Electronic Journal of Statistics
Vol. 12 (2018) 1782–1821
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1446

Dimension reduction and estimation in the secondary analysis of case-control studies

Liang Liang*

Department of Biostatistics, Harvard University, Boston, MA 02115, USA
e-mail: liliang@hsph.harvard.edu

Raymond Carroll*

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843, USA, and School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123 Broadway NSW 2007, Australia
e-mail: carroll@stat.tamu.edu

Yanyuan Ma†

Department of Statistics, Penn State University, University Park, PA 16802, USA
e-mail: yanyuanma@gmail.com

Abstract: Studying the relationship between covariates based on retrospective data is the main purpose of secondary analysis, an area of increasing interest. We examine the secondary analysis problem when multiple covariates are available, while only a regression mean model is specified. Despite the completely parametric modeling of the regression mean function, the case-control nature of the data requires special treatment and semiparametric efficient estimation generates various nonparametric estimation problems with multivariate covariates. We devise a dimension reduction approach that fits with the specified primary and secondary models in the original problem setting, and use reweighting to adjust for the case-control nature of the data, even when the disease rate in the source population is unknown. The resulting estimator is both locally efficient and robust against the misspecification of the regression error distribution, which can be heteroscedastic as well as non-Gaussian. We demonstrate the advantage of our method over several existing methods, both analytically and numerically.

MSC 2010 subject classifications: 62G05.

Keywords and phrases: Biased samples, case-control study, dimension reduction, heteroscedastic error, secondary analysis, semiparametric estimation.

Received February 2017.

*Research was supported by grants from the National Cancer Institute (U01-CA057030).
†Research was supported by the National Science Foundation (DMS-1206693) and the National Institute of Neurological Disorders and Stroke (R01-NS073671).

1. Introduction

Case-control studies are popular tools in investigating risk factors associated with various uncommon diseases, such as cancer and myocardial infarction, often because these studies are relatively less expensive and more convenient to implement compared with designs such as cross-sectional and prospective cohort studies (Chatterjee et al., 2009). Typically, a population-based case-control study employs a random sample of cases (diseased subjects) and a separate random sample of controls (non-diseased subjects). It also collects covariate information on the exposure of interest and other risk factors. The primary task of case-control studies lies in understanding the relationship between disease rates and covariates, usually via a prospective logistic regression analysis, which gives an efficient estimator of all parameters except the intercept, under the conditions that the disease rate is unknown and no parametric model for the predictors is available in the underlying source population (Prentice and Pyke, 1979). It is however now well-recognized in gene-environment interaction studies that estimation of interactions can be made much more efficient if the distribution of the gene given the environment is modeled parametrically (Piegorsch et al., 1994; Chatterjee and Carroll, 2005; Chen et al., 2008, 2009).

Recently, there has been considerable interest in using case-control data for a separate task, namely examining the interrelationship between covariates, say Y and $\mathbf{X}$, where Y is a scalar and $\mathbf{X}$ is potentially multivariate (Jiang et al., 2006; Lin and Zeng, 2009; Li et al., 2010; Wei et al., 2013; Tchetgen, 2014). For example, in Section 6, we describe a case-control study involving breast cancer. Mammographic density, age at first live birth, age at menarche and body mass index are all known to be predictors of breast cancer, but it is also of interest to examine the effects of age at first live birth, age at menarche and body mass index, $\mathbf{X}$, on mammographic density, Y in this case-control study.

The main difficulty of such *secondary analysis* is that the case-control data is not a random sample from the underlying source population, which we refer to as *true population* throughout the paper. In fact the case-control samples are taken separately from the case subpopulation and the control subpopulation. As a consequence, the relationship between covariates Y and $\mathbf{X}$ in the secondary analysis under the case-control context can be very different from the relationship in the true population. Hence, simply regressing Y on $\mathbf{X}$ and ignoring the case-control sampling scheme can be grossly misleading.

A simple approach to secondary analysis is using only controls if the disease rate is rare, say less than 1%. This type of approach is widely used, because if the disease rate is $< 1\%$, the controls make up more than 99% of the population, and analysis of them is close to that of the entire population. However, this approach can have relatively low efficiency because it ignores the information carried by the cases. A more efficient approach is to adopt a semiparametric framework, assuming a parametric distribution for Y given $\mathbf{X}$, e.g., linear regression with normally distributed and homoscedastic regression errors, as well as known or rare disease rate (Jiang et al., 2006; Lin and Zeng, 2009; Li et al., 2010; Wei et al., 2013; Tchetgen, 2014). This approach improves estimation efficiency compared

with the controls only method because both cases and controls are taken into account.

However, the disease rate in the source population being sampled is often unknown and some diseases may not be so rare as less than 1%, so that the controls-only analysis can have considerable bias. This prompted Ma and Carroll (2016) to propose a further improved approach, which does not require a known or a rare disease assumption, and also, unlike the papers referenced above, does not assume normality or homoscedasticity of the regression error. In fact, they only specify a mean model to describe the relationship between covariates. Their semiparametric estimator involves positing density functions for $\mathbf{X}$ and Y given $\mathbf{X}$ that may or may not be true. The resulting estimator is (a) consistent and asymptotically normally distributed even if the posited functions are incorrectly specified; and (b) it is efficient if the posited functions are correctly specified. An estimator with the properties (a) and (b) will be called locally efficient throughout this article.

Because the approach of Ma and Carroll (2016) was developed by adopting a superpopulation concept and viewing case-control samples as independent and identically distributed observations sampled from the superpopulation, they need to link the quantities in the superpopulation to the ones in the true population. As a consequence, several additional conditional distributions arise in the likelihood formulation, including quantities conditional on the covariates. This leads to the need to perform several nonparametric regressions on the covariates in their estimator. When the covariate dimension increases, such nonparametric regressions inevitably suffer from the curse of dimensionality.

In this paper, we work in the superpopulation framework and handle the potential dimensionality problem using a dimension reduction modeling approach. We assume several quantities of interest depend on the covariates $\mathbf{X}$ only through linear combinations of $\mathbf{X}$ and/or known functions of $\mathbf{X}$. This allows us to avoid multivariate nonparametric regression. However, because of the inherent relation between the covariates assumed in the original true population, the dimension reduction structure is not completely arbitrary. Instead, it is subject to various constraints, which makes the problem different from the classical dimension reduction modeling and estimation. Taking these various special features into consideration, we construct asymptotically consistent estimators for the regression parameters in the true population model. These estimators have a parametric convergence rate and are robust to the misspecification of the conditional distribution of Y given $\mathbf{X}$.

We emphasize that ours is not a paper about advancing dimension reduction modeling, which already has a massive literature (Ma and Zhu, 2013b; Li, 1991; Li and Duan, 1989; Li, 1992; Li and Dong, 2009; Li and Wang, 2007; Li et al., 2008, 2005; Dong and Li, 2010; Ma and Zhu, 2012b, 2013a; Zhu et al., 2010; Cook, 2009; Cook and Li, 2002; Yin and Cook, 2002; Cook, 1994; Setodji and Cook, 2004; Cook and Setodji, 2003; Yin and Bura, 2006; Xia, 2007). Instead it is about *using* dimension reduction ideas for solving a semiparametric problem in the secondary analysis of case-control studies when the dimensionality of the covariates is potentially large.

2. Methodology

2.1. Background

Let D be disease status, where $D = 1$ denotes a case and $D = 0$ denotes a control. Also let $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}}$ be a $(p+1) \times 1$ vector of covariates, where $\mathbf{X}$ is a p-dimensional vector and Y is a scalar. We assume that both $\mathbf{X}$ and Y are continuous and they are related to disease status D via a logistic regression model

$$\begin{aligned} \mathrm{pr}(D = d|\mathbf{X} = \mathbf{x}, Y = y) &= f^{\mathrm{true}}_{D|X,Y}(d, \mathbf{x}, y) = H(d, \mathbf{x}, y, \boldsymbol{\alpha}) \\ &= \frac{\exp\{d(\alpha_c + \mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}_1 + y\alpha_2)\}}{1 + \exp(\alpha_c + \mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}_1 + y\alpha_2)}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\alpha} = (\alpha_c, \boldsymbol{\alpha}_1^{\mathrm{T}}, \alpha_2)^{\mathrm{T}}$.

As mentioned before, the goal of secondary analysis is to investigate the relationship between $\mathbf{X}$ and Y in the source population, which we assume is of the form

$$Y = m(\mathbf{X}, \boldsymbol{\beta}) + \epsilon, \tag{2.2}$$

where $m(\cdot)$ is a smooth function known up to a parameter $\boldsymbol{\beta}$. The error term ϵ satisfies $E_{\mathrm{true}}(\epsilon|\mathbf{X}) = 0$, but no other assumptions about ϵ are made, especially normality or homoscedasticity or independence from $\mathbf{X}$. Under mild conditions, the parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ defined in (2.1) and (2.2) are identifiable (Ma and Carroll, 2016).

2.2. Superpopulation Model Framework and Efficient Estimator

From model (2.2), the conditional distribution of Y given $\mathbf{X}$ and the marginal distribution of $\mathbf{X}$ with respect to the true population are

$$f^{\mathrm{true}}_{Y|\mathbf{X}}(y, \mathbf{x}, \boldsymbol{\beta}) = \eta_2\{y - m(\mathbf{x}, \boldsymbol{\beta}), \mathbf{x}\} = f^{\mathrm{true}}_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x}), \tag{2.3}$$

$$f^{\mathrm{true}}_{\mathbf{X}}(\mathbf{x}) = \eta_1(\mathbf{x}). \tag{2.4}$$

Here η_2 is an unknown probability density function with mean 0, which is free of the unknown parameters $\boldsymbol{\beta}$, ϵ is the error term defined in (2.2), i.e., $\epsilon = Y - m(\mathbf{X}, \boldsymbol{\beta})$ and η_1 is another probability density function which is also unknown. The superscript "true" emphasizes that the probability densities in (2.3) - (2.4) are defined under the true population.

Suppose we draw a case-control sample with N_1 cases and N_0 controls. Because of the sampling design, classical large-sample asymptotic theory does not work here. The idea of a superpopulation is to construct a hypothetical population with infinite sample size and a fixed ratio of cases to controls, N_1/N_0,

then treat the case-control sample as a random sample from the superpopulation with sample size $N = N_0 + N_1$ (Ma, 2010). The explicit form of the joint density of $(\mathbf{X}, Y, D)$ in such a superpopulation is

$$\begin{aligned} f_{\mathbf{X},Y,D}(\mathbf{x}, y, d) &= (N_d/N) f^{\text{true}}_{X,Y|D}(\mathbf{x}, y, d) \\ &= \frac{N_d}{N} \frac{\eta_1(\mathbf{x})\eta_2(\epsilon, \mathbf{x}) H(d, \mathbf{x}, y, \boldsymbol{\alpha})}{\int \eta_1(\mathbf{x})\eta_2(\epsilon, \mathbf{x}) H(d, \mathbf{x}, y, \boldsymbol{\alpha}) d\mu(\mathbf{x}) d\mu(y)}. \end{aligned}$$

Here we use the fact that the distribution of $(\mathbf{X}, Y)$ conditional on the disease status D in the superpopulation and in the true population are identical, which links the distributions in these two populations.

Ma and Carroll (2016) derived the semiparametric efficient score function corresponding to the above superpopulation, $\mathbf{S}_{\text{eff}}(\mathbf{X}_i, Y_i, D_i) = \{\mathbf{S}(\mathbf{X}_i, Y_i, D_i) - \mathbf{g}\{Y_i - m(\mathbf{X}_i, \boldsymbol{\beta}), \mathbf{X}_i)\} - (1 - D_i)\mathbf{v}_0 - D_i\mathbf{v}_1$. The resulting efficient estimating equation is

$$\sum_{i=1}^{N} \{\mathbf{S}(\mathbf{X}_i, Y_i, D_i) - \mathbf{g}\{Y_i - m(\mathbf{X}_i, \boldsymbol{\beta}), \mathbf{X}_i)\} - (1 - D_i)\mathbf{v}_0 - D_i\mathbf{v}_1 = \mathbf{0}, \quad (2.5)$$

where

$$\mathbf{S}(\mathbf{x}, y, d, \boldsymbol{\theta}) = \begin{Bmatrix} \partial \log\{H(d, \mathbf{x}, y, \boldsymbol{\alpha})\}/\partial \boldsymbol{\alpha} \\ \partial \log\{\eta_2(\epsilon, \mathbf{x})\}/\partial \boldsymbol{\beta} \end{Bmatrix}. \quad (2.6)$$

Although as a function, η_2 does not depend on $\boldsymbol{\beta}$, its first argument ϵ contains $\boldsymbol{\beta}$. Other quantities used in (2.5) are defined in (2.7).

$$\begin{aligned} &\pi_0 \equiv p_D^{\text{true}}(0) = \textstyle\int \eta_1(\mathbf{x})\eta_2(\epsilon, \mathbf{x}) H(0, \mathbf{x}, y) d\mu(\mathbf{x}) d\mu(y); \\ &\pi_1 \equiv p_D^{\text{true}}(1) = \textstyle\int \eta_1(\mathbf{x})\eta_2(\epsilon, \mathbf{x}) H(1, \mathbf{x}, y) d\mu(\mathbf{x}) d\mu(y); \\ &b_0 \equiv E\{f_{D|\mathbf{X},Y}(1, \mathbf{X}, y) \mid D = 0\}; b_1 \equiv E\{f_{D|\mathbf{X},Y}(0, \mathbf{X}, y) \mid D = 1\}; \\ &\boldsymbol{\mu}_s(\mathbf{x}, y) \equiv E(\mathbf{S} \mid \epsilon, \mathbf{X} = \mathbf{x});\ \mathbf{c}_0 \equiv E(\mathbf{S} \mid D = 0) - E\{\boldsymbol{\mu}_s(\mathbf{X}, Y) \mid D = 0\}; \\ &\mathbf{c}_1 \equiv E(\mathbf{S} \mid D = 1) - E\{\boldsymbol{\mu}_s(\mathbf{X}, Y) \mid D = 1\}; \\ &\kappa(\mathbf{x}, y) \equiv \left[\sum_{d=0}^{1} \{N_d H(d, \mathbf{x}, y)\}/(N\pi_d)\right]^{-1}; \\ &t_1(\mathbf{X}) \equiv \left[E_{\text{true}}\left\{\epsilon^2 \kappa(\mathbf{X}, Y) \mid \mathbf{X}\right\}\right]^{-1}; \\ &\mathbf{t}_2(\mathbf{X}) \equiv E_{\text{true}}\left\{\epsilon \boldsymbol{\mu}_s(\mathbf{X}, Y) \mid \mathbf{X}\right\} - (\mathbf{c}_0/b_0) E_{\text{true}}\left\{\epsilon f_{D|\mathbf{X},Y}(0, \mathbf{X}, Y) \mid \mathbf{X}\right\}; \\ &t_3(\mathbf{X}) \equiv -b_0^{-1} E_{\text{true}}\left\{\epsilon f_{D|\mathbf{X},Y}(0, \mathbf{X}, Y) \mid \mathbf{X}\right\}; \\ &\mathbf{a}(\mathbf{x}) \equiv t_1(\mathbf{x})\{\mathbf{t}_2(\mathbf{x}) + t_3(\mathbf{x})\mathbf{u}_0\}; \\ &\mathbf{u}_0 \equiv (1 - E[\epsilon t_1(\mathbf{X}) t_3(\mathbf{X}) \kappa(\mathbf{X}, Y) \mid D = 0])^{-1} E[\epsilon t_1(\mathbf{X}) \mathbf{t}_2(\mathbf{X}) \kappa(\mathbf{X}, Y) \mid \\ &D = 0];\ \mathbf{u}_1 \equiv -(N_0/N_1)\mathbf{u}_0; \\ &\mathbf{v}_0 \equiv (\pi_1/b_0)(\mathbf{u}_0 + \mathbf{c}_0);\ \mathbf{v}_1 \equiv -(\pi_0/b_0)(\mathbf{u}_0 + \mathbf{c}_0); \\ &\mathbf{g}(\epsilon, \mathbf{x}) \equiv \boldsymbol{\mu}_s(\mathbf{x}, y) - \epsilon \mathbf{a}(\mathbf{x}) \kappa(\mathbf{x}, y) - \mathbf{v}_0 f_{D|\mathbf{X},Y}(0, \mathbf{x}, y) - \mathbf{v}_1 f_{D|\mathbf{X},Y}(1, \mathbf{x}, y). \end{aligned} \quad (2.7)$$

3. Approach via Dimension Reduction

3.1. Background

The estimating equation (2.5) contains three expectations conditional on covariates $\mathbf{X}$, i.e., $E_{\text{true}}\left\{\epsilon^2\kappa(\mathbf{X},Y) \mid \mathbf{X}\right\}, E_{\text{true}}\left\{\epsilon\boldsymbol{\mu}_s(\mathbf{X},Y) \mid \mathbf{X}\right\}$ and $E_{\text{true}}\{\epsilon f_{D|\mathbf{X},Y}(0, \mathbf{X},Y) \mid \mathbf{X}\}$, which need to be estimated nonparametrically. However, such estimation may be extremely hard when the covariates $\mathbf{X}$ are multivariate. To bypass the potential curse of dimensionality problem caused by the multivariate nature of $\mathbf{X}$, we use a dimension reduction modeling strategy, i.e., we assume all three quantities in the conditional expectations depend on $\mathbf{X}$ only through several linear combinations $\mathbf{X}^{\mathrm{T}}\boldsymbol{\gamma}$ or several linear combinations of functions of $\mathbf{X}$. Under such a dimension reduction structure, we can construct nonparametric regression estimators for high dimensional covariates $\mathbf{X}$ in a way similar to the univariate case with desired bias and MSE order, hence facilitating the estimation procedure via solving the estimating equation (2.5).

