n_1 = 50$	$n_0 = 500$ $n_1 = 100$	$n_0 = 1000$ $n_1 = 200$	$n_0 = 2000$ $n_1 = 400$
Deconvolution Estimator	0.1067 (1.2195)	0.0631 (0.9906)	0.0587 (0.8709)	0.0349 (0.4929)
Naive Estimator	0.2195 (0.6209)	0.2063 (0.4985)	0.2271 (0.4225)	0.1966 (0.3198)

Table 6: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.6$, sample size ratio $5:1$, and $(a, b) = (0, 1)$

Since it is not feasible to use the true mean of the covariate groups for the values of

a and b in practice, we investigate the option of using sample means of the covariate groups.

Table 7 through 9 show the simulation results of when the covariate sample is contaminated with Laplace error with parameter $\sigma_\varepsilon = 0.3$ and (a, b) as the sample means of the observed control and case groups, respectively. Obviously, both estimators do not perform as well as they do when using true means for a and b. Regardless, the deconvolution ISD estimator still outperforms the naive estimator in terms of obtaining a smaller bias.

	$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
Deconvolution Estimator	0.0810 (0.7132)	0.0734 (0.4626)	0.0695 (0.3643)	0.0408 (0.2869)
Naive Estimator	0.1193 (0.5597)	0.1114 (0.3853)	0.0978 (0.3043)	0.0801 (0.2421)

Table 7: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.3$, equal sample size, and (a, b) chosen to be sample means of control and case group respectively

	$n_0 = 300$ $n_1 = 100$	$n_0 = 600$ $n_1 = 200$	$n_0 = 900$ $n_1 = 300$	$n_0 = 1200$ $n_1 = 400$
Deconvolution Estimator	0.0661 (0.5733)	0.0606 (0.4288)	0.0582 (0.3814)	0.0513 (0.3524)
Naive Estimator	0.0931 (0.4706)	0.0924 (0.3672)	0.0902 (0.3168)	0.0850 (0.2929)

Table 8: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.3$, sample size ratio 3 : 1, and (a, b) chosen to be sample means of control and case group respectively

	$n_0 = 250$ $n_1 = 50$	$n_0 = 500$ $n_1 = 100$	$n_0 = 1000$ $n_1 = 200$	$n_0 = 2000$ $n_1 = 400$
Deconvolution Estimator	0.0883 (0.7287)	0.0784 (0.5314)	0.0694 (0.4138)	0.0485 (0.3191)
Naive Estimator	0.1129 (0.5466)	0.1014 (0.4326)	0.1118 (0.3478)	0.0853 (0.2670)

Table 9: Comparison of deconvolution ISD estimator and naive ISD estimator for $\sigma_\varepsilon = 0.3$, sample size ratio 5 : 1, and (a, b) chosen to be sample means of control and case group respectively

IV.2 Real Data Application

We apply the method on the Framingham Heart Study data to investigate the relationship between the systolic blood pressure and the presence of cardiovascular disease. As mentioned, the dataset has been the subject of study in various previous papers studying effects of covariate measurement error in logistic regression. The Framingham Heart Study is one of the longest running epidemiologic studies and is conducted under the direction of the National Heart, Lung, and Blood Institute. The dataset was retrieved from Buonaccorsi (2012). The dataset includes two systolic blood pressure measurements from two different visits, the age, and the presence or absence of cardiovascular disease of 1615 individuals. There are 128 individuals with cardiovascular disease (case group) and 1487 individuals without disease (control group). Thus, the sample size is severely imbalanced between the two groups. The variables of our interest are the two systolic blood pressure as covariate X and the presence or absence of cardiovascular disease as binary reponse Y. We assume the average of the two blood pressure measurements is the "true" measurement and the difference between the measurements and the average as the measurement errors. Figure (7) shows the histogram of the "true" systolic blood pressure measurement is right-skewed. Figure (8) and (9) display the histogram of the error sample and the Laplace theoretical quantiles versus sample quantiles of the errors, respectively. Since the histogram shows the distribution is symmetrical and peaky and the quantile plot shows the sample follow a straigt line, we assume the error sample is from the Laplace distribution.

