A FUNCTIONAL SINGLE-INDEX MODEL

Fei Jiang, Seungchul Baek, Jiguo Cao and Yanyuan Ma

*University of Hong Kong, University of Maryland, Baltimore County,
Simon Fraser University and Penn State University*

Abstract: We propose a semiparametric functional single-index model for studying
the relationship between a univariate response and multiple functional covariates.
The parametric part of the model integrates a functional linear regression model
and a sufficient dimension-reduction structure. The nonparametric part of the
model allows the response-index dependence or the link function to be unspecified.
The B-spline method is used to approximate the coefficient function, which leads
to a dimension-folding-type model. A new kernel regression method is developed to
handle the dimension-folding model, allowing us to estimate the index vector and
the B-spline coefficients efficiently. We also establish the asymptotic properties and
semiparametric optimality for the estimators.

Key words and phrases: B-spline, dimension folding, dimension reduction, functional data analysis, functional linear model, kernel estimation.

1. Introduction

The National Morbidity, Mortality, and Air Pollution Study (NMMAP) is
an important study aiming to address the uncertainty in the association between
pollution and health (Samet et al. (2000)). In this study, daily measurements of
the air pollutants 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 year.
The annual death rate caused by cardiovascular disease (CVD) is also collected
for these cities. Each pollutant has been studied individually in the past with
no significant effect detected on the CVD death rate (Cox and Popken (2015);
Turner et al. (2016)).

This motivates us to develop a single air pollution index that combines pol-
lutants in a way that best describes the severity of the overall air pollution level
in terms of the CVD death rate. At the same time, we examine the possible
time-varying effect of the single air pollution index on the CVD death rate. To
achieve these goals, we propose a functional single-index model and proceed to
devise a novel class of estimators.

More specifically, the NMMAP data contain measurements of 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 severity of the air pollution, we define the single air pollution index as follows:

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

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^{\mathrm{T}}$ is a vector of weights for various air pollutants, capturing the relative importance of the pollutants in determining the pollution severity. We assume that Y is linked to $W(t)$ through

$$f\{Y|\mathbf{X}(t)\} = f\left\{Y, \int_0^1 W(t)\alpha(t)dt\right\}, \tag{1.1}$$

where f is a conditional probability density function (pdf) or probability mass function (pmf); and is left unspecified. Here, the functional parameter $\alpha(t)$ captures the time-varying effect of the air pollution index on the annual CVD death rate. Note that: unlike the usual single-index model, as discussed in Chen, Hall and Müller (2011) and Ma (2016), we make an assumption on the conditional distribution, rather than on the conditional mean.

Estimations and inferences using the functional single-index model (1.1) are not simple. The complexity is due to the unspecified bivariate link function f, unknown coefficient function $\alpha(t)$, and unknown index vector $\boldsymbol{\beta}$. If $\alpha(t)$ were known, (1.1) would reduce to a central space estimation problem; here various methods exist to estimate $\boldsymbol{\beta}$ (Li (1991); Cook and Weisberg (1991); Li and Wang (2007); Ma and Zhu (2013)). If $\boldsymbol{\beta}$ were known, (1.1) would reduce to a functional dimension reduction problem (Ferré and Yao (2003, 2005, 2007)). Qu, Wang and Wang (2016) handle the unknown $\alpha(t)$ using the reproducing kernel method, where $\alpha(t)$ is approximated by a function in the reproducing kernel space. Based on a similar idea, we approximate $\alpha(t)$ by a spline function, which facilitates the estimation and inference procedures on $\alpha(t)$.

The functional single-index model (1.1) is closely related to sufficient dimension-reduction modeling, where a response depends on the covariate vector through its linear transformation (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. Moreover, the proposed method forms an alternative solution for the dimension-folding problem (Li, Kim and Altman (2010)), and does not require conditions on the covariates; as other methods in the literature do.

In summary, the proposed model and estimators have the following features. First, the model contains a single air pollution index that summarizes the pollution severity level. Second, the time-varying coefficient helps to provide timely adjusted health advice to the general public. Third, the flexible relation between the CVD death rate and the overall pollution effect avoids possible model misspecification. Fourth, the model extends the sufficient dimension-reduction model to handle multi-functional covariates.

2. Methodology

2.1. Model and identifiability

The identifiability of model (1.1) is shown in Proposition 1, which is justified in the Supplementary Material.

Proposition 1. *Assume $\alpha(t)$ is continuous and set $\alpha(0) = 1$ and $\beta_J = 1$. Then model (1.1) is identifiable.*

Under the assumptions in Proposition 1, we approximate $\alpha(t)$ using a B-spline function $\sum_{k=1}^{d_\gamma} \gamma_k B_{rk}(t)$, where $B_{rk}(t)$ is the kth B-spline basis function, with order r $(r \geq 0)$. Then, the functional single-index model (1.1) reduces to the model

$$f(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) = f \left\{ Y, \sum_{j=1}^{J} \sum_{k=1}^{d_\gamma} \beta_j \gamma_k \int_0^1 X_j(t) B_{rk}(t) dt \right\}, \qquad (2.1)$$

where $\mathbf{Z}$ is a $J \times d_\gamma$ matrix with (j, k)th entry $Z_{jk} \equiv \int_0^1 X_j(t) B_{rk}(t) dt$, and f is an unknown density function. Next, we 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 $\gamma_1 = 1$.

Model (2.1) is in the form of the dimension-folding models described in Li, Kim and Altman (2010); in which the predictors (i.e., $\mathbf{Z}$) are matrix-valued. The covariate matrix is sandwiched between the left and right coefficient vectors, that is $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in our setting, to generate a univariate quantity. The dimension-folding structure reduces the number of parameters of interest. In our setting, we need to estimate $d_\gamma + J - 2$ free coefficients, owing to the multiplication from both the left and the right sides of $\mathbf{Z}$. The dimension-folding model has two main advantages over the standard single-index model (Li (1991); Cook and Weisberg (1991); Li and Wang (2007)), where the matrix covariates $\mathbf{Z}$ are vectorized and

$\beta_j\gamma_k$, for $j = 1,\ldots,J$, $k = 1,\ldots,d_\gamma$, are Jd_γ new coefficients (Ma (2016)): (1) it automatically takes into account the relations between the coefficients; and (2) it reduces the dimension of the unknown coefficients from Jd_γ to $d_\gamma + J - 2$, hence avoiding possible extra high-dimensional problems when the number of covariates is merely moderately large.