Let $f_0(\mathbf{X},Y,\boldsymbol{\alpha}) = f_{D|\mathbf{X},Y}(0,\mathbf{X},Y)$. All three functions $\kappa(\mathbf{x},y), \boldsymbol{\mu}_s(\mathbf{x},y)$ and $f_0(\mathbf{x},y)$ depend on $\pi_d = \pi_d(\boldsymbol{\alpha})$. To emphasize this, we replace π_d with $\pi_d(\widetilde{\boldsymbol{\alpha}})$ in those three functions and we use the notation $\kappa(\mathbf{x},y,\widetilde{\boldsymbol{\alpha}}), \boldsymbol{\mu}_s(\mathbf{x},y,\widetilde{\boldsymbol{\alpha}}), f_0(\mathbf{x},y,\widetilde{\boldsymbol{\alpha}})$ to distinguish them from the ones using the true parameter value $\boldsymbol{\alpha}$. In addition, we define $\epsilon(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}}) = Y - m(\mathbf{X},\widetilde{\boldsymbol{\beta}})$ to distinguish it from the true $\epsilon = Y - m(\mathbf{X},\boldsymbol{\beta})$.

There are two cases that need to be considered, namely that (i) $m(\cdot)$ defined in (2.2) is a linear function of $\mathbf{X}$; and (ii) that $m(\cdot)$ is not a linear function of $\mathbf{X}$. In case (i), we set $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} = \mathbf{X}$, while in case (ii), we set $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} = \{\mathbf{X}^{\mathrm{T}}, m(\mathbf{X},\widetilde{\boldsymbol{\beta}})\}^{\mathrm{T}}$.

Regardless of whether $m(\mathbf{X},\boldsymbol{\beta})$ is linear or nonlinear, our dimension reduction models are

$$E_{\text{true}}\{\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}}) \mid \mathbf{X}\} = \zeta_1(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\alpha}}), \tag{3.1}$$

$$E_{\text{true}}\{\epsilon(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\boldsymbol{\mu}_s(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}}) \mid \mathbf{X}\} = \boldsymbol{\zeta}_2(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2,\mathbf{X},\widetilde{\boldsymbol{\alpha}}), \tag{3.2}$$

$$E_{\text{true}}\{\epsilon(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})f_0(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}}) \mid \mathbf{X}\} = \zeta_3(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3,\widetilde{\boldsymbol{\alpha}}), \tag{3.3}$$

for $\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\beta}}$ that are in a neighborhood of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Here $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}$ is a finite dimensional vector, each element of which is a function of $\mathbf{X}$. The subscript $\widetilde{\boldsymbol{\beta}}$ indicates $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}$ may depend on the unknown parameter $\widetilde{\boldsymbol{\beta}}$. The three indices $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3$ are vectors or matrices that have the same row size as the length of $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}$ and with ℓ columns. The lower square blocks of all three matrices $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3$ are set to be identity to ensure identifiability. Throughout the text, we use the notation $\boldsymbol{\gamma}_{-1}$ to denote the submatrix of $\boldsymbol{\gamma}$ without the lower square block for any matrix $\boldsymbol{\gamma}$. $\zeta_1(\cdot),\boldsymbol{\zeta}_2(\cdot),\zeta_3(\cdot)$ are three unknown functions. Strictly speaking, model (3.2) is not a standard dimension reduction model. However, in Appendix A.1, we describe its actual form, which in general consists of three different standard dimension reduction models.

3.2. Data Generating Mechanisms for Which (3.1)-(3.3) Are Valid

The dimension reduction models (3.1)-(3.3) are used here only as working models to facilitate the estimation procedure. We do not intend to include these models as part of our original model assumptions and thereby take these structures into account to further improve estimation efficiency.

There are at least two simple and important data generating mechanisms for which (3.1)-(3.3) hold: (a) when ϵ is independent of $\mathbf{X}$; and (b) when, as in equation (1) of Lian et al. (2015), $\epsilon = v(\mathbf{X}^{\mathrm{T}}\omega)\epsilon^*$, where $\mathbf{v}(\cdot)$ is an unknown smooth function and ϵ^* is independent of $\mathbf{X}$ with mean 0 and variance 1. More generally, we have the following result, proved in Appendix Section A.4, and including the two special cases given above.

Proposition 1. Suppose $\epsilon = Q(\mathbf{X}^{\mathrm{T}}\omega, \epsilon^*)$, where $Q(\cdot)$ is an arbitrary smooth function and ϵ^* is independent of $\mathbf{X}$. Then the dimension reduction models (3.1)-(3.3) hold for any $m(\mathbf{X}, \boldsymbol{\beta})$ model.

3.3. Estimation

As stated in Section 3.2, models (3.1)-(3.3) can often be used as working models to facilitate the multivariate nonparametric regression. Therefore, in the rest of the derivation, we use the general model (3.1)-(3.3) without specifying the particular form of $\mathbf{Z}_{\tilde{\boldsymbol{\beta}}}$. Of course, we need to estimate $\boldsymbol{\gamma}_j$ and $\zeta_j(\cdot)$ for $j = 1, 2, 3$. To resolve the issue of estimating conditional expectations in the true population while we only have a random sample from the superpopulation, the key point is to recognize the connection between the two populations and to adjust the case-control data in the context of conditional expectations via

$$E_{\mathrm{true}}\{h(D, \mathbf{X}, Y)\} = \sum_{d=0}^{1} \pi_d E\{h(D, \mathbf{X}, Y) \mid D = d\},$$

where $h(\cdot)$ is any function such that $h(D, \mathbf{X}, Y)$ has finite mean. Hence we can simply weight cases by π_1/N_1 and controls by π_0/N_0 and this will give us the $\zeta_j(\cdot)$'s. Take $\zeta_1(\cdot)$ as an example. A valid estimating equation for $\zeta_1(\cdot)$ is

$$\begin{aligned} 0 \quad = \quad & \sum_{d=0}^{1} (\pi_d/N_d) \sum_{i=1}^{N} I(D_i = d)\{\epsilon_i^2 \kappa(\mathbf{X}_i, Y_i) - \\ & \zeta_1(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1)\} K_h(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_1), \end{aligned} \tag{3.4}$$

since

$$\begin{aligned} & E[\sum_{d=0}^{1} (\pi_d/N_d) \sum_{i=1}^{N} I(D_i = d)\{\epsilon_i^2 \kappa(\mathbf{X}_i, Y_i) - \zeta_1(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1)\} K_h(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_1)] \\ & \quad = \sum_{d=0}^{1} \pi_d E_{\mathrm{true}}[\{\epsilon^2 \kappa(\mathbf{X}, Y) - \zeta_1(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1)\} K_h(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_1) | D = d] \\ & \quad = E_{\mathrm{true}}[\{\epsilon^2 \kappa(\mathbf{X}, Y) - \zeta_1(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1)\} K_h(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_1)] = 0. \end{aligned}$$

Here $K_h(\mathbf{u}) = \prod_{i=1}^{\ell} K(u_i/h)/h^{\ell}$ for any ℓ-dimensional vector $\mathbf{u} = (u_1, \cdots, u_\ell)^{\mathrm{T}}$.

Of course π_d is not known. Thus, to implement the idea stated in (3.4), we need an estimator of $\pi_d = \pi_d(\boldsymbol{\alpha})$. As an equation for π,

$$E\left[\frac{H(0,\mathbf{X},Y,\boldsymbol{\alpha})}{(N_0/N)H(0,\mathbf{X},Y,\boldsymbol{\alpha}) + (N_1/N)H(1,\mathbf{X},Y,\boldsymbol{\alpha})\{\pi/(1-\pi)\}}\right] = 1 \qquad (3.5)$$

has a solution $\pi = \pi_0(\boldsymbol{\alpha})$. It is the unique solution as long as $\mathrm{pr}\{H(0,\mathbf{X},Y,\boldsymbol{\alpha}) > 0\} > 0$, since $\pi/(1-\pi)$ is strictly increasing, ranging from 0 to ∞. Based on (3.5), we can construct a root-N consistent estimator of π_0 and plug it into (3.4). The resulting estimators of the $\zeta_1(\cdot)$ have the same bias and mean squared error order as the usual nonparametric estimator. The proof is provided in Appendix A.5.

For simplicity, one may use the same index in (3.1)-(3.3), i.e. assuming $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_3 = \boldsymbol{\gamma}$. As before, we restrict the lower square block of $\boldsymbol{\gamma}$ to be identity. We provide detailed estimation procedures and algorithms for both cases, with the algorithm for different indices in Appendix A.2.1 and that for the same index in Appendix A.2.2.

Remark 1. It is worth pointing out that the estimation of π via (3.5) originates from

$$\begin{aligned}
\pi_0 &= \int H(0,\mathbf{X},Y,\boldsymbol{\alpha}) f^{\mathrm{true}}_{Y|\mathbf{X}}(y,\mathbf{x},\boldsymbol{\beta}) f^{\mathrm{true}}_{\mathbf{X}}(\mathbf{x}) d\mu(\mathbf{x}) d\mu(y) \\
&= \int \frac{H(0,\mathbf{X},Y,\boldsymbol{\alpha})}{\sum_d N_d/(N\pi_d) H(d,\mathbf{X},Y,\boldsymbol{\alpha})} \\
&\qquad \times \sum_d N_d/(N\pi_d) H(d,\mathbf{X},Y,\boldsymbol{\alpha}) f^{\mathrm{true}}_{Y|\mathbf{X}}(y,\mathbf{x},\boldsymbol{\beta}) f^{\mathrm{true}}_{\mathbf{X}}(\mathbf{x}) d\mu(\mathbf{x}) d\mu(y) \\
&= \int \frac{H(0,\mathbf{X},Y,\boldsymbol{\alpha})}{\sum_d N_d/(N\pi_d) H(d,\mathbf{X},Y,\boldsymbol{\alpha})} f_{\mathbf{X},Y}(y,\mathbf{x},\boldsymbol{\beta}) d\mu(\mathbf{x}) d\mu(y).
\end{aligned}$$

Thus, the estimator takes into account the difference between the superpopulation and the population from which the case-control sample is drawn, and thus leads to a consistent estimator of π_0.

3.4. Estimation Algorithm Using Different Indices

The estimating equation in (2.5) relies on the unknown probability density function η_2. Here, we use a posited model η_2^*, which is not necessarily the truth, to calculate the efficient score and other related quantities. The resulting estimating function is denoted by $\mathbf{S}^*_{\mathrm{eff}}$. We will show that the resulting estimator is still consistent, and it is efficient if the posited model η_2^* is the correct one.

The main difficulty in calculating $\mathbf{S}^*_{\mathrm{eff}}$ lies in approximating functions $\mathbf{g}, \mathbf{v}_0$, and $\mathbf{v}_1$, because they depend on three expectations conditional on covariates $\mathbf{X}$, which need to be estimated nonparametrically. We bypass this difficulty via the dimension reduction strategy described in Section 3.1-3.3. A sketch of the algorithm is the following.

1. Posit a model for $\eta_2(\epsilon, \mathbf{x})$ which has mean zero. Under this posited model, calculate S^* from (2.6).
2. Solve $\widehat{\pi}_0(\boldsymbol{\alpha}) = \sum_{i=1}^{N} H(0, \mathbf{X}_i, Y_i, \boldsymbol{\alpha})[N_0 H(0, \mathbf{X}_i, Y_i, \boldsymbol{\alpha})/\widehat{\pi}_0(\boldsymbol{\alpha}) + N_1 H(1, \mathbf{X}_i, Y_i, \boldsymbol{\alpha})/\{1 - \widehat{\pi}_0(\boldsymbol{\alpha})\}]^{-1}$, and set $\widehat{\pi}_1(\boldsymbol{\alpha}) = 1 - \widehat{\pi}_0(\boldsymbol{\alpha})$.
3. Estimate the indices $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3$ and the corresponding functions $\zeta_1, \boldsymbol{\zeta}_2, \zeta_3$ defined in (3.1)-(3.3) respectively by following the procedure in Section 3.3.
4. Plug the estimation from Step 3 into the expression of functions $\mathbf{g}, \mathbf{v}_0$ and v_1 in (2.7) to get $\widehat{\mathbf{g}}, \widehat{\mathbf{v}}_0$ and $\widehat{\mathbf{v}}_1$.
5. Form $\widehat{\mathbf{S}}^*_{\text{eff}}(D_i, \mathbf{X}_i, Y_i) = \mathbf{S}^*_i - \widehat{\mathbf{g}}_i - \widehat{\mathbf{v}}_{D_i}$ and solve the corresponding estimating equation.

For convenience, we adopt $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_3 = \boldsymbol{\gamma}$ in all the simulations, where the lower square block of $\boldsymbol{\gamma}$ is set to be identity to ensure identifiability. The algorithm in this simplified case is identical to the one described above except step 3. The detailed algorithms for cases using different indices and using a common index are given in Appendix A.2.

4. Distribution Theory

We now establish the asymptotic distribution theory of our estimators, stated as Theorem 1 below, with necessary regularity conditions C1-C11 listed in Appendix A.3. The proof of Theorem 1 is detailed and lengthy and is thus sketched in the Appendix Section A.5. While Theorem 1 holds for both the estimator using different indices and the estimator using a common index, we only provide the proof and regularity conditions for the algorithm with different indices. One can easily adapt the conditions and proof to the case of a common index.

Under the regularity conditions C1-C11 listed in Appendix A.3, the following theorem holds. The proof is in the Appendix Section A.5.

Theorem 1. *Define*

$$\mathbf{A} = E\left\{\partial \mathbf{S}^*_{\text{eff}}(D, \mathbf{X}, Y, \boldsymbol{\theta})/\partial \boldsymbol{\theta}\right\}$$

and

$$\mathbf{B} = cov\left\{\mathbf{S}^*_{\text{eff}}(D, \mathbf{X}, Y, \boldsymbol{\theta})\right\}.$$

The estimator $\widehat{\theta}$ obtained from solving the estimating equation

$$\sum_{i=1}^{N} \widehat{\mathbf{S}}^*_{\text{eff}}(D_i, \mathbf{X}_i, Y_i, \widehat{\boldsymbol{\theta}}) = \mathbf{0} \tag{4.1}$$

satisfies $\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to Normal\{0, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}}\}$ and $\widehat{\boldsymbol{\theta}}$ is locally efficient, see the definition of locally efficient in Section 1.

5. Simulations

5.1. Setup

We performed a series of simulations to understand the behaviour of our method and compare it to competitors. The simulations displayed in this section are for the case that the regression errors ϵ are Gaussian or centered Gamma, both homoscedastic and heteroscedastic.

In these simulations, we considered different disease rates, different dimensions and distributions for $\mathbf{X}$ and different error variance structures. The results indicate that our methods have small bias and good coverage probability in all the cases we examined. Here, due to space limitations, we only list the results for two typical scenarios, where the first one is homoscedastic and the second one is heteroscedastic. In both cases, we chose a balanced design with $N_1 = 1000$ cases and $N_0 = 1000$ controls, set the disease rate to be approximately 4.5% and let $\mathbf{X}$ be exchangeable with $p = \dim(\mathbf{X}) = 4$.

More specifically, we generated $\mathbf{X} = (X_1, \cdots, X_4)^{\mathrm{T}}$ in the following way.

1. Generate $\mathbf{X}^* = \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\Sigma_{i,j})_{1 \le i,j \le 4}$ and $\Sigma_{i,j} = 1$ if $i = j$ and $\Sigma_{i,j} = \rho$ for $|\rho| < 1$ if $1 \le i \ne j \le 4$.
2. Let $\mathbf{X} = \Phi(\mathbf{X}^*) = \{\Phi(\mathbf{X}_1^*), \cdots, \Phi(\mathbf{X}_4^*)\}^{\mathrm{T}}$, where Φ is the cumulative distribution function of a standard normal random variable.

Hence, $\mathbf{X}$ is an exchangeable vector of i-th random variables satisfying $X_i = \text{Uniform}[0,1]$ for $i = 1, \cdots, 4$ and $\text{corr}(X_i, X_j) = \text{corr}(X_k, X_l)$ for all $i \ne j, k \ne l$. In our simulation studies, we used $\rho = 0.2$, which resulted in $\text{corr}(X_i, X_j) \approx 0.191$ for all $1 \le i \ne j \le 4$.

As mentioned in the opening paragraph of this section, in this section we display results when the regression errors are Gaussian or Gamma. Specifically, we generated homoscedastic errors ϵ as $\text{Normal}(0,1)$ and we generated heteroscedastic errors ϵ such that $\epsilon \mid \mathbf{X} \sim \text{Normal}(0, [1 + \{\mathbf{X}^{\mathrm{T}}(\boldsymbol{\alpha}_1 + \alpha_2 \boldsymbol{\beta}_1)\}^2]^{3/2}/4)$. In the Gamma case, we generated homoscedastic errors ϵ from a Gamma distribution with shape parameter 0.4, scale parameter 1.8 and then normalized it to have mean 0 and variance 1; we generated heteroscedastic errors ϵ using the same distribution except that ϵ was multiplied by $[1 + \{\mathbf{X}^{\mathrm{T}}(\boldsymbol{\alpha}_1 + \alpha_2 \boldsymbol{\beta}_1)\}^2]^{3/4}/2$.

