Journal of Multivariate Analysis 171 (2019) 382–396

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Large-sample estimation and inference in multivariate single-index models

Jingwei Wu [a,*], Hanxiang Peng [b], Wanzhu Tu [c]

[a] Department of Epidemiology and Biostatistics, College of Public Health, Temple University, Philadelphia, PA 19122, United States
[b] Department of Mathematical Science, Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202, United States
[c] Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202, United States

ARTICLE INFO

Article history:
Received 14 February 2018
Available online 21 January 2019

Keywords:
Asymptotic normality
Consistency
Mixed effect model
Multivariate outcomes
P-splines
Single-index models

ABSTRACT

By optimizing index functions against different outcomes, we propose a multivariate single-index model (SIM) for development of medical indices that simultaneously work with multiple outcomes. Fitting of a multivariate SIM is not fundamentally different from fitting a univariate SIM, as the former can be written as a sum of multiple univariate SIMs with appropriate indicator functions. What have not been carefully studied are the theoretical properties of the parameter estimators. Because of the lack of asymptotic results, no formal inference procedure has been made available for multivariate SIMs. In this paper, we examine the asymptotic properties of the multivariate SIM parameter estimators. We show that, under mild regularity conditions, estimators for the multivariate SIM parameters are indeed $\sqrt{n}$-consistent and asymptotically normal. We conduct a simulation study to investigate the finite-sample performance of the corresponding estimation and inference procedures. To illustrate its use in practice, we construct an index measure of urine electrolyte markers for assessing the risk of hypertension in individual subjects.

1. Introduction

Composite indices are frequently used in medical research to quantify latent characteristics of study subjects. In most situations, indices are constructed as simple aggregations of factors observed in individual subjects. By combining multiple factors into a composite measure, the index provides a single-valued summary of specific health risks in individual patients. For example, the commonly used body mass index (BMI) aggregates the height and weight information into a composite measure that quantifies the "fatness" of an individual, without actually measuring the person's body composition.

In practice, the construction of indices has always been heuristic, and the derivation is usually based on "insights" from content experts. The acceptance of an index, however, depends on the strength of associations between the index measure and health outcomes that the index predicts. In the case of BMI, ample evidence affirms that the formulation Weight/Height^2 indeed correlates well with many adiposity-related adverse outcomes, although there is no assurance that this particular formulation is the best functional form. In modeling practice, analysts often want to know how to find an optimal scalar function η satisfying $\mathrm{E}(Y|\mathbf{X}) = \eta(\boldsymbol{\alpha}^\top \mathbf{X})$ for outcome Y and a given set of risk factors $\mathbf{X}$. Single-index model (SIM) provides the necessary tools for estimating η and thus directly meeting this need. But indices that work with only one outcome usually have limited appeal. To remedy, Wu and Tu [32] put forward a multivariate SIM that allows the estimation of η for multiple

* Corresponding author.
E-mail address: tug30693@temple.edu (J. Wu).

https://doi.org/10.1016/j.jmva.2019.01.003

pre-selected outcomes Y, which could be used in the development of multi-purpose index measures. The ability to predict multiple health outcomes is considered a hallmark of generally applicable indices, as in the case of BMI.

Using SIM for index development is a relatively recent practice [31,32]. In the early literature of SIM, the method is primarily used as a dimension reduction tool, and the existing literature has largely focused on the estimation of η and $\boldsymbol{\alpha}$; see, e.g., [3,7,9,13]. Along that vein, a number of authors have investigated the algorithmic efficiency [2,10,14], and asymptotic properties of univariate SIMs [2,23,33–35,37,39–41].

What has not been well developed is the theory of multivariate SIM. For the proposed model, an asymptotic theory is essential for the purpose of statistical inference. Extending the theoretical results from univariate to multivariate SIM, however, is no trivial task. In this paper we present a multivariate SIM with partially linear effects and heterogeneous random errors. Along this line, we also incorporate a working covariance matrix to improve the estimation efficiency and to account for data heterogeneity. We demonstrate in the present article that the proposed multivariate SIM inherits the nice theoretical properties possessed by the univariate SIMs, while offering enhanced modeling flexibility for heterogeneous errors, as well as for deriving indices that work for multiple outcomes. The theoretical development extends the results of Yu and Ruppert [37] and Tian et al. [23] to a multivariate modeling setting, under more relaxed conditions.

The article is organized as follows. In Section 2, we introduce the multivariate partially linear single-index model, give the parameter estimation, and present the theoretical results. We then present a simulation study in Section 3. We illustrate the use of the method in a real data application in Section 4. We provide a discussion in Section 5. Details of the proofs of main theoretical results are provided in the Appendix.

2. The model and theoretical results

In this section, we first introduce the model and the related parameter estimation method. We then present the theoretical results.

2.1. Multivariate single-index models

2.1.1. The basic model

We consider a situation in which there are multiple outcomes of interest. Let $Y_{\ell,i}$ be the ℓth outcome for the ith subject, for $\ell \in \{1,\ldots,L\}$ and $i \in \{1,\ldots,n\}$. Let $\mathbf{x}_i$ be a d-dimensional non-random vector of independent variables from the ith subject. The multivariate SIM can be written, for all $\ell \in \{1,\ldots,L\}$ and $i \in \{1,\ldots,n\}$, as

$$Y_{\ell,i} = f_\ell(\boldsymbol{\alpha}^\top \mathbf{x}_i) + \epsilon_{\ell,i}, \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^d$ is an index parameter vector, $\epsilon_{\ell,i}$ is a random error, and each f_ℓ is assumed to be an unknown smooth scalar function that is twice continuously differentiable with bounded first two moments. Such an assumption assures that the unknown link function is arbitrarily well approximated by spline models. We note that most of the SIMs in the existing literature have been expressed in such a form [17,28,36,37,39,41]. The main task is therefore to approximate each f_ℓ with a spline function and to estimate the spline coefficients. Here we let the index functions be outcome-specific for maximal flexibility. In most applications, however, it suffices to use a common index function f for all outcomes. Here we use a common set of index coefficients $\boldsymbol{\alpha}$ so that the basic composition of the resultant index itself does not vary across outcomes.

Following the standard SIM, we assume $\|\boldsymbol{\alpha}\| = 1$ with its first component $\alpha_1 > 0$ so that $\boldsymbol{\alpha}$ is fully identifiable. Let $\mathbf{f} = (f_1,\ldots,f_L)^\top$ be the vector of link functions. For $\mathbf{u} = (u_1,\ldots,u_L)^\top$, we define the Hadamard product operator $\circ$ on $\mathbf{f}$ as

$$\mathbf{f}_\circ(\mathbf{u}) = \mathbf{f} \circ \mathbf{u} = (f_1(u_1),\ldots,f_L(u_L))^\top.$$

Let $\mathbf{X}_i = (\mathbf{x}_i,\ldots,\mathbf{x}_i)^\top$ be the $L \times d$ covariate matrix on the ith individual, $\mathbf{y}_i = (Y_{1,i},\ldots,Y_{L,i})^\top$ be the ith outcome vector, and $\boldsymbol{\epsilon}_i = (\epsilon_{1,i},\ldots,\epsilon_{L,i})^\top$ be the ith random error vector. We then rewrite (1) in a vector form, by setting, for all $i \in \{1,\ldots,n\}$,

$$\mathbf{y}_i = \mathbf{f}_\circ(\mathbf{X}_i\boldsymbol{\alpha}) + \boldsymbol{\epsilon}_i.$$

We further express the multivariate SIM as a sum of univariate SIM with indicators $\mathbf{1}(j = \ell)$ of $j = \ell$. To do so, we let $\mathbf{e}_1,\ldots,\mathbf{e}_L$ be the usual orthonormal basis of $\mathbb{R}^L$. Using these, model (1) can be written, for all $\ell \in \{1,\ldots,L\}$ and $i \in \{1,\ldots,n\}$, as

$$Y_{\ell,i} = \mathbf{f}_\circ^\top(\mathbf{X}_i\boldsymbol{\alpha})\mathbf{e}_\ell + \epsilon_{\ell,i}.$$

In practice, different smoothing methods have been used in SIMs, including kernel smoothing, local linear methods, and average derivatives methods [3,9,21]. In this research, we chose to model the unknown link functions f_ℓ as a pth order polynomial spline with K knots $\kappa_{\ell,k}$ [18,37], so that for all $v \in \mathbb{R}$ and $\ell \in \{1,\ldots,L\}$,

$$f_\ell(v) = \sum_{q=0}^{p} \beta_{\ell,q} v^q + \sum_{k=1}^{K} \beta_{\ell,p+k}(v - \kappa_{\ell,k})_+^p, \tag{2}$$

where $v_+ = \max(0, v)$. For simplicity, we assume that the knots are fixed and p is a constant with $p \geq 3$. For the remainder of the paper, we operate under the assumption that for some p, f_ℓ is well approximated by a pth-order spline function with knots $\kappa_{\ell,k}$. Such polynomial splines are mathematically simple and computationally efficient. Such an assumption results in a finite-dimensional linear space, within which we derive the theoretical results. In most applications, the assumption is not unreasonable because the order of splines and the number and location of knots are typically determined by the analyst based on the degree of smoothness needed. We note, however, that several authors have investigated situations where numbers of knots depend on sample sizes [5,29].

Let $\mathbf{g}_\ell(v) = (1, v, \dots, v^p, (v-\kappa_{\ell,1})_+^p, \dots, (v-\kappa_{\ell,K})_+^p)^\top$ be the spline bases with K knots $\kappa_{\ell,k}$, and $\boldsymbol{\beta}_\ell = (\beta_{\ell,0}, \dots, \beta_{\ell,p}, \beta_{\ell,p+1}, \dots, \beta_{\ell,p+K})^\top$ be the corresponding coefficients. Using the splines, we express f_ℓ in (2) as

$$f_\ell(v) = \boldsymbol{\beta}_\ell^\top \mathbf{g}_\ell(v)$$

for all $v \in \mathbb{R}$ and $\ell \in \{1, \dots, L\}$. Let $\mathbf{b}_\ell(v) = \mathbf{g}_\ell(v) \otimes \mathbf{e}_\ell$ and $m_\ell(v; \boldsymbol{\beta}) = \boldsymbol{\beta}^\top \mathbf{b}_\ell(v)$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_L^\top)^\top$ and $\otimes$ is the Kronecker product operator. Letting $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$, we write the mean function for the ℓth outcome in the ith individual as

$$\mathrm{E}(Y_{\ell,i}) = m_{\ell,i}(\boldsymbol{\theta}) = m_\ell(\mathbf{x}_i^\top \boldsymbol{\alpha}; \boldsymbol{\beta}) = \boldsymbol{\beta}^\top \mathbf{b}_\ell(\mathbf{x}_i^\top \boldsymbol{\alpha}), \quad (3)$$

where $\ell \in \{1, \dots, L\}$ and $i \in \{1, \dots, n\}$.

Instead of (1), the multivariate single index model can be expressed, for all $\ell \in \{1, \dots, L\}$ and $i \in \{1, \dots, n\}$, as

$$Y_{\ell,i} = \boldsymbol{\beta}^\top \mathbf{b}_\ell(\mathbf{x}_i^\top \boldsymbol{\alpha}) + \epsilon_{\ell,i}. \quad (4)$$

Note that the optimization of the penalized least squares with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in (4) is subject to the constraint $\|\boldsymbol{\alpha}\| = 1$ with its first component being positive. For convenience, we use the "delete-one-component" to re-parameterize $\boldsymbol{\alpha}$ [28,37,41]. Let $\boldsymbol{\phi}$ be a $(d-1)$-dimensional parameter vector with $\|\boldsymbol{\phi}\| \le 1$ and

$$\boldsymbol{\alpha}(\boldsymbol{\phi}) = (\sqrt{1 - \|\boldsymbol{\phi}\|^2},\ \boldsymbol{\phi}^\top)^\top.$$

It then follows that $\boldsymbol{\alpha}(\boldsymbol{\phi})$ is d-dimensional and satisfies the identifiability conditions $\|\boldsymbol{\alpha}(\boldsymbol{\phi})\| = 1$ with the first component being positive. With the re-parameterization, optimization of the restricted penalized least squares now becomes an unrestricted optimization problem for (4) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$.

