A Functional Single Index Model

Fei Jiang, Seungchul Baek, Jiguo Cao and Yanyuan Ma
Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China
Department of Statistics, University of South Carolina, Columbia, SC 29208, USA
Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
Department of Statistics, Penn State University, University Park, PA 16802, USA
feijiang@hku.hk, sbaek@stat.sc.edu, jiguo_cao@sfu.ca, and yzm63@psu.edu

Abstract

Motivated by the National Morbidity, Mortality and Air Pollution Study, we propose a semiparametric functional single index model to study the relation between a univariate response and multiple functional covariates. The parametric part of the model integrates the functional linear regression model and the sufficient dimension reduction structure. The nonparametric part of the model further allows the response-index dependence or the link function to be unspecified. We use B-splines to approximate the coefficient function in the functional linear regression model part and reduce the problem to a dimension folding model. We develop a new method to handle the subsequent dimension folding model by using kernel regression in combination with semiparametric treatment. The new method does not impose any special requirement on the inner product between the covariate function and the B-spline bases, and allows efficient estimation of both the index vector and the B-spline coefficients. The estimation method is general and applicable to both continuous and discrete response variables. We further derive asymptotic properties of the class of methods for both the index vector and the coefficient function. We establish the semiparametric optimality, which involves highly unconventional analysis

and has never been done before in a functional semiparametric model where both kernel and B-spline estimation are involved.

Key Words: B-spline, Dimension reduction, Dimension folding, Functional linear model, Infinite dimension, Kernel estimation, Single index model.

1 Introduction

The National Morbidity, Mortality and Air Pollution Study (NMMAP) is an important environmental study aiming at addressing the uncertainties regarding the association between pollution and health (Samet et al., 2000). In this study, daily measurements on various air pollutants such as carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and ozone (O_3) are collected in different cities over the course of a whole year. The annual death rate caused by cardiovascular diseases (CVD) is also collected in these cities. Long-term epidemiological studies have been conducted to examine the effect of O_3 on the CVD rate individually with no significant effect detected (Cox and Popken, 2015; Turner et al., 2016). We also use a simple functional linear regression model to study the time varying effect of each air pollutant on the CVD death rate. The result shows that no individual air pollutant has a significant effect on the CVD death rate. We would like to further investigate whether various kinds of air pollutants combined together will have a significant effect on the CVD death rate.

This motivates us to develop a single air pollution index which combines the pollutants in the way that best describes the severity of the overall air pollution level to the CVD death rate. At the same time, we are also interested in discovering possible time-varying effect of the single air pollution index to the CVD death rate. To achieve these goals, in this article we propose a functional single index model and proceed to devise a novel class of estimators.

More specifically, the NMMAP data contains the measurements of the daily air pollutant concentrations $\mathbf{X}(t) \equiv \{X_1(t), \ldots, X_J(t)\}^{\mathrm{T}}$, a J-dimensional functional covariate of $t \in [0, 1]$, and the annual CVD death rate in the subsequent year as the response Y. To measure the overall severeness of the air pollution, we define the single air pollution index as

$$W(t) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}(t) = \sum_{i=1}^{J} \beta_i X_i(t),$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^{\mathrm{T}}$ is the vector of weights to various air pollutants. We assume that Y is linked to $W(t)$ through

$$f\{Y|\mathbf{X}(t)\} = f\left\{Y, \int_0^1 W(t)\alpha(t)dt\right\} = f\left\{Y, \int_0^1 \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}(t)\alpha(t)dt\right\}, \qquad (1)$$

where f is a conditional probability density function (pdf) or probability mass function (pmf) and is left unspecified. Therefore, f satisfies $f \geq 0$, $\int f(y, \cdot)d\mu(y) = 1$, where $\mu(y)$ is the probability measure of Y, but the functional form of f is unknown. Note that unlike the usual single index model, we make the assumption on the conditional distribution, instead of the conditional mean.

In the functional single index model (1), the components in $\boldsymbol{\beta}$ capture the relative importance of various pollutants in determining the pollution severity, and the functional parameter $\alpha(t)$ captures the time-varying effect of the air pollution index on the annual CVD death rate. The construction of a single air pollution index $W(t)$ allows the public who are interested in the relation between CVD death and air pollution to simply focus on a singular air pollution index instead of viewing multiple air pollutants over time. Our objective is to estimate $\boldsymbol{\beta}$ and $\alpha(t)$ for the functional single index model (1) from the observations $\{\mathbf{X}_1(t), Y_1\}, \ldots, \{\mathbf{X}_n(t), Y_n\}$.

Model (1) is of interest in its own form and is very general. For example, when $X(t)$ is univariate, $\boldsymbol{\beta}$ is eliminated from the model and (1) simplifies to contain only the unknown index function $\alpha(t)$. This can be used to capture the less structured modeling of the air pollution problem that motivated our study. Specifically, consider combining the pollutants effect via $\int_0^1 \mathbf{X}(t)^{\mathrm{T}}\boldsymbol{\alpha}(t)dt$, where now $\boldsymbol{\alpha}(t) = \{\alpha_1(t), \ldots, \alpha_J(t)\}^{\mathrm{T}}$ is a vector of functions. This can be rewritten as $\int_0^J X_L(t)\alpha_L(t)dt$, $X_L(t) = X_{[t]}(t - [t])$, $\alpha_L(t) = \alpha_{[t]}(t - [t])$, where $[t]$ is the largest integer that is smaller than or equal to t. We can of course further rescale the time through $X_L^*(t) = X_L(Jt)$ and $\alpha_L^*(t) = \alpha_L(Jt)J$ to rewrite the model into $\int_0^1 X_L^*(t)\alpha_L^*(t)dt$. In other words, we can stack the covariates and index functions to rewrite the less structured model $\int_0^1 \mathbf{X}(t)^{\mathrm{T}}\boldsymbol{\alpha}(t)dt$ into the functional single index model $\int_0^1 X_L^*(t)\alpha_L^*(t)dt$, which is a special case of model (1) corresponding to $J = 1$ and hence $\boldsymbol{\beta}$ drops out of the model. This indicates that although for a specific problem, a less structured functional index model is more flexible, mathematically, such model falls into the general framework described by (1) hence the methodology developed here will also apply to such model automatically.

The estimation and inference of the functional single index model (1) is of course not simple. The complexity is due to the unspecified bivariate link function f as well as the unknown coefficient function $\alpha(t)$, in addition to the unknown index vector $\boldsymbol{\beta}$. If $\alpha(t)$ had been known, (1) would reduce to a central space estimation problem and various methods exist to estimate $\boldsymbol{\beta}$ (Li, 1991; Cook and Weisberg, 1991; Li and Wang, 2007; Ma and Zhu, 2013). If $\boldsymbol{\beta}$ had been known, (1) would reduce to a functional dimension reduction problem (Ferré and Yao, 2003, 2005, 2007). In Section 2.1, we point out that our proposed estimation method also forms an alternative solution for the dimension folding problem (Li et al., 2010), and our method does not require any conditions on the covariates as other methods in the literature do.

We establish the asymptotic properties and provides inference tools by simultaneously taking into account spline approximation and multiple kernel estimation. We show that the estimation procedure is not only consistent but also efficient under the original nonparametric modeling components, despite that the estimation procedure is devised post the replacement of the nonparametric function with its spline approximation. The theoretical derivation heavily relies on the properties of a second moment linear operator based on the integral of the covariates, which is not common in the functional data analysis.

The functional single index model (1) is closely related to the sufficient dimension reduction modeling, where a response is assumed to be related to the linear combinations of a multivariate covariate while the exact functional relation is left unspecified (Cook, 1998). To this end, we can view $\int_0^1 \mathbf{X}(t)\alpha(t)dt$ as the covariate vector in the classical sufficient dimension reduction model. On the other hand, since at each time point t, we can also view $W(t)$ as a single function, model (1) can be classified as a nonparametric model with a functional linear regression core. Note that (1) is different from the model considered in Jiang et al. (2014), where $\boldsymbol{\beta}$ does not occur, only univariate covariate function $X(t)$ is considered, and several coefficient functions are considered. Further, as we have pointed out, (1) has more structure and is more complex than the stack model (Ramsay and Silverman, 2005).

In summary, the new model and estimators have the following features. (i). The model contains a single air pollution index to summarize the pollution severity level. (ii). The model also contains a time varying coefficient which helps to provide timely and seasonally adjusted health advices to the general public. (iii). The model allows the relation between the CVD death rate and the overall pollution effect to be arbitrary and unspecified, which avoids possible model mis-specification. (iv). The model can be viewed as an extension of dimension reduction from handling vector of variables to vector of functions.

The rest of the article is organized as the following. In Section 2, we develop a class of semiparametric methods to estimate both $\boldsymbol{\beta}$ and $\alpha(t)$, and derive the asymptotic properties of the estimators. We discuss the relation between the proposed method with the semiparametric sufficient dimension reduction in Section 3. We illustrate the finite sample performance of the proposed method through simulation in Section 4. The analysis of the pollution-cardiovascular disease relation is carried out in Section 5. We conclude the article with some discussions in Section 6.

2 Methodology

2.1 Model and Identifiability

Before proceeding with developing estimation methods, we first need to establish the identifiability of the model described in (1). In fact, without more specific parameterization requirement, (1) is indeed not identifiable. For example, we can easily multiply $\alpha(t)$ or $\boldsymbol{\beta}$ by a constant, and adjust the link function f to capture the same relation between $\mathbf{X}(t)$ and Y. In Proposition

1, we provide one way of parameterization that guarantees the identifiability. We provide its proof in the Supplement.