We use the Gaussian kernel and the bootstrap bandwidth selection method to construct the deconvolution kernel density estimator, similar to the simulation study. We also
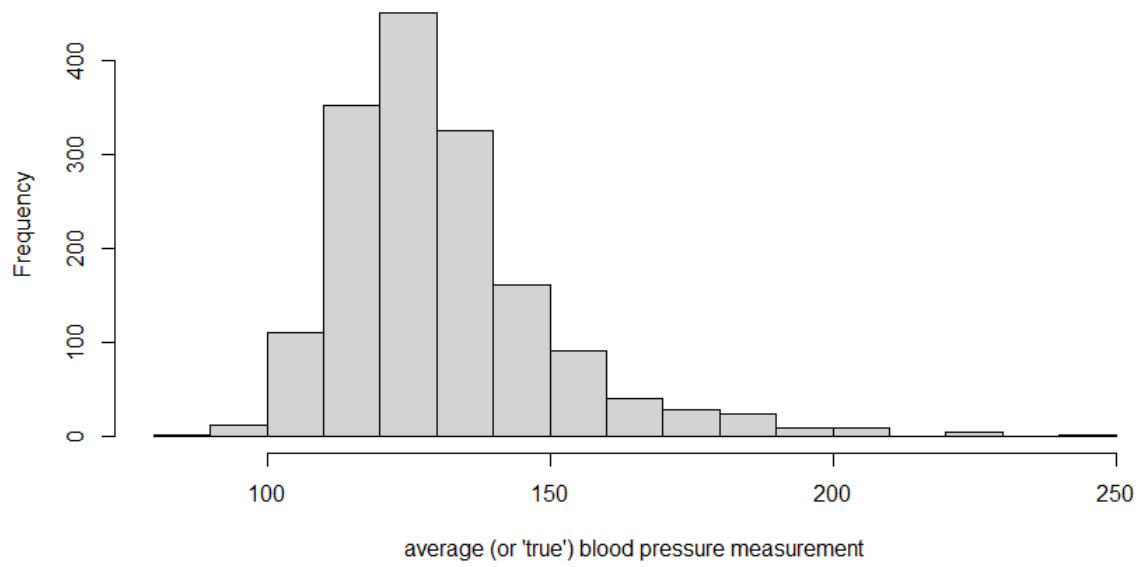
Figure 7: Histogram of the average ("true") systolic blood pressure measurement
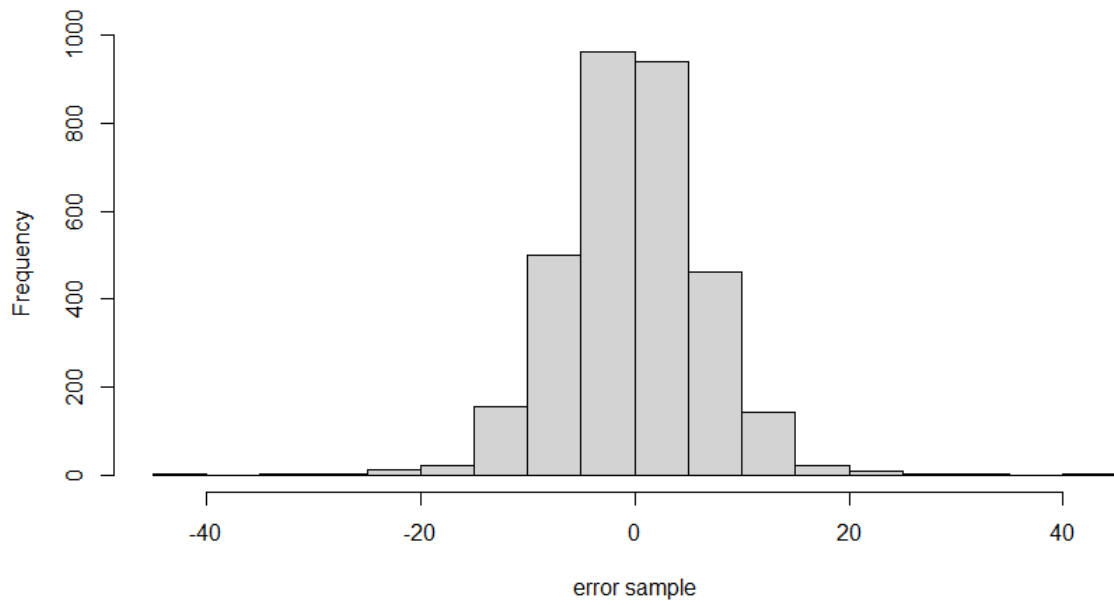
Figure 8: Histogram of the error sample
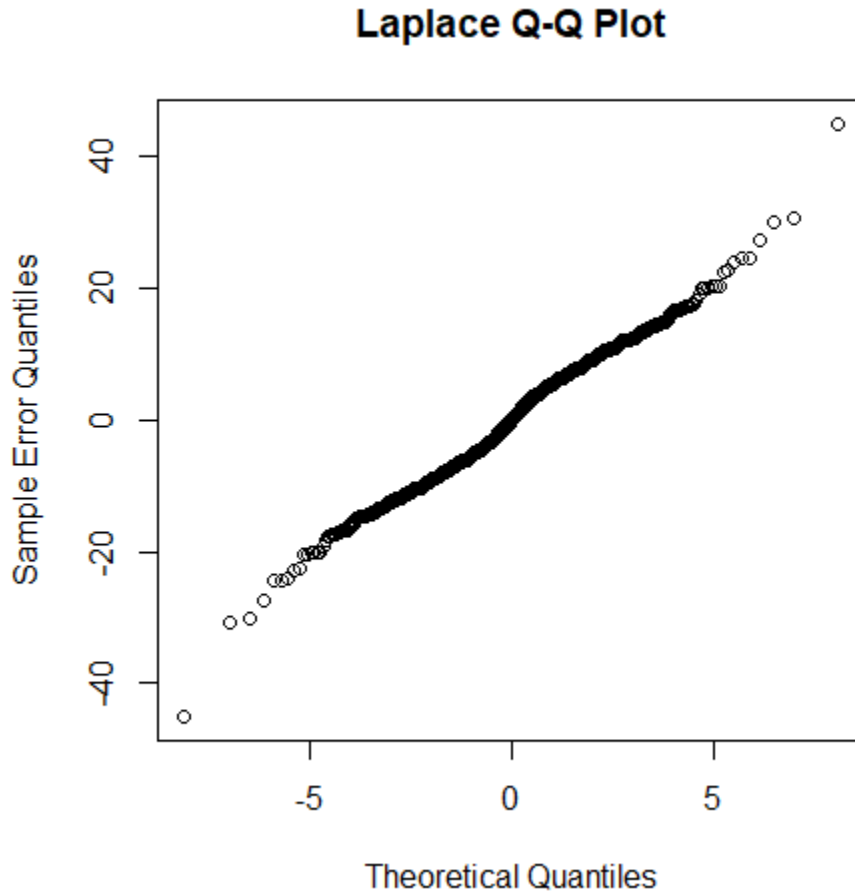
Laplace Q-Q Plot

Figure 9: Laplace Q-Q plot of the error distribution

use Gaussian kernel to construct the naive kernel density estimator. The bandwidth for the naive kernel density estimator is scaled by the sample standard deviation $h_i = \hat{\sigma}_i * n_i^{-1/3}$. Table 10 display the result of the data application. We compute $\hat{\beta}$ using the "true" systolic blood pressure measurement and the naive ISD and the naive MLE estimators. We use the naive ISD estimator $\hat{\beta} = 0.0316$ to as the benchmark to compare the result of the other estimators. When using the first systolic blood pressure measurements as observed covariate values, the deconvolution ISD method outperforms compared to the naive ISD method, but underperforms compared to the naive MLE method. When using the second systolic

blood pressure measurements as observed covariate values, the deconvolution ISD method

outperforms other two methods.

	"True" measurement	First measurement	Second measurement
Naive MLE Estimator	0.0223	0.0201	0.0204
Naive Estimator	0.0316	0.0787	0.0570
Deconvolution Estimator	N/A	0.0503	0.0329

Table 10: Comparing data application results using naive maximum likelihood, naive ISD, and deconvolution ISD methods

CHAPTER V:DISCUSSION

In this paper, we discussed a method to estimate parameters of logistic regression when the covariate is measure with additive error. The estimator is derived by the minimizing the estimated logarithm ratio of the densities of case and control groups to the true one. The unknown densities are estimated by the deconvolution kernel estimators as the true measures of the covariate cannot be observed. We established the consistency and the asymptotic normality of the estimator. Our simulation study showed the estimates obtained by this method have smaller bias than ones obtained by ignoring measurement error.