We assume that the true parameter vector $\boldsymbol{\phi}_0$ satisfies the constraint $\|\boldsymbol{\phi}_0\| < 1$ so that $\boldsymbol{\alpha}(\boldsymbol{\phi})$ is infinitely differentiable in a neighborhood of $\boldsymbol{\phi}_0$. We write $\boldsymbol{\theta}_{\boldsymbol{\phi}} = (\boldsymbol{\phi}^\top, \boldsymbol{\beta}^\top)^\top$. Obviously $\boldsymbol{\theta}_{\boldsymbol{\phi}}$ is one dimension lower than $\boldsymbol{\theta}_{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$. In terms of the new parameter $\boldsymbol{\phi}$, the mean function (3) can now be written, for all $\ell \in \{1, \dots, L\}$ and $i \in \{1, \dots, n\}$, as

$$m_{\ell,i}(\boldsymbol{\theta}_{\boldsymbol{\phi}}) = m_\ell\{\mathbf{x}_i^\top \boldsymbol{\alpha}(\boldsymbol{\phi}); \boldsymbol{\beta}\} = \boldsymbol{\beta}^\top \mathbf{b}_\ell\{\mathbf{x}_i^\top \boldsymbol{\alpha}(\boldsymbol{\phi})\}.$$

Let $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_L)^\top$. Similar to the Hadamard notation introduced above, we define the Hadamard product operator $\circ$ on a matrix function $\mathbf{B}$ by letting, for $\mathbf{u} = (u_1, \dots, u_L)^\top$,

$$\mathbf{B}_\circ(\mathbf{u}) = (\mathbf{b}_1(u_1), \dots, \mathbf{b}_L(u_L))^\top.$$

With such a notation, the penalized spline approximation to the multivariate SIM (4) can be written, for all $i \in \{1, \dots, n\}$, as

$$\mathbf{y}_i = \mathbf{B}_\circ\{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\}\boldsymbol{\beta} + \boldsymbol{\epsilon}_i. \quad (5)$$

Let $\mathbf{m}_i = (m_{1,i}, \dots, m_{L,i})^\top$ be the vector of the mean functions for the ith individual. We have, for all $i \in \{1, \dots, n\}$,

$$\mathbf{m}_i(\boldsymbol{\beta}, \boldsymbol{\phi}) = \mathbf{B}_\circ\ \{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\}\boldsymbol{\beta}.$$

2.1.2. Partially linear multivariate SIMs

In regression analysis, testing of linear effects remains one of the mainstay inference practices. To this end, we add linear effects to the proposed single-index model. The concept has been described previously in [31,32,37]. In the context of multivariate single-index model, we write the extended model, for all $\ell \in \{1, \dots, L\}$ and $i \in \{1, \dots, n\}$, as

$$Y_{\ell,i} = f_\ell(\boldsymbol{\alpha}^\top \mathbf{x}_i) + \boldsymbol{\gamma}_\ell^\top \mathbf{z}_i + \epsilon_{\ell,i}, \quad (6)$$

where $\mathbf{z}_i \in \mathbb{R}^{d_\gamma}$ are fixed observed covariate vectors, and each $\boldsymbol{\gamma}_\ell \in \mathbb{R}^{d_\gamma}$ is an unknown linear parameter vector for ℓth outcome.

If we define a fixed covariate design matrix $\mathbf{Z}_i = \mathbf{z}_i^\top \otimes \iota_L^\top$ and corresponding linear parameter vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \dots, \boldsymbol{\gamma}_L^\top)^\top$, where ι_L is a vector of 1s of length L, and if we further define $\mathbf{B}_\mathbf{z}\{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\} = (\mathbf{B}_\circ\{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\}, \mathbf{Z}_i)$ and $\boldsymbol{\beta}_\mathbf{z} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, we can express multivariate partially linear single-index model (6), for each $i \in \{1, \dots, n\}$, as

$$\mathbf{y}_i = \mathbf{B}_\mathbf{z}\{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\}\boldsymbol{\beta}_\mathbf{z} + \boldsymbol{\epsilon}_i,$$

which is similar to the multivariate SIM (5).

It is important to note that the addition of the linear terms does not fundamentally change the estimation of the index function, nor does it alter the theoretical properties of the index parameters. Therefore, without loss of generality, we consider the model (5) in the parameter estimation and theoretical development. The same proof remains valid for partially linear single-index models when we replace $\mathbf{B}_\circ\{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\}$ by $\mathbf{B}_\mathbf{z}\{\mathbf{X}_i \boldsymbol{\alpha}(\boldsymbol{\phi})\}$ and $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_\mathbf{z}$.

2.2. Parameter estimation

For parameter estimation, we write the working covariance matrix for the ith subject as $\mathbf{W}_i$, which is an $L \times L$ positive-definite matrix. The diagonal elements of $\mathbf{W}_i$ are $\sigma^2_{\epsilon_\ell}$, where for each $\ell \in \{2, \ldots, L\}$, $\sigma^2_{\epsilon_\ell}$ is expressed as product of a common variance component σ^2_ϵ and outcome-specific scale parameters δ_ℓ, i.e., $\sigma^2_{\epsilon_\ell} = \delta_\ell \sigma^2_\epsilon$. The off-diagonal elements are $\rho_{st}\sigma_{\epsilon_s}\sigma_{\epsilon_t}$, where, ρ_{st} is the correlation of the paired outcomes $Y_{s,i}$ and $Y_{t,i}$, whereas $\sigma^2_{\epsilon_s}$ and $\sigma^2_{\epsilon_t}$ are the corresponding variance components.

Let the objective quadratic function be

$$Q_n(\boldsymbol{\theta}_{\boldsymbol{\phi}}) = Q_n(\boldsymbol{\phi}, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\phi}, \boldsymbol{\beta})\}^\top \mathbf{W}_i^{-1}\{\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\phi}, \boldsymbol{\beta})\}. \tag{7}$$

The penalized objective function is

$$Q_{n,\lambda}(\boldsymbol{\theta}_{\boldsymbol{\phi}}) = Q_{n,\lambda}(\boldsymbol{\phi}, \boldsymbol{\beta}) = Q_n(\boldsymbol{\phi}, \boldsymbol{\beta}) + \sum_{\ell=1}^{L}\lambda_\ell \boldsymbol{\beta}_\ell^\top \mathbf{D}_\ell \boldsymbol{\beta}_\ell, \tag{8}$$

where each $\mathbf{D}_\ell$ is an appropriate positive semi-definite symmetric matrix, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L)^\top$ is a vector of nonnegative penalty parameters.

For the penalized splines considered in the current paper, we choose $\mathbf{D}_\ell$ to be a diagonal matrix with the last K diagonal elements equal to 1 and the rest equal to 0, so that it penalizes the sum of squares of the parameters to the pth order [20]. Solving the estimating equations, we have a weighted penalized spline least square estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\phi}}_n^\top, \hat{\boldsymbol{\beta}}_n^\top)^\top$ of $\boldsymbol{\theta}_{\boldsymbol{\phi}} = (\boldsymbol{\phi}^\top, \boldsymbol{\beta}^\top)^\top$. Parameters in the covariance matrix $\mathbf{W}_i$ are estimated along with the spline smoothing parameters (also treated as regular variance component parameters) within the linear mixed effect model framework [15,32]. Finally, we note that all parameters can be estimated explicitly by minimizing $Q_{n,\lambda}(\boldsymbol{\theta}_{\boldsymbol{\phi}})$ via the nonlinear least squares, and the selection of λ_ℓ can be obtained by using the generalized cross validation (GCV), and the model written in the form of a mixed model and fitted through commonly used statistical software [20,32].

2.3. Theoretical results

In this section, we lay out the conditions with which the estimators of the multivariate SIM achieve the desired asymptotic properties. Sketches of the proofs of Theorems 1 and 3 are presented in the Appendix.

Let $[\mathbf{A}_1 : \mathbf{A}_2]$ be a column-binding matrix of $\mathbf{A}_1, \mathbf{A}_2$ of compatible dimension. Suppose $\boldsymbol{\phi}$ is in a neighborhood of $\boldsymbol{\phi}_0$ such that $\|\boldsymbol{\phi}\| < 1$. The gradient of the $\dim(\boldsymbol{\theta}_{\boldsymbol{\alpha}})$-dimensional mean function is then given by

$$\dot{\mathbf{m}}_{\ell,i}(\boldsymbol{\theta}_{\boldsymbol{\phi}}) = \begin{pmatrix} \boldsymbol{\beta}^\top \dot{\mathbf{b}}_\ell\{\mathbf{x}_i^\top \boldsymbol{\alpha}(\boldsymbol{\phi})\}[-(1-\|\boldsymbol{\phi}\|^2)^{-1/2}\boldsymbol{\phi} : \mathbf{I}_{d-1}]\mathbf{x}_i \\ \mathbf{b}_\ell\{\mathbf{x}_i^\top \boldsymbol{\alpha}(\boldsymbol{\phi})\} \end{pmatrix},$$

where $\mathbf{I}_{d-1}$ is the $(d-1)\times(d-1)$ identity matrix, and $\dot{\mathbf{b}}_\ell$ is the first gradient of $\mathbf{b}_\ell$.

The $\dim(\boldsymbol{\theta}_{\boldsymbol{\alpha}}) \times \{\dim(\boldsymbol{\theta}_{\boldsymbol{\alpha}}) - 1\}$ Jacobian matrix of the mapping $\boldsymbol{\theta}_{\boldsymbol{\phi}} \to \boldsymbol{\theta}_{\boldsymbol{\alpha}}$ is

$$\mathbf{J}(\boldsymbol{\phi}) = \frac{\partial}{\partial \boldsymbol{\theta}_{\boldsymbol{\phi}}^\top}\boldsymbol{\theta}_{\boldsymbol{\alpha}} = \begin{pmatrix} -(1-\|\boldsymbol{\phi}\|^2)^{-1/2}\boldsymbol{\phi}^\top & \mathbf{0} \\ \mathbf{I}_{d-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(p+K+1)\times L} \end{pmatrix}. \tag{9}$$

For narrative convenience, we simplify the notation for the rest of the paper. We drop $\boldsymbol{\phi}$ and write $\boldsymbol{\theta} = \boldsymbol{\theta}_{\boldsymbol{\phi}}$ unless otherwise specified, while keeping the subscript in $\boldsymbol{\theta}_{\boldsymbol{\alpha}}$ to emphasize its relation to the original parameters. Similarly, we let $\boldsymbol{\theta}_0 = (\boldsymbol{\phi}_0^\top, \boldsymbol{\beta}_0^\top)^\top$ be the true values of parameters, while $\boldsymbol{\theta}_{\boldsymbol{\alpha}_0}$ and $\boldsymbol{\theta}_{\boldsymbol{\phi}_0}$ are used to remind the true parameter values in $\mathbb{R}^d$ and $\mathbb{R}^{d-1}$, respectively. We suppress the subscript n and write $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_n$, when there is no confusion.

We derive the asymptotics under the following conditions.

