Statistical Methods & Applications
https://doi.org/10.1007/s10260-023-00737-7

ORIGINAL PAPER

Parameter estimation for Logistic errors-in-variables regression under case–control studies

Pei Geng[1] · **Huyen Nguyen**[2]

Accepted: 18 November 2023

Abstract

The article develops parameter estimation in the Logistic regression when the covariate is observed with measurement error. In Logistic regression under the case–control framework, the logarithmic ratio of the covariate densities between the case and control groups is a linear function of the regression parameters. Hence, an integrated least-square-type estimator of the Logistic regression can be obtained based on the estimated covariate densities. When the covariate is precisely measured, the covariate densities can be effectively estimated by the kernel density estimation and the corresponding parameter estimator was developed by Geng and Sakhanenko (2016). When the covariate is observed with measurement error, we propose the least-square-type parameter estimators by adapting the deconvolution kernel density estimation approach. The consistency and asymptotic normality are established when the measurement error in covariate is ordinary smooth. Simulation study shows robust estimation performance of the proposed estimator in terms of bias reduction against the error variance and unbalanced case–control samples. A real data application is also included.

Keywords Case–control study · Deconvolution kernel density estimators · Integrated square distance · Bias reduction

Mathematics Subject Classification 62J12 · 62G07

✉ Pei Geng
pei.geng@unh.edu

[1] Department of Mathematics and Statistics, University of New Hampshire, Durham, NH 03824, USA

[2] Department of Statistics, University of Connecticut, Storrs, CT 06269, USA

Published online: 13 December 2023

1 Introduction

Case-control studies are frequently used for studying the association between health outcomes and risk factors. Logistic regression is one of the most commonly applied statistical methods to model the relationship between binary health outcomes and the factors of interest. In the univariate logistic regression analysis, the response Y is binary ($Y = 0$ or 1), the predictor X is a real value, and the corresponding conditional regression function is

$$\Psi(x) = E(Y = 1|X = x) = \frac{\exp(\alpha^* + \beta x)}{1 + \exp(\alpha^* + \beta x)}. \tag{1.1}$$

Here α^* and β are scalar parameters. Let π denote $P(Y = 1)$ and f be the marginal density of X, then $\pi = \int \Psi(x) f(x) dx$. We denote the conditional density functions of the covariate X as f_0 and f_1 corresponding to the response $Y = 0$ and $Y = 1$, respectively. Then under (1.1), the density functions obey the relation:

$$f_1(x) = \exp(\alpha + \beta x) f_0(x), \quad \alpha = \alpha^* + \ln\{(1 - \pi)/\pi\}. \tag{1.2}$$

This case–control sampling scheme was first introduced by Prentice and Pyke (1979). Since then, various parameter estimation methods were developed by Qin and Zhang (1997), Bondell (2005), and Geng and Sakhanenko (2016). Particularly, Geng and Sakhanenko (2016) observed the linear relationship between the logarithm of the density ratio and the parameters from (1.2) and further proposed the estimation approach based on an integrated square distance (ISD). The ISD estimators were shown to achieve superior performance for the cases of small sample sizes and severely unbalanced samples.

However, the predictor variable X may not be accurately observed in practice. Instead, one observes a surrogate Z of X with measurement error. Specifically, we adapt the errors-in-variables (EIVs) model:

$$Z = X + U. \tag{1.3}$$

Here we assume that U and (X, Y) are independent. Moreover, the density f_U of the measurement error random variable U is known for model identification purpose. See Carroll et al. (2006) for details. In practice, when the measurement error density f_U is unknown, it is natural to estimate the error density from additional data such as validation data, auxiliary variables and instrumental variables. For example, additional negative control data are available in Lumina Bead microarray studies (Xie et al. 2009) for the background noise. In the cases without additional data, parametric assumptions of the distribution of U can be made such as Laplace distribution with unknown variance and the parameters can be estimated under different criteria. See Sects. 4 and 5 for more details. For the Logistic EIVs regression, Stefanski and Carroll (1985) proposed an effective bias-adjusted maximum likelihood estimation approach for normally distributed measurement errors and small variances. Carroll and Wand (1991) developed a semiparametric estimation method when there is a validation data set available for the covariate.

In this paper, we propose the ISD estimator for the Logistic EIVs models based on the deconvolution kernel density estimation and investigate the asymptotic properties as well as simulation performance. We first demonstrate the bias of naive estimators when ignoring measurement errors in covariate, then present the deconvolution ISD estimation in Sect. 2.1. Furthermore, we show the consistency and the asymptotic normality of the proposed estimator and the needed assumptions in Sect. 2.2. In Sect. 3, we conduct simulation study to address the estimation performance under various settings. Moreover, we include a real data application of the Framingham Heart Study in Sect. 4. Practical guidelines and limitations of the proposed method are included in Sect. 5. The proofs of the main results can be found in Sect. 6. Programming R codes for the simulation study and real data application are available in supplemental material.

2 Nonparametric estimation

We first briefly review the ISD estimation method by Geng and Sakhanenko (2016) when the covariate X is error free. When the covariate is observed with measurement error, we show that the naive ISD estimator by ignoring the measurement error is biased in the Gaussian distribution setting. We further propose the ISD estimation approach based on the deconvolution kernel density estimators in Stefanski and Carroll (1990).

2.1 Bias of naive ISD estimators

In the Logistic regression under case–control study, Geng and Sakhanenko (2016) discovered the linear relationship between the log-ratio of the covariate densities and the parameters in (2.1), constructed the integrated square distance in (2.2), and obtained the parameter estimators by minimizing the ISD in (2.3), i.e.,

$$\ln\left(\frac{f_1(x)}{f_0(x)}\right) = \alpha + \beta x, \tag{2.1}$$

$$\tilde{T}_n(s,t) := \int_a^b \left[\ln\left\{\frac{\tilde{f}_1(x)+b_{n_1}}{\tilde{f}_0(x)+b_{n_0}}\right\} - s - tx\right]^2 dx, \quad s \in \mathbb{R}, t \in \mathbb{R}, \tag{2.2}$$

$$(\tilde{\alpha}, \tilde{\beta}) := \mathrm{argmin}_{s,t}\tilde{T}_n(s,t), \tag{2.3}$$

where $\tilde{f}_0(x)$ and $\tilde{f}_1(x)$ are the kernel density estimators of covariate X in the control and case group, respectively, a and b are pre-determined real value constants, b_{n_0} and b_{n_1} are small positive values to ensure the log-ratio is well defined.

In the Logistic EIVs regression under the case–control study, a naive ISD estimator of Geng and Sakhanenko (2016) can be obtained based on the kernel density estimators of samples of Z by ignoring the measurement error.

However, there are two main issues about the naive estimation. First, the yielded naive estimator can be biased. For example, consider the Gaussian distribution setting for both covariate and measurement error. Specifically, when X in the control group is normally distributed $f_0 = N(\mu_X, \sigma_X^2)$, then (1.2) implies that $f_1 = N(\mu_X + \beta\sigma_X^2, \sigma_X^2)$. We denote g_0 and g_1 as the density of Z under the control and case group, respectively. Assume the measurement error U follows a normal distribution $N(0, \sigma_e^2)$ and U is independent of X, then $g_0 = N(\mu_X, \sigma_X^2 + \sigma_e^2)$ and $g_1 = N(\mu_X + \beta\sigma_X^2, \sigma_X^2 + \sigma_e^2)$. It further implies that

$$\ln\left(\frac{g_1(z)}{g_0(z)}\right) = -\frac{(z - \mu_X - \beta\sigma_X^2)^2 - (z - \mu_X)^2}{2(\sigma_X^2 + \sigma_e^2)} = \tilde{\alpha} + \frac{\sigma_X^2}{\sigma_X^2 + \sigma_e^2}\beta z,$$

where $\tilde{\alpha} = \{-2\mu_X\beta\sigma_X^2 - \beta^2\sigma_X^4\}/\{2(\sigma_X^2 + \sigma_e^2)\}$. Therefore, the naive estimator $\hat{\beta}_{\text{naive}}$ based on the integrated least square idea converges to $\frac{\sigma_X^2}{\sigma_X^2+\sigma_e^2}\beta$. This variance ratio factor is well known in the parameter estimation bias of the traditional linear errors-in-variables models. This discovery is also consistent with the finding of bias and attenuation in Stefanski and Carroll (1985). Second, the linear relationship in (2.1) may simply no longer hold for the density of the surrogate Z. For instance, when X is exponentially distributed under the case–control study, $f_0 = \text{Exp}(\lambda)$ with mean $1/\lambda$ and $f_1 = \text{Exp}(\lambda - \beta)$, and U follows Laplace distribution $f_U(u) = \exp\{-|u|/\gamma\}/(2\gamma)$, then g_0 and g_1 are no longer exponentially distributed. In fact,

$$g_0(z) = \begin{cases} \frac{\lambda}{2(1+\lambda\gamma)}e^{z/\gamma}, & z < 0 \\ \frac{\lambda}{1-\lambda^2\gamma^2}e^{-\lambda z} + \frac{\lambda}{2(\lambda\gamma-1)}e^{-z/\gamma}, & z \geq 0, \end{cases} \quad g_1(z) = \begin{cases} \frac{\lambda}{2[1+(\lambda-\beta)\gamma]}e^{z/\gamma}, & z < 0 \\ \frac{\lambda}{1-(\lambda-\beta)^2\gamma^2}e^{-(\lambda-\beta)z} + \frac{\lambda-\beta}{2[(\lambda-\beta)\gamma-1]}e^{-z/\gamma}, & z \geq 0 \end{cases}.$$

Hence, it is critical to take the measurement error model (1.3) into consideration in the ISD estimation. In the following subsection, we adapt the deconvolution kernel density estimation by Stefanski and Carroll (1990) to effectively estimate the densities of the covariate X under case and control groups, and obtain the modified ISD estimators according to (2.2).