Since the method includes the use of the deconvolution kernel density estimator, the method suffers similar limitations of the deconvolution kernel density estimator. Though one advantage of the method is that it does not make any assumptions on the underlying distribution of the covariate, it still requires that the distribution of the error is known to estimate the density of the true covariate. Practitioners could derive knowledge on the error distribution using past experiments and data in practice, however, misspecification of the error distribution could still occur. Another limitation of the method is that its performance degrades when the variance is large, as shown by the simulation study. This is due to the fact that the deconvolution is more difficult when the error density has large variance.

In this paper, we only considered the case in which the error distribution is ordinary smooth. We did not derive results for the case in which the error distribution is super smooth but we do hope to do so in future work. The normal distribution belongs to this group of distributions and it is common in practice that the error distribution is assumed to be normal distribution. In addition, as mentioned, in deconvolution kernel density estimation,

the smoother the error distribution, the more difficult the deconvolution is. Thus, future work studying the method with this type of error is needed. We also hope to discuss the goodness-of-fit testing procedure of the model in future work.

REFERENCES

[AmWH] Armstrong, B. G., Whittemore, A. S., & Howe, G. R. (1989). Analysis of case-control data with covariate measurement error: application to diet and colon cancer. *Statistics in Medicine,* 8(9), 1151-1163.

[Bu1] Buonaccorsi, J. P. (2010). *Measurement error : Models, methods, and applications.* ProQuest Ebook Central https://ebookcentral.proquest.com

[Bu2] Buonaccorsi, J. P. (2012). ST697M/797M: Measurement Error and Misclassification. Retrieved June 9, 2021 from https://people.math.umass.edu/ johnpb/s697m.html.

[CaH] Carroll, R. J., & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association, 83(404)*, 1184-1186.

[CSLBA] Carroll, R. J., Spiegelman, C. H., Lan, K. G., Bailey, K. T., & Abbott, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika, 71(1)*, 19-25.

[DeG1] Delaigle, A., & Gijbels, I. (2004a). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics, 56(1)*, 19-47.

[DeG2] Delaigle, A., & Gijbels, I. (2004b). Practical bandwidth selection in deconvolution kernel density estimation. *Computational statistics and data analysis, 45(2)*, 249-267.

[DeH]   Delaigle, A., & Hall, P. (2006). On optimal kernel choice for deconvolution. *Statistics and Probability Letters, 76(15)*, 1594-1602.

[De]   Devroye, L. (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 235-239.

[Fa1]   Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, 97-110.

[Fa2]   Fan, J. (1991b). Global behavior of deconvolution kernel estimates. *Statistica Sinica*, 541-551.

[Fa3]   Fan, J. (1991c). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 1257-1272.

[GeS]   Geng, P., & Sakhanenko, L. (2016). Parameter estimation for the logistic regression model under case-control study. *Statistics and Probability Letters, 109*, 168-177.

[KoN]   Koul, H. L., & Ni, P. (2004). Minimum distance regression model checking. Journal of Statistical Planning and Inference, 119(1), 109-141.

[LiT]   Liu, M. C., & Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics, 17(4)*, 427-438.

[Sc]   Schafer, D. W. (1987). Covariate measurement error in generalized linear models. *Biometrika, 74(2)*, 385-391.

[St]   Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters, 9(3)*, 229-235.

[StC1]   Stefanski, L. A., & Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics*, 1335-1351.

[StC2]   Stefanski, L. A., & Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics, 21(2)*, 169-184.

[TaZ]   Taylor, R. L., & Zhang, H. M. (1990). On a strongly consistent nonparametric density estimator for the deconvolution problem. *Communications in Statistics-Theory and Methods, 19(9)*, 3325-3342.

[WaW]   Wang, X. F., & Wang, B. (2011). Deconvolution estimation in measurement error models: the R package decon. *Journal of statistical software, 39(10)*.

[YiYLS]   Yi, G., Yan, Y., Liao, X., & Spiegelman, D. (2019). Parametric Regression Analysis with Covariate Misclassification in Main Study/Validation Study Designs. *The International Journal of Biostatistics, 15(1)*, 20170002. https://doi.org/10.1515/ijb-2017-0002

APPENDIX: DETAILED PROOFS OF THEOREM 1 AND TECHNICAL LEMMAS

Proof of Lemma 1.

$$
\begin{aligned}
&E\left[\frac{1}{h_1}K^*\frac{(x-U)}{h_1}\right] \\
&= \frac{1}{h_1}EK^*\left[\frac{x-(X+\varepsilon)}{h_1}\right] \\
&= \frac{1}{h_1}\int_z\int_\epsilon K^*\left[\frac{x-(z+\epsilon)}{h_1}\right]f_1(z)f_\varepsilon(\epsilon)d\epsilon dz \\
&= \frac{1}{h_1}\int_z\int_\epsilon \frac{1}{2\pi}\int_t \exp\left[i\frac{(z+\epsilon)-x}{h_1}t\right]\frac{\phi_K(t)}{\phi_\varepsilon(t/h)}dt f_\varepsilon(\epsilon)d\epsilon f_1(z)dz \\
&= \frac{1}{h_1}\int_z \frac{1}{2\pi}\int_t \exp\left[i\frac{(z-x)}{h_1}t\right]\frac{\phi_K(t)}{\phi_\varepsilon(t/h)}\int_\epsilon \exp\left[i\frac{\epsilon}{h_1}t\right]f_\varepsilon(\epsilon)d\epsilon dt f_1(z)dz \\
&= \frac{1}{h_1}\int_z \frac{1}{2\pi}\int_t \exp\left[i\frac{(z-x)}{h_1}t\right]\frac{\phi_K(t)}{\phi_\varepsilon(t/h)}\phi_\varepsilon(t/h)dt f_1(z)dz \\
&= \frac{1}{h_1}\int_z \frac{1}{2\pi}\int_t \exp\left[i\frac{(z-x)}{h_1}t\right]\phi_K(t)dt f_1(z)dz \\
&= \frac{1}{h_1}\int_z K\left(\frac{z-x}{h_1}\right)f_1(z)dz.
\end{aligned}
\tag{48}
$$