Condition 1. There exist constants $r > 0$ and $c_0 > 0$ such that

$$\sup_n \max_{i\in\{1,\ldots,n\}} \max_{\ell\in\{1,\ldots,L\}} \mathrm{E}|\epsilon_{\ell,i}|^{2+r} \le c_0 < \infty.$$

Condition 2. Matrices $\boldsymbol{\Sigma}_i = \mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top)$ and $\mathbf{W}_i$ are positive definite with eigenvalues bounded away from zero and infinity uniformly in $i \in \{1, \ldots, n\}$ such that, for some $\tau \in (0, \infty)$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{tr}(\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i) = \tau.$$

Here, $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, \ldots, \epsilon_{L,i})$ for each $i \in \{1, \ldots, n\}$, and $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ are mutually independent with mean zero.

Condition 3. There exists a matrix function $\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*)$ on a compact subspace Φ of $\mathbb{R}^{d-1}$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\mathbf{W}_i^{-1}\mathbf{B}_\circ\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi}^*)\} \to \mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*) \tag{10}$$

uniformly in $\boldsymbol{\phi}, \boldsymbol{\phi}^* \in \Phi$, that $\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi})$ is positive definite on Φ, and that $\mathbf{T}(\boldsymbol{\phi}) = \mathbf{S}(\boldsymbol{\phi}_0, \boldsymbol{\phi}_0) - \mathbf{S}(\boldsymbol{\phi}_0, \boldsymbol{\phi})\mathbf{S}^{-1}(\boldsymbol{\phi}, \boldsymbol{\phi})\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}_0)$ has a unique zero solution at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

Condition 4. There exists a neighborhood of $\boldsymbol{\theta}_0$ such that each $\dot{\mathbf{m}}_i(\boldsymbol{\theta}) = (\dot{\mathbf{m}}_{1,i}(\boldsymbol{\theta}), \ldots, \dot{\mathbf{m}}_{L,i}(\boldsymbol{\theta}))^\top$ is of full rank,

$$\boldsymbol{\Delta}_{n,1}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\dot{\mathbf{m}}_i^\top(\boldsymbol{\theta})\mathbf{W}_i^{-1}\dot{\mathbf{m}}_i(\boldsymbol{\theta}), \quad \boldsymbol{\Delta}_{n,2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\dot{\mathbf{m}}_i^\top(\boldsymbol{\theta})\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{W}_i^{-1}\dot{\mathbf{m}}_i(\boldsymbol{\theta}), \tag{11}$$

where $\boldsymbol{\Delta}_1(\boldsymbol{\theta}_0) = \lim_{n\to\infty}\boldsymbol{\Delta}_{n,1}(\boldsymbol{\theta}_0)$ is non-singular and $\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0) = \lim_{n\to\infty}\boldsymbol{\Delta}_{n,2}(\boldsymbol{\theta}_0)$, and for all $s, t \in \{1, \ldots, \dim(\boldsymbol{\theta})\}$,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{L}\left\{\frac{\partial^2}{\partial\theta_s\partial\theta_t}m_{\ell,i}(\boldsymbol{\theta})\right\}^2 \tag{12}$$

converges uniformly in $\boldsymbol{\theta}$ over the neighborhood of $\boldsymbol{\theta}_0$.

Remark 1. As the order p of the spline is such that $p \geq 3$, the vector mean function $\mathbf{m} = (\mathbf{m}_1, \ldots, \mathbf{m}_n)^\top$ has Lipschitz continuous second order partial derivatives since the second derivative $(x_+^3)'' = 6x_+$ is Lipschitz continuous in view of the inequality $|y_+ - x_+| \leq |y - x|$ for all $x, y \in \mathbb{R}$. This fact will be applied in the proof of Lemma A without explicitly mentioning.

The following condition is shown to be sufficient for Condition 3 in the Appendix.

Remark 2. Suppose that $\max_i \|\mathbf{X}_i\| \leq M_0 < \infty$ and that the eigenvalues of $\mathbf{W}_i$ have a common lower bound $m_0 > 0$ for constants M_0 and m_0. Suppose Φ is compact and is contained in a ball centered at the origin with radius r_0 for some $r_0 \in (0, 1)$. Then the point-wise convergence in (10) over Φ implies the uniform convergence.

We write $\lambda_{n,\ell} = \lambda_\ell$ for all $\ell \in \{1, \ldots, L\}$ to emphasize the parameter's dependence on the sample size n. We now state the main theorems.

Theorem 1 (*Strong Consistency*)**.** *Suppose Conditions 1–3 hold. If $\lambda_{n,\ell} = o(1)$ for all $\ell \in \{1, \ldots, L\}$ as $n \to \infty$, then there exists a sequence of weighted penalized spline least squares estimators $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$ that are strongly consistent, i.e.,* $\Pr(\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0) = 1$ *as $n \to \infty$.*

Remark 3. Theorem 1 states a result that includes Theorem 1 of [37] and Theorem 1′ of [38] as special cases. It extends the results of the two theorems from scalar responses and independent and identically distributed (iid) random errors to vector responses and independent but not identically distributed errors. Additionally, working covariances are used in the estimating procedure. In particular, this generalizes the previously published results from scalar responses with iid random errors to scalar responses with independent heterogeneous random errors. Such extensions are important practically as heterogeneity is common in real life data.

Remark 4. For the proof of consistency, Yu and Ruppert [37] assumed that the all parameters, i.e., both the linear and spline coefficients, resided in a compact space. We show that consistency in Theorem 1 can be achieved under Condition 3. In other words, we only require a compactness assumption for the subspace that contains the reparameterized single-index parameters. Since the partially linear coefficients and spline coefficients lie within a general Euclidean space, they can be handled separately through linear ridge regression analysis. Therefore, the compactness assumption of the parameter space that Yu and Ruppert [37] used is no longer necessary. For details, see [39] and our proof of Theorem 1.

We note that Conditions 1–2 are only used to prove Lemma A. If we relax the conditions with the following Condition 2′, we can prove that the result of Lemma A still holds by applying Theorem 5.4.2 of [4].

Condition 2′. *The random errors $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ are mutually independent and identically distributed with zero mean and (component-wise) finite positive definite covariance matrix $\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top)$. Furthermore, the working covariance matrices satisfy $\mathbf{W}_1 = \cdots = \mathbf{W}_n = \mathbf{W}$ for some positive definite matrix $\mathbf{W}$.*

Under this altered condition, we have the following Theorem 2.

Theorem 2 (*Strong Consistency*)**.** *Suppose Conditions 2′ and 3 hold. If $\lambda_{n,\ell} = o(1)$ for all $\ell \in \{1, \ldots, L\}$, then the result in Theorem 1 holds.*

Theorem 3 (*Asymptotic Normality*)**.** *Suppose that Conditions 1–4 hold. If $\lambda_{n,\ell} = o(n^{-1/2})$ for all $\ell \in \{1, \ldots, L\}$, then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightsquigarrow \mathcal{N}[\mathbf{0},\ \boldsymbol{\Delta}_1^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)\boldsymbol{\Delta}_1^{-1}(\boldsymbol{\theta}_0)]$$

as $n \to \infty$, where $\boldsymbol{\Delta}_1(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)$ are given in Condition 4.

Recall that in our notation, $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\phi}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top$ is an estimator of $\boldsymbol{\theta}_{\boldsymbol{\phi}} = (\boldsymbol{\phi}^\top, \boldsymbol{\beta}^\top)^\top$. We can estimate the original parameter $\boldsymbol{\theta}_{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ with the plug-in estimator $\hat{\boldsymbol{\theta}}_{\hat{\boldsymbol{\alpha}}} = (\boldsymbol{\alpha}^\top(\hat{\boldsymbol{\phi}}), \hat{\boldsymbol{\beta}}^\top)^\top$, where $\boldsymbol{\alpha}(\boldsymbol{\phi}) = (\sqrt{1 - \|\boldsymbol{\phi}\|^2}, \boldsymbol{\phi}^\top)^\top$. Asymptotic normality can be easily derived by combining Theorems 1 and 3 and by using the multivariate Delta method. This is stated below.

Theorem 4. *Suppose the Conditions in Theorem 3 hold. Then there exists a sequence of generalized weighted penalized spline least squares estimators $\hat{\boldsymbol{\theta}}_{\hat{\boldsymbol{\alpha}}} = (\hat{\boldsymbol{\alpha}}^\top, \hat{\boldsymbol{\beta}}^\top)$ with $\|\hat{\boldsymbol{\alpha}}\| = 1$ and positive first component that are strongly consistent and asymptotically normal, i.e., as $n \to \infty$, $\hat{\boldsymbol{\theta}}_{\hat{\boldsymbol{\alpha}}} \to \boldsymbol{\theta}_{\boldsymbol{\alpha}_0}$ a.s. and*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\hat{\boldsymbol{\alpha}}} - \boldsymbol{\theta}_{\boldsymbol{\alpha}_0}) \rightsquigarrow \mathcal{N}[\mathbf{0},\ \mathbf{J}(\boldsymbol{\phi}_0)\boldsymbol{\Delta}_1^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)\boldsymbol{\Delta}_1^{-1}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\phi}_0)^\top],$$

where $\mathbf{J}$ is given in (9).

Theorem 5 (*Asymptotic Normality*)**.** *Suppose Conditions 2′ and 3–4 hold. If $\lambda_{n,\ell} = o(n^{-1/2})$ for all $\ell \in \{1, \ldots, L\}$, then the result in Theorem 3 holds.*

Theorem 6. *Suppose Conditions in Theorem 5 hold. If $\lambda_{n,\ell} = o(n^{-1/2})$ for all $\ell \in \{1, \ldots, L\}$, then the result in Theorem 4 holds.*

Remark 5. In the proofs, we considered the asymptotic properties of the estimators for a fixed number of knots. The theoretical results presented in the current paper directly correspond to models specified in Eqs. (2)–(5). This type of the asymptotic theory, as argued in [39], is more useful than that of an increasing number of knots for the purpose of statistical inference. Clearly, too few knots do not provide a satisfactory fit, and too many knots tend to reduce the efficiency of the model fitting. Ruppert [18] compared the minimization of the average mean squared errors for different selections of knots and recommended default fixed numbers of knots for all sample sizes, as long as the smooth regression functions do not oscillate excessively. We note, however, that a number of authors have studied the theory of changing the numbers of knots. For example, Huang [8] gave the rates of convergence for the case of a growing number of knots. We take the view that the asymptotic theory for a fixed number of knots provides a sufficiently sound basis for practical statistical inference.

Remark 6. In Theorems 3 and 4, the asymptotic variance–covariance matrices are unknown, as in most cases in practice. They can be estimated with the usual plug-in estimates $\boldsymbol{\Delta}_{1,n}(\hat{\boldsymbol{\theta}})$, $\boldsymbol{\Delta}_{2,n}(\hat{\boldsymbol{\theta}})$ and $\mathbf{J}(\hat{\boldsymbol{\phi}})$. These affect the estimation of the smoothing parameters; see e.g., [26,27]. If the variance–covariance matrix is known, we can easily modify the existing results to obtain the more simplified results. For example, if $\mathbf{W} = \boldsymbol{\Sigma}$, then the asymptotic variance–covariance matrix in Theorem 3 becomes $\boldsymbol{\Delta}_1^{-1}(\boldsymbol{\theta}_0)$ and, as a result, the one in Theorem 4 changes to $\mathbf{J}(\boldsymbol{\phi}_0)\boldsymbol{\Delta}_1^{-1}(\boldsymbol{\phi}_0)\mathbf{J}(\boldsymbol{\phi}_0)^\top$.

Remark 7. While the penalized spline approach for fitting multivariate SIMs has yielded promising analytical and theoretical results, a generalization of the method to situations of high dimensions remains uncertain. Theoretical issues aside, computational burden could be overwhelming. For example, the number of the basis functions can easily become larger than the number of observations, rendering the algorithms unusable. To analyze larger datasets, we suggest to use the penalty approach as it downplays the number and location of knots, whereas the use of a low-rank smoother mitigates the computational problems better than other available approaches; see e.g., [6].