Proposition 1. *Assume $\alpha(t)$ is continuous and set $\alpha(0) = 1$ and $\beta_J = 1$, then model (1) is identifiable.*

We perform the analysis of model (1) under the specification in Proposition 1. Without any parametric assumption on $\alpha(t)$, we approximate $\alpha(t)$ as a linear combination of B-spline basis function: $\alpha(t) \approx \sum_{k=1}^{d_\gamma} \gamma_k B_{rk}(t)$, where $B_{rk}(t)$ is the kth B-spline basis function with order r ($r \geq 0$). Then the functional single index model (1) reduces to

$$f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = f\left\{Y, \sum_{j=1}^{J}\sum_{k=1}^{d_\gamma} \beta_j\gamma_k \int_0^1 X_j(t)B_{rk}(t)dt\right\}, \tag{2}$$

where $\mathbf{Z}$ is a $J \times d_\gamma$ matrix with the (j,k)th entry $Z_{jk} \equiv \int_0^1 X_j(t)B_{rk}(t)dt$. We then propose methods to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{d_\gamma})^{\mathrm{T}}$ simultaneously. Because the B-spline property ensures $B_{r1}(0) = 1$ and $B_{rk}(0) = 0$ for $k > 1$ and our parameterization fixes $\alpha(0) = 1$, we automatically obtain $\gamma_1 = 1$.

Treating each Z_{jk} as a single index and $\beta_j\gamma_k$ for $j = 1, \ldots, J$, $k = 1, \ldots, d_\gamma$, as new coefficients, model (2) can be considered as a standard single index model. Hence, the procedures discussed in Li (1991); Cook and Weisberg (1991); Li and Wang (2007) and Ma and Zhu (2013) can be applied for the estimation. However, this will require additional care to take into consideration the inherent relation between these coefficients. In addition, this practice will result in a model with dimension Jd_γ which can be extra high-dimensional even if the number of covariates are moderately large.

Weighing the prices of converting (2) into a single index model described above, we decide to proceed with an independent analysis of (2) and give up pursuing available procedures in the single index model literature. Since model (2) assumes a form of the dimension folding models described in Li et al. (2010), we introduce an efficient procedure where we only have $d_\gamma + J - 2$ free coefficients to estimate due to the multiplication from both left and right sides of $\mathbf{Z}$.

As an improvement of the dimension folding method, our proposed estimation procedure relaxes the additional constraints on the covariate matrix $\mathbf{Z}$. For example, our procedure does not require that $E\{\mathbf{X} \mid (\boldsymbol{\gamma} \otimes \boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})\}$ is a linear function of $(\boldsymbol{\gamma} \otimes \boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})$ and that $\mathrm{var}\{\mathbf{X} \mid (\boldsymbol{\gamma} \otimes \boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})\}$ does not depend on $(\boldsymbol{\gamma} \otimes \boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})$ as being enforced in Li et al. (2010). Here, $\otimes$ stands for the Kronecker product and $\mathrm{vec}(\mathbf{X})$ is the vector formed by the concatenation of the columns of $\mathbf{X}$. These constraints may be violated and are not assumed to hold for model (2) in general.

2.2 Doubly Robust Local Efficient Score

To estimate the parameters, we derive the efficient score components as

$$\begin{aligned} \mathbf{S}_{\mathrm{eff}\boldsymbol{\beta}}(Y, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, g) &= \left[g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) - E\left\{g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right\}\right] \\ &\quad \times(\mathbf{I}_{J-1}, \mathbf{0})\{\mathbf{Z} - E(\mathbf{Z} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\boldsymbol{\gamma}, \\ \mathbf{S}_{\mathrm{eff}\boldsymbol{\gamma}}(Y, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, g) &= \left[g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) - E\left\{g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right\}\right] \\ &\quad \times(\mathbf{0}, \mathbf{I}_{d_\gamma-1})\{\mathbf{Z} - E(\mathbf{Z} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}^{\mathrm{T}}\boldsymbol{\beta}, \end{aligned} \tag{3}$$

where the function $g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = f_2'(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})/f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ and $f_2'(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ is the partial derivative of f with respect to its second argument. Let $\mathbf{S}_{\mathrm{eff}}(Y_i, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, g) = \{\mathbf{S}_{\mathrm{eff}\boldsymbol{\beta}}(Y, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, g)^{\mathrm{T}}, \mathbf{S}_{\mathrm{eff}\boldsymbol{\gamma}}(Y, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, g)^{\mathrm{T}}\}^{\mathrm{T}}$, we then essentially use the Newton-Raphson procedure to solve for $\boldsymbol{\beta}, \boldsymbol{\gamma}$ from

$$\sum_{i=1}^{n} \mathbf{S}_{\mathrm{eff}}(Y_i, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, g) = \mathbf{0}. \tag{4}$$

When $g(\cdot)$ and $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ are correctly specified, $\mathbf{S}_{\mathrm{eff}}$ is a score function which falls in the space that orthogonal to the nuisance tangent space induced by the unknown conditional density $f(\cdot)$ defined in Proposition 3 in the Supplement. Hence, as shown in Bickel et al. (1993); Tsiatis (2004), $\mathbf{S}_{\mathrm{eff}}$ is an efficient score (see Proposition S.4.1 in the Supplement) which yields the optimal estimators with the smallest asymptotic variances. In addition, $\mathbf{S}_{\mathrm{eff}}$ is also doubly robust function so that the estimation consistency holds whenever $E\left\{g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right\}$ or $E\left(\mathbf{Z} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right)$ is correctly specified.

In reality, the functional form for the conditional density $f(\cdot)$ is often unknown, this leads to the difficulty of obtaining $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ and $f_2'(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})/f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ in (3). To retain the best estimation efficiency without imposing more assumptions, we use the standard kernel smoothing method in nonparametric regression to estimate $E(\cdot \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, i.e.

$$\widehat{E}\{m(Y_i, \mathbf{Z}_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\} = \frac{\sum_{i=1}^{n} m(Y_i, \mathbf{Z}_i) K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})}{\sum_{i=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})}, \tag{5}$$

for arbitrary function $m(Y_i, \mathbf{Z}_i)$. Here $K(\cdot)$ is a kernel function and $K_h(\cdot) = h^{-1}K(\cdot/h)$. To estimate $f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma})$, we obtain the estimators $\widehat{f}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}) = c_0$, and $\partial\widehat{f}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma})/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}) = c_1$ from minimizing

$$\sum_{j=1}^{J}\{K_b(Y_j - Y) - c_0 - c_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma})\}^2 K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}) \tag{6}$$

with respect to c_0, c_1. (5) and (6) allow us to obtain the unknown quantities in $\mathbf{S}_{\mathrm{eff}}$ consistently, which requires heavy computation especially in solving (6) for each value of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j\boldsymbol{\gamma}$. To ease the computation, we may posit a model for f, f_2, say f^* and f_2^*. Let $g^* = f_2^{*\prime}/f^*$, $\mathbf{S}_{\mathrm{eff}}(Y_i, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, g^*)$ is a locally efficient score function which guarantees estimation consistency while resulting in efficient estimators when g^* is the truth.

It is also worth pointing out the issue of obtaining $\mathbf{Z}_i$'s. Unlike in the usual dimension reduction problems, $\mathbf{Z}_i$'s are not directly observed and need to be constructed from the observed $X_j(t)$'s. This involves a numerical approximation of the integrals $\int_0^1 B_{rk}(t)X_j(t)dt$. The composite Simpson's rule (Atkinson, 1989) can be used to approximate the numerical integration, which has the form

$$\begin{aligned}\int_0^1 B_{rk}(t)X_j(t)dt \quad = \quad & \frac{1}{3Q}\Bigg[B_{rk}(t_0)X_j(t_0) + 2\sum_{q=1}^{Q/2-1}\{B_{rk}(t_{2q})X_j(t_{2q})\} \\ & +4\sum_{q=1}^{Q/2}\{B_{rk}(t_{2q-1})X_j(t_{2q-1})\} + B_{rk}(t_Q)X_j(t_Q)\Bigg],\end{aligned}$$

where $t_q = q/Q$, $q = 0, 1, \ldots, Q$, and Q is an even number.

The estimation procedures can be summarized as follows:

Step 1: Choose f and f_2' by minimizing (6) or positing f and f_2'. Denote the choices by f^* and $f_2'^*$ and let $g^* = f_2'^*/f^*$.

Step 2: Replace g in (4) by g^* according to the **Step 1** choice.

Step 3: Let $\widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\gamma}}$ be the version of $\mathbf{S}_{\text{eff}\boldsymbol{\gamma}}$ when replacing $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ by $\widehat{E}(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ defined in (5). Treating $\boldsymbol{\gamma}$ as a function of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\gamma}(\boldsymbol{\beta})$, we solve

$$\sum_{i=1}^{n} \widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\gamma}}(Y_i, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}(\boldsymbol{\beta}), g^*) = \mathbf{0}$$

for $\boldsymbol{\gamma}(\boldsymbol{\beta})$, and denote the estimator as $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})$.

Step 4: Let $\widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\beta}}$ be the version of $\mathbf{S}_{\text{eff}\boldsymbol{\beta}}$ when replacing $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ by $\widehat{E}(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$. We solve

$$\sum_{i=1}^{n} \widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\beta}}(Y_i, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), g^*) = \mathbf{0}$$

for $\boldsymbol{\beta}$, and denote the estimator as $\widehat{\boldsymbol{\beta}}$.