Using change of variable $s=\frac{z-x}{h_1}$, Taylor's expansion, and the assumptions of the ordinary kernel K,

$$\frac{1}{h_1} \int_z K\left(\frac{z-x}{h_1}\right) f_1(z) dz$$

$$= \int K(s) f_1(x + sh_1) ds$$

$$= \int K(s)[f_1(x) + f_1'(x)(sh_1) + \frac{1}{2!} f_1^{(2)}(x)(s^2 h_1^2) + \frac{1}{3!} f_1^{(3)}(x)(s^3 h_1^3)$$

$$+ \frac{1}{4!} f_1^{(4)}(x)(s^4 h_1^4) + ... + \frac{1}{m!} f_1^{(m)}(x)(s^m h_1^m) + o(h^m)] ds \tag{49}$$

$$= f_1(x) \int K(s) ds + h_1 f_1'(x) \int s K(s) ds + \frac{1}{2!} h_1^2 f_1^{(2)}(x) \int s^2 K(s) ds$$

$$+ \frac{1}{3!} h_1^3 f_1^{(3)}(x) \int s^3 K(s) ds + \frac{1}{4!} h_1^4 f_1^{(4)}(x) \int s^4 K(s) ds + ... +$$

$$+ \frac{1}{m!} h_1^m f_1^{(m)}(x) \int s^m K(s) ds + o(h_1^m)$$

$$= f_1(x) + O(h_1^m).$$

By the similar argument and continuity of $g(x)$,

$$\int_a^b E\left[\frac{1}{h_1} K^*(x - U)\right] g(x) dx = \int_a^b f_1(x) g(x) dx + O(h_1^m). \tag{50}$$

Applying similar argument, we have $E[h_0^{-1} K^*\{(x-U)/h_0\}] = f_0(x) + O(h_0^m)$ and $\int_a^b E[h_0^{-1} K^*(x - U)] g(x) dx = \int_a^b f_0(x) g(x) dx + O(h_0^m).$ $\qquad \square$

Proof of Lemma 2. We use similar procedure as in Geng and Sakhanenko (2016). Applying Taylor's expansion with integral remainder for a function f with at least two continuous derivative

$$f(x) = f(a) + f'(a)(x - a) + \int_{t=a}^x f''(t)(x - t) dt, \tag{51}$$

we have

$$\ln\{\hat{f}_{n_i}(x) + b_{n_i}\} - \ln f_i(x) = \frac{1}{f_i(x)}\{\hat{f}_{n_i}(x) - f_i(x) + b_{n_i}\} + \int_{f_i(x)}^{\hat{f}_{n_i}(x)+b_{n_i}} \frac{1}{t^2}\{\hat{f}_{n_i}(x) + b_{n_i} - t\}dt.$$

(52)

Let

$$R_{n_i}(x) = \int_{f_i(x)}^{\hat{f}_{n_i}(x)+b_{n_i}} \frac{1}{t^2}\{\hat{f}_{n_i}(x) + b_{n_i} - t\}dt$$

(53)

then,

$$\hat{\alpha} - \alpha = \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{11}(x)dx - \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g_{10}(x)dx$$

$$+ \int_a^b R_{n_1}(x)f_1(x)g_{11}(x)dx - \int_a^b R_{n_0}(x)f_0(x)g_{10}(x)dx$$

$$+ b_{n_1}\int_a^b g_{11}(x)dx - b_{n_0}\int_a^b g_{10}(x)dx.$$

$$\hat{\beta} - \beta = \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g_{21}(x)dx - \int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g_{20}(x)dx$$

$$+ \int_a^b R_{n_1}(x)f_1(x)g_{21}(x)dx - \int_a^b R_{n_0}(x)f_0(x)g_{20}(x)dx$$

$$+ b_{n_1}\int_a^b g_{21}(x)dx - b_{n_0}\int_a^b g_{20}(x)dx.$$

(54)

Since g_{1i} and g_{2i} for $i = 0, 1$ are bounded over $[a, b]$, the last two terms of the two equations

are $o(n_i^{-1/2})$ by assumption (9). We derive the upper bound for $R_{n_i}(x)$.

$$\sup_{x\in[a,b]} |R_{n_i}(x)| = \sup_{x\in[a,b]} \frac{|\hat{f}_{n_i}(x) + b_{n_i} - f_i(x)|^2}{[\min\{f_i(x), \hat{f}_{n_i}(x) + b_{n_i}\}]^2}$$

$$\leq 2 \sup_{x\in[a,b]} \frac{|\hat{f}_{n_i}(x) - f_i(x)|^2 + |b_{n_i}|^2}{[\min\{f_i(x), \hat{f}_{n_i}(x) + b_{n_i}\}]^2}$$

(55)

Corollary 3.2 of Liu and Taylor (1989) yields $\sup_{x\in[a,b]} |\hat{f}_{n_i}(x) - f_i(x)| \to 0$. Then for a large

40

enough n_i,

$$\sup_{x\in[a,b]} |\hat{f}_{n_i}(x) - f_i(x)| \leq \eta \tag{56}$$

for any $0 < \eta < b_{n_i}$.

$$
\begin{aligned}
\sup_{x\in[a,b]} |R_{n_i}(x)| &\leq 2 \sup_{x\in[a,b]} \frac{|\hat{f}_{n_i}(x) - f_i(x)|^2 + |b_{n_i}|^2}{[\min\{f_i(x), f_i(x) - \eta + b_{n_i}\}]^2} \\
&\leq 2 \sup_{x\in[a,b]} \frac{|\hat{f}_{n_i}(x) - f_i(x)|^2 + |b_{n_i}|^2}{f_i^2(x)} \\
&= O_p(\sup_{x\in[a,b]} |\hat{f}_{n_i}(x) - f_i(x)|^2 + |b_{n_i}|^2).
\end{aligned}
\tag{57}
$$

From Stefanski and Carroll (1990), by assumptions (10), K^* is bouded and square integrable.