We provided proofs of Theorems 1 and 3 in the Appendix.

3. A simulation study

3.1. Data generation

We conducted a simulation study to evaluate the numerical characteristics of the proposed method. We chose $L = 2$ and let the two index functions be $f_1(v) = v^2 \sin(v)$ and $f_2(v) = ve^v$. We generated two correlated response outcomes $\mathbf{y}_1, \mathbf{y}_2$ from the following model. For each $i \in \{1, \ldots, n\}$,

$$\begin{cases} y_{1,i} = (\alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \alpha_3 x_{3,i})^2 \sin(\alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \alpha_3 x_{3i}) + \beta_1 z_i + \epsilon_{1i}, \\ y_{2,i} = (\alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \alpha_3 x_{3,i}) \exp(\alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \alpha_3 x_{3i}) + \beta_2 z_i + \epsilon_{2,i}, \end{cases}$$

where the independent variables within the index functions, $x_{1,i}$, $x_{2,i}$ and $x_{3,i}$, are assumed to be independent and following a uniform distribution $\mathcal{U}(0, \pi)$. We generated a binary variable z_i from a Bernoulli distribution with $\Pr(z_i = 1) = 0.3$.

The random errors $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, \epsilon_{2,i})^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where the correlation coefficient between the two outcomes was ρ, and σ_1^2 and $\sigma_2^2 = \delta\sigma_1^2$, which represented a heteroscedasticity in the two outcomes.

Table 1
Summary of parameter estimates over 200 simulation runs: true parameters $(\alpha_1, \alpha_2, \alpha_3) = 1/\sqrt{14}(2, -1, 3) = (0.5345, -0.2673, 0.8018)$, $\beta_1 = 10$, $\beta_2 = -30$, $\rho = 0.50$, $\delta = 0.9$, $\sigma = 2$.

Parameter	n	Mean	SE	Bias	MSE
α_1	50	0.5369	0.0005	2.3635e−03	5.8726e−06
	100	0.5349	0.0004	3.4925e−04	2.5836e−07
	200	0.5344	0.0003	−1.7117e−04	9.6444e−08
	500	0.5344	0.0002	−8.8799e−05	4.1073e−08
α_2	50	−0.2696	0.0005	−2.3795e−03	5.9286e−06
	100	−0.2657	0.0005	1.5940e−03	2.8512e−06
	200	−0.2679	0.0002	−6.7369e−04	5.1232e−07
	500	−0.2679	0.0002	−6.0122e−04	3.8904e−07
α_3	50	0.7993	0.0004	−2.4725e−03	6.3077e−06
	100	0.8020	0.0003	2.3071e−04	1.3925e−07
	200	0.8016	0.0002	−1.5116e−04	4.7859e−08
	500	0.8017	0.0001	−1.3019e−04	4.0289e−08
β_1	50	9.9404	0.0501	−5.9577e−02	3.9919e−03
	100	10.0175	0.0286	1.7495e−02	2.8136e−03
	200	9.9847	0.0210	−1.5350e−02	1.0037e−03
	500	9.9865	0.0138	−1.3528e−02	4.2729e−04
β_2	50	−30.0126	0.0517	−1.2617e−02	2.6680e−03
	100	−29.2858	0.0283	1.1419e−02	8.3198e−04
	200	−29.9943	0.0224	5.6557e−03	6.6280e−04
	500	−30.0002	0.0121	−2.1758e−04	3.4807e−04
ρ	50	0.5177	0.0089	1.7664e−02	3.9054e−04
	100	0.5055	0.0055	5.4806e−03	3.6987e−05
	200	0.5026	0.0037	2.6031e−03	4.4087e−05
	500	0.4988	0.0025	−1.1626e−03	7.8169e−06
δ	50	0.9630	0.0099	6.2954e−02	3.9967e−03
	100	0.9614	0.0058	6.1426e−02	3.7872e−03
	200	0.9538	0.0038	5.3801e−02	3.7872e−03
	500	0.9522	0.0026	5.2163e−02	2.9938e−03
σ	50	1.9703	0.0147	−2.9682e−02	7.6571e−04
	100	1.9766	0.0103	−2.3439e−02	9.8797e−04
	200	1.9810	0.0073	−1.8969e−02	4.1309e−04
	500	1.9966	0.0046	−3.3625e−03	3.2598e−05

Point estimates for the regression parameters and variance components were averaged over 200 simulation runs. Claeskens et al. [5] studied the class of penalized spline estimates, and showed that penalized splines using truncated polynomial basis functions leads to the optimal rate of convergence independent of the assumption made on the number of knots, even though their main theoretical results are derived under the assumption that the number of knots tends to infinity at a rate of $o(n)$ (Assumption 3 in Claeskens et al. [5]).

In this simulation study, we used 20 knots to fit cubic spline models. This number of knots was chosen to ensure the capture of the basic features of the regression functions [20]. We also conducted a sensitivity analysis to assess the influence of the numbers of knots to model performance. In each model fit, we used equally spaced interior knots. The smoothing parameters were chosen by ratio of variance components estimated from restricted maximum likelihood (REML) [12,20].

3.2. Simulation results

Simulation results were presented for the case where $\boldsymbol{\alpha} = 1/\sqrt{14}(2, -1, 3)^\top$, $\beta_1 = 10$, $\beta_2 = -30$, $\rho = 0.5$, $\sigma = 2$, $\delta = 0.9$. Four different sample sizes were considered: $n \in \{50, 100, 200, 500\}$. For evaluation, we compared the estimated values of parameters against the true values. The parameter estimation results, including the mean values of the parameter estimates (Mean), standard error (SE), bias and mean squared error (MSE), are summarized in Table 1. The estimated coefficient values are close to the true values, and the standard errors estimated based on the multivariate model are consistently smaller. In addition, the empirical standard errors of the estimates of the variance–covariance components are close to the true values. Clearly, both bias and MSE of each parameter estimates decrease with an increasing number of subjects.

Fig. 1 depicts the average cubic-spline estimates fit to two correlated outcomes with 500 subjects based on 200 simulated datasets, as well as the corresponding 2.5% and 97.5% quantiles. The proposed estimates are well behaved, with a minimal estimation bias, although the bias is slightly larger in the tails. Meanwhile, both 2.5% and 97.5% quantiles are close to the true curves, showing very small variation in the estimates.

We also conducted a sensitivity analysis to compare the model performance, when three different numbers of knots, 10, 15 and 20 were used. Point estimates for the single index parameters were averaged over 200 simulation runs. A sample size $n = 500$ was considered.

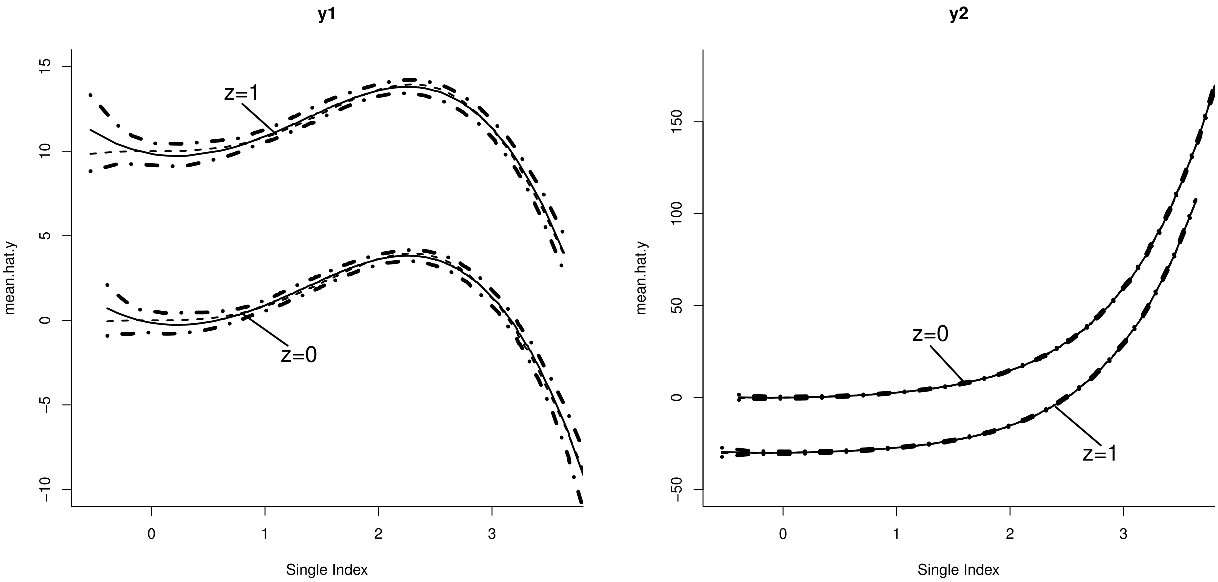

Fig. 1. Curve estimates and confidence bands for the simulated data with bivariate outcomes. The solid curves are the true mean functions; the dashed curves are the average cubic-spline fit over 200 simulations. The dot-dashed curves are the corresponding 2.5% and 97.5% quantiles.

Table 2
Summary of parameter estimates over 200 simulation runs: true parameters $(\alpha_1, \alpha_2, \alpha_3) = 1/\sqrt{14}(2, -1, 3) = (0.5345, -0.2673, 0.8018)$, $\rho = 0.50$, $\delta = 0.9$, $\sigma = 2$.

Parameter	Number of Knots	Mean	SE	Bias	MSE
α_1	10	0.5344	0.0002	−3.8961e−04	3.3118e−07
	15	0.5343	0.0002	−2.0427e−04	7.6622e−08
	20	0.5343	0.0002	−5.1607e−05	3.7761e−08
α_2	10	−0.2676	0.0002	−3.6114e−04	1.5356e−07
	15	−0.2674	0.0002	−1.7599e−04	5.5722e−08
	20	−0.2674	0.0002	−1.0535e−04	3.6330e−08
α_3	10	0.8016	0.0001	−1.0325e−04	2.6445e−08
	15	0.8018	0.0001	6.7844e−05	2.2531e−08
	20	0.8018	0.0001	−1.0371e−05	1.7531e−08

Table 2 presents mean values of the parameter estimates, standard error, bias and mean squared error. The results indicate that parameter estimates were only slightly impacted when using different knots. Thus, our method is quite robust against the variation in the number of knots.

A complete estimation algorithm, including sample R code, and more comprehensive simulation studies have been discussed in detail previously [32].

4. An application

To illustrate the use of the proposed method, we analyzed the data from an observational study of healthy young adults. The study protocol was described elsewhere [25]. Briefly, healthy young volunteers between 21 and 39 years of age were recruited for participation in this observational study. Enrolled subjects were followed prospectively with semiannual assessments, for up to two years. During the course of follow-up, the participants' blood pressure was recorded; blood and overnight urine samples were collected at each follow-up visit.

In the current analysis, we evaluated the influence of dietary sodium and potassium intake on systolic and diastolic blood pressure (SBP and DBP). Here we used the logarithm of the calculated urinary sodium (UNACR) and potassium (UKCR) excretion rates (adjusted for creatinine) to approximate the dietary intake. The focus of the analysis is to derive an index measure of urinary sodium and potassium excretion rates for prediction of systolic and diastolic blood pressure. As a result, we considered the two urinary metrics as the main ingredients of the index measure. Other known correlates of blood pressure, such as age, sex, race and body mass index, were included in the model as covariates (denoted by **Z** vector). The model is as follows:

$$\begin{cases} \mathrm{SBP}_i = \eta_s(\alpha_1 \ln \mathrm{UNACR}_i + \alpha_2 \ln \mathrm{UKCR}_i) + \boldsymbol{\beta}_s^\top \mathbf{Z}_i + \epsilon_i^s, \\ \mathrm{DBP}_i = \eta_d(\alpha_1 \ln \mathrm{UNACR}_i + \alpha_2 \ln \mathrm{UKCR}_i) + \boldsymbol{\beta}_d^\top \mathbf{Z}_i + \epsilon_i^d. \end{cases}$$

Table 3
Summary of linear covariates in the fitted index model.