2.2 Deconvolution ISD estimation

In this subsection, we describe the data structure for the Logistic EIVs models, present the explicit expression of the proposed ISD estimators and establish their asymptotic normality. We assume that $\{z_i, i = 1, \ldots, n_0\}$ are the observed surrogate values of Z from (1.3) in the control group ($Y_i = 0$); and $\{\tilde{z}_j, j = 1, \ldots, n_1\}$ are the observed surrogate values of Z in the case group ($Y_j = 1$). Let $n = n_0 + n_1$. The measurement error U has known density f_U and $E(U) = 0$. Let K be a kernel density and denote $h_i := h_i(n_i)$, $i = 0, 1$ as the bandwidths for the control and case group, respectively. We adapt the deconvolution kernel density estimators $\hat{f}_0$ and $\hat{f}_1$ by Stefanski and Carroll (1990) for the estimation of f_0 and f_1 as follows. For each $x \in \mathbb{R}$,

$$\hat{f}_0(x) = (n_0 h_0)^{-1} \sum_{i=1}^{n_0} K^*_{h_0}((z_i - x)/h_0), \quad \hat{f}_1(x) = (n_1 h_1)^{-1} \sum_{j=1}^{n_1} K^*_{h_1}((\tilde{z}_j - x)/h_1), \tag{2.4}$$

$$K^*_{h_j}(y) = (2\pi)^{-1} h_j^{-1} \int \exp(-ity) \frac{\phi_K(t)}{\phi_U(t/h_j)} dt, \quad j = 0, 1, \tag{2.5}$$

where ϕ_K and ϕ_U are the characteristic functions of a chosen kernel density K and the measurement error density f_U, respectively. Furthermore, we define the integrated square distance and the modified ISD estimators $\hat{\alpha}$ and $\hat{\beta}$ as

$$T_n(s,t) := \int_a^b \left[\ln \left\{ \frac{(\hat{f}_1(x) \vee 0) + b_{n_1}}{(\hat{f}_0(x) \vee 0) + b_{n_0}} \right\} - s - tx \right]^2 dx, \quad s \in \mathbb{R},\ t \in \mathbb{R},$$
$$(\hat{\alpha}, \hat{\beta}) := \text{argmin}_{s,t} T_n(s,t).$$

Here $x \vee y = \max(x, y)$ for any $x, y \in \mathbb{R}$. Due to the fact that the deconvolution density estimator based on finite samples may take negative values when $|x|$ is large, $\hat{f}_i(x) \vee 0$ is used to truncate those negative values to 0. The finite constants a, b and the positive sequences b_{n_i} are chosen similarly as in Geng and Sakhanenko (2016). Details can be found in the Sect. 3.

Similar to the ISD estimation in Geng and Sakhanenko (2016), the unique solution to the optimization problem can be written as

$$\begin{aligned} \hat{\beta} &= \frac{12}{(b-a)^3} \int_a^b \left[\ln \left\{ \frac{(\hat{f}_1(x) \vee 0) + b_{n_1}}{(\hat{f}_0(x) \vee 0) + b_{n_0}} \right\} \right] \left\{ x - \frac{a+b}{2} \right\} dx, \\ \hat{\alpha} &= \frac{1}{b-a} \int_a^b \left[\ln \left\{ \frac{(\hat{f}_1(x) \vee 0) + b_{n_1}}{(\hat{f}_0(x) \vee 0) + b_{n_0}} \right\} - \hat{\beta} x \right] dx. \end{aligned} \tag{2.6}$$

Note that the deconvolution kernel density estimators $\hat{f}_0$ and $\hat{f}_1$ play a key role in the proposed estimation. As established in Carroll and Hall (1988) and Fan (1991), the asymptotic behavior of the deconvolution kernel density estimators heavily depend on the tail of ϕ_U. Hence the measurement errors were separated into two cases:

(1) ordinary smooth case: $\phi_U(t) = O(t^{-\tau})$ as $t \to \infty$ for some $\tau \geq 0$;
(2) supersmooth case: $|\phi_U(t)| = O(t^{-\alpha_0} \exp(-t^{\beta_0}/\gamma_0))$ for some $\beta_0 > 0$, $\gamma_0 > 0$ and real number α_0.

In this paper, we focus on the ordinary smooth case as shown in the assumptions below. Examples of ordinary smooth distributions include gamma, uniform, and Laplace distributions. For instance, the characteristic function of a uniform distribution on $(-a, a)$ is $\phi_U(t) = \sin(at)/(at) = O(t^{-1})$ which belongs to the ordinary smooth case with $\tau = 1$. A second example of ordinary smooth errors is Laplace distribution whose density $f_U(u) = \exp\{-|u|/\gamma\}/(2\gamma)$ and characteristic function

$\phi_U(t) = (1+\gamma^2 t^2)^{-1} = O(t^{-2})$ with $\tau = 2$. The supersmooth case which includes the normal and Cauchy errors is worth of future study.

We now state the assumptions needed for establishing the asymptotic normality of $(\hat{\alpha}, \hat{\beta})$.

$$f_0 \text{ and } f_1 \text{ have } m \text{ continuous derivatives, for } m \geq 2. \tag{2.7}$$

$$\phi_K \text{ is a symmetric function; } \phi_K \text{ has } m+2 \text{ bounded integrable derivatives;} \tag{2.8}$$

$$\begin{aligned} &\phi_K(0) = 1; \phi_K(t) = 1 + O(t^m) \text{ as } t \to 0. \\ &|\phi_U(t)| > 0 \text{ for all real } t; d_0|t|^{-\tau} \leq |\phi_U(t)| \leq d_1|t|^{-\tau} \text{ as } t \to \infty \text{ for some constants} \end{aligned} \tag{2.9}$$

$$\begin{aligned} &0 < d_0 \leq d_1 \text{ and } \tau > 0. \text{ Moreover, } \int_{-\infty}^{\infty} \{|\phi_K(t)| + |\phi_K'(t)|\}|t|^{\tau} dt < \infty. \\ &\text{There exists } \delta_i > 0 \text{ such that } n_i^{\delta_i/2} h_i^{\delta_i + \tau(2+\delta_i)} \to \infty \text{ for } i = 0, 1. \end{aligned} \tag{2.10}$$

$$h_i \to 0 \text{ and } n_i h_i \to \infty \text{ for } i = 0, 1. \tag{2.11}$$

$$n_i h_i^2 \to \infty \text{ and } n_i h_i^{2m} \to 0 \text{ for } i = 0, 1. \tag{2.12}$$

$$n_i h_i / \log n_i \to \infty \text{ and } |\log h_i| / \log(\log n_i) \to \infty \text{ for } i = 0, 1. \tag{2.13}$$

$$n_i^{1/2} b_{n_i} \to 0 \text{ for } i = 0, 1. \tag{2.14}$$

$$n_1/n \to \rho,\ 0 \leq \rho \leq 1,\ n = n_1 + n_0. \tag{2.15}$$

Assumption (2.8) implies that the kernel $K(x)$ is symmetric and of m-th order, i.e, all moments less than m of the kernel are equal to 0. The assumption of the higher-order kernel is used to ensure the bias of the deconvolution kernel density estimators converges to zero sufficiently fast. The assumption of ϕ_U in (2.9) indicates that the error density is ordinary smooth. The assumption of ϕ_K in (2.9), along with the bandwidth assumption (2.10), is used to verify the Lyapunov's Central Limit Theorem conditions for the proposed estimator $(\hat{\alpha}, \hat{\beta})$ in Theorems 2.1 and 2.2. The bandwidth assumption (2.11) is used in Lemma 5.1 below to establish the consistency of the deconvolution kernel density estimators $\hat{f}_i, i = 0, 1$. The assumption (2.12) is imposed to eliminate the bias of $\hat{f}_i, i = 0, 1$ as shown in Lemma 5.2. The assumption (2.13) is used in Lemma 5.2 to obtain the upper bound of $\sup_x |\hat{f}_i(x) - E\hat{f}_i(x)|, i = 0, 1$ by Giné and Guillou (2002). To make the assumptions more explicit, consider the case that the true covariate $f_i, i = 0, 1$ follows a normal distribution and the measurement error f_U follows a Laplace distribution. In this case, $m = \infty$ in (2.7) and $\tau = 2$ in (2.9). One possible bandwidth choice that satisfies (2.11)–(2.13) is

$h_i = O(n_i^{-a}), i = 0, 1$ with $a < 1/6$. Consequentially, there exists $\delta_i > 8a/(1-6a)$ that satisfies (2.10).

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

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

Now we present the asymptotic results of the proposed estimator. Proof details can be found in Sect. 6. Denote $\to_D$ as the convergence in distribution.

