Electronic Journal of Statistics
Vol. 18 (2024) 1603–1642
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2239

Penalized estimation of panel count data using generalized estimating equation

Minggen Lu

School of Public Health, University of Nevada, Reno, NV, U.S.A.,
e-mail: minggenl@unr.edu

Abstract: This manuscript discusses the regression analysis of a semiparametric proportional mean model for panel count data. A spline-based generalized estimating estimation (GEE) approach is applied to account for the correlation among cumulative counts. To avoid the potential issue of overfitting, a penalization technique is applied to regularize the spline estimation. An easy-to-implement and computationally efficient two-stage iterative algorithm is developed to accomplish the penalized estimation. The proposed methodology does not specify the stochastic model of the underlying counting process and hence provides great flexibility for model fitting. Theoretically, the uniform convergence and the optimal rate of convergence for the functional estimator are established, and the asymptotic normality for regression parameter estimators is shown to be valid even if the working covariance matrix is misspecified. The semiparametric efficiency for regression parameter estimators can be achieved if the working matrix is correctly specified. Further, to address the issue of the underestimation of the variance-covariance matrix of regression parameter estimators for small sample sizes, which is brought up by GEE methodology, we propose a novel approach based on the modified sandwich estimator to compensate for the deficiency in variance-covariance estimation. Numerically, an extensive Monte Carlo study was conducted to evaluate the finite-sample performance of penalized spline estimators and the impact of the selection of the working matrix on the estimation, along with the robustness of the methodology to the underlying counting process. The proposed penalized approach was further applied to analyze data from a non-melanoma skin cancer chemoprevention study.

MSC2020 subject classifications: Primary 62G05, 62G20; secondary 62G08.
Keywords and phrases: B-spline, generalized estimating equation, panel count data, penalized estimation.

Received October 2022.

Contents

1. Introduction

Panel count data are frequently observed in long-term cohort or experimental studies. The distinctive feature of this type of data is that only the number of recurrent events of interest is accessible at each examination time. The exact time to event cannot be measured directly, but rather is known to be relative to two adjacent examination times. For example, in a randomized placebo-controlled skin cancer clinical trial (Bailey et al. [1]), study subjects with a history of non-melanoma skin cancer were randomly assigned to either a chemoprevention therapy (DFMO) group or a placebo group. The primary aim of this study was to evaluate whether DFMO can effectively reduce the recurrence of two types of non-melanoma skin cancers, including basal carcinoma (BCC) and squamous cell carcinoma (SCC). The participants were scheduled to have an assessment of the efficacy of the treatment every 6 months. At each examination time, newly developed tumors were recorded and removed. Throughout the study, the number of follow-up visits and examination times varied greatly from patient to patient due to the flexible scheduled times and different entry times. In this study, only the number of recurrent tumors between two clinic visits was observed, but the exact time of the occurrence of the tumor is unavailable; see, Chiou et al. [4], Li et al. [11], and Sun and Zhao [31] for more information.

Statistical methodology for panel count data has been extensively studied in the literature, for example, Sun and Kalbfleisch [28], Sun and Wei [29], and Zhang [38, 39], among many others; for a comprehensive review, see Sun and Zhao [31] and the references therein. To alleviate the computational burden, Lu et al. [13, 14] applied monotone spline techniques to study nonparametric and semiparametric pseudo-likelihood and maximal likelihood estimations under a proportional mean model proposed by Wellner and Zhang [36, 37]. A potentially prohibitive assumption made in all the above works is that the stochastic model for the underlying counting process is a (mixed) non-homogeneous Poisson process. This restrictive assumption may lead to biased estimation and incorrect inference. Moreover, the Poisson process-based methodology fails to account for the overdispersion problem that commonly arises in regression analysis for count data. To address these issues and provide a more general estimation framework, Hua and Zhang [7] developed a spline-based semiparametric estimation approach that does not specify the stochastic model for the underlying counting process under the proportional mean model using GEE methodology. Several working covariance matrices were selected to accommodate the stochastic models of the underlying counting process, and the impact of the selection of a working covariance matrix on the estimation was also investigated. Further, A hybrid algorithm including the Newton method and the weighed isotonic regression approach was employed for estimation.

In this article, we endeavor to develop an easy-to-implement and computationally efficient approach that can be employed to conduct regression analysis for panel count data under the proportional mean model within the framework of GEE. In particular, we utilized a monotone B-spline to approximate the unknown baseline mean function to facilitate the model fitting. To address the potential issue of overfitting that frequently occurs in spline estimation, the penalization technique was used to regularize the estimation of the unknown baseline cumulative mean function. Within the carefully constructed framework of the proposed penalized model, a two-stage hybrid iterative algorithm was developed to fulfill the model fitting. A desirable feature of this algorithm is that it simultaneously estimates regression parameters and spline coefficients and incorporates the update of the smoothing parameter in the model fitting process, which remarkably relieves the computational burden. Further, by making use of spline approximation, a simple and consistent variance-covariance estimation approach was proposed to provide valid inference for regression parameters. The novelty of this variance-covariance estimation approach is that it atones for the underestimation of the variance-covariance matrix that originates from the standard sandwich estimator for small sample sizes, and hence the proposed approach can provide more accurate Wald-type inference for regression parameters. As demonstrated through extensive simulation studies, the proposed penalized methodology is computationally efficient and robust to the misspecification of the working variance-covariance structure under a variety of settings. Theoretically, by integrating the spline approximation and penalization technique innovatively and employing modern empirical process theory, we established the large-sample properties for penalized spline GEE estimators. In particular, the uniform consistency and the optimal rate of convergence for the functional estimator, which is the best attainable rate in the context of semiparametric regression, are established if the smoothing parameter and the dimension of spline space are specified in an appropriate order. Moreover, the regression parameter estimators are shown to be asymptotically normal even if the working covariance matrix is misspecified and the semiparametric efficiency of regression parameter estimators (i.e., they attain the semiparametric information bound) can be achieved if the working covariance matrix is correctly specified. Furthermore, the estimated variance-covariance matrix of regression parameter estimators is shown to be consistent and exhibits better finite-sample performance compared to the one based on the standard sandwich method.

The remainder of the paper is organized as follows. The methodological details of penalized spline GEE estimators are provided, and the two-stage hybrid algorithm and the selection of the smoothing parameter and spline knots as well as the working covariance matrix are discussed in Section 2. The large-sample properties of penalized spline GEE estimators are presented in Section 3. A novel variance-covariance estimation procedure for regression parameters is developed in Section 4. The finite-sample performance of the proposed methodology is evaluated through extensive simulation studies in Section 5. The methodology is further illustrated by analyzing data from a randomized non-melanoma skin cancer study in Section 6. A summary of the work and future research are discussed in Section 7. Finally, the proofs of the asymptotic results are available in Section 8.

2. Model setup and numerical algorithm

2.1. Model

Let $\{\mathbb{N}(t) : t \geq 0\}$ denote a underlying counting process with $\mathbb{N}(0) = 0$. Consider a proportional mean model in which the conditional mean of $\mathbb{N}(\cdot)$ given a covariate vector Z takes the form:

$$E\{\mathbb{N}(t)|Z\} = \Lambda(t)\exp(\beta^\intercal Z), \tag{2.1}$$

where $\Lambda(\cdot)$ is an unknown baseline cumulative mean of $\mathbb{N}(\cdot)$ and β is a d-dimensional vector of regression parameters corresponding to the possibly time-dependent covariate vector Z. For panel count data $X = (\mathbb{N}, K, T, Z)$ of a counting process $\mathbb{N}(\cdot)$, K is the total number of random examination times and $T = (T_{K,1}, \ldots, T_{K,K})^\intercal$ is a vector of examination times with $0 < T_{K,1} < \ldots < T_{K,K}$. Let $\mathbb{N}(T) = \{\mathbb{N}(T_{K,1}), \ldots, \mathbb{N}(T_{K,K})\}^\intercal$ denote a vector of the cumulative numbers of the recurrent events corresponding to examination times T. Obviously, $0 \leq \mathbb{N}(T_{K,1}) \leq \ldots \leq \mathbb{N}(T_{K,K})$. It is assumed that the examination times of panel count data are non-informative, i.e., (K, T) are conditionally independent of the underlying counting process $\mathbb{N}(\cdot)$, given the covariate vector Z. In what follows, the observed data comprise of a random sample $X_1, \ldots, X_n$, where $X_i = \{\mathbb{N}^{(i)}, K_i, T_i, Z_i\}$ for $T_i = \{T^{(i)}_{K_i,1}, \ldots, T^{(i)}_{K_i,K_i}\}^\intercal$, $Z_i = (Z_{i1}, \ldots, Z_{iK_i})^\intercal$, and $\mathbb{N}^{(i)} = (\mathbb{N}_{i1}, \ldots, \mathbb{N}_{iK_i})^\intercal$ with $\mathbb{N}_{ij} = \mathbb{N}^{(i)}(T^{(i)}_{K_i,j})$. Also, denote by $\mu_i = (\mu_{K_i,1}, \ldots, \mu_{K_i,K_i})^\intercal$ the conditional mean of $\mathbb{N}^{(i)}$ given (Z_i, T_i) with $\mu_{K_i,j} = \Lambda(T^{(i)}_{K_i,j})\exp(Z_i^\intercal \beta)$, for $j \in \{1, \ldots, K_i\}$.

For ith underlying counting process $\mathbb{N}^{(i)}$, define $\Sigma_i = \text{var}\ (\mathbb{N}^{(i)}|K_i, Z_i, T_i)$ and $V_i = V_i(K_i, Z_i, T_i)$ as the true covariance matrix of $\mathbb{N}^{(i)}$ and as a working covariance matrix that may depend on a finite-dimensional vector of nuisance parameters, respectively. In the sequel, to simplify the presentation and facilitate the development of asymptotic results, the vector of the unknown parameters is defined as $\tau = (\beta, \varphi)$, where $\varphi(\cdot) = \log \Lambda(\cdot)$. It is well-known that the penalized methodology is competent in control of the balance between the fidelity and the smoothness of the fitted curve in semiparametric estimation. Thus, following the works of Ma and Kosorok [17], Lu and Li [15], and Lu et al. [16], we propose to estimate the unknown parameter τ via identifying the minimizer of the following penalized weighted least squares objective function, namely,

$$W_{n,\lambda}(\tau) = \frac{1}{2n} \sum_{i=1}^{n} \left\{ \mathbb{N}^{(i)} - \mu_i \right\}^{\intercal} V_i^{-1} \left\{ \mathbb{N}^{(i)} - \mu_i \right\} + \frac{1}{2} \lambda^2 J^2(\varphi), \tag{2.2}$$

where $J^2(\varphi) = \int \{\varphi^{(m)}(t)\}^2 dt$ is the penalized term for a fixed integer $m \geq 1$ and $\lambda > 0$ is the smoothing parameter used to administer the smoothness of the estimated function.

In general, it is considerably challenging to estimate the unknown function $\varphi(\cdot)$ directly from the penalized objective function (2.2). To tackle the difficulty, following the proposal of Lu and Li [15] and Lu et al. [14], we approximated $\varphi(\cdot)$ via a monotone B-spline. In particular,

$$\varphi(\cdot) \approx \sum_{j=1}^{q_n} \gamma_j b_j(\cdot) \equiv \gamma^{\intercal} b(\cdot), \tag{2.3}$$

where $b(\cdot) = \{b_1(\cdot), \ldots, b_{q_n}(\cdot)\}^{\intercal}$ is a vector of B-spline basis functions and $\gamma = (\gamma_1, \ldots, \gamma_{q_n})^{\intercal}$ is a vector of spline coefficients under the constraints $\gamma_1 \leq \ldots \leq \gamma_{q_n}$. According to Theorem 5.9 of Schumaker [25], the nondecreasing constraints on spline coefficients γ guarantee that the resulting spline is nondecreasing. Let $\mathcal{T} = [d_1, d_2]$ with $0 \leq d_1 < d_2 < \infty$ be the support of T. Denote by $\mathcal{M}_n$ the space of monotone splines defined on $\mathcal{T}$ with degree $m+1$ and knots $\mathcal{K}_n$. The penalized spline estimator of τ is defined as one that minimizes $W_{n,\lambda}(\tau)$ on $\Phi \times \mathscr{N}$, where the regression parameter space Φ is a compact subset of $\mathbb{R}^d$ and the nonparametric space is defined as $\mathscr{N} = \{\varphi : \varphi \in \mathcal{M}_n, J(\varphi) < \infty\}$.

Under the spline approximation (2.3), we obtain the proposed spline model, namely,

$$E\{\mathbb{N}(t)|Z\} = \exp\{\beta^{\intercal} Z + \gamma^{\intercal} b(t)\}. \tag{2.4}$$

Let $\theta = (\beta, \gamma)$ denote the unknown parameters to be estimated under the spline model (2.4) and define the conditional mean of $\mathbb{N}^{(i)}$ as $\overline{\mu}_i = (\overline{\mu}_{K_i,1}, \ldots, \overline{\mu}_{K_i,K_i})^{\intercal}$ with $\overline{\mu}_{K_i,j} = \exp\{\beta^{\intercal} Z_i + \gamma^{\intercal} b(T^{(i)}_{K_i,j})\}$, for $j \in \{1, \ldots, K_i\}$. Under the above specifications, the penalized spline objective function for θ can be expressed as

$$\overline{W}_{n,\lambda}(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left\{ \mathbb{N}^{(i)} - \overline{\mu}_i \right\}^{\intercal} V_i^{-1} \left\{ \mathbb{N}^{(i)} - \overline{\mu}_i \right\} + \frac{1}{2} \lambda^2 \gamma^{\intercal} \mathcal{D} \gamma, \tag{2.5}$$

where $\mathcal{D}$ is a band matrix with (j,k)th element $\int_{\mathcal{T}} b_j^{(m)}(t) b_k^{(m)}(t) dt$. The penalized spline estimator $\widehat{\theta} = (\widehat{\beta}, \widehat{\gamma})$ is defined as the minimizer of the penalized spline objective function (2.5) under nondecreasing constraints on γ. Accordingly, $\widehat{\varphi}(\cdot) = \widehat{\gamma}^\intercal b(\cdot)$ is defined as the penalized spline estimator of $\varphi(\cdot)$. Notice that, for a fixed λ, the regression parameters β and spline coefficients γ can be estimated jointly by minimizing the penalized spline weighted least squares objective function (2.5). Clearly, to calculate the penalized spline estimator $\widehat{\tau}$ is equivalent to deriving $\widehat{\theta}$ under monotone constraints. To further facilitate the model fitting, we employ the P-spline approach (e.g., Eilers and Marx [6]) to approximate the penalty matrix $\mathcal{D}$ by the difference matrix $D^\intercal D$, where D is the matrix induced by the difference operator of order m. As discussed in Wood ([34], p. 206), the difference matrix can be easily implemented with any order of B-spline basis, and hence the P-spline method provides a great deal of flexibility for penalized estimation. In Monte Carlo study and real application, we use monotone cubic splines to approximate $\varphi(\cdot)$, i.e., m is set to be 2.

Under the above specifications, the closed forms for the gradient and expected Hessian matrix of $\overline{W}_{n,\lambda}$ are given by

$$\nabla \overline{W}_{n,\lambda}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} E_i^\intercal \Delta_i V_i^{-1}\left\{\mathbb{N}^{(i)} - \overline{\mu}_i\right\} + \lambda^2 \begin{pmatrix} 0 \\ D^\intercal D\gamma \end{pmatrix}$$

and

$$\mathcal{I}_{n,\lambda}(\theta) = \frac{1}{n}\sum_{i=1}^{n} E_i^\intercal \Delta_i V_i^{-1}\Delta_i E_i + \lambda^2 \begin{pmatrix} 0 & 0 \\ 0 & D^\intercal D \end{pmatrix},$$

respectively, where $E_i = (1_{K_i} Z_i^\intercal, B_i)$ for 1_{K_i} being the K_i-dimensional vector with ones, $B_i = \{b(T^{(i)}_{K_i,1}), \ldots, b(T^{(i)}_{K_i,K_i})\}^\intercal$, and $\Delta_i = \operatorname{diag}\left\{\overline{\mu}_{K_i,1}, \ldots, \overline{\mu}_{K_i,K_i}\right\}$. Obviously, $\mathcal{I}_{n,\lambda}(\theta)$ is positive definite. As will be discussed in Section 2.3, $\mathcal{I}_{n,\lambda}(\theta)$ plays an important role in updating the smoothing parameter λ in the proposed two-stage algorithm. Notice that the penalized spline estimator $\widehat{\theta}$ solves the estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \overline{\mu}_i}{\partial \theta}\right)^\intercal V_i^{-1}\{\mathbb{N}^{(i)} - \overline{\mu}_i\} - \lambda^2 \begin{pmatrix} 0 \\ D^\intercal D\gamma \end{pmatrix} = 0.$$

Thus, the penalized spline estimators can be regarded as penalized semiparametric versions of the standard GEE estimators. By carefully taking into consideration the spline approximation and properly choosing the order of λ, we can justify that the large-sample properties of parametric GEE estimators such as the consistency and asymptotic normality can be extended in semiparametric context. The asymptotic properties of the penalized spline estimators will be discussed in Section 3.

2.2. Selection of working covariance matrix

As is widely discussed in the literature, the selection of the working covariance matrix has a significant influence on the efficiency of GEE-type estimator. Hua and Zhang [7] discussed three different working covariance matrices to accommodate the underlying counting process for panel count data. In particular, define $V_i^{(1)} = (\sigma_{kl})_{K_i \times K_i}$ with $\sigma_{kl} = \overline{\mu}_{K_i,l}$, for $k = l$, and 0 otherwise, and $V_i^{(2)} = (\sigma_{kl})$ with $\sigma_{kl} = \overline{\mu}_{K_i,\min(k,l)}$. The selection of $V_i^{(1)}$ implies that the cumulative accounts are independent, and hence cov $\{\mathbb{N}(t_1), \mathbb{N}(t_2)\} = 0$, for $t_1 \neq t_2$, while the choice of $V_2^{(i)}$ suggests that $\text{cov}\{\mathbb{N}(t_1), \mathbb{N}(t_2)\} = E\{\mathbb{N}(t_1)\}$, for $t_1 \leq t_2$. The spline GEE estimator based on $V_i^{(1)}$ or $V_i^{(2)}$ is equivalent to the pseudo-likelihood estimator or full likelihood estimator under the non-homogeneous Poisson model proposed in Lu et al. [14], respectively; see Hua and Zhang [7] for more details. The full likelihood estimator based on $V_i^{(2)}$ is more computationally efficient compared to the pseudo-likelihood estimator based on $V_i^{(1)}$, which ignores the positive correlation among cumulative counts. It is observed that the spline estimation based on working covariance matrix $V_i^{(1)}$ or $V_i^{(2)}$ fails to account for overdispersion, which arises naturally from the count data and may result in loss of efficiency and inflation of type I error. To address this issue, Hua and Zhang [7] proposed a working covariance matrix $V_i^{(3)} = V_i^{(2)} + \sigma^2 \overline{\mu}_i^{\otimes 2}$, where the parameter σ^2 is used to account for overdispersion. The spline estimator based on $V_i^{(3)}$ is equivalent to one under the gamma frailty non-homogeneous Poisson model, i.e., given the frailty variable $\gamma \sim \Gamma(1/\sigma^2, 1/\sigma^2)$, the underlying counting process $\mathbb{N}(t)$ is a non-homogeneous Poisson process with mean $\gamma \Lambda(t) e^{\beta^\intercal Z}$. Under this frailty model, the marginal mean of $\mathbb{N}(\cdot)$ still satisfies the proportional mean model (2.1), but the distribution of the cumulative count is marginally negative binomial rather than Poisson. Obviously, the estimator based on $V_i^{(3)}$ reduces to the one based on $V_i^{(2)}$ if $\sigma^2 = 0$, i.e., the underlying counting process is indeed a non-homogeneous Poisson process. The numerical properties of penalized spline estimators based on working covariance matrices will be discussed in Section 5.