Response	Parameter estimate (95% confidence interval)			
	Age	Male	Black	BMI
SBP	0.0932 (−0.2342, 0.4207)	9.2706 (6.6056, 11.9356)	−2.0343 (−4.8590, 0.7902)	0.6242 (0.4563, 0.7920)
DBP	0.6570 (0.3649, 0.9490)	4.0721 (−0.6953, 8.4490)	0.0493 (−2.4699, 2.5685)	0.4546 (0.3049, 0.6043)

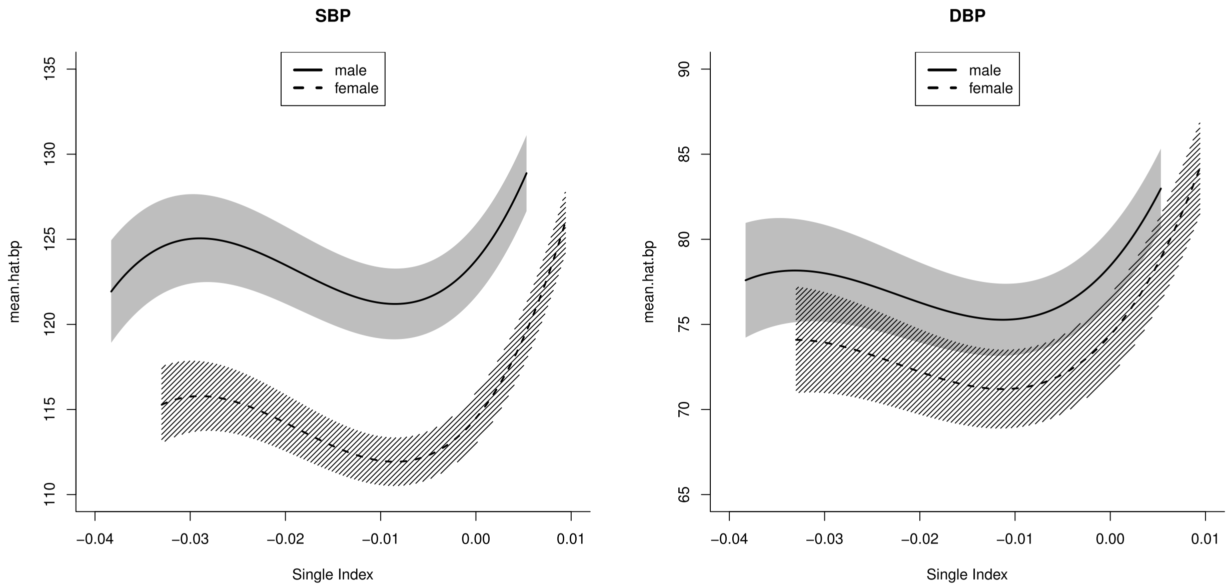

Fig. 2. Relations between the index and blood pressure.

Our analysis was based on data from 228 study participants. The sample included 109 males and 88 blacks. The mean age at study entry was 30 years (SD = 4.12 years); average BMI was 29.63 (SD = 8.22). The mean values for urinary sodium and potassium excretion rates were 0.079 and 0.02 mmol/mg Creatinine, respectively. The estimated index parameters are 0.1156 and −0.9933 for urinary sodium and potassium excretion rates, respectively. Model fitting results for the linear covariates are presented in Table 3.

We present the relations between the index function value as systolic and diastolic blood pressure in the following figures. Fig. 2 clearly shows that the index values were gradually positively associated with both systolic and diastolic blood pressure, although the relations were not linear. In particular, the non-negative index values were more strongly associated with the increase of blood pressure.

A more careful examination of the index coefficients revealed that urine sodium is positively associated with blood pressure and that urine potassium is negatively associate with blood pressure. This finding is consistent with the observations from recent studies that sodium-rich and potassium-poor diet increase the risk of hypertension [1,16]. The finding also corroborates the results from the latest mechanistic investigations on the roles that potassium plays in kidney reabsorption of sodium and its impact on blood pressure elevation [24].

5. Discussion

Use of medical indices has been ubiquitous in clinical practice and research investigation. A main motivation for using index measures is to succinctly quantify the latent characteristics in individual subjects. Towards that end, calculated values of the indices are linked to patient outcomes for associational or causal interpretations. Despite the popularity of medical indices, few have carefully examined the methodological issues concerning index development. To the best of our knowledge, most of the existing medical indices are constructed in a subjective fashion, based primarily on opinions of clinical experts. The resulting index may or may not perform in ways that its designer had hoped.

In this paper, we described a method for analytical derivation of index measures, through the use of multivariate single-index model. By linking subject characteristics to a set of pre-specified outcomes, we were able to derive index measures that were optimally linked to the outcomes. We would like to think that this multivariate extension of the single-index model has enhanced the possibility of achieving optimally performing, multi-purpose index measures. We showed that the method was easily implementable and had a good performance. The simulations suggested that the method could also reliably recover the true shapes of the nonlinear link functions, which further assured us of its potential.

One implicit assumption used in this research is that the unknown link function is adequately described by a polynomial spline function, so that the analyst can closely approximate the link function by estimating the spline coefficients. As in all analyses, the validity of parameter estimation and statistical inference depends to a large extent on the appropriateness of the underlying model specification. This is true for the proposed multivariate SIM as it is for the simple linear regression, where the analyst assumes the relationship is adequately depicted by a linear function. But compared to other statistical models, spline models are less vulnerable to misspecification of the functional relationship, because with appropriately chosen knots and estimated coefficient from the data, they could accommodate virtually all types and shapes of nonlinear relationships [19]. Penalized spline estimates can be achieved with expedient computational algorithms [18]. This said, as in other statistical models, there is always an uncertainty associated with model specification. But all things considered, such an uncertainty is generally much smaller than those associated with fixed functional forms [37].

Along with the model fitting methods, we describe theoretical properties of the model parameters in this paper. In principle, the proposed model provides an extension of single-index model from scalar response to multivariate response, and generalizes homogeneous random errors to heterogeneous errors with working covariance matrices. We showed that under mild regularity conditions, the parameter estimators were consistent and asymptotically normally distributed. As a result, large sample inference becomes possible for this class of models. Collectively, these theoretical results have provide a valid foundation for using the model in index development.

Finally, we note that the ultimate test of an index measure is validation through independent samples. This said, the proposed method gives a good starting place for constructing a useful measure that is linked to pre-specified outcomes. The method itself does not preclude the need for independent validation. Notwithstanding this limitation, we contend that this research has added a more detailed theoretical description of this new index development tool.

Acknowledgments

This work was supported by NIH grants R01 095086 and AA R01 025208. The authors thank the Editor-in-Chief, Christian Genest, an Associate Editor and two anonymous referees for detailed comments and suggestions that improved the paper.

Appendix. Sketches of proofs

We establish two lemmas in preparation for the proof of the main theorems.

Lemma A. *If Conditions 1 and 2 hold, then, as* $n \to \infty$, $(\boldsymbol{\epsilon}_1^\top \mathbf{W}_1^{-1}\boldsymbol{\epsilon}_1 + \cdots + \boldsymbol{\epsilon}_n^\top \mathbf{W}_n^{-1}\boldsymbol{\epsilon}_n)/n \to \tau$ *a.s., where* τ *is given in Condition 2.*

Lemma B. *If Conditions 2 and 4 hold, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i \rightsquigarrow \mathcal{N}[\mathbf{0}, \boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)], \tag{A.1}$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *denotes the multivariate normal distribution with mean* $\boldsymbol{\mu}$ *and variance–covariance matrix* $\boldsymbol{\Sigma}$, *and* $\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)$ *is given in Condition 4.*

Proof of Lemma A. We approach the proof in a way similar to that of Lemma 4.1 of Tian et al. [23]. First we note that

$$\mathrm{E}(\boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i) = \mathrm{E}\{\mathrm{tr}(\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top)\} = \mathrm{tr}\{\mathbf{W}_i^{-1}\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^\top)\} = \mathrm{tr}(\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i).$$

Let $X_i = \boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i - \mathrm{tr}(\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i)$. Then $X_1, \ldots, X_n$ are mutually independent random variables with zero mean. From Condition 2, there exist constants c_1, c_2 such that

$$0 < c_1 \le \min_{i\in\{1,\ldots,n\}} \xi_{(1)i} \le \max_{i\in\{1,\ldots,n\}} \xi_{(L)i} \le c_2 < \infty,$$
$$0 < c_1 \le \min_{i\in\{1,\ldots,n\}} \zeta_{(1)i} \le \max_{i\in\{1,\ldots,n\}} \zeta_{(L)i} \le c_2 < \infty,$$

where $\xi_{(1)i}, \xi_{(L)i}$ and $\zeta_{(1)i}, \zeta_{(L)i}$ are the minimal, maximal eigenvalues of $\boldsymbol{\Sigma}_i$, $\mathbf{W}_i$ for $i \in \{1, \ldots, n\}$, respectively. Thus

$$\mathrm{tr}(\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i) \le L \max_{i\in\{1,\ldots,n\}} \xi_{(L)i} / \min_{i\in\{1,\ldots,n\}} \zeta_{(1)i} \le L c_2/c_1,$$

and, for all $i \in \{1, \ldots, n\}$, $0 \le \boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i \le \zeta_{(1)i}^{-1}\boldsymbol{\epsilon}_i^\top\boldsymbol{\epsilon}_i \le c_1^{-1}\boldsymbol{\epsilon}_i^\top\boldsymbol{\epsilon}_i$.

Let $p = (2 + r)/2 > 1$, where $r > 0$ is given in Condition 1. Using the inequality $\mathrm{E}\{(X+Y)^p\} \le 2^{p-1}\{\mathrm{E}(X^p) + \mathrm{E}(Y^p)\}$ for $X \ge 0, Y \ge 0$ (see, e.g., Inequality (19) on p. 50 in [4]), we deduce that, for all $i \in \{1, \ldots, n\}$,

$$\mathrm{E}\{(\boldsymbol{\epsilon}_i^\top\boldsymbol{\epsilon}_i)^p\} = \mathrm{E}\left(\sum_{\ell=1}^{L} \epsilon_{\ell,i}^2\right)^p \le 2^{(L-1)(p-1)} \sum_{\ell=1}^{L} \mathrm{E}(\epsilon_{\ell,i}^{2p}) \le 2^{(L-1)(p-1)} L \max_{1\le\ell\le L} \mathrm{E}(\epsilon_{\ell,i}^{2p}) \le 2^{(L-1)(p-1)} c_0 L,$$

where c_0 is given in Condition 1. Hence, for all $i \in \{1, \ldots, n\}$,

$$\mathrm{E}|X_i|^p \le 2^{p-1}\{\mathrm{E}|\boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i|^p + |\mathrm{tr}(\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i)|^p\} \le 2^{p-1}[\zeta_{(1)i}^{-p}\mathrm{E}\{(\boldsymbol{\epsilon}_i^\top\boldsymbol{\epsilon}_i)^p\} + (Lc_2/c_1)^p] \le 2^{p-1}\{c_1^{-p}2^{(L-1)(p-1)}c_0L + (Lc_2/c_1)^p\}.$$