To obtain an approximately 4.5% disease rate in both Gaussian and Gamma cases with both homoscedastic and heteroscedastic errors, we first set $\alpha_c = -3.6, \boldsymbol{\alpha}_1 = (-1.0, 0.3, 0.5, 0.7)^{\mathrm{T}}$ and $\alpha_2 = 0.6$ in the logistic model $\text{pr}(D = 1 | \mathbf{X}, Y) = H(\alpha_c + \mathbf{X}^{\mathrm{T}} \boldsymbol{\alpha}_1 + Y \alpha_2)$. Then we set the regression model for Y to be linear, i.e., $Y = \beta_0 + \mathbf{X}^{\mathrm{T}} \boldsymbol{\beta}_1 + \epsilon$ and let $\beta_0 = -1.1, \boldsymbol{\beta}_1 = (0.5, 1.0, 0.3, 0.5)^{\mathrm{T}}$. For each setting, we generated 1,000 simulated data sets.

We set the posited model η_2^* to be $\text{Normal}(0,1)$ and adopted the estimation algorithm discussed in Section 3.4 and Appendix A.2 for the three important conditional expectations $E_{\text{true}}\{\epsilon^2 \kappa(\mathbf{X}, Y) \mid \mathbf{X}\}$, $E_{\text{true}}\{\epsilon \boldsymbol{\mu}_s(\mathbf{X}, Y) \mid \mathbf{X}\}$ and $E_{\text{true}}\{\epsilon f_{D|\mathbf{X},Y}(0, \mathbf{X}, Y) \mid \mathbf{X}\}$. In steps (a)-(c) in Appendix A.2 that involves nonparametric calculations, we used the asymptotically justified band-

width $h = cn_0^{-1/5}$: we found that when $c \in [1, 6]$, the estimation results are very similar.

5.2. Results

We contrasted three methods. The first one is ordinary least squares using controls only. The second one is the semiparametric efficient method that assumes the regression error ϵ to be normally distributed with homoscedastic variance and $E(Y \mid \mathbf{X})$ to be linear in $\mathbf{X}$, or equivalently in our notation, $m(\mathbf{X}, \boldsymbol{\beta}) = \beta_0 + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1$ (Lin and Zeng, 2009). This method also requires a rare or known disease rate, which was set to 0.1% in the simulations. The third is our method described in Section 3.4, which does not require the rare disease assumption and does not put any restriction on ϵ other than that $E(\epsilon|\mathbf{X}) = 0$.

To implement Lin and Zeng's method, we used their software SPREG provided on http://dlin.web.unc.edu/software/spreg-2/, which adopts the rare disease assumption if the input disease rate is less than 1%. This software was designed to work in a semiparametric framework where it assumes a fully parametric Gaussian model for ϵ but the distribution of $\mathbf{X}$ is nonparametric. However, through multiple attempts we found that their software can only handle the case where components of $\mathbf{X}$ are independent. Thus, before running SPREG, we decorrelated $\mathbf{X}$ by multiplying it by L^{-1}, where L is the Cholesky decomposition of the $\mathrm{cov}(\mathbf{X}) = \Sigma$ satisfying $LL^{\mathrm{T}} = \Sigma$. In the simulations, we used the true covariance matrix Σ to fulfill the restriction of SPREG. However when dealing with the mammographic density data in Section 6, the true covariance matrix Σ is unknown. We estimated it using only the controls.

The results are summarized in Tables 1-2. In the homoscedastic Gaussian scenario (Table 1), the approach using only controls ("Ctrl") is asymptotically valid with small bias and near nominal coverage. Lin and Zeng's method ("Param"), which assumes normality and homoscedasticity, has the smallest standard deviation among the three methods since it is efficient if the errors are normal. However, it suffers from slight bias since the true disease rate is 4.5%, larger than 1%. Our method ("Semi"), which assumes neither normality nor rare disease, is superior considering overall performance. It has the smallest bias compared with the other two methods. In addition, its mean-squared error efficiency is from 60.0% to 79.9% greater than using only controls and is comparable to Lin and Zeng's method. In the homoscedastic Gamma case (Table 2), Lin and Zeng's methods has considerable bias, under-coverage and loss of mean squared error efficiency.

In the heteroscedastic scenario, for both Gaussian and Gamma errors, both the "Ctrl" and the "Param" methods suffered from low coverage probabilities while our approach ("Semi") maintains nominal coverage. The approach using only controls is reasonably unbiased in the Gaussian case but suffers from much larger bias in the Gamma case. In both cases, Lin and Zeng's parametric method gives badly biased estimates, low coverage probabilities and low mean squared error efficiency. Taking β_{13}, the third element in $\boldsymbol{\beta}_1$, as an example, while the

		Homoscedastic Gaussian error				Heteroscedastic Gaussian error			
	β_1	0.5	1.0	0.3	0.5	0.5	1.0	0.3	0.5
Ctrl	mean	0.514	0.977	0.280	0.480	0.543	0.940	0.254	0.433
	s.d.	0.113	0.115	0.114	0.111	0.106	0.103	0.100	0.101
	est. sd	0.114	0.113	0.113	0.113	0.102	0.102	0.102	0.102
	95%	0.957	0.941	0.951	0.947	0.922	0.910	0.937	0.900
Param	mean	0.523	0.970	0.273	0.461	0.264	1.257	0.495	0.781
	s.d.	0.082	0.085	0.087	0.084	0.089	0.083	0.088	0.086
	est. sd	0.083	0.084	0.087	0.087	0.089	0.082	0.088	0.088
	95%	0.948	0.942	0.933	0.932	0.250	0.115	0.406	0.101
	MSE Eff	1.759	1.717	1.618	1.484	0.204	0.196	0.263	0.170
Semi	mean	0.507	0.992	0.292	0.493	0.510	0.986	0.289	0.484
	s.d	0.089	0.088	0.086	0.087	0.102	0.093	0.092	0.098
	est. sd	0.091	0.093	0.091	0.094	0.093	0.095	0.089	0.100
	95%	0.960	0.964	0.961	0.975	0.932	0.957	0.936	0.950
	MSE Eff	1.600	1.755	1.799	1.666	1.240	1.606	1.396	1.490

TABLE 1

Simulation study in Section 5 with $N_1 = 1,000$ cases and $N_0 = 1,000$ controls, disease rate of approximately 4.5% and 4-dimensional correlated covariates $\mathbf{X}$ over 1,000 simulated data sets. The results for the homoscedastic normal error model are listed on the left and the results for the heteroscedastic normal error model are listed on the right. The three analyses performed are "Ctrl", which is ordinary least squares using only controls, "Param", which is semiparametric efficient method proposed by Lin and Zeng (2009) assuming normality and homoscedasticity, and "Semi", which is our new estimator described in Section 3.4. Here, we list the sample mean ("mean"), the sample standard deviation ("s.d."), the mean estimated standard deviation ("est. sd") and the coverage for the nominal 95% confidence intervals ("95%") for all three methods. In addition, we computed the mean squared error efficiency compared to using only controls for the "Param" and "Semi" methods.

		Homoscedastic Gamma error				Heteroscedastic Gamma error			
	β_1	0.5	1.0	0.3	0.5	0.5	1.0	0.3	0.5
Ctrl	mean	0.522	0.967	0.277	0.470	0.581	0.902	0.228	0.394
	s.d.	0.102	0.101	0.103	0.099	0.086	0.087	0.084	0.090
	est. sd	0.100	0.100	0.100	0.100	0.087	0.087	0.087	0.087
	95%	0.942	0.939	0.934	0.938	0.858	0.782	0.876	0.751
Param	mean	0.630	0.830	0.165	0.301	0.173	1.393	0.585	0.922
	s.d.	0.135	0.135	0.135	0.135	0.144	0.124	0.127	0.137
	est. sd	0.131	0.134	0.135	0.136	0.138	0.127	0.133	0.130
	95%	0.820	0.750	0.831	0.691	0.368	0.124	0.427	0.105
	MSE Eff	0.307	0.239	0.307	0.186	0.110	0.100	0.125	0.098
Semi	mean	0.502	0.995	0.299	0.501	0.513	0.981	0.291	0.482
	s.d	0.068	0.068	0.067	0.068	0.084	0.081	0.073	0.088
	est. sd	0.066	0.068	0.066	0.069	0.087	0.096	0.085	0.105
	95%	0.948	0.958	0.947	0.955	0.948	0.958	0.953	0.946
	MSE Eff	2.314	2.449	2.528	2.345	1.922	2.463	2.261	2.388

TABLE 2

Simulation study in Section 5 with $N_1 = 1,000$ cases and $N_0 = 1,000$ controls, disease rate of approximately 4.5% and 4-dimensional correlated covariates $\mathbf{X}$ over 1,000 simulated data sets. The results for the homoscedastic gamma error model are listed on the left and the results for the heteroscedastic gamma error model are listed on the right. The three analyses performed are "Ctrl", which is ordinary least squares using only controls, "Param", which is semiparametric efficient method proposed by Lin and Zeng (2009) assuming normality and homoscedasticity, and "Semi", which is our new estimator described in Section 3.4. Here, we list the sample mean ("mean"), the sample standard deviation ("s.d."), the mean estimated standard deviation ("est. sd") and the coverage for the nominal 95% confidence intervals ("95%") for all three methods. In addition, we computed the mean squared error efficiency compared to using only controls for the "Param" and "Semi" methods.

nominal coverage is 95%, the actual coverage rates are 40.6% and 43.7% in the Gaussian and Gamma case, respectively. Our approach has no larger than 4% bias compared with the truth, which is the best among three methods. It also achieves the best coverage probabilities and smallest mean-squared errors.

We have done other simulations with different disease rates, and the overall picture remains the same as what we have described above. For example, in the Appendix Section A.7, we display results for the case that the intercept was adjusted to make the disease rate in the source population $\approx 10\%$.

We have also done simulations when the dimension of $\mathbf{X}$ is 6, 8 and 10 with an approximate 4.5% disease rate, and found results similar to the ones previously described. Of course the computation takes longer as the dimension of $\mathbf{X}$ increases. Please see the Appendix Section A.8 for numerical results.

Remark 2. *While a number of methods on secondary analysis exist in the literature, none of them is applicable in our setting. For example, Jiang et al. (2006) and Li et al. (2010) focused on binary Y, for which a logistic regression model for Y and $\mathbf{X}$ or Y and $(\mathbf{X}, D)$ was considered. Ma and Carroll (2016) adopted kernel density regression in their estimation procedure, and thus it is not applicable to the cases with multivariate $\mathbf{X}$ due to the curse of dimensionality. Wei et al. (2013) requires the known or rare disease assumption as well as homoscedastic regression errors, and hence is also not applicable in our model setting. Likewise, Lin and Zeng (2009) requires the known or rare disease assumption and is applicable only when the secondary model is parametric. Thus, we have compared our approach to only two methods, the control only method for its simplicity and sometimes surprisingly good result when the disease is truly rare, and Lin and Zeng's method for its gold standard status in practice, for parametric models.*

6. Analysis of Mammographic Density Data

Here we apply our methodology in a case-control study of breast cancer, where the data were collected from women in the breast cancer detection demonstration project (BCDDP), see Chen et al. (2006) and Chen et al. (2008). The study recruited a total of 284,780 women, starting from January 1, 1973 and ended December 31, 1995. Then in the following five years, follow-up annual screening was performed for each subject. Here the period from 1973-1980 is referred to as the "screening phase" of the study. At the end of the screening phase, the study selected all cases, i.e. women who developed breast cancer, and sampled from the controls. All the selected women were included in a further extended follow-up study from 1980 to 1995. Standard risk factors, including age at menarche, age at first live birth and body mass index, were available in this study. However, we were only able to retrieve mammographic density measurements at baseline in 1973-1975 for $N_1 = 2092$ cases and $N_0 = 3295$ controls.

Mammographic density is a measure of the average of dense tissue percentage in both breasts. Women's breasts consist of fat, breast tissue, nerves, veins, arteries and connective tissue that holds everything in place. Both breast tissue

and connective tissue are denser than fat. Previous studies showed that higher mammographic density is a strong risk factor for breast cancer. In addition, age at menarche and age at first live birth are both known to be associated with breast cancer. Women who have their first menstruation before age 12 have a slightly higher chance of developing breast cancer compared with those who have their first period after 14; women who give birth to their first child at a young age tend to have a relatively lower risk of developing breast cancer. Body mass index is another risk factor for breast cancer. Before menopause, being slightly overweight can reduce breast cancer risk. However, there is little existing work discussing the interrelationship between mammographic density, age at menarche, age at first live birth and body mass index. The goal of our analysis is to investigate this interrelationship. Before implementing our method, we used an inverse logistic transformation on mammographic density and rescaled the other three risk factors to [0,1] by subtracting their minimums and dividing by the ranges.

Preliminary analysis based on only the controls data showed that mammographic density is reasonably linear in age at menarche, age at first live birth and body mass index. To check this, we fit both a linear regression model and a quadratic regression model using controls and compared these two models via analysis of variance. The p-value is about .78, which indicates the linear model is preferred over the quadratic model. Hence, we adopted a linear $m(\cdot)$ in the secondary analysis. The diagnostic plots of linear regression are given in Figure 1. The left plot is the kernel density estimate of the residuals from a linear fit on the controls, with an overlaid normal density. It shows that the regression error almost follows a normal distribution but with slightly negative skewness. The right plot is the LOWESS smoother of fitted values versus the square roots of absolute values of residuals, which indicates the regression error is homoscedastic.

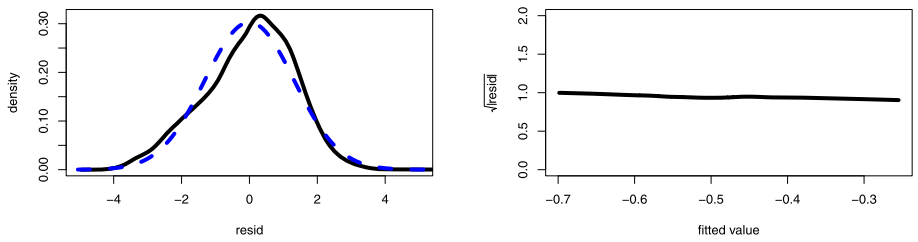

FIG 1. *Mammographic density data in Section 6. The left plot is the kernel density estimate (solid black line) of the residuals from a linear fit on the controls, with an overlaid normal density (dashed blue line). The right plot is the LOWESS smoother of fitted values versus the square roots of absolute values of residuals: the fact that it is flat indicates little heteroscedasticity.*

The results of secondary analysis using only controls, Lin and Zeng's parametric method and our semiparametric approach based on 1000 bootstrap samples

		MENARCHE	1STLB	BMI
Ctrl	mean	-0.047	0.428	-0.105
	boot. sd	0.164	0.139	0.172
	est. sd	0.165	0.144	0.176
	Lower	-0.371	0.146	-0.449
	Upper	0.277	0.710	0.240
Param	mean	-0.054	0.356	-0.121
	boot. sd	0.131	0.106	0.138
	est. sd	0.127	0.107	0.135
	Lower	-0.302	0.147	-0.385
	Upper	0.195	0.565	0.144
	Eff	1.550	1.710	1.547
Semi	mean	-0.061	0.363	-0.135
	boot. sd	0.129	0.107	0.137
	est. sd	0.130	0.113	0.140
	Lower	-0.315	0.142	-0.410
	Upper	0.194	0.584	0.140
	Eff	1.606	1.698	1.575

TABLE 3

Analyses of the mammographic density data from the breast cancer detection demonstration project (BCDDP) in Section 6, which has $N_1 = 2092$ cases and $N_0 = 3295$ controls, using only controls ("Ctrl"), Lin and Zeng's method ("Param") and our approach ("Semi"). Displayed are the mean estimates of the coefficients for age at menarche (MENARCHE), age at first live birth (1STLB) and body mass index (BMI), their bootstrap standard deviation ("boot. sd"), the mean estimated bootstrap standard deviation ("est. sd") and the lower and upper end values of the 95% confidence intervals ("Lower" and "Upper"). Also displayed is the efficiency ("Eff"), which is the square of the ratio of bootstrap standard deviation to that using only controls.

are given in Table 3. All three methods have fairly consistent results as expected, due to the fact that the regression error is homoscedastic and close to normal. For all three methods, age at first live birth is highly statistically significant with a positive effect on mammographic density. That is women who gave birth to their first children earlier tend to have a lower mammographic density, and hence obtain some protective effect from developing breast cancer. Both age at menarche and body mass index have negative coefficients, which indicates that having a relatively late first period or being moderately overweight can slightly reduce mammographic density. However, neither of them is statistically significant.