2.3. A two-stage hybrid algorithm

To accomplish the model fitting, we develop a two-stage iterative procedure to identify the penalized spline estimate of θ. In particular, the smoothing parameter λ is updated during the outer iteration, while the inner iteration attempts to identify the minimizer of the constrained penalized spline objective function for a *fixed* λ. Under the gamma frailty non-homogeneous Poisson model, i.e., $V_i = V_i^{(3)}$, var $\{\mathbb{N}(t)\} = \overline{\mu}_t + \sigma^2 \overline{\mu}_t^2$ with $\overline{\mu}_t = E\{\mathbb{N}(t)\}$. In order to identify the penalized spline estimate of θ, we need to estimate the nuisance parameter σ^2 in the working covariance. Following the proposal of Hua and Zhang [7], we apply an adjusted moment estimation method which accounts for penalization to estimate σ^2. In particular,

$$\widehat{\sigma}^2 = \frac{n}{n - \rho} \frac{\sum_{i=1}^n \sum_{j=1}^{K_i} \left\{ (\mathbb{N}_{ij} - \widehat{\mu}_{ij})^2 - \widehat{\mu}_{ij} \right\}}{\sum_{i=1}^n \sum_{j=1}^{K_i} \widehat{\mu}_{ij}^2}, \tag{2.6}$$

where $\rho = \mathcal{I}_{n,\lambda}^{-1}\mathcal{I}_n$ with $\mathcal{I}_n = \mathcal{I}_{n,\lambda}|_{\lambda=0}$ is the effective degree of freedom for penalized estimation and $\widehat{\mu}_t$ is an estimate of $\overline{\mu}_t$. In case of negative, $\widehat{\sigma}^2$ is set to be 0. The adjusted term $n/(n-\rho)$ allows for bias correction when the effective degree of freedom is relatively large compared to the sample size n. Our numerical experiments reveal this ad hoc approach provides an accurate estimation of σ^2 for small sample sizes. The adjustment diminishes as the sample size is increased.

During each inner iteration, the current estimate of θ, say $\theta_\lambda^{(l)}$, is updated to $\overline{\theta}_\lambda^{(l)}$ via the Fisher's scoring method, i.e.,

$$\overline{\theta}_\lambda^{(l)} = \theta_\lambda^{(l)} - \mathcal{I}_{n,\lambda}^{-1}(\theta_\lambda^{(l)})\nabla\overline{W}_{n,\lambda}(\theta_\lambda^{(l)}). \tag{2.7}$$

In general, the update of $\overline{\gamma}_\lambda^{(l)}$ available in $\overline{\theta}_\lambda^{(l)}$ is not necessarily non-decreasing. To enforce the constraints imposed on spline coefficients, $\overline{\gamma}_\lambda^{(l)}$ is then projected into a constrained space called $\Gamma_n = \{\gamma : \gamma_1 \leq \ldots \leq \gamma_{q_n}\}$. This can be achieved via solving the quadratic optimization problem, i.e.,

$$\widetilde{\gamma}_\lambda^{(l)} = \underset{\gamma\in\Gamma_n}{\arg\min}(\gamma - \overline{\gamma}_\lambda^{(l)})^\intercal \mathcal{W}_n(\gamma - \overline{\gamma}_\lambda^{(l)}), \tag{2.8}$$

for a positive-definite weighted matrix $\mathcal{W}_n$. As discussed in Cheng and Zhang [2] and Hua and Zhang [7], if $\mathcal{W}_n$ is chosen to be a diagonal matrix with elements as the diagonal elements of $\mathcal{I}_{n,\lambda}^{-1}(\overline{\theta}_\lambda^{(l)})$ with respect to γ, the quadratic programming problem (2.8) reduces to the standard isotonic regression (i.e., Robertson et al. [23]), which can be solved via the procedure pava in R package Iso. Substituting $\overline{\gamma}_\lambda^{(l)}$ with $\widetilde{\gamma}_\lambda^{(l)}$ in $\overline{\theta}_\lambda^{(l)}$ yields the constrained update of $\widetilde{\theta}_\lambda^{(l)}$, and $\widehat{\sigma}^2$ is then updated via equation (2.6). These two optimization procedures (i.e., Fisher's scoring method and isotonic regression) iterate in turn until convergence, and the inner iteration is complete. Notice that after each inner iteration, we obtain the penalized spline estimate $\theta_\lambda^{(l+1)}$ of θ for a fixed λ. Then the outer iteration restarts and the smoothing parameter λ is updated via the generalized Fellner-Schall method; see Wood and Fasiolo [35] for more information. In particular, the algorithm makes the update of λ as

$$\overline{\lambda}^2 = \frac{\mathrm{tr}\{A_\lambda^- A\} - \mathrm{tr}\left\{\mathcal{I}_{n,\lambda}^{-1}(\theta_\lambda^{(l+1)})A\right\}}{\theta_\lambda^{(l+1)\intercal} A \theta_\lambda^{(l+1)}}\lambda^2, \tag{2.9}$$

where A is a diagonal matrix with diagonal elements $0_{d\times d}$ and $D^\intercal D$ and A_λ^- is the generalized inverse of $A_\lambda = \lambda A$. Since $\mathcal{I}_{n,\lambda}(\theta_\lambda^{(l+1)})$ is positive definite, the update of λ in (2.9) is guaranteed to be positive; see Theorem 4 of Wood and Fasiolo [35] for further details. Once the updated $\overline{\lambda}$ is available, the inner process restarts to identify the constrained minimizer $\theta_{\overline{\lambda}}^{(l+1)}$ for the given $\overline{\lambda}$. The two-stage iterative algorithm converges if the difference between $\theta_{\overline{\lambda}}^{(l+1)}$ and $\theta_\lambda^{(l+1)}$ is less than a pre-specified value. The implementation of the proposed algorithm is outlined as follows:

Step 1 (inner iteration). For a given $\lambda > 0$, (a) update the current $\theta_\lambda^{(l)}$ to $\overline{\theta}_\lambda^{(l)}$ through the Fisher's scoring approach (2.7); (b) the monotonic update $\widetilde{\gamma}_\lambda^{(l)}$ of $\overline{\gamma}_\lambda^{(l)}$ is acquired via the isotonic regression (2.8) and set $\widetilde{\theta}_\lambda^{(l)} = (\overline{\beta}_\lambda^{(l)\intercal}, \widetilde{\gamma}_\lambda^{(l)\intercal})^\intercal$; and (c) update σ^2 via equation (2.6) at $\theta = \widetilde{\theta}_\lambda^{(l)}$. Go to **Step 2** if $\|\widetilde{\theta}_\lambda^{(l)} - \theta_\lambda^{(l)}\| < 10^{-6}$ and set $\theta_\lambda^{(l+1)} = \widetilde{\theta}_\lambda^{(l)}$. Otherwise go back to (a).

Step 2 (outer iteration). Update λ to $\overline{\lambda}$ via the generalized Fellner-Schall method (2.9) and go to **Step 1** to identify $\theta_{\overline{\lambda}}^{(l+1)}$.

The two-stage algorithm converges if $\|\theta_{\overline{\lambda}}^{(l+1)} - \theta_\lambda^{(l+1)}\| < 10^{-6}$.

To accelerate the convergence of the proposed algorithm, the spline full likelihood estimator $\theta^{(0)}$ (i.e., $V_i = V_i^{(2)}$) is used as the initial value of θ and the initial value of σ^2 can be obtained via the equation (2.6) at $\theta = \theta^{(0)}$ accordingly. As shown in Lu et al. [14], the spline full likelihood estimator $\theta^{(0)}$ is $n^{1/2}$-consistent, and hence the initial value of σ^2 is a consistent estimator of σ^2. Further, as will be shown in Section 3, the penalized spline estimator $\widehat{\theta}$ is $n^{1/2}$-consistent. Thus, the moment estimator $\widehat{\sigma}^2$ via equation (2.6) at $\theta = \widehat{\theta}$ is also consistent. The proposed two-stage iterative algorithm performed very well in our simulation settings. In particular, the issue of divergence is very rare, and the algorithm usually converges in a few steps by using the full spline estimator as the initial value. Further, the optimization procedure does not depend on the good starting value of λ.

2.4. Selection of spline knots

As is extensively discussed in the literature, the selection of spline knots is not as crucial as that of smoothing parameters; see, for example, Ruppert et al. [24]. Our numerical experiments reinforce this assertion and reveal that the proposed methodology is robust to the selection of knots, i.e., the penalized spline estimates are almost identical in terms of the bias and standard deviation for different numbers of knots. Therefore, for practical computation, it is reasonable to choose the number of inner knots as $\lceil n^{1/3} \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. The location of knots is then selected as equally spaced percentiles of examination times. This empirical rule was applied in Monte Carol studies and the real application.

3. Asymptotic results

3.1. Assumptions

In this section we discuss the asymptotic properties of the penalized spline GEE estimator $\widehat{\tau}$. Denote by $\tau_0 = (\beta_0, \varphi_0)$ the true value of τ. Let $\mathscr{B}$ denote the collection of Borel sets in $\mathbb{R}$. For any $B \in \mathscr{B} \cap \mathcal{T}$, define the measure μ as follows:

$$\mu(B) = \int_{\mathbb{R}^d} \sum_{k=1}^{K} \Pr(K = k | Z = z) \sum_{j=1}^{k} \Pr(T_{k,j} \in B | K = k, Z = z) dF(z),$$

where $F(\cdot)$ is the distribution function of Z. Let $\mathscr{F} = \{\varphi : \varphi(\cdot)$ is montoneincreasing on $\mathcal{T}\}$. Notice that the nonparametric space $\mathscr{N} \subset \mathscr{F}$. Based on the measure μ, define the L_2-metric $\|\cdot\|_{L_2(\mu)}$ on $\Phi \times \mathscr{F}$ as

$$\begin{aligned}\|\tau_2 - \tau_1\|_{L_2(\mu)}^2 &= \|\beta_2 - \beta_1\|^2 + \|\varphi_2 - \varphi_1\|_{L_2(\mu)}^2 \\ &= \|\beta_2 - \beta_1\|^2 + \int_{\mathcal{T}} \{\varphi_2(t) - \varphi_1(t)\}^2 \, d\mu(t),\end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm. The following regularity conditions are sufficient to establish asymptotic results of $\widehat{\tau}$.

C1. The true parameters β_0 and φ_0 are in the interiors of Φ and $\mathscr{F}$, respectively. Further, φ_0 is strictly increasing and its mth derivative satisfies the Lipschitz condition on $\mathcal{T}$.

C2. (a) The examination time $T_{K_i,j}^{(i)}$ is uniformly bounded on $\mathcal{T}$, for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, K_i\}$; (b) the joint distribution of any pair $T_{K_i,j_1}^{(i)}$ and $T_{K_i,j_2}^{(i)}$ $(j_1 \neq j_2)$ is uniformly bounded, for $j_1, j_2 \in \{1, \ldots, K_i\}$.

C3. The eigenvalues of the true covariance matrix Σ_i and the working covariance matrix V_i are uniformly bounded away from 0 and infinity, for $i \in \{1, \ldots, n\}$.

C4. The covariate Z_i is uniformly bounded, for $i \in \{1, \ldots, n\}$.

C5. The total number of the observation times K_i is uniformly bounded, for $i \in \{1, \ldots, n\}$.

C6. Let $e_{ij} = \mathbb{N}_{ij} - \exp\left\{Z_i^\intercal \beta_0 + \varphi_0(T_{K_i,j}^{(i)})\right\}$, for $j \in \{1, \ldots, K_i\}$. The error term $e_i = (e_{i1}, \ldots, e_{iK_i})^\intercal$ is assumed to satisfy uniformly sub-Gaussian condition, for $i \in \{1, \ldots, n\}$, i.e., there exist some fixed positive constants M_0 and σ_0 such that

$$\max_{i=1,\ldots,n} M_0^2 E\left\{\exp\left(\|e_i\|^2/M_0^2\right) - 1\right\} \leq \sigma_0^2,$$

almost surely, for all n.

C7. The smoothing parameter λ is of the order

$$\lambda = o_p(n^{-1/4}) \quad \text{and} \quad \lambda^{-1} = O_p(n^{m/(1+2m)}).$$

C8. The maximum spacing of knots is of the order $O(n^{-m/(1+2m)})$ and the ratio of maximum and minimum spacings is uniformly bounded.

C9. The efficient information $\mathscr{I}_0$ defined in (3.1) and the matrix $\mathscr{I}_1$ defined in (3.2) are positive definite.

Remark 1. Condition C1 assumes that $\varphi_0(\cdot)$ is smooth enough such that it can be well approximated by a B-spline. Conditions C2 and C3 are regularity conditions used in the literature for longitudinal/clustered data; see, for example, Huang et al. [9] for partially linear models and Cheng et al. [3] for generalized partially linear additive models. The uniformly bounded assumptions for the covariate Z_i (C4) and the total number of examinations K_i (C5) as well as the sub-Gaussian condition (C6) are used in entropy calculation to derive the rate of convergence of $\widehat{\varphi}(\cdot)$ and the asymptotic normality of $\widehat{\beta}$. Condition C7 is a standard assumption in penalized estimation; see, for example, Mammen and van de Geer [18] and Murphy and van der Vaart [21]. Condition C8 specifies the appropriate order of the dimension of the monotone spline space $\mathcal{M}_n$ to derive the rate of convergence of $\widehat{\varphi}(\cdot)$. Finally, condition C9 is required to establish the asymptotic normality of $\widehat{\beta}$.

3.2. Semiparametric efficient score and efficiency bound

In this section, we discuss the semiparametric efficient score and efficient information matrix when the covariance structure is *correctly specified*, i.e., $V = \Sigma_0$, where Σ_0 represents Σ evaluated at $\tau = \tau_0$. It is well known that the information bound induced by the efficient information matrix plays an important role in establishing the asymptotic properties of regression parameter estimators. In other words, it serves as a benchmark to evaluate the asymptotic behavior of regression parameter estimators. Further, it is worthwhile to point out that the derivation of the information bound does not involve the distributional assumptions on the data other than the proportional mean assumption (2.1).

Let Δ_0 be a diagonal matrix with jth diagonal element $\exp\{Z^\intercal\beta_0+\varphi_0(T_{K,j})\}$, for $j \in \{1,\dots,K\}$, and $\overline{Z} = 1_K Z^\intercal$. Denote by Z_l the lth column of $\overline{Z}$, for $l \in \{1,\dots,d\}$. In the sequel, for any $\psi \in L_2(\mathcal{T})$, let $\psi(T) = \{\psi(T_{K,1}),\dots,\psi(T_{K,K})\}^\intercal$. Define the inner product for a positive definite matrix W as

$$\langle \xi_1, \xi_2\rangle_W = E(\xi_1^\intercal W \xi_2)$$

and the corresponding norm as $\|\xi\|_W^2 = \langle \xi, \xi\rangle_W$. Let $W_0 = \Delta_0\Sigma_0^{-1}\Delta_0$, $W_1 = \Delta_0 V^{-1}\Delta_0$, and $W_2 = \Delta_0 V^{-1}\Sigma_0 V^{-1}\Delta_0$. The least favorable direction is defined as $\kappa^*(\cdot) = (\kappa_1^*(\cdot),\dots,\kappa_d^*(\cdot))$ that satisfies

$$\langle Z_l - \kappa_l^*(T), \kappa(T)\rangle_{W_0} = 0, \quad l \in \{1,\dots,d\},$$

for any $\kappa(\cdot) \in L_2(\mathcal{T})$, and hence has a closed form

$$\kappa^*(T) = \{E(W_0|T)\}^{-1}E(W_0\overline{Z}|T).$$

Theorem 1. *Under the proportional mean model (2.1), the efficient score for β at $\tau = \tau_0$ is given by*

$$\ell_\beta^*(\tau_0) = \{\overline{Z} - \kappa^*(T)\}^\intercal \Delta_0\Sigma_0(\mathbb{N} - \mu_0),$$

where $\mathbb{N} = \mathbb{N}(T)$ and μ_0 represents μ evaluated at $\tau = \tau_0$. Accordingly, the semiparametric efficient information matrix for β at $\tau = \tau_0$ is given by

$$\mathscr{I}_0 \equiv E\{\ell^*_\beta(\tau_0)\}^{\otimes 2} = \|\overline{Z} - \kappa^*(T)\|^2_{W_0}. \tag{3.1}$$

Define $h^*(\cdot) = (h^*_1(\cdot), \ldots, h^*_d(\cdot))$, where $h^*_l(\cdot) \in L_2(\mathcal{T})$ minimizes $\|Z_l - h(T)\|^2_{W_1}$, for $l \in \{1, \ldots, d\}$, or, equivalently, $\langle Z_l - h^*_l(T), h(T)\rangle_{W_1} = 0$, for any $h(\cdot) \in L_2(\mathcal{T})$. It follows that

$$h^*(T) = \{E(W_1|T)\}^{-1} E(W_1\overline{Z}|T).$$

It can be shown that $\kappa^*(\cdot)$ and $h^*(\cdot)$ are bounded and smooth on $\mathcal{T}$ under the regularity conditions C2–C5; see supplementary material S.1 in Cheng et al. [3] for further information. These properties are crucial to establish the asymptotic normality and the semiparametric efficiency for regression parameter estimators. Define

$$\mathscr{I}_1 \equiv \|\overline{Z} - h^*(T)\|^2_{W_1} \quad \text{and} \quad \mathscr{I}_2 \equiv \|\overline{Z} - h^*(T)\|^2_{W_2}. \tag{3.2}$$

If the covariance structure is correctly specified, i.e., $V = \Sigma_0$, $h^*(\cdot)$ turns into the least favorable direction $\kappa^*(\cdot)$, and $\mathscr{I}_1$ and $\mathscr{I}_2$ reduce to the efficient information matrix $\mathscr{I}_0$ accordingly.