Theorem 2.1 $(0 < \rho < 1)$. *Under models* (1.1) *and* (1.3), *when assumptions* (2.7)–(2.14) *hold, and* $0 < \rho < 1$ *in* (2.15), *we have that* $\hat{\alpha}$ *and* $\hat{\beta}$ *are consistent estimators of* α *and* β, *respectively. Moreover,*

$$\sqrt{n}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix} \to_D N(0,\Sigma), \Sigma = \rho^{-1}\Sigma_1 + (1-\rho)^{-1}\Sigma_0,$$

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

$$\Sigma_i^{(kl)} = \int_a^b g_{ki}(x)g_{li}(x)g_i(x)dx - \int_a^b f_i(x)g_{ki}(x)dx \int_a^b f_i(x)g_{li}(x)dx, \quad \text{for } i = 0, 1,$$
$$g_i(z) = \int f_i(x)f_U(z-x)dx. \tag{2.16}$$

Theorem 2.2 $(\rho = 0$ **or** $1)$. *Under models* (1.1) *and* (1.3), *when assumptions* (2.7)–(2.14) *hold, we have*

$$\sqrt{n_1}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix} \to_D N(0,\Sigma_1), \text{ for } \rho = 0; \quad \sqrt{n_0}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix} \to_D N(0,\Sigma_0), \text{ for } \rho = 1,$$

where Σ_0 *and* Σ_1 *are defined as in Theorem* 2.1.

The covariance matrix Σ in Theorem 2.1 is the weighted sum of Σ_1 and Σ_0 with the coefficients ρ^{-1} and $(1-\rho)^{-1}$ determined by the sample size ratio of case and control groups. Theorem 2.2 shows the asymptotic result of the estimators when the sample sizes of the case and control groups are severely unbalanced. In each *kl*-th entry of the matrix Σ_i, the integrand of the first integration term includes the surrogate density $g_i, i = 0, 1$. When the covariate is observed free of error, the surrogate density g_i coincides with the true covariate density f_i, hence our covariance structure degenerates to that of Geng and Sakhanenko (2016).

3 Simulation study

In this section, we consider the Logistic EIVs regression in (1.1) and (1.3) with the measurement error distribution chosen as the Laplace distribution $L(0, \gamma)$ where

$$f_U(u) = \exp\{-|u|/\gamma\}/(2\gamma), \quad \phi_U(t) = (1+\gamma^2 t^2)^{-1}.$$

One can see that the Laplace distribution $L(0,\gamma)$ satisfies the ordinary smooth assumption in (2.9) with $\tau = 2$. Moreover, the measurement error variance $\sigma_U^2 = 2\gamma^2$.

This simulation study aims to investigate the effectiveness and robustness of the proposed deconvolution ISD estimator $\hat{\beta}$ in estimation bias reduction under different choices of the covariate distribution, sample sizes and error variance. Various scenarios are simulated to evaluate the estimation performance based on the following factors: 1) increasing sample sizes (n_0, n_1) when $n_0/n_1 = 1, 2$) small or large error variance σ_U^2, 3) the sample size ratio n_1/n_0 for unbalanced case control studies.

To generate the case–control data, we specifically consider two distribution cases of covariate X with Case 1 as Gaussian and Case 2 as Exponential. For each case and each combination of chosen values of (n_0, n_1) and σ_U^2, we generate $\{x_i, i = 1, \ldots, n_0\}$ from f_0 and measurement error $\{u_i, i = 1, \ldots, n_0\}$ from $L(0, \gamma = \sigma_U/\sqrt{2})$ $L(0, \gamma = \sigma_U/\sqrt{2})$, then we form the observed surrogate $\{z_i, i = 1, ...n_0\}$ with $z_i = x_i + u_i$ for the control group. Similarly, $\{\tilde{z}_j, j = 1, ...n_1\}$ is simulated with $\tilde{z}_j = \tilde{x}_j + \tilde{u}_j$ for the case group, where $\tilde{x}_j \sim f_1$ and $\tilde{u}_j \sim L(0, \gamma = \sigma_U/\sqrt{2})$. We simulate 500 replicates for each scenario to present the estimation results.

Four estimation methods are computed based on the simulated case–control data $\{z_i, i = 1, ...n_0\}$ and $\{\tilde{z}_j, j = 1, ...n_1\}$. Note that the parameter α is not identifiable due to the unknown parameter π in (1.2), therefore we only focus on the estimation of β. Specifically, we compared the estimation bias and root mean square error (RMSE) of the proposed estimator $\hat{\beta}$ with three existing estimators: the naive ISD estimator $\hat{\beta}_{ISD}$, the naive MLE estimator $\hat{\beta}_{MLE}$ described in Dobson and Barnett (2018), and the bias-corrected estimator $\hat{\beta}_{BC}$ by Stefanski and Carroll (1985). The proposed estimator $\hat{\beta}$ is computed by (2.6) and (2.4) while the naive ISD estimator is calculated by

$$\hat{\beta}_{ISD} = \frac{12}{(b-a)^3} \int_a^b \left[\ln \left\{ \frac{\tilde{f}_1(x) + b_{n_1}}{\tilde{f}_0(x) + b_{n_0}} \right\} \right] \left\{ x - \frac{a+b}{2} \right\} dx,$$

in which $\tilde{f}_0(x)$ and $\tilde{f}_1(x)$ are the standard kernel density estimators using the observed contaminated covariate data in the control and case group, respectively. The naive MLE estimator, $\hat{\beta}_{MLE}$ is the solution that maximizes the log-likelihood function

$$\ell = \sum_{i=1}^{n} \Big[y_i \ln\{\Psi(z_i)\} + (1-y_i)(1 - \ln\{\Psi(z_i)\}) \Big],$$

in which $\Psi(x)$ is defined as in (1.1). The bias-corrected estimator can be calculated by

$$\hat{\beta}_{BC} = (1 - \sigma_U^2 \hat{B}_n)\hat{\beta}_{MLE},$$

in which σ_U^2 is the known measurement error variance and

$$\hat{B}_n = \frac{n}{\sum_{i=1}^{n} \Psi'(z_i\hat{\beta}_{MLE})z_i^2}\left[-\frac{1}{2n}\sum_{i=1}^{n}\Psi''(z_i\hat{\beta}_{MLE})z_i\hat{\beta}_{MLE} - \frac{1}{n}\sum_{i=1}^{n}\Psi'(z_i\hat{\beta}_{MLE})\right].$$

Since the proposed ISD estimation relies on the deconvolution kernel density estimators as defined in (2.4), the choices of kernel function K and bandwidth $h_i, i = 0, 1$ are two key elements. First, the kernel function K should be carefully chosen so that the integral in (2.5) exists. For ordinary smooth errors, as addressed in Yi et al. (2021), commonly used kernel functions include the standard Gaussian kernel, the sinc kernel $K_1(x) = \sin(x)/(\pi x)$ with $\phi_{K_1}(t) = I\{t| \le 1\}$ and the kernel $K_2(x)$ defined through its characteristic function $\phi_{K_2}(t) = (1-t^2)^3 I\{t| \le 1\}$. In this simulation study, we set the density kernel $K(x)$ to be the Gaussian kernel so that the deconvolution kernel $K^*(x)$ takes the following form when the measurement error follows Laplace$(0, \gamma)$

$$K_h^*(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right)\left[1 + \frac{\gamma^2(1-x^2)}{h^2}\right]. \tag{3.1}$$

Note that $K^*(x)$ may take negative values when $|x|$ is large, for the log ratio of the densities to be well-defined, we use truncate $\hat{f}_i(x)$ to zero when it is negative and further set $b_{n_i} = n_i^{-1}$.

Second, for the bandwidth selection in the deconvolution kernel estimation, we adopted the bootstrap bandwidth selection method proposed by Delaigle and Gijbels (2004a). The method first requires obtaining a pilot bandwidth using the rule of thumb method $h_{\text{pilot}} = O(n^{-1/9})$ for Laplace error. This bandwidth and the contaminated observed data are then used to obtain the pseudo deconvolution kernel density estimator $\hat{f}_X(x;h_{\text{pilot}})$. Next, a bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$ is drawn from $\hat{f}_X(x;h_{\text{pilot}})$ and the error U 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 mean integrated square error between $\hat{f}_X(x, h_{pilot})$ and $\hat{f}_X^*(x;h)$. The function *bw.dboot2* in the R package *decon* is available to obtain the bandwidths $h_i, i = 0, 1$, under the Gaussian kernel. The bootstrap bandwidth selection method is proven to be consistent by Delaigle and Gijbels (2004a) and its performance in deconvolution density estimation is shown via simulation study by Delaigle and Gijbels (2004b) to be superior to other commonly used methods such as cross-validation. For the naive ISD estimator, we follow the setting in Geng and Sakhanenko (2016) and used the bandwidth $w_i = n_i^{-1/3}$. The integral limits (a, b) are chosen as the sample means of the two contaminated samples.