As an improvement to the dimension-folding method, our proposed estimation procedure relaxes the additional constraints on the covariate matrix $\mathbf{Z}$: our procedure does not require that $E\{\mathbf{X} \mid (\boldsymbol{\gamma} \otimes \boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})\}$ be a linear function of $(\boldsymbol{\gamma}\otimes\boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})$, or that $\mathrm{var}\{\mathbf{X} \mid (\boldsymbol{\gamma}\otimes\boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})\}$ not depend on $(\boldsymbol{\gamma}\otimes\boldsymbol{\beta})^{\mathrm{T}}\mathrm{vec}(\mathbf{X})$, as enforced in Li, Kim and Altman (2010). Here, $\otimes$ stands for the Kronecker product, and $\mathrm{vec}(\mathbf{X})$ is the vector formed by concatenating the columns of $\mathbf{X}$. These constraints may be violated, and are not assumed to hold for model (2.1), in general.

2.2. Doubly robust local efficient score

To estimate the parameters, we first derive the analytic form of the efficient score. Let

$$\mathbf{S}_{\mathrm{eff}\boldsymbol{\beta}}(Y,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\gamma},g) = \left[g(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) - E\left\{g(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right\}\right] \quad (2.2)$$
$$\times(\mathbf{I}_{J-1},\mathbf{0})\{\mathbf{Z} - E(\mathbf{Z} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\boldsymbol{\gamma},$$
$$\mathbf{S}_{\mathrm{eff}\boldsymbol{\gamma}}(Y,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\gamma},g) = \left[g(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) - E\left\{g(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\right\}\right]$$
$$\times(\mathbf{0},\mathbf{I}_{d_\gamma-1})\{\mathbf{Z} - E(\mathbf{Z} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}^{\mathrm{T}}\boldsymbol{\beta},$$

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. Then, the efficient score is $\mathbf{S}_{\mathrm{eff}}(Y_i,\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\gamma},g) \equiv \{\mathbf{S}_{\mathrm{eff}\boldsymbol{\beta}}(Y,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\gamma},g)^{\mathrm{T}}, \mathbf{S}_{\mathrm{eff}\boldsymbol{\gamma}}(Y,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\gamma},g)^{\mathrm{T}}\}^{\mathrm{T}}$. Our hope is to use the efficient score to construct an estimating equation that we can use to solve for $\boldsymbol{\beta},\boldsymbol{\gamma}$ from

$$\sum_{i=1}^{n}\mathbf{S}_{\mathrm{eff}}(Y_i,\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\gamma},g) = \mathbf{0}. \quad (2.3)$$

When $g(\cdot)$ and $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ are correctly specified, $\mathbf{S}_{\mathrm{eff}}$ is indeed a function that falls in the space orthogonal to the nuisance tangent space induced by the unknown conditional density $f(\cdot)$ defined in Proposition 3. Hence, as shown in Bickel et al. (1993); Tsiatis (2004), $\mathbf{S}_{\mathrm{eff}}$ is the efficient score (see Proposition S4.1 in the Supplementary Material) that yields the optimal estimators with the smallest asymptotic variances. In addition, $\mathbf{S}_{\mathrm{eff}}$ is a doubly robust function, such 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 (Ma and Zhu (2012, 2013)).

In reality, the functional form for the conditional density $f(\cdot)$ is usually unknown, making it difficult to obtain $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ and $f_2'(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})/f(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ in (2.2). Hence, the efficient score cannot be used directly. To retain the best estimation efficiency without imposing additional assumptions, we adopt a nonparametric estimation to estimate the unknown components in the efficient score function. Specifically, we use the standard kernel smoothing method in a nonparametric regression to estimate $E(\cdot \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, that is,

$$\widehat{E}\{m(Y_i,\mathbf{Z}_i) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}\} = \frac{\sum_{i=1}^n m(Y_i,\mathbf{Z}_i)K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})}{\sum_{i=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})}, \qquad (2.4)$$

for an 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 use a local linear estimator. Specifically, we obtain the estimators $\widehat{f}(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}) = c_0$ and $\partial\widehat{f}(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma})/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}) = c_1$ by minimizing

$$\sum_{j=1}^{J}\{K_b(Y_j - Y) - c_0 - c_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma})\}^2 K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j\boldsymbol{\gamma} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}) \quad (2.5)$$

with respect to c_0, c_1. (2.4) and (2.5) allow us to obtain the unknown quantities in $\mathbf{S}_{\mathrm{eff}}$ consistently, given any parameter $\boldsymbol{\beta}, \boldsymbol{\gamma}$.

This is clearly a profiling procedure, where unknown nuisance components are estimated as functions of the parameters of interest, and then the estimating equations are solved to obtain the final estimator. This procedure yields the optimal estimator for $\boldsymbol{\beta}, \boldsymbol{\gamma}$, but requires a relatively heavy computation, especially when solving (2.5) for each value of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j\boldsymbol{\gamma}$. Thus, when the estimation variability is not of great concern, to ease the computation burden, we may aim for a possibly nonoptimal estimator. Specifically, we posit working models for f and f_2, say f^* and f_2^*. Let $g^* = f_2^{*\prime}/f^*$; then, $\mathbf{S}_{\mathrm{eff}}(Y_i,\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\gamma},g^*)$ is a locally efficient score function. Using this function to construct estimating equations guarantees estimation consistency, and can result in an efficient estimator when g^* is the truth.

Note that there is a difficulty in obtaining $\mathbf{Z}_i$. Unlike in the usual dimension-reduction problems, $\mathbf{Z}_i$ is not directly observed and, thus, needs to be constructed from the observed $X_j(t)$. 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

$$\int_0^1 B_{rk}(t)X_j(t)dt = \frac{1}{3Q}\left[B_{rk}(t_0)X_j(t_0) + 2\sum_{q=1}^{Q/2-1}\{B_{rk}(t_{2q})X_j(t_{2q})\}\right.$$

$$\left. + 4\sum_{q=1}^{Q/2}\{B_{rk}(t_{2q-1})X_j(t_{2q-1})\} + B_{rk}(t_Q)X_j(t_Q)\right],$$

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

The estimation procedures can be summarized as follows:

Step 1: Choose f and f_2' by minimizing (2.5) or positing f and f_2'. Denote the choices by $f^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and $f_2'^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, respectively, and let $g^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = f_2'^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\gamma})/f^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

Step 2: Replace g in (2.3) with g^*, according to the **Step 1** choice.