3.3. Large sample properties

Parallel to the parametric setting as discussed in Liang and Zeger [12], we show that, under the regularity conditions and the appropriate selection of the dimension of the spline space and the order of the smoothing parameter, the functional estimator attains the optimal rate of convergence, and the asymptotic normality of regression parameter estimators can be established even if the working covariance matrix is misspecified. Further, semiparametric efficiency can be achieved if the working covariance matrix is identical to the true covariance matrix.

Theorem 2. *(Consistency and the rate of convergence) Under conditions C1–C8, the penalized estimator $\widehat{\beta}$ is consistent for β_0, $\|\widehat{\varphi}\|_\infty = O_p(1)$, $\|\widehat{\varphi} - \varphi_0\|_\infty = o_p(1)$, $J(\widehat{\varphi}) = O_p(1)$, and $\|\widehat{\varphi} - \varphi_0\|_{L_2(\mu)} = O_p(n^{-m/(1+2m)})$, which is the optimal rate of convergence for $\widehat{\varphi}$.*

Theorem 3. *(Asymptotic normality and efficiency) Under conditions C1–C9,*

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathscr{I}_1^{-1}\mathscr{I}_2\mathscr{I}_1^{-1}), \quad \text{as } n \to \infty.$$

If the working covariance structure is correctly specified, i.e., $V = \Sigma_0$, then

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \mathscr{I}_0^{-1}\sqrt{n}\mathbb{P}_n\ell^*_\beta(\widehat{\tau}) + o_p(1) \xrightarrow{d} N(0, \mathscr{I}_0^{-1}), \quad \text{as } n \to \infty,$$

where $\mathbb{P}_n$ is the empirical measure, i.e., $\widehat{\beta}$ achieves the semiparametric efficiency bound.

4. Variance estimation

To consistently estimate the variance-covariance matrix of $\widehat{\beta}$, or, equivalently, to consistently estimate $\mathscr{I}_1$ and $\mathscr{I}_2$, we need to first derive the consistent estimator of $h^*(\cdot)$. Again we employ the spline approximation technique. In particular, approximate $h_l^*(\cdot)$ via a B-spline $h_{nl}^*(\cdot)$, for $l \in \{1, \ldots, d\}$, i.e., $h_l^*(\cdot) \approx h_{nl}^*(\cdot) = \sum_{j=1}^{q_n} \zeta_{lj} b_j(\cdot) \equiv \zeta_l^\intercal b(\cdot)$ such that $\|h_l^* - h_{nl}^*\|_\infty = o(1)$, where $\zeta_l = (\zeta_{l1}, \ldots, \zeta_{lq_n})^\intercal$. Let Z_{il} be the lth column of $\overline{Z}_i = (Z_i, \ldots, Z_i)^\intercal$ and $B_i = \{b(T_{K_i,1}^{(i)}), \ldots, b(T_{K_i,K_i}^{(i)})\}^\intercal$, for $i \in \{1, \ldots, n\}$. In view of the definition of $h_l^*(\cdot)$, the spline estimator $\widehat{\zeta}_l$ is defined as the minimizer of the function

$$1/n \sum_{i=1}^n \{Z_{il} - h_{nl}^*(T_i)\}^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i \{Z_{il} - h_{nl}^*(T_i)\},$$

where $\widehat{\Delta}_i$ is a diagonal matrix with jth diagonal element $\exp\{Z_i^\intercal \widehat{\beta} + \widehat{\varphi}(T_{K_i,j}^{(i)})\}$, for $j \in \{1, \ldots, K_i\}$, and consequently the spline estimator of $h_l^*(\cdot)$ is defined as $\widehat{h}_l^*(\cdot) = \widehat{\zeta}_l^\intercal b(\cdot)$. Define

$$\widehat{\mathcal{H}} = \begin{pmatrix} \widehat{\mathcal{H}}_{11} & \widehat{\mathcal{H}}_{12} \\ \widehat{\mathcal{H}}_{21} & \widehat{\mathcal{H}}_{22} \end{pmatrix} \equiv \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n \overline{Z}_i^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i \overline{Z}_i & \sum_{i=1}^n \overline{Z}_i^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i B_i \\ \sum_{i=1}^n B_i^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i \overline{Z}_i & \sum_{i=1}^n B_i^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i B_i \end{pmatrix}.$$

Notice that $\widehat{\mathcal{H}}$ is equivalent to the expected Hessian matrix of the penalized spline objective function (2.5) without the penalized term, i.e., $\widehat{\mathcal{H}}$ is equivalent to $\mathcal{I}_{n,\lambda}(\theta)$ evaluated at $\theta = \widehat{\theta}$ and $\lambda = 0$. It concludes from matrix algebra that

$$\widehat{h}_l^*(T_i) = B_i \widehat{\zeta}_l = \frac{1}{n} B_i \widehat{\mathcal{H}}_{11}^{-1} \sum_{i=1}^n B_i^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i Z_{il},$$

and hence

$$\widehat{h}^*(T_i) = \left\{\widehat{h}_1^*(T_i), \ldots, \widehat{h}_d^*(T_i)\right\} = B_i \widehat{\mathcal{H}}_{22}^{-1} \widehat{\mathcal{H}}_{21}.$$

Let

$$\widehat{\mathscr{I}_{n1}} = 1/n \sum_{i=1}^n \{\overline{Z}_i - \widehat{h}^*(T_i)\}^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Delta}_i \{\overline{Z}_i - \widehat{h}^*(T_i)\}$$

and

$$\widehat{\mathscr{I}_{n2}} = 1/n \sum_{i=1}^n \{\overline{Z}_i - \widehat{h}^*(T_i)\}^\intercal \widehat{\Delta}_i V_i^{-1} \widehat{\Sigma}_i V_i^{-1} \widehat{\Delta}_i \{\overline{Z}_i - \widehat{h}^*(T_i)\},$$

for $\widehat{\Sigma}_i = (\mathbb{N}^{(i)} - \widehat{\mu}_i)^{\otimes 2}$. It follows from matrix algebra that $\widehat{\mathscr{I}_{n1}} = \widehat{\mathcal{H}}_{11} - \widehat{\mathcal{H}}_{12} \widehat{\mathcal{H}}_{22}^{-1} \widehat{\mathcal{H}}_{21}$. Further, as shown in Theorem 4, $\widehat{\mathscr{I}_{n1}}$ and $\widehat{\mathscr{I}_{n2}}$ are the consistent estimators of $\mathscr{I}_1$ and $\mathscr{I}_2$, respectively. Therefore, the asymptotic variance-covariance matrix of $\widehat{\beta}$ can be consistently estimated by $\mathscr{V}_R = \widehat{\mathscr{I}_{n1}}^{-1} \widehat{\mathscr{I}_{n2}} \widehat{\mathscr{I}_{n1}}^{-1}$.

Theorem 4. *(Variance estimation) Under conditions C1–C9, $\widehat{\mathscr{I}_{n1}}$ and $\widehat{\mathscr{I}_{n2}}$ are consistent for $\mathscr{I}_1$ and $\mathscr{I}_2$, respectively.*

Our numerical experiments reveal that the asymptotic consistency does not hold in small sample settings. The variance-covariance estimator tends to underestimate the standard error of $\widehat{\beta}$, even if the working covariance matrix is correctly specified. It has been extensively discussed in the literature that the sandwich-type estimator tends to yield biased variance estimation for small sample sizes. To cope with this issue, many researchers have proposed methodologies to improve the performance of the sandwich estimator under these circumstances; see, for example, Mancl and DeRouen [19] and Pan [22], among many others. To adjust for the bias incurred by the sandwich estimator, Morel et al. [20] recommended an inflated estimator by adding a scaled version of trace to the sandwich estimator. Incorporating the suggestion by Morel et al. [20], we proposed an adjusted estimator to atone for the deficiency in variance estimation. Define

$$\widehat{\mathcal{B}} = 1/n \sum_{i=1}^{n} E_i^{\intercal} \widehat{\Delta}_i V_i \left(\mathbb{N}^{(i)} - \overline{\mu}_i \right)^{\otimes 2} V_i \widehat{\Delta}_i E_i.$$

The modified variance-covariance estimator is referred to as

$$\mathscr{V}_M = \mathscr{V}_R + \frac{1}{n} \delta \phi \widehat{\mathscr{I}_{n1}^{-1}},$$

where $\delta = \min\{0.5, \rho/(n-\rho)\}$ and $\phi = \max\{1, \rho^{-1}\mathrm{trace}(\widehat{\mathcal{H}}^{-1}\widehat{\mathcal{B}})\}$. The performance of the proposed variance-covariance estimation method was evaluated in Monte Carlo studies and the results indicate that the modified estimator provides more accurate variance-covariance estimation compared to the standard sandwich estimator in various settings. The approach was further employed for inference on regression parameters in the real application.

5. Numerical illustration

In this section, we investigate the performance of the proposed methodology under a variety of simulation settings. Each participant is scheduled for 6 follow-up visits and for each visit the subject might choose to skip with a nonzero probability. As a result, it is assumed that the total number of visits for the ith individual takes each of the values $\{1, 2, \ldots, 6\}$ with equal probability, i.e., $P(K_i = k) = 1/6$, for $k \in \{1, \ldots, 6\}$, and the time between two consecutive visits is generated from Uniform $(0, 10)$. In addition, the vector of covariates for the ith individual, $Z_i = (Z_{i1}, Z_{i2})^{\intercal}$, is simulated as $Z_{i1} \sim \mathrm{Normal}(0, 1)$ and $Z_{i2} \sim \mathrm{Bernoulli}(0.5)$ with regression parameters given by $\beta = (-1, 1)^{\intercal}$. Assume a proportional mean model $E\{\mathbb{N}(t)|Z)\} = \Lambda(t)\exp(Z^{\intercal}\beta)$ with the baseline cumulative mean function $\Lambda(t) = 2t^{1/2}$.

To demonstrate the wide-ranging applicability of the proposed methodology, we consider 5 different simulation settings in which the underlying counting processes to generate panel counts differ. In Simulation I (S1), the panel count data were generated from a Poisson process with the conditional mean $\Lambda(t)\exp(Z^{\intercal}\beta)$, i.e., given the covariate Z_i, $\Delta\mathbb{N}_{ij} \sim$ Poisson $\{\Delta\Lambda_{ij} \exp(Z_i^\intercal \beta)\}$, where $\Delta\mathbb{N}_{ij} = \mathbb{N}_{ij} - \mathbb{N}_{i(j-1)}$ and $\Delta\Lambda_{ij} = \Lambda(T^{(i)}_{K_i,j}) - \Lambda(T^{(i)}_{K_i,j-1})$, for $j \in \{1, \ldots, K_i\}$. In Simulation II (S2) and Simulation III (S3), the panel count data were generated from gamma frailty Poisson processes, i.e., given the covariate Z_i and the gamma frailty variable γ_i, $\Delta\mathbb{N}_{ij} \sim$ Poisson $\{\gamma_i \Delta\Lambda_{ij} \exp(Z_i^\intercal \beta)\}$. To assess the impact of overdispersion on the proposed methodology, the frailty variable γ_i is assumed to be Gamma(1,1) or Gamma(1/2,1/2) in S2 or S3, yielding the overdispersion parameter $\sigma^2 = 1$ or 2, respectively. To further investigate the robustness of the proposed methodology, we consider scenarios in which the working matrices (i.e., $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$) are all misspecified. Specifically, in Simulation IV (S4), the panel count data were generated from a mixture Poisson process, i.e., given the covariate Z_i and the discrete random variable γ_i from $(-0.4,0,0.4)$ with corresponding probability (1/4,1/2/,1/4), $\Delta\mathbb{N}_{ij} \sim$ Poisson$\{(1 + \gamma_i)\Delta\Lambda_{ij} \exp(Z_i^\intercal \beta)\}$, and in Simulation V (S5), the panel count data were generated from a log-normal frailty Poisson process, i.e., given the covariate Z_i and the log-normal variable γ_i with mean 1 and variance 1, $\Delta\mathbb{N}_{ij} \sim$ Poisson $\{\gamma_i \Delta\Lambda_{ij} \exp(Z_i^\intercal \beta)\}$. Clearly, the panel counts are *not* marginally Poisson distributed in settings S2 through S5, but conditional on the covariates, the means of panel count data still satisfy the proportional mean model (2.1). In all these settings, cubic monotone B-splines were applied to approximate $\varphi(\cdot)$. The selection of the spline knots and the smoothing parameter follows the discussion presented in Sections 2.3 and 2.4. For each simulation configuration, 1,000 Monte Carlo samples were generated with $n = 50$ or $n = 100$.

Table 1 summarizes the results for estimating regression parameters with working matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$ under settings S1–S3, including the empirical bias, standard deviation (SD), relative efficiency (RE) defined as the ratio of mean squared errors between the estimator with $V_i^{(3)}$ and that with $V_i^{(1)}$ or $V_i^{(2)}$, averages of estimated standard errors based on the sandwich estimator (SSE) and the modified estimator (MSE), and estimated coverage probabilities denoted by CP1 and CP2, which are based on the sandwich estimator and the modified estimator, respectively. These results indicate that all empirical biases of the estimators with all three working matrices are negligible compared to standard deviations under all settings, and the biases and standard deviations decrease as the sample size is increased, which suggests the asymptotic consistency for the GEE penalized spline estimators. Moreover, histograms and quantile-quantile (Q-Q) plots (results not shown) reveal that the GEE penalized estimators with all three working matrices are approximately normally distributed, which provides the numerical justification for the asymptotic normality of regression parameter estimators established in Theorem 3, even in the case that the working matrix is misspecified. Also, it is observed that under all simulation configurations the estimation with $V_i^{(2)}$ or $V_i^{(3)}$ tends to provide smaller mean squared errors and more satisfactory 95% coverage probabilities than that with $V_i^{(1)}$, which fails to account for the positive association among the panel counts.

Table 1. *Summary of simulation results for penalized spline GEE estimators of the regression coefficients* $\beta = (\beta_1, \beta_2)^\intercal = (-1, 1)^\intercal$ *with working covariance matrices* $V_i^{(1)}$, $V_i^{(2)}$, *and* $V_i^{(3)}$ *under simulation settings S1–S3; SD, standard deviation of the estimates; RE, relative efficiency (%) (defined as the ratio of mean squared errors between the estimator with* $V_i^{(3)}$ *and that with* $V_i^{(1)}$ *or* $V_i^{(2)}$*); SSE and MSE are the averages of standard errors based on the sandwich estimator and modified estimator, respectively; CP1 and CP2 are the empirical coverage probabilities (%) based on the sandwich estimator and modified estimator, respectively.*

			$n = 50$							$n = 100$						
			Bias	SD	RE	SSE	MSE	CP1	CP2	Bias	SD	RE	SSE	MSE	CP1	CP2
S1	$V_i^{(1)}$	β_1	−0.003	0.051	76.0	0.042	0.044	89.0	90.6	−0.001	0.031	81.8	0.028	0.029	93.1	93.2
		β_2	0.000	0.102	76.0	0.086	0.091	90.3	91.8	−0.001	0.064	77.1	0.062	0.064	92.2	93.6
	$V_i^{(2)}$	β_1	−0.001	0.044	101.7	0.037	0.044	89.4	94.7	−0.000	0.027	104.5	0.025	0.028	93.5	95.1
		β_2	0.001	0.088	101.6	0.078	0.092	91.7	94.8	−0.000	0.056	100.3	0.055	0.060	92.9	95.7
	$V_i^{(3)}$	β_1	−0.001	0.045	100.0	0.038	0.045	89.5	94.7	−0.001	0.028	100.0	0.025	0.028	93.7	95.2
		β_2	−0.001	0.089	100.0	0.078	0.093	91.4	94.8	−0.000	0.056	100.0	0.055	0.060	93.3	95.6
S2	$V_i^{(1)}$	β_1	0.039	0.244	47.3	0.171	0.212	77.9	85.2	0.033	0.183	41.1	0.135	0.152	81.9	86.1
		β_2	0.006	0.414	56.7	0.350	0.434	88.6	94.9	−0.013	0.328	46.8	0.279	0.315	88.6	92.5
	$V_i^{(2)}$	β_1	0.043	0.238	49.6	0.161	0.210	73.9	85.9	0.034	0.177	43.6	0.129	0.149	79.4	85.1
		β_2	0.004	0.390	63.9	0.331	0.434	89.0	96.1	−0.012	0.306	54.0	0.261	0.304	90.0	93.8
	$V_i^{(3)}$	β_1	0.003	0.170	100.0	0.142	0.170	89.5	94.2	0.004	0.119	100.0	0.106	0.117	91.0	93.5
		β_2	0.012	0.312	100.0	0.292	0.345	93.0	96.1	0.003	0.225	100.0	0.213	0.233	93.1	94.8
S3	$V_i^{(1)}$	β_1	0.098	0.320	49.2	0.218	0.282	75.8	86.8	0.053	0.237	40.7	0.179	0.211	79.3	84.5
		β_2	0.013	0.512	59.9	0.457	0.589	90.8	96.5	0.012	0.414	50.9	0.374	0.442	91.6	95.5
	$V_i^{(2)}$	β_1	0.094	0.314	51.3	0.203	0.278	73.5	85.2	0.053	0.236	41.1	0.169	0.203	77.6	84.5
		β_2	0.004	0.490	65.4	0.431	0.585	90.3	97.8	−0.003	0.395	55.9	0.349	0.421	91.6	95.8
	$V_i^{(3)}$	β_1	0.028	0.233	100.0	0.183	0.221	88.1	92.0	0.011	0.155	100.0	0.139	0.157	92.0	93.2
		β_2	0.018	0.396	100.0	0.387	0.457	93.6	96.6	−0.001	0.295	100.0	0.284	0.319	93.5	95.8

When panel count data follow a Poisson process as specified in S1, i.e., the overdispersion parameter $\sigma^2 = 0$, the penalized estimation with $V_i^{(2)}$ or $V_i^{(3)}$ displays almost identical finite-sample performance due to the fact that the overdispersion estimator $\widehat{\sigma}^2$ obtained from (2.6) is almost equal to 0 in this scenario. When the underlying process to generate panel count data is a gamma frailty Poisson process with $\sigma^2 = 1$ or $\sigma^2 = 2$ as described in S2 or S3, respectively, due to accounting for the overdispersion, the GEE estimator with $V_i^{(3)}$ exceeds that with $V_i^{(1)}$ or $V_i^{(2)}$ in terms of the smaller mean squared errors and estimated standard errors as well as more reasonable coverage probabilities. In all simulation settings, even in the case that the working matrix is correctly specified, the estimation based on the sandwich method underestimates the standard error regardless of the sample size, and consequently, the corresponding coverage probability tends to be less than the nominal level. On the other hand, the proposed method provides superior variance-covariance estimation. In particular, the averages of estimated standard errors agree with the standard deviations of the estimates, and the corresponding empirical coverage probabilities are close to the nominal level. As displayed in Table 1, with extra overdispersion involved in the underlying counting process, the estimation is liable to be less accurate with increased mean squared error and less favorable empirical coverage probability, regardless of the specification of the working matrix.