Consequently,

$$\sum_{i=1}^{\infty} \mathrm{E}|X_i|^p/i^p < \infty.$$

By the Strong Law of Large Numbers (see, e.g., the corollary on p. 132 in [4], Theorem 3.1 in [22], and Corollary 8.2 in [11]), we have, as $n \to \infty$, $(X_1 + \cdots + X_n)/n \xrightarrow{\text{a.s.}} 0$. This yields the desired result in view of Condition 2. □

Proof of Lemma B. Let $\boldsymbol{Z}_n = \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i/\sqrt{n}$. As $\boldsymbol{\epsilon}_i$ has mean zero and covariance matrix $\boldsymbol{\Sigma}_i$, one gets

$$\mathrm{E}(\boldsymbol{Z}_n) = \mathbf{0}, \quad \mathrm{var}(\boldsymbol{Z}_n) = \frac{1}{n}\sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^\top \mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{W}_i^{-1}\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0).$$

By the Cramér–Wold device, in order to prove (A.1), it suffices to show

$$\mathbf{u}^\top\mathbf{Z}_n \rightsquigarrow \mathcal{N}[\mathbf{0}, \mathbf{u}^\top\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)\mathbf{u}], \tag{A.2}$$

for an arbitrary unit vector $\mathbf{u}$. For each $i \in \{1, \ldots, n\}$, let $c_i^2 = \mathbf{u}^\top\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^\top\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{W}_i^{-1}\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)\mathbf{u}$. Since $\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)$ has full rank and $\mathbf{W}_i$, $\boldsymbol{\Sigma}_i$ are positive definite, $c_i > 0$, we can write

$$\mathbf{u}^\top\boldsymbol{Z}_n = n^{-1/2}\sum_{i=1}^n c_ie_i$$

with $e_i = c_i^{-1}\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^\top\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i$, whence $e_1, \ldots, e_n$ are mutually independent with zero mean and unit variance. By Condition 4, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n c_i^2 = \mathbf{u}^\top\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)\mathbf{u},$$

where $\boldsymbol{\Delta}_2(\boldsymbol{\theta})$ is given in Condition 4. It follows from Lemma 3 of [30] that

$$\max_{i\in\{1,\ldots,n\}} c_i^2/(c_1^2 + \cdots + c_n^2) \to 0.$$

This implies the Lindeberg condition. We now invoke the Lindeberg–Feller Central Limit Theorem to claim (A.2). □

Proof of Theorem 1. Our proof is a modification of that of Theorem 1′ of Yu and Ruppert [38]. The main differences between our Theorem 1 and theirs are (i) we generalize their theory from scalar to multivariate responses; (ii) we relax the homogeneous random errors; and (iii) we use working matrices. We shall omit the details and only give sketches to those where exist important differences.

We first show the existence. With known $\boldsymbol{\phi}$, the minimization of (8) becomes a problem of ridge regression, which has the explicit solution for the penalized least squares estimator $\hat{\boldsymbol{\beta}}$, given by

$$\hat{\beta}(\boldsymbol{\phi}) = \left[\frac{1}{n}\sum_{i=1}^n \mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\mathbf{W}_i^{-1}\mathbf{B}_\circ\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\} + \mathbb{D}(\boldsymbol{\lambda})\right]^{-1}\frac{1}{n}\sum_{i=1}^n \mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\mathbf{W}_i^{-1}\mathbf{y}_i, \tag{A.3}$$

where $\mathbb{D}(\boldsymbol{\lambda})$ is the diagonal matrix consisting of the diagonal matrices $\lambda_\ell\mathbf{D}_\ell$ for $\ell \in \{1, \ldots, L\}$, i.e., $\mathbb{D}(\boldsymbol{\lambda}) = \mathrm{diag}(\lambda_1\mathbf{D}_1, \ldots, \lambda_L\mathbf{D}_L)$. Using this explicit solution, the minimization problem in (8) with respect to both $\boldsymbol{\beta} \in \mathbb{R}^{\dim(\boldsymbol{\beta})}$ and $\boldsymbol{\phi} \in \Phi$ boils down to the following minimization with respect to $\boldsymbol{\phi} \in \Phi$:

$$Q_n(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^n[\mathbf{y}_i - \mathbf{B}_\circ\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\hat{\boldsymbol{\beta}}(\boldsymbol{\phi})]^\top\mathbf{W}_i^{-1}[\mathbf{y}_i - \mathbf{B}_\circ\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\hat{\boldsymbol{\beta}}(\boldsymbol{\phi})].$$

It is clear from this explicit formula that the compactness for the regression parameter $\boldsymbol{\beta}$ is not needed. Meanwhile, the existence of the least squares estimator $\hat{\boldsymbol{\phi}}$ follows immediately from the usual existence theorem in calculus as the objective function $Q_n(\boldsymbol{\phi})$ is continuous over the compact set Φ.

We now show that $\hat{\boldsymbol{\phi}} \to \boldsymbol{\phi}_0$ a.s., where $Q_n(\hat{\boldsymbol{\phi}}) = \inf_{\boldsymbol{\phi}\in\Phi} Q_n(\boldsymbol{\phi})$. For each $i \in \{1, \ldots, n\}$, let

$$\mathbf{a}_{n,i}(\boldsymbol{\phi}) = \mathbf{B}_\circ\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi}_0)\}\boldsymbol{\beta}_0 - \mathbf{B}_\circ\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}).$$

Simple algebra leads to

$$Q_n(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i + 2\frac{1}{n}\sum_{i=1}^{n}\mathbf{a}_{n,i}^\top(\boldsymbol{\phi})\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i + \frac{1}{n}\sum_{i=1}^{n}\mathbf{a}_{n,i}^\top(\boldsymbol{\phi})\mathbf{W}_i^{-1}\mathbf{a}_{n,i}(\boldsymbol{\phi}) = A_1 + 2A_2(\boldsymbol{\phi}) + A_3(\boldsymbol{\phi}),$$

say. It follows from Lemma A that $A_1 \to \tau$ a.s. as $n \to \infty$. We show below that, as $n \to \infty$,

$$A_2(\boldsymbol{\phi}) \to 0, \quad A_3(\boldsymbol{\phi}) \to \boldsymbol{\beta}_0^\top \mathbf{T}(\boldsymbol{\phi})\boldsymbol{\beta}_0, \text{ a.s.} \tag{A.4}$$

uniformly over Φ. Here $\mathbf{T}(\boldsymbol{\phi})$ is given in Condition 3 and has the unique zero at $\boldsymbol{\phi}_0$. Thus, as $n \to \infty$,

$$Q_n(\boldsymbol{\phi}) \to \boldsymbol{\beta}_0^\top \mathbf{T}(\boldsymbol{\phi})\boldsymbol{\beta}_0 + \tau \text{ a.s.}$$

uniformly in $\boldsymbol{\phi} \in \Phi$. The rest of the proof for $\hat{\boldsymbol{\phi}} \to \boldsymbol{\phi}_0$ a.s. is similar to the proof of the theorem of Yu and Ruppert [37].

To compute the first limit in (A.4), we write $A_2(\boldsymbol{\phi}) = A_{21} + A_{22}(\boldsymbol{\phi})$, where

$$A_{21} = \boldsymbol{\beta}_0^\top \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi}_0)\}\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i\right], \quad A_{22}(\boldsymbol{\phi}) = \hat{\boldsymbol{\beta}}(\boldsymbol{\phi})^\top \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i\right].$$

It can be shown that under Condition 3, as $n \to \infty$,

$$\boldsymbol{\xi}_n(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i \to 0 \text{ a.s.} \tag{A.5}$$

uniformly in $\boldsymbol{\phi} \in \Phi$. In fact, fix $\boldsymbol{\phi} \in \Phi$ and $\varepsilon > 0$, as $n \to \infty$. Under Condition 3, there exist an event Ω_ϕ with $\Pr(\bar{\Omega}_\phi) = 0$ and an integer n_ϕ such that $\|\boldsymbol{\xi}_n(\boldsymbol{\phi})\| < \varepsilon/2$ holds over Ω_ϕ for $n > n_\phi$, using a proof similar to that of Theorem 5.1.2 of Chung [4]. Let

$$\bar{\mathbf{G}}_n(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_i(\boldsymbol{\phi})\mathbf{W}_i^{-1}\mathbf{G}_i^\top(\boldsymbol{\phi}),$$

where $\mathbf{G}_i(\boldsymbol{\phi}) = \mathbf{B}_\circ^\top\{\mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})\}$. It follows from Condition 3 that $\bar{\mathbf{G}}_n(\boldsymbol{\varphi}) - \bar{\mathbf{G}}_n(\boldsymbol{\phi}) \to \mathbf{S}(\boldsymbol{\varphi},\boldsymbol{\varphi}) - \mathbf{S}(\boldsymbol{\phi},\boldsymbol{\phi})$ uniformly in $\boldsymbol{\varphi}$ as $n \to \infty$. Since each $\mathbf{G}_i$ is continuous, the uniform convergence in Condition 3 implies that $\mathbf{S}(\boldsymbol{\varphi},\boldsymbol{\varphi})$ is continuous. As a result, given an arbitrary $\varepsilon > 0$, there exists a neighborhood N_ϕ of $\boldsymbol{\phi}$, such that for all $\boldsymbol{\varphi} \in N_\phi$, $\|\bar{\mathbf{G}}_n(\boldsymbol{\varphi}) - \bar{\mathbf{G}}_n(\boldsymbol{\phi})\| < \varepsilon/(4\tau)$. By Lemma A, there exist an event Ω_0 with $P(\bar{\Omega}_0) = 0$ and an integer n_0 such that over Ω_0, when $n > n_0$,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i\right\| \le 2\tau.$$

Consequently, for all $\boldsymbol{\varphi} \in N_\phi$, over $\Omega'_\phi = \Omega_0 \cap \Omega_\phi$ with $P(\bar{\Omega}'_\phi) = 0$, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\|\boldsymbol{\xi}_n(\boldsymbol{\varphi})\| &\le \|\boldsymbol{\xi}_n(\boldsymbol{\varphi}) - \boldsymbol{\xi}_n(\boldsymbol{\phi})\| + \|\boldsymbol{\xi}_n(\boldsymbol{\phi})\| \\ &\le \|\bar{\mathbf{G}}_n(\boldsymbol{\varphi}) - \bar{\mathbf{G}}_n(\boldsymbol{\phi})\| \times \left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^\top \mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i\right\| + \|\boldsymbol{\xi}_n(\boldsymbol{\phi})\| < \varepsilon/2 + \varepsilon/2 = \varepsilon,\end{aligned} \tag{A.6}$$

for $n > n'_\phi = \max(n_0, n_\phi)$. As $\{\Omega'_\phi\}$ is an open cover of the compact Φ, it follows from the Heine–Borel Theorem that one can choose a finite subcover $\Omega'_{\phi_1}, \ldots, \Omega'_{\phi_K}$. Let $\Omega' = \cap_k \Omega'_{\phi_k}$ and $n' = \max(n'_{\phi_k})$. Then $\Pr(\bar{\Omega}') = 0$, and for all $\boldsymbol{\phi} \in \Phi$, when $n > n'$, we have $\|\boldsymbol{\xi}_n(\boldsymbol{\phi})\| < \varepsilon$ over Ω'. This shows the almost sure uniform convergence in (A.5).