Step 3: Let $\widehat{\mathbf{S}}_{\mathrm{eff}\boldsymbol{\gamma}}$ be the version of $\mathbf{S}_{\mathrm{eff}\boldsymbol{\gamma}}$ 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 (2.4). Treating $\boldsymbol{\gamma}$ as a function of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\gamma}(\boldsymbol{\beta})$, we solve

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

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

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

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

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

In the algorithm, we use the last two arguments in $g^*(Y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ to emphasize its possible dependence on $\boldsymbol{\beta}, \boldsymbol{\gamma}$. Obviously, when we posit a model for f, the functional form does not have to depend on $\boldsymbol{\beta}, \boldsymbol{\gamma}$. However, when we estimate the model f, the functional form certainly depends on $\boldsymbol{\beta}, \boldsymbol{\gamma}$, as in all profiling estimators. The resulting estimators are consistent, as discussed in Proposition 2, because the expectations of the score functions have a zero mean when the parameters are specified correctly. When estimating $\widehat{E}(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, we use the variance of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma}$ times $n^{-1/5}$ as the bandwidth. Our results are robust in the range between half of this bandwidth to double the bandwidth.

2.3. Asymptotic results

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

holds, by the following proposition.

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

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

where

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

$$= \left\{ g^*(Y_i, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}}) - \frac{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})g^*(Y_j, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}})}{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})} \right\} \boldsymbol{\Theta}_{\boldsymbol{\beta}}$$

$$\times \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})\mathbf{Z}_j}{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})} \right\} \widehat{\boldsymbol{\gamma}},$$

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

$$= \left\{ g^*(Y_i, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}}) - \frac{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})g^*(Y_j, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}})}{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})} \right\} \boldsymbol{\Theta}_{\boldsymbol{\gamma}}$$

$$\left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})\mathbf{Z}_j}{\sum_{j=1}^{J} K_h(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i\widehat{\boldsymbol{\gamma}})} \right\}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}.$$

Let $\boldsymbol{\beta}_0$ *be the true* $\boldsymbol{\beta}$. *Further, let* $\boldsymbol{\gamma}_0$ *be a spline coefficient satisfying* $\sup_{t\in[0,1]} |\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\gamma}_0 - \alpha_0(t)| = O_p(h_b^q)$, *as stated in Condition (A5). Then,* $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$, $\sup_{t\in[0,1]} |\mathbf{B}_r(\cdot)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}} - \mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_0| = o_p(1)$.

In Step 3, at the point of convergence, we show that $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0)$ achieves the non-parametric spline regression convergence rate and derive its asymptotic variation, as follows.

Theorem 1. *Assume Conditions* (A1)–(A8) *hold, and let* $\mathbf{B}_r(\cdot)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0)$ *satisfy*

$$\sum_{i=1}^{n} \widehat{\mathbf{S}}_{\mathrm{eff}\boldsymbol{\gamma}}(Y_i, \mathbf{Z}_i, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0), g^*) = \mathbf{0}.$$

Then, $n^{1/2}\{\widehat{\boldsymbol{\gamma}}^-(\boldsymbol{\beta}_0) - \boldsymbol{\gamma}_0^-\} = \mathbf{L} + o_p(\mathbf{L})$, *where*

$$\mathbf{L} = -\bigg(\bigg\{ E\bigg(\boldsymbol{\Theta}_{\boldsymbol{\gamma}} \left[\mathbf{Z}_i - E\{\mathbf{Z}_i | \boldsymbol{\beta}_0 \alpha_0(\mathbf{X}_i)\} \right]^{\mathrm{T}} \boldsymbol{\beta}_0$$

$$\times \frac{\partial \left(g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\alpha_0(\mathbf{X}_i)\} - E[g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\alpha_0(\mathbf{X}_i)\} | \boldsymbol{\beta}_0^{\mathrm{T}}\alpha_0(\mathbf{X}_i)] \right)}{\partial \boldsymbol{\beta}_0^{\mathrm{T}}\alpha_0(\mathbf{X}_i)}$$

$$\times \boldsymbol{\beta}_0^{\mathrm{T}}[\mathbf{Z}_i - E\{\mathbf{Z}_i | \boldsymbol{\beta}_0^{\mathrm{T}}\alpha_0(\mathbf{X}_i)\}]\boldsymbol{\Theta}_{\boldsymbol{\gamma}}^{\mathrm{T}} \bigg) \bigg\}^{-1} \bigg) \bigg(n^{-1/2}\sum_{i=1}^{n} [g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\alpha_0(\mathbf{X}_i)\}$$

$$-E\{g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}|\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}]\boldsymbol{\Theta}_{\boldsymbol{\gamma}}\left[\mathbf{Z}_i - E\{\mathbf{Z}_i|\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}\right]^{\mathrm{T}}\boldsymbol{\beta}_0\bigg).$$

Here, $\boldsymbol{\gamma}_0^- = (\gamma_{02}, \ldots, \gamma_{0d_\gamma})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\gamma}}^-(\boldsymbol{\beta}_0) = (\widehat{\gamma}_2(\boldsymbol{\beta}_0), \ldots, \widehat{\gamma}_{d_\gamma}(\boldsymbol{\beta}_0))^{\mathrm{T}}$. Further, for an arbitrary $d_{\gamma-1}$-dimensional vector with $\|\mathbf{a}\|_2 < \infty$, we have $\mathbf{a}^{\mathrm{T}}\{\widehat{\boldsymbol{\gamma}}^-(\boldsymbol{\beta}_0) - \boldsymbol{\gamma}_0^-\} = O_p\{(nh_b)^{-1/2}\}$.

In addition, we show $\widehat{\boldsymbol{\beta}}$ from Step 4 is not only root-n consistent, but is also efficient and achieves the information lower bound $\{E(\mathbf{S}_{\mathrm{oeff}}^{\otimes 2})\}^{-1}$. Here, $\mathbf{S}_{\mathrm{oeff}}$ is the efficient score for $\boldsymbol{\beta}$ in the original model (1.1), which contains $\alpha(\cdot)$ instead of $\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}$; Hence, it is different from $\mathbf{S}_{\mathrm{eff}\boldsymbol{\beta}}$. The precise expression is given in Proposition S4.2 in the Supplementary Material.