As expected, Lin and Zeng's parametric method has a much smaller bootstrap standard deviation compared with the ordinary least squares using only controls, with an average efficiency of 1.60. Here the efficiency is defined as the square of the ratio of bootstrap standard deviation compared with using only controls. Our semiparametric approach, which assumes neither homoscedasticity nor normality, has almost the same bootstrap standard deviation as Lin and Zeng's method. The bootstrap standard errors of Lin and Zeng's parametric approach for age at menarche, age at first live birth and body mass index are 0.131, 0.106, 0.138, respectively, while that of our semiparametric approach are 0.129, 0.107 and 0.137 respectively. The average efficiency of our approach is 1.63, which is even slightly larger than that of Lin and Zeng's method.

7. Discussion

We have extended the work of Ma and Carroll (2016) and have overcome the potential dimensionality issue involved in their nonparametric kernel regression. Multivariate kernel regression is avoided by using dimension reduction modeling ideas. We repeat that our work is not about fitting dimension reduction models per se, but to use them in the secondary analysis of case-control studies. Our method makes no assumptions about the regression errors, and we do not need to make a rare disease assumption or require known disease rate.

The dimension reduction assumptions stated in (3.1)-(3.3) are mild in general, see Proposition 1, and are applicable in many practical situations. An interesting topic for future work would be to consider using regularization to further reduce the dimension of $\mathbf{Z}_{\boldsymbol{\beta}}$ so as to obtain an even more parsimonious model.

Alternative dimension reduction modeling approaches could exist, although it is not easy to identify them based on our preliminary analysis along this line. For example, generalized additive models do not appear to be suitable in the common regression error structures described in Section 3.2. For example, in (3.1),

$$E\{\epsilon^2\kappa(\mathbf{X},Y,\boldsymbol{\alpha})\mid\mathbf{X}\}=E(\epsilon^2G[\{\mathbf{X}^{\mathrm{T}},m(\mathbf{X},\boldsymbol{\beta})\}(\boldsymbol{\alpha}_1^{\mathrm{T}},\alpha_2)^{\mathrm{T}}+\epsilon\alpha_2]\mid\mathbf{X}).$$

where G is a function of the logistic distribution function, i.e., a function of several exponential functions. It is not clear that this can be written as a generalized additive model. Even if it can be done, using such a dimension reduction approach will still require careful exploration and new methodology development because off-the-shelf results on generalized additive models may not apply due to the case-control sampling nature.

Finally, in some cases, it might be possible to posit a parametric form for $\mathrm{var}(\epsilon\mid\mathbf{X})$. We believe that our approach can be extended to this case, and would further improve efficiency in estimating $\boldsymbol{\beta}$. This will be pursued in future work.

Appendix A: Sketch of Technical Arguments and Additional Numerical Results

A.1. Dimension Reduction Model (3.2)

The dimension reduction assumption (3.2) on $\epsilon\boldsymbol{\mu}_s$ is more complicated than (3.1) or (3.3), and requires careful attention.

In the usual case, we have that

$$\begin{aligned}S^*&=\begin{bmatrix}\partial\log\{H(d,\mathbf{X},Y,\boldsymbol{\alpha}))\}/\partial\boldsymbol{\alpha}\\ \partial\log\{\eta_2^*(\epsilon,\mathbf{X}))\}/\partial\boldsymbol{\beta}\end{bmatrix}\\&=\begin{bmatrix}\{d-H(1,\mathbf{X},Y,\boldsymbol{\alpha})\}\,(1,\mathbf{X}^{\mathrm{T}},Y)^{\mathrm{T}}\\ -\{\eta_2^*(\epsilon,\mathbf{X})^{-1}\partial\eta_2^*(\epsilon,\mathbf{X})/\partial\epsilon\}\times\{\partial m(\mathbf{X},\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}\end{bmatrix},\end{aligned}$$

where $y = m(\mathbf{x}, \boldsymbol{\beta}) + \epsilon$. Here η_2^* is the posited conditional density of ϵ given $\mathbf{X}$, not necessarily the true model. Let $w(d, \mathbf{x}, y; \boldsymbol{\alpha}) = d - H(1, \mathbf{X}, Y, \boldsymbol{\alpha})$, so that

$$\boldsymbol{\mu}_s = E(S^* \mid \epsilon, \mathbf{X}) = \begin{bmatrix} r(\mathbf{x}, y; \boldsymbol{\alpha})(1, \mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}} \\ -\{\eta_2^*(\epsilon, \mathbf{X})^{-1} \partial \eta_2^*(\epsilon, \mathbf{X})/\partial \epsilon\} \times \{\partial m(\mathbf{X}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}\} \end{bmatrix},$$

where

$$\begin{aligned} r(\mathbf{x}, y; \boldsymbol{\alpha}) &= E\{w(D, \mathbf{X}, Y) \mid \mathbf{X}, Y\} \\ &= \sum_{d=0}^{1} N_d H(d, \mathbf{X}, Y) w(d, \mathbf{X}, Y) \kappa(\mathbf{X}, Y)/(N\pi_d), \\ &= N^{-1} \left(N_1/\pi_1 - N_0/\pi_0\right) H(0, \mathbf{X}, Y) H(1, \mathbf{X}, Y) \kappa(\mathbf{X}, Y); \\ \kappa(\mathbf{X}, Y) &= \left\{ \sum_{d=0}^{1} N_d H(d, \mathbf{X}, Y)/(N\pi_d) \right\}^{-1}. \end{aligned}$$

Hence,

$$\begin{aligned} &E_{\mathrm{true}}\{\epsilon \boldsymbol{\mu}_s(\mathbf{X}, Y) \mid \mathbf{X}\} \\ &\quad = \begin{bmatrix} E_{\mathrm{true}}\{\epsilon r(\mathbf{X}, Y; \boldsymbol{\alpha}) \mid \mathbf{X}\}(1, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}} \\ E_{\mathrm{true}}\{\epsilon r(\mathbf{X}, Y; \boldsymbol{\alpha}) m(\mathbf{X}, \boldsymbol{\beta}) + \epsilon^2 r(\mathbf{X}, Y; \boldsymbol{\alpha}) \mid \mathbf{X}\} \\ -E_{\mathrm{true}}\left\{\epsilon \eta_2^*(\epsilon, \mathbf{X})^{-1} \partial \eta_2^*(\epsilon, \mathbf{X})/\partial \epsilon \mid \mathbf{X}\right\} \{\partial m(\mathbf{X}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}\} \end{bmatrix}. \end{aligned}$$

We assume the following models hold.

$$E_{\mathrm{true}}\{\epsilon r(\mathbf{X}, Y; \boldsymbol{\alpha}) \mid \mathbf{X}\} = \zeta_{21}(\mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{\gamma}_{21}); \tag{A.1}$$

$$E_{\mathrm{true}}\{\epsilon^2 r(\mathbf{X}, Y; \boldsymbol{\alpha}) \mid \mathbf{X}\} = \zeta_{22}(\mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{\gamma}_{22}); \tag{A.2}$$

$$E_{\mathrm{true}}\left\{\epsilon \eta_2^*(\epsilon, \mathbf{X})^{-1} \partial \eta_2^*(\epsilon, \mathbf{X})/\partial \epsilon \mid \mathbf{X}\right\} = \zeta_{23}(\mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{\gamma}_{23}), \tag{A.3}$$

where $\mathbf{Z} = \{\mathbf{X}^{\mathrm{T}}, m(\mathbf{X}, \boldsymbol{\beta})\}^{\mathrm{T}}$ when m is nonlinear while $\mathbf{Z} = \mathbf{X}$ when m is linear. For identifiability, the lower square blocks of $\boldsymbol{\gamma}_{2j}, j = 1, 2, 3$ are fixed to be identity.

In models (A.1)-(A.3), $\zeta_{21}, \zeta_{22}, \zeta_{23}$ can be estimated by

$$\begin{aligned} &\widehat{\zeta}_{21}(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma}_{21}) \\ &\quad = \frac{\sum_{d=0}^{1} \widehat{\pi}_d/N_d \sum_{i=1}^{N} I\{D_i = d\} \epsilon_i r(\mathbf{X}_i, Y_i; \boldsymbol{\alpha}) K_h(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\gamma}_{21} - \mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma}_{21})}{\sum_{d=0}^{1} \widehat{\pi}_d/N_d \sum_{i=1}^{N} I\{D_i = d\} K_h(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\gamma}_{21} - \mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma}_{21})}; \end{aligned} \tag{A.4}$$

$$\begin{aligned} &\widehat{\zeta}_{22}(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma}_{22}) \\ &\quad = \frac{\sum_{d=0}^{1} \widehat{\pi}_d/N_d \sum_{i=1}^{N} I\{D_i = d\} \epsilon_i^2 r(\mathbf{X}_i, Y_i; \boldsymbol{\alpha}) K_h(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\gamma}_{22} - \mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma}_{22})}{\sum_{d=0}^{1} \widehat{\pi}_d/N_d \sum_{i=1}^{N} I\{D_i = d\} K_h(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\gamma}_{22} - \mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma}_{22})}; \end{aligned} \tag{A.5}$$

$$\widehat{\zeta}_{23}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_{23}) = \frac{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}\epsilon_i\frac{\partial\eta_2^*(\epsilon_i,\mathbf{X}_i)/\partial\epsilon_i}{\eta_2^*(\epsilon_i,\mathbf{X}_i)}K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_{23}-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_{23})}{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_{23}-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_{23})}. \quad \text{(A.6)}$$

To get a consistent estimate of $\boldsymbol{\gamma}_{21,-1}$, we solve

$$\begin{aligned}\mathbf{0} &= \sum_{d=0}^{1}\frac{\widehat{\pi}_d}{N_d}\sum_{j=1}^{N}I(D_j=d)\times\{\epsilon_j(\mathbf{X}_j,Y_j,\boldsymbol{\beta})r(\mathbf{X}_j,Y_j,\boldsymbol{\alpha})\\ &\qquad -\widehat{\zeta}_{21}(\mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\gamma}_{21})\Big\}\Big\{\mathbf{Z}^*_{\boldsymbol{\beta},j}-\widehat{E}^{\widehat{\pi}}_{\mathrm{true}}(\mathbf{Z}^*_{\boldsymbol{\beta},j}\mid\mathbf{Z}^*_{\boldsymbol{\beta},j}\boldsymbol{\gamma})\Big\}.\end{aligned}$$

Similar results work for $\boldsymbol{\gamma}_{22,-1}$ and $\boldsymbol{\gamma}_{23,-1}$. Denote the resulting estimators by $\widehat{\boldsymbol{\gamma}}_{2j,-1}$ and let $\widehat{\boldsymbol{\gamma}}_{2j}=(\widehat{\boldsymbol{\gamma}}_{2j,-1}^{\mathrm{T}},1)^{\mathrm{T}}$ for $j=1,2,3$. Then $E_{\mathrm{true}}\{\epsilon\boldsymbol{\mu}_s(\mathbf{X},Y)\mid\mathbf{X}\}$ can be estimated by

$$\widehat{E}_{\mathrm{true}}\{\epsilon\widehat{\boldsymbol{\mu}}_s(\mathbf{X},Y)\mid\mathbf{X}\}=\begin{bmatrix}\widehat{\zeta}_{21}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_{21})(1,\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\\ \widehat{\zeta}_{21}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_{21})m(\mathbf{X},\boldsymbol{\beta})+\widehat{\zeta}_{22}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_{22})\\ -\widehat{\zeta}_{23}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_{23})\{\partial m(\mathbf{X},\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}\end{bmatrix}.$$

In all of our simulations, $m(\mathbf{X},\boldsymbol{\beta})=\beta_0+\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1$. In addition, the posited model is standard normal, and simplifications result. Thus, $\partial\{\log\eta_2^*(\epsilon,\mathbf{X})\}/\partial\epsilon$ is simply $-\epsilon$. In our simulations, we further take $\boldsymbol{\gamma}_{21}=\boldsymbol{\gamma}_{22}=\boldsymbol{\gamma}_{23}=\boldsymbol{\gamma}_2$ for computational and programming simplicity. As a result, we have that

$$\begin{aligned}S^* &= \left[\{d-H(1,\mathbf{X},Y,\boldsymbol{\alpha})\}(1,\mathbf{X}^{\mathrm{T}},Y),\epsilon(1,\mathbf{X}^{\mathrm{T}})\right]^{\mathrm{T}};\\ \boldsymbol{\mu}_s &= E\{S^*\mid\epsilon,\mathbf{X}\}=\{r(\mathbf{X},Y;\boldsymbol{\alpha})(1,\mathbf{X}^{\mathrm{T}},Y),\epsilon(1,\mathbf{X}^{\mathrm{T}})\}^{\mathrm{T}}.\end{aligned}$$

Then

$$E_{\mathrm{true}}\{\epsilon\boldsymbol{\mu}_s(\mathbf{X},Y)\mid\mathbf{X}\}=\begin{bmatrix}E_{\mathrm{true}}\{\epsilon r(\mathbf{X},Y;\boldsymbol{\alpha})\mid\mathbf{X}\}(1,\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\\ E_{\mathrm{true}}\{\epsilon r(\mathbf{X},Y;\boldsymbol{\alpha})\mid\mathbf{X}\}m(\mathbf{X},\boldsymbol{\beta})+E_{\mathrm{true}}\{\epsilon^2 r(\mathbf{X},Y;\boldsymbol{\alpha})\mid\mathbf{X}\}\\ E_{\mathrm{true}}\{\epsilon^2\mid\mathbf{X}\}(1,\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\end{bmatrix}.$$

Under the assumption $\boldsymbol{\gamma}_{21}=\boldsymbol{\gamma}_{22}=\boldsymbol{\gamma}_{23}=\boldsymbol{\gamma}_2$ in (A.4) -(A.6), $\zeta_{21},\zeta_{22},\zeta_{23}$ can be estimated by

$$\widehat{\zeta}_{21}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2) = \frac{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}\epsilon_i r(\mathbf{X}_i,Y_i;\boldsymbol{\alpha})K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_2-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2)}{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_2-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2)};$$

$$\widehat{\zeta}_{22}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2) = \frac{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}\epsilon_i^2 r(\mathbf{X}_i,Y_i;\boldsymbol{\alpha})K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_2-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2)}{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_2-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2)};$$

$$\widehat{\zeta}_{23}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2) = \frac{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}\epsilon_i^2K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_2-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2)}{\sum_{d=0}^{1}\widehat{\pi}_d/N_d\sum_{i=1}^{N}I\{D_i=d\}K_h(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}_2-\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}_2)},$$

where again the lower square block of $\boldsymbol{\gamma}_2$ is fixed to be identity. The consistent estimator of $\boldsymbol{\gamma}_{2,-1}$ can be obtained through solving

$$\begin{aligned}\mathbf{0} = \sum_{d=0}^{1}\frac{\widehat{\pi}_d}{N_d}\sum_{j=1}^{N}I(D_j=d)\Big[&\{\epsilon_j(\mathbf{X}_j,Y_j,\boldsymbol{\beta})r(\mathbf{X}_j,Y_j,\boldsymbol{\alpha})-\widehat{\zeta}_{21}(\mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\gamma}_2)\}\\&+\{\epsilon_i^2r(\mathbf{X}_i,Y_i;\boldsymbol{\alpha})-\widehat{\zeta}_{22}(\mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\gamma}_2)\}+\{\epsilon_i^2-\widehat{\zeta}_{23}(\mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\gamma}_2)\}\Big]\\&\times\Big\{\mathbf{Z}^*_{\boldsymbol{\beta},j}-\widehat{E}^{\widehat{\pi}}_{\mathrm{true}}(\mathbf{Z}^*_{\boldsymbol{\beta},j}\mid\mathbf{Z}^*_{\boldsymbol{\beta},j}\boldsymbol{\gamma})\Big\}.\end{aligned}$$

Denote the resulting estimators $\widehat{\boldsymbol{\gamma}}_{2,-1}$ and let $\widehat{\boldsymbol{\gamma}}_2=(\widehat{\boldsymbol{\gamma}}_{2,-1}^{\mathrm{T}},1)^{\mathrm{T}}$. $E_{\mathrm{true}}\{\epsilon\boldsymbol{\mu}_s(\mathbf{X},Y)\mid\mathbf{X}\}$ can be estimated by

$$\widehat{E}_{\mathrm{true}}\{\epsilon\widehat{\boldsymbol{\mu}}_s(\mathbf{X},Y)\mid\mathbf{X}\}=\begin{bmatrix}\widehat{\zeta}_{21}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_2)(1,\mathbf{X}^{\mathrm{T}})^{\mathrm{T}}\\ \widehat{\zeta}_{21}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_2)m(\mathbf{X},\boldsymbol{\beta})+\widehat{\zeta}_{22}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_2)\\ -\widehat{\zeta}_{23}(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_2)\{\partial m(\mathbf{X},\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}\end{bmatrix}.$$