2.3 Asymptotic Results

The profiling procedures Step 3 and 4 yield the estimators which are asymptotically equivalent to the ones from solving the estimating equation based on the estimating function $(\widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\gamma}}^{\mathrm{T}}, \widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\beta}}^{\mathrm{T}})^{\mathrm{T}}$. Hence the estimation consistency readily holds by the following proposition.

Proposition 2. *Let $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}$ satisfy*

$$\sum_{i=1}^{n} \{\widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\beta}}(Y_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, g^*)^{\mathrm{T}}, \widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\gamma}}(Y_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, g^*)^{\mathrm{T}}\}^{\mathrm{T}} = \mathbf{0}$$

where

$$\widehat{\mathbf{S}}_{\text{eff}\boldsymbol{\beta}}(Y_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, g^*)$$

$$
\begin{aligned}
&= \left\{ g^*(Y_i, \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma}) - \frac{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma}) g^*(Y_j, \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma})}{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma})} \right\} \boldsymbol\Theta_{\boldsymbol\beta} \\
&\quad \times \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma})\mathbf{Z}_j}{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma})} \right\} \widehat{\boldsymbol\gamma}, \\
&\widehat{\mathbf{S}}_{\rm eff\boldsymbol\gamma}(Y_i, \mathbf{Z}_i, \widehat{\boldsymbol\beta}, \widehat{\boldsymbol\gamma}, g^*) \\
&= \left\{ g^*(Y_i, \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma}) - \frac{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma}) g^*(Y_j, \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma})}{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma})} \right\} \boldsymbol\Theta_{\boldsymbol\gamma} \\
&\quad \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma})\mathbf{Z}_j}{\sum_{j=1}^J K_h(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_j\widehat{\boldsymbol\gamma} - \widehat{\boldsymbol\beta}^{\rm T}\mathbf{Z}_i\widehat{\boldsymbol\gamma})} \right\}^{\rm T} \widehat{\boldsymbol\beta}.
\end{aligned}
$$

Let $\boldsymbol\beta_0$ be the true $\boldsymbol\beta$. Further let $\boldsymbol\gamma_0$ be a spline coefficient which satisfies $\sup_{t\in[0,1]} |\mathbf{B}_r(t)^{\rm T}\boldsymbol\gamma_0 - \alpha_0(t)| = O_p(h_b^q)$ as stated in Condition (A5). Then $\widehat{\boldsymbol\beta} - \boldsymbol\beta_0 = o_p(1)$, $\sup_{t\in[0,1]} |\mathbf{B}_r(\cdot)^{\rm T}\widehat{\boldsymbol\gamma} - \mathbf{B}_r(\cdot)^{\rm T}\boldsymbol\gamma_0| = o_p(1)$.

In Step 3, at the point of convergence, we show that $\widehat{\boldsymbol\gamma}(\boldsymbol\beta_0)$ achieves the nonparametric spline regression convergence rate, and derive its asymptotic variation as follows.

Theorem 1. *Assume Conditions (A1)–(A8) hold, and let $\mathbf{B}_r(\cdot)^{\rm T}\widehat{\boldsymbol\gamma}(\boldsymbol\beta_0)$ satisfy*

$$
\sum_{i=1}^n \widehat{\mathbf{S}}_{\rm eff\boldsymbol\gamma}(Y_i, \mathbf{Z}_i, \boldsymbol\beta_0, \widehat{\boldsymbol\gamma}(\boldsymbol\beta_0), g^*) = \mathbf{0}.
$$

Then $n^{1/2}\{\widehat{\boldsymbol\gamma}^-(\boldsymbol\beta_0) - \boldsymbol\gamma_0^-\} = \mathbf{L} + o_p(\mathbf{L})$, where

$$
\begin{aligned}
\mathbf{L} &= -\Bigg(\Bigg\{E\Bigg(\boldsymbol\Theta_{\boldsymbol\gamma}\left[\mathbf{Z}_i - E\{\mathbf{Z}_i|\boldsymbol\beta_0\boldsymbol\alpha_0(\mathbf{X}_i)\}\right]^{\rm T}\boldsymbol\beta_0 \\
&\quad \times \frac{\partial\left(g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\} - E[g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}|\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)]\right)}{\partial\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)} \\
&\quad \times\boldsymbol\beta_0^{\rm T}[\mathbf{Z}_i - E\{\mathbf{Z}_i|\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}]\boldsymbol\Theta_{\boldsymbol\gamma}^{\rm T}\Bigg)\Bigg\}^{-1}\Bigg)\Bigg(n^{-1/2}\sum_{i=1}^n[g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\} \\
&\quad -E\{g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}|\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}]\boldsymbol\Theta_{\boldsymbol\gamma}\left[\mathbf{Z}_i - E\{\mathbf{Z}_i|\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}\right]^{\rm T}\boldsymbol\beta_0\Bigg).
\end{aligned}
$$

Here $\boldsymbol\gamma_0^- = (\gamma_{02}, \ldots, \gamma_{0d_{\boldsymbol\gamma}})^{\rm T}$, and $\widehat{\boldsymbol\gamma}^-(\boldsymbol\beta_0) = (\widehat\gamma_2(\boldsymbol\beta_0), \ldots, \widehat\gamma_{d_{\boldsymbol\gamma}}(\boldsymbol\beta_0))^{\rm T}$. Further, for arbitrary $d_{\boldsymbol\gamma-1}$-dimensional vector with $\|\mathbf{a}\|_2 < \infty$, we have $\mathbf{a}^{\rm T}\{\widehat{\boldsymbol\gamma}^-(\boldsymbol\beta_0) - \boldsymbol\gamma_0^-\} = O_p\{(nh_b)^{-1/2}\}$.

In addition, we show $\widehat{\boldsymbol\beta}$ from Step 4 is not only root-n consistent, but also efficient and achieve the information lower bound $\{E(\mathbf{S}_{o\text{eff}}^{\otimes 2})\}^{-1}$. Here $\mathbf{S}_{o\text{eff}}$ is the efficient score for $\boldsymbol\beta$ of the original model (1), which contains $\alpha(\cdot)$ instead of $\mathbf{B}_r(\cdot)^{\rm T}\boldsymbol\gamma$, hence it is different from $\mathbf{S}_{\text{eff}\boldsymbol\beta}$. Its precise expression is given in Proposition S.4.2 in the Supplement.

To show the asymptotic properties of $\widehat{\boldsymbol\beta}$, we first define

$$\begin{aligned}&\Delta g_c^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}\\ \equiv\ & \frac{\partial[g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\} - E\{g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}|\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}]}{\partial\{\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}}\end{aligned}$$

and $\mathbf{w}_0^*(t)$ as a function that satisfies

$$\begin{aligned}&\boldsymbol\Theta_{\boldsymbol\beta} E[\boldsymbol\alpha_{c0}(\mathbf{X}_i)\Delta g_c^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}\mathbf{X}_{ic}^{\rm T}(t)]\boldsymbol\beta_0 \\ =\ & \int_0^1 E[\boldsymbol\beta_0^{\rm T}\mathbf{X}_{ic}(s)\Delta g_c^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}\mathbf{X}_{ic}^{\rm T}(t)\boldsymbol\beta_0]\mathbf{w}_0^*(s)ds,\end{aligned} \tag{7}$$

where $\boldsymbol\alpha_{c0}(\mathbf{X}) \equiv \boldsymbol\alpha_0(\mathbf{X}) - E\{\boldsymbol\alpha_0(\mathbf{X}) \mid \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X})\}$. We have the following results.

Theorem 2. *Assume Conditions (A1)–(A8) hold, and let* $\widehat{\boldsymbol\beta}$ *satisfy*

$$\sum_{i=1}^n \widehat{\mathbf{S}}_{\text{eff}\boldsymbol\beta}(Y_i, \mathbf{Z}_i, \widehat{\boldsymbol\beta}, \widehat{\boldsymbol\gamma}(\widehat{\boldsymbol\beta}), g^*) = \mathbf{0}.$$

Then $\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta_0) = \mathbf{A}^{-1}\mathbf{B} + o_p(1)$, *where*

$$\begin{aligned}\mathbf{A} &= -E\Big[\Delta g_c^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}\{\boldsymbol\Theta_{\boldsymbol\beta}\boldsymbol\alpha_{c0}(\mathbf{X}_i)\}^{\otimes 2} \\ &\qquad -\Delta g_c^*\left\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\right\}\int_0^1 \boldsymbol\beta_0^{\rm T}\mathbf{X}_{ic}(t)\mathbf{w}_0^*(t)dt\,\boldsymbol\alpha_{c0}(\mathbf{X}_i)^{\rm T}\boldsymbol\Theta_{\boldsymbol\beta}^{\rm T}\Big]\end{aligned}$$

and

$$\begin{aligned}\mathbf{B} &= n^{-1/2}\sum_{i=1}^n \left\{g^*\left\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\right\} - E\left[g^*\{Y_i, \boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\}|\boldsymbol\beta_0^{\rm T}\boldsymbol\alpha_0(\mathbf{X}_i)\right]\right\} \\ &\quad \times\left\{\boldsymbol\Theta_{\boldsymbol\beta}\boldsymbol\alpha_{c0}(\mathbf{X}_i) - \int_0^1 \boldsymbol\beta_0^{\rm T}\mathbf{X}_{ic}(t)\mathbf{w}_0^*(t)dt\right\}.\end{aligned}$$