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

$$\Delta g_c^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}$$
$$\equiv \frac{\partial[g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\} - E\{g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}|\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}]}{\partial\{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}},$$

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

$$\boldsymbol{\Theta}_{\boldsymbol{\beta}}E[\boldsymbol{\alpha}_{c0}(\mathbf{X}_i)\Delta g_c^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}\mathbf{X}_{ic}^{\mathrm{T}}(t)]\boldsymbol{\beta}_0 \qquad (2.6)$$
$$= \int_0^1 E[\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_{ic}(s)\Delta g_c^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}\mathbf{X}_{ic}^{\mathrm{T}}(t)\boldsymbol{\beta}_0]\mathbf{w}_0^*(s)ds,$$

where $\boldsymbol{\alpha}_{c0}(\mathbf{X}) \equiv \boldsymbol{\alpha}_0(\mathbf{X}) - E\{\boldsymbol{\alpha}_0(\mathbf{X}) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X})\}$. We have the following results.

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

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

Then, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{A}^{-1}\mathbf{B} + o_p(1)$, *where*

$$\mathbf{A} = -E\left[\Delta g_c^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}\{\boldsymbol{\Theta}_{\boldsymbol{\beta}}\boldsymbol{\alpha}_{c0}(\mathbf{X}_i)\}^{\otimes 2}\right.$$
$$\left. -\Delta g_c^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}\int_0^1 \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_{ic}(t)\mathbf{w}_0^*(t)dt\boldsymbol{\alpha}_{c0}(\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\Theta}_{\boldsymbol{\beta}}^{\mathrm{T}}\right]$$

and

$$\mathbf{B} = n^{-1/2}\sum_{i=1}^n \left\{g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\} - E\left[g^*\{Y_i, \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\}|\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\alpha}_0(\mathbf{X}_i)\right]\right\}$$
$$\times \left\{\boldsymbol{\Theta}_{\boldsymbol{\beta}}\boldsymbol{\alpha}_{c0}(\mathbf{X}_i) - \int_0^1 \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_{ic}(t)\mathbf{w}_0^*(t)dt\right\}.$$

Hence, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *converges to a normal distribution with mean* $\mathbf{0}$ *and variance*

$\boldsymbol{\Sigma}$, *where* $\boldsymbol{\Sigma} \equiv \mathbf{A}^{-1}E(\mathbf{B}^{\otimes 2})\mathbf{A}^{-1^{\mathrm{T}}}$. *Here,* $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ *for an 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* $\{E(\mathbf{S}_{\mathrm{oeff}}^{\otimes 2})\}^{-1}$; *that is,* $\widehat{\boldsymbol{\beta}}$ *is the semiparametric efficient estimator of* $\boldsymbol{\beta}$ *for model* (1.1).

Theorem 2 indicates that $\widehat{\boldsymbol{\beta}}$ is a consistent estimator. Furthermore, it is semiparametric efficient when g^* is correctly specified, even though the estimation of $\widehat{\boldsymbol{\beta}}$ is devised under the approximate model (2.1). In general, we can replace g^* with 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 from the result in Theorem 2 and the consistency of $\widehat{g}$. We omit the details. In practice, we can use the 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 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 (3.1), 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 theoretical results are provided in the Supplementary Material.

3. Relation to Semiparametric Sufficient Dimension Reduction

Although performed for the functional model and the dimension-folding model, the proposed estimation procedure is in line 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 $\mathbf{f}$ in the description of $\Lambda^\perp$.

Proposition 3. *In the Hilbert space $\mathcal{H}$ of all mean-zero finite-variance functions associated with (2.1), that is, $\mathcal{H} = \{\mathbf{a}(\mathbf{Z}, Y) : \int \mathbf{a}(\mathbf{z}, y) f(y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{z} \boldsymbol{\gamma}) f_{\mathbf{Z}}(\mathbf{z}) d\mu(\mathbf{z}, y) = \mathbf{0}, \int \mathbf{a}^{\mathrm{T}}(\mathbf{z}, y) \mathbf{a}(\mathbf{z}, y) f(y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{z} \boldsymbol{\gamma}) f_{\mathbf{Z}}(\mathbf{z}) d\mu(\mathbf{z}, y) < \infty, \mathbf{a}(\mathbf{z}, y) \in \mathcal{R}^{d_\gamma + J - 2}\}$, 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 (2.1), 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 Supplementary Material. Let $\mathbf{f}(Y, \mathbf{Z}) = [\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) - E\{\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}\}] \mathbf{a}(\mathbf{Z})$, 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 $\mathbf{f} - E(\mathbf{f} \mid Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) = \{\mathbf{g} - E(\mathbf{g} \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma})\} \{\mathbf{a} - E(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma})\}$. Thus, we can construct an estimating equation based on the sample version of

$$E\left([\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) - E\{\mathbf{g}(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}\}][\mathbf{a}(\mathbf{Z}) - E\{\mathbf{a}(\mathbf{Z}) \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}\}]\right)$$
$$= \mathbf{0}, \tag{3.1}$$

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

We now perform a set of analyses, somewhat in the spirit of Ma and Zhu (2012), to illustrate that a different choice of $\mathbf{g}$ and $\mathbf{a}$ in (3.1) leads to the classical dimension-reduction estimators.

3.1. The relation with a sliced inverse regression

As a first choice of $\mathbf{g}$ and $\mathbf{a}$, let $\mathbf{V} = \mathrm{vec}(\mathbf{Z})$, 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 similar to a sliced inverse regression (SIR, Li (1991)) in the classical dimension-reduction framework. Specifically, under this choice of $\mathbf{g}, \mathbf{a}$, (3.1) has the form

$$E\left([E(\mathbf{V} \mid Y) - E\{E(\mathbf{V} \mid \mathbf{Y}) \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}\}]\{\mathbf{V}^{\mathrm{T}} - E(\mathbf{V}^{\mathrm{T}} \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma})\}\right) = \mathbf{0}. \tag{3.2}$$

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

3.2. The relation with a sliced average variance estimator

As a second choice of $\mathbf{g}, \mathbf{a}$, we select $\mathbf{g}_1(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) = \mathbf{I}_{J d_\gamma} - \mathrm{cov}(\mathbf{V} \mid Y)$, $\mathbf{g}_2(Y, \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma}) = \mathbf{g}_1 E(\mathbf{V} \mid Y)$, $\mathbf{a}_1(\mathbf{Z}) = -\mathbf{V}\{\mathbf{V} - E(\mathbf{V} \mid \boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\gamma})\}^{\mathrm{T}}$, and $\mathbf{a}_2(\mathbf{Z}) =$