By assumption (8), using theorem 2.3 of Gine and Guillou (2002),

$$\sup_{x\in[a,b]} |\hat{f}_{n_i}(x) - E[\hat{f}_{n_i}(x)]|^2 = O_p\left(\frac{\log(h_i)^{-1}}{n_i h_i}\right). \tag{58}$$

From lemma 1, we have

$$\sup_{x\in[a,b]} |E[\hat{f}_{n_i}(x)] - f_i(x)|^2 = O_p(h_i^m). \tag{59}$$

Then,

$$
\begin{aligned}
\sup_{x\in[a,b]} |R_{n_i}(x)| &\leq O_p(\sup_{x\in[a,b]} |\hat{f}_{n_i}(x) - f_i(x)|^2 + |b_{n_i}|^2) \\
&\leq O_p(\sup_{x\in[a,b]} |\hat{f}_{n_i}(x) - E[\hat{f}_{n_i}(x)]|^2 + \sup_{x\in[a,b]} |E[\hat{f}_{n_i}(x)] - f_i(x)|^2 + |b_{n_i}|^2) \\
&= O_p\left(\frac{\log(h_i)^{-1}}{n_i h_i} + h_i^m + b_{n_i}^2\right).
\end{aligned}
\tag{60}
$$

Then,

$$n^{1/2} \int_a^b R_{n_i} f_i(x) g_{2i}(x) dx \leq O_p \left(\frac{\log h_i^{-1}}{n_i^{1/2} h_i} + n_i h_i^{2m} + n_i^{1/2} b_{n_i}^2 \right) \to 0. \tag{61}$$

$\square$

Proof of Lemma 3.

$$E \left[\int_a^b \left\{ \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) - f_1(x) \right\} g_1(x) dx \int_a^b \left\{ \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) - f_1(x) \right\} g_2(x) dx \right]$$

$$= E \left[\int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) g_1(x) dx \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) g_2(x) dx \right.$$

$$- \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) g_1(x) dx \int_a^b f_1(x) g_2(x) dx \tag{62}$$

$$- \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) g_2(x) dx \int_a^b f_1(x) g_1(x) dx$$

$$\left. + \int_a^b f_1(x) g_1(x) dx \int_a^b f_1(x) g_2(x) dx \right].$$

By lemma 1, the two middle terms converge to $\int_a^b f_1(x) g_1(x) dx \int_a^b f_1(x) g_2(x) dx$ as $h_1 \to 0$.

We consider the first term

$$E \left[\int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) g_1(x) dx \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) g_2(x) dx \right]$$

$$= E \int_a^b \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) \frac{1}{h_1} K^* \left(\frac{y-U}{h_1} \right) g_1(x) g_2(y) dx dy$$

$$= \int \int_a^b \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) \frac{1}{h_1} K^* \left(\frac{y-U}{h_1} \right) g_1(x) g_2(y) f_{1U}(u) dx dy du$$

$$= \left(\int_{-\infty}^a + \int_a^b + \int_b^\infty \right) \int_a^b \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) \frac{1}{h_1} K^* \left(\frac{y-U}{h_1} \right) g_1(x) g_2(y) f_{1U}(u) dx dy du$$

$$=: M_1 + M_2 + M_3.$$

$$\tag{63}$$

We will show that $M_1 \to 0$, $M_2 \to \int_a^b g_1(x)g_2(x)f_{1U}(x)dx$, and $M_3 \to 0$ as $h_1 \to 0$.

$$
\begin{aligned}
M_1 &= \int_{-\infty}^a \int_a^b \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) \frac{1}{h_1} K^* \left(\frac{y-U}{h_1} \right) g_1(x)g_2(y)f_{1U}(u)dxdydu \\
&= \int_{-\infty}^a \int_{(a-u)/h}^{(b-u)/h} \int_{(a-u)/h}^{(b-u)/h} K^*(s)K^*(t)g_1(u+sh_1)g_2(u+th_1)f_{1U}(u)dsdtdu \\
&= \int_{-\infty}^a \int_{(a-u)/h}^{(b-u)/h} \int_{(a-u)/h}^{(b-u)/h} [K^{*+}(s)-K^{*-}(s)][K^{*+}(t)-K^{*-}(t)]g_1(u+sh_1)g_2(u+th_1)f_{1U}(u)dsdtdu \\
&=: M_{11} + M_{12} + M_{13} + M_{14}.
\end{aligned}
$$

$$(64)$$

We consider M_{11}. Because g_1, g_2 are bounded on [a,b], K^* is bounded and integrable on $\mathbb{R}$, M_{11} is bounded by, up to a constant C,

$$
\int_{-\infty}^a \int_{(a-u)/h}^{(b-u)/h} \int_{(a-u)/h}^{(b-u)/h} K^{*+}(s)K^{*+}(t)f_{1U}(u)dsdtdu
\tag{65}
$$

From Stefanski and Carroll (1990), we know that $K^*(x)$ is integrable and $\int K^*(x)dx = 1$. Define $F^{*+}(x) = \int_{-\infty}^x F^{*+}(t)dt$. Then the bound in (65) can be rewritten as

$$
\int_{-\infty}^a \left[F^{*+} \left(\frac{b-u}{h_1} \right) - F^{*+} + \left(\frac{a-u}{h_1} \right) \right]^2 f_{1U}(u)du
\tag{66}
$$

For $\forall u < a$,

$$
F^{*+} + \left(\frac{b-u}{h_1} \right) - F^{*+} \left(\frac{a-u}{h_1} \right) \to F^{*+}(\infty) - F^{*+} + (\infty) = 0, \text{ as } h \to 0,
\tag{67}
$$

By dominated convergence theorem, (66) converges to 0. Similarly, M_{12}, M_{13}, and M_{14} all converges to 0. Therefore, $M_1 \to 0$ as $h_1 \to 0$. By similar argument, $M_3 \to 0$. We now

consider M_2. As $h_1 \to 0$,

$$
\begin{aligned}
M_2 &= \int_a^b \int_a^b \int_a^b \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) \frac{1}{h_1} K^* \left(\frac{y-U}{h_1} \right) g_1(x) g_2(y) f_{1U}(u) dx\, dy\, du \\
&= \int_a^b \int_{(a-u)/h}^{(b-u)/h} \int_{(a-u)/h}^{(b-u)/h} K^{*+}(s) K^{*+}(t) f_{1U}(u) ds\, dt\, du \\
&\to \int_a^b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^*(s) K^*(t) g_1(u) g_2(u) f_{1U}(u) ds\, dt\, du \\
&= \int_a^b g_1(u) g_2(u) f_{1U}(u) du.
\end{aligned}
\tag{68}
$$

Thus,

$$
\begin{aligned}
E &\left[\int_a^b \left\{ \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) - f_1(x) \right\} g_1(x) dx \int_a^b \left\{ \frac{1}{h_1} K^* \left(\frac{x-U}{h_1} \right) - f_1(x) \right\} g_2(x) dx \right] \\
&\to \int_a^b g_1(x) g_2(x) f_{1U}(x) dx - \int_a^b g_1(x) f_1(x) dx \int_a^b g_2(x) f_1(x) dx.
\end{aligned}
$$