To demonstrate the robustness of the proposed methodology in the case where the underlying counting process is misspecified, panel count data were generated from a mixture Poisson process in S4 or a log-normal frailty Poisson process in S5. As presented in Table 2, the proposed methodology still displays adequate numerical properties in the matter of the empirical bias, standard deviation, and empirical coverage probability. As expected, the estimator with $V_i^{(3)}$ outperforms that with $V_i^{(1)}$ or $V_i^{(2)}$, and the variance-covariance estimation based on the modified estimator provides more satisfactory coverage probability compared to the one based on the sandwich formula. It concludes that the proposed methodology is robust to the specification of the underlying counting process up to a point.

The pointwise mean estimates of $\varphi(\cdot)$ and the corresponding 2.5th and 97.5th percentiles of 1,000 Monte Carlo samples as well as the true curve of the function with all three different working covariance matrices under settings S1–S5 are presented in Figures 1 and 2. It is observed that all fitted spline curves are reasonably close to the true function. The proposed penalized method with $V_i^{(3)}$ provides less biased functional estimates and narrower pointwise confidence intervals compared to that with $V_i^{(1)}$ or $V_i^{(2)}$ when the overdispersion is presented as displayed in Figure 1, and is robust to the specification of the underlying counting process as presented in Figure 2. Further, the bias and variability are decreased when the sample size is increased.

The penalized spline method was also compared to the regression spline approach proposed by Hua and Zhang [7] (i.e., $\lambda = 0$) under S1–S5. As presented in Tables 3 and 4 and Figures 3 and 4, the two methods provide comparable results in terms of bias and standard deviation of the estimates. One explana-

Table 2. *Summary of simulation results for penalized spline GEE estimators of the regression coefficients $\beta = (\beta_1, \beta_2)^\intercal = (-1, 1)^\intercal$ with working covariance matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$ under simulation settings S4 and S5; SD, standard deviation of the estimates; RE, relative efficiency (%) (defined as the ratio of mean squared errors between the estimator with $V_i^{(3)}$ and that with $V_i^{(1)}$ or $V_i^{(2)}$); SSE and MSE are the averages of standard errors based on the sandwich estimator and modified estimator, respectively; CP1 and CP2 are the empirical coverage probabilities (%) based on the sandwich estimator and modified estimator, respectively.*

			$n = 50$							$n = 100$						
			Bias	SD	RE	SSE	MSE	CP1(%)	CP2	Bias	SD	RE(%)	SSE	MSE	CP1	CP2
S4	$V_i^{(1)}$	β_1	0.003	0.095	59.1	0.065	0.071	80.7	83.0	0.000	0.067	53.5	0.050	0.053	83.1	84.4
		β_2	0.007	0.158	66.5	0.133	0.144	89.3	91.0	−0.006	0.115	62.7	0.101	0.107	92.4	91.9
	$V_i^{(2)}$	β_1	0.004	0.090	68.6	0.062	0.073	82.4	87.1	0.001	0.063	60.4	0.048	0.053	84.5	87.8
		β_2	0.007	0.146	78.2	0.126	0.148	89.4	93.8	−0.006	0.104	75.8	0.095	0.104	92.3	94.5
	$V_i^{(3)}$	β_1	0.000	0.075	100.0	0.063	0.075	90.0	93.0	0.000	0.047	100.0	0.046	0.050	91.8	95.2
		β_2	0.006	0.133	100.0	0.123	0.144	93.6	96.0	−0.004	0.092	100.0	0.088	0.096	94.7	96.0
S5	$V_i^{(1)}$	β_1	0.048	0.219	52.4	0.148	0.179	75.9	82.3	0.027	0.173	41.0	0.122	0.137	80.4	84.4
		β_2	0.010	0.384	58.8	0.311	0.376	88.5	93.7	0.002	0.308	49.5	0.251	0.285	89.1	92.4
	$V_i^{(2)}$	β_1	0.046	0.214	54.5	0.140	0.179	74.2	83.8	0.029	0.169	42.2	0.118	0.135	78.3	83.6
		β_2	0.001	0.361	66.6	0.296	0.379	89.7	95.3	0.003	0.290	55.8	0.238	0.277	89.5	94.1
	$V_i^{(3)}$	β_1	0.010	0.162	100.0	0.131	0.158	85.9	92.2	0.009	0.109	100.0	0.101	0.111	91.3	94.6
		β_2	0.008	0.294	100.0	0.273	0.322	92.8	96.7	−0.000	0.216	100.0	0.202	0.223	92.5	95.1

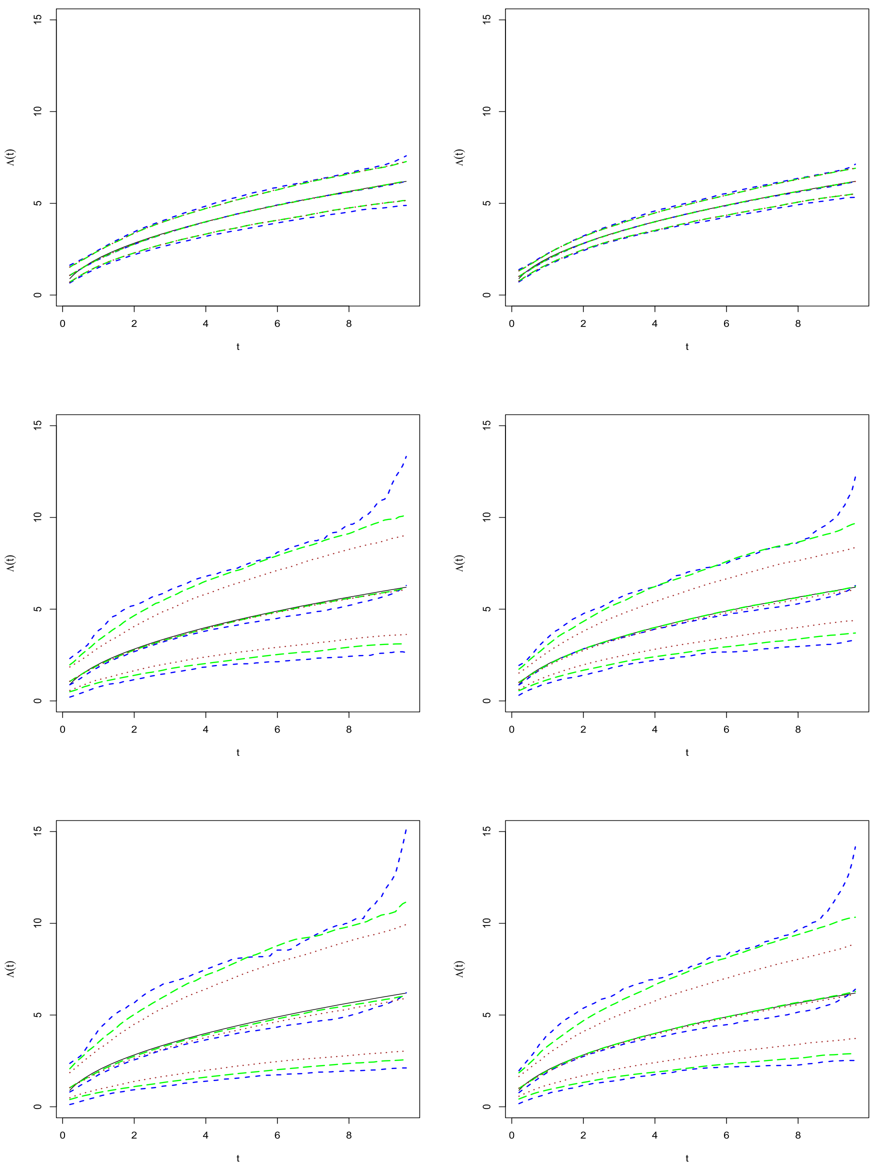

FIG 1. *Penalized spline estimates and corresponding 95% pointwise confidence intervals of* $\Lambda(t) = 2t^{1/2}$ *under simulation settings S1–S3. From top to bottom, the panels correspond to S1, S2 and S3, respectively, while left and right panels correspond to* $n = 50$ *and* $n = 100$*, respectively, in each simulation setting. The solid curves correspond to the true value of the function. The dashed curves, longdashed curves, and dotted curves are the averages of estimates along with* $2.5th$ *and* $97.5th$ *percentiles of the estimates from 1,000 Monte Carlo samples with* $V_i^{(1)}$, $V_i^{(2)}$, *and* $V_i^{(3)}$, *respectively.*

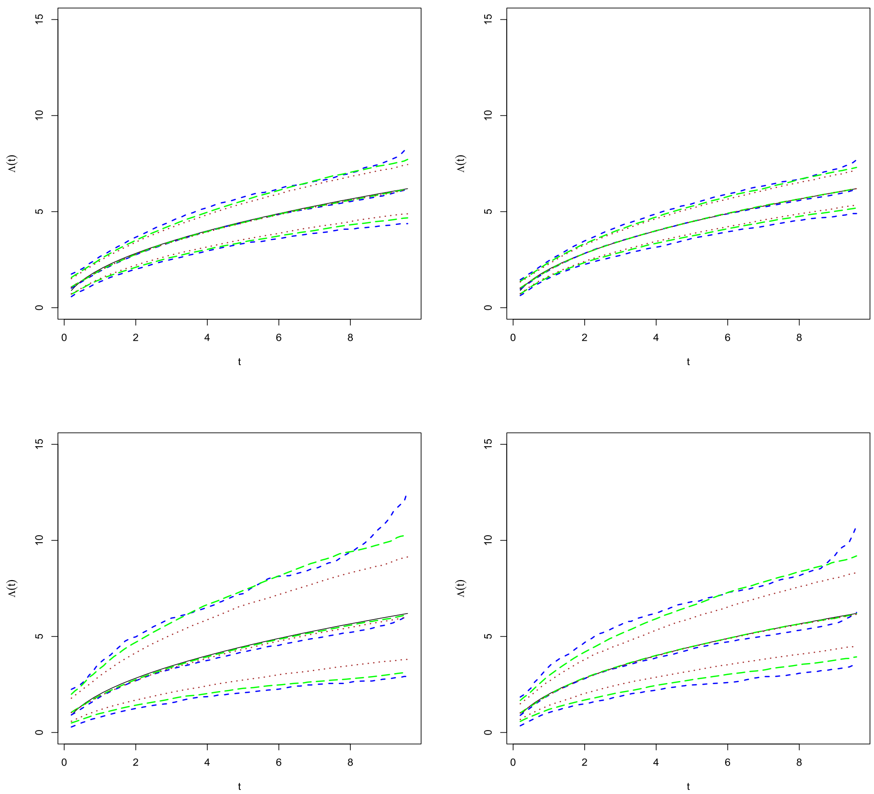

FIG 2. *Penalzied spline estimates and corresponding 95% pointwise confidence intervals of* $\Lambda(t) = 2t^{1/2}$ *under simulation settings S4 and S5. From top to bottom, the panels correspond to S4 and S5, respectively, while left and right panels correspond to* $n = 50$ *and* $n = 100$*, respectively, in each simulation setting. The solid curves correspond to the true value of the function. The dashed curves, longdashed curves, and dotted curves are the averages of estimates along with* $2.5th$ *and* $97.5th$ *percentiles of the estimates from 1,000 Monte Carlo samples with* $V_i^{(1)}$, $V_i^{(2)}$, *and* $V_i^{(3)}$, *respectively.*

tion for this phenomenon is that the monotone spline estimator is constrained, i.e., under monotone constraints, and hence, the impact of the penalization is not as significant as in the case of non-constrained spline estimation. Although the penalized estimators do not outperform the spline alternatives under the proposed simulation settings, the penalized technique is more computationally stable compared to the spline alternative. Further, the penalized method displays preferable properties in variance-covariance estimation. More specifically, the variance estimator has less variability and the corresponding empirical coverage probabilities are close to the nominal level.

Table 3

Summary of simulation results for regression spline and penalized spline GEE estimators of the regression coefficients $\beta = (\beta_1, \beta_2)^\intercal = (-1, 1)^\intercal$ with working covariance matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$ under simulation settings S1–S3; SD, standard deviation of the estimates; SSE and MSE are the averages of standard errors based on the sandwich estimator and modified estimator, respectively; CP1 and CP2 are the empirical coverage probabilities (%) based on the sandwich estimator and modified estimator, respectively.

				Regression Spline				Penalized Spline			
				Bias	SD	SSE	CP1	Bias	SD	MSE	CP2
S1	$n = 50$	$V_i^{(1)}$	β_1	−0.003	0.052	0.043	89.3	−0.003	0.051	0.044	90.6
			β_2	0.000	0.103	0.087	90.0	0.000	0.102	0.091	91.8
		$V_i^{(2)}$	β_1	−0.001	0.044	0.037	89.7	−0.001	0.044	0.044	94.7
			β_2	0.001	0.088	0.078	91.3	0.001	0.088	0.092	94.8
		$V_i^{(3)}$	β_1	−0.001	0.045	0.038	89.9	−0.001	0.045	0.045	94.7
			β_2	0.001	0.089	0.078	91.2	0.001	0.089	0.093	94.8
	$n = 100$	$V_i^{(1)}$	β_1	−0.000	0.032	0.029	92.4	−0.001	0.031	0.029	93.2
			β_2	−0.001	0.065	0.062	93.1	−0.001	0.064	0.064	93.6
		$V_i^{(2)}$	β_1	−0.007	0.027	0.025	93.0	−0.000	0.027	0.028	95.1
			β_2	−0.000	0.056	0.055	93.2	−0.000	0.056	0.060	95.7
		$V_i^{(3)}$	β_1	−0.001	0.028	0.025	92.9	−0.001	0.028	0.028	95.2
			β_2	−0.000	0.056	0.055	93.2	−0.000	0.056	0.060	95.6
S2	$n = 50$	$V_i^{(1)}$	β_1	0.039	0.244	0.166	77.0	0.039	0.244	0.212	85.2
			β_2	0.007	0.415	0.334	88.1	0.006	0.414	0.434	94.9
		$V_i^{(2)}$	β_1	0.043	0.238	0.161	74.1	0.043	0.238	0.210	85.9
			β_2	0.004	0.390	0.331	89.2	0.004	0.390	0.434	96.1
		$V_i^{(3)}$	β_1	0.001	0.171	0.143	89.5	0.003	0.170	0.170	94.2
			β_2	0.014	0.311	0.293	93.0	0.012	0.312	0.345	96.1
	$n = 100$	$V_i^{(1)}$	β_1	0.032	0.181	0.132	82.0	0.033	0.183	0.152	86.1
			β_2	−0.012	0.328	0.271	88.3	−0.013	0.328	0.315	92.5
		$V_i^{(2)}$	β_1	0.034	0.177	0.129	79.6	0.034	0.177	0.149	85.1
			β_2	−0.013	0.306	0.261	90.0	−0.012	0.306	0.304	93.8
		$V_i^{(3)}$	β_1	0.038	1.069	0.107	90.7	0.004	0.119	0.117	93.5
			β_2	0.046	1.391	0.213	92.8	0.003	0.225	0.233	94.8
S3	$n = 50$	$V_i^{(1)}$	β_1	0.096	0.319	0.208	74.0	0.098	0.320	0.282	86.8
			β_2	0.015	0.511	0.427	88.8	0.013	0.512	0.589	96.5
		$V_i^{(2)}$	β_1	0.094	0.313	0.203	73.3	0.094	0.314	0.278	85.2
			β_2	0.005	0.490	0.431	90.4	0.004	0.490	0.585	97.8
		$V_i^{(3)}$	β_1	0.027	0.233	0.183	87.9	0.028	0.233	0.221	92.0
			β_2	0.019	0.397	0.387	93.5	0.018	0.396	0.457	96.6
	$n = 100$	$V_i^{(1)}$	β_1	0.053	0.236	0.172	79.1	0.053	0.237	0.211	84.5
			β_2	0.012	0.414	0.355	90.5	0.012	0.414	0.442	95.5
		$V_i^{(2)}$	β_1	0.053	0.236	0.169	77.5	0.053	0.236	0.203	84.5
			β_2	−0.002	0.395	0.349	91.6	−0.003	0.395	0.421	95.8
		$V_i^{(3)}$	β_1	0.037	0.841	0.139	91.8	0.011	0.155	0.157	93.2
			β_2	0.009	0.447	0.284	93.3	−0.001	0.295	0.319	95.8

6. Real data analysis

The proposed penalized methodology was applied to the skin cancer chemoprevention data. In this clinical trial, study subjects were randomly assigned

Table 4
Summary of simulation results for regression spline and penalized spline GEE estimators of the regression coefficients $\beta = (\beta_1, \beta_2)^\intercal = (-1, 1)^\intercal$ with working covariance matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$ under simulation settings S4 and S5; SD, standard deviation of the estimates; SSE and MSE are the averages of standard errors based on the sandwich estimator and modified estimator, respectively; CP1 and CP2 are the empirical coverage probabilities (%) based on the sandwich estimator and modified estimator, respectively.