$\mathbf{V}^{\mathrm{T}}$. We then construct a classical sliced average variance estimator (SAVE, Cook and Weisberg (1991)), based on

$$E\left[\{\mathbf{g}_1 - E(\mathbf{g}_1 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\{\mathbf{a}_1 - E(\mathbf{a}_1 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\right] \tag{3.3}$$
$$+E\left[\{\mathbf{g}_2 - E(\mathbf{g}_2 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\{\mathbf{a}_2 - E(\mathbf{a}_2 \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\right] = \mathbf{0}.$$

3.3. The relation with a directional regression

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

$$\sum_{i=1}^{3} E\left[\{\mathbf{g}_i - E(\mathbf{g}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\{\mathbf{a}_i - E(\mathbf{a}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})\}\right] = \mathbf{0}. \tag{3.4}$$

The three estimators given in (3.2), (3.3) and (3.4) are similar to the SIR, SAVE, and DR, respectively. This is 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)). Furthermore, the choices of $\mathbf{g}, \mathbf{a}$ in (3.2), (3.3), and (3.4) depend only on the moments of $\mathbf{Z}$, instead of on the conditional density, as used in $\mathbf{S}_{\mathrm{eff}}$ defined in (2.2). 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 under the following settings in order to assess the finite-sample performance of our estimation method. In each simulation, we generate 1,000 data sets with a sample size $n = 500$.

Simulation 1

(1) $J = 9$, $\boldsymbol{\beta} = (1, 1.2, 1.5, 0.5, -0.5, -1.5, -1.2, -1, 1)^{\mathrm{T}}$, and $\alpha_0(t) = \sin(\pi t) + 1$, for $t \in [0, 1]$;

(2) $X_{ji}(t), j = 1 \ldots, 4$ follows $U(-5, 5)$, where $U[a, b]$ denotes a random variable from the uniform distribution in the range $[a, b]$;

(3) Y_i follows a normal distribution with mean $\int_0^1 W_i(t)\alpha_0(t)dt$ and variance 1, where $W_i(t) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i(t)$.

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

Simulation 2 investigates the ability of our methods to handle a nonlinear mean and variance.

(1) Y_i follows 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 $0.5[1 + \{\boldsymbol{\beta}^{\text{T}}\int_0^1 \mathbf{X}_i(t)\alpha_0(t)dt\}^2]^{1/5}$.

Simulation 3 resembles the air pollution data structure in Section 5.

(1) $J = 4$, $\boldsymbol{\beta} = (-0.2, -1, -1.5, 1)^{\text{T}}$, and $\alpha_0(t) = 1 - 26.76t + 145.3t^2 - 227.27t^3 + 107.99t^4$;

(2) $X_{i1}(t) = 0.66 - 4.84t + 5.12t^2 + U[-4, 6]$, $X_{i2}(t) = 0.43 - 2.95t + 3.11t^2 + U[-5, 5]$, $X_{i3}(t) = -1.61 + 10.40t - 10.85t^2 + U[-4, 4]$, and $X_{i4}(t) = 0.58 - 3.59t + 3.52t^2 + U[-4, 8]$;

(3) Y_i is taken 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.

We applied the proposed method to estimate both $\boldsymbol{\beta}$ and $\alpha(t)$, where $\alpha(t)$ is approximated using cubic B-spline basis functions, with three equally spaced internal knots. For comparison, we implemented three estimators: the oracle, efficient, and locally efficient estimators. In the oracle estimator, we specified $f(Y, \boldsymbol{\beta}^{\text{T}}\mathbf{Z}\boldsymbol{\gamma})$ using the normal pdf form, and used the true $g(Y, \boldsymbol{\beta}^{\text{T}}\mathbf{Z}\boldsymbol{\gamma})$ in the

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

estimation. 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 using a nonparametric method. In the local estimator, we specified an incorrect model of $f(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$, hence using a misspecified $g^*(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ function, and estimated $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ nonparametrically. Note that the form of $f(Y,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\gamma})$ is unknown, in general. Hence, the oracle estimator is unrealistic and is only included here as a benchmark for comparison.

The numerical performance of the estimation of $\boldsymbol{\beta}$ in Simulations 1, 2, and 3 is summarized in Tables 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 values, and the 95% coverage is 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 that of 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.

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

The performance shown in Figure 1 is rather typical for spline approximations.

In a 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

$$E\{Y|\mathbf{X}(t)\} = \int_0^1 \mathbf{X}(t)^{\text{T}} \boldsymbol{\eta}(t) dt, \tag{4.1}$$

where $\boldsymbol{\eta}(t) = \{\eta_1(t), \ldots, \eta_p(t)\}^{\text{T}}$. The stacking approach is a special case of the proposed functional single-index model. We thus implemented the stacking approach and compared the two estimators in Figure 2. It is easy to see that our estimator performs better than the stacking approach, with narrower confidence bands. This pattern also applies to simulations 2 and 3. We provide the corresponding plots in Figures S1 and S2 of the Supplementary Material.