By similar argument, we can show

$$
\begin{aligned}
E &\left[\int_a^b \left\{ \frac{1}{h_0} K^* \left(\frac{x-U}{h_0} \right) - f_0(x) \right\} g_1(x) dx \int_a^b \left\{ \frac{1}{h_0} K^* \left(\frac{x-U}{h_0} \right) - f_0(x) \right\} g_2(x) dx \right] \\
&\to \int_a^b g_1(x) g_2(x) f_{0U}(x) dx - \int_a^b g_1(x) f_0(x) dx \int_a^b g_2(x) f_0(x) dx.
\end{aligned}
$$

$\square$

Proof of Lemma 4. We first show

$$
\frac{\int_a^b \hat{f}_{n_1}(x) g(x) dx - E \int_a^b \hat{f}_{n_1}(x) g(x) dx}{\sqrt{Var(\hat{f}_{n_1}(x) g(x) dx)}} \to N(0,1).
\tag{69}
$$

Let

$$
Z_{n_{1i}}(x) = \frac{1}{h_1} K^* \left(\frac{x - U_{1i}}{h_1} \right) \quad \text{and} \quad W_{n_{1i}} = \int_a^b Z_{n_{1i}}(x) g(x) dx.
\tag{70}
$$

44

Note that $\int_a^b \hat{f}_{n_1}(x)g(x)dx$ is the sum of an i.i.d sequence. Then, the Lyapunov's condition for the asymptotic normality in (69) is that for some $\delta > 0$

$$\frac{E|W_{n_{11}} - EW_{n_{11}}|^{2+\delta}}{n^{\delta/2}[Var(W_{n_{11}})]^{1+\delta/2}} \to 0 \tag{71}$$

as $n \to \infty$. Since $g(x)$ is continous and bounded over $[a, b]$, by similar argument to Lemma 1 and Fubini's theorem, we have

$$\begin{aligned}
EW_{n_{11}} &= E \int_a^b Z_{n_{11}}(x)g(x)dx \\
&= \int_a^b EZ_{n_{11}}(x)g(x)dx \\
&= \int_a^b E\left[\frac{1}{h_1}K^*(x-U)\right]g(x)dx \\
&= \int_a^b f_1(x)g(x)dx + O(h^m).
\end{aligned} \tag{72}$$

By lemma 3, we have

$$Var(W_{n_{11}}) \le E(W_{n_{11}})^2 \to \int_a^b g^2(x)f_{1U}(x)dx. \tag{73}$$

Then, we need to show for some $\delta > 0$,

$$n^{-\delta/2}E|W_{n_{11}} - EW_{n_{11}}|^{2+\delta} \to 0. \tag{74}$$

Using similar argument to Geng and Sakhanenko (2016) and Koul and Ni (2004), by Holder's

inequality and $g(x)$ bounded on $[a, b]$,

$$n^{-\delta/2}E|W_{n_{11}} - EW_{n_{11}}|^{2+\delta}$$

$$\leq n^{-\delta/2}2^{2+\delta}(E|W_{n_{11}}|^{2+\delta} + |EW_{n_{11}}|^{2+\delta})$$

$$\leq n^{-\delta/2}2^{2+\delta}E|W_{n_{11}}|^{2+\delta} + o(1)$$

$$= n^{-\delta/2}2^{2+\delta}E\left(\int_a^b Z_{n_{11}}(x)g(x)dx\right)^{2+\delta} + o(1) \qquad (75)$$

$$\leq n^{-\delta/2}2^{2+\delta}E\left(\int_a^b Z_{n_{11}}(x)\sup_{x\in[a,b]}|g(x)|dx\right)^{2+\delta} + o(1)$$

$$= n^{-\delta/2}2^{2+\delta}CE\left(\int_a^b Z_{n_{11}}(x)dx\right)^{2+\delta} + o(1)$$

$$\leq n^{-\delta/2}2^{2+\delta}CE\left(\int_a^b Z_{n_{11}}^{1+\delta/2}(x)dx\right)^2 + o(1)$$

for constant $C = \sup_{x\in[a,b]}$.

$$n^{-\delta/2}E\left(\int_a^b Z_{n_{11}}^{1+\delta/2}(x)dx\right)$$

$$= n^{-\delta/2}E\int_a^b Z_{n_{11}}^{1+\delta/2}(x)dx \int_a^b Z_{n_{11}}^{1+\delta/2}(y)dy$$

$$= \frac{1}{n^{\delta/2}h_1^{2+\delta}}\int_{-\infty}^\infty \int_a^b \int_a^b \left[K^*\left(\frac{x-u}{h_1}\right)\right]^{(1+\delta/2)}\left[K^*\left(\frac{y-u}{h_1}\right)\right]^{(1+\delta/2)} f(u)dxdydu$$

$$= \frac{1}{n^{\delta/2}h_1^{2+\delta}}\left(\int_{-\infty}^a + \int_a^b + \int_b^\infty\right)\int_a^b \int_a^b \left[K^*\left(\frac{x-u}{h_1}\right)\right]^{(1+\delta/2)}\left[K^*\left(\frac{y-u}{h_1}\right)\right]^{(1+\delta/2)} f_{1U}(u)dxdydu$$

$$:= N_1 + N_2 + N_3.$$

$$(76)$$

46

$$N_2 = \frac{1}{n^{\delta/2} h_1^{2+\delta}} \int_a^b \int_a^b \int_a^b \left[K^* \left(\frac{x-u}{h_1} \right) \right]^{(1+\delta/2)} \left[K^* \left(\frac{y-u}{h_1} \right) \right]^{(1+\delta/2)} f_{1U}(u) dx\, dy\, du$$

$$= \frac{1}{n^{\delta/2} h_1^{\delta}} \int_a^b \int_{(a-u)/h_1}^{(b-u)/h_1} \int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(s)]^{(1+\delta/2)} [K^*(t)]^{(1+\delta/2)} f_{1U}(u) ds\, dt\, du \qquad (77)$$

$$= \frac{1}{n^{\delta/2} h_1^{\delta}} \int_a^b \int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(s)]^{(1+\delta/2)} ds \int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(t)]^{(1+\delta/2)} dt\, f_{1U}(u) du.$$