Case 1: Gaussian covariates. In this case, we consider that the true covariate is symmetrically distributed with Gaussian distributions, $f_0 = N(0, 1)$ for the control group and $f_1 = N(\beta, 1)$ for the case group obeying (1.2). Particularly, we set $\beta = 2$. To investigate the effect of the error variance, two choices of $\sigma_U^2 = 0.5^2$ and $\sigma_U^2 = 1$ are chosen to represent small error and large error, respectively. The sample size

Table 1 Bias and RMSE comparison of estimators with error variance $\sigma_U^2 = 0.5^2$ and $n_0/n_1 = 1$ under Case 1

Estimator		$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
$\hat{\beta}$	|Bias|	0.1462	0.0652	0.0708	0.0851
	RMSE	0.5663	0.4349	0.3458	0.2505
$\hat{\beta}_{ISD}$	|Bias|	0.3556	0.3563	0.3556	0.3628
	RMSE	0.5199	0.4260	0.3950	0.3810
$\hat{\beta}_{MLE}$	|Bias|	0.3459	0.3760	0.3798	0.3885
	RMSE	0.4212	0.3989	0.3908	0.3952
$\hat{\beta}_{BC}$	|Bias|	0.0233	0.0757	0.0822	0.0959
	RMSE	0.3572	0.2043	0.1548	0.1394

Table 2 Bias and RMSE comparison of estimators with error variance $\sigma_U^2 = 1$ and $n_0/n_1 = 1$ under Case 1

Estimator		$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
$\hat{\beta}$	|Bias|	0.1866	0.0398	0.0319	0.0279
	RMSE	0.8953	0.7695	0.7097	0.6044
$\hat{\beta}_{ISD}$	|Bias|	0.7699	0.8067	0.7920	0.8048
	RMSE	0.8580	0.8348	0.8101	0.8139
$\hat{\beta}_{MLE}$	|Bias|	0.9378	0.9673	0.9661	0.9720
	RMSE	0.9537	0.9723	0.9692	0.9734
$\hat{\beta}_{BC}$	|Bias|	0.4488	0.5140	0.5101	0.5246
	RMSE	0.5553	0.5441	0.5294	0.5327

ratio is chosen as $n_0/n_1 = \{1/10, 1/5, 1/3, 1, 5, 10\}$ where $n_0/n_1 = 1$ represents the balanced case and $n_0/n_1 = \{1/10, 1/5, 1/3, 5, 10\}$ represent the unbalanced cases.

Tables 1 and 2 display the bias and RMSE of the four estimators with increased sample sizes for the sample size ratio $n_0/n_1 = 1$ when $\sigma_U^2 = 0.5^2$ and $\sigma_U^2 = 1$, respectively. When the measurement error is small $\sigma_U^2 = 0.5^2$, Table 1 indicates that both the proposed $\hat{\beta}$ and the bias-corrected estimator $\hat{\beta}_{BC}$ attain minimal bias compared to the other two estimators. Table 2 shows that when the measurement error variance is large $\sigma_U^2 = 1$, for each fixed sample size combination, the proposed deconvolution ISD estimator $\hat{\beta}$ achieves the smallest bias among the four estimators while other three estimators show dramatically large bias. Particularly, compared to the naive ISD estimator $\hat{\beta}_{ISD}$, the bias reduction shown in $\hat{\beta}$ is significant for both small and large sample sizes. Moreover, the bias and RMSE of $\hat{\beta}$ decrease as the sample sizes increase for each chosen ratio setting. Overall, we can see that the proposed deconvolution ISD estimator $\hat{\beta}$ shows robust performance against the error variance σ_U^2, however, the bias-corrected estimator $\hat{\beta}_{BC}$ performs poorly when the error is large.

Tables 3 and 4 present the estimation performance for different unbalanced choices of (n_0, n_1) with $n_0/n_1 = \{1/10, 1/3, 1/5, 5, 10\}$ when $\sigma_U^2 = 0.5^2$ and

Table 3 Bias and RMSE comparison of estimators with $\sigma_U^2 = 0.5^2$ and unbalanced sample sizes under Case 1

Estimator		$n_0 = 50$ $n_1 = 150$	$n_0 = 50$ $n_1 = 500$	$n_0 = 100$ $n_1 = 500$	$n_0 = 100$ $n_1 = 1000$	$n_0 = 500$ $n_1 = 100$	$n_0 = 1000$ $n_1 = 100$
$\hat{\beta}$	|Bias|	0.1011	0.1375	0.0770	0.0823	0.0958	0.0862
	RMSE	0.5897	0.5500	0.4891	0.4434	0.4271	0.4799
$\hat{\beta}_{ISD}$	|Bias|	0.3201	0.3599	0.3457	0.3484	0.3576	0.3638
	RMSE	0.5379	0.5307	0.4578	0.4458	0.4455	0.4618
$\hat{\beta}_{MLE}$	|Bias|	0.3398	0.3871	0.3827	0.3881	0.3873	0.4014
	RMSE	0.4355	0.4408	0.4151	0.4123	0.4207	0.4215
$\hat{\beta}_{BC}$	|Bias|	0.1234	0.2569	0.2227	0.2606	0.0642	0.0796
	RMSE	0.3875	0.3570	0.2982	0.3072	0.2589	0.2145

Table 4 Bias and RMSE comparison of estimators with $\sigma_U^2 = 1$ and unbalanced sample sizes under Case 1

Estimator		$n_0 = 50$ $n_1 = 150$	$n_0 = 50$ $n_1 = 500$	$n_0 = 100$ $n_1 = 500$	$n_0 = 100$ $n_1 = 1000$	$n_0 = 500$ $n_1 = 100$	$n_0 = 1000$ $n_1 = 100$
$\hat{\beta}$	|Bias|	0.1947	0.1316	0.1065	0.0853	0.0986	0.1282
	RMSE	0.8299	0.8309	0.7714	0.7575	0.7935	0.7414
$\hat{\beta}_{ISD}$	|Bias|	0.7619	0.7670	0.7979	0.7899	0.7951	0.7964
	RMSE	0.8576	0.8575	0.8458	0.8332	0.8369	0.8368
$\hat{\beta}_{MLE}$	|Bias|	0.9514	1.0065	0.9881	1.0103	0.9913	1.0183
	RMSE	0.9735	1.0171	0.9946	1.015	0.9979	1.0227
$\hat{\beta}_{BC}$	|Bias|	0.5988	0.7602	0.7068	0.7695	0.3415	0.3438
	RMSE	0.6941	0.7882	0.7269	0.7810	0.4181	0.3984

$\sigma_U^2 = 1$, respectively. The proposed estimator $\hat{\beta}$ shows well controlled bias for all the chosen unbalanced scenarios compared to the other three estimators especially when the error variance is large $\sigma_U^2 = 1$ as shown in Table 4. The bias-corrected estimator $\hat{\beta}_{BC}$ performs fairly comparable to $\hat{\beta}$ only when the error variance is small $\sigma_U = 0.5^2$ and n_0/n_1 is large ($n_0/n_1 = 5$ or 10) as shown in Table 3.

Case 2: Exponential covariates. In this case, we consider that the true covariate is skewed with exponential distributions, i.e., $f_0 = \text{Exp}(\lambda)$ with $\lambda = 3$ in the control group and $f_1 = \text{Exp}(\lambda - \beta)$ in the case group. We set $\beta = 2$. The Laplace measurement error variance is chosen as $\sigma_U^2 = 0.1^2$ and $\sigma_U^2 = 0.2^2$. The sample size ratio is chosen as $n_0/n_1 = \{1/10, 1/5, 1, 5, 10\}$. Similar to Case 1, Tables 5 and 6 present the performance of bias and RMSE of the four estimators with increased sample sizes for the balanced case when $\sigma_U^2 = 0.1^2$ and $\sigma_U^2 = 0.2^2$, respectively. Tables 7 and 8 show the estimation performance for different unbalanced choices of (n_0, n_1) with $n_0/n_1 = \{1/10, 1/5, 1, 5, 10\}$ when $\sigma_U^2 = 0.1^2$ and $\sigma_U^2 = 0.2^2$, respectively. For both balanced and unbalanced sample sizes, when the error variance is smaller $\sigma_U^2 = 0.1^2$, both $\hat{\beta}$ and $\hat{\beta}_{BC}$ give small bias than the other two methods. When the error variance

Table 5 Bias and RMSE comparison of estimators with error variance $\sigma_U^2 = 0.1^2$ and $n_0/n_1 = 1$ under Case 2

Estimator		$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
$\hat{\beta}$	|Bias|	0.0390	0.0016	0.0445	0.0179
	RMSE	0.9879	0.6789	0.5722	0.3872
$\hat{\beta}_{ISD}$	|Bias|	0.2398	0.0980	0.0238	0.05467
	RMSE	0.6253	0.4868	0.4262	0.3112
$\hat{\beta}_{MLE}$	|Bias|	0.0413	0.0810	0.0758	0.0909
	RMSE	0.3903	0.2338	0.1765	0.1458
$\hat{\beta}_{BC}$	|Bias|	0.0308	0.0141	0.0088	0.0254
	RMSE	0.4190	0.2351	0.1710	0.1246

Table 6 Bias and RMSE comparison of estimators with error variance $\sigma_U^2 = 0.2^2$ and $n_0/n_1 = 1$ under Case 2

Estimator		$n_0 = 100$ $n_1 = 100$	$n_0 = 300$ $n_1 = 300$	$n_0 = 500$ $n_1 = 500$	$n_0 = 1000$ $n_1 = 1000$
$\hat{\beta}$	|Bias|	0.0073	0.0207	0.0623	0.0102
	RMSE	1.1089	0.7773	0.6311	0.4748
$\hat{\beta}_{ISD}$	|Bias|	0.4241	0.2840	0.2074	0.2278
	RMSE	0.7163	0.5510	0.4489	0.3775
$\hat{\beta}_{MLE}$	|Bias|	0.2800	0.3070	0.3025	0.3172
	RMSE	0.4366	0.3640	0.3329	0.3326
$\hat{\beta}_{BC}$	|Bias|	0.0730	0.1121	0.1074	0.1264
	RMSE	0.4209	0.2643	0.2011	0.1757

Table 7 Bias and RMSE comparison of estimators with error variance $\sigma_U^2 = 0.1^2$ under severely unbalanced cases under Case 2

Estimator		$n_0 = 100$ $n_1 = 500$	$n_0 = 100$ $n_1 = 1000$	$n_0 = 200$ $n_1 = 1000$	$n_0 = 1000$ $n_1 = 200$	$n_0 = 1000$ $n_1 = 100$	$n_0 = 500$ $n_1 = 100$
$\hat{\beta}$	|Bias|	0.0633	0.0155	0.0046	0.0030	0.0150	0.0441
	RMSE	0.8707	0.8493	0.6688	0.5213	0.6663	0.7321
$\hat{\beta}_{ISD}$	|Bias|	0.3089	0.3573	0.1711	0.0034	0.0447	0.0514
	RMSE	0.5626	0.5715	0.4541	0.3975	0.4540	0.4955
$\hat{\beta}_{MLE}$	|Bias|	0.0955	0.1258	0.1114	0.0530	0.0388	0.0416
	RMSE	0.3115	0.3213	0.2337	0.1799	0.2284	0.2436
$\hat{\beta}_{BC}$	|Bias|	0.0507	0.0864	0.0682	0.0500	0.0766	0.0634
	RMSE	0.3157	0.3208	0.2260	0.1942	0.2587	0.2701

is larger $\sigma_U^2 = 0.2^2$, the proposed estimator $\hat{\beta}$ outperforms all other methods with the least bias for all choices of (n_0, n_1).