				Regression Spline				Penalized Spline			
				Bias	SD	SSE	CP1	Bias	SD	MSE	CP2
S4	$n = 50$	$V_i^{(1)}$	β_1	0.002	0.094	0.067	81.4	0.003	0.095	0.071	83.0
			β_2	0.007	0.157	0.134	89.2	0.007	0.158	0.144	91.0
		$V_i^{(2)}$	β_1	0.004	0.090	0.063	80.0	0.004	0.090	0.073	87.1
			β_2	0.007	0.146	0.126	89.7	0.007	0.146	0.148	93.8
		$V_i^{(3)}$	β_1	0.000	0.075	0.064	89.1	0.009	0.075	0.075	93.0
			β_2	0.006	0.133	0.123	92.3	0.006	0.133	0.144	96.0
	$n = 100$	$V_i^{(1)}$	β_1	0.000	0.067	0.051	84.1	0.000	0.067	0.053	84.4
			β_2	−0.006	0.115	0.102	90.6	−0.006	0.115	0.107	91.9
		$V_i^{(2)}$	β_1	0.001	0.063	0.049	85.1	0.001	0.063	0.053	87.8
			β_2	−0.006	0.104	0.096	92.7	−0.006	0.104	0.104	94.5
		$V_i^{(3)}$	β_1	−0.000	0.047	0.046	93.4	−0.000	0.047	0.050	95.2
			β_2	−0.004	0.092	0.088	94.0	−0.004	0.092	0.096	96.0
S5	$n = 50$	$V_i^{(1)}$	β_1	0.047	0.216	0.145	76.1	0.048	0.219	0.179	82.3
			β_2	0.012	0.382	0.300	87.0	0.010	0.384	0.376	93.7
		$V_i^{(2)}$	β_1	0.046	0.214	0.140	74.2	0.046	0.214	0.179	83.8
			β_2	0.001	0.361	0.296	89.5	0.001	0.361	0.379	95.3
		$V_i^{(3)}$	β_1	0.010	0.162	0.132	86.0	0.010	0.162	0.158	92.2
			β_2	0.008	0.294	0.273	92.7	0.008	0.294	0.322	96.7
	$n = 100$	$V_i^{(1)}$	β_1	0.026	0.172	0.120	79.9	0.027	0.173	0.137	84.4
			β_2	0.002	0.308	0.247	88.9	0.002	0.308	0.285	92.4
		$V_i^{(2)}$	β_1	0.029	0.169	0.117	78.7	0.029	0.169	0.135	83.6
			β_2	0.003	0.290	0.239	89.4	0.003	0.290	0.277	94.1
		$V_i^{(3)}$	β_1	0.009	0.109	0.100	92.0	0.009	0.109	0.111	94.6
			β_2	−0.000	0.216	0.202	92.3	−0.000	0.216	0.223	95.1

to either a DFMO group or a placebo group. The primary goal was to assess the efficacy of DFMO in the reduction of the recurrence of two types of non-melanoma skin cancers, BCC and SCC. The data include 290 study subjects with at least one follow-up visit. The number and the times of examinations vary greatly from individual to individual. In particular, the number of follow-up visits ranges from 1 to 17, while the examination times range from 11 to 1,879 days. Further, the cumulative count of the recurrence of combined BCC and SCC tumors ranges from 0 to 29. Among these patients, the majority were males ($n = 174$, 60%) and aged 65 years old or above ($n = 119$, 62.6%), around half of participants ($n = 143$, 49.4%) were randomly assigned to the DFMO group, and the average count of prior skin cancers from first diagnosis to randomization is 4.33. We fit the data via the following proportional mean model, namely,

$$E\{\mathbb{N}(t)|Z\} = \Lambda(t)\exp(Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_4\beta_4),$$

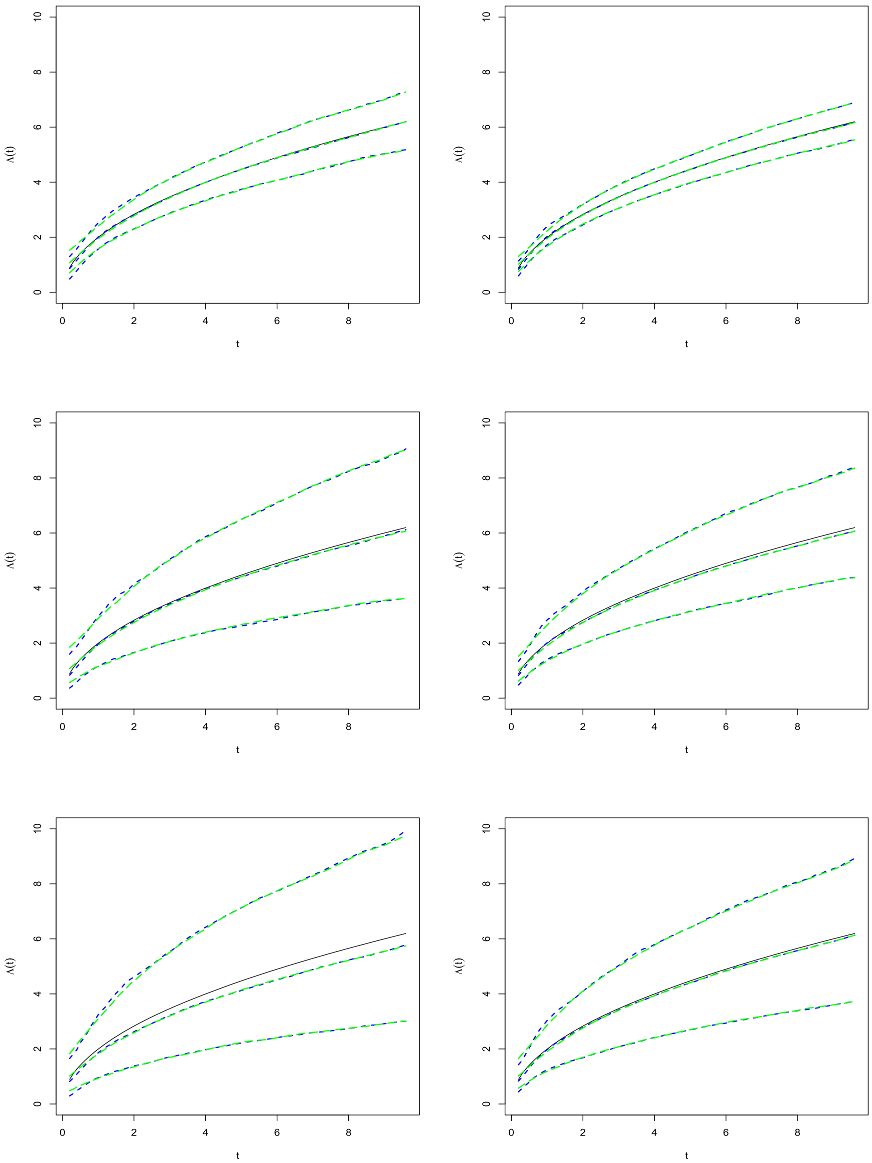

FIG 3. *Regression spline and penalized spline estimates and corresponding 95% pointwise confidence intervals of* $\Lambda(t) = 2t^{1/2}$ *under simulation settings S1–S3. From top to bottom, the panels correspond to S1, S2 and S3, respectively, while left and right panels correspond to* $n = 50$ *and* $n = 100$*, respectively, in each simulation setting. The solid curves correspond to the true value of the function. The dashed curves and longdashed curves are the averages of regression spline and penalized spline estimates, respectively, along with* 2.5*th and* 97.5*th percentiles of the estimates from 1,000 Monte Carlo samples with* $V_i^{(3)}$.

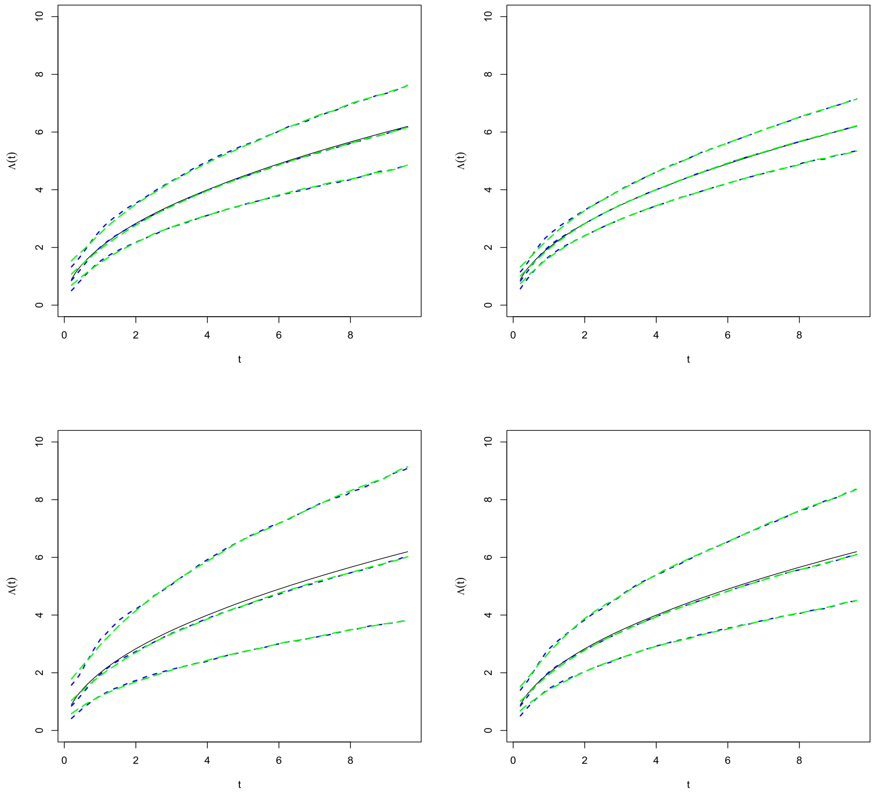

FIG 4. *Regression spline and penalized spline estimates and corresponding 95% pointwise confidence intervals of* $\Lambda(t) = 2t^{1/2}$ *under simulation settings S4 and S5. From top to bottom, the panels correspond to S4 and S5, respectively, while left and right panels correspond to* $n = 50$ *and* $n = 100$*, respectively, in each simulation setting. The solid curves correspond to the true value of the function. The dashed curves and longdashed curves are the averages of regression spline and penalized spline estimates, respectively, along with* $2.5th$ *and* $97.5th$ *percentiles of the estimates from 1,000 Monte Carlo samples with* $V_i^{(3)}$.

where Z_1 is the age at enrollment (1 for those aged 65 years old or above and 0 otherwise), Z_2 is gender (1 for male and 0 for female), Z_3 is a treatment indicator (1 for DFMO and 0 for placebo), and Z_4 is the number of prior non-melanoma tumors. In this analysis the outcome variable is defined as the combined panel count of two non-melanoma skin cancers. The baseline cumulative mean function $\Lambda(\cdot)$ was approximated by a monotone cubic B-spline. We applied the empirical rule discussed in Section 2.4 to select the number and the location of knots and used the generalized Fellnee-Schall approach and the moment estimation method to determine the smoothing parameter λ^2 and the overdispersion parameter σ^2, respectively. The model fitting was carried out via the proposed two-stage iterative algorithm.

Table 3 summarizes the results of data analysis, including the estimates of

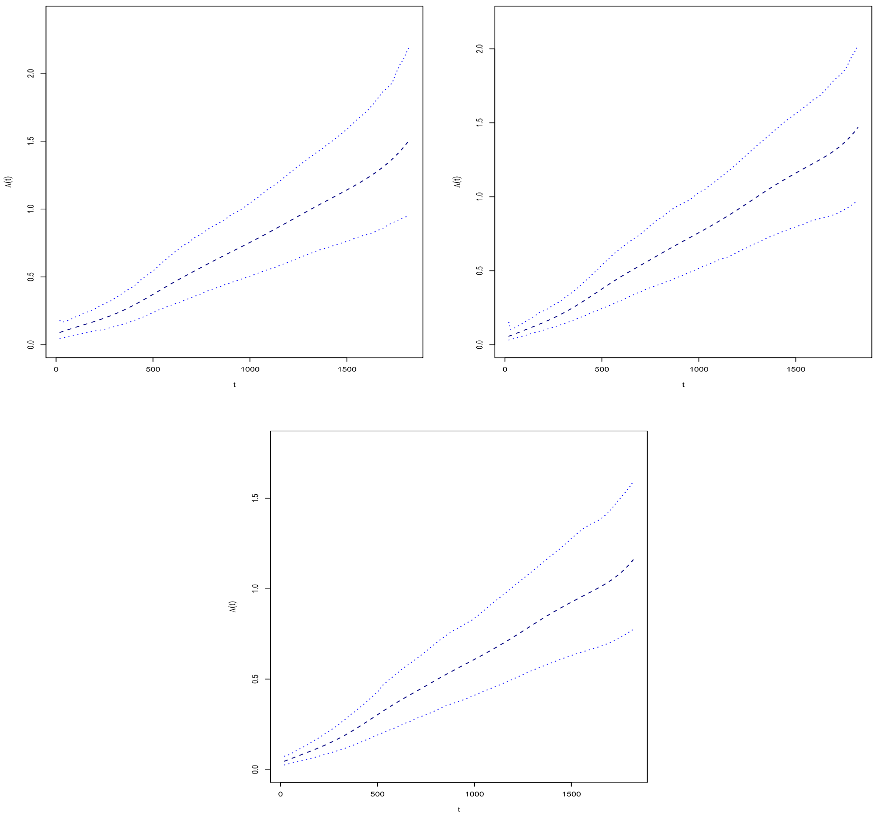

FIG 5. *Plots of the penalized spline estimates of $\Lambda(\cdot)$ for skin cancer chemoprevention data; the dashed curves correspond to the functional estimates, while the dotted curves are 2.5th and 97.5th percentiles of the estimates from 1,000 bootstrap samples. From left to right, the panels correspond to estimations with working covariance matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$, respectively.*

the parameters with working matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$, the estimated standard errors based on the sandwich estimator and the modified estimator, and the corresponding p-values, as well as the estimated values of λ^2 and σ^2. The analyses with the three different working covariance matrices yield the consistent results that being male or increasing of age is prone to be susceptible to a higher number of tumor recurrences on average and the chemoprevention treatment suppresses the tumor recurrence, but none of the three factors is statistically significant. The number of prior skin cancers is shown to be significantly associated with the recurrence. In particular, with every additional prior skin cancer, the newly developed tumor count was estimated to increase by 8.2%, 8.0%, or 11.5% with $V_1^{(1)}$, $V_i^{(2)}$, or $V_i^{(3)}$, respectively. These results agree with those discussed in Chiou et al. [4] under the accelerated mean model. One explanation for the consistent results with different working matrices is that the

TABLE 5
Summary of penalized spline estimation with working matrices $V_i^{(1)}$, $V_i^{(2)}$, and $V_i^{(3)}$ for skin cancer chemoprevention data; estimate: point estimate; SSE and MSE are estimated standard errors based on the sandwich estimator and modified estimator, respectively; λ^2 and σ^2 are the estimated values for the smoothing parameter and overdispersion parameter, respectively.

Method	λ^2	σ^2	Covariate	Estimate	SSE	p-value	MSE	p-value
$V_i^{(1)}$	270.935	N/A	Age	0.121	0.159	0.447	0.161	0.452
			Gender	0.303	0.188	0.106	0.190	0.110
			DFMO	-0.185	0.169	0.274	0.171	0.281
			Prior Count	0.079	0.008	< 0.001	0.008	< 0.001
$V_i^{(2)}$	178.593	N/A	Age	0.162	0.141	0.248	0.144	0.260
			Gender	0.228	0.160	0.155	0.164	0.164
			DFMO	-0.209	0.149	0.160	0.152	0.169
			Prior Count	0.077	0.007	< 0.001	0.007	< 0.001
$V_i^{(3)}$	187.214	0.519	Age	0.067	0.142	0.638	0.148	0.650
			Gender	0.238	0.156	0.128	0.161	0.140
			DFMO	-0.058	0.149	0.686	0.154	0.705
			Prior Count	0.109	0.009	< 0.001	0.009	< 0.001

effect of overdispersion is estimated as small as 0.519 and the correlation among observed panel counts seems to be not very strong. Further, it is worthwhile to mention that as discussed in Chiou et al. [4] and Li et al. [11], the recurrence rate was related to the number of follow-up visits, which violates the assumption of non-informative observation, suggesting that the proposed methodology could be misused for this data. Further, the estimated curves of the baseline cumulative mean function $\Lambda(\cdot)$, along with 95% pointwise confidence intervals with the three working covariance matrices are presented in Figure 3. The lower and upper limits of pointwise confidence intervals were constructed as 2.5th and 97.5th percentiles of 1,000 bootstrap samples.

7. Summary and future work

In this study, we consider a computationally efficient penalized approach for a semiparametric proportional mean model with panel count data using GEE methodology. The method provides great flexibility for model fitting by not specifying the stochastic model of the underlying counting process. The asymptotic properties of the penalized spline estimators including the uniform convergence and the optimal rate of convergence for the functional estimator and the asymptotic normality for the regression parameter estimators were rigorously proved by applying model empirical process theory. By addressing the underestimation of the variance-covariance matrix of the regression parameter estimators in GEE methodology, the proposed method yields less biased variance-covariance estimation and provides more reliable inference for the regression parameters compared to the standard sandwich approach. The simulation results indicate that the proposed methodology that accounts for the overdispersion enhances the accuracy of estimation and the reliability of inference, and still performs well even if the stochastic model of the underlying counting process is not correctly specified for practical sample sizes.

The proposed methodology assumes that the examination process is independent of the underlying counting process. This non-informative assumption is not always valid in practice. For instance, the frequency of the clinic visit is related to the tumor recurrence rate in the chemoprevention skin cancer trial; see more discussion in Chiou [4] and Li et al. [11]. Statistical analysis of panel count data with informative observation times has been discussed in the literature; see Huang et al. [8], Sun et al, [30], Li et al. [11], and Chiou et al. [4], among many others. The analysis of panel count data with dependent observation under GEE framework is challenging and has not been discussed in the literature yet. It will be our future work to tackle the challenge using penalized methodology. The other fundamental assumption in this study is that the regression coefficients are constant over time. This restrictive proportional mean assumption can be relaxed in various ways. Specifically, a time-varying coefficient model that allows the covariate effects to change over time offers great flexibility to capture the temporal dynamics of covariate effects. It will be our future work to consider the efficient estimation of time-varying coefficient models for panel count data using penalization methodology. In particular, the time-varying effects will be modeled via penalized splines. The proposed penalization methodology can be easily adapted to time-varying coefficient models. The asymptotic results such as uniform consistency and the optimal rate of convergence are still valid. The asymptotic normality for regression parameter estimators, especially the calculation of information bound, needs further investigation. Following the line of Gray [5], a hypothesis that the regression parameters are time-independent will be considered to test the proportional mean assumption.

Further, the proposed variance-covariance estimation is based on Morel estimator (Morel et al. [20]). Another future research direction is to develop a new estimator to accommodate the nature of panel count data in variance-covariance estimation by incorporating other estimators such as Pan estimator (Pan [22]) and investigate its numerical properties via extensive Monte Carlo studies for practical sample sizes. Finally, it is worthwhile to further inspect the impact of the selection of other working covariance matrices commonly used in correlated data analysis.