5. Application

We apply the proposed method to study the effect of various air pollutants on the rate of death caused by CVD, where we adopt the model in (1.1) without specifying any special link function.
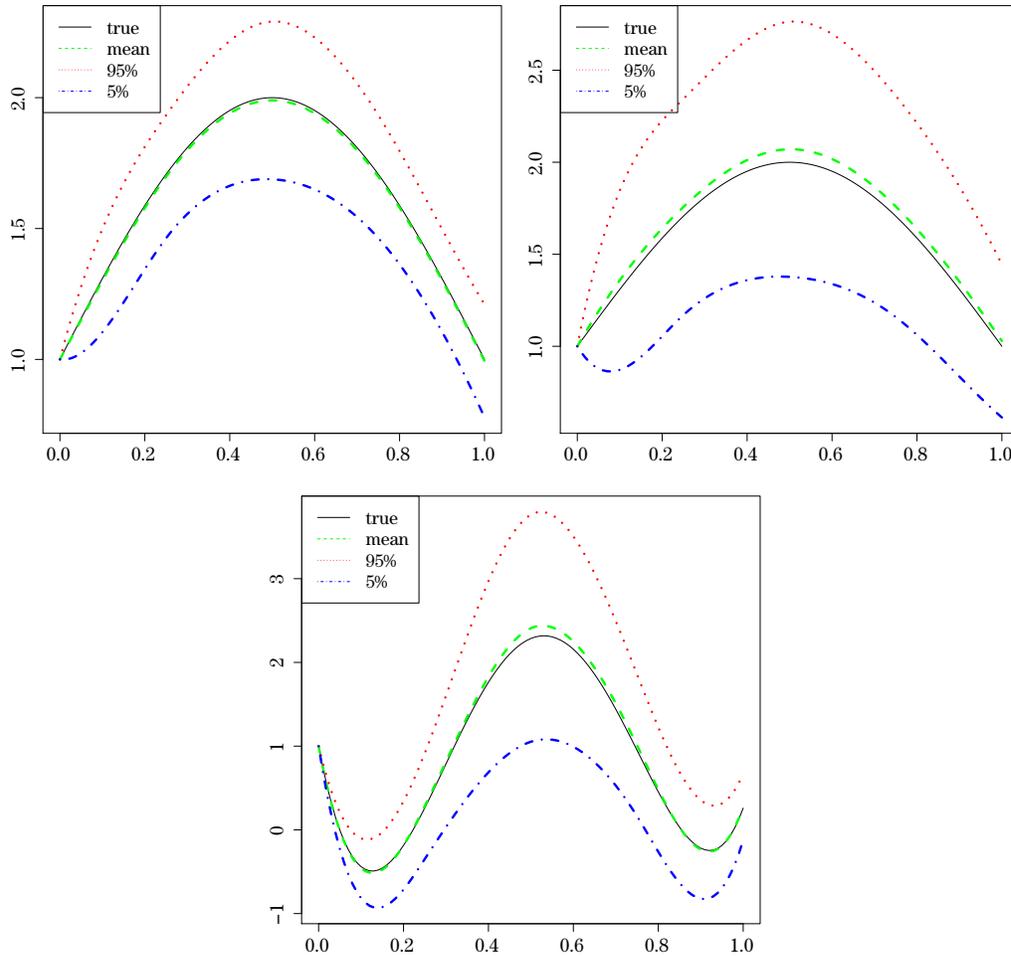
Figure 1. The mean and pointwise 90% confidence bands of the estimated $\widehat{\alpha}(t)$ for the functional single-index model (1.1) in Simulations 1 (left), 2 (middle), and 3 (right). The true $\alpha_0(t)$ is plotted in the solid curve.

In the NMMAP data (Peng and Welty (2004)), all four pollutants (CO, NO_2, SO_2, and O_3) were recorded on a daily basis in 108 U.S. cities. The measurements unit is parts per billion (ppb) by volume, and covers the period 1987 to 2000. We use 400 observations, with a relatively small portion of missing values, for the analysis. Each observation has 365 daily median measurements of the four air pollutants, where we imputed a few missing days in some observations using linear interpolation. We also standardize each pollutant across the whole year so that the 365 observations yield a sample mean of zero and a sample variance of
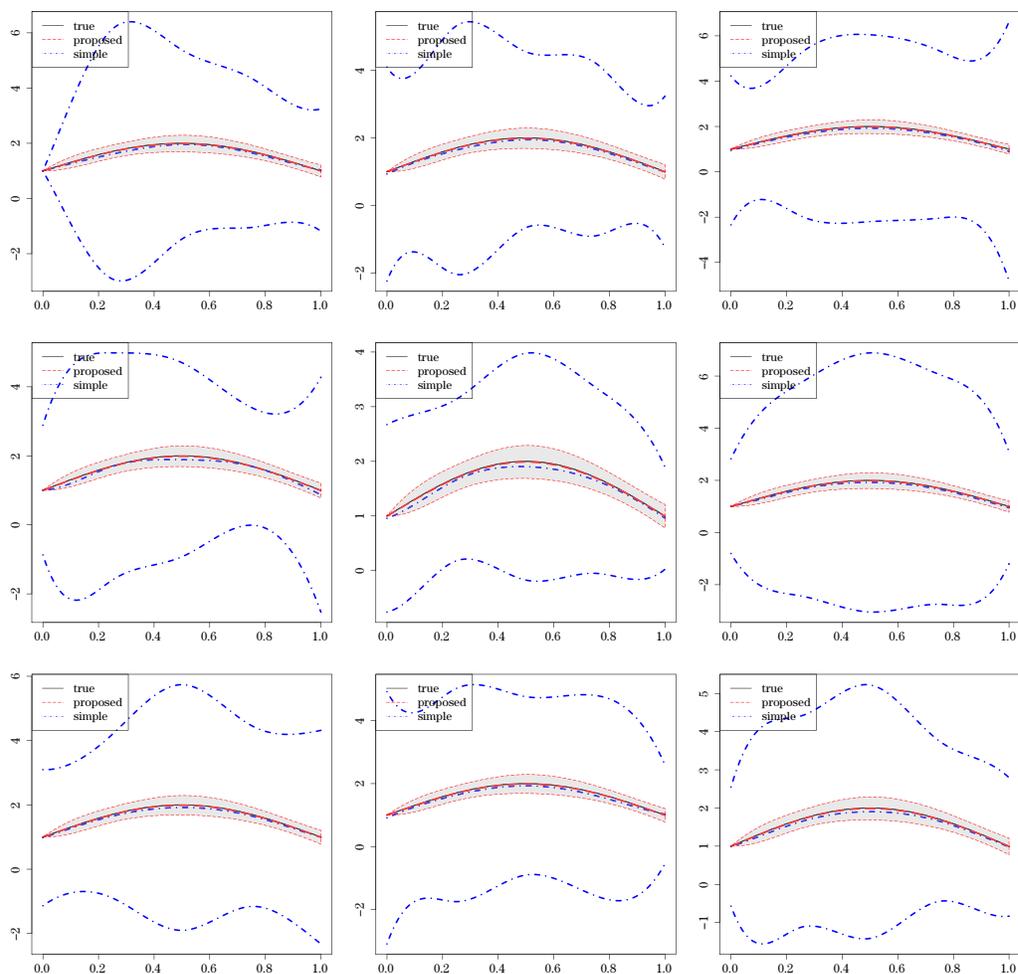
Figure 2.   The estimated $\beta_k \alpha(t)$, for $k = 1, \ldots, 9$, and their pointwise 90% confidence bands for the proposed functional single-index model (1.1), compared with the estimated $\widehat{\eta}_k(t)$ for the simple stacking functional linear model (4.1) in Simulation 1.

one. The time interval is normalized to $[0, 1]$. Figure S3 of the Supplementary Material displays the mean trajectories for the four pollutants.