Table 8 Bias and RMSE comparison of estimators with error variance $\sigma_U^2 = 0.2^2$ under severely unbalanced cases under Case 2

Estimator		$n_0 = 100$ $n_1 = 500$	$n_0 = 100$ $n_1 = 1000$	$n_0 = 200$ $n_1 = 1000$	$n_0 = 1000$ $n_1 = 200$	$n_0 = 1000$ $n_1 = 100$	$n_0 = 500$ $n_1 = 100$
$\hat{\beta}$	|Bias|	0.0597	0.1013	0.0278	0.0433	0.0836	0.0801
	RMSE	0.9167	0.9546	0.7709	0.6153	0.6939	0.8012
$\hat{\beta}_{ISD}$	|Bias|	0.5171	0.5619	0.3945	0.1594	0.0870	0.1279
	RMSE	0.6998	0.7143	0.5654	0.4332	0.4534	0.5003
$\hat{\beta}_{MLE}$	|Bias|	0.3850	0.4276	0.3989	0.2221	0.1918	0.2193
	RMSE	0.4537	0.4855	0.4336	0.2743	0.2869	0.3124
$\hat{\beta}_{BC}$	|Bias|	0.2635	0.3221	0.2808	0.0946	0.1666	0.1006
	RMSE	0.3825	0.4146	0.3422	0.2293	0.3267	0.3071

In summary, the proposed deconvolution ISD estimator shows robust performance in bias reduction for different choices of covariate distributions, error variance and sample sizes. Particularly, the proposed estimator shows superior performance when the error variance is large for both balanced and severely unbalanced case control studies.

4 Data Application

In this section, we apply our proposed method to the Framingham Heart Study data to investigate the relationship between the systolic blood pressure (SBP) X and the presence of cardiovascular disease Y. The Framingham Heart Study was begun in 1948 to explore risk factors and consequences of cardiovascular disease in a longitudinal population-based cohort and it is one of the longest running epidemiological studies conducted under the National Heart, Lung, and Blood Institute. The variables of interest are the repeated systolic blood pressure measurements and the presence or absence of cardiovascular disease of 1615 patients. This example was first described by Carroll et al. (2006) to study the effect of measurement error. The data is available as *framingham* in the software R package *deconvolve*. During the first visit, each subject had two SBP measurements Z_1 and Z_2. There are 128 individuals with cardiovascular disease and 1487 individuals without the disease. To fit the case–control framework, we generated a nested case–control dataset (Clayton and Hills 2013) by matching 5 controls for each case with cardiovascular disease according to their age and smoking status. In the case–control dataset, there are $n_1 = 127$ cases and $n_0 = 566$ controls due to some cases could not be matched. The two SBP measurements Z_1 and Z_2 are treated as the surrogate of the true value. We use the average of the two blood pressure measurements of each subject as the "true" measurement X of the subject and the difference between Z_1, Z_2 and the average as the measurement error U. The Normal Q-Q plot of the measurement error sample in Fig. 1 indicates a heavier tail than the normal distribution (The p-value $= 1.7 \times 10^{-9}$ for the Shapiro-Wilk

Table 9 Parameter estimation using the deconvolution ISD method, naive ISD, naive maximum likelihood and bias-corrected methods in the Framingham Heart Study data

	First measurement Z_1	Second meausrement Z_2
$\hat{\beta}$	0.022709	0.023068
$\hat{\beta}_{ISD}$	0.025736	0.027160
$\hat{\beta}_{MLE}$	0.014023	0.012568
$\hat{\beta}_{BC}$	0.014032	0.012578

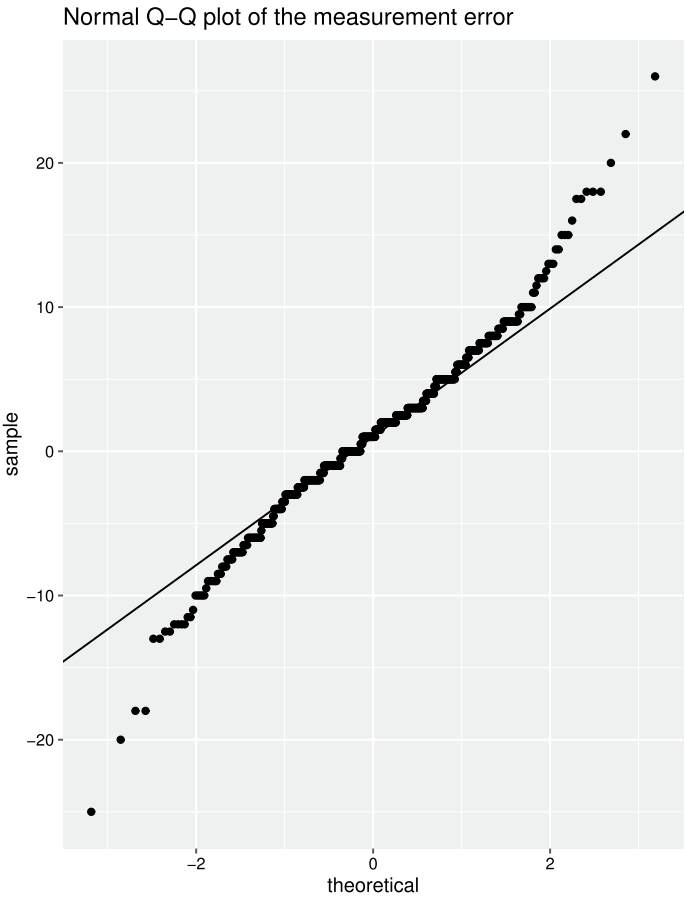

Fig. 1 Normal Q-Q plot for the measurement error in SBP

normality test). Hence we assume the error U follows a Laplace distribution. The estimated measurement error variance is $\hat{\sigma}_U^2 = 5.5386^2$. In Table 9, the four estimators described in Sect. 3 are shown based on the two SBP measurements. It appears that the ISD methods detect larger effects of SBP on the cardiovascular

disease compared to other two methods. The bias-corrected estimation is very close to the naive MLE estimation for both measurements while the deconvolution ISD estimator is reduced from the naive ISD estimator for both measurements.

5 Discussion

In practice, there are great potential application opportunities of the proposed method in many real data studies. For example, the National Health and Nutrition Examination Survey (NHANES) was designed to assess the health and nutritional status of adults and children in the United States. Many variables in NHANES were collected with measurement errors and raised attention to researchers and practitioners such as physical activity level (Tooze et al. 2013), Body Mass Index measures (Stommel and Schoenborn 2009) and sodium intake (Va et al. 2019). With the concerns in measurement error, many studies started collecting validation samples to investigate the errors. The validation data make it possible to confirm the error distribution as required in the proposed method. Additionally, numerous papers have focused on the estimation correction using validation data (Lee and Sepanski 1995; Thürigen et al. 2000; Siddique et al. 2019).

If no extra data is available for the measurement error distribution, an alternative approach is to assume the error distribution with unknown parameters such as Laplace with unknown variance σ_U^2. A grid search of the parameter values can be performed to select the optimal value to minimize the approximated mean integrated squared error (MISE) of the deconvolution density estimator in Stefanski and Carroll (1990). For instance, if the covariate X is normally distributed, the estimated MISE is

$$\widehat{MISE}(h,\sigma_U^2) = \frac{1}{2\pi nh}\int \frac{|\phi_K(t)|^2}{|\phi_U(t/h)|^2}dt + 0.375\pi^{-1/2}(s_Z^2-\sigma_U^2)^{-5/2}\frac{h^4}{4}\int x^2K(x)dx.$$

The optimal bandwidth h and the variance σ_U^2 can be selected iteratively. If the normality assumption of X is violated, a bootstrap MISE proposed by Delaigle and Gijbels (2004a) can be used to select h and σ_U.

Despite the proposed method is developed based on the case–control framework, as demonstrated in Bondell (2005), yet it is also applicable under prospective sampling. However, some limitations in the proposed methods include the ordinary smoothness assumption of the measurement error. This excludes the commonly used normal error which belongs to the super smooth case. The main reason is the pessimistic slow convergence rate of the deconvolution density estimation for the super smooth error (Fan 1991). As explained in page 192 of Yi et al. (2021), despite the slow convergence rate, deconvolution in practice works reasonably well even if the error is super smooth. Hence it holds promises for the proposed deconvolution ISD estimation to work fairly well even if the error is normally distributed.

6 Proofs

We need the following four lemmas to prove Theorems 2.1 and 2.2.