Recall that the weighted penalized spline least squares estimators depend on the values of $\boldsymbol{\lambda}_n = (\lambda_{n,1}, \ldots, \lambda_{n,L})^\top$. To simplify the notation, we write the response vector as $\mathbf{Y}_n = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ and the error vector as $\mathcal{E}_n = (\boldsymbol{\epsilon}_1^\top, \ldots, \boldsymbol{\epsilon}_n^\top)^\top$. We then define $\mathbb{W}_n = \operatorname{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_n)$ and

$$\mathbb{B}_n = \mathbb{B}_n(\boldsymbol{\phi}) = (\mathbf{B}_\circ^\top\{\mathbf{X}_1\boldsymbol{\alpha}(\boldsymbol{\phi})\}, \ldots, \mathbf{B}_\circ^\top\{\mathbf{X}_n\boldsymbol{\alpha}(\boldsymbol{\phi})\})^\top, \quad \mathbb{B}_{n,0} = \mathbb{B}_n(\boldsymbol{\phi}_0).$$

Using a matrix notation and noting $\mathbf{Y}_n = \mathbb{B}_{n,0}\boldsymbol{\beta}_0 + \mathcal{E}_n$, we write (A.3) as

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}) &= \{\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbb{B}_n(\boldsymbol{\phi})/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbf{Y}_n/n \\ &= \{\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbb{B}_n(\boldsymbol{\phi})/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbb{B}_{n,0}\boldsymbol{\beta}_0/n \\ &\quad + \{\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbb{B}_n(\boldsymbol{\phi})/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathcal{E}_n/n.\end{aligned} \tag{A.7}$$

By Condition 3, one has, for $\boldsymbol{\phi}, \boldsymbol{\phi}_0 \in \Phi$, and as $n \to \infty$,

$$\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbb{B}_n(\boldsymbol{\phi})/n \to \mathbf{S}(\boldsymbol{\phi},\boldsymbol{\phi}), \quad \mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathbb{B}_n(\boldsymbol{\phi}_0)/n \to \mathbf{S}(\boldsymbol{\phi},\boldsymbol{\phi}_0) \text{ a.s.} \tag{A.8}$$

By (A.5), one also has, as $n \to \infty$,

$$\mathbb{B}_n^\top(\boldsymbol{\phi})\mathbb{W}_n^{-1}\mathcal{E}_n/n \to 0 \text{ a.s.} \tag{A.9}$$

Both (A.8) and (A.9) hold uniformly over Φ. As $\boldsymbol{\lambda}_n = o(1)$, it follows that, as $n \to \infty$,

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}) \to \mathbf{S}^{-1}(\boldsymbol{\phi}, \boldsymbol{\phi})\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}_0)\boldsymbol{\beta}_0 \text{ a.s.} \tag{A.10}$$

uniformly in $\boldsymbol{\phi} \in \Phi$. Thus, $A_{22}(\boldsymbol{\phi}) \to 0$, which takes care of the first limit in (A.4).

Using simple algebra, one finds

$$\begin{aligned} A_3(\boldsymbol{\phi}) = \frac{1}{n}\Big[&[\mathbf{I} - \mathbb{B}_n\{\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top \mathbb{W}_n^{-1}/n]\mathbb{B}_{n,0}\boldsymbol{\beta}_0 \\ &-\mathbb{B}_n\{\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathcal{E}_n/n\Big]^\top \mathbb{W}_n^{-1} \\ &\times \Big[[\mathbf{I} - \mathbb{B}_n\{\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top \mathbb{W}_n^{-1}/n]\mathbb{B}_{n,0}\boldsymbol{\beta}_0 \\ &- \mathbb{B}_n\{\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1}\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathcal{E}_n/n\Big]. \end{aligned}$$

Let $\mathbf{H}_n = \mathbb{B}_n(\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1}\mathbb{B}_n^\top$. Because the dimensions of $\mathbb{D}$ and $\boldsymbol{\lambda}_n$ are fixed (independent of n) and $\boldsymbol{\lambda}_n = o(1)$, we have

$$\begin{aligned} \{\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n + \mathbb{D}(\boldsymbol{\lambda}_n)\}^{-1} &= (\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1} + (\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1}\mathbb{D}(\boldsymbol{\lambda}_n)(\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1} \\ &+ (\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1}\mathbb{D}(\boldsymbol{\lambda}_n)(\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1}\mathbb{D}(\boldsymbol{\lambda}_n)(\mathbb{B}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n)^{-1} + o(\|\boldsymbol{\lambda}_n\|^2). \end{aligned}$$

Thus

$$\begin{aligned} A_3(\boldsymbol{\phi}) &= \boldsymbol{\beta}_0^\top\{\mathbb{B}_{n,0}^\top \mathbb{W}_n\mathbb{B}_{n,0}/n - \mathbb{B}_{n,0}^\top \mathbb{W}_n\mathbb{B}_n/n(\mathbb{B}_n^\top \mathbb{W}_n\mathbb{B}_n/n)^{-1}\mathbb{B}_n^\top \mathbb{W}_n\mathbb{B}_{n,0}/n\}\boldsymbol{\beta}_0 \\ &+ \mathcal{E}_n^\top \mathbb{W}_n^{-1}\mathbf{H}_n\mathbb{W}_n^{-1}\mathbf{H}_n\mathbb{W}_n^{-1}\mathcal{E}_n/n + o(n^{-1/2}) = A_{31}(\boldsymbol{\phi}) + A_{32}(\boldsymbol{\phi}) + o(n^{-1/2}). \end{aligned}$$

By (A.5), we have $\mathcal{E}_n^\top \mathbb{W}_n^{-1}\mathbb{B}_n/n \to 0$ as $n \to \infty$. By Condition 3, $\mathbb{B}_n^\top \mathbb{W}_n\mathbb{B}_n/n \to \mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi})$ as $n \to \infty$. Both of these limits hold uniformly almost surely over Φ. Consequently, $A_{32}(\boldsymbol{\phi}) \to 0$ as $n \to \infty$, which takes care of the second limit in (A.4).

Finally, we write

$$\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\phi}}) - \boldsymbol{\beta}_0 = \{\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\phi}}) - \hat{\boldsymbol{\beta}}(\boldsymbol{\phi}_0)\} + \{\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}_0) - \boldsymbol{\beta}_0\}.$$

As in the proof of Theorem 1′ of [38], we argue that the first difference converges to zero almost surely in view of the uniform convergence in (A.10). To deal with the second difference, we write

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}_0) - \boldsymbol{\beta}_0 = [\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}_0) - \mathrm{E}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}_0)\}] + [\mathrm{E}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}_0)\} - \boldsymbol{\beta}_0].$$

The second difference is the bias of the ridge regression estimator and tends to zero as $\boldsymbol{\lambda}_n = o(1)$, while the first difference tends to zero almost surely in view of (A.7)–(A.9). Combining the above yields $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\phi}}) \to \boldsymbol{\beta}_0$ almost surely. This completes the proof of Theorem 1. □

Proof of Remark 2. The result can be proved similar to the uniform convergence in (A.5). Let $\mathbf{T}_i(\boldsymbol{\phi}) = \mathbf{X}_i\boldsymbol{\alpha}(\boldsymbol{\phi})$. By assumption, $\|\mathbf{T}_i(\boldsymbol{\phi})\| \le \|\mathbf{X}_i\| \le LM_0$ as $\|\boldsymbol{\alpha}(\boldsymbol{\phi})\| = 1$. As $\mathbf{B}_\circ(\mathbf{u})$ is continuous, there is a constant $K_1 > 0$ such that $\|\mathbf{B}_\circ\{\mathbf{T}_i(\boldsymbol{\phi})\}\| \le K_1$ for $\boldsymbol{\phi} \in \Phi$ for all $i \in \{1, \ldots, n\}$. By assumption, $(1 - \|\boldsymbol{\phi}\|^2)^{-1} \ge (1 - r_0^2)^{-1}$ for $\boldsymbol{\phi} \in \Phi$. Hence the derivative of $\mathbf{T}_i(\boldsymbol{\phi})$, which contains $(1 - \|\boldsymbol{\phi}\|^2)^{-1}$, is bounded over Φ for all $i \in \{1, \ldots, n\}$. This implies that the derivative of each entry of $\mathbf{B}_\circ\{\mathbf{T}_i(\boldsymbol{\phi})\}$ is bounded. Therefore, for all $i \in \{1, \ldots, n\}$ and $\boldsymbol{\phi}, \boldsymbol{\varphi} \in \Phi$,

$$\|\mathbf{B}_\circ\{\mathbf{T}_i(\boldsymbol{\varphi})\} - \mathbf{B}_\circ\{\mathbf{T}_i(\boldsymbol{\phi})\}\| \le K_2\|\boldsymbol{\varphi} - \boldsymbol{\phi}\|,$$

for some constant $K_2 > 0$.

Let $\mathbf{S}_n(\boldsymbol{\phi}, \boldsymbol{\phi}^*)$ be the average in (10). Then

$$\|\mathbf{S}_n(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*) - \mathbf{S}_n(\boldsymbol{\phi}, \boldsymbol{\phi}^*)\| \le K_2 m_0^{-1}(\|\boldsymbol{\phi} - \boldsymbol{\varphi}\| + \|\boldsymbol{\phi}^* - \boldsymbol{\varphi}^*\|), \tag{A.11}$$

where m_0 is the common lower bound of the eigenvalues of $\mathbf{W}_i$. This implies that $\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*)$ is continuous as

$$\begin{aligned} \|\mathbf{S}(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*) - \mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*)\| &\le \|\mathbf{S}(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*) - \mathbf{S}_n(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*)\| + \|\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*) - \mathbf{S}_n(\boldsymbol{\phi}, \boldsymbol{\phi}^*)\| \\ &+ \|\mathbf{S}_n(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*) - \mathbf{S}_n(\boldsymbol{\phi}, \boldsymbol{\phi}^*)\| \to 0, \end{aligned}$$

with $\|\boldsymbol{\varphi} - \boldsymbol{\phi}\| + \|\boldsymbol{\varphi}^* - \boldsymbol{\phi}^*\| \to 0$ and $n \to \infty$. Here the above first two terms go to zero by assumption while the third term goes to zero by (A.11). The uniform convergence in (10) can be proved by applying the Heine–Borel Theorem in a similar way to the uniform convergence in (A.5) with a similar decomposition to (A.6) as follows:

$$\begin{aligned} \|\mathbf{S}_n(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*) - \mathbf{S}(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*)\| &\le \|\mathbf{S}_n(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*) - \mathbf{S}_n(\boldsymbol{\phi}, \boldsymbol{\phi}^*)\| + \|\mathbf{S}_n(\boldsymbol{\phi}, \boldsymbol{\phi}^*) - \mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*)\| \\ &+ \|\mathbf{S}(\boldsymbol{\phi}, \boldsymbol{\phi}^*) - \mathbf{S}(\boldsymbol{\varphi}, \boldsymbol{\varphi}^*)\| \to 0, \end{aligned}$$

as $\|\boldsymbol{\varphi} - \boldsymbol{\phi}\| + \|\boldsymbol{\varphi}^* - \boldsymbol{\phi}^*\| \to 0$ and $n \to \infty$. This concludes the proof. □

Proof of Theorem 3. Note first that the consistent estimator $\hat{\boldsymbol{\theta}}_n$ minimizes (8), i.e.,

$$Q_{n,\lambda_n}(\boldsymbol{\theta}) = Q_n(\boldsymbol{\theta}) + \sum_{\ell=1}^{L} \lambda_{n,\ell}\boldsymbol{\beta}_\ell^\top \mathbf{D}_\ell\boldsymbol{\beta}_\ell, \tag{A.12}$$

where $Q_n(\boldsymbol{\theta})$ is given in (7). Recall our convention $\boldsymbol{\theta} = \boldsymbol{\theta}_{\boldsymbol{\phi}} = (\boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top)^\top$. Expanding the derivative $\dot{Q}_{n,\lambda_n}(\hat{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$ to the first order, we derive that $\hat{\boldsymbol{\theta}}_n$ must satisfy

$$0 = \dot{Q}_{n,\lambda_n}(\hat{\boldsymbol{\theta}}_n) = \dot{Q}_{n,\lambda_n}(\boldsymbol{\theta}_0) + \ddot{Q}_{n,\lambda_n}(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where $\bar{\boldsymbol{\theta}}$ is a vector lying between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$ and $\ddot{Q}_{n,\lambda_n}(\boldsymbol{\theta})$ denotes the second order partial derivative matrix of $Q_{n,\lambda_n}(\boldsymbol{\theta})$.