Hence, $\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta_0)$ *converges to a normal distribution with mean* $\mathbf{0}$ *and variance* $\boldsymbol\Sigma$, *where* $\boldsymbol\Sigma \equiv \mathbf{A}^{-1}E(\mathbf{B}^{\otimes 2})\mathbf{A}^{-1^{\rm T}}$. *Here* $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\rm T}$ *for arbitrary vector or* *matrix* $\mathbf{a}$*. In addition, when* $g^* = g$*,* $\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta_0)$ *converges to a normal distribution with mean* $\mathbf{0}$ *and variance* $\{E(\mathbf{S}_{\mathrm{oeff}}^{\otimes 2})\}^{-1}$*, i.e.,* $\widehat{\boldsymbol\beta}$ *is the semiparametric efficient estimator of* $\boldsymbol\beta$ *for model (1).*

Theorem 2 indicates that $\widehat{\boldsymbol\beta}$ is the semiparametric efficient estimator in model (1) when g^* is correctly specified, even though the estimation of $\widehat{\boldsymbol\beta}$ is devised under model (2). Generally, we can replace g^* by a consistent estimator of g. The following corollary ensures the asymptotic efficiency of the resulting $\widehat{\boldsymbol\beta}$.

Corollary 1. *Assume Conditions (A1)–(A8) hold, and let* $\widehat{\boldsymbol\beta}$ *satisfy*

$$\sum_{i=1}^{n} \widehat{\mathbf{S}}_{\mathrm{eff}\boldsymbol\beta}(Y_i, \mathbf{Z}_i, \widehat{\boldsymbol\beta}, \widehat{\boldsymbol\gamma}(\widehat{\boldsymbol\beta}), \widehat{g}) = \mathbf{0},$$

where $\widehat{g}$ *is a uniformly consistent estimator for the true function* g*, and* $\widehat{\mathbf{S}}_{\mathrm{eff}\boldsymbol\beta}$ *is defined in Proposition 2, then* $\widehat{\boldsymbol\beta}$ *is semiparametric efficient.*

The corollary readily holds by using the result in Theorem 2 and the consistency of $\widehat{g}$. We omit the details. In practice, we can use kernel method to estimate $f(Y, \boldsymbol\beta^{\mathrm{T}}\mathbf{Z}\boldsymbol\gamma)$ and in turn to obtain $\widehat{g}$, which is guaranteed to be uniformly consistent to g (Mack and Silverman, 1982). Combining the results of Theorems 1 and 2, we are able to establish the theoretical properties of the estimation of $\widehat{\alpha}(t)$ in Theorem 3. Specifically, Theorem 3 shows that the spline approximation $\widehat{\alpha}(t) = \mathbf{B}_r(t)^{\mathrm{T}}\widehat{\boldsymbol\gamma}(\widehat{\boldsymbol\beta})$ with $\boldsymbol\beta, \boldsymbol\gamma$ estimated using the estimating equation set (8) indeed achieves the usual nonparametric spline regression convergence rate.

Theorem 3. *Assume Conditions (A1) – (A8) hold, then*

$$\sup_{t\in[0,1]} |\mathbf{B}_r(t)^{\mathrm{T}}\widehat{\boldsymbol\gamma}(\widehat{\boldsymbol\beta}) - \alpha_0(t)| = O_p(n^{-1/2}h_b^{-1/2}).$$

The proofs for the theoretic results are elaborated in the supplement.

3 Relation to the semiparametric sufficient dimension reduction

Although performed for the functional model and the dimension folding model, the proposed estimation procedure is inline with the semiparametric sufficient dimension reduction techniques discussed in Ma and Zhu (2012). To illustrate the similarity, following Bickel et al. (1993) and Tsiatis (2004), we first develop the nuisance tangent space $\Lambda^{\perp}$ in the following proposition, which allows us to construct estimators of $\boldsymbol{\beta}, \boldsymbol{\gamma}$ from various choices of the function $\mathbf{f}$ appeared in the description of $\Lambda^{\perp}$.

Proposition 3. *In the Hilbert space $\mathcal{H}$ of all the mean zero finite variance functions associated with (2), i.e. $\mathcal{H} = \{\mathbf{a}(\mathbf{Z}, Y) : \int \mathbf{a}(\mathbf{z}, y) f(y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}\boldsymbol{\gamma}) f_{\mathbf{Z}}(\mathbf{z}) d\mu(\mathbf{z}, y) = \mathbf{0}, \int \mathbf{a}^{\mathrm{T}}(\mathbf{z}, y)\mathbf{a}(\mathbf{z}, y) f(y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}\boldsymbol{\gamma}) f_{\mathbf{Z}}(\mathbf{z}) d\mu(\mathbf{z}, y) < \infty, \mathbf{a}(\mathbf{z}, y) \in \mathcal{R}^{d_\gamma+J-2}\}$, where $\mu(\mathbf{z}, y)$ is the probability measure of $(\mathbf{Z}, Y)$, $f_{\mathbf{Z}}(\mathbf{z})$ is the pdf of $\mathbf{Z}$ and $f(y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}\boldsymbol{\gamma})$ is given in (2), the orthogonal complement of the nuisance tangent space is*

$$\Lambda^{\perp} = \{\mathbf{f}(Y, \mathbf{Z}) - E(\mathbf{f} \mid Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) : E(\mathbf{f} \mid \mathbf{Z}) = E(\mathbf{f} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}), \forall\, \mathbf{f}\}.$$

The proof of Proposition 3 is given in the supplement. Let $\mathbf{f}(Y, \mathbf{Z}) = \left[\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) - E\left\{\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right\}\right]\mathbf{a}(\mathbf{Z})$, where $\mathbf{g}, \mathbf{a}$ can be chosen arbitrarily as long as the resulting $\mathbf{f}$ contains sufficiently many equations. Obviously, $E(\mathbf{f} \mid \mathbf{Z}) = \mathbf{0}$ hence $E(\mathbf{f} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = \mathbf{0}$, and $\mathbf{f} - E(\mathbf{f} \mid Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = \left\{\mathbf{g} - E\left(\mathbf{g} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right)\right\}\left\{\mathbf{a} - E(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\right\}$. Thus, we can construct estimating equation based on the sample version of

$$E\left([\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) - E\{\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\}][\mathbf{a}(\mathbf{Z}) - E\{\mathbf{a}(\mathbf{Z}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\}]\right) = \mathbf{0}, \quad (8)$$

and it provides a class of estimators for $\boldsymbol{\beta}, \boldsymbol{\gamma}$.

We now perform a set of analysis somewhat in the spirit of Ma and Zhu (2012) to illustrate that by different choice of $\mathbf{g}$ and $\mathbf{a}$, (8) leads to the classical dimension reduction estimators.

3.1 The relation with the sliced inverse regression

As a first choice of $\mathbf{g}$ and $\mathbf{a}$, let $\mathbf{V} = \text{vec}(\mathbf{Z})$, and select $\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = E(\mathbf{V} \mid Y), \mathbf{a}(\mathbf{Z}) = \mathbf{V}^{\mathrm{T}}$. This provides an estimator with the flavor of the sliced inverse regression (SIR, Li (1991)) in the classical dimension reduction framework. Specifically, under this choice of $\mathbf{g}, \mathbf{a}$, (8) has the form

$$E\left(\left[E(\mathbf{V} \mid Y) - E\{E(\mathbf{V} \mid \mathbf{Y}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\}\right]\left\{\mathbf{V}^{\mathrm{T}} - E(\mathbf{V}^{\mathrm{T}} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\right\}\right) = \mathbf{0}. \quad (9)$$

The above estimating equation set contains $J^2 {d_\gamma}^2$ equations while we only have $J + d_\gamma - 2$ free parameters. We can use GMM to reduce the number of equations in practice. Of course we can also construct $\mathbf{g}$ or $\mathbf{a}$ or both using only a subset of $\mathbf{V}$.