Lemma 6.1 *Let $g(x)$ be a continuous function over $[a, b]$. Under assumptions (2.7)–(2.9), we have*

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

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

Proof For brevity, we only show the calculation of $E[h_1^{-1}K^*_{h_1}\{(x-Z)/h_1\}]$.

$$\begin{aligned}
&E\left[\frac{1}{h_1}K^*_{h_1}\frac{(x-Z)}{h_1}\right] \\
&= \frac{1}{h_1}\int\int \frac{1}{2\pi}\int_t \exp\left[i\frac{(y+u)-x}{h_1}t\right]\frac{\phi_K(t)}{\phi_U(t/h)}dtf_U(u)duf_1(y)dy \\
&= \frac{1}{h_1}\int \frac{1}{2\pi}\int_t \exp\left[i\frac{(y-x)}{h_1}t\right]\frac{\phi_K(t)}{\phi_U(t/h)}\phi_U(t/h)dtf_1(y)dy \\
&= \frac{1}{h_1}\int K\left(\frac{y-x}{h_1}\right)f_1(y)dy = \int K(u)f_1(x+h_1u)du = f_1(x) + O(h_1^m).
\end{aligned}$$

Denote $S_{ni} = \int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{i1}(x)dx + \int_a^b\{\hat{f}_0(x) - f_0(x)\}g_{i0}(x)dx$ for $i = 1, 2$. □

Lemma 6.2 *Under models (1.1) and (1.3), if assumptions (2.7)–(2.9), (2.11)–(2.14) hold, then*

$$\begin{aligned}
&n^{1/2}(\hat{\alpha} - \alpha - S_{n1}) = o_p(1), \quad n^{1/2}(\hat{\beta} - \beta - S_{n2}) = o_p(1), \text{ for } 0 < \rho < 1; \\
&n_1^{1/2}(\hat{\alpha} - \alpha - S_{n1}) = o_p(1), \quad n_1^{1/2}(\hat{\beta} - \beta - S_{n2}) = o_p(1), \text{ for } \rho = 0; \\
&n_0^{1/2}(\hat{\alpha} - \alpha - S_{n1}) = o_p(1); \quad n_0^{1/2}(\hat{\beta} - \beta - S_{n2}) = o_p(1), \text{ for } \rho = 1.
\end{aligned}$$

Proof Taylor's expansion implies that

$$\hat{\alpha} - \alpha - S_{n1} = \int_a^b R_{n_1}(x)f_1(x)g_{11}(x)dx - \int_a^b R_{n_0}(x)f_0(x)g_{10}(x)dx \tag{6.1}$$

$$+ b_{n_1}\int_a^b g_{11}(x)dx - b_{n_0}\int_a^b g_{10}(x)dx, \tag{6.2}$$

$$\hat{\beta} - \beta - S_{n2} = \int_a^b R_{n_1}(x)f_1(x)g_{21}(x)dx - \int_a^b R_{n_0}(x)f_0(x)g_{20}(x)dx \tag{6.3}$$

$$+ b_{n_1} \int_a^b g_{21}(x)dx - b_{n_0} \int_a^b g_{20}(x)dx. \tag{6.4}$$

in which

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

Since g_{1i} and g_{2i} for $i = 0, 1$ are bounded over $[a, b]$, the two terms in (6.2) and (6.4) are $o(n_i^{-1/2})$ by (2.14), respectively. Similar argument as in the the proof of Lemma 4.2 in Geng and Sakhanenko (2016), using Corollary 3.2 of Liu and Taylor (1989) and Theorem 2.3 of Giné and Guillou (2002), we derive the upper bound for $R_{n_i}(x)$ as

$$\begin{aligned}\sup_{x\in[a,b]} |R_{n_i}(x)| &\le O_p(\sup_{x\in[a,b]} |\hat{f}_i(x) - f_i(x)|^2 + |b_{n_i}|^2) \\ &\le O_p(\sup_{x\in[a,b]} |\hat{f}_i(x) - E[\hat{f}_i(x)]|^2 + \sup_{x\in[a,b]} |E[\hat{f}_i(x)] - f_i(x)|^2 + |b_{n_i}|^2) \\ &= O_p\left(\frac{\log(h_i^{-1})}{n_i h_i} + h_i^m + b_{n_i}^2\right).\end{aligned} \tag{6.5}$$

Then, (2.12)–(2.14) imply that

$$n_i^{1/2} \int_a^b R_{n_i}(x) f_i(x) g_{2i}(x)dx = O_p\left(\frac{\log(h_i^{-1})}{n_i^{1/2} h_i} + n_i h_i^{2m} + n_i^{1/2} b_{n_i}^2\right) = o_p(1), \ i = 0, 1.$$

Therefore the two terms in (6.1) and (6.3) are $o_p(n_i^{-1/2})$ respectively. This completes the proof. □

Lemma 6.3 *Let $\tilde{g}_1(x)$ and $\tilde{g}_2(x)$ be continuous functions over $[a, b]$. Under the assumptions (2.7)–(2.9), (2.11), and (2.12), we have*

$$\begin{aligned}&E\left[\int_a^b \left\{\frac{1}{h_i}K_{h_i}^*\left(\frac{x-Z}{h_i}\right) - f_i(x)\right\}\tilde{g}_1(x)dx \int_a^b \left\{\frac{1}{h_i}K_{h_i}^*\left(\frac{x-Z}{h_i}\right) - f_i(x)\right\}\tilde{g}_2(x)dx\right] \\ &\quad \to \int_a^b \tilde{g}_1(x)\tilde{g}_2(x)g_i(x)dx - \int_a^b \tilde{g}_1(x)f_i(x)dx \int_a^b \tilde{g}_2(x)f_i(x)dx \quad \text{as } h_i \to 0 \quad \text{for } i = 0, 1,\end{aligned} \tag{6.6}$$

in which $g_i(x)$ is defined as in (2.16).

Proof For $i = 1$, by Lemma 6.1, we have

$$
\begin{aligned}
&E\left[\int_a^b \left\{\frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)-f_1(x)\right\}\tilde{g}_1(x)dx\int_a^b \left\{\frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)-f_1(x)\right\}\tilde{g}_2(x)dx\right]\\
&\quad= E\left[\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\tilde{g}_1(x)dx\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\tilde{g}_2(x)dx\right.\\
&\quad\left.-\int_a^b f_1(x)\tilde{g}_1(x)dx\int_a^b f_1(x)\tilde{g}_2(x)dx\right]+O(h_1^m).
\end{aligned}
\tag{6.7}
$$

First, we rewrite the term in (6.7)

$$
\begin{aligned}
&E\left[\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\tilde{g}_1(x)dx\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\tilde{g}_2(x)dx\right]\\
&\quad= E\int_a^b\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\frac{1}{h_1}K_{h_1}^*\left(\frac{y-Z}{h_1}\right)\tilde{g}_1(x)\tilde{g}_2(y)dxdy\\
&\quad= \left(\int_{-\infty}^a+\int_a^b+\int_b^\infty\right)\int_a^b\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\frac{1}{h_1}K_{h_1}^*\left(\frac{y-Z}{h_1}\right)\tilde{g}_1(x)\tilde{g}_2(y)g_1(z)dxdydz\\
&\quad=: M_1+M_2+M_3.
\end{aligned}
$$

We will show that $M_1\to 0, M_2\to\int_a^b \tilde{g}_1(x)\tilde{g}_2(x)g_1(x)dx$, and $M_3\to 0$ as $h_1\to 0$. Rewrite

$$
\begin{aligned}
M_1 &= \int_{-\infty}^a\int_a^b\int_a^b \frac{1}{h_1}K_{h_1}^*\left(\frac{x-Z}{h_1}\right)\frac{1}{h_1}K_{h_1}^*\left(\frac{y-Z}{h_1}\right)\tilde{g}_1(x)\tilde{g}_2(y)g_1(z)dxdydz\\
&= \int_{-\infty}^a\int_{(a-z)/h_1}^{(b-z)/h_1}\int_{(a-z)/h_1}^{(b-z)/h_1} K_{h_1}^*(s)K_{h_1}^*(t)\tilde{g}_1(z+sh_1)\tilde{g}_2(z+th_1)g_1(z)dsdtdz\\
&= \int_{-\infty}^a\int_{(a-z)/h_1}^{(b-z)/h_1}\int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^{*+}(s)-K_{h_1}^{*-}(s)][K_{h_1}^{*+}(t)-K_{h_1}^{*-}(t)]\tilde{g}_1(z+sh_1)\tilde{g}_2(z+th_1)g_1(z)dsdtdz\\
&=: M_{11}+M_{12}+M_{13}+M_{14}.
\end{aligned}
$$

We consider M_{11}. Because $\tilde{g}_1,\tilde{g}_2$ are bounded on [a,b], K^* is bounded and integrable on $\mathbb{R}$, M_{11} is bounded by, up to a constant C,

$$
\int_{-\infty}^a\int_{(a-z)/h_1}^{(b-z)/h_1}\int_{(a-z)/h_1}^{(b-z)/h_1} K_{h_1}^{*+}(s)K_{h_1}^{*+}(t)g_1(z)dsdtdz. \tag{6.8}
$$

From Stefanski and Carroll (1990), we know that $K_{h_1}^*(x)$ is integrable and $\int K_{h_1}^*(x)dx=1$. Define $F^{*+}(x)=\int_{-\infty}^x K_{h_1}^{*+}(t)dt$. Then (6.8) can be rewritten as

$$
\int_{-\infty}^a\left[F^{*+}\left(\frac{b-z}{h_1}\right)-F^{*+}\left(\frac{a-z}{h_1}\right)\right]^2 g_1(z)du. \tag{6.9}
$$