Define

$$F(x) = \int_{-\infty}^x [K^*(s)]^{(1+\delta/2)} ds. \qquad (78)$$

Then,

$$\int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(s)]^{(1+\delta/2)} ds = F \left(\frac{b-u}{h_1} \right) - F \left(\frac{a-u}{h_1} \right). \qquad (79)$$

As $h_1 \to 0, \forall a \le u \le b$,

$$F \left(\frac{b-u}{h_1} \right) \to F(+\infty); \quad F \left(\frac{a-u}{h_1} \right) \to F(-\infty) = 0. \qquad (80)$$

Then,

$$N_2 \sim \frac{1}{n^{\delta/2} h_1^{\delta}} \int_a^b \int_{-\infty}^{\infty} [K^*(s)]^{(1+\delta/2)} ds \int_{-\infty}^{\infty} [K^*(t)]^{(1+\delta/2)} dt\, f_{1U}(u) du. \qquad (81)$$

Here, we employ the results of Fan (1991a). Since $|\phi_\varepsilon(t)| > 0$, there exists a large enough (but fixed) M such that

$$|\phi_\varepsilon(t/h_1)| \ge \min_{|t| \le M} |\phi_\varepsilon(t)| > 0, \text{ when } |t| \le M h_1. \qquad (82)$$

By the assumption of the ordinary smooth error distribution, $|\phi_\varepsilon(t) t^\tau \ge c$ for some positive

constant c as $t \to \infty$. Then, when $|t| \geq Mh$

$$|\phi_\varepsilon(t/h_1)| \geq \frac{c}{2}(t/h_1)^{-\tau}$$
$$\Rightarrow \left| \frac{\phi_K(t)}{\phi_\varepsilon(t/h_1)} \right| \leq \left| \frac{2t^\tau \phi_K(t)}{ch_1^\tau} \right|. \tag{83}$$

Combining (82) and (83) we have,

$$\left| \frac{h_1^\tau \phi_K(t)}{\phi_\varepsilon(t/h_1)} \right| \leq \max \left\{ \left| \frac{2t^\tau \phi_K(t)}{c} \right| 1_{|t| \geq Mh_1}, \frac{\max |\phi_K(t)|}{\min_{|t| \leq M} |\phi_\varepsilon(t)|} 1_{|t| \geq Mh_1} \right\} \leq g_0(t), \tag{84}$$

where g_0 is a positive integrable function. This implies

$$|h_1^\tau K^*(s)| \leq \int_{-\infty}^\infty g_0(t)dt := C_1. \tag{85}$$

On the other hand, equation (2.7) of Fan (1991a) implies

$$|h_1^\tau K^*(s)| \leq \frac{C_2}{|s|}. \tag{86}$$

Combining these two bounds, we have

$$|h_1^\tau K^*(s)| \leq \min \left\{ C_1, \frac{C_2}{|s|} \right\} \tag{87}$$

for some constants C_1 and C_2 independent of n and s. Let $M(s) = \min\{C_1, C_2/|s|\}$. Then,

48

$|M(s)|^{1+\delta/2}$ is integrable for $\delta > 0$.

$$\int_{-\infty}^{\infty} [K^*(s)]^{(1+\delta/2)} ds$$

$$= \frac{1}{h_1^{\tau(1+\delta/2)}} \int_{-\infty}^{\infty} [h^\tau K^*(s)]^{1+\delta/2} ds$$

$$\leq \frac{1}{h_1^{\tau(1+\delta/2)}} \int_{-\infty}^{\infty} |M(s)|^{1+\delta/2} ds \qquad (88)$$

$$= O(h_1^{-\tau(1+\delta/2)}).$$

Then,

$$N_2 \sim \frac{1}{n^{\delta/2} h_1^{\delta+\tau(2+\delta)}} \to 0. \qquad (89)$$

We then show $N_1 \to 0$ and $N_3 \to 0$.

$$N_1 = \frac{1}{n^{\delta/2} h_1^{2+\delta}} \int_{-\infty}^{a} \int_{a}^{b} \int_{a}^{b} \left[K^* \left(\frac{x-u}{h_1} \right) \right]^{(1+\delta/2)} \left[K^* \left(\frac{y-u}{h_1} \right) \right]^{(1+\delta/2)} f_{1U}(u) dx dy du$$

$$= \frac{1}{n^{\delta/2} h_1^{\delta}} \int_{-\infty}^{a} \int_{(a-u)/h_1}^{(b-u)/h_1} \int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(s)]^{(1+\delta/2)} [K^*(t)]^{(1+\delta/2)} f_{1U}(u) ds dt du \qquad (90)$$

$$= \frac{1}{n^{\delta/2} h_1^{\delta}} \int_{-\infty}^{a} \int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(s)]^{(1+\delta/2)} ds \int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(t)]^{(1+\delta/2)} dt f_{1U}(u) du.$$

As $h_1 \to 0, \forall u \leq a$,

$$\int_{(a-u)/h_1}^{(b-u)/h_1} [K^*(s)]^{(1+\delta/2)} ds \to F(\infty) - F(\infty) = 0. \qquad (91)$$

Then, $N_1 \to 0$. Using similar argument, $N_3 \to 0$. By (72),

$$E \int_a^b \hat{f}_{n1}(x)g(x)dx = \frac{1}{n_1} \sum_{i=1}^{n_1} EW_{n_1}$$

$$= EW_{n_1} \tag{92}$$

$$= \int_a^b f_1(x)g(x)dx + O(h_1^m).$$

By Lemma 3, we have

$$Var\left(\int_a^b \hat{f}_{n1}(x)g(x)dx\right)$$

$$= Var\left(\frac{1}{n_1} \sum_{i=1}^{n_1} \int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g(x)dx\right)$$

$$= \frac{1}{n_1} Var\left(\int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g(x)dx\right) \tag{93}$$

$$= \frac{1}{n_1}\left[E\left(\int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g(x)dx\right)^2 - \left(E\int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g(x)dx\right)^2\right]$$

$$= \frac{1}{n_1}\left[\int_a^b g^2(x)f_{1U}(x)dx - \left(\int_a^b f_1(x)g(x)dx + O(h_1^m)\right)^2\right].$$

Replacing $E\int_a^b \hat{f}_{n1}(x)g(x)dx$ and $Var(\hat{f}_{n_1}(x)g(x)dx)$ in (69) and by assumption of h_1

$$\frac{\sqrt{n_1}\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\}g(x)dx}{s_1} \to N(0,1). \tag{94}$$