A.2. Details for the Algorithm in Section 3.4

A.2.1. Algorithm Using Different Indices

1. Posit a model for $\eta_2(\epsilon,\mathbf{x})$ which has mean zero. Under this posited model, calculate S^* from (2.6).
2. Solve $\widehat{\pi}_0(\boldsymbol{\alpha})=\sum_{i=1}^{N}H(0,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})[N_0H(0,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})/\widehat{\pi}_0(\boldsymbol{\alpha})+N_1H(1,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})/\{1-\widehat{\pi}_0(\boldsymbol{\alpha})\}]^{-1}$, and set $\widehat{\pi}_1(\boldsymbol{\alpha})=1-\widehat{\pi}_0(\boldsymbol{\alpha})$.
3. Obtain

$$\begin{aligned}\widehat{\kappa}_i &= \widehat{\kappa}(\mathbf{X}_i,Y_i,\boldsymbol{\alpha})=\Big[\sum_d N_dH(d,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})/\{N\widehat{\pi}_d(\boldsymbol{\alpha})\}\Big]^{-1}\\ \widehat{f}_{0i} &= \widehat{f}_{D|X,Y}(0,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})=N_0H(0,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})\widehat{\kappa}_i/\{N\widehat{\pi}_0(\boldsymbol{\alpha})\}\\ \widehat{f}_{1i} &= \widehat{f}_{D|X,Y}(1,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})=N_1H(1,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})\widehat{\kappa}_i/\{N\widehat{\pi}_1(\boldsymbol{\alpha})\}\\ \widehat{\boldsymbol{\mu}}_{si} &= \widehat{E}(\mathbf{S}_i^*\mid\epsilon_i,\mathbf{X}_i,\boldsymbol{\alpha})\\ &= \sum_d N_dH(d,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})\mathbf{S}^*(d,\mathbf{X}_i,Y_i,\boldsymbol{\alpha})\widehat{\kappa}_i/\{N\widehat{\pi}_d(\boldsymbol{\alpha})\}\\ \widehat{b}_0 &= \sum_{i=1}^{N}\widehat{f}_{1i}\widehat{f}_{0i}/\sum_{i=1}^{N}\widehat{f}_{0i}\end{aligned}$$

$$\begin{aligned}
\widehat{b}_1 &= \sum_{i=1}^{N} \widehat{f}_{0i}\widehat{f}_{1i} / \sum_{i=1}^{N} \widehat{f}_{1i} \\
\widehat{\mathbf{c}}_0 &= \sum_{i=1}^{N} \{\mathbf{S}^*(0, \mathbf{X}_i, Y_i, \boldsymbol{\alpha}) - \widehat{\boldsymbol{\mu}}_{si}\} \widehat{f}_{0i} / \sum_{i=1}^{N} \widehat{f}_{0i} \\
\widehat{\mathbf{c}}_1 &= \sum_{i=1}^{N} \{\mathbf{S}^*(1, \mathbf{X}_i, Y_i, \boldsymbol{\alpha}) - \widehat{\boldsymbol{\mu}}_{si}\} \widehat{f}_{1i} / \sum_{i=1}^{N} \widehat{f}_{1i}.
\end{aligned}$$

4. Estimate $E_{\text{true}}\{\epsilon^2 \widehat{\kappa}(\mathbf{X}, Y) \mid \mathbf{X}\}$ using nonparametric regression under the dimension reduction model assumption (3.1).

   (a) Let

$$\begin{aligned}
&\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \boldsymbol{\theta}) \\
&\quad \equiv \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i \neq j} I(D_i = d) \epsilon_i^2(\mathbf{X}_i, Y_i, \boldsymbol{\beta}) \widehat{\kappa}(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) \times K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}} \boldsymbol{\gamma}_1 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}} \boldsymbol{\gamma}_1)}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i \neq j} I(D_i = d) K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}} \boldsymbol{\gamma}_1 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}} \boldsymbol{\gamma}_1)},
\end{aligned}$$

   for $j = 1, \cdots, N$. Here $\epsilon_i(\mathbf{X}_i, Y_i, \boldsymbol{\beta}) = Y_i - m(\mathbf{X}_i, \boldsymbol{\beta})$. $\mathbf{Z}_{\boldsymbol{\beta},i} = \mathbf{X}$ if $m(\cdot)$ is linear in $\mathbf{X}$; $\mathbf{Z} = \{\mathbf{X}_i^{\mathrm{T}}, m(\mathbf{X}_i, \boldsymbol{\beta})\}^{\mathrm{T}}$, otherwise.

   (b) Estimate $\boldsymbol{\gamma}_{1,-1}$ through solving

$$\begin{aligned}
\mathbf{0} &= \sum_{d=0}^{1} \widehat{\pi}_d(\boldsymbol{\alpha}) / N_d \sum_{j=1}^{N} I(D_j = d) \{\epsilon_j^2(\mathbf{X}_j, Y_j, \boldsymbol{\beta}) \widehat{\kappa}(\mathbf{X}_j, Y_j, \boldsymbol{\alpha}) \\
&\qquad - \widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \boldsymbol{\theta})\} \{\mathbf{Z}_{\boldsymbol{\beta},j}^* - \widehat{E}_{\text{true}}^{\widehat{\pi}}(\mathbf{Z}_{\boldsymbol{\beta},j}^* \mid \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}} \boldsymbol{\gamma}_1)\},
\end{aligned}$$

   where $\mathbf{Z}_{\boldsymbol{\beta},j}^*$ is the subvector or submatrix of $\mathbf{Z}_{\boldsymbol{\beta},j}$ without the lower square block and

$$\begin{aligned}
&\widehat{E}_{\text{true}}^{\widehat{\pi}}(\mathbf{Z}_{\boldsymbol{\beta},j}^* \mid \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}} \boldsymbol{\gamma}_1) \\
&\quad = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i \neq j} I(D_i = d) \mathbf{Z}_{\boldsymbol{\beta},i}^* K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}} \boldsymbol{\gamma}_1 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}} \boldsymbol{\gamma}_1)}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i \neq j} I(D_i = d) K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}} \boldsymbol{\gamma}_1 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}} \boldsymbol{\gamma}_1)}.
\end{aligned}$$

   Let the solution be $\widehat{\boldsymbol{\gamma}}_{1,-1}$. Denote $\widehat{\boldsymbol{\gamma}}_1 = (\widehat{\boldsymbol{\gamma}}_{1,-1}^{\mathrm{T}}, 1)^{\mathrm{T}}$.

   (c) Form

$$\widehat{E}_{\text{true}} \left\{\epsilon^2(\mathbf{X}, Y, \boldsymbol{\beta}) \widehat{\kappa}(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\right\}$$

$$= \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i=1}^{N} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \beta)\widehat{\kappa}(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) \times K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\widehat{\gamma}_1 - \mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\gamma}_1)}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i=1}^{N} I(D_i = d)K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\widehat{\gamma}_1 - \mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\gamma}_1)}.$$

5. Estimate $E_{\mathrm{true}}\{\epsilon\widehat{\boldsymbol{\mu}}_s(\mathbf{X}, Y) \mid \mathbf{X}\}$ using nonparametric regression under the dimension reduction model assumption (3.2). Because $E_{\mathrm{true}}\{\epsilon\widehat{\boldsymbol{\mu}}_s(\mathbf{X}, Y) \mid \mathbf{X}\}$ actually consists of three separate dimension reduction models, its estimation is slightly complex. We give the estimation details in Appendix A.1 and denote the resulting estimator by $\widehat{E}_{\mathrm{true}}\{\epsilon(\mathbf{X}, Y, \boldsymbol{\beta})\widehat{\boldsymbol{\mu}}_s(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\}$.
6. Estimate $E_{\mathrm{true}}\{\epsilon\widehat{f}_0(\mathbf{X}, Y) \mid \mathbf{X}\}$ using nonparametric regression under the dimension reduction model assumption (3.3).
   (a) Let

$$\widehat{E}_3^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_3, \boldsymbol{\theta}) \equiv \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I(D_i = d)\epsilon_i(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) \times K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma}_3 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma}_3)}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I(D_i = d)K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma}_3 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma}_3)},$$

   for $j = 1, \cdots, N$.
   (b) Estimate $\boldsymbol{\gamma}_{3,-1}$ by solving

$$\mathbf{0} = \sum_{d=0}^{1} \widehat{\pi}_d(\boldsymbol{\alpha})/N_d \sum_{j=1}^{N} I(D_j = d)\{\epsilon_j(\mathbf{X}_j, Y_j, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}_j, Y_j, \boldsymbol{\alpha}) - \widehat{E}_3^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_3, \boldsymbol{\theta})\}\{\mathbf{Z}_{\boldsymbol{\beta},j}^{*} - \widehat{E}_{\mathrm{true}}^{\widehat{\pi}}(\mathbf{Z}_{\boldsymbol{\beta},j}^{*} \mid \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma}_3)\},$$

   where

$$\widehat{E}_{\mathrm{true}}^{\widehat{\pi}}(\mathbf{Z}_{\boldsymbol{\beta},j}^{*} \mid \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma}_3) = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I(D_i = d)\mathbf{Z}_{\boldsymbol{\beta},i}^{*}K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma}_3 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma}_3)}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I(D_i = d)K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma}_3 - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma}_3)}.$$

   Let the minimizer be $\widehat{\boldsymbol{\gamma}}_{3,-1}$. Denote $\widehat{\boldsymbol{\gamma}}_3 = (\widehat{\boldsymbol{\gamma}}_{3,-1}^{\mathrm{T}}, 1)^{\mathrm{T}}$.
   (c) Form

$$\widehat{E}_{\mathrm{true}}\left\{\epsilon(\mathbf{X}, Y, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\right\}$$

$$= \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i=1}^{N} I(D_i = d)\epsilon_i(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) \times K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_3 - \mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_3)}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i=1}^{N} I(D_i = d)K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_3 - \mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_3)}.$$

7. (a) Form $\widehat{t}_1(\mathbf{X}) = \{\widehat{E}_{\mathrm{true}}(\epsilon^2\widehat{\kappa}(\mathbf{X}, Y)|\mathbf{X})\}^{-1}$, $\widehat{\mathbf{t}}_2(\mathbf{X}) = \widehat{E}_{\mathrm{true}}(\epsilon\widehat{\boldsymbol{\mu}}_s \mid \mathbf{X}) - (\widehat{\mathbf{c}}_0/\widehat{b}_0)\widehat{E}_{\mathrm{true}}(\epsilon\widehat{f}_0 \mid \mathbf{X})$ and $\widehat{t}_3(\mathbf{x}) = -\widehat{b}_0^{-1}\widehat{E}_{\mathrm{true}}(\epsilon\widehat{f}_0 \mid \mathbf{x})$.

   (b) Form $\widehat{E}\{\epsilon t_1(\mathbf{X})t_3(\mathbf{X})\kappa(\mathbf{X}, Y) \mid D = 0\} = \sum_{i=1}^{N} \epsilon_i\widehat{t}_1(\mathbf{X}_i)\widehat{t}_3(\mathbf{X}_i)\widehat{\kappa}(\mathbf{X}_i, Y_i)\widehat{f}_{0i}/\sum_{i=1}^{N} \widehat{f}_{0i}$, $\widehat{E}\{\epsilon t_1(\mathbf{X})\mathbf{t}_2(\mathbf{X})\kappa(\mathbf{X}, Y) \mid D = 0\} = \sum_{i=1}^{N} \epsilon_i\widehat{t}_1(\mathbf{X}_i)\widehat{\mathbf{t}}_2(\mathbf{X}_i) \times\widehat{\kappa}(\mathbf{X}_i, Y_i)\widehat{f}_{0i}/\sum_{i=1}^{N} \widehat{f}_{0i}$ and $\widehat{\mathbf{u}}_0 = (1-\widehat{E}[\epsilon t_1(\mathbf{x})t_3(\mathbf{x})\kappa(\mathbf{x}, y) \mid D = 0])^{-1} \times\widehat{E}\left[\epsilon t_1(\mathbf{x})\mathbf{t}_2(\mathbf{x})\kappa(\mathbf{x}, y) \mid D = 0\right]$.

   (c) Form $\widehat{\mathbf{u}}_1 = -(N_0/N_1)\widehat{\mathbf{u}}_0$, $\widehat{\mathbf{v}}_0 = (\widehat{\pi}_1/\widehat{b}_0)(\widehat{\mathbf{u}}_0 + \widehat{\mathbf{c}}_0)$ and $\widehat{\mathbf{v}}_1 = -(\widehat{\pi}_0/\widehat{b}_0) \times(\widehat{\mathbf{u}}_0 + \widehat{\mathbf{c}}_0)$.

   (d) Form $\widehat{\mathbf{a}}(\mathbf{x}) = \widehat{t}_1(\mathbf{x})\{\widehat{\mathbf{t}}_2(\mathbf{x}) + \widehat{t}_3(\mathbf{x})\widehat{\mathbf{u}}_0)$.

   (e) Form $\widehat{\mathbf{g}}_i = \widehat{\boldsymbol{\mu}}_{si} - \epsilon_i\widehat{\mathbf{a}}(\mathbf{X}_i)\widehat{\kappa}_i - \widehat{\mathbf{v}}_0\widehat{f}_{0i} - \widehat{\mathbf{v}}_1\widehat{f}_{1i}$.

   (f) Form $\widehat{\mathbf{v}}_{D_i} = (1 - D_i)\widehat{\mathbf{v}}_0 + D_i\mathbf{v}_1$.

   (g) Form $\widehat{\mathbf{S}}_{\mathrm{eff}}^*(D_i, \mathbf{X}_i, Y_i) = \mathbf{S}_i^* - \widehat{\mathbf{g}}_i - \widehat{\mathbf{v}}_{D_i}$ and solve the corresponding estimating equation.

A.2.2. Algorithm Using A Common Index

Specifically, we replace the steps 4-6 of Appendix A.2.1 with the following three steps.

(a) Define

$$\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I\{D_i = d\}\epsilon_i^2(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{\kappa}(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) \times K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma} - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma})}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I\{D_i = d\}K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma} - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma})};$$

$$\widehat{E}_3^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I\{D_i = d\}\epsilon_i(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) \times K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma} - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma})}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j} I\{D_i = d\}K_h(\mathbf{Z}_{\boldsymbol{\beta},i}^{\mathrm{T}}\boldsymbol{\gamma} - \mathbf{Z}_{\boldsymbol{\beta},j}^{\mathrm{T}}\boldsymbol{\gamma})}.$$

Construct $\widehat{E}_2^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_2, \boldsymbol{\theta}) = \widehat{E}_{\text{true}}\{\epsilon_j \widehat{\boldsymbol{\mu}}_s(\mathbf{X}_j, Y_j) \mid \mathbf{X}_j\}$ for $j = 1, \cdots, N$, with the method given in Appendix A.1.

(b) Estimate $\boldsymbol{\gamma}_{-1}$ by solving

$$\begin{aligned}\mathbf{0} \;=\; & \sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{j=1}^{N} I(D_j = d) \\ & \times \Big[\epsilon_j^2(\mathbf{X}_j, Y_j, \boldsymbol{\beta})\widehat{\kappa}(\mathbf{X}_j, Y_j, \boldsymbol{\alpha}) - \widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}, \boldsymbol{\theta}) \\ & \quad + \mathbf{1}_{\dim(\boldsymbol{\theta})}^{\mathrm{T}} \Big\{\epsilon_j(\mathbf{X}_j, Y_j, \boldsymbol{\beta})\widehat{\boldsymbol{\mu}}_s(\mathbf{X}_j, Y_j, \boldsymbol{\alpha}) - \widehat{E}_2^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}, \boldsymbol{\theta})\Big\} \\ & \quad + \epsilon_j(\mathbf{X}_j, Y_j, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}_j, Y_j, \boldsymbol{\alpha}) - \widehat{E}_3^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}, \boldsymbol{\theta})\Big] \\ & \times \Big\{\mathbf{Z}^*_{\boldsymbol{\beta},j} - \widehat{E}_{\text{true}}^{\widehat{\pi}}(\mathbf{Z}^*_{\boldsymbol{\beta},j} \mid \mathbf{Z}^*_{\boldsymbol{\beta},j}\boldsymbol{\gamma})\Big\},\end{aligned}$$

where

$$\begin{aligned}&\widehat{E}_{\text{true}}^{\widehat{\pi}}(\mathbf{Z}^*_{\boldsymbol{\beta},j} \mid \mathbf{Z}^*_{\boldsymbol{\beta},j}\boldsymbol{\gamma}) \\ &\quad = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j, 1\le i\le N} I(D_i = d)\mathbf{Z}^*_{\boldsymbol{\beta},i} K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\boldsymbol{\gamma} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},j}\boldsymbol{\gamma})}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\boldsymbol{\alpha})}{N_d} \sum_{i\neq j, 1\le i\le N} I(D_i = d) K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\boldsymbol{\gamma} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},j}\boldsymbol{\gamma})}.\end{aligned}$$

Denote the solution by $\widehat{\boldsymbol{\gamma}}_{-1}$ and let $\widehat{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\gamma}}_{-1}^{\mathrm{T}}, 1)^{\mathrm{T}}$.