3.2 The relation with the sliced average variance estimator

As a second choice of $\mathbf{g}, \mathbf{a}$, we select $\mathbf{g}_1(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = \mathbf{I}_{Jd_\gamma} - \text{cov}(\mathbf{V} \mid Y)$, $\mathbf{g}_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = \mathbf{g}_1 E(\mathbf{V} \mid Y)$, $\mathbf{a}_1(\mathbf{Z}) = -\mathbf{V}\{\mathbf{V} - E(\mathbf{V} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}^{\mathrm{T}}$, $\mathbf{a}_2(\mathbf{Z}) = \mathbf{V}^{\mathrm{T}}$. We then construct a classical sliced average variance estimator (SAVE, Cook and Weisberg (1991)) flavored estimator based on

$$\begin{aligned} & E\left[\{\mathbf{g}_1 - E(\mathbf{g}_1 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\{\mathbf{a}_1 - E(\mathbf{a}_1 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\right] \\ & + E\left[\{\mathbf{g}_2 - E(\mathbf{g}_2 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\{\mathbf{a}_2 - E(\mathbf{a}_2 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\right] = \mathbf{0}. \end{aligned} \quad (10)$$

3.3 The relation with the directional regression

The third choice of $\mathbf{g}, \mathbf{a}$ that we would like to present is $\mathbf{g}_1(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = \mathbf{I}_{Jd_\gamma} - E(\mathbf{V}\mathbf{V}^{\mathrm{T}} \mid Y)$, $\mathbf{g}_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = E\{E(\mathbf{V} \mid Y)E(\mathbf{V}^{\mathrm{T}} \mid Y)\}E(\mathbf{V} \mid Y)$, $\mathbf{g}_3(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) = E\{E(\mathbf{V}^{\mathrm{T}} \mid Y)E(\mathbf{V} \mid Y)\}E(\mathbf{V} \mid Y)$, $\mathbf{a}_1(\mathbf{Z}) = -\mathbf{V}\{\mathbf{V} - E(\mathbf{V} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}^{\mathrm{T}}$, $\mathbf{a}_2(\mathbf{Z}) = \mathbf{a}_3(\mathbf{Z}) = \mathbf{V}^{\mathrm{T}}$. This leads to a classic directional regression (DR, Li and Wang (2007)) flavored estimator from

$$\sum_{i=1}^{3} E\left[\{\mathbf{g}_i - E(\mathbf{g}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\{\mathbf{a}_i - E(\mathbf{a}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\right] = \mathbf{0}. \quad (11)$$

The three estimators given in (9), (10) and (11) are respectively of the flavors of SIR, SAVE and DR, because if we had worked in the classical sufficient dimension reduction context, and if further equipped with the additional linearity condition and constant variance condition, the choices of $\mathbf{g}$ and $\mathbf{a}$ that led to the three estimating equations above would have further led to SIR, SAVE and DR (Ma and Zhu, 2012). Further, the choices of $\mathbf{g}, \mathbf{a}$ in (9), (10) and (11) only depend on the moments of $\mathbf{Z}$ instead of the conditional density as the one used in $\mathbf{S}_{\text{eff}}$ defined in (3). Hence, these estimators can serve as alternatives to the proposed efficient estimators when the conditional density is difficult to obtain.

4 Simulation Studies

We carry out three simulation studies in order to assess the finite sample performance of our estimation method. In each simulation, we generate 1000 data sets with the sample size $n = 500$. In Simulation 1, we set $J = 9$, $\boldsymbol{\beta} = (1, 1.2, 1.5, 0.5, -0.5, -1.5, -1.2, -1, 1)^{\mathrm{T}}$ and $\alpha_0(t) = \sin(\pi t) + 1$, $t \in [0, 1]$. We generated each component of $\mathbf{X}_i(t)$ from a uniform distribution on $(-5, 5)$. We then generated the response Y_i from a normal distribution with mean $\int_0^1 W_i(t)\alpha_0(t)dt$ and variance 1, where $W_i(t) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i(t)$. In Simulation 2, we modify Simulation 1 so that the response Y_i is generated from a normal distribution with mean $\sin\{2\int_0^1 W_i(t)\alpha_0(t)dt\} + \log[1 + \{\int_0^1 W_i(t)\alpha_0(t)dt\}^2] - 3$ and variance $0.5[1 + \{\boldsymbol{\beta}^{\mathrm{T}} \int_0^1 \mathbf{X}_i(t)\alpha_0(t)dt\}^2]^{1/5}$. The purpose of Simulation 2 is to investigate the ability of our methods in handling the nonlinear mean and variance. In Simulation 3, we set $J = 4$, $\boldsymbol{\beta} = (-0.2, -1, -1.5, 1)^{\mathrm{T}}$ and $\alpha_0(t) = 1 - 26.76t + 145.3t^2 - 227.27t^3 + 107.99t^4$. The $\mathbf{X}_i(t)$'s are generated from the models: $X_{i1}(t) = 0.66 - 4.84t + 5.12t^2 + U[-4, 6]$, $X_{i2}(t) = 0.43 - 2.95t + 3.11t^2 + U[-5, 5]$, $X_{i3}(t) = -1.61 + 10.40t - 10.85t^2 + U[-4, 4]$, $X_{i4}(t) = 0.58 - 3.59t + 3.52t^2 + U[-4, 8]$, where $U[a, b]$ denotes a random variable from the uniform distribution in the range of $[a, b]$. The response Y_i is generated from a normal distribution with mean $0.075 + 0.53(\int_0^1 W_i(t)\alpha_0(t)dt + 1.23)$ and variance 0.05. Our Simulation 3 is designed to resemble the air pollution data structure in Section 5.

We applied the semiparametric methods proposed in Section 2 to estimate both $\boldsymbol{\beta}$ and $\alpha(t)$, where $\alpha(t)$ is approximated with cubic B-spline basis functions using three equally spaced internal knots. Specifically, we implemented three estimators: the oracle, the efficient and the locally efficient estimators. In the oracle estimator, we specified $f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ using the normal pdf form, hence proposed a true $g(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ function, while estimated all the conditional expectations $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$. Note that the form of $f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ is generally unknown, hence the oracle estimator is unrealistic and is only included here as a benchmark for comparison purpose. In the efficient estimator, $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, $f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, and $f_2'(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ are estimated via nonparametric method. In the local estimator, we specified an incorrect model of $f(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, hence a misspecified $g^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ function was used, and estimated $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ nonparametrically.

The numerical performances of the estimation of $\boldsymbol{\beta}$ in Simulation 1, 2 and 3 are summarized in Table 1, 2 and 3, respectively. Based on the asymptotic results in Theorem 2, the average of the estimated standard error is obtained and the coverage of the 95% confidence interval is also provided. As expected, both the efficient and the locally efficient estimators have very small bias, the estimated variances are close to their empirical ones, and the 95% coverage are also reasonably close to the nominal level. The variances of the efficient estimators are smaller than those of the locally efficient estimators. In fact, the performance of the efficient estimators is very close to the oracle estimators. We illustrate the performance of the estimation of $\alpha_0(t)$ in Figure 1, where we show the mean estimated curves and the pointwise 90% confidence bands. The performance shown in Figure 1 is rather typical for spline approximations.

As we pointed out, in functional data analysis, a simple stacking approach is often used to study the effect of the functional covariates (Ramsay and

Silverman, 2005) in a less structured model

$$E\{Y|\mathbf{X}(t)\} = \int_0^1 \mathbf{X}(t)^{\mathrm{T}} \boldsymbol{\eta}(t) dt, \tag{12}$$

where $\boldsymbol{\eta}(t) = \{\eta_1(t), \ldots, \eta_p(t)\}^{\mathrm{T}}$. We also explained that the stacking approach is a special case of the functional single index model we study here. We thus also implemented the stacking approach and compared the two estimators in Figure 2. It is easy to see that the our estimator performs much better than the general model fitting, with the 90% confidence band much narrower. This pattern also applies to simulations 2 and 3, and we provide the corresponding plots in the supplementary document. This is of course not a surprise as the functional linear model (12) does not fully take advantage of the form in the functional index model (1).

5 Application

We now analyze the data set to study the effect of various air pollutants on the rate of death caused by CVD, which motivated this work. There has been some studies regarding whether or not air pollution forms a risk factor for CVD occurrence and death (Brook et al., 2004). Because it is unclear how the annual CVD death rate links to air pollution, we adopt the model in (1) which does not need to specify any special link function.

In the data set, four pollutants with sufficiently high quality measurements are available. Carbon monoxide (CO) is a colorless, odorless gas that can result in health problems by diminishing oxygen delivery to the body's organs and tissues. Nitrogen dioxide (NO_2) and Sulfur dioxide (SO_2) are connected with many harmful effects on the respiratory system. Both are highly reactive gases. Ground-level ozone (O_3) is made by chemical reactions between oxides of nitrogen (NO_X) and volatile organic compounds (VOC) in the presence of sunlight. A number of health problems can be caused by breathing O_3.

All pollutants were recorded on a daily basis in 108 US cities. The measurements unit is parts per billion (ppb) by volume and span the range from 1987 to 2000. 400 observations with relatively small portion of missing values are used for the analysis and each observation has 365 daily median measurements of four air pollutants, where we imputed a few missing days in some observations by linear interpolation. The dataset is organized in the NMMAPSdata package (Peng and Welty, 2004) from public sources such as the United States Environmental Protection Agency and the United States Census Bureau. Each of the four pollutants is standardized to have mean zero and variance one and the time interval is normalized to $[0, 1]$.

We fit model (1) to estimate the air pollution index directly related to the following year's CVD death rate. Throughout the implementation, we set the bandwidth used in (6) to be $h = n^{-1/5}\mathrm{range}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}_{\mathrm{i}}\boldsymbol{\gamma})$ and $b = n^{-1/7}\mathrm{range}(\mathrm{y}_{\mathrm{i}})$, where the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are replaced with their most updated estimates during each iteration. The functional parameter $\alpha(t)$ is estimated using a linear combination of cubic B-splines with three equally spaced internal knots in [0,1]. We calculated the confidence band for $\alpha(t)$ using the asymptotic results in Theorem 1. This is justified because the bias term of the estimated $\widehat{\alpha}(t)$ is of higher order according to Theorem 3 and hence can be ignored in computing confidence band.

Figure 3 shows the time-varying effect $\hat{\alpha}(t)$ of the estimated air pollution index to the CVD death rate. The time-varying effect of the air pollution index is significantly positive in the spring, summer, and fall seasons but insignificant in the winter. The air pollution index has the largest positive effect on the CVD death rate in the summer.