8. Technical details

8.1. Technical lemmas

In this section we provide some empirical process results for panel count data that are useful to establish the asymptotic properties of the penalized spline GEE estimators. Denote by $\mathscr{G}$ a class of K-dimensional vector functions g with jth element $g_j = Z^\intercal\beta + \varphi(T_{K,j})$, for $j \in \{1, \dots, K\}$, where $\beta \in \Phi$, $\varphi \in \mathcal{M}_n$, and $J(\varphi) < \infty$. Define $\mu = (\mu_{K,1}, \ldots, \mu_{K,K})^\intercal$ with $\mu_{K,j} = \exp(g_j)$. In view of Lemma A1 of Lu et al. [13], $\varphi_0(\cdot)$ can be well approximated by a monotone spline; i.e., there exists $\varphi_n \in \mathcal{M}_n$ such that $\|\varphi_n - \varphi_0\|_\infty = O(n^{-m/(1+2m)})$. Denote $\tau_n = (\beta_0, \varphi_n)$, and let g_0, g_n, and $\widehat{\mathrm{g}}$ denote g evaluated at $\tau = \tau_0$, τ_n, and $\widehat{\tau}$, respectively. Define the empirical inner product with respect to working matrices V as

$$\langle \xi_1, \xi_2 \rangle_{v,n} = \mathbb{P}_n(\xi_1^\intercal V^{-1} \xi_2),$$

for $\xi_1, \xi_2 \in \mathbb{R}^K$ and the corresponding empirical norm as $\|\xi\|_{v,n}^2 = \langle \xi, \xi \rangle_{v,n}$. In view of the definition of $\widehat{\tau}$, we have,

$$\widehat{\mathrm{g}} = \arg\min_{g \in \mathscr{G}} \left\{ \frac{1}{2} \|\mathbb{N} - \mu\|_{v,n}^2 + \frac{1}{2} \lambda^2 J^2(g) \right\}.$$

For $R > 0$, define $\mathscr{G}(R) = \{g \in \mathscr{G}, \|g - \mathrm{g}_0\|_{v,n} \le R\}$. This definition implies $\sup_{\mathrm{g} \in \mathscr{G}(R)} \|\mathrm{g}\|_{v,n} \le R$, which is required in Lemmas 1, and 2. For every probability measure Q and a class of measurable functions $\mathscr{L}$, let $H(\varepsilon, \mathscr{L}, L_2(Q))$ denote the entropy number of $\mathscr{L}$. In the sequel, it is assumed that the eigenvalues of positive definite matrices V_i are uniformly bounded away from 0 and infinite.

Lemma 1. *Suppose that $e_1, \ldots, e_n$ are independent random vectors with expectation 0 satisfying the uniformly sub-Guassian condition, i.e.,*

$$\max_{i=1,\ldots,n} M_0^2 \{E \exp(\|e_i\|^2 / M_0^2) - 1\} \le \sigma_0^2,$$

for some fixed positive constants M_0 and σ_0, and that

$$H(\varepsilon, \mathscr{L}, \|\cdot\|_{v,n}) \le A_0 \varepsilon^{-\alpha}, \tag{8.1}$$

for all $\varepsilon > 0$ and some constants $A_0 > 0$ and $0 < \alpha < 2$. Further, assume $\sup_{\xi \in \mathscr{L}} \|\xi\|_{v,n} \le R$. Then for some constant C depending on A_0, α, R, M_0 and σ_0, we have for all $T \ge C$,

$$P\left(\sup_{\xi \in \mathscr{L}} \frac{n^{1/2} |\langle e, \xi \rangle_{v,n}|}{\|\xi\|_{v,n}^{1-\frac{\alpha}{2}}} \ge T \right) \le C \exp(-T^2/C^2).$$

Lemma 2. *For every $\varepsilon > 0$,*

$$H\left(\varepsilon, \left\{ \frac{A(\mathrm{g})}{J(\varphi) + J(\varphi_n)} : \mathrm{g} \in \mathscr{G}(R) \right\}, \|\cdot\|_{v,n} \right) \le (1/\varepsilon)^{1/m},$$

up to a constant constant, where $A(\mathrm{g}) = \exp(\mathrm{g}) - \exp(\mathrm{g}_n)$.

Lemma 3. *For a sufficiently small $\varepsilon > 0$, $\|\varphi\|_\infty \le C\{1 + J(\varphi)\}$, for $C > 0$, whenever $J(\varphi) < \infty$ and $\|\varphi - \varphi_0\|_{L_2(\mu)} < \varepsilon$.*

Lemma 4. $\|\widehat{h}_l^* - h_l^*\|_{L_2(\mu)} = o_p(1)$, *for $l \in \{1, \ldots, d\}$.*

Remark 2. Lemma 1 that is the extension of Lemma 8.4 of van de Geer [32] for panel count data is applied to derive the modulus of continuity for the random process $n^{1/2}\langle e, \xi\rangle_{v,n}$. Lemma 2 establishes a useful result of the entropy number for a class of spline functions. It is related to Lemma 11.3 of van de Geer [32]. The above lemmas are key results to establish the consistency and rate of convergence of $\widehat{g}$. Lemma 3 is used to derive the uniform boundedness of $\widehat{\varphi}(\cdot)$ in probability once the consistency of $\widehat{\varphi}(\cdot)$ is established. Lemma 4 is employed to establish the consistency of the variance-covariance estimator of $\widehat{\beta}$.

8.2. Proof of Lemma 1

The proof is similar to that of Lemma 8.4 of van de Geer [32], and thus is omitted.

8.3. Proof of Lemma 2

Let $\zeta = Z^{\intercal}\beta + \varphi(t)$. Define $\zeta_n = Z^{\intercal}\beta + \varphi_n(t)$ and $\zeta_0 = Z^{\intercal}\beta_0 + \varphi_0(t)$. First we show that

$$H\left(\varepsilon, \left\{\frac{A(\zeta)}{J(\zeta)+J(\zeta_n)} : \beta \in \Phi, \varphi \in \mathcal{M}_n, J(\varphi) < \infty, \|\zeta - \zeta_0\|_n \le R\right\}, \|\cdot\|_n\right) \\ \le (1/\varepsilon)^{1/m}, \tag{8.2}$$

up to a constant, where $\|\cdot\|_n$ is the standard empirical norm. For any function $\varphi(\cdot)$ defined on $\mathcal{T}$ with $J(\varphi) < \infty$, by a Taylor expansion, $\varphi(t)$ can be written as $\varphi_1(t) + \varphi_2(t)$, where $\varphi_1(\cdot)$ is a polynomial of degree $m-1$ and $\sup_{t\in\mathcal{T}} |\varphi_2(t)| \le C_0 J(\varphi)$, for $C_0 > 0$, with $J(\varphi) = J(\varphi_2)$. Without loss of generality, assume $0 < C_0 < 1$, i.e., $\sup_{t\in\mathcal{T}} |\varphi_2(t)| \le J(\varphi)$. Let $\zeta_1 = Z^{\intercal}\beta + \varphi_1(t)$ and $\zeta_2 = \varphi_2(t)$. For a fixed bounded function h, by Example 3.7.4d of van de Geer [32], the class of uniformly bounded functions $\zeta_1 + h$ is a Vapnik-Chervonenkis subgraph class of index bounded by $d+m+2$. It concludes from Lemma 19.15 of van der Vaart [33] that the entropy number with $\|\cdot\|_n$ of the class of uniformly bounded functions $\zeta_1 + h$ is bounded by $\ln(1/\varepsilon)$, up to a constant. Since $A(\zeta) = \exp(\zeta) - \exp(\zeta_n)$ is Lipschitz with respect to ζ, we read

$$H(\varepsilon, \{A(\zeta_1 + h)\}, \|\cdot\|_n) \le C_1 \ln(1/\varepsilon), \tag{8.3}$$

for some constant $C_1 > 0$. For a small $\varepsilon > 0$, write $q(\zeta) = \lfloor 1/[\{J(\zeta) + J(\zeta_n)\}\varepsilon]\rfloor \varepsilon$, where $\lfloor\cdot\rfloor$ is the floor function. In view of the inequality $1/2 < a\lfloor 1/a\rfloor < 1$, for $0 < a < 1$, we read $J\{q(\zeta)\zeta_2\} < 1$ and $\|q(\zeta)\zeta_2\|_\infty < 1$. Theorem 2.4 of van de Geer [32] applies and yields

$$H(\varepsilon, \{q(\zeta)\zeta_2 : J(\zeta) < \infty\}, \|\cdot\|_\infty) \le C_2 (1/\varepsilon)^{1/m}, \tag{8.4}$$

for some constant $C_2 > 0$. Thus, in view of entropy results (8.3) and (8.4), there exists $g_l(\cdot)$ for $1 \le l \le \exp\{C_2(1/\varepsilon)^{1/m}\} \equiv M_2$ such that $\|q(\zeta)\zeta_2 - g_l\|_n \le$

$\|q(\zeta)\zeta_2 - g_l\|_\infty \le \varepsilon$ and $f_k(\cdot)$ for $1 \le k \le (1/\varepsilon)^{C_1} \equiv M_1$ such that

$$\|A(\zeta_1 + g_l/q(\zeta)) - A(f_k + g_l/q(\zeta))\|_n \le \varepsilon.$$

It follows that

$$\|A(\zeta_1 + \zeta_2)/\{J(\zeta) + J(\zeta_n)\} - A(f_k + g_l/q(\zeta))\, q(\zeta)\|_n \le C_3\varepsilon,$$

for some constant $C_3 > 0$. The entropy result (8.2) follows. Thus, for $\xi(t) = A\{\zeta(t)\}/\{J(\zeta) + J(\zeta_n))\}$, there exists $\rho_{kl}(t) = A\{f_k(t) + g_l(t)/q(\zeta)\}q(\zeta)$, for $k \in \{1, \ldots, M_1\}$ and $l \in \{1, \ldots, M_2\}$, such that

$$\sum_{j=1}^{K} \|\xi(T_{K,j}) - \rho_{kl}(T_{K,j})\|_n^2 \le C_4\varepsilon^2, \tag{8.5}$$

for some constant $C_4 > 0$. Let $\xi(T) = (\xi(T_{K,1}), \ldots, \xi(T_{K,K}))^\intercal$ and $\rho_{kl}(T) = (\rho_{kl}(T_{K,1}), \ldots, \rho_{kl}(T_{K,K}))^\intercal$. It follows from the uniformly bounded assumption of the eigen values of V and (8.5) that

$$\mathbb{P}_n \{\xi(T_K) - \rho_{kl}(T_K)\}^\intercal V^{-1} \{\xi(T_K - \rho_{kl}(T_K)\} \le C_5\varepsilon^2,$$

for some constants $C_5 > 0$. Hence, the entropy result (8.2) also holds for norm $\|\cdot\|_{v,n}$. Lemma 2 follows.

8.4. Proof of Lemma 3

As discussed in the proof of Lemma 2, every $\varphi(\cdot)$ with $J(\varphi) < \infty$ can be written as $\varphi_1(\cdot)+\varphi_2(\cdot)$, where $\varphi_1(\cdot)$ is a polynomial of degree $m-1$ and $\sup_{t\in\mathcal{T}} |\varphi_2(t)| \le C_0 J(\varphi)$, for $0 < C_0 < 1$. Write $\varphi_1(t) = a^\intercal t$ for $t = (1, t, \ldots, t^{m-1})^\intercal$ and $a = (a_0, a_1 \ldots, a_{m-1})^\intercal$ with $a_l = \varphi^{(l)}(t_1)/l!$, for $t_1 \in \mathcal{T}$ and $l \in \{0, \ldots, m-1\}$. By convention, $\varphi^{(0)}(\cdot) = \varphi(\cdot)$. Under the bounded assumption of K and $\varphi_0(\cdot)$, we have $\|\varphi_0\|_{L_2(\mu)} \le M$ for $M > 0$, and hence $\|\varphi\|_{L_2(\mu)} \le \|\varphi_0\|_{L_2(\mu)} + \|\varphi - \varphi_0\|_{L_2(\mu)} \le M + \varepsilon$. Thus,

$$\frac{\|\varphi_1\|_{L_2(\mu)}}{1 + J(\varphi)} \le \frac{\|\varphi\|_{L_2(\mu)}}{1 + J(\varphi)} + \frac{\|\varphi_2\|_{L_2(\mu)}}{1 + J(\varphi)} \le M + \varepsilon + 1.$$

The non-singularity of matrix $E\left(\sum_{j=1}^K \eta_{K,j}\eta_{K,j}^\intercal\right)$ for $\eta_{K,j}=(1, T_{K,j}, \ldots, T_{K,j}^{m-1})^\intercal$ implies that $E\left(\sum_{j=1}^K \eta_{K,j}\eta_{K,j}^\intercal\right)$ is positive definite, and hence $\|a\|/\{1+J(\varphi)\} = O(1)$. Thus, $\|\varphi_1\|_\infty/\{1 + J(\varphi)\} = O(1)$ since $\varphi_1(\cdot)$ is defined on a bounded set $\mathcal{T}$. Lemma 3 follows from the inequality $\|\varphi\|_\infty \le \|\varphi_1\|_\infty + \|\varphi_2\|_\infty$.

8.5. Proof of Lemma 4

Because $h_l^*(\cdot)$ is bounded and smooth, according to the standard spline approximation result, there exists $h_{nl}^* \in \mathcal{S}_n$ such that $\|h_l^* - h_{nl}^*\|_\infty = O(n^{-m/(1+2m)})$

and $J(h^*_{nl}) \leq M$ for $M > 0$, where $\mathcal{S}_n$ is a class of B-splines of degree $m+1$ defined on $\mathcal{T}$. For a positive definite matrix V^{-1}, there exists an orthogonal matrix $U = (u_{ij})_{K\times K}$ such that $U^\intercal V^{-1}U$ =diag $\{\lambda_1, \dots, \lambda_K\}$, for $0 < \lambda_1 \leq \dots \leq \lambda_K$. Define

$$\rho_l(\tau;\mu,h) = \{Z_l - h(T)\}^\intercal \, \Delta V^{-1} \Delta \, \{Z_l - h(T)\} \, .$$

It is easy to verify that

$$\rho_l(\tau;\mu,h) = \sum_{m=1}^{K} \lambda_m \left[\sum_{j=1}^{K} u_{mj}\mu_j \left\{ z_l - h(T_{K,j}) \right\} \right]^2 .$$

Let $\mu = \exp\{Z^\intercal\beta + \varphi(t)\}$. It concludes from Example 19.7 of van der Vaart [33] and Theorem 2.4 of van de Geer [32] together with the Donsker preservation theorem (e.g., Corollary 9.32 of Kosorok [10]) that

$$H(\varepsilon, \{\mu : \|\tau - \tau_0\|_{L_2(\mu)} \leq M, \|\varphi\|_\infty \leq M, J(\varphi) \leq M\}, \|\cdot\|_\infty) \leq C(1/\varepsilon)^{1/m},$$

for some constant $C > 0$. That is, for any $\varepsilon > 0$, there exits $\mu^{(k)} = \exp\{Z^\intercal\beta^{(k)} + \varphi^{(k)}(t)\}$, for $1 \leq k \leq \exp\{C(1/\varepsilon)^{1/m}\}$, such that $\|\mu^{(k)} - \mu\|_\infty \leq \varepsilon$. Obviously, by the triangle inequality, $\|\mu^{(k)}\|_\infty$ is bounded. Construct a class of functions $\rho_l(\tau;\mu^{(k)},h)$, where $\mu^{(k)}$ is a K-vector with jth element $\mu_j^{(k)} = \exp\{Z^\intercal\beta^{(k)} + \varphi^{(k)}(T_{K,j})\}$, for $j \in \{1,\dots,K\}$. Thus, under conditions C3, C4, and C5, for any $h \in \mathcal{S}_n$, we have $\|\rho_l(\tau;\mu,h) - \rho_l(\tau;\mu^{(k)},h)\|_2 \leq \varepsilon$, up to a constant. Thus, the class of functions $\rho_l(\tau;\mu,h)$ with $\|\tau-\tau_0\|_{L_2(\mu)} \leq M$, $J(\varphi) \leq M$, and $\|\varphi\|_\infty \leq M$ is a Donsker class. It follows from the consistency of $\widehat{\tau}$ and Glivenko-Cantelli theorem that

$$(\mathbb{P}_n - P)\rho_l(\widehat{\tau};\widehat{\mu},h_l^*) = o_p(1) \quad \text{and} \quad (\mathbb{P}_n - P)\rho_l(\widehat{\tau};\widehat{\mu},h) = o_p(1), \tag{8.6}$$

for $h \in \mathcal{S}_n$. Further, under conditions C3, C4, and C5, the continuous mapping theorem and the dominated convergence theorem together with the consistency of $\widehat{\tau}$ yield

$$P\{\rho_l(\widehat{\tau};\widehat{\mu},h_l^*) - \rho_l(\tau_0;\mu_0,h_l^*)\} = o_p(1) \text{ and } P\{\rho_l(\widehat{\tau};\widehat{\mu},h) - \rho_l(\tau_0;\mu_0,h)\} = o_p(1). \tag{8.7}$$

Combing (8.6) and (8.7) yields

$$\mathbb{P}_n \left\{\rho_l(\widehat{\tau};\widehat{\mu},h) - \rho_l(\widehat{\tau};\widehat{\mu},h_l^*)\right\} = P\left\{\rho_l(\tau_0;\mu_0,h) - \rho_l(\tau_0;\mu_0,h_l^*)\right\} + o_p(1). \tag{8.8}$$

For any $\varepsilon > 0$ and $h \in \mathcal{S}_n$ with $\|h - h_l^*\|_\infty \geq \varepsilon$, it concludes from the uniqueness of $h_l^*(\cdot)$ that $P\left\{\rho_l(\tau_0;\mu_0,h) - \rho_l(\tau_0;\mu_0,h_l^*)\right\} > 0$, and hence (8.8) yields

$$\mathbb{P}_n \left\{\rho_l(\widehat{\tau};\widehat{\mu},h) - \rho_l(\widehat{\tau};\widehat{\mu},h_l^*)\right\} > 0 \tag{8.9}$$

in probability. In view of Theorem 2.4 of van de Geer [32] and Donsker preservation theorem, the class of functions $\rho_l(\tau;\mu,h) - \rho_l(\tau;\mu,h_l^*)$ with $\|\tau - \tau_0\|_{L_2(\mu)} \leq$

M, $\|h - h_l^*\|_\infty \le M$, $\|h\|_\infty \le M$, $J(h) \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ is a Donsker class. Further, by the spline approximation $\|h_l^* - h_{nl}^*\|_\infty = o(1)$ and $\|h_l^*\|_\infty \le M$, we have $\|h_{nl}^*\|_\infty \le M$ and $J(h_{nl}^*) \le M$. Thus, Theorem 2 and Glivenko-Cantelli theorem apply and yield

$$(\mathbb{P}_n - P)\left\{\rho_l(\widehat{\tau};\widehat{\mu},h_{nl}^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_l^*)\right\} = o_p(1). \tag{8.10}$$

Further, applying the continuous mapping theorem and dominated convergence theorem together with the consistency of $\widehat{\tau}$ yields $P\left\{\rho_l(\widehat{\tau};\widehat{\mu},h_{nl}^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_l^*)\right\} = o_p(1)$, and it follows from (8.10) that $\mathbb{P}_n\left\{\rho_l(\widehat{\tau};\widehat{\mu},h_{n,l}^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_l^*)\right\} = o_p(1)$. Thus,

$$\mathbb{P}_n\left\{\rho_l(\widehat{\tau};\widehat{\mu},\widehat{h}_l^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_l^*)\right\} = \mathbb{P}_n\left\{\rho_l(\widehat{\tau};\widehat{\mu},\widehat{h}_l^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_{nl}^*)\right\} + o_p(1). \tag{8.11}$$

In view of (8.9), $\|\widehat{h}_l^* - h_l^*\|_\infty \ge \varepsilon$ implies $\mathbb{P}_n\left\{\rho_l(\widehat{\tau};\widehat{\mu},\widehat{h}_l^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_{nl}^*)\right\} \ge 0$ in probability. Thus, it concludes from (8.11) and the definition of $\widehat{h}_l^*(\cdot)$ that

$$P(\|\widehat{h}_l^* - h_l^*\|_\infty \ge \varepsilon) \le P\left[\mathbb{P}_n\left\{\rho_l(\widehat{\tau};\widehat{\mu},\widehat{h}_l^*) - \rho_l(\widehat{\tau};\widehat{\mu},h_{nl}^*)\right\} \ge 0\right] \to 0,$$

as $n \to \infty$. The consistency of $\widehat{h}_l^*(\cdot)$ follows.