(c) Form

$$\begin{aligned}&\widehat{E}_{\text{true}}\left\{\epsilon^2(\mathbf{X}, Y, \boldsymbol{\beta})\widehat{\kappa}(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\right\} \\ &\quad = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d}{N_d} \sum_{i=1}^{N} I\{D_i = d\}\epsilon_i^2(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{\kappa}(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\widehat{\boldsymbol{\gamma}} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{\gamma}})}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d}{N_d} \sum_{i=1}^{N} I\{D_i = d\} K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\widehat{\boldsymbol{\gamma}} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{\gamma}})}; \\ &\widehat{E}_{\text{true}}\left\{\epsilon(\mathbf{X}, Y, \boldsymbol{\beta})\widehat{\boldsymbol{\mu}}_s(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\right\} \\ &\quad = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d}{N_d} \sum_{i=1}^{N} I\{D_i = d\}\epsilon_i(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{\boldsymbol{\mu}}_s(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\widehat{\boldsymbol{\gamma}} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{\gamma}})}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d}{N_d} \sum_{i=1}^{N} I\{D_i = d\} K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\widehat{\boldsymbol{\gamma}} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{\gamma}})}; \\ &\widehat{E}_{\text{true}}\left\{\epsilon(\mathbf{X}, Y, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\right\} \\ &\quad = \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d}{N_d} \sum_{i=1}^{N} I\{D_i = d\}\epsilon_i(\mathbf{X}_i, Y_i, \boldsymbol{\beta})\widehat{f}_0(\mathbf{X}_i, Y_i, \boldsymbol{\alpha}) K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\widehat{\boldsymbol{\gamma}} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{\gamma}})}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d}{N_d} \sum_{i=1}^{N} I\{D_i = d\} K_h(\mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta},i}\widehat{\boldsymbol{\gamma}} - \mathbf{Z}^{\mathrm{T}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{\gamma}})}.\end{aligned}$$

A.3. Regularity Conditions

Let ℓ be the dimensionality of the kernel regressions in our method after dimension reduction. In our simulations and example, we took $\ell = 1$. The set of regularity conditions required by Theorem 1 is listed below.

C1. The univariate kernel function is a function that integrates to 1 and has support $(-1, 1)$ and order r, i.e., $\int K(u)u^t du = 0$ if $1 \le t < r$ and $\int K(u)u^r du \neq 0$. The ℓ-dimensional kernel function, still represented with K, is a product of ℓ univariate kernel functions, that is, $K(\mathbf{u}) = \prod_{i=1}^{\ell} K(u_i)$ for a ℓ-dimensional $\mathbf{u}$.

C2. Let $\xi_{i,\widetilde{\boldsymbol{\beta}}}^{\text{true}}$ be the true population density of $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_i$ for $i = 1, 2, 3$ and $\widetilde{\boldsymbol{\beta}}$ in a local neighborhood of $\boldsymbol{\beta}$. Assume that $\xi_{i,\widetilde{\boldsymbol{\beta}}}^{\text{true}}$'s are bounded away from 0 and they all have third order bounded and continuous derivatives.

C3. At any fixed $\widetilde{\boldsymbol{\alpha}}$ in a local neighborhood of $\boldsymbol{\alpha}$, $\zeta_1(\cdot, \widetilde{\boldsymbol{\alpha}})$, $\boldsymbol{\zeta}_2(\cdot, \widetilde{\boldsymbol{\alpha}})$ and $\zeta_3(\cdot, \widetilde{\boldsymbol{\alpha}})$ are functions of $\cdot$ with second order bounded and continuous derivatives.

C4. $E_{\text{true}}\{\epsilon^4(\mathbf{X}, Y, \widetilde{\boldsymbol{\beta}})\kappa^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})|\mathbf{X}\}$, $E_{\text{true}}\{\epsilon^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\beta}})\boldsymbol{\mu}_s(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})^{\otimes 2}|\mathbf{X}\}$ and $E_{\text{true}}\{\epsilon^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\beta}})f_0^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})|\mathbf{X}\}$ are bounded for any $\widetilde{\boldsymbol{\theta}}$ in a local neighborhood of $\boldsymbol{\theta}$.

C5. $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^* \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^* \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^* \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2\kappa(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}}) \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon\boldsymbol{\mu}_s(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}}) \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2\right\}$ and $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon f_0(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}}) \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3\right\}$ have $(r+1)$th order bounded and continuous derivatives for any $\widetilde{\boldsymbol{\theta}}$ in a local neighborhood of $\boldsymbol{\theta}$.

C6. $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2\kappa(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}$, $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon\boldsymbol{\mu}_s(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2\right\}$, and $E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon f_0(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3\right\}$ have $(r+1)$th order bounded and continuous derivatives for any $\widetilde{\boldsymbol{\theta}}$ in a local neighborhood of $\boldsymbol{\theta}$.

C7. $E_{\text{true}}\left[\epsilon^4(\mathbf{X}, Y, \widetilde{\boldsymbol{\beta}})\kappa^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}^{\mathrm{T}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1, \mathbf{X}\right]$, $E_{\text{true}}\left[\epsilon^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\beta}})\boldsymbol{\mu}_s(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}^{\mathrm{T}}\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}\boldsymbol{\mu}_s(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})^{\mathrm{T}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_2, \mathbf{X}\right]$, $E_{\text{true}}\left[\epsilon^2 f_0^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}})\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}^{\mathrm{T}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_3, \mathbf{X}\right]$, and $E_{\text{true}}\left[\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}\{\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}} - \mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}\}^{\mathrm{T}} \mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_i, \mathbf{X}\right]$, for $i = 1, 2, 3$, all have bounded entries for any $\widetilde{\boldsymbol{\theta}}$ in a local neighborhood of $\boldsymbol{\theta}$, where $\mathbf{Z}'_{\widetilde{\boldsymbol{\beta}}}$ is an independent and identically distributed copy of $\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}$.

C8. $\pi_d(\widetilde{\boldsymbol{\alpha}})/(N_d/N)$ are bounded for $d = 0, 1$.

C9. $\pi_d(\widetilde{\boldsymbol{\alpha}})/\pi_d(\boldsymbol{\alpha})$ are bounded for $d = 0, 1$.

C10. The bandwidth $h = N^{-\tau}$ where $1/(2\ell) > \tau > 1/(4r)$. This includes the optimal bandwidth $h = O(N^{-1/(2r+\ell)})$ as long as we choose a kernel of order $2r > \ell$.

C11. There exists a positive constant C such that $\lim_{N\to\infty} N_0/N_1 = C < \infty$.

Conditions C1 and C10 are standard requirements on an rth order kernel function and on the bandwidth in the kernel smoothing literature (Ma and Zhu, 2013c).

A.4. Proof of Proposition 1

We provide a detailed proof that the first dimension reduction model (3.1) satisfies Proposition 1. Proving that the other two dimension reduction models (3.2) and (3.3) also satisfy Proposition 1 is similar.

In (3.1), $\kappa(\mathbf{x}, y, \boldsymbol{\alpha})$ is a function of the weighted sum of $H(d, \mathbf{x}, y)$ with $d = 0, 1$. As a result,

$$\begin{aligned}\kappa(\mathbf{x}, y, \boldsymbol{\alpha}) &= h\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}_1 + m(\mathbf{x}, \boldsymbol{\beta})\alpha_2 + \epsilon\alpha_2\} \\ &= h[\{\mathbf{x}^{\mathrm{T}}, m(\mathbf{x}, \boldsymbol{\beta})\}(\boldsymbol{\alpha}_1^{\mathrm{T}}, \alpha_2)^{\mathrm{T}} + \epsilon\alpha_2],\end{aligned}$$

where $h(\cdot)$ is a differentiable function.

For $\epsilon = Q(\mathbf{X}^{\mathrm{T}}\omega, \epsilon^*)$, where $Q(\cdot)$ is an arbitrary function,

$$\begin{aligned}&E\{\epsilon^2\kappa(\mathbf{X}, Y, \boldsymbol{\alpha}) \mid \mathbf{X}\} \\ &\quad= E(\epsilon^2 h[\{\mathbf{X}^{\mathrm{T}}, m(\mathbf{X}, \boldsymbol{\beta})\}(\boldsymbol{\alpha}_1^{\mathrm{T}}, \alpha_2)^{\mathrm{T}} + \epsilon\alpha_2] \mid \mathbf{X}) \\ &\quad= E\{Q(\mathbf{X}^{\mathrm{T}}\omega, \epsilon^*)^2 h[\{\mathbf{X}^{\mathrm{T}}, m(\mathbf{X}, \boldsymbol{\beta})\}(\boldsymbol{\alpha}_1^{\mathrm{T}}, \alpha_2)^{\mathrm{T}} + Q(\mathbf{X}^{\mathrm{T}}\omega, \epsilon^*)\alpha_2] \mid \mathbf{X}\} \\ &\quad= \zeta_1(\mathbf{X}^{\mathrm{T}}\omega, \{\mathbf{X}^{\mathrm{T}}, m(\mathbf{X}, \boldsymbol{\beta})\}(\boldsymbol{\alpha}_1^{\mathrm{T}}, \alpha_2)^{\mathrm{T}}) \\ &\quad= \zeta_1(\mathbf{Z}_{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\gamma}_1),\end{aligned}$$

where $\zeta_1(\cdot)$ is a smooth function, $\mathbf{Z}_{\boldsymbol{\beta}} = \mathbf{X}$ and $\boldsymbol{\gamma}_1 = (\omega, \boldsymbol{\alpha}_1 + \alpha_2\boldsymbol{\beta})$ is a $p \times 2$ matrix if $m(\cdot)$ is linear in $\mathbf{X}$; otherwise, $\mathbf{Z}_{\boldsymbol{\beta}} = \{\mathbf{X}^{\mathrm{T}}, m(\mathbf{X}, \boldsymbol{\beta})\}^{\mathrm{T}}$ and $\boldsymbol{\gamma}_1 = \{(\omega^{\mathrm{T}}, 0)^{\mathrm{T}}, (\boldsymbol{\alpha}_1^{\mathrm{T}}, \alpha_2)^{\mathrm{T}}\}$ is a $(p+1) \times 2$ matrix.

A.5. Background and Technical Results

A.5.1. Introduction

Following Ma and Carroll (2016), we divide the N observations randomly into three sets, where the first set contains $n_1 = N - N^{1-\delta} - N^{1-2\delta}$ observations, the second set contains $n_2 = N^{1-\delta}$ observations and the third set contains $n_3 = N^{1-2\delta}$ observations, where δ is a small positive number. For convenience of proof, we require the disease proportion in the third data set to be the same as the whole data set. That is, $n_{30}/n_{31} = N_0/N_1$, where n_{30} and n_{31} are the numbers of controls and cases in the third set of data, respectively. We form and solve the estimating equation (2.5) using data in the first set while calculating all the estimated quantities described in Appendix A.2 steps 1-3 using data in the second set and the other estimated quantities defined in Appendix A.2 steps 4-6 using the data in the third set.

A.5.2. Lemmas

Before proving Theorem 1, we first state several lemmas, which ensure the quantities defined in Appendix A.2 steps 4-6 have the desired orders of bias and mean square error, i.e., the same as that of the usual nonparametric estimators.

From (3.5), we can easily show that

Lemma 1. *For some* $\sigma^2_{\pi_d(\widetilde{\boldsymbol{\alpha}})} < \infty$, $\sqrt{n_2}\{\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}}) - \pi_d(\widetilde{\boldsymbol{\alpha}})\} \xrightarrow{d} N(0, \sigma^2_{\pi_d(\widetilde{\boldsymbol{\alpha}})})$, *as* $N \to \infty$.

We now analyze the property of our estimators defined in Appendix A.2 steps 4-6. For notational brevity, we only focus on the first conditional expectation $E_{\text{true}}\{\epsilon^2\kappa(\mathbf{X}, Y)|\mathbf{X}\}$. The other two conditional expectations have similar properties. We split the analysis into three parts: i) analyze the properties of $\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}})$; ii) analyze the properties of $\widehat{\boldsymbol{\gamma}}_1(\widetilde{\boldsymbol{\theta}})$ for $\widetilde{\boldsymbol{\theta}}$ near $\boldsymbol{\theta}$; iii) show that $\widehat{E}_{\text{true}}\{\epsilon^2(\mathbf{X}, Y, \widetilde{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X}, Y, \widetilde{\boldsymbol{\alpha}}) \mid \mathbf{X}\}$ has desired bias order and standard deviation order.

For the first part of the analysis, we establish the following lemma.

Lemma 2. *Under the regularity conditions C1-C10,*

$$\begin{aligned}\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) &= \widehat{E}_1(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) + O_p(n_2^{-1/2}) \\ &= E_1(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) + O_p(h^r) + O_p\left(n_3^{-1/2}h^{-\ell/2}\right),\end{aligned}$$

where

$$\begin{aligned}&\widehat{E}_1(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) \\ &= \frac{\sum_{d=0}^{1} \frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3} \sum_{i\neq j, 1\le i\le n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_i, Y_i, \widetilde{\boldsymbol{\alpha}}) \times K_h(\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},i}\boldsymbol{\gamma}_1 - \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)}{\sum_{d=0}^{1} \frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3} \sum_{i\neq j, 1\le i\le n_3} I(D_i = d)K_h(\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},i}\boldsymbol{\gamma}_1 - \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)}; \\ &E_1(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) \\ &= \frac{E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\epsilon_j^2(\mathbf{X}_j, Y_j, \widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j, Y_j, \widetilde{\boldsymbol{\alpha}})|\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1\right\}}{E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1\right\}}.\end{aligned}$$

Proof. Denote the numerator and denominator of $\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}})$ by q_{num} and q_{den} respectively. We can replace $\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}})$ in q_{num} and q_{den} with $\pi_d(\widetilde{\boldsymbol{\alpha}})$ without changing the error order due to the data partition scheme we use. That is,

$$\begin{aligned}&q_{\text{num}} \\ &= (n_3 - 1)^{-1}\sum_{d=0}^{1}\frac{\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j, 1\le i\le n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widetilde{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X}_i, Y_i, \widetilde{\boldsymbol{\alpha}})\end{aligned}$$

$$
\begin{aligned}
&\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&= (n_3-1)^{-1}\sum_{d=0}^{1}\frac{\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\{\widehat{\kappa}(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})\\
&\quad -\kappa(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})\}\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&\quad +(n_3-1)^{-1}\sum_{d=0}^{1}\frac{\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}})-\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\\
&\quad \times\kappa(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&\quad +(n_3-1)^{-1}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})\\
&\quad \times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1).
\end{aligned}
$$

With further calculations, this means that

$$
\begin{aligned}
&q_{\mathrm{num}}\\
&= O_p(n_2^{-1/2})(n_3-1)^{-1}\sum_{d=0}^{1}\frac{\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\\
&\quad \times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&\quad +O_p(n_2^{-1/2})(n_3-1)^{-1}\sum_{d=0}^{1}\frac{1}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\\
&\quad \times\kappa(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&\quad +(n_3-1)^{-1}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})\\
&\quad \times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&= (n_3-1)^{-1}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_i,Y_i,\widetilde{\boldsymbol{\alpha}})\\
&\quad \times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1) + O_p(n_2^{-1/2}).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
q_{\mathrm{den}} \quad = \quad & (n_3-1)^{-1}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=d)K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
& +O_p(n_2^{-1/2}).
\end{aligned}
$$

We now analyze the conditional expectations of q_{num} and q_{den} given $\mathbf{X}_j$ one by one. First,

$$
\begin{aligned}
&E(q_{\text{num}}|\mathbf{X}_j)\\
&\quad=\sum_{d=0}^{1}\pi_d(\widetilde{\boldsymbol{\alpha}})E\left\{\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid D=d,\mathbf{X}_j\right\}\\
&\qquad+O_p(n_2^{-1/2})\\
&\quad=E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid\mathbf{X}_j\right\}\\
&\qquad+O_p(n_2^{-1/2})\\
&\quad=E_{\text{true}}\left[E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\mid\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1,\mathbf{X}_j\right\}\right.\\
&\qquad\left.\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid\mathbf{X}_j\right]+O_p(n_2^{-1/2})\\
&\quad=E_{\text{true}}\left[E_{\text{true}}\left\{\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\mid\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\right.\\
&\qquad\left.\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid\mathbf{X}_j\right]+O_p(n_2^{-1/2})\\
&\quad=E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}})\mid\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\xi_1^{\text{true}}(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\\
&\qquad+O_p(h^r)+O_p(n_2^{-1/2}).
\end{aligned}
$$

Here we used the regularity conditions C1-C2, C5, C8-C10.

In addition, with the regularity conditions C1-C4 and C8-C10, we have

$$
\begin{aligned}
&\text{var}\,(q_{\text{num}}\mid\mathbf{X}_j)\\
&\quad=(n_3-1)^{-1}\text{var}\left\{\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}I(D=d)\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\right.\\
&\qquad\left.\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid\mathbf{X}_j\right\}+O_p(n_2^{-1})\\
&\quad=(n_3-1)^{-1}\left(E\left[\left\{\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}I(D=d)\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\right.\right.\right.\\
&\qquad\left.\left.\left.\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\right\}^2\mid\mathbf{X}_j\right]\right.\\
&\qquad\left.-E\left\{\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}I(D=d)\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\right.\right.\\
&\qquad\left.\left.\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid\mathbf{X}_j\right\}^2\right)+O_p(n_2^{-1})\\
&\quad=(n_3-1)^{-1}\left(E\left[\left\{\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}I(D=d)\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\right.\right.\right.\\
&\qquad\left.\left.\left.\times K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\right\}^2\mid\mathbf{X}_j\right]\right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad -E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}})\mid \mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}^2\\
&\qquad \times\xi_1^{\text{true}}(\mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\gamma}_1)^2\big)\\
&\quad +O_p\{(n_3-1)^{-1}h^r\}+O_p(n_2^{-1})\\
&=O_p\left(n_3^{-1}h^{-\ell}\right).
\end{aligned}
$$

The last equality is because

$$
\begin{aligned}
&E\left[\left\{\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}I(D=d)\epsilon^2(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})K_h(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\right\}^2\mid \mathbf{X}_j\right]\\
&\le 2E\left[\sum_{d=0}^{1}\frac{\pi_d^2(\widetilde{\boldsymbol{\alpha}})}{n_{3d}^2/n_3^2}I(D=d)\epsilon^4(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa^2(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})K_h^2(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid \mathbf{X}_j\right]\\
&=2E_{\text{true}}\left[\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{n_{3D}/n_3}\frac{\pi_D(\widetilde{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^4(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa^2(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})K_h^2(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid \mathbf{X}_j\right]\\
&\le CE_{\text{true}}\left\{\epsilon^4(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa^2(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})K_h^2(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid \mathbf{X}_j\right\}\\
&=CE_{\text{true}}\left[E_{\text{true}}\left\{\epsilon^4(\mathbf{X},Y,\widetilde{\boldsymbol{\beta}})\kappa^2(\mathbf{X},Y,\widetilde{\boldsymbol{\alpha}})\mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}K_h^2(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid \mathbf{X}_j\right]\\
&\le C'E_{\text{true}}\left\{K_h^2(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)\mid \mathbf{X}_j\right\}\\
&=O_p(h^{-\ell}),
\end{aligned}
$$

where C, C' are constants.