For $\forall z<a$,

$$
F^{*+}\left(\frac{b-z}{h_1}\right)-F^{*+}\left(\frac{a-z}{h_1}\right)\to F^{*+}(\infty)-F^{*+}(\infty)=0,\ \text{as } h_1\to 0, \tag{6.10}
$$

By the dominated convergence theorem, (6.9) 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, $M_3 \to 0$. As $h_1 \to 0$,

$$\begin{aligned} M_2 &= \int_a^b \int_a^b \int_a^b \frac{1}{h_1} K_{h_1}^*\left(\frac{x-Z}{h_1}\right) \frac{1}{h_1} K_{h_1}^*\left(\frac{y-Z}{h_1}\right) \tilde{g}_1(x)\tilde{g}_2(y) g_1(z) dxdydz \\ &= \int_a^b \int_{(a-z)/h_1}^{(b-z)/h_1} \int_{(a-z)/h_1}^{(b-z)/h_1} K_{h_1}^{*+}(s) K_{h_1}^{*+}(t) g_1(z) dsdtdz \\ &\to \int_a^b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_{h_1}^*(s) K_{h_1}^*(t) \tilde{g}_1(z)\tilde{g}_2(z) g_1(z) dsdtdz = \int_a^b \tilde{g}_1(z)\tilde{g}_2(z) g_1(z) dz. \end{aligned}$$

Thus, (6.6) holds for $i = 1$. Similar argument shows that (6.6) holds for $i = 0$. □

Lemma 6.4 *Let $g(x)$ be a continuous function over $[a, b]$. Under (2.7)–(2.12), we have*

$$\sqrt{n_i} \int_a^b \{\hat{f}_i(x) - f_i(x)\} g(x) dx \to N(0, s_i^2), \tag{6.11}$$

where $s_i^2 = \int_a^b g^2(x) g_i(x) dx - \left\{\int_a^b f_i(x) g(x) dx\right\}^2, i = 0, 1.$

Proof We first show that

$$\frac{\int_a^b \hat{f}_1(x) g(x) dx - E\int_a^b \hat{f}_1(x) g(x) dx}{\sqrt{Var(\int_a^b \hat{f}_1(x) g(x) dx)}} \to N(0, 1). \tag{6.12}$$

Let

$$T_{1i}(x) = \frac{1}{h_1} K_{h_1}^*\left(\frac{x - Z_{1i}}{h_1}\right), \quad W_{1i} = \int_a^b T_{1i}(x) g(x) dx.$$

Note that $\int_a^b \hat{f}_1(x) g(x) dx$ is the sum of an i.i.d sequence. It suffices to show the Lyapunov's condition for the asymptotic normality in (6.12), i.e., for some $\delta_1 > 0$,

$$\frac{E|W_{11} - EW_{11}|^{2+\delta_1}}{n_1^{\delta_1/2}[Var(W_{11})]^{1+\delta_1/2}} \to 0, \tag{6.13}$$

as $n_1 \to \infty$. Since $g(x)$ is continuous and bounded over $[a, b]$, by similar argument to Lemma 6.1 and Fubini's theorem, we have

$$EW_{11} = \int_a^b f_1(x) g(x) dx + O(h_1^m).$$

Moreover, by Lemma 6.3, we have

$$Var(W_{11}) \le E(W_{11})^2 \to \int_a^b g^2(x)g_1(x)dx.$$

Then, to show the Lyapunov's condition in (6.13), it suffices to show for some $\delta_1 > 0$,

$$n_1^{-\delta_1/2} E|W_{n_{11}} - EW_{n_{11}}|^{2+\delta_1} \to 0. \tag{6.14}$$

Using similar argument to the proof of Lemma 4.4 of Geng and Sakhanenko (2016) and Koul and Ni (2004), by Hölder's inequality and $g(x)$ bounded on $[a, b]$, we get the upper bound

$$\begin{aligned} n_1^{-\delta_1/2} E|W_{11} - EW_{11}|^{2+\delta_1} &\le n_1^{-\delta_1/2} 2^{2+\delta_1} (E|W_{11}|^{2+\delta_1} + |EW_{11}|^{2+\delta_1}) \le n_1^{-\delta_1/2} 2^{2+\delta_1} E|W_{11}|^{2+\delta_1} + o(1) \\ &\le n_1^{-\delta_1/2} 2^{2+\delta_1} E\left(\int_a^b T_{11}(x) \sup_{x\in[a,b]} |g(x)|dx \right)^{2+\delta_1} + o(1) \le C n_1^{-\delta_1/2} 2^{2+\delta_1} E\left(\int_a^b T_{11}^{1+\delta_1/2}(x)dx \right)^2 + o(1) \end{aligned}$$

with constant $C = (\sup_{x\in[a,b]} |g(x)|)^{2+\delta_1}$ and

$$\begin{aligned} n_1^{-\delta_1/2} E\left(\int_a^b T_{11}^{1+\delta_1/2}(x)dx \right) &= n_1^{-\delta_1/2} E \int_a^b T_{11}^{1+\delta_1/2}(x)dx \int_a^b T_{11}^{1+\delta_1/2}(y)dy \\ &= \frac{1}{n_1^{\delta_1/2} h_1^{2+\delta_1}} \int_{-\infty}^{\infty} \int_a^b \int_a^b \left[K_{h_1}^*\left(\frac{x-z}{h_1} \right) \right]^{(1+\delta_1/2)} \left[K_{h_1}^*\left(\frac{y-z}{h_1} \right) \right]^{(1+\delta_1/2)} g_1(z)dxdydz \\ &= \frac{1}{n_1^{\delta_1/2} h_1^{2+\delta_1}} \left(\int_{-\infty}^{a} + \int_a^b + \int_b^{\infty} \right) \int_a^b \int_a^b \left[K_{h_1}^*\left(\frac{x-z}{h_1} \right) \right]^{(1+\delta_1/2)} \left[K_{h_1}^*\left(\frac{y-z}{h_1} \right) \right]^{(1+\delta_1/2)} g_1(z)dxdydz \\ &:= N_1 + N_2 + N_3. \end{aligned}$$

Then, showing (6.14) is equivalent to showing $N_1 \to 0$, $N_2 \to 0$, and $N_3 \to 0$. We consider N_2. Rewrite

$$\begin{aligned} N_2 &= \frac{1}{n_1^{\delta_1/2} h_1^{2+\delta_1}} \int_a^b \int_a^b \int_a^b \left[K_{h_1}^*\left(\frac{x-z}{h_1} \right) \right]^{(1+\delta_1/2)} \left[K_{h_1}^*\left(\frac{y-z}{h_1} \right) \right]^{(1+\delta_1/2)} g_1(z)dxdydz \\ &= \frac{1}{n_1^{\delta_1/2} h_1^{\delta_1}} \int_a^b \int_{(a-z)/h_1}^{(b-z)/h_1} \int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(s)]^{(1+\delta_1/2)} [K_{h_1}^*(t)]^{(1+\delta_1/2)} g_1(z)dsdtdz \\ &= \frac{1}{n_1^{\delta_1/2} h_1^{\delta_1}} \int_a^b \int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds \int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(t)]^{(1+\delta_1/2)} dt g_1(z)dz. \end{aligned}$$

Define

$$F(x) = \int_{-\infty}^{x} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds.$$

Then,

$$\int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds = F\left(\frac{b-z}{h_1} \right) - F\left(\frac{a-z}{h_1} \right).$$

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

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

Therefore,

$$N_2 = O\Big(\frac{1}{n_1^{\delta_1/2} h_1^{\delta_1}} \int_a^b \int_{-\infty}^{\infty} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds \int_{-\infty}^{\infty} [K_{h_1}^*(t)]^{(1+\delta_1/2)} dt g_1(z) dz\Big).$$

Here, we employ equation (3.2) of Theorem 2.1 of Fan (1991),

$$|h_1^{\tau} K_{h_1}^*(s)| \le \min\left\{C_1, \frac{C_2}{|s|}\right\} := M(s).$$

for some constants C_1 and C_2 independent of n_1 and s. Then, $|M(s)|^{1+\delta_1/2}$ is integrable for $\delta_1 > 0$. We have

$$\int_{-\infty}^{\infty} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds = \frac{1}{h_1^{\tau(1+\delta_1/2)}} \int_{-\infty}^{\infty} [h^{\tau} K_{h_1}^*(s)]^{1+\delta_1/2} ds$$
$$\le \frac{1}{h_1^{\tau(1+\delta_1/2)}} \int_{-\infty}^{\infty} |M(s)|^{1+\delta_1/2} ds = O(h_1^{-\tau(1+\delta_1/2)}).$$

Therefore,

$$N_2 = O\Big(\frac{1}{n_1^{\delta_1/2} h_1^{\delta_1+\tau(2+\delta_1)}}\Big) \to 0.$$

We then show that $N_1 \to 0$. Rewrite

$$N_1 = \frac{1}{n_1^{\delta_1/2} h_1^{2+\delta_1}} \int_{-\infty}^{a} \int_a^b \int_a^b \left[K_{h_1}^*\left(\frac{x-z}{h_1}\right)\right]^{(1+\delta_1/2)} \left[K_{h_1}^*\left(\frac{y-z}{h_1}\right)\right]^{(1+\delta_1/2)} g_1(z) dx dy dz$$
$$= \frac{1}{n_1^{\delta_1/2} h_1^{\delta_1}} \int_{-\infty}^{a} \int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds \int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(t)]^{(1+\delta_1/2)} dt g_1(z) dz.$$