As usual, we formally get

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\ddot{Q}_{n,\lambda_n}(\bar{\boldsymbol{\theta}})^{-1}\sqrt{n}\,\dot{Q}_{n,\lambda_n}(\boldsymbol{\theta}_0).$$

By Slusky's Lemma it suffices to show that, as $n \to \infty$,

$$\sqrt{n}\,\dot{Q}_{n,\lambda_n}(\boldsymbol{\theta}_0) \rightsquigarrow \mathcal{N}[\mathbf{0}, 4\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)] \tag{A.13}$$

and

$$\ddot{Q}_{n,\lambda_n}(\bar{\boldsymbol{\theta}}) \xrightarrow{\text{P}} 2\boldsymbol{\Delta}_1(\boldsymbol{\theta}_0). \tag{A.14}$$

We first prove (A.13). To this end, taking the partial derivative with respect to $\boldsymbol{\theta}$ in (A.12) yields

$$\begin{aligned}\sqrt{n}\,\dot{Q}_{n,\lambda_n}(\boldsymbol{\theta}_0) &= \sqrt{n}\,\dot{Q}_n(\boldsymbol{\theta}_0) + 2\sqrt{n}\,(\mathbf{0}_{d-1}^\top,\ \lambda_{n,1}(\mathbf{D}_1\boldsymbol{\beta}_{0,1})^\top,\ \ldots,\ \lambda_{n,L}(\mathbf{D}_L\boldsymbol{\beta}_{0,L})^\top)\\ &= -\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^\top\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i + 2\sqrt{n}\,(\mathbf{0}_{d-1}^\top,\ \lambda_{n,1}(\mathbf{D}_1\boldsymbol{\beta}_{0,1})^\top,\ \ldots,\ \lambda_{n,L}(\mathbf{D}_L\boldsymbol{\beta}_{0,L})^\top)^\top.\end{aligned}$$

The second term on the last line tends to $\mathbf{0}_{\dim(\boldsymbol{\theta})-1}$ as $\lambda_{n,\ell} = o(n^{-1/2})$ for all $\ell \in \{1,\ldots,L\}$, while the first term converges in distribution to the multivariate normal with mean zero and variance–covariance matrix $4\boldsymbol{\Delta}_2(\boldsymbol{\theta}_0)$ under Condition 4. This proves (A.13).

It remains to show (A.14). Note that

$$\ddot{Q}_{n,\lambda_n}(\bar{\boldsymbol{\theta}}) = \ddot{Q}_n(\bar{\boldsymbol{\theta}}) + 2\,\text{Diag}(\mathbf{0}_{(d-1)\times(d-1)}, \lambda_{n,1}\mathbf{D}_1, \ldots, \lambda_{n,L}\mathbf{D}_L).$$

Again, the second term on the right-hand side of the above equation tends to $\mathbf{0}_{\dim(\boldsymbol{\theta})\times\dim(\boldsymbol{\theta})}$ as each $\lambda_{n,\ell} = o(n^{-1/2})$.

We write

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{\epsilon}_i + \mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta})\}^\top\mathbf{W}_i^{-1}\{\boldsymbol{\epsilon}_i + \mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta})\}.$$

Taking partial derivatives with respect to $\boldsymbol{\theta}$ on both sides, we have

$$\ddot{Q}_n(\boldsymbol{\theta}) = \frac{2}{n}\sum_{i=1}^{n}\dot{\mathbf{m}}_i(\boldsymbol{\theta})^\top\mathbf{W}_i^{-1}\dot{\mathbf{m}}_i(\boldsymbol{\theta}) - 2\mathbf{B}_n(\boldsymbol{\theta}) - 2\mathbf{C}_n(\boldsymbol{\theta}),$$

where $\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{C}(\boldsymbol{\theta})$ are $\dim(\boldsymbol{\theta}) \times \dim(\boldsymbol{\theta})$ matrices whose (s,t) entry $\mathbf{B}(\boldsymbol{\theta})_{s,t}$ and $\mathbf{C}(\boldsymbol{\theta})_{s,t}$ are, respectively, given by

$$\mathbf{B}_n(\boldsymbol{\theta})_{s,t} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\mathbf{m}_i(\boldsymbol{\theta})^\top}{\partial\theta_s\partial\theta_t}\mathbf{W}_i^{-1}\boldsymbol{\epsilon}_i, \quad \mathbf{C}_n(\boldsymbol{\theta})_{s,t} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\mathbf{m}_i(\boldsymbol{\theta})^\top}{\partial\theta_s\partial\theta_t}\mathbf{W}_i^{-1}\{\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta})\}.$$

By the uniform convergence in (11)–(12) and the consistency of the $\hat{\boldsymbol{\theta}}_n$ in Theorem 1,

$$\mathbf{B}_n(\bar{\boldsymbol{\theta}})_{s,t} \xrightarrow{\text{P}} 0, \quad \mathbf{C}_n(\bar{\boldsymbol{\theta}})_{s,t} \xrightarrow{\text{P}} 0.$$

These yield the desired (A.14). This concludes the proof of Theorem 3. □

References

[1] H. Adrogue, N. Madias, Sodium and potassium in the pathogenesis of hypertension, New Engl. J. Med. 356 (2007) 1966–1978.
[2] Y. Bai, K. Fung, Y. Zhu, Penalized quadratic inference functions for single-index models with longitudinal data, J. Multivariate Anal. 100 (2009) 152–161.
[3] R.J. Carroll, J. Fan, I. Gijbels, M. Wand, Generalized partially linear single-index models, J. Amer. Statist. Assoc. 92 (1997) 477–489.
[4] K.L. Chung, A Course in Probability Theory, Academic Press, San Diego, CA, 2001.
[5] G. Claeskens, T. Krivobokova, J.D. Opsomer, Asymptotic properties of penalized spline estimators, Biometrika 96 (2009) 529–544.
[6] M. Durban, J. Harezlak, M. Wand, R.J. Carroll, Simple fitting of subject-specific curves for longitudinal data, Stat. Med. 24 (2005) 1153–1167.
[7] W.K. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, Ann. Statist. 21 (1993) 157–178.
[8] J. Huang, Asymptotics for polynomial spline regression under weak conditions, Statist. Probab. Lett. 65 (2003) 207–216.
[9] H. Ichimura, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, Econometrica 58 (1993) 71–120.
[10] C. Jiang, J. Wang, Functional single index models for longitudinal data, Ann. Statist. 39 (2011) 362–388.
[11] T. Klesov, Limit Theorems for Multi-Indexed Sums of Random Variables, Springer, New York, 2014.
[12] T. Krivobokova, G. Kauermann, A note on penalized spline smoothing with correlated errors, J. Amer. Statist. Assoc. 102 (2007) 1328–1337.
[13] K. Li, Sliced inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (1991) 316–342.

[14] H. Lin, R. Zhang, J. Shi, J. Liu, Y. Liu, A new local estimation method for single index models for longitudinal data, J. Nonparametr. Stat. 28 (2016) 644–658.
[15] M.L. Lindstrom, D.M. Bates, Nonlinear mixed effects models for repeated measures data, Biometrics 46 (1990) 673–687.
[16] A. Mente, M.J. O'Donnell, S. Rangarajan, M.J. McQueen, P. Poirier, A. Wielgosz, H. Morrison, W. Li, X. Wang, C. Di, P. Mony, A. Devanath, et al., Association of urinary sodium and potassium excretion with blood pressure, New Engl. J. Med. 371 (2014) 601–611.
[17] Z. Pang, L. Xue, Estimation for the single-index models with random effects, Comput. Statist. Data Anal. 56 (2012) 1837–1853.
[18] D. Ruppert, Selecting the number of knots for penalized splines, J. Comput. Graph. Statist. 11 (2002) 735–757.
[19] D. Ruppert, R.J. Carroll, Spatially adaptive penalties for spline fitting, Austr. N. Z. J. Statist. 42 (2000) 205–223.
[20] D. Ruppert, M. Wand, R.J. Carroll, Semiparametric Regression, Cambridge University Press, Cambridge, 2002.
[21] T. Stoker, Consistent estimation of scaled coefficients, Econometrica 54 (1986) 1461–1481.
[22] L. Thanh, On the strong law of large numbers for d-dimensional arrays of random variables, Electron. Commun. Probab. 12 (2007) 434–441.
[23] P. Tian, L. Yang, L. Xue, Asymptotic properties of estimators in partially linear single-index model for longitudinal data, Acta Math. Sci. 30B (2010) 677–687.
[24] W. Tu, G. Eckert, B. Decker, J. Pratt, Varying influences of aldosterone on the plasma potassium concentration in blacks and whites, Amer. J. Hypertension 30 (2017) 490–494.
[25] W. Tu, G. Eckert, T. Hannon, H. Liu, L. Pratt, M. Wagner, L. DiMeglio, J. Jumg, J. Pratt, Racial differences in sensitivity of blood pressure to aldosterone novelty and significance, Hypertension 63 (2014) 1212–1218.
[26] Y. Wang, Smoothing spline models with correlated random errors, J. Amer. Statist. Assoc. 93 (1998) 341–348.
[27] Y. Wang, W. Guo, M. Brown, Spline smoothing for bivariate data with applications to association between hormones, Statist. Sinica 10 (2000) 377–397.
[28] J. Wang, L. Xue, L. Zhu, Y. Chong, Estimation for a partial-linear single-index model, Ann. Statist. 38 (2010) 246–272.
[29] L. Wang, L. Yang, Spline estimation of single-index models, Statist. Sinica 19 (2009) 765–783.
[30] C. Wu, Asymptotic theory of nonlinear least squares estimation, Ann. Statist. 9 (1981) 501–513.
[31] J. Wu, W. Tu, Development of a pediatric body mass index using longitudinal single-index models, Stat. Methods Med. Res. 25 (2016) 872–884.
[32] J. Wu, W. Tu, A multivariate single-index model for longitudinal data, Stat. Modell. 16 (2016) 392–408.
[33] Y. Xia, W.K. Härdle, Semi-parametric estimation of partially linear single-index models, J. Multivariate Anal. 97 (2006) 1162–1184.
[34] Y. Xia, W.K. Härdle, O. Linton, Optimal smoothing for a computationally and statistically efficient single index estimator, in: I. Van Keilegom, P.W. Wilson (Eds.), Exploring Research Frontiers in Contemporary Statistics and Econometrics, Physica-Verlag HD, Springer, Berlin, Heidelberg, 2012, pp. 229–261.
[35] Y. Xia, H. Tong, W. Li, L. Zhu, An adaptive estimation of dimension reduction space, J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2002) 363–410.
[36] L. Xue, L. Zhu, Empirical likelihood for single-index models, J. Multivariate Anal. 97 (2006) 1295–1312.
[37] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, J. Amer. Statist. Assoc. 97 (2002) 1042–1054.
[38] Y. Yu, D. Ruppert, Root-n consistency of penalized spline estimator for partially linear single-index models under general Euclidean space, in: Technical Report #1385, Cornell University, Ithaca, NY, 2003.
[39] Y. Yu, D. Ruppert, Root-n consistency of penalized spline estimator for partially linear single-index models under general Euclidean space, Statist. Sinica 14 (2004) 449–455.
[40] R. Zhang, Tests for nonparametric parts on partially linear single index models, Sci. China A 50 (2007) 439–449.
[41] L. Zhu, L. Xue, Empirical likelihood confidence regions in a partially linear single-index model, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2006) 549–570.