We fit model (1.1) to estimate the air pollution index directly related to the subsequent year's CVD death rate. Throughout the implementation, we set the kernel bandwidth h to $n^{-1/5}\mathrm{range}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{z_i}\boldsymbol{\gamma})$ and $b = n^{-1/7}\mathrm{range}(y_i)$, where the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are updated 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 $[0, 1]$, where the optimal number of inter-
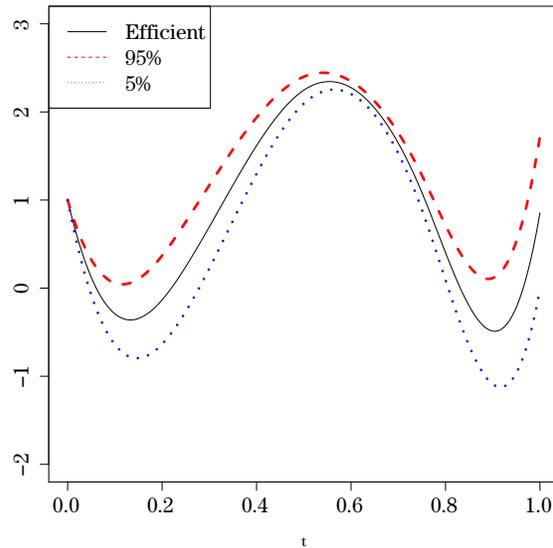
Figure 3.   The estimated $\widehat{\alpha}(t)$ for the functional single-index model (1.1) from the air pollution data. This captures the time-varying effect of the air pollution index on the annual CVD death rate. The pointwise 90% confidence band of the estimated $\widehat{\alpha}(t)$ is also provided.

nal knots is determined through a ten-fold cross-validation. We calculated the confidence band for $\alpha(t)$ using the asymptotic results in Theorem 1.

Figure 3 shows the time-varying effect $\widehat{\alpha}(t)$ of the estimated air pollution index to the CVD death rate. The time-varying effect is significantly positive in the spring, summer, and fall seasons, but is statistically 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 summertime, New York has the largest CVD death rate. On the other hand, Boston has the lowest air pollution index in summer and, hence, the CVD death rate is smallest in Boston, despite it having the highest air pollution index in winter.

Table 4 displays the estimated coefficients for the four air pollutants $\boldsymbol{\beta}$. The standard errors and the p-values are obtained based on the asymptotic normality of $\widehat{\boldsymbol{\beta}}$ shown in Theorem 2. 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 in the air pollution index related to the CVD death rate. This reaffirms that all pollu-
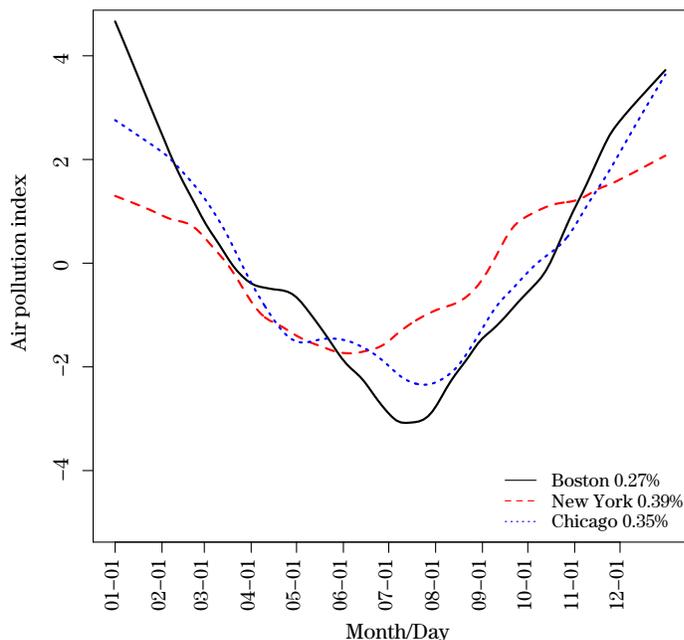
Figure 4.   The pollution indices for Boston, New York, and Chicago. The CVD death rates are shown in the legend.

Table 4.   The estimated coefficients for the air pollutants CO, NO_2, and SO_2, and the standard errors for the functional single-index model (1.1) in the air pollution data using the efficient method. The coefficient for O_3 is fixed at one 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	-