Figure 4 displays the air pollution index for three major cities: Boston, New York and Chicago together with their CVD annual death rates. With the largest air pollution index in the summer time, New York has the largest CVD death rate. On the other hand, Boston has the lowest air pollution index in the summer and hence the CVD death rate is the smallest in Boston, although Boston has the highest air pollution index in the winter.

Table 4 displays the estimated coefficients for all four air pollutants $\boldsymbol{\beta}$. All estimated coefficients $\widehat{\boldsymbol{\beta}}$ are statistically significant, which indicates that

CO, NO_2, O_3 and SO_2 are all significant risk factors for the air pollution index related to the CVD death rate. The estimated coefficients for CO, NO_2, and SO_2 are negative, which is caused by the correlation of these three air pollutants to O_3.

For comparison, we implemented the stacking approach to estimate the functional linear model (12). Figure 5 compares the estimated $\widehat{\eta}_k(t)$ for the stacking functional linear model (12) and the estimated $\widehat{\beta}_k\widehat{\alpha}(t)$ for our functional index model (1), where $k = 1, \ldots, 4$. While there is slight disagreement between the two sets of estimations from the two models, it is clear that the unstructured model has very large variability and can hardly deliver any statistically significant results.

We further assessed the prediction performance of our proposed method in comparison with three other methods, including the stacking functional linear model (12), the functional additive model (Müller and Yao, 2008), and a single index model where each covariate is simply the yearly average of each pollutant. The evaluation is conducted through a ten-fold cross-validation. Table 5 displays the mean squared prediction errors (MSPE) of our proposed method and the three comparison methods. It shows that our proposed functional single index model has the smallest MSPE among all four methods. For instance, MSPE is decreased by 31% when using our proposed functional single index in comparison with the stacking functional linear model (12).

6 Discussion

We proposed a functional single index model to study the relation between the pollutants and CVD death rate. The model contains a single index which summarizes the pollution severity level and a time-varying coefficient which captures the seasonality of the pollution effects. Furthermore, the model is robust against the misspecification of the conditional density function $f_{Y|\mathbf{X}(t)}(\cdot)$. Interestingly, when replacing the function $\alpha(\cdot)$ by its B-spline approximation, the model reduces to a dimension folding model, and the our estimator yields a new estimator as a by-product. This new estimator requires much relaxed conditions on the covariates while at the same time performs much better than all existing methods. Finally, the model and method can be used in the high dimensional settings thanks to the fact that the numbers of covariate functions and spline basis are added. In contrast, the traditional functional single index described in (12) would result in multiplication of these two numbers.

We conclude our work with some remarks. First, in pollution measurements and in many other problems, due to various reasons such as the imprecise or sparse measurements, only the discrete measurements for functional covariates $X_i(t)$ may be available and these measurements may further contain measurement errors. In an initial data analysis step, we can estimate the functional covariates $X_i(t)$ from the data with measurement errors by using some nonparametric smoothing methods such as smoothing splines or local polynomial regression. We then analyze the functional single index model (1) using the estimated $X_i(t)$. In our theoretical results, we assume the functional covariates $X_i(t)$ are known to simplify the problem. To properly take into account the fact that the covariate function we constructed may not be the true $X_i(t)$, the model in (1) should be further augmented with an errors in variable model $\tilde{\mathbf{X}}(t) = \mathbf{X}(t) + \mathbf{U}(t)$, where $\mathbf{U}(t)$ is measurement error, and $\tilde{\mathbf{X}}(t)$, instead of $\mathbf{X}(t)$, is observed. This will bring the model to the measurement error problem framework. Because measurement error problems in much simpler models, for example when $\alpha(t)$ does not appear at all, have not been solved or even studied thoroughly, the problem deserves much careful investigation and is beyond the scope of the paper.

Supplementary Materials.

The asymptotic properties of both the finite dimensional parameter and the function estimates, and the comprehensive proofs of the theoretic results are provided in the supplementary document online.

References

Atkinson, K. E. (1989), *An Introduction to Numerical Analysis (2nd ed.)*, New York: John Wiley & Sons.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993), *Efficient and adaptive estimation for semiparametric models*, New York: Springer.

Brook, R., Franklin, B., Cascio, W., Hong, Y., G, H., Lipsett, M., Luepker, R., Mittleman, M., Samet, J., SC, S., and Tager, I. (2004), "Air pollution and cardiovascular disease: A statement for healthcare professionals from the expert panel on populaton and prevention science of the american heart association," *Circulation*, 109, 2655–2671.

Cook, D. (1998), *Regression Graphics: Ideas for Studying Regressions through Graphics*, New York: Wiley.

Cook, R. and Weisberg, S. (1991), "Discussion of Sliced inverse regression for dimension reduction," *Journal of the American Statistical Association*, 86, 28–33.

Cox, L. A. T. and Popken, D. A. (2015), "Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States?" *Annals of epidemiology*, 25, 162–173.

Ferré, L. and Yao, A. (2003), "Functional sliced inverse regression analysis," *Statistics*, 37, 475–488.

— (2005), "Smoothed functional inverse regression," *Statistica Sinica*, 15, 665–683.

— (2007), "Reply to the paper 'A note on smoothed functional inverse regression' by Liliana Forzani and R. Dennis Cook," *Statistica Sinica*, 17, 1534–1544.

Jiang, C., Yu, W., and Wang, J.-L. (2014), "Inversed regression for longitudinal data," *Annals of Statistics*, 42, 563–591.

Li, B., Kim, M. K., and Altman, N. (2010), "On dimension folding of matrix or array valued statistical objects," *Annals of Statistics*, 38, 1094–1121.

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

Li, K. C. (1991), "Sliced inverse regression for dimension reduction (with discussion)," *Journal of the American Statistical Association*, 86, 316–342.

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

— (2013), "Efficient Estimation in Sufficient Dimension Reduction," *Annals of Statistics*, 41, 250–268.

Mack, Y. and Silverman, B. W. (1982), "Weak and strong uniform consistency of kernel regression estimates," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61, 405–415.

Müller, H. and Yao, F. (2008), "Functional Additive Models," *Journal of the American Statistical Association*, 103, 1683.

Nogueira, J. B. (2009), "Air pollution and cardiovascular disease." *Revista portuguesa de cardiologia: orgão oficial da Sociedade Portuguesa de Cardiologia= Portuguese journal of cardiology: an official journal of the Portuguese Society of Cardiology*, 28, 715–733.

Peng, R. and Welty, L. (2004), "The NMMAPSdata package," *R News*, 4, 10–14.

Ramsay, J. and Silverman, B. W. (2005), *Functional data analysis*, New York: Springer-Verlag, 2nd ed.

Samet, J., Dominici, F., Zeger, S., Schwartz, J., and Dockery, D. (2000), "The National Morbidity, Mortality, and Air Pollution Study. Part I: Methods and methodologic issues." *Research report (Health Effects Institute)*, 5–14.

Tsiatis, A. A. (2004), *Semiparametric Theory and Missing Data*, New York: Springer.

Turner, M. C., Jerrett, M., Pope III, C. A., Krewski, D., Gapstur, S. M., Diver, W. R., Beckerman, B. S., Marshall, J. D., Su, J., Crouse, D. L., et al. (2016), "Long-term ozone exposure and mortality in a large prospective study," *American journal of respiratory and critical care medicine*, 193, 1134–1142.

Table 1: The average (AVE), the sample standard deviations (STD), the average of the estimated standard deviation ($\widehat{\text{STD}}$), the square root of the mean squared error (MSE) and the coverage of the estimated 95% confidence interval (CI) from the oracle (Ora), efficient (Eff) and Locally efficient (Loc) estimates of $\boldsymbol{\beta}$,$\boldsymbol{\gamma}$ in Simulation 1.

		β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
		1	1.2	1.5	0.5	-0.5	-1.5	-1.2	-1
Ora	AVE	1.0089	1.2070	1.5134	0.5067	-0.5033	-1.5081	-1.2141	-1.0062
	STD	0.1294	0.1364	0.1588	0.1020	0.1007	0.1553	0.1412	0.1304
	$\widehat{\text{STD}}$	0.1244	0.1365	0.1577	0.0994	0.0996	0.1570	0.1373	0.1242
	MSE	0.0168	0.0186	0.0254	0.0104	0.0101	0.0242	0.0201	0.0170
	CI	0.9470	0.9520	0.9540	0.9420	0.9470	0.9530	0.9420	0.9330
Eff	AVE	1.0264	1.2279	1.5398	0.5160	-0.5122	-1.5344	-1.2347	-1.0240
	STD	0.1368	0.1460	0.1719	0.1065	0.1056	0.1669	0.1522	0.1389
	$\widehat{\text{STD}}$	0.1349	0.1495	0.1744	0.1051	0.1052	0.1734	0.1503	0.1349
	MSE	0.0194	0.0221	0.0311	0.0116	0.0113	0.0290	0.0243	0.0198
	CI	0.9640	0.9650	0.9660	0.9500	0.9510	0.9630	0.9580	0.9520
Loc	AVE	1.0335	1.2427	1.5533	0.5194	-0.5139	-1.5479	-1.2444	-1.0341
	STD	0.1535	0.1674	0.1962	0.1229	0.1192	0.1903	0.1728	0.1559
	$\widehat{\text{STD}}$	0.1544	0.1700	0.1997	0.1208	0.1207	0.1969	0.1702	0.1538
	MSE	0.0247	0.0298	0.0413	0.0155	0.0144	0.0385	0.0318	0.0254
	CI	0.9550	0.9540	0.9580	0.9520	0.9570	0.9640	0.9550	0.9600

Table 2: The average (AVE), the sample standard deviations (STD), the average of the estimated standard deviation ($\widehat{\text{STD}}$), the square root of the mean squared error (MSE) and the coverage of the estimated 95% confidence interval (CI) from the oracle (Ora), efficient (Eff) and Locally efficient (Loc) estimates of $\boldsymbol{\beta}$,$\boldsymbol{\gamma}$ in Simulation 2.