By similar argument we can show

$$\frac{\sqrt{n_0}\int_a^b \{\hat{f}_{n_0}(x) - f_0(x)\}g(x)dx}{s_0} \to N(0,1). \tag{95}$$

□

Proof of Theorem 1. We consider the first term of the sum in (33). By lemma 1 and 3, we have

$$E\left[\int_a^b \{\hat{f}_{n_1} - f_1(x)\}g_{21}(x)dx\right]^2$$

$$= E\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right]^2$$

$$= \frac{1}{n_1^2}E\left[\sum_{i=1}^{n_1}\left(\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right)^2\right.$$

$$\left. + 2\sum_{i=1}^{n_1}\sum_{\substack{j=1\\i\neq j}}^{n_1}\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1j}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right]$$

$$= \frac{1}{n_1^2}\left[n_1 E\left(\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right)^2\right.$$

$$+ n_1(n_1-1)\left(E\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right)$$

$$\left.\left(E\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1j}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right)\right]$$

$$= O\left(n_1^{-1} + n_1^{-1}h_1^{2m} + h_1^{2m}\right) \to 0.$$

(96)

By similar argument, the second term is $O(n_0^{-1} + n_0^{-1}h_0^{2m} + h_0^{2m})$ and converges to 0. By independence of the case and control groups

$$E\left[\int_a^b \{\hat{f}_{n_1} - f_1(x)\}g_{21}(x)dx\int_a^b \{\hat{f}_{n_0} - f_0(x)\}g_{20}(x)dx\right]$$

$$= E\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\int_a^b \left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right) - f_1(x)\right\}g_{21}(x)dx\right]$$

(97)

$$E\left[\frac{1}{n_0}\sum_{j=1}^{n_0}\int_a^b \left\{\frac{1}{h_0}K^*\left(\frac{x-U_{0j}}{h_0}\right) - f_0(x)\right\}g_{20}(x)dx\right].$$

By lemma 1,

$$E\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\int_a^b\left\{\frac{1}{h_1}K^*\left(\frac{x-U_{1i}}{h_1}\right)-f_1(x)\right\}g_{21}(x)dx\right]\to 0 \text{ as } h_1\to 0,$$

$$E\left[\frac{1}{n_0}\sum_{j=1}^{n_0}\int_a^b\left\{\frac{1}{h_0}K^*\left(\frac{x-U_{0j}}{h_0}\right)-f_0(x)\right\}g_{20}(x)dx\right]\to 0 \text{ as } h_0\to 0.$$

Thus, $\hat{\beta}$ is a consistent estimator of β. Using similar argument, we can show $\hat{\alpha}$ is a consistent estimator of α.

We now show the asymptotic normality. We first show

$$\sqrt{n_1}\begin{pmatrix}\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}g_{11}(x)dx\\[2mm]\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}g_{21}(x)dx\end{pmatrix}$$

is bivariate normal. For $a_{11}, a_{21}\in\mathbb{R}$,

$$\begin{aligned}&\sqrt{n_1}\left[a_{11}\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}g_{11}(x)dx+a_{21}\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}g_{21}(x)dx\right]\\&=\sqrt{n_1}\left[\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}a_{11}g_{11}(x)dx+\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}a_{21}g_{21}(x)dx\right]\\&=\sqrt{n_1}\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}\{a_{11}g_{11}(x)+a_{21}g_{21}(x)\}dx\\&=\sqrt{n_1}\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}g(x)\}dx\end{aligned}\tag{98}$$

in which $g(x)=a_{11}g_{11}(x)+a_{12}g_{12}(x)$. Then $g(x)$ is a continuous function. By lemma 4,

$$\sqrt{n_1}\int_a^b\{\hat{f}_{n_1}(x)-f_1(x)\}g(x)dx\to N(0,s_1^2).$$

in which $s_1^2 = \int_a^b g^2(x) f_{1U}(x) dx - \left\{ \int_a^b f_1(x) g(x) dx \right\}^2$. We now find the covariance matrix Σ_1

$$
\begin{aligned}
&Var\left(\sqrt{n_1} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{11}(x) dx \right) \\
&= n_1 Var\left(\int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{11}(x) dx \right) \\
&= n_1 Var\left(\int_a^b \hat{f}_{n_1}(x) g_{11}(x) dx \right) \\
&= Var\left(\int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g_{11}(x) dx \right) \\
&= \left[E\left(\int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g_{11}(x) dx \right)^2 - \left(E \int_a^b \frac{1}{h_1} K^* \frac{(x-U)}{h_1} g_{11}(x) dx \right)^2 \right] \\
&\to \int_a^b g_{11}^2(x) f_{1U}(x) dx - \left(\int_a^b f_1(x) g_{11}(x) dx \right)^2
\end{aligned}
\tag{99}
$$

by lemma 3 and lemma 1. Similarly,

$$
\begin{aligned}
&Var\left(\sqrt{n_1} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{21}(x) dx \right) \\
&\to \int_a^b g_{21}^2(x) f_{1U}(x) dx - \left(\int_a^b f_1(x) g_{21}(x) dx \right)^2
\end{aligned}
$$

and

$$
\begin{aligned}
&Cov\left(\sqrt{n_1} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{11}(x) dx, \sqrt{n_1} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{21}(x) dx \right) \\
&\to \int_a^b g_{11}(x) g_{21}(x) f_{1U}(x) dx - \int_a^b f_1(x) g_{11}(x) dx \int_a^b f_1(x) g_{21}(x) dx.
\end{aligned}
$$

Then the kl-th entry of Σ_1 for $k, l = 1, 2$ is

$$
\Sigma_1^{(kl)} = \int_a^b g_{k1}(x) g_{l1}(x) f_{1U}(x) dx - \int_a^b f_1(x) g_{k1}(x) dx \int_a^b f_1(x) g_{l1}(x) dx
$$

By similar argument,

$$\sqrt{n_1} \begin{pmatrix} \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{11}(x) dx \\ \int_a^b \{\hat{f}_{n_1}(x) - f_1(x)\} g_{21}(x) dx \end{pmatrix} \to N(0, \Sigma_0),$$

in which

$$\Sigma_0^{(kl)} = \int_a^b g_{k0}(x) g_{l0}(x) f_{0U}(x) dx - \int_a^b f_0(x) g_{k0}(x) dx \int_a^b f_0(x) g_{l0}(x) dx.$$

□