tants have a significant effect on 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 with O_3. The correlation coefficients between these four air pollutants are provided in Section S5 of the Supplementary Material. We also study the time-varying effect of each individual pollutant on the CVD death rate by fitting a simple functional linear regression $E(Y) = \int_0^1 \eta(t)X(t)dt$ to the air pollution data. Here, the response variable Y is the annual CVD death rate, and the functional covariate $X(t)$ is the daily concentration of the air pollutants CO, NO_2, SO_2, and O_3. Figure S4 in the Supplementary Material displays the
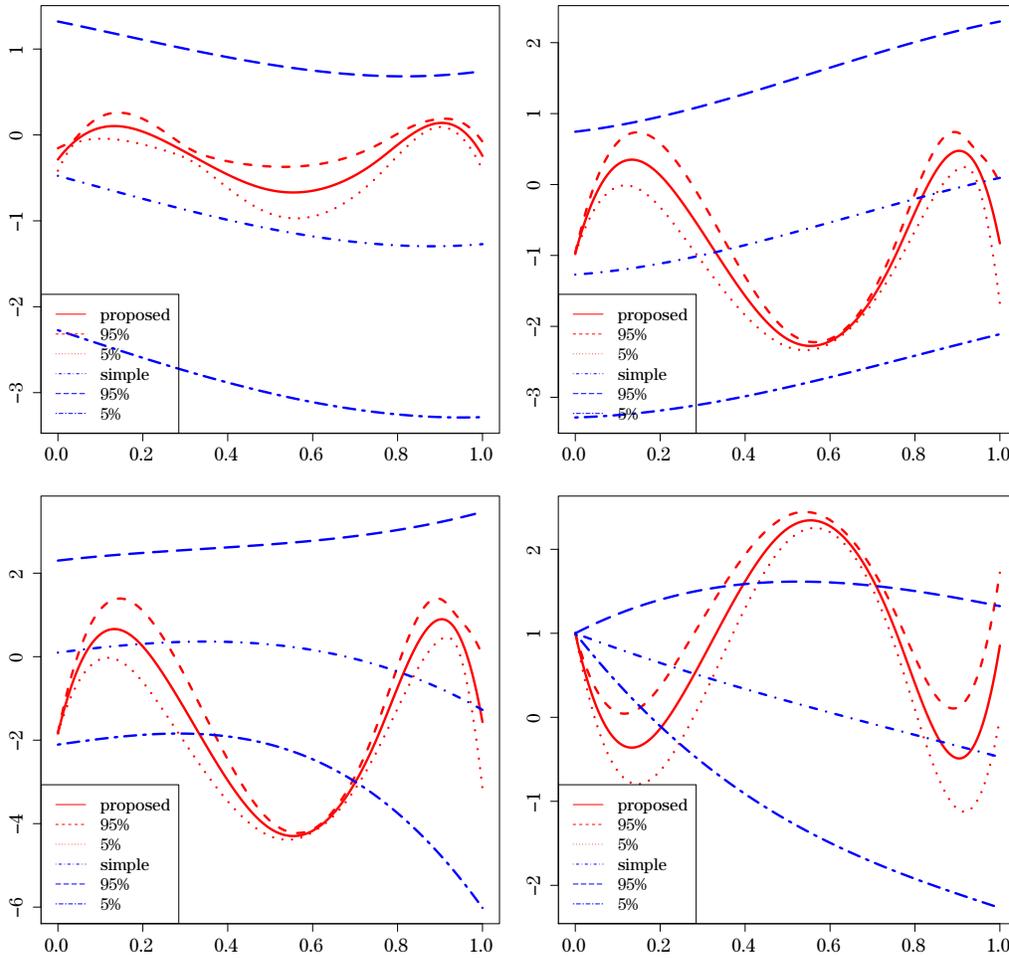
Figure 5. Comparison of the estimated $\widehat{\beta}_k\widehat{\alpha}(t)$ in the proposed functional single-index model (1.1), with 90% confidence bands, and the estimated $\widehat{\eta}_k(t)$ with 90% confidence bands in the simple stacking functional linear model (4.1), for $k = 1,\ldots,4$. The top, left panel shows $\widehat{\beta}_1\widehat{\alpha}(t)$ and $\widehat{\eta}_1(t)$, the top, right panel shows $\widehat{\beta}_2\widehat{\alpha}(t)$ and $\widehat{\eta}_2(t)$, the bottom, left panel shows $\widehat{\beta}_3\widehat{\alpha}(t)$ and $\widehat{\eta}_3(t)$, and the bottom, right panel shows $\widehat{\beta}_4\widehat{\alpha}(t)$ and $\widehat{\eta}_4(t)$.

estimated functional coefficient $\hat{\eta}(t)$ with the 95% pointwise confidence interval. It shows there is no significant effect of each individual pollutant on the CVD death rate. This is another motivation for us to estimate a comprehensive air pollution index to measure the contributions of air pollutants simultaneously.

For comparison, we implemented the stacking approach to estimate the functional linear model (4.1). Figure 5 compares the estimated $\widehat{\eta}_k(t)$ for the stacking

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}

functional linear model (4.1) and the estimated $\widehat{\beta}_k\widehat{\alpha}(t)$ for our functional index model (1.1), 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: the stacking functional linear model (4.1), 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 using a 10-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 the four methods. For instance, the MSPE decreases by 31% when using our proposed functional single-index in comparison with using the stacking functional linear model (4.1).

6. Discussion

We have proposed a functional single-index model for examing the relation between pollutants and the CVD death rate. The model contains a single-index that summarizes the pollution severity level, and a time-varying coefficient that captures the seasonality of the pollution effects. Furthermore, the model is robust against the misspecification of the conditional density function $f_{Y|\mathbf{X}(t)}(\cdot)$. When replacing the function $\alpha(\cdot)$ with its B-spline approximation, the model reduces to a dimension-folding model, and our estimator yields a new estimator as a by-product. This new estimator requires more relaxed conditions on the covariates, but still outperforms existing methods. Finally, the model and method can be used in high-dimensional settings because the numbers of covariate functions and spline bases are added. In contrast, the traditional functional single index described in (4.1) results in a multiplication of these two numbers.

In our analysis, to simplify the problem, we assume the functional covariate $X_i(t)$ is known. However, in practice, measurements for the functional covariate $X_i(t)$ may contain errors. To take into account such errors, model (1.1) needs to be extended. The resulting model falls within the measurement error framework and deserves careful investigation in future work.

Supplementary Material

The online Supplementary Material includes comprehensive proofs of all theoretical results. The computing code for our simulation studies and application can be downloaded from `https://github.com///sbaek306/FSIM`.

Acknowledgements

References

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

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*. Springer, New York.

Chen, D., Hall, P. and Müller, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics* **39**, 1720–1747.

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

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.

Ferré, L. and Yao, A. (2005). Smoothed functional inverse regression. *Statistica Sinica* **15**, 665–683.

Ferré, L. and Yao, A. (2007). Reply to the paper 'a note on smoothed functional inverse regression' by liliana forzani and r. dennis cook. *Statistica Sinica* **17**, 1534–1544.

Li, B., Kim, M. K. and Altman, N. (2010). On dimension folding of matrix or array valued statistical objects. *The 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, S. (2016). Estimation and inference in functional single-index models. *Annals of the Institute of Statistical Mathematics* **68**, 181–208.

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

Ma, Y. and Zhu, L. (2013). Efficient estimation in sufficient dimension reduction. *The 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.

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

Qu, S., Wang, J.-L. and Wang, X. (2016). Optimal estimation for the functional cox model. *The Annals of Statistics* **44**, 1708–1738.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd Edition. Springer-Verlag, New York.

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*. Springer, New York.

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. and 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.

Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China.

E-mail: feijiang@hku.hk

Department of Mathematics and Statistics, University of Maryland, Baltimore County, MD 21250, USA.

E-mail: baek@umbc.edu

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

E-mail: jiguo_cao@sfu.ca

Department of Statistics, Penn State University, University Park, PA 16802, USA.

E-mail: yzm63@psu.edu