Similarly, we have that

$$
\begin{aligned}
E(q_{\text{den}}|\mathbf{X}_j) &= E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\xi_1^{\text{true}}(\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1)+O_p(h^r)\\
&\quad +O_p(n_2^{-1/2});\\
\operatorname{var}(q_{\text{den}}|\mathbf{X}_j) &= O_p\left(n_3^{-1}h^{-\ell}\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}}) &= \frac{E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}})|\mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}}{E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\mid \mathbf{Z}_{\widetilde{\boldsymbol{\beta}},j}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}}\\
&\quad +O_p(h^r)+O_p\left(n_3^{-1/2}h^{-\ell/2}\right).
\end{aligned}
$$

When $\widetilde{\boldsymbol{\theta}}=\boldsymbol{\theta}$, we have $\widehat{E}_1^{\widehat{\pi}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta}) = E_{\text{true}}\{\epsilon_j^2\kappa(\mathbf{X}_j,Y_j)|\mathbf{X}_j\}+O_p(h^r)+O_p\left(n_3^{-1/2}h^{-\ell/2}\right)$. □

Lemma 3. *Under the regularity conditions C1-C10,*

$$\begin{aligned}\widehat{E}^{\widehat{\pi}}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1) &= \widehat{E}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1) + O_p(n_2^{-1/2}) \\ &= E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) + O_p(h^r) + O_p\left(n_3^{-1/2}h^{-\ell/2}\right),\end{aligned}$$

where

$$\begin{aligned}&\widehat{E}_{\text{true}}\left(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1\right) \\ &= \frac{n_3^{-1}\sum_{r=0}^{1}\frac{\pi_r(\widetilde{\boldsymbol{\alpha}})}{n_{3r}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=r)\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},i}K_h(\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},i}\boldsymbol{\gamma}_1 - \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)}{n_3^{-1}\sum_{r=0}^{1}\frac{\pi_r(\widetilde{\boldsymbol{\alpha}})}{n_{3r}/n_3}\sum_{i\neq j,1\le i\le n_3} I(D_i=r)K_h(\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},i}\boldsymbol{\gamma}_1 - \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)};\end{aligned}$$

$$\begin{aligned}&E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}}) \\ &= \frac{E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})}\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}|\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1\right\}}{E_{\text{true}}\left\{\frac{\pi_{D_j}(\widetilde{\boldsymbol{\alpha}})}{\pi_{D_j}(\boldsymbol{\alpha})} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1\right\}}.\end{aligned}$$

We skip the proof of the Lemma 3 here since it is similar to the proof of Lemma 2. Next, we establish the root-n_3 consistency of $\widehat{\boldsymbol{\gamma}}_{j,-1}$ for $j = 1, \cdots, 3$.

Lemma 4. *Under the regularity conditions C1-C10,*

$$\begin{aligned}&\sqrt{n_3}\{\widehat{\boldsymbol{\gamma}}_{1,-1}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\gamma}_{1,-1}\} \to \mathit{Normal}(0, \Sigma_{\boldsymbol{\gamma}_{1,-1}}), \\ &\sqrt{n_3}\{\widehat{\boldsymbol{\gamma}}_{2,-1}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\gamma}_{2,-1}\} \to \mathit{Normal}(0, \Sigma_{\boldsymbol{\gamma}_{2,-1}}), \\ &\sqrt{n_3}\{\widehat{\boldsymbol{\gamma}}_{3,-1}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\gamma}_{3,-1}\} \to \mathit{Normal}(0, \Sigma_{\boldsymbol{\gamma}_{3,-1}}),\end{aligned}$$

when $N \to \infty$. Here $\Sigma_{\boldsymbol{\gamma}_{1,-1}}$, $\Sigma_{\boldsymbol{\gamma}_{2,-1}}$ and $\Sigma_{\boldsymbol{\gamma}_{3,-1}}$ are positive definite matrices.

Proof. Here we only provide the proof of the root-n_3 consistency of $\widehat{\boldsymbol{\gamma}}_{1,-1}$ below. Similar derivations can be used to prove the results regarding $\boldsymbol{\gamma}_{2,-1}$ and $\boldsymbol{\gamma}_{3,-1}$. The estimator $\widehat{\boldsymbol{\gamma}}_{1,-1}$ solves

$$\begin{aligned}\mathbf{0} &= n_3^{-1/2}\sum_{d=0}^{1}\frac{\widehat{\pi}_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\left\{\epsilon_j^2(\mathbf{X}_j, Y_j, \widetilde{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X}_j, Y_j, \widetilde{\boldsymbol{\alpha}})\right. \\ &\qquad \left. -\widehat{E}^{\widehat{\pi}}_1(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}})\right\}\left\{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} - \widehat{E}^{\widehat{\pi}}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)\right\} \\ &= n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\left\{\epsilon_j^2(\mathbf{X}_j, Y_j, \widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j, Y_j, \widetilde{\boldsymbol{\alpha}})\right. \\ &\qquad \left. -\widehat{E}_1(\mathbf{X}_j, \boldsymbol{\gamma}_1, \widetilde{\boldsymbol{\theta}})\right\}\left\{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} - \widehat{E}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} \mid \mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)\right\} \\ &\quad +O_p\{(n_3/n_2)^{1/2}\},\end{aligned}$$

where we used Lemma 2 and 3. Simple calculation shows that the above equation can be further expanded as

$$
\begin{aligned}
0 &= n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\Big\{\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}}) \\
&\quad -E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})+E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})-\widehat{E}_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\}\Big\{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j} \\
&\quad -E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})+E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})-\widehat{E}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}\mid\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)\Big\} \\
&\quad +o_p(1) \\
&= n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\Big\{\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}}) \\
&\quad -E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\}\Big\{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}-E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\} \\
&\quad +n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d) \\
&\quad \times\Big\{\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}})-E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\} \\
&\quad \times\Big\{E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})-\widehat{E}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}\mid\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)\Big\} \\
&\quad +n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d) \qquad \text{(A.7)} \\
&\quad \times\Big\{E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})-\widehat{E}_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\}\Big\{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}-E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\} \\
&\quad +n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\Big\{E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}}) \\
&\quad -\widehat{E}_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\}\Big\{E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})-\widehat{E}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}\mid\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)\Big\} \\
&\quad +o_p(1).
\end{aligned}
$$

Using Lemmas 2 and 3 and the regularity condition C10, we have that the fourth term in (A.7)

$$
\begin{aligned}
&\Bigg\| n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\Big\{E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}}) \\
&\qquad -\widehat{E}_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\Big\}\Big\{E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})-\widehat{E}_{\text{true}}(\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}\mid\mathbf{Z}^{\mathrm{T}}_{\widetilde{\boldsymbol{\beta}},j}\boldsymbol{\gamma}_1)\Big\}\Bigg\| \\
&= \Big| n_3^{1/2}\Big\{O_p(h^r)+O_p\Big(n_3^{-1/2}h^{-\ell/2}\Big)\Big\}^2\Big| = o_p(1).
\end{aligned}
$$

By applying Lemma A1 in Ma and Zhu (2012a), we obtain that the second and third terms in (A.7) are of order $O_p(h^r+n_3^{1/2}h^{2r}+\log^2 n_3/\sqrt{n_3h^{2\ell}})=o_p(1)$.

Hence, the estimating equation can be written as

$$\begin{aligned}\mathbf{0} &= n_3^{-1/2}\sum_{d=0}^{1}\frac{\pi_d(\widetilde{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{j=1}^{n_3} I(D_j=d)\left\{\epsilon_j^2(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\beta}})\kappa(\mathbf{X}_j,Y_j,\widetilde{\boldsymbol{\alpha}})\right.\\ &\quad\left.-E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\right\}\left\{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}},j}-E_{\mathbf{Z}^*_{\widetilde{\boldsymbol{\beta}}}}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\widetilde{\boldsymbol{\theta}})\right\}\\ &\quad+o_p(1). \qquad\text{(A.8)}\end{aligned}$$

We now show that the influence function given in (A.8) has mean 0 at $\widetilde{\boldsymbol{\theta}}=\boldsymbol{\theta}$.

$$\begin{aligned}&E\left[\sum_{d=0}^{1}\frac{\pi_d(\boldsymbol{\alpha})}{n_{3d}/n_3}I(D_j=d)\left\{\epsilon_j^2(\mathbf{X}_j,Y_j,\boldsymbol{\beta})\kappa(\mathbf{X}_j,Y_j,\boldsymbol{\alpha})\right.\right.\\ &\qquad\left.\left.-E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta})\right\}\left\{\mathbf{Z}_j^*-E_{\mathbf{Z}^*}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta})\right\}\right]\\ &=E_{\text{true}}\left[\left\{\epsilon_j^2(\mathbf{X}_j,Y_j,\boldsymbol{\beta})\kappa(\mathbf{X}_j,Y_j,\boldsymbol{\alpha})-E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta})\right\}\left\{\mathbf{Z}_j^*\right.\right.\\ &\qquad\left.\left.-E_{\mathbf{Z}^*}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta})\right\}\right]\\ &=E_{\text{true}}\left(E_{\text{true}}\left[\left\{\epsilon_j^2(\mathbf{X}_j,Y_j,\boldsymbol{\beta})\kappa(\mathbf{X}_j,Y_j,\boldsymbol{\alpha})-E_1(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta})\right\}\mid\mathbf{X}_j\right]\right.\\ &\qquad\left.\times\left\{\mathbf{Z}_j^*-E_{\mathbf{Z}^*}(\mathbf{X}_j,\boldsymbol{\gamma}_1,\boldsymbol{\theta})\right\}\right)\\ &=\mathbf{0}.\end{aligned}$$

The last equality is because of the single index model assumption (3.1). In practical operation, we will replace $\widetilde{\boldsymbol{\theta}}$ by $\widehat{\boldsymbol{\theta}}$, the solution of the estimating equation defined in (4.1). As long as $\widehat{\boldsymbol{\theta}}\to\boldsymbol{\theta}$ in probability, the above expectation approaches $\mathbf{0}$.

Hence, we have that

$$\sqrt{n_3}\{\widehat{\boldsymbol{\gamma}}_{1,-1}(\widehat{\boldsymbol{\theta}})-\boldsymbol{\gamma}_{1,-1}\}\to\text{Normal}(0,\Sigma_{\boldsymbol{\gamma}_{1,-1}})$$

when $N\to\infty$, where $\Sigma_{\boldsymbol{\gamma}_{1,-1}}$ is a positive definite matrix. □

We now analyze $\widehat{E}_{\text{true}}\{\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{X}\}$. We will show that it has bias order $O_p(h^r)$ and standard deviation $O_p\left(n_3^{-1/2}h^{-\ell/2}\right)$ as given in the following lemma.

Lemma 5. *Under the regularity conditions C1-C10,*

$$\begin{aligned}&\widehat{E}_{\text{true}}\{\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{X}\}\\ &\quad=\frac{E_{\text{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{Z}^{\mathrm{T}}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{\gamma}_1\right\}}{E_{\text{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mid\mathbf{Z}^{\mathrm{T}}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{\gamma}_1\right\}}+O_p(h^r)\\ &\qquad+O_p\left(n_3^{-1/2}h^{-l/2}\right)+O_p\left(n_3^{-1}h^{-\ell/2-1}\right).\end{aligned}$$

Proof. Similar to the proof of Lemma 2, we have that

$$\widehat{E}_{\text{true}}\{\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{X}\}$$

$$
\begin{aligned}
&\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\alpha}}) \\
&\qquad \times K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\} \\
=\ & \frac{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\}}{\sum_{d=0}^{1} \frac{\widehat{\pi}_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\}} \\
&\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\beta}})\kappa(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\alpha}}) \\
&\qquad \times K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\} \\
=\ & \frac{\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\}}{\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\}} \\
&+O_p(n_2^{-1/2}).
\end{aligned}
$$

We first inspect the numerator.

$$
\begin{aligned}
&n_3^{-1}\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}} \sum_{i=1}^{n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\beta}})\kappa(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\alpha}}) \\
&\qquad \times K_h\left\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\right\} \\
&= n_3^{-1}h^{-(\ell+1)}\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}/n_3} \sum_{i=1}^{n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\beta}})\kappa(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\alpha}}) \\
&\qquad \times K'\left\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1^*/h - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1^*/h\right\}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i} - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}})^{\mathrm{T}}\{\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\gamma}_1\} \\
&\quad +n_3^{-1}\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}/n_3} \sum_{i=1}^{n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\beta}})\kappa(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\alpha}}) \\
&\qquad \times K_h\left(\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right) \\
&= n_3^{-1}\sum_{d=0}^{1} \frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}/n_3} \sum_{i=1}^{n_3} I(D_i = d)\epsilon_i^2(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\beta}})\kappa(\mathbf{X}_i, Y_i, \widehat{\boldsymbol{\alpha}}) \\
&\qquad \times K_h\left(\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1 - \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right) + O_p(n_3^{-1/2}) + O_p(n_3^{-1}h^{-\ell/2-1}) \\
&= E_{\mathrm{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X}, Y, \widehat{\boldsymbol{\beta}})\kappa(\mathbf{X}, Y, \widehat{\boldsymbol{\alpha}}) \mid \mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\xi_1^{\mathrm{true}}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1) + O_p(h^r) \\
&\quad +O_p\left(n_3^{-1/2}h^{-l/2}\right) + O_p(n_3^{-1}h^{-\ell/2-1}).
\end{aligned}
$$

Here $\boldsymbol{\gamma}_1^*$ is on the interval connecting $\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{\gamma}_1$. In the second equality above, we used condition C10, the root-n_3 consistency of $\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})$, the regularity conditions C5-C7 and the fact that

$$
\begin{aligned}
& n_3^{-1}h^{-(\ell+1)}\sum_{d=0}^{1}\frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}/n_3}\sum_{i=1}^{n_3} I(D_i=d)\epsilon_i^2(\mathbf{X}_i,Y_i,\widehat{\boldsymbol{\beta}}) \\
& \qquad \times\kappa(\mathbf{X}_i,Y_i,\widehat{\boldsymbol{\alpha}})K'\{h^{-1}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\boldsymbol{\gamma}_1-\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1)\}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}-\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}) \\
& = -\frac{\partial}{\partial\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1}\left[E_{\mathrm{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}\mid\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\xi_1^{\mathrm{true}}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1)\right] \\
& \quad +\frac{\partial}{\partial\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1}\left[E_{\mathrm{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\xi_1^{\mathrm{true}}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1)\right]\mathbf{Z}_{\widehat{\boldsymbol{\beta}}} \\
& \quad +O_p(h^2)+O_p\{(n_3h^{\ell+2})^{-1/2}\}.
\end{aligned}
$$

Similarly, for the denominator, we have that

$$
\begin{aligned}
& \sum_{d=0}^{1}\frac{\pi_d(\widehat{\boldsymbol{\alpha}})}{n_{3d}}\sum_{i=1}^{n_3} I(D_i=d)K_h\{\mathbf{Z}_{\widehat{\boldsymbol{\beta}},i}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})-\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_1(\widehat{\boldsymbol{\theta}})\} \\
& \qquad = E_{\mathrm{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mid\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}\xi_1^{\mathrm{true}}(\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1)+O_p(h^r)+O_p\left(n_3^{-1/2}h^{-l/2}\right) \\
& \qquad\quad +O_p\left(n_3^{-1}h^{-\ell/2-1}\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \widehat{E}_{\mathrm{true}}(\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\widehat{\kappa}(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{X}) \\
& \qquad = \frac{E_{\mathrm{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\epsilon^2(\mathbf{X},Y,\widehat{\boldsymbol{\beta}})\kappa(\mathbf{X},Y,\widehat{\boldsymbol{\alpha}})\mid\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}}{E_{\mathrm{true}}\left\{\frac{\pi_D(\widehat{\boldsymbol{\alpha}})}{\pi_D(\boldsymbol{\alpha})}\mid\mathbf{Z}_{\widehat{\boldsymbol{\beta}}}^{\mathrm{T}}\boldsymbol{\gamma}_1\right\}}+O_p(h^r) \\
& \qquad\quad +O_p\left(n_3^{-1/2}h^{-l/2}\right)+O_p\left(n_3^{-1}h^{-\ell/2-1}\right).
\end{aligned}
$$