		β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
		1	1.2	1.5	0.5	-0.5	-1.5	-1.2	-1
Ora	ave	1.0037	1.2019	1.5031	0.5015	-0.5030	-1.5015	-1.2001	-1.0014
	std	0.0611	0.0677	0.0752	0.0503	0.0493	0.0750	0.0682	0.0568
	$\widehat{\text{std}}$	0.0603	0.0662	0.0764	0.0484	0.0485	0.0763	0.0662	0.0604
	mse	0.0037	0.0046	0.0057	0.0025	0.0024	0.0056	0.0046	0.0032
	CI	0.9500	0.9430	0.9430	0.9370	0.9480	0.9550	0.9310	0.9580
Eff	ave	1.0027	1.2018	1.5042	0.5024	-0.5037	-1.5005	-1.2004	-1.0011
	std	0.0752	0.0835	0.0930	0.0571	0.0564	0.0966	0.0801	0.0718
	$\widehat{\text{std}}$	0.0734	0.0805	0.0931	0.0579	0.0578	0.0937	0.0805	0.0730
	mse	0.0057	0.0070	0.0087	0.0033	0.0032	0.0093	0.0064	0.0051
	CI	0.9460	0.9460	0.9470	0.9470	0.9500	0.9410	0.9470	0.9500
Loc	ave	0.9944	1.1907	1.4877	0.4963	-0.4973	-1.4868	-1.1889	-0.9913
	std	0.1494	0.1720	0.2098	0.0830	0.0842	0.2126	0.1736	0.1426
	$\widehat{\text{std}}$	0.1571	0.1855	0.2282	0.0893	0.0895	0.2296	0.1854	0.1568
	mse	0.0223	0.0296	0.0441	0.0069	0.0071	0.0453	0.0302	0.0204
	CI	0.9440	0.9400	0.9440	0.9530	0.9470	0.9460	0.9430	0.9460

Table 3: The average (AVE), the sample standard deviations (STD), the average of the estimated standard deviation ($\widehat{\text{STD}}$), the square root of the mean squared error (MSE) and the coverage of the estimated 95% confidence interval (CI) from the oracle, efficient and Locally efficient estimates of $\boldsymbol{\beta}$,$\boldsymbol{\gamma}$ in Simulation 3.

		β_1	β_2	β_3
	TRUE	-0.2	-1.0	-1.5
Oracle	AVE	-0.2009	-1.0015	-1.5005
	STD	0.0497	0.0650	0.0842
	$\widehat{\text{STD}}$	0.0493	0.0634	0.0860
	MSE	0.0025	0.0042	0.0071
	CI	0.9520	0.9480	0.9520
Efficient	AVE	-0.2017	-1.0057	-1.5058
	STD	0.0502	0.0662	0.0851
	$\widehat{\text{STD}}$	0.0497	0.0642	0.0871
	MSE	0.0025	0.0044	0.0071
	CI	0.9480	0.9440	0.9540
Locally	AVE	-0.2020	-0.9893	-1.4849
Efficient	STD	0.0769	0.1002	0.1246
	$\widehat{\text{STD}}$	0.0790	0.1069	0.1508
	MSE	0.0059	0.0101	0.0157
	CI	0.9630	0.9420	0.9530

Table 4: The estimated coefficients for the air pollutants CO, NO_2, SO_2 and standard errors for the functional single index model (1) in the air pollution data using the efficient method. The coefficient for O_3 is fixed to be 1 for identifiability, as introduced in Section 2.1.

	$\widehat{\beta}_1$ (CO)	$\widehat{\beta}_2$ (NO_2)	$\widehat{\beta}_3$ (SO_2)	β_4 (O_3)
Coefficients	-0.286	-0.971	-1.833	1.000
Standard Errors	0.080	0.006	0.002	-
p-values	3e-4	$<$5e-5	$<$5e-5	-

Table 5: The mean squared prediction errors of the four methods for the CVD death rate.

Methods	Mean Squared Prediction Errors
Functional single index Model (1)	2.14×10^{-6}
Stacking functional linear model (12)	3.11×10^{-6}
Functional additive model	2.56×10^{-6}
Single index model	2.44×10^{-6}

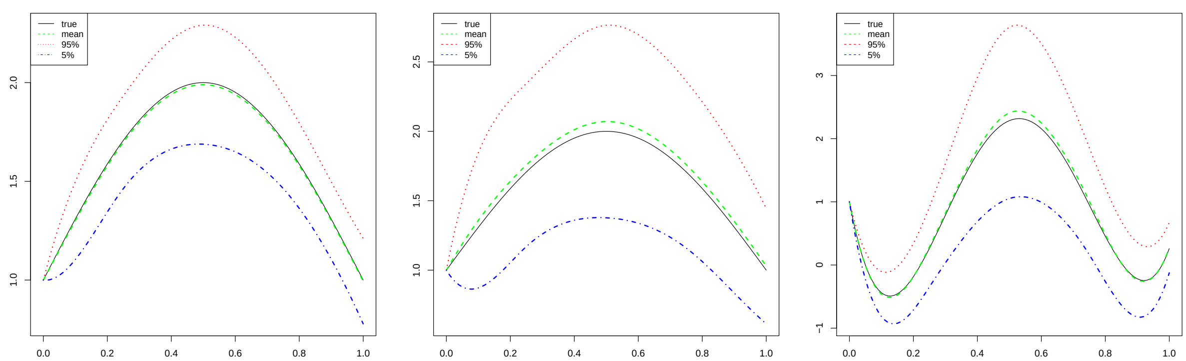

Figure 1: The mean and point wise 90% confidence bands of the estimated $\widehat{\alpha}(t)$ for the functional single index model (1) in Simulations 1 (left), 2 (middle) and 3 (right). The true $\alpha_0(t)$ is plotted in the solid curve.

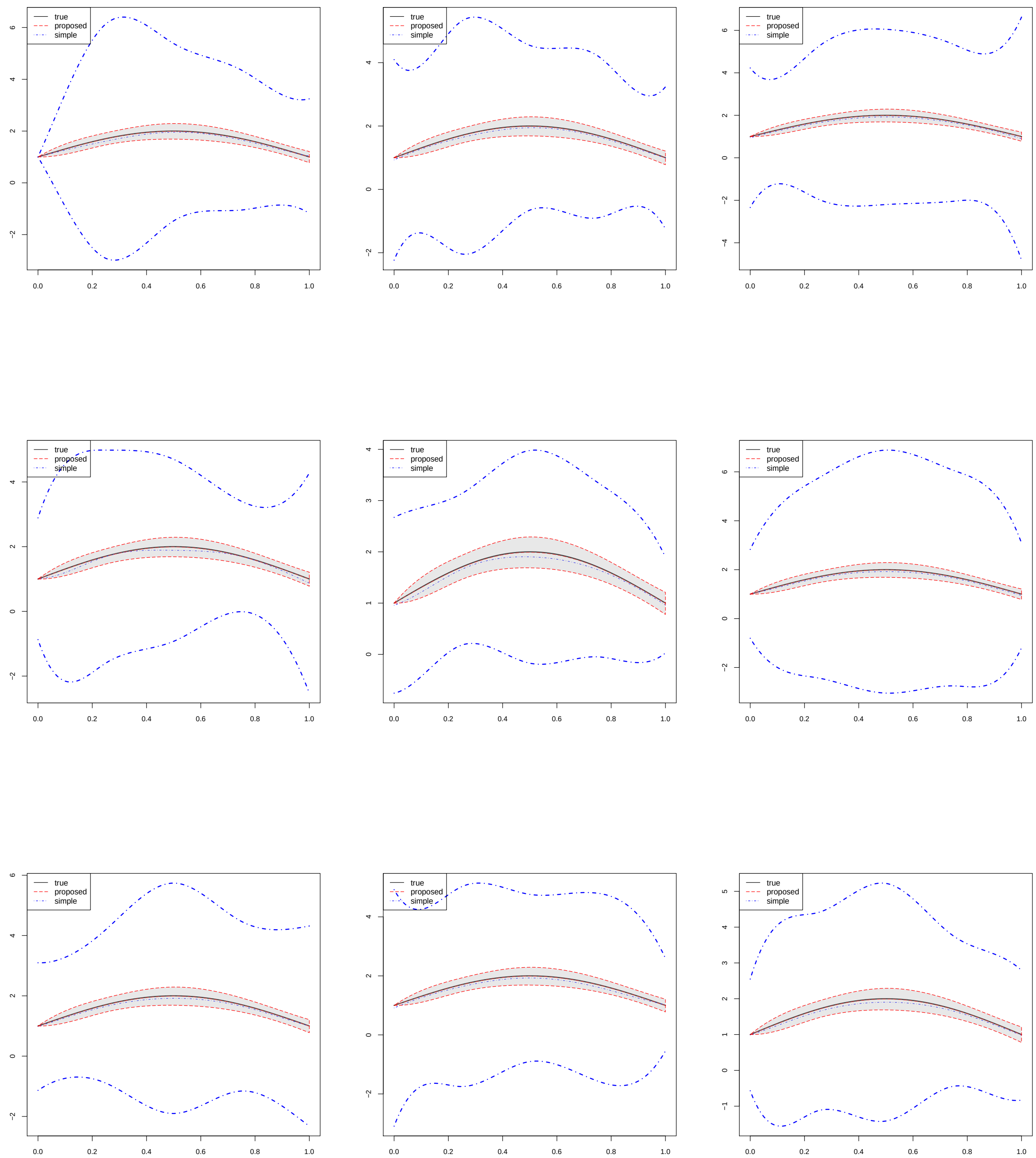

Figure 2: The estimated $\beta_k \alpha(t), k = 1, \ldots, 9$, and their point-wise 90% confidence bands for the proposed functional single index model (1), in comparison with the estimated $\widehat{\eta}_k(t)$ for the simple stacking functional linear model (12) in Simulation 1.