8.6. Proof of Theorem 1

The proof of Theorem 1 follows the similar arguments to those in the proof of Lemma A4 of Huang et al. [9], and thus is omitted.

8.7. Proof of Theorem 2

Define the error term $e = \mathbb{N} - \mu_0$. According to the definition of $\widehat{\tau}$,

$$\|\mathbb{N} - \widehat{\mu}\|_{v,n}^2 + \lambda^2 J^2(\widehat{g}) \le \|\mathbb{N} - \mu_n\|_{v,n}^2 + \lambda^2 J^2(g_n).$$

Write $\mathbb{N} - \widehat{\mu} = \mathbb{N} - \mu_n - (\widehat{\mu} - \mu_n)$ and the above inequality reduces to

$$\|\widehat{\mu} - \mu_n\|_{v,n}^2 + \lambda^2 J^2(\widehat{g}) \le 2\langle \mathbb{N} - \mu_n, \widehat{\mu} - \mu_n\rangle_{v,n} + \lambda^2 J^2(g_n). \tag{8.12}$$

In view of the uniformly bounded assumptions of the eigen values of V and the total number of examinations K, the spline approximation applies and yields $\|\mu_n - \mu_0\|_{v,n} = o(1)$, and it concludes from Cauchy-Schwarz inequality that

$$\langle \mu_n - \mu_0, \widehat{\mu} - \mu_n\rangle_{v,n} = o_p(1)\|\widehat{\mu} - \mu_n\|_{v,n}. \tag{8.13}$$

Combing (8.12) and (8.13), we have the following basic inequality

$$\{1 + o_p(1)\}\|\widehat{\mu} - \mu_n\|_{v,n}^2 + \lambda^2 J^2(\widehat{g}) \le 2\langle e, \widehat{\mu} - \mu_n\rangle_{v,n} + \lambda^2 J^2(g_n). \tag{8.14}$$

It is observed that $e_1, \dots, e_n$ satisfy the sub-Gaussian condition under condition C5. Further, by Lemma 2, the entropy result (8.1) in Lemma 1 holds with $\alpha = 1/m$. Thus, Lemma 1 applies and yields

$$\frac{|\langle e, \widehat{\mu} - \mu_n \rangle_{v,n}|}{\|\widehat{\mu} - \mu_n\|_{v,n}^{1-\frac{1}{2m}} \{J(\widehat{g}) + J(g_n)\}^{\frac{1}{2m}}} = O_p(n^{-1/2}). \tag{8.15}$$

If $J(\widehat{g}) \geq J(g_n)$ in probability, then in view of the basic inequality (8.14) along with (8.15), we have

$$\{1 + o_p(1)\} \|\widehat{\mu} - \mu_n\|_{v,n}^2 + \lambda^2 J^2(\widehat{g}) \leq O_p(n^{-1/2}) \|\widehat{\mu} - \mu_n\|_{v,n}^{1-\frac{1}{2m}} J^{\frac{1}{2m}}(\widehat{g}) + \lambda^2 J^2(g_n).$$

Under the assumption of the order of λ, applying the similar arguments to those in the proof of Theorem 10.2 of [32] yields $J(\widehat{g}) = O_P(1)$ and $\|\widehat{\mu} - \mu_n\|_{v,n} = O_p(n^{-\frac{m}{1+2m}})$. Similarly, in the case of $J(\widehat{g}) < J(g_n)$ in probability (i.e., $J(\widehat{g}) = O_p(1)$), it follows from (8.14) and (8.15) that

$$\|\widehat{\mu} - \mu_n\|_{v,n}^2 \leq O_p(n^{-1/2}) \|\widehat{\mu} - \mu_n\|_{v,n}^{1-\frac{1}{2m}} + O_p(\lambda^2).$$

Hence, $\|\widehat{\mu} - \mu_n\|_{v,n} = O_p(n^{-\frac{m}{1+2m}})$, and it results from the assumption of the order of λ and the spline approximation $\|\varphi_n - \varphi_0\|_\infty = O_p(n^{-\frac{2m}{1+2m}})$ that

$$\|\widehat{\mu} - \mu_0\|_{v,n}^2 \leq \|\widehat{\mu} - \mu_n\|_{v,n}^2 + \|\mu_n - \mu_0\|_{v,n}^2 = O_p(n^{-\frac{2m}{1+2m}}).$$

Under conditions C3 and C6, a Taylor expansion yields

$$\mathbb{P}_n \sum_{j=1}^{K} \left\{ Z^\intercal(\widehat{\beta} - \beta_0) + (\widehat{\varphi} - \varphi_0)(T_{K,j}) \right\}^2 = O_p(n^{-\frac{2m}{2m+1}}).$$

According to the bracketing entropy calculations for the class of monotone functions and the parametric class with bounded index set, we can verify that the bracketing entropy of the class $\{\zeta - \zeta_0 : \|\zeta - \zeta_0\|_{v,n} \leq M, \beta \in \Phi, \varphi \in \mathcal{M}_n, J(\varphi) < \infty\}$ is bounded by $1/\varepsilon$, up to a constant, for every $\varepsilon > 0$ and every probability measure; see Examples of 19.7 and 19.11 of van der Vaart [33]. Thus, Theorem 2.3 of Mammen and van de Geer [18] applies and yields

$$E \sum_{j=1}^{K} \left\{ Z^\intercal(\widehat{\beta} - \beta_0) + (\widehat{\varphi} - \varphi_0)(T_{K,j}) \right\}^2 = O_p(n^{-\frac{2m}{1+2m}}).$$

According to Lemma 3.1 of Stone [27], $\|\widehat{\varphi} - \varphi_0\|_{L_2(\mu)} = O_p(n^{-m/(1+2m)})$. Further, under the assumptions that $E(ZZ^\intercal)$ and K are uniformly bounded, $\|\widehat{\beta} - \beta_0\| = O_p(n^{-m/(1+2m)})$. According to Lemma 3, we can establish that $\|\widehat{\varphi}\|_\infty = O_p(J(\widehat{\varphi}) + 1)$, and hence $\|\widehat{\varphi}\|_\infty = O_p(1)$ on account of the fact that $J(\widehat{\varphi}) = O_p(1)$. In view of Cauchy-Schwarz inequality and $J(\widehat{\varphi}) = O_p(1)$, for any $t \in \mathcal{T} = [d_1, d_2]$, $|\widehat{\varphi}^{(m-1)}(t) - \widehat{\varphi}^{(m-1)}(d_1)| = \left| \int_{d_1}^t \widehat{\varphi}^{(m)}(u) du \right| \leq J(\widehat{\varphi}) = O_p(1)$, and hence $\widehat{\varphi}^{(m-1)}(\cdot)$ is uniformly bounded in probability. Integrating $\widehat{\varphi}^{(m-1)}(\cdot)$ $(m-2)$ times yields the first derivative of $\widehat{\varphi}(\cdot)$ is uniformly bounded in probability. Thus, $\widehat{\varphi}(\cdot)$ is uniformly equicontinuous in probability on the compact set $\mathcal{T}$, which implies the uniform convergence of $\widehat{\varphi}(\cdot)$ by Arzel-Ascoli theorem.

8.8. Proof of Theorem 3

Inserting $(\widehat{\beta}+\xi, \widehat{\varphi}(T)-h^*(T)\xi)$ with a vector of functions $h^*(\cdot)=(h_1^*(\cdot), \dots, h_d^*(\cdot))$ satisfying $J(h_l^*) < \infty$, for $l \in \{1, \dots, d\}$, into the penalized weighted least squares function (2.2), differentiating it with respect to ξ, and evaluating it at $\xi = 0$, we obtain the following stationary equations

$$0 = \mathbb{P}_n Q^\intercal \widehat{\Delta} V^{-1}(\mathbb{N} - \widehat{\mu}) + \lambda^2 \int \begin{pmatrix} \widehat{\eta}^{(m)}(u) h_1^{*(m)}(u) \\ \vdots \\ \widehat{\eta}^{(m)}(u) h_d^{*(m)}(u) \end{pmatrix} du, \tag{8.16}$$

where $Q = \overline{Z} - h^*(T)$ and $\widehat{\Delta} = \text{diag}\,\{\widehat{\mu}_{K,1}, \dots, \widehat{\mu}_{K,K}\}$. Rewrite the stationary equation (8.16) as

$$\begin{aligned} 0 &= \mathbb{P}_n Q^\intercal (\widehat{\Delta} - \Delta_0) V^{-1}(\mathbb{N} - \mu_0) + \mathbb{P}_n Q^\intercal \Delta_0 V^{-1}(\mathbb{N} - \mu_0) \\ &- \mathbb{P}_n Q^\intercal (\widehat{\Delta} - \Delta_0) V^{-1}(\widehat{\mu} - \mu_0) - \mathbb{P}_n Q^\intercal \Delta_0 V^{-1}(\widehat{\mu} - \mu_0) \\ &+ \lambda^2 \int \begin{pmatrix} \widehat{\eta}^{(m)}(u) h_1^{*(m)}(u) \\ \vdots \\ \widehat{\eta}^{(m)}(u) h_d^{*(m)}(u) \end{pmatrix} du \\ &\equiv I_1 + I_2 - I_3 - I_4 + I_5. \end{aligned}$$

To derive the asymptotic normality, we need to show that $\|I_1\| = \|I_3\| = \|I_5\| = o_p(n^{-1/2})$. Let $\widehat{g}_j$ and g_{0j} be g_j evaluated at $\widehat{\tau}$ and τ_0, respectively, for $j \in \{1, \dots, K\}$. Denote by $\mathsf{h}_l(X, h^*; g)$ a K-vector of functions for $g \in \mathscr{G}(R)$ with jth element $\{z_l - h_l^*(T_{K,j})\}\left\{\exp(g_j) - \exp(g_{0j})\right\}$. Clearly, the lth element of I_1 can be written as $\mathbb{P}_n \mathsf{h}_l^\intercal(X, h^*; \widehat{g}) V^{-1} e$, for $l \in \{1, \dots, d\}$. For any $g \in \mathscr{G}(R)$, it follows from the uniformly bounded assumptions of Z, $h^*(\cdot)$, and the eigen values of V that $\|\mathsf{h}_l(X, h^*; g)\|_{v,n} \le \|g - g_0\|_{v,n}$, up to a constant, and hence Theorem 2 applies and yields

$$\|\mathsf{h}_l(X, h^*; \widehat{g})\|_{v,n} \le O_p(1)\|\widehat{g} - g_0\|_{v,n} = O_p(n^{-m/(1+2m)}) \equiv \rho_n.$$

Similarly, it can be shown that $\|\mathsf{h}_l(X, h^*; \widetilde{g}) - \mathsf{h}_l(X, h^*; g)\|_{v,n} \le \|\widetilde{g} - g\|_{v,n}$, up to a constant, for g, $\widetilde{g} \in \mathscr{G}(R)$. In view of Examples 19.7 and 19.10 of van der Vaart [33], the bracketing entropy of the class of functions $Z^\intercal\beta + \varphi(t)$ for $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ with $\|\cdot\|_\infty$ is bounded by $(1/\varepsilon)^{1/m}$, up to a constant, for any $\varepsilon > 0$. In view of the uniformly bounded assumptions of the total number of examinations K and the eigen values of V, the entropy result also holds for the class of functions $g \in \mathscr{G}(R)$ with $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ with $\|\cdot\|_{v,n}$. Further, since $\mathsf{h}_l(X, h^*; g)$ is Lipschitz with respect to g, it follows that

$$\begin{aligned} &H(\varepsilon, \left\{\mathsf{h}_l : \|\beta - \beta_0\| \le M, \|\varphi - \varphi_0\|_{L_2(\mu)} \le M, \|\varphi\|_\infty \le M, J(\varphi) \le M\right\}, \|\cdot\|_{v,n}) \\ &\le (1/\varepsilon)^{1/m}, \end{aligned}$$

up to a constant. Therefore, it concludes from Theorem 2 and Lemma 1 that

$$|\mathbb{P}_n \mathsf{h}_l^\intercal(X, h^*; \widehat{g}) V^{-1} e| = \|\mathsf{h}_l(X, h^*; \widehat{g})\|_{v,n}^{1-1/2m} O_p(n^{-1/2}),$$

which implies

$$|\mathbb{P}_n \mathsf{h}_l^\intercal(X, h^*; \widehat{g}) V^{-1} e| = \rho_n^{1-1/2m} O_p(n^{-1/2}) = O_p(n^{-2m/(1+2m)}),$$

and hence $\|I_1\| = o_p(n^{-1/2})$. Obviously, the lth element of I_3 is $\mathbb{P}_n \mathsf{h}_l^\intercal(U, h^*; \widehat{g}) V^{-1}$ $(\widehat{\mu} - \mu_0)$, for $l \in \{1, \ldots, d\}$. Since Z, $h^*(\cdot)$, and the eigen values of V are uniformly bounded, mean value theorem and Cauchy-Schwarz inequality apply and yield that the lth element of I_3 can be bounded by $\|\widehat{g} - g_0\|_{v,n}^2$, up to a constant. It concludes from Theorem 2 that $\|I_3\| = O_p(n^{-2m/(1+2m)}) = o_p(n^{-1/2})$. Further, under conditions C3 and C5, applying a Taylor expansion and the rate of convergence of $\widehat{g}$, as well as the uniform boundedness of $\widehat{g}$ and $h^*(\cdot)$ yields

$$I_4 = \mathbb{P}_n Q^\intercal \Delta_0 V^{-1} \Delta_0 (\widehat{g} - g_0) + o_p(n^{-1/2}).$$

Since Z and $h^*(\cdot)$ along with the eigen values of V are uniformly bounded, the jth element of $\mathbb{P}_n Q^\intercal \Delta_0 V^{-1} \Delta_0 (\widehat{g} - g_0)$ can be written as $\mathbb{P}_n \sum_{j=1}^K a_j (\widehat{g}_j - g_{0j})$ with $|a_j| \le M$ for a constant $M > 0$. Under the uniformly bounded assumption of Z and K, it concludes from Examples 19.7 and 19.10 of van der Vaart [33] that the bracketing entropy of the class of functions $\sum_{j=1}^K a_j (g_j - g_{0j})$ for $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ with $\|\cdot\|_\infty$ is bounded by $(1/\varepsilon)^{1/m}$, up to a constant, for every $\varepsilon > 0$, and hence the above class is a Donsker class. Further, in view of the rate of convergence of $\widehat{g}$ and Cauchy-Schwarz inequality, $E\left\{\sum_{j=1}^K a_j (\widehat{g}_j - g_{0j})\right\}^2 = O_p(n^{-2m/(1+2m)})$. In view of Theorem 2, $\|\widehat{\beta} - \beta_0\| = o_p(1)$, $\|\widehat{\varphi} - \varphi_0\|_{L_2(\mu)} = o_p(1)$, $\|\widehat{\varphi}\|_\infty = O_p(1)$, and $J(\widehat{\varphi}) = O_p(1)$, and hence Lemma 19.24 of van der Vaart [33] applied and yields

$$(\mathbb{P}_n - P) Q^\intercal \Delta_0 V^{-1} \Delta_0 (\widehat{g} - g_0) = o_p(n^{-1/2}).$$

Thus,

$$I_4 = E\{Q^\intercal \Delta_0 V^{-1} \Delta_0 \overline{Z} (\widehat{\beta} - \beta_0)\} + E\{Q^\intercal \Delta_0 V^{-1} \Delta_0 (\widehat{\varphi} - \varphi_0)(T)\} + o_p(n^{-1/2}).$$

According to the characteristic of $h^*(\cdot)$, $E\{Q^\intercal \Delta_0 V^{-1} \Delta_0 (\widehat{\varphi} - \varphi_0)(T)\} = 0$ and $E\{Q^\intercal \Delta_0 V^{-1} \Delta_0 h_l^*(T)\} = 0$, for $l \in \{1, \ldots, d\}$, and thus, $I_4 = \mathscr{I}_1 (\widehat{\beta} - \beta_0) + o_p(n^{-1/2})$. In view of Cauchy-Schwarz inequality and the assumption of the order of λ, the lth element of I_5 is bounded by $\lambda^2 J(\widehat{\varphi}) J(h_l^*) = O_p(n^{-2m/(1+2m)}) O_p(1) = o_p(n^{-1/2})$, and hence $\|I_5\| = o_p(n^{-1/2})$. It concludes that

$$n^{1/2}(\widehat{\beta} - \beta_0) = \mathscr{I}_1^{-1} n^{1/2} \mathbb{P}_n Q^\intercal \Delta_0 V^{-1} (\mathbb{N} - \mu_0) + o_p(1) \to_d N(0, \mathscr{I}_1^{-1} \mathscr{I}_2 \mathscr{I}_1^{-1}),$$

as $n \to \infty$. We conclude the asymptotic normality of $\widehat{\beta}$.