□

A.6. Proof of Theorem 1

Through the analyses in the lemmas, we proved that all the estimated quantities defined in Appendix A.2.1 have desired bias order and standard deviation orders. Specifically, the difference between the quantities with hat and without hat either have mean zero, standard deviation $O_p(n_2^{-1/2})=O_p(n_3^{-1/2})$ or have bias $O_p(h^r)$ and standard deviation $O_p\left(n_3^{-1/2}h^{-\ell/2}\right)$ or $O_p\left(n_3^{-1/2}h^{-\ell/2}\right)+O_p\left(n_3^{-1}h^{-(\ell+2)/2}\right)$. Now we are ready to prove our main theorem.

$$
\begin{aligned}
\mathbf{0} &= n_1^{-1/2}\sum_{i=1}^{n_1}\widehat{\mathbf{S}}_{\mathrm{eff}}^{*}\left(D_i,\mathbf{X}_i,Y_i,\widehat{\boldsymbol{\theta}}\right) \\
&= n_1^{-1/2}\sum_{i=1}^{n_1}\mathbf{S}_{\mathrm{eff}}^{*}\left[D_i,\mathbf{X}_i,Y_i,\widehat{\boldsymbol{\theta}},\widehat{\pi}_d(\widehat{\boldsymbol{\alpha}}),\widehat{E}\{\widehat{\pi}_d(\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\theta}})\}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= n_1^{-1/2}\sum_{i=1}^{n_1}\mathbf{S}^*_{\text{eff}}\left[D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta},\widehat{\pi}_d(\boldsymbol{\alpha}),\widehat{E}\{\widehat{\pi}_d(\widehat{\boldsymbol{\alpha}}),\widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta})\}\right]\\
&\quad +n_1^{-1}\sum_{i=1}^{n_1}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{S}^*_{\text{eff}}\left[D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta}^*,\widehat{\pi}_d(\boldsymbol{\alpha}^*),\widehat{E}\{\widehat{\pi}_d(\boldsymbol{\alpha}^*),\widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta}^*)\}\right]\sqrt{n_1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\
&= n_1^{-1/2}\sum_{i=1}^{n_1}\mathbf{S}^*_{\text{eff}}(D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta})+n_1^{-1/2}\sum_{i=1}^{n_1}\left\{\widehat{\mathbf{S}}^*_{\text{eff}}(D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta})\right.\\
&\qquad \left.-\mathbf{S}^*_{\text{eff}}(D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta})\right\}\\
&\quad +E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{S}^*_{\text{eff}}(D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta})+o_p(1)\right\}\sqrt{n_1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\
&= n_1^{-1/2}\sum_{i=1}^{n_1}\mathbf{S}^*_{\text{eff}}(D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta})+E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{S}^*_{\text{eff}}(D_i,\mathbf{X}_i,Y_i,\boldsymbol{\theta})\right.\\
&\quad \left.+o_p(1)\right\}\sqrt{n_1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})+o_p(1),
\end{aligned}
$$

where $\boldsymbol{\alpha}^*$ is a point on the line connecting $\boldsymbol{\alpha}$ and $\widehat{\boldsymbol{\alpha}}$. Simple calculation lead to the proof of Theorem 1. □

A.7. Simulations with 10% Disease Rate

Ctrl	$\boldsymbol{\beta}_1$	Homoscedastic Gaussian error				Heteroscedastic Gaussian error			
		0.5	1.0	0.3	0.5	0.5	1.0	0.3	0.5
Param	mean	0.534	0.958	0.263	0.458	0.591	0.890	0.225	0.386
	s.d.	0.115	0.116	0.117	0.115	0.098	0.091	0.106	0.098
Semi	mean	0.543	0.945	0.251	0.432	0.353	1.167	0.423	0.679
	s.d.	0.086	0.082	0.084	0.091	0.086	0.084	0.083	0.089
	MSE Eff	1.557	1.557	1.565	1.157	1.937	2.019	1.798	1.794
	mean	0.504	0.992	0.297	0.496	0.517	0.983	0.285	0.482
	s.d	0.098	0.082	0.078	0.087	0.096	0.101	0.092	0.105
	MSE Eff	1.497	2.247	2.457	2.001	1.877	1.931	1.956	1.997

TABLE 4

500 simulations, 1000 cases/1000 controls,10% disease rate, correlated covariates $\mathbf{X}$ *with dimension 4, Gaussian error. See Table 1 for definitions.*

Ctrl	$\boldsymbol{\beta}_1$	Homoscedastic Gamma error				Heteroscedastic Gamma error			
		0.5	1.0	0.3	0.5	0.5	1.0	0.3	0.5
Param	mean	0.535	0.960	0.267	0.456	0.607	0.866	0.201	0.354
	s.d.	0.087	0.095	0.096	0.093	0.082	0.079	0.078	0.080
Semi	mean	0.633	0.833	0.170	0.299	0.343	1.205	0.436	0.706
	s.d.	0.122	0.122	0.114	0.124	0.114	0.105	0.112	0.109
	MSE Eff	0.272	0.249	0.341	0.188	0.481	0.460	0.524	0.516
	mean	0.511	0.991	0.287	0.491	0.526	0.977	0.281	0.458
	s.d	0.063	0.066	0.064	0.062	0.081	0.085	0.075	0.085
	MSE Eff	2.173	2.377	2.386	2.665	2.491	3.106	2.661	3.080

TABLE 5

500 simulations, 1000 cases/1000 controls,10% disease rate, correlated covariates $\mathbf{X}$ *with dimension 4, Gamma error. See Table 1 for definitions.*

A.8. Simulation with Higher Dimensional Covariates

		Homoscedastic Gaussian error					
Ctrl	β_1	0	0	0.5	1.0	0.3	0.5
	mean	-0.001	0.006	0.518	0.974	0.285	0.467
Param	s.d.	0.115	0.117	0.118	0.121	0.115	0.107
	mean	-0.007	0.000	0.523	0.968	0.273	0.460
	s.d.	0.087	0.089	0.087	0.087	0.088	0.092
Semi	MSE Eff	1.735	1.728	1.776	1.778	1.575	1.252
	mean	-0.010	0.005	0.506	0.992	0.299	0.497
	s.d	0.080	0.086	0.102	0.099	0.095	0.083
	MSE Eff	2.020	1.846	1.372	1.564	1.477	1.827
		Heteroscedastic Gaussian error					
Ctrl	β_1	0	0	0.5	1.0	0.3	0.5
	mean	-0.006	0.006	0.553	0.937	0.248	0.444
Param	s.d.	0.107	0.108	0.104	0.104	0.104	0.100
	mean	-0.001	-0.002	0.262	1.258	0.500	0.777
	s.d.	0.088	0.092	0.089	0.084	0.086	0.086
Semi	MSE Eff	1.482	1.373	0.211	0.201	0.282	0.156
	mean	0.003	0.005	0.503	0.982	0.295	0.497
	s.d.	0.086	0.087	0.109	0.106	0.097	0.097
	MSE Eff	1.555	1.538	1.158	1.282	1.417	1.397

TABLE 6

500 simulations, 1000 cases/1000 controls, 4.5% disease rate, correlated covariates $\mathbf{X}$ *with dimension 6, Gaussian error. See Table 1 for definitions.*

		Homoscedastic Gaussian error							
Ctrl	β_1	0	0	0	0	0.5	1.0	0.3	0.5
	mean	-0.011	-0.005	0.004	0.003	0.517	0.978	0.288	0.479
Param	s.d.	0.117	0.114	0.114	0.116	0.117	0.120	0.114	0.116
	mean	-0.001	0.004	-0.005	0.003	0.519	0.967	0.274	0.460
	s.d.	0.087	0.086	0.091	0.091	0.085	0.089	0.091	0.087
Semi	MSE Eff	1.844	1.760	1.570	1.620	1.836	1.629	1.445	1.519
	mean	-0.010	0.005	-0.002	-0.002	0.512	0.994	0.308	0.507
	s.d	0.086	0.085	0.087	0.082	0.094	0.104	0.104	0.092
	MSE Eff	1.854	1.800	1.724	1.980	1.538	1.360	1.210	1.656
		Heteroscedastic Gaussian error							
Ctrl	β_1	0	0	0	0	0.5	1.0	0.3	0.5
	mean	-0.002	-0.005	0.003	0.005	0.545	0.941	0.259	0.444
Param	s.d.	0.106	0.106	0.103	0.105	0.105	0.102	0.109	0.113
	mean	-0.008	-0.009	0.001	0.001	0.266	1.249	0.496	0.776
	s.d.	0.093	0.086	0.092	0.093	0.094	0.086	0.088	0.086
Semi	MSE Eff	1.295	1.535	1.254	1.265	0.207	0.200	0.294	0.189
	mean	-0.005	-0.002	0.009	0.005	0.498	1.001	0.304	0.494
	s.d.	0.084	0.091	0.079	0.084	0.099	0.105	0.108	0.102
	MSE Eff	1.594	1.367	1.660	1.541	1.337	1.261	1.159	1.529

TABLE 7

500 simulations, 1000 cases/1000 controls, 4.5% disease rate, correlated covariates $\mathbf{X}$ *with dimension 8, Gaussian error. See Table 1 for definitions.*

Ctrl	β_1	Homoscedastic Gaussian error									
		0	0	0	0	0	0	0.5	1.0	0.3	0.5
Param	mean	-0.003	-0.000	-0.002	0.005	-0.001	-0.005	0.517	0.984	0.282	0.472
	s.d.	0.112	0.107	0.117	0.121	0.116	0.120	0.120	0.120	0.126	0.117
Semi	mean	0.006	-0.000	-0.003	0.000	-0.002	-0.004	0.521	0.967	0.277	0.459
	s.d.	0.088	0.086	0.085	0.087	0.089	0.086	0.091	0.088	0.091	0.088
	MSE Eff	1.607	1.558	1.911	1.954	1.681	1.971	1.663	1.655	1.849	1.513
	mean	-0.003	-0.001	-0.003	0.002	0.003	-0.002	0.497	0.997	0.299	0.502
	s.d.	0.083	0.080	0.085	0.083	0.087	0.089	0.090	0.094	0.089	0.091
	MSE Eff	1.803	1.778	1.921	2.113	1.759	1.806	1.812	1.640	2.046	1.742
Ctrl	β_1	Heteroscedastic Gaussian error									
		0	0	0	0	0	0	0.5	1.0	0.3	0.5
Param	mean	0.003	-0.015	-0.001	-0.001	0.001	-0.001	0.552	0.947	0.247	0.444
	s,d.	0.111	0.108	0.105	0.104	0.103	0.105	0.115	0.105	0.106	0.109
Semi	mean	-0.004	-0.001	-0.013	-0.009	0.001	-0.003	0.267	1.252	0.495	0.780
	s.d.	0.092	0.096	0.091	0.092	0.091	0.089	0.098	0.088	0.093	0.088
	MSE Eff	1.439	1.294	1.314	1.266	1.288	1.415	0.252	0.196	0.298	0.174
	mean	-0.001	-0.001	-0.008	0.003	0.011	-0.002	0.509	1.000	0.292	0.496
	s.d.	0.088	0.085	0.081	0.078	0.079	0.087	0.104	0.100	0.101	0.108
	MSE Eff	1.572	1.631	1.681	1.766	1.699	1.457	1.480	1.384	1.364	1.281

TABLE 8. *500 simulations, 1000 cases/1000 controls, 4.5% disease rate, correlated covariates* $\mathbf{X}$ *with dimension 10, Gaussian error. See Table 1 for definitions.*

References

Chatterjee, N. and Carroll, R. J. (2005). Semiparametric maximum likelihood estimation in case-control studies of gene-environment interactions. *Biometrika*, 92, 399–418. MR2201367

Chatterjee, N., Chen, Y.-H., Luo, S., and Carroll, R. J. (2009). Analysis of case-control association studies: SNPs, imputation and haplotypes. *Statistical Science*, 24, 489–502. MR2779339

Chen, J., Ayyagari, R., Chatterjee, N., Pee, D. Y., Schairer, C., Byrne, C., Benichou, J., and Gail, M. H. (2008). Breast cancer relative hazard estimates from case–control and cohort designs with missing data on mammographic density. *Journal of the American Statistical Association*, 103, 976–988. MR2528822

Chen, J., Pee, D., Ayyagari, R., Graubard, B., Schairer, C., Byrne, C., Benichou, J., and Gail, M. H. (2006). Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute*, 98, 1215–1226.

Chen, Y. H., Chatterjee, N., and Carroll, R. J. (2008). Retrospective analysis of haplotype-based case-control studies under a flexible model for gene-environment association. *Biostatistics*, 9, 81–99.

Chen, Y. H., Chatterjee, N., and Carroll, R. J. (2009). Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *Journal of the American Statistical Association*, 104, 220–233. MR2663041

Cook, R. D. (1994). On the interpretation of regression plots. *Journal of the American Statistical Association*, 89, 177–189. MR1266295

Cook, R. D. (2009). *Regression Graphics: Ideas for Studying Regressions Through Graphics*, volume 482. John Wiley & Sons. MR1645673

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Annals of Statistics*, pages 455–474. MR1902895

Cook, R. D. and Setodji, C. M. (2003). A model-free test for reduced rank in multivariate regression. *Journal of the American Statistical Association*, 98, 340–351. MR1995710

Dong, Y. and Li, B. (2010). Dimension reduction for non-elliptically distributed predictors: second-order methods. *Biometrika*, 97, 279–294. MR2650738

Jiang, Y., Scott, A. J., and Wild, C. J. (2006). Secondary analysis of case-control data. *Statistics in Medicine*, 25, 1323–1339. MR2226789

Li, B. and Dong, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Annals of Statistics*, pages 1272–1298. MR2509074

Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102, 997–1008. MR2354409

Li, B., Wen, S., and Zhu, L. (2008). On a projective resampling method for dimension reduction with multivariate responses. *Journal of the American Statistical Association*, 103, 1177–1186. MR2462891

Li, B., Zha, H., and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *Annals of Statistics*, pages 1580–1616. MR2166556

Li, H., Gail, M. H., Berndt, S., and Chatterjee, N. (2010). Using cases to

strengthen inference on the association between single nucleotide polymorphisms and a secondary phenotype in genome-wide association studies. *Genetic Epidemiology*, 34, 427–433. MR2504371

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86, 316–327. MR1137117

Li, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *Journal of the American Statistical Association*, 87, 1025–1039. MR1209564

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *Annals of Statistics*, 17, 1009–1052. MR1015136

Lian, H., Liang, H., and Carroll, R. J. (2015). Variance function partially linear single-index models. *Journal of the Royal Statistical Society: Series B*, 77, 171–194. MR3299404

Lin, D. Y. and Zeng, D. (2009). Proper analysis of secondary phenotype data in case-control association studies. *Genetic Epidemiology*, 33, 256–265.

Ma, Y. (2010). A semiparametric efficient estimator in case-control studies. *Bernoulli*, 16, 585–603. MR2668916

Ma, Y. and Carroll, R. J. (2016). Semiparametric estimation in the secondary analysis of case–control studies. *Journal of the Royal Statistical Society, Series B*, 78, 127–151. MR3453649

Ma, Y. and Zhu, L. (2012a). Efficiency loss caused by linearity condition in dimension reduction. *Biometrika*, 99, 1–13.

Ma, Y. and Zhu, L. (2012b). A semiparametric approach to dimension reduction. *Journal of the American Statistical Association*, 107, 168–179. MR2949349

Ma, Y. and Zhu, L. (2013a). Efficient estimation in sufficient dimension reduction. *Annals of Statistics*, 41, 250–268. MR3059417

Ma, Y. and Zhu, L. (2013b). A review on dimension reduction. *International Statistical Review*, 81, 134–150. MR3047506

Ma, Y. and Zhu, L. P. (2013c). Efficient estimation in sufficient dimension reduction. *Annals of Statistics*, 41, 250–268. MR3059417

Piegorsch, W. W., Weinberg, C. R., and Taylor, J. A. (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population based case-control studies. *Statistics in Medicine*, 13, 153–162.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66, 403–411. MR0556730

Setodji, C. M. and Cook, R. D. (2004). K-means inverse regression. *Technometrics*, 46, 421–429. MR2101510

Tchetgen, E. J. T. (2014). A general regression framework for a secondary outcome in case–control studies. *Biostatistics*, 15, 117–128.

Wei, J., Carroll, R. J., Müller, U. U., Van Keilegom, I., and Chatterjee, N. (2013). Robust estimation for homoscedastic regression in the secondary analysis of case–control data. *Journal of the Royal Statistical Society, Series B*, 75, 185–206. MR3008277

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Annals of Statistics*, pages 2654–2690. MR2382662

Yin, X. and Bura, E. (2006). Moment-based dimension reduction for multivari-

ate response regression. *Journal of Statistical Planning and Inference*, 136, 3675–3688. MR2256281

Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional kth moment in regression. *Journal of the Royal Statistical Society: Series B*, 64, 159–175. MR1904698

Zhu, L., Wang, T., Zhu, L., and Ferré, L. (2010). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika*, 97, 295–304. MR2650739