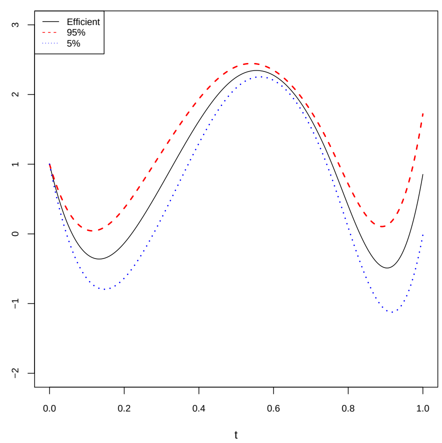

Figure 3: The estimated $\widehat{\alpha}(t)$ for the functional single index model (1) from the air pollution data. It captures the time-varying effect of the air pollution index on the annual CVD death rate. The pointwise 90% confidence band of the estimated $\widehat{\alpha}(t)$ is also provided.

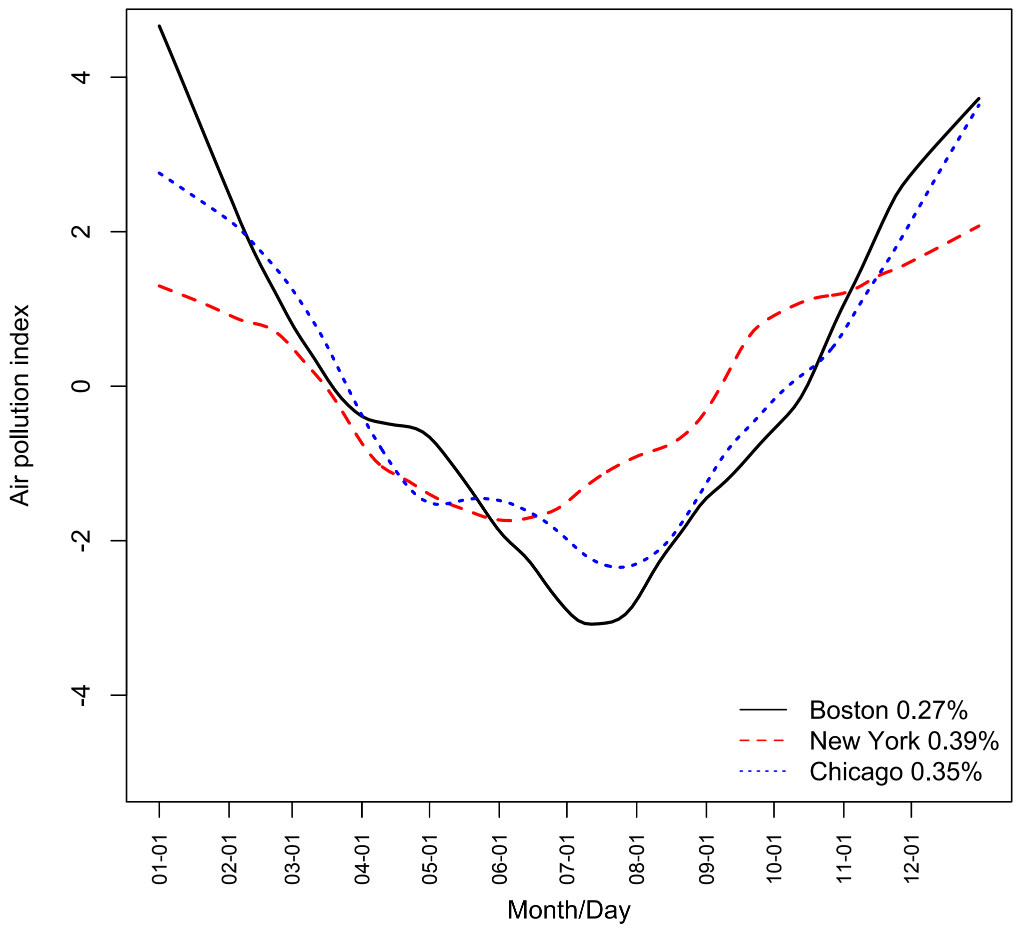

Figure 4: The pollution indices for Boston, New York and Chicago. The CVD death rates are shown in the legend.

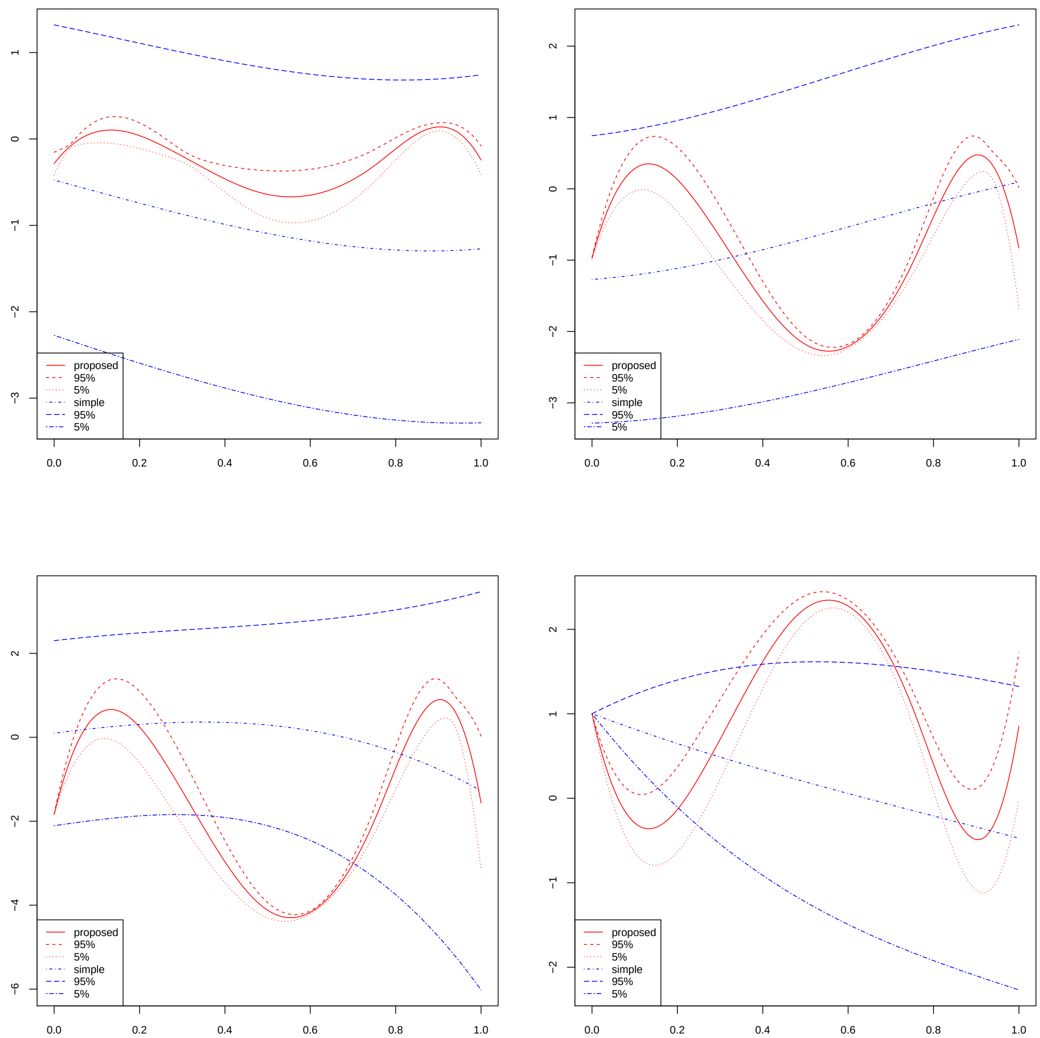

Figure 5: Comparison of the estimated $\widehat{\beta}_k\widehat{\alpha}(t)$ in the proposed functional single index model (1) with their 90% confidence bands, and the estimated $\widehat{\eta}_k(t)$ with their 90% confidence bands in the simple stacking functional linear model (12), for $k = 1, \ldots, 4$. The top left panel is $\widehat{\beta}_1\widehat{\alpha}(t)$ and $\widehat{\eta}_1(t)$, the top right panel is $\widehat{\beta}_2\widehat{\alpha}(t)$ and $\widehat{\eta}_2(t)$, the bottom left panel is $\widehat{\beta}_3\widehat{\alpha}(t)$ and $\widehat{\eta}_3(t)$, and the bottom right panel is $\widehat{\beta}_4\widehat{\alpha}(t)$ and $\widehat{\eta}_4(t)$.