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

$$\int_{(a-z)/h_1}^{(b-z)/h_1} [K_{h_1}^*(s)]^{(1+\delta_1/2)} ds \to F(\infty) - F(\infty) = 0.$$

Hence, $N_1 \to 0$. Using similar argument, $N_3 \to 0$. By Lemma 6.1,

$$E\int_a^b \hat{f}_1(x)g(x)dx = \frac{1}{n_1}\sum_{i=1}^{n_1} EW_{11} = \int_a^b f_1(x)g(x)dx + O(h_1^m).$$

By Lemma 6.3, we have

$$\begin{aligned} Var\left(\int_a^b \hat{f}_1(x)g(x)dx\right) &= \frac{1}{n_1}\left[E\left(\int_a^b \frac{1}{h_1}K_{h_1}^*\frac{(x-Z)}{h_1}g(x)dx\right)^2 - \left(E\int_a^b \frac{1}{h_1}K_{h_1}^*\frac{(x-Z)}{h_1}g(x)dx\right)^2\right] \\ &= \frac{1}{n_1}\left[\int_a^b g^2(x)g_1(x)dx - \left(\int_a^b f_1(x)g(x)dx + O(h_1^m)\right)^2\right]. \end{aligned}$$

Replacing $E\int_a^b \hat{f}_1(x)g(x)dx$ and $Var(\hat{f}_1(x)g(x)dx)$ in (6.12), we get

$$\frac{\sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g(x)dx}{s_1} \to N(0,1).$$

Similar argument implies that (6.11) holds for $i = 0$.

Proof of Theorem 2.1 By Lemmas 6.1, 6.2, and 6.3, we have $E(\hat{\beta} - \beta)^2$ and $E(\hat{\alpha} - \alpha)^2$ converge to 0. Therefore, $(\hat{\beta}, \hat{\alpha})$ are consistent estimators of (β, α). Furthermore, by Lemma 6.4, we have for all $a_{11}, a_{21} \in \mathbb{R}$,

$$a_{11}\sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{11}(x)dx + a_{21}\sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{21}(x)dx$$

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

$$\sqrt{n_1}\begin{pmatrix}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{11}(x)dx \\ \int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{21}(x)dx\end{pmatrix}$$

is a bivariate normal random variable. By Lemmas 6.1 and 6.3, we obtain

$$\begin{aligned} Var\left(\sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{11}(x)dx\right) &\to \int_a^b g_{11}^2(x)g_1(x)dx - \left(\int_a^b f_1(x)g_{11}(x)dx\right)^2 \\ Var\left(\sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{21}(x)dx\right) &\to \int_a^b g_{21}^2(x)g_1(x)dx - \left(\int_a^b f_1(x)g_{21}(x)dx\right)^2 \end{aligned}$$

and

$$\begin{aligned} &Cov\left(\sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{11}(x)dx, \sqrt{n_1}\int_a^b\{\hat{f}_1(x) - f_1(x)\}g_{21}(x)dx\right) \\ &\to \int_a^b g_{11}(x)g_{21}(x)g_1(x)dx - \int_a^b f_1(x)g_{11}(x)dx\int_a^b f_1(x)g_{21}(x)dx. \end{aligned}$$

Therefore,

$$\sqrt{n_1}\begin{pmatrix}\int_a^b\{\hat{f}_1(x)-f_1(x)\}g_{11}(x)dx\\ \int_a^b\{\hat{f}_1(x)-f_1(x)\}g_{21}(x)dx\end{pmatrix}\to N(0,\Sigma_1).$$

Similarly,

$$\sqrt{n_0}\begin{pmatrix}\int_a^b\{\hat{f}_0(x)-f_0(x)\}g_{10}(x)dx\\ \int_a^b\{\hat{f}_0(x)-f_1(x)\}g_{20}(x)dx\end{pmatrix}\to N(0,\Sigma_0).$$

The asymptotic normality above, the independence between case and control samples and $0<\rho<1$ in (2.15) complete the proof. □

Proof of Theorem 2.2 When $\rho=0$, we have $n_1/n_0\to 0$. By Lemmas 6.1–6.4, we obtain

$$\begin{aligned}\sqrt{n_1}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix}&=\sqrt{n_1}\begin{pmatrix}\int_a^b\{\hat{f}_1(x)-f_1(x)\}g_{11}(x)dx\\ \int_a^b\{\hat{f}_1(x)-f_1(x)\}g_{21}(x)dx\end{pmatrix}+\frac{\sqrt{n_1}}{\sqrt{n_0}}\sqrt{n_0}\begin{pmatrix}\int_a^b\{\hat{f}_0(x)-f_0(x)\}g_{10}(x)dx\\ \int_a^b\{\hat{f}_0(x)-f_0(x)\}g_{20}(x)dx\end{pmatrix}\\&\quad+o_p(1)+o_p(\sqrt{n_1}/\sqrt{n_0})\\&=\sqrt{n_1}\begin{pmatrix}\int_a^b\{\hat{f}_1(x)-f_1(x)\}g_{11}(x)dx\\ \int_a^b\{\hat{f}_1(x)-f_1(x)\}g_{21}(x)dx\end{pmatrix}+o_p(1)\to N(0,\Sigma_1).\end{aligned}$$

By similar argument when $\rho=1$,

$$\sqrt{n_0}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix}\to\sqrt{n_0}\begin{pmatrix}\int_a^b\{\hat{f}_0(x)-f_0(x)\}g_{10}(x)dx\\ \int_a^b\{\hat{f}_0(x)-f_0(x)\}g_{20}(x)dx\end{pmatrix}.$$

□

Acknowledgements The authors would like to thank the editor and two referees for their help in improving the manuscript. This research was supported by the National Science Foundation under Grant No. 2349860.

References

Bondell H (2005) Minimum distance estimation for the logistic regression model. Biometrika 92:724–731
Bondell H (2007) Testing goodness-of-fit in logistic case-control studies. Biometrika 94:487–495
Carroll RJ, Hall P (1988) Optimal rates of convergence for deconvolving a density. J Am Stat Assoc 83(404):1184–1186
Carroll RJ, Wand MP (1991) Semiparametric estimation in logistic measurement error models. J Roy Stat Soc B (Methodol) 53(3):573–585
Carroll RJ, Ruppert D, Stefanski L A, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC
Clayton D, Hills M (2013) Statistical models in epidemiology. OUP Oxford
Delaigle A, Gijbels I (2004) Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. Ann Inst Stat Math 56(1):19–47
Delaigle A, Gijbels I (2004) Practical bandwidth selection in deconvolution kernel density estimation. Comput Stat Data Anal 45(2):249–267

Dobson AJ, Barnett AG (2018) Binary Variables and Logistic Regression. In: Press CRC (ed) An introduction to generalized linear models, 3rd edn. Taylor and Francis Group, pp 123–143

Fan J (1991) Asymptotic normality for deconvolution kernel density estimators. Sankhyā Indian J Stat A 97–110

Geng P, Sakhanenko L (2016) Parameter estimation for the logistic regression model under case-control study. Stat Probabil Lett 109:168–177

Giné E, Guillou A (2002) Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincare (B) Probabil Stat. 38(6): 907–921

Koul HL, Ni P (2004) Minimum distance regression model checking. J Stat Plann Inference 119(1):109–141

Lee LF, Sepanski JH (1995) Estimation of linear and nonlinear errors-in-variables models using validation data. J Am Stat Assoc 90(429):130–140

Liu MC, Taylor RL (1989) A consistent nonparametric density estimator for the deconvolution problem. Canadian J Stat 17(4):427–438

Prentice RL, Pyke R (1979) Logistic disease incidence models and case-control studies. Biometrika 66:403–411

Qin J, Zhang B (1997) A goodness-of-fit test for logistic regression models based on case-control data. Biometrika 84(3):609–618

Siddique J, Daniels MJ, Carroll RJ, Raghunathan TE, Stuart EA, Freedman LS (2019) Measurement error correction and sensitivity analysis in longitudinal dietaryintervention studies using an external validation study. Biometrics 75(3):927–937

Stefanski LA, Carroll RJ (1985) Covariate measurement error in logistic regression. Annals Stat, 1335–1351

Stefanski LA, Carroll RJ (1990) Deconvolving kernel density estimators. Statistics 21(2):169–184

Stommel M, Schoenborn CA (2009) Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001–2006. BMC Public Health 9:1–10

Thürigen D, Spiegelman D, Blettner M, Heuer C, Brenner H (2000) Measurement error correction using validation data: a review of methods and their applicability in case-control studies. Stat Methods Med Res 9(5):447–474

Tooze JA, Troiano RP, Carroll RJ, Moshfegh AJ, Freedman LS (2013) A measurement error model for physical activity level as measured by a questionnaire with application to the 1999–2006 NHANES questionnaire. Am J Epidemiol 177(11):1199–1208

Va P, Dodd KW, Zhao L, Thompson-Paul AM, Mercado CI, Terry AL, Cogswell ME (2019) Evaluation of measurement error in 24-hour dietary recall for assessing sodium and potassium intake among US adults-National Health and Nutrition Examination Survey (NHANES). Am J Clin Nutrit 109(6):1672–1682

Xie Y, Wang X, Story M (2009) Statistical methods of background correction for Illumina BeadArray data. Bioinformatics 25(6):751–757

Yi GY, Delaigle A, Gustafson P (2021) Handbook of measurement error models. CRC Press, Boca Raton