8.9. Proof of Theorem 4

We first show that $\widehat{\mathscr{I}_{n1}} = \mathscr{I}_1 + o_p(1)$. For a positive definite matrix V^{-1}, there exists an orthogonal matrix $C = (c_{ij})_{K\times K}$ such that $C^\intercal V^{-1}C$ =diag $\{\lambda_1, \dots, \lambda_K\}$ for $0 < \lambda_1 \le \dots \le \lambda_K$. Define

$$\mathscr{I}_1^{(jk)} \equiv \sum_{v=1}^{K} \lambda_v \sum_{p,q=1}^{K} c_{vp}c_{vq}\{z_j - h_j(T_{K,p})\}\{z_k - h_k(T_{K,q})\}\mu_p\mu_q,$$

and let $\widehat{\mathscr{I}_1}^{(jk)}$ denote $\mathscr{I}_1^{(jk)}$ evaluated at $\tau = \widehat{\tau}$ and $h_l = \widehat{h}_l^*$, for $l = j, k$. Clearly, the (j, k)th element of $\widehat{\mathscr{I}_{n1}}$ can be written as $\mathbb{P}_n\widehat{\mathscr{I}_1}^{(jk)}$, for $j, k \in \{1, \dots, K\}$. As shown in the proof of Theorem 3, the class of uniformly bounded functions $\exp\{Z^\intercal\beta + \varphi(t)\}$ for $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ with $\|\cdot\|_\infty$ is a Donsker class. Further, in view of the bracketing entropy calculation of spline (e.g., Shen and Wong ([26], p. 597)), the class of uniformly bounded functions $z_l - h_l$ for $h_l \in \mathcal{S}_n$ and $\|h_l - h_l^*\|_\infty \le M$ is also a Donsker class. Thus, under conditions C3, C4, and C5, it concludes from the Donsker preservation theorem that the class of uniformly bounded functions $\mathscr{I}_1^{(jk)}$ for $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, $J(\varphi) \le M$, $h_l \in \mathcal{S}_n$, and $\|h_l - h_l^*\|_\infty \le M$, for $l = j, k$, with $\|\cdot\|_\infty$ is a Donsker class, and hence a Glivenko-Cantelli class. Thus, it results from Glivenko-Cantelli Theorem along with Theorem 2 and Lemma 4 that $(\mathbb{P}_n - P)\widehat{\mathscr{I}_1}^{(jk)} = o_p(1)$. Further, combining the continuous mapping Theorem and dominated convergence Theorem with the consistency of $\widehat{\tau}$ and $\widehat{h}_l^*$ yields $E\widehat{\mathscr{I}_1}^{(jk)} = E\mathscr{I}_{1,0}^{(jk)} + o_p(1)$, where $\mathscr{I}_{10}^{(jk)}$ denotes $\mathscr{I}_1^{(jk)}$ evaluated at $\tau = \tau_0$ and $h_l = h_l^*$, for $l = j, k$. This completes the proof of the consistency of $\widehat{\mathscr{I}_{n1}}$.

Next we shall show that $\widehat{\mathscr{I}_{n2}} = \mathscr{I}_2 + o_p(1)$. Write $\{Z_l - \widehat{h}_l^*(T)\}^\intercal\widehat{\Delta}V^{-1}$ as

$$\begin{aligned}
&\{Z_l - h_l^*(T)\}^\intercal\Delta_0V^{-1} + \{(h_l^* - \widehat{h}_l^*)(T)\}^\intercal\Delta_0V^{-1} \\
&+\{Z_l - h_l^*(T)\}^\intercal(\widehat{\Delta} - \Delta_0)V^{-1} + \{(h_l^* - \widehat{h}_l^*)(T)\}^\intercal(\widehat{\Delta} - \Delta_0)V^{-1} \\
&\equiv \xi_{l1} + \xi_{l2} + \xi_{l3} + \xi_{l4}, \quad l = j, k.
\end{aligned}$$

Then the (j, k)th element of $\mathscr{I}_{n2}$ can be written as

$$\mathbb{P}_n \sum_{r,s=1}^{4} \xi_{jr}\,(\mathbb{N} - \widehat{\mu})^{\otimes 2}\,\xi_{ks}^\intercal \equiv \mathbb{P}_n \sum_{r,s=1}^{4} \delta_{rs}^{(jk)}.$$

Let $\mathscr{I}_2^{(jk)}$ denote (j, k)th element of $\mathscr{I}_2$. Observe that $\mathscr{I}_2^{(jk)} = E\delta_{11}^{(jk)}$. In the following we shall show that $\mathbb{P}_n\delta_{11}^{(jk)} = \mathscr{I}_2^{(jk)} + o_p(1)$ and $\mathbb{P}_n\delta_{rs}^{(jk)} = o_p(1)$, for $r \ne 1$ or $s \ne 1$. Define

$$\begin{aligned}
\omega_{pq}^{(jk)} = \lambda_p\lambda_q \sum_{a_1,a_2,b_1,b_2=1}^{K} &C_{a_1a_2b_1b_2}^{(pq)} \left\{z_j - h_j^*(T_{K,a_1})\right\}\left\{\mathbb{N}(T_{K,b_1}) - \mu_{b_1}\right\} \\
&\left\{z_k - h_k^*(T_{K,a_2})\right\}\left\{\mathbb{N}(T_{K,b_2}) - \mu_{b_2}\right\},
\end{aligned}$$

where $\mu_b = \exp\{Z^\intercal\beta + \varphi(T_{K,b})\}$, $\mu_{0a} = \exp\{Z^\intercal\beta_0 + \varphi_0(T_{K,a})\}$, and $C^{(pq)}_{a_1a_2b_1b_2} = c_{pa_1}c_{pb_1}c_{qa_2}c_{qb_2}\mu_{0a_1}\mu_{0a_2}$, for $a = a_1, a_2$ and $b = b_1, b_2$. Let $\widehat{\omega}^{(jk)}_{pq}$ denote $\omega^{(jk)}_{pq}$ evaluated at $\tau = \widehat{\tau}$. Observe that $\sum_{p,q=1}^K \widehat{\omega}^{(jk)}_{pq} = \xi_{j1}\,(\mathbb{N} - \widehat{\mu})^{\otimes 2}\,\xi_{k1}^\intercal$. As discussed above, for any $\varepsilon > 0$, the bracketing entropy of the class of uniformly bounded functions $\mathsf{m}(t) = \exp\{Z^\intercal\beta + \varphi(t)\}$ for $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ with $\|\cdot\|_\infty$ is bounded by $C_1(1/\varepsilon)^{1/m}$, i.e., there exists $\mathsf{m}^{(r)}(t) = \exp\{Z^\intercal\beta^{(r)} + \varphi^{(r)}(t)\}$ for $1 \le r \le \exp\{C_1(1/\varepsilon)^{1/m}$ such that $\|\mathsf{m}^{(r)} - \mathsf{m}\|_\infty \le C_2\varepsilon$, for some constants $C_1, C_2 > 0$. Let $\mu^{(r)}_{K,j}$ and $\omega^{(jk,r)}_{pq}$ denote $\mu_{K,j}$ and $\omega^{(jk)}_{pq}$ evaluated at $\beta = \beta^{(r)}$ and $\varphi(\cdot) = \varphi^{(r)}(\cdot)$, respectively. Combing the sub-Gaussian condition of $\mathbb{N}(\cdot)$ with the uniformly bounded conditions of K, Z, and the eigen values of V along with $\|h^*_l\|_\infty \le M$ for $l = j, k$ yields

$$
\begin{aligned}
E\left\{\omega^{(jk,r)}_{pq} - \omega^{(jk)}_{pq}\right\}^2 &\le C_3 E \sum_{b_1,b_2=1}^{K} \left[\{\mu^{(r)}_{K,b_1} - \mu_{K,b_1}\}^2 + \{\mu^{(r)}_{K,b_2} - \mu_{K,b_2}\}^2\right] \\
&\le C_4\varepsilon^2,
\end{aligned}
$$

for some constants $C_3, C_4 > 0$, which implies that the class of uniformly bounded functions $\sum_{p,q=1}^K \omega^{(jk)}_{pq}$ for $\|\beta - \beta_0\| \le M$, $\|\varphi - \varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, and $J(\varphi) \le M$ is a Donsker class, and hence is a Glivenko-Cantelli class. Therefore, it concludes from the Glivenko-Cantelli theorem and Theorem 2 that $(\mathbb{P}_n - P)\sum_{p,q=1}^K \widehat{\omega}^{(jk)}_{pq} = o_p(1)$. Further, applying the continuous mapping theorem and dominated convergence theorem together with the sub-Gaussian condition of $\mathbb{N}(\cdot)$ and the consistency of $\widehat{\tau}$ yields that $E\sum_{p,q=1}^K \widehat{\omega}^{(jk)}_{pq} = \mathscr{I}^{(jk)}_2 + o_p(1)$. Thus,

$$
\mathbb{P}_n\delta^{(jk)}_{11} = \mathbb{P}_n \sum_{p,q=1}^{K} \widehat{\omega}^{(jk)}_{pq} = \mathscr{I}^{(jk)}_2 + o_p(1).
$$

Next, let

$$
\begin{aligned}
\zeta^{(jk)}_{pq} = \lambda_p\lambda_q \sum_{a_1,a_2,b_1,b_2=1}^{K} & C^{(pq)}_{a_1a_2b_1b_2}(h - h^*_j)(T_{K,a_1})\,\{\mathbb{N}(T_{K,b_1}) - \mu_{b_1}\} \\
& \{z_k - h^*_k(T_{K,a_2})\}\,\{\mathbb{N}(T_{K,b_2}) - \mu_{b_2}\}\,.
\end{aligned}
$$

Let $\widehat{\zeta}^{(jk)}_{pq}$ denote $\zeta^{(jk)}_{pq}$ evaluated at $\tau = \widehat{\tau}$ and $h(\cdot) = \widehat{h}^*_j(\cdot)$. Observe that $\xi_{j2}\,(\mathbb{N} - \widehat{\mu})^{\otimes 2}\,\xi_{k1}^\intercal = \sum_{p,q=1}^K \widehat{\zeta}^{(jk)}_{pq}$. As discussed above, there exists $\mathsf{m}^{(r)}(t) = \exp\{Z^\intercal\beta^{(r)} + \varphi^{(r)}(t)\}$ such that $\|\mathsf{m}^{(r)} - \mathsf{m}\|_\infty \le C_6\varepsilon$, for $1 \le r \le \exp\{C_5(1/\varepsilon)^{1/m}\}$ and some constants $C_5, C_6 > 0$. Further, in view of Example 19.10 of van der Vaart [33], the bracketing entropy of the class of functions $h - h^*_j$ for $\|h\|_\infty \le M$ and $J(h) \le M$ with $\|\cdot\|_\infty$ is $C_7(1/\varepsilon)^{1/m}$, i.e., there exists $h^{(s)}$ for $1 \le s \le \exp\{C_7(1/\varepsilon)^{1/m}\}$ such that $\|h - h^{(s)}\|_\infty \le C_8\varepsilon$, for some constants $C_7, C_8 > 0$. Notice that both $\mathsf{m}^{(r)}(\cdot)$ and $h^{(s)}(\cdot)$ are uniformly bounded. Construct $\zeta^{(jk,rs)}_{pq}$ as $\zeta^{(jk)}_{pq}$ evaluated at $\beta = \beta^{(r)}$, $\varphi(\cdot) = \varphi^{(r)}(\cdot)$, and $h(\cdot) = h^{(s)}(\cdot)$.

In view of the sub-Gaussian condition of $\mathbb{N}(\cdot)$ and the uniformly bounded conditions of K, Z, and the eigen values of V together with $\|h_j\|_\infty \le M$, we have $E\{\zeta_{pq}^{(jk,rs)} - \zeta_{pq}^{(jk)}\}^2 \le C_9\varepsilon^2$, for some constant $C_9 > 0$, which implies that the class of functions $\sum_{p,q=1}^{K} \zeta_{pq}^{(jk)}$ for $\|\beta-\beta_0\| \le M$, $\|\varphi-\varphi_0\|_{L_2(\mu)} \le M$, $\|\varphi\|_\infty \le M$, $J(\varphi) \le M$, $h \in \mathcal{S}_n$, $\|h\|_\infty \le M$, and $J(h) \le M$ with $\|\cdot\|_\infty$ is a Donsker class, and hence a Glivenko-Cantelli class. It follows from Theorem 2 and Lemma 4 along with Glivenko-Cantelli theorem that $(\mathbb{P}_n - P)\sum_{p,q=1}^{K} \widehat{\zeta}_{pq}^{(jk)} = o_p(1)$. Further, applying the continuous mapping theorem and dominated convergence theorem along with the sub-Gaussian condition of $\mathbb{N}(\cdot)$ and the consistency of $\widehat{\tau}$ and $\widehat{h}_j^*(\cdot)$ yields that $E\left\{\sum_{p,q=1}^{K} \widehat{\zeta}_{pq}^{(jk)}\right\} = o_p(1)$. Therefore, $\mathbb{P}_n\left\{\sum_{p,q=1}^{K} \widehat{\zeta}_{pq}^{(jk)}\right\} = o_p(1)$. Similarly, we can show that for $r \ne 1$ or $s \ne 1$, $\mathbb{P}_n\left\{\xi_{jr}(\mathbb{N} - \widehat{\mu})^{\otimes 2}\xi_{ks}^{\intercal}\right\} = o_p(1)$. Hence, $\mathbb{P}_n\left\{\sum_{r,s=1}^{4} \delta_{rs}^{(jk)}\right\} = \mathscr{I}_2^{(jk)} + o_p(1)$. The consistency of $\widehat{\mathscr{I}_{n2}}$ follows.

Acknowledgments

The author is grateful to the Editor, the Associate Editor, and the two referees for their helpful comments and constructive suggestions.

References

[1] Bailey, H. H., Kim, K., Verma, A. K., Sielaff, K., Larson, P. O., Snow, S. and et al. (2010). A randomized, double-blind, placebo-controlled phase 3 skin cancer prevention study of α-difluoromethylornithine in subjects with previous history of skin cancer. *Cancer Prevention Research.* **3** 35–47.

[2] Cheng, G., Zhang, Y. and Lu, L. (2011). Efficient algorithms for computing the non and semiparametric maximum likelihood estimates with panel count data. *Journal of nonparametric statistics.* **23** 567–579. MR2801310

[3] Cheng, G., Zhou, L. and Huang, J. Z. (2014). Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data. *Bernoulli.* **20** 141–163. MR3160576

[4] Chiou, S. H., Xu, G., Yan, J. and Huang, C. Y. (2018). Semiparametric estimation of the accelerated mean model with panel count data under informative examination times. *Biometrics.* **74** 944–953. MR3860715

[5] Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association.* **87** 942–951.

[6] Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science.* **11** 89–102. MR1435485

[7] Hua, L. and Zhang, Y. (2012). Spline-based semiparametric projected generalized estimating equation method for panel count data. *Biostatistics.* **13** 440–454. MR2941442

[8] Huang, C. Y., Wang, M. C., and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika*, **93** 763–775. MR2285070

[9] Huang, J. Z., Zhang, L. and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scandinavian Journal of Statistics*. **34** 451–477. MR2368793

[10] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Verlag, Dordrecht. MR2724368

[11] Li, N., Park, D. H., Sun, J. and Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *Canadian Journal of Statistics*. **39** 458–474. MR2842424

[12] Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*. **73** 13–22. MR0836430

[13] Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*. **94** 705–718. MR2410018

[14] Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *Journal of American Statistical Association*. **104** 1060–1070. MR2750237

[15] Lu, M. and Li, C. S. (2017). Penalized estimation for proportional hazards models with current status data. *Statistics in Medicine*. **36** 4893–4907. MR3734481

[16] Lu, M., Liu, Y., Li, C. S., and Sun, J. (2022). An efficient penalized estimation approach for semiparametric linear transformation models with interval-censored data. *Statistics in Medicine*, **41** 1829–1845. MR4408614

[17] Ma, S. and Kosorok, M. R. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *The Annals of Statistics*. **33** 2256–2290. MR2211086

[18] Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *The Annals of Statistics*. **25** 1014–1035. MR1447739

[19] Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics*. **57** 126–134. MR1833298

[20] Morel, J. G., Bokossa, M. C. and Neerchal, N. K. (2003). Small sample correction for the variance of GEE estimators. *Biometrical Journal*. **45** 395–409. MR1984617

[21] Murphy, S. A. and van der Vaart, A. W. (1999). Observed information in semiparametric models. *Bernoulli*. **5** 381–412. MR1693616

[22] Pan, W. (2001). On the robust variance estimator in generalised estimating equations. *Biometrika*. **88** 901–906. MR1859421

[23] Robertson, T., Write, F. T. and Dykstra, R. L. (1986). *Order Restricted Statistical Inference*, 1st ed. Wiley, New York.

[24] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge. MR1998720

[25] Schumaker, L. (1981). *Spline Functions: Basic Theory*, 1st ed. Wiley,

New York MR0606200

[26] Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics.* **22** 580–615. MR1292531

[27] Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics.* **22** 118–171. MR1272079

[28] Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica.* **5** 279–289. MR1329298

[29] Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* **62** 293–302. MR1749540

[30] Sun, J., Tong, X., and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics*, **63** 1053–1059. MR2414582

[31] Sun, J. and Zhao, X. (2013). *Statistical analysis of panel count data.* Springer, New York. MR3136574

[32] van de Geer, S. (2000). *Empirical Processes in M-estimation.* Cambridge University Press, Cambridge.

[33] van der Vaart, A. W. (2000). *Asymptotic Statistics*, 1st paperback ed. Cambridge University Press, Cambridge. MR1652247

[34] Wood, S. N. (2017). *Generalized additive models: an introduction with R.* chapman and hall/CRC. MR2206355

[35] Wood, S. N. and Fasiolo, M. (2017). A generalized Fellner-Schall method for smoothing parameter optimization with application to Tweedie location, scale and shape models. *Biometrics.* **73** 1071–1081. MR3744521

[36] Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of statistics.* **28** 779–814. MR1792787

[37] Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics.* **35** 2106–2142. MR2363965

[38] Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika.* **89** 39–48. MR1888344

[39] Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika.* **93** 777–790. MR2285071

[40] Zhang, Y., Hua, L. and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics.* **37** 338–354